IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 13, 2020, accepted March 18, 2020, date of publication March 23, 2020, date of current version April 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982434

Automatic Sleep Stage Classification With Single Channel EEG Signal Based on Two-Layer Stacked Ensemble Model

JINJIN ZHOU[1], GUANGSHENG WANG[1], JUNBIAO LIU[2,4], DUANPO WU[3,4], WEIFENG XU[4], ZIMENG WANG[3], JING YE[5], MING XIA[1], YING HU[1], AND YUANYUAN TIAN[1]

[1]Department of Neurology, Shuyang People's Hospital, Xuzhou Medical University, Xuzhou 223000, China
[2]College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China
[3]School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China
[4]Hangzhou Neuro Science and Technology Company Ltd., Hangzhou 310000, China
[5]Department of Surgical ICU, Children's Hospital, Zhejiang University School of Medicine, Hangzhou 310052, China

Corresponding author: Duanpo Wu (wuduanpo@hdu.edu.cn)

This work was supported in part by the Natural Science Foundation of Zhejiang Province under Grant LGG19F030013 and Grant LGF18F010007, in part by the Key Research and Development Programs in Zhejiang Province under Grant 2020C03038, in part by the NSFC-Zhejiang Integration Joint Fund under Grant U1909209, in part by the Special Funds for Information Development in Shanghai under Grant 201801050, and in part by the Scientific Research Project of Zhejiang Provincial Department of Education under Grant Y201942165 and Grant Y201737787.

ABSTRACT Sleep stage classification, including wakefulness (W), rapid eye movement (REM), and non-rapid eye movement (NREM) which includes three sleep stages that describe the depth of sleep, is one of the most critical steps in effective diagnosis and treatment of sleep-related disorders. Clinically, sleep staging is performed by domain experts through visual inspection of polysomnography (PSG) recordings, which is time-consuming, labor-intensive and often subjective in nature. Therefore, this study develops an automatic sleep staging system, which uses single channel electroencephalogram (EEG) signal, for convenience of wearing and less interference in the sleep, to do automatic identification of various sleep stages. To achieve the automatic sleep staging system, this study proposes a two-layer stacked ensemble model, which combines the advantages of random forest (RF) and LightGBM (LGB), where RF focuses on reducing the variance of the proposed model while LGB focuses on reducing the bias of the proposed model. Particularly, the proposed model introduces a class balance strategy to improve the N1 stage recognition rate. In order to evaluate the performance of the proposed model, experiments are performed on two datasets, including Sleep-EDF database (SEDFDB) and Sleep-EDF Expanded database (SEDFEDB). In the SEDFDB, the overall accuracy (ACC), weight F1-score (WF1), Cohen's Kappa coefficient (Kappa), sensitivity of N1 (SEN-N1) obtained by proposed model are 91.2%, 0.916, 0.864 and 72.52% respectively using subject-non-independent test (SNT). In parallel, the ACC, WF1, Kappa, SEN-N1 obtained by proposed model are 82.4%, 0.751, 0.719 and 27.15% respectively using subject-independent test (SIT). Experimental results show that the performance of the proposed model are competitive with the state-of-the-art methods and results, and the recognition rate of N1 stage is significantly improved. Moreover, in the SEDFEDB, the experimental results indicate the robustness and generality of the proposed model.

INDEX TERMS Sleep stage classification, single channel EEG signal, two-layer stacked ensemble model, random forest, LightGBM.

I. INTRODUCTION

Sleep is one of the most important circadian rhythms of human physiological activities [1]. The quality of sleep impacts the performance of many basic activities, such as learning, memorization, and concentration [2]. Sustained deprivation of sleep can induce the rising risk of hypertension [3], sleep apnea syndrome [4], obesity [5], cardiovascular disease [6], Alzheimer's disease [7], Parkinson's desease [8], heart disease [9], [10] and a decrease in the efficiency in the immunitary system. Better understanding of the above mentioned sleep-related diseases and disorders relies

The associate editor coordinating the review of this manuscript and approving it for publication was Ali Salehzadeh-Yazdi.

on accurate detection of sleep stages and the sleep cycle. In sleep studies on clinical diagnosis and treatment of sleep disturbances, polysomnography (PSG) is a gold standard method for sleep evaluation, which requires the recording of many physiological signals such as electroencephalogram (EEG), electrocardiogram (ECG), electromyography (EMG), electrooculogram (EOG), pulse oximetry and respiration. PSG recordings are generally divided into 30 s epochs and each epoch is assigned with a certain sleep stage by domain experts using guidelines developed by Rechtschaffen and Kales (R&K) [11]. 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). A more recent classification manual proposed by the American Academy of Sleep Medicine (AASM) in 2007 (updated in 2017) [12], combines N3 and N4 into a single stage of deep sleep to be slow wave sleep (SWS). These distinct sleep stages are associated with distinct physiological and neuronal features which are generally used to identify the sleep stage a person is in. This process can be called sleep scoring, or sleep staging. However, manual stage scoring on PSG by domain experts is time-consuming, labor-intensive and often subjective in nature. Moreover, PSG is expensive and it may be uncomfortable for the patients as several adhesive electrodes and wires are attached to them to acquire physiological during sleep. Thus, developing a simple and reliable automatic sleep stage scoring system could be of great help. Among all the PSG signals, EEG signal plays a crucial role in recognizing sleep stages no matter manual scoring by human experts or automatic classification systems. In order to design a convenient wearable system, this study only employs single channel EEG signal, Pz-Oz channel, to classify sleep stages, since it has been proved by previous studies [13], [14] with significant advantages. The sleep stage classification methods, especially machine learning methods, based on single channel EEG signal have been widely investigated so far.

Machine learning methods follow the typical path of data preprocessing, feature extraction and feature selection before feeding the data into a classifier. Normally, EEG signals are composed of alpha (α), beta (β), theta (θ), delta (δ), sawtooth, sigma (σ) and K-complex characteristic waves. For different sleep stages, EEG are characterized by different waves. The characteristic waves of EEG signals for the 5 stages and the frequency range of each characteristic wave are shown in Table 2. Clearly, Table 2 suggests that σ wave and δ wave are dominant during N2 and SWS, respectively. In addition, α and β waves are dominant in W and REM, while θ wave is dominant in N1. Therefore, for EEG preprocessing, most studies [16]–[18] divided EEG signals into different frequency bands. Lsu *et al.* [19] used six finite impulse response band-pass filter in δ wave (0.5-2 Hz), sawtooth wave (2-6 Hz), θ wave (4-8 Hz), α wave (8-13 Hz), σ wave (12-14 Hz) and β wave (12-30 Hz), to separate the characteristic waves of EEG signals, respectively. Moreover, Memar and Faradji [17] claimed that the gamma (γ, 30-49.5 Hz) wave has a significant effect on sleep stage classification and the evidence that not using γ wave results in significant degradation in performance. However, these studies have ignored the effect of age on EEG signals. In fact, the frequency of α wave is closely related to age. Generally, α wave is formed at about 3 years old and the frequency is about 8 Hz, while the main frequency band of α wave for adults is 9-11 Hz, and the α wave slows down after 60 years old [20]–[22]. In addition, β wave increases with age, but decreases after 60 years old. The relationship between wave frequencies and ages can be summarized in Table 1.

TABLE 1. Relationship between wave frequencies and ages.

Age	Relationship between wave frequencies and ages
20-60	Slow wave components (EEG with low frequencies) decrease with age, fast wave components (EEG with high frequencies) tend to increase with age
>60	Slow wave components increase with age

For feature extraction of EEG signals, the existing studies have been discussed in detail by many studies [15], [17], [18], [23]–[30], including statistical features (standard deviation, mean, maximum, minimum, skewness, kurtosis, etc.), entropy features (kraskov entropy, spectral entropy, renyi entropy, etc.), dimension features (fractal dimension, katz fractal dimension, petrosian fractal dimension, higuchi fractal dimension, etc.), hjorth parameters (hjorth mobility, hjorth complexity, hjorth mobility) and etc. However, EEG physicians make judgments when interpreting EEG signals based on age, lead position and other factors including genetic factor, consciousness state, mental activity, drug and etc., because the same pattern may differ due to other factors. Therefore, this study includes the age of subject in the feature selection scheme. For feature selection, several feature selection techniques, including recursive feature elimination (RFE) [31], ReliefF [32], minimum redundancy maximum relevance (mRMR) [33] and information gain (IG) [34], have been discussed. In this study, mRMR is employed since it can achieve the best performance in the final sleep stage. Beyond the specific electrophysiological features used, existing classification methods mainly include support vector machine (SVM) [30], decision tree (DT) [35], random forest (RF) [18] and adaptive boosting [23], etc. Moreover, in order to improve the performance of staging results, some works have paid attention to correcting the results after classification, including smoothing rules [36], path probability rules [37] and hidden markov model (HMM)-based strategy [18]. However, in the benchmark dataset, the proportion of W stage data reaches up to 50%. Therefore, this study is an unbalanced classification problem. Unfortunately, the existing studies rarely consider this problem. In addition, most of existing studies use a single algorithm for sleep staging, but do not integrate these algorithms and then combine the advantages of each algorithm to predict sleep staging. For RF, it can improve the variance of the model prediction results through

TABLE 2. Characteristic waves of EEG signals for each sleep stage.

Sleep stage	Characteristic wave	Definition of bandwidth by IFCN (Hz)
W	α and β	$8\leq \alpha<13$, $13\leq \beta<30$
N1	α and θ	$5\leq \theta<7$
N2	σ and K-complex	$12\leq \sigma<15$, $0.5\leq$K-complex <1
SWS	δ	$0.5\leq \delta<4$
REM	α, β, θ and sawtooth	$2\leq$sawtooth<6

IFCN: International Federation of Clinical Neurophysiology.

bootstrap sampling and random selection of features, while the boosting-type algorithm can improve the bias of the model prediction results through optimizing the residual. Therefore, this study will combine the advantages of these two types of algorithms through stacking method to further improve the predictive performance of the model.

Although the current sleep staging algorithms have achieved promising results with machine learning, there is still room and need for improvement, especially the sensitivity of N1 (SEN-N1) stage due to imbalanced data. Hence, the aim of the present study is to develop an automatic sleep staging system based on single channel EEG signal which can be realized in a portable device and improve the SEN-N1. In order to improve the SEN-N1, this study proposes a two-layer stacked ensemble model, which can effectively deal with the class imbalance in sleep staging, and thus improve the performance of the proposed model, especially the SEN-N1. Moreover, the existing studies most focused on analyzing EEG signal features without considering the features of the subject, such as age. Therefore, this study first attempt to apply age information to sleep staging. In order to verify the performance of the proposed model and the effectiveness of introducing age information into the subjects, the paper conducted experiments in two published sleep staging databases, including Sleep-EDF database (SEDFDB) and Sleep-EDF Expanded database (SEDFEDB), and the test schemes use subject-non-independent test (SNT) and subject-independent test (SIT), respectively. The SEDFDB is designed to compare with existing studies of sleep staging, while the SEDFEDB is used to verify the robustness and versatility of the proposed model.

The rest of this paper is structured as follows. Materials and proposed methods are explained in full detail in Section II. Experimental results with comparison are described in Section III. Discussion is given in Sections IV, and Section V is the conclusion.

II. MATERIALS AND METHODS

This paper develops a simple and reliable automatic sleep staging system based on single channel EEG signal. Figure 1 shows the steps of proposed method and the following subsections describe the details of each step. Particularly, the orange box in Figure 1 highlights some innovative measures taken in this study in order to distinguish between various stages of sleep. For the step of ''Extracting feature from subject'', this study is the first to include age in the feature extraction scheme of sleep staging, taking into account that α wave and β wave show a large difference with age. For the step of ''class balance strategy'', this study introduces class balance strategy to address the unbalanced classification problem. Moreover, this study uses stacking method to integrate RF algorithm and lightGBM (LGB) [38] algorithm to further improve sleep staging performance.

A. SLEEP DATASETS

In this study, the first experimental data come from the SEDFDB [39] provided by Physionet, which is an open-source, benchmark, and extensively utilized database in sleep scoring literature [17], [18], [37]. The second dataset, SEDFEDB [39], is the extended version of the SEDFDB, which was updated to version 2 in March 2018. The two most common public sleep datasets for this study can be found in [40] and [41].

1) SLEEP-EDF DATABASE

In this dataset, 8 full sleep PSG recordings from Caucasian (4 males and 4 females), aged from 21 to 35, who were not on any medication at the time of the data collection. Theses recordings were grouped into 2 subsets (marked as sc* and st*). The sc* includes 4 recordings (sc4002e0, sc4012e0, sc4102e0, and sc4112e0) from ambulatory healthy subjects, which were collected during a typical 24h period of daily life. The st* includes 4 recordings (st7022j0, st7052j0, st7121j0, and st7132j0) from the subjects with mild difficultly in falling asleep but were otherwise healthy, which were collected during a night stay in a hospital. For both of the subsets, data of 2 EEG channels (EEG Pz-Oz, EEG Fpz-Cz) and 1 horizontal EOG signal have been collected. These signals were obtained with a sampling rate of 100 Hz, and then manually annotated by domain experts into different stages based on the R&K manual, which are named as W, N1, N2, N3, N4 and REM. More details about the database is described in [42].

2) SLEEP-EDF EXPANDED DATABASE

This dataset is the extended version of the SEDFDB with 197 whole-night sleep PSG recordings which contain EEG (from Pz-Oz and Fpz-Cz electrode locations), EOG (horizontal), submental chin EMG, and an event marker. These recordings are divided into 2 different groups, including 153 files marked as sc* and 44 files marked as st*. sc* files were obtained in a 1987-1991 study of age effects on sleep in 78 healthy subjects (34 males and 44 females) aged from 25 to 101, without any sleep-related medication [43]. PSG recordings of 2 subsequent day-night periods were available for each subject, except for subject 13, 36 and 52. st* files were obtained in a 1994 study of temazepam effects on sleep in 22 Caucasian males and females without other medication but had mild difficulty falling asleep. The PSGs of about 9 h were recorded in the hospital during 2 nights, one of which was after temazepam intake, and the other of which was after placebo intake [44]. Same as SEDFDB, these signals were obtained with a sampling rate of 100 Hz, and then manually

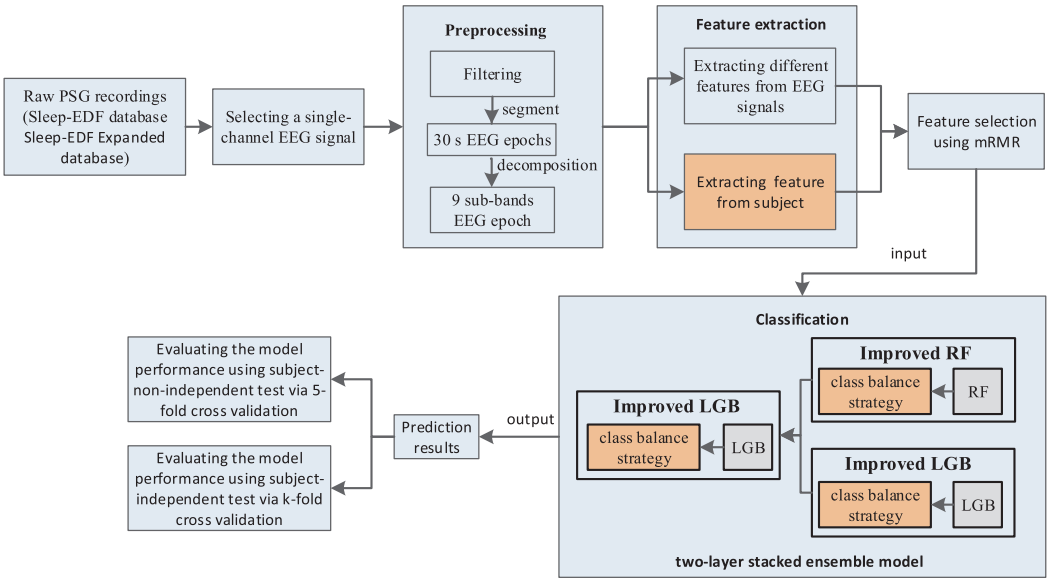

FIGURE 1. Block diagram of the automatic sleep stage classification.

annotated by domain experts into different stages based on the R&K manual, which are named as W, N1, N2, N3, N4 and REM.

B. EEG PREPROCESSING

In this study, the single channel EEG Pz-Oz signal is used. EEG Pz-Oz signal is obtained by recording electrode Pz and recording electrode Oz, which are located in the back of the brain. Therefore, the eye movement artifacts common in EEG signals almost does not interfere with the collected signals. Moreover, some studies [45], [46] have shown the existence of γ wave during waking and sleep, and the study of Memar and Faradji [17] also has shown that γ wave has a significant effect on sleep stage classification. Therefore, the single channel EEG signal is filtered with a band-pass filter with cutoff frequencies of 0.5 Hz and 49.5 Hz. In this study, the lower pass-band edge selection of 0.5 Hz is used to filter out some DC components, which are widely used in sleep staging [19], [29], [30], [47]. The upper pass-band edge selection of 49.5 Hz is based on the results of research by Memar and Faradji [17]. Then, all EEG signals are segmented into 30 s epochs with no overlap in accordance with the stage annotation of PSG recordings. The number of epochs of various sleep stages used in this study are shown in Table 3. Clearly, when both sc* files and st* files exist, the proportion of N1 is quite low while the proportion of W is quite high. When the sc* files exist alone, the proportion of W is quite high while the other stages are relatively low. When the st* files exist alone, the proportion of N2 is quite high while the other stages are relatively low. Therefore, there is a serious data imbalance problem in sleep staging, and should be noticed fully in model design.

According to Table 2, EEG signals contain several characteristic patterns with different frequency ranges, which

TABLE 3. The number of epochs of various sleep stages.

Databases	Subs(sc*/st*)	W	N1	N2	SWS	REM	Total
SEDFDB	4/4	8,027	604	3,621	1,299	1,609	15,170
SEDFEDB	153/-	285,433	21,522	69,132	13,029	25,835	414,961
SEDFEDB	-/44	4,423	3,653	19,851	6,415	8,349	42,691
SEDFEDB	44/44	78,858	6,496	38,189	12,122	16,243	162,777

Subs: it represents the Subjects.
-: it represents a placeholder.

are highly correlated to sleep states. Therefore, these discriminative information extracted from different sub-bands (rhythms) can be included in the scheme for developing an automatic sleep staging methods. In fact, the subject in SEDFDB and SEDFEDB are all adults, so in this study, the α wave of 8-13 Hz is modified to 9-11 Hz, because the dominant frequency of α rhythm in adults is between 9-11 Hz. Moreover, the β wave of different brain parts has different characteristics. The β wave in the posterior head usually ranges from 13 Hz to 20 Hz, but in the REM stage, the β wave of 20 Hz to 30 Hz gradually increases. Therefore, in order to better distinguish the W stage and REM stage, the β wave is divided into β_1 wave (13-20 Hz) and β_2 wave (20-30 Hz). As for the γ wave, since the study of Memar and Faradji [17] has shown that the γ wave has a significant effect on the classification of sleep stages, this study also divides it into γ_1 (30-40 Hz) wave and γ_2 (40-49.5 Hz) wave according to the processing method in their paper. For the rest of waves, this study is consistent with previous studies [17]–[19], that is, the boundary between different waves is defined as follows: δ wave: 0.5-4 Hz, θ wave: 4-8 Hz, σ wave: 12-15 Hz and K-complex wave: 0.5-1 Hz. Next, 9 Finite Impulse Response (FIR) band-pass filters are used in 0.5-4 Hz, 4-8 Hz, 9-11 Hz, 12-15 Hz, 14-20 Hz, 20-30 Hz, 30-40 Hz, 40-49.5 Hz and 0.5-1 Hz to separate the δ, θ, α, σ, β_1, β_2, γ_1, γ_2 and

TABLE 4. Extracted features.

Number	Features	Formulas	Counts
1	Standard deviation (SD)	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2}$	9×1
2	Spectral entropy (SE)	$-\sum_{i=1}^{N} f_t(i) log f_t(i)$	9×1
3	Kraskov entropy (KE)	$\varphi(N)-\varphi(k)+log(C_d)+\frac{d}{N}\sum_{i=1}^{N} log(\varepsilon_i^k)$	9×1
4	Renyi entropy (RE)	$-log_2(\sum_{i=1}^{N} p_i^2)$	9×1
5	Hjorth activity (HA)	$\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2$	9×1
6	Hjorth mobility (HM)	$\frac{\sigma_x'}{\sigma_x}$	9×1
7	Hjorth complexity (HC)	$\frac{\sigma_x''/\sigma_x'}{\sigma_x'/\sigma_x}$	9×1
8	Katz fractal dimension (KFD)	$\frac{log_2 N}{log_2 N+log_2(d/L)}$	9×1
9	Petrosian fractal dimension (PFD)	$\frac{log_2 N}{log_2 N+log_2(N/(N+0.4M))}$	9×1
10	Maximum-minimum distance (MMD)	$\sum_{k=1}^{N}\sqrt{\Delta x_k^2+\Delta y_k^2}$	9×1
11	Log root sum of sequential variations (LRSSV)	$log_{10}\sqrt{\sum_{i=1}^{N-1}(x_i-x_{i-1})^2}$	9×1
12	Generalized hurst exponent (GHE)	$\frac{\langle\lvert x(n+d)-x(n)\rvert^m\rangle}{\langle\lvert x(n)\rvert^m\rangle}\propto d^{mH(m)}$	9×1
13	Age	-	1×1

K-complex waves of each 30 s epochs EEG signals, and this process is accomplished using package MNE-python [48], wherein the optimal results were obtained by using an FIR filter with blackman window.

C. FEATURE EXTRACTION

Feature extraction extracts the characteristic patterns of EEG signals corresponding to different sleep stages, which is an important step for automatic classification of different sleep stages. However, the existing studies on sleep staging most focused on analyzing signal features without considering the feature of the subject, such as age. Therefore, this study not only extracts the signal features that already reported in previous studies [17], [18], [49], but also extracts the age information of the subjects. The features extracted from sub-bands epoch are presented in Table 4.

In this study, $x(n)$ represents a characteristic wave (sub-band) in 30 s epoch, which contains a total of 3000 samples since the sampling rate of EEG signals equals to 100 Hz. In Table 4:

- SD is used to quantify the amount of variation or dispersion of the sub-bands EEG data, where x_i represents the i-th sample of $x(n)$, and $\bar{x}$ represents the mean of $x(n)$ and N equals to 3000.
- SE is widely used to evaluate the flatness of the acoustic spectrum shape, and it was suggested by Inouye *et al.* [50]. This study firstly uses N-point Fast Fourier Transform (FFT) to calculate the frequency spectrum of each frame, then normalize the spectrum of each frame and let $f_t(i)$ be the normalized magnitude of the i-th frequency bin in the spectrum of frame f_t.
- KE is an unbiased estimator of Shannon entropy of a d-dimensional random sample $x(n)$ using the k-nearest neighbors sample, where φ represents the digamma function, C_d represents the volume of the d-dimensional until ball, and ε_i^k is the distance between sample x_i and its k-NN sample (in this paper, $k = 3$) points in d-dimensional sample space, and more detailed information of KE can be found in [51]. In the formula of RE, where p_i is the probability of x_i belong to possible outcome.
- The Hjorth parameters, including HA, HM and HC, provide dynamic temporal information of the EEG signal [52], where $x'(n)$ and $x''(n)$ are the first and second derivatives of $x(n)$, σ_x, σ_x' and σ_x'' are the SD of $x(n)$, $x'(n)$ and $x''(n)$, respectively.
- KFD and PFD measure the complexity of a EEG signal, where L refers to the sum of distances between two successive data samples, d represents the maximum Euclidean distance between the first sample and any other sample on the waveform, and M is the number of the sign changes in the signal derivative.
- MMD is to find the distance between the maximum and minimum samples in each subwindow, where Δx_k and Δy_k refer to the x-axis difference and the y-axis difference of the maximum and minimum samples in the k-th window.
- LRSSV is proposed in [17], which is used to measure the sequential variations between the samples of the signal.
- GHE computes long memory in a given signal at different scales [53], where d is a time lag, $\langle . \rangle$ is the sample average operator, $H(m)$ is the GHE of $x(n)$. In this study, $5 \leq d \leq 19$, $m = 1$, and more detailed information on GHE can be found in [53], [54].
- In fact, age specifically affects the neurophysiological slow-wave-generating mechanism [43]. Therefore, age is also included in the feature extraction scheme of this study.

By this feature extraction, each 30 s epoch in a sub-band is represented by 12-dimensional vector. Therefore, 108 features are extracted from the 9 sub-bands, and together with age feature of the subject, a 109-dimensional feature vector is obtained for each EEG signal epoch.

D. FEATURE SELECTION

Since the overall performance of the classifier was deeply affected by the process of feature selection, several feature selection techniques, including recursive feature elimination (RFE) [31], ReliefF [32], minimum redundancy maximum relevance (mRMR) [33] and information gain (IG) [34], were tested to find an optimum subset of features with superior classification, and finally mRMR is employed.

mRMR is proposed by Peng *et al.* [33] for dealing with the redundancy problem. It selects a subset of features by maximizing the relevance of each feature to the target class and minimizes the redundancy between the selected features. The redundancy and relevance are calculated using mutual information. For more information about mRMR can be found in [55] and [33].

E. PROPOSED MODEL

As shown in Table 2, W stage accounts for up to 50% while N1 stage accounts for less than 4% in SEDFDB. Clearly, there is an unbalanced classification problem in sleep staging, which will mislead the classification algorithm to produce biased results, especially for N1 stage. Therefore, this study introduces a class balance strategy to deal with this situation, which can improve the prediction performance of sleep staging, especially the recall rate of N1 stage, without increasing computing resources and it is inspired by the study of cost sensitive [56]–[58]. Specifically, the class balance strategy punishes W stage by adjusting the weight, i.e. giving W stage a lower weight while giving N1 a higher weight. Eq. (1) shows the formula of the weight of various classes. In particular, in order to further improve the performance of sleep staging prediction, the grid search method is employed to adjust the balance coefficient p_i.

$$w_i = p_i \frac{NT}{NC_i}, \quad (i = 1, 2, .., 5) \tag{1}$$

where NT represents the number of all epochs in the dataset, NC_i represents the number of epoch which belongs to i-th class, and p_i represents the balance coefficient which belongs to the i-th class and calculated by grid search and cross validation method based on sleep datasets which is described in Part III in detail.

In addition to the innovative introduction of class balance strategy to improve the recall rate of N1 stage, this study proposes a 2-layer stacked ensemble model to further improve the prediction performance of sleep staging.

Figure 2 shows the framework of the proposed 2-layer stacked ensemble model. In the first layer, the improved RF and the improved LGB by class balance strategy are fit to the training set that is used to prepare the inputs for the second layer classifier. RF is an ensemble method of bagging type, which has shown good performance in sleep staging [18], [37]. In this study, it mainly deals with the impact of data perturbation on the performance of the proposed model. LGB is an ensemble method of boosting type, which is a gradient boosting framework that uses tree-based learning algorithm. Considering a large number of comparison experiments on public datasets have shown that LGB can outperform existing boosting frameworks on both efficiency and accuracy, with significantly lower memory consumption, this study employs the LGB to handle sleep staging task and it mainly focuses on reducing the bias of the proposed model. To the best knowledge of authors, this study is the first to apply LGB to sleep stage task. In order to avoid overfitting, the first layer introduces the cross validation (this study uses 5-fold cross validation). As shown in Figure 2, the training set is split randomly into 5 equal-size folds, 4 folds are used to fit the improved RF algorithm and the improved LGB algorithm, and then uses the remaining 1 fold that is not used for the improved RF algorithm and the improved LGB algorithm fitting to validating. At the same time, the improved RF algorithm and the improved LGB algorithm are trained by 4 folds EEG data to predict the testing set. This process is repeated 5 times with each fold used exactly once for validating. In the second layer, the improved LGB is employed again. This study uses the combination of the validation results obtained from the first layer as the new training dataset and the combination of test results obtained from the first layer as the new testing set. Finally, the improved LGB in second layer first uses the new training set to train, then uses testing set as the input to the LGB algorithm to get the final sleep stages results.

In summary, the proposed model first introduces a class balance strategy to deal with imbalance in sleep staging, and then integrates the RF algorithm and LGB algorithm in a stacking manner to further improve the prediction performance of sleep staging.

F. EVALUATION METRICS

In order to evaluate and compare the performance of different methods, this study uses sensitivity (SEN), precision (P) and F1-score (F1) to evaluate the classification performance for each individual sleep stage, and uses Accuracy (ACC), Weighted F1-score (WF1) and Kappa to evaluate the overall performance for all the classes according to the study of [18]. Eq. (2) to Eq. (7) show the calculation methods for these evaluation metrics.

$$SEN = \frac{TP}{TP + FN} \tag{2}$$

$$P = \frac{TP}{TP + FP} \tag{3}$$

$$F1 = \frac{2 \times R \times P}{R + P} \tag{4}$$

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

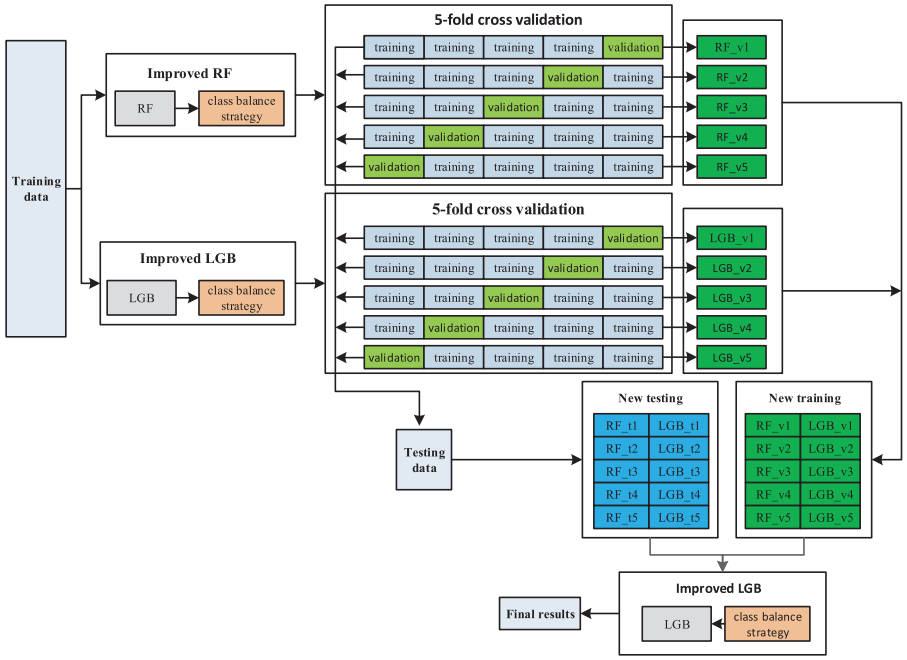

FIGURE 2. Two-layer stacked ensemble model.

where TP, TN, FP and FN represent the number of true positive, true negative, false positive and false negative samples, respectively.

$$WF1 = \sum_{i=1}^{5} w_i \times F_{1i} \tag{6}$$

$$Kappa = \frac{ACC - P_e}{1 - P_e} \tag{7}$$

where w_i and F_{1i} represent the weight and F1-score of the i-th class, respectively. P_e is the hypothetical probability of agreement by chance.

All the above evaluation metrics are obtained on the SEDFDB and the SEDFEDB because this can quantify the performance of the proposed model and it is easy to compare the performance of different sleep staging models on the public database. Certainly, the proposed model in this study can also be compared with existing devices commercially available in the market, but this will face a knotty problem, that is, the existing devices commercially available in the market cannot obtain the raw EEG signals or ECG signals, so that the performance of sleep staging of each product cannot be quantified.

G. TEST SCHEMES

According to the study of [18], this study also employs 2 kinds of test schemes, including SNT and SIT, to evaluate the performance of the proposed model.

1) SUBJECT-NON-INDEPENDENT TEST

In this test scheme, a 5-fold cross validation that uses all mixed together to evaluate the performance of proposed model. For this test scheme, testing data and training data could come from the same subject, so the performance may be overly-optimistic compared to SIT.

2) SUBJECT-INDEPENDENT TEST

In this test scheme, a k-fold cross validation based on k subjects is used to evaluate the performance of proposed model. For k subjects, this test scheme is repeated k times where the data of each subject is used to test in turn when the data of the other $k - 1$ subjects are used for training the proposed model. Particularly, for the experiments on the SEDFEDB, considering the large amount of data, the paper uses 2-fold cross validation to evaluate the performance of proposed model.

III. EXPERIMENTS AND RESULTS

This study performs 9 groups of experiments according to different purposes. Table 5 shows the different settings for all experiments and all experiments focus on the 5-class classification. Experiment 1 to Experiment 4 use 5-fold cross validation, Experiment 5 and Experiment 6 use 8-fold cross validation based on 8 subjects, and the remaining experiments, including Experiment 7, Experiment 8 and Experiment 9, all use 2-fold cross validation based on 153 files, 44 files and 88 files, respectively.

In order to make a fair comparison with previous studies on the SEDFDB, the conditions are set as similar as possible, including dataset sizes, EEG channels and test schemes. Experiment 1 and Experiment 2 perform SNT on the SEDFDB (4 sc* files and 4 st* files, 15,170 samples with 109-dimensional features) using the EEG Pz-Oz channel and

TABLE 5. Experiments under different conditions.

Experiments	Databases	Channels	Subs (sc*/st*)	No. of features	Test schemes
Experiment 1	SEDFDB	Pz-Oz	4/4	109	SNT
Experiment 2	SEDFDB	Fpz-Cz	4/4	109	SNT
Experiment 3	SEDFDB	Pz-Oz	4/4	108	SNT
Experiment 4	SEDFDB	Fpz-Cz	4/4	108	SNT
Experiment 5	SEDFDB	Pz-Oz	4/4	108	SIT
Experiment 6	SEDFDB	Fpz-Cz	4/4	108	SIT
Experiment 7	SEDFEDB	Pz-Oz	153/-	108	SIT
Experiment 8	SEDFEDB	Pz-Oz	-/44	108	SIT
Experiment 9	SEDFEDB	Pz-Oz	44/44	108	SIT

EEG Fpz-Cz channel, respectively. Particularly, in order to illustrate the effect of the age of the subject on the classification performance, Experiment 3 and Experiment 4 perform SNT on the SEDFDB (4 sc* files and 4 st* files, 15,170 samples with 108-dimensional features that removing age) using the EEG Pz-Oz channel and EEG Fpz-Cz channel, respectively. In fact, SIT is more in line with medical scenarios, using existing subject data to predict the sleep stages of unknown individual subject. Hence, Experiment 5 and Experiment 6 also perform SIT on the SEDFDB (4 sc* files and 4 st* files, 15,170 samples with 108-dimensional features that removing age) using the EEG Pz-Oz channel and EEG Fpz-Cz channel, respectively. Experiment 7 and Experiment 8 perform SIT using 153 sc* files and 44 st* files in the SEDFEDB to explore the performance of the proposed model in different subgroups. Moreover, in order to further test the robustness and generality of the proposed model, Experiment 9 uses a mixture of 44 st* files and 44 sc* files from SEDFEDB to perform SIT. The 9 groups of experiments are implemented using python 3.5 with a workstation: Intel(R) Core (TM) i5-8400 CPU @ 2.58 GHz and 8 GB of RAM.

A. EXPERIMENTS AND RESULTS OF SLEEP STAGING ON SLEEP-EDF DATABASE

Experiment 1 to Experiment 6 are performed in this section to make a fair comparison with previous studies.

1) SUBJECT-NON-INDEPENDENT TEST OF SLEEP STAGING ON SLEEP-EDF DATABASE

For Experiment 1 and Experiment 2, Firstly, mRMR is used to perform feature selection on 15,170 samples with 109-dimensional features. Next, 5-fold cross validation combined with grid search method are employed for tuning parameter. In order to find the optimal number of the features from the 109-dimensional features, Figure 3 shows the change of classification performance when features sorted by score according to mRMR. It can be found that the evaluation metrics, including ACC, WF1 and Kappa, reach their maxima when the number of feature is close to 25 while the SEN-N1 reach its maximum when the number of feature is close to 45. Therefore, the top-45 features are employed in the proposed model. Finally, in the proposed two-layer stacked ensemble model, the number of trees in RF algorithm is set to be 100, the weight of W, N1, N2, SWS and REM are set to be 1, and the other parameters in RF algorithm use the default values in the scikit-learn package. The number of trees in

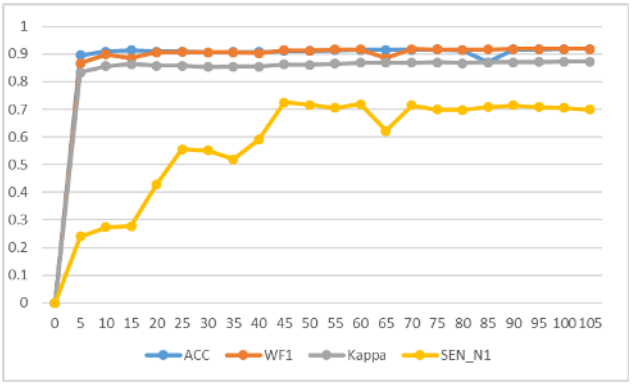

FIGURE 3. The performance versus the number of sorted features by mRMR.

LGB algorithm is set to be 200, the weight of W, N1, N2, SWS and REM are set to be 1, 115, 1, 1 and 1, respectively. The learning rate is set to be 0.09, and the other parameters in LGB algorithm are set to the default values in the scikit-learn package.

Table 6 and Table 7 show the confuse matrix and evaluation metrics for Experiment 1 and Experiment 2, respectively. For the EEG Pz-Oz channel, the ACC, WF1 and Kappa of the proposed model are 91.2%, 0.916 and 0.864, respectively. In parallel, For the EEG Fpz-Cz channel, the ACC, WF1 and Kappa of the proposed model are 91.8%, 0.919 and 0.872, respectively. In general, the value of Kappa from 0 to 0.20 is considered bad, the value from 0.21 to 0.40 is poor, the value from 0.41 to 0.60 is reasonable, the value form 0.61 to 0.80 is good, and the value overs 0.80 is outstanding. The Kappa value of both channels exceed 0.80, indicating that the agreement between manual and automatic scoring is outstanding. Moreover, ACC and WF1 also achieves an outstanding performance. Hence, the proposed model achieved excellent performance on two individual EEG channel signals. Especially, it is worth noticing the SEN-N1 stage for both channels has improved significantly, reaching 72.52% on the EEG Pz-Oz channel and 63.74% on the EEG Fpz-Cz channel.

2) COMPARISON OF THE PROPOSED MODEL WITH AND WITHOUT AGE ON SLEEP-EDF DATABASE

In order to illustrate the effect of the age of the subject on sleep staging task, the parameters of proposed model in Experiment 3 and Experiment 4 are consistent with Experiment 1 and Experiment 2. The performance of the proposed

TABLE 6. Confuse matrix and classification performance for Experiment 1 on SEDFDB using EEG Pz-Oz channel.

Exp./Mod.	W	N1	N2	SWS	REM	SEN	P	F1
W	7,876	135	9	3	14	98.00%	98.71%	98.35%
N1	42	438	58	2	64	**72.52%**	43.71%	54.55%
N2	25	181	3,268	92	55	90.25%	88.25%	89.24%
SWS	13	7	216	1,063	0	81.83%	91.64%	86.46%
REM	23	241	152	0	1,193	74.15%	89.97%	81.29%
Overall						**ACC:91.2%**	**WF1:0.916**	**Kappa:0.864**

Exp./Mod.: Exp. indicates the column labels annotated by the experts, and Mod. indicates the row labels classified by the proposed model.

TABLE 7. Confuse matrix and classification performance for Experiment 2 on SEDFDB using EEG Fpz-Cz channel.

Exp./Mod.	W	N1	N2	SWS	REM	SEN	P	F1
W	7,918	91	8	2	18	98.52%	96.68%	97.59%
N1	69	385	68	1	81	**63.74%**	49.94%	56.00%
N2	84	117	3,289	74	57	90.83%	90.43%	90.63%
SWS	67	3	139	1,090	0	83.91%	93.24%	88.33%
REM	52	175	133	2	1,247	77.50%	88.88%	82.80%
Overall						**ACC:91.8%**	**WF1:0.919**	**Kappa:0.872**

TABLE 8. Confuse matrix and classification performance for Experiment 3 on SEDFDB using EEG Pz-Oz channel.

Exp./Mod.	W	N1	N2	SWS	REM	SEN	P	F1
W	7,829	182	7	7	12	97.41%	98.35%	97.88%
N1	64	402	67	1	70	**66.56%**	40.61%	50.44%
N2	27	165	3,233	127	69	89.28%	87.69%	88.48%
SWS	16	5	217	1061	0	81.68%	88.71%	85.05%
REM	24	236	163	0	1,186	73.71%	88.71%	80.52%
Overall						**ACC:90.4%**	**WF1:0.908**	**Kappa:0.851**

TABLE 9. Confuse matrix and classification performance for Experiment 4 on SEDFDB using EEG Fpz-Cz channel.

Exp./Mod.	W	N1	N2	SWS	REM	SEN	P	F1
W	7,926	76	12	2	21	98.62%	96.15%	97.37%
N1	93	345	76	1	89	57.12%	**51.34%**	54.08%
N2	95	99	3,290	78	59	90.86%	88.54%	89.68%
SWS	71	4	185	1,039	0	79.98%	92.60%	85.83%
REM	58	148	153	2	1248	77.56%	88.07%	82.49%
Overall						**ACC:91.3%**	**WF1:0.912**	**Kappa:0.863**

model without age on SEDFDB using the EEG Pz-Oz channel and EEG Fpz-Cz channel is reported in Table 8 and Table 9. Comparing Table 6 and Table 8, it can be found that the ACC increases from 90.4% to 91.2%, the WF1 increases from 0.908 to 0.916, the Kappa increases from 0.851 to 0.864, and the SEN-N1 increases from 66.56% to 72.52% for the EEG Pz-Oz channel when using age. Similarly, comparing Table 7 and Table 9, it can be found that the ACC increases from 91.3% to 91.8%, the WF1 increases from 0.912 to 0.919, the Kappa increases from 0.863 to 0.872, and the SEN-N1 increases from 51.34% to 63.74% for the EEG Fpz-Cz channel when using age. Clearly, the difference in classification performance between using and not using age shows that the age of subject plays an important role in sleep staging.

For Experiment 8, according to the experimental steps of Experiment 1, mRMR is used to determine the number of features as 76, 2-fold cross validation based on 44 st* files and grid search method are used to determine the hyperparameters of proposed model, where the weight of W, N1, N2, SWS and REM in LGB are set to be 1, 200, 1, 1 and 150, respectively, and the other parameters in proposed model are consistent with the Experiment 1. Table 14 shows the confuse matrix and evaluation metrics for Experiment 8.

3) SUBJECT-INDEPENDENT TEST OF SLEEP STAGING ON SLEEP-EDF DATABASE

Considering the differences in sc* files and st* files, this study separately performs the SIT for sc* files and st* files. For Experiment 5, firstly, 4 sc* files are used for the SIT, mRMR is used to determine the number of features as 71, 4-fold cross validation based on 4 sc* files and grid search method are used to determine the hyperparameters of proposed model, where the weight of W, N1, N2, SWS and REM in LGB are set to be 1, 120, 1, 1 and 1, respectively, and the other parameters in proposed model are consistent with the Experiment 1. Next, 4 st* files are used for the SIT, the number of features is set to be 66, the weight of W, N1, N2, SWS and REM in LGB are set to be 1, 220, 1, 1 and 220, respectively, and the other parameters in proposed model are consistent with the Experiment 1. Finally, the confusion

TABLE 10. Confuse matrix and classification performance for Experiment 5 on SEDFDB using EEG Pz-Oz channel.

Exp./Mod.	W	N1	N2	SWS	REM	SEN	P	F1
W	7,878	72	66	5	16	98.02%	93.16%	95.53%
N1	224	164	110	1	105	**27.15%**	23.60%	25.25%
N2	87	132	3,089	192	121	85.31%	73.43%	78.92%
SWS	18	11	581	689	0	53.04%	77.68%	63.04%
REM	249	316	361	0	683	42.45%	73.84%	53.91%
Overall						**ACC:82.4%**	**WF1:0.751**	**Kappa:0.719**

TABLE 11. Confuse matrix and classification performance for Experiment 6 on SEDFDB using EEG Fpz-Cz channel.

Exp./Mod.	W	N1	N2	SWS	REM	SEN	P	F1
W	7,592	133	183	44	85	94.46%	86.79%	90.46%
N1	126	263	72	0	143	**43.54%**	42.35%	42.94%
N2	392	70	2,782	235	142	76.83%	82.36%	79.50%
SWS	316	3	188	773	19	59.51%	72.92%	65.54%
REM	322	152	153	8	974	60.53%	71.46%	65.55%
Overall						**ACC:81.6%**	**WF1:0.727**	**Kappa:0.706**

TABLE 12. The performance for 5-class classification compared with state-of-the-art works on SEDFDB.

Scheme tests	Methods	Channel	Subs (sc*/st*)	No. of epochs	ACC	WF1	Kappa	SEN-N1
SNT	Spectral features & Adaboost [61]	Pz-Oz	4/4	15,188	82.0%	0.816	0.717	22.85%
	Statistical features & EMD [62]	Pz-Oz	4/4	15,188	90.1%	0.897	0.844	39.74%
	CEEMDAN & Bagging [23]	Pz-Oz	4/4	15,188	90.7%	0.905	0.855	47.02%
	TQWT & RF [26]	Pz-Oz	4/4	15,188	91.5%	0.908	0.865	37.42%
	TQWT & Adaboost [63]	Pz-Oz	4/4	15,188	91.4%	0.910	0.864	39.74%
	RUSBoost [24]	Pz-Oz	4/4	15,188	83.5%	-	0.841	42.05%
	RF & gamma rhythm [17]	Pz-Oz	4/4	15,175	**98.4%**	-	-	67.88%
	Statistical features & TQWT [64]	Pz-Oz	4/4	15188	90.8%	0.904	0.854	38.74%
	RF & hidden Markov model [18]	Pz-Oz	4/4	15,160	92.0%	0.914	**0.874**	-
	RF & hidden Markov model [18]	Fpz-Cz	4/4	15,160	**92.7%**	**0.922**	**0.884**	-
	Energy features & recurrent neural [19]	Fpz-Cz	4/4	960	90.3%	0.898	0.820	36.70%
	Proposed method	Pz-Oz	4/4	15,170	91.2%	**0.916**	0.864	**72.52%**
	Proposed method	Fpz-Cz	4/4	15,170	91.8%	0.919	0.872	**63.74%**
SIT	RF & hidden Markov model [18]	Pz-Oz	10/10	36,972	**86.9%**	**0.860**	**0.793**	9.80%
	RF & hidden Markov model [18]	Fpz-Cz	10/10	36,972	**88.4%**	**0.879**	**0.817**	25.30%
	Proposed method	Pz-Oz	4/4	15,170	82.4%	0.751	0.719	**27.15%**
	Proposed method	Fpz-Cz	4/4	15,170	81.6%	0.727	0.706	**43.54%**

TABLE 13. Confuse matrix and classification performance for Experiment 7 on SEDFEDB using EEG Pz-Oz channel.

Exp./Mod.	W	N1	N2	SWS	REM	SEN	P	F1
W	280,554	786	2,002	27	2,064	98.29%	94.25%	96.23%
N1	8,226	162	675	30	4,896	**7.53%**	29.39%	11.98%
N2	4,184	1,219	56,338	1,619	5,772	81.49%	68.80%	74.61%
SWS	320	1	7,493	5,177	48	39.70%	75.41%	52.02%
REM	4,388	1,887	9,305	12	10,243	39.65%	44.49%	41.93%
Overall						**ACC:85.3%**	**WF1:0.834**	**Kappa:0.686**

matrix of the sc* files and st* files is added together, as shown in Table 10. Similarly, for Experiment 6, the parameters in proposed model are consistent with the Experiment 5, the confuse matrix is shown in Table 11.

For the EEG Pz-Oz channel, the ACC, WF1 and Kappa of the proposed model are 82.4%, 0.751 and 0.719, respectively. In parallel, For the EEG Fpz-Cz channel, the ACC, WF1 and Kappa of the proposed model are 81.6%, 0.727 and 0.706, respectively. Comparing Table 6 and Table 10 or Table 7 and Table 11, it reveals that the performance of using the SIT is worse than the SNT, which may be due to the large differences between the subjects and the testing data not trained in the proposed model. Furthermore, it can be found that the proposed model is likely to mislabel N1 stage as W stage or N2 stage, especially REM stage from Table 10 or Table 11, and more detailed information is discussed in section IV.

4) COMPARISON OF THE PROPOSED MODEL WITH OTHER EXISTING METHODS ON SLEEP-EDF DATABASE

The performance of some of the existing sleep stage classification systems are given and compared with the performance of the proposed model in Table 12. As illustrated in Table 12, the proposed model has the best performance in SEN-N1 under SNT or SIT. For SNT, Memar and Faradji [17]

TABLE 14. Confuse matrix and classification performance for Experiment 8 on SEDFEDB using EEG Pz-Oz channel.

Exp./Mod.	W	N1	N2	SWS	REM	SEN	P	F1
W	3,973	290	85	1	74	89.83%	74.46%	81.42
N1	826	925	856	5	1041	**25.32%**	36.43%	29.88%
N2	286	611	16,377	1,346	1,231	82.50%	73.28%	77.62%
SWS	34	19	2,402	3,867	93	60.28%	73.18%	66.11%
REM	217	694	2628	65	4,745	56.83%	66.05%	61.10%
Overall						**ACC:70.0%**	**WF1:0.689**	**Kappa:0.564**

TABLE 15. Confuse matrix and classification performance for Experiment 9 on SEDFEDB using EEG Pz-Oz channel.

Exp./Mod.	W	N1	N2	SWS	REM	SEN	P	F1
W	85,031	355	518	8	631	98.25%	93.83%	95.99%
N1	2,383	1,003	1,798	10	1,836	**14.27%**	33.41%	20.00
N2	1,478	677	33,722	2,081	2,100	84.18%	74.01%	78.77%
SWS	201	24	4,014	7,938	99	64.66%	78.56%	70.94%
REM	1,529	943	5,515	68	8,815	52.25%	65.39%	58.09%
Overall						**ACC:83.9%**,	**WF1:0.782**,	**Kappa:0.740**

achieved the best ACC for 98.4%, which is much better than other studies. However, by observing the confusion matrix in the study of Memar and Faradji it can be found that the ACC has been wrongly reported. Furthermore, the study employed a nested k-fold cross validation to evaluate the system performance, which may be overly-optimistic for sleep staging. Compared with the rest of the existing sleep stage classification systems in SNT, the evaluation metrics (including ACC, WF1 and Kappa) obtained by the proposed model are competitive with the state-of-the-art results, and the proposed method significantly improves the SEN-N1, reaching 72.52% and 63.74% by using EEG Pz-Oz channel and EEG Fpz-Cz channel, respectively. For SIT, this study still has significant advantages in SEN-N1, but ACC, WF1 and Kappa have no advantage over the state-of-the-art results, which is left for future work.

B. EXPERIMENTS AND RESULTS OF SLEEP STAGING ON SLEEP-EDF EXPANDED DATABASE

Experiment 7, Experiment 8 and Experiment 9 are performed in this section to explore the performance of the proposed model in different subgroups and test the robustness and generality of the proposed model. Since the EEG signals of different age are quite different and the number of subjects is little, which in turn affects the performance of the proposed model in SIT. Hence, the 108-dimensional features that removing age are used to represent EEG signals.

For Experiment 7, according to the experimental steps of Experiment 1, mRMR is used to determine the number of features as 71, 2-fold cross validation based on 153 sc* files and grid search method are used to determine the hyperparameters of proposed model, where the weight of W, N1, N2, SWS and REM in LGB are set to be 1, 200, 1, 1 and 1, respectively, and the other parameters in proposed model are consistent with the Experiment 1. Table 13 shows the confuse matrix and evaluation metrics for Experiment 7.

For Experiment 9, according to the experimental steps of Experiment 3, for 44 st* files, the parameters in proposed model are consistent with the Experiment 8. For 44 sc* files, mRMR is used to determine the number of features as 51, 2-fold cross validation based on 44 sc* files and grid search method are used to determine the hyperparameters of proposed model, where the weight of W, N1, N2, SWS and REM in LGB are set to be 1, 200, 1, 1 and 1, respectively, and the other parameters in proposed model are consistent with the Experiment 1. Table 15 shows the confuse matrix and evaluation metrics for Experiment 9.

Comparing Table 13 and Table 14, it can be found that ACC, WF1 and Kappa on sc* subgroup are better than st* subgroup, but SEN-N1 is worse, which may be due to the smaller proportion of N1 in sc* subgroup. Furthermore, Table 13, Table 14 and Table 15 again show that the proposed model is prone to misidentify N1 stage as W stage, N2 stage or REM stage, which may imply that the features extracted in this study are insufficient, and the performance of sleep stages in SIT can not be improved by the model alone, which is the focus of future research. In fact, Experiment 7, Experiment 8 and Experiment 9 achieved Kappa values of 0.686, 0.564 and 0.740, respectively. It indicates that the agreement between manual and automatic scoring is reasonable in sc* subgroup, good in st* subgroup and good in mix group.

IV. DISCUSSION

To achieve a simple and more accurate automatic sleep staging system, this study proposes a two-layer stacked ensemble model to distinguish between different sleep stages using a single EEG channel. The two-layer stacked ensemble model combines the advantages of RF and LGB, where RF focuses on reducing the variance of the proposed model while LGB focuses on reducing the bias of the proposed model. Compared with existing studies on the SEDFDB, the proposed model can provide a promising performance. Particularly, the proposed model introduces a balance strategy to improve the N1 stage recognition rate, which is a common challenge for automatic sleep staging system. There are two main

reasons for the difficulty of N1 stage recognition. Firstly, considering that the N1 stage is the transitional stage between the W stage and N2 stage, the automatic sleep staging system is likely to mislabel it as W stage or N2 stage. Moreover, there exists strong similarity between the N1 stage and the REM stage, making them hard to differentiate even by domain experts without other information [18]. Secondly, since human sleep is composed of several stages with uneven distribution, the number of N1 epochs are much less than that of the other 4 stages, making the sleep staging an unbalanced classification problem, which seriously affects the SEN-N1.

In this study, a class balance strategy by adjusting the weights of each class is proposed, which significantly improves the N1 stage recognition rate and keeps the overall performance of proposed model. Moreover, the age information, a new feature that the existing studies on sleep staging has not included in their feature extraction scheme, plays an important role in proposed sleep staging, and the comparison of Table 6 and Table 8, or Table 7 and Table 9 has confirmed that it has a positive effect on the classification of sleep stages. Furthermore, the results shown in Table 13 to Table 15 reveal that the proposed model also achieves good performance in SIT, which is more in line with the real world medical scene. In addition, the single channel EEG signal greatly simplifies the automatic sleep staging system. In fact, all 9 groups of experiments performed in different conditions show the robustness and generality of the proposed model.

However, in the present case of study, there are several limitations of the current study that need to be recognized. Firstly, the experimental data are derived from healthy subjects and the performance of the proposed model is not tested by those who had a sleep-breathing disorders, which reduced the persuasiveness of the robustness and generality of proposed model. Secondly, this study does not analyze EEG signals in detail and in depth to extract more representative or distinctive signal features. This study only decomposes a single channel EEG signal into 9 sub-bands to extract some common features in literature. Particularly, for the difficulty of N1 stage recognition, researchers may be able to extract some representative signal features by deep analysis of the differences between N1 and REM, which may further improve the recognition rate of N1. Thirdly, for the performance of SIT, there have been some studies that have paid attention to correct the classification results, including smoothing rules [36], path probability rules [37], and HMM-based refinement [18]. However, this study does not make further research on it, so future work can focus on improving the performance of SIT.

V. CONCLUSION

This study undertakes to develop an automatic sleep stage classification system using a single EEG channel, which can provide a wearable home sleep monitoring system. In order to improve the ACC, WF1, Kappa and SEN-N1 of the proposed method, a two-layer stacked ensemble model is proposed. The performance of the model is evaluated in two datasets. Particularly, the SEN-N1 reaches 72.52% in SEDFDB dataset.

ACKNOWLEDGMENT

(Jinjin Zhou1 and Guangsheng Wang contributed equally to this work.)

REFERENCES

[1] S. Panda, J. B. Hogenesch, and S. A. Kay, "Circadian rhythms from flies to human," *Nature*, vol. 417, no. 6886, pp. 329–335, May 2002, doi: 10.1038/417329a.

[2] J. M. Ellenbogen, J. D. Payne, and R. Stickgold, "The role of sleep in declarative memory consolidation: Passive, permissive, active or none?" *Current Opinion Neurobiol.*, vol. 16, no. 6, pp. 716–722, Dec. 2006, doi: 10.1016/j.conb.2006.10.006.

[3] K. Lu, J. Chen, S. Wu, J. Chen, and D. Hu, "Interaction of sleep duration and sleep quality on hypertension prevalence in adult Chinese males," *J. Epidemiol.*, vol. 25, no. 6, pp. 415–422, Apr. 2015, doi: 10.2188/jea.je20140139.

[4] K. A. Franklin and E. Lindberg, "Obstructive sleep apnea is a common disorder in the population—A review on the epidemiology of sleep apnea," *J. Thoracic Disease*, vol. 7, no. 8, p. 1311, Aug. 2015, doi: 10.3978/j.issn.2072-1439.2015.06.11.

[5] J. Liu, J. Hay, and B. E. Faught, "The association of sleep disorder, obesity status, and diabetes mellitus among US adults—The NHANES 2009–2010 survey results," *Int. J. Endocrinol.*, vol. 2013, no. 12, p. 234129, Jul. 2013, doi: 10.1155/2013/234129.

[6] S. M. Bertisch, B. D. Pollock, M. A. Mittleman, D. J. Buysse, L. A. Bazzano, D. J. Gottlieb, and S. Redline, "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, Jun. 2018, Art. no. zsy047, doi: 10.1093/sleep/zsy047.

[7] M. Olsson, J. Ärlig, J. Hedner, K. Blennow, and H. Zetterberg, "Sleep deprivation and CSF biomarkers for Alzheimer disease," *Sleep*, vol. 41, no. 1, pp. 1–8, May 2018, doi: 10.1093/sleep/zsy025.

[8] K. Wulff, S. Gatti, J. G. Wettstein, and R. G. Foster, "Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease," *Nature Rev. Neurosci.*, vol. 11, no. 8, pp. 589–599, Aug. 2010, doi: 10.1038/nrn2868.

[9] D. J. Gottlieb, S. Redline, F. J. Nieto, C. M. Baldwin, A. B. Newman, H. E. Resnick, and N. M. Punjabi, "Association of usual sleep duration with hypertension: The sleep heart health study," *Sleep*, vol. 29, no. 8, pp. 1009–1014, Aug. 2006, doi: 10.1093/sleep/29.8.1009.

[10] A. M. Arruda-Olson, L. J. Olson, A. Nehra, and V. K. Somers, "Sleep apnea and cardiovascular disease: Implications for understanding erectile dysfunction," *Herz*, vol. 28, no. 4, pp. 298–303, Jun. 2003, doi: 10.1007/s00059-003-2482-z.

[11] E. A. Wolpert, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," *Arch. Gen. Psychiatry*, vol. 20, no. 2, p. 246, Feb. 1969, doi: 10.1001/archpsyc.1969.01740140118016.

[12] R. B. Berry, R. Brooks, C. Gamaldo, S. M. Harding, R. M. Lloyd, S. F. Quan, M. T. Troester, and B. V. Vaughn, "AASM scoring manual updates for 2017 (version 2.4)," *J. Clin. Sleep Med.*, vol. 13, no. 05, pp. 665–666, May 2017, doi: 10.5664/jcsm.6576.

[13] S.-F. Liang, C.-E. Kuo, Y.-H. Hu, Y.-H. Pan, and Y.-H. Wang, "Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 6, pp. 1649–1657, Jun. 2012, doi: 10.1109/tim.2012.2187242.

[14] G. Zhu, Y. Li, and P. Wen, "Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 6, pp. 1813–1821, Nov. 2014, doi: 10.1109/JBHI.2014.2303991.

[15] J. Dong, D. Liu, C. Zhang, J. Ma, G. Wang, D. Guo, Y. Liu, H. Zhong, J. Zhang, C.-K. Peng, and J. Fang, "Automated sleep staging technique based on the empirical mode decomposition algorithm: A preliminary study," *Adv. Adapt. Data Anal.*, vol. 2, no. 2, pp. 267–276, Apr. 2010, doi: 10.1142/S1793536910000483.

[16] C.-S. Huang, C.-L. Lin, L.-W. Ko, S.-Y. Liu, T.-P. Su, and C.-T. Lin, "Knowledge-based identification of sleep stages based on two forehead electroencephalogram channels," *Frontiers Neurosci.*, vol. 8, p. 263, Sep. 2014, doi: 10.3389/fnins.2014.00263.

[17] P. Memar and F. Faradji, "A novel multi-class EEG-based sleep stage classification system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 1, pp. 84–95, Jan. 2018, doi: 10.1109/TNSRE.2017.2776149.

[18] D. Jiang, Y.-N. Lu, Y. Ma, and Y. Wang, "Robust sleep stage classification with single-channel EEG signals using multimodal decomposition and HMM-based refinement," *Expert Syst. Appl.*, vol. 121, pp. 188–203, May 2019, doi: 10.1016/j.eswa.2018.12.023.

[19] Y.-L. Hsu, Y.-T. Yang, J.-S. Wang, and C.-Y. Hsu, "Automatic sleep stage recurrent neural classifier using energy features of EEG signals," *Neurocomputing*, vol. 104, pp. 105–114, Mar. 2013, doi: 10.1016/j.neucom.2012.11.003.

[20] M. Matoušek, J. Volavka, and J. Roubíček, "EEG frequency analysis related to age in normal adults," *Electroencephalogr. Clin. neurophysiology*, vol. 23, no. 6, pp. 162–167, Aug. 1967, doi: 10.1016/0013-4694(67)90106-X.

[21] J. Carrier, S. Land, D. J. Buysse, D. J. Kupfer, and T. H. Monk, "The effects of age and gender on sleep EEG power spectral density in the middle years of life (ages 20–60 years old)," *Psychophysiology*, vol. 38, no. 2, pp. 232–242, Mar. 2001, doi: 10.1111/1469-8986.3820232.

[22] A. J. Silverman, E. W. Busse, R. H, Barnes, "Studies in the processes of aging: Electroencephalographic findings in 400 elderly subjects," *Electroencephalogr. Clin. Neurophysiol.*, vol. 7, pp. 67–74, 1955, doi: 10.1016/0013-4694(55)90060-2.

[23] A. R. Hassan and M. I. H. Bhuiyan, "Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating," *Biomed. Signal Process. Control*, vol. 24, pp. 1–10, Feb. 2016, doi: 10.1016/j.bspc.2015.09.002.

[24] A. R. Hassan and M. I. H. Bhuiyan, "Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting," *Comput. Methods Programs Biomed.*, vol. 140, pp. 201–210, Mar. 2017, doi: 10.1016/j.cmpb.2016.12.015.

[25] A. Tsanas and G. D. Clifford, "Stage-independent, single lead EEG sleep spindle detection using the continuous wavelet transform and local weighted smoothing," *Frontiers Hum. Neurosci.*, vol. 9, p. 181, Apr. 2015, doi: 10.3389/fnhum.2015.00181.

[26] A. R. Hassan and M. I. H. Bhuiyan, "A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features," *J. Neurosci. Methods*, vol. 271, pp. 107–118, Sep. 2016, doi: 10.1016/j.jneumeth.2016.07.012.

[27] F. Ebrahimi, M. Mikaeili, E. Estrada, and H. Nazeran, "Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, doi: 10.1109/IEMBS.2008.4649365.

[28] N. Kannathal, J. Chee, K. Er, K. Lim, and O. H. Tat, "Chaotic analysis of epileptic EEG signals," *IFMBE Proc.*, vol. 43, pp. 652–654, 2014, doi: 10.1007/978-3-319-02913-9_166.

[29] B. Koley and D. Dey, "An ensemble system for automatic sleep stage classification using single channel EEG signal," *Comput. Biol. Med.*, vol. 42, no. 12, pp. 1186–1195, Dec. 2012, doi: 10.1016/j.compbiomed.2012.09.012.

[30] T. Lajnef, S. Chaibi, P. Ruby, P.-E. Aguera, J.-B. Eichenlaub, M. Samet, A. Kachouri, and K. Jerbi, "Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines," *J. Neurosci. Methods*, vol. 250, pp. 94–105, Jul. 2015, doi: 10.1016/j.jneumeth.2015.01.022.

[31] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, nos. 1–3, pp. 389–422, Jan. 2002, doi: 10.1023/a:1012487302797.

[32] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," in *Proc. 10th Nat. Conf. Artif. Intell.*, 1992, pp. 129–134.

[33] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005, doi: 10.1109/TPAMI.2005.159.

[34] J. R. Quinlan, *C4.5: Programs for Machine Learning*, 1st ed. San Mateo, CA, USA: Morgan Kaufmann, 1992, pp. 313–320.

[35] S. A. Imtiaz and E. Rodriguez-Villegas, "Automatic sleep staging using state machine-controlled decision trees," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 378–381, doi: 10.1109/EMBC.2015.7318378.

[36] S.-F. Liang, C.-E. Kuo, Y.-H. Hu, and Y.-S. Cheng, "A rule-based automatic sleep staging method," *J. Neurosci. Methods*, vol. 205, no. 1, pp. 169–176, Mar. 2012, doi: 10.1016/j.jneumeth.2011.12.022.

[37] X. Li, L. Cui, S. Tao, J. Chen, X. Zhang, and G.-Q. Zhang, "HyCLASSS: A hybrid classifier for automatic sleep stage scoring," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 2, pp. 375–385, Mar. 2018, doi: 10.1109/JBHI.2017.2668993.

[38] G. Ke, Q. Meng, T. W. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3149–3157.

[39] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, p. E215, Jun. 2000, doi: 10.1161/01.CIR.101.23.e215.

[40] A. L. Goldberger, L. A. N. Amaral, L.Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, and Moody. (2007). *The Sleep-EDF Database*. [Online]. Available: https://physionet.org/physiobank/database/sleep-edf/

[41] A. L. Goldberger, L. A. N. Amaral, L.Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, and Moody. (2018). *The Sleep-EDF Database [Expanded]*. [Online]. Available: https://physionet.org/pn6/sleep-edfx/

[42] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. C. Kamphuisen, and J. J. L. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 9, pp. 1185–1194, Sep. 2000, doi: 10.1109/10.867928.

[43] M. S. Mourtazaev, B. Kemp, A. H. Zwinderman, and H. A. C. Kamphuisen, "Age and gender affect different characteristics of slow waves in the sleep EEG," *Sleep*, vol. 18, no. 7, pp. 557–564, Sep. 1995, doi: 10.1093/sleep/18.7.557.

[44] B. Kemp, A. Janssen, and M. J. Roessen, "A digital telemetry system for ambulatory sleep recording," *Sleep-Wake Res. The Netherlands*, vol. 18, no. 4, pp. 129–132, 1993.

[45] J. L. Cantero, M. Atienza, J. R. Madsen, and R. Stickgold, "Gamma EEG dynamics in neocortex and hippocampus during human wakefulness and sleep," *NeuroImage*, vol. 22, no. 3, pp. 1271–1280, Jul. 2004, doi: 10.1016/j.neuroimage.2004.03.014.

[46] S. M. Montgomery, A. Sirota, and G. Buzsaki, "Theta and gamma coordination of hippocampal networks during waking and rapid eye movement sleep," *J. Neurosci.*, vol. 28, no. 26, pp. 6731–6741, Jun. 2008, doi: 10.1523/JNEUROSCI.1227-08.2008.

[47] S. Khalighi, T. Sousa, G. Pires, and U. Nunes, "Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels," *Expert Syst. Appl.*, vol. 40, no. 17, pp. 7046–7059, Dec. 2013, doi: 10.1016/j.eswa.2013.06.023.

[48] A. Gramfort, M. Luessi, and E. Larson, "MEG and EEG data analysis with MNE-Python," *Frontiers Neurosci.*, vol. 7, no. 267, pp. 1–13, Dec. 2013, doi: 10.3389/fnins.2013.00267.

[49] K. Šušmáková and A. Krakovská, "Discrimination ability of individual measures used in sleep stages classification," *Artif. Intell. Med.*, vol. 44, no. 3, pp. 261–277, Nov. 2008, doi: 10.1016/j.artmed.2008.07.005.

[50] T. Inouye, K. Shinosaki, H. Sakamoto, S. Toi, S. Ukai, A. Iyama, Y. Katsuda, and M. Hirano, "Quantification of EEG irregularity by use of the entropy of the power spectrum," *Electroencephalogr. Clin. Neurophysiol.*, vol. 79, no. 3, pp. 204–210, Sep. 1991, doi: 10.1016/0013-4694(91)90138-t.

[51] K. A. Veselkov, V. I. Pahomov, J. C. Lindon, V. S. Volynkin, D. Crockford, G. S. Osipenko, D. B. Davies, R. H. Barton, J.-W. Bang, E. Holmes, and J. K. Nicholson, "A metabolic entropy approach for measurements of systemic metabolic disruptions in patho-physiological states," *J. Proteome Res.*, vol. 9, no. 7, pp. 3537–3544, Jul. 2010, doi: 10.1021/pr1000576.

[52] S. J. Redmond and C. Heneghan, "Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 485–496, Mar. 2006, doi: 10.1109/TBME.2005.869773.

[53] T. Di Matteo, "Multi-scaling in finance," *Quant. Finance*, vol. 7, no. 1, pp. 21–36, Feb. 2007, doi: 10.1080/14697680600969727.

[54] T. D. Matteo, T. Aste, and M. M. Dacorogna, "Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development," *J. Banking Finance*, vol. 29, no. 4, pp. 827–851, Apr. 2005, doi: 10.1016/j.jbankfin.2004.08.004.

[55] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *J. Bioinf. Comput. Biol.*, vol. 3, no. 2, pp. 185–205, Apr. 2005, doi: 10.1142/S0219720005001004.

[56] A. Braytee, W. Liu, and P. Kennedy, "A cost-sensitive learning strategy for feature extraction from imbalanced data," in *Proc. Int. Conf. Neural Inf. Process.*, Kyoto, Japan, 2016, pp. 78–86.
[57] Z. Qin, T. Wang, and S. Zhang, "Incorporating medical history to cost sensitive classification with lazy learning strategy," in *Proc. IEEE Int. Conf. Prog. Informat. Comput.*, Shanghai, China, Dec. 2010, pp. 19–23.
[58] F. Min and W. Zhu, "A competition strategy to cost-sensitive decision trees," in *Proc. Int. Conf. Rough Sets Knowl. Technol.*, Chengdou, China, 2012, pp. 359–368.
[59] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[60] Q. Meng, G. Ke, T. Wang, W. Chen, Q. Ye, Z.-M. Ma, and T.-Y. Liu, "A communication-efficient parallel algorithm for decision tree," in *Proc. Neural Inf. Process. Syst.*, 2016, pp. 1279–1287.
[61] A. R. Hassan, S. K. Bashar, and M. I. H. Bhuiyan, "Automatic classification of sleep stages from single-channel electroencephalogram," in *Proc. Annu. IEEE India Conf. (INDICON)*, Dec. 2015, pp. 1–6, doi: 10.1109/INDICON.2015.7443756.
[62] A. R. Hassan and M. I. Hassan Bhuiyan, "Automatic sleep scoring using statistical features in the EMD domain and ensemble methods," *Biocybern. Biomed. Eng.*, vol. 36, no. 1, pp. 248–255, Nov. 2016, doi: 10.1016/j.bbe.2015.11.001.
[63] A. R. Hassan and M. I. H. Bhuiyan, "An automated method for sleep staging from EEG signals using normal inverse Gaussian parameters and adaptive boosting," *Neurocomputing*, vol. 219, pp. 76–87, Jan. 2017, doi: 10.1016/j.neucom.2016.09.011.
[64] A. R. Hassan and A. Subasi, "A decision support system for automated identification of sleep stages from single-channel EEG signals," *Knowl.-Based Syst.*, vol. 128, pp. 115–124, Jul. 2017, doi: 10.1016/j.knosys.2017.05.005.

JINJIN ZHOU received the B.S. degree from the College of Medical, Yichun University, in 2009. He has been working with the Department of Neurology, Shuyang People's Hospital, Xuzhou Medical University, since 2009. He is good at clinical application of EEG.

GUANGSHENG WANG received the B.S. and M.S. degrees in neurology from Xuzhou Medical University, in 1994 and 2011, respectively. He has been working with the Department of Neurology, Shuyang People's Hospital, Xuzhou Medical University, since 1994. He is good at neurological intervention (including intracranial aneurysm, arteriovenous malformation, and intravascular treatment of cerebral embolism) and clinical application of EEG.

JUNBIAO LIU received the B.S. degree in electronic communication engineering from Zhejiang University, in 2003, and the M.S. degree in control theory and control engineering from Hangzhou Dianzi University, in 2008. Since 2015, he has been the CTO with Hangzhou Neuro Science and Technology Company Ltd., China. His research interests include signal processing and wireless sensor networks.

DUANPO WU received the B.S. degree from the College of Electronics and Information, Hangzhou Dianzi University, in 2009, and the Ph.D. degree from the College of Information Science and Electronic Engineering, Zhejiang University, China, in 2014. Since 2014, he has been a Lecturer with Hangzhou Dianzi University, China. His research interests include intelligent signal processing, biologic data analysis, and machine learning.

WEIFENG XU received the B.S. degree from the College of Information Engineering, West Anhui University, in 2015, and the M.S. degree from the Key Laboratory of Advanced Design and Intelligent Computing (Dalian University), Ministry of Education, China, in 2018. Since 2018, he has been a Machine Learning Engineer with Hangzhou Neuro Science and Technology Company Ltd, China. His research interests include biologic data analysis, modeling, and machine learning.

ZIMENG WANG received the B.S. degree in electronic information engineering from the Qingdao University of Technology, in 2018. She is currently pursuing the master's degree with Hangzhou Dianzi University. Her main research interests include automatic sleep staging, epileptic seizure detection, and prediction.

JING YE received the bachelor's degree in clinical medicine from Sun Yat-sen University, and the master's degree in pediatrics from Zhejiang University, in 2003. She is currently the Deputy Chief Pediatrician with the Surgical ICU Department, Children's Hospital, Zhejiang University School of Medicine. Her research interests include brain damage, brainwave, and pathological changes in children.

MING XIA received the B.S. degree from Nanjing Medical University, in 2014. She has been working with the Department of Neurology, Shuyang People's Hospital, Xuzhou Medical University, since 2006. In 2019, she studied EEG at the Second Affiliated Hospital, Zhejiang University School of Medicine. She is good at clinical application of EEG.

YING HU received the B.S. degree from Xuzhou Medical University, in 2016. She has been working with the Department of Neurology, Shuyang People's Hospital, Xuzhou Medical University, since 2013. In 2019, she studied EEG at the Second Affiliated Hospital, Zhejiang University School of Medicine. She is good at clinical application of EEG.

YUANYUAN TIAN received the B.S. degree from the Medical College, Yangzhou University. She has been working with the Department of Neurology, Shuyang People's Hospital, Xuzhou Medical University, since 2011. In 2019, she studied EEG at the Second Affiliated Hospital, Zhejiang University School of Medicine. She is good at clinical application of EEG.