Automatic Sleep Staging Algorithm Based on Random Forest and Hidden Markov Model

Junbiao Liu[1, 6], Duanpo Wu[2, 3], Zimeng Wang[2], Xinyu Jin[1, *], Fang Dong[4], Lurong Jiang[5]

and Chenyi Cai[6]

Abstract: In the field of medical informatics, sleep staging is a challenging and time-consuming task undertaken by sleep experts. According to the new standard of the American Academy of Sleep Medicine (AASM), the stages of sleep are divided into wakefulness (W), rapid eye movement (REM) and non-rapid eye movement (NREM) which includes three sleep stages (N1, N2 and N3) that describe the depth of sleep. This study aims to establish an automatic sleep staging algorithm based on the improved weighted random forest (WRF) and Hidden Markov Model (HMM) using only the features extracted from double-channel EEG signals. The WRF classification model focuses on reducing the bias of the imbalance data, while the HMM model focuses on improving the detection rate of sleep staging through the relationship between adjacent sleep stages. In particular, the improved weighted RF classification model can increase the recognition rate of the N1 stage. In addition, the method of removing features with low variance is used to select meaningful and contributing feature parameters for model training. This is an innovative content of this paper. The sleep EEG data are first segmented into 30 s epochs, and the feature parameters of the epoch data are extracted from the double-channel by applying continuous wavelet packet transform (WPT). Each epoch is then segmented into 29 subepochs of 2 s long with 1 s overlap, and the frequency domain features and statistical features of each subepoch are extracted. The performance of the proposed method is tested by evaluating the accuracy (AC), Kappa coefficient, Recall (R), Precision (P) and F1-score (F1). In the Sleep-EDF database, the overall AC and Kappa coefficient obtained by WRF

---

[1]College of Information Science & Electronic Engineering, Zhejiang University, Hangzhou, China.

[2]School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, China.

[3]Zhejiang Provincial Key Laboratory of Information Processing, Communication and Networking, Hangzhou, China.

[4]College of Information and Electric Engineering, Zhejiang University City College, Hangzhou, China.

[5]School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou, China.

[6]Hangzhou Neuro Science and Technology Co., Ltd., Hangzhou, China.

*Corresponding Author: Xinyu Jin. Email: jinxy@zju.edu.cn.

are 93.20% and 0.890, respectively using the subject-non-independent test. In the 10 sc* and 10 st* Sleep-EDF Expanded database, the overall AC and Kappa coefficient obtained by proposed method are 91.97% and 0.874, respectively using the subject-independent test. The best AC and Kappa coefficient of single subject can reach 96.3% and 0.912, respectively. Experimental results show that the performance of the proposed method is competitive with the most current methods and results, and the recognition rate of N1 stage is significantly improved.

Keywords: Sleep staging, wavelet packet transform, random forest, HMM.

1 Introduction

Sleep is an important physiological activity of the human body, and sufficient sleep will guarantee the work of the next day [Liu, Hay and Faught (2013); Bertisch, Pollock, Mittleman et al. (2018)]. After continuous exploration, the researchers found that sleep contains specific physiological laws, and many disease-related information can be obtained from sleep electroencephalogram (EEG) [Hood, Wilson, Gorenz et al. (2018); Olsson, Ärlig, Hedner et al. (2018); Gonzalez, Small, Cases et al. (2018)]. Therefore, studying the information contained in sleep EEG signals and analyzing the changes of sleep cycle is of great significance for the study of sleep-related diseases.

Overnight polysomnography (PSG) with EEG is the primary diagnostic test for evaluating patients with sleep problems [Malhotra, Kirsch, Kristo et al. (2018)]. Typically, sleep specialists use PSG for clinical diagnosis of sleep staging. PSG contains a set of signals, including EEG, electromyogram (EMG), electrocardiogram (ECG) and electrooculogram (EOG). These signals are recorded by sensors connected to different parts of the body [Tarek, Sahbi, Perrine et al. (2015)]. Sleep experts or doctors will classify these epochs into different sleep stages according to the guidelines developed by Rechtschaffen and Kales (R&K) or the classification manual proposed by the American Academy of Sleep Medicine (AASM) [da Silveira, Kozakevicius and Rodrigues (2017)]. This process is called sleep stage scoring or sleep staging. Basing on the R&K rules, sleep recordings can be classified into six sleep stages: wakefulness (W), non-rapid eye movement (NREM) sleep stage 1 (N1), NREM sleep stage 2 (N2), NREM sleep stage 3 (N3), NREM sleep stage 4 (N4), and rapid eye movement (REM) [Wolpert (1969)]. According to the standardization rules of AASM in recent years, N4 has been merged into the N3 sleep stage, so this paper will also study five sleep stages [Berry, Brooks, Gamaldo et al. (2017)].

EEG is a kind of bioelectrical signal with randomness, time variant, non-stationarity and non-linearity. Accurate sleep staging is a crucial part of the diagnosis and treatment of some disease [Bachmann, Päeske, Kalev et al. (2018)]. So far, doctors have been accustomed to visually observe brain waves for sleep stage marking, which takes a lot of time. Therefore, an effective automatic sleep stage classification algorithm will alleviate the doctor's scoring work.

In order to classify the sleep stage of EEG signals, feature extraction from sleep EEG signals is one of the most critical technique in the process of sleep staging. Some researchers used feature extraction methods that combined time and frequency domain

features, but most studies only extracted time-domain features or frequency-domain features [Li, Wang, Luo et al. (2017); Hernández, Trujillo, Z-Flores et al. (2018); Colominas, Jomaa, Jrad et al. (2017); TM and M (2018)]. Cabir et al. [Cabir and Murat (2010)] adopted principal component analysis (PCA) dimensional reduction techniques and compared features in time domain, frequency domain and the combination of both in perforemance evaluation of sleep staging. According to the experimental results, time-domain features alone provided the worst classification while the mixed features provided the best classification performance. Salih et al. [Salih, Kemal and Yosunkaya (2010)] proposed a new hybrid system comprising of Welch spectral analysis, k-means clustering based feature weighting and classifier algorithms (k-NN classifier and decision tree classifier) to automatically score the sleep stages. A total samples of 35-hours EEG signals collected from five male subjects were used to achieve 82% classification accuracy (AC) through this system. Fraiwan et al. [Fraiwan, Lweesy and Khasawneh (2012)] researched sleep staging algorithm using three time-frequency techniques (Choi-Williams distribution, continuous wavelet transform and Hilbert-Huang Transform) and entropy measures based on the sleep EEG, the experimental results showed that the continuous wavelet transform analysis method provided the best performance with accuracy of 83%.

Although the state-of-the-art research has fully captured the time domain and frequency domain information of sleep EEG signals, it still has certain limitations in feature extraction. Memar et al. [Memar and Faradji (2017)] used the bandpass filter to decompose 8 subbands waves (δ, θ, α, σ, $\beta 1$, $\beta 2$, $\gamma 1$ and $\gamma 2$) from each EEG epoch. Baha et al. [Baha, Musa, Abdullah et al. (2014)] decomposed the rhythm bands of EEG data by 5-level discrete wavelet transform. Since the wavelet transform only decomposes the low-frequency bands of the signal and the high-frequency bands are ignored, the large amount of information about the sleep cycle contained in the EEG signals cannot be fully explored. Fortunately, wavelet packet transform (WPT) [Shahnaz, Sarker and Paul (2017); Reddy, Fahim, Nikhil et al. (2018)] can provide more detailed EEG signal decomposition including low frequency part and high frequency part, so it can perform better time-frequency localization analysis on EEG signals containing a large amount of medium and high frequency information.

The choice of classification methods greatly affects the performance of sleep staging. In recent years, the application of deep learning methods in biomedical signal processing has become more and more common. For example, the convolutional neural networks (CNN), the one-dimensional convolutional neural networks (1D-CNN), recurrent neural networks (RNN) and the fast discriminative complex-valued convolutional neural network (FDCCNN) have been proposed recently to deal with the sleep stage classification tasks [Sokolovsky, Guerrero, Paisarnsrisomsuk et al. (2019); Malek, Melgani and Bazi (2018); Yulita, Fanany and Arymuthy (2017); Zhang and Wu (2017)]. In addition, the method of feature extraction using deep learning instead of manual feature extraction is also explored for automatic sleep staging. However, Ghimatgar et al. [Ghimatgar, Kazemi, Helfroush et al. (2019)] proposed that deep learning is limited in the study of automatic sleep staging. Firstly, deep learning needs a lot of data to train complex models, which is less meaningful for the analysis of a small number of special patients clinically. Secondly, compared to

traditional methods, deep learning usually takes hours or even days for training which requires huge computing resources. Thirdly, no relevant research can prove the clinical significance of auto-learning characteristics from a physiological point of view, and it cannot explain the role of these features in each sleep stage.

In this paper, an innovative method for automatic sleep EEG staging based on improved weighted random forest (WRF) and hidden markov model (HMM) is proposed. Compared to traditional shallow learning methods, deep learning utilizes multiple layers of linear and nonlinear processing units to learn hierarchical representations or features from input data. Compared with using the WRF classifier alone, the algorithm combining WRF with HMM can take the sleep stage state of neighboring epoch into account, which can smooth the classification result and improve the accuracy of the model. After transforming the preprocessed EEG signals by WPT, long-term features which are extracted by decomposing the EEG data into different 30 s epoch data and short-term features which are extracted by segmenting the 30 s epoch data into 2 s epoch data with 1 s overlap are used to train a sleep staging model by WRF. And then, HMM is adopted to smooth the sleep staging model and reduce misclassification.

The rest of the paper is organized as follows. In Section 2, materials and methods are described. The experimental results are given in Section 3. Discussion is described in Section 4 and Section 5 is the conclusions.

2 Materials and methods

The paper proposes a sleep staging method based on WRF and HMM, following the step of preprocessing, feature extraction, WRF training, WRF testing and HMM smoothing. Fig. 1 shows the procedure of the method. The algorithm simulation in this paper is done on python 3.7 platform, and EDF data waveform can be viewed through EDFbrowser.
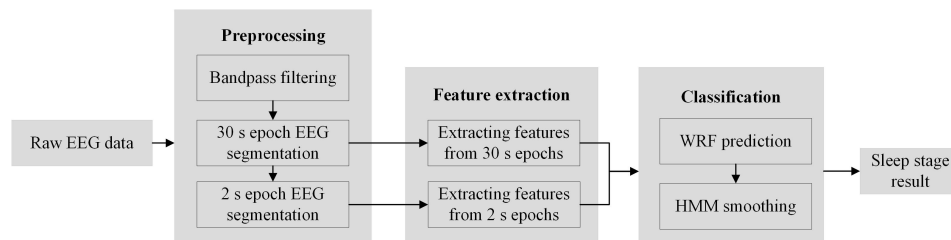

Figure 1: Flowchart of the automatic sleep stage classification

2.1 Sleep datasets

The experimental data in this study comes from the Sleep-EDF database (website is `https://www.physionet.org/physiobank/database/sleep-edf/`) and the Sleep-EDF Expanded database (website is `https://www.physionet.org/physiobank/database/sleep-edfx/`) which are provided by Kemp [Kemp, Zwinderman, Tuk et al. (2000)]. The Sleep-EDF database includes 8 EEG files recording from 4 males and 4 females between the age of 21 and 35. These 8 EEG recordings are

divided into healthy group (sc4002e0, sc4012e0, sc4102e0, and sc4112e0, marked as sc*) and mild sleep disorder group (st7022j0, st7052j0, st7121j0, and st7132j0, marked as st*) according to sleep status. The 4 sc* recordings were obtained in 1989 from ambulatory healthy volunteers during 24 hours in their normal daily life, using a modified cassette tape recorder. The 4 st* recordings were obtained in 1994 from subjects who had mild difficulty falling asleep but were healthy, during a night in the hospital. Each EEG recording contains 1 horizontal EOG and 2 EEG channels called Fpz-Cz and Pz-Oz with the sampling rate of 100 Hz. All of these sleep EEG data are annotated by experts with different stages according to R&K guideline, which are named as W, REM, N1, N2, N3 and N4 and unscored. However, sleep EEG is now commonly staged according to the AASM guideline in clinical studies, including W, N1, N2, N3 and REM. Therefore, this paper states that the class N3 includes two stages in this paper, N3 and N4.

Sleep-EDF Expanded database is the extended version of the Sleep-EDF database, which was updated to version 1 in 2013 with a total of 61 subjects, including 39 sc* files and 22 st* files. The 39 SC* files were obtained from 10 healthy males and 10 healthy females aged from 25 to 34, without any sleep-related medication [Mourtazaev, Kemp, Zwinderman et al. (1995)]. The ST* files were obtained from a study of sleep effects in 7 males and 15 females who had a mild difficulty falling asleep. Same as Sleep-EDF database, the sampling rate of these signals is 100 Hz. In this paper, channel Fpz-Cz and channel Pz-Oz are analyzed.

2.2 EEG preprocessing

The frequency range of Sleep-EDF database and Sleep-EDF Expanded database EEG is generally 0.1-50 Hz. In this study, a 5-th order IIR Butterworth bandpass filter with the frequency range of 0.5-32 Hz is used to remove noise and artifacts from EEG signals. Then, the filtered EEG signals will be segmented into 30 s epochs without overlap.

According to the above division method, 15172 sleep epochs are obtained from the 8 EEG recordings in which there are 8039, 1609, 604, 3621, 1299 epochs of W, REM, N1, N2 and N3 stage respectively. Tab. 1 shows the epoch numbers for different sleep stages in the 3 EEG groups (Sleep-EDF database, 20 sc* files in the Sleep-EDF Expanded database and 10 sc* + 10 st* files in the Sleep-EDF Expanded database) used in this study.

Table 1: The number of epochs of various sleep stages

Databases	Subjects	W	REM	N1	N2	N3	Total
Sleep-EDF	4 sc* + 4 st*	8039	1609	604	3621	1299	15172
Sleep-EDF Expanded 1	20 sc*	37410	3756	1240	9200	2981	54587
Sleep-EDF Expanded 2	10 sc*+ 10 st*	19973	3761	2002	8910	2758	37404

2.3 Wavelet packet transform

The EEG signal in W stage is characterized by low amplitude β wave in state of eyes closed and α wave in state of eyes opened. The EEG signal in N1 stage is characterized by the

presence of vertex sharp and θ wave with high amplitude. Stage N2 is characterized by the appearance of spindles and K-complex. The deepest sleep stage N3 is characterized by δ wave with high amplitude. The REM stage is characterized by low amplitude sawtooth wave. Therefore, WPT is used to process EEG signals, and the time-frequency features of different spontaneous EEG rhythms(δ wave, θ wave, α wave, β wave, spindles and K-complex) are analyzed in this paper. Choosing db6 as wavelet basis function to transform each 30 s epoch in WPT. The paper decomposes the signal into 7 frequency bands: δ wave (1.6-4 Hz), θ wave (4-8 Hz), α wave (8-13 Hz), $\beta 1$ wave (13-22 Hz), $\beta 2$ wave (22-32 Hz), spindles (12-14 Hz) and K-complex wave (0.5-1.5 Hz). The wavelet transform only decomposes the low-frequency part of the signal step by step, and does not perform any processing on the high-frequency part of the signal. The WPT further decomposes the low-frequency component and the high-frequency component of the previous stage.

Let $\varphi(n)$ and $\Psi(n)$ be scale function and wavelet function, respectively, so that:

$$\begin{cases} \Psi_0(n) = \varphi(n) \\ \Psi_1(n) = \Psi(n) \end{cases} \tag{1}$$

$$\begin{cases} \Psi_{2i}^{j,k} = \displaystyle\sum_{k=-\infty}^{+\infty} h(m)\Psi_i^{j-1,2k-m}(n) \\ \Psi_{2i+1}^{j,k} = \displaystyle\sum_{k=-\infty}^{+\infty} g(m)\Psi_i^{j-1,2k-m}(n) \end{cases} \tag{2}$$

where i is the node number, j is the level of decomposition. $h(m)$ and $g(m) = (-1)^{1-m}h(1-m)$ are a pair of quadrature mirror filters. $\Psi_{2i}^{j,k}$ and $\Psi_{2i+1}^{j,k}$ are wavelet packet basis functions based on two-scale equations.

The wavelet transform coefficients of the EEG signal $x(n)$ at j-th level and k-th point are computed via the following recursive equations:

$$\begin{cases} C_{2i}^j(k) = \displaystyle\sum_m h(m)C_i^{j-1}(2k-m) \\ C_{2i+1}^j(k) = \displaystyle\sum_m g(m)C_i^{j-1}(2k-m) \end{cases} \tag{3}$$

For the M-layer wavelet packet decomposition, the EEG signal $x(n)$ have 2^M sample points, then $x(n)$ can be reconstructed by the Eq. (4).

$$x(n) = \sum_{i=0}^{2^{j-1}-1}\sum_{k=0}^{2^{M-j}-1} C_{2i}^j \Psi_{2i}^{j,k}(n) + \sum_{i=0}^{2^{j-1}-1}\sum_{k=0}^{2^{M-j}-1} C_{2i+1}^j \Psi_{2i+1}^{j,k}(n) \tag{4}$$

For the 5-layer wavelet packet decomposition of the signal, there are 2^5 nodes in the 5-th layer. In the first layer decomposition, the EEG signal (0, 0) is decomposed into a low frequency part (1, 0) and a high frequency part (1, 1), in the second layer decomposition,

the signal of (1, 0) is decomposed into a low frequency part (2, 0) and a high frequency part (2, 1), and the signal of (1, 1) is decomposed into a low frequency part (2, 3) and high frequency part (2, 2), the third to fifth layers are sequentially decomposed according to this law. In order to show the law of wavelet packet decomposition more intuitively, 0 is used to represent the low-frequency part obtained after each decomposition, and 1 is used to indicate the high-frequency part obtained by each decomposition, and then a binary number b_n which contains n digits can be defined, which means it is decomposed into n layers. For example, (1, 0) can be expressed as 0, and (2, 2) can be expressed as 11. If the low-frequency portion of the upper layer is decomposed, the low-frequency portion is decomposed to the left side, and if the high-frequency portion of the upper layer is decomposed, the high-frequency portion is decomposed to the left side. So the frequency distribution of the nodes of the last layer is not in order. In this paper, the EEG signal is decomposed into five layers. The fifth layer contains 32 nodes. According to the above-mentioned wavelet packet decomposition law, the frequencies of these 32 nodes are arranged in order from low to high as: (5, 0), (5, 1), (5, 3), (5, 2), (5, 6), (5, 7), (5, 5), (5, 4), (5, 12), (5, 13), (5, 15), (5, 14), (5, 10), (5, 11), (5, 9), (5, 8), (5, 24), (5, 25), (5, 27), (5, 26), (5, 30), (5, 31), (5, 29), (5, 28), (5, 20), (5, 21), (5, 23), (5, 22), (5, 18), (5, 19), (5, 17), (5, 16). The sampling frequency of the EEG signal in this paper is 100 Hz, and the rhythm waves obtained by WPT are shown in Tab. 2.

Table 2: Rhythm waves extracted from WPT

Nodes	Bands/Hz	Marks	Rhythm waves
(4, 0)	0-3.125	$\sqrt{}$	δ wave
(4, 1)	3.125-6.25	$\sqrt{}$	θ wave
(3, 1)	6.25-12.5	$\sqrt{}$	α wave
(3, 2)	18.75-25	$\sqrt{}$	$\beta1$ wave
(3, 3)	12.5-18.75	$\sqrt{}$	$\beta1$ wave
(3, 6)	25-31.25	$\sqrt{}$	$\beta2$ wave
(5, 0)	0-1.5625	$\sqrt{}$	K-complex
(5, 12)	12.5-14.0625	$\sqrt{}$	spindles

2.4 Feature extraction

Sun et al. [Sun, Jia, Goparaju et al. (2017)] claimed that the duration of each epoch (30 s) is relatively long compared to transient events, including sleep spindle waves and K complex waves. So the spectral information of the epoch may not be fully extracted. Therefore, each 30 s epoch will be segmented into 29 subepochs of 2 s with 1 s overlap. The rhythms of the different frequency bands of the wavelet packet decomposition have been obtained, and the overall features of the 30 s epochs and the statistical features of the 29 subepochs are respectively calculated to represent the feature differences between the EEG signals from different sleep stages. The features extracted from rhythm waves are presented in Tab. 3.

In this study, $x(n)$ represents a kind of rhythm wave in 30 s epoch. Since the sampling rate of the EEG signal is 100 Hz, the length of $x(n)$ is $N = 3000$.

2.4.1 Time-domain features

- Standard deviation(SD): SD is a time domain feature that measures the amount of variation of the EEG data which can be calculated by Eq. (5).

$$SD = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} (x(n) - \bar{x})^2} \tag{5}$$

where $\bar{x}$ is the mean of all samples in $x(n)$, expressed as Eq. (6).

$$\bar{x} = \frac{1}{N} \sum_{n=1}^{N} x(n) \tag{6}$$

- Skewness (Skew): Skew is a time domain feature used to measure the degree of asymmetry of data distribution which can be calculated by Eq. (7).

$$Skew = \frac{\frac{1}{N} \sum_{n=1}^{N} (x(n) - \bar{x})^3}{[\frac{1}{N-1} \sum_{n=1}^{N} (x(n) - \bar{x})^2]^{3/2}} \tag{7}$$

where $\bar{x}$ is the mean of $x(n)$, and $N = 3000$ is the number of samples in $x(n)$.

- Kurtosis (Kur): Kur is used to measure the level of the peak of the data frequency distribution curve.

$$Kur = \frac{\frac{1}{N} \sum_{n=1}^{N} (x(n) - \bar{x})^4}{[\frac{1}{N} \sum_{n=1}^{N} (x(n) - \bar{x})^2]^2} \tag{8}$$

where $\bar{x}$ is the mean of $x(n)$, and $N = 3000$ is the number of samples in $x(n)$.

- Line length: The measurement of the amplitude and frequency of EEG oscillation can be realized by line length. The line length is represented by the variable L, and the formula is as follows.

$$LL = \frac{1}{N-1} \sum_{n=1}^{N-1} |x(n+1) - x(n)| \tag{9}$$

- Hjorth Parameters: The Hjorth parameters, including Hjorth mobility (HM) and Hjorth complexity (HC), provide dynamic temporal information of the EEG signal. They are calculated by Eqs. (10)-(11).

$$HM = \frac{\sigma_{x'}}{\sigma_x} \tag{10}$$

$$HC = \frac{\sigma_{x''}/\sigma_{x'}}{\sigma_{x'}/\sigma_x} \tag{11}$$

where σ_x, $\sigma_{x'}$ and $\sigma_{x''}$ are the SD of the $x(n)$ and its first and second derivatives, $x(n)'$ and $x(n)''$, respectively.

- Katz fractal dimension (KFD): KFD is used to describe the geometric scale characteristics of EEG signals and is an indicator for measuring signal complexity.

$$KFD = \frac{logN}{logN + log(d/L)} \tag{12}$$

where L refers to the sum of distances between two successive points, and d represents the maximum Euclidean distance between the first point and any other point on the waveform.

- Hurst exponent (HE): HE is used to study the scaling properties of a signal. The HE of $x(n)$ is defined as follows:

$$HE = \frac{log(R/SD)}{logN} \tag{13}$$

where R represents the biggest difference, that is, the difference between the maximum and the minimum of the deviation of $x(n)$.

2.4.2 Frequency-domain features

The frequency-domain features extracted by the proposed method include two parts: the overall features of the 30 s epochs and the statistical features of the 2 s subepochs. the overall features of the 30 s epochs include the ratio of δ, θ, α, $\beta1$, $\beta2$, K-complex and spindles energy to total energy, the ratio of α energy to θ energy, and the ratio of δ energy to θ energy. The statistical features of the 2 s subepochs include the 95th percentile, min, mean and variance of E_i/E_{all}, where E_i is the energy of $i\text{-}th$ subepoch, $i = 1, 2, ..., 29$ and the E_{all} is the overall energy of the 30 s epoch.

By this feature extraction method, a 109-dimensional feature vector is obtained for one channel of each EEG signal epoch.

Table 3: Features extracted from one EEG channel

Types	Feature parameters	Number
	Mean/SD/Skew/Kur	8×4
	Line length ($\frac{1}{(N-1)} \sum_{i-1}^{N} |x_{i+1} - x_i|$)	8×1
Time	HM ($\frac{\sigma_{x'}}{\sigma_x}$)	8×1
	HC ($\frac{\sigma_{x''}/\sigma_{x'}}{\sigma_{x'}/\sigma_x}$)	8×1
	KFD ($\frac{logN}{logN + log(d/L)}$)	8×1
	HE ($\frac{log(R/SD)}{logN}$)	8×1
Frequency	Ratio of 7 waves energy to total energy	7×1
	Ratio of α energy to θ energy	1
	Ratio of δ energy to θ energy	1
	Mean, min, variance and 95th percentile of $\{E_i/E_{all}\}$, $i = 1, 2, ..., 29$	7×4

2.5 Removing features with low variance

As can be seen from Tab. 3, each channel has 109 features, so the double-channel epoch contains 218 features. In order to eliminate the features with lower contribution, reduce the dimension of the feature vector, and reduce the difficulty of model training, the method of removing features with low variance is adopted in this study. The method of removing features with low variance is simple in principle and fast in operation. Assuming that the features vector of an epoch is F, each column in F is a feature, and the variance corresponding to each feature in F is calculated, and a threshold is set, if the variance is lower than this threshold, the feature is removed.

2.6 Weighted random forest algorithm

It can be seen from Tab. 1 that the number of epochs in W stage is much higher than that in other stages (N1, N2, N3 and REM) and the number of N1 epochs is the least. In view of the problem of data imbalance in sleep staging, the typical RF classification method is optimized in this paper, and a WRF model is established to improve the recognition rate of W, NREM and REM stage.

It is assumed that the number of feature vectors for each class in the training set is NUM_i, $i = 1, 2, ..., 5$, the number of feature vectors of all classes is NUM_{all}, the number of class $N = 5$, and the weight of each class W_i can be calculated by Eq. (14)

$$W_i = \frac{NUM_{all}}{N * NUM_i} \tag{14}$$

RF classifier proposed by Leo Breiman contains multiple decision trees and its output category is determined by the highest number of votes among all trees' results [Breiman (2001)]. Through weighted bootstrap resampling technique, a new training sample set is generated by randomly selecting k samples repeatedly from the original training sample

set M, and then a RF is generated from k individual decision tree classifier. The essence of RF classifier is an improvement of the decision tree algorithm. Multiple decision trees are merged together, and the establishment of each tree depends on an independently extracted sample. Each tree in the forest has the same distribution, and the classification error depends on each tree's classification ability and their correlation.

In this paper, the subject-independent test is performed on the sleep-EDF database, and over-fitting is avoided by the 10-fold cross-validation method. The feature data extracted from each 30 s epoch EEG listed in Tab. 3 are trained through the WRF method after removing features with low variance.
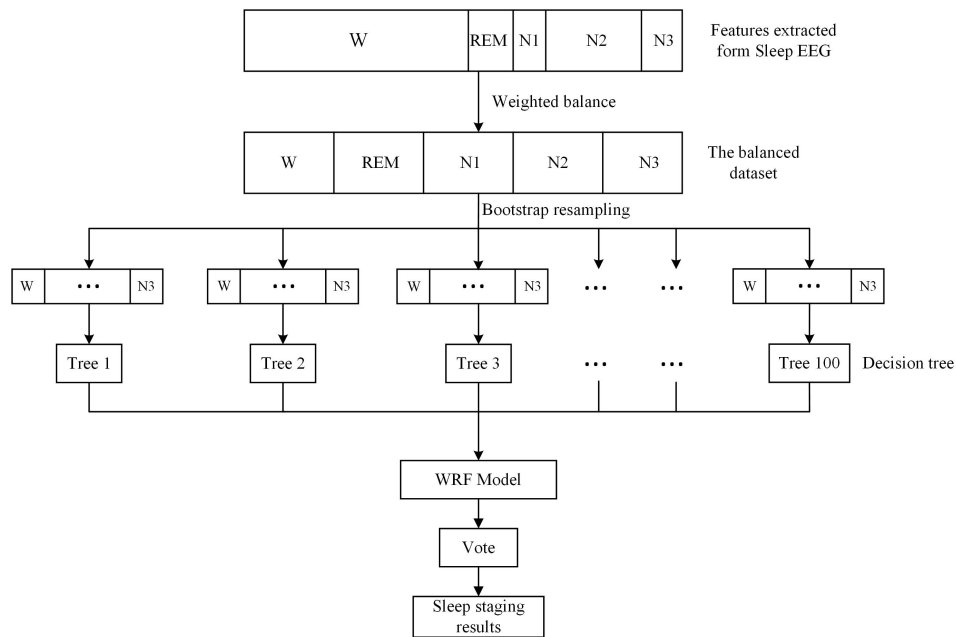

Figure 2: The weighted random forest model

Fig. 2 shows the structure of the WRF classifier and the different steps of the training procedure. For each of the 100 trees, the training features are fed to the bootstrapping and then are selected. Then all decision trees are integrated into a WRF classification model that automatically balances the data, and the output of each tree is fed to the vote part to obtain the classification output.

2.7 HMM smoothing method

The sleep state and time series in the EEG data are closely related. The WRF classification result has no correlation from the time axis and may appear from the W stage to the N3 stage, which will lead to the wrong sleep stage switching. While HMM can consider the neighboring sleep stage handoff information and improve the classification accuracy of sequence sleep EEG signal. So this paper uses HMM to smooth the output of the WRF.

HMM is a probabilistic model of time series which describes the process of generating the sequence of unobservable states randomly from a hidden markov chain, and then generating the sequence of observations randomly from each state. There is an important property of HMM called non-aftereffect property, which means the current state of system is related only to the past state, not to the future state of system.

Generally, HMM is mainly used to solve three kinds of problems: evaluation problem, decoding problem and learning problem. The HMM decoding problem refers to find the most likely path through a model given an observed sequence. Smoothing by HMM's Viterbi algorithm needs the stage transition matrix which describes the probability of transition from previous state to the next state and the emission matrix for the observations which describe the probability distribution of the stages assigned by WRF classifier as a function of the underlying true stage.

In this work, the state transition matrix is obtained by calculating the probability of the handoff relation of neighboring epoch's sleep stage state. After obtaining the sleep state of a single subject in the test set after passing through the random forest, we get the emission matrix. The accuracy of the model is improved by using the Viterbi algorithm.

2.8 Evaluation criteria

In order to evaluate the effectiveness of the proposed method, confusion matrix (CM) will be calculated to show the difference between the result given by the proposed method and the experts' marks as shown in Eq. (15). S_{ij} represents the number of epochs that are marked to be sleep stage i and are classified as sleep stage j. The value of i or j means the epoch is in sleep stage W, REM, N1, N2 and N3 when it is equal to 1, 2 , 3, 4 and 5, respectively.

$$CM = \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\ S_{21} & S_{22} & S_{23} & S_{24} & S_{25} \\ S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\ S_{41} & S_{42} & S_{43} & S_{44} & S_{45} \\ S_{51} & S_{52} & S_{53} & S_{54} & S_{55} \end{bmatrix} \tag{15}$$

Basing on the Eq. (15), Cohen's Kappa coefficient which can describe the agreement between the proposed method and the experts' marks [Sara, Leila, Ina et al. (2018); Sabil, Vanbuis, Baffet et al. (2017)] can be calculated from Eq. (16) to Eq.(18).

$$Kappa \quad = \quad \frac{P_0 - P_e}{1 - P_e} \tag{16}$$

$$P_0 \quad = \quad \frac{\sum_{i=1}^{5} S_{ii}}{\sum_{i=1}^{5} \sum_{j=1}^{5} S_{ij}} \tag{17}$$

$$P_e \quad = \quad \frac{\sum_{i=1}^{5} (\sum_{j=1}^{5} S_{ij} \sum_{j=1}^{5} S_{ji})}{(\sum_{i=1}^{5} \sum_{j=1}^{5} S_{ij})^2} \tag{18}$$

AC, Recall (R), Precision (P) and F1-score(F1) can also be calculated from the CM to show the performance of the proposed method as shown in Eq. (19) to Eq. (22). Among them, AC represents the percentage of correctly classified number of epochs to the number of all testing epochs. In this paper, the AC, R and F1-score are used as indicators to evaluate the classification effect of each sleep stage. AC and Kappa coefficient are used to evaluate the general performance of all stages and measure the agreement between the algorithm prediction results and the expert score results.

$$AC = \frac{\sum_{i=1}^{5} S_{ii}}{\sum_{i=1}^{5} \sum_{j=1}^{5} S_{ij}} \tag{19}$$

$$R = \frac{S_{ii}}{\sum_{j=1}^{5} S_{ij}} \tag{20}$$

$$P = \frac{S_{ii}}{\sum_{j=1}^{5} Sji} \tag{21}$$

$$F1 = \frac{2 \times R \times P}{R + P} \tag{22}$$

2.9 Test scheme

According to the different ways of dividing training set and test set, the study uses two test methods to evaluate the performance of the model, which includs subject-non-independent test and subject-independent test.

(1) Subject-non-independent test

The test scheme mixes the epochs of 8 subjects from the Sleep-EDF database for 10-fold cross-validation and assesses performance based on the level of epochs rather than the subjects. In addition, the HMM smoothing method cannot be used for subject-non-independent test because the temporal relationship between epochs has disappeared. In this type of testing, some of the test data and training data may come from the same EDF data, so the actual application is less meaningful than subject-independent test. However, in order to compare the results to some of the latest studies under similar conditions, we also provides test results of subject-non-independent test.

(2) Subject-independent test

In the subject-independent test, the data in the test set and the training set are all independent of the subjects. The data of 19 people in the Sleep-EDF Expanded dataset are used to train the random forest classifier, and the data of one person is used for testing and the experiment is repeated 20 times. Clearly, subject-independent test avoids the self-correlation of EEG from the same subject to affect performance, and thus can provide more meaningful results for practical applications and clinical studies. This paper mainly conducts multi-research direction experimental verification of the proposed model based on subject-independent test.

3 Experiment results and discussion

In this study, experiments are performed according to different purposes using 2 subsets of Sleep-EDF database (4 sc* and 4 st*) and the Sleep-EDF Expanded database (20 sc*, 10 sc* + 10 st*). The data in Sleep-EDF database is often studied in the past work. To conduct a fair and comprehensive comparison with previous studies on Sleep-EDF database, experiments are performed from multi-conditions which are set to be as similar as possible, including dataset size, EEG channel, and the number of sleep stages. At the same time, in order to verify the robustness and versatility of the proposed model, subject-independent test is performed using 20 sc* files and 10 sc* + 10 st* files in Sleep-EDF Expanded database, respectively. If there is no special explanation, the experiments in this paper are for 5-class of sleep staging, of which the expected sleep stages are W, REM, N1, N2 and N3. In Section 3.1, this paper explores the performance of a kind of classifiers to verify that the proposed WRF model has advantages in sleep staging. In Section 3.2, E1, E2 and E3 indicate the advantages of the dual channel selected in this paper in automatic sleep staging. In this paper, the importance of HMM smoothing method in automatic sleep staging is shown in Section 3.3, which proves that the proposed method has strong generalization ability and competitiveness.

3.1 Comparison of different classifiers

It can be known from the no free lunch theorem that there is no universally applicable optimal classification model that is independent of the specific application. An algorithm (Algorithm A) performs better on a particular dataset than another algorithm (Algorithm B), and must be accompanied by the performance of Algorithm A on another particular dataset that is not as good as Algorithm B. Therefore, in order to compare the performance of different classification algorithms in sleep staging, this paper also lists the classification results of different classifiers to select a best classifier. As shown in Tab. 4, in addition to WRF, this paper also uses 2 classifiers that are widely used in various classification problems: Support Vector Machines (SVM) and eXtreme Gradient Boosting (XGBoost). The results show that the performance of the WRF classifier is the best.

Table 4: The overall AC and Kappa for five-class classification using different classifiers

Classifiers	AC(%)	Kappa
XGBoost	90.75	0.853
SVM	80.00	0.662
WRF	**93.20**	**0.893**

3.2 Experiments and results of sleep staging on Sleep-EDF database

This section includes 3 experiments to compare with previous studies. E1, E2 and E3 perform subject-non-independent test on the Sleep-EDF database (4 sc* files and 4 st* files, 15,172 epochs with 218 features) using the Fpz-Cz channel, Pz-Oz channel and the double-channel EEG (Fpz-Cz and Pz-Oz), respectively.

Tab. 5 to Tab. 7 give the confusion matrices of the 3 experiments and the evaluation indicators of each sleep stage. Each class evaluation metrics are shown in the last three columns, while the overall performance metrics are listed in the last row. For E1, the AC and Kappa coefficient of the EEG Fpz-Cz channel are 91.15% and 0.860 respectively. In parallel, For E2, the AC and Kappa of the proposed model are 91.47% and 0.865, respectively. For E3, the AC and Kappa coefficient of the double-channel EEG are 93.20% and 0.893, respectively. The proposed model is universal for double-channel and single-channel EEG, and the overall result of using double-channel EEG signals is optimal.

Table 5: CM and classification performances for E1

Actual	Results of Fpz-Cz channel EEG							
	W	REM	N1	N2	N3	R(%)	P(%)	F1
W	7958	53	11	12	5	98.99	95.56	0.972
REM	92	1359	13	145	0	84.46	80.89	0.826
N1	214	179	105	101	5	17.38	75.00	0.282
N2	55	89	11	3322	144	91.74	87.77	0.897
N3	9	0	0	205	1085	83.53	87.57	0.855
Overall	**AC: 91.15%**					**Kappa: 0.860**		

Table 6: CM and classification performances for E2

Actual	Results of Pz-Oz channel EEG							
	W	REM	N1	N2	N3	R(%)	P(%)	F1
W	7964	26	41	8	0	99.07	95.38	0.972
REM	113	1314	47	135	0	81.67	81.97	0.818
N1	132	155	235	82	0	38.91	69.32	0.498
N2	94	108	16	3276	127	90.47	89.41	0.899
N3	47	0	0	163	1089	83.83	89.56	0.866
Overall	**AC: 91.47%**					**Kappa: 0.865**		

Table 7: CM and classification performances for E3

Actual	Results of double-channel EEG							
	W	REM	N1	N2	N3	R(%)	P(%)	F1
W	7991	17	23	6	2	99.40	97.71	0.986
REM	35	1400	41	132	1	87.01	83.83	0.854
N1	107	166	232	95	4	38.41	74.36	0.507
N2	111	99	20	3275	116	92.85	90.06	0.914
N3	9	87	16	3362	117	88.91	90.30	0.896
Overall	**AC: 93.20%**					**Kappa: 0.893**		

In order to visually show the results of different channel EEG signals, Fig. 3(a) shows Recall, P and F1 of the N1 stage for subject-non-independent test in double-channel EEG,

Fpz-Cz channel EEG and Pz-Oz channel EEG. The Recall and F1 of the Fpz-Cz channel are the lowest. It can be seen from Fig. 3(b) that the overall performance of the double-channel EEG is optimal, and at the same time comparing Tabs. 5-7, it proves that the double-channel EEG selected in this study has strong superiority.
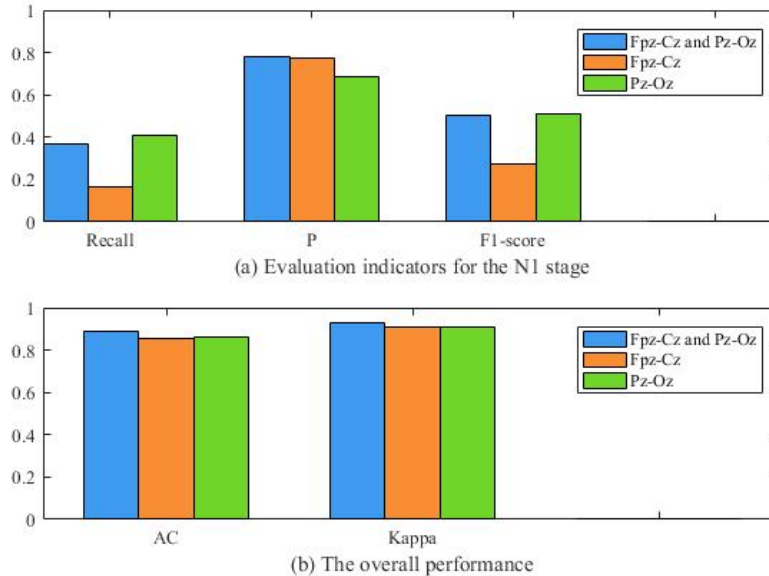

(a) Evaluation indicators for the N1 stage

(b) The overall performance

Figure 3: Comparison of performance of the different EEG channel

3.3 Experiments and results of sleep staging on Sleep-EDF Expanded database

This section contains a total of 8 experiments to explore the performance of the proposed model in different files and test the robustness and generality of the proposed model. E4 to E7 perform the subject-independent test of 20 sc* files in the Sleep-EDF Expanded database. E8 to E11 perform the subject-independent test of 10 sc* files and 10 st* files in the Sleep-EDF Expanded database. For Pz-Oz channel and double-channel signals from the sc* group, the confusion matrix and statistical parameters for each stage without and with HMM are listed in Tab. 8. For Pz-Oz channel and double-channel signals from the 10 sc* + 10 st* group, the statistical parameters for each stage without and with HMM are listed in Tab. 9.

As shown in Tabs. 8 and 9, the classification results calculated by WRF have achieved high consistency with the standard labels by experts before HMM-based smoothing. In Sleep-EDF Expanded 1 database, this paper improves the AC of the double-channel from 92.02% to 94.34% by using HMM smoothing, and the total Kappa coefficient from 0.835 to 0.886. The overall AC of the Pz-Oz channel is improved from 90.88% to 94.04% and the Kappa is improved from 0.810 to 0.880. It is also worth noting that with the improvement of AC and Kappa, the R of the N1 stage has been significantly improved. In Tab. 9, The

R for Pz-Oz channel signals and the R for double-channel signals increases by 22.27% and 21.32%, respectively. This provides a new way to increase the N1 stage recognition rate, which is a common challenge for automatic sleep staging systems.

Table 8: Classification performance in Sleep-EDF Expanded database 1

Channels	Indicators	Results without HMM					Results with HMM				
		W	REM	N1	N2	N3	W	REM	N1	N2	N3
Pz-Oz	R(%)	98.32	65.97	9.11	84.03	83.97	98.64	90.79	40.97	86.20	86.72
	P(%)	95.60	67.06	28.68	84.56	96.97	99.18	74.52	50.00	89.81	87.39
	F1	0.969	0.665	0.138	0.843	0.854	0.989	0.819	0.450	0.880	0.870
	Overall	**AC: 90.88% Kappa: 0.810**					**AC: 94.04% Kappa: 0.880**				
Fpz-Cz and Pz-Oz	R(%)	98.74	72.44	10.08	85.71	85.98	98.63	90.60	40.24	88.00	87.35
	P(%)	96.74	67.20	37.99	86.33	88.65	99.23	77.98	48.12	89.46	88.24
	F1	0.977	0.697	0.159	0.860	0.873	0.989	0.838	0.438	0.887	0.878
	Overall	**AC: 92.02% Kappa: 0.835**					**AC: 94.34% Kappa: 0.886**				

Table 9: Classification performances in Sleep-EDF Expanded database 2

Channels	Indicators	Results without HMM					Results with HMM				
		W	REM	N1	N2	N3	W	REM	N1	N2	N3
Pz-Oz	R(%)	97.66	69.56	24.23	86.96	79.51	98.13	90.48	46.50	88.74	84.70
	P(%)	93.62	70.42	39.66	86.06	83.48	97.43	83.02	65.52	88.53	82.37
	F1	0.956	0.670	0.300	0.865	0.815	0.978	0.866	0.544	0.886	0.835
	Overall	**AC: 87.01% Kappa: 0.793**					**AC: 91.40% Kappa: 0.865**				
Fpz-Cz and Pz-Oz	R(%)	98.44	78.30	24.13	88.61	81.44	98.38	91.49	45.45	90.71	84.01
	P(%)	95.86	72.75	46.49	85.97	85.66	97.68	85.22	64.27	88.22	86.65
	F1	0.971	0.754	0.318	0.873	0.835	0.980	0.882	0.533	0.895	0.853
	Overall	**AC: 88.84% Kappa: 0.823**					**AC: 91.97% Kappa: 0.874**				

In order to visually show the performance of HMM, Figs. 4(a) and 4(b) show the R, P and F1 of each stage for double-channel signals in Sleep-EDF Expanded 2 database without and with HMM smoothing, respectively. It can be clearly seen that R, P and F1 of each stage are improved, and the increase of R in the N1 stage is most obvious. The WRF algorithm based on HMM smoothing mainly improves the recognition rate of N1 stage and REM stage, which greatly improves the overall performance of the proposed method. After HMM smoothing, the overall AC for the double-channel EEG is optimized from 88.84% to 91.97%, and Kappa coefficient is optimized from 0.823 to 0.874. Similarly, the overall performance of Pz-Oz channel EEG with HMM smoothing has also been greatly improved, which fully demonstrates the competitiveness and robustness of the proposed model.

The WRF without HMM classifier ignores contextual information between neighboring epoch during sleep staging, which assigns stages independent of other epochs. This method occasionally leads to the improvement of the false recognition rate of sleep staging classification, as seen in the example shown in the middle panel of Figs. 5(a) and 5(b). It
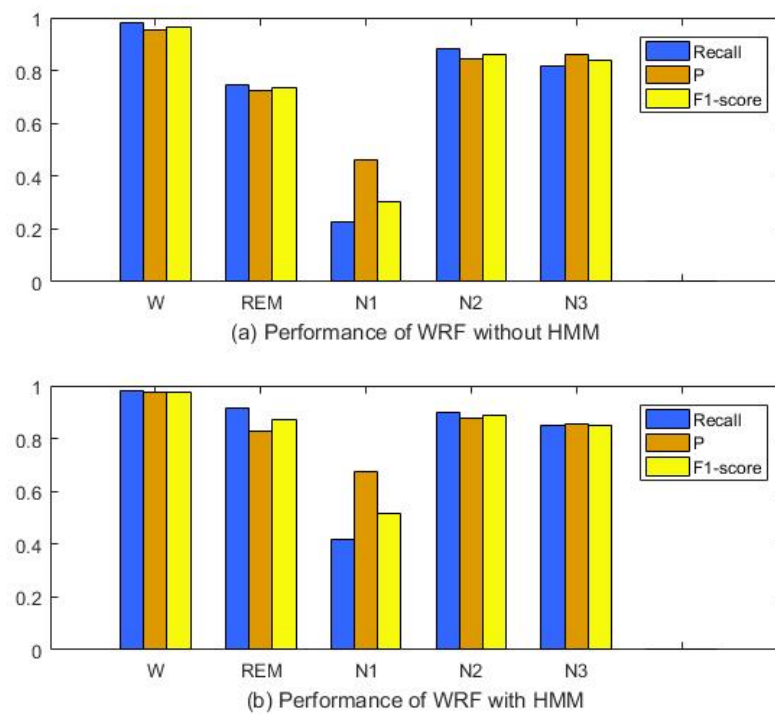
Figure 4: Comparison of performance of WRF with HMM and without HMM

can be seen that occasionally the N1 stage is identified as the REM stage. Fig. 5 shows the sleep stage changes of the SC4001E0-PSG.edf. The best classification result as shown in Fig. 5 can be gained with the value of Kappa coefficient and accuracy to be 97.47% and 0.939, respectively. Fig. 5(a) shows the actual sleep stage of each epoch in this subject, Fig. 5(b) shows the sleep stage after only WRF classification, and Fig. 5(c) shows the sleep stage after classification of RF with HMM. Obviously, after HMM, the classification results are more accurate.
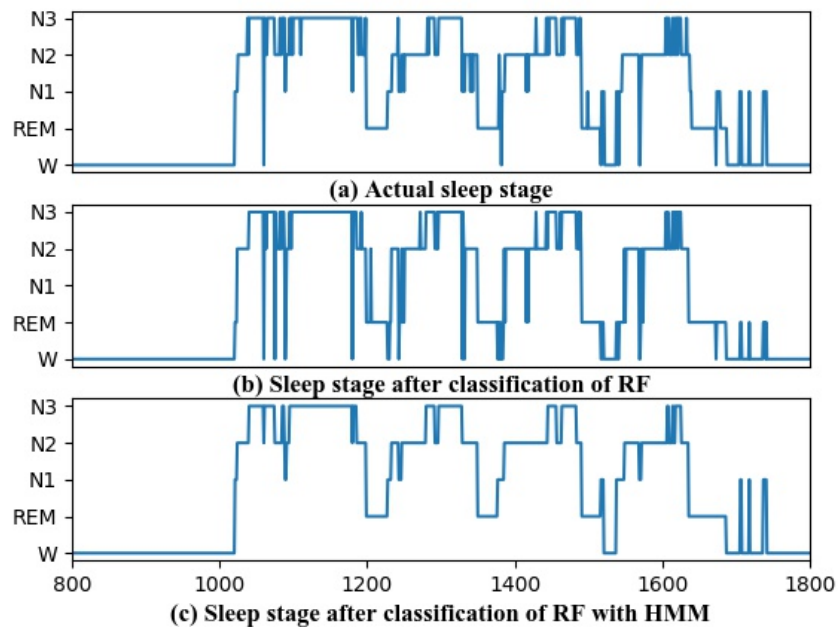

Figure 5: Hypnograms before and after HMM smoothing for the single subject with the best classification

In terms of signal filtering research, traditional filtering algorithms such as Butterworth filters often do not fully understand the time-frequency characteristics of actual signals, and the EEG signal is an unsteady signal, which is more prominent. The wavelet packet transform method can identify and determine the frequency components contained in the signal, thereby filtering out noise or unwanted frequency components, and retaining the required rhythm waves for filtering purposes. As can be seen from Tab. 10, in Sleep-EDF Expanded database 1 and Sleep-EDF Expanded database 2, the sleep staging result after WPT is superior to the result of extracting the rhythm waves by the bandpass filter, regardless of whether or not the HMM smoothing method is employed.

Table 10: Comparison of performance of the method without WPT and with WPT

Databases	Methods	WRF without HMM		WRF with HMM	
		AC(%)	Kappa	AC(%)	Kappa
Sleep-EDF Expanded 1	Bandpass filtering	90.60	0.805	93.71	0.873
	WPT	92.02	0.835	94.34	0.886
Sleep-EDF Expanded 2	Bandpass filtering	86.18	0.780	89.56	0.836
	WPT	88.84	0.823	91.97	0.874

3.4 Overall classification performance compared with state-of-the-art works

The performance of the proposed method is compared with the recent work available from the literature listed in Tab. 11. In comparison with state-of-the-art methods, the conditions are set as similar as possible, like dataset sizes, EEG channels and the test schemes. As shown in Tab. 11, the proposed method gains the best performance on both the Fpz-Cz channel and the Pz-Oz channel in the subject-non-independent test from 8 subjects of the Sleep-EDF database compared to the recent machine learning method. Seral [Seral (2013)] used the EEG, EMG and EOG recordings of 5 healthy subjects to establish an artificial neural network (ANN) classification model, and improved the sequential feature selection method to classify the sleep stages step by step. However, the AC is about 2.3% lower than the proposed method.

For subject-independent test works on the Sleep-EDF Expanded database 1, Jiang et al. [Jiang, Lu, MA et al. (2019)] achieved the best AC and Kappa of 91.4% and 0.862, which is much better than other studies, but is still lower than the proposed method by 3.9% and 0.018. Compared to the performance for Pz-Oz channel signals, the overall AC and Kappa for the double-channel signals of the proposed method reached 94.34% and 0.886.

At the same time, even though the results of the proposed method for Sleep-EDF Expanded database 2 are better than those for other methods of the sc* group topic which indicates that the method has good generalization capabilities for different databases. Jiang et al. [Jiang, Lu, MA et al. (2019)] also used RF and HMM model for subject-independent test on 10sc* and 10st* files, but the overall AC and Kappa were only 87.80% and 0.810 which is far lower than the proposed method of 91.97% and 0.874. In general, the Kappa coefficient of less than 0.4 is considered to be bad, the value between 0.41 and 0.60 is normal, the value from 0.61 to 0.80 is good, and the value higher than 0.80 is excellent. The proposed method has higher Kappa coefficients for different channel signals of different datasets which indicates that our algorithm results are in good agreement with the expert labels.

4 Discussion

An automatic sleep staging algorithm, based on the AASM standard, using WRF classifier and HMM is implemented successfully. This paper uses double-channel (Fpz-Cz and Pz-Oz) EEG from the Sleep-EDF database and the Sleep-EDF Expanded database to study the classification performances of different sleep stages. The datasets used in the proposed method is public, which facilitates repeatability studies to facilitate comparison with other papers.

Table 11: Performance comparison of recent automatic sleep scoring methods

Authors	Methods	Channels	Subjects	AC(%)	Kappa
[Rashik, Khairul and Hassan (2015)]	Spectral features & Adaboost	Pz-Oz	4 sc*+4 st*	82.00	0.717
[Liang, Kuo, Hu et al. (2012)]	Multiscale entropy & LDA	Pz-Oz	4 sc*+4 st*	83.60	0.745
[Rashik and Hassan (2016)]	Statistical features & EMD	Pz-Oz	4 sc*+4 st*	90.10	0.840
[Hsu, Yang, Wang et al. (2013)]	Elman recurrent neural classifier	Pz-Oz	4 sc*+4 st*	87.20	-
[Seral (2013)]	ANN	Pz-Oz, EOG and EMG	4 sc*+4 st*	90.90	-
[Hassan and Bhuiyan (2016)]	CEEMDAN & Bagging	Pz-Oz	4 sc*+4 st*	90.60	0.855
Proposed method	WRF & HMM	Pz-Oz	4 sc*+4 st*	**91.47**	**0.865**
[Hsu, Yang, Wang et al. (2013)]	Energy features & RNN	Fpz-Cz	4 sc*+4 st*	90.30	0.820
[Jiang, Lu, MA et al. (2019)]	RF & HMM	Fpz-Cz	4 sc*+4 st*	92.70	0.882
Proposed method	WRF & HMM	Fpz-Cz	4 sc*+4 st*	**91.15**	**0.860**
Proposed method	WRF & HMM	Fpz-Cz and Pz-Oz	4 sc*+4 st*	**93.20**	**0.893**
[Supratak, Dong, Wu et al. (2017)]	CNN & bidirect-LSTM	Pz-Oz	20 sc*	79.80	0.720
[Jiang, Lu, MA et al. (2019)]	RF & HMM	Pz-Oz	20 sc*	91.40	0.862
Proposed method	WRF & HMM	Pz-Oz	20 sc*	**94.04**	**0.880**
Proposed method	WRF & HMM	Fpz-Cz and Pz-Oz	20 sc*	**94.34**	**0.886**
[Jiang, Lu, MA et al. (2019)]	RF & HMM	Pz-Oz	10 sc*+10 st*	87.80	0.810
Proposed method	WRF & HMM	Pz-Oz	10 sc*+10 st*	**91.40**	**0.865**
Proposed method	WRF & HMM	Fpz-Cz and Pz-Oz	10 sc*+10 st*	**91.97**	**0.874**

The rhythm waves in different sleep stages are significantly different. For example, the δ wave plays a major role in the recognition of the N3 stage, while the N1 stage EEG signal is characterized by the appearance of vertex sharp waves and high amplitude θ waves. Therefore, rhythm wave extraction can effectively identify different sleep stages. The study of Sun et al. [Sun, Jia, Goparaju et al. (2017)] shows that the duration of the rhythm wave is less than 2 s, so this study divides each 30 s epoch into 29 subepochs of 2 s long with 1 s overlap. The contributions of this paper are WPT-based rhythm wave decomposition method and the method of feature extraction. Unlike most studies, this paper uses a 5-layer wavelet packet transform method to decompose each epoch into δ, θ, α, β, spindles and K-complex waves. As shown in Tab. 10, the AC of the proposed method increases by 2% to 3%, and the Kappa coefficient increases by about 0.02, compareing with the method of extracting the rhythm wave using a bandpass filter. In this paper, two types of features are extracted including long-term features and short-term features, and the features having higher correlation with sleep staging are selected by the method of removing features with low variance. Subepoch segmentation and feature selection are the main reasons for the good performance of the proposed model. As a validation, the method has been tested under different conditions, such as different datasets, EEG channels and training models. The results not only indicate the good applicability of the proposed method under various conditions, but also provide the most suitable conditions for potential applications.

Another contribution of this method is the HMM smoothing method. This is a fully automatic correction algorithm that does not require pre-defined rules. There is a great correlation between adjacent sleep stages. For example, it is impossible to directly convert the sleep state from N1 to N3, as can be seen from the example of the sleep stage change in Fig. 5(a). Contextual information from neighboring epochs can be interpreted using HMM. The predicted hypnogram after smoothing is shown in the bottom panel of Fig. 5(c). This paper solves the problem that the sleep staging algorithm can only give the classification results of each stage separately without considering the contact between time series and adjacent sleep stages. The innovation of this paper is to propose a novel method of sleep staging, and establish a model combining RF and HMM. There are a lot of literatures about sleep staging algorithms either using RF classification model alone or applying HMM alone, the contribution of this paper lies in the combination of these two methods. The proposed method can take the sleep stage state of neighboring epoch into account, which can smooth the classification model and improve the accuracy of it.

At present, the research on sleep staging has been very extensive, but all methods face the same problem, that is, the recognition rate of the N1 stage is much lower than other stages. Because N1 stage is a transitional stage between the W and the N2 stage, and has the characteristics of these two stages. The duration of the N1 stage is much less than other four stages, so it may be wrongly marked with these two stages. In addition, there is a strong similarity between the N1 stage and the REM stage, it is difficult for experts to distinguish the two stages. To solve this problem, this paper proposes an improved WRF model to balance the training set by setting different weights so that the number of epochs in each classes is equal.

Fraiwan et al. [Fraiwan, Lweesy and Khasawneh (2012)] used the time-frequency analysis

methods of CWD, CWT and HHT to study the sleep staging of single-channel EEG signals, The accuracy and Kappa coefficients for identifying sleep stages using CWD, CWT, and HHT were 78% and 70%, 83% and 76%, 75% and 65%, respectively. Compared with the proposed method, their research is based on single-channel EEG data, but the accuracy and Kappa coefficient have no obvious advantages. Another study was done by Masaaki et al. [Masaaki, Masaki and Haruaki (2002)] which obtained an accuracy of 80% according to AASM sleep staging rules. Their work was based on spectral analysis of the EEG waveforms and they used decision-tree learning for the classification procedure. The proposed method in this paper uses a classification model combining WRF and HMM , and the AC reaches 91.97% when using double-channel EEG data in Sleep-EDF Expanded 2 database.

5 Conclusions

As accurate sleep staging is helpful to diagnose some disease related to sleep disorder. Doctors need to manually observe the EEG data to mark different sleep stages, which will take a lot of time according to AASM standards. But the variability of sleep EEG waves, accurate sleep staging is hard work for most of the doctors. In order to solve the problem of intelligent sleep staging, the paper proposes an automatic sleep staging method based on RF and HMM with double-channel EEG data. The experimental data are coming from Sleep-EDF database and Sleep-EDF Expanded database. The sleep EEG data are divided into 5 classes (W, REM, N1, N2 and N3) and are segmented into 30 s epochs and 2 s epochs to extract the time domain and frequency domain features based on the WPT coefficient and some statistical features of different rhythm waves. The performance of the proposed method is evaluated by Kappa coefficient, AC, Recall, P and F1-score of the Sleep-EDF database and the Sleep-EDF Expanded database. The experimental results of using the Sleep-EDF Expanded 2 database show that the AC and Kappa coefficient reaches 91.97% and 0.874, respectively. The best AC and Kappa coefficient of single subject reaches 99.68% and 0.919, respectively.

Further study on diagnosis of sleep-related disease like depression can be researched based on the automatic sleep staging algorithm proposed in this paper.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Bachmann, M.; Päeske, L.; Kalev, K.; Aarma, K.; Lehtmets, A. et al. (2018): Methods for classifying depression in single channel eeg using linear and nonlinear signal analysis.

Computer Methods and Programs in Biomedicine, vol. 155, pp. 11-17.

Baha, Ş.; Musa, P.; Abdullah, Ç.; Çelebi, F. V. (2014): A comparative study on classification of sleep stage based on eeg signals using feature selection and classification algorithms. *Journal of Medical Systems*, vol. 38, no. 3, pp. 18.

Berry, R. B.; Brooks, R.; Gamaldo, C.; Harding, S. M.; Lloyd, R. M. et al. (2017): Aasm scoring manual updates for 2017 (version 2.4). *Journal of Clinical Sleep Medicine*, vol. 13, no. 05, pp. 665-666.

Bertisch, S. M.; Pollock, B. D.; Mittleman, M. A.; Buysse, D. J.; Bazzano, L. A. et al. (2018): Insomnia with objective short sleep duration and risk of incident cardiovascular disease and all-cause mortality: sleep heart health study. *Sleep*, vol. 41, no. 6.

Breiman, L. (2001): Random forests. *Machine Learning*, vol. 45, no. 1, pp. 5-32.

Cabir, V.; Murat, Y. (2010): Determination of sleep stage separation ability of features extracted from eeg signals using principle component analysis. *Journal of Medical Systems*, vol. 34, no. 1, pp. 83-89.

Colominas, M. A.; Jomaa, M. E. S. H.; Jrad, N.; Humeau, H. A.; Bogaert, P. V. (2017): Time-varying time-frequency complexity measures for epileptic eeg data analysis. *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 8, pp. 1681-1688.

da Silveira, T. L. T.; Kozakevicius, A. J.; Rodrigues, C. R. (2017): Single-channel eeg sleep stage classification based on a streamlined set of statistical features in wavelet domain. *Medical & Biological Engineering & Computing*, vol. 55, no. 2, pp. 343-352.

Fraiwan, L.; Lweesy, K.; Khasawneh, N. (2012): Automated sleep stage identification system based on time-frequency analysis of a single eeg channel and random forest classifier. *Computer Methods and Programs in Biomedicine*, vol. 108, no. 1, pp. 10-19.

Ghimatgar, H.; Kazemi, K.; Helfroush, M. S.; Aarabi, A. (2019): An automatic single-channel EEG-based sleep stage scoring method based on hidden markov model. *Journal of Neuroscience Methods*.vol. 324, 108320.

Gonzalez, B. D.; Small, B. J.; Cases, M. G.; Williams, N. L.; Fishman, M. N. et al. (2018): Sleep disturbance in men receiving androgen deprivation therapy for prostate cancer: the role of hot flashes and nocturia. *Cancer*, vol. 124, no. 3, pp. 499-506.

Hassan, A. R.; Bhuiyan, M. I. H. (2016): Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating. *Biomedical Signal Processing and Control*, vol. 24, pp. 1-10.

Hernández, D.; Trujillo, L.; Z-Flores, E.; Villanueva, O.; Fewell, O. R. (2018): Detecting epilepsy in eeg signals using time, frequency and time-frequency domain features. *Computer Science and Engineering-Theory and Applications*, pp. 167-182.

Hood, M. M.; Wilson, R.; Gorenz, A.; Jedel, S.; Raeisi, S. et al. (2018): Sleep quality in ulcerative colitis: associations with inflammation, psychological distress, and quality of life. *International Journal of Behavioral Medicine*, vol. 25, no. 5, pp. 517-525.

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*, vol. 104, pp. 105-114.

Jiang, D.; Lu, Y.; MA, Y.; WANG, Y. (2019): Robust sleep stage classification with single-channel eeg signals using multimodal decomposition and hmm-based refinement. *Expert Systems with Applications*, vol. 121, pp. 188-203.

Kemp, B.; Zwinderman, A. H.; Tuk, B.; Kamphuisen, H. A. C.; Oberye, J. J. L. (2000): Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the eeg. *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 9, pp. 1185-1194.

Li, Y.; Wang, X.; Luo, M.; Li, K.; Yang, X. et al. (2017): Epileptic seizure classification of eegs using time-frequency analysis based multiscale radial basis functions. *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 2, pp. 386-397.

Liang, S. F.; Kuo, C. E.; Hu, Y. H.; Yu, H. P.; Wang, Y. H. (2012): Automatic stage scoring of single-channel sleep eeg by using multiscale entropy and autoregressive models. *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 6, pp. 1649-1657.

Liu, J.; Hay, J.; Faught, B. E. (2013): The association of sleep disorder, obesity status, and diabetes mellitus among us adults-the nhanes 2009-2010 survey results. *International Journal of Endocrinology*, vol. 2013.

Malek, S.; Melgani, F.; Bazi, Y. (2018): One-dimensional convolutional neural networks for spectroscopic signal regression. *Journal of Chemometrics*, vol. 32, no. 5, pp. e2977.

Malhotra, R. K.; Kirsch, D. B.; Kristo, D. A.; Olson, E. J.; Aurora, R. N. et al. (2018): Polysomnography for obstructive sleep apnea should include arousal-based scoring: an american academy of sleep medicine position statement. *Journal of Clinical Sleep Medicine*, vol. 14, no. 07, pp. 1245-1247.

Masaaki, H.; Masaki, K.; Haruaki, Y. (2002): Automatic sleep stage scoring based on waveform recognition method and decision-tree learning. *Systems and Computers in Japan*, vol. 33, no. 11, pp. 1-13.

Memar, P.; Faradji, F. (2017): A novel multi-class eeg-based sleep stage classification system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 1, pp. 84-95.

Mourtazaev, M. S.; Kemp, B.; Zwinderman, A. H.; Kamphuisen, H. A. C. (1995): Age and gender affect different characteristics of slow waves in the sleep eeg. *Sleep*, vol. 18, no. 7, pp. 557-564.

Olsson, M.; Ärlig, J.; Hedner, J.; Blennow, K.; Zetterberg, H. (2018): Sleep deprivation and csf biomarkers for alzheimer disease. *Sleep*, vol. 41, no. 5, pp. 1-8.

Rashik, H. A.; Hassan, B. M. I. (2016): Automatic sleep scoring using statistical features in the emd domain and ensemble methods. *Biocybernetics and Biomedical Engineering*, vol. 36, no. 1, pp. 248-255.

Rashik, H. A.; Khairul, B. S.; Hassan, B. M. I. (2015): Automatic classification of sleep stages from single-channel electroencephalogram. *Annual IEEE India Conference*, pp. 1-6.

Reddy, E. D.; Fahim, A. M.; Nikhil, C.; Shubham, D. (2018): Classification of facial expressions from eeg signals using wavelet packet transform and svm for wheelchair control operations. *Procedia Computer Science*, vol. 132, pp. 1467-1476.

Sabil, A.; Vanbuis, J.; Baffet, G.; Feuilloy, M.; Meslier, N. et al. (2017): Automatic estimation of sleep and wakefulness using a single-channel eeg and home polygraphy signals. *Sleep Medicine*, vol. 40, pp. e287.

Salih, G.; Kemal, P.; Yosunkaya, Ṡ. (2010): Efficient sleep stage recognition system based on eeg signal using -means clustering based feature weighting. *Expert Systems with Applications*, vol. 37, no. 12, pp. 7922-7928.

Sara, M.; Leila, T.; Ina, D.; E, C. B. (2018): Evaluation of an automated pipeline for large scale eeg spectral analysis: the anational sleep research resource. *Sleep Medicine*, vol. 47, pp. 126-136.

Seral, Ö. (2013): Classification of sleep stages using class-dependent sequential feature selection and artificial neural network. *Neural Computing and Applications*, vol. 23, no. 5, pp. 1239-1250.

Shahnaz, C.; Sarker, T.; Paul, S. (2017): Teager energy operated perceptual wavelet packet transform coefficients for arithmetic task classification from eeg signals. *IEEE 60th International Midwest Symposium on Circuits and Systems*, pp. 176-179.

Sokolovsky, M.; Guerrero, F.; Paisarnsrisomsuk, S.; Ruiz, C.; Alvarez, S. A. (2019): Deep learning for automated feature discovery and classification of sleep stages. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Sun, H.; Jia, J.; Goparaju, B.; Huang, G.; Sourina, O. et al. (2017): Large-scale automated sleep staging. *Sleep*, vol. 40, no. 10.

Supratak, A.; Dong, H.; Wu, C.; Guo, Y. (2017): Deepsleepnet: A model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1998-2008.

Tarek, L.; Sahbi, C.; Perrine, R.; Pierre-Emmanuel, A.; Jean-Baptiste, E. et al. (2015): Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multi-class support vector machines. *Journal of Neuroscience Methods*, vol. 250, pp. 94-105.

TM, W. A.; M, B. E. (2018): Effects of reward context on feedback processing as indexed by time-frequency analysis. *Psychophysiology*, vol. 55, no. 9, pp. e13195.

Wolpert, E. A. (1969): A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. *Archives of General Psychiatry*, vol. 20, no. 2, pp. 246-247.

Yulita, I. N.; Fanany, M. I.; Arymuthy, A. M. (2017): Bi-directional long short-term memory using quantized data of deep belief networks for sleep stage classification. *Procedia Computer Science*, vol. 116, pp. 530-538.

Zhang, J.; Wu, Y. (2017): A new method for automatic sleep stage classification. *IEEE Transactions on Biomedical Circuits and Systems*, vol. 11, no. 5, pp. 1097-1110.