Analyzing EEG Signals Using Graph Entropy Based Principle Component Analysis and J48 Decision Tree

Shuaifang Wang, Yan Li, Pen Wen, and Guohun Zhu
University of Southern Queensland, Toowoomba, Australia
E-mail: {U1049801, yan.li, peng.wen, U1017412}@umail.usq.edu.au

***Abstract*—This paper proposed a method using principle component analysis based on graph entropy (PCA-GE) and J48 decision tree on electroencephalogram (EEG) signals to predict whether a person is alcoholic or not. Analysis is performed in two stages: feature extraction and classification. The principle component analysis (PCA) chooses the optimal subset of channels based on graph entropy technique and the selected subset is classified by the J48 decision tree in Weka. K-nearest neighbor (KNN) and support vector machine (SVM) in R package are also used for comparison. Experimental results show that the proposed PCA-GE method is successful in selecting a subset of channels, which contributes to the high accuracy and efficiency in the classification of alcoholics and non-alcoholics.**

***Index Terms*—EEG, graph entropy, Horizontal Visibility Graph (HVG), Support Vector Machine (SVM), Principle Component Analysis (PCA), J48 decision tree**

I. Introduction

Discovered by Hans Berger [1] in 1924, EEG is 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], characterization of sleep phenomena [7], [8], encephalopathy [9], Creutzfeldt-Jakob disease [10], monitoring the depth of anesthesia [11], [12] and the depression diagnosis. Instead of making pictures of the brain's anatomy like computed tomography or magnetic resonance imaging, EEG evaluates the brain's physiology with a millisecond-range temporal resolution in a convenient and relatively inexpensive way, which makes it continue to play a central role in the diagnosis and management of patients with brain disorders working in conjunction with the now remarkable variety of other diagnostic techniques developed over the last 30 or so years.

People who drink alcohol excessively suffer from blurred vision, difficult 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, identify and mobility impairments but also damage the brain systems [14]. Clinical evidence of using advanced signal processing methods has proven that detecting alcoholism from the EEG can be effective [15]-[17]. Therefore, an increasing number of researchers are studying the connection between EEGs and alcoholics.

Automatic EEG classification systems are the trend in both research and clinical areas because the traditional visual inspection of EEG signals requires highly trained medical professionals and it is time consuming, error prone and not sufficient enough for reliable detection and prediction. In automatic classification, the amount of data needed increases exponentially 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. Therefore, how to reduce the number of dimensionalities needed for classification but preserve the critical information to classify the subjects accurately is the major problem in EEG signal research. The innovation of classifiers also contributes to the improvement of classification accuracy. To solve this problem, some countermeasures in both feature extraction and classification aspects have been proposed so far. Adbulhamit Subasi [18] decomposed EEG signals into frequency sub-bands using Discrete Wavelet Transform and classified normal and epileptic EEGs with a mixture of expert mode. Güler, İnan 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 T, Bu N, Fukuda O, Kaneko M [20] employed a Gaussian mixture model to conduct EEG pattern classification. Oldrich Vasicek [21] had tested 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. Chandaka, Suryannarayana, Amitava Chatterjee and Sugata Munshi [23] introduced a most promising pattern recognition technique called cross-correlation aided SVM

Manuscript received August 1, 2014; revised November 24, 2014.

©2016 Int. J. Sig. Process. Syst.

doi: 10.12720/ijsps.4.1.67-72

based classifier and achieved classification accuracy on normal and epileptic EEGs as high as 95.96%.

This study uses the principle component analysis based on the graph entropy (PCA-GE) to extract features and discriminate EEG signals. It transfers the original data into horizontal visibility graphs (HVGs) and gains the graph entropy of each sample. After that, the PCA is used to select the subset of channels to do classifications by J48 decision tree.

The paper is organized as follows: In Section II, the experimental dataset is briefly introduced. The proposed PCA-GE algorithm from HVGs is described in Section III. In Section IV, the classification results of the proposed method with graph entropy based on HVGs, KNN and SVM are presented for comparison. Finally, the conclusion is drawn in Section V.

II. Experimental Data

The alcoholic EEG dataset (SMNI_CMI_TRAIN.tar.gz & SMNI_CMI_TEST.tar.gz) used in this paper was published by Henri Begleiter from State University of New York Health Center [24]. The data was digitized at 256 Hz for one second from 64 electrodes placed on subjects' scalps. The whole dataset consists of 10 alcoholic and 10 control subjects (denoted as alcoholic (a) and control (c)) with ten runs per subject completed with 120 trials. In this paper, the data from SMNI_CMI_TRAIN.tar.gz are used as the training samples and that from SMNI_CMI_TEST.tar.gz are used as the testing data. That is to say, the alcoholic EEG signal dataset we use is a 600*64 matrix for training and another 600*64 matrix for testing.

III. Methodology

The proposed algorithm extracts features by the PCA technique based on the graph entropy (GE) from HVGs. Then all the selected features are classified by J48 decision tree. The diagram of the proposed method is shown in Fig. 1.

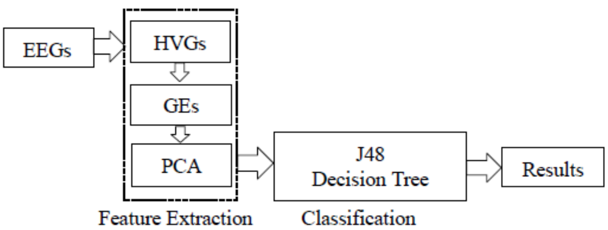

Figure 1. The diagram of the proposed algorithm.

Feature extraction aims at using subsets of optimal channels instead of original vectors while remaining as much useful information as possible. The implementation details are described below.

A. Feature Extraction

1) Horizontal visibility graph

A horizontal visibility graph (HVG) is defined in association with an ordered set of nonnegative reals [25]. HVGs realize a methodology in the analysis of time series and their degree distribution is a good discriminator between randomness and chaos. A graph is an ordered pair G = (V, E), where V is a set of elements called nodes and E is a set of unordered pairs of nodes called edges. Let X = ($x_i \in$ R≥0: i = 1, 2, . . . , n) be an ordered set (or, equivalently, a sequence) of non-negative real numbers. The horizontal visibility graph (HVG) of X is the graph G = (V, E), with V = X and an edge. The definition of it is shown in (1) as follows:

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

where every $k \in (i, j)$. A HVG is a subset of complex networks [26]. The node degree and sequence are used to describe the characteristics of the graph. In graph theory, the degree of a node of a graph is the number of edges connecting to the node, with loops counted twice [27]. The degree of a vertex is denoted as deg (x_i). The degree sequence (DS) is the sequence of the degree of a graph. In this paper, the time series of alcoholic EEGs are mapped into a graph G (V, E). To make it clear, let's take the data set co2a0000368 from electrode FP1 in the forementioned database for 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 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 to non-negative real by adding the absolute value of the smallest value which is negative. e.g.,

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}

Which 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 as Fig. 2.

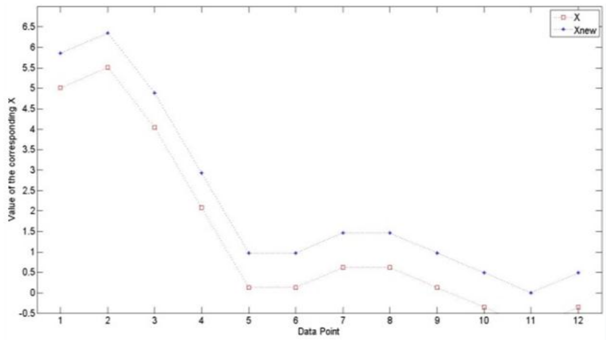

Figure 2. Nonnegative transform of X.

b. Horizontal visibility check is used to calculate the degree of each node, which is showed as Fig. 3.

c. Get the degree sequence.

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

2) Graph entropy

The graph entropy is the entropy of the frequency distribution of connections of the nodes in an undirected, unweighted HVG. It is a function from information theory on a graph G with a probability distribution P on its node set. It was introduced by Janos Korner in [28]. Shannons entropy [29] is used in this paper, which is shown by (2) as follows:

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

where $p(k)$ 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 (3) as follows:

$$p(k) = n_k / n \tag{3}$$

In the above case, $p(k)$ of DS is (0, 1/12, 8/12, 3/12). The graph entropy is 0.824 when it takes the logarithm base two. The plot of 64 channel EEG signals is shown in Fig. 4 [30]. From Fig. 4, it is clear that the differences between the alcoholics and control subjects are indeed different from channels to channels. That's the reason why optimal subset of channel selection is required. In this paper, the principle component analysis is adopted.

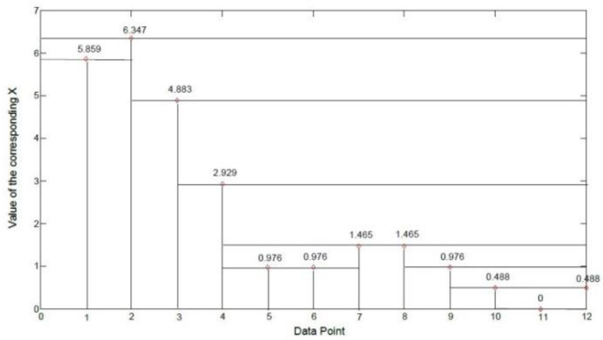

Figure 3. The degree of each node from HVGs.

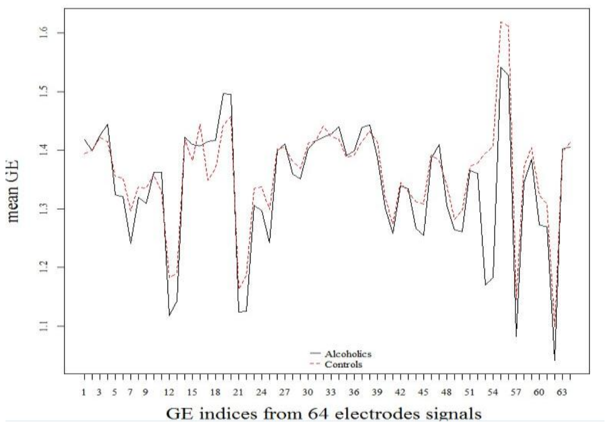

Figure 4. Mean GE from 64 electrode signals.

3) *Principle Component Analysis (PCA)*

We briefly recap the concept of PCA to introduce our usage of it. PCA was invented by Karl Pearson [31] in 1901 as an analogue of the principal axes theorem in mechanics and later independently developed (and named) by Harold Hotelling in the 1930s [32]. The method is mostly used as a tool in exploratory data analysis and for making predictive models. The faithful transformation T = XW maps a data vector X (i) from an original space to a new space of p variables which are uncorrelated over the dataset. However, not all the principal components need to be kept. Keeping only the first L principal components, it gives the truncated transformation (4) as follows:

$$T_L = XW_L \tag{4}$$

where matrix T_L now has n rows but only L columns. By reconstruction, all the transformed data matrices reserve only columns out of the original data. Such dimensionality reduction can be a very useful step for visualizing and processing high-dimensional data by keeping as much of the useful information as possible. The PCA is implemented in Matlab2013b.

TABLE I. THE CORRESPONDENCE POWER PERCENTAGES OF THE NUMBER OF ELECTRODES FROM ORIGINAL SAMPLES.

Eletrodes No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Power %	0.479	0.637	0.743	0.785	0.817	0.841	0.858	0.873
Eletrodes No.	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Power %	0.884	0.892	0.899	0.906	0.912	0.917	0.922	0.926
Eletrodes No.	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
Power %	0.929	0.932	0.935	0.938	0.940	0.943	0.945	0.948
Eletrodes No.	**25**	**25**	**27**	**28**	**29**	**30**	**31**	**32**
Power %	0.950	0.952	0.954	0.956	0.958	0.960	0.962	0.963
Eletrodes No.	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**
Power %	0.965	0.967	0.968	0.970	0.971	0.973	0.974	0.976
Eletrodes No.	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**
Power %	0.977	0.978	0.980	0.981	0.982	0.983	0.984	0.986
Eletrodes No.	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**
Power %	0.988	0.988	0.989	0.989	0.991	0.992	0.993	0.994
Eletrodes No.	**57**	**58**	**59**	**60**	**61**	**62**	**63**	**64**
Power %	0.995	0.995	0.996	0.997	0.998	0.999	0.999	1

The experimental EEG dataset consists of two classes (denoted as alcoholic (a) and control (c)). Every class contains 600 samples with 64 channels data each, which is [600*64] for each class). In this paper, graph entropy is used to extract features based on HVGs and then the PCA is implemented in choosing the subset of channels for alcoholic EEG classification. The selected channels are to preserve as much information as compared to the full set of 64 channels as possible. The distribution of the percentage of total power is shown in Table I as follows.

TABLE II. THE CORRESPONDENCE POWER PERCENTAGES OF THE PRINCIPLE COMPONENTS FROM ORIGINAL DATA

Set ID	Channel No.	Power Percentage
Set 1	1	0.479
Set 2	2	0.637
Set 3	19	0.935
Set 4	64	1

Using percentage of total power retained, the number of principle components to restrict the data is determined. This is more suitable for automated procedures since it does not involve manual inspection. To compare with the current literature classification results, we design the following four groups of experiments with particular power percentages which are shown in Table II:

B. *Classification*

To evaluate the performances of the proposed method, different classifiers are used for comparison.

1) J48 decision tree

The J48 decision tree (Weka implementation of C4.5) was published by Ross Quinlan in 1993 [33]. 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. The J48 decision tree also gives a variety of options available, which can make a significant difference in the quality of classification results. In this paper, the default settings provided by Weka are used because they are proven to be adequate in many cases. 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 data sets.

Using this interface, several test-domains were experimented to gain an insight into the effectiveness of different methods.

2) K-Nearest Neighbor (KNN)

The KNN algorithm is selected to conduct the binary classification. The KNN algorithm is a statistical supervised classification which is widely used in traditional pattern recognition [34]. The idea is that given a new test data t, the algorithm obtains the K nearest neighbor from the training set based on the distance between t and the training set. The most dominated class amongst these K neighbors is assigned as the class of t. In this study, the KNN algorithm is implemented in R package FNN [35], where K is assigned as 3.

3) Support Vector Machine (SVM)

The SVM is also selected to conduct the binary classification because it has been successfully used to classify the HVG features associated with sleep EEG signals [36]. It can perform both the linear space discrimination and nonlinear classification by using different kernel functions which can be linear, polynomial kernel, radical basis function (RBF) and sigmoid. In this paper, the SVM algorithm with RBF kernel is implemented in R package e1071 [37].

TABLE III. 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]

For the classification part, there are 600 samples making up of 64 channels data in SMNI_CMI_TRAIN.tar.gz and 600 samples making up of 64 channels data in SMNI_CMI_TEST.tar.gz, respectively. Therefore, both the training data and testing data are 600*64 matrixes. When it comes to the PCA, each original sample with 64 channels data is transferred to samples with single channel, two channels, 19 channels and 64 channels, respectively. The distributions of the sample sets and the PCA extracted features are summarized in Table III. The classifier used is J48 decision tree.

IV. Experimental Results

To evaluate the performance of the proposed PCA-GE algorithm based on HVGs presented in Section III, the classification results based on different numbers of channels using J48 decision tree are displayed in Fig. 5. It is clear that the rate of accuracy increase is zero when the number of channels is larger than 21. That is to say, the increase of data to more than 21 channels does nothing better to the classification accuracy. On the other hand, it decreases the efficiency of the classification. For big data processing, it is meaningful to select the informative data and eliminate the redundant and misleading data to improve the analysis and classification performance.

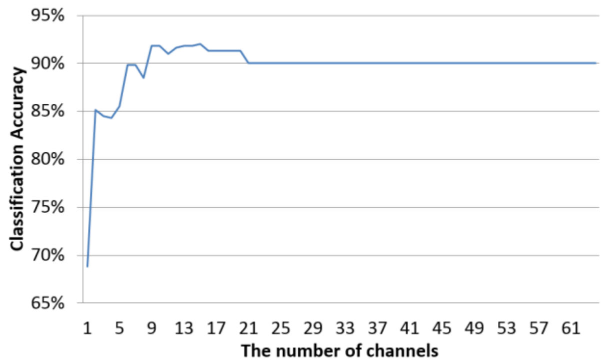

Figure 5. Classification accuracy based on different number of channels.

Besides that, different classifiers (KNN and SVM) are also used to test the classification accuracy of extracted data with and without using the proposed method as follows: (1) Set 1 (single channel) (2) Set 2 (two channels) (3) Set 3 (19 channels) (4) Set 4 (64 channels).

Graph entropy is extracted by R and the implementation of the PCA is done by matlab2013b. Classification is performed using J48 decision tree in Weka version 3.7.10, KNN and SVM in R *64 3.1.0. All the experiments are run on a 3.00GHz Intel(R) Core(TM) 2 Duo CPU processor PC with 4.00G RAM. The operation system is Microsoft Windows 7.

A. *Performance Comparison*

TABLE IV. The Comparison between the Proposed PCA Involved Method and Graph Entropy Based on HVGs Mehtod

Method / Group	KNN Accuracy	SVM Accuracy	J48 Accuracy
1 channel	77.0%	**79.3%**	78.7%
2 channels	81.6%	82.5%	**85.2%**
19 channels	89.2%	89.8%	**91.3%**
64 channels	**96.8%**	91.3%	87.5%

In this section, the performance comparison between graph entropy based on HVGs only method and the proposed PCA involved method with the experimental EEG database is presented by Table IV in terms of classification accuracy.

From Table IV, the feature extraction algorithm is efficient, which contributes to the improvement of the accuracy using subsets of the channels because it is clear that the classification accuracy on the full database of 64 channels is only 87.5%. It implies the proposed method can be used in channel selection. Using particular representative principle components to represent all channels' data to do classification is an EEG signal research field. Further study is needed because the high dimensional EEG data is of large size and the analysis is time-consuming. From the point of mathematics, there are C_n^m possibilities when select m electrode from n electrodes. How to select electrodes is still challenging.

Whether there is a standard selecting rule which different EEG databases can follow is still a research question. It is important for online EEG signal analysis and classification, which needs further study.

V. Conclusion

For multi-channel real EEG signals, a few principle components might be sufficient to represent all the signals and the correct number of principle components is crucial to the following analysis and classification. The proposed algorithm in this study uses a proper number of principle components to represent the full dataset successfully transforming the high dimensional original data into a smaller size of representations by using the PCA technique based on graph entropy from HVGs.

References

[1] L. F. Haas, "Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 74, no. 1, p. 9, 2003.

[2] K. Lehnertz and C. E. Elger, "Can epileptic seizures be predicted? Evidence from nonlinear time series analyses of brain electrical activity," *Phys. Rev. Lett*., vol. 80, pp. 5019-5023, June 1998.

[3] J. Martinerie, C. Adam, M. Le van Quyen, M. Baulac, B. Renault, and F. J. Varela, "Epileptic seizures can be anticipated by non-linear analysis," *Nature Medicine*, vol. 4, pp. 1173-1176, Oct. 1998.

[4] Siuly, Y. Li, and P. Wen, "Clustering technique-based least square support vector machine for EEG signal classification," *Computer Methods and Programs in Biomedicine*, vol. 104, no. 3, pp. 358-372.

[5] Siuly, Y. Li, and P. Wen, 'EEG signal classification based on simple random sampling technique with least square support vector machines," *International Journal of Biomedical Engineering and Technology*, vol. 7, no. 4, pp. 390-409.

[6] G. H. Zhu, Y. Li, and P. Wen, "Evaluating functional connectivity in alcoholics based on maximal weight matching," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 15, no. 9, pp. 1221-1227, 2011.

[7] J. Wackermann, "Beyond mapping: Estimating complexity of multi-channel EEG recordings," *Acta Neurobiol. Exp.*, vol. 56, pp. 197-208, 1996.

[8] G. H. Zhu, Y. Li, and P. P. Wen, "An efficient visibility graph similarity algorithm and its application for sleep stages classification," in *Proc. 2012 International Conference on Brain Informatics*, Dec 4-7, 2012, Macao.

[9] C. J. Stam, E. M. H. V. D. Leij, R. W. M. Keunen, D. L. J. Tavy, "Nonlinear EEG changes in postanoxic encephalopathy," *Theor. Biosci*, vol. 118, pp. 209-218, 1999.

[10] C. J. Stam, T. C. A. M. V. Woerkom, and R. W. M. Keunen, "Nonlinear analysis of the electroencephalogram in Creutzfeldt-Jakob disease," *Biol. Cybern*., vol. 77, pp. 247-256, Oct. 1997.

[11] I. A. Rezek and S. J. Roberts, "Stochastic complexity measures for physiological signal analysis," *IEEE Trans. on Biomedical Engineering*, vol. 45, no. 9, pp. 1186-1191, Sep. 1998.

[12] T. N. Li, P. Wen, and J. Sophie, "Anaesthetic EEG signal denoise using improved nonlocal mean methods," *Australasian Physical and Engineering Science in Medicine*, vol. 37, no. 2, pp. 431-437, 2014.

[13] K. E. Misulis, *Spehlmann's Evoked Potential Primer: Visual, Auditory, and Somatosensory Evoked Potentials in Clinical Diagnosis*, Boston: Butterworth-Heinemann, 1994.

[14] M. Oscar-Berman and K. Marinkovi, "Alcohol: Effects on neurobehavioral functions and the brain," *Neuropsychology Review*, vol. 17, pp. 239-257,2007.

[15] J. S. Richman and J. R. Moorman. "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, 2000.

[16] D. Wu, Z. H. Chen, R. F. Feng, G. Y. Li, and T. Luan, "Study on human brain after consuming alcohol based on EEG signal," in *Proc. 2010 3rd IEEE International Conference on Computer Science and Information Technology*, vol. 5, 2010.

[17] Y. G. Sun, N. Ye, and X. H. Xu, "EEG analysis of alcoholics and controls based on feature extraction," *Signal Processing*, in *Proc. 8th International Conference*, vol. 1, 2006.

[18] S. Abdulhamit, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084-1093, May 2007.

[19] G. İnan and E. D. Übeyli, "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients," *Journal of Neuroscience Methods*, vol. 148, no. 2, pp. 113-121, Oct. 2005.

[20] T. Toshio, N. Bu, O. Fukuda, and M. Kaneko, "A recurrent log-linearized gaussian mixture network," *Neural Networks, IEEE Trans. on Neural Networks*, vol. 14, pp. 304-316, 2003.

[21] O. Vasicek, "A test for normality based on sample entropy," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 38, no. 1, pp. 54-59, 1976.

[22] P. Kemal 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, April 2007.

[23] S. Chandaka, A. Chatterjee, and S. Munshi, "Cross-correlation aided support vector machine classifier for classification of EEG signals," *Expert Systems with Applications*, vol. 36, no. 2, part 1, pp. 1329-1336, March 2009.

[24] K. Bache and M. Lichman. (2013). UCI machine learning repository, [Online]. Available: http://archive.ics.uci.edu/ml

[25] B. Luque, L. Lacasa, F. Ballesteros, and J. Luque, "Horizontal visibility graphs: Exact results for random time series," *Phys. Rev. E 80 046103*, 2009.

[26] G. Gutina, T. Mansourb, and S. Severinic, "A characterization of horizontal visibility graphs and combinatorics on words," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 12, pp. 2421-2428.

[27] D. Reinhard, *Graph Theory*, 3rd ed., Berlin, New York: Springer-Verlag, 2005.

[28] J. Korner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," *Transactions of the 6th Prague conference on Information Theory*, etc., Academia, Prague, pp. 411-425, 1973.

[29] C. Shannon, "A mathematical theory of communication," *Bell System Tech. J*. vol. 27, pp. 379–423, 1948.

[30] G. H. Zhu, Y. Li, S. Seneweera, P. P. Wen, and S. F. Wang, "Analysis of alcoholic EEG signals based on graph entropy," *Brain Informatics: Brain Data Computing and Health Studies*.

[31] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, no. 11, pp. 559–572, 1901.

[32] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, pp. 417-441, and 498-520, 1933.
[33] L. Sehgal, N. Mohan, and P. S. Sandhu, "Quality prediction of function based software using decision tree approach," presented at the International Conference on Computer Engineering and Multimedia Technologies (ICCEMT'2012), Bangkok (Thailand), September 8-9, 2012.
[34] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, New York: Wiley, 1973.
[35] Fast Nearest Neighbor Search Algorithms and Applications. [Online]. Available: http://cran.r-project.org/web/packages/FNN/index.html
[36] G. Zhu, Y. Li, and P. P. Wen, "Analysis and classification of sleep stages based on difference visibility graphs from a single channel EEG signal," *IEEE Journal of Biomedical and Health Informatics*, vol. 00, no. 00, 2014.
[37] A. Karatzoglou, D. Meyer, and K. Hornik, "Support vector machines in R," *Journal of Statistical Software*, vol. 15, no. i09, pp. 1–28, 2006.

Shuaifang Wang received the B.S. and M.S. degrees in Computer Science from Henan University and Information Technology from the University of New South Wales, in 2011 and 2013, respectively. She is currently a Phd student of USQ in the Faculty of Health, Engineering and Sciences, University of Southern Queensland Toowoomba. Her research interests include biomedical signal analysis and data mining.

Yan Li is currently an Associate Professor of Computer Sciences and Deputy Associate Dean in Research, Faculty of Health, Engineering and Sciences, University of Southern Queensland. Her research interests are in the areas of Biomedical Engineering, Artificial Intelligent, Blind Signal Separation, Signal/Image Processing, Computer Communications and Internet Technologies, etc.

Peng Wen is currently an Associate Professor of Systems & Computer Control, Faculty of Health, Engineering and Sciences, University of Southern Queensland. His research interests are in the areas of Biomedical Engineering (EEG Research), Complex Medical Engineering, Networked System and Intelligent control.

Guohun Zhu is an Associate Professor of Guilin University of electronic technology. He is currently a Phd student of USQ in the Faculty of Health, Engineering and Sciences, University of Southern Queensland Toowoomba. His research interests include biomedical signal analysis, networking and Computational Intelligence.