An Improved kNN Classifier for Epilepsy Diagnosis

ZHIPING WANG, JUNYING NA, AND BAOYOU ZHENG
College of Science, Dalian Maritime University, Dalian 116000, China
Corresponding author: Zhiping Wang (wzp@dlmu.edu.cn)

This work was supported by the Fundamental Research Funds for the Central Universities under Contract 3132019323.

ABSTRACT The electroencephalogram (EEG) signals are important for reflecting seizures and the diagnosis of epilepsy. In this paper, a weighted k-nearest neighbor classifier based on Bray Curtis distance (WBCKNN) is proposed to implement automatic detection of epilepsy. The Fourier transform can transform the time-domain characteristics of the signal into frequency domain, which can display more useful information. The WBCKNN classifier can well overcome the sensitivity of the neighborhood size k and has good robustness. Therefore, it can classify EEG signals more accurately for different situations. WBCKNN is applied on public dataset and tested by k-fold cross-validation. Experimental results show that the best accuracy of the two-classification problems and three-classification problems is 99.67% and 99%, respectively. Compared to other classifiers, the accuracy of classification is also improved. In addition, this method is superior to traditional methods in sensitivity, specificity and false alarm rate of epilepsy classification. This method can be applied to the medical market to help doctors diagnose epilepsy.

INDEX TERMS EEG, epilepsy, Bray Curtis distance, Fourier transform, k-nearest-neighbor.

I. INTRODUCTION
Epilepsy is the most common brain disease in the world. It affects about one in 200 people and may last a person’s whole life [1]. This neurological disease is caused by sudden disturbances of brain function. The frequent occurrence of the seizure causes epilepsy can directly threaten the patient’s life but the transition from the interictal stage to the next interictal is not sudden. Therefore, the diagnosis and treatment of epilepsy seizures are imperative to prevent the epilepsy or death of the patient.

EEG is the measure of the electrical activity of the brain due to the postsynaptic potentials of Cortic nerve cells representing the superposition of the potentials produced due to synchronous firing of the fibers or neurons [2]. It is an important tool for studying the temporal dynamics of the human brain’s large-scale neuronal circuits [3]. At the same time, the EEG contains a wealth of neurophysiological and pathological information, which is an effective tool for evaluating epilepsy and other brain disorders. Experienced experts can even predict upcoming seizure by examining changes in the EEG record. But the visual detection of epileptic seizures from EEG signals is error-prone and time-consuming. This method of visual examination mainly relies on the knowledge and experience of doctors, which leads too many uncertainties. Computers have long been proposed to solve this problem and thus, automated systems to recognize electroencephalograhic changes have been under study for several years [4] and the automated detection techniques based on EEG signal also have been widely used in the diagnosis of epilepsy. Because there are fewer uncertainties in automated inspection technology, automated inspection technology can get more accurate results faster than visual inspection.

In recent years, a large amount of literature describes classifiers that attempt to identify seizures through automatic monitoring. These EEG classifiers can be roughly divided into the following five categories in general: linear classifiers, neural networks, nonlinear Bayesian classifier, nearest neighbor classifiers, and combinations of classifiers [5]. For example, in [6] a new approach for feature extraction and pattern recognition of ictal EEG, based upon EMD and SVM was proposed. Petrosian et al. [7] applied recurrent neural networks combined with signal wavelet decomposition to the problem. In 2013, Labate et al. [8] implemented multiscale permutation entropy complexity measure and SVM classifier for automatic seizure detection. A method that finds the median in time series data and processing the lower and upper parts of the data separately for classification have been developed by Paulose and Bedeuzaman [9]. In 2017, a novel seizure detection algorithm using mutual information
and SVM have been researched [10]. A new algorithm was used in 2018 to classify and detect epilepsy risk levels in EEG signals with the help of expectation maximization and improved expectation maximization [11]. Alam et.al in [12] derived the IMF patterns by extracting amplitude frequency modulated oscillatory patterns and further applied the higher order statistics. In [13], Long Short-Term Memory (LSTM) networks are introduced in epileptic seizure prediction using EEG signals, expanding the use of deep learning algorithms with convolutional neural networks (CNN). Samiee et.al classified the EEG signals for epileptic seizure detection by the width of the largest local Gabor binary patterns by using the sparse rational decomposition as a feature [14]. Among the several algorithms for epilepsy classification, some method based on feature in time-domain, while others based on frequency domain. Some algorithm classification steps are complicated and have the disadvantage of insufficient diagnosis rate.

In this article, our research aims to reduce the cost of epilepsy diagnosis. On the other hand, it can reduce the workload of doctors, help doctors improve the efficiency of diagnosis, and reduce the possibility of misdiagnosis, which can enable patients to receive timely and effective treatment and reduce the burden on doctors and patients. To solve the above problems, we proposed a new epilepsy classification system, which consists of two parts: processing the EEG signal using a Fourier transform and WBCKNN classifier. Our main contributions include:

1. In order to get more efficient information from the EEG signal, the data is processed using Fourier transform. The classification situation before and after date processing was compared experimentally.
2. The WBCKNN algorithm is proposed and used in the diagnosis of epilepsy. WBCKNN method has better robustness and is not easily affected by outliers.
3. Detection of WBCKNN method by preforming two sets of experiments and compared with other algorithms. The possibility and reliability of its practical application for seizure detection are verified.

The contents of the paper are organized as follows. The EEG dataset, data processing and WBCKNN technique are given in Section 2. Simulation experiments and evaluation methods are introduced in Section 3. The experimental results, comparison with other classifiers and the advantages of WBCKNN are given in Section 4. At last, we propose our conclusion in Section 5.

II. THE FOUNDATION OF THE WBCKNN AND MATERIALS

A. DATA INTERPRETATION AND PREPROCESSING

1) DATASET

The EEG dataset used in this paper if from the work of Andrzejak et al. [15]. There are five sets(A, B, C, D and E). Each class containing 100 single-channel EEG segments. The length of the EEG signals is 23.6s and it were digitized at 173.61HZ. Each sample is a time series with 4097 numbers.

Set A and B contain segments extracted from surface EEG recordings of five health volunteers in an awake state with eyes open(A) and eyes closed(B). Set C, D and E are taken from EEG archive of pre-surgical diagnosis from five epileptics patients. Segments in Set D were recorded from within the epileptogenic zone, and those in Set C from the hippocampal formation of the opposite hemisphere of the brain. While Set C and D contained only activity measured during seizure free intervals, Sets E only contained seizure activity. The amplitude change of each category is shown in Fig. 1.

Numerous previous studies focused on A, E classification [16]–[21]. {C, D}, E classification [22]–[25]; A, D, E classification [26]–[30] and {A, B, C, D} E classification [23], [26], [31]. In this paper, we mainly consider {C, D}, E classification and A, D, E classification problems.

2) DATA PREPROCESSING

In analysis of time series or signal information, frequency domain analysis and time domain analysis are more commonly used methods. The method used in this paper is based on frequency domain analysis. By analyzing the signal from the perspective of frequency domain, it is easy to find some features that are not easy to find in the time domain. This analysis is not only simple but also superior to analysis of complex signals.

The discrete Fourier transform (DFT) is the Fourier transform that is discrete in both the time domain and the frequency domain. We used the following formula to process EEG samples

\[
X(k) = \sum_{i=1}^{N} x(i) \exp \left( -j \frac{2\pi nk}{N} \right), \quad k = 0, 1, \ldots, N-1
\]
where $X(k)$ is express as the discrete Fourier coefficient, $N$ is the length of the transformed data and $x(i)$ is a each input signal on the time domain.

Let us consider the input the EEG signal dataset that contains several time series sample as $x_i = x_1, x_2, \ldots, x_n$ where $i = 1, 2, \ldots, n$. Here ‘$n$’ indicates the total number of samples in a given dataset. Each sample $x_i$ can be represented as $x_i = \{x_{i1}, x_{i2}, \ldots, x_{in}\}$. Here ‘$N$’ indicates the total number of time series features contained in the sample. By means Eq.(1) we convert the information of each sample from the time domain $x_i$ to the frequency domain to obtain the new sample $X_i, X_i = X_1, X_2, \cdots, X_n$. And each $X_i$ also represented as $X_i = \{X_{i1}, X_{i2}, \cdots, X_{IN}\}$.

After the above processing, we transform the time domain feature to the frequency domain feature. Then we proceed with the data as follow

$$F_{ij} = \frac{|X_{ij}|^2}{N}, \quad (j = 1, 2, \cdots, N) \quad (2)$$

By means (2), we reconvert the sample characteristics from the complex number domain to the real number domain. The resulting sample are expressed as $F_i = \{F_{i1}, F_{i2}, \cdots, F_{iN}\}$.

### B. DESCRIPTION OF WBCKNN

The k-Nearest-Neighbours (kNN) is a nonparametric classification method, which is simple but effective in many cases [32]. The major drawbacks with respect to kNN are its low efficiency-being a lazy learning method, and its dependency on the selection of a “good value” for k. Frequently, to apply kNN we need to choose an appropriate value for k, and the success of classification is very much dependent on this value. In addition, Euclidean distance is commonly used in the kNN algorithm. Euclidean distance is often used to measure the similarity of two low-dimensional samples. However, the characteristic parameters used in the diagnosis of epilepsy are often more and related. Euclidean distance cannot characterize the similarity of data in high-dimensional space. Therefore, in the process of identifying and classifying epilepsy signals, this distance measurement method has certain limitations.

In order to make kNN classifier overcome the sensitivity of k and improve the sensitivity of kNN algorithm in the task of classifying epilepsy. We proposed a weight kNN classification method on Bray Curtis distance. An introduction of Bray Curtis distance is in paper in [33]. The Bray Curtis distance between data in a high-dimensional space can directly represent the similarity between data. The core of the kNN algorithm is to achieve the classification purpose by calculating the distance between the classification sample and each training sample in space. Therefore, to ensure that there is a clear difference between the distance between patients of the same type and the distance between patients of different types, the kNN algorithm can accurately classify. Euclidean distance can’t be used to find the correlation. However, the Bray Curtis distance has the characteristic of mining the correlation before the feature parameters, and it has significance to measure the similarity of data in high-dimensional space. Therefore, the Bray Curtis distance can be used as the distance measurement standard in the kNN algorithm to effectively improve the recognition accuracy in EEG signal recognition.

The WBCKNN classifier was used to classify epileptic seizures. In epilepsy classification, $n$ training samples with $m$ different classes are defined as $X = \{x_i \in R^n_{1}, \cdots, x_n\}$, each class is represented as set $C = \{C_1, C_2, \cdots, C_m\}$, and the class label of $x_i$ is $C_i \in C$. The training subset from $j$-th class is recorded as $X_j = \{x_{ij} \in R^m\}_{i=1}^N$. For a given sample $y$ to be tested its class label is as follow:

(a) Calculate the distance between $y(y = \{y_1, y_2, \cdots, y_N\})$ and the sample $x_i$ in the training set at first by using Bray Curtis distance.

$$b(y, x_i) = \frac{\sum_{h=1}^{N} |x_{ih} - y_h|}{N}, \quad 0 \leq b(y, x_i) \leq 1 \quad (3)$$

where $N$ is the number of characteristics of the sample. Obviously, $b(y, x_i) = 0$ means the samples are exactly the same, while $b(y, x_i) = 1$ is the maximum difference that can be observed between two samples.

(b) Find the k nearest neighbors of the query sample $y$ in each class according to the $b(y, x_i)$. Therefore, a total of $k \times m$ samples are selected. The $k$ samples from class $C_j$ closest to sample $y$ are represented as $X_{jNN} = \{x_{j1}^{NN} \in X_j\}_{i=1}^k$.

(c) Calculate k local mean vectors by using the k nearest neighbors in each class. Use $v_{j}^{NN} = \{v_{ij}^{NN} \in R^m\}_{i=1}^k$ to denote the k local mean vectors from class $C_j$, where $v_{j}^{NN}$ is defined as

$$v_{ij}^{NN} = \frac{1}{k} \sum_{i=1}^{k} x_{ij}^{NN}, \quad 1 \leq i \leq k \quad (4)$$

(d) Using (6) to calculate the Bray Curtis distance between $y$ and the local mean vectors of each class, and mark $k$ distances in each class as

$$B_j = \left\{b'(y, v_{ij}^{NN}), b'(y, v_{j2}^{NN}), \cdots, b'(y, v_{jN}^{NN})\right\} \quad (5)$$

where $j = 1, 2, \cdots, m$

$$b'(y, v_{ij}) = \frac{\sum_{h=1}^{N} |v_{ij}^{h} - y_h|}{\sum_{h=1}^{N} |v_{ij}^{h} + y_h|} \quad (6)$$

(e) Calculate generalized mean distances based on the $B_j$. The definition of the generalized mean is given in [34]. The $r$-th generalized mean distances is calculated by
the Bray Curtis distance from the first $r$ local vectors in each class to using (7).

$$g^j_r(y) = \left( \frac{1}{r} \sum_{i=1}^{r} \left( b^j_i(y, y^{NN}_i^j) \right)^p \right)^{\frac{1}{p}}, \quad 1 \leq r \leq k \quad (7)$$

where $p$ is a non-zero real number. When $p=1$, the generalized mean becomes the arithmetic mean. When $p \to 0$, the generalized mean becomes the geometric mean. And when $p = -1$, the generalized mean becomes the harmonic mean. The $k$ generalized mean distances in class $C_j$ are denoted as $G_j = \{g^1_j(y), g^2_j(y), \ldots, g^k_j(y)\}$.

(f) Calculate the weight distance of each class. The weights for $k$ generalized mean distance $g^j_r(y)$ from the class $C_j$ is defined as in (8)

$$w^j_i = \frac{1}{i}, \quad 1 \leq i \leq k \quad (8)$$

The weighted distance between the $k$ nearest neighbor $X_{NN}^j = \{x_{NN}^j \in X^j\}_{i=1}^k$ and $y$ in class $C_j$ is calculates as

$$D_j(y) = \left( w^1_j \cdot g^1_j(y) + w^2_j \cdot g^2_j(y) + \cdots + w^k_j \cdot g^k_j(y) \right) \quad (9)$$

(g) Finally, the weighted distance in each classes is expressed as

$$D(y) = \{D_1(y), D_2(y), \ldots, D_m(y)\} \quad (10)$$

where $m$ is the number of categories. Using (11) to determine the category of sample $y$ to be classified.

$$C = \text{arg} \min_{C_j} D(y) \quad (11)$$

The above is the specific process of the WBCKNN classifier for classification operation. Figure 2 shows the WBCKNN classifier process.

The pseudo codes of the proposed WBCKNN method are intuitively summarized in Algorithm 1.

### III. EXPERIMENTAL SETTINGS

The WBCKNN technology is implemented in Python 3.7. The main purpose of WBCKNN technology is to improve the accuracy of early diagnosis of epilepsy. It can achieve the effect of accurately classifying the state of epilepsy patients to help doctors give patients timely treatment. Therefore, WBCKNN technique are estimated in terms of accuracy (ACC), sensitivity (SEN), specificity (SPE), and false alarm rate (FAR). Among them, ACC is the main evaluation index, which is the ratio of the number of patients correctly classified as normal or abnormal to the total number of patient data evaluated by experiments. ACC can most intuitively reflect whether the WBCKNN technology achieves its purpose of accurate classification, and also serves as the main basis for our comparison with other advanced classifiers. In statistical learning theory, k-fold cross-validation, independent test and self-consistency are commonly used to test the validity of model. This paper also used k-fold cross-validation to test the performance of the WBCKNN.

In addition, we conducted comparative experiments with 6 widely used classification models. They are k-nearest neighbor (kNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), eXtreme Gradient Boosting (XG-boost) and deep neural network (DNN). These classifiers are all representative in the field of machine learning, integrated learning, and deep learning. In medical diagnosis, these kinds of classifiers are also widely used. Not only that, we also conducted a comparative test before and after data processing. The evaluation standards and experimental settings will be described in detail below.

### A. INTRODUCTION OF EVALUATION INDICATORS

The classification test performance of the WBCKNN classifier can be determined by computation of FAR, SEN and SPE along with ACC. They are defined as:

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

$$\text{SEN} = \frac{TP}{TP + FP} \quad (13)$$

$$\text{SPE} = \frac{TN}{TN + FP} \quad (14)$$

$$\text{FAR} = \frac{FP + FN}{TP + TN + FP + FN} \quad (15)$$

![FIGURE 2. WBCKNN classification process.](image-url)
Algorithm 1 WBCKNN Classifier

//WBCKNN Classifier Algorithm

Require: \( X = \{ x_i \in R^d \}_{i=1}^n \) the training EEG data with \( n \) train samples.
\( C = \{ C_1, C_2, \ldots, C_m \} \) : the class set with \( m \) class.
\( X^j = \{ x_{ij} \in R^d \}_{i=1}^n \) : the training subset from \( j \)-th class
\( y \) : the given query sample.
\( k \) : the neighborhood size
\( p \) : the size of generalized average calculation coefficient

Input: Processed data \( X \); query sample \( y \); \( k \); \( p \)

Output: Query of category \( C \) entered in sample \( y \)

Step1: Begin

Step2: Calculated distance \( b (y, x_i) \) between \( y \) and \( x_i \) using (3)

Step3: Find the \( k \) nearest neighbors of \( y \) in each class by \( b (y, x_i) \), say \( X_{NN}^{ij} = \{ x_{ij} \in X^j \}_{i=1}^k \)

Step4: Calculate \( k \) local mean vectors from every class. For \( i = 1 \) to \( k \) using (4), say \( V_{NN}^{ij} = \{ y_{ij} \in R^d \}_{i=1}^k \)

Step5: For \( i = 1 \) to \( k \) calculate the distance \( B_j = \{ b' (y, y_{NN}^{ij}), b' (y, y_{NN}^{i2}), \ldots, b' (y, y_{NN}^{im}) \} \) using (6)

Step6: Calculate \( k \) generalized mean distances \( G_j = \{ g_1(y), g_2(y), \ldots, g_k(y) \} \) based on the \( B_j \)

Step7: Calculate the weighted distance of each class \( D(y) = \{ D_1(y), D_2(y), \ldots, D_m(y) \} \), using (10)

Step8: Get the category \( C \) of \( C = \arg \min_{C_j} D(y) \)

Step9: Return \( C \)

Step10: End

Here TP and TN represent the total number of correctly detected true positive events and true negative events respectively. The FP and FN represent the total number of erroneously positive events and erroneously negative events.

B. K-FOLD CROSS-VALIDATION

K-fold cross validation is a method to ensure that each subsample is trained and tested, reducing the generalization error. This method divides the data into \( k \) groups. Each time one of the \( k \) subsets is used as the test set and the other \( k - 1 \) subsets are put together to form a training set. The final classification results is the average of \( k \)-fold cross-validation. In each set of experiments, we performed \( k = 2, 5, 10 \) fold cross-validation.

C. EXPERIMENTAL EXPLANATION

In this paper, we conducted two sets of experiments. They are defined as follows:

Experiment 1: Classify patients with epilepsy.

The purpose of this experiment is to classify patients with seizures from those who have not. It can help to detect when it’s happening and notify doctors to rescue the patient, which may provide an alert as soon as possible.

Experiment 2: Classify normal, seizure free intervals and ictal.

There are differences in EEG signals between healthy people and patients with seizures, and classification can help to diagnose whether the people have the disease. At the same time, the EEG signals of clinically healthy people are similar to those without seizure intervals. This experiment can also help to distinguish these two groups of people, thereby improving the accuracy of the diagnosis.

The experimental process is shown in Fig. 3.

IV. RESULTS AND DISCUSSION

In this section, the experimental results obtained by the WBCKNN classifier will be introduced. At the same time, we will also analyze the characteristics and advantages of WBCKNN system. By analyzing the WBCKNN system, it provides a theoretical basis for the effectiveness of WBCKNN.

A. RESULTS OF THE EXPERIMENT

The results of the first experiment are shown in Table 1, Table 2 and Table 3. Table 1 shows result of \{C, D\} E classification after 2-fold cross validation. Table 2 shows the result of \{C, D\} E classification after 5-fold cross validation. Table 3 shows the result of \{C, D\} E classification after 10-fold cross validation.

The experimental results of the three classification A, D and E are given in Table 4, 5 and 6. Table 4, 5 and 6 are the result of 2, 5 and 10 fold cross validation respectively.

B. ANALYSIS OF EXPERIMENTAL RESULTS AND ANALYSIS OF WBCKNN

The proposed EEG signal classification system has three characteristics: (a) the relationship between samples is mea-
TABLE 1. Classification results of 2-fold cross-validation of \{C, D\}, E.

| Cross-validation | Classifier | Data unprocessed | The data is processed using FT |
|------------------|------------|------------------|------------------------------|
|                  |            | ACC   | SEN   | SPE   | FAR   | ACC   | SEN   | SPE   | FAR   |
| 2-fold           | kNN        | 73.033%| 55.64%| 11.282%| 26.967%| 90.990%| 87%   | 75    | 9.015%|
|                  | SVM        | 78.333%| 67.548%| 35.096%| 21.667%| 98.000%| 97.587%| 96.153%| 2.000%|
|                  | DT         | 82.667%| 80.456%| 73.878%| 17.333%| 90.999%| 88.598%| 81.884%| 9.000%|
|                  | RF         | 93.667%| 83.993%| 91.239%| 6.333% | 97.333%| 96.939%| 97.227%| 2.667%|
|                  | DNN        | 77.667%| 66.667%| 33.37% | 22.333%| 97.667%| 96.963%| 94.887%| 2.333%|
|                  | XG-BOOST   | 88.999%| 84.286%| 70.101%| 11.001%| 98%   | 97.486%| 95.974%| 2.000%|
|                  | WBCKNN(k=20, p=2) | 95.667%| 87.825%| 84.949%| 4.333% | 99.667%| 99.743%| 100% | 0.333%|

TABLE 2. Classification results of 5-fold cross-validation of \{C, D\}, E.

| Cross-validation | Classifier | Data unprocessed | The data is processed using FT |
|------------------|------------|------------------|------------------------------|
|                  |            | ACC   | SEN   | SPE   | FAR   | ACC   | SEN   | SPE   | FAR   |
| 5-fold           | kNN        | 71.726%| 58.333%| 16.667%| 28.274%| 92.191%| 88.943%| 78.49%| 7.809%|
|                  | SVM        | 82.333%| 73.448%| 46.696%| 17.667%| 98.333%| 97.675%| 95.394%| 1.667%|
|                  | DT         | 82.667%| 81.432%| 67.780%| 17.333%| 95.667%| 95.356%| 94.212%| 4.333%|
|                  | RF         | 93.667%| 84.545%| 91.284%| 6.333% | 97.333%| 96.08% | 97.054%| 2.667%|
|                  | DNN        | 77.667%| 83.334%| 66.688%| 22.333%| 99.333%| 99%   | 98%   | 0.667%|
|                  | XG-BOOST   | 88.999%| 90.641%| 82.842%| 11.001%| 96.333%| 95.611%| 93.77%| 0.667%|
|                  | WBCKNN(k=20, p=2) | 97.666%| 95.221%| 95.221%| 2.334% | 99.667%| 99.744%| 100% | 0.333%|

TABLE 3. Classification results of 10-fold cross-validation of \{C, D\}, E.

| Cross-validation | Classifier | Data unprocessed | The data is processed using FT |
|------------------|------------|------------------|------------------------------|
|                  |            | ACC   | SEN   | SPE   | FAR   | ACC   | SEN   | SPE   | FAR   |
| 10-fold          | kNN        | 73.078%| 60.187%| 20.374%| 26.922%| 92.666%| 89.609%| 79.733%| 7.334%|
|                  | SVM        | 82.667%| 74.192%| 48.384%| 17.333%| 98.333%| 97.849%| 95.69%| 1.667%|
|                  | DT         | 85% | 80.901%| 72.082%| 15.000%| 93.667%| 92.409%| 89.143%| 6.333%|
|                  | RF         | 95.667%| 90.397%| 94.421%| 4.333% | 97.333%| 96.691%| 97.047%| 2.667%|
|                  | DNN        | 87%   | 81.041%| 62.087%| 13.000%| 99.333%| 99.231%| 98.462%| 0.667%|
|                  | XG-BOOST   | 93% | 90.404%| 81.974%| 7.000% | 97.333%| 97.58% | 97.024%| 2.667%|
|                  | WBCKNN(k=20, p=2) | 97% | 95.105%| 95.105%| 3.000% | 99.667%| 99.528%| 100% | 0.333%|

sured by Bray Curtis distance. (b) introducing a generalized mean. (c) the distance weighting mechanism is introduced.

Before classification, we performed a Fourier transform on the data. This allow us to get more useful information from EEG signals. In Table 1-6, the classification results of each classifier after data processing are better than the corresponding results obtained without data processing. Experimental results also prove this advantage.

In the WBCKNN classifier, we used the Bray Curtis distance to measure the similarity between the two samples. Replacing the original Euclidean distance in kNN classifier with Bray Curtis distance can better reflect the relationship between samples. This method of calculating the correlation lays a good foundation for the following classification. In addition, we also used the local mean vector. The local mean vectors is also an effective method to overcome the influence of outliers on the classification results and overcome the sensitivity of k.

The proposed method used generalized mean distance. Given a set of k positive real number \{z_1, z_2, \cdots, z_k\},
the generalized mean is defined as

\[ G_P(z_1, z_2, \cdots, z_k) = \left( \frac{1}{k} \sum_{i=1}^{k} z_i^p \right)^{\frac{1}{p}} \]  

where \( p \) is a non-zero real number. It’s worth noting that the generalized mean is affected by the value \( p \).

With the decrease of \( p \), the generalized mean values are greatly affected by the numbers with smaller values. As \( p \) increases, the generalized mean will be more influenced by the processed different according to the importance of samples in kNN classifier. Therefore, generalized mean values are introduced into epilepsy classification can strengthen the classification contribution from the neighbors with small Bray Curtis distance. This also has a good effect on overcoming the negative effects of outliers in the \( k \) domain.

Finally, we assign weights to each generalized mean distance in such a way. In this way, those samples closer to the test sample contribute more to the classification. Therefore, the proposed classification method can not only overcome the sensitivity of \( k \) but also have better robustness in the classification.

The following conclusions:

1) Compared with other six traditional methods, the proposed WBCKNN has significant improvements in accuracy, sensitivity and specificity.
2) The classification specificity of WBCKNN technology has reached 100%, and the accuracy of classification has also reached 99.67%.
Based on the results of Experiment 1, we can consider that WBCKNN has a better performance in classifying the activities measured in the seizure phase. Therefore, WBCKNN technology can help doctors to classify patients with epilepsy. According to the results shown in Table 4, 5 and 6, we can also obtain similar conclusion as in Experiment 1:

1) The classification effect of WBCKNN method is better than other traditional methods.
2) Classification accuracy reached 99% in 10-fold cross validation. Sensitivity and specificity have also reached a relatively good level.

The above experimental results also fully show that compared with the traditional kNN classifier, the proposed method has greatly improved the classification performance and has better robustness.

We also compared our results with previous the results reported by earlier methods. Table 7 gives the classification accuracy of our method and previous methods. As we can see from these results, for the \{C, D\}, E classification task our method using 5-fold cross-validation obtains the highest classification accuracy, 99.67%, reported so far. For A, D, E classification tasks, we used the 10-fold cross-validation method to obtain the highest classification accuracy, reaching 98.72%. This accuracy is the same as that proposed by Hassan.

**TABLE 7. Algorithmic performance comparison of the proposed method with various state-of-the-art ones.**

| Authors        | Classification | Accuracy |
|----------------|----------------|----------|
| Hassan[26]     | A-D-E          | 99%      |
| Orhan[27]      | A-D-E          | 96.67%   |
| Tzallas[28]    | A-D-E          | 99.28%   |
| Wen[29]        | A-D-E          | 97.3%    |
| Nihal[30]      | A-D-E          | 96.79%   |
| This Method    | A-D-E          | 99%      |
| Sharma[22]     | {CD}-E         | 98.67%   |
| Lu[23]         | {CD}-E         | 98.2%    |
| Pachori[24]    | {CD}-E         | 97.75%   |
| Joshi[25]      | {CD}-E         | 95.33%   |
| This method    | {CD}-E         | **99.67%** |

**V. CONCLUSION**

Detecting epilepsy is a difficult task. Effective analysis of seizure activity is critical to help to patients choose the right treatment. This article presents a new epilepsy classification system. The system consists of two parts, one is to process the data through Fourier transform, and the other is to use WBCKNN classifier for decision. Using this hybrid system we interpret EEG signals and make decisions. Through experimental verification, the accuracy rates of the two types of problems and the three types of problems are 99.67% and 99%, respectively. Compared with the existing classification methods of seizure and non-seizure EEG signals, this method has better classification accuracy. Therefore, we consider it as a diagnostic decision support mechanism for epilepsy treatment.

**ACKNOWLEDGMENT**

The findings achieved herein are solely the responsibility of the authors.

**REFERENCES**

[1] J. A. Ure and M. Perassolo, “Update on the pathophysiology of the epilepsies,” J. Neurolog. Sci., vol. 177, no. 1, pp. 1–17, Aug. 2000.
[2] M. N. Tibdewal and S. A. Tale, “Multichannel detection of epilepsy using SVM classifier on EEG signal,” in Proc. Int. Conf. Comput. Commun. Control Autom. (ICCC/REA), Pune, India, Aug. 2016, pp. 1–6.
[3] C. M. Michel, M. M. Murray, G. Lantz, S. Gonzalez, L. Spinelli, and R. G. de Peralta, “EEG source imaging,” Clin. Neurophysiol., vol. 115, no. 10, pp. 2195–2222, Oct. 2004.
[4] K. Polat and S. Güneş, “Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform,” Appl. Math. Comput., vol. 187, no. 2, pp. 1017–1026, Apr. 2007.
[5] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, “A review of classification algorithms for EEG-based brain-computer interfaces,” J. Neural Eng., vol. 4, no. 2, pp. R1–R13, Jun. 2007.
[6] S. Li, W. Zhou, Q. Yuan, S. Geng, and D. Cai, “Feature extraction and recognition of ictal EEG using EMD and SVM,” Comput. Biol. Med., vol. 43, no. 7, pp. 807–816, Aug. 2013.
[7] A. Petrosian, D. Prokhorov, R. Homan, R. Dasheiff, and D. Wunsch, “Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG,” Neurocomputing, vol. 30, nos. 1–4, pp. 201–218, Jan. 2000.
[8] D. Labate, I. Pulamara, N. Mannmone, G. Morabito, F. L. Foresta, and F. C. Morabito, “SVM classification of epileptic EEG recordings through multiscale permutation entropy,” in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Dallas, TX, USA, Aug. 2013, pp. 1–5.
[9] A. Paulose and M. Bedeeuzzaman, “Seizure detection using median based feature,” in Proc. 1st Int. Conf. Comput. Syst. Commun. (ICCSSC), Trivandrum, India, Dec. 2014, pp. 328–332.
[10] Z. Chen, L. Huang, Y. Shen, J. Wang, R. Zhao, and J. Dai, “A new algorithm for classification of ictal and pre-ictal epilepsy EC0G using MI and SVM,” in Proc. Int. Conf. Signals Syst. (ICSigSyS), May 2017, pp. 211–216.
[11] S. K. Prabhakar and H. Rajaguru, “Comparison of fuzzy output optimization with expectation maximization algorithm and its modification for epilepsy classification,” in Proc. Int. Conf. Cognition Recognit., vol. 14, Sep. 2017, pp. 263–272.
[12] S. M. S. Alam and M. I. H. Bhuivyan, “Detection of seizure and epilepsy using higher order statistics in the EMD domain,” IEEE J. Biomed. Health Informat., vol. 17, no. 2, pp. 312–318, Mar. 2013.
[13] K. M. Tsouris, V. C. Pezoulas, M. Zervakis, S. Koniotiotis, D. D. Koutsouris, and D. I. Fotiadis, “A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals,” Comput. Biol. Med., vol. 99, pp. 24–37, Aug. 2018.
[14] K. Samiee, P. Kovacs, and M. Gabbouj, “Epileptic seizure detection in long-term EEG records using sparse rational decomposition and local Gabor binary patterns feature extraction,” Knowled.-Based Syst., vol. 118, pp. 228–240, Feb. 2017.
[15] R. G. Andrejak, K. Lehnhertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, “Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 64, no. 6, Nov. 2001, Art. no. 061907.
[16] V. P. Nigam and D. Graupe, “A neural-network-based detection of epilepsy,” Neurolog. Res., vol. 26, no. 1, pp. 55–60, Jan. 2004.
[17] V. Srinivasan, C. Eswaran, and A. N. Sriraam, “Artificial neural network based epileptic detection using time-domain and frequency-domain features,” J. Med. Syst., vol. 29, no. 6, pp. 647–660, Dec. 2005.
[18] N. Kannathal, M. L. Choo, U. R. Acharya, and P. K. Sadasivan, “Entropies for detection of epilepsy in EEG,” Comput. Methods Programs Biomed., vol. 80, no. 3, pp. 187–194, Dec. 2005.

[19] S. Altunay, Z. Telatar, and O. Erogul, “Epileptic EEG detection using the linear prediction error energy,” Expert Syst. Appl., vol. 37, no. 8, pp. 5661–5665, Aug. 2010.

[20] L. Guo, D. Rivero, and A. Pazos, “Epileptic seizure detection using multivariate transform based approximate entropy and artificial neural networks,” J. Neurosci. Methods, vol. 193, no. 1, pp. 156–163, Oct. 2010.

[21] A. Subasi and M. Ismail Gursoy, “EEG signal classification using PCA, ICA, LDA and support vector machines,” Expert Syst. Appl., vol. 37, no. 12, pp. 8659–8666, Dec. 2010.

[22] R. Sharma and R. B. Pachori, “Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions,” Expert Syst. Appl., vol. 42, no. 3, pp. 1106–1117, Feb. 2015.

[23] Y. Lu, Y. Ma, C. Chen, and Y. Wang, “Classification of single-channel EEG signals for epileptic seizures detection based on hybrid features,” Technol. Health Care, vol. 26, pp. 337–346, May 2018.

[24] A. R. Hassan and A. Subasi, “Automatic identification of epileptic seizures from EEG signals using linear programming boosting,” Comput. Methods Programs Biomed., vol. 136, pp. 65–67, Nov. 2016.

[25] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, “Automatic seizure detection based on time-frequency analysis and artificial neural networks,” Comput. Intell. Neurosci., vol. 2007, pp. 1–13, Jan. 2007.

[26] T. Wen and Z. Zhang, “Effective and extensible feature extraction method using genetic algorithm-based frequency-domain feature search for epileptic EEG multiclassification,” Medicine, vol. 96, no. 19, May 2017, Art. no. e6879.

[33] ZHIPING WANG received the B.E. degree in science from Hubei Normal University, China, and the M.E. degree in maritime transportation management and the Ph.D. degree in marine engineering from Dalian Maritime University, Dalian, China, in 1998 and 2003, respectively. From August 2006 to February 2007, he was with the Department of Mathematics and Statistics, University of Alberta, as a Visiting Scholar. He is currently a Professor and a Ph.D. Supervisor with Dalian Maritime University. His current research interests include network topology and reliability and super network theory.

[34] JUNYING NA received the B.E. degree in science from the University of Science and Technology Liaoning, China, in 2019. She is currently pursuing the M.E. degree in applied mathematics with Dalian Maritime University, China. Her current research interests include machine learning and biomathematics.

[35] BAOYOU ZHENG received the B.E. degree in science from Heilongjiang Bayi Agricultural University, China, in 2018. He is currently pursuing the M.E. degree in applied mathematics with Dalian Maritime University, China. His current research interests include deep learning and biomathematics.

* * *