Heliyon 8 (2022) e12136

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A novel approach to automatic sleep stage classification using forehead electrophysiological signals

Hengyan Guo [a,†], Yang Di [a,†], Xingwei An [a,c,*], Zhongpeng Wang [b,c,*], Dong Ming [a,b,c]

[a] Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, China
[b] Department of Biomedical Engineering, College of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin, China
[c] TianJin Center for Brain Science, Tianjin, China

ARTICLE INFO

Keywords:
Sleep stages
Light gradient boosting machine (LGB)
Forehead electrophysiological signal
Electrooculogram (EOG)
Discrete wavelet transform (DWT)

ABSTRACT

Background: Sleep stage scoring is very important for the effective diagnosis and intervention of sleep disorders. However, the current automatic sleep staging methods generally have the problems of poor model generalization ability and non-portable acquisition equipment.

Method: In this paper, we propose a novel automatic sleep scoring system based on forehead electrophysiological signals that is more effective and convenient than other systems. We extract 3 channel signals from the forehead, named forehead electroencephalogram 1 (Fh1), forehead electroencephalogram 2 (Fh2), and forehead center electrooculogram (Fhz). Spectral features, statistical features, and entropy features are extracted using the discrete wavelet transform (DWT) method. Light gradient boosting machine (LGB), random forest (RF), and support vector machine (SVM) are employed to classify four stages: awake, light sleep (LS), deep sleep (DS), and rapid eye movement (REM).

Result: The performance of the proposed method is validated using databases of the Sleep-EDFX and our Own-data, which include polysomnograms (PSGs) and forehead signals of 28 subjects. The overall classification accuracy of using the combination of Fh1, Fh2, and Fhz can reach up to 90.25% accuracy with a kappa coefficient of 0.857.

Conclusions: The proposed method could provide state-of-art multichannel sleep stage scoring performance with higher portability. This will facilitate the application of long-term monitoring of sleep quality in the future.

1. Introduction

Sleep plays a critical role in a person's health as a fundamental function of the brain. Sleep-related disorders such as insomnia, lethargy, and apnea syndrome may lead to high blood pressure, stroke, myocardial infarction, and arrhythmia (Zarei and Asl, 2019). The sleep scoring is defined as a gold standard for analyzing human sleep, and it can be used as a diagnostic tool in the detection of various sleep disorders. Human sleep, according to the Rechtschaffen and Kales (R&K) scoring criteria, is divided into stages N1 to N4 (from light to deep) and stage REM (rapid eye movement) and later developed by the American Academy of Sleep Medicine (AASM) (Berry et al., 2012), which combines N3 and N4 into a single class as a deep sleep stage. Manual sleep stage scoring is often conducted by sleep experts by visually inspecting patients' polysomnography signals collected in sleep labs. Overnight PSG signals include electroencephalogram (EEG), electrooculography (EOG), electromyogram (EMG), electrocardiogram (ECG), airflow, thoracic and abdominal movements, and oximetry. Manual classification of sleep stages is time-consuming and prone to human error. Some research reported that experts' agreement on sleep stage classification was less than 90% (Mora et al., 2010). Because the PSG recording includes multiple signal channels, the amount of data recorded overnight is large, and the adhesive electrodes and wires attached to the head will also affect the sleep efficiency of the subjects to a certain extent. Therefore, a variety of methods have been proposed for automatic sleep stage classification (ASSC). The ASSC system would decrease the workload for clinicians, increase the accuracy of sleep stage classification and improve the diagnosis and treatment of sleep disorders.

Many methods have been proposed for the automatic classification of sleep stages. Machine learning algorithms along with different signal

* Corresponding authors.
E-mail addresses: Anxingwei@tju.edu.cn (X. An), tunerl_wzp1@tju.edu.cn (Z. Wang).
† Hengyan Guo and Yang Di are the co-first authors.

https://doi.org/10.1016/j.heliyon.2022.e12136
Received 17 February 2022; Received in revised form 29 June 2022; Accepted 28 November 2022

processing techniques are adopted to obtain useful information from physiological signals. Such methods are divided into two categories, i.e., multichannel and single-channel processing. In the former approach, the combination of various physiological signals, such as multichannel EEG signals, electromyograms (EMGs), and electrooculograms (EOGs), is utilized to extract informative features. Phan et al. combined EEG and EOG channels to classify sleep stages and obtained an overall accuracy of 82,30% using the multitask 1-max CNN method (Huy Phan et al., 2013). Huang et al. extracted spectral features from two foreheads (FP1 and FP2) of EEG signals by using short-term fast Fourier transform (Huang et al., 2014) The relevant vector machine classification algorithm was used to classify sleep stages using these features. Ozal et al (Yildirim et al., 2019) developed a flexible deep learning model to classify sleep stages using electroencephalogram (EEG) and electrooculogram (EOG) signals. Several studies have shown that the use of multichannel signals leads to higher performance (Khalighi et al., 2013). Pattern recognition may benefit from multimodal physiological data that provide more information. However, the equipment needed for multiple EEG channels limits the subject's movement, which limits the portability and feasibility of the wearable sleep monitor device. Furthermore, excessive wire connections during the recording process might cause sleep disturbance.

Therefore, the limitations of multichannel have escalated the demand for automatic sleep scoring based on single channel physiological signal. Among multiple physiological signals, EEG signals are often used for the analysis of sleep staging, regardless of whether the scoring is manual or automatic. Researchers employ time, frequency, time-frequency domain-based transformations, and nonlinear feature extraction methods in the feature extraction stage for EEG signals (Ronzhina et al., 2012). Hassan and Bhuiyan developed a sleep classification system, which decomposed EEG signals using the ensemble empirical mode decomposition method and the RUSBoost classifier with an average accuracy of 88.1% (Hassan and Bhuiyan, 2017). Hsu et al. proposed a system to classify sleep stages based on EEG signal energy features and a recurrent neural classifier, resulting in 87.2% accuracy (Hsu et al., 2013). Relevant research on automatic sleep staging based on single-channel EOG has also been presented. Rahman et al. used discrete wavelet transform (DWT) for feature extraction on EOG signals, claiming EOG signals were superior to EEG signals for the classification of sleep stages (Rahman et al., 2018).

Although the single-channel EEG has been relatively successful in sleep scoring, the electrodes need to be placed above the hairline and expert assistance is necessary for system setup and data collection. In addition, it is difficult to conduct ambulatory sleep stage studies with EEG- or EOG-based methods (Rahman et al., 2018). In this regard, the forehead is a promising placement to record the physiological signal, which is more convenient for practical applications in sleep soring. Related research has explored a more practical and portable position of electrooculogram electrodes. Cai et al. (Hao-Yu Cai et al., 2011) first proposed the concept of forehead EOG in their research. EOG features are extracted from the forehead EOG signals using both independent component analysis and support vector machines. The signals separated by blind source separation are very similar to traditional EOG signals. On this basis, Zhang et al. proposed a novel electrode placement on the forehead and proposed a corresponding algorithm to extract HEO and VEO from forehead EOG to detect driving fatigue (Zhang et al., 2015). In addition, Huo et al. fused EEG and forehead EOG to detect drivers' fatigue level by using a discriminative graph regularized extreme learning machine, and the results showed that fusing forehead EOG data and EEG data can effectively improve the performance of driving fatigue detection (Xue-Qin Huo et al., 2016).

In this study, we propose an automatic sleep stage classification system based on forehead electrophysiological signals. Basic PSG and three additional electrophysiological signals located on the forehead are collected for sleep stage scoring. We compare the performance of the forehead electrophysiological signal and the public dataset sleep-EDF using single-channel EEG, single-channel EOG, and a combination of EEG and EOG signals. Different dimensional features from the time domain, frequency domain, and time-frequency domain were extracted. The discriminating ability of the features was studied using one-way analysis of variance (ANOVA). The proposed system further incorporates the light gradient boosting machine (LGB) as the basic classifier. For comparison, classification of the features is also carried out using a support vector machine (SVM) and random forest (RF). The results of sleep staging indicated that the overall accuracy of the proposed method of using forehead electrophysiological signals for sleep staging is either comparable or superior to the others. Thus, we propose a more portable and efficient sleep monitoring electrode configuration, which proves the feasibility of using forehead electrophysiological signals for sleep monitoring and provides a more efficient and convenient sleep and pre-diagnostic system.

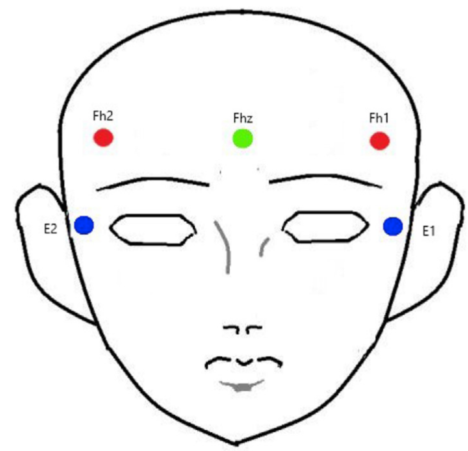

Figure 1. Electrode placement of Fhz, Fh-EOG (Fh1-Fh2) and traditional EOG (E1-E2).

The rest of the article is organized as follows. Section 2 contains the details of the experimental data and the methodology of the present work. The experimental results are described in Section 3. Section 4 presents the discussion of experiment. Finally, the conclusion is given in section 5.

2. Materials and methods

2.1. Sleep datasets

2.1.1. Acquisition of sleep data for forehead electrophysiological

In this study, we also conducted a sleep data acquisition experiment. Twenty-eight subjects (mean age 23 $\pm$ 2 years) participated in this study. None of the participants reported having a history of psychological disorders. All subjects were informed of the experimental matters needing attention and signed the consent form before the experiment. The participants were subjected to a sleep collection experiment at noon (11:30 am -1:30 pm) and night (11:00 pm–7:00 am). The first-night effect (FNE) affects the accuracy of polysomnography findings. The purpose of sleep collection experiments scheduled at noon is to enable the subjects to adapt to the sleep environment and reduce the interference with the subjects' normal sleep. Two subjects were removed due to signal quality

Table 1. Detailed information about the sleep database records used in this study.

Database	Sleep Stages				Total Samples
	Wake	Light sleep	Deep sleep	REM	
Sleep-EDF	4214 (16.11%)	9148 (34.96%)	7998 (30.57%)	4805 (18.36%)	26165
Own-data	4742 (17.81%)	11261 (42.28%)	5877 (22.07%)	4754 (17.85%)	26634

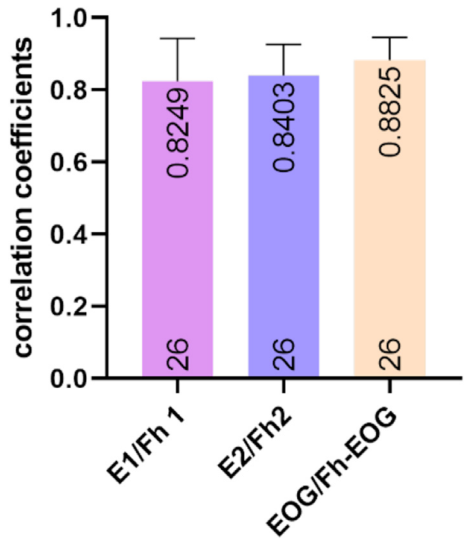

Figure 2. The correlation coefficients of E1 and Fh1, E2 and Fh2, traditional EOG and forehead EOG on different subjects.

issues. All experiment procedures were approved by the Ethics Review Committee of Tianjin University.

We recorded the sleep PSG signals using a Gral 4k PSG amplifier with a sampling rate of 128 Hz. The complete PSG recording contains twelve channels EEG, two channels EOG, chin EMG, leg EMG, airflow signals, lead-II ECG, oximetry, nasal pressure, snoring sounding, and body position. Three new forehead electrophysiological channels were also recorded using the same PSG system for further channel optimization. The locations of the three forehead channels are shown in Figure 1. Fh1 and Fh2 are regarded as alternative electrodes for traditional EOG that contain eye movement information, which has been proven in a previous study (Zhang et al., 2015). The third forehead channel, which is referred to as 'Fhz', is located in the middle of the forehead and parallel to the positions of Fh1 and Fh2. The Fhz channel was recorded to provide forehead electrophysiological information. The contact impedance between each electrode and scalp was controlled to be lower than 5 kΩ.

2.1.2. Sleep-EDF

Sleep-EDF (expanded) as a common public sleep dataset has been used in this study. The dataset contains 61 polysomnograms (PSGs) records, which consist of 2 EEG (Fpz-Cz and Pz-Oz), 1 horizontal EOG, 1 submental chin EMG, and event annotations. EEG and EOG signals were obtained with a sampling rate of 100 Hz. The recordings were divided into 30s-fragments and manually classified into six sleep stages: W, S1, S2, S3, S4, REM, M (Movement time), and ? (not scored) by sleep experts based on the R&K manual. In this work, the Pz-Oz channel and horizontal EOG channel are used. We selected sleep records and labeled recordings from 26 healthy subjects (SC, mean age 25 ± 3.5 years). In addition, S1 and S2 are combined into light sleep (LS) and S3 and S4 are combined into deep sleep (DS) in this system.

2.2. Data processing and algorithm

2.2.1. Feature extraction

As AASM recommendations, the data were segmented into epochs of 30 s after being filtered with 0.5–50 Hz. All epochs were manually classified into four stages (W, LS, DS, and REM). Table 1 shows the detailed information of the sleep records used.

We decompose the forehead electrical signal using Daubechies-4 (db4) as the mother wavelet in the DWT domain. Detailed Coefficients D = {D1, D2, D3, D4, D5, D6} and Approximate Coefficients A = {A1, A2, A3, A4, A5, A6} were extracted. Statistical features and differential entropy are then extracted from these coefficients. In addition, according to

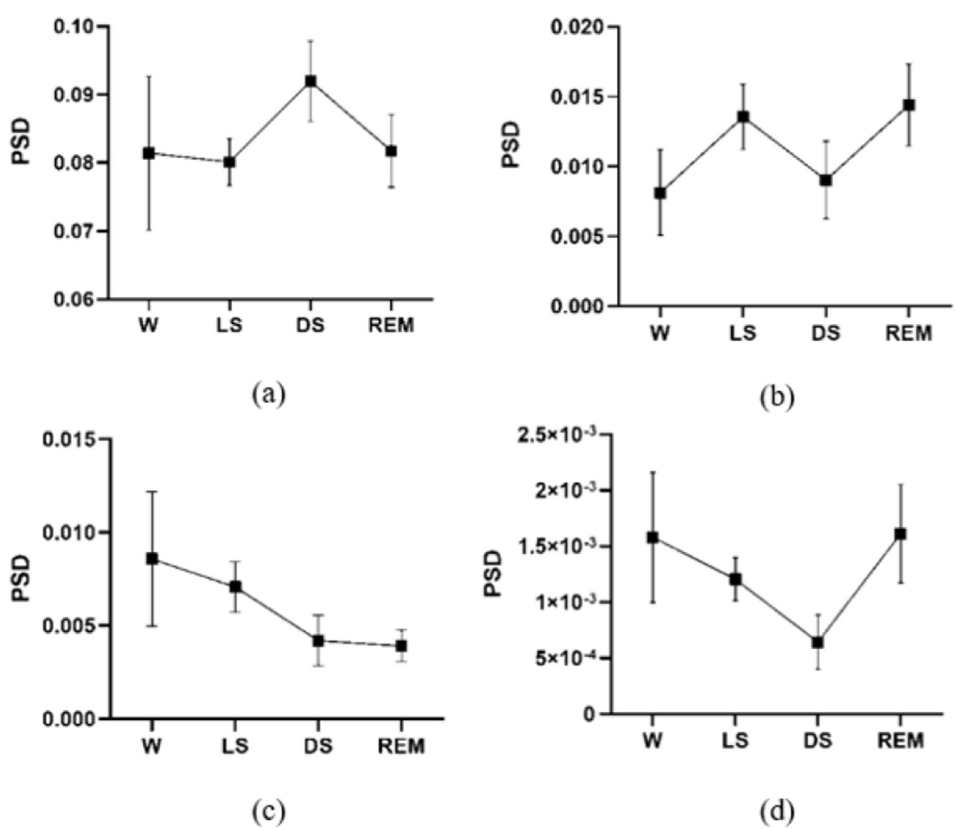

Figure 3. The PSD of the forehead electrical signal Fhz in each sleep stage. (a) delta rhythm, (b) theta rhythm, (c) alpha rhythm, (d) beta rhythm.

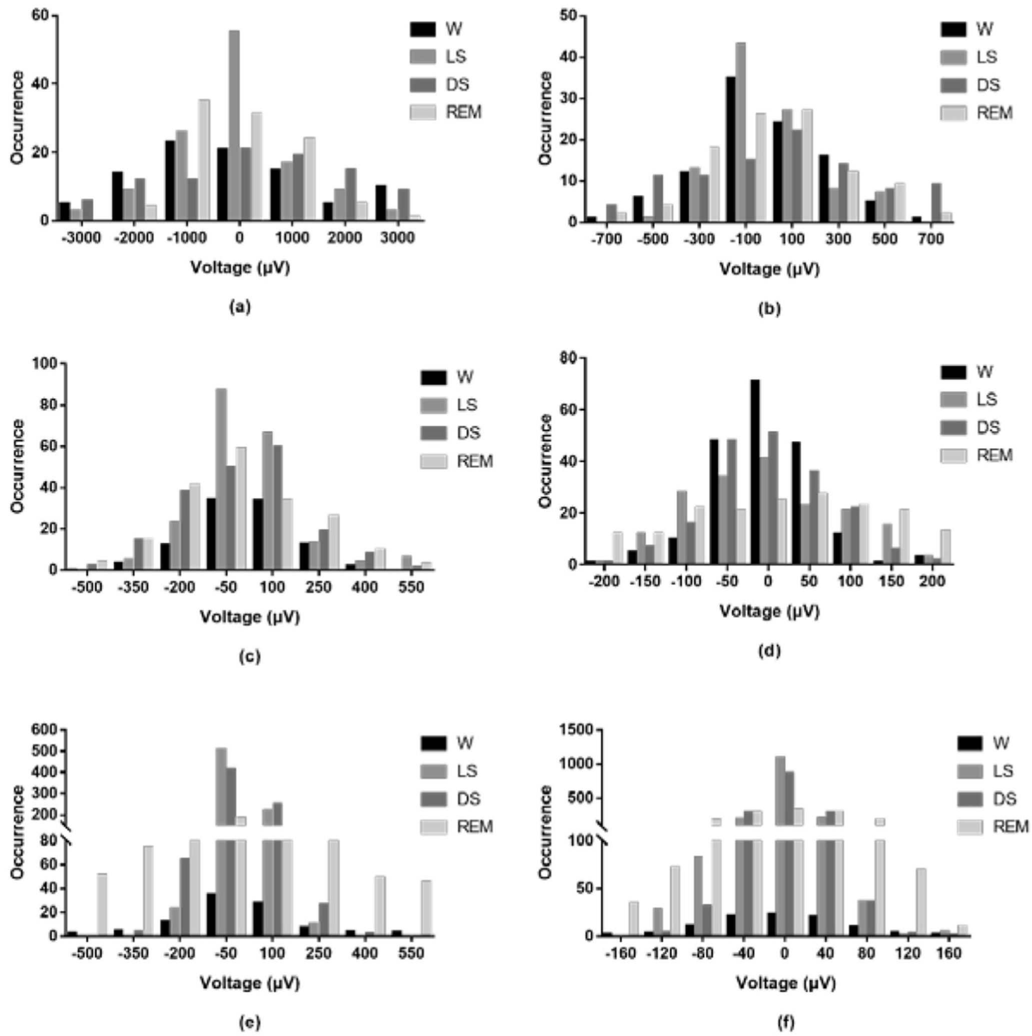

Figure 4. Frequency distribution histogram of wavelet subband coefficients of forehead electrical signals in each sleep stage. (a) A_5 subband, (b) D_5 subband, (c) D_4 subband, (d) D_3 subband, (e) D_2 subband, (f) D_1 subband.

the physiological characteristics of the EEG signal, we also extracted the power spectral density of the forehead electrophysiological signal based on four frequency bands.

PSD. Nonparametric methods for spectral estimation are the most general form of spectral analysis. It has been by far the most frequently used technique in sleep EEG analysis, largely because they are simple to implement and interpret (Motamedi-Fakhr et al., 2014). In this work, 'PSD' estimated the power spectral densities using the Welch method by fast Fourier transform (FFT) within the entire 30-s signals. The signal is decomposed into four frequency bands: delta (0.5–4 Hz), theta (4.5–8 Hz), alpha (8–12 Hz), beta (12–30 Hz). A Hamming window was employed to reduce the side lobe effect. The FFT length was set to 0.256 s, resulting in a frequency resolution of approximately 4 Hz. Power spectral densities were smoothed from segments with 50% overlap.

Statistical feature. From each frequency band, five statistics were extracted: 1) mean of coefficients' absolute values; 2) average power of the coefficients; 3) standard deviation of the coefficients; 4) ratio of absolute mean values of adjacent subbands; 5) kurtosis of each subband.

Entropy. Chaos is characterized by disorder, which is quantified by the entropy of a signal. The differential entropy measures the degree of uncertainty in the continuous probability distribution and is a continuum of Shannon's entropy (Hsu et al., 2013). In sleep recordings, entropy can directly reflect the differences in the complexity of the EEG in each sleep stage. The differential entropy of each subband is equivalent to the logarithm of this subband's energy spectrum, which also provides a physiological interpretation of the energy spectrum. Thus, in this study, the differential entropy of each subband is selected as the entropy feature.

2.2.2. Classifiers

The proposed system uses the light gradient boosting machine (LGB) as its basic classifier. As a comparison, support vector machine (SVM) and random forest (RF) are applied to analyze and classify the features. LGB is an ensemble method of boosting type, which is a gradient boosting framework that uses a tree-based learning algorithm. Algorithms such as XGBoost and RUSboost in the boosting collection framework have achieved good results in sleep scoring. Considering that a large number of comparative experiments on public data sets show that LGB, proposed by Ke et al., can outperform the existing boosting framework in terms of efficiency and accuracy (Ke et al., 2017). Thus, LGB is employed to perform the sleep stage classification in this study. Scikit-learn's GridSearchcv function is used to perform an evaluative search and optimize the parameters, including num_leaves, max_depth, bagging_fraction, feature_fraction, etc.

2.2.3. Performance evaluation methods

In this study, the sleep PSG data of 26 subjects are collected. Leave-one-subject-out cross validation (LOSO) is used to evaluate the classification performance of the proposed system. The reason is to consider that when dealing with models aiming to provide clinical predictions (e.g. sleep stage), for cross-validation to be meaningful the process of splitting

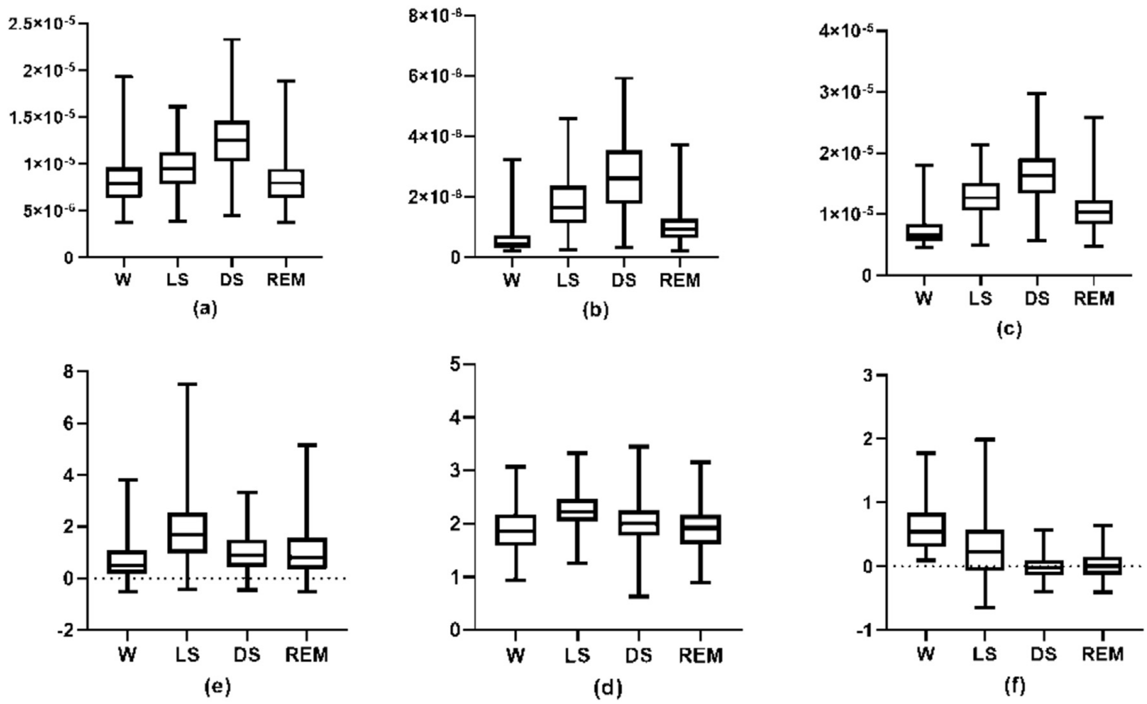

Figure 5. Box plot of different features for the D3 subband of DWT: (a) the kurtosis, (b) mean of coefficients' absolute values, (c) average power of the coefficients, (d) standard deviation of the coefficients, (e) ratio of absolute mean values of adjacent subbands, (f) differential entropy.

should be conducted subject-wise (not sample wise). Using CV sample may create data leakage and limits the interpretation of the results. The LOSO method is implemented by taking the data from one subject as the testing set and the data from other remaining subjects as the training set. Repeat the above steps until all subjects have been included in the test set, and finally average the validation set evaluation metrics.

We adopted the precision, accuracy, sensitivity, and Cohen's kappa coefficient to assess the classification performance. Precision refers to the repeatability or reproducibility of measurement results. Accuracy indicates the proximity of measurement results to the true value. Sensitivity is a statistical measure of the performance of a binary classification test, which reflects the ability to identify positive results. Cohen's kappa is a statistical measure of inter-rater agreement or inter-annotator agreement for qualitative (categorical) items (Hassan and Mohammed, 2016). It is generally considered a more robust measure than a simple percent agreement calculation since the kappa coefficient takes into account the agreement occurring by chance.

3. Results

3.1. Forehead EOG extraction and correlation analysis

In this study, signal of Fh1-Fh2 are regarded as alternative EOG, which are referred to as Fh-EOG. We conducted the correlation coefficient between each commonly used EOG channel (E1, E2 and E1-E2) and the Fh-EOG channels of each subject, as shown in Figure 2. The results show that there is a high correlation between the traditional EOG and the forehead EOG, reaching 0.8825. The correlation coefficients between electrode channels E1 and Fh1, E2 and Fh2 are 0.824 and 0.840, respectively.

3.2. Feature analysis

For the forehead electrophysiological signals collected by Fhz electrodes, frequency domain features based on power spectral density are extracted, as shown in Figure 3. It can be seen that the power of delta rhythm increases with deeper sleep (a). Theta rhythm is observed in deep relaxation and stress-free subconscious states. It is found that theta rhythm shows higher power during the three stages of sleep (light sleep, deep sleep and REM) (b). Alpha rhythm is the main manifestation of the cerebral cortex when it is in a awake and relaxed state. It can be seen that in the awake stage, the alpha rhythm shows the highest power. As the frequency value of the spindle signal in the light sleep stage overlaps with the alpha band, the light sleep stage also displays an alpha rhythm as compared to deep sleep and REM (c). Beta rhythm typically shows electrophysiological activity when the cerebral cortex is in a state of excitement. It is apparent that beta rhythm only shows high power in the awake stage, and that power decreases further as a result of deeper sleep (d).

Frequency distribution histogram of wavelet subband coefficients of forehead electrical signals is shown in Figure 4 (a-f). The signals showed more significant differences in the low frequency subband. The distribution of high frequency subbands D1 and D2 is more concentrated in light sleep, deep sleep and REM. The shapes show significant differences between sleep stages, indicating that the statistics are different for each sleep stage. In this regard, DWT subband statistical features could provide good discrimination functions.

After extracting the features from different DWT subbands, the statistical significance of the discriminating capabilities of the features is evaluated by employing one-way analysis of variance (ANOVA). All two databases are tested with a 99% level of confidence. Each of the features has a very low p-value, indicating the statistical significance of their ability to discriminate between sleep stages. Visualization of the features using box plots confirmed this. Figure 5 (a-f) shows the different features extracted from D3 subband, which shows that the DWT subband statistical features are highly discriminatory.

Table 2 show the ANOVA result of subband coefficient features of forehead signal in each sleep stage. Feature 4 (ratio of absolute means in adjacent subbands), feature 5 (skewness of each subband), and feature 6 (kurtosis of each subband) all have significant differences among the four sleep stages. Feature 1 (average of absolute values of coefficients in each subband) and feature 2 (average power of coefficients in each subband) were not significantly different during light sleep and REM stage. Feature 3 was significantly different between wakefulness and sleep stage, but no significant difference between the other three sleep stages.

Table 2. ANOVA results of subband coefficient features of forehead signal in each sleep stage.

Feature	Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Adjusted P Value
1	W vs. LS	-3.124E-07	-4.223e-007 to -2.025e-007	Yes	<0.0001
	W vs. DS	-1.627E-06	-1.739e-006 to -1.516e-006	Yes	<0.0001
	W vs. R	-2.424E-07	-3.725e-007 to -1.123e-007	Yes	<0.0001
	LS vs. DS	-1.315E-06	-1.390e-006 to -1.239e-006	Yes	<0.0001
	LS vs. R	7E-08	-3.063e-008 to 1.706e-007	No	0.2794
	DS vs. R	1.385E-06	1.282e-006 to 1.488e-006	Yes	<0.0001
2	W vs. LS	-7.153E-09	-7.753e-009 to -6.554e-009	Yes	<0.0001
	W vs. DS	-1.887E-08	-1.948e-008 to -1.826e-008	Yes	<0.0001
	W vs. R	-7.14E-09	-7.846e-009 to -6.434e-009	Yes	<0.0001
	LS vs. DS	-1.172E-08	-1.213e-008 to -1.131e-008	Yes	<0.0001
	LS vs. R	1.366E-11	-5.294e-010 to 5.567e-010	No	>0.9999
	DS vs. R	1.173E-08	1.118e-008 to 1.228e-008	Yes	<0.0001
3	W vs. LS	-4.787E-06	-4.968e-006 to -4.607e-006	Yes	<0.0001
	W vs. DS	-4.732E-06	-4.944e-006 to -4.520e-006	Yes	<0.0001
	W vs. R	-4.732E-06	-4.944e-006 to -4.520e-006	Yes	<0.0001
	LS vs. DS	5.569E-08	-1.074e-007 to 2.187e-007	No	0.8165
	LS vs. R	5.569E-08	-1.074e-007 to 2.187e-007	No	0.8165
	DS vs. R	0	-1.977e-007 to 1.977e-007	No	>0.9999
4	W vs. LS	-0.1933	-0.2136 to -0.1731	Yes	<0.0001
	W vs. DS	-0.7269	-0.7474 to -0.7063	Yes	<0.0001
	W vs. R	0.04632	0.02241 to 0.07024	Yes	<0.0001
	LS vs. DS	-0.5335	-0.5474 to -0.5196	Yes	<0.0001
	LS vs. R	0.2397	0.2212 to 0.2581	Yes	<0.0001
	DS vs. R	0.7732	0.7543 to 0.7920	Yes	<0.0001
5	W vs. LS	2.194	2.125 to 2.264	Yes	<0.0001
	W vs. DS	4.286	4.215 to 4.357	Yes	<0.0001
	W vs. R	1.608	1.526 to 1.690	Yes	<0.0001
	LS vs. DS	2.091	2.044 to 2.139	Yes	<0.0001
	LS vs. R	-0.5868	-0.6502 to -0.5234	Yes	<0.0001
	DS vs. R	-2.678	-2.743 to -2.614	Yes	<0.0001
6	W vs. LS	-0.9722	-1.083 to -0.8610	Yes	<0.0001
	W vs. DS	-0.4503	-0.5634 to -0.3372	Yes	<0.0001
	W vs. R	0.1342	-0.09899 to 0.1636	Yes	<0.0001
	LS vs. DS	0.5219	0.4456 to 0.5982	Yes	<0.0001
	LS vs. R	1.005	0.9032 to 1.106	Yes	<0.0001
	DS vs. R	0.4826	0.3793 to 0.5860	Yes	<0.0001

Table 3. Performance of Fpz-Cz and Pz-Oz channels on the Sleep-EDF Dataset.

Channel	LGB	RF	SVM
Fpz-Cz	87.71 ± 0.7	86.52 ± 0.4	75.61 ± 0.1
Pz-Oz	88.13 ± 0.3	84.16 ± 0.3	83.80 ± 0.5

3.3. Sleep stage results

The experimental results were based on the sleep-EDF dataset and acquired sleep dataset. We reported the results of sleep staging based on three different channel combinations. The proposed method for feature extraction and classification was applied to each experiment and the

Table 4. F1-score, average accuracy, and kappa coefficient of the proposed method to classify each sleep stage for two databases using a single EEG signal (highest accuracy values are shown in bold).

Dataset	Algorithm	F1-score of Each Class (%)				Average Accuracy （%）	Kappa coefficient
		W	LS	DS	REM		
Own-data (Fhz)	LGB	90.1 ±1.3	87.9 ±0.7	91.3 ±1.6	73.2 ±0.2	86.61 ±1.51	0.804
	RF	87.7 ±0.8	86.0 ±0.2	91.0 ±1.0	75.3 ±0.7	86.28 ±1.48	0.802
	SVM	80.5 ±0.1	82.7 ±0.8	89.3 ±0.7	71.1 ±0.2	82.25 ±1.10	0.746
Sleep-EDF (Pz-Oz)	LGB	89.3 ±0.7	90.3 ±1.0	93.4 ±1.7	80.6 ±0.5	88.65 ±2.24	0.822
	RF	86.1 ±0.3	87.8 ±0.9	83.3 ±0.2	74.4 ±0.1	84.16 ±1.05	0.749
	SVM	84.0 ±0.7	86.3 ±1.4	90.2 ±0.3	71.0 ±0.5	83.80 ±1.92	0.745

Table 5. The performance values obtained for the 4-class data using the sleep-EDF dataset with a single EEG channel (Pz-Oz).

Classes	Sleep Stages	Sensitivity	Specificity	Precision
4-class	Wake	0.93	0.99	0.93
	Light Sleep	0.91	0.92	0.88
	Deep Sleep	0.93	0.97	0.94
	REM	0.77	0.98	0.85

Table 7. The performance values obtained for the 4-class data using the Own-data dataset with a single Fh-EOG channel.

Classes	Sleep Stages	Sensitivity	Specificity	Precision
4-class	Wake	0.89	0.98	0.87
	Light Sleep	0.87	0.90	0.85
	Deep Sleep	0.91	0.96	0.91
	REM	0.72	0.97	0.78

performance of the three classifiers was compared. The experiments were performed in our study using a Python 3.8.1 environment on a Windows x64 platform using an Intel(R) Core(TM) i5-9400 CPU 2.90 GHz and 12 GB of RAM.

We also investigated the algorithm performance of the two EEG channels in the sleep-EDF database, namely, Fpz-Cz and Pz-Oz, to select a channel for the next exploration. In the four classification tasks, it is evident from Table 3 that the results obtained from the EEG signals of Pz-Oz outperform those of the Fpz-Cz channel in all the classifiers. This finding is also consistent with the results reported in prior works in the literature (Alickovic and Subasi, 2018; Lee et al., 2014), wherein EEG signals from the Pz-Oz channel yielded higher accuracy values than those of the Fpz-Cz channel.

3.3.1. Single channel results

Table 4 shows the results of the performance of the Pz-Oz channel in the sleep EDF database and the Fhz channel in Own-data. Among the three classifiers, LGB achieves higher classification accuracy. For Fhz, the overall accuracy and kappa coefficient reached 86.61% and 0.804, respectively. For Pz-Oz, the average overall accuracy was 88.65%, with a kappa coefficient of 0.7527. The F1-scores of the W, LS, SWS, and REM stages were 89.3%, 90.3%, 93.4%, and 80.6%, respectively. The classification of each sleep stage was better than that of the Fhz channel in the Own data. Table 5 shows the sensitivity, specificity, and accuracy of various sleep categories and Pz-Oz channels classified based on the LGB algorithm. In terms of evaluation performance, the highest values were obtained in the deep sleep stage, where the sensitivity, specificity, and precision values were 93%, 97%, and 94%, respectively. The REM stages produced the lowest performance values, with an accuracy value of 80.6%.

3.3.2. EOG results

For single-channel horizontal EOG, compared with the sleep EDF database, Fh-EOG has a higher average accuracy and kappa coefficient, which are 87.37% and 0.770, respectively. We think this may be because forehead EOG contains other information. There are related works in the literature that show that EOG signals have more valuable information, which can help to distinguish between rapid eye movement and non-rapid eye movement. It can be seen from Table 6 that the classification accuracy of the REM stage is relatively improved compared to the forehead Fhz channel and the EEG Pz-Oz channel. The highest F1 score value achieved 83.7%. In each stage of the two databases, LGB consistently produces a satisfactory performance. For other evaluable indexes in Table 7, similar to the use of single-channel EEG, the sensitivity and precision of the REM phase are also significantly lower than those of other sleep stages. According to the confusion matrix data, some REM stages are misclassified as LS stages.

3.3.3. Multichannel results

In this part, multimodal data for sleep staging was used. For our dataset, we combine the Fhz and Fh-EOG data. For the sleep EDF dataset, we select the data of the Pz-Oz channel and EOG channel for sleep staging. The former uses the LGB algorithm to achieve higher classification accuracy and kappa coefficient, which are 88.15% and 0.8265, respectively, and the accuracies of W, LS DS, and REM were 89.3, 87.2, 92.5, and 82.7%. The latter also uses the LGB algorithm to achieve the highest classification accuracy of 85.04%, with a kappa coefficient of 0.7863. The detailed numerical values are given in Table 8. Notably, the accuracy for REM was relatively low. The classification effect of the REM stage in the sleep EDF dataset is significantly better than that of Own-data.

4. Discussion

4.1. Forehead electrophysiological signal in sleep stage scoring

In this work, we proposed a new method for sleep stage scoring based on forehead electrophysiological signals. The system provides researchers with a more portable and easy-to-operate sleep monitoring method and makes long-term sleep monitoring possible. Most previous automatic sleep staging systems are based on multichannel or multiple physiological signals, but the subject's sleep position often interferes with the wire, thereby degrading the signal quality (Liang et al., 2012). Therefore, in terms of current research trends, the automatic sleep staging system based on single-channel physiological signals has received more attention from the academic community. Although single-channel

Table 6. F1-score, average accuracy, and kappa coefficient of the proposed method to classify each sleep stage for two databases using a single EOG channel (highest accuracy values are shown in bold).

Dataset	Algorithm	F1-score of Each Class (%)				Average Accuracy (%)	Kappa coefficient
		W	LS	DS	REM		
Own-data (Fh-EOG)	LGB	87.7 ±0.7	85.2 ±0.4	89.5 ±1.1	81.9 ±0.9	87.37 ±1.70	0.770
	RF	85.6 ±0.6	84.7 ±1.4	89.2 ±0.7	66.4 ±0.2	83.53 ±1.33	0.756
	SVM	70.1 ±0.3	76.4 ±0.2	85.2 ±0.9	54.5 ±0.5	75.02 ±1.67	0.643
Sleep-EDF (EOG)	LGB	82.3 ±0.9	86.1 ±0.3	81.5 ±0.7	83.7 ±1.4	85.81 ±1.65	0.722
	RF	77.0 ±1.0	84.8 ±0.2	77.3 ±0.7	70.0 ±0.1	82.18 ±1.18	0.665
	SVM	81.7 ±1.5	84.2 ±0.7	82.9 ±0.9	69.3 ±1.3	80.25 ±2.01	0.685

Table 8. F1-score, average accuracy, and kappa coefficient of the proposed method to classify each sleep stage for two databases using multimodal signals (highest accuracy values are shown in bold).

Dataset	Algorithm	F1-score of Each Class (%)				Average Accuracy (%)	Kappa coefficient
		W	LS	DS	REM		
Own-data (Fhz + Fh-EOG)	LGB	93.4 ±0.7	91.5 ±0.4	94.1 ±1.7	82.7 ±0.1	90.25 ±2.30	0.857
	RF	91.0 ±0.1	88.1 ±0.7	92.4 ±0.3	78.5 ±1.1	88.64 ±1.45	0.833
	SVM	82.7 ±1.0	83.6 ±0.3	88.8 ±0.5	70.3 ±0.2	82.47 ±1.56	0.746
Sleep-EDF (Pz-Oz + EOG)	LGB	91.5 ±0.2	90.5 ±0.7	91.7 ±1.3	86.9 ±0.9	90.06 ±1.76	0.848
	RF	82.3 ±0.4	86.1 ±0.4	81.5 ±0.6	83.7 ±0.7	85.81 ±1.17	0.722
	SVM	90.4 ±0.7	89.0 ±0.5	92.6 ±1.1	80.5 ±0.1	88.23 ±1.54	0.819

Table 9. Summary of results obtained for various combinations of data used and sleep stages.

Dataset	PSG Signals	F1-score of Each Class (%)				Average Accuracy (%)
		W	LS	DS	REM	
Own-data	Single-channel Fhz	90.1 ±0.7	87.9 ±0.2	91.3 ±0.8	73.2 ±1.1	86.61 ± 1.7
	Single-channel Fh-EOG	87.7 ±0.6	85.2 ±0.7	89.5 ±0.6	81.9 ±0.1	87.37 ± 1.2
	Fhz + Fh-EOG	93.4 ±1.3	91.5 ±0.6	94.1 ±1.5	82.7 ±0.2	90.25 ± 2.0
Sleep-EDF	Single-channel EEG	89.3 ±0.8	90.3 ±1.1	93.4 ±1.0	80.6 ±0.5	88.65 ± 1.3
	Single-channel EOG	82.3 ±0.9	86.1 ±0.3	81.5 ±0.7	83.7 ±1.4	85.81 ± 1.6
	EEG + EOG	91.5 ±1.3	90.5 ±0.7	91.7 ±1.4	86.9 ±0.9	90.06 ± 2.1

EEG or EOG is effective in sleep scoring, electrodes must be placed above the hairline and often affected by subject movement. The ambulatory study of sleep stage monitoring based on these methods is difficult to achieve. The method we proposed is different from the traditional EOG that collects signals from both sides of the eye, and the signal is collected on the subjects' forehead. It is also different from traditional EEG, which requires electrode recording above the hairline. This method will not affect the subjects' normal activities but is also conducive to the integration and practical use of wearable devices. Compared with previous sleep monitoring studies, the electrodes used in this study are placed on hairless skin and fixed with solid conductive paste and tape, which can be considered a real forehead electrical signal. It is worth noting that under the same sample size and classification model, the classification performance based on forehead electrophysiological signals is better than traditional physiological signals, which illustrates the potential value of the former in the field of sleep staging. It provides a more portable and comfortable collection method and better classification accuracy than traditional monitoring electrodes.

4.2. Single channel vs Multichannel

There are various studies on sleep stage scoring with physiological signals. Among these signals, EEG patterns show different characteristics during sleep stages. These features have been used for the development of numerous sleep-stage classification systems. On the other hand, EOG results from the continuous measurement of the corneal-

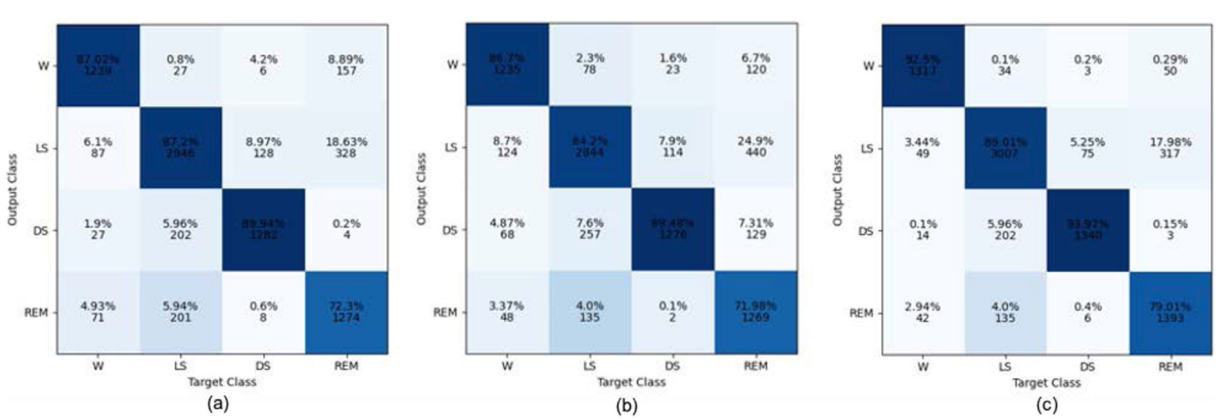

Figure 6. Confusion matrix obtained for different channels of the Own-data: (a) Fhz, (b) Fh-EOG, (c) Fhz + Fh-EOG.

Table 10. The accuracy obtained in the proposed method and with previous methods based on Sleep-EDF.

Work	Channel	Primary method	4-Class
Zhu et al. [22]	EEG	(VG-HVG) + SVM	89.30%
Ozal et al. [23]	Single-channel EEG + EOG	1D-CNN	92.33%
Michielli et al. [24]	EEG	mRMR + LSTM	90.8%
Pan et al. [21]	EEG + EOG + EMG + Resp + Temp	ICA + Relief + SVM	90.1%
Hassan et al. [7]	EEG	QTWT + RF	92.11%
Rahman et al. [9]	EOG	DWT + RUSBoost	91.11%
Hassan et al. [28]	EEG	CEEMDAN + Bagging Decision Tree	92.14%
Proposed	EEG	DWT + LGB	92.17%

retinal standing potential, which provides important information for REM stage detection. Hence, the classification accuracy of sleep scores can be significantly improved. In this study, the EEG and EOG data provided by the sleep-EDF dataset, the Fhz and Fh-EOG extracted from the collected forehead electrophysiological signals, and the multi-channel data obtained from the respective combination of these channels were used for sleep stage staging. The results showed that multimodal signals produced relatively high overall accuracy on both datasets. For sleep-EDF, reaching 90.06%, the kappa coefficient was 0.848. For our data, the accuracy was 90.25%, with a kappa coefficient of 0.857. This result offers researchers valuable information about the effectiveness of methods based on multimodal signals because the combination of multiple physiological signals provides redundant information (Faust et al., 2019). In this work, the experiment results show that LGB always maintains satisfactory performance, so Table 9 shows the experimental results based on the LGB method. The classification results of using multimodal channel combinations in the two data sets are better than using single-channel data in each stage. In addition, the proposed method of sleep staging using the Forehead electrophysiological system achieved better staging results than traditional datasets except for the REM stage. In the other two-channel combinations, both the Fhz channel and the Pz-Oz channel are better than the Fh-EOG and EOG. Multimodal signals have an overall accuracy of 3.26% higher than single-channel signals.

4.3. Performance of proposed system.

It is worth mentioning that, in this research, we combine S1 and S2 as the stage of light sleep. Identifying S1 stage exactly is an enormous challenge in sleep scoring research as S1 stage is intermediate, or transitional between awake and sleep that the number of S1 stage data is less than 10% in the whole set. Therefore, considering the imbalance of data and the main purpose of this work is to explore the feasibility of using forehead electrophysiological signals in sleep stage scoring, the sleep state is divided into four stages in this system. It should be noted that the forehead sleep electrophysiological signal channel proposed in this study did not achieve a significant improvement in the classification performance of the LS stage compared with the traditional EEG and EOG channels. The channel combination of Fhz and Fh-EOG achieves the best classification performance with a classification accuracy of 91.5%. Tables 4, 6, and 8 show that the accuracy of the proposed algorithm for REM is significantly lower than that of the other three stages. The results showed that both the Fh-EOG and EOG channels, which contain electrooculographic information, are significantly better than the forehead electrophysiological channel Fhz and the traditional EEG signal Pz-Oz in the classification performance of the REM stage. This reflects why this study uses channel combination to improve sleep staging performance. Considering that REM sleep occurs approximately 90 min after falling asleep and occurs at 90-minute intervals of 5–30 min (Faust et al., 2019), which results in a slight imbalance in the number of collected data samples. In addition, the cortex shows gamma waves of 40–60 Hz during REM as well as during waking (Horne, 2013). This similarity with awake causes it to be misclassified by both computerized sleep scoring systems. Compared to the awake records of subjects collected throughout the day in the public dataset, we only recorded wake stage signals of the subjects in the resting state, so it is relatively prone to be misclassified, as shown in Figure 6 (a-b). In addition, S1 is a phase transition phase between wakefulness and sleep, similar to awake neuronal oscillations from a neurophysiological perspective. Since we merged S1 into the LS, this may also be the reason that the REM is more misclassified as W and LS in the confusion matrix in Figure 6 (a-c). The detection of REM detection is crucial for the diagnosis of various sleep-related ailments, including REM behavior disorder (RBD) and narcolepsy, as well as an indicator of various neurological diseases, such as Parkinson's disease (Hassan and Bhuiyan, 2016). Table 8 shows that the proposed method-based forehead electrophysiological signal gives 82.7% accuracy for REM detection, whereas the method proposed in Pan et al. (Khalighi et al., 2013) and Zhu et al. (Zhu et al., 2014)give 78.10% and 76.20% Rem detection accuracy, respectively. Therefore, the REM detection accuracy of our automatic sleep stage scoring-based forehead electrophysiological signal is comparable to or better than that of existing works.

The class imbalance of sleep EEG data presents a significant challenge for automatic sleep scoring. This situation can be found in most studies using the sleep-EDF database because the amount of data in the awake period is too large, even more than half of the epoch. Therefore, without the treatment of category balance, the reported accuracy cannot be achieved in actual clinical monitoring. In our study, for each selected subject in the sleep-EDF database, we only selected the awake data of 30 min before and after sleep. While the data in the Own-data, the awake period and the rapid eye movement period accounted for relatively little, and there was no data imbalance. In addition, the method we proposed shows a relatively stable classification performance on both datasets, indicating that the algorithm has a certain degree of portability and robustness. The methods listed in Table 10 used the sleep-EDF dataset, and only the methods with the highest accuracy were presented. In this study, the proposed method achieved the highest classification accuracy of 90.25% and 0.857 kappa when using a combination of Fhz and Fh-EOG for 28 normal subjects According to the review of relevant literature, Sheng et al and Huang et al reported a study on automatic sleep staging using Fp1 and Fp2 as forehead signals (Sheng et al., 2014; Huang et al., 2014). Clinical tests of the former show that the proposed automatic sleep stage classification system can classify the four sleep stages with about 77% accuracy and 67% kappa for 12 normal subjects. The latter system correctly classified the five sleep stages with an average accuracy of 76.7% and 0.68 kappa for 10 normal subjects. In addition, regarding the sleep monitoring devices, such as Dreem, Sleep profile (Levendowski et al., 2017), etc. The former uses 5 EEG sensor to measure EEG, including 2 frontal sensors in F7 and F8 locations to measure frontal brain activity and 1 ground sensor on the frontal band on Fp2 location. The electrode layout is quite different from our proposed method. The latter uses a similar electrode layout. The mean overall interscorer agreements between the 5 technologists were 75.9%, and the mean kappa score was 0.70. Our proposed method outperforms the related studies mentioned above.

4.4. Limitations of the study and comments on prospects

The primary limitation of the study is that the sleep data used are from normal subjects. Sleep staging in patients with sleep disorders is considered to be further discussed on the basis of this work. In addition, it may be worthwhile to investigate further forehead signals performance for different classifiers, different features etc. In addition, deep learning holds great potential for transforming clinical studies of human sleep. Some intelligent classification processing algorithms, such as: Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Graph Convolutional Network (GCN) and Attention Mechanism (Attention) can also apply to the electrophysiological physiological signals proposed in this study. These deep learning-based approaches recently been reported to achieve comparable performance to sleep experts on the sleep-scoring task, at least in healthy individuals. However, it is important to note that these deep-learning-based methods have to overcome a number of challenges in order to be able to prove their clinical effectiveness. Besides sleep disorders, issues related to data heterogeneity, model explainability, and subjectivity will require additional consideration (Phan and Kaare, 2022).

5. Conclusion

In this study, a novel forehead electrode configuration was introduced, which used three dry electrodes placed on hairless skin to record forehead electrophysiological signals. The Fhz and Fh-EOG extracted from electrophysiological signals were used to construct an automatic sleep stage scoring system. For the evaluations, a public sleep dataset and a self-collected dataset were used. Performance comparisons using different channels and combinations are given. The results showed that the proposed method achieved the highest classification accuracy of 90.25% and 0.857 kappas when using a combination of Fhz and Fh-EOG for 26 normal subjects. Our work provides a more convenient and practical wearing mode for sleep stage detection in the clinical environment and mobile devices. It is a meaningful attempt to use forehead electrophysiological signals for sleep stage monitoring in practical applications and make long-term sleep monitoring possible.

Declarations

Author contribution statement

Hengyan Guo: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Yang Di: Conceived and designed the experiments.

Xingwei An: Conceived and designed the experiments; Wrote the paper.

Zhongpeng Wang, Dong Ming: Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported by the National Key Research & Development Program of China [2022YFF1202400] and National Natural Science Foundation of China [62276181, 62006171], Tianjin Science and Technology Project of China [20JCYBJC00930].

Data availability statement

The authors do not have permission to share data.

Declaration of interest's statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

Alickovic, E., Subasi, A., 2018. Ensemble SVM method for automatic sleep stage classification. IEEE Trans. Instrum. Meas. 67, 1258–1265.
Berry, R.B., Budhiraja, R., Gottlieb, D.J., Gozal, D., Iber, C., Kapur, V.K., et al., 2012. Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events: deliberations of the sleep apnea definitions task force of the American Academy of sleep medicine. J. Clin. Sleep Med. 8, 597–619.
Faust, O., Razaghi, H., Barika, R., Ciaccio, E.J., Acharya, U.R., 2019. A review of automated sleep stage scoring based on physiological signals for the new millennia. Comput. Methods Progr. Biomed. 176, 81–91.
Hassan, Ahnaf Rashik, Mohammed, Imamul Hassan Bhuiyan, 2016. Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating. Biomed. Signal Process Control 24, 1–10.
Hassan, A.R., Bhuiyan, M.I.H., 2016. A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features. J. Neurosci. Methods 271, 107–118.
Hassan, A.R., Bhuiyan, M.I.H., 2017. Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting. Comput. Methods Progr. Biomed. 140, 201–210.
Hao-Yu Cai, Ma, Jia-Xin, Shi, Li-Chen, Lu, Bao-Liang, 2011. A novel method for EOG features extraction from the forehead. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, Boston, MA, pp. 3075–3078.
Horne, J., 2013. Why REM sleep? Clues beyond the laboratory in a more challenging world. Biol. Psychol. 92, 152–168.
Hsu, Y.-L., Yang, Y.-T., Wang, J.-S., Hsu, C.-Y., 2013. Automatic sleep stage recurrent neural classifier using energy features of EEG signals. Neurocomputing 104, 105–114.
Huang, C.-S., Lin, C.-L., Ko, L.-W., Liu, S.-Y., Su, T.-P., Lin, C.-T., 2014. Knowledge-based identification of sleep stages based on two forehead electroencephalogram channels. Front. Neurosci. 8.
Ke, G., Meng, Q., Finley, T., et al., 2017. Lightgbm: a highly efficient gradient boosting decision tree[J]. Adv. Neural Inform. Process. Syst. 30.
Khalighi, S., Sousa, T., Pires, G., Nunes, U., 2013. Automatic sleep staging: a computer assisted approach for optimal combination of features and polysomnographic channels. Expert Syst. Appl. 40, 7046–7059.
Lee, B.-G., Lee, B.-L., Chung, W.-Y., 2014. Mobile healthcare for automatic driving sleep-onset detection using wavelet-based EEG and respiration signals. Sensors 14, 17915–17936.
Levendowski, D.J., Ferini-Strambi, L., Gamaldo, C., et al., 2017. The accuracy, night-to-night variability, and stability of frontopolar sleep electroencephalography biomarkers[J]. J. Clin. Sleep Med. 13 (6), 791–803.
Liang, S.-F., Kuo, C.-E., Hu, Y.-H., Pan, Y.-H., Wang, Y.-H., 2012. Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models. IEEE Trans. Instrum. Meas. 61, 1649–1657.
Mora, A.M., Fernandes, C.M., Herrera, L.J., Castillo, P.A., Merelo, J.J., Rojas, F., et al., 2010. Sleeping with ants, SVMs, multilayer perceptrons, and SOMs. In: 2010 10th International Conference on Intelligent Systems Design and Applications. IEEE, Cairo, Egypt, pp. 126–131.
Motamedi-Fakhr, S., Moshrefi-Torbati, M., Hill, M., Hill, C.M., White, P.R., 2014. Signal processing techniques applied to human sleep EEG signals—a review. Biomed. Signal Process Control 10, 21–33.
Phan, Huy, Do, Quan, Do, The-Luan, Vu, Duc-Lung, 2013. Metric learning for automatic sleep stage classification. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Osaka: IEEE), pp. 5025–5028.
Phan, Huy, Kaare, Mikkelsen, 2022. Automatic Sleep Staging of EEG Signals: Recent Development, Challenges, and Future Directions. Physiological Measurement.
Rahman, M.M., Bhuiyan, M.I.H., Hassan, A.R., 2018. Sleep stage classification using single-channel EOG. Comput. Biol. Med. 102, 211–220.
Ronzhina, M., Janoušek, O., Kolářová, J., Nováková, M., Honzík, P., Provazník, I., 2012. Sleep scoring using artificial neural networks. Sleep Med. Rev. 16, 251–263.
Xue-Qin Huo, Zheng, W.-L., Lu, B.-L., 2016. Driving fatigue detection with fusion of EEG and forehead EOG. In: 2016 International Joint Conference on Neural Networks (IJCNN). IEEE, Vancouver, BC, Canada, pp. 897–904.
Yildirim, O., Baloglu, U., Acharya, U., 2019. A deep learning model for automated sleep stages classification using PSG signals. IJERPH 16, 599.
Zarei, A., Asl, B.M., 2019. Automatic detection of obstructive sleep apnea using wavelet transform and entropy-based features from single-lead ECG signal. IEEE J. Biomed. Health Inform. 23, 1011–1021.
Zhang, Y.-F., Gao, X.-Y., Zhu, J.-Y., Zheng, W.-L., Lu, B.-L., 2015. A novel approach to driving fatigue detection using forehead EOG. In: 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE, Montpellier, France, pp. 707–710.
Zhu, G., Li, Y., Wen, P., 2014. Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal. IEEE J. Biomed. Health Inform. 18, 1813–1821.