Biomedical Signal Processing and Control 79 (2023) 104118

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Classification of seizure types based on multi-class specific bands common spatial pattern and penalized ensemble model

Duanpo Wu [a,b,c,*], Jie Li [a], Fang Dong [d,*], Junbiao Liu [e], Lurong Jiang [f], Jiuwen Cao [c,e], Xunyi Wu [g], Xin Zhang [h]

[a] School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China
[b] Zhejiang Provincial Key Laboratory of Information Processing, Communication and Networking, Hangzhou 310027, China
[c] Machine Learning and I-health International Cooperation Base of Zhejiang Province, Hangzhou Dianzi University, Hangzhou 310018, China
[d] College of Information and Electric Engineering, Zhejiang University City College, Hangzhou 310015, China
[e] Artificial Intelligence Institute, Hangzhou Dianzi University, Hangzhou 310018, China
[f] School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou, China
[g] Department of Neurology, Huashan Hospital, Fudan University, Shanghai 200040, China
[h] Department of Neurology, The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou 310003, China

ARTICLE INFO

Keywords:
Seizure type
Penalized strategy
Common spatial pattern
Joint approximation diagonalization

ABSTRACT

Seizures are induced by sudden abnormal discharge of brain electrical activity. Accurate judgment of seizure type is a prerequisite for correct treatment, rational use of drugs and prognosis. In this paper, a novel feature extraction algorithm based on multi-class specific bands common spatial pattern (MSBCSP), which applies the original CSP algorithm to multi-classification tasks with the help of joint approximation diagonalization (JAD), is presented. Complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) is applied to extract the energy of intrinsic mode functions (IMFs). Next, this study employs linear discriminant analysis (LDA) and random forest (RF) for feature selection. A penalized ensemble method was established to address imbalanced data and improve model generalization. Specifically, this study first introduces penalized strategy to handle imbalanced data, which punishes the majority classes by giving them lower weights while giving minority classes higher weights. Furthermore, this study introduces weighted voting strategy to combine logistic regression (LR) algorithm with LightGBM (LGB) algorithm. For evaluating the performance of proposed model, Temple University Hospital EEG dataset v1.5.0 is adopted, accuracy, kappa score, precision, recall and f1-score are computed with the value of 96.14%, 0.9335, 97.24%, 96.36% and 0.9679, respectively.

1. Introduction

Epilepsy is a chronic disease that causes sudden abnormal discharges of brain neurons [1]. There are about 65 million patients in the world. At present, although structural neuroimaging, functional neuroimaging and other technologies including diffusion optical imaging (DOI), magnetic resonance imaging (MRI) and position emission tomography (PET) have made great progress [2], clinicians still cannot quickly and accurately judge the progress of the disease due to the complexity of electroencephalogram (EEG) signals. Correct identification of seizure type can help clinicians to develop diagnosis and treatment plans, thus reduce the hidden dangers of future seizures and complications of epilepsy. In fact, most clinical decision are based on subjective judgement and experience of clinicians, rather than strict data analysis. Therefore, an efficient and cost-effective seizure type classification method has attracted wide concern of researchers in recent years.

Many seizure type classification methods have been investigated so far. Inung Wijayanto et al. [3] calculated the mean, variance, skewness, kurtosis, standard deviation, and interquartile range. They classified four types including complex partial seizure (CPSZ), focal non-specific seizure (FNSZ), generalized non-specific seizure (GNSZ) and tonic clonic seizure (TCSZ) by support vector machine (SVM), which achieved 95% of accuracy by using quadratic SVM kernel. Saputro et al. [4] observed the combination of three feature extraction methods, mel frequency cepstral coefficients (MFCC), hjorth descriptor and independent component analysis (ICA). The best result obtained by

* Corresponding authors.
E-mail addresses: wuduanpo@hdu.edu.cn (D. Wu), zjcxlijie@163.com (J. Li), dongf@zucc.edu.cn (F. Dong), liujunbiao@zju.edu.cn (J. Liu), jianglurong@zstu.edu.cn (L. Jiang), jwcao@hdu.edu.cn (J. Cao), dr.xunyiwu@163.com (X. Wu), 6510039@zju.edu.cn (X. Zhang).

https://doi.org/10.1016/j.bspc.2022.104118
Received 25 February 2022; Received in revised form 15 June 2022; Accepted 15 August 2022
Available online 26 August 2022
1746-8094/© 2022 Elsevier Ltd. All rights reserved.

combining MFCC and Hjorth descriptor can detect seizure type with 91.4% of average accuracy. In [5], a feature vector consisting of 10 time-domain components was selected, including the absolute sum, second order norm, third order norm, fourth order norm, infinity norm, maximum value, minimum value, variance, mean value and root mean squared value of the elements in each channel. They proposed the variable weight convolutional neural networks (VWCNNs) for the classification of seven seizure types, which achieved 100%, 97%, 98%, 88%, 88%, 71% and 100% of accuracy for each seizure type, respectively. The overall accuracy and F1 score reached 95.00% and 0.9400. S. Raghu et al. [6] attempted to classify seven variants of seizures with non-seizure EEG through the application of convolutional neural networks (CNN) and transfer learning by making use of the Temple University Hospital EEG corpus. They used pretrained network and classify using the support vector machine classifier by extracting image features. Khalid Abualsaud et al. [7] applied the discrete wavelet transform (DWT), including wavelet families and decomposition levels, to extract statistical features from each wavelet sub-band, which contains maximum, minimum, mean and standard deviation. In [8], they presented the compressive sensing integrated with the discrete cosine transform (DCT) and measurement matrix, which achieved 90% accuracy for noiseless data. Moreover, they combined DCT with the best basis function neural networks for EEG signals classification [9], which achieved an optimal classification rate of about 95%.

Most studies on seizure type classification mainly extract time or frequency domain features, but rarely pay attention to the influence of spatial domain features. One of the most effective algorithms of spatial domain is common spatial pattern (CSP) [10]. CSP transforms the data into a new time series in which the variance of one class of signals is maximized and the variance of the other class of signals is minimized [11]. However, feature extraction in EEG signals using CSP is highly dependent on frequency band selection. Since the frequency band is specific, it is difficult to determine the optimal filter frequency band. Improperly selected bands will most likely not be able to capture the changes in band power, reducing the efficiency of the CSP [12].

Several approaches have been proposed to solve the problem of manually selecting a specific frequency band for operation of the CSP algorithm. One method was the common space spectral pattern (CSSP). It optimized a simple filter using a single delayed sample in the CSP algorithm [13]. Another method was the common sparse spectral spatial pattern (CSSSP). It improved the CSSP algorithm by performing concurrent optimizations on arbitrary finite impulse response (FIR) filters in the CSP algorithm [14]. However, due to the inherent nature of the optimization problem, the solution of filter coefficients was also highly dependent on the selection of initial parameters [15]. Recently, an alternative method called subband generic spatial patterns (SBCSP) has been proposed, and it has been shown to achieve higher classification accuracy compared to CSSP and CSSSP in publicly available datasets [16]. Rather than optimizing a single arbitrary FIR filter in the CSP algorithm, SBCSP used a set of heirloom filters that decomposed the EEG measurement into multiple subbands. A spatial filter using the CSP algorithm was used for each of these subbands. After obtaining the subband scores, a recursive band removal or classification algorithm was used to fuse the subband scores. The fused subband scores were then classified using another classification algorithm [16]. All of the above methods are summarized in Table 1.

In this paper, we propose a multi-class specific bands common spatial pattern (MSBCSP) method. The contribution and novelty of our approach, which makes it different from regular spatial domain algorithms are as follows. Firstly, as different seizure types have different performance in different frequency bands, which may be overlapped, this paper extracts six frequency bands with overlapping bands, which are 3–5 Hz, 5–7 Hz, 4–8 Hz, 10–20 Hz, 14–20 Hz and 10–25 Hz, which is different from the traditional methods that extract some non-overlapping bands [17,18]. Secondly, to apply CSP from binary classification to multi-classification tasks, joint approximation diagonalization (JAD) is adopted to improve it. Thirdly, a penalized ensemble method was established to address imbalanced data and improve model generalization.

The structure of the rest of the paper is described below. In Section 2, the dataset used and the method of seizure type classification are described in detail. In Section 3, the standards for evaluating the model are introduced, and the experimental results are given according to the methods in Section 2. Finally, discussion and conclusion are presented in Sections 4 and 5.

2. Dataset and methodology

This study develops an efficient and cost-effective seizure type classification method based on the Temple University Hospital EEG (TUH-EEG) dataset. Fig. 1 shows the flows of seizure type classification algorithm based on MSBCSP and penalized ensemble model, whose details of each step are described in the following subsections. In particular, the orange box in Fig. 1 indicates that the innovations employed to improve the performance of seizure type classification. Different seizure types not only are differentiated in the time–frequency domain, but also have different performance in the spatial domain. We found that each seizure type may have more than one frequency band in the seizure stage, and some of them even overlap. Therefore, we select six frequency bands, including the overlapping bands, and then extract features in the spatial domain. The traditional CSP algorithm is applied to multi-classification tasks by JAD algorithm, which can obtain six corresponding spatial filters, so as to acquire the required spatial feature vectors, which are fused with the time–frequency domain features for classification.

2.1. Dataset description and preliminary analysis

TUH-EEG dataset, which contains 25,000 EEG data with detailed records of physiological information of patients including gender, age and pathological information. Recordings are sampled by 250 Hz, 256 Hz, 400 Hz and 1000 Hz with 32 channels. The dataset contains several seizure types, including FNSZ, GNSZ, Simple Partial Seizure (SPSZ), CPSZ, Tonic Seizure (TNSZ) and TCSZ. These types of epileptic seizure are contained in 1227 EEG files from 115 subjects, and each seizure event is recorded in detail with start time and end time by the professional doctor. In Table 2, the description of different type of seizure, the meaning of the markers, the number of events and the duration of every type are described in detail. In addition, the data is not artifact-free, so we take two measures to eliminate the influence of artifacts in Section 2.2.

This study mainly researches on the classification of FNSZ, GNSZ, SPSZ, CPSZ, TNSZ and TCSZ, whose typical waveforms are selected with 10 s long and shown in Fig. 2.

2.2. Data pre-processing

The original EEG signals are not standardized during data collection, so it is necessary to preprocess the original signals. Data preprocessing mainly includes four techniques: standardization of sampling rate, channel standardization, filtering and data discretization.

TUH-EEG contains signals of four different sampling frequencies. In order to ensure the same operation of different signals, all EEG signals are standardized to a sampling frequency of 100 Hz after being downsampled by interval values. For example, for the sampling frequency of 400 Hz, one point is reserved for every four sampling points.

The original signals are mainly recorded through the reference channel, and the average (AV) reference channel is obtained by subtracting the average of all channels from one channel, but this mode often brings artifacts, so it needs to be configured as bipolar (BP) channel, which can eliminate common interference from common reference electrodes.

Table 1
Algorithm summary.

Authors	Year	The proposed approach	The type of algorithm (s)
Inung Wijayanto et al. [3]	2019	EMD+SVM	Machine learning
Saputro et al. [4]	2019	SVM	Machine learning
Jia et al. [5]	2021	VWCNNs	Machine learning
S. Raghu et al. [6]	2020	CNN+Transfer learning	Machine learning
H. Ramoser et al. [10]	2000	CSP	Spatial feature extraction
C. Guger et al. [11]	2000	CSP	Spatial feature extraction
G. Dornhege et al. [12]	2006	CSP	Spatial feature extraction
S. Lemm et al. [13]	2005	CSSP	Spatial feature extraction
Bhatti et al. [14]	2019	CSSSP	Spatial feature extraction
Kirar et al. [15]	2016	SBCSP	Spatial feature extraction
Zhang et al. [16]	2015	SBCSP	Spatial feature extraction

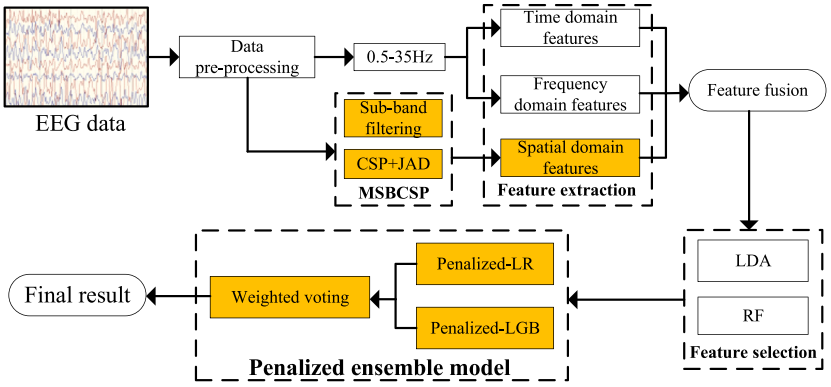


Fig. 1. The flowchart of seizure type classification algorithm based on MSBCSP and penalized ensemble model.

Table 2
Details of the dataset.

Seizure type	Description	Events	Duration (s)
CPSZ	Temporal lobe seizures or psychomotor seizures, which show partial seizures with varying degrees of consciousness disturbance	79	11179
FNSZ	Focal seizures with unspecified type	814	54086
GNSZ	Generalized seizures with unspecified type	255	27848
SPSZ	Partial seizures without disturbance	49	1941
TNSZ	Continuous tonic contraction of the whole body or part of the muscle, without clonic period	18	455
TCSZ	Continuous tonic contraction of the whole body or part of the muscle, with clonic period	12	626

In this paper, BP channel is obtained by a subtraction of two AV channels. Through this operation, 20 BP channels are finally obtained, including 'F7-T3', 'T3-T5', 'T5-O1', 'F8-T4', 'T4-T6', 'T6-O2', 'F3-C3', 'C3-P3', 'P3-O1', 'F4-C4', 'C4-P4', 'P4-O2', 'FZ-CZ', 'CZ-PZ', 'F7-F3', 'T3-C3', 'T5-P3', 'F4-F8', 'C4-T4' and 'P4-T6'. These 20 BP channels do not include the leads at positions FP1 and FP2, which can effectively avoid the artifact interference from the eye area. In addition, the Butterworth filter is used to filter the original signal from 0.5 Hz to 35 Hz for muscle artifact removal [19].

Through the above three steps, a time sequence of 100 Hz sampling rate and 0.5 Hz to 35 Hz frequency band is obtained. Seizure type classification needs to identify type of seizure from a period of time. But the duration of each patient's seizure is also different, so it is necessary to discretize the data to obtain several epochs. Moreover, the EEG signals during the seizure are divided into epochs of 5 s in this paper. Table 3 shows the epoch number of each seizure type.

Table 3
Epoch numbers of six seizure types.

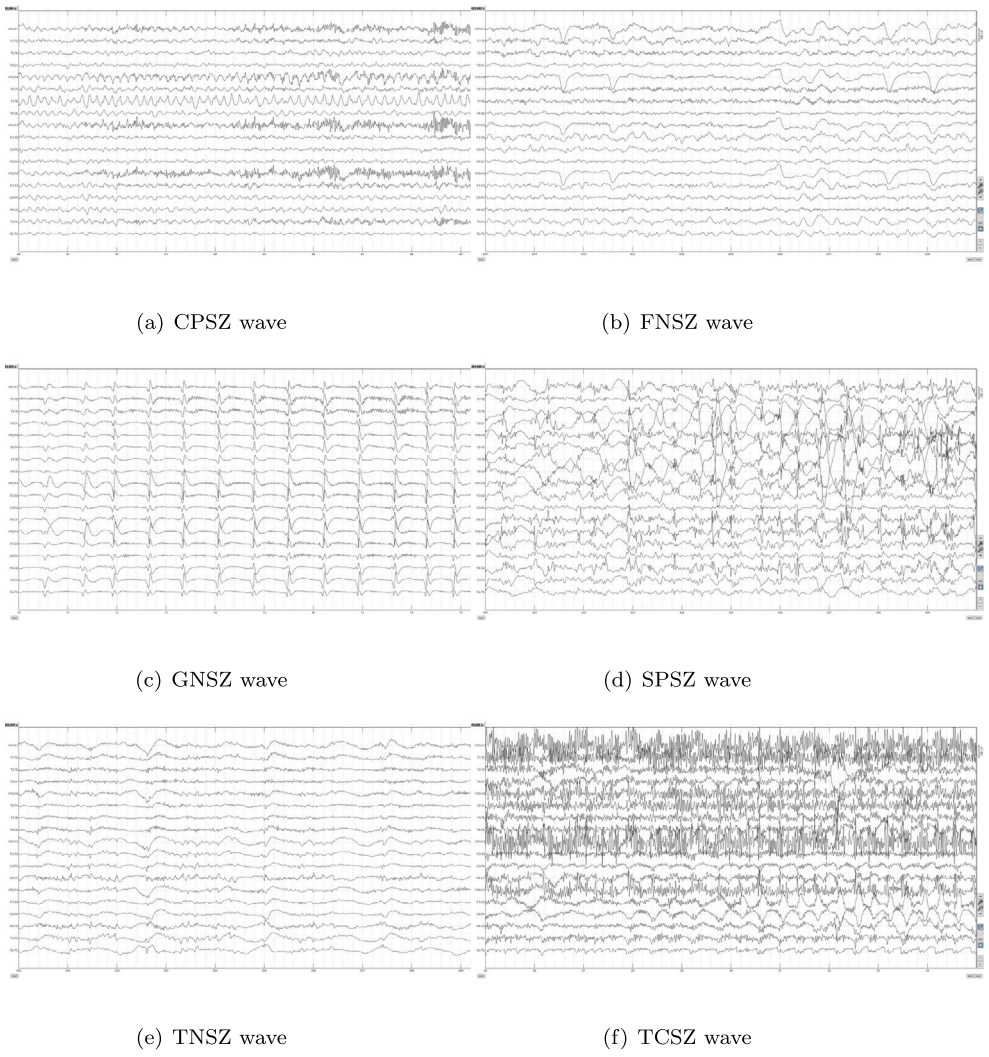
Seizure type	CPSZ	FNSZ	GNSZ	SPSZ	TNSZ	TCSZ
Epoch number	2764	13123	6852	462	108	145

2.3. Feature extraction

EEG signals not only reflect unique characteristics in the time and frequency domain, but also have different performances in the spatial domain. So this article use some algorithms including complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and MSBCSP to extract features, which are mainly include three parts: time domain feature, frequency domain feature and spatial domain feature. All of the features extracted from EEG signals are recapitulated in Table 4:

- In time domain, we extract six statistical features, including: maximum value, minimum value, average value, skewness, kurtosis and timeline length, with a total of 20 channels, so there are 6 * 20 features in total. Then the correlation coefficient between any 2 channels in 20 channels is extracted, which is 190 in total, and the Hjorth descriptor has 3 parts, including activity, mobility and complexity with a total of 3 * 20 features. Then we calculate the MI between any 2 channels in 20 channels with a total of 190 features.
- In frequency domain, we use the CEEMDAN algorithm to calculate the energy of 6 IMFs decomposed on 20 channels, with a total of 6 * 20 features. Correlation coefficients, Hjorth and MI are with the same number of features as in the time domain.
- In spatial domain, we use MSBCSP to obtain 6 spatial domain filters. For 20 channels, a total of 6 * 20 features can be obtained.

The calculations of some conventional features are shown in Table 5. For the six statistical features [20], which contain max, min, mean, kurtosis, skewness and time line length, x represents a 5 s epoch



(a) CPSZ wave (b) FNSZ wave

(c) GNSZ wave (d) SPSZ wave

(e) TNSZ wave (f) TCSZ wave

Fig. 2. The waveforms of each seizure type.

Table 4
Features extracted in this study.

Domain	Feature parameters	Count
Time	Max/Min/Mean/Kurtosis/Skewness/Time line length	6 × 20
	Correlation coefficient between any 2 channels in 20 channels	190
	Hjorth descriptor	3 × 20
	MI between any 2 channels in 20 channels	190
Frequency	Energy extracted through CEEMDAN	6 × 20
	Correlation coefficient between any 2 channels in 20 channels	190
	Hjorth descriptor	3 × 20
	MI between any 2 channels in 20 channels	190
Spatial	MSBCSP features	6 × 20

with 20 channels, N equals to 500, $\bar{x}$ represents the mean of x, and $x(i)$ represents the ith sample of x. For the correlation coefficient [20], assuming that $X_1 = [x_{11}, x_{12}, \ldots, x_{1N}]$, $X_2 = [x_{21}, x_{22}, \ldots, x_{2N}]$ are two channels of one EEG epoch, $\bar{X}_1$ and $\bar{X}_2$ represent the mean of X_1 and X_2, respectively. The Hjorth parameters, including activity, mobility and complexity [4], provide dynamic temporal information of the EEG signal, where $x'(i)$ and $x''(i)$ are the first and second derivatives of $x(i)$, σ, σ_1 and σ_2 are the standard deviation of $x(i)$, $x'(i)$ and $x''(i)$, respectively. Moreover, the calculation of CEEMDAN, MI and MSBCSP features will be described in detail in Sections 2.3.1–2.3.3.

Table 5
The formulas of conventional EEG features.

Feature type	Calculation
Max	$\max(x)$
Min	$\min(x)$
Mean	$\frac{1}{N}\sum_{i=1}^{N} x(i)$
Kurtosis	$\frac{\frac{1}{N}\sum_{i=1}^{N}(x(i)-\bar{x})^4}{(\frac{1}{N}\sum_{i=1}^{N}(x(i)-\bar{x})^2)^2}$
Skewness	$\frac{\frac{1}{N}\sum_{i=1}^{N}(x(i)-\bar{x})^3}{(\frac{1}{N}\sum_{i=1}^{N}(x(i)-\bar{x})^2)^{\frac{3}{2}}}$
Time line length	$\frac{1}{N-1}\sum_{i-1}^{N}\lvert x(i+1)-x(i)\rvert$
Correlation coefficient	$\frac{\sum_{i=1}^{N}(x_{1i}-\overline{X_1})(x_{2i}-\overline{X_2})}{\sqrt{\sum_{i=1}^{N}(x_{1i}-\overline{X_1})^2\sum_{i=1}^{N}(x_{2i}-\overline{X_2})^2}}$
Hjorth activity	$\frac{\sum_{i=1}^{N}(x(i)-\bar{x})^2}{N}$
Hjorth mobility	$\frac{\sigma_1}{\sigma}$
Hjorth complexity	$\frac{\sigma_2/\sigma_1}{\sigma_1/\sigma}$

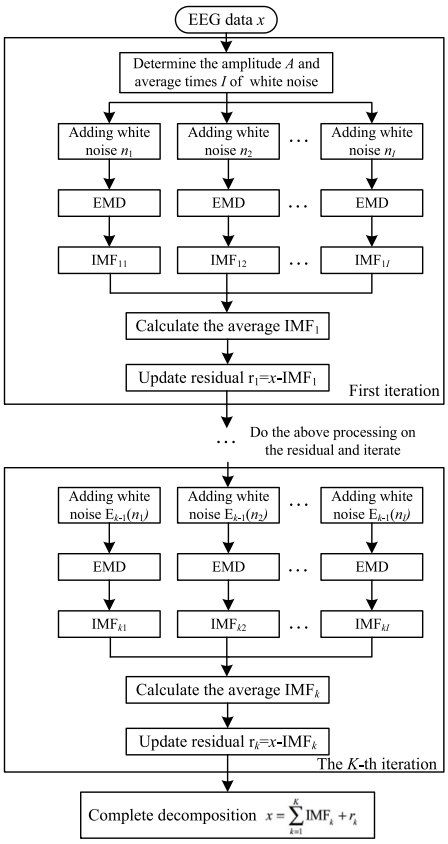


Fig. 3. The decomposition steps of CEEMDAN.

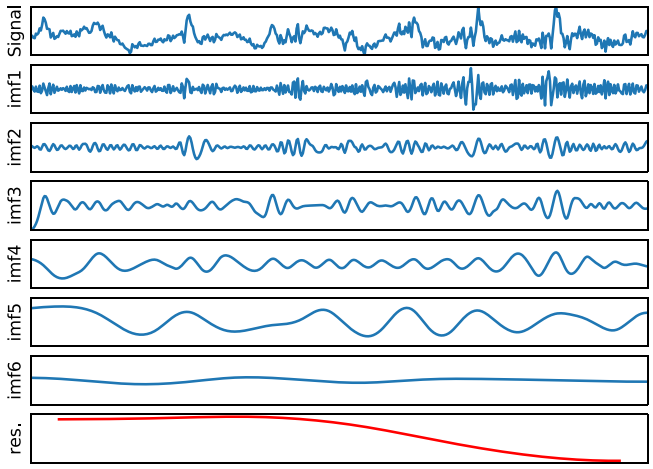


Fig. 4. The decomposition details of CEEMDAN.

2.3.1. Complete ensemble empirical mode decomposition with adaptive noise

Compared with wavelet algorithms such as wavelet packet transform (WPT), empirical mode decomposition (EMD) does not need to use wavelet basis, it can adaptively decompose the signal according to different time scales, and can well extract the characteristics of non-stationary signal changes. However, it is prone to the problem of modal aliasing. Aiming at the problem in the signal decomposition, we apply an improved algorithm of complete ensemble empirical mode decomposition with Adaptive Noise (CEEMDAN) algorithm, whose decomposition steps are shown in Fig. 3.

Repeat iterations until the obtained residual signal is a monotonic function, which cannot continue to be decomposed, and the algorithm ends. The number of Intrinsic Mode Functions (IMFs) obtained at this time is K, then the original signal x is decomposed as Eq. (1).

$$x = \sum_{k=1}^{K} IMF_k + r_k \tag{1}$$

In this study, the decomposition details of a certain segment of EEG signal is shown in Fig. 4. This EEG signal is decomposed into six IMFs and the residual part. Then calculate the energy of these six IMFs as part of the feature set, we can get 120 features through this way.

2.3.2. MSBCSP

As a spatial filtering algorithm, MSBCSP can be used to extract spatial features of multi-channel EEG signals. The details of MSBCSP algorithm are shown in Fig. 5. The MSBCSP algorithm is improved on the basis of the CSP algorithm, which adopts the sub-band filtering of six specific frequency bands and uses the JAD algorithm to extend it to six-classified application.

First, the raw EEG data is filtered with six frequency bands, and then performs a six-class supervised classification. Suppose X_1 to X_6 are six types of data samples respectively. Next, calculate the covariance matrix R_1 to R_6 of the six types of data as Eq. (2).

$$R_l = \frac{X_l X_l^T}{trace(X_l X_l^T)}, l = 1, 2, \ldots, 6. \tag{2}$$

where $trace(\cdot)$ represents the sum of the diagonal elements of the matrix, l represents lth seizure type.

Apply the principal component analysis method to orthogonally whiten the positive definite matrix R_i and perform diagonal processing:

$$R_l = U_l \Lambda_l U_l^T \tag{3}$$

where U_l represents the eigenvector matrix, and Λ_l is the diagonal matrix formed by eigenvalues in descending order. After conversion, the whitening value matrix P_l can be obtained as Eq. (4).

$$P_l = \frac{1}{\sqrt{\Lambda_l}} U_l^T \tag{4}$$

The next step is to perform the JAD algorithm on P_1 to P_6. Then Jacobi rotation is used to perform multi-matrix joint diagonalization, which use the second-order correlation delay matrix and the fourth-order cumulant characteristic matrix to obtain the corresponding rotation angles respectively. Before deciding whether to update the matrix, judge whether the obtained rotation angles are equal. If they are equal, update the matrix. After the above process, six rotation matrices are obtained, which are denoted as P_1' to P_6'.

Finally, the spatial filter is constructed by Eqs. (5)–(8).

$$S_l = P_l' R_l P_l'^T \tag{5}$$

$$S_l = B \Lambda_l B^T \tag{6}$$

$$\sum_{l=1}^{6} \Lambda_l = I \tag{7}$$

$$W_l = (B^T P_l')^T \tag{8}$$

where B and Λ_l represent the eigenvector matrix and diagonal matrix of S_l, respectively, I is the identity matrix, S_l represents transitional diagonalizable matrix, and W_l is the spatial filter obtained.

In this paper, a multi-class specific bands common spatial pattern (MSBCSP) method is proposed. The difference from the traditional CSP algorithm is that multiple sub-bands under professional opinions are used and apply CSP to multi-category through the JAD algorithm instead of the original two-category. The JAD algorithm is to obtain a rotation matrix P_l' so that all real symmetric matrices are approximately

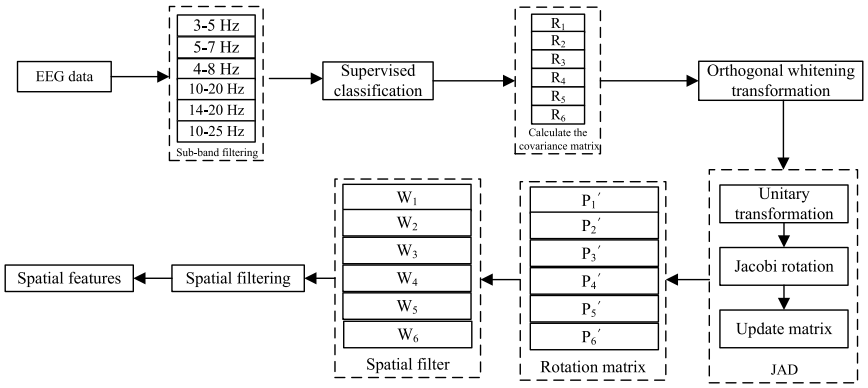


Fig. 5. The specific flow of the MSBCSP algorithm.

diagonal at the same time [21–23]. The specific process of the MSBCSP algorithm is shown in Fig. 5.

Through JAD algorithm, the related diagonal matrix λ_1, λ_2, …, λ_6 and eigenvector matrix V_1, V_2, …, V_6 of the six types of seizure can be obtained. According to the principle of the two-classified CSP algorithm, the relationship of the six eigenvector matrices $V_1 = V_2 = V_3 = V_4 = V_5 = V_6$ can be obtained by extension, and then the spatial filter can be designed accordingly, which is shown as Eq. (9).

$$Z_l = V_l^T P_l', l = 1, 2, \ldots, 6. \tag{9}$$

2.3.3. MI

For each EEG BP channel in each frequency band, the Shannon entropy is calculated by Eq. (10).

$$H(a) = -\sum_{i=1}^{n} P(a_i) \ln P(a_i) \tag{10}$$

where $P(a_i)$ represents the relative probability of a signal from channel a falling in histogram bin i, and n is the number of histogram bins. If a histogram bin is empty, $P(a_i) \ln P(a_i)$ is zero.

Use of the same histogram bin parameters across all frequency bands to calculate entropy and MI is associated with decreased sensitivity to individual subject and group differences within each frequency band. Thus, this study opts to use histogram parameters optimized for each frequency band. EEG signals falling outside the histogram amplitude range are not included in calculation of entropy and MI.

Joint entropy for each pair of channels in each frequency band can be calculated by Eq. (11).

$$H(a, b) = -\sum_{i=1}^{n} \sum_{j=1}^{n} P(a_i, b_j) \ln P(a_i, b_j) \tag{11}$$

where $P(a_i, b_j)$ is a 2-dimensional normalized probability distribution histogram of the EEG signals from channels a and b, which is the relative probability of a signal from channel a falling in histogram row i and the contemporaneous signal from channel b falling in histogram column j. Joint entropy calculation uses the same histogram bin parameters in each dimension for each frequency band. If a histogram bin is empty, $P(a_i, b_j) \ln P(a_i, b_j)$ is zero. And the calculation formula of MI is shown as Eq. (12).

$$MI(a, b) = H(a) + H(b) - H(a, b) \tag{12}$$

2.4. Feature selection

Through the previous step, a large number of features can be extracted, which will not only bring about redundancy of features, but also reduce the running efficiency of the program, so it is necessary to reduce the dimensionality of the original features.

Table 6
The experimental results of validation under different k values.

Value of k	AC	KA
2	95.86%	0.9286
3	95.87%	0.9290
5	**96.14%**	**0.9335**
8	95.89%	0.9292
10	95.82%	0.9280

In this paper, linear discriminant analysis (LDA) is used for feature dimensionality reduction, which is also known as Fisher Linear Discriminant. The feature data is six-dimensional, there are six categories in total, then the dimensions after LDA are 1, 2, …, 5 dimensions, so this paper forms five dimensional feature. In addition, visualizing the contribution of all the features is also needed so that the importance of each feature can be obtained for subsequent improvements and operations, instead of selecting a certain number of features. In this study, Random Forest (RF) algorithm is applied to calculate the contribution of all the features. The feature vectors are selected with contribution more than 0.1 and the feature vectors obtained by LDA as two feature sets to the subsequent classifier.

2.5. Penalized ensemble model

Fig. 6 shows the diagram of penalized ensemble model. First, this paper uses penalized strategy to ameliorate the performance of LR algorithms and LGB algorithms, respectively, and Section 2.5.1 shows the details of the application of the penalized strategy in seizure type classification. Second, k-fold cross validation method is used to train and test model that can avoid the random problems caused by a train-test-split method. The experimental results of cross validation under different k values are shown in Table 6. It can be found that when $k = 5$, AC and kappa score are the highest. Particularly, the studies used for comparison in Section 3.5 all use 5-fold cross-validation, so we use 5-fold cross validation finally. Data are arbitrarily divided into 5 disjoint sections, 4 folds are used to train proposed model, and the remaining 1 fold is utilized to test the performance of the proposed model. Finally, the average of 5 iterations determines the performance of proposed model. In order to facilitate comparison, this paper uses weighted voting strategy to enhance the performance of classification and Section 2.5.2 shows the details of the application of the penalized strategy in seizure type classification.

2.5.1. Penalized strategy

Data analysis in Table 3 shows that the samples are imbalanced between six types of seizure, which misleads the algorithm of seizure type classification to cause biased consequences, especially for the minority. This study puts forward an effective penalized strategy to improve

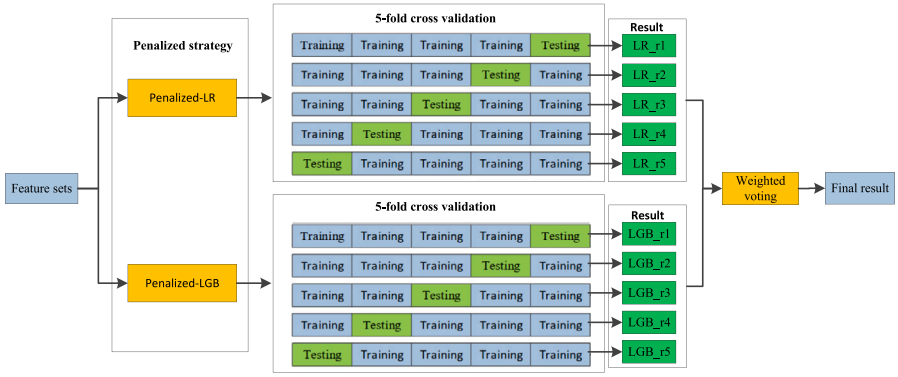


Fig. 6. Penalized ensemble model.

the performance of seizure type classification [24–26]. Expressly, this strategy punishes the majority by giving a lower weight, comparing with the minority. The weights of different classes can be adjusted by Eq. (13).

$$W_l = \beta_l \cdot \frac{NT}{NC_l}, (l = 1, 2, \ldots, 6) \tag{13}$$

where NT is the amount of epochs of all seizure types used in the dataset, NC is the amount of epochs of lth class, and β_l is the weight of the control coefficient, which can be adjusted by grid search method to further improve the performance of classification.

2.5.2. Weighted voting strategy

With the view of acquiring better predictive performance than constituent learning algorithms, ensemble machine learning methods are adopted. Typically, the greater the difference between the individual models that constitute the ensemble model, the better the ensemble model works. Therefore, this study uses the weighted voting strategy to combine the commonly used linear model LR algorithm with the nonlinear model LGB algorithm with excellent performance on multiple public datasets [27,28]. In addition, this study uses two feature sets for LR model and LGB model training, which further increases the differences between separate models to improve the performance of the ensemble model. Eq. (14) shows formula of the weighted voting strategy.

$$\begin{aligned} P(\mathbf{x}) &= C_{\arg\max_l \sum_{m=1}^{2} \gamma_m h_m^l(\mathbf{x})} \\ s.t. &\sum_{m=1}^{2} \gamma_m = 1 \end{aligned} \tag{14}$$

where $\mathbf{x}$ represents the feature vector of an epoch, and $P(\mathbf{x})$ represents the prediction result of the weighted voting strategy on $\mathbf{x}$; $h_m^l(\mathbf{x})$ represents the probability that the result of lth class predicted by the learning algorithm h_m; γ_m represents the weight of h_m.

3. Experiments and results

3.1. Performance evaluation

To appraise and contradistinguish the performance of different methods, confusion matrix (CM) is a commonly used method including TN, FP, FN, TP, which represent true negative, false positive, false negative and true positive, respectively. Table 7 shows the structure of CM. Several statistical parameters are computed including accuracy, kappa score, precision, recall and F1-score [29] based on CM, which are denoted as AC, KA, P, RE and $F1$, respectively. AC is the percentage of correct predicted samples, and the expression is shown in formula Eq. (15). Although AC can determine the overall classification performance, it cannot guarantee that the classification effect of each category is good while the samples are imbalanced. Therefore, KA, which is calculated according to Eqs. (16) and (17), is introduced to measure the results, where a_l represents the amount of positive samples in the lth category, b_l represents the amount of prediction in the lth category, and n indicates the amount of all samples. The formula of P and RE are shown as Eqs. (18) and (19). However, AC and RE are in conflict with each other and cannot achieve double high. So $F1$ is introduced, and the formula is shown as Eq. (20).

Table 7
Structure of CM.

Actual status	Predicted status	
	Negative	Positive
Negative	TN	FP
Positive	FN	TP

$$AC = \frac{TP + TN}{TP + FN + FP + TN} \tag{15}$$

$$KA = \frac{AC - p_e}{1 - p_e} \tag{16}$$

$$p_e = \frac{\sum_{l=1}^{6} a_l \cdot b_l}{\sum_{l=1}^{6} a_l} \tag{17}$$

$$P = \frac{TP}{TP + FP} \tag{18}$$

$$RE = \frac{TP}{TP + FN} \tag{19}$$

$$F1 = \frac{2 \cdot P \cdot RE}{P + RE} \tag{20}$$

3.2. Contribution of features

Through the RF algorithm, the contribution of all feature vectors can be calculated. First of all, it can be seen from Fig. 7 that the number of eigenvectors in each type of feature vector with a contribution degree greater than 0.001. Among them, the feature vectors belonging to the correlation coefficient and MSBCSP are the most with 28 and 24 respectively. When calculating the average number of features in each category, it can be found that the average contribution of MSBCSP is the highest with 0.2, and CEEMDAN is the second with 0.1667.

The average contribution of each type of feature is shown in Fig. 8. The average contribution of MSBCSP is only lower than the statistical characteristics. Then the average contribution of the features are obtained in the time domain, frequency domain and spatial domain at the same time. The average contribution of the 120 features in the spatial domain is the highest of the three as shown in Fig. 9, which can also reflect the effectiveness of the algorithm in this paper.

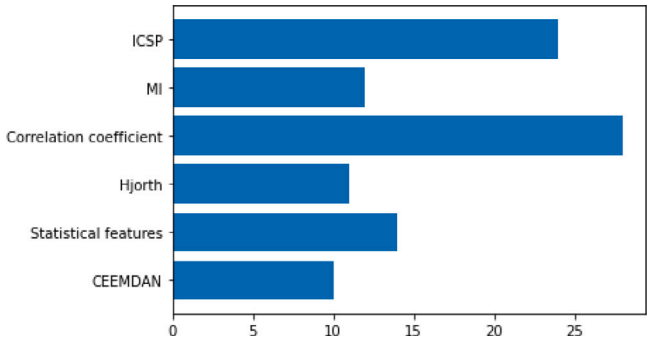


Fig. 7. The number of eigenvectors with a contribution degree greater than 0.001 in each type of feature vector.

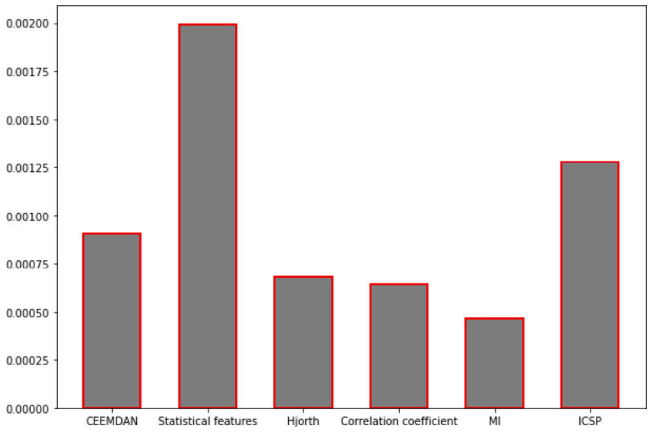


Fig. 8. The average contribution of each type of feature.

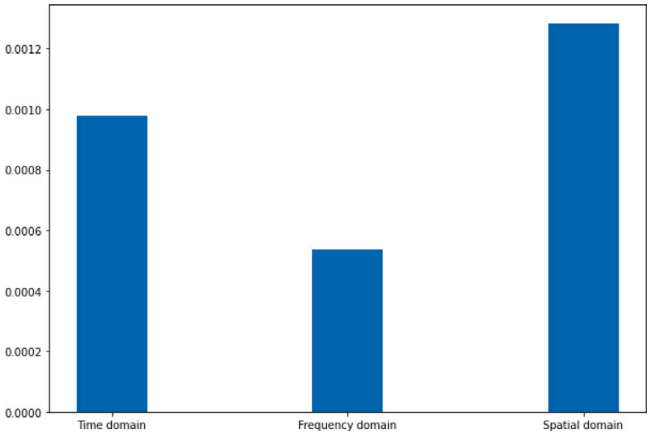


Fig. 9. The average contribution of features in time domain, frequency domain and spatial domain.

3.3. Impact of CEEMDAN, Hjorth descriptor, MI and MSBCSP

In order to deal with the complexity of seizure types, this paper particularly extracts Hjorth descriptor, MI, CEEMDAN and MSBCSP, whose contributions are shown in Fig. 10. By comparing Fig. 10(a) with Figs. 10(b)–10(e), it can be clearly seen that after adding the four features separately, the classification accuracy of each seizure type has been improved to varying extent. Especially after adding only MSBCSP features, the number of correctly classified samples from each class has been significantly increased. The features of CEEMDAN contributed the most to the differentiation of CPSZ. Hjorth and MSBCSP are able to maximize the differentiation of FZSZ from other seizure types. When all features are added, it can be seen that the accuracy of each class is further increased, especially CPSZ, FPSZ and GNSZ. Moreover, the number of misidentifications of one class as another is significantly reduced. In particular, it should be mentioned that this dataset is biased towards FNSZ and GNSZ due to the imbalanced number of events. Therefore, it is very difficult to distinguish the minority class. However, after using the method proposed in this paper, the improvements in discrimination for the minority classes (TCSZ and TNSZ) are quite large by comparing Fig. 10(a) with Fig. 10(f), from 122 and 91 to 139 and 104, respectively.

Table 8
Experiments list.

Experiment	Algorithm	Strategy
1	LR	–
2	LGB	–
3	LR	Penalized
4	LGB	Penalized
5	LR + LGB	Penalized + weighted voting

Table 9
Performance of experiment 1–5.

Experiment	Evaluation metrics				
	AC	*KA*	*P*	*RE*	*F*1
1	95.38%	0.9215	95.41%	95.38%	0.9539
2	95.42%	0.9221	88.34%	95.42%	0.9206
3	95.82%	0.9284	96.79%	95.83%	0.9554
4	95.75%	0.9268	95.79%	95.76%	0.9441
5	96.14%	0.9335	**97.24%**	96.36%	**0.9679**

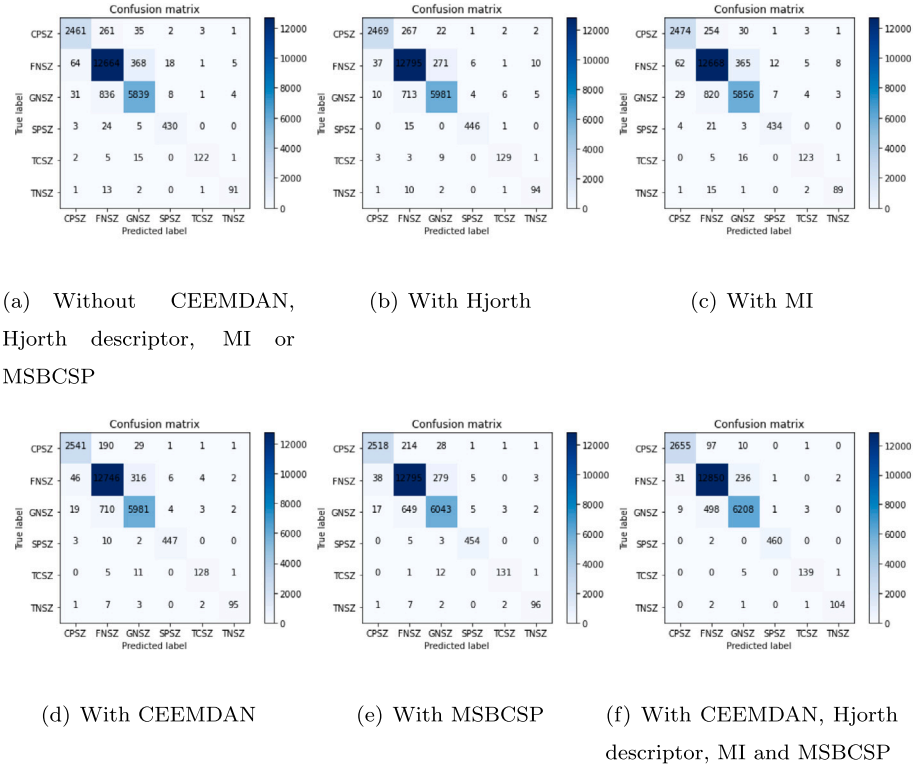
3.4. Outcomes of penalized ensemble model

Table 8 shows the differences among 5 experiments, which are performed to demonstrate the advantages of penalized strategy in dealing with imbalanced data. Especially, the method of 5-fold cross validation is used with grid search to determine that the β_1 and β_2 of LR algorithm in Experiment 3 are both 1, the β_1 and β_2 of LGB algorithm in Experiment 4 are 1 and 8, respectively.

Table 9 shows the evaluation metrics of Experiment 1 to Experiment 5. By observing Table 9, it can be found that the penalized strategy significantly improves the P and $F1$ of seizure type classification. Similarly, Table 9 also shows that penalized strategy can significantly improve the P of LGB algorithm. As a result, the penalized strategy improves the P of the LR model and the LGB model from 95.41% and 88.34% to 96.79% and 95.79%, respectively, which confirms the superiority of the penalized strategy in dealing with imbalanced data. Experiment 5 implements the proposed penalized ensemble model on seizure type classification, which first uses the penalized strategy to improve the P of the LR model and LGB model to process the imbalanced data, and then further introduces the weighted voting strategy to combine the two separate models so as to further improve the performance of seizure type classification. To make full use of the advantages of the ensemble model, 5-fold cross validation is employed to find the appropriate values of γ_1 and γ_2, which are 0.6 and 0.4, respectively.

3.5. Comparison of different methods and studies

The performance of some existing seizure type classification studies [3,4,6,30–33] are summarized in Table 10. As illustrated in Table 10, the proposed model has significant improvements in KA (0.9335) and P (97.24%) over most other existing seizure type classification works and other commonly used models established in this study. In Table 10, these seven studies use only single model.



(a) Without CEEMDAN, Hjorth descriptor, MI or MSBCSP (b) With Hjorth (c) With MI

(d) With CEEMDAN (e) With MSBCSP (f) With CEEMDAN, Hjorth descriptor, MI and MSBCSP

Fig. 10. Confusion matrixes comparison on the proposed model.

Table 10
Performance comparison of proposed model with other works/models.

Works (Models)	Year	Datasets	AC	KA	P	RE	$F1$
SVM [3]	2019	TUH-EEG	95.00%	–	–	–	–
SVM [4]	2019	TUH-EEG	91.40%	–	–	–	–
CNN and transfer learning [6]	2020	TUH-EEG	88.30%	–	–	–	–
Neural memory networks [30]	2019	TUH-EEG	–	–	–	–	0.9450
Corresponding hyperparameters [31]	2020	TUH-EEG	–	–	–	–	0.9010
Deep learning framework [32]	2020	TUH-EEG	–	–	–	–	0.9400
VWCNNs [33]	2021	TUH-EEG	95.00%	–	–	–	0.9400
SVM on proposed model	–	TUH-EEG	94.45%	0.9039	95.19%	94.46%	0.9442
Random forest on proposed model	–	TUH-EEG	94.41%	0.9034	94.97%	94.41%	0.9438
KNN on proposed model	–	TUH-EEG	94.52%	0.9055	95.00%	94.53%	0.9450
Proposed model	–	TUH-EEG	**96.14%**	**0.9335**	**97.24%**	**96.36%**	**0.9679**

Moreover, study [3] and Saputro et al. [4] just classify the seizure data into three types and four types, respectively. Therefore, these researches lack generalization. Particularly, these seven literature use single evaluation criteria so that the performance of the algorithm cannot be comprehensively evaluated. However, the proposed model classify the data into six types, which use the penalized strategy and weighted voting strategy. Particularly, this model obtains the best AC (96.14%), KA (0.9335), P (97.24%), RE (96.36%) and $F1$ (0.9679). In summary, the proposed penalized ensemble model achieves a high performance in all evaluations, and KA is obviously better than other existing seizure type classification studies, which indicates the strength and suitability of the proposed model in seizure type classification.

4. Discussion

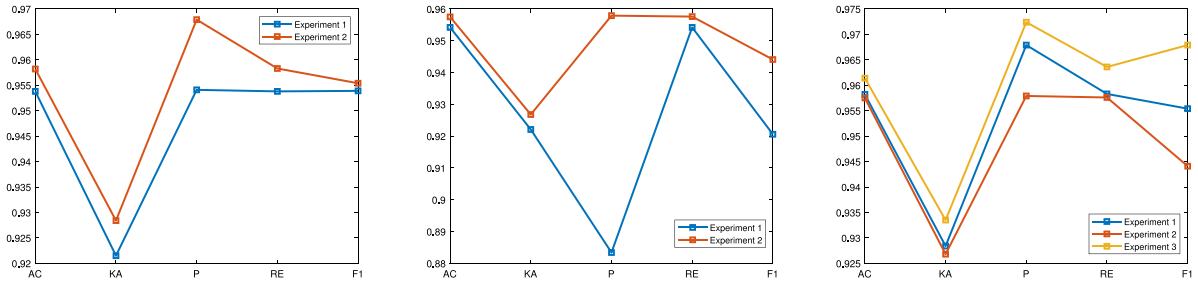
The foremost purpose of the study is to construct a seizure type classification algorithm, thereby to assist physicians in reducing their burden, and in turn to lead to reductions in misdiagnosis of epilepsy patients. To achieve an accurate seizure type classification model, based on the time domain and frequency domain, this study uses the MSBCSP algorithm to extract the features in the spatial domain. It can be seen from the experimental results that the effect of spatial features on the classification of seizure types is significantly improved, especially for CPSZ, FNSZ and GNSZ. After adding all the features of CEEMDAN, Hjorth descriptor, MI and MSBCSP, the discrimination of each type has been greatly improved compared to the previous ones. Besides, the number of fully distinguishable categories has also increased a lot.

Simultaneously, this study introduces two strategies, namely penalized strategy and weighted voting strategy. For penalized strategy, it can be found from Figs. 11(a) and 11(b) that it can effectively handle the imbalanced data. For weighted voting strategy, it can be found from Fig. 11(c) that it can improve the performance of seizure type classification. Furthermore, the comparison results also confirm that proposed penalized ensemble model has significant advantages in seizure type classification. Therefore, clinical team may be able to take some possible interventions to improve the treatment of epilepsy patients.

Conclusively, this study believes that the proposed penalized ensemble model can assist clinician in making clinical decisions and improve the therapeutic schedules.

5. Conclusion

In this study, a novel method, refers to MSBCSP and penalized ensemble model, is proposed and applied to the seizure type



(a) Comparison between experiment 1 and experiment 2

(b) Comparison between experiment 2 and experiment 4

(c) Comparison among experiment 3, experiment 4 and experiment 5

Fig. 11. Performance comparison between proposed model and single models.

classification. In the penalized ensemble model, penalized strategy is introduced to deal with imbalance data. For the purpose of improving the performance of classification, weighted voting strategy is introduced. Compared with the various models commonly used in the studies and the existing works of seizure type classification, it promotes the performance of classification significantly.

In summary, the proposed model reveals high-performance quality that has potential application value for clinical systems and thus helps to upgrade the level of treatment for patients with epilepsy. Furthermore, the proposed penalized ensemble framework can be used to handle imbalanced data in similar clinical prediction problem.

CRediT authorship contribution statement

Duanpo Wu: Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition. **Jie Li:** Conceptualization, Methodology, Investigation, Writing – original draft. **Fang Dong:** Conceptualization, Formal analysis, Writing – review & editing, Funding acquisition. **Junbiao Liu:** Methodology, Formal analysis. **Lurong Jiang:** Validation, Formal analysis. **Jiuwen Cao:** Methodology, Investigation, Funding acquisition. **Xunyi Wu:** Validation, Investigation. **Xin Zhang:** Validation, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Duanpo Wu reports financial support was provided by Zhejiang Province Natural Science Foundation. Jiuwen Cao reports financial support was provided by National Natural Science Foundation of China. Jiuwen Cao reports financial support was provided by Ministry of Science and Technology of the People's Republic of China. Duanpo Wu reports financial support was provided by Zhejiang Province Science and Technology Department. Fang Dong reports financial support was provided by Hangzhou Science and Technology Bureau. Xin Zhang reports financial support was provided by Zhejiang Province Science and Technology Department. Jiuwen Cao reports financial support was provided by Zhejiang University - Zhijiang Campus. Duanpo Wu reports financial support was provided by Zhejiang University.

Acknowledgments

This project has received funding from the National Key R & D Program, China (2021YFE0100100), Key R & D Programs in Zhejiang Province, China (2020C03038), Joint Fund of Zhejiang Provincial Natural Science Foundation of China (LBY21H090001), NSFC-Zhejiang Integration Joint Fund, China (U1909209), Hangzhou Agricultural and Social Development Scientific Research Project, China (20201203B145), the Open Research Projects of Zhejiang Lab (2021MC0AB04), Public Welfare Technology Application Research Plan Project of Zhejiang Province in China (LGF19H090020) and the open project of Zhejiang Provincial Key Laboratory of Information Processing, Communication and Networking, Zhejiang, China, China.

References

[1] P.J. Karoly, V.R. Rao, N.M. Gregg, et al., Cycles in epilepsy, Nature Rev. Neurol. 17 (2021) 267–284.
[2] Britta Wandschneider, Seok-Jun Hong, Boris C. Bernhardt, Fatemeh Fadaie, Christian Vollmar, Matthias J. Koepp, Neda Bernasconi, Andrea Bernasconi, Developmental MRI markers cosegregate juvenile patients with myoclonic epilepsy and their healthy siblings, Neurology 93 (13) (2019) e1272–e1280.
[3] Inung Wijayanto, et al., Seizure type detection in epileptic EEG signal using empirical mode decomposition and support vector machine, in: 2019 International Seminar on Intelligent Technology and Its Applications, ISITIA, 2019, pp. 314–319.
[4] Inggi Ramadhani Dwi Saputro, et al., Seizure type classification on EEG signal using support vector machine, J. Phys. Conf. Ser. 1201 (1) (2019) 012065.
[5] Guangyu Jia, Hak-Keung Lam, Kaspar Althoefer, Variable weight algorithm for convolutional neural networks and its applications to classification of seizure phases and types, Pattern Recognit. 121 (2021) 108226.
[6] Raghu Shivarudhrappa, et al., EEG based multi-class seizure type classification using convolutional neural network and transfer learning, Neural Netw. 124 (2020) 202–212.
[7] K. Abualsaud, M. Mahmuddin, M. Saleh, et al., Performance comparison of classification algorithms for EEG-based remote epileptic seizure detection in wireless sensor networks, in: 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications, AICCSA, 2014, pp. 633–639.
[8] K. Abualsaud, M. Mahmuddin, M. Saleh, et al., Ensemble classifier for epileptic seizure detection for imperfect EEG data, Sci. World J. (2015) 2015.
[9] K. Abualsaud, M. Mahmuddin, R. Hussein, et al., Performance evaluation for compression-accuracy trade-off using compressive sensing for EEG-based epileptic seizure detection in wireless tele-monitoring, in: 2013 9th International Wireless Communications and Mobile Computing Conference, IWCMC, 2013, pp. 231–236.
[10] H. Ramoser, J. Muller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, IEEE Trans. Rehabil. Eng. 8 (4) (2000) 441–446.
[11] C. Guger, H. Ramoser, G. Pfurtscheller, Real time EEG analysis with subject specific spatial patterns for a brain-computer interface, IEEE Trans. Rehabil. Eng. 8 (4) (2000) 447–456.
[12] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, K.R. Muller, Combined optimization of spatial and temporal filters for improving brain-computer interfacing, IEEE Trans. Biomed. Eng. 53 (11) (2006) 2274–2281.
[13] S. Lemm, B. Blankertz, G. Curio, K.-R. Muller, Spatio-spectral filters for improving the classification of single trial EEG, IEEE Trans. Biomed. Eng. 52 (9) (2005) 1541–1548.
[14] Muhammad Hamza Bhatti, et al., Soft computing-based EEG classification by optimal feature selection and neural networks, IEEE Trans. Ind. Inf. 15 (10) (2019) 5747–5754.
[15] Jyoti Singh Kirar, R.K. Agrawal, Optimal spatio-spectral variable size subbands filter for motor imagery brain computer interface, Procedia Comput. Sci. 84 (2016) 14–21.
[16] Yu Zhang, et al., Optimizing spatial patterns with sparse filter bands for motor-imagery based brain–computer interface, J. Neurosci. Methods 255 (2015) 85–91.

[17] Gregory D. Cascino, Joseph I. Sirven, William O. Tatum, Epilepsy, second ed., Wiley-Blackwell, America, 2021.
[18] Vibhangini S. Wasade, Marianna V. Spanaki, Understanding Epilepsy: A Study Guide for the Boards, Cambridge University Press, Cambridge, 2019.
[19] S. Motamedi-Fakhr, M. Moshrefi-Torbati, M. Hill, et al., Signal processing techniques applied to human sleep EEG signals—A review, Biomed. Signal Process. Control 10 (2014) 21–23.
[20] D. Wu, Z. Wang, H. Huang, et al., Epileptic seizure detection system based on multi-domain feature and spike feature of EEG, Int. J. Hum. Robot. 16 (04) (2019) 1950016.
[21] W. Zhou, D. Chelidze, Blind source separation based vibration mode identification, Mech. Syst. Signal Process. 21 (8) (2007) 3072–3087.
[22] A. Belouchrani, K. Abed-Meraim, J.F. Cardoso, et al., A blind source separation techinique based on second order statistics, IEEE Trans. Signal Process. 45 (2) (1997) 434–444.
[23] S.R. Liyanage, J.-X. Xu, C.T. Guan, K.K. Ang, T.H. Lee, EEG signal separation for multi-class motor imagery using common spatial patterns based on Joint Approximate Diagonalization, in: The 2010 International Joint Conference on Neural Networks, IJCNN, 2010, pp. 1–6.
[24] Ali Braytee, Wei Liu, Paul J. Kennedy, A cost-sensitive learning strategy for feature extraction from imbalanced data, in: International Conference on Neural Information Processing, 2016, pp. 78–86.
[25] Zhenxing Qin, Tao Wang, Shichao Zhang, Incorporating medical history to cost sensitive classification with lazy learning strategy, in: IEEE International Conference on Progress in Informatics and Computing, vol. 1, 2010, pp. 19–23.
[26] Fan Min, William Zhu, A competition strategy to cost-sensitive decision trees, in: Rough Sets and Knowledge Technology, 2012, pp. 359–368.
[27] W. Wu, Y. Xia, W. Jin, Predicting bus passenger flow and prioritizing influential factors using multi-source data: Scaled stacking gradient boosting decision trees, IEEE Trans. Intell. Transp. Syst. 22 (4) (2021) 2510–2523.
[28] Z. Zhang, C. Jung, GBDT-MO: Gradient-boosted decision trees for multiple outputs, IEEE Trans. Neural Netw. Learn. Syst. 32 (7) (2021) 3156–3167.
[29] Ye Yuan, Guangxu Xun, Kebin Jia, Aidong Zhang, A multi-view deep learning framework for EEG seizure detection, IEEE J. Biomed. Health Inf. 23 (1) (2018) 83–94.
[30] David Ahmedt-Aristizabal, et al., Neural memory networks for seizure type classification, in: IEEE International Conference of Engineering in Medicine and Biology Society, 2019.
[31] Subhrajit Roy, et al., Seizure type classification using EEG signals and machine learning: Setting a benchmark, in: 2020 IEEE Signal Processing in Medicine and Biology Symposium, SPMB, 2020.
[32] Umar Asif, et al., SeizureNet: Multi-spectral deep feature learning for seizure type classification, in: Machine Learning in Clinical Neuroimaging and Radiogenomics in Neuro-Oncology, 2020, pp. 77–87.
[33] Guangyu Jia, Hak-Keung Lam, Kaspar Althoefer, Variable weight algorithm for convolutional neural networks and its applications to classification of seizure phases and types, Pattern Recognit. 121 (2021) 108226.

Duanpo Wu

Jie Li

Fang Dong

Junbiao Liu

Lurong Jiang

Jiuwen Cao

Xunyi Wu

Xin Zhang