



# Enhancing automatic sleep stage classification with cerebellar EEG and machine learning techniques



Wang Manli<sup>a</sup>, Guan Junwen<sup>b</sup>, Sun Tong<sup>b</sup>, Wang Junjie<sup>b</sup>, Yuan Yikai<sup>b</sup>, Zhou Yicheng<sup>b</sup>, Zhang Yi<sup>b</sup>, Yang Xiaoyu<sup>b</sup>, Li Xuepei<sup>b</sup>, Yang Jingguo<sup>b</sup>, Zhou Xuebin<sup>b</sup>, Yu Hang<sup>a,\*</sup>

<sup>a</sup> Clinical Research Department, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, Affiliated Cancer Hospital of University of Electronic Science and Technology of China, Chengdu, China

<sup>b</sup> Department of Neurosurgery, West China Hospital, Sichuan University, China

## ARTICLE INFO

### Keywords:

Cerebellar electroencephalography  
Automatic sleep stage classification  
Machine learning  
Polysomnography

## ABSTRACT

Sleep disorders have become a significant health concern in modern society. To investigate and diagnose sleep disorders, sleep analysis has emerged as the primary research method. Conventional polysomnography primarily relies on cerebral electroencephalography (EEG) and electromyography (EMG) for sleep stage scoring, but manual scoring is time-consuming and subjective. This study investigated the potential application of cerebellar EEG combined with machine learning in automatic sleep stage classification. Twenty-five male mice underwent 24-h cerebral EEG/cerebellar EEG/EMG recording, and manual sleep staging was performed. Various machine learning models, including Light Gradient Boosting (LGBoost), Extreme Gradient Boosting, Categorical Boosting, Support Vector Machine, Logistic Regression, Random Forest, Long Short-Term Memory and Convolutional Neural Network, were applied for automatic sleep stage classification. The performance of different models and the efficacy of cerebellar EEG, cerebral EEG, and EMG were compared under different training:test set ratios. Cerebellar EEG exhibited significant differences in power spectral density across wakefulness, non-rapid eye movement sleep stages, and rapid eye movement sleep stages, particularly at frequencies  $>7$  Hz. LGBoost, Extreme Gradient Boosting, and Categorical Boosting models showed comparable performance, with LGBoost being selected for further analyses due to its shorter computation time. Cerebral EEG consistently demonstrated the highest precision, recall/sensitivity, and specificity in classifying sleep stages across all training:test set ratios, followed by cerebellar EEG, which outperformed EMG. Combining the top 5 cerebellar EEG features with cerebral EEG features yielded better classification performance than combining EMG features with cerebral EEG features. Using the top 20 features, the model achieved mean area under the receiver operating characteristic curve values of  $0.98 \pm 0.08$ ,  $0.98 \pm 0.10$ , and  $0.99 \pm 0.07$  for wakefulness, non-rapid eye movement sleep stages, and rapid eye movement sleep stages, respectively. The cerebellum may play a unique and important role in sleep-wake regulation. Incorporating cerebellar EEG into polysomnography has the potential to enhance the accuracy and efficiency of sleep stage classification.

## 1. Introduction

Sleep is a vital physiological process that fulfills a pivotal function in human beings, primarily contributing to the restoration and consolidation of physical and cognitive functions [1]. However, sleep disturbances have become a pressing health issue in contemporary society, largely due to the fast-paced nature of modern life [2]. This has led to an increasing prevalence of sleep disorders, which stem from multiple factors such as societal influences, environmental conditions, and

various medical conditions or medications [3]. Furthermore, there has been a consistent decline in the global average duration of sleep [4]. The American Academy of Sleep Medicine (AASM) and the Sleep Research Society recommend that adults should aim for a daily sleep duration of 7–9 h [5]. Nevertheless, sleep deprivation remains a common issue among millions of individuals [6]. Studies have also underscored the deleterious effects of sleep deprivation, which include diminished alertness and cognitive memory, as well as an elevated risk of psychological disorders like depression and chronic diseases such as heart

\* Corresponding author.

E-mail address: [yuhang@scslyy.org.cn](mailto:yuhang@scslyy.org.cn) (Y. Hang).

disease and diabetes [7–9]. The World Health Organization estimates that sleep problems affect approximately one-third of the global population [10]. Consequently, the investigation of sleep patterns has become a significant issue in clinical medicine. Researchers commonly use electrophysiological signal acquisition and analysis to monitor sleep patterns, with the goal of understanding the sleep status and overall health of individuals, and providing timely interventions for associated ailments [11]. The quantitative and precise assessment of sleep quality plays a crucial role in the clinical research of sleep disorders. Sleep staging involves detecting sleep patterns of an individual throughout the night. Polysomnography (PSG) is the primary method used for signal acquisition in sleep staging, collecting various physiological signals such as Electroencephalogram (EEG), Electrooculogram (EOG), Electromyogram (EMG), Electrocardiogram, and oxygen saturation [12]. Following the collection of these physiological signals, sleep experts interpret and analyze the PSG recordings in 30-s intervals, assigning labels to each epoch based on the criteria proposed by AASM [13]. Each sleep stage is classified according to the AASM criteria, including five distinct sleep stages: wakefulness (W), Non-rapid eye movement (NREM, including N1, N2, N3), and Rapid eye movement (REM) [14]. Each sleep stage is linked to specific neurological activity within the brain, as the autonomic activity of the brain undergoes alterations throughout different sleep stages [15]. The process of manually classifying sleep recordings, which typically involves a substantial amount of data spanning approximately 7–8 h, is both laborious and time-consuming. Moreover, manual classification is highly subjective, as evidenced by studies indicating an approximately 83 % agreement rate between manual classifiers [16]. The emergence of feature extraction techniques has facilitated the utilization of carefully crafted manual features in conjunction with classical classifiers for the purpose of automatic sleep staging [17]. This approach partially mitigates the issue of the considerable expense associated with manual labeling. However, the efficacy of manual feature design heavily depends on expert knowledge, and the ultimate accuracy of classification is directly influenced by the quality of the features, rendering them incapable of adapting to dataset variations [18].

To address these issues, numerous experts and scholars have been actively engaged in research aimed at enhancing the precision of sleep staging and improving clinical efficiency. With the increasing utilization of artificial intelligence across diverse domains, a growing number of scholars have employed deep/machine learning techniques within the realm of automatic sleep staging [2,19]. Current research on automatic sleep staging can be categorized into two main approaches: single-channel and multi-channel. Single-channel primarily focuses on EEG, EOG, and EMG, with the majority of studies relying on single-channel EEG data [20]. In contrast, the multi-channel approach involves the integration of these three types of data [21]. The cerebellum plays a crucial role in various motor functions, including motor planning, execution, and learning [22,23]. However, individuals with cerebellar disorders not only encounter motor impairments but also experience non-motor symptoms [24]. Cerebellar malfunction can lead to alterations in the sleep-wake cycle and sleep disturbances [25]. Studies involving cats with surgically removed cerebellum have demonstrated reduced wakefulness and heightened REM sleep [26]. The present study indicates that individuals with primary sleep disorders typically exhibit a decrease in cerebellar volume [27]. These findings corroborate the hypothesis that the cerebellum plays a pivotal role in the sleep-wake regulation. Furthermore, the correlation between the cerebellum and sleep intimates that cerebellar EEG could potentially serve as a valuable instrument for enhancing the precision of automatic sleep stage classification. Consequently, the objective of this paper is to probe the feasibility of cerebellar EEG, in conjunction with cerebral EEG, EMG, and machine learning techniques, for sleep staging. This investigation harbors the potential to propel sleep research forward by furnishing novel insights and opening up new avenues for further exploration.

## 2. Materials and methods

### 2.1. Animals and experimental designs

Twenty-five male C57BL/6 mice, aged 3–5 months and weighing 20–25 g, were individually housed in the laboratory animal facility under a 12-h light-dark cycle. The lights were switched on at 08:00, and the mice were provided with ad libitum access to food and water. All procedures related to animal care and use were conducted in compliance with the guidelines and regulations established by the Institutional Animal Care and Use Committee at Sichuan Cancer Hospital. For EEG/EMG monitoring, each mouse was individually housed in a 10-inch diameter plexiglass cylinder equipped with bedding.

EEG and EMG electrode implantation surgeries were performed as previously described [28]. Mice were anesthetized using 2 % isoflurane in air, and their body temperature was maintained at 37 °C by a feedback-controlled thermoregulation system. The head was then secured in a stereotaxic apparatus, and a midline scalp incision was made to expose the skull. Three craniotomies were performed using a high-speed drill with a 0.25 mm diameter round burr. The first drilling site was located in the right primary motor cortex (coordinates relative to bregma: anterior-posterior (AP) +0.38 mm, medial-lateral (ML) +0.96 mm, and dorsal-ventral (DV) –1.5 mm from the dura). The second drilling site was in the right 6th cerebellar lobule (AP -6.96 mm, ML +1.44 mm, DV -1.5 mm from the dura), and the third was in the left 6th cerebellar lobule (AP -6.96 mm, ML -1.44 mm, DV -1.5 mm from the dura). Immediately following the craniotomies, perfluoroalkoxy (PFA) polymer-coated silver wire electrodes (bare diameter: 127 µm, coated diameter: 177.8 µm, A-M Systems) were implanted. An EEG recording electrode was inserted into the right primary motor cortex, and another EEG recording electrode was placed in the right 6th cerebellar lobule. A ground electrode was positioned in the left 6th cerebellar lobule. For EMG recordings, an EMG recording electrode and a reference electrode were carefully inserted into the neck muscles. All implanted electrodes were connected to an adapter, which was subsequently secured to the skull with dental acrylic. The surgical wound was closed with sutures, and the mice were allowed to recover in their home cages for a minimum of two weeks.

Subsequent recordings were conducted in 10-inch diameter plexiglass chambers. EEG/EMG signals were sampled at 500 Hz using a 0.5 Hz high-pass filter. Synchronous video recordings of freely moving mice were obtained over a 24-h period using a 4-channel EEG/EMG tethered system from Pinnacle Technology. The acquired signal was segmented into 30-s intervals, and each segment was filtered using a third-order Butterworth bandpass filter with cutoff frequencies of 0.5–40 Hz. The filtering process was implemented using a Python program (Python Software Foundation). In this study, spectrum analysis and time-frequency analysis were performed on the EEG segments to extract information in both the frequency and time-frequency domains using the Welch periodogram method. To standardize the data and facilitate comparison, the signals were normalized by calculating the difference between the EEG data and the mean of the dataset, then dividing by the standard deviation. For each segment, we computed the power spectral density (PSD) for the EEG signals from the cerebrum and cerebellum, as well as the EMG signals, across the 0–30 Hz range. The PSD was averaged at each frequency, resulting in a total of 90 features per segment, which comprehensively captured the distribution of power across the specified frequency bands. This approach allowed for an in-depth characterization of the electrophysiological patterns associated with different states, providing a robust set of features for further analysis.

### 2.2. Machine learning models

Extreme Gradient Boosting (XGBoost) is an optimized gradient boosting library, which combines weak learner models (decision trees) under the Gradient Boosting framework, allowing for the optimization

of a differentiable loss function. Key advantages of XGBoost include its ability to handle missing values, support parallel computing, and employ a sparsity-aware algorithm for sparse data. It also has a built-in cross-validation method and supports various objective functions for regression, classification, and ranking tasks. In this study, XGBoost was implemented using the Python library XGBoost (<https://github.com/dmlc/xgboost>).

Light Gradient Boosting (LGBBoost) is a gradient boosting framework that utilizes tree-based learning algorithms. It is designed to be distributed and efficient, offering advantages such as faster training speed, higher efficiency, lower memory usage, better accuracy, and the ability to handle large-scale data. LGBBoost uses the leaf-wise tree growth algorithm, which can reduce more loss than a level-wise algorithm when growing the same leaf. It also employs a histogram-based algorithm for computing the best split, which accelerates training. In this study, LGBBoost was implemented using the Python library LGBBoost (<https://github.com/microsoft/LightGBM>).

Categorical Boosting (CatBoost) is an open-source gradient boosting library. Based on gradient boosted decision trees, it provides categorical feature support and an innovative algorithm for processing categorical features. CatBoost can automatically handle categorical features without any explicit pre-processing. It utilizes a permutation-driven approach to calculate leaf values, which helps to reduce overfitting. CatBoost was implemented using the Python library CatBoost (<https://github.com/catboost/catboost>).

Support Vector Machine (SVM) are a set of supervised learning methods used for classification, regression, and outlier detection. The main concept of SVM is to construct a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for various tasks. SVM are effective in high-dimensional spaces and versatile in the choice of the decision function. In this study, SVM was implemented using the Python library scikit-learn (<https://scikit-learn.org/>).

Logistic Regression (LR) is a statistical model that employs a logistic function to model a binary dependent variable. Despite its name, it is used for classification rather than regression. LR estimates the probability of an event occurring based on a given dataset of independent variables. It is simple, easy to implement, and efficient to train. LR was implemented using the Python library scikit-learn.

Random Forest (RF) is an ensemble learning method for classification, regression, and other tasks that operate by constructing multiple decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. RF addresses the tendency of decision trees to overfit their training set. It is a fast, scalable, and easy-to-use algorithm. RF was implemented using the Python library scikit-learn.

Long Short-Term Memory (LSTM) network is a type of Recurrent Neural Network (RNN) designed to learn long-term dependencies, effectively addressing the vanishing gradient problem common in traditional RNN. It achieves this through a sophisticated cell state and a set of gates: the input gate, the forget gate, and the output gate. These mechanisms enable the LSTM to selectively retain or discard information over time, which is particularly advantageous for tasks involving sequential data, such as automatic sleep staging.

Convolutional Neural Network (CNN) is widely recognized for its effectiveness in image recognition and processing. A CNN typically consists of convolutional layers, pooling layers, and fully connected layers. The convolutional layer applies a set of learnable filters to the input, generating feature maps that highlight specific patterns within the data. Pooling layers then reduce the spatial dimensions of these feature maps, decreasing computational complexity while preserving essential information. The fully connected layers at the end of the network perform the final classification. In this research, the LSTM and CNN models were implemented utilizing the Keras library (<https://keras.io/>), with TensorFlow (<https://www.tensorflow.org/>) serving as the computational backend.

### 2.3. Automatic sleep staging

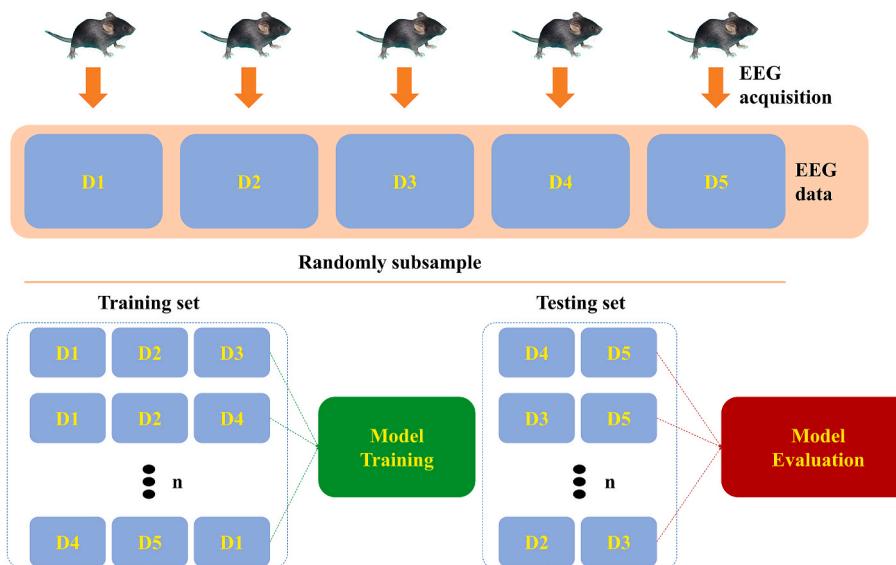
In this experiment, a total of 77,760 segments of sleep data were collected. The manual scoring of sleep stages was conducted according to established criteria, with specific differentiation made between NREM sleep, REM sleep, and wakefulness. The EEG and EMG signals were amplified and recorded using Sirenia Acquisition software (Pinnacle, USA) over a 24-h period. To ensure consistency and reliability in the scoring process, all scorers underwent thorough training prior to the analysis. Scoring was performed by a blinded investigator using Sirenia Sleep Pro software (Pinnacle, USA), as previously described [29, 30]. The epochs for scoring were set at 30 s, and visual scoring was complemented by spectral analysis through fast Fourier transform to assist in the identification of distinct sleep stages. To minimize potential bias, the scorers were kept unaware of the experimental conditions and had access only to the raw EEG and EMG data during the scoring process. The EEG processing methodology employed in this study is consistent with previously published literature [28]. Data normalization was performed by dividing the difference between the EEG data and the data mean by the standard error. Data analysis and implementation of the model algorithms were conducted using Python 3.9. At first, the data from each experimental mouse were individually partitioned into training and testing sets at ratios of 3:2. GridSearchCV or Bayesian optimization, incorporating 5-fold cross-validation, was employed to fine-tune the model hyperparameters for each mouse-specific dataset. The negative log-likelihood loss function was employed as the metric for assessing model performance across the various EEG data partitions within each mouse. After determining the optimal hyperparameters, we compared the performance of LGBBoost, XGBoost, CatBoost, SVM, LR, RF, LSTM and CNN models in classifying wakefulness, NREM, and REM sleep stages using cerebral EEG and EMG data. The best-performing model was selected based on these comparisons. Subsequently, we analyzed the effectiveness of cerebellar EEG, cerebral EEG, and EMG in classifying sleep stages using the best-performing model, and computed the feature importance for the top 20 features. To further evaluate the model, data from each experimental mouse were randomly subsampled 30 times, maintaining training and testing set ratios of 1:4, 2:3, and 3:2 for each iteration. The model was trained and evaluated on each subsampled dataset to assess its classification performance using various metrics, including accuracy, precision, recall, F1-score, and AUC. This repeated subsampling approach aimed to provide a more robust evaluation of the model performance (Fig. 1).

### 2.4. Statistical analysis

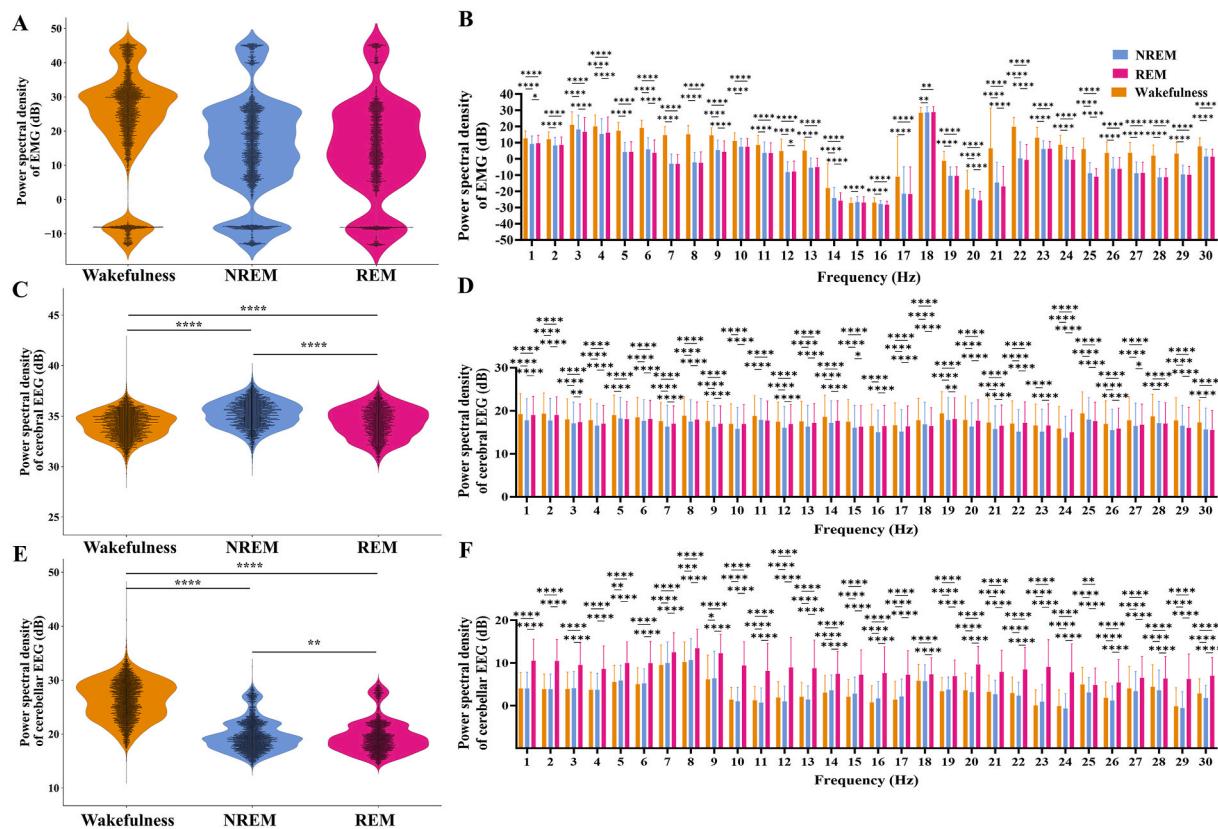
All statistical analyses were performed using GraphPad Prism 9 or Python. When applicable, data were analyzed using the Student's t-test or one-way/two-way analysis of variance, followed by the Tukey multiple comparison test. All tests were conducted using a two-tailed approach. The significance levels were set as follows: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , and \*\*\*\* $p < 0.0001$ . Sample sizes in this study were determined based on the Animal Research: Reporting In Vivo Experiments guidelines. Assuming a normal data distribution, all graphs presented in this study include both individual data points and means  $\pm$  standard error.

## 3. Results

We first compared the PSD of Cerebellar EEG, Cerebral EEG, and EMG in the wakefulness, NREM, and REM sleep stages (Fig. 2). The results revealed distinct patterns of PSD across the three states, with cerebral EEG exhibiting the highest PSD during NREM sleep and the lowest during wakefulness, while cerebellar EEG showed the opposite pattern. EMG PSD did not differ significantly across the three states. Frequency-specific analysis revealed significant differences in PSD among the three states for both cerebral and cerebellar EEG at a wide



**Fig. 1.** The electroencephalogram (EEG) dataset from each experimental mouse was randomly subsampled 30 times, maintaining training to testing set ratios of 1:4, 2:3, and 3:2 for each iteration. The model was trained and evaluated on each subsampled dataset. The repeated subsampling approach was employed to provide a more robust evaluation of the model performance across different data splits and sample sizes.



**Fig. 2.** Power spectral density (PSD) of cerebellar electroencephalography (EEG), cerebral EEG, and electromyography (EMG) during wakefulness, non-rapid eye movement (NREM), and rapid eye movement (REM) sleep stages (A, C, E). Cerebral EEG PSD was highest in NREM sleep, followed by REM sleep and lowest in wakefulness. Conversely, cerebellar EEG PSD was highest during wakefulness, followed by REM sleep and lowest in NREM sleep. EMG PSD showed no significant differences across the three states. Frequency-specific analysis (B, D, F) revealed significant differences in EMG PSD between NREM and REM sleep at 1, 3, 4, 6, 9, 12, 14, 20, 21, 22, and 25 Hz. Moreover, the state of wakefulness exhibited pronounced differences from both NREM and REM sleep across the entire frequency spectrum. Cerebral EEG PSD differed significantly among all three sleep states at most frequencies, except at 5, 10, 11, 16, 23, 28, and 30 Hz, where NREM and REM sleep did not differ. The PSD of cerebellar EEG exhibited significant differences between NREM and REM sleep stages at frequencies of 5 Hz, 7–17 Hz, and 19–30 Hz. Furthermore, the wakefulness stage showed distinct differences from both REM and NREM sleep stages across the entire frequency spectrum.

range of frequencies, while EMG showed differences at a limited number of frequencies. Cerebral EEG PSD differed significantly among all three states at most frequencies, except at a few specific frequencies where NREM and REM sleep did not differ.

We then compared the performance of LGBoost, XGBoost, CatBoost, SVM, LR, RF, LSTM and CNN models in classifying wakefulness, NREM, and REM sleep stages using Cerebral EEG and EMG (Fig. 3). Due to the long computation time for SVM on large datasets, we downsampled the data to 2000, 1000, and 200 samples for wakefulness, NREM, and REM sleep stages, respectively. The LR model had lower precision in classifying REM sleep compared to other models. The SVM model had lower recall/sensitivity in classifying NREM sleep, while SVM and RF models had lower recall/sensitivity in classifying REM sleep. The SVM model also had lower specificity in classifying wakefulness. The LSTM model performed poorly in classifying both NREM and REM sleep, and the CNN model was slightly less effective in classifying REM sleep. The LGBoost, XGBoost, and CatBoost models demonstrated similar performance levels. Notably, LGBoost and LR models were distinguished by their minimal computation time. Furthermore, the area under the receiver operating characteristic curve (AUC) was found to be larger for LGBoost, XGBoost, and CatBoost. Consequently, LGBoost was chosen for subsequent analyses.

We further analyzed the performance of cerebellar EEG, cerebral EEG, and EMG in classifying sleep stages using LGBoost with training: test set ratios of 3:2 (Fig. 4). Feature importance analysis revealed that cerebral EEG features dominated the top 20 rankings, with cerebellar EEG features contributing the second most, and EMG features contributing the least to the model.

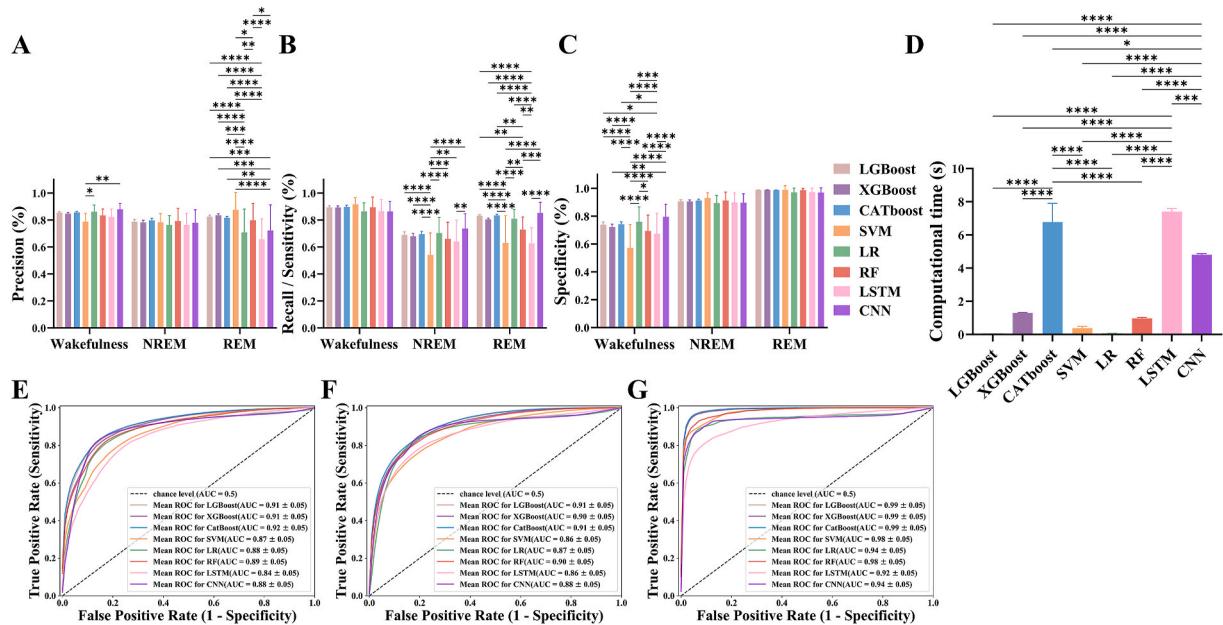
The performance of LGBoost models in classifying sleep stages using

cerebral EEG, cerebellar EEG, and EMG signals was evaluated using precision, recall/sensitivity, and specificity scores (Fig. 5). Cerebral EEG demonstrated the best overall performance in distinguishing wakefulness, NREM, and REM sleep stages, while cerebellar EEG outperformed EMG in most aspects, particularly in differentiating REM sleep. The results suggest that Cerebral and Cerebellar EEG signals contribute significantly to the precision of sleep stage classification.

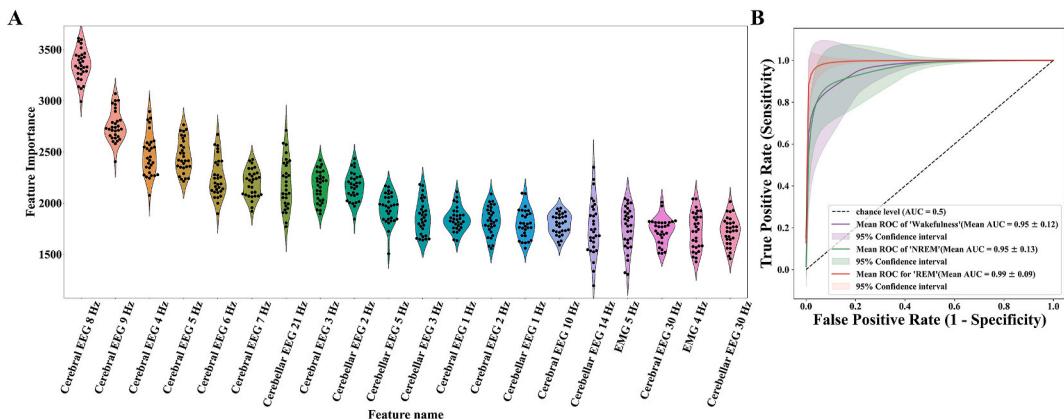
The performance of cerebral EEG, cerebellar EEG, and EMG in classifying sleep stages was evaluated using AUC metric across different training:test set ratios (Fig. 6). Cerebral EEG consistently demonstrated the highest performance in classifying wakefulness, NREM, and REM sleep stages across all ratios. Cerebellar EEG exhibited the second-best performance, showing notable improvements in classifying wakefulness at higher training:test set ratios, with performance comparable to that of cerebral EEG. EMG showed the lowest overall performance across all sleep stages and training:test set ratios. These findings highlight the primacy of cerebral EEG in sleep stage classification accuracy. Cerebellar EEG emerges as a potential adjunct, especially effective in wakefulness identification at higher training:test set ratios.

Feature importance analysis showed that among the top 20 features, 11 belonged to cerebral EEG, 7 to cerebellar EEG, and 2 to EMG (Fig. 4). We investigated the performance of two feature sets (Fig. 7): the top 5 cerebellar EEG features combined with cerebral EEG features (Cerebellar EEG & Cerebral EEG) and the top 5 EMG features combined with cerebral EEG features (EMG & Cerebral EEG). Cerebellar EEG & Cerebral EEG demonstrated better precision, recall/sensitivity, and specificity than EMG & Cerebral EEG in classifying all stages.

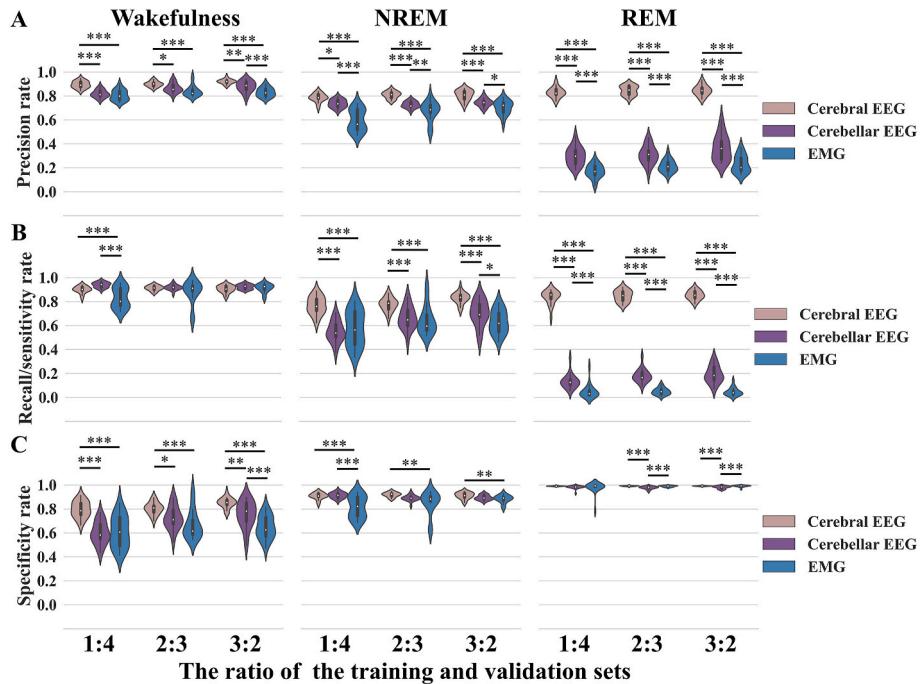
Lastly, we analyzed the performance of the 20 most important cerebellar EEG, cerebral EEG, and EMG features in classifying sleep



**Fig. 3.** Results of sleep staging using common machine learning algorithms. The precision (A), recall/sensitivity (B), and specificity (C) scores for the Light Gradient Boosting (LGBoost), Extreme Gradient Boosting (XGBoost), Categorical boosting (CatBoost), Support Vector Machines (SVM), Logistic Regression (LR), Random Forest (RF), Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) models are presented. The SVM model exhibits lower precision and specificity when classifying wakefulness and lower recall/sensitivity when classifying non-rapid eye movement (NREM) and rapid eye movement (REM). The LR model shows lower precision when classifying REM. The RF model shows slightly lower recall/sensitivity when classifying REM sleep and lower specificity than LR and CNN model when classifying wakefulness. The LSTM model exhibits lower specificity when classifying wakefulness and lower recall/sensitivity when classifying NREM sleep. Additionally, it has lower precision and recall/sensitivity when classifying REM sleep. The CNN model shows lower precision when classifying REM sleep (D) The computation time results for the LGBoost, XGBoost, CatBoost, SVM, LR, RF, LSTM and CNN models. The LGBoost and LR models have shorter computation times compared to the other models. The receiver operating characteristic (ROC) curves for the classification of wakefulness (E), NREM sleep (F), and REM sleep (G) by the LGBoost, XGBoost, CatBoost, SVM, LR, RF, LSTM and CNN models are depicted. The area under the ROC (AUC) for the classification of wakefulness by the LGBoost, XGBoost, and CatBoost models is slightly higher than that of the other models. The AUC for the classification of NREM sleep by the LGBoost and CatBoost models is slightly higher than that of the other models. Except for the LR, LSTM and CNN models, the other models exhibit good AUCs for the classification of REM sleep.



**Fig. 4.** Feature importance and receiver operating characteristic (ROC) curve of the Light Gradient Boosting model for wakefulness, non-rapid eye movement (NREM), and rapid eye movement (REM) sleep stage classification using cerebellar electroencephalogram (EEG), cerebral EEG, and electromyogram (EMG) features. (A) Top 20 feature importance rankings for cerebellar EEG, cerebral EEG, and EMG signals. Cerebral EEG features at 8, 9, 4, 5, 6, 7, 3, 1, 2, 10, and 30 Hz were among the top 20 most important features. In addition, cerebellar EEG features at 21, 2, 5, 3, 1, 14, and 30 Hz, as well as EMG features at 5 and 4 Hz, were also included in the top 20 most important features. (B) ROC curve of the LGBM model for classifying wakefulness, NREM, and REM sleep stages. The model achieved mean area under the ROC curve (AUC) values of  $0.95 \pm 0.12$ ,  $0.95 \pm 0.13$ , and  $0.99 \pm 0.09$  for classifying wakefulness, NREM, and REM sleep stages, respectively.



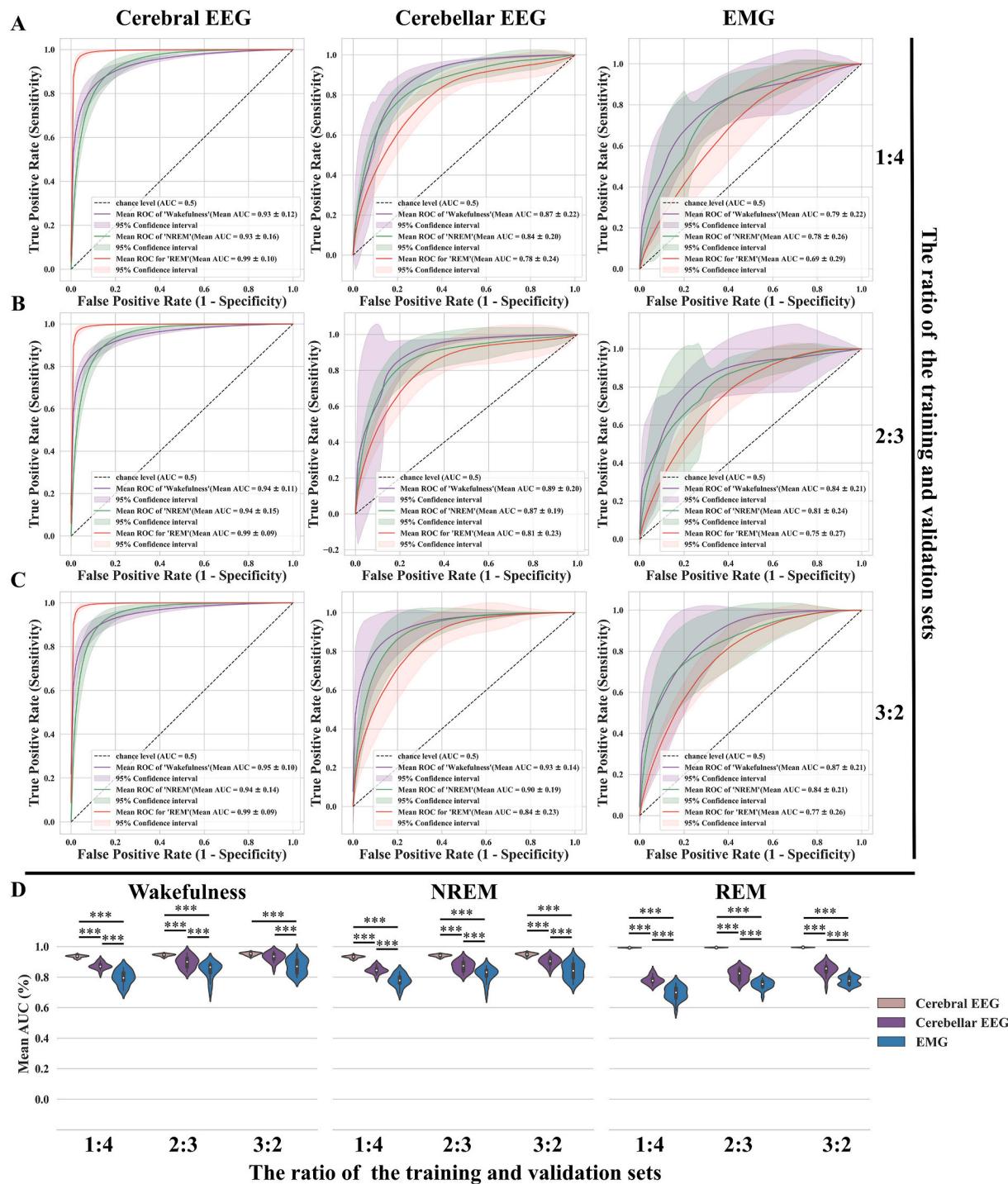
**Fig. 5.** The precision (A), recall/sensitivity (B), and specificity (C) scores for the Light Gradient Boosting models are presented. Cerebral electroencephalogram (EEG) has the highest precision rate and specificity in distinguishing wakefulness, followed by cerebellar EEG. Cerebellar EEG is higher than cerebral EEG and electromyogram (EMG) in recall/sensitivity rate in distinguishing wakefulness set at a ratio of 1:4. Cerebral EEG performs best in all aspects in distinguishing non-rapid eye movement (NREM) and rapid eye movement (REM). Cerebellar EEG has slightly higher precision rate than EMG in distinguishing NREM and REM sleep. In differentiating REM sleep, cerebellar EEG performs better than EMG in all aspects.

stages using LGBM with training:test set ratios of 1:4, 2:3, and 3:2 (Fig. 8). The model performance improved with increasing proportions of the training set, as evidenced by the increasing mean AUC values for distinguishing wakefulness, NREM, and REM sleep stages.

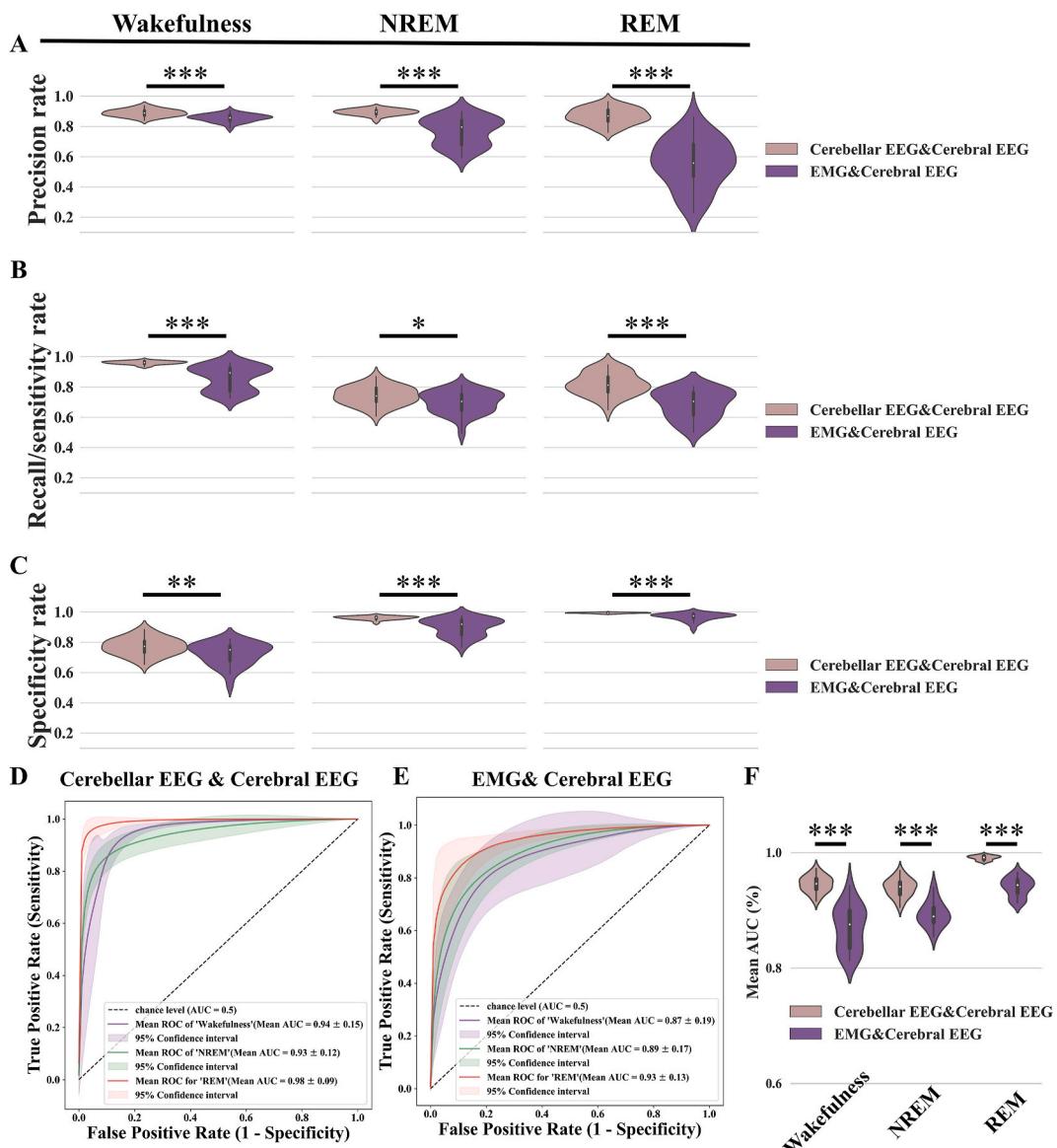
#### 4. Discussion

The use of cerebellar EEG recording represents a novel technique, requiring careful exclusion of alternative artifacts and signal sources. Clinical studies have demonstrated the feasibility of acquiring cerebellar EEG in humans while also establishing the potential for excluding

interference from EMG and occipital EEG [31]. The results of this study demonstrate that cerebellar EEG outperforms EMG in automatic sleep stage classification. The power spectral density of cerebellar EEG showed significant differences across wakefulness, NREM, and REM sleep stages, particularly in frequencies above 7 Hz. When using the LGBM model for sleep stage classification, cerebellar EEG surpassed EMG in terms of precision, recall/sensitivity, and specificity in distinguishing between sleep stages. Feature importance analysis also revealed that among the top 20 features contributing to sleep stage classification, 7 belonged to cerebellar EEG, while only 2 were from EMG. Combining the top 5 cerebellar EEG features with cerebral EEG



**Fig. 6.** Comparison of the performance of cerebral electroencephalogram (EEG), cerebellar EEG, and electromyogram (EMG) in classifying sleep stages using receiver operating characteristic (ROC) curves metric across different training test set ratios. (A) At 1:4 ratio, cerebral EEG achieved the highest mean area under the ROC curve (AUC) of  $0.93 \pm 0.12$ ,  $0.93 \pm 0.16$ , and  $0.99 \pm 0.10$  for classifying wakefulness, non-rapid eye movement (NREM), and rapid eye movement (REM) sleep stages, respectively. Cerebellar EEG demonstrated lower mean AUCs of  $0.87 \pm 0.22$ ,  $0.84 \pm 0.20$ , and  $0.78 \pm 0.24$ , while EMG showed the lowest mean AUCs of  $0.79 \pm 0.22$ ,  $0.78 \pm 0.26$ , and  $0.69 \pm 0.29$  for the respective sleep stages. (B) When the training test set ratio was increased to 2:3, cerebral EEG maintained its superior performance, with mean AUCs of  $0.94 \pm 0.11$ ,  $0.94 \pm 0.15$ , and  $0.99 \pm 0.09$  for classifying wakefulness, NREM, and REM sleep stages, respectively. Cerebellar EEG exhibited slight improvements, with mean AUCs of  $0.89 \pm 0.20$ ,  $0.87 \pm 0.19$ , and  $0.81 \pm 0.23$ , while EMG also showed marginal enhancements, with mean AUCs of  $0.84 \pm 0.21$ ,  $0.81 \pm 0.24$ , and  $0.75 \pm 0.27$  for the respective sleep stages. (C) At the highest training:test ratio of 3:2, cerebral EEG continued to outperform the other features, achieving mean AUCs of  $0.95 \pm 0.10$ ,  $0.94 \pm 0.14$ , and  $0.99 \pm 0.09$  for classifying wakefulness, NREM, and REM sleep stages, respectively. Notably, cerebellar EEG demonstrated a substantial improvement in classifying wakefulness, with a mean AUC of  $0.93 \pm 0.14$ , closely matching that of cerebral EEG. Cerebellar EEG also showed enhanced performance in classifying NREM and REM sleep stages, with mean AUCs of  $0.90 \pm 0.19$  and  $0.84 \pm 0.23$ , respectively. EMG exhibited the lowest performance, with mean AUCs of  $0.87 \pm 0.21$ ,  $0.84 \pm 0.21$ , and  $0.77 \pm 0.26$  for the respective sleep stages. (D) Cerebral EEG consistently demonstrated the highest mean AUCs in classifying wakefulness, NREM, and REM sleep stages across all training test set ratios, followed by cerebellar EEG, with EMG showing the lowest mean AUCs. The mean AUCs of cerebellar EEG in classifying wakefulness at the 3:2 training test set ratio was notably comparable to that of cerebral EEG.

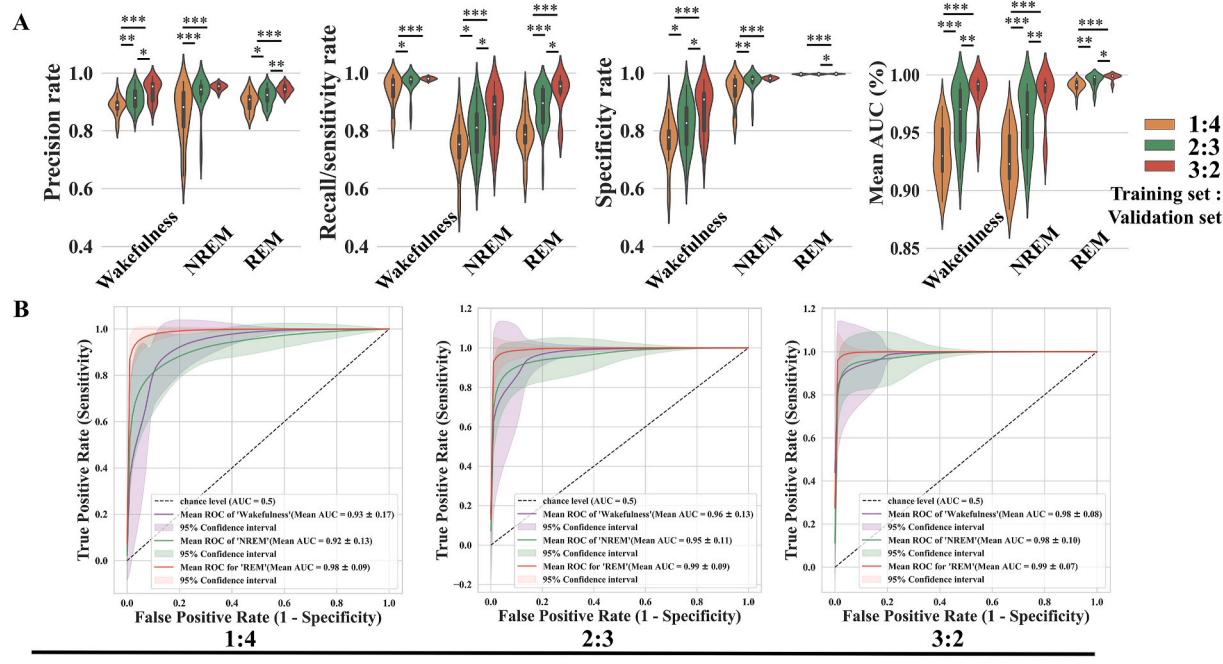


**Fig. 7.** Performance comparison of Light Gradient Boosting models in classifying sleep stages using cerebellar electroencephalogram (EEG) & cerebral EEG versus electromyogram (EMG) & cerebral EEG features. Precision (A), recall/sensitivity (B), and specificity (C) scores demonstrate that cerebellar EEG & cerebral EEG consistently outperforms EMG & cerebral EEG in distinguishing wakefulness, non-rapid eye movement (NREM), and rapid eye movement (REM) sleep stages. Receiver operating characteristic (ROC) curves (D, E) and mean area under the ROC curve (AUC) values (F) further confirm the superior performance of cerebellar EEG & cerebral EEG (Mean AUC:  $0.94 \pm 0.15$ ,  $0.93 \pm 0.12$ ,  $0.98 \pm 0.09$ ) compared to EMG & cerebral EEG (Mean AUC:  $0.87 \pm 0.19$ ,  $0.89 \pm 0.17$ ,  $0.93 \pm 0.13$ ) in classifying sleep stages.

features resulted in better sleep stage classification performance compared to the combination of the top 5 EMG and cerebral EEG features. These findings suggest that the cerebellum may play a unique and important role in sleep-wake regulation. The integration of cerebellar EEG into clinical practices for personalized medicine opens promising avenues by leveraging advancements in non-invasive data acquisition. This enhanced understanding of cerebellar function and connectivity could improve therapeutic approaches through targeted neuro-modulation techniques like deep brain stimulation or responsive neurostimulation [32]. Furthermore, insights into cerebellar physiology, particularly the ultrastructural synaptic changes during sleep and wakefulness, offer deeper comprehension of its role in learning and memory processes [33].

Traditionally, the cerebellum is commonly attributed to its involvement solely in the domains of motor functions and balance, but recent study provides evidence for its involvement in the regulation of sleep-

wake regulation [34]. Despite constituting a mere 10 % of the overall brain volume, the cerebellum boasts four times the number of neurons found in the cerebrum [35]. Evolutionary investigations further unveil a progressive increase in the ratio of granule cells to Purkinje cells within the cerebellum, with the human cerebellum exhibiting a ratio 3000 times greater [36]. Recent electrophysiological studies have further mapped sensorimotor representations within the human cerebellum, highlighting its crucial role in integrating sensory input with motor output during varying states of consciousness. Additionally, research has revealed that stronger delta-theta oscillations around visual stimuli indicate the function of cerebellum as a state estimator for rhythmic patterns, which may contribute to sleep stage differentiation [37]. The regulation of sleep and wakefulness is primarily governed by neural circuits comprising orexinergic neurons, melanocortin neurons in the lateral hypothalamus, cholinergic neurons in the basal forebrain [38], dopaminergic neurons in the ventral tegmental area, glutamatergic



The ratio of the training and validation sets

**Fig. 8.** Performance metrics of the Light Gradient Boosting model for wakefulness, non-rapid eye movement (NREM), and rapid eye movement (REM) sleep stages classification using the 20 most important cerebellar electroencephalogram (EEG), cerebral EEG, and electromyogram (EMG) features. (A) Precision, recall/sensitivity, specificity scores, and mean area under the receiver operating characteristic curve (AUC) for the LGBBoost model. As the training set proportion increased from 1:4 to 2:3 to 3:2, the classification performance metrics for wakefulness, NREM, and REM sleep stages progressively improved. (B) Receiver operating characteristic (ROC) curve of the LGBBoost model for wakefulness, NREM, and REM sleep stage. At the 1:4 training:test ratio, the model achieved mean AUCs of  $0.93 \pm 0.17$ ,  $0.92 \pm 0.13$ , and  $0.98 \pm 0.09$  for classifying wakefulness, NREM, and REM stages, respectively. At the 2:3 ratio, the mean AUCs were  $0.96 \pm 0.13$ ,  $0.95 \pm 0.11$ , and  $0.99 \pm 0.07$ . Furthermore, at the 3:2 ratio, the mean AUCs were  $0.98 \pm 0.08$ ,  $0.98 \pm 0.10$ , and  $0.99 \pm 0.07$  for the respective sleep stages.

neurons in the paraventricular nucleus of the thalamus, and GABAergic neurons in the substantia nigra reticulata. These circuits work in concert with the complex interplay between brain regions such as the hippocampal gyrus and the cerebellum, which contribute to spatial navigation, early visual processing stages, and overall sensory-motor integration during various states of consciousness [39]. The deep cerebellar nucleus (DCN) is the ultimate integrating and output nucleus of the cerebellum. Previous research has indicated that the DCN innervates the ventral thalamus, a brain region involved in the transition from NREM sleep to wakefulness [40]. Furthermore, the DCN exhibits interconnections with the hypothalamus, which houses various neurons responsible for the regulation of arousal and sleep [41]. Additionally, certain investigations have identified disparities in cerebellar electrical activity across distinct sleep states [26]. Moreover, functional magnetic resonance imaging revealed a notable augmentation in cerebellar activity during slow-wave sleep [42]. In the context of REM sleep, the cerebellum actively contributes to the generation of REM atonia and the phasic movement of the lateral rectus muscles of the eyes [43]. It has been proposed that, during REM sleep, the cerebellum governs autonomic inputs originating from the amygdala, periaqueductal gray, and thalamus, while concurrently manifesting parasympathetic and sympathetic outputs to the brainstem [44].

Recent literature suggests that the cerebellum plays a role in the regulation of sleep by establishing communication with various brain regions in the cerebral cortex through the superior cerebellar peduncle [41]. Conversely, the cerebral cortex can establish communication with the cerebellum through the pontine nuclei and the middle cerebellar peduncle [45]. Moreover, the fastigial nucleus at DCN is correlated with K-complexes and sleep spindles during NREM sleep. The activity of the cerebellum during sleep is likely influenced by climbing fibers and mossy fibers, which elicit complex spikes and simple spike activity in Purkinje cells, respectively [46].

Traditional PSG primarily relies on cerebral EEG and EMG for sleep staging. However, the results of this study indicate that incorporating cerebellar EEG into PSG has the potential to improve the accuracy and efficiency of automatic sleep stage classification. The neural activity patterns reflected by cerebellar EEG may contain important information distinct from cerebral EEG and EMG, which could help better differentiate between sleep stages. Current research suggests that the cerebellum is involved in the regulation of sleep, but the specific mechanisms remain to be fully elucidated [26,40–46]. The cerebellum has extensive anatomical connections with the cerebral cortex, brainstem, and other regions, which may allow it to interact with sleep-wake regulatory centers through these pathways [38,41,45]. This study suggests that analyzing the features of cerebellar EEG may help reveal the functional role of the cerebellum in sleep regulation and advance our understanding of the neural circuitry underlying sleep. Moreover, applying cerebellar EEG to clinical PSG examinations may provide new perspectives for the diagnosis and treatment of sleep disorders. Many neurodegenerative diseases, such as Parkinson's disease and Alzheimer's disease, are associated with significant sleep disturbances and often involve the cerebellum [28,47]. Analyzing changes in cerebellar EEG in these patients may help identify disease-specific patterns of sleep disruption and guide clinical management.

## 5. Conclusion

In conclusion, this study underscores the significance and potential benefits of incorporating cerebellar EEG in automatic sleep stage classification. The findings not only introduce a novel approach to optimizing PSG technology but also pave the way for extensive research into the role of cerebellum in sleep-wake regulation. The superior performance of cerebellar EEG compared to EMG in distinguishing between sleep stages suggests that it may contain unique neural activity patterns

that can enhance the accuracy and efficiency of sleep stage classification. To further validate these results, future studies should aim to replicate the findings using larger sample sizes and explore the clinical utility of cerebellar EEG in the diagnosis and treatment of sleep disorders and neurodegenerative diseases. Combining cerebellar EEG with other electrophysiological, imaging, and molecular biological techniques may provide a comprehensive understanding of the neural mechanisms underlying sleep and contribute to advancements in the field of sleep medicine. Given the promising results of this study, it is conceivable that the incorporation of cerebellar EEG acquisition and analysis in sleep EEG devices may be considered in the future. The widespread adoption of cerebellar EEG equipment has the potential to serve as a valuable tool in the early detection of sleep disturbances and the enhancement of diagnostic capabilities for sleep disorders. By providing a more accurate and detailed assessment of sleep stages, cerebellar EEG could facilitate the development of targeted interventions and personalized treatment plans for individuals with sleep-related issues.

#### CRediT authorship contribution statement

**Wang Manli:** Writing – original draft, Formal analysis, Conceptualization. **Guan Junwen:** Writing – original draft, Formal analysis. **Sun Tong:** Formal analysis. **Wang Junjie:** Writing – review & editing, Formal analysis. **Yuan Yikai:** Writing – review & editing, Formal analysis. **Zhou Yicheng:** Writing – review & editing, Formal analysis. **Zhang Yi:** Writing – review & editing, Formal analysis. **Yang Xiaoyu:** Writing – review & editing, Formal analysis. **Li Xuepei:** Writing – review & editing, Formal analysis. **Yang Jingguo:** Writing – review & editing, Formal analysis. **Zhou Xuebin:** Writing – review & editing, Formal analysis. **Yu Hang:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

#### Data availability statement

The datasets generated and/or analyzed during this study can be obtained from GitHub: <https://github.com/yuhang316/Automatic-Sleep-Stage-Classification>. All the sleep stage classification and analysis scripts are freely available for academic use on GitHub: <https://github.com/yuhang316/Automatic-Sleep-Stage-Classification>.

#### Ethics declarations

The animal experiments designed for this study were approved by the Institutional Animal Care and Use Committee at Sichuan Cancer Hospital, with the ethical approval number SCCHEC-04-2023-032. All methods were carried out in accordance with relevant guidelines and regulations to ensure the welfare and ethical treatment of the animals involved.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

The research was supported by the Chengdu technological innovation research and development project (2022-YF05-02034-SN).

#### References

- [1] I. Lambert, et al., Hippocampal interictal spikes during sleep impact long-term memory consolidation, *Ann. Neurol.* 87 (2020) 976–987.
- [2] F. L, et al., Automated sleep scoring: a review of the latest approaches, *Sleep Med. Rev.* 48 (2019).
- [3] H. Yu, M.-L. Wang, X.-L. Xu, R. Zhang, W.-D. Le, Changes in electroencephalography and sleep architecture as potential electrical biomarkers for Alzheimer's disease, *Chin. Med. J.* 134 (2021) 662–664.
- [4] M. Lancel, M. Stroebe, M.C. Eisma, Sleep disturbances in bereavement: a systematic review, *Sleep Med. Rev.* 53 (2020) 101331.
- [5] Y.-H. Wang, et al., Association of longitudinal patterns of habitual sleep duration with risk of cardiovascular events and all-cause mortality, *JAMA Netw. Open* 3 (2020) e205246.
- [6] C.G. Vecsey, et al., Sleep deprivation impairs cAMP signaling in the hippocampus, *Nature* 461 (2009) 1122–1125.
- [7] W.V. Lee, J. Shaman, Heat-coping strategies and bedroom thermal satisfaction in New York City, *Sci. Total Environ.* 574 (2017) 1217–1231.
- [8] B. Gee, et al., The effect of non-pharmacological sleep interventions on depression symptoms: a meta-analysis of randomised controlled trials, *Sleep Med. Rev.* 43 (2019) 118–128.
- [9] R.C. Cox, B.O. Olatunji, Sleep in the anxiety-related disorders: a meta-analysis of subjective and objective research, *Sleep Med. Rev.* 51 (2020) 101282.
- [10] A. Bosquillon de Jenlis, et al., Impacts of subchronic, high-level noise exposure on sleep and metabolic parameters: a juvenile rodent model, *Environ. Health Perspect.* 127 (2019) 57004.
- [11] E.A. Nofzinger, Neuroimaging and sleep medicine, *Sleep Med. Rev.* 9 (2005) 157–172.
- [12] M. Tamaki, et al., Complementary contributions of NREM and REM sleep to visual learning, *Nat. Neurosci.* 23 (2020) 1150–1156.
- [13] D.C. Lim, et al., Reinventing polysomnography in the age of precision medicine, *Sleep Med. Rev.* 52 (2020) 101313.
- [14] J.B. Stephansen, et al., Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy, *Nat. Commun.* 9 (2018) 5229.
- [15] L. D, D. Y, A motor theory of sleep-wake control: arousal-action circuit, *Annu. Rev. Neurosci.* 42 (2019).
- [16] R. Vallat, M.P. Walker, An open-source, high-performance tool for automated sleep staging, *Elife* 10 (2021) e70092.
- [17] P. Chirkos, et al., A review on current trends in automatic sleep staging through bio-signal recordings and future challenges, *Sleep Med. Rev.* 55 (2021) 101377.
- [18] A. Guillot, V. Thorey, RobustSleepNet: transfer learning for automated sleep staging at scale, *IEEE Trans. Neural Syst. Rehabil. Eng. Publ. IEEE Eng. Med. Biol. Soc.* 29 (2021) 1441–1451.
- [19] P. Chirkos, C.A. Frantzidis, P.T. Gkivoglou, P.D. Bamidis, C. Kourtidou-Papadeli, Automatic sleep staging employing convolutional neural networks and cortical connectivity images, *IEEE Transact. Neural Networks Learn. Syst.* 31 (2020) 113–123.
- [20] S. Zhao, et al., Evaluation of a single-channel EEG-based sleep staging algorithm, *Int. J. Environ. Res. Publ. Health* 19 (2022) 2845.
- [21] M. Perslev, et al., U-Sleep: resilient high-frequency sleep staging, *NPJ Digit. Med.* 4 (2021) 72.
- [22] A. Sathyanesan, et al., Emerging connections between cerebellar development, behavior, and complex brain disorders, *Nat. Rev. Neurosci.* 20 (2019) 298–313.
- [23] A.A. Sokolov, R.C. Miall, R.B. Ivry, The cerebellum: adaptive prediction for movement and cognition, *Trends Cognit. Sci.* 21 (2017) 313–332.
- [24] M. Ma, et al., Molecular layer interneurons in the cerebellum encode for valence in associative learning, *Nat. Commun.* 11 (2020) 4217.
- [25] C.B. Canto, Y. Onuki, B. Bruinsma, Y.D. van der Werf, C.I. De Zeeuw, The sleeping cerebellum, *Trends Neurosci.* 40 (2017) 309–323.
- [26] L.-B. Zhang, et al., Neuronal activity in the cerebellum during the sleep-wakefulness transition in mice, *Neurosci. Bull.* 36 (2020) 919–931.
- [27] E.Y. Joo, et al., Brain gray matter deficits in patients with chronic primary insomnia, *Sleep* 36 (2013) 999–1007.
- [28] H. Yu, et al., The electrophysiological and neuropathological profiles of cerebellum in APPswe/PS1ΔE9 mice: a hypothesis on the role of cerebellum in Alzheimer's disease, *Alzheimers Dement. J. Alzheimers Assoc.* 19 (2023) 2365–2375.
- [29] D. Yao, et al., Melatonin alleviates depression-like behaviors and cognitive dysfunction in mice by regulating the circadian rhythm of AQP4 polarization, *Transl. Psychiatry* 13 (2023) 310.
- [30] J.-F. Xie, et al., Neuropeptide S counteracts paradoxical sleep deprivation-induced anxiety-like behavior and sleep disturbances, *Front. Cell. Neurosci.* 12 (2018) 64.
- [31] S.-B. Wong, et al., Cerebellar oscillations in familial and sporadic essential tremor, *Cerebellum Lond. Engl.* 21 (2022) 425–431.
- [32] R.J. Piper, et al., Towards network-guided neuromodulation for epilepsy, *Brain* 145 (2022) 3347.
- [33] S.S. Loschky, et al., Ultrastructural effects of sleep and wake on the parallel fiber synapses of the cerebellum, *Elife* 11 (2022) e84199.
- [34] K.E. Cullen, Internal models of self-motion: neural computations by the vestibular cerebellum, *Trends Neurosci.* 46 (2023) 986–1002.
- [35] S. Herculanou-Houzel, The human brain in numbers: a linearly scaled-up primate brain, *Front. Hum. Neurosci.* 3 (2009) 31.
- [36] W. Lange, Cell number and cell density in the cerebellar cortex of man and some other mammals, *Cell Tissue Res.* 157 (1975) 115–124.
- [37] A. Leroy, et al., Short-term EEG dynamics and neural generators evoked by navigational images, *PLoS One* 12 (2017) e0178817.
- [38] S.O. Irmak, L. de Lecea, Basal forebrain cholinergic modulation of sleep transitions, *Sleep* 37 (2014) 1941–1951.
- [39] R. Ashida, et al., Sensory and motor electrophysiological mapping of the cerebellum in humans, *Sci. Rep.* 12 (2022) 177.

- [40] S. Honjoh, et al., Regulation of cortical activity and arousal by the matrix cells of the ventromedial thalamic nucleus, *Nat. Commun.* 9 (2018) 2100.
- [41] S. Çavdar, M. Özgür, Y. Kuvvet, H.H. Bay, The cerebello-hypothalamic and hypothalamo-cerebellar pathways via superior and middle cerebellar peduncle in the rat, *Cerebellum Lond. Engl.* 17 (2018) 517–524.
- [42] B. Song, J.-C. Zhu, A narrative review of cerebellar malfunctions and sleep disturbances, *Front. Neurosci.* 15 (2021) 590619.
- [43] M. Gadea-Ciria, J. Fuentes, Analysis of phasic activities in the lateral rectus muscle of the eyes (PALRE) during paradoxical sleep in chronic cerebellectomized cats, *Brain Res.* 111 (1976) 416, 415.
- [44] N.E. Dharani, The role of vestibular system and the cerebellum in adapting to gravitoinertial, spatial orientation and postural challenges of REM sleep, *Med. Hypotheses* 65 (2005) 83–89.
- [45] J.D. Cunhillos, I. De Andrés, Participation of the cerebellum in the regulation of the sleep-wakefulness cycle. Results in cerebellectomized cats, *Electroencephalogr. Clin. Neurophysiol.* 53 (1982) 549–558.
- [46] G.F. Marchesi, P. Strata, Mossy and climbing fiber activity during phasic and tonic phenomena of sleep, *Pflügers Archiv* 323 (1971) 219–240.
- [47] T. Li, W. Le, J. Jankovic, Linking the cerebellum to Parkinson disease: an update, *Nat. Rev. Neurol.* 19 (2023) 645–654.