BUILD SIMUL (2023) 16: 1409–1423
https://doi.org/10.1007/s12273-023-1049-6

A sleep staging model for the sleep environment control based on machine learning

Ting Cao, Zhiwei Lian (✉), Heng Du, Jingyun Shen, Yilun Fan, Junmeng Lyu

School of Design, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract

To date, dynamic sleep environment has been attracted the focus of researchers. Owing to the individual difference on sleep phase and thermal comfort, changes in sleep environment should be occupant-centered, and precise regulation of the environment required current sleep stages. However, few studies connected occupants and the environment through physiological signal-based model of sleep staging. Therefore, this study tried to develop a data driven sleep staging model with higher accuracy through sleep experiments collecting information. Raw database was processed and selected efficiently according to the characteristics of physiological signals. Finally, the sleep staging model with an average accuracy of 93.9% was built, and other mean indicators (precision: 82.5%, recall: 83.1%, F1 score: 82.8%) performed well. The features adopted by model were found to come from different brain regions, and the global brain signals were suggested to play an important role in the construction of sleep staging model. Moreover, the computational processing of physiology signals should consider their characteristics, i.e., time domain, frequency domain, time-frequency domain and nonlinear characteristics. The model obtained in this study may deliver a credible reference to advance the research on control of sleep environment.

Keywords

sleep environment;
sleep staging;
model;
physiological signals;
machine learning;
environmental control

Article History

Received: 10 April 2023
Revised: 09 May 2023
Accepted: 24 May 2023

1 Introduction

Automatic regulation of air conditioner in thermal environment for comfort (Zhou et al. 2022; Jaffal 2023) and energy saving (Fan et al. 2022; Tang et al. 2022) has been receiving increasing interest from researchers (Gan et al. 2022). Ghahramani et al. (2018) captured personal thermal comfort for predicting uncomfortable states (i.e., warm, cool) to communicate to the air conditioner, followed by Wu et al. (2023) developing the intelligent air conditioners based on individual thermal comfort. They have addressed the environmental control in the waking state. However, during sleep, people cannot actively regulate the environment. Any changes outside of their comfort zone would disrupt sleep and seriously reduce sleep quality (Haskell et al. 1981). Sleep environmental control thus needed more attention as well.

So far, some studies have investigated the sleep status in dynamic environments. Lan et al. (2016) designed three different ways of changing the ambient temperature at night (condition 1: constant 26 °C; condition 2: rise-fall (25 °C – 26 °C – 27 °C – 28 °C – 27 °C – 26 °C); condition 3: fall-rise (28 °C – 27 °C – 26 °C – 27 °C – 28 °C), and found that the starting from 25 °C might severely affected sleep efficiency. Miyake et al. (1996) set three different fluctuations between 28 and 30 °C (baseline condition: constant; condition 1: 45 min/cycle; condition 2: 90 min/cycle), and results showed that the deep sleep duration was significantly longer in changing environment. Although the above studies proposed that appropriate environmental fluctuations throughout the night would be conducive to sleep, the regulation of environment relied on operators and might not be available in daily life. In addition, sleep periods can be divided into two main stages: rapid eye movement (REM) and non-rapid eye movement (NREM), where NREM was subdivided into stage 1 (N1), stage 2 (N2), and stage 3 (N3) (Rechtschaffen and Kales 1968). Possible individual differences were reflected in the random occurrence of different sleep stages. When in light sleep (N1, N2, or REM stage), improper environment may give rise to the sleep disruption (Lan et al. 2017; Zhang et al. 2023).

E-mail: zwlian@sjtu.edu.cn

Therefore, creating a comfortable dynamic environment at night needed to be occupant-centered (Yang et al. 2022). Current information of sleep stages should be transmitted to the environment in time, the accuracy of which would directly determine to the response of air-conditioner. Ngarambe et al. (2019) attempted to connect the sleep phase with body movements in order to change the room temperature, but objective results were not clearly given. American Academy of Sleep Medicine (AASM) manual, one of the authoritative standards in sleep research, stated that physiological signals were important during sleep (Berry et al. 2020), such as electroencephalogram (EEG), electromyography (EMG), and electrooculogram (EOG). These signals have been found to appear in different shapes due to the rhythms in different sleep periods, thus providing potential and accurate information for sleep classification. Especially, five basic waves of brain activity were dominated at different phases (Berry et al. 2020): delta (0.3–3.99 Hz, 75 uV), theta (7–4.99 Hz), alpha (8–13 Hz), beta (12–14 Hz), and gamma (30–40 Hz), which acquired from several signal channels in different brain regions. Six recommended channels of EEG in Figure 1(a) were named C4A1, C3A2, F4A1, F3A2, O2A1, and O1A2, referring to the 10–20 electrode placement system. The locations of EOG were E1A2 (LOC: EOG detected on left side) and E2A1 (ROC: EOG detected on right) in Figure 1(b), and EMG probes were located at the left (EMG_L) and right (EMG_R) mandible in Figure 1(c) (Rechtschaffen and Kales 1968).

Traditional sleep staging was done by experts who manually classified according to the various frequencies and shapes of waves. However, in the auto-controlled sleep environment, real-time sleep stage information provided to the environmental equipment required the exploration of the automatic sleep staging techniques. Advances in artificial intelligence and the availability of large datasets have driven the development of sleep staging (Craik et al. 2019). Most researchers preferred the path of machine learning based on feature engineering starting from feature extraction (Motamedi-Fakhr et al. 2014). Physiological signals have four different properties, including time domain (Motamedi-Fakhr et al. 2014), frequency domain (Fell et al. 1996), time-frequency domain (Oropesa et al. 1999), and nonlinear analysis (Pincus 1991; Richman and Moorman 2000; Acharya et al. 2005). The analysis and calculation on the characteristics could effectively extract useful information from different sleep periods and achieve the goal of automatic staging. Table 1 listed some relevant studies on automatic sleep staging from 2017.

Although the mentioned studies worked on automatic staging from physiological electrical signals, there were several problems. (1) Almost all studies have used public database signals for analysis, and the performance of models based on actual experimental data remained to be considered. (2) Few studies have focused on EMG and EOG. In fact, sleep stages such as N1 and REM stages were difficult to discriminate, and Šušmáková and Krakovská (2008) suggested that the addition of EMG information would possibly solve the dilemma. (3) They did not conduct the feature selection before modeling. A portion of redundant features might be overfitted to mask the performance of classification and increase the computational load. (4) Only parts of the properties and signal channels were analyzed, which can ignore a lot of useful information.

In conclusion, for the objective of automatic sleep environment control, sleep staging model linked the environment to the occupant, and provided the important sleep information to air conditioners. Here was an example about the relationship between an air conditioner and the sleep stage determination. Assuming a user was in a deep sleep stage, there was an inappropriate disturbance factor which may cause the environment too warm or too cold. This may give rise to the conversion to light sleep phases or

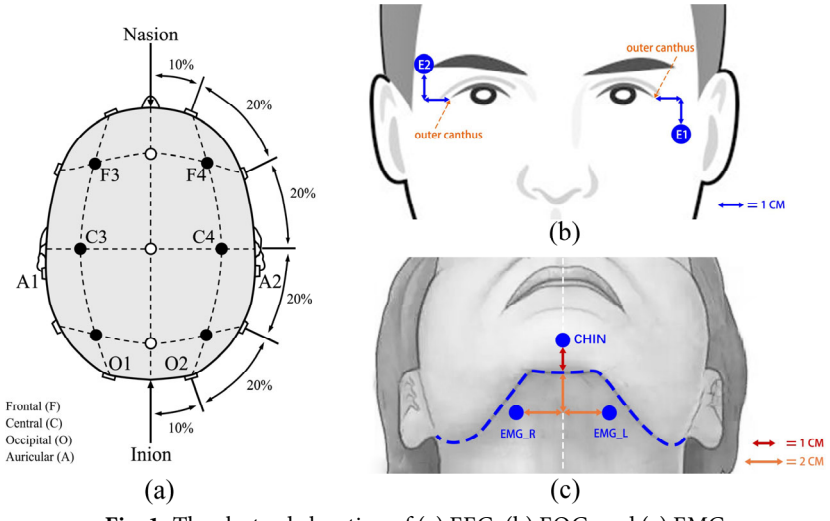

Fig. 1 The electrode location of (a) EEG; (b) EOG; and (c) EMG

Table 1 Previous studies on sleep staging

Reference	Signal	Characteristic	Data resource	Process	Note
Gharbali et al. 2018	EEG, EOG, EMG	Time domain	Open access dataset	Data process & modeling	Single EEG channel
Zhao et al. 2019	EEG, EOG, EMG	Time domain, time-frequency domain, nonlinear	Open access dataset	Data process & modeling	Single EEG channel
Alickovic and Subasi 2018	EEG	Time-frequency domain	Open access dataset	Data process & modeling	Single EEG channel
Hassan and Bhuiyan 2016	EEG	Frequency domain	Open access dataset	Data process & modeling	Single EEG channel
Fu et al. 2021	EEG	Using deep learning	Open access dataset	Modeling	Two EEG channels
Grossi et al. 2021	EEG	Time domain	From hospital	Data process & modeling	Single EEG channel
Vilamala et al. 2017	EEG	Using deep learning	Open access dataset	Modeling	Two channels of EEG
Supratak et al. 2017	EEG	Raw data	Open access dataset	Modeling	Single EEG channel
Bresch et al. 2018	EEG	Using deep learning	Open access dataset	Modeling	Single EEG channel
Zhang et al. 2020	EEG	Using deep learning	Open access dataset	Data process & modeling	Single EEG channel

wakefulness and the decrement of sleep quality. The air conditioner equipped with the sleep staging model proposed in the study can detect the corresponding information and act immediately to adjust the ambient temperature into a range suitable for deep sleep. Throughout the process, the judgement of sleep state was important. Therefore, to improve the performance of sleep staging model, this study conducted the sleep experiment, and collected signal data from six EEG channels, EMG, and EOG. The feature screening was processed as well. We expected to discover new conclusions from the optimal feature sets as the conditions for the realization of automatic environmental control, and suggested that the physiological signals should continue to be applied in the automatic control of sleep environment. The potential applications and prospects in building field research was that this study can propose the automatic sleep staging technology for further transmitting these sleep information (including sleep stages and physiological electrical signal characteristics, etc.) into the environmental control devices in real time, and the relevant devices may regulate the physical environment according to the current sleep stage. This automatic approach can make the auto control in the sleep environment more closely related to people's sleep quality.

2 Method

In this section, it was clear that the data were derived from sleep experiments and contained information on physiological electrical signals (Section 2.1). In addition, data processing would follow the classical steps of machine learning (from preprocessing to feature extraction) (Sections 2.1 and 2.2), and a comparison of three feature selection algorithms (Section 2.3) was originally added to reduce the negative impact of feature redundancy on the results before training the model (Section 2.4).

2.1 Data

Data resource. Through questionnaire screening, 28 young volunteers (17 males and 11 females) with similar schedules (age: 23 ± 2 years; BMI: males: 21.6 ± 1.6 kg/m^2, females: 19.5 ± 1.1 kg/m^2; bedtime 22:50–23:30, wake up time 7:00–7:30) were recruited. Before the experiment, Pittsburg Sleep Quality Index (PSQI), as a self-assessment index to judge sleep quality over a one-month time, was counted (Cao et al. 2022; Xu et al. 2022). Their PSQI scores less than four suggested that they were not troubled by sleep issues. During the experiment, participants should report their health status, and intake of caffeine, as well as psychotropic drugs, was strictly controlled. The experiment was supported by the Ethical Committee of Shanghai Jiao Tong University, and participants signed written informed consents. The experiment was conducted in two identical and adjacent sleep chambers in Shanghai Jiao Tong University (Figure 2). The laboratories were decorated like a daily bedroom to weaken the psychological impact (Figure 3). During the experimental period, two subjects were required to arrive at the preparation room at 21:00 to change into pajamas. 15 min later, they were guided into the chambers, and wore Somté PSG (Compumedics, Australia; Accuracy) to record EEG, EOG, and EMG following the position recommended in Figure 1. At 23:00, subjects began to sleep until 7:00 in the next morning, and took off the devices. After completing the whole experiment, 212 nights of sleep raw data were collected, including six channels of EEG (location noted as F3A2, C3A2, O1A2, F4A1, C4A1, O2A1 in Figure 1), two channels of EMG (EMG_L and EMG_R), and two channels of EOG (LOC and ROC). The sample rate was set at 256 Hz.

Data process. The continuous sleep data of 2–3 hours were randomly selected from each night of data, and finally, 442.7 hours were extracted. Generally, experts choose a

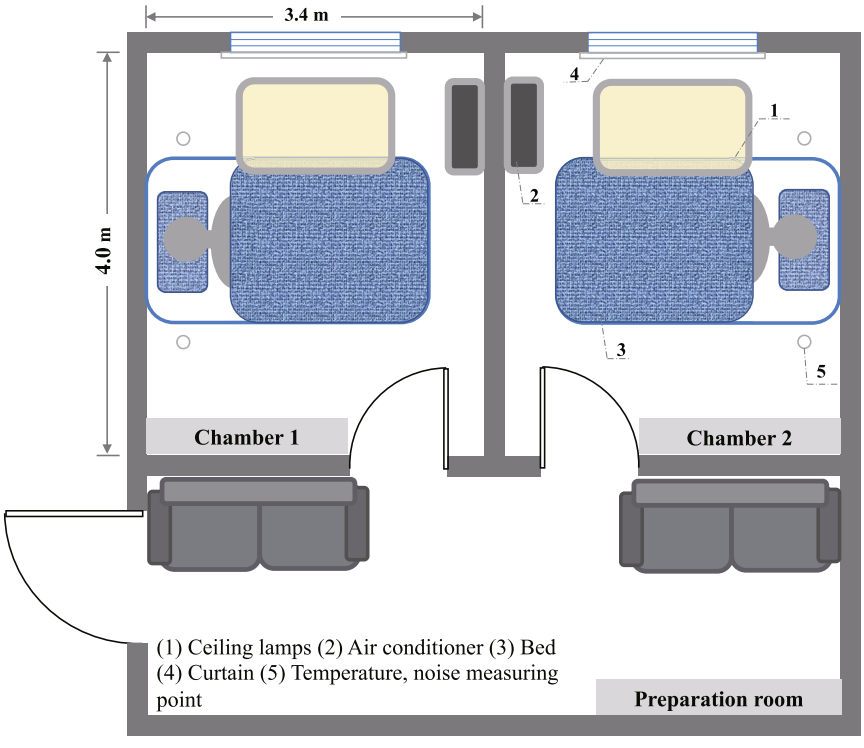

Fig. 2 Map of the experimental room

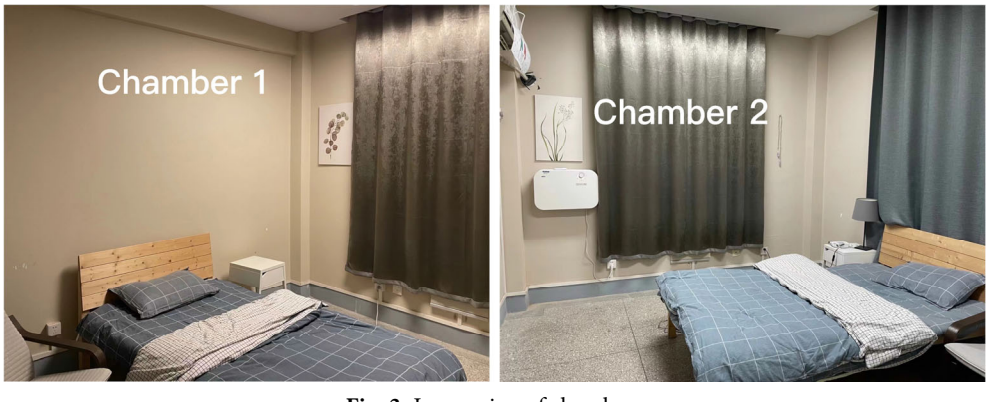

Fig. 3 Impression of chambers

30-second window (one epoch) to scan the whole night signals to classify the stages (Rechtschaffen and Kales 1968). Thus, the total numbers of sleep epochs were 53120 (males: 32398, females: 20722), including 8744 stages of N1, 11968 stages of N2, 17939 stages of N3, 8386 stages of REM, and 6083 stages of waking. All epochs were randomly divided 7:2:1 into a training, validation, and test set. The stages were independently scored by two experts referring to the AASM standard. Considering the inter-scorer agreement between the two experts in sleep stage scoring (N1, N2, N3, REM, Wake), Weighted Kappa (Fleiss and Cohen 1973) was calculated as 0.745 ($p < 0.05$), suggesting the significant consistency between two experts. If there was an ambiguity, the third independent specialist made the decision according to the standard. The frequency at 50 Hz in all raw data was cut off to eliminate the interference of environmental noise. A 3rd order IIR was designed to do the bandpass filter work for EEG and EOG with the cut-off frequency of 0.1–40 Hz, whereas the other 3rd order IIR was applied to pass the EMG signal at the frequency of 10–100 Hz (Pan et al. 2012).

2.2 Feature extraction

Feature extraction was worked on four methods corresponding to the different characteristics of the signal. First was time domain analysis, the information in signals, such as waveform, amplitude, etc., can be directly observed. Secondly, frequency analysis was related to the Fast Fourier Transform with Welch's method to convert into frequency domain relationship between frequency and power density.

Every 30 s of time domain signal was split into five non-overlapping segments, and 6-s Hamming window scanned each segment to calculate the mean spectrum. Moreover, time-frequency domain calculated the wavelet entropy of each epoch to assess the non-stationarity (Rosso et al. 2001). Finally, fractal dimension, entropy measures and Lyapunov exponents were commonly used in the nonlinear analysis of EEG. Table 2 listed all the features, including their functions and references. Some features were inclined to occur during specific period. Considering that the distribution of feature might affect the model, at the beginning of the feature calculation, we adopted the important properties of each sleep stage as much as possible and quantified these characteristics effectively by some methods. For example, delta waves dominated in the N3 phase, so we tried to fully express all the features of delta waves through multiple

Table 2 The list of feature name

No.	Feature name	Function	Characteristic	Reference	Note
1	Slow wave activity	Count delta waves for N3 stage	Time	Bersagliere and Achermann2010	SWA
2	Spectral edge frequency	Detect the REM period	Frequency	Imtiaz et al. 2014	SEFd
3	Central frequency	Reflect the multiple EEG bands	Frequency	Pop-Jordanova and Pop-Jordanov2005	f_{CF}
4	Power density of delta	Give delta information for deep sleep	Frequency	Berry et al. 2020	Delta
5	Power density of theta	Give theta information for light sleep	Frequency	Berry et al. 2020	Theta
6	Power density of alpha	Give alpha information for light sleep	Frequency	Berry et al. 2020	Alpha
7	Power density of gamma	Give gamma information for light sleep	Frequency	Berry et al. 2020	Gamma
8	Power density of beta	Give beta information for light sleep	Frequency	Berry et al. 2020	Beta
9	Power density of spindles	Give spindles information for N2 stages	Frequency	Berry et al. 2020	Spindles
10	Ratios of power density	Staged the different periods	Frequency	Šušmáková and Krakovská 2008	Sigma/Beta
11	Ratios of power density	Staged the different periods	Frequency	Šušmáková and Krakovská 2008	Alpha/Delta
12	Ratios of power density	Staged the different periods	Frequency	Šušmáková and Krakovská 2008	Beta/Theta
13	Ratios of power density	Staged the different periods	Frequency	Šušmáková and Krakovská 2008	Delta/Beta
14	Ratios of power density	Staged the different periods	Frequency	Šušmáková and Krakovská 2008	Alpha/Theta
15	Ratios of power density	Staged the different periods	Frequency	Šušmáková and Krakovská 2008	Delta/Theta
16	Ratios of power density	Staged the different periods	Frequency	Louis et al. 2004	(Delta×Alpha)/(Beta×Gamma)
17	Ratios of power density	Staged the different periods	Frequency	Louis et al. 2004	(Theta×Theta)/(Delta×Alpha)
18	Ratios of power density	Detect eye activity during N3 and REM	Frequency	Malafeev et al. 2018	powEOG/Delta
19	Wavlet entropy	Analyze the distribution of signal energy	Time-frequency	Rosso et al. 2001	WavPackEn
20	Spectral entropy	Recognize the different frequency waves	Nonlinear	Şen et al. 2014	SpectEn
21	Approximate entropy	Recognize the different frequency waves	Nonlinear	Şen et al. 2014	ApEn
22	Sample entropy	Recognize the different frequency waves	Nonlinear	Şen et al. 2014	SampEn
23	Fractal dimension	Distinguish well between N3 and others	Nonlinear	Acharya et al. 2005	FD
24	Lyapunov exponents	Distinguish well between N1 and N2	Nonlinear	Fell et al. 1996	Le
25	Blink wave	Eye movement for waking	Time	Carl et al. 2012	Blink_w
26	Slow eye movement	Eye movement for N1 stage	Time	Magosso et al. 2006	SEM_w
27	Rapid eye movement	Eye movement for REM stage	Time	Malafeev et al. 2018	REM_w
28	EOG artifacts	Check normal situation of data	Time	Imtiaz and Rodriguez-Villegas 2014	EOG_art
29	Power density of EMG	Muscle activity during sleep	Frequency	Stanus et al. 1987	pow_EMGR
30	Power density of EMG	Muscle activity during sleep	Frequency	Stanus et al. 1987	pow_EMGL
31	Fractal dimension of EMG	Identification of EMG for REM	Nonlinear	Šušmáková 2008	FD_EMGL
32	Fractal dimension of EMG	Identification of EMG for REM	Nonlinear	Šušmáková and Krakovská 2008	FD_EMGR

Note: No.1 to No. 24 belonged to six EEG channel (F3A2, C3A2, O1A2, F4A1, C4A1, O2A1) in Figure 1(a); No.25 to No. 28 were from EOG channel, and No. 29 to No. 32 were from EMG channel.

analysis in order to distinguish well between the N3 and other phases. The process was repeated as well in REM, N1, N2, waking stages.

All the above features were listed in Table 2, and the formulas for calculating them could be checked in the Appendix. Among them, No. 1 to No. 24 were applied in the six EEG channels, and features 25 to 32 were extracted from EOG and EMG channel. Finally, 152 features (24 × 6 channels + 4 EOG features + 4 EMG features = 152 features) were initially extracted.

Before starting the following contents, the pipeline of Sections 2.3 and 2.4 was shown in Figure 4 for better understanding. After feature extraction in Section 2.2, feature selection should be performed in order to input the valid feature information into the model and reduce the computational load. However, there were many algorithms for filtering (step 1), and they were required to be compared. Therefore, in step 2, when filtering features, the training set was fed into several classification models that were built by three different selection algorithms separately combined with random forest. The validation set was then brought into each of these classification models to give overall accuracy results (step 3). After choosing the best model, the test set was input directly into the model to verify the final performance of the model (step 4).

2.3 Feature selection algorithm

When selecting features, there were too many combinations of features related to the different distributions, and all these complex connections could address the multi-classification task. However, it was not possible to choose the best manually, but relevant algorithms can better decide which distribution would be the most effective for staging tasks. Feature selection was classified into three categories: wrappers, embedded, and filter (Roffo et al. 2015). Wrapper and embedded methods depended on the types of classifiers, meaning that the classifier changed along with the adjustment of the features. For the generalization property, this study employed three types of filtering algorithms, namely ReliefF (Kononenko 1994), mRMR (Ding and Peng 2003), and Fisher (Šušmáková and Krakovská 2008) through MATLAB 2021a (Mathworks, America). They were more available for multi-classification tasks and extraction the most relevant set of features from the raw data.

2.4 Classifier

The classifier for sleep staging in this study was random forest, supported by Şen et al. (2014) recommending the random forest owing to the best performance for sleep classification. Especially, random forest here was involved to compare among three filtering algorithms in Section 2.3 through MATLAB 2021a (Mathworks, America), and all potential models with the overall accuracy and performance was obtained. Our study was the result-oriented selection of the best feature set and better model, which can avoid the influence of complex distributional relationships. The performance of the final model would be exhibited in confusion matrix and evaluated from the accuracy, precision, recall, and F1 score. We would give the evaluation results of the final model in N1, N2, N3, REM and W stage. Taking the results at N3 stage as an example, accuracy referred to correctly predicting in all samples whether this stage belonged to N3; precision involved the actual N3 stage in the predicted samples of N3; recall was connected with the predicted correctly samples in actually N3 period; F1 score was the summed average of precision and recall.

2.5 Statistic analysis

In order to compare three feature filtering algorithms, all features were sorted by the results of important ranks from algorithms, and incrementally inputted in the classifier to plot the curve between the number of features and the accuracy of the model. After determining the number of features, the accuracy changes of the three algorithms were required to be analyzed by significances tests through SPSS R26.0.0.0. Firstly, the normality was calculated through Shapiro Wilk's W test. Then, for normally distributed data, a repeated measures design and a paired sample t-test were suitable, whereas Wilcoxon Signed-Ranks test was used for non-normally distributed data. The statistic results were marked with p value ($p < 0.05$), and if $p < 0.1$, effect size should be signed with Cohen's d for testing the degree of deviation (Small: $d = 0.2$, medium: $d = 0.5$, and large:

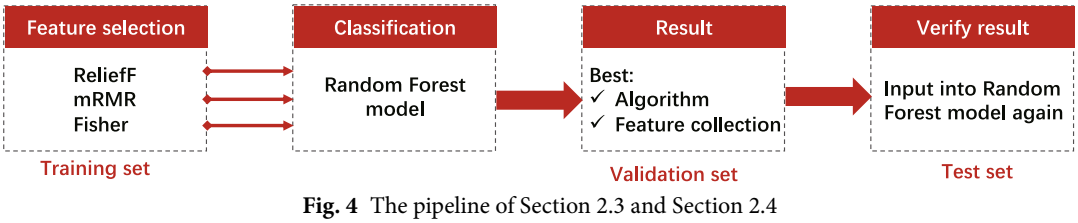

Fig. 4 The pipeline of Section 2.3 and Section 2.4

$d = 0.8$) (Lan and Lian 2010). The larger effect size related to the more meaningful statistically significance.

3 Results

3.1 Feature selection

Figure 5 reports the feature ordering results of features 1 to 20 calculated by mRMR, ReliefF, and Fisher, in which the abscissa represented the feature number, and the ordinate was the No. ranking result. The complete sorting results of features 1 to 152 were placed in the Appendix, which is available in the Electronic Supplementary Material in the online version of this article. Taking the feature 4 ($Delta_{\_C3A2}$) as an example (marked with a blue line), the ranking order by three algorithms were the 72th (mRMR), 26th (ReliefF), and 90th (Fisher). The different result occurred in feature 13 ($Delta/Beta_{C3A2}$) as well. Furthermore, Table 3 lists the

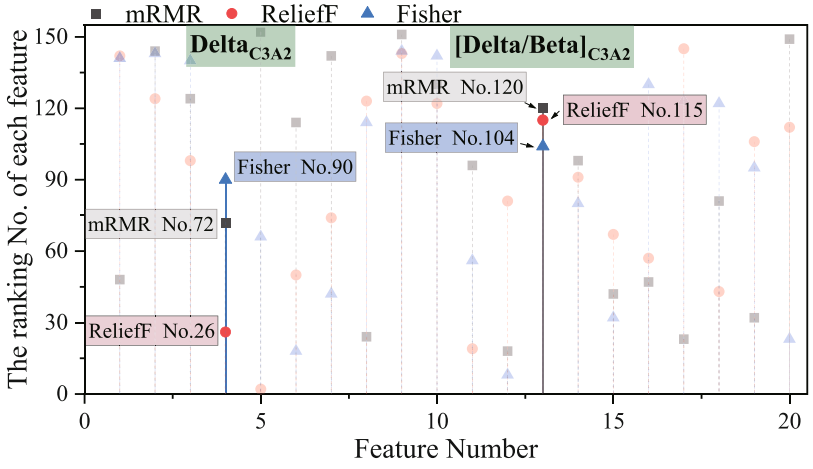

Fig. 5 The ranking results of features 1 to 20. Note: abscissa was related to the feature before ranking, and the specific information could be found in the Appendix in the online version of this article

Table 3 The list of the top 20 feature names through three algorithms

	mRMR		ReliefF		Fisher
1	*Delta*$_{\_F3A2}$	1	*SpectEn*$_{\_F4A1}$	1	*SpectEn*$_{\_O2A1}$
2	*SampEn*$_{\_C4A1}$	2	*Theta*$_{\_C3A2}$	2	*ApEn*$_{\_O1A2}$
3	*powEOG/Delta*$_{\_O2A1}$	3	*SampEn*$_{\_F3A2}$	3	[*Theta*×*Theta/Delta* ×*Alpha*]$_{C4A1}$
4	[*Sigma/Beta*] $_{F4A1}$	4	[*Delta/Theta*] $_{F4A1}$	4	*SEFd*$_{\_O2A1}$
5	*Spindles*$_{\_F4A1}$	5	*Spindles*$_{\_F4A1}$	5	[*Delta/Theta*]$_{C4A1}$
6	*Delta*$_{\_F4A1}$	6	*Le*$_{\_F3A2}$	6	[*Sigma/Beta*]$_{C4A1}$
7	*Alpha*$_{\_F4A1}$	7	*Alpha*$_{\_F4A1}$	7	*Spindles*$_{\_C4A1}$
8	*Spindles*$_{\_C4A1}$	8	*FDH*$_{\_C4A1}$	8	[*Delta/Theta*] $_{C3A2}$
9	[*Beta/Theta*] $_{C4A1}$	9	*Gamma*$_{\_O1A2}$	9	*WavlPackEn*$_{\_O2A1}$
10	*ApEn*$_{\_C4A1}$	10	*powEOG/Delta*$_{\_C4A1}$	10	*Theta*$_{\_F4A1}$
11	*Beta*$_{\_O2A1}$	11	*Spindles*$_{\_O2A1}$	11	*Alpha*$_{\_O2A1}$
12	*Le*$_{\_F3A2}$	12	*SWA* $_{\_F4A1}$	12	*SWA*$_{\_F4A1}$
13	*SWA*$_{\_F3A2}$	13	*FD*$_{\_O2A1}$	13	*SpectEn*$_{\_F3A2}$
14	*SEFd*$_{\_O1A2}$	14	[*Delta*×*Alpha/Beta* ×*Gamma*] $_{F4A1}$	14	*ApEn*$_{\_F4A1}$
15	*SampEn*$_{\_F3A2}$	15	*Delta*$_{\_O1A2}$	15	*powEOG/Delta*$_{\_F4A1}$
16	*Gamma*$_{\_F3A2}$	16	*Le*$_{\_O1A2}$	16	*Alpha*$_{\_F3A2}$
17	[*Alpha/Theta*] $_{O2A1}$	17	*Central freq.*$_{\_F3A2}$	17	*Le*$_{\_F3A2}$
18	[*Delta/Theta*] $_{C3A2}$	18	*Delta*$_{\_F4A1}$	18	*Alpha*$_{\_C3A2}$
19	*pow_EMGR*	19	[*Alpha/Delta*]$_{C3A2}$	19	*Spindles*$_{\_O2A1}$
20	*Delta*$_{\_O1A2}$	20	*SpectEn*$_{\_C4A1}$	20	[*Beta/Theta*] $_{O1A2}$

Note: EEG channel name listed at the bottom right of the feature name.

top 20 feature names by the three algorithms with EEG channel information at the bottom right. The inconsistent sorting results indicated that the algorithm should be selected first to determine the best feature sets.

The features sorted by the three algorithms were preliminarily trained with 456 (152 times 3) random forest models. The total accuracy of classification in all 456 models was verified with the validation sets, and the trend of accuracy along with the number of features was depicted as the folding line in Figure 6. As the number of features increased, accuracy raised to a steady-state. The detail of the knee point was amplified in Figure 6(a) due to the large scale of the *x*-axis. Results showed that Fisher algorithm provided the highest accuracy with the minimum quantity of features. For large datasets without affecting the accuracy of the staging model, effectively censoring numbers could improve computational efficiency. Additionally, the accuracy difference in the smooth segments from knee points to the last features by the three algorithms was statistically analyzed in Figure 6(b), and significant results showed that the accuracy of Fisher was much better than the other two algorithms. When calculating the rate of change in accuracy after the knee point, values for Fisher performed the lowest (1.40%). Thus, Fisher algorithm was adopted to effectively filter out 11.18% (17/152) of features and get an optimal feature set consisting of 17 features.

The final features decided by Fisher algorithm were brought into the random forest for sleep staging classification. A confusion matrix can exhibit the potential of a classifier, and answer how many predictions were correct and incorrect per class. The confusion matrix result of this study on test set was given in Figure 7 suggesting that the final model had good performance for sleep classification on five different periods. Furthermore, the specific indices (i.e., accuracy, precision, recall, F1 score) were listed in Table 4. The accuracy for five different sleep stages in row one was between 90.1% and 96.5%. With regard to precision, recall, and F1 score, values exceeding 90% were at N3 stage classification, while N2 stage, REM stage, and Wake stage classification results were just around 80%. However, the performance indices for N1 stage were lower than those of other sleep stages, which needed to be discussed. The mean accuracy (93.9%), precision (82.5%), recall (83.1%), F1 score (82.8%) of the ultimate model were proposed, suggesting that one epoch predicted by the model was actually true with the probability of 82.5%, and the model has an 83.06% probability of successfully predicting an actual stage. Additionally, the model could judge whether one epoch belonged to the sleep period with an accuracy of 93.9%. For the sleep staging task, the above four metrics were equally important to prove the better staging performance of the model in this study.

3.2 Feature sets at different sleep stage

The names of 17 features selected and sorted by Fisher were listed in Table 5. Many features appeared repeatedly but from different EEG channels, such as *ApEn*, *SpectEn*, *Alpha*, *Delta/Theta*, illustrating that Fisher algorithm preferred to extract some features contributing to classification instead

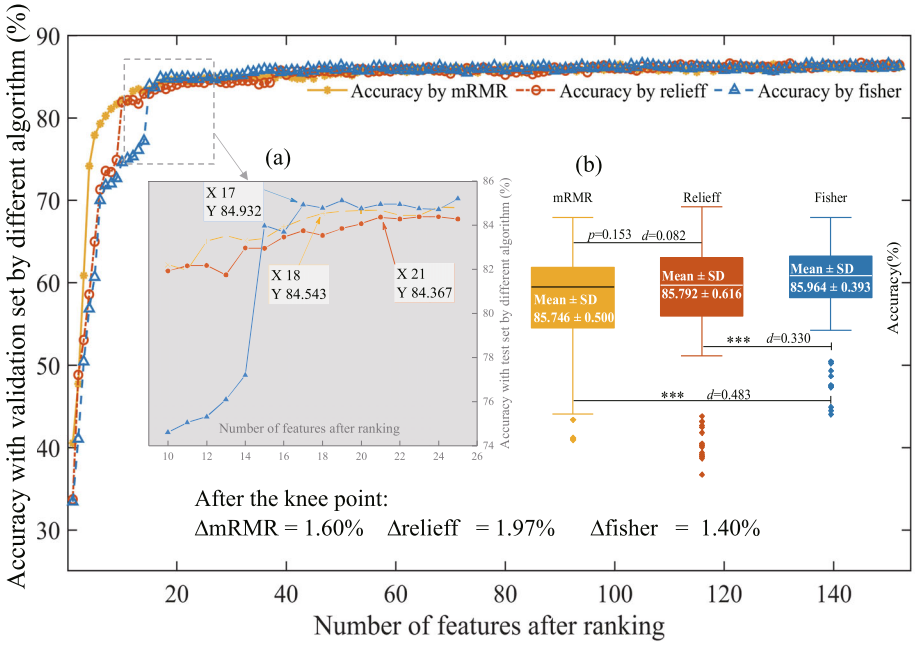

Fig. 6 Relationship between the number of features and accuracy: (a) the detail of knee points; (b) the significant difference of three algorithm after the knee points

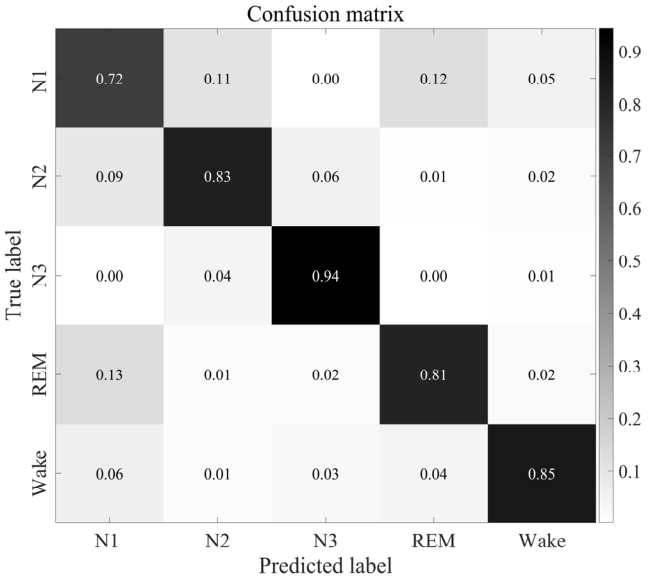

Fig. 7 Confusion matrix of sleep classification through the final model

Table 4 The performance of classification on test sets

	Test set				
	N1	N2	N3	REM	Wake
Accuracy	90.1%	92.9%	96.3%	93.9%	96.5%
Precision	74.7%	80.0%	95.2%	82.2%	80.5%
Recall	71.7%	82.7%	94.4%	81.4%	85.2%
F1 score	73.2%	81.3%	94.8%	81.8%	82.8%

of filtering out features according to channels. The statistical results showed that four features were extracted from O2A1, F4A1, and C4A1, three from F3A2, and one from O1A2 and C3A2 respectively. It meant that signals from each channel were indispensable for the assignment of the sleep staging task. Moreover, in the optimal feature sets, ten features (58.8%) were based on frequency domain analysis, five features (29.4%) related to nonlinear characteristics, and time domain, as well as time-frequency domain features, accounted for 5.9% respectively. In frequency analysis, four features were the spectrum ratio in different signal bands, emphasizing the importance of proportional combinations during EEG staging tasks.

The value distributions from the optimal feature sets were exhibited in the Appendix, and here only some of the features were described for analysis. Firstly, all features corresponded to the close distribution and mean value between the training and test sets, indicating the consistent results of machine prediction staging and manual label. Spectral entropy, taking the first ranked feature as an example, was shown in Figure 8(a). The horizontal short line meant the average value of the scatter distribution, and the spectral entropy value of each sample was represented with gray (the manual labels on the training set) and red points (the predicted labels on the test set). Both means of spectral entropy (0.38 and 0.42) at N3 was significantly lower than that at other stages, similar with $ApEn\_{O1A2}$ (No.2), $[Delta/Theta]_{C4A1}$ (No.5), $WavPackEn\_{O2A1}$ (No.9), $SpectEn\_{F3A2}$ (No.13), $ApEn\_{F4A1}$ (No.14), Le_{F3A2} (No.17) in the Appendix. For the waking stage in Figure 8(b), the mean of $Alpha\_{O2A1}$ (No.11) could be obviously discriminated between waking and other stages, and the average of $powEOG/Delta\_{F4A1}$ in REM stage was significantly higher (Figure 8(c)). Other features appeared the identical means of distribution among multiple sleep stages, such as the uniform distribution of $[Sigma\ /\ Beta]_{C4A1}$ in N2 and N3 stages (Figure 8(d)).

3.3 Results of classification for individual

This section displays a sample of the overnight classification illustrating the predicted accuracy of the staging model in Figure 9. Figure 9(a) plots the manual staging by experts, and Figure 9(b) shows the results judged by the trained model. The predictions were generally close to the manual labels, except for the differences in the N1 and N2 stage at 23:00-23:15, 02:00-03:00, and 05:00-06:00. In addition, the changes of the top three features during the whole night were depicted in Figures 9(c)–(e), and the eigenvalues fluctuated significantly along with the variation of sleep phases.

4 Discussion

4.1 Comparison of the staging model with other studies

This section compared and discussed other studies from the perspective of data sources, feature input, and staging performance.

Data source. Most of the current models had the original EEG data from public open-source datasets (Hassan and Bhuiyan 2016; Alickovic and Subasi 2018; Gharbali et al. 2018), like Sleep-EDF datasets (Zhao et al. 2019; Supratak

Table 5 The list of Feature sets

No.	Feature sets name					
1–6	$SpectEn\_{O2A1}$	$ApEn\_{O1A2}$	$[Theta\times Theta/Delta\times Alpha]_{C4A1}$	$SEFd\_{O2A1}$	$[Delta/Theta]_{C4A1}$	$[Sigma/Beta]_{C4A1}$
7–12	$Spindles\_{C4A1}$	$[Delta/Theta]_{C3A2}$	$WavlPackEn\_{O2A1}$	$Theta\_{F4A1}$	$Alpha\_{O2A1}$	SWA_{F4A1}
13–17	$SpectEn\_{F3A2}$	$ApEn\_{F4A1}$	$powEOG/Delta\_{F4A1}$	$Alpha\_{F3A2}$	Le_{F3A2}	

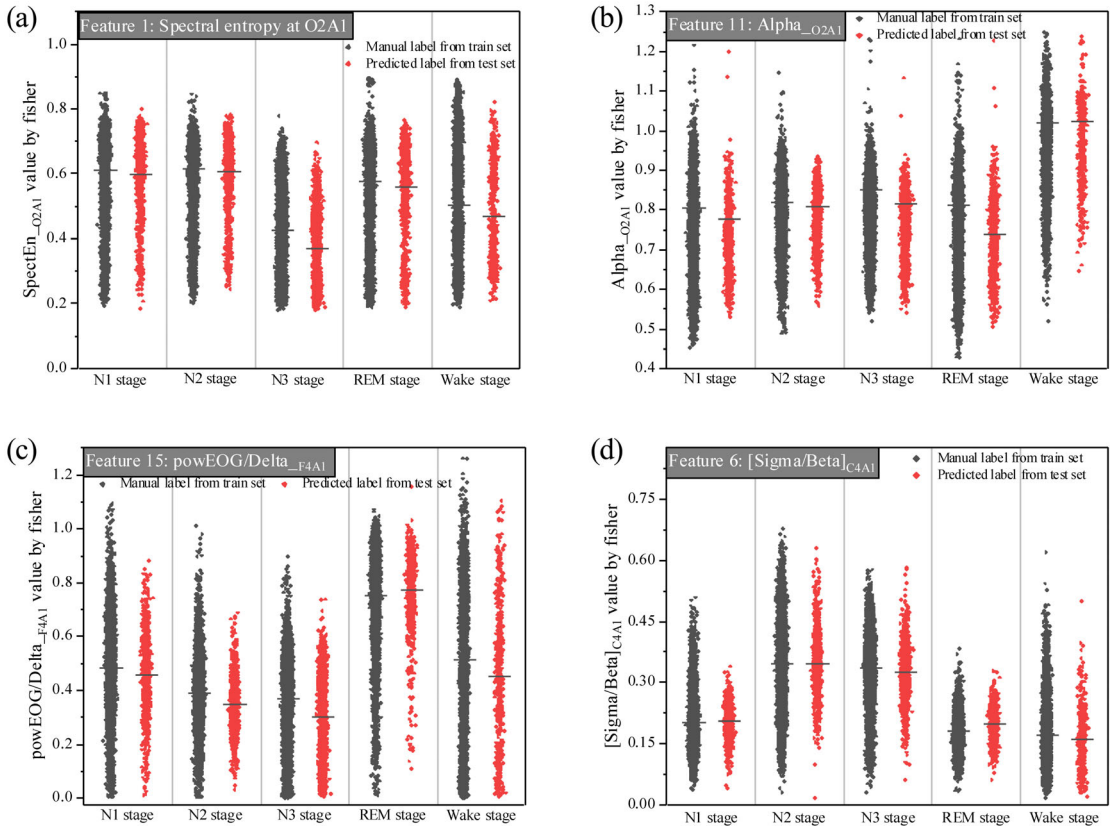

Fig. 8 The examples of feature distribution: (a) spectral entropy at O2A1; (b) Alpha at O2A1; (c) powEOG/Delta at F4A1; (d) Sigma/Beta at C4A1. Note: the horizontal line was represented by the average value

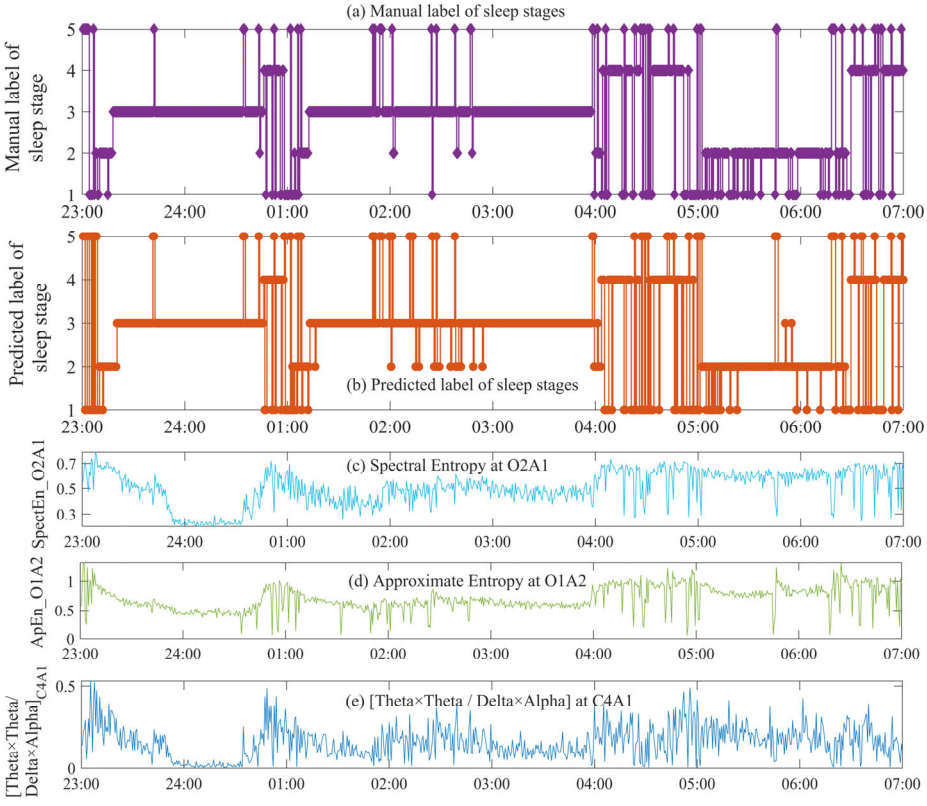

Fig. 9 The staging results of an example overnight. Note in (a) and (b): 1-N1 stage, 2-N2 stage, 3-N3 stage, 4-REM stage, 5-Wake stage

et al. 2017; Hassan and Subasi 2017). Researchers might be unable to conduct the sleep experiment limited by the experimental sites and equipment. Open-source data would help develop the types of algorithms, but meanwhile, two problems should be noted: (1) The process of collection was unknown. During the experiments, the operator could grasp the sleep information to eliminate the abnormal situation (the test probes fell off due to sweating or body movement, and occasional midnight insomnia) for the data purity. (2) The basic information of data was limited. Firstly, the data capacity depended on the size of the dataset, which greatly limited the input entrance. Secondly, many single-channel signals caused the loss of useful information at other channels. Finally, EEG signals were significantly different owing to the age (Mourtazaev et al. 1995) and BMI (Morselli et al. 2018), which would not be strictly collected and controlled in public data.

Feature input. The screened features belonged to four different properties in the study, and more than half of the features came from the frequency domain analysis which emphasized the frequency domain analysis, supported by Hassan and Subasi (2017). They considered that spectral features could effectively distinguish the various sleep phases. Among the spectral features proposed in the study, only three features (i.e., *Theta*$_{\_F4A1}$, *Alpha*$_{\_O2A1}$, and *Alpha*$_{\_F3A2}$) targeted the specific frequency domain. Theta waves appeared in the late N1 and REM period (Lomas et al. 2015), and similar mean values of N1 and REM were observed in Figure 10(a). Alpha waves occurred in the waking stage (Rechtschaffen and Kales 1968), and distribution at the waking stage was significantly higher in Figures 8(b) and 10(b). Fell et al. (1996) noted that *Le* distinguished well between N1 and N2 stage, but the phenomenon was obscured in Figure 10(c). Motamedi-Fakhr et al. (2014) explained that *Le* was still controversial in EEG analysis. In addition, AASM manual regarded spindles as one of the important features to recognize the N2 stage (Rechtschaffen and Kales 1968), and thus the study calculated the power density of spindles at the required range from 11 to 16 Hz. However, Figure 10(d) points out that the power density distribution of spindles wave in N2 period was not significantly different from the other phases. The reason might be that spindles vibrated rapidly in the range of 11–16 Hz but with a short duration (1–2 s), and were easily ignored by a 6 s long sliding window. The identification of spindles needed to be further investigated.

Šušmáková and Krakovská (2008) found that *Delta/Beta* was the best index for discrimination between N3 and the waking stage after comparing several ratios. In our study, seven features clearly distinguished deep sleep from other sleep stages shown in the Appendix (*SpectEn*$_{\_O2A1}$ (No. 1), *ApEn*$_{\_O1A2}$ (No. 2), [*Delta/Theta*]$_{C4A1}$ (No. 5), *WavPackEn*$_{\_O2A1}$

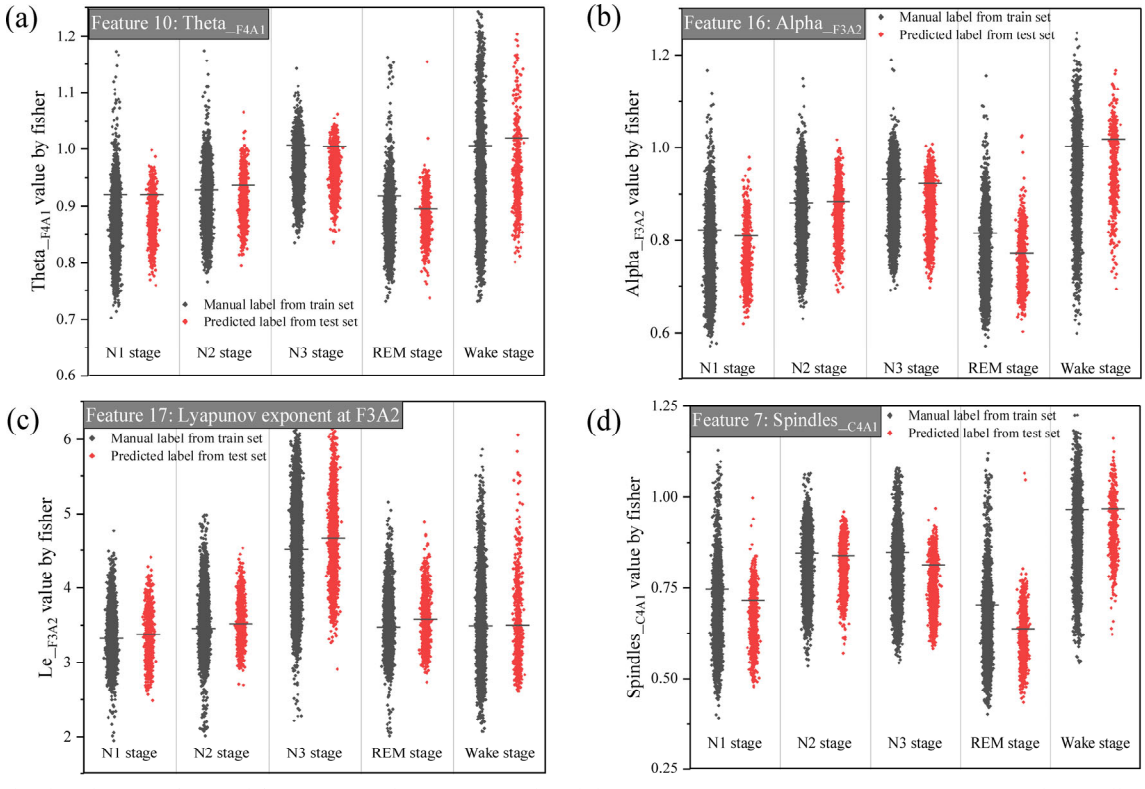

Fig. 10 The distribution of several features: (a) Theta at F4A1; (b) Alpha at F3A2; (c) Lyapunov exponent at F3A2; (d) Spindles at C4A1. Note: the horizontal line was represented by the average value

(No. 9), *SpectEn*_F3A2 (No. 13), *ApEn*_F4A1 (No. 14), *Le*_F3A2 (No. 17)), but *Delta/Beta* was filtered out. Two-class classification (Šušmáková and Krakovská 2008) task, which is different from the multi-classification task in this study, might lead to the biased results of the optimal features. Zhao et al. (2019) presented that *ApEn* and *SampEn* could achieve ideal staging, but FD was calculated simply with low accuracy. *SpectEn*, *ApEn*, and *Le* were the great nonlinear indices in this study, and *FD* was ranked in 62nd place, indicating the low contribution to sleep staging, which was consistent with Zhao et al. (2019).

The relation between the optimal feature sets and EEG channels was investigated. Firstly, features associated with the delta waves in deep sleep, such as *SWA* in time domain and *powEOG/Delta* in frequency domain, were both drawn out from the F4A1 channel, consistent with the opinion of delta waves frontally in adults (Lomas et al. 2015). Relative power density ratio of waves (i.e., [*Theta*×*Theta/Delta*× *Alpha*], [*Delta/Theta*], [*Sigma/Beta*]) were observed in the central part or sensorimotor cortex. Alpha generally appeared in the frontal region with obvious amplitude (Rechtschaffen and Kales 1968), similar to *Alpha*_F3A2 (the 11th feature) in Figure 8(b). Theta could be observed in the hippocampus (Lomas et al. 2015), and we drew the theta power from the frontal lobe. Other time-frequency and nonlinear features came from occipital and frontal lobes.

Furthermore, when comparing the accuracy of the three algorithms in Figure 6, though only the first ranked feature was inputted, the models were underperforming with an accuracy between 30% and 40%. For the sleep staging task, a single feature might be impossible to separate the phase satisfactorily, supported by Šušmáková and Krakovská (2008) that the classification error approached above 40% even for the best measure. Additionally, some features repetitively in the optimal sets in our research came from different EEG channels, inferring that the principle of selection for Fisher depended on the contribution to classification rather than the brain regions. To date, some studies still focused on the only single-channel, mostly with Parietal lobes-centered signal data (Zhao et al. 2019; Alickovic and Subasi 2018; Hassan and Bhuiyan 2016; Fu et al. 2021). Although single-channel offered irreplaceable convenience in data acquisition, much more useful information was lost. The trade-off between the convenience of acquisition and the performance of classifiers was required for further exploration.

Staging performance. Most studies directly adopted one certain selection algorithm, such as Fisher (Šušmáková and Krakovská 2008; Hassan and Subasi 2017). However, Gharbali et al. (2018) compared several algorithms for feature screening, and the close accuracy of mRMR, ReliefF, and Fisher was at 50%–85%. But, in our study, the data in the smoothing segment for Fisher was significantly superior to the other two algorithms, shown in Figure 6(b). Possible causes were as follows: (1) Sample source. EEG signals for sleep disorders were different from healthy people. Our dataset was derived from actual experiments conducted by us targeting the healthy young people, whereas healthy and sleep-disordered persons from a public dataset were mixed up in Gharbali et al. (2018). (2) Different features. 8.3% of the frequency domain features on EMG, 35.4% of the time domain features on EEG, and nonlinear features dominated by entropy measures were focused on in Gharbali et al.'s study (2018). Differences in feature inputs led to the final decision on the selection of the algorithm. (3) The channel of EEG signal. Gharbali et al. (2018) extracted data only from C3A2 without any information on other channels, but in this study, the number of features from C3A2 was relatively small. (4) The method of judgment. When comparing the three algorithms, Gharbali et al. (2018) brought all the filtered results into one classifier which could not clarify the changes in accuracy, while we trained a total of 456 classifiers to check the situation before and after the knee point.

Figure 11 shows the statistical results of the prior sleep staging models with the mean accuracy from 73.4% to 97.0%. Şen et al. (2014) derived all combinations of feature sets by contrasting the five different screening algorithms and obtained the staging model with the highest accuracy (97%). However, compared with our study, if the features filtered by the three algorithms were inputted into the random forest classifier simultaneously, the amount of entry features increased by three times, which was redundant for the classifier and generated the overfitting phenomenon. Vilamala et al. (2017) provided the accuracy for the specific sleep stage, especially only 44% at N1 stage. Alickovic and Subasi (2018) built an updated framework with SVM at the mean accuracy of 87.8%, but the sensitivity of the N1 period was only 49.62%. Šušmáková and Krakovská (2008) considered the most difficult task was to distinguish the N1 and REM stages, supported by Hassan and Subasi (2017). In our study, precision, recall, and F1 score values for N1 stage were between 71.7% and 74.7%. Possible reasons were as follows: (1) None of the extracted features in the optimal sets fully represented the N1 period in Figure 8 or in the Appendix, implying that the classification of N1 period could rely on the relationships of distributions among 17 features. This might lead to the poor performance of the N1 phase on the test set. (2) In many cases, the EEG signals in the N1 and REM phases were too close to differentiate without considering the EOG and EMG. In the experiment, however, EOG was easily disturbed by surroundings, and EMG probes occasionally appeared to loosen and dislodge, which unconsciously shrunk the features for distinction.

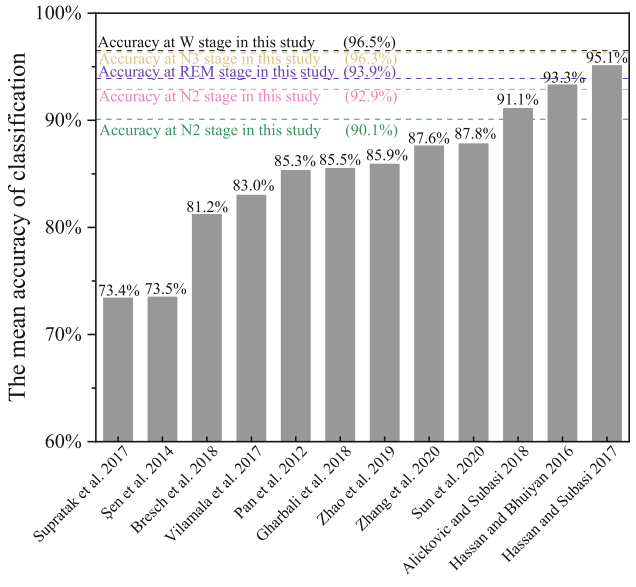

Fig. 11 The statistic results for prior models

Many studies only provided accuracy as an indicator, except that Motamedi-Fakhr et al. (2014), Alickovic and Subasi (2018), Grossi et al. (2021), and Supratak et al. (2017) gave the comprehensive performance of staging models, like confusion matrix. In our study, although the accuracy of classification ranged from 90.1% to 96.4%, the mean precision (82.5%), recall (83.1%), and F1 score (82.8%) were relatively low. Therefore, it was recommended that the multiple indices would evaluate the model in all directions. Moreover, deep learning was a more popular field in machine learning, in which staging tasks could take it into account as well. In particular, the whole process (from extraction to selection) might be handled by deep learning. Vilamala et al. (2017) attempted to involve the spectral EEG in the deep convolution neural network for calculation. Although N1 staging did not meet expectations, the relevant method inspired the subsequent study.

4.2 Limitation

The study tried to process and calculate the physiological signals features by adopting machine learning tools in the field of sleep environments. After collecting the signal information from different regions of the brain and adding the step of feature selection, the accuracy of the sleep staging model was improved, which laid the basis for a reliable physiological indicator for subsequent automatic sleep environment control. However, only the random forest classifier has been applied to the classification of fragmented data. In the future, we would consider using long-term data information by time-sequential data processing models such as LSTM. Furthermore, although EMG and EOG received attention in this study, they were accounted for a relatively small proportion in the optimal feature set. We would further focus on the data acquisition and feature calculation of EMG and EOG for the better performance of sleep staging model.

Applications. Although there was not much discussion on environment, the physiological characteristics presented in this study could provide a physiological basis for the regulation of automatic sleep environment control. As shown in Figure 12, the trained model and the physiological features can effectively carry out the sleep staging task. These features can be applied as part of the further sleep environment automatic control, allowing that the night environment responded in time with the feedback of occupants' sleep.

5 Conclusions

In order to precisely regulate the sleep environment, this study attempted to develop a data driven sleep staging model based on physiological signals through conducting sleep experiments and collecting raw data. The conclusions were as follows:

(1) In the sleep environment applications of the process on physiological signals, the different characteristics, including time domain, frequency domain, time-frequency domain and nonlinear characteristics, should be comprehensively

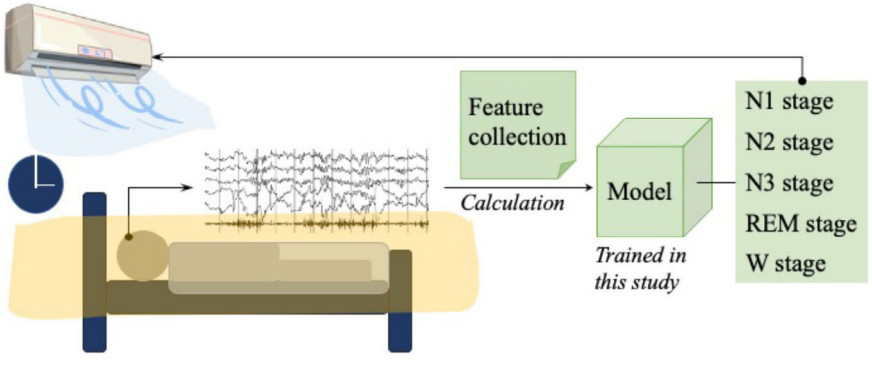

Fig. 12 The application of this work

considered. In addition, after comparing the preliminary accuracies of different screening algorithms, the combination of Fisher and Random Forest algorithms could effectively improve the performance of sleep staging models.

(2) Among the features screening in the optimal sets, 58.8% were mainly from frequency domain analysis, and 29.4% were based on nonlinearity. Time domain and time-frequency domain features accounted for 5.9% respectively. Additionally, the selected features came from different regions of brain, emphasizing that the raw signal of each channel would deliver the important information for sleep staging.

(3) In this study, a well-performing sleep staging model with an accuracy of 96.4% was established, and it would provide an effective and reliable feature basis for the automatic regulation and control of the sleep environment.

Electronic Supplementary Material (ESM): the appendix is available in the online version of this article at https://doi.org/10.1007/s12273-023-1049-6.

Acknowledgements

This work was supported by the National Key R&D Program of China (2022YFC3803201) and the National Natural Science Foundation of China (52078291).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Author contribution statement

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Ting Cao, Zhiwei Lian, Heng Du, Jingyun Shen, Yilun Fan and Junmeng Lyu. The first draft of the manuscript was written by Ting Cao and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

Acharya UR, Faust O, Kannathal N, et al. (2005). Non-linear analysis of EEG signals at various sleep stages. *Computer Methods and Programs in Biomedicine*, 80: 37–45.

Alickovic E, Subasi A (2018). Ensemble SVM method for automatic sleep stage classification. *IEEE Transactions on Instrumentation and Measurement*, 67: 1258–1265.

Berry RB, Quan SF, Abreu AR, et al. (2020). The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.6. Darien, IL, USA: American Academy of Sleep Medicine.

Bersagliere A, Achermann P (2010). Slow oscillations in human non-rapid eye movement sleep electroencephalogram: Effects of increased sleep pressure. *Journal of Sleep Research*, 19: 228–237.

Bresch E, Großekathöfer U, Garcia-Molina G (2018). Recurrent deep neural networks for real-time sleep stage classification from single channel EEG. *Frontiers in Computational Neuroscience*, 12: 85.

Cao T, Lian Z, Zhu J, et al. (2022). Parametric study on the sleep thermal environment. *Building Simulation*, 15: 885–898.

Carl C, Açık A, König P, et al. (2012). The saccadic spike artifact in MEG. *NeuroImage*, 59: 1657–1667.

Craik A, He Y, Contreras-Vidal JL (2019). Deep learning for electroencephalogram (EEG) classification tasks: A review. *Journal of Neural Engineering*, 16: 031001.

Ding C, Peng H (2003). Minimum redundancy feature selection from microarray gene expression data. In: Proceedings of the 2003 IEEE Bioinformatics Conference, Stanford, CA, USA.

Fan C, Chen M, Tang R, et al. (2022). A novel deep generative modeling-based data augmentation strategy for improving short-term building energy predictions. *Building Simulation*, 15: 197–211.

Fell J, Röschke J, Mann K, et al. (1996). Discrimination of sleep stages: A comparison between spectral and nonlinear EEG measures. *Electroencephalography and Clinical Neurophysiology*, 98: 401–410.

Fleiss JL, Cohen J (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33: 613–619.

Fu M, Wang Y, Chen Z, et al. (2021). Deep learning in automatic sleep staging with a single channel electroencephalography. *Frontiers in Physiology*, 12: 628502.

Gan VJL, Wang B, Chan CM, et al. (2022). Physics-based, data-driven approach for predicting natural ventilation of residential high-rise buildings. *Building Simulation*, 15: 129–148.

Ghahramani A, Castro G, Karvigh SA, et al. (2018). Towards unsupervised learning of thermal comfort using infrared thermography. *Applied Energy*, 211: 41–49.

Gharbali AA, Najdi S, Fonseca JM (2018). Investigating the contribution of distance-based features to automatic sleep stage classification. *Computers in Biology and Medicine*, 96: 8–23.

Grossi E, Valbusa G, Buscema M (2021). Detection of an autism EEG signature from only two EEG channels through features extraction and advanced machine learning analysis. *Clinical EEG and Neuroscience*, 52: 330–337.

Haskell EH, Palca JW, Walker JM, et al. (1981). The effects of high and low ambient temperatures on human sleep stages. *Electroencephalography and Clinical Neurophysiology*, 51: 494–501.

Hassan AR, Bhuiyan MIH (2016). A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features. *Journal of Neuroscience Methods*, 271: 107–118.

Hassan AR, Subasi A (2017). A decision support system for automated identification of sleep stages from single-channel EEG signals. *Knowledge-Based Systems*, 128: 115–124.

Imtiaz SA, Rodriguez-Villegas E (2014). A low computational cost algorithm for REM sleep detection using single channel EEG. *Annals of Biomedical Engineering*, 42: 2344–2359.

Jaffal I (2023). Physics-informed machine learning for metamodeling thermal comfort in non-air-conditioned buildings. *Building Simulation*, 16: 299–316.

Kononenko I (1994). Estimating attributes: Analysis and extensions of RELIEF. In: Bergadano F, De Raedt L (eds), Machine Learning: ECML-94. Berlin, Heidelberg: Springer.

Lan L, Lian Z (2010). Application of statistical power analysis - How to determine the right sample size in human health, comfort and productivity research. *Building and Environment*, 45: 1202–1213.

Lan L, Lian ZW, Lin YB (2016). Comfortably cool bedroom environment during the initial phase of the sleeping period delays the onset of sleep in summer. *Building and Environment*, 103: 36–43.

Lan L, Tsuzuki K, Liu YF, et al. (2017). Thermal environment and sleep quality: A review. *Energy and Buildings*, 149: 101–113.

Lomas T, Ivtzan I, Fu CHY (2015). A systematic review of the neurophysiology of mindfulness on EEG oscillations. *Neuroscience & Biobehavioral Reviews*, 57: 401–410.

Louis RP, Lee J, Stephenson R (2004). Design and validation of a computer-based sleep-scoring algorithm. *Journal of Neuroscience Methods*, 133: 71–80.

Magosso E, Provini F, Montagna P, et al. (2006). A wavelet based method for automatic detection of slow eye movements: A pilot study. *Medical Engineering & Physics*, 28: 860–875.

Malafeev A, Laptev D, Bauer S, et al. (2018). Automatic human sleep stage scoring using deep neural networks. *Frontiers in Neuroscience*, 12: 781.

Miyake S, Sato N, Akatsu J, et al. (1996). The effects of fluctuating room temperature on night-sleep in human. *The Japanese Journal of Ergonomics*, 32: 239–249.

Morselli LL, Temple KA, Leproult R, et al. (2018). Determinants of slow-wave activity in overweight and obese adults: roles of sex, obstructive sleep apnea and testosterone levels. *Frontiers in Endocrinology*, 9: 377.

Motamedi-Fakhr S, Moshrefi-Torbati M, Hill M, et al. (2014). Signal processing techniques applied to human sleep EEG signals—A review. *Biomedical Signal Processing and Control*, 10: 21–33.

Mourtazaev MS, Kemp B, Zwinderman AH, et al. (1995). Age and gender affect different characteristics of slow waves in the sleep EEG. *Sleep*, 18: 557–564.

Ngarambe J, Yun G, Lee K, et al. (2019). Effects of changing air temperature at different sleep stages on the subjective evaluation of sleep quality. *Sustainability*, 11: 1417.

Oropesa E, Cycon HL, Jobert M (1999). Sleep stage classification using wavelet transform and neural network. International Computer Science Institute.

Pan ST, Kuo C, Zeng J, et al. (2012). A transition-constrained discrete hidden Markov model for automatic sleep staging. *Biomedical Engineering Online*, 11: 52.

Pincus SM (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 88: 2297–2301.

Pop-Jordanova N, Pop-Jordanov J (2005). Spectrum-weighted EEG frequency ("brain-rate") as a quantitative indicator of mental arousal. *Prilozi*, 26: 35–42.

Rechtschaffen A, Kales A (1968). A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. US Department of Health, Education, and Welfare; National Institutes of Health.

Richman JS, Moorman JR (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology Heart and Circulatory Physiology*, 278: H2039–H2049.

Roffo G, Melzi S, Cristani M (2015). Infinite feature selection. In: Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile.

Rosso OA, Blanco S, Yordanova J, et al. (2001). Wavelet entropy: A new tool for analysis of short duration brain electrical signals. *Journal of Neuroscience Methods*, 105: 65–75.

Şen B, Peker M, Çavuşoğlu A, et al. (2014). A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms. *Journal of Medical Systems*, 38: 18.

Stanus E, Lacroix B, Kerkhofs M, et al. (1987). Automated sleep scoring: a comparative reliability study of two algorithms. *Electroencephalography and Clinical Neurophysiology*, 66: 448–456.

Sun C, Chen C, Li W, et al. (2020). A hierarchical neural network for sleep stage classification based on comprehensive feature learning and multi-flow sequence learning. *IEEE Journal of Biomedical and Health Informatics*, 24: 1351–1366.

Supratak A, Dong H, Wu C, et al. (2017). DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25: 1998–2008.

Šušmáková K, Krakovská A (2008). Discrimination ability of individual measures used in sleep stages classification. *Artificial Intelligence in Medicine*, 44: 261–277.

Tang R, Fan C, Zeng F, et al. (2022). Data-driven model predictive control for power demand management and fast demand response of commercial buildings using support vector regression. *Building Simulation*, 15: 317–331.

Vilamala A, Madsen KH, Hansen LK (2017). Deep convolutional neural networks for interpretable analysis of EEG sleep stage scoring. In: Proceedings of 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP).

Wu Y, Cao B, Hu M, et al. (2023). Development of personal comfort model and its use in the control of air conditioner. *Energy and Buildings*, 285: 112900.

Xu X, Zhu J, Chen C, et al. (2022). Application potential of skin temperature for sleep-wake classification. *Energy and Buildings*, 266: 112137.

Yang T, Bandyopadhyay A, O'Neill Z, et al. (2022). From occupants to occupants: A review of the occupant information understanding for building HVAC occupant-centric control. *Building Simulation*, 15: 913–932.

Zhang J, Yao R, Ge W, et al. (2020). Orthogonal convolutional neural networks for automatic sleep stage classification based on single-channel EEG. *Computer Methods and Programs in Biomedicine*, 183: 105089.

Zhang N, Cao B, Zhu Y (2023). An effective method to determine bedding system insulation based on measured data. *Building Simulation*, 16: 121–132.

Zhao D, Wang Y, Wang Q, et al. (2019). Comparative analysis of different characteristics of automatic sleep stages. *Computer Methods and Programs in Biomedicine*, 175: 53–72.

Zhou X, Xu L, Zhang J, et al. (2022). Development of data-driven thermal sensation prediction model using quality-controlled databases. *Building Simulation*, 15: 2111–2125.