Data Selection in EEG Signals Classification

Shuaifang Wang, Yan Li and Peng Wen，David Lai

Abstract The alcoholism can be detected by analyzing electroencephalogram (EEG) signals. However, analyzing multi-channel EEG signals is a challenging task, which often requires complicated calculations and long execution time. This paper proposes three data selection methods to extract representative data from the EEG signals of alcoholics. The methods are the principal component analysis based on graph entropy (PCA-GE), the channel selection based on graph entropy (GE) difference, and the mathematic combinations channel selection, respectively. For comparison purposes, the selected data from the three methods are then classified by three classifiers: the J48 decision tree, the K-nearest neighbor (KNN) and the Kstar, separately. The experimental results show that the proposed methods are successful in selecting data without compromising the classification accuracy in discriminating the EEG signals from alcoholics and non-alcoholics. Among them, the proposed PCA-GE method uses only 29.69% of the whole data and 29.5% of the computation time but achieves a 94.5% classification accuracy. The channel selection method based on the GE difference also gains a 91.67% classification accuracy by using only 29.69% of the full size of the original data. Using as little data as possible without sacrificing the final classification accuracy is useful for online EEG analysis and classification application design.

Keywords EEG, data selection, horizontal visibility graph (HVG), principal component analysis (PCA).

1. INTRODUCTION

Discovered by Hans Berger [1] in 1924, EEGs are recorded using multiple electrodes placed on the scalp to measure voltage fluctuations resulting from ionic current flows within the neurons of the brain. The brain electrochemical activity is widely used in the detection of epilepsy [2-5] as well as the assessment of alcoholism [6], the characterization of sleep phenomena[7,8], the diagnosis of encephalopathy [9], depression and Creutzfeldt-Jakob disease [10], and monitoring the depth of anesthesia [11,12]. The advantages, such as having short time constants, less environmental limits and inexpensive equipment, ensure the wide practical uses of EEGs. Instead of making visual presentations of the brain's anatomy like computed tomography (CT) or magnetic resonance imaging (MRI), EEGs evaluate the brain's physiology with a millisecond-range temporal resolution in a convenient and relatively inexpensive way. EEG signals play a central role in the diagnosis and management of patients with brain disorders, working in conjunction with other diagnostic techniques developed over the last 30 or so years.

People who drink alcohol excessively suffer from blurred vision, difficulty walking, slurred speech, slow reaction, impaired memory and sleep [13]. Long-term alcohol abuse is called alcoholism. Alcoholism is a common neurological disease which may not only lead to cognitive, identification and mobility impairments, but may also damage the brain systems [14]. Clinical evidences of using advanced signal processing methods have proven that detecting alcoholism from the EEG signals can be effective [15-17]. Therefore, an increasing number of researchers are studying the connections between EEGs and alcoholics.

Currently, most of the diagnoses are done by traditional visual inspections in the clinical settings. However, it is time-consuming, error prone and highly trained medical professionals are needed. Therefore, automatic EEG analysis and classification systems are the trend in both research and clinical areas. In automatic EEG classification, the amount of data needed increases exponentially

Shuaifang Wang,
Faculty of Health, Engineering and Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia (e-mail: Shuaifang.Wang@usq.edu.au).

Yan Li,
Faculty of Health, Engineering and Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia (e-mail: yan.li@usq.edu.au).

Peng (Paul) Wen
Faculty of Health, Engineering and Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia (e-mail: peng.wen@usq.edu.au).

David Lai
Faculty of Health, Engineering and Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia (e-mail: david.lai@usq.edu.au).

with the dimensionality of the feature vectors to gain high classification accuracy. It is recommended to use, at least, five to ten times as many training samples per class as the dimensionality. The analysis and classification of EEG signals require a large amount of data when dealing with high dimensional EEG data by supervised classification. Besides, considering the computation time of the classification, data reduction is essential. Therefore, how to reduce the amount of data while still preserving the original critical information is one of the major problems in EEG research. Of course, better classifiers also contribute to the improvement of classification accuracy.

There has been a host of related work on automatic EEG classification published in the literature. Siuly [5] chose nine statistical features instead of using all the data points from each channel. Subasi [18] decomposed EEG signals into frequency sub-bands using discrete wavelet transform and classified normal and epileptic EEGs with a mixture of expert modes. İnan Güler and Elif Derya Übeyli [19] extracted features using wavelet transform and the adaptive neuro-fuzzy inference system trained with the backpropagation gradient descent method in combination with the least squares method. Toshio et al. [20] employed a Gaussian mixture model to conduct EEG pattern classification. Vasicek [21] tested the normality using sample entropy. Kemal [22] detected epileptic seizures in EEG signals using a hybrid system based on a decision tree classifier and fast Fourier transform with 98.72% classification accuracy. Suryannarayana et al. [23] introduced cross-correlation aided SVM based classifier, and achieved 95.96% classification accuracy with normal and epileptic EEG data. Guohun Zhu et al. [24] analysed alcoholic EEG signals based on HVG entropy, which dramatically decreased the data size to be processed. Naoki Tomida et al. [25] used an active data selection method for motor imagery EEG data classification. Most of the studies aim at improving the classification accuracy only while my work is evaluated on terms of both classification accuracy and execution time.

This study applies three different data selection methods and compares their performances on EEG signals from alcoholics. The first method is the PCA based on GE features. The second one is the channel selection based on GE difference. The third one is the mathematic combinations channel selection, which chooses the corresponding numbers of channels randomly to get a subset of the extracted data. All of the three methods perform the features extracted based on the HVGs mapped from the original data. After that, all the selected data are classified by the J48 decision tree, the KNN and the Kstar.

2. EXPERIMENTAL DATA

The EEG signals (SMNI_CMI_TRAIN.tar.gz and SMNI_CMI_TEST.tar.gz) from alcoholics and the control subjects used in this paper were published by Henri Begleiter from State University of New York Health Center [26]. The large data sets contain data from 10 alcoholic and 10 control subjects, with 10 runs per subject. There are 600 samples making up of 64 channels of data in SMNI_CMI_TRAIN.tar.gz and 600 samples making up of 64 channels of data in SMNI_CMI_TEST.tar.gz, respectively. Each data sample contains the signals digitized at 256 Hz for one second. The indices of the 64 electrodes are "FP1", "FP2", "F7", "F8", "AF1", "AF2", "FZ", "F4", "F3", "FC6", "FC5", "FC2", "FC1", "T8", "T7", "CZ", "C3", "C4", "CP5", "CP6", "CP1", "CP2", "P3", "P4", "PZ", "P8", "P7", "PO2", "PO1", "O2", "O1", "X", "AF7", "AF8", "F5", "F6", "FT7", "FT8", "FPZ", "FC4", "FC3", "C6", "C5", "F2", "F1", "TP8", "TP7", "AFZ", "CP3", "CP4", "P5", "P6", "C1", "C2", "PO7", "PO8", "FCZ", "POZ", "OZ", "P2", "P1", "CPZ", "nd" and "Y". The electrodes, "X" and "Y", are EOG signals; and "nd" is the reference electrode. The locations of the EEG electrodes used for data acquisition are shown in Fig. 1. In this paper, the data from SMNI_CMI_TRAIN.tar.gz are used as the training data, and those from SMNI_CMI_TEST.tar.gz are used as the testing data, respectively.

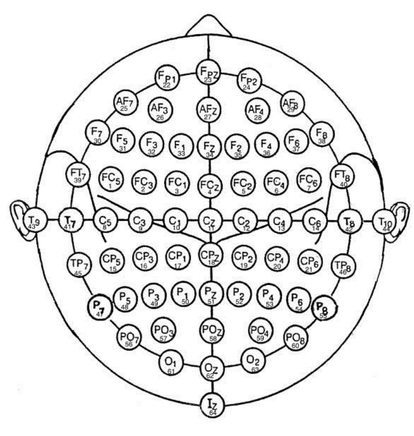

Fig. 1 Electrode Location.

3. METHODOLOGY

The workflow of the three proposed data selection methods is shown in Fig. 2.

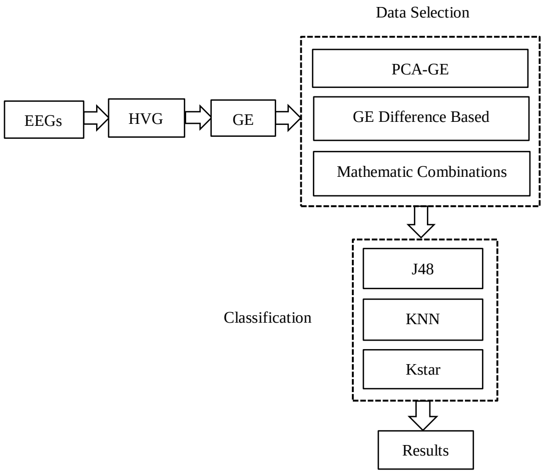

Fig. 2 The workflow of the proposed methods.

Data selection aims at using optimal subsets of variables, while retaining as much useful information as possible. The implementation details are described below.

- *HVG*

A HVG is a mapping between time series and complex network [27] according to a specific geometric criterion to make use of methods of complex network theory for characterizing time series. Each datum in the time series corresponds to a node in the graph, such that two nodes are connected if their corresponding data heights are larger than all the data heights between them [28]. Its degree distribution is a good discriminator between randomness and chaos. Let **X** = ($x_i \in R \geq 0$: i = 1, 2, . . . , n) be an ordered set (or, equivalently, a sequence) of non-negative real numbers. The HVG of **X** is graph **G = (X, E),** where **X** is a set of elements called nodes and **E** is a set of unordered pairs of nodes called edges. Its definition is shown in equation (1):

$$e_{ij} = \begin{cases} 1, (x_k < x_i) \wedge (x_k < x_j) \\ 0, otherwise \end{cases} \quad (1)$$

where every $k \in (i, j)$. In graph theory, the degree of a node (or vertex) of a graph is the number of edges connecting to the node, with loops counted twice [29]. The degree of a vertex is denoted as deg(x_i). The degree sequence (***DS***) is the sequence of the degree of a graph. The node degree and its sequence can be used to describe the characteristics of the graph.

In this paper, the time series of EEGs are mapped into graphs (***G (X, E)***). Each EEG sample is mapped to a HVG, and each HVG has a GE value. There are 1200 samples to be analyzed for each electrode from 10 different trails. Totally 76800 features are extracted for 64 electrodes. All GE features are evaluated with groups of alcoholics or non-alcoholics. To illustrate the data transformation process, let us take the dataset co2a0000368 from electrode FP1 in the forementioned database for an example. Given ***X***= {5.015, 5.503, 4.039, 2.085, 0.132, 0.132, 0.621, 0.621, 0.132, -0.356, -0.844, -0.356}, we can get the degree sequence ***DS***= {1, 2, 2, 3, 2, 2, 3, 2, 2, 3, 2, 2} by the following implementation:

(a). Transform ***X*** into ***Xnew***, making every element be a non-negative real number by adding the absolute value of the smallest value which is negative. For example,

X= {5.015, 5.503, 4.039, 2.085, 0.132, 0.132, 0.621, 0.621, 0.132, -0.356, -0.844, -0.356}

should be transformed as

Xnew = {5.859, 6.347, 4.883, 2.929, 0.976, 0.976, 1.465, 1.465, 0.976, 0.488, 0, 0.488}, which is demonstrated in Fig. 3.

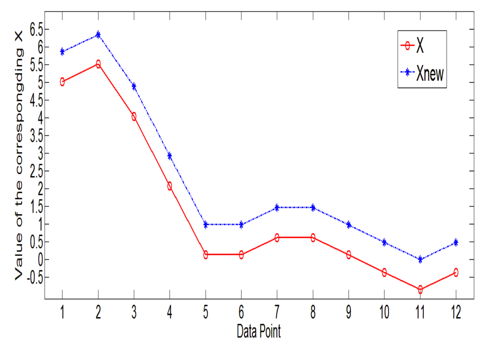

Fig. 3 Nonnegative transform of X.

(b). Horizontal visibility check is used to calculate the degree of each node, which is shown in Fig. 4. Two nodes *i* and *j* in the graph are connected if one can draw a horizontal line in the time series joining x_i and x_j that does not intersect any intermediate data height.

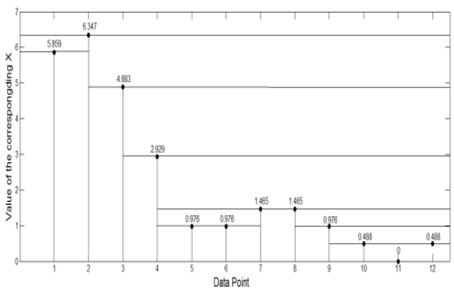

Fig. 4 The degree of each node from HVGs.

(c). Degree sequence. For arbitrary datum in the time series, we calculate the visibility with all the other corresponding nodes and record the number of edges connecting to it as the degree of the node. In the above example, the degree sequence is as follows:

DS= {1, 2, 2, 3, 2, 2, 3, 2, 2, 3, 2, 2}.

- *GE*

The GE is the entropy of the frequency distribution of the node connections in an undirected and unweighted HVG. It is a function from information theory on a graph ***G***, with a probability distribution $p(k)$ on its node set. It was introduced by Janos Korner in [30]. Shannon entropy [31] is used in this paper, which is shown in equation (2):

$$h = -\sum_{k=1}^{n} p(k)\log(p(k)) \qquad (2)$$

where $entropy$ is the degree distribution of graph **G**. The degree distribution $p(k)$ of a network is defined to be the fraction of nodes in the network with degree k. Thus if there are n nodes in total in a network and n_k of them have degree k, we have equation (3) below:

$$p(k) = n_{\mathrm{k}} / n \qquad (3)$$

In the above case, $p(k)$ of DS is (0, 1/12, 8/12, 3/12). The GE is 0.824 when it takes the logarithm base two. The Mean GE plot from 64 electrodes is shown in Fig. 5. From Fig. 5, it is clear that the differences between the alcoholics and the control subjects are indeed different from channel to channel. That is the reason why optimal subsets of channel selection are possible. In this paper, the principal component analysis, the GE difference and the mathematic combinations based on GE channel selection are proposed. The details of the proposed methods are demonstrated below.

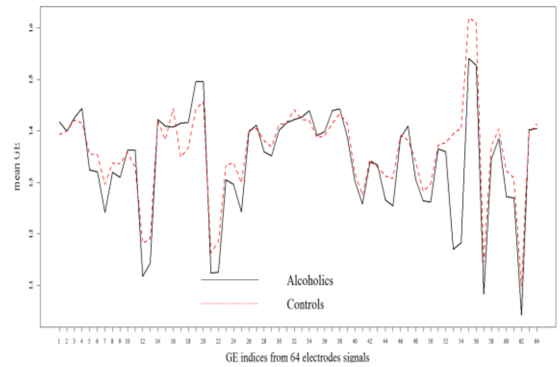

Fig. 5 Mean GE from 64 electrode signals.

3.1 The PCA Based on GE from HVG (PCA-GE)

Invented by Pearson [32] in 1901, the PCA was widely used in mechanics and independently developed (and named) by Harold Hotelling later in the 1930s [33]. Nowadays, it is used as a tool in exploratory data analysis and for making predictive models. The faithful transformation T = XW maps a data vector X from an original space to a new space of p variables which are uncorrelated over the dataset. However, not all the principal components are kept. Keeping only the first L principal components, it gives the truncated transformation as shown in equation (4):

$$T_L = XW_L \qquad (4)$$

where matrix T_L now has n rows but only L columns. By reconstruction, all the transformed data matrices reserve only L columns out of the original data. Such dimensionality reduction can be a very useful step for visualizing and processing high-dimensional data while keeping as much useful information as possible. In order to keep the same size of input data for the further classification process, here, the corresponding number of the principal components which are the same as that of channels has been chosen. Therefore, the dimensionalities of all the samples are the same. The PCA is implemented in Matlab2013b. The distribution percentage of the total power is shown in Fig.6 as follows.

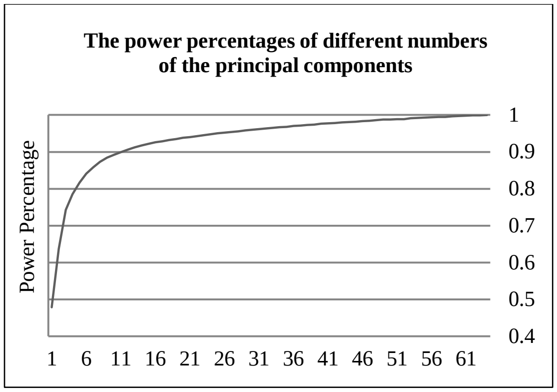

Fig. 6 The corresponding power percentages of different numbers of the principal components from the full size of data.

The PCA-GE technique is applied to extracted representative data transformed from the dataset without specific channel selection investigation. For the alcoholic database, there are 64 electrodes of signals per trial. The inconvenient data preparation and complicated calculations are still challenging for an online analysis and classification system. In the following section, how to gain an optimal subset of specific channels is discussed.

3.2 The GE Difference Based Channel Selection

From Fig. 5, it is clear that the mean GE differs from electrode to electrode between alcoholic subjects and non-alcoholic ones. Therefore, the channel selection based on the GE difference is proposed. Firstly, the electrodes should be ordered degressively according to the mean GE gap values. They are C1, C2, PO8, PO7, C3, FC2, FCZ, CP2, CPZ, PZ, FZ, CP5, F1, P2, C4, FC1, F2, P4, CP1, P1, CP6, CZ, CP4, AFZ, FC5, AF2, AF1, F8, P3, TP7, T7, POZ, F3, FPZ, FT7, FP1, PO2, AF8, OZ, X, F4, CP3, P6, FC3, PO1, FC4, FT8, O2, Y, F6, P7, P5, nd, C6, C5, TP8, AF7, F7, F5, FC6, T8, P8, FP2, and O1. For comparison reasons, the corresponding specific numbers of channels are selected to generate the optimal subsets for classification. For example, C1 is selected to gain the one-channel subset because the mean GE gap is the largest among all the channels. Similarly, C1 and C2 are chosen to gain the two-channel subset, and so on. After that, all the selected data are forwarded to three different classifiers for classification separately. The performance of the proposed channel selection method is demonstrated in the experimental results section.

3.3 The Mathematic Combinations Channel Selection

In mathematics, a combination is a way of selecting members from a group, and the order of members does not matter. In smaller cases, it is possible to count the number of combinations. More formally, a *k*-combination of a set ***S*** is a subset of *k* distinct elements of ***S***. If the set has *n* elements, the number of *k*-combination is equal to the binomial coefficient.

$$C_n^k = \frac{n(n-1)\dots(n-k+1)}{k(k-1)\dots 1} \qquad (5)$$

which can be written using factorials as $\frac{n!}{k!(n-k)!}$ if $k \le n$, and is zero when $k > n$. The set of all *k*-combination of a set ***S*** is sometimes denoted by C_n^k.

Here, mathematic combinations can also be used to select channels from the original 64 electrodes. The proposed method ignores the importance of the individual channels and treats them equally. The main idea of this method is to introduce a simple computer-assisted-mathematic-method for medical signals analysis. It seems inefficient to do random mathematic combinations. However, it can easily find out the optimal subsets in a dataset by computers through C_n^k runs. In this paper, the average classification accuracy of the ten-time trials with specific numbers of channels chosen by mathematic combinations is demonstrated in the experimental results section for comparison.

During the classification process in this paper, the extracted data from the previous data selection stage are classified by three different classifiers, namely: the J48 decision tree, the K-nearest neighbor (KNN) and the Kstar. The details of the classifiers are introduced in this section.

- *J48 Decision Tree*

The J48 decision tree (Weka implementation of C4.5) was published by Ross Quinlan in 1993 [34]. It is a classic method to represent information from a machine learning algorithm and offers a fast and powerful means to express structures in data [35]. In this paper, the J48 algorithm provided by Weka is used. Weka is an open-source Java application produced by the University of Waikato in New Zealand. This software offers an interface through which many algorithms can be utilized on pre-formatted datasets. Using this interface, several test domains are experimented to gain an insight into the effectiveness of the above three different data selection methods.

- *K-nearest neighbor (KNN)*

The KNN algorithm is also selected to conduct the binary classification. The KNN algorithm is a statistical supervised classification which is widely used in traditional pattern recognition techniques [36]. The idea is that given a set of data t, the algorithm obtains the K nearest neighbors from the training set based on the distance between t and the training set. The most dominating class amongst these K neighbors is assigned as class t. In this study, the KNN algorithm is implemented as IBK package in Weka 3.7.11.

- *Kstar*

The Kstar algorithm is used to evaluate the efficiency of the proposed data selection methods. It can be defined as a method of clustering analysis which aims at partitioning n observations into k clusters in which each observation belongs to a cluster with the nearest mean. The algorithm provides a consistent approach to handle real valued attributes, symbolic attributes and missing values. It uses $entropy$ as a distance measure. In this study, the Kstar algorithm is also implemented in Weka 3.7.11.

4. EXPERIMENTAL RESULTS

- *Experimental Environment*

GE is extracted by R x64 3.1.0 and the implementation of the PCA is done by Matlab2013b. The classification is performed using the J48 decision tree, the KNN and the Kstar in Weka 3.7.10. All experiments are performed on a 3.40GHz Intel(R) Core(TM) i7-3770 CPU processor PC, with 8.00G RAM and 64-bit Operation System. The operation system of the PC is Microsoft Windows 7.

- *Data Set Selection*

The experimental EEG datasets consist of two classes (denoted as alcoholic (a) and control (c)). There are 600 samples in SMNI_CMI_TRAIN.tar.gz and 600 samples in SMNI_CMI_TEST.tar.gz from 64 different channels, respectively. In this paper, GE is used to extract features based on HVGs and then the PCA, GE differential based selection or mathematic combinations selection are implemented in choosing the subset of the EEG signals. Each channel data in one second from one sample is mapped to a HVG, and each HVG is extracted as one GE value. Therefore, 76,800 GE features are extracted from the 1200 samples, with each sample having 64 channels. That is to say, both the training data and the testing data are transferred into a [600*64] matrix.

Then different subsets of both the training data and the testing data used during the experiments are determined as: (1). Set 1 (1/64 of data), (2). Set 2 (2/64 of data), (3). Set 3 (19/64 of data), and (4). Set 4 (64/64 of data). The reason why adopt the above mentioned sets is illustrated as follows: The classification results based on different percentages of the whole data, which is 1/64, 2/64, 19/64, 26/64, 32/64, 39/64, 46/64, 53/64, 64/64, using the J48 decision tree, the KNN and the Kstar are displayed in Fig. 7. The classification accuracy increases dramatically when increasing the amount of data used between 1/64 and 19/64. But the accuracy decreases slightly after that and rises again after using 53/64 of the whole data. For online EEG analysis and classification system design, both classification accuracy and the computation time are critical. The redundancy and noise often cause the decrease of the classification efficiency. Therefore, it is significant to select the informative data and eliminate the redundant and misleading data to reduce the computation time.

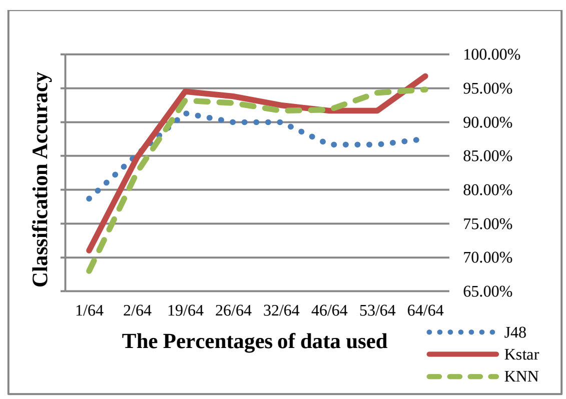

Fig. 7 Classification accuracies based on the different percentages of data used.

Data selection is expected to preserve as much information as those in the whole database. This paper proposes three data selection methods: (1). the PCA-GE, (2). the GE difference based channel selection, and (3). the mathematic combinations channel selection. In PCA-GE, the four groups of experiments with their power percentages are shown in Table 1. The distributions of the data sets and the PCA selected data are summarized in Table 2.

Table 1
The corresponding power percentages of different numbers of principal components from original data.

Set ID	No. of Principal Components	Power Percentage
Set 1	1	0.479
Set 2	2	0.637
Set 3	19	0.935
Set 4	64	1

Table 2
The distribution of sample sets and the PCA extracted features.

Set ID	Training Set	Testing Set	Total
Set 1	[600 x 1]	[600 x 1]	[1200 x 1]
Set 2	[600 x 2]	[600 x 2]	[1200 x 2]
Set 3	[600 x 19]	[600 x 19]	[1200 x 19]
Set 4	[600 x 64]	[600 x 64]	[1200 x 64]

- *Performance Comparisons*

The performances of the PCA-GE method with the experimental EEG datasets using the three different classifiers are evaluated with the aspect of the classification accuracy as shown in Table 3 and the computation time in Table 4. From Tables 3 and Table 4, it is apparent that using 19 out of 64 original data can achieve as high as 94.5% accuracy by costing only 29.52% of the computation time, compared to the 96.8% accuracy by using the whole data through the Kstar classifier. Besides, it is interesting to see the improvement of the accuracy from 87.5% to 91.3% by using 19/64 data through the J48 decision tree classifier for the PCA-GE data selection method. It is probably due to the filtering of the noise, so that the remaining data are more representative but with much smaller amount. To evaluate the wide applicability of the selected data, three different classifiers are adopted and the one having the highest classification accuracy among the three classifiers is denoted as Bold in the following tables (e.g., Tables 3, 5 and 6).

Table 3
The classification accuracy of the proposed PCA-GE method.

Group \ Classifier	Kstar	KNN	J48
1 component	71.0%	68.0%	**78.7%**
2 components	84.8%	82.5%	**85.2%**
19 components	**94.5%**	93.2%	91.3%
64 components	**96.8%**	94.8%	87.5%

Table 4
The computation time of the proposed PCA-GE method.

Group \ Classifier	Kstar	KNN	J48
1 component	0.63s	0.01s	0.01s
2 components	1.28s	0.01s	0.02s
19 components	11.51s	0.06s	0.03s
64 components	38.99s	0.10s	0.04s

Apparently, less data means less computation time. The computation times of all the three proposed methods are reduced significantly when the number of data used decreases as shown in Table 4. In the meantime, the performances of the selected channels subsets based on the mean GE gap values are presented by Table 5 in terms of the classification accuracy. The one-channel signal is from electrode C1. The two-channel data are from electrodes C1 and C2. The 19 channels data are from electrodes C1, C2, PO8, PO7, C3, FC2, FCZ, CP2, CPZ, PZ, FZ, CP5, F1, P2, C4, FC1, F2, P4, CP1; and the 64 channels signals are all the recorded signals from the HVG GEs, respectively. According to our experiment, the proposed GE difference based channel selection method achieves as high as 91.67% classification accuracy by using only 19 out of 64 channels of data for the Kstar classifier. Therefore, it can significantly enhance the efficiency of the EEG data collection. Instead of using all the 64 electrodes placed on the scalp of the subjects, 19 electrodes are enough to gain satisfactory classification results.

Table 5
The classification accuracy of the GE difference based channel selection.

Group \ Classifier	Kstar	KNN	J48
1 channel	68.17%	57.67%	**68.83%**
2 channels	**68.5%**	65.5%	64.5%
19 channels	**91.67%**	90.17%	88.33%
64 channels	**96.8%**	94.83%	87.5%

The performances of the selected channels subsets from mathematic combinations are presented by Table 6 in terms of the classification accuracy. Compared to the data selection based on PCA-GE or GE difference, this method neglects the possible different impacts of the individual channels. The method yields an 83.83% classification accuracy when the channel number is 19 through the KNN classifier.

Table 6
The classification accuracy of the mathematic combinations based channel selection.

Group \ Classifier	Kstar	KNN	J48
1 channel	**58%**	54.18%	**58%**
2 channels	58.33%	59.5%	**61%**
19 channels	81.33%	**83.83%**	78.5%
64 channels	**96.8%**	94.83%	87.5%

In summary, all the proposed methods have been proved to yield an acceptable classification accuracy using significantly reduced amount of data. The results validate the efficiency of the proposed methods in the EEG data reduction. Using as less as possible data to gain high classification performances could significantly reduce the processing time as well as the data collection hardware requirements.

5. DISCUSSION

According to experimental results, the proposed PCA-GE algorithm can achieve the comparable accuracy 94.5% by costing only 29.52% of the computation time and using 19 out of 64 original data, compared to the 96.8% accuracy by using the whole 64 channels of the data through the Kstar classifier. Similarly, the proposed GE difference based channel selection method also gets 91.67% classification accuracy by using only 19 out of 64 channels of data for the Kstar classifier. They are of high efficiency in terms of both the classification accuracy and the computation time. It is demonstrated that the proposed methods can gain relatively high classification accuracies with a significantly reduced running time during the EEG analysis and classification process. Data selection opens the possibility of using much less representative data to gain satisfactory analysis and classification results

6. CONCLUSSION

For multi-channel real EEG signals, using optimal data subsets instead of all the original data and achieving relatively satisfactory classification accuracies with much less computation time are important for EEG analysis and classification. How to get the optimal subsets from the original data is crucial to the following classification performance. In this paper, firstly the GE features from HVGs of the EEG data from alcoholics are calculated. Based on the GE features, the proposed data selection methods are the PCA-GE, the GE difference based channel selection and the mathematic combinations channel selection. It is apparent that less running time is needed by the analysis and classification system if less data are used. Instead of using original data, we extracted features using GE based on HVG. The PCA is successfully used in data selection. Meantime, channel selections based on GE difference and mathematic combinations are proposed for the purpose of comparisons. All of them can gain high classification accuracy as well as decrease the computation time, which is important for the design of the online EEG signals analysis and classification system. Data selection using PCA-GE algorithm was found to be more efficient and beneficial.

REFERENCES

[1] Haas LF (2003) Hans Berger (1873–1941), Richard Caton (1842–1926), and electroencephalography. Journal of Neurology, Neurosurgery & Psychiatry 74 (1):9-9

[2] Lehnertz K, Elger CE (1998) Can Epileptic Seizures be Predicted? Evidence from Nonlinear Time Series Analysis of Brain Electrical Activity. Physical Review Letters 80 (22):5019-5022

[3] Martinerie J, Adam C, Quyen MLV, Baulac M, Clemenceau S, Renault B, Varela FJ (1998) Epileptic seizures can be anticipated by non-linear analysis. Nat Med 4 (10):1173-1176

[4] Siuly S, Kabir E, Wang H, Zhang Y (2015) Exploring Sampling in the Detection of Multicategory EEG Signals. Computational and Mathematical Methods in Medicine 2015:576437. doi:10.1155/2015/576437

[5] Siuly, Li Y, Wen P (2011) EEG signal classification based on simple random sampling technique with least square support vector machine. International Journal of Biomedical Engineering and Technology 7 (4):390-409. doi:10.1504/IJBET.2011.044417

[6] Zhu G, Li Y, Wen P (2011) Evaluating functional connectivity in alcoholics based on maximal weight matching. Journal of Advanced Computational Intelligence and Intelligent Informatics 15 (9):1221-1227

[7] Wackermann J (1995) Beyond mapping: estimating complexity of multichannel EEG recordings. Acta neurobiologiae experimentalis 56 (1):197-208

[8] Zhu G, Li Y, Wen PP (2012) An efficient visibility graph similarity algorithm and its application on sleep stages classification. In: Brain Informatics. Springer, pp 185-195

[9] Stam C, Lelj EHvd, Keunen R, Tavy D (1999) Nonlinear EEG changes in postanoxic encephalopathy. Theory in Biosciences-Theorie in den Biowissenschaften 118 (3-4):209-218

[10] Stam CJ, Van Woerkom T, Keunen R (1997) Non-linear analysis of the electroencephalogram in Creutzfeldt-Jakob disease. Biological cybernetics 77 (4):247-256

[11] Nguyen-Ky T, Wen P, Li Y, Malan M (2012) Measuring the hypnotic depth of anaesthesia based on the EEG signal using combined wavelet transform, eigenvector and normalisation techniques. Computers in biology and medicine 42 (6):680-691

[12] Li T, Wen P, Jayamaha S (2014) Anaesthetic EEG signal denoise using improved nonlocal mean methods. Australasian Physical & Engineering Sciences in Medicine 37 (2):431-437

[13] Misulis KE, Spehlmann R (1994) Spehlmann's evoked potential primer: visual, auditory, and somatosensory evoked potentials in clinical diagnosis. Butterworth-Heinemann Medical,

[14] Oscar-Berman M, Marinković K (2007) Alcohol: effects on neurobehavioral functions and the brain. Neuropsychology review 17 (3):239-257

[15] Richman JS, Moorman JR (2000) Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology 278 (6):H2039-H2049

[16] Di W, Zhihua C, Ruifang F, Guangyu L, Tian L Notice of Retraction Study on human brain after consuming alcohol based on EEG signal. In: Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on, 2010. IEEE, pp 406-409

[17] Sun Y, Ye N, Xu X EEG analysis of alcoholics and controls based on feature extraction. In: Signal Processing, 2006 8th International Conference on, 2006. IEEE,

[18] Subasi A (2007) EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Systems with Applications 32 (4):1084-1093

[19] Güler I, Übeyli ED (2005) Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. Journal of neuroscience methods 148 (2):113-121

[20] Tsuji T, Bu N, Fukuda O, Kaneko M (2003) A recurrent log-linearized Gaussian mixture network. Neural Networks, IEEE Transactions on 14 (2):304-316

[21] Vasicek O (1976) A test for normality based on sample entropy. Journal of the Royal Statistical Society Series B (Methodological):54-59

[22] Polat K, Güneş S (2007) Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. Applied Mathematics and Computation 187 (2):1017-1026

[23] Chandaka S, Chatterjee A, Munshi S (2009) Cross-correlation aided support vector machine classifier for classification of EEG signals. Expert Systems with Applications 36 (2):1329-1336

[24] Zhu G, Li Y, Wen PP, Wang S (2014) Analysis of alcoholic EEG signals based on horizontal visibility graph entropy. Brain Informatics:1-7

[25] Tomida, Naoki, et al. "Active Data Selection for Motor Imagery EEG Classification." Biomedical Engineering, IEEE Transactions on 62.2 (2015): 458-467.

[26] Bache K, Lichman M (2013) UCI machine learning repository. URL http://archive. ics. uci. edu/ml, vol 901.

[27] Gutin G, Mansour T, Severini S (2011) A characterization of horizontal visibility graphs and combinatorics on words. Physica A: Statistical Mechanics and its Applications 390 (12):2421-2428

[28] Luque B, Lacasa L, Ballesteros F, Luque J (2009) Horizontal visibility graphs: Exact results for random time series. Physical Review E 80 (4):046103

[29] Diestel R (2005) Graph Theory (3rd ed'n).

[30] Körner J Coding of an information source having ambiguous alphabet and the entropy of graphs. In: 6th Prague conference on information theory, 1973. pp 411-425

[31] Shannon CE (2001) A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review 5 (1):3-55

[32] Person K (1901) On lines and planes of closest fit to systems of points in space. philosophical magazine 2 (6):559-572

[33] Hotelling H (1933) Analysis of a complex of statistical variables into principal components. Journal of educational psychology 24 (6):417

[34] Salzberg SL (1994) C4. 5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993. Machine Learning 16 (3):235-240

[35] Sehgal L, Mohan N, Sandhu PS Quality prediction of function based software using decision tree approach. In: International Conference on Computer Engineering and Multimedia Technologies (ICCEMT), 2012. pp 43-47

[36] Duda RO, Hart PE (1973) Pattern classification and scene analysis. vol 3. Wiley New York,