Journal of Network Intelligence ©2025 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **10**, Number **1**, February **2025**

Machine Learning-Based Early Intelligent Diagnosis of Youth Mental Health Conditions

Ya-Nan Wang*

Anhui Vocational College of Defense Technology, Lu'an 237005, P. R. China
wxwszyxywyn@163.com

Jing Yang

West Anhui University, Lu'an 237000, P. R. China
skysmemory@163.com

Jing Gao

College of Engineering
Pukyong National University, Nam-gu, Busan 61005, South Korea
eg3906@163.com

*Corresponding author: Ya-Nan Wang
Received February 1, 2024, revised June 19, 2024, accepted August 6, 2024.

Abstract. *Subjective judgement of EEG (Electroencephalogram) using a priori knowledge and related theories is currently the dominant technical tool for diagnosing mental health status in youth. However, this subjective judgement method is highly error-prone and inefficient. Advanced signal processing technology, machine learning technology, etc. provide new methods for fast and accurate intelligent mental consultation. Therefore, an early intelligent diagnosis method based on wavelet Bispectrum and CatBoost is proposed to address the problem of automatic identification of abnormal EEG information in youth. Firstly, wavelet Bispectrum estimation algorithm is proposed by combining wavelet and Bispectrum analysis methods for the characteristics of EEG signals. Wavelet Bispectrum and sparse learning are combined to achieve signal feature extraction. The sparse learning method is used to extract sparse features at each scale to obtain a more discriminative feature set. Then, a CatBoost model optimised by a genetic algorithm is proposed and used for the EEG signal classification task. Finally, the schizophrenic clinical EEG signals were tested as an example. The results show that the proposed diagnostic method avoids diagnostic bias due to subjective factors or differences in judgement criteria, with an average Sensitivity of 94.37%, an average Specificity of 88.72%, and an average Accuracy of 96.33%.*
Keywords: Mental health; Electroencephalogram; wavelet Bispectrum; sparse learning; CatBoost; genetic algorithm

---

1. **Introduction.** For the diagnosis of mental health status of youth, EEG (Electroencephalogram) is an important technical tool [1, 2]. EEG can reflect an individual's functional state of the brain, including information on cognition, emotion, and attention by recording the electrical activity of the cerebral cortex. Subjective judgement of EEG by combining a priori knowledge and relevant theories is a common diagnostic method nowadays [3, 4].

In subjective judgements of EEG, a priori knowledge may include psychological theory and clinical experience of mental health conditions in young people. For example, knowledge of the characteristics of mental health problems such as anxiety, depression, and schizophrenia in EEG signals, as well as patterns of brain region activity associated with specific mental health problems. This knowledge can help psychological counseling teachers to be more astute in identifying abnormal patterns and features for higher vocational college students and making subjective judgements [5, 6].

However, subjective judgement also has a certain degree of subjectivity and uncertainty, so there are also a number of studies devoted to the objective processing and automated analysis of EEG data. For example, an automated diagnostic system can be developed to reduce the uncertainty and error of subjective judgement by combining a large amount of EEG data and relevant mental health labels [7] through machine learning and deep learning techniques. In the future, with the development of artificial intelligence and data science, objectivised analysis of EEG will become an important technical tool for diagnosing mental health conditions in youth [8, 9].

An automated diagnostic system based on machine learning can objectively analyse EEG data, extract features related to mental health status, and correlate these features with existing mental health labels, ultimately enabling automated diagnosis of an individual's mental health status [10]. Machine learning algorithms can be used to process large amounts of EEG data. Through supervised learning, the system can learn the patterns and features in the EEG signals that are related to the mental health status, and thus be able to automatically classify and diagnose new EEG data. Meanwhile, deep learning techniques can be used to extract complex feature representations in EEG data, such as time-domain and frequency-domain features in time series [11], and correlation features between different regions of the brain. The extraction of these features is crucial for accurate diagnosis of mental health states. The aim of this work is to design an automated diagnostic technique for ECG signals with high accuracy to greatly reduce the uncertainty and error of subjective judgement, which will help to detect mental health problems in young people at an early stage and treat them in a timely manner.

1.1. **Related Work.** Automated diagnosis of mental disorders based on machine learning is a research area of great interest and many advances have been made in recent years [12, 13, 14]. Among them, EEG plays an important role in the diagnosis of psychological disorders. The research in the related field is mainly divided into two aspects: (1) feature extraction and selection (2) classification and diagnostic models.

In terms of feature extraction and selection, researchers use machine learning techniques to explore how to extract features related to mental illness from EEG data. These features can include features in the time, frequency, and spatial domains, as well as correlation features between different regions of the brain. In addition, there are studies focusing on how to select the features with the most diagnostic value to improve the accuracy and interpretability of diagnostic models. Acharya et al. [15] explored a comparative study of different feature extraction methods in EEG classification. The study used several common feature extraction methods, including time domain, frequency domain and time-frequency domain feature extraction methods, and compared the classification performance. Through the experimental results, it is found that the time-frequency domain feature extraction methods have significant advantages in classification performance and can better exploit the time-frequency characteristics of EEG signals. Soufineyestani et al. [16] study compares various EEG feature extraction methods such as power spectral density, wavelet transform, etc., and uses a feature selection algorithm to select the most relevant features. The experimental results showed that the accuracy of EEG classification can be significantly improved by choosing appropriate feature extraction and selection methods. Liu et al. [17] proposed a Bispectrum based EEG feature extraction method. The authors used Bispectrum analysis to extract phase coupling features of EEG signals as input feature vectors. The experimental results show that the Bispectrum features can better characterise the nonlinear dynamics of the EEG signals than the traditional power spectrum features, and the classification performance is significantly improved.

In terms of classification and diagnostic models, researchers use machine learning algorithms to build automated models for the diagnosis of mental illnesses, including Support Vector Machines (SVM), Random Forest, XGBoost, and so on. These models can be trained with labelled EEG data, and then classify and diagnose unknown samples, thus enabling automated diagnosis of mental disorders. Edla et al. [18] combined the time-frequency distribution characteristics of EEG signals to achieve automatic classification of schizophrenic patients by Random Forest. The results showed that the model could distinguish between normal people and schizophrenic patients, demonstrating the potential in assisting schizophrenia screening. Wang et al. [19] extracted the time-frequency features of the EEG signals by using wavelet transform, which, combined with the clinical assessment results, was used as the input feature vector. Then the XGBoost model was used to classify patients with schizophrenia. The results show that compared with other machine learning models, the classification performance of XGBoost is superior and can effectively assist the screening and diagnosis of schizophrenia.

1.2. **Motivation and contribution.** Although the traditional Bispectrum features can better characterise the nonlinear dynamics of EEG signals, they cannot well reveal the non-smoothness of EEG signals [20]. The wavelet transform can autonomously adjust the time window and frequency window, which can effectively extract the hidden transient information in the signal, and is suitable for non-smooth feature extraction of EEG signals containing noise.

In addition, the XGBoost algorithm is more favourable to deal with data with balanced category distribution. However, in the EEG classification task, the amount of data in different categories is often unbalanced, which will reduce the classification performance of the XGBoost model. In contrast, CatBoost can effectively solve the problem of category imbalance by using category smoothing technique, which improves the classification effect of samples from a few categories. Therefore, this work proposes an intelligent diagnosis method based on wavelet Bispectrum and CatBoost. The main innovations and contributions of this work include:

(1) Aiming at the limitations of the traditional Bispectrum estimation method in the analysis of EEG signals of brain diseases, a wavelet Bispectrum estimation algorithm is proposed, thus effectively exploiting the advantages of the wavelet transform in analysing non-smooth signals. The combination of wavelet Bispectrum and sparse learning is used to achieve signal feature extraction, which lays the foundation for improving the total recognition rate of subsequent classification and diagnosis.

(3) Although CatBoost can effectively solve the problem of category imbalance compared to the XGBoost algorithm, the CatBoost model has more parameters. Therefore, it is proposed to perform parameter optimisation by Genetic Algorithm (GA) to build GA-Catboost model, so as to improve the stability and generalisation performance of EEG classification and diagnosis model.

2. **EEG feature extraction based on wavelet Bispectrum and sparse learning.**

2.1. **Bispectrum Fundamentals.** A Bispectrum is a complex-valued spectrum with amplitude and phase [21, 22]. For a discrete-time deterministic signal, its Bispectrum is defined as:

$$B(\omega_1, \omega_2) = X(\omega_1)X(\omega_2)X^*(\omega_1 + \omega_2) \tag{1}$$

where $X(\omega)$ is the Fourier transform of the signal $X(t)$ at ω_1, and $X^*(\omega_1 + \omega_2)$ is the conjugate function of $X(\omega_1 + \omega_2)$.

According to the definition of Bispectrum, its main properties are as follows: 1) Bispectrum is a complex function, which can be expressed by amplitude Bispectrum and phase Bispectrum.

$$B_x(\omega_1, \omega_2) = |B_x(\omega_1, \omega_2)|e^{j\phi_B(\omega_1,\omega_2)} \tag{2}$$

2) The Bispectrum appears in the dual frequency domain as a biperiodic function with period π.

$$B_x(\omega_1, \omega_2) = B_x(\omega_1 + 2\pi, \omega_2 + 2\pi) \tag{3}$$

3) Bispectrum has symmetry.

$$\begin{aligned} B_x(\omega_1, \omega_2) &= B_x(\omega_2, \omega_1) = B_x^*(-\omega_1, -\omega_2) = B_x^*(-\omega_2, -\omega_1) \\ &= B_x(\omega_1, -\omega_1 - \omega_2) = B_x(-\omega_1 - \omega_2, \omega_2) \end{aligned} \tag{4}$$

4) If $x(t)$ is a deterministic signal, its Bispectrum can be expressed as:

$$B_x(\omega_1, \omega_2) = X(\omega_1)X(\omega_2)X^*(\omega_1 + \omega_2) \tag{5}$$

If $x(t)$ is a smooth random signal, its Bispectrum can be expressed as:

$$B_x(\omega_1, \omega_2) = E\left[X(\omega_1)X(\omega_2)X^*(\omega_1 + \omega_2)\right] \tag{6}$$

2.2. **Bispectrum estimation algorithm.** Higher-order spectral estimation is an effective method for analysing non-Gaussian signals. The Bispectrum estimation algorithm is one of the most commonly used signal analysis methods for higher-order spectral estimation [23]. In practical analysis, since the data are all of finite length, the calculation of Bispectrum cannot precisely apply the defining equation to solve the third-order cumulative quantity of the signal, and only a specific method can be used to estimate the Bispectrum for the data of finite length. Traditional Bispectrum estimation algorithms utilise FourierTransform (FT) for signal analysis.

Let $x(n)$ be a third-order real smooth random sequence whose third-order correlation function is:

$$R_x(\omega_1, \omega_2) = E\left[x(n)x(n + m_1)x(n + m_2)\right] \tag{7}$$

Then the Bispectrum expression is:

$$B_x(\omega_1, \omega_2) = \sum_{m_1}\sum_{m_2} R_x(m_1, m_2)e^{-j(\omega_1 m_1 + \omega_2 m_2)} \tag{8}$$

Firstly, the FFT of the sequence is calculated before finding its frequency domain correlation signal. Let $\{x(0), x(1), \dots, x(N-1)\}$ be the test sequence and f_s be the sampling frequency. The raw data is divided into K segments, each segment contains M samples. Repetition is allowed between segments. For each segment of data, subtract the mean value of the segment to make each segment a zero-mean sequence [24].

Then, the discrete Fourier transform (DFT) coefficients are calculated for each sequence.

$$X^{(k)}(\lambda) = \frac{1}{M}\sum_{n=0}^{M-1} x^{(k)}(n)e^{-i2\pi n\lambda/M} \tag{9}$$

Based on the DFT coefficients, the Bispectrum estimation is derived separately for each segment of data.

$$b_k(\lambda_1, \lambda_2) = \frac{1}{\Delta_0^2} \sum_{i_1=-L_1}^{L_1} \sum_{i_2=-L_1}^{L_1} X^{(k)}(\lambda_1 + i_1) X^{(k)}(\lambda_2 + i_2) X^{(k)}(-\lambda_1 - \lambda_2 - i_1 - i_2) \quad (10)$$

where λ_1 and λ_2 correspond to the points of ω_1 and ω_2 axes after DFT transformation, and L_1 denotes the number of smooth points.

Finally, the Bispectrum estimates for each segment of data are statistically averaged to obtain a Bispectrum estimate of the signal.

$$B_D(\omega_1, \omega_2) = \frac{1}{K} \sum_{k=1}^{K} \hat{b}(\omega_1, \omega_2) \quad (11)$$

2.3. **Wavelet Bispectrum estimation.** The traditional Bispectrum estimation algorithm uses FT for signal analysis. Since FT has limited ability to deal with non-smooth signals, the traditional Bispectrum estimation algorithm cannot reveal the non-smoothness of the signal well. Wavelet transform can effectively extract the hidden transient information in the signal and is suitable for non-smooth feature extraction of EEG signals containing noise.

The wavelet Bispectrum estimation algorithm is defined as shown below:

$$B^w(a_1, a_2) = \int_T W_f^*(a, \tau) W_f(a_1, \tau) W_f(a_2, \tau) d\tau \quad (12)$$

where a is the scale factor in the wavelet, $a > 0$; τ is the time shift factor, W_f denotes the wavelet transform of the function $f(t)$ (over the time interval $T : \tau_0 \leq \tau \leq \tau_1$), and W_f^* is the covariance function of W_f.

First, define the original data $f(t)$, divide it into N segments and subtract the mean of the data. Then, the continuous wavelet function, bandwidth coefficient f_b and centre frequency f_c are set. The wavelet transform is performed in the set time and frequency range to obtain the time series. In this paper, Morlet wavelet is used for wavelet transform.

$$W_f(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \psi^* \left(\frac{t - \tau}{a} \right) dt = \left\langle f(t), \psi \left(\frac{t - \tau}{a} \right) \right\rangle \quad (13)$$

$$\psi(t) = \frac{1}{\sqrt{\pi f_b}} \exp \left[i 2\pi f_c t - \frac{t^2}{f_b} \right] \quad (14)$$

Finally, the wavelet Bispectrum is calculated according to Equation (12).

The wavelet Bispectrum energy entropy feature extraction algorithm adopts the theory of "information entropy" to describe the distribution of the matrix energy obtained after the original data is analysed by the wavelet Bispectrum algorithm. Assuming that the Bispectrum of the received signal is $B(\omega_1, \omega_2)$, its energy matrix E is described as follows.

$$E_{ij} = \int_{(i-1)\Delta\omega_1}^{i\Delta\omega_1} \int_{(j-1)\Delta\omega_2}^{j\Delta\omega_2} B(\omega_1, \omega_2) d\omega_1 d\omega_2 \quad (15)$$

where E_{ij} denotes the value of point (i, j) in E.

Energy entropy is defined as follows.

$$E_n = -\sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} \log p_{ij}, \quad p_{ij} = \frac{E_{ij}}{E}, \quad E = \sum_{i=1}^{I} \sum_{j=1}^{J} E_{ij} \quad (16)$$

where p_{ij} represents the ratio of the energy of each point in the energy moment to the total energy.

The energy entropy of wavelet Bispectrum can reflect the distribution of EEG energy matrix to some extent. Under different brain functional states, EEG signals show different wavelet Bispectrum structures. When the brain is in different functional states, the wavelet Bispectrum spectral peak positions and wave amplitudes of EEG signals are significantly different.

2.4. **Sparse learning.** Sparse learning and wavelet Bispectrum estimation are both tools used for feature extraction and they can be used in combination to improve the effectiveness of feature extraction. Sparse learning is a method based on sparse representation [25], which can be used to extract sparse features from data. Through robustness and robustness, sparse learning is able to extract discriminative features from high dimensional data. This can make full use of the local and frequency features extracted by wavelet Bispectrum estimation, and then sparse learning can be used to further extract sparse features to better characterise the EEG data.

Firstly, for each EEG signal, scale decomposition is performed using Equation (12) and Equation (16) to obtain the subband signals at different scales.

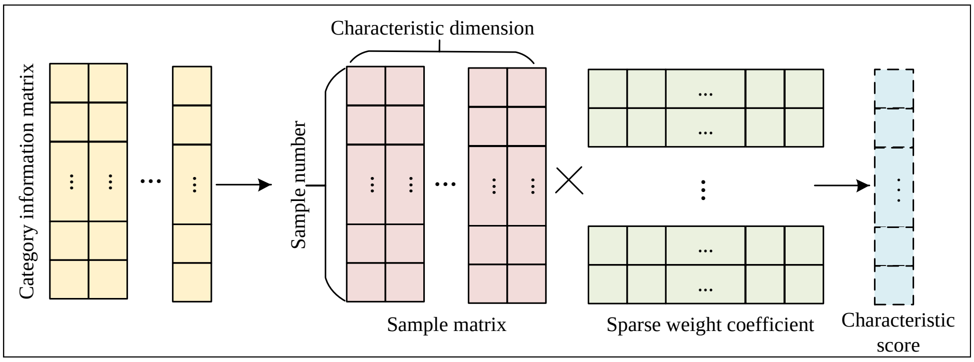


Figure 1. Feature selection based on sparse learning

Then, a sparse representation of the subband signals at each scale is performed, as shown in Figure 1. Here in this paper, Lasso regression with L1-paradigm regularisation is used. The sparse representation can be obtained by the following optimisation problem:

$$\hat{x} = \min \|y - Dx\|^2 + \lambda\|x\|_1 \tag{17}$$

where y is the subband signal, D is the dictionary, x is the sparsity coefficient, and λ is the regularisation parameter.

By solving the above optimisation problem, the sparse coefficient $\hat{x}$ of each subband signal is obtained. Based on the obtained sparse coefficients $\hat{x}$, feature selection can be performed using sparsity, and the part with larger sparse coefficients is selected as the sparse feature.

3. GA-CatBoost based EEG signal diagnostic model.

3.1. **CatBoost.** Similar to the XGBoost algorithm, the CatBoost algorithm is also obtained by improving on the GBDT [26, 27]. The algorithm chooses a base classifier with a fully symmetric tree, and this constraint on the tree structure also has some regularity effect. More importantly, it makes the model inference process faster. For the classification process of the CatBoost model, each feature split is independent and non-sequential, and multiple samples can be predicted together. The CatBoost algorithm has two important benefits, one is that it can handle category type features well without the need to convert them in advance at the data processing stage. The second is the use of a rank-boosting type approach, which reduces the prediction bias problem in the XGBoost algorithm. CatBoost is able to significantly increase the computational rate while solving the overfitting problem.

When the traditional gradient boosting algorithm analyses the categorical features, it generally uses the labeled mean value of the corresponding target variable to replace the categorical feature value in the sample set, and uses this value as the split criterion of the nodes in the decision tree. Suppose x_k^i denotes the i-th category feature value of the k-th training sample, and $\hat{x}_k^i$ denotes the value used for replacement.

$$\hat{x}_k^i = \frac{\sum_{j=1}^{n} I_{\{x_j^i = x_k^i\}} y_j}{\sum_{j=1}^{n} I_{\{x_j^i = x_k^i\}}} \tag{18}$$

where I denotes the indicator function, i.e., indicator function is 1 if the condition in the parentheses holds, 0 otherwise.

$$I_{\{x_j^i = x_k^i\}} = \begin{cases} 1, & x_j^i = x_k^i \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

If a nominal value is x_k^i, and there are few records or only one record, then converting this nominal value to a number corresponds to the label value of the record. Using such a method will often cause overfitting problems. Therefore, it is noisy for low-frequency categories, so the general choice is to add the priority term p and its corresponding weight coefficient value $\alpha(\alpha > 0)$ to smooth it out.

$$\hat{x}_k^i = \frac{\sum_{j=1}^{n} I_{\{x_j^i = x_k^i\}} \cdot y_j + \alpha \cdot p}{\sum_{j=1}^{n} I_{\{x_j^i = x_k^i\}} + \alpha} \tag{20}$$

p is usually taken as the mean of the target variable over all the data. However, such a greedy approach is also accompanied by the risk of target information leakage, where the target value y_k of x_k is used to compute $\hat{x}_k^i$, which may also result in conditional bias, i.e., differences in the data distributions of the training set and the test set.

The CatBoost algorithm, on the other hand, chooses a more efficient strategy, the sorting principle, which is also the core idea of the CatBoost algorithm [28]. Firstly, all sample data are randomly ranked. And multiple sets of random sequences are generated. Then, for each sample in any set of sequences, the mean of similar samples is estimated. The type characteristics of each sample data were processed by taking the mean and converting it to a numeric variable using the category label value that ranked before that sample, and adding the priority term p and its corresponding weight coefficient value $\alpha(\alpha > 0)$. This method can effectively reduce the noise effect of low base class features, where the priority term p is the a priori probability of the positive class samples in the EEG dataset in this paper. In short, it means that the calculation of $\hat{x}_k^i$ values relies on the sample set that has been observed so far.

Let the sample dataset be $D = \{(X_k, Y_k)\}_{k=1}^n$, and the number of features be m. Where $X_k = (x_k^1, x_k^2, \cdots, x_k^m)$, Y_i is the labelled value of the target variable. Randomly sort the samples in D and generate multiple sets of sequences, where any set of random sequences σ, $\sigma = (\sigma_1, \sigma_2, \cdots, \sigma_n)$, is replaced by $x_{\sigma_p}^i$, as follows.

$$\hat{x}^i_{\sigma_p} = \frac{\sum_{j=1}^{p-1} I_{\{x^i_{\sigma_j}=x^i_{\sigma_p}\}} y_j + \alpha \cdot p}{\sum_{j=1}^{p-1} I_{\{x^i_{\sigma_j}=x^i_{\sigma_p}\}} + \alpha} \tag{21}$$

The CatBoost algorithm combines classification features, especially high-level correlations. In EEG datasets, the number of classification features increases exponentially, making it impossible to process all combinations simultaneously. The CatBoost algorithm solves this difficulty with a greedy combination. Without performing any combinations, the CatBoost algorithm combines all the classification features (and their combinations) in the current tree with all the features in the dataset, while the combinations are dynamically transformed into alternative values.

3.2. **GA-CatBoost based EEG feature classification.** Although CatBoost can effectively solve the problem of category imbalance compared to the XGBoost algorithm, the CatBoost model has more parameters, resulting in the final classification performance of the classifier being completely determined by the parameters.

Genetic algorithms are a class of optimisation techniques with both stochastic, parallel and full-domain characteristics, whose individual adaptation (i.e., the objective function of the solution problem) automatically determines and reduces the search range and scope of the optimal parameter population. At the same time, the use of cross replication and information mutation techniques can make the network jump out of the local search and avoid falling into the local optimal state. The optimal solution can be efficiently derived by performing multiple search information points simultaneously, thus obtaining the optimal parameter set of CatBoost model and the best EEG feature classification model. The pseudo-code of GA-CatBoost classification model is shown in Algorithm 1.

Algorithm 1 GA-CatBoost classification model

Input: EEG signals X, labels y
Output: Optimised CatBoost model

```
1: Preprocess EEG signals X.
2: Extract features from preprocessed signals.
3: Initialize population P of n CatBoost models with random hyperparameters.
4: Repeat for G generations.
5: for each model M_i in P do
6:     Train M_i on training set.
7:     Evaluate fitness F_i of M_i on validation set.
8: end for
9: Select top k models based on validation fitness F_i.
10: Create offspring models via crossover and mutation.
11: Crossover: combine hyperparameters from 2 models.
12: Mutation: randomly alter hyperparameters.
13: Add offspring models to P.
14: Remove lowest scoring models from P to maintain size n.
15: Return best model catboost_model from final population P that maximises validation
    fitness.
16: Train catboost_model on full dataset using best hyperparameters.
17: Evaluate performance of catboost_model on held-out test set.
18: Return final optimised CatBoost model.
```

4. **Experimental results and analyses.**

4.1. **Experimental data.** This work validates the proposed diagnostic model using early schizophrenic EEG signals as an example. The data used in the experiment was obtained from UPenn and Mayo Clinic's Seizure Detection Challenge (data download address: `https://www.kaggle.com/c/seizure-detection/data`), and the age range of the selected young subjects was from 12 to 18 years old. The age range of the selected young subjects was 12 to 18 years. The distribution of electrode locations for this dataset using EEG signals is shown in Figure 2. Letters indicate the location of the electrodes, F for frontal, C for central, T for temporal, P for parietal, and O for occipital.

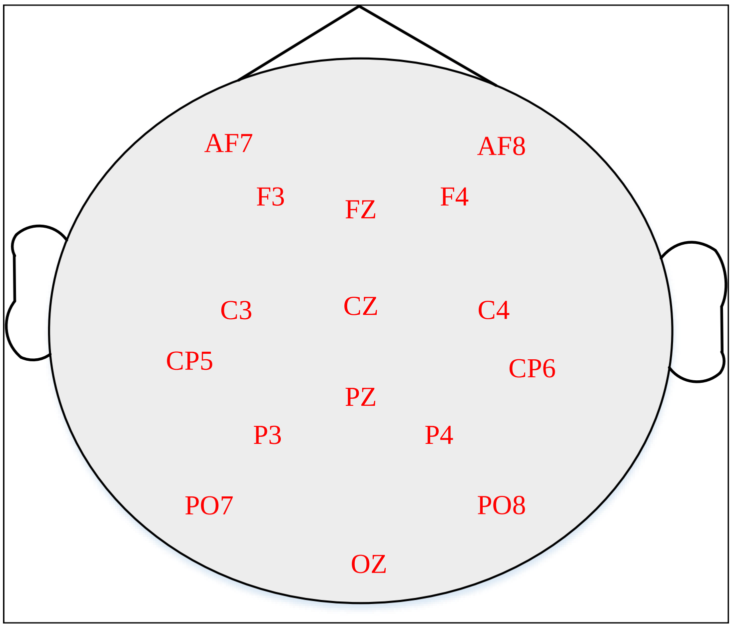
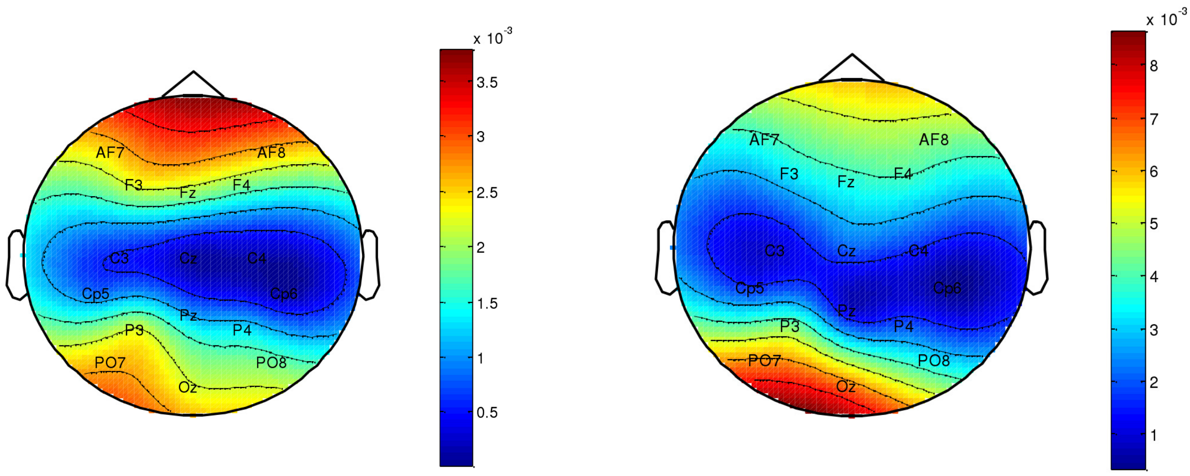


Figure 2. Schematic diagram of electrode position distribution

The data include training and test data. The training data are labelled "ES" for epileptic seizure data segments or "NES" for non-epileptic seizure data segments. A comparison of the distribution of early schizophrenic EEG signals is shown in Figure 3.



(a) "ES" (b) "NES"

Figure 3. Comparison of the distribution of EEG signals in early schizophrenia

The training data were arranged sequentially, while the test data were arranged randomly. The sampling rate of the data varies from 500Hz to 5000Hz. When constructing the classifier, "ES" and "NES" of Patient 1∼Patient 10 are used as the training data, and "tcst" of a single patient is used as the test data. The performance of the classifier was examined using the test samples. The parameters of the early schizophrenia EEG signal dataset are shown in Table 1.

Table 1. Parameters of the Early Schizophrenia EEG Signal Dataset

	ES	**NES**	**test**
Patient 1	68	102	2018
Patient 2	149	2988	3862
Patient 3	325	712	1249
Patient 4	18	188	511
Patient 5	133	2608	2954
Patient 6	223	2770	2965
Patient 7	280	3237	3569
Patient 8	178	11708	1890
Patient 9	209	893	1680
Patient 10	93	202	2319

The main parameters of the GA-CatBoost model during the experiment are shown in Table 2.

Table 2. The parameters of the GA-CatBoost model

Parameters	**Numerical value**
Iterations	420
Depth	13
Learning_rate	0.09
L2_leaf_reg	0.35
Random strength	0.45
Subsample	0.55
Colsample_bylevel	0.65
Min_child_samples	20
Loss_function	Logloss

4.2. **Assessment metrics.** Currently, there are three commonly used and important metrics in evaluating the performance of brain disease prediction methods, namely Sensitivity, Specificity and Accuracy.

(1) Sensitivity is used to identify how sensitive the algorithm is to the onset of schizophrenia.

$$S_1 = \frac{N_{PS}}{N_{TN}} \tag{22}$$

where S_1 is Sensitivity, N_{PS} is the number of times seizures were correctly forecast, and N_{TN} is the total number of schizophrenic seizures.

(2) Specificity is used to identify the degree of difference between the characteristics of the "NES" EEG signal and those of the "ES" EEG signal.

$$S_2 = \frac{N_{CS}}{N_{TS}} \tag{23}$$

where S_2 is Specificity, N_{CS} is the number of correctly identified "ES" EEG signals, and N_{TS} is the total number of "ES" EEG signals.

(3) Accuracy is used to evaluate the strength of the recognition algorithm's recognition ability.

$$S_3 = \frac{N_{PS} + N_{CS}}{N_{TN} + N_{TS}} \tag{24}$$

where S_3 is Accuracy.

4.3. **Classification Diagnosis Results.** Firstly, the feature extraction method based on wavelet Bispectrum and sparse learning is used to get the EEG feature extraction of each patient. Then, the above EEG features are inputted into GA-CatBoost model for classification judgment, and the results are shown in Table 3.

Table 3. Classification results of the proposed diagnostic model

	Sensitivity/%	**Specificity/%**	**Accuracy/per cent**
Patient 1	92.94	84.23	96.58
Patient 2	96.92	88.04	98.23
Patient 3	94.49	88.77	97.48
Patient 4	95.71	87.06	98.56
Patient 5	96.22	90.11	97.68
Patient 6	94.66	88.82	96.31
Patient 7	95.46	87.88	97.37
Patient 8	93.37	89.54	95.12
Patient 9	94.61	90.37	92.48
Patient 10	89.28	92.41	93.53
Average	94.37	88.72	96.33

It can be seen that the proposed diagnostic model shows good results on all data, with an average Sensitivity of 94.37%, an average Specificity of 88.72%, and an average Accuracy of 96.33%. In order to verify the sophistication of the proposed model, Random Forest, XGBoost, and CatBoost were compared with GA-CatBoost using the same EEG features as inputs, and the results are shown in Figure 4.

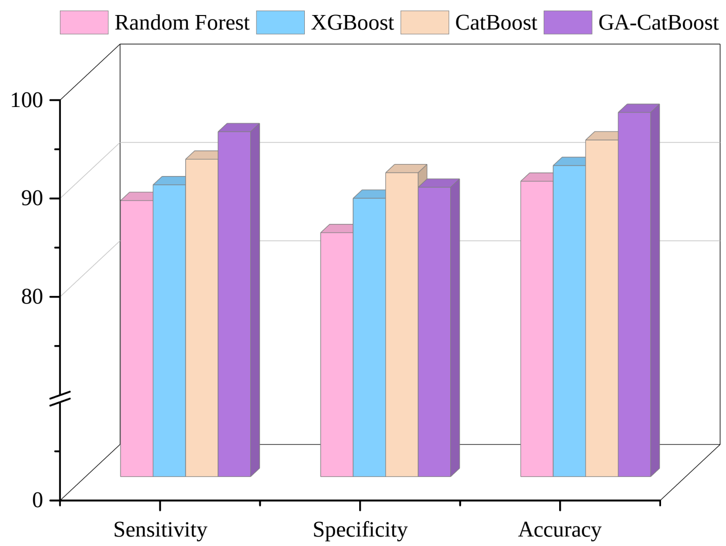


Figure 4. Comparison of different machine learning models for classification diagnosis

It can be seen that compared to Random Forest, XGBoost, and CatBoost, GA-CatBoost has a significant improvement in both Sensitivity and Accuracy, but a slight decrease in Specificity. This can be attributed to the fact that using GA to adjust the parameters of the model during the classification of the EEG feature signals makes the CatBoost model more suitable for accurately predicting the positive samples, but the prediction performance for the negative samples decreases.

5. **Conclusion.** In this work, an automatic recognition algorithm for mental abnormality is constructed with EEG signals, which can be effectively used for early schizophrenia diagnosis in young people. An EEG feature extraction method based on wavelet Bispectrum and sparse learning is proposed. The proposed GA-CatBoost model embodies a better generalisation ability and also an efficient classification ability.

The experimental results show that the average Sensitivity of the proposed diagnostic model is 94.37%, the average Specificity is 88.72%, and the average Accuracy is 96.33%. The highest recognition rate for a single patient, on the other hand, is 98.56%. Although the specificity is slightly decreased, the sensitivity and recognition rate are higher than that of Random Forest, XGBoost, and CatBoost, possessing good experimental results.

The proposed model is useful for quantitative analysis of mental health conditions and can extract the features of EEG signals under different mental health conditions with high quality, which provides a solid foundation for accurate judgment and prevention of mental diseases.

REFERENCES

[1] T. Kirschstein, and R. Köhling, “What is the source of the EEG?,” *Clinical EEG and Neuroscience*, vol. 40, no. 3, pp. 146-149, 2009.

[2] N. Kannathal, U. R. Acharya, C. M. Lim, and P. Sadasivan, “Characterization of EEG—a comparative study,” *Computer Methods and Programs in Biomedicine*, vol. 80, no. 1, pp. 17-23, 2005.

[3] R. Elul, “The genesis of the EEG,” *International Review of Neurobiology*, vol. 15, pp. 227-272, 1972.

[4] D. P. Subha, P. K. Joseph, R. Acharya U, and C. M. Lim, “EEG signal analysis: a survey,” *Journal of Medical Systems*, vol. 34, pp. 195-212, 2010.

[5] R. W. Thatcher, D. North, and C. Biver, “EEG and intelligence: relations between EEG coherence, EEG phase delay and power,” *Clinical Neurophysiology*, vol. 116, no. 9, pp. 2129-2141, 2005.

[6] I. J. Rampil, “A primer for EEG signal processing in anesthesia,” *The Journal of the American Society of Anesthesiologists*, vol. 89, no. 4, pp. 980-1002, 1998.

[7] C. M. Michel, and M. M. Murray, “Towards the utilization of EEG as a brain imaging tool,” *Neuroimage*, vol. 61, no. 2, pp. 371-385, 2012.

[8] T.-Y. Wu, L. Wang, and C.-M. Chen, ”Enhancing the Security: A Lightweight Authentication and Key Agreement Protocol for Smart Medical Services in the IoHT,” *Mathematics*, vol. 11, no. 17, 3701, 2023.

[9] T.-Y. Wu, Q. Meng, L. Yang, S. Kumari, and M. Pirouz, ”Amassing the Security: An Enhanced Authentication and Key Agreement Protocol for Remote Surgery in Healthcare Environment,” *Computer Modeling in Engineering & Sciences*, vol. 134, no. 1, pp. 317-341, 2023.

[10] T.-Y. Wu, L. Yang, Q. Meng, X. Guo, and C.-M. Chen, ”Fog-Driven Secure Authentication and Key Exchange Scheme for Wearable Health Monitoring System,” *Security and Communication Networks*, vol. 2021, pp. 1-14, 2021.

[11] C.-M. Chen, S. Liu, S. Ashraf Chaudhry, Y.-C. Chen, and M. Asghar Khan, ”A Lightweight and Robust User Authentication Protocol with User Anonymity for IoT-Based Healthcare,” *Computer Modeling in Engineering & Sciences*, vol. 131, no. 1, pp. 307-329, 2022.

[12] R. Abd Rahman, K. Omar, S. A. M. Noah, M. S. N. M. Danuri, and M. A. Al-Garadi, “Application of machine learning methods in mental health detection: a systematic review,” *IEEE Access*, vol. 8, pp. 183952-183964, 2020.

[13] H. Allende-Cid, J. Zamora, P. Alfaro-Faccio, and M. F. Alonso-Sánchez, “A machine learning approach for the automatic classification of schizophrenic discourse,” *IEEE Access*, vol. 7, pp. 45544-45553, 2019.

[14] W. Mumtaz, and A. Qayyum, “A deep learning framework for automatic diagnosis of unipolar depression,” *International Journal of Medical Informatics*, vol. 132, 103983, 2019.

[15] U. R. Acharya, Y. Hagiwara, S. N. Deshpande, S. Suren, J. E. W. Koh, S. L. Oh, N. Arunkumar, E. J. Ciaccio, and C. M. Lim, “Characterization of focal EEG signals: a review,” *Future Generation Computer Systems*, vol. 91, pp. 290-299, 2019.

[16] M. Soufineyestani, D. Dowling, and A. Khan, “Electroencephalography (EEG) technology applications and available devices,” *Applied Sciences*, vol. 10, no. 21, 7453, 2020.

[17] C. Liu, J. Jin, I. Daly, H. Sun, Y. Huang, X. Wang, and A. Cichocki, “Bispectrum-based hybrid neural network for motor imagery classification,” *Journal of Neuroscience Methods*, vol. 375, 109593, 2022.

[18] D. R. Edla, K. Mangalorekar, G. Dhavalikar, and S. Dodia, “Classification of EEG data for human mental state analysis using Random Forest Classifier,” *Procedia Computer Science*, vol. 132, pp. 1523-1532, 2018.

[19] F. Wang, Y.-C. Tian, X. Zhang, and F. Hu, “An ensemble of Xgboost models for detecting disorders of consciousness in brain injuries through EEG connectivity,” *Expert Systems with Applications*, vol. 198, 116778, 2022.

[20] W. Collis, P. White, and J. Hammond, “Higher-order spectra: the bispectrum and trispectrum,” *Mechanical Systems and Signal Processing*, vol. 12, no. 3, pp. 375-394, 1998.

[21] J. Newman, A. Pidde, and A. Stefanovska, “Defining the wavelet bispectrum,” *Applied and Computational Harmonic Analysis*, vol. 51, pp. 171-224, 2021.

[22] C. K. Kovach, H. Oya, and H. Kawasaki, “The bispectrum and its relationship to phase-amplitude coupling,” *Neuroimage*, vol. 173, pp. 518-539, 2018.

[23] H. Alquran, A. M. Alqudah, I. Abu-Qasmieh, A. Al-Badarneh, and S. Almashaqbeh, “ECG classification using higher order spectral estimation and deep learning techniques,” *Neural Network World*, vol. 29, no. 4, pp. 207-219, 2019.

[24] V. Anandan, G. R. Reddy, and P. Rao, “Spectral analysis of atmospheric radar signal using higher order spectral estimation technique,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 9, pp. 1890-1895, 2001.

[25] X. Li, Y. Wang, and R. Ruiz, “A survey on sparse learning models for feature selection,” *IEEE Transactions on Cybernetics*, vol. 52, no. 3, pp. 1642-1660, 2020.

[26] J. T. Hancock, and T. M. Khoshgoftaar, “CatBoost for big data: an interdisciplinary review,” *Journal of Big Data*, vol. 7, no. 1, pp. 1-45, 2020.

[27] G. Huang, L. Wu, X. Ma, W. Zhang, J. Fan, X. Yu, W. Zeng, and H. Zhou, “Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions,” *Journal of Hydrology*, vol. 574, pp. 1029-1041, 2019.

[28] D. Niu, L. Diao, Z. Zang, H. Che, T. Zhang, and X. Chen, “A machine-learning approach combining wavelet packet denoising with Catboost for weather forecasting,” *Atmosphere*, vol. 12, no. 12, 1618, 2021.