

Machine-Learning-Based-Approaches for Sleep Stage Classification Utilising a Combination of Physiological Signals: A Systematic Review

Haifa Almutairi * , Ghulam Mubashar Hassan *  and Amitava Datta * 

Department of Computer Science & Software Engineering, The University of Western Australia, Crawley, WA 6009, Australia

* Correspondence: haifa.almutairi@research.uwa.edu.au (H.A.); ghulam.hassan@uwa.edu.au (G.M.H.); amitava.datta@uwa.edu.au (A.D.)

Abstract: Increasingly prevalent sleep disorders worldwide significantly affect the well-being of individuals. Sleep disorder can be detected by dividing sleep into different stages. Hence, the accurate classification of sleep stages is crucial for detecting sleep disorders. The use of machine learning techniques on physiological signals has shown promising results in the automatic classification of sleep stages. The integration of information from multichannel physiological signals has shown to further enhance the accuracy of such classification. Existing literature reviews focus on studies utilising a single channel of EEG signals for sleep stage classification. However, other review studies focus on models developed for sleep stage classification, utilising either a single channel of physiological signals or a combination of various physiological signals. This review focuses on the classification of sleep stages through the integration of combined multichannel physiological signals and machine learning methods. We conducted a comprehensive review spanning from the year 2000 to 2023, aiming to provide a thorough and up-to-date resource for researchers in the field. We analysed approximately 38 papers investigating sleep stage classification employing various machine learning techniques integrated with combined signals. In this study, we describe the models proposed in the existing literature for sleep stage classification, discuss their limitations, and identify potential areas for future research.

Keywords: classification; EEG; EMG; EOG; deep learning; machine learning; sleep stage



Citation: Almutairi, H.; Hassan, G.M.; Datta, A. Machine-Learning-Based Approaches for Sleep Stage Classification Utilising a Combination of Physiological Signals: A Systematic Review. *Appl. Sci.* **2022**, *13*, 13280. <https://doi.org/10.3390/app132413280>

Academic Editors: Asiya Khan and Gloria Iyawa

Received: 3 August 2023

Revised: 12 December 2023

Accepted: 13 December 2023

Published: 15 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sleep is a fundamental human function that involves a series of changes in the heart, brain, muscles, eyes, and respiratory activities. Besides aiding in mental and physical health recovery, sleep contributes to healthy brain functionality during the day [1,2]. However, drowsiness and sleep disorders, such as sleep apnoea [3] and periodic leg movement [4], can adversely affect daily activities [5]. Worldwide, more than 400 million adults have sleep apnoea [6]. A study conducted in the United States found that up to 24% of adults are affected by sleep issues [7]. In Australia, 33% of the population is affected by insomnia [8]. In addition, the Sleep Heart Health Study found that people who have trouble falling asleep may be affected by health issues that include neurocognitive deficits, cardiovascular problems, diabetes, recurrent heart attacks, and stroke [9,10]. Therefore, to protect human health, it is essential to monitor sleep.

Sleep specialists, who are experts trained in sleep medicine, follow the guidelines of the American Academy of Sleep Medicine (AASM) [11] to classify sleep into three primary stages: wake (W), non-rapid eye movement (NREM) sleep encompassing three substages (N1, N2, and N3), and rapid eye movement (REM) sleep. During the NREM stage, parasympathetic activity rises, while heart rate (HR), sympathetic activity, blood pressure, and metabolic rate fall. Neuronal activity is higher during the REM stage than

during the NREM stage [12,13]. In a typical sleep cycle, 50–60% of sleep is spent in N1 and N2, which are considered light sleep stages; 15–20% in N3, which is considered a deep sleep stage; 20–25% in the REM sleep stage; and 5% or less in the W stage [14]. The absence of certain sleep stages in a typical sleep cycle may suggest the presence of sleep disorders.

The activities mentioned above for sleep stages are recorded by a traditional method called polysomnography (PSG) [15], which is considered as the gold standard for classifying sleep stages following the guidelines established by AASM. The physiological signals recorded during PSG include respiratory effort [16], electroencephalogram (EEG) [17], electrocardiogram (ECG) [18], electrooculogram (EOG) [19], and electromyogram (EMG) [20]. Each physiological signal recording is divided by sleep specialists into 30-second segments. These segments are then classified into W, NREM (including N1, N2, and N3), or REM sleep stages [21]. This classification is based on visual analysis, which is not only time-consuming but also error prone. Furthermore, patients need to be connected to sensors and equipment for extended periods during sleep studies, which can have an impact on the quality of recorded data [22].

Artificial intelligence (AI) has recently been employed in a wide range of clinical medical applications, including surgeries [23], classifications of types of seizure [24], and stage classifications [25]. AI research aims to build intelligent tools that can help medical specialists make clinical decisions in the medical field [26]. Machine learning (ML) is a subfield of AI that employs algorithms and approaches to create models that learn from data and make predictions or decisions based on that learning. The key drawback of these conventional ML techniques is the need for feature engineering methods to extract features from input data. The model's performance may be restricted by the time-consuming process of designing and selecting relevant features from the input data. However, recently introduced deep learning (DL) has overcome the limitations of conventional machine learning algorithms by employing multilayered neural networks to extract and learn key features from raw data [27].

The utilisation of machine learning (ML) techniques for sleep stage classification has been extensively examined in literature reviews [25,28–33]. The systematic reviews in [25,29,31] focus on models of sleep stage classification based on single-channel EEG signals. The advantages of using single-channel EEG signals include convenience and ease of use, and they can be adapted for use in the patient's home using wearable sensors. Other reviews [28,30,32,33] have focused on models developed using single-channel EEG signals and a combination of physiological signals to classify sleep stages. The benefits of using a combination of physiological signals with ML models can increase accuracy because the model has more information and can extract more discriminated features.

This systematic review focuses on machine learning investigations involving multiple channels of physiological signals, including EEG, ECG, EMG, EOG, and respiratory data for sleep stage classifications. The physiological signals were employed, as either a multichannel or a combination of multiple signals, to develop a model for the classification of sleep stages. We selected research studies that encompassed multiple physiological signals from reviews [30,32,33]. Furthermore, we conducted a thorough literature search to identify additional publications aligned with our criteria. Our emphasis on signal combinations sets this review apart and offers a valuable resource for researchers exploring this specialised domain.

The remainder of the article is organised as follows: Section 2 provides a detailed description of the research article selection methodology. Section 3 describes a conceptual framework for the classification of sleep stages. Section 4 reviews different existing models of sleep stage classification. Section 5 highlights the limitations of existing approaches. Section 6 discusses and analyses the existing research used in the classification of sleep stages. Finally, Section 7 summarises our conclusions and identifies directions for future research.

2. Methodology of Selection Papers

In this review article, we followed a systematic review methodology proposed by Dixon et al. [34].

2.1. Data Sources

We systematically searched literature databases, including Scopus, Google Scholar, and PubMed, from the last few decades (approximately from 2000) up to present, with English language restriction. In this systematic review, we focused on studies that employed multiple channels of EEG, ECG, EMG, EOG, and respiratory signals or a combination of these signals to propose models for the classification of sleep stages. We conducted a comprehensive search of the literature to identify additional studies that met our inclusion criteria. We screened the titles and abstracts of the retrieved articles and included studies that employed the specified physiological signals for sleep stage classification.

2.2. Data Extraction

A data extraction protocol was defined and evaluated by all authors. The inclusion criteria for this study encompassed studies with keywords related to (“Classification of sleep stage”) AND (“Combined physiological signals” OR “Combined EEG ECG EMG EOG respiratory”) AND (“Machine learning” OR “Deep learning” OR “Artificial intelligence” OR “Big data”). The included document types were indexed journal papers, conference papers, book chapters, and books. Exclusion criteria were applied to filter out studies that did not fall under the subareas of interest, and those that were not in English or did not meet the predefined criteria.

2.3. Data Analyses

This article primarily concentrates on conducting a systematic review, rather than a meta-analysis, to explore the classification of sleep stages using intelligent data analysis techniques in the medical field. However, it does not extensively delve into specific details and results obtained from individual case studies. Therefore, the utilization of data analysis techniques within this specific context is not the main focus.

2.4. Results

In our analysis, we incorporated a total of 38 papers that fulfilled the predefined inclusion criteria. Figure 1 represents the comprehensive search and selection process, outlining the reasons for excluding certain studies.

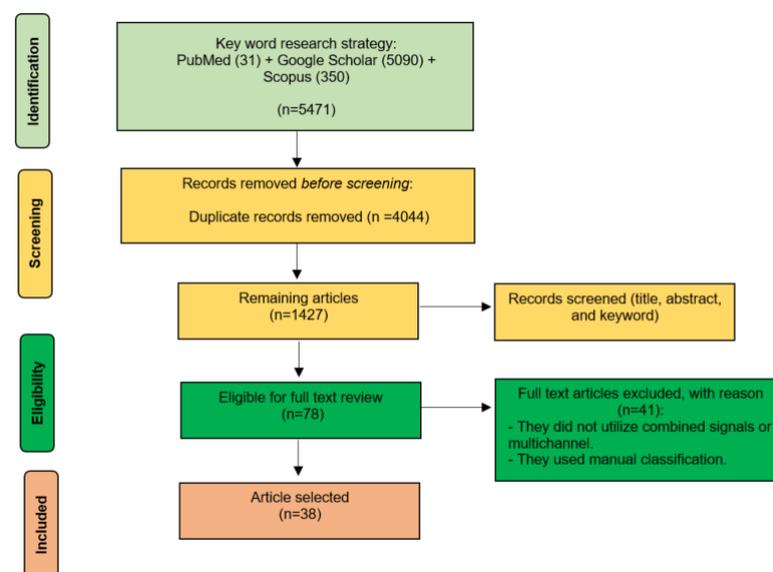


Figure 1. Comprehensive search and selection process for systematic review.

3. Conceptual Framework for the Classification of Sleep Stages

Figure 2 shows the conceptual framework for sleep stage classification. Researchers in this area use various datasets, such as Sleep-edf [35], Sleep-edfx [36], MASS [37], MIT-BIH [38], ISRUC-Sleep [39], SHHS [40], UCD [41], and PhysioNet Challenge [42]. In the preprocessing step, datasets are cleaned to exclude missing values and eliminate noise, artefacts, and other distortions. The most common preprocessing methods are filtering [43], normalization [44], and signal conditioning [45]. After that, the feature selection step is used to select the most important features that have a significant impact on the model's performance. The two common methods for feature selection are feature engineering and statistical methods. Most of the existing work on sleep stage classification employs feature engineering methods, such as fast Fourier transform (FFT) [46], wavelet transform (WT) [47], short-time Fourier transform (STFT) [48], Pan–Tompkins algorithm [49], and temporal decomposition (TD) [50]. Common statistical methods include dynamic wrapping, dispersion entropy, max, mean, skew, and variance features [51]. However, recently, some studies proposed to use raw data as input without feature engineering or statistical methods to reduce the complexity of the proposed model. This approach allows deep learning algorithms to learn directly from raw data [52].

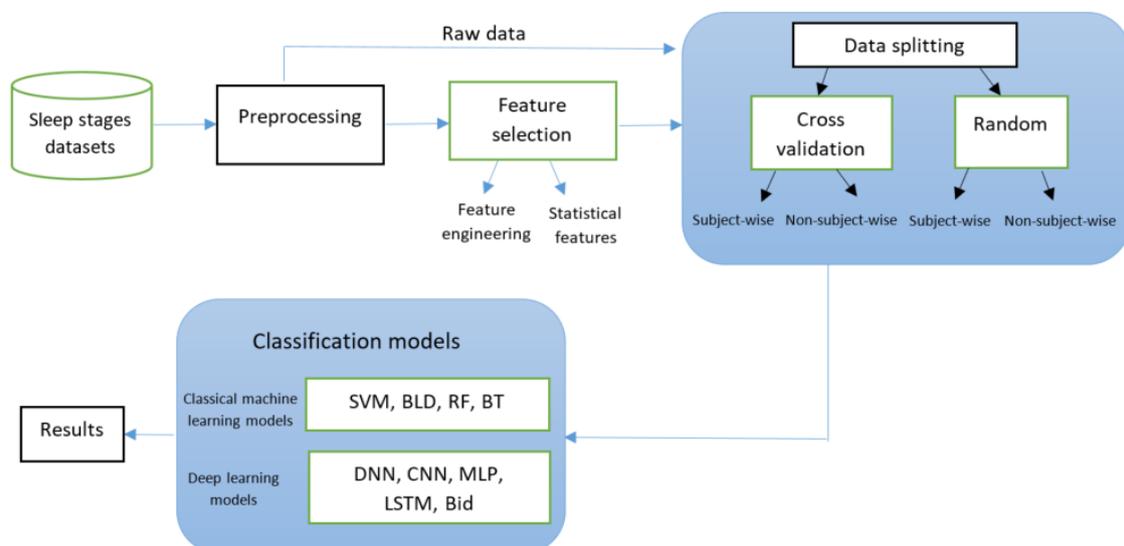


Figure 2. Conceptual framework for the classification of sleep stages.

The next step is the data splitting strategy, which includes cross-validation and random splitting approaches used to split datasets into training and testing sets. The cross-validation approach divides the dataset into 3, 5, or 10 parts to evaluate the classification models [53]. The random splitting approach employs a small fraction of the testing set to evaluate the model (e.g., 70% training set, 15% validation set, and 15% testing set). In addition, some studies use the above splitting strategies with either a subject-wise or non-subject-wise approach. In a subject-wise approach, the training and testing sets do not share any patient's recording samples. Therefore, the patients whose recordings are used in the training set are excluded from the testing set. In contrast, a non-subject-wise approach uses the same patients' recording samples in the training and testing sets.

The final step involves classification using machine learning or deep learning models. In the sleep stage classification models, three types of categorizations for the sleep stages have been used. In the first type, a binary classification has been implied, distinguishing between wake stage (W) and sleep stage (combining REM and NREM). In the second type, it categorises sleep into three stages: wake (W), non-rapid eye movement (NREM), and rapid eye movement (REM), while the third type involves categorization into five stages: W, N1, N2, N3, and REM. However, it is important to note that binary and three-sleep stage

classifications do not completely align with traditional AASM guidelines, while a five-sleep stage classification is consistent with the AASM guidelines.

4. Literature Review

For this literature review, the selected papers were categorised into five subsections based on the utilised input signals. According to a preliminary survey of the literature, the signals most frequently used in the classification of sleep stages were found to be EEG, EMG, EOG, ECG, and respiratory effort. Each subsection includes a definition of the signals and describes all associated models and their performances.

4.1. Electroencephalogram (EEG)

EEG techniques capture the brain's electrical activity. Electrical signals in the brain can be measured by observing changes in the electrical activity between two electrodes over time. The standard method for measuring EEG signals is commonly known as the 10–20 system, which employs a minimum of 21 electrodes [54]. The EEG signal displays the diverse properties of brain activity. These activities help to classify sleep stages. Stage W is characterised by alpha activity in the occipital area [55]. The N1 stage is the transition between the W and N2 stages, and theta activity is one of the characteristics of EEG activity of the N1 stage [55]. Stage N2 is characterised by distinctive features known as spindles and K-complexes. Spindles are brief bursts of high-frequency brain activities, while K-complexes are characterised by sharp, high-amplitude brain activity with a unique appearance in EEG signals [56]. Delta activity is indicative of N3, which is a deep sleep stage [57]. The REM stage is characterised by rapid, low-voltage theta waves [58]. The distinct frequency ranges of EEG signals corresponding to each sleep stage are shown in Table 1. Figure 3 shows a sample of time series EEG data for the five sleep stages.

Table 1. Distinct frequency ranges of EEG signals corresponding to each sleep stage.

Sleep Stage	Characteristic Frequency
W	Alpha (8–12 Hz)
N1	Theta (4–8 Hz)
N2	Spindle and K-complexes (12–15 Hz)
N3	Delta (0.5–4 Hz)
REM	Alpha (8–12 Hz) Theta (4–8 Hz)

Table 2 lists studies that used two or more physiological signals for sleep stage classification. Few studies of sleep stage classification models have been developed utilising multiple channels of EEG signals as inputs. Blanco et al. [59] proposed a deep learning model to select important features from EEG signals for sleep stage classification, aiming to reduce reliance on sleep experts. They used two channels of EEG signals (Fpz-Cz and Pz-Oz) as input to the 1D-CNN model. The architecture of 1D-CNN included seven layers of 1D-CNN, a max-pooling layer followed by a fully connected layer to classify five sleep stages. Their model achieved an accuracy of 92.60%. Similarly, Satapathy et al. [60] proposed a deep learning model, which used two channels of EEG signals (C3-A2 and C4-A1) as input to the 1D-CNN model. The architecture of 1D-CNN contained seven blocks that included 1D-CNN, batch normalization, and ReLU layers. The last layer was fully connected with softmax to classify the segments into five sleep stages. Their model was tested on subgroup 1 and subgroup 2 of the ISRUC-Sleep dataset, and the accuracies achieved for the classification of five sleep stages were 97.22% and 95.06%, respectively.

Another study by Delimayanti et al. [61] extracted features from two channels of EEG signals (Pz-Oz, Fpz-Cz) by using an FFT method to improve the accuracy of the classification. The features extracted passed to SVM to classify three sleep stages and

five sleep stages. Their model achieved an accuracy of 94.14% and 91.73% for three- and five-sleep-stage classification, respectively. Dequidt et al. [62] conducted a study to explore the utilization of time–frequency representations, such as spectrograms, as input for a fine-tuned VGG-16 network. Their research focused on comparing various spectrograms encoding multiple EEG channels to facilitate the recognition of visual patterns in images. The study reported an achieved accuracy of 82.96% with five-sleep-stage classification.

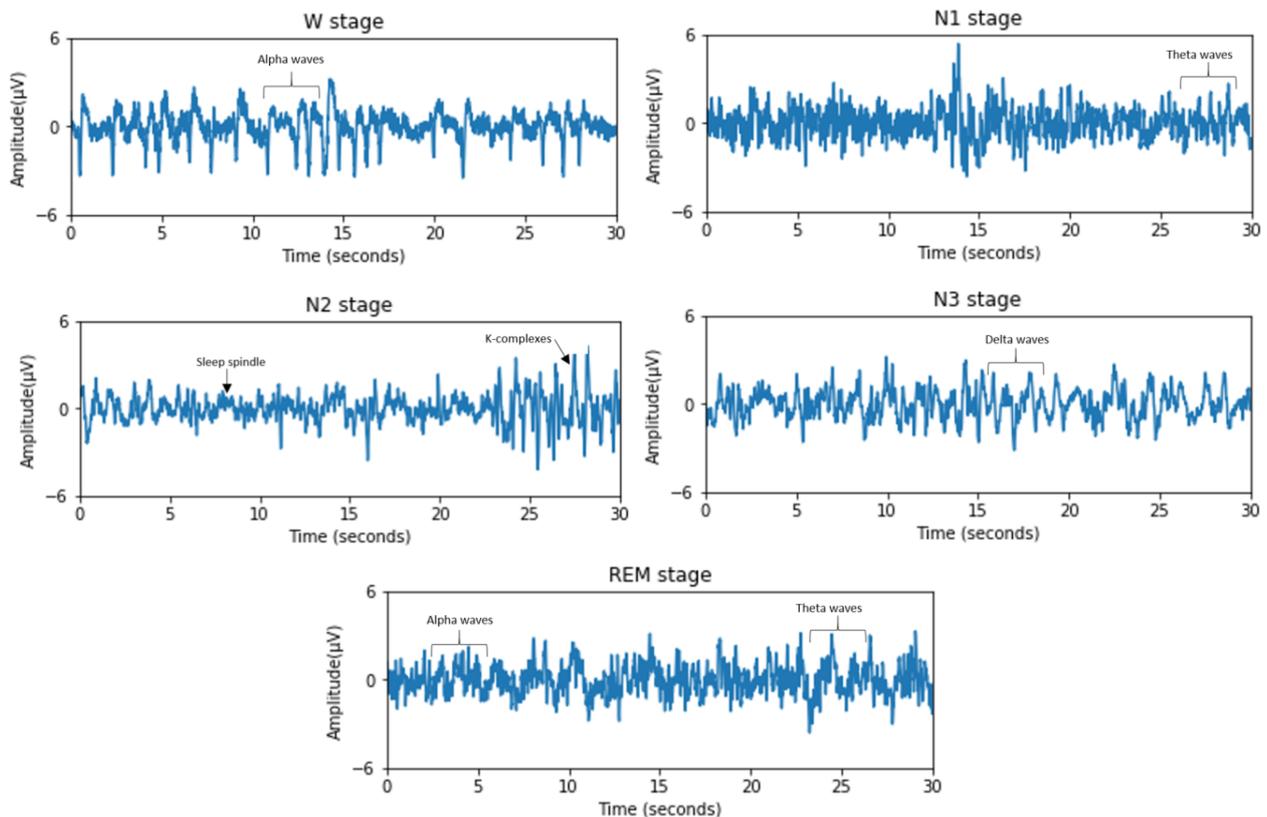


Figure 3. Samples of EEG patterns in five sleep stages from the Sleep-edfx dataset [36].

4.2. Electromyogram (EMG)

The EMG technique records muscles' electrical activity during contraction and relaxation during sleep, which makes EMG a significant signal for classifying sleep stages [63]. Figure 4 presents a sample of time series EMG data for the five sleep stages, illustrating that EMG activity reaches its peak during the W stage. As we progress from stage N1 to stage N3, EMG activity gradually decreases as the muscles begin to relax. In the REM stage, the EMG activity is at its lowest point as the muscles are inactive and relaxed.

Few studies have explored the use of combining EEG and EMG signals to enhance the accuracy of sleep stage classification. Tautan et al. [64] used a combination of one channel of EEG (F3-M2) and one channel of EMG signals as input. They extracted statistical and FFT features from the raw data to pass to RF and MLP classifiers to classify five sleep stages, achieving accuracies of 88.65% and 66.70%, respectively. Akin et al. [65] proposed a machine learning model that used one channel of EMG and one channel of EEG signals (C3-A2) as input, and applied a wavelet transform (WT) to extract features from the signals. The deep neural network (DNN) model they developed achieved a 98% accuracy in classifying five sleep stages.

Kim et al. [66] used a temporal decomposition method to extract features from one channel of EEG signal (Fpz-Cz) and one channel of EMG signal, and used SVM as a classifier to classify five sleep stages, achieving an accuracy of 93.8%. Almutairi et al. [67] selected multichannel EEG signals (Fpz-Cz and Pz-Cz) and one channel of EMG signal as input, and passed them through a deep learning model named as SSNet model containing

two deep learning architectures. The first architecture contained five 1D-CNN layers, and the second architecture contained two LSTM layers. The features extracted from the two architectures were combined and passed to a fully connected layer to classify three sleep stages, achieving an accuracy of 95.46%.

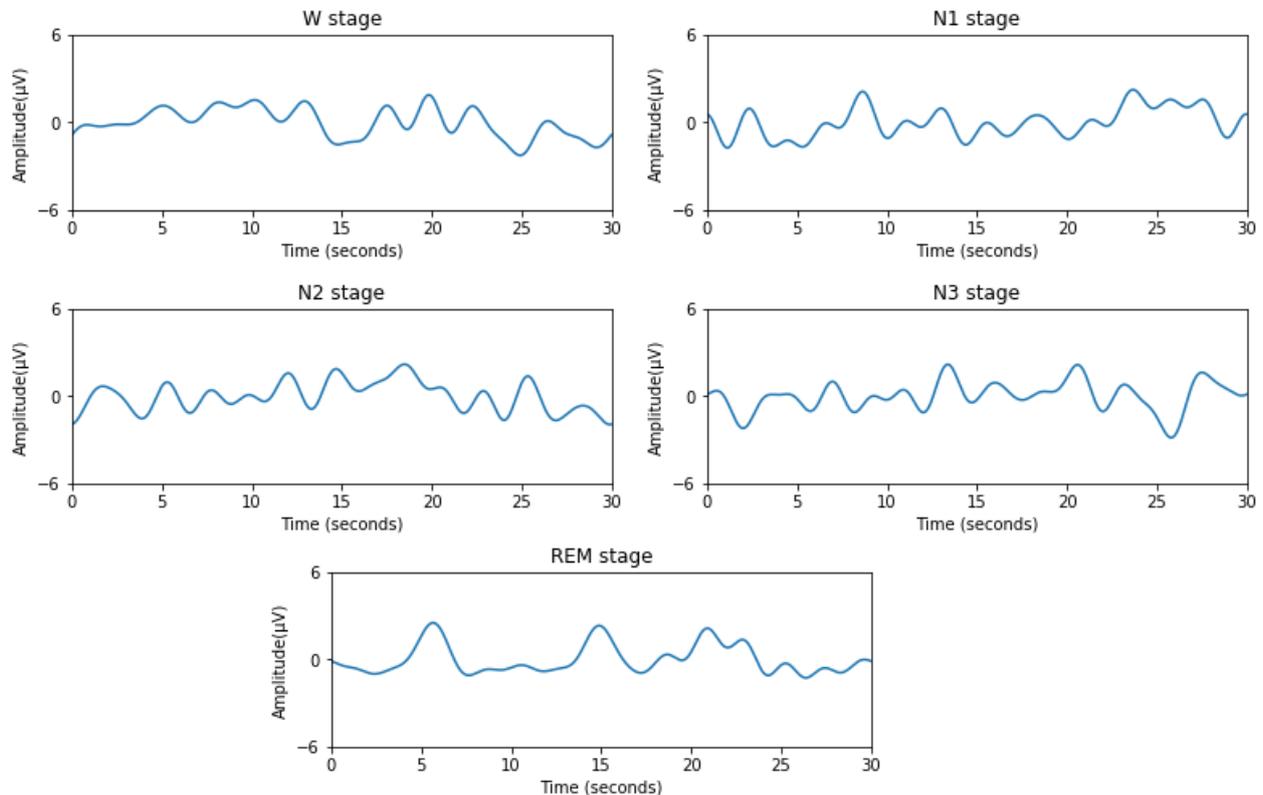


Figure 4. Samples of EMG patterns in five sleep stages from the Sleep-edfx dataset [36]. Muscle activity exhibits a gradual reduction from the wake (W) stage to the REM (rapid eye movement) stage.

4.3. Electrooculogram (EOG)

In sleep research, the measurement of eye movements is crucial for evaluating sleep quality and identifying sleep stages [68]. EOG recording is used to capture eye movements: Surface electrodes are positioned around the eye to measure a potential gap between its anterior and posterior poles [69]. During the W stage, EOG signals are used to detect rapid eye movements, providing an indication of this stage's characteristics. In the NREM stage, EOG signals record slow eye movements. In contrast, during the REM stage, EOG signals can capture bursts of rapid eye movements, which are a distinctive feature of this stage [70]. Figure 5 presents a sample of time series EOG data for the five sleep stages.

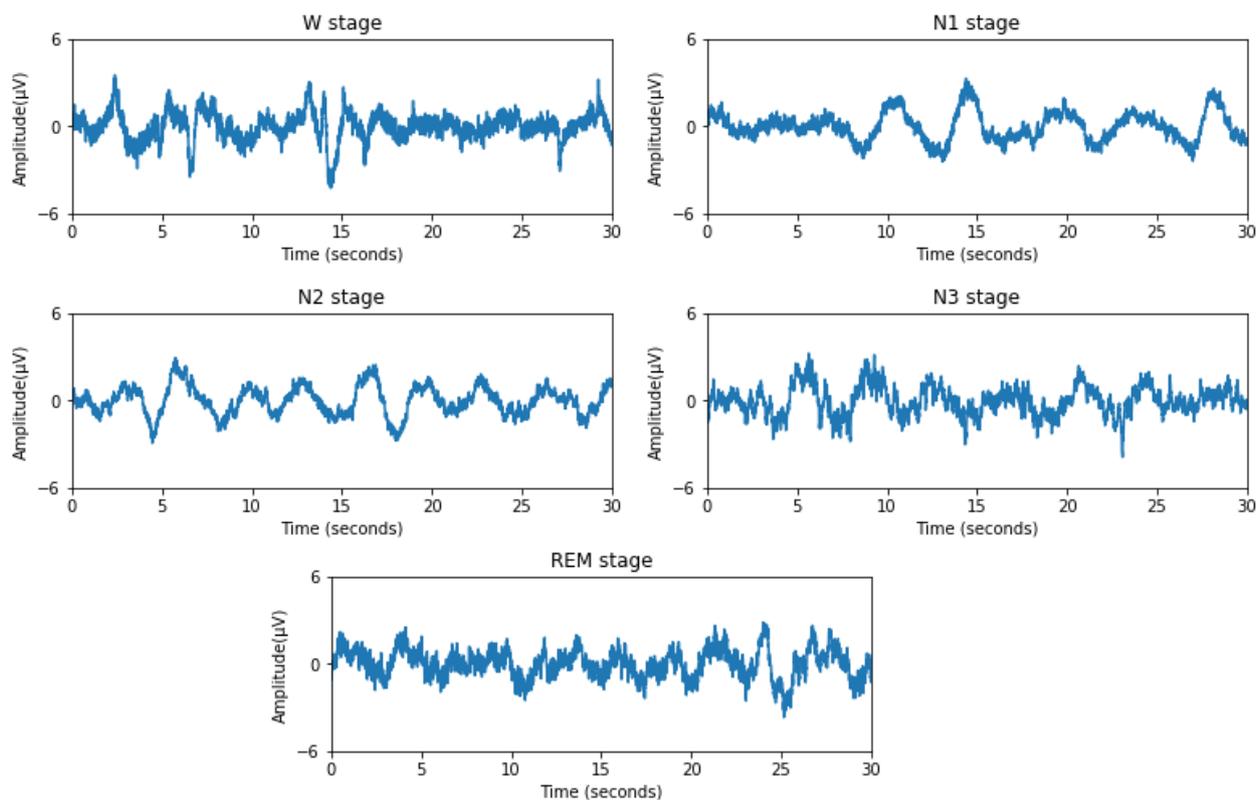


Figure 5. Samples of EOG patterns in five sleep stages from the Sleep-edfx dataset [36]. Wake (W) shows frequent eye movements, the NREM stages display sporadic eye movements and unique patterns, and the REM stage exhibits rapid distinct eye movements.

The characteristics of EOG signals have been utilised in proposing models for classifying sleep stages. Estrada et al. [71] used a feature engineering method with fuzzy rules for the classification. Two channels of EOG and EMG signals were combined and passed to an FFT method to extract features. Fuzzy rules were used to predict the final results for the classification of five sleep stages. The study did not report the performance of their proposed model. Yildirim et al. [72] proposed a deep learning model to extract features from a combination of EEG and EOG signals. They selected one channel of EEG signal (Fpz-Cz) and one channel of EOG signal (horizontal). These features passed to a 1D-CNN model. Their architecture consisted of two layers of 1D-CNN and max-pooling layers, and the order of these layers was repeated five times. The final layers were two fully connected layers to classify the segments into three and five sleep stages. They tested their model on two datasets (Sleep-edf and Sleep-edfx). The model achieved an accuracy of 94.64% for three sleep stages and 91.22% for five sleep stages on the Sleep-edf dataset. Similarly, the model achieved an accuracy of 94.34% for three sleep stages and 90.98% for five stages on the Sleep-edfx dataset.

A study by Sokolovsky et al. [73] proposed a deep learning model that contained a deep network to improve the classification accuracy. Their model inputs two channels of EEG signals (Fpz-Cz and Pz-Cz) and one channel of EOG signal. Their architecture consisted of six layers of 1D-CNN, followed by batch-normalization and max-pooling layers. After that, they added three layers of 1D-CNN, followed by a batch normalization and max-pooling layer. This structure was repeated three times. In the end, they added two max-pooling layers, followed by a 1D-CNN layer, a max-pooling layer, a 1D-CNN layer, and two fully connected layers. Their model achieved an accuracy of 81% for the classification of five sleep stages. Phan et al. [74] used two different datasets (Sleep-edf and MASS) for evaluating their proposed model. They selected a combination of one channel of

EEG signal (Fpz-Cz) and one channel of EOG (horizontal) signal from the Sleep-edf dataset. Similarly, they used a combination of one channel of EEG (C4-A1) and one channel of EOG signal (ROC-LOC) from the MASS dataset. They used a short-time Fourier transform method and a 2D-CNN model to classify five sleep stages. The architecture of their model consisted of one layer of 2D-CNN, a max-polling layer, and a multitask softmax layer. Their model achieved an accuracy of 82.30% on the Sleep-edf dataset and 82.50% on the MASS dataset.

Almutairi et al. [67] proposed an SSNet model using a combination of signals of two channels of EEG (Fpz-Cz and Pz-Cz) and one channel of EOG as input. Their model classified the segments into the three sleep stages with an accuracy of 95.65%. Sekkal et al. [75] used two channels of EEG signals (Fpz-Cz and Pz-Cz) and one channel of EOG signal as input. They extracted statistical features from raw data to pass to different machine learning classifiers, such as SVM, RF, and KNN. Their model with an SVM classifier achieved the highest accuracy of 89.1%. Toma et al. [76] proposed an end-to-end CRNN (convolutional recurrent neural network) model for five-sleep-stage classification. The model takes three channels of EEG signals (Pz-Oz and Fpz-Cz) and EOG signal as input. It consists of two branches of 1D-CNN to extract spatial features, followed by several RNN layers. Dropout layers are inserted between the RNN layers to prevent overfitting. Two types of dropout layers, regular dropout and spatial dropout, are used in the model. The study reported an accuracy of 90.30% for five-sleep-stage classification.

4.4. *Electrocardiogram (ECG) and Respiratory*

ECG and respiratory signals can assist with sleep stage classification because the heart rate and respiratory effort change throughout the sleep stages [77]. An ECG is a recording of the heart's electrical activity over a timespan. The ECG signal consists of several beats comprising the P wave, QRS complex, and T wave, depending on the individual's heart condition [78,79]. The properties of the ECG signal can change during both NREM and REM sleep stages. The heart rate decreases during the NREM sleep stage, while heart rate variability can either increase or decrease depending on the NREM sleep stage. Conversely, both the heart rate and heart rate variability increase during the REM sleep stage, as reported in [80].

Respiratory inductance plethysmography (RIP) is a noninvasive technique for measuring airflow and respiratory effort. Changes in respiratory patterns can help classify sleep stages. For example, during the REM sleep stage, the respiratory system can become more irregular, and the upper airway muscles may become more relaxed, leading to more frequent disruptions in breathing and potential sleep apnoea [81].

We have categorised studies into two subgroups below based on the type of signal combination used as input. The first group includes studies utilising a combination of ECG and respiratory as input. The second group encompasses studies employing a combination of EEG and ECG or a combination of EEG and respiratory as input.

As mentioned earlier, the first group includes studies that utilised a combination of ECG and respiratory as input. Long et al. [82] used statistical features of dynamic wrapping to extract features from ECG and respiratory signals and achieved a 95% accuracy in the binary classification of sleep stages using the LDA classifier. Fonseca et al. [83] applied the Pan–Tompkins algorithm to extract an R-R interval from ECG signals and the mean and variance of respiratory signals to classify three sleep stages with an accuracy of 80% by using a BLD classifier. Casal et al. [84] utilised a combination of signals from ECG and respiratory effort to classify the segments into the binary classification of sleep stages using a two-layered gated recurrent unit (GRU) neural network. They reported achieving an accuracy of 90.13%.

The second group includes studies that utilised a combination of EEG and ECG or a combination of EEG and respiratory as input. For example, Tripathy et al. [85] used R-R intervals from the ECG signal and the dispersion entropy method for extracting statistical features from the EEG signal, achieving a 73.70% accuracy in classifying five sleep stages using multi-fully-connected layers. Yu et al. [86] used a fast Fourier transform method

to extract features from one channel of EEG and ECG signals, achieving an accuracy of 99% in classifying five sleep stages using SVM. Tautan et al. [64] proposed a model that takes one EEG (F3-M2) and one respiratory signal channel as input. The model uses an FFT method to extract features from the EEG signal and extract statistical features such as mean, skew, and variance from the respiratory signal. The model achieved accuracies of 93.72% and 52.27% using RF and MLP classifiers, respectively. Moreover, they tested their model by combining one channel of EEG and ECG signals, using an FFT method to extract features from the EEG signal and R-R intervals from the ECG signal. Their proposed model achieved accuracies of 72.52% and 60.28% using RF and MLP classifiers, respectively. Zhao et al. [87] used a combination of two channels of EEG and ECG as input. They passed these signals separately to a 1D-CNN model, which contains five layers of 1D-CNN. The model they developed achieved an accuracy of 98.84% in a binary sleep stage classification.

4.5. Combination of Signals

The combination of more than two types of physiological signals provides complementary information that can improve sleep stage classification. This approach can be beneficial because certain features of a sleep stage might be missed by one signal but detected by another [30].

We have categorised studies into two further subcategories below based on the type of signal combination used as input. The first category includes studies that utilise a combination of four types of signals as input: EEG, EMG, ECG, and respiratory. The second category encompasses studies that employ a combination of three types of signals as input: EEG, EMG, and ECG, or a combination of EEG, EMG, and EOG.

As mentioned earlier, the first category includes studies that utilise a combination of four types of signals: EEG, EMG, ECG, and respiratory as input to classify sleep stages. Only two studies were found in this category. Willemen et al. [88] proposed a model that extracted statistical features and utilised an SVM classifier, achieving an accuracy of 69% in classifying five sleep stages. Furthermore, Helland et al. [89] extracted mean and variance features from raw data and employed a BLD classifier, resulting in an 80% accuracy for the classification of five sleep stages.

The second category encompasses studies that employ a combination of three types of signals: EEG, EMG, and ECG/EOG as inputs to classify five sleep stages. Takatani et al. [90] extracted R-R features from ECG signals and applied fast Fourier transform (FFT) to extract frequency domain features from EEG and EMG signals. The selected features were then evaluated using a linear discriminant analysis (LDA) classifier, resulting in an accuracy of 80%. Biswal et al. [91], on the other hand, used a short-time Fourier transform method to extract frequency domain features. These features were evaluated by passing them through a model that included a combination of 1D-CNN layers and a bidirectional LSTM (Bi-LSTM) layer. Their model achieved a classification accuracy of 87.5%.

In a study by Choi et al. [92], the researchers investigated the utilization of five signal combinations, namely, ECG, EEG, EMG, left-eye EOG, and right-eye EOG. They explored all possible combinations of these signals and determined that the combination of EEG, EMG, and ECG exhibited the most promising outcomes. Statistical features were extracted from these signals, taking into account different window sizes and signal lengths, and an XGBoost classifier was employed to evaluate the performance. The proposed model achieved an accuracy of 85%.

Cui et al. [93] used two channels of EEG signals (C3-A2 and C4-A1), two channels of EOG signals (O1-A2 and LOC-A2), and one channel of EMG signals (X1). They applied fine-grained segmentation and a 2D-CNN model to classify five sleep stages. The architecture of the 2D-CNN model included two 2D-CNN layers, max-pooling layers, and a fully connected layer. The classification of five sleep stages by their model resulted in an accuracy of 90.12%. Zhang et al. [94] proposed a method that combined short-time Fourier transform features with raw data to classify five sleep stages using a 2D-CNN model. The architecture of their model consisted of two 2D-CNN layers, followed by a max-pooling

layer, an LSTM layer, and a fully connected layer. Their approach achieved an accuracy of 86%.

Chambon et al. [95] proposed a 2D-CNN model to extract features from a combination of six channels of EEG signals and two channels of EOG signals. The 2D-CNN architecture consisted of three layers of 2D-CNN and max-pooling layers. Additionally, they utilised a separate 2D-CNN model to extract features from three channels of EMG signals. All the extracted features were then combined and passed to a fully connected layer for the classification of five sleep stages. The proposed model achieved an accuracy of 79%. Phan et al. [74] proposed a model that combined a short-time Fourier transform method with multitask CNN layers for the classification of five sleep stages. The multitask CNN architecture comprised one layer of 2D-CNN, a max-pooling layer, and a multitask softmax layer. The model used a combination of one EEG signal channel (C4-A1), one EOG signal channel (ROC-LOC), and two EMG signal channels (CHIN1-CHIN2) as input, and the model achieved an accuracy of 81.2%.

Xu et al. [96] utilised a combination of EEG signals (Fpz-Cz and Pz-Oz), an EOG signal (Horizontal), and an EMG signal as input. To extract features from raw data, they employed a 1D-CNN model consisting of a convolution block (four 1D-CNN layers and max-pooling layers) and a reduction block (input passed to two max-pooling layers and three layers of 1D-CNN). These blocks were repeated four times, followed by two 1D-CNN layers and a fully connected layer. The model's performance was evaluated on two datasets, achieving an accuracy of 85.40% on the Sleep-edf dataset and 81.60% on the Sleep-edfx dataset for classifying five sleep stages. Similarly, Sharma et al. [97] utilised two channels of EEG signals (C3-A2 and C4-A1), one channel of an EMG signal, and two channels of EOG signals (EOG-L and EOG-R) as input. They applied a wavelet decomposition method to extract frequency domain features and evaluated the performance of these features by using a bagging tree classifier for the classification of three and five sleep stages. The model achieved an accuracy of 95.44% and 95.20% for the classification of three and five sleep stages, respectively.

Yan et al. [98] utilised a combination of EEG, EMG, and EOG signals to feed into a 1D-CNN model with four layers of 1D-CNN, followed by max-pooling layers, achieving an accuracy of 73% for the classification of five sleep stages. Then, they applied an STFT method to extract frequency domain features from the raw data. These features were passed to the 1D-CNN model, which resulted in an improved accuracy of 74.24%. Almutairi et al. [67] utilised two datasets, Sleep-edfx and ISRUC-Sleep, to evaluate their SSNet model's performance for the classification of three and five sleep stages. From the Sleep-edfx dataset, they chose two EEG channels (Fpz-Cz and Pz-Cz), one EMG channel, and one EOG channel as input. Meanwhile, from the ISRUC-Sleep dataset, they selected two EEG channels (C3-A2 and C4-A1), one EMG channel (X1), and two EOG channels (O1-A2 and LOC-A2) as input. Their SSNet model achieved accuracies of 94.64% and 91.22% for three- and five-sleep-stage classification on the Sleep-edfx dataset, respectively. On the ISRUC-Sleep dataset, the SSNet model reported accuracies of 94.34% and 90.98% for three and five sleep stages, respectively.

Satapathy et al. [99] utilised EEG (C3-A2), EMG (X1), and EOG (ROC-A2) signals as input, and passed them to a 1D-CNN model consisting of nine layers. Their model demonstrated high accuracy for the classification of three and five sleep stages of subgroup 1 of the ISRUC-Sleep dataset, with reported accuracies of 98.61% and 98.46%, respectively. Furthermore, the model was tested on subgroup 2 of the ISRUC-Sleep dataset, yielding accuracies of 98.78% and 98.46% for the classification of three and five sleep stages, respectively. Another study by Satapathy et al. [100] employed the same signals as in their previous study [99] and extracted statistical features to pass to an RF classifier. Their model classified five sleep stages on subgroups 1 and 2 of the ISRUC-Sleep dataset, achieving accuracies of 98.52% and 98.46%, respectively. Toma et al. [101] proposed a model for sleep stage classification, which aims to classify five sleep stages using features extracted from four distinct channel signals, namely, EEG (Fpz-Cz, Pz-Oz), EOG, and EMG signals

obtained from PSG recording. The model architecture consists of two key building blocks: the “Conv Block” and the “Bi-LSTM Block”. The Conv Block includes two consecutive 1D convolutional layers, a max-pooling layer, and a dropout layer for extracting spatial features from the input signals. On the other hand, the Bi-LSTM Block comprises a Bi-LSTM layer, a max-pooling layer, and a dropout layer to capture and learn temporal correlations in the data. By concatenating the outputs of these dual-channel convolutional Bi-LSTM network modules, the model classifies the five sleep stages and reported an accuracy of 91.44% in their study.

Later, Pei et al. [102] proposed a hybrid model that combined multiple signals, including EEG (C4-A1), EOG (EOGL and EOGR), and EMG signals. They fed the signals to a model architecture that consisted of seven layers of 1D-CNN and GRU. They tested their model on the SHHS dataset, utilising 717,883 segments, and their model achieved an accuracy of 83.15%. Huang et al. [103] proposed a DeConvolution- and Self-Attention-based Model (DCSAM) as a novel approach for the classification of five sleep stages. DCSAM has the capability to reverse the feature map of a hidden layer, mapping it back to the input space. The DCSAM model comprises five layers of 1D-CNN, followed by max-pooling layers. The final two layers consist of an attention layer and a fully connected layer. Their model achieved an accuracy rate of 90.26%.

Table 2. Selected studies used two or more physiological signals for sleep stage classification.

S.No	Author/Year	Dataset	Number of Samples/Recordings	Signals	Number of Channels	Input	Classification	Number of Classes	Accuracy	Kappa	Splitting Strategy
1	Esteevez et al., 2002 [104]	Private	11 recordings	EOG, EMG, EEG	-	FFT	Fuzzy rule	5	-	-	-
2	Estrada et al., 2006 [71]	Private	10 recordings	EOG, EMG	2	FFT	Fuzzy rule	5	-	-	-
3	Akin et al., 2008 [65]	Private	30 recordings	EEG, EMG	2	WT	DNN	3	-	98.00	50% training + 50% testing
4	Yu et al., 2012 [86]	Private	4 recordings	EEG, ECG	2	FFT	SVM	5	99.00	-	-
5	Long et al., 2014 [82]	Private	115 recordings	ECG, respiratory	-	Statistic features	LD	2	95.00	59.00	-
6	Willemen et al., 2014 [88]	Private	35,124 samples	EEG, EMG, respiratory	-	WT	SVM	5	69.00	69.50	-
7	Helland et al., 2015 [89]	Private	10 recordings	EEG, ECG, respiratory	3	Statistic features	BLD	3	80.00	-	-
8	Fonseca et al., 2015 [83]	Private	-	ECG, respiratory	2	Statistical features	BLD	3	80.00	49.00	-
9	Kim et al., 2018 [66]	Sleep-edf	5 recordings	EEG, EMG	2	TD	SVM	5	93.80	94.00	10-fold
10	Takatani et al., 2018 [90]	Private	431 recordings	EEG, ECG, EMG	-	RR+FFT	LD	5	80.00	-	-
11	Cui et al., 2018 [93]	ISRUC-Sleep	106 recordings	EEG, EOG, EMG	5	Fine-grained	2D-CNN	5	90.12	81.00	10-fold subject-wise
12	Tripathy et al., 2018 [85]	MIT-BIH	18 recordings	EEG, ECG	2	Statistic features	DNN	5	73.70	-	10-fold subject-wise
13	Yuan et al., 2018 [98]	UCD	25 recordings	EEG, ECG, EMG	-	Raw data STFT	1D-CNN 2D-CNN	5	73.00 74.22	-	-
14	Bisawal et al., 2018 [91]	Private	10,000 samples	EEG, EMG, ECG	6	FFT	1D-CNN+ Bi-LSTM	5	87.50	80.50	Train 90%, testing 10% subject-wise
15	Zhang et al., 2018 [94]	SHHS	5804 recordings	EEG, EMG, EOG	5	TD+FFT	2D-CNN	5	86.00	82.00	Train 90%, testing 10% subject-wise
16	Chambon et al., 2018 [95]	MASS	62 recordings	EEG, EOG, EMG	11	Raw data	2D-CNN	5	79.00	70.00	5-fold subject-wise
17	Phan et al., 2019 [74]	MASS	200 recordings	EEG, EOG	2	FFT	2D-CNN	5	87.10	81.50	20-fold subject-wise
18	Yildirim et al., 2019 [72]	Sleep-edf	15,188 samples	EEG, EOG	2	Raw data	1D-CNN	3 5	94.64 91.22	-	Training 70%, validation 15%, testing 15% non-subject-wise
	Sleep-edfx	127,512 samples	2		3 5			94.34 90.98			
19	Blanco et al., 2019 [59]	Sleep-edfx	20 recordings	EEG	2	Raw data	1D-CNN	5	92.60	84.00	20-fold subject-wise
20	Phan et al., 2019 [105]	Sleep-edf	20 recordings	EEG, EMG, EOG	2	FFT	2D-CNN	5	82.30	75.00	Training 19 subjects, validation 4 subjects, testing 4 subjects
	MASS	200 recordings	82.50						75.00	20-fold cross-validation	

Table 2. Cont.

S.No	Author/Year	Dataset	Number of Samples/Recordings	Signals	Number of Channels	Input	Classification	Number of Classes	Accuracy	Kappa	Splitting Strategy																																																																																																																																																																
21	Satapathy et al., 2020 [60]	ISRUC-Sleep Subgroup 1	6000 samples	EEG	2	Raw data	1D-CNN	5	97.22	-	Training 70%, testing 30%																																																																																																																																																																
		ISRUC-Sleep Subgroup 2							95.06	-		22	Tautan et al., 2020 [64]	PhysioNet Challenge	994 recordings	EEG, ECG	2	Statistic features+FFT	RF	5	72.52	-	10-fold subject-wise	EEG, EMG	88.65	-	EEG, respiratory	93.72	-	EEG, ECG	60.28	-	EEG, EMG	66.70	-	EEG, respiratory	52.27	-	23	Sokolovsky et al., 2020 [73]	Sleep-edfx	20 recordings	EEG, EOG	3	Raw data	1D-CNN	5	81.00	-	10-fold subject-wise	24	Xu et al., 2020 [96]	Sleep-edf	37,628 samples	EEG, EMG, EOG	4	Raw data	1D-CNN	5	85.40	78.90	5-fold subject-wise	Sleep-edfx	213,695 samples	81.60	74.70	25	Delimayanti et al., 2020 [61]	Sleep-edfx	127,663 samples	EEG	2	FFT	SVM	3	94.14	-	10-fold	5	91.37	-	26	Casal et al., 2021 [84]	SHHS	5000 recordings	ECG, respiratory	2	Raw data	GRU	2	90.13	74.00	Training 50%, validation 25%, testing 25% subject-wise	27	Zhao et al., 2021 [87]	MIT-BIH	10,127 samples	EEG, ECG	2	Raw data	1D-CNN	2	98.84	-	10 fold	28	Sharma et al., 2022 [97]	SHHS visit 1	5,861,304 samples	EEG, EOG, EMG	5	WT	BT	3	95.05	83.80	Training 90%, testing 10%	SHHS visit 2	3,037,838 samples	5	94.79	86.00	29	Satapathy et al., 2022 [99]	ISRUC-Sleep Subgroup 1	3750 samples	EEG, EOG, EMG	3	Raw data	1D-CNN	3	98.61	-	Training 70%, testing 30%	ISRUC-Sleep Subgroup 2	3750 samples	5	89.46	-	30	Satapathy et al., 2022 [100]	ISRUC-Sleep Subgroup 1	3750 samples	EEG, EOG, EMG	3	Statistic features	RF	3	98.78	-	Training 70%, testing 30%	ISRUC-Sleep Subgroup 3	3750 samples	5	98.46	-	31	Pie et al., 2022 [102]	SHHS visit 1	717,883 samples	EEG, EMG, EOG	4	Raw data	1D-CNN	5	83.15	89.00	Training 50%, validation 20%, testing 30%	32	Sekkal et al., 2022 [75]	Sleep-edfx
22	Tautan et al., 2020 [64]	PhysioNet Challenge	994 recordings	EEG, ECG	2	Statistic features+FFT	RF	5	72.52	-	10-fold subject-wise																																																																																																																																																																
				EEG, EMG					88.65	-																																																																																																																																																																	
				EEG, respiratory					93.72	-																																																																																																																																																																	
				EEG, ECG	60.28	-																																																																																																																																																																					
				EEG, EMG	66.70	-																																																																																																																																																																					
EEG, respiratory	52.27	-																																																																																																																																																																									
23	Sokolovsky et al., 2020 [73]	Sleep-edfx	20 recordings	EEG, EOG	3	Raw data	1D-CNN	5	81.00	-	10-fold subject-wise																																																																																																																																																																
24	Xu et al., 2020 [96]	Sleep-edf	37,628 samples	EEG, EMG, EOG	4	Raw data	1D-CNN	5	85.40	78.90	5-fold subject-wise																																																																																																																																																																
		Sleep-edfx	213,695 samples						81.60	74.70																																																																																																																																																																	
25	Delimayanti et al., 2020 [61]	Sleep-edfx	127,663 samples	EEG	2	FFT	SVM	3	94.14	-	10-fold																																																																																																																																																																
								5	91.37	-																																																																																																																																																																	
26	Casal et al., 2021 [84]	SHHS	5000 recordings	ECG, respiratory	2	Raw data	GRU	2	90.13	74.00	Training 50%, validation 25%, testing 25% subject-wise																																																																																																																																																																
27	Zhao et al., 2021 [87]	MIT-BIH	10,127 samples	EEG, ECG	2	Raw data	1D-CNN	2	98.84	-	10 fold																																																																																																																																																																
28	Sharma et al., 2022 [97]	SHHS visit 1	5,861,304 samples	EEG, EOG, EMG	5	WT	BT	3	95.05	83.80	Training 90%, testing 10%																																																																																																																																																																
		SHHS visit 2	3,037,838 samples					5	94.79	86.00																																																																																																																																																																	
29	Satapathy et al., 2022 [99]	ISRUC-Sleep Subgroup 1	3750 samples	EEG, EOG, EMG	3	Raw data	1D-CNN	3	98.61	-	Training 70%, testing 30%																																																																																																																																																																
		ISRUC-Sleep Subgroup 2	3750 samples					5	89.46	-																																																																																																																																																																	
30	Satapathy et al., 2022 [100]	ISRUC-Sleep Subgroup 1	3750 samples	EEG, EOG, EMG	3	Statistic features	RF	3	98.78	-	Training 70%, testing 30%																																																																																																																																																																
		ISRUC-Sleep Subgroup 3	3750 samples					5	98.46	-																																																																																																																																																																	
31	Pie et al., 2022 [102]	SHHS visit 1	717,883 samples	EEG, EMG, EOG	4	Raw data	1D-CNN	5	83.15	89.00	Training 50%, validation 20%, testing 30%																																																																																																																																																																
32	Sekkal et al., 2022 [75]	Sleep-edfx	21,265 samples	EEG, EOG	3	Statistic features	SVM	5	89.10	82.00	Training 80%, testing 15%																																																																																																																																																																

Table 2. Cont.

S.No	Author/Year	Dataset	Number of Samples/Recordings	Signals	Number of Channels	Input	Classification	Number of Classes	Accuracy	Kappa	Splitting Strategy
33	Almutairi et al., 2023 [67]	Sleep-edfx	72,000 samples	EEG, EMG	3	Raw data	1D-CNN + LSTM	3	95.46	90.12	Training 70%, validation 15%, testing 15%, non-subject-wise
				EEG, EOG	3			5	95.65	89.70	
		ISRUC-Sleep Subgroup 1	56,515 samples	EEG, EMG, EOG	4	Raw data	1D-CNN + LSTM	3	96.36	93.40	
				EEG, EMG, EOG	5			5	96.57	83.05	
34	Choi et al., 2023 [92]	SHHS	9736 recordings	EKG, EMG, EEG	3	Statistic features	XGBoost	5	85.00	-	10-fold non-subject-wise
35	Dequidt et al., 2023 [62]	MASS	62 recordings	EEG	8	FFT	VGG-16	5	82.96	80.90	31-fold subject-wise
36	Toma et al., 2023 [101]	Sleep-edf	20 recordings	EEG, EMG, EOG	4	Raw data	1D-CNN + Bi-LSTM	5	91.44	89.00	Training 85%, testing 15% non-subject-wise
37	Toma et al., 2023 [76]	Sleep-edf	20 recordings	EEG, EOG	3	Raw data	1D-CNN + RNN	5	90.30	86.86	Training 85%, testing 15% non-subject-wise
38	Huang et al., 2023 [103]	Sleep-edfx	20 recordings	EEG, EOG, EMG	3	Raw data	1D-CNN + attention	5	90.30	86.86	-

5. Gaps in Literature

This section aims to provide a valuable resource for scholars who are seeking to gain a comprehensive understanding of the limitations within the current state of the literature. The classification of sleep stages is crucial as it helps in detecting sleep disorders, which can have significant implications for life-threatening conditions. The literature review indicates that the existing framework used for sleep stage classification encounters one or more of the following limitations.

5.1. Testing of Multiple Datasets

Enhanced reliability and generalizability of a proposed model can be achieved by evaluating a model with multiple datasets collected through the implementation of diverse recording equipment and laboratory practices. As a result, it guarantees the model's strength and effectiveness in handling various scenarios [106]. A considerable amount of the existing literature assesses models using a single dataset. However, a growing body of literature as highlighted in [60,67,96,97,99,100] tested their models across multiple datasets with variations in data collection environments. Results from these studies highlight the critical role that the diversity in datasets and data collection methods plays in improving a model's robustness.

5.2. Splitting Strategy

The splitting strategy used to divide a dataset into training, validation, and testing sets can impact a model's performance in the classification of sleep stages [107]. Many studies have employed random allocation, where data in the dataset are randomly divided into fixed percentages of training, validation, and testing, such as 70% training, 15% validation, and 15% testing, or 90% training and 10% testing. Another cross-validation approach has been utilised to calculate the average accuracy of the entire dataset. Another factor that can significantly influence model performance is the choice between a subject-wise or a non-subject-wise strategy. In a subject-wise approach, the model may recognise patterns in the test data in a more effective and generalisable way, as training and testing sets do not include the same subjects. Conversely, a non-subject-wise approach may result in the model being unable to adequately generalise, as it recognises similar patterns in training and testing data [108]. These considerations highlight the importance of carefully selecting the splitting strategy and considering the subject-wise or non-subject-wise approach to ensure an accurate, reliable classification of sleep stages.

5.3. Computational Complexity

Computational complexity is mainly associated with training and deploying deep learning models. Deep neural networks often exhibit several parameters, leading to increased computational demands and longer training times [109]. Researchers have tackled this problem by prioritising the development of low-parameter deep learning models. Therefore, there is a need to propose new, efficient models that are less computationally complex while maintaining high performance standards.

5.4. Imbalanced Dataset

The classification of sleep stages is hindered by the limitations posed by imbalanced datasets. Sleep stage classification involves training machine learning models to accurately identify sleep stages based on physiological signals. However, imbalanced datasets arise due to the uneven distribution of samples across sleep stages. Sleep time is predominantly spent in the N2 stage, while other stages, such as N1, N3, and REM, are comparatively less frequent. This inherent class imbalance leads to a bias in the model's performance, with a tendency to favour dominant classes. Consequently, minority classes such as N1, N3, and REM sleep stages may be poorly classified [110]. The insufficient representation of these under-represented classes makes it challenging for models to learn their distinctive characteristics.

5.5. Scarcity of Studies Using a Combination of Signals for Sleep Stage Classification

In this study, we identified only 38 out of 1427 studies that utilised a combination of signals and machine learning models. This finding underscores the prevalent focus of researchers on utilising single-channel EEG signals with machine learning. The utilization of multiple signals available from PSG is reported in the identified studies to provide additional features that aid in accurately classifying sleep stages. Hence, we suggest that future studies explore the use of combined physiological signals to enhance the accuracy of sleep stage classification.

6. Discussion

This systematic review investigated the effectiveness of using machine learning on a combination of physiological signals in sleep stage classification. The combination of signals from multiple physiological sources has gained attention, as it has been found to be a promising approach for enhancing the accuracy and reliability of sleep stage classification. By leveraging complementary information captured by signals, researchers aim to improve the overall performance of sleep stage classification models. The studies included in the literature review were characterised based on the type and number of physiological signals used, the classification models employed, and the accuracy achieved. Figure 6 presents the distribution of the total number of studies that utilised either multiple channels of a single type of physiological signal or a combination of signal types for the classification of sleep stages.

EEG signals	EOG signals	Combination of signals
Multiple channels of EEG (4)	EOG+EMG (1) EOG+EEG (6)	EEG, EMG, ECG (3) EEG, EMG, EOG (13) EEG, ECG, EMG, respiratory (2)
EMG signals EMG+EEG (4)	ECG and respiratory signals with combination of different signals ECG+respiratory (2) ECG+EEG (4) respiratory+EEG (1)	

Figure 6. Distribution of studies using multiple channels of a single type of physiological signal or a combination of different types of physiological signals for the classification of sleep stages.

Most of the reviewed studies utilised a combination of EEG + EOG + EMG signals for sleep stage classification, as these signals provided more accurate discrimination between sleep stages. They capture both brain activity and eye movement patterns that characterise each sleep stage. Additionally, incorporating EMG signals provides valuable information about muscle activity and helps differentiate sleep stages with varying muscle tone [67].

The selection of a signal's channels is critical for the performance of the sleep stage classification model. Typically, multiple signal channels are used instead of using a single channel. However, utilising additional channels can increase the costs associated with the recording configuration and impose a greater computational complexity on machine learning models. Thus, the selection of channels balancing the accuracy and efficiency is an important research area. For example, Cui et al. [93] observed that increasing the number of channels correlated with enhanced model performance. Chambon et al. [95] demonstrated that their research using a set of six EEG channels produced results comparable to those obtained with a larger set of 20 EEG channels. Sharma et al. [97] conducted a comprehensive investigation that explored 15 signal configurations. Notably, among this array of combinations, the one that incorporated the specific set of five channels, as proposed in

their research, consistently demonstrated superior performance in the classification of sleep stages. Furthermore, Sekkal et al. [75] comprehensively compared signal combinations and a single-channel EEG. They compared combinations of signals with a single-channel EEG. The study found that when specific classifiers were used, there was only a small decrease in accuracy, even when using a single EEG channel or different signal combinations. This suggests that the choice of classifiers plays a significant role in maintaining the accuracy of sleep stage classification. Investigations by Almutairi et al. [67] and Dequidt et al. [62] exploring numbers of channels of signal combinations revealed the potential for significant improvements in model results by utilising all available channels from the dataset used in their proposed models.

It is also observed from the literature that majority of the studies for the classification of sleep stages used feature extraction methods as a pivotal step in their data processing pipelines. Figure 7 presents the distribution of studies utilising different feature extraction methods and raw signals as inputs to machine learning models. A total of 27 studies have chosen to employ feature extraction techniques to extract crucial information or features from raw data. These methods are designed to condense and represent underlying patterns in a more informative manner [111]. In parallel, an alternative approach has been embraced by 14 studies, wherein they directly utilise raw, unprocessed data as input for their classification models. This distinction highlights the variety of methodologies used within this research domain, where some researchers prioritise feature engineering, while others take advantage of deep learning and sophisticated machine learning techniques with raw data input [111].

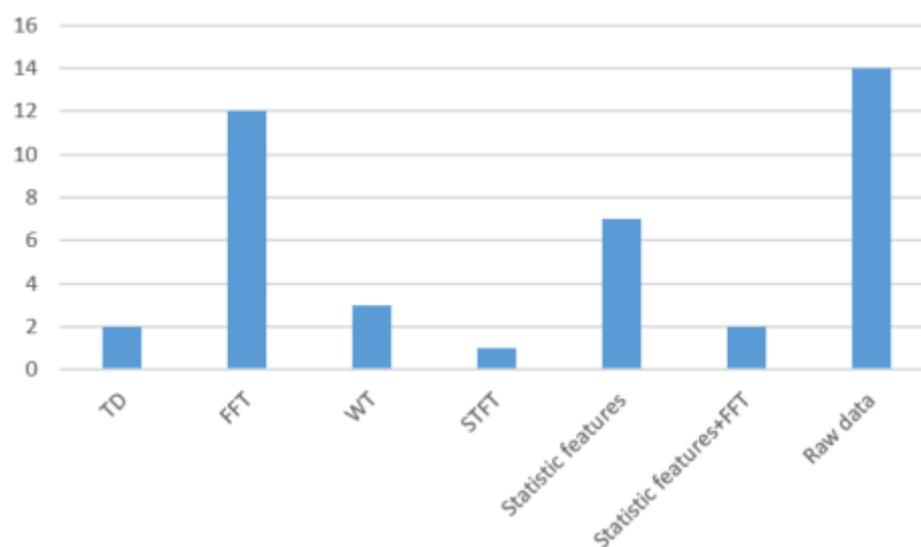


Figure 7. The distribution of studies used feature extraction methods or raw data as input to machine learning.

Figure 8 illustrates the distribution of the utilisation of sleep datasets in studies dedicated to sleep stage classification through the utilisation of ML techniques. The most frequently used open-source datasets in sleep research are ISRUC-Sleep, Sleep-edf, Sleep-edfx, and MASS. These datasets stand out due to their availability and diversity. In contrast, the usage of the PhysioNet Challenge 2018 and the MIT-BIH, UCD, and SHHS datasets is comparatively low due to restrictive access and small size. However, when comparing the performances of machine learning models across studies, a significant challenge arises due to variations in datasets and sample sizes. These dataset differences, including data source, diversity, and size, introduce confounding factors that complicate direct model comparisons. A model trained on a small, specialised dataset may excel within that context but might not generalise to another dataset with distinct characteristics [112]. Therefore, it is essential to recognise that identifying the ‘best’ model is highly context dependent,

and meaningful comparisons necessitate careful consideration of the data and sample sizes underlying each model's evaluation.

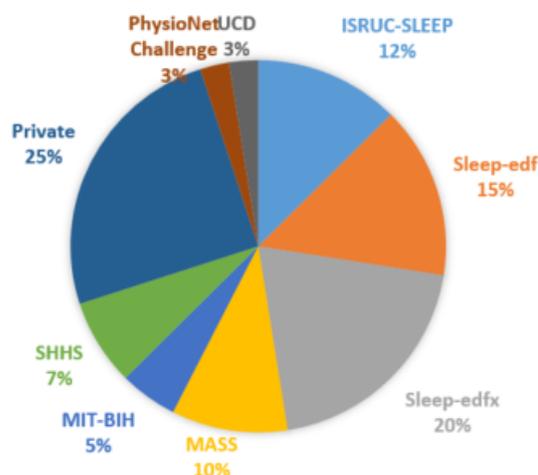


Figure 8. The distribution of the utilization of each sleep dataset in studies employing ML techniques for sleep stage classification.

Figure 9 illustrates the ML models proposed in the literature to classify sleep stages. CNN-based architectures are the most popular models for classifying sleep stages. In our review, we found that 14 studies proposed models based on 1D-CNN, and 6 studies proposed models based on 2D-CNN. These 1D-CNN models have the advantage of being computationally less complex than 2D-CNN models. In addition, 2D-CNN-based models require input signals to be converted from 1D to 2D. This conversion process must be carefully handled to prevent the potential loss of important information [113]. Therefore, 1D-CNN models are well suited for real-time applications, such as home-based sleep stage classification.

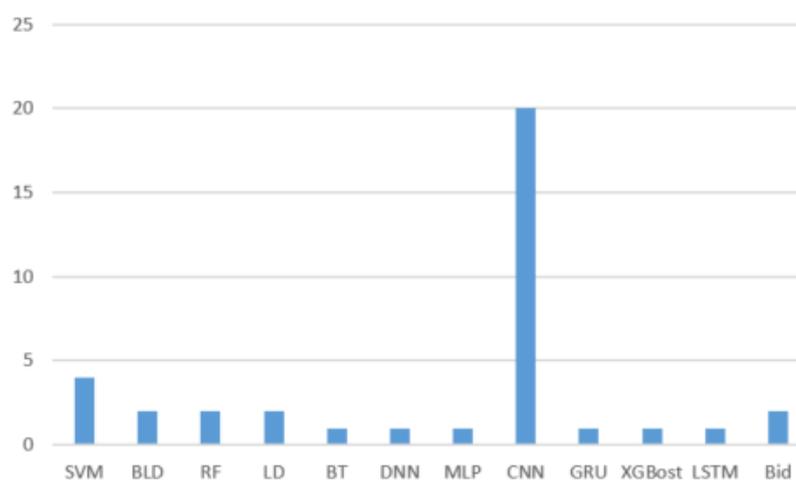


Figure 9. Distribution of different machine learning models for the classification of sleep stages.

7. Conclusions

This systematic review targets studies that employ machine learning techniques for sleep stage classification using combined multichannel physiological signals. These studies utilise signal combinations to enhance classification accuracy, with EEG, EMG, and EOG signals being the most frequently used inputs for machine learning models. Most reviewed studies proposed a variety of machine learning models for both three- and five-sleep-stage classifications. Additionally, a prevailing preference was observed for feature engineering methods over raw data utilisation. Furthermore, the review highlights an emerging trend that underscores the potential benefits of leveraging combined signals and deep learning

algorithms to achieve improved sleep stage classification. This trend represents a promising direction for future research and application in the field of sleep medicine.

To further advance sleep stage classification, future studies are recommended to consider additional metrics such as specificity, sensitivity, F1 score, and kappa for evaluating ML models. These metrics are especially beneficial when dealing with imbalanced datasets. Moreover, researchers are encouraged to evaluate model performance using both subject-wise and non-subject-wise evaluation approaches. This comparative analysis will yield valuable insights into the generalisability and effectiveness of the models across diverse data distributions. In the context of addressing imbalanced datasets, future research should also consider implementing data augmentation techniques to improve models' performance. Class imbalance difficulties can be overcome by creating synthetic samples from the minority class through data augmentation, thereby enhancing classification accuracy. Future research in the domain of sleep stage classification must prioritise the investigation of the most effective combination of physiological channels required for accurate and efficient classification.

Author Contributions: H.A.: investigation, writing—original draft, writing—review and editing, validation. G.M.H.: conceptualization, validation, writing—review and editing, supervision. A.D.: conceptualization, writing—review and editing, project administration, resources, supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Laposky, A.D.; Bass, J.; Kohsaka, A.; Turek, F.W. Sleep and circadian rhythms: Key components in the regulation of energy metabolism. *FEBS Lett.* **2008**, *582*, 142–151. [[CrossRef](#)] [[PubMed](#)]
- Cho, J.W.; Duffy, J.F. Sleep, sleep disorders, and sexual dysfunction. *World J. Men Health* **2019**, *37*, 261–275. [[CrossRef](#)] [[PubMed](#)]
- Ohayon, M.M. Epidemiological overview of sleep disorders in the general population. *Sleep Med. Res.* **2011**, *2*, 1–9. [[CrossRef](#)]
- Ohayon, M.M.; Smirne, S. Prevalence and consequences of insomnia disorders in the general population of Italy. *Sleep Med.* **2002**, *3*, 115–120. [[CrossRef](#)] [[PubMed](#)]
- Ohayon, M.M. Epidemiology of insomnia: What we know and what we still need to learn. *Sleep Med. Rev.* **2002**, *6*, 97–111. [[CrossRef](#)] [[PubMed](#)]
- Fietze, I.; Laharnar, N.; Bargiotas, P.; Basoglu, O.K.; Dogas, Z.; Drummond, M.; Fanfulla, F.; Gislason, T.; Gouveris, H.; Grote, L.; et al. Management of obstructive sleep apnea in Europe—A 10-year follow-up. *Sleep Med.* **2022**, *97*, 64–72. [[CrossRef](#)] [[PubMed](#)]
- Li, X.; Sotres-Alvarez, D.; Gallo, L.C.; Ramos, A.R.; Aviles-Santa, L.; Perreira, K.M.; Isasi, C.R.; Zee, P.C.; Savin, K.L.; Schneiderman, N.; et al. Associations of sleep-disordered breathing and insomnia with incident hypertension and diabetes. The Hispanic community health study/study of Latinos. *Am. J. Respir. Crit. Care Med.* **2021**, *203*, 356–365. [[CrossRef](#)] [[PubMed](#)]
- Streatfeild, J.; Smith, J.; Mansfield, D.; Pezzullo, L.; Hillman, D. The social and economic cost of sleep disorders. *Sleep* **2021**, *44*, zsab132. [[CrossRef](#)]
- Pennings, N.; Golden, L.; Yashi, K.; Tondt, J.; Bays, H.E. Sleep-disordered breathing, sleep apnea, and other obesity-related sleep disorders: An Obesity Medicine Association (OMA) Clinical Practice Statement (CPS) 2022. *Obes. Pillars* **2022**, *4*, 100043. [[CrossRef](#)]
- Yan, B.; Yang, J.; Zhao, B.; Fan, Y.; Wang, W.; Ma, X. Objective sleep efficiency predicts cardiovascular disease in a community population: The sleep heart health study. *J. Am. Heart Assoc.* **2021**, *10*, e016201. [[CrossRef](#)]
- Silber, M.H.; Ancoli-Israel, S.; Bonnet, M.H.; Chokroverty, S.; Grigg-Damberger, M.M.; Hirshkowitz, M.; Kapen, S.; Keenan, S.A.; Kryger, M.H.; Penzel, T.; et al. The visual scoring of sleep in adults. *J. Clin. Sleep Med.* **2007**, *3*, 121–131. [[CrossRef](#)]
- Obal, F., Jr.; Krueger, J.M. Biochemical regulation of non-rapid-eye-movement sleep. *Front. -Biosci.-Landmark* **2003**, *8*, 520–550.
- Somers, V.K.; Dyken, M.E.; Mark, A.L.; Abboud, F.M. Sympathetic-nerve activity during sleep in normal subjects. *N. Engl. J. Med.* **1993**, *328*, 303–307. [[CrossRef](#)]
- Penzel, T.; Kantelhardt, J.W.; Lo, C.C.; Voigt, K.; Vogelmeier, C. Dynamics of heart rate and sleep stages in normals and patients with sleep apnea. *Neuropsychopharmacology* **2003**, *28*, S48–S53. [[CrossRef](#)] [[PubMed](#)]
- Bloch, K.E. Polysomnography: A systematic review. *Technol. Health Care* **1997**, *5*, 285–305. [[CrossRef](#)] [[PubMed](#)]

16. Coronel, C.; Wiesmeyr, C.; Garn, H.; Kohn, B.; Wimmer, M.; Mandl, M.; Glos, M.; Penzel, T.; Klösch, G.; Stefanic-Kejik, A.; et al. Detection of respiratory events by respiratory effort and oxygen desaturation. *J. Med. Biol. Eng.* **2020**, *40*, 517–525. [[CrossRef](#)]
17. Campbell, I.G. EEG recording and analysis for sleep research. *Curr. Protoc. Neurosci.* **2009**, *49*, 10–12. [[CrossRef](#)] [[PubMed](#)]
18. Kesper, K.; Canisius, S.; Penzel, T.; Ploch, T.; Cassel, W. ECG signal analysis for the assessment of sleep-disordered breathing and sleep pattern. *Med. Biol. Eng. Comput.* **2012**, *50*, 135–144. [[CrossRef](#)]
19. Jammes, B.; Sharabty, H.; Esteve, D. Automatic EOG analysis: A first step toward automatic drowsiness scoring during wake-sleep transitions. *Somnologie-Schlafforschung Schlafmed.* **2008**, *12*, 227–232. [[CrossRef](#)]
20. Shokrollahi, M.; Krishnan, S. Sleep EMG analysis using sparse signal representation and classification. In Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 28 August–1 September 2012; pp. 3480–3483.
21. Steriade, M.M.; McCarley, R.W. *Brainstem Control of Wakefulness and Sleep*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
22. Collop, N.A. Scoring variability between polysomnography technologists in different sleep laboratories. *Sleep Med.* **2002**, *3*, 43–47. [[CrossRef](#)]
23. Zhou, X.Y.; Guo, Y.; Shen, M.; Yang, G.Z. Application of artificial intelligence in surgery. *Front. Med.* **2020**, *14*, 417–430. [[CrossRef](#)] [[PubMed](#)]
24. Albaqami, H.; Hassan, G.M.; Datta, A. Wavelet-Based Multi-Class Seizure Type Classification System. *Appl. Sci.* **2022**, *12*, 5702. [[CrossRef](#)]
25. Aboalayon, K.A.I.; Faezipour, M.; Almuhammadi, W.S.; Moslehpour, S. Sleep stage classification using EEG signal analysis: A comprehensive survey and new investigation. *Entropy* **2016**, *18*, 272. [[CrossRef](#)]
26. Zhang, Y.; Weng, Y.; Lund, J. Applications of explainable artificial intelligence in diagnosis and surgery. *Diagnostics* **2022**, *12*, 237. [[CrossRef](#)] [[PubMed](#)]
27. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
28. Ronzhina, M.; Janoušek, O.; Kolářová, J.; Nováková, M.; Honzík, P.; Provazník, I. Sleep scoring using artificial neural networks. *Sleep Med. Rev.* **2012**, *16*, 251–263. [[CrossRef](#)] [[PubMed](#)]
29. Movahedi, F.; Coyle, J.L.; Sejdić, E. Deep belief networks for electroencephalography: A review of recent contributions and future outlooks. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 642–652. [[CrossRef](#)]
30. Loh, H.W.; Ooi, C.P.; Vicnesh, J.; Oh, S.L.; Faust, O.; Gertych, A.; Acharya, U.R. Automated detection of sleep stages using deep learning techniques: A systematic review of the last decade (2010–2020). *Appl. Sci.* **2020**, *10*, 8963. [[CrossRef](#)]
31. Mishra, S.; Birok, R. Literature review: Sleep stage classification based on EEG signals using artificial intelligence technique. *Recent Trends Commun. Electron.* **2021**, *10*, 241–244.
32. Faust, O.; Razaghi, H.; Barika, R.; Ciaccio, E.J.; Acharya, U.R. A review of automated sleep stage scoring based on physiological signals for the new millennia. *Comput. Methods Programs Biomed.* **2019**, *176*, 81–91. [[CrossRef](#)]
33. Fiorillo, L.; Puiatti, A.; Papandrea, M.; Ratti, P.L.; Favaro, P.; Roth, C.; Bargiotas, P.; Bassetti, C.L.; Faraci, F.D. Automated sleep scoring: A review of the latest approaches. *Sleep Med. Rev.* **2019**, *48*, 101204. [[CrossRef](#)] [[PubMed](#)]
34. Dixon-Woods, M.; Bonas, S.; Booth, A.; Jones, D.R.; Miller, T.; Sutton, A.J.; Shaw, R.L.; Smith, J.A.; Young, B. How can systematic reviews incorporate qualitative research? A critical perspective. *Qual. Res.* **2006**, *6*, 27–44. [[CrossRef](#)]
35. Kemp, B.; Zwinderman, A.H.; Tuk, B.; Kamphuisen, H.A.; Obery, J.J. Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG. *IEEE Trans. Biomed. Eng.* **2000**, *47*, 1185–1194. [[CrossRef](#)] [[PubMed](#)]
36. Kemp, B.; Zwinderman, A.; Tuk, B.; Kamphuisen, H.; Obery, J. Sleep-EDF Database Expanded. Available online: <https://www.physionet.org> (accessed on 17 July 2018).
37. O’reilly, C.; Gosselin, N.; Carrier, J.; Nielsen, T. Montreal Archive of Sleep Studies: An open-access resource for instrument benchmarking and exploratory research. *J. Sleep Res.* **2014**, *23*, 628–635. [[CrossRef](#)] [[PubMed](#)]
38. Ichimaru, Y.; Moody, G. Development of the polysomnographic database on CD-ROM. *Psychiatry Clin. Neurosci.* **1999**, *53*, 175–177. [[CrossRef](#)] [[PubMed](#)]
39. Khalighi, S.; Sousa, T.; Santos, J.M.; Nunes, U. ISRUC-Sleep: A comprehensive public dataset for sleep researchers. *Comput. Methods Programs Biomed.* **2016**, *124*, 180–192. [[CrossRef](#)] [[PubMed](#)]
40. Quan, S.F.; Howard, B.V.; Iber, C.; Kiley, J.P.; Nieto, F.J.; O’Connor, G.T.; Rapoport, D.M.; Redline, S.; Robbins, J.; Samet, J.M.; et al. The sleep heart health study: Design, rationale, and methods. *Sleep* **1997**, *20*, 1077–1085.
41. Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **2000**, *101*, e215–e220. [[CrossRef](#)]
42. Ghassemi, M.M.; Moody, B.E.; Lehman, L.W.H.; Song, C.; Li, Q.; Sun, H.; Mark, R.G.; Westover, M.B.; Clifford, G.D. You snooze, you win: The physionet/computing in cardiology challenge 2018. In Proceedings of the 2018 Computing in Cardiology Conference (CinC), Maastricht, The Netherlands, 23–26 September 2018; Volume 45, pp. 1–4.
43. Redfern, M.S.; Hughes, R.E.; Chaffin, D.B. High-pass filtering to remove electrocardiographic interference from torso EMG recordings. *Clin. Biomech.* **1993**, *8*, 44–48. [[CrossRef](#)]
44. Mohamad, I.B.; Usman, D. Standardization and its effects on K-means clustering algorithm. *Res. J. Appl. Sci. Eng. Technol.* **2013**, *6*, 3299–3303. [[CrossRef](#)]

45. Karthik, G.V.S.; Fathima, S.Y.; Rahman, M.Z.U.; Ahamed, S.R.; Lay-Ekuakille, A. Efficient signal conditioning techniques for brain activity in remote health monitoring network. *IEEE Sensors J.* **2013**, *13*, 3276–3283. [[CrossRef](#)]
46. Nussbaumer, H.J. Fast convolution algorithms. In *Fast Fourier Transform and Convolution Algorithms*; Springer: Berlin/Heidelberg, Germany, 1982; pp. 32–79.
47. Sundararajan, D. *Discrete Wavelet Transform: A Signal Processing Approach*; John Wiley & Sons: Hoboken, NJ, USA, 2016.
48. Phan, H.; Andreotti, F.; Cooray, N.; Chén, O.Y.; De Vos, M. DNN filter bank improves 1-max pooling CNN for single-channel EEG automatic sleep stage classification. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 453–456.
49. Pan, J.; Tompkins, W.J. A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng.* **1985**, *3*, 230–236. [[CrossRef](#)] [[PubMed](#)]
50. Koles, Z.; Lind, J.; Soong, A. Spatio-temporal decomposition of the EEG: A general approach to the isolation and localization of sources. *Electroencephalogr. Clin. Neurophysiol.* **1995**, *95*, 219–230. [[CrossRef](#)] [[PubMed](#)]
51. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [[CrossRef](#)]
52. Yunita, A.; Santoso, H.B.; Hasibuan, Z.A. Deep Learning for Predicting Students' Academic Performance. In Proceedings of the 2019 Fourth International Conference on Informatics and Computing (ICIC), Semarang, Indonesia, 16–17 October 2019; pp. 1–6.
53. Yadav, S.; Shukla, S. Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In Proceedings of the 2016 IEEE 6th International Conference on Advanced Computing (IACC), Bhimavaram, Andhra Pradesh, India, 27–28 February 2016; pp. 78–83.
54. Morley, A.; Hill, L.; Kaditis, A. *10–20 System EEG Placement*; European Respiratory Society: Lausanne, Switzerland, 2016; p. 34.
55. Lee, M.; Song, C.B.; Shin, G.H.; Lee, S.W. Possible effect of binaural beat combined with autonomous sensory meridian response for inducing sleep. *Front. Hum. Neurosci.* **2019**, *13*, 425. [[CrossRef](#)] [[PubMed](#)]
56. Lechat, B.; Hansen, K.; Catcheside, P.; Zajamsek, B. Beyond K-complex binary scoring during sleep: Probabilistic classification using deep learning. *Sleep* **2020**, *43*, zsa0077. [[CrossRef](#)] [[PubMed](#)]
57. Garcia-Molina, G.; Tsoneva, T.; Jasko, J.; Steele, B.; Aquino, A.; Baher, K.; Pastoor, S.; Pfundtner, S.; Ostrowski, L.; Miller, B.; et al. Closed-loop system to enhance slow-wave activity. *J. Neural Eng.* **2018**, *15*, 066018. [[CrossRef](#)]
58. Nir, Y.; Massimini, M.; Boly, M.; Tononi, G. Sleep and consciousness. In *Neuroimaging of Consciousness*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 133–182.
59. Fernandez-Blanco, E.; Rivero, D.; Pazos, A. Convolutional neural networks for sleep stage scoring on a two-channel EEG signal. *Soft Comput.* **2020**, *24*, 4067–4079. [[CrossRef](#)]
60. Satapathy, S.K.; Loganathan, D.; Narayanan, P.; Sharathkumar, S. Convolutional neural network for classification of multiple sleep stages from dual-channel EEG signals. In Proceedings of the 2020 IEEE 4th Conference on Information & Communication Technology (CICT), Chennai, India, 3–5 December 2020; pp. 1–16.
61. Delimayanti, M.K.; Laya, M.; Faisal, M.R.; Naryanto, R.F.; Satou, K. The Effect of Feature Selection on Automatic Sleep Stage Classification Based On Multichannel EEG Signals. In Proceedings of the 2021 IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Purwokerto, Indonesia, 24–25 November 2021; pp. 272–276.
62. Dequidt, P.; Seraphim, M.; Lechervy, A.; Gaez, I.I.; Brun, L.; Etard, O. Automatic Sleep Stage Classification on EEG Signals Using Time-Frequency Representation. In Proceedings of the International Conference on Artificial Intelligence in Medicine, Portoroz, Slovenia, 12–15 June 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 250–259.
63. Levendowski, D.J.; Louis, E.K.S.; Strambi, L.F.; Galbiati, A.; Westbrook, P.; Berka, C. Comparison of EMG power during sleep from the submental and frontalis muscles. *Nat. Sci. Sleep* **2018**, *10*, 431. [[CrossRef](#)]
64. Tăutan, A.M.; Rossi, A.C.; de Francisco, R.; Ionescu, B. Automatic sleep stage detection: A study on the influence of various PSG input signals. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 5330–5334.
65. Akin, M.; Kurt, M.B.; Sezgin, N.; Bayram, M. Estimating vigilance level by using EEG and EMG signals. *Neural Comput. Appl.* **2008**, *17*, 227–236. [[CrossRef](#)]
66. Kim, H.; Choi, S. Automatic sleep stage classification using eeg and emg signal. In Proceedings of the 2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN), Prague, Czech Republic, 3–6 July 2018; pp. 207–212.
67. Almutairi, H.; Hassan, G.M.; Datta, A. Classification of sleep stages from EEG, EOG and EMG signals by SSNet. *arXiv* **2023**, arXiv:2307.05373.
68. Banerjee, A.; Pal, M.; Tibarewala, D.; Konar, A. Electrooculogram based blink detection to limit the risk of eye dystonia. In Proceedings of the 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR), Kolkata, India, 4–7 January 2015; pp. 1–6.
69. Banerjee, A.; Pal, M.; Datta, S.; Tibarewala, D.; Konar, A. Eye movement sequence analysis using electrooculogram to assist autistic children. *Biomed. Signal Process. Control.* **2014**, *14*, 134–140. [[CrossRef](#)]
70. Leclair-Visonneau, L.; Oudiette, D.; Gaymard, B.; Leu-Semenescu, S.; Arnulf, I. Do the eyes scan dream images during rapid eye movement sleep? Evidence from the rapid eye movement sleep behaviour disorder model. *Brain* **2010**, *133*, 1737–1746. [[CrossRef](#)] [[PubMed](#)]
71. Estrada, E.; Nazeran, H.; Barragan, J.; Burk, J.R.; Lucas, E.A.; Behbehani, K. EOG and EMG: Two important switches in automatic sleep stage classification. In Proceedings of the 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, USA, 30 August–3 September 2006; pp. 2458–2461.

72. Yildirim, O.; Baloglu, U.B.; Acharya, U.R. A deep learning model for automated sleep stages classification using PSG signals. *Int. J. Environ. Res. Public Health* **2019**, *16*, 599. [[CrossRef](#)]
73. Sokolovsky, M.; Guerrero, F.; Paisarnsriromsuk, S.; Ruiz, C.; Alvarez, S.A. Deep learning for automated feature discovery and classification of sleep stages. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *17*, 1835–1845. [[CrossRef](#)] [[PubMed](#)]
74. Phan, H.; Andreotti, F.; Cooray, N.; Chén, O.Y.; De Vos, M. SeqSleepNet: End-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 400–410. [[CrossRef](#)] [[PubMed](#)]
75. Sekkal, R.N.; Bereksi-Reguig, F.; Ruiz-Fernandez, D.; Dib, N.; Sekkal, S. Automatic sleep stage classification: From classical machine learning methods to deep learning. *Biomed. Signal Process. Control.* **2022**, *77*, 103751. [[CrossRef](#)]
76. Toma, T.I.; Choi, S. An End-to-End Convolutional Recurrent Neural Network with Multi-Source Data Fusion for Sleep Stage Classification. In Proceedings of the 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), Bali, Indonesia, 20–23 February 2023; pp. 564–569.
77. Kaplan, V.; Zhang, J.; Russi, E.; Bloch, K. Detection of inspiratory flow limitation during sleep by computer assisted respiratory inductive plethysmography. *Eur. Respir. J.* **2000**, *15*, 570–578. [[CrossRef](#)]
78. Seena, V.; Yomas, J. A review on feature extraction and denoising of ECG signal using wavelet transform. In Proceedings of the 2014 2nd International Conference on Devices, Circuits and Systems (ICDCS), Coimbatore, India, 6–8 March 2014; pp. 1–6.
79. Silva, C.V.; Philominraj, A.; del Río, C. A DSP Practical Application: Working on ECG Signal. In *Applications of Digital Signal Processing*; IntechOpen: London, UK, 2011.
80. Snyder, F.; Hobson, J.A.; Morrison, D.F.; Goldfrank, F. Changes in respiration, heart rate, and systolic blood pressure in human sleep. *J. Appl. Physiol.* **1964**, *19*, 417–422. [[CrossRef](#)]
81. Gaiduk, M.; Rodríguez, J.J.P.; Seepold, R.; Madrid, N.M.; Penzel, T.; Glos, M.; Ortega, J.A. Estimation of sleep stages analyzing respiratory and movement signals. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 505–514. [[CrossRef](#)]
82. Long, X.; Fonseca, P.; Foussier, J.; Haakma, R.; Aarts, R.M. Sleep and wake classification with actigraphy and respiratory effort using dynamic warping. *IEEE J. Biomed. Health Inform.* **2013**, *18*, 1272–1284. [[CrossRef](#)]
83. Fonseca, P.; Long, X.; Radha, M.; Haakma, R.; Aarts, R.M.; Rolink, J. Sleep stage classification with ECG and respiratory effort. *Physiol. Meas.* **2015**, *36*, 2027. [[CrossRef](#)] [[PubMed](#)]
84. Casal, R.; Di Persia, L.E.; Schlotthauer, G. Classifying sleep–wake stages through recurrent neural networks using pulse oximetry signals. *Biomed. Signal Process. Control.* **2021**, *63*, 102195. [[CrossRef](#)]
85. Tripathy, R.; Acharya, U.R. Use of features from RR-time series and EEG signals for automated classification of sleep stages in deep neural network framework. *Biocybern. Biomed. Eng.* **2018**, *38*, 890–902. [[CrossRef](#)]
86. Yu, S.; Chen, X.; Wang, B.; Wang, X. Automatic sleep stage classification based on ECG and EEG features for day time short nap evaluation. In Proceedings of the 10th World Congress on Intelligent Control and Automation, Beijing, China, 6–8 July 2012; pp. 4974–4977.
87. Zhao, R.; Xia, Y.; Wang, Q. Dual-modal and multi-scale deep neural networks for sleep staging using EEG and ECG signals. *Biomed. Signal Process. Control.* **2021**, *66*, 102455. [[CrossRef](#)]
88. Willemen, T.; Van Deun, D.; Verhaert, V.; Vandekerckhove, M.; Exadaktylos, V.; Verbraecken, J.; Van Huffel, S.; Haex, B.; Vander Sloten, J. An evaluation of cardiorespiratory and movement features with respect to sleep-stage classification. *IEEE J. Biomed. Health Inform.* **2013**, *18*, 661–669. [[CrossRef](#)] [[PubMed](#)]
89. Helland, V.F.; Gapelyuk, A.; Suhrbier, A.; Riedl, M.; Penzel, T.; Kurths, J.; Wessel, N. Investigation of an automatic sleep stage classification by means of multiscored hypnogram. *Methods Inf. Med.* **2010**, *49*, 467–472. [[CrossRef](#)] [[PubMed](#)]
90. Takatani, T.; Takahashi, Y.; Yoshida, R.; Imai, R.; Uchiike, T.; Yamazaki, M.; Shima, M.; Nishikubo, T.; Ikada, Y.; Fujimoto, S. Relationship between frequency spectrum of heart rate variability and autonomic nervous activities during sleep in newborns. *Brain Dev.* **2018**, *40*, 165–171. [[CrossRef](#)]
91. Biswal, S.; Sun, H.; Goparaju, B.; Westover, M.B.; Sun, J.; Bianchi, M.T. Expert-level sleep scoring with deep neural networks. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 1643–1650. [[CrossRef](#)]
92. Choi, J.; Kwon, S.; Park, S.; Han, S. Validation of the influence of biosignals on performance of machine learning algorithms for sleep stage classification. *Digit. Health* **2023**, *9*, 20552076231163783. [[CrossRef](#)]
93. Cui, Z.; Zheng, X.; Shao, X.; Cui, L. Automatic sleep stage classification based on convolutional neural network and fine-grained segments. *Complexity* **2018**, *2018*, 9248410. [[CrossRef](#)]
94. Zhang, L.; Fabbri, D.; Upender, R.; Kent, D. Automated sleep stage scoring of the Sleep Heart Health Study using deep neural networks. *Sleep* **2019**, *42*, zsz159. [[CrossRef](#)] [[PubMed](#)]
95. Chambon, S.; Galtier, M.N.; Arnal, P.J.; Wainrib, G.; Gramfort, A. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 758–769. [[CrossRef](#)] [[PubMed](#)]
96. Xu, M.; Wang, X.; Zhang, X.; Bin, G.; Jia, Z.; Chen, K. Computation-Efficient Multi-Model Deep Neural Network for Sleep Stage Classification. In Proceedings of the 2020 Asia Service Sciences and Software Engineering Conference, Nagoya, Japan, 13–15 May 2020; pp. 1–8.
97. Sharma, M.; Yadav, A.; Tiwari, J.; Karabatak, M.; Yildirim, O.; Acharya, U.R. An Automated Wavelet-Based Sleep Scoring Model Using EEG, EMG, and EOG Signals with More Than 8000 Subjects. *Int. J. Environ. Res. Public Health* **2022**, *19*, 7176. [[CrossRef](#)]

98. Yuan, Y.; Jia, K.; Ma, F.; Xun, G.; Wang, Y.; Su, L.; Zhang, A. A hybrid self-attention deep learning framework for multivariate sleep stage classification. *BMC Bioinform.* **2019**, *20*, 586. [[CrossRef](#)] [[PubMed](#)]
99. Satapathy, S.K.; Loganathan, D. Automated classification of multi-class sleep stages classification using polysomnography signals: A nine-layer 1D-convolution neural network approach. *Multimed. Tools Appl.* **2023**, *82*, 8049–8091. [[CrossRef](#)]
100. Satapathy, S.K.; Loganathan, D. Multimodal multiclass machine learning model for automated sleep staging based on time series data. *SN Comput. Sci.* **2022**, *3*, 276. [[CrossRef](#)]
101. Toma, T.I.; Choi, S. An End-to-End Multi-Channel Convolutional Bi-LSTM Network for Automatic Sleep Stage Detection. *Sensors* **2023**, *23*, 4950. [[CrossRef](#)]
102. Pei, W.; Li, Y.; Siuly, S.; Wen, P. A hybrid deep learning scheme for multi-channel sleep stage classification. *Comput. Mater. Contin.* **2022**, *71*, 889–905.
103. Huang, X.; Shirahama, K.; Irshad, M.T.; Nisar, M.A.; Piet, A.; Grzegorzec, M. Sleep Stage Classification in Children Using Self-Attention and Gaussian Noise Data Augmentation. *Sensors* **2023**, *23*, 3446. [[CrossRef](#)]
104. Estévez, P.; Held, C.; Holzmann, C.; Perez, C.; Pérez, J.; Heiss, J.; Garrido, M.; Peirano, P. Polysomnographic pattern recognition for automated classification of sleep-waking states in infants. *Med. Biol. Eng. Comput.* **2002**, *40*, 105–113. [[CrossRef](#)] [[PubMed](#)]
105. Phan, H.; Andreotti, F.; Cooray, N.; Chén, O.Y.; De Vos, M. Joint classification and prediction CNN framework for automatic sleep stage classification. *IEEE Trans. Biomed. Eng.* **2018**, *66*, 1285–1296. [[CrossRef](#)] [[PubMed](#)]
106. Özdemir, A.; Yavuz, U.; Dael, F.A. Performance evaluation of different classification techniques using different datasets. *Int. J. Electr. Comput. Eng.* **2019**, *9*, 3584–3590. [[CrossRef](#)]
107. Laber, E.S.; Pereira, F.d.A.M. Splitting criteria for classification problems with multi-valued attributes and large number of classes. *Pattern Recognit. Lett.* **2018**, *111*, 58–63. [[CrossRef](#)]
108. Tougui, I.; Jilbab, A.; El Mhamdi, J. Impact of the choice of cross-validation techniques on the results of machine learning-based diagnostic applications. *Healthc. Inform. Res.* **2021**, *27*, 189–199. [[CrossRef](#)] [[PubMed](#)]
109. Shinde, P.P.; Shah, S. A review of machine learning and deep learning applications. In Proceedings of the 2018 Fourth International Conference on Computing Communication Control and Automation (ICCCUBEA), Pune, India, 16–18 August 2018; pp. 1–6.
110. Utomo, O.K.; Surantha, N.; Isa, S.M.; Soewito, B. Automatic sleep stage classification using weighted ELM and PSO on imbalanced data from single lead ECG. *Procedia Comput. Sci.* **2019**, *157*, 321–328. [[CrossRef](#)]
111. Page, A.; Turner, J.; Mohsenin, T.; Oates, T. Comparing raw data and feature extraction for seizure detection with deep learning methods. In Proceedings of the Twenty-Seventh International Flairs Conference, Pensacola Beach, FL, USA, 21–23 May 2014.
112. Jain, A.; Zongker, D. Feature selection: Evaluation, application, and small sample performance. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 153–158. [[CrossRef](#)]
113. Ahmed, H.O.A.; Nandi, A.K. Vibration Image Representations for Fault Diagnosis of Rotating Machines: A Review. *Machines* **2022**, *10*, 1113. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.