Research Article

EEG Eye State Identification Using Incremental Attribute Learning with Time-Series Classification

Ting Wang,[1,2] Sheng-Uei Guan,[2] Ka Lok Man,[2] and T. O. Ting[3]

[1] Department of Computer Science, University of Liverpool, Liverpool L69 3BX, UK
[2] Department of Computer Science and Software Engineering, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China
[3] Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China

Correspondence should be addressed to Ting Wang; tingwang1981@163.com

Eye state identification is a kind of common time-series classification problem which is also a hot spot in recent research. Electroencephalography (EEG) is widely used in eye state classification to detect human's cognition state. Previous research has validated the feasibility of machine learning and statistical approaches for EEG eye state classification. This paper aims to propose a novel approach for EEG eye state identification using incremental attribute learning (IAL) based on neural networks. IAL is a novel machine learning strategy which gradually imports and trains features one by one. Previous studies have verified that such an approach is applicable for solving a number of pattern recognition problems. However, in these previous works, little research on IAL focused on its application to time-series problems. Therefore, it is still unknown whether IAL can be employed to cope with time-series problems like EEG eye state classification. Experimental results in this study demonstrates that, with proper feature extraction and feature ordering, IAL can not only efficiently cope with time-series classification problems, but also exhibit better classification performance in terms of classification error rates in comparison with conventional and some other approaches.

1. Introduction

Nowadays, electroencephalography (EEG) eye state classification is a very hot research area. Many studies about EEG signals have been implemented. The findings from these studies are important and useful for human cognitive state classification, which are not only crucial to medical care but also significant to some daily life tasks. For example, EEG eye state classification has been successfully applied in the areas of infant sleep-waking state identification [1], driving drowsiness detection [2], epileptic seizure detection [3], classification of bipolar mood disorder (BMD) and attention deficit hyperactivity disorder (ADHD) patients [4], stress features identification [5], human eye blinking detection [6], and so on. These occurrences indicate the importance of research on EEG eye state signal analysis.

In usual case, the data describing EEG eye state belong to continuous type of time-series data. There are a number of machine learning and statistical approaches which can be employed to solve the classification problems with this time-series data. Moreover, previous research has validated that EEG eye state signals can be successfully analysed by some machine learning or statistical approaches.

In this paper, incremental attribute learning (IAL), a novel machine learning approach, is proposed to solve the EEG eye state classification problem. IAL is a "divide-and-conquer" machine learning strategy, which can be implemented by almost all machine learning algorithms, such as neural networks (NNs), particle swarm optimization (PSO), and genetic algorithms (GAs). In IAL, features are gradually imported into the system to predict the class labelling one after the other. Because of this unique process, IAL is able to effectively reduce the interference among features. As a result of that, this approach can not only reduce the noise from data but also enhance the classification accuracy and obtain better results than some conventional approaches [7–9].

The rest of the paper is organized as follows. In Section 2, a brief introduction to EEG eye state classification is given.

Section 3 reviews IAL and presents the preprocessing works for the proposed EEG eye state classification approach. The proposed approach to solve time-series EEG eye state classification problem is presented in Section 4. Section 5 reviews the experimental benchmark problem and compares the results with those derived by some other approaches. Lastly, conclusions are drawn in Section 6.

2. EEG Eye State Classification

EEG signals of eye state monitoring are usually time-series data. Therefore, in order to classify different eye states, time-series pattern recognition approaches should be employed for EEG eye state classification. In previous studies, a number of different time-series approaches have been applied in EEG eye state identification. For example, Fukuda et al. employed a log-linearized Gaussian mixture neural network for EEG eye state classification [10]. Yeo et al. successfully used support vector machines (SVMs) to detect drowsiness during car driving by eye blink [2]. Furthermore, a hybrid system based on decision tree classifier and fast Fourier transform was applied to the detection of epileptic seizure by Polat and Güneş [3]. Sulaiman et al. also employed K-nearest neighbor (K-NN) for stress features identification [5]. In addition, Hall et al. built a 117-second EEG eye state corpus and employed 42 different machine learning and statistical approaches based on Weka [11] to predict eye states. They found that KStar is the best approach among these different methods [12]. Their eye state corpus is now a benchmark problem saved by Machine Learning Repository, University of California, Irvine (UCI) [13]. All these works showed that machine learning and statistical methods are feasible in solving time-series classification for EEG eye state identification.

3. IAL and Feature Ordering

3.1. Incremental Attribute Learning. IAL is a novel "divide-and-conquer" machine learning strategy, which gradually imports features into predictive systems according to some sequential orderings. IAL is a kind of supervised machine learning. Such an approach is designed to avoid dimensional disasters of high dimensional problem. Moreover, it is also able to cope with high dimensional problems where almost all of the features are significant and cannot be discarded by dimensional reduction approaches like feature selection. Moreover, because features are separately imported into the system, they are also computed in isolation. Such a process can effectively reduce the interference among different features. Previous research has validated that IAL can not only cope with problems with large feature dimension space [14] but also reduce the interference during the process and exhibit better performance in the final pattern recognition results [7, 15].

So far, IAL has been widely employed for pattern recognition based on a number of different predictive machine learning algorithms. In previous studies, IAL has been shown as an applicable approach in solving machine learning problems like classification using genetic algorithm (GA) [16, 17],
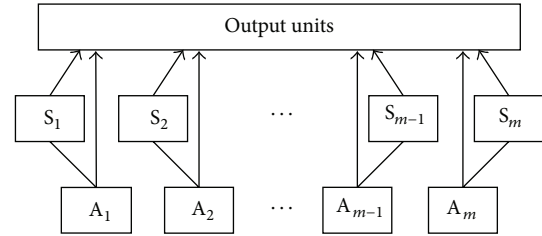

FIGURE 1: This is the basic neural network structure of ITID, where S_m stands for mth subnetworks and A_m refers to the mth attribute, which is also called features in pattern recognition and machine learning.

neural network (NN) [8, 18], support vector machine (SVM) [19], particle swarm optimization (PSO) [20], decision tree [21], and so on. These previous studies also showed that IAL can exhibit better performance than conventional methods that train all pattern features in one batch.

In this study, incremental neural network training with an increasing input dimension (ITID) [8] is employed for EEG eye state classification. ITID is one of IAL neural network approaches with ordered feature training. It was developed based on incremental learning in terms of input attributes (ILIA) [18]. However, different from ILIA which often trains features with the original feature orderings in problem dataset, ITID prefers to adjust feature orderings according to some criteria and trains features according to the adjusted sequential feature orderings [22].

Previous research has shown that ITID is applicable for classification. It divides the whole input space into several subdimensions, each of which corresponds to an input feature. Instead of learning input features altogether as an input vector in a training instance, ITID learns input features one after another through their corresponding subnetworks, and the structure of NN gradually grows with an increasing input dimension. During training, information obtained by a new subnetwork is merged together with the information obtained by the old network. With less internal interference among input features, ITID achieves higher generalization accuracy than conventional methods [8]. Figure 1 shows the basic neural network structure of ITID.

3.2. Feature Ordering. IAL gradually imports features for pattern recognition according to some orderings; thus in IAL classification, it is necessary to decide which feature should be imported into the predictive system in an earlier phase and which one can be computed later. Therefore, feature ordering is a compulsory step in IAL preprocessing. In this unique step, features with greater discrimination ability are arranged in the earlier place and those with weak discrimination ability are imported in later steps. This is similar to feature selection, a well-known dimensional reduction preprocessing, where features with greater discrimination ability are selected into a subset for further computing, while those with weak discrimination ability are discarded. However, IAL has a continuously growing feature space; previous research has

validated that a proper feature ordering is a key to lower classification error rates based on IAL [22, 23].

In previous studies, a number of feature ordering estimation methods have been developed for IAL. These methods can be divided into two types. One is based on each feature's single classification error rates, such as contribution-based methods [8, 24], while the other ranks features according to some metrics on feature discrimination ability, like mRMR [25, 26], entropy [27], Fisher's linear discriminant (FLD) [28], single discriminability (SD) [23], and accumulative discriminability (AD) [9]. Previous experimental results showed that feature orderings derived by AD often outperform those derived by other approaches, because AD is a global metric which aims to ensure that the whole growing feature space always has the largest discrimination ability during the IAL feature importing process, while others are local metrics, which only concentrate on finding the feature with the largest discrimination ability in each single step.

3.3. AD and Maximum Mean Discriminative Criterion. In IAL, it is necessary to ensure that datasets always have the greatest discrimination ability relative to each feature importing step. Namely, comparing among all the different feature orderings, the optimum feature ordering should have the largest value of feature discrimination ability in average. When a new feature is imported into the predictive system accordingly, the feature dimension is increased from d to $d + 1$. Therefore, the metric of feature discrimination ability should be the largest all the time, as only in this way it can guarantee that different classes can be separated in the easiest way. Therefore, with the aim of optimal classification results, each intermediate step will identify an optimal feature with the greatest discrimination ability for each round of feature importing. Obviously, after all features are imported, the resulting feature ordering will have the largest sum or mean of accumulative feature discrimination ability calculated in each step of the process. Here, as an efficient metric of feature discrimination ability, AD is employed for the criterion to obtain the optimum feature ordering. The criterion can be given with maximum discrimination ability mean by

$$\max \frac{1}{d} \sum_{d=1}^{d} AD (f_{1:d}), \quad (1 \le d \le m), \tag{1}$$

where $f_{1:d}$ is the feature subset of $\{f_1, f_2, \ldots, f_m\}$ during the feature importing process. The mean with a larger value indicates that the corresponding feature ordering has greater discrimination ability than the others. This criterion is called the maximum mean discriminative criterion (MMDC), which has the capacity to select the optimum feature ordering for IAL.

In (1), AD refers to the accumulative discrimination ability of the dimensional feature space with all imported features, which is the ratio in d-feature space between the multidimensional standard deviation of all class centers and the sum of all multidimensional standard deviations of all patterns in each class.

If $\{f_1, f_2, \ldots, f_m\}$ is the pool of input features, $f = \{f_{k,d}\}_{k=1,d=1}^{k=r,d=m} \in R_{\text{feature}}^{r \times m}$, when the dth ($1 \le d \le m$) feature is imported, AD is

$$AD (f_1, f_2, \ldots, f_d) = \frac{\text{std} \left[\left(\widetilde{\boldsymbol{\mu}}_j \right)_{j=1}^{j=n} \right]}{\sum_{j=1}^{j=n} \text{std} \left[\left(f_i \right)_{i=1}^{i=d} \right]_j}, \tag{2}$$

$$(1 \le d \le m),$$

where $\widetilde{\mu}_j$ is the centroid of vector $(f_1, f_2, \ldots, f_d)$ with patterns belonging to j.

Therefore, the results of (2) are calculated on the run when new features are gradually imported into training. To obtain better classification results, it is necessary to ensure that the result of (2) is the maximum in every step of feature importing. Here, std denotes the standard deviation in multidimensional space, which is derived by the standard deviation and Euclidean norm.

Let $\mathbf{x}$ be the vector for standard deviation calculation; the standard deviation of $\mathbf{x}$ is

$$\text{std} (\mathbf{x}) = \sqrt{\frac{\sum_{k=1}^{k=r} (x_k - \mu)^2}{r - 1}}, \tag{3}$$

where the vector $\mathbf{x} = \{x_k\}_{k=1}^{k=r}$, x_k is the value of kth pattern, and r is the total number of patterns. Obviously, in (3), the component $(x_k - \mu)$ is a distance between kth pattern and its mean. This distance can be written as $\|D_{x_k,\widetilde{\mu}}\|$, the Euclidean norm of d-dimensional feature space;

$$\left\| D_{x_k,\widetilde{\mu}} \right\| = \sqrt{\sum_{i=1}^{d} (x_{k,i} - \mu_i)^2}. \tag{4}$$

Here d is the total number of features imported so far. Therefore, to calculate the standard deviation of r patterns in two dimensions, (3) can be written as

$$\text{std} (\mathbf{x}) = \sqrt{\frac{\sum_{k=1}^{k=r} \left[(x_{k,1} - \mu_1)^2 + (x_{k,2} - \mu_2)^2 \right]}{r - 1}}, \tag{5}$$

and for a tridimensional space, the equation is

$$\text{std} (\mathbf{x})$$
$$= \sqrt{\frac{\sum_{k=1}^{k=r} \left[(x_{k,1} - \mu_1)^2 + (x_{k,2} - \mu_2)^2 + (x_{k,3} - \mu_3)^2 \right]}{r - 1}}. \tag{6}$$

Accordingly, multidimensional standard deviation used in (2) of r patterns in an m-dimensional space is

$$\text{std} (\mathbf{x}) = \sqrt{\frac{\sum_{k=1}^{k=r} \sum_{i=1}^{i=m} (x_{k,i} - \mu_i)^2}{r - 1}}, \tag{7}$$

$$\mathbf{x} = \{x_{k,d}\}_{k=1,d=1}^{k=r,d=m} \in R_{\text{feature}}^{r \times m}.$$

For EEG eye state identification, feature ordering can be derived based on (1) and MMDC in the first place, and then time-series data with sorted features are imported into the predictive systems according to the feature ordering. Moreover, it is necessary to notice that feature ordering should be obtained only from the training data, so that it is possible to get rid of the influence from the testing data during the preprocessing and training, and the classification results will be closer to the real situation. However, in the validation and testing stages, all the data should be sorted according to the feature ordering derived in the preprocessing by training dataset. Such an operation can ensure that all the features are trained, validated, and tested in the same ordering; and all the features are sorted according to their discrimination abilities. If the feature orderings in the training, validation, and testing phases are different, features with weak discrimination ability will be trained in an early stage. However, this will reduce the accuracy of the classification.

4. Time-Series Classification Approach Based on IAL

4.1. Time-Series Classification.
There are two kinds of different times-series classification approaches: instance-based and feature-based [29] approaches. The instance-based approach predicts the classification results for the testing instances based on the similarity to the training instances. In this approach, the nearest-neighbor classifiers with Euclidean distance (NN-Euclidean) [30] or dynamic time warping (NN-DTW) [31, 32] were widely employed. In another aspect, the feature-based approach builds temporal features extracted from the original features and potentially can outperform instance-based classifiers. Feature-based classifiers commonly consist of two steps: (1) defining the temporal features and (2) training a classifier based on the temporal features defined. Due to the fact that IAL employs features one by one to the predictive system, it has little linkage to the instance-based approaches. Therefore, when IAL is employed to solve time-series classification, feature-based approach is more suitable than instance-based approach and should be employed by IAL during the time-series classification process.

4.2. Feature Extraction for Time-Series Classification.
Before the formal time-series classification, it is necessary to preprocess the experimental data in two stages: firstly, feature extraction from original data and, secondly, feature ordering for IAL. In comparison with some other classification problems, apart from feature ordering which is a special preprocessing in IAL, temporal feature extraction from the original data is another unique step in time-series classification problems.

Temporal feature extraction aims to classify instances based on the original feature and the state difference in the time distance. Usually, the original features and those directly derived from original features are called first-order features, while features extracted from the state difference in different time distance are called second-order features. Equation (8) is the formula for the second-order features:

$$y'(t) = F[y(t+D), y(t)], \tag{8}$$

where $y(t)$ is the first-order feature, D is the time distance length between two instances, and F is the function with $y(t+D)$ and $y(t)$. Theoretically, F can be any calculation rules. Moreover, some statistical metrics like mean and standard deviation are often used for feature extraction [33].

In the process of time-series classification, feature extraction and feature ordering should be carried out before training. However, different from feature ordering which is derived only by training data, temporal feature extraction is a process for new feature building; thus it is necessary for it to be implemented with all datasets. Figure 2 demonstrates the working flow of a time-series classification system based on ITID.

5. Experiments

5.1. Benchmark.
In this study, the EEG eye state corpus from UCI machine learning repository is employed for the experiments [13]. This EEG eye state dataset was donated by Rösler and Suendermann from Baden-Wuerttemberg Cooperative State University (DHBW), Stuttgart, Germany [12]. All data were derived from one continuous EEG measurement with the Emotiv EEG neuroheadset, which is shown in Figure 3. There are 14980 patterns and 14 features in the dataset, where the 14 features are the data obtained by 14 sensors shown in Figure 4. The duration of the measurement was 117 seconds. The eye state was detected via a camera during the EEG measurement and added later manually to the file after analyzing the video frames.

In this eye state corpus, there are three instances with the numbers 899, 10387, and 11510 having obvious errors. Their values are out of almost 15 k, so they are outliers, which should be deleted before the experiments. Therefore, only 14977 instances are employed in Rösler's experiments. To compare the results derived from our experiments with those obtained in the previous studies, those three error instances are also discarded in our experiments.

For the output of the corpus, "1" indicates the eye-closed and "0" the eye-open state. There are 6722 legal eye-closed instances and 8255 legal eye-open instances. All values are in chronological order with the first measured value at the top of the data.

Table 1 is the overview of 14977 legal values obtained by those 14 sensors. It presents the minimum and maximum values and means and standard deviations of eye-closed and eye-open states, respectively. According to this table, it is manifested that although the minimum and maximum values are quite different in both eye-closed and eye-open states, the means of these two states of the same feature are almost the same. Nevertheless, the standard deviations of both states have a large difference which is obvious in the eye-open state. Moreover, it can be easily found that the maximum values of eye-open state of features F7, F3, F4, FC6, T8, P8, and O2 are much higher than the same feature's maximum values in
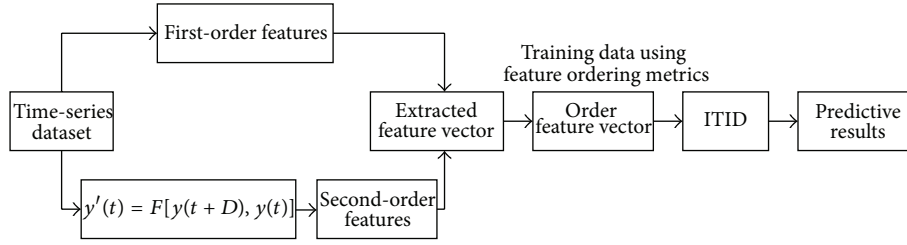
FIGURE 2: The working flow of time-series classification based on ITID.

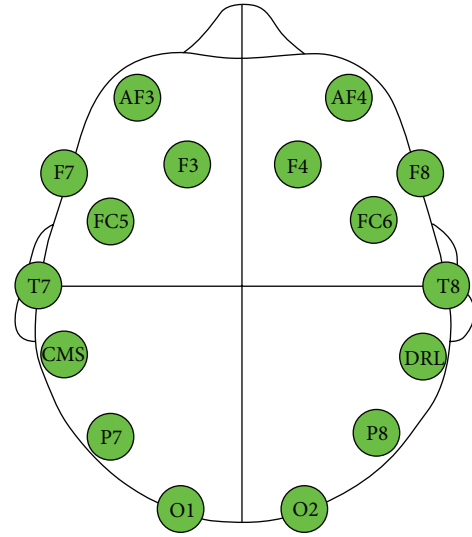FIGURE 3: Emotiv EEG neuroheadset (http://www.emotiv.com).

FIGURE 4: Emotiv EEG neuroheadset sensor position and corresponding behavior groups. AF3, AF4, FC5, F8, T7, P7, and O1 correspond to a maximum increase while F7, F3, F4, FC6, T8, P8, and O2 correspond to a minimum decrease when the tester opens eyes.

the eye-closed state; and for features AF3, AF4, FC5, F8, T7, P7, and O1, the minimum values of eye-open state are much lower than those in eye-closed state. Therefore, both the mean and standard deviation should be extracted as new features in the eye state classification process.

5.2. Experiments. During the experiments, all the patterns were partitioned for training, validation, and testing with the divisions of 50%, 25%, and 25%, respectively, and sorted according to the time-series sequence. Six different approaches for eye state classification were employed in the experiments. They are designed as shown in Table 2, while Table 3 presents the final classification results by different approaches.

(1) IAL with Feature Extraction and Time-Series 1. This is a time-series approach based on IAL. All the features used in this approach are second-order features, which are the averages and standard deviations derived from every 12 instances of original features. These features have been shown in (9), which were improved from (8),

$$y'(t) = \frac{\sum_{D=0}^{D=11} y(t+D)}{12},$$

$$y'(t) = \sqrt{\frac{\sum_{D=0}^{D=11} \left[y(t+D) - \mu\right]^2}{12}}. \tag{9}$$

The time distance here is 12, where $0 \leq D \leq 11$, $D \in N$. This time distance is equal to the shortest length of a blink, which is 100 milliseconds, because the total number of instances in the eye state corpus is 14980, which was recorded in 117 seconds. As a result of that, there are 14965 instances and 28 features. Before the classification, feature ordering was derived by AD and MMDC, which has been shown in Table 4. It is obvious that the standard deviations played a more important role in the classification, because all the standard deviations were imported earlier than the averages except the thirteenth feature which is an average in the first place. Such a phenomenon coincides with the data shown in Table 1, where the averages are approximately the same, while the standard deviations are quite different. In addition, the number of features in this approach is 28. It needs feature selection; otherwise the computational cost will be large. Therefore an IAL approach with feature selection is employed [34]. At last only two features were used in this time-series classification, and it obtained a good result in classification error rate, which is 27.3991%.

TABLE 1: Overview of sensor values.

Eye state		Eye-closed				Eye-open			
	Sensor	Min.	Mean	Max.	Std.	Min.	Mean	Max.	Std.
1	AF3	4198	4305	4445	33.46	1030	4297	4504	54.27
2	F7	3905	4005	4138	27.54	3924	4013	7804	52.37
3	F3	4212	4265	4367	20.13	4197	4263	5762	27.66
4	FC5	4058	4121	4214	21.31	2453	4123	4250	27.58
5	T7	4309	4341	4435	18.08	2089	4341	4463	29.54
6	P7	4574	4618	4708	17.44	2768	4620	4756	28.06
7	O1	4026	4073	4167	24.14	3581	4071	4178	18.60
8	O2	4567	4616	4695	18.44	4567	4615	7264	34.38
9	P8	4147	4202	4287	18.55	4152	4200	4586	17.69
10	T8	4174	4233	4323	19.36	4152	4229	6674	33.48
11	FC6	4130	4204	4319	23.71	4100	4200	5170	27.08
12	F4	4225	4281	4368	18.57	4201	4277	7002	36.62
13	F8	4510	4610	4811	32.79	86	4601	4833	59.88
14	AF4	4246	4367	4552	34.82	1366	4356	4573	52.28

TABLE 2: Design of experiments.

	IAL	Batch-training	Time-series	First-order feature	Second-order feature	$F[y(t + D), y(t)]$	Feature selection
Approach 1	●		●		●	Equations (9), (10)	●
Approach 2	●		●	●	●	Equation (10)	
Approach 3	●			●			
Approach 4	●			●(pure original)			
Approach 5		●	●	●		Equation (10)	
Approach 6		●		●			

(2) IAL with Feature Extraction and Time-Series 2. This is also an IAL time-series classification approach. With the feature ordering based on AD and MMDC and feature extraction based on average and standard deviation, this approach also follows the model shown in Figure 2. To investigate the microinfluence brought by time-series, the time duration is set to be 1; namely, $D = 1$. Equation (8) here is

$$y'(t) = y(t + 1) - y(t). \tag{10}$$

Four extracted features were employed in this approach: the first-order means and the first-order standard deviations of all the values of each original instance and the second-order average and second-order standard deviation of all the values of each instance calculated by (10). Therefore, the total instance number in this approach is 14976. Based on the IAL feature ordering metric using AD, these features were sorted following this order: the first-order means of instances, the first-order standard deviations of instances, the second-order standard deviation, and the second-order average. This feature ordering has been shown in Table 4. The error rate obtained in the final classification result is 27.4573%.

(3) IAL with Feature Extraction. This approach is IAL classification without time-series. In comparison with the first and second approaches, this experiment is designed to show the effect of time-series factors, namely, $y'(t)$ in (8), which is employed in the previous experiments but will not be used in this experiment. Thus there are no time-series factors considered in this approach. As a result, there are no second-order features. Consequently, there are no second-order average and second-order standard deviation. Therefore, this experiment only retains the first-order feature extraction process. During the experiment, IAL with feature extraction is employed merely based on the first-order average and first-order standard deviation, but without considering time-series factors. In this way, all the original features are trained according to the original ordering with newly built average and standard deviation of each instance following behind. This has been shown in Table 4. The final classification error rate is 27.4793%, which is slightly worse than the first approach.

(4) Pure IAL Approach. This approach only employs IAL with feature ordering derived by AD and MMDC. There is no new feature extracted from the original data. Moreover, no time-series impact factors are used in this approach. Such a design aims to find out the influence brought by time-series and first- and second-order feature extraction. The feature ordering is shown in Table 4. The classification error rate is 27.4693%, which approximates to the results obtained in the first and second approaches.

TABLE 3: EEG eye state time-series classification results.

Approaches	Description	Classification error rate (%)
1	AD feature ordering with feature extraction from means and standard deviations	27.3991
2	AD feature ordering with feature extraction where $y'(t) = y(t+1) - y(t)$	27.4573
3	Feature extraction without time-series and original ordering	27.4793
4	AD feature ordering without time-series and feature extraction	27.4693
5	Conventional batch-training method, where $y'(t) = y(t+1) - y(t)$	29.5046
6	Conventional batch-training method without time-series	30.6328

TABLE 4: Feature orderings for IAL approaches.

Approaches	Description	Feature orderings
1	AD feature ordering with feature selection (in bold)	**AVG(F8)-STD(FC5)**-*STD(T7)-STD(F3)-STD(F7)-STD(F4)-STD(P7)-STD(AF4)-STD(FC6)-STD(O2)-STD(F8)-STD(T8)-STD(AF3)-STD(P8)-STD(O1)-AVG(F7)-AVG(F3)-AVG(FC6)-AVG(F4)-AVG(FC5)-AVG(AF4)-AVG(T7)-AVG(P8)-AVG(AF3)-AVG(T8)-AVG(P7)-AVG(O2)-AVG(O1)*
2	AD feature ordering	AVG1-STD1-STD2-AVG2
3	Original ordering	AF3-F7-F3-FC5-T7-P7-O1-O2-P8-T8-FC6-F4-F8-AF4-AVG1-STD1
4	AD feature ordering	F8-F7-FC6-F3-F4-FC5-T7-P8-T8-AF4-O2-P7-AF3-O1

(5) Time-Series Classification with Batch-Training. This method has the same consideration of time-series, which is similar to the second approach, except that all features are trained using the conventional batch-training method. The time duration is the same as that in approach 2, where $D = 1$. Therefore, the total instance number in this approach is 14976. All the features are extracted according to (10) and trained in one batch. As such, there is no average or standard deviation extracted from the extract vector derived by (10). The objective of this experiment is to check the effect of IAL and second-order feature extraction compared using the first and second approaches. The error rate in the final time-series classification is 29.5046%, which is much higher than in the previous IAL approaches.

(6) Conventional Batch-Training Method. The last approach is the conventional method based on back propagation neural networks without considering time-series, whereby all the original features are directly trained in one batch. This approach obtains the highest error rate among all four experiments. The error rate is about 30.6328% in the final classification result.

5.3. Result Analysis. According to the experimental results of these six different approaches shown and compared in Table 3, the first approach obtains the lowest classification error rate and the conventional batch-training method exhibits the worst and highest classification error rate of 30.63%. In comparison with Rösler's experimental results using multilayer perceptron, where the error rate is more than 30% [12], the results derived by IAL approaches are much better. All of these classification error rates are lower than 30%. This merely indicates that, firstly, IAL approach can outperform conventional batch-training methods; secondly, feature extraction with time-series properties is very useful in the improvement of the classification results for time-series problems; thirdly, feature ordering is very important to IAL. Moreover, feature orderings derived by AD and MMDC can outperform the original feature ordering in IAL.

6. Conclusions

In this paper, a time-series classification approach based on IAL is proposed for EEG eye state identification. The approach is novel in a way that it firstly extracts features from the raw data and then sorts these features using IAL feature ordering approach according to feature's discrimination ability. During the training process, the newly extracted features are imported into the neural predictive system in a sequential order based on the feature ordering. In comparison with the conventional batch-training methods and feature extraction method without considering the relation between time-series data, the experimental results of time-series IAL showed that such a machine learning approach can not only cope with time-series classification problems but also improve the accuracy of the classification results. Moreover, the experimental results also imply that the relation among time-series data is crucial to the data analysis in such classification problems.

In future, some issues remain as open topics for further research. For instance, besides the mean and standard deviation, whether there exist any other time-series related features for time-series classification is still an open problem. Secondly, the approach to extract new second-order features from raw data is also significant. Thirdly, the correlations between second-order features and first-order features often vary. Hence, the influence existing in the time-series classification process is still unknown. Last but not the least, whether there exists an optimal method for both first- and second-order feature ordering in IAL training is also a challenging problem for time-series classification.

In general, the feasibility of IAL-based time-series classification approach proposed in this paper has been validated

by EEG eye state identification experiments. The final results indicated that IAL is applicable for EEG time-series classification. However, there are still a number of work items remaining for the future studies.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] P. A. Estévez, C. M. Held, C. A. Holzmann et al., "Polysomnographic pattern recognition for automated classification of sleep-waking states in infants," *Medical and Biological Engineering and Computing*, vol. 40, no. 1, pp. 105–113, 2002.

[2] M. V. M. Yeo, X. Li, K. Shen, and E. P. V. Wilder-Smith, "Can SVM be used for automatic EEG detection of drowsiness during car driving?" *Safety Science*, vol. 47, no. 1, pp. 115–124, 2009.

[3] K. Polat and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," *Applied Mathematics and Computation*, vol. 187, no. 2, pp. 1017–1026, 2007.

[4] K. Sadatnezhad, R. Boostani, and A. Ghanizadeh, "Classification of BMD and ADHD patients using their EEG signals," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1956–1963, 2011.

[5] N. Sulaiman, M. N. Taib, S. Lias, Z. H. Murat, S. A. M. Aris, and N. H. A. Hamid, "Novel methods for stress features identification using EEG signals," *International Journal of Simulation: Systems, Science and Technology*, vol. 12, no. 1, pp. 27–33, 2011.

[6] T. Nguyen, T. H. Nguyen, K. Q. D. Truong, and T. van Vo, "A mean threshold algorithm for human eye blinking detection using EEG," in *Proceedings of the 4th International Conference on the Development of Biomedical Engineering in Vietnam*, pp. 275–279, Ho Chi Minh City, Vietnam, 2013.

[7] J. Hua Ang, S. U. Guan, K. C. Tan, and A. A. Mamun, "Interference-less neural network training," *Neurocomputing*, vol. 71, no. 16-18, pp. 3509–3524, 2008.

[8] S.-U. Guan and J. Liu, "Incremental neural network training with an increasing input dimension," *Journal of Intelligent Systems*, vol. 13, no. 1, pp. 45–69, 2004.

[9] T. Wang, S. U. Guan, T. O. Ting, K. L. Man, and F. Liu, "Evolving linear discriminant in a continuously growing dimensional space for incremental attribute learning," in *Proceedings of the 9th IFIP International Conference on Network and Parallel Computing (NPC '12)*, pp. 482–491, Gwangju, Republic of Korea, 2012.

[10] O. Fukuda, T. Tsuji, and M. Kaneko, "Pattern classification of EEG signals using a log-linearized Gaussian mixture neural network," in *Proceedings of the IEEE International Conference on Neural Networks. Part 1 (of 6)*, pp. 2479–2484, Perth, Australia, December 1995.

[11] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, pp. 10–18, 2009.

[12] O. Rösler and D. Suendermann, "First step towards eye state prediction using EEG," in *Proceedings of the International Conference on Applied Informatics for Health and Life Sciences (AIHLS '13)*, Istanbul, Turkey, 2013.

[13] A. Frank and A. Asuncion, "UCI machine learning repository," 2010, http://archive.ics.uci.edu/ml.

[14] T. Wang and S. U. Guan, "Feature ordering for neural incremental attribute learning based on Fisher's linear discriminant," in *Proceedings of the 5th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC '13)*, Hangzhou, China, 2013.

[15] S. U. Guan and J. H. Ang, "Incremental training based on input space partitioning and ordered attribute presentation with backward elimination," *Journal of Intelligent Systems*, vol. 14, no. 4, pp. 321–351, 2005.

[16] S. U. Guan and F. Zhu, "An incremental approach to genetic-algorithms-based classification," *IEEE Transactions on Systems, Man, and Cybernetics B: Cybernetics*, vol. 35, no. 2, pp. 227–239, 2005.

[17] F. Zhu and S. U. Guan, "Ordered incremental training with genetic algorithms," *International Journal of Intelligent Systems*, vol. 19, no. 12, pp. 1239–1256, 2004.

[18] S. U. Guan and S. Li, "Incremental learning with respect to new incoming input attributes," *Neural Processing Letters*, vol. 14, no. 3, pp. 241–260, 2001.

[19] X. Liu, G. Zhang, Y. Zhan, and E. Zhu, "An incremental feature learning algorithm based on least square support vector machine," in *Proceedings of the 2nd International Frontiers in Algorithmics Workshop (FAW '08)*, pp. 330–338, Changsha, China, 2008.

[20] W. Bai, S. Cheng, E. M. Tadjouddine, and S. U. Guan, "Incremental attribute based particle swarm optimization," in *Proceedings of the 8th International Conference on Natural Computation (ICNC '12)*, pp. 669–674, Chongqing, China, 2012.

[21] S. Chao and F. Wong, "An incremental decision tree learning methodology regarding attributes in medical data mining," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, pp. 1694–1699, Baoding, China, July 2009.

[22] S. U. Guan and J. Liu, "Incremental ordered neural network training," *Journal of Intelligent Systems*, vol. 12, no. 3, pp. 137–172, 2002.

[23] T. Wang, S. U. Guan, and F. Liu, "Feature discriminability for pattern classification based on neural incremental attribute learning," in *Foundations of Intelligent Systems*, vol. 122 of *Advances in Intelligent and Soft Computing*, pp. 275–280, 2012.

[24] S. U. Guan and J. Liu, "Feature selection for modular networks based on incremental training," *Journal of Intelligent Systems*, vol. 14, no. 4, pp. 353–383, 2005.

[25] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[26] T. Wang and Y. Wang, "Pattern classification with ordered features using mRMR and neural networks," in *Proceedings of the International Conference on Information, Networking and Automation (ICINA '10)*, pp. V2128–V2131, chn, October 2010.

[27] T. Wang, S. U. Guan, and F. Liu, "Entropic feature discrimination ability for pattern classification based on neural IAL," in *Proceedings of the 9th International Symposium on Neural Networks (ISNN '12)*, pp. 30–37, Shenyang, China, 2012.

[28] T. Wang and S. U. Guan, "Feature ordering for neural incremental attribute learning based on Fisher's linear discriminant," in *Proceedings of the 5th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC '13)*, pp. 507–510, 2013.

[29] H. Deng, G. Runger, E. Tuv, and M. Vladimir, "A time series forest for classification and feature extraction," *Information Sciences*, vol. 239, pp. 142–153, 2013.

[30] Z. Xing, J. Pei, and P. S. Yu, "Early prediction on time series: s nearest neighbor approach," in *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI '09)*, pp. 1297–1302, Pasadena, Calif, USA, July 2009.

[31] D. Yu, X. Yu, Q. Hu, J. Liu, and A. Wu, "Dynamic time warping constraint learning for large margin nearest neighbor classification," *Information Sciences*, vol. 181, no. 13, pp. 2787–2796, 2011.

[32] Y.-S. Jeong, M. K. Jeong, and O. A. Omitaomu, "Weighted dynamic time warping for time series classification," *Pattern Recognition*, vol. 44, no. 9, pp. 2231–2240, 2011.

[33] A. Nanopoulos, R. Alcock, and Y. Manolopoulos, "Feature-based classification of time-series data," in *Information Processing and Technology*, M. Nikos and D. N. Stavros, Eds., pp. 49–61, Nova Science, 2001.

[34] T. Wang, S. U. Guan, and F. Liu, "Feature selection in growing dimensional space for classification based on neural incremental attribute learning," in *Proceedings of the 9th International Symposium on Linear Drives for Industry Applications*, X. Liu and Y. Ye, Eds., vol. 270, pp. 501–507, Springer, 2014.