CLASSIFICATION OF ELECTROOCULOGRAPH SIGNALS: COMPARING CONVENTIONAL CLASSIFIERS USING CBFS FEATURE SELECTION ALGORITHM

S. Mala
Research Scholar,
Department of Computer Science and Engineering,
Anna University::Regional Centre, Tiruchirappalli
Tiruchirappalli – 620 024.
ursmala@gmail.com

Dr. K. Latha
Assistant Professor,
Department of Computer Science and Engineering,
Anna University::Regional Centre, Tiruchirappalli
Tiruchirappalli – 620 024.
erklatha@gmail.com

***Abstract* - This work select the features in high dimensional data using CBFS Feature selection algorithm by ElectroOculoGraph (EOG) signals using eye movements of reading and writing task. EOG measures the changes in the electric potential field caused by eye movements. This work has three phases; the first phase identifies and removes noise from the signal. The second phase involves analysis of EOG signals by CBFS Feature Selection method and the third phase classifies EOG signals using various conventional classifiers.**

Keywords - ElectroOculoGraph (EOG); Eye Movements; Clearness Based Feature Selection; classifiers.

I. INTRODUCTION

Traditionally, activity recognition research has focused on gait, posture, and gesture. The recognition of activities such as reading and writing are investigated during stationary and mobile settings using different eye tracking techniques. These studies aimed to model visual behavior during specific tasks using small number of well known eye movement characteristics. All the researchers tried to investigate the relationship between the task and eye movements, but they did not recognize the activity using eye movement characters.

Emerging number of research use video-based eye trackers to study eye movements in natural environments [25]. This shows advances on how the brain processes tasks and the role of visual system in processing the task. To investigate the visual behaviour eye movements are used as a tool. Hacisalihzade transformed fixation sequences into strings and used Markov processes to model visual fixations of observers to recognize an object [13]. Salvucci assessed means for automated analysis of eye movements using Hidden-Markov models and sequence matching [33]. Ruo-Fei Du, Ren-Jie Liu, Tian-Xiang Wu and Bao-Liang Lu designed a system to analysis vigilance level combining both video and Electrooculography (EOG) features [41].

Andreas Bulling is the first to describe and apply a general-purpose architecture for Eye-based Activity Recognition to the problem of recognising everyday activities for context inference and cognitive awareness [9].

The ElectroOculography (EOG) is a technique for measuring the resting potential of the retina. The signal from this technique is called the ElectroOculoGram. EOG is the electrical signal measurable around the eyes and can be used to detect eye movements with careful signal processing. Vertical movements are detected by placing electrodes above and below the eye and horizontal movements are detected by placing the electrodes to the left and right of the eye. Fig.1 denotes the electrode placements to acquire EOG signals.

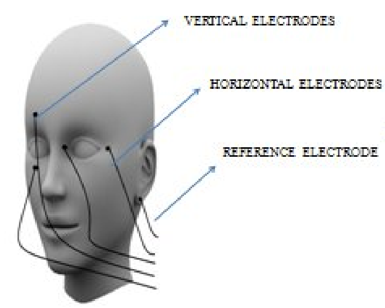

Fig. 1 A Sample ElectroOculoGraphy

EOG amplitude varies as the eyeball rotates, and thus can be used to determine horizontal and vertical eye movements. Fig. 2 denotes a sample Electrooculogram signal while rotating the eyes towards right by 30^0 and left by 15^0.

*https://www.andreas-bulling.de/datasets/recognition-of-office-activities/

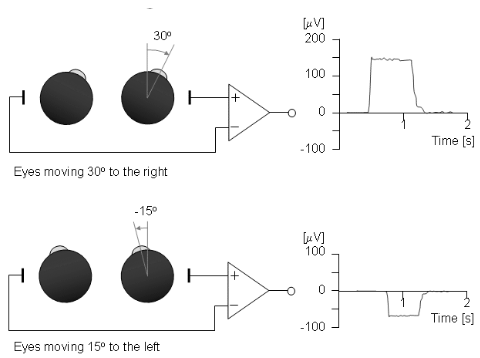

Fig. 2 A Sample ElectroOculoGram

Activity recognition has become an important topic for a broad range of real-life applications such as patient monitoring, vigilance systems, and a variety of systems that involve interactions between persons and electronic devices such as human-computer interfaces. A rich source of information for activity recognition is the movement of the eyes [1]. This work demonstrates the machine recognition of reading and writing activities using EOG signals and reduces the computational complexity which mainly occurs during the feature selection.

II. METHODOLOGY

A. *EOG Data Collection*

The first process is data collection. The data used in this study is collected from the Andreas Bulling "RECOGNITION OF OFFICE ACTIVITIES" * data set [9]. In this for each subject, ~8 hours of eye movement data using a wearable Electrooculography (EOG) system were collected. The data collection for this work involved two major office-based activities - reading a printed paper, taking handwritten notes. We used the columns of data found in mat files representing the time, voltage readings of EOG signals, during reading and writing activities.

B. *Preprocessing*

- Denoising

EOG signal characteristics are needed to be preserved by denoising method. First, the sharpness of signal needs to be retained to detect blinks and saccades. Second, EOG signal amplitudes. The EOG signal amplitudes able to distinguish between different types and directions of saccadic eye movements [5]. Finally, denoising filters must not introduce signal artefacts that may be misinterpreted as saccades or blinks in subsequent signal processing steps. The median filter performed denoising best in EOG signals and it preserved the edge sharpness of saccadic eye movements.

- *Feature Extraction*

The mission of the feature extraction and feature selection is to obtain the most relevant information from the original data and represent that information in a lower dimensionality space.

- *Basic* Eye *Movements*

Basic eye movements such as Saccades, Fixations and Blinks can be detected from the processed EOG signals. Basic eye movement's features such as Mean, Trim mean (Trimmed Mean), Median, Variance, Standard deviation, z-score, sign-rank, kurtosis, sample entropy, HFD (Higuchi's Fractional Dimension) can be detected from the processed Horizontal and Vertical EOG signals.

C. *Feature Selection*

Feature selection is important in many pattern recognition problems for excluding irrelevant and redundant features. The feature selection improves the recognition accuracy by reducing computational time and system complexity. Feature selection is a search problem for finding an optimal or suboptimal subset of m features out of original M features. Many feature subset selection algorithms have been proposed. These algorithms can generally be classified as wrapper or filter algorithms according to the criterion function used in searching for good features.

In a wrapper algorithm, the performance of the classifier is used to evaluate the feature subsets. In a filter algorithm, some feature evaluation function is used rather than optimizing the classifier's performance. Many feature evaluation functions have been used particularly functions that measure distance, information, dependency, and consistency. Wrapper methods are usually slower than filter methods but offer better performance. The simplest feature selection methods select best individual features and a feature evaluation criterion is used to rank individual features. Then the highest ranked n features are selected. Feature ranking methods can exclude irrelevant features, they often include redundant features. The minimal-redundancy-maximal-relevance (mRMR) algorithm is another sequential forward selection algorithm [7]. It uses mutual information to analyze relevance and redundancy. The mRMR scheme selects the features that correlate the strongest with a classification variable and combined with selection features that are mutually different from each other have high correlation and it is denoted by the equation 1.

$$J(X_n) = I(X_n;Y) - \frac{1}{|S|}\sum_{X_i \in S} I(X_n;X_i) \qquad (1)$$

Where $I(X_n;Y)$ -measure of dependence between feature X_n and target Y.

$J = I(X_{1:n};Y) - I(X_{1:n-1};Y)$ - difference in information with and without the feature X_n.

S- Feature Set |S| - number of features

However, mRMR grows the selected subset by adding the feature that has the maximum difference between its relevance measure and its aggregate redundancy measure with the already selected features. The mRMR scores the features based on mutual information.

In this paper, we concentrate on improving the feature extraction (in EOG signals) stage by selecting efficient subset of features. We extract 13 statistical features from a database of EOG signals for a particular subject. These features are used in reading and writing Activity recognition. We use clearness based feature selection technique to select and recommend good features for recognizing activities. We analyse the recognition accuracy as a function of the feature subset size using various conventional classifiers.

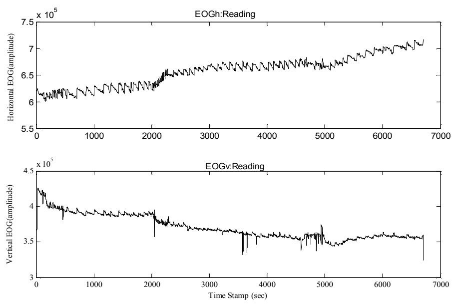

Fig. 3 EOG signals: SUBJECT10 during Reading.

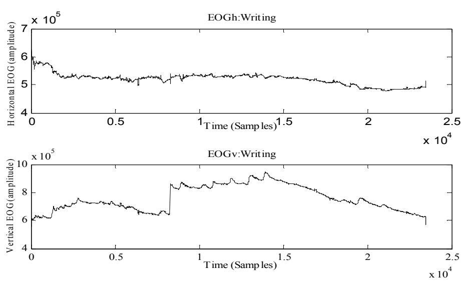

Fig. 4 EOG signals: SUBJECT10 during Writing

- *Clearness Based Feature Selection (CBFS)*

CBFS calculates the distance between the target sample and centroid of each class. It compares the class of the nearest centroid with the class of the target sample. The matching ratio becomes a clearness value for the individual feature.

Clearness based feature selection (CBFS) algorithm which can be classified as a filter method. Clearness means the separability between classes in a feature. If (clearness of feature f_2) > (clearness of feature f_1), then f_2 is more advantageous to classification than f_1.

Step 1.

The centroid for read and write is calculated by average operation. It is the median point of a class. Med (f_i, j) denotes the median point of class j in the feature f_i, which is calculated by equation 2:

$$Med(fi, j) = \frac{1}{k}\sum_{r=1}^{k}(Xri \in classj) \qquad (2)$$

Where k is a number of samples of class j.

The Table 1 list the median point of class Reading and Writing for each feature.

Table 1: Med (f_i,j) for each class

Feature	Class(read)	Class(write)
Mean	15.11973	16.24934
Trim mean	15.12032	16.24328
Inter quartile range	0.048505	0.116869
median absolute deviation	0.028197	0.069294
Standard deviation	0.035405	0.084569
Variance	0.002126	0.011101
Median	15.12132	16.25746
z-score	0.553406	0.353346
sign-rank	1.84E-10	3.63E-10
Hurst Exponent	0.899231	0.943017
HFD*	1.584791	1.510738
Sample Entropy	0.748001	5.011675
Kurtosis	0.352229	3.17936

*Higuchi's Fractional Dimension

Step 2.

For each x_{ij} in sample predicted class label is calculated. After calculating the distance between x_{ij} and Med (f_j,c_i) for all classes, we take the nearest centroid Med(f_j, s) and s is a predicted class label for x_{ij}. The distance from x_{ij} to Med(f_j, t) is calculated using:

$$D(x_{ij}, Med(f_j, t) = | x_{ij} - Med(f_j, t) | \qquad (3)$$

Step 3.

Calculate $n \times m$ matrix M_2 which contains a matching result of predicted class label and correct class label in *CS*.
$M_2(i,j)$ is calculated by:

$$M_2(i,j) = \begin{Bmatrix} 1 if M_1(i,j) = C_i \\ 0 if M_1(i,j) \neq C_i \end{Bmatrix} \qquad (4)$$

Step 4.

Calculate CScore (f_i).initially we calculated CScore (f_i) by:

$$CScore(f_i) = \frac{1}{n}\sum_{r=1}^{n} M_2(r,i) \qquad (5)$$

The range of CScore (f_i) is [0, 1]. If CScore (f_i) is close to 1, this shows that classes in feature f_i are clustered well and elements in f_i can be clearly classified. The Table 2 list all the features with its clearness score.

Table 2: Clearness Scoring for each features

Feature	**CScore(f_i)**
Mean	**0.75**
Trim mean	**0.75**
Inter quartile range	0.675
Median absolute deviation	0.625
Standard deviation	0.625
Variance	0.675
Median	**0.75**
z-score	0.625
sign-rank	0.5
Hurst Exponent	0.575
HFD	0.6
Sample Entropy	0.625
Kurtosis	0.725

The Table 3 list the score for selected features by CBFS feature selection and mRMR feature selection.

Table 3: Feature selection method CBFS with its selected features

Feature Selection Method	**Order**	**Feature Number**	**Name**	**C Score**
CBFS	1	1	Mean	0.75
	2	2	Trim mean	0.75
	3	7	Median	0.75

CScore (f_i) measures the clearness of the feature. The selected features by CBFS feature selection algorithm is shown in Table 3 with its CScore. The clearness score for Mean, Trim mean and Median has same values and these features alone can clearly classify the physical activities such as read and write in EOG signals effectively and efficiently.

III. CONVENTIONAL CLASSIFICATION METHODS

For classification without CBFS we used all 13 features and with CBFS we selected only 3 features (Mean, Trim mean, Median) with high scores as 0.75 CScore (f_i).

Decision tree is a tree structure where non-terminal nodes represent tests on one or more attributes and terminal nodes reflect decision outcomes. The strengths of decision trees are easy to understand, map nicely to set of production rules, can able to process both numerical and categorical data. We use Random Tree for tree based classifier. Fig. 5 illustrates the tree structure of the random tree classifier with all 13 features.

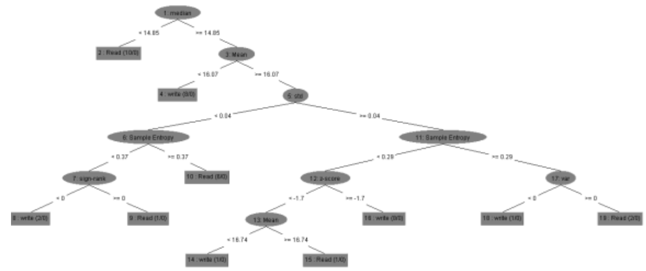

Fig. 5 Tree structure of RandomTree classifier (with all features)

Naïve Bayes classifier relates the conditional and marginal probabilities of stochastic events *A* and *C*. The strength of Bayes algorithm are simplicity, efficiency, convenience. We use Naïve Bayes and Bayes Net classifiers based on Bayes theorem.

$$P(C \mid A) = \frac{P(A \mid C) \times P(C)}{P(A)} \tag{6}$$

where *P* is the probability of variable
C is the hypothesis to be tested
A is evidence associated with C

Neural Network based classification extract patterns or trends from data which is too complex or imprecise to be noticed by humans or other computer algorithms. Strength of Neural Network based classifiers are works well with noisy data, can process numerical and categorical data, can perform well in several domains. We use multilayer perceptron for artificial neural network based classification in EOG dataset.

SVM is a margin classifier that draws an optimal hyper plane in the feature vector space and defines a boundary that maximizes the margin between data samples in two classes. This leads leading to good generalization properties. A key factor in SVM is to use kernels to construct nonlinear decision boundary. We use weka software which implements John Platt's Sequential Minimal Optimization (SMO) algorithm for training a support vector classifier.

Bagging is the process of developing a model for each random sample. Bagging classifies an unknown sample based on what majority of the models predict. In boosting, increased focus is given to misclassified instances to capture their behaviour well. We use AdaBoostM1 performance with J48 classifier.

IV. EMPIRICAL RESULTS

The basic problem of classification is to classify a given instance (40 instances with 6705 reading samples and 23492 writing samples) to one of the known classes (Read/Write). A set of features presumably contains enough information to distinguish among the classes. The problem of classification is defined by all its features. The number of features can be quite large, many of which can be irrelevant or redundant in classification task. Here we use various conventional classifiers and the results are tabulated.

Table 4: Detailed Accuracy by Class with conventional classifiers

Method	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
Naive Bayes	0.667	0.500	0.500	0.500	0.500	0.708	Read
(With All Features)	0.500	0.333	0.667	0.667	0.667	0.708	Write
	0.571	0.405	0.595	0.571	0.571	0.708	Average
Naive Bayes	0.500	0.125	0.750	0.7500	0.750	0.708	Read
(With 3 CBFS Features)	0.875	0.500	0.700	0.700	0.700	0.708	Write
	0.714	**0.339**	**0.721**	**0.714**	**0.702**	**0.708**	Average
BayesNet	0.500	0.000	1.000	1.000	1.000	0.823	Read
(With All Features)	1.000	0.500	0.727	0.727	0.727	0.823	Write
	0.786	0.286	0.844	0.786	0.767	0.823	Average
BayesNet	0.500	0.000	1.000	1.000	1.000	0.750	Read
(With 3 CBFS Features)	1.000	0.500	0.727	0.727	0.727	0.750	Write
	0.786	**0.286**	**0.844**	**0.786**	**0.767**	**0.750**	Average
SMO (With All Features)	0.500	0.250	0.600	0.600	0.600	0.625	Read
	0.750	0.500	0.667	0.667	0.667	0.625	Write
	0.643	0.393	0.638	0.643	0.637	0.625	Average
SMO (With 3 CBFS Features)	0.500	0.000	1.000	1.000	1.000	0.750	Read
	1.000	0.500	.727	0.727	0.727	0.750	Write
	0.786	**0.286**	**0.844**	**0.786**	**0.767**	**0.750**	Average
Multilayer Perceptron	0.667	0.500	0.500	0.500	0.500	0.646	Read
(With all features)	0.500	0.333	0.667	0.667	0.667	0.646	Write
	0.571	0.405	0.595	0.571	0.571	0.646	Average
Multilayer Perceptron	0.500	0.000	1.000	1.000	1.000	0.500	Read
(With 3 CBFS features)	1.000	0.500	0.727	0.727	0.727	0.500	Write
	0.786	**0.286**	**0.844**	**0.786**	**0.767**	**0.500**	Average
Lazy K-Star (With all Features)	0.667	0.250	0.667	0.667	0.667	0.813	Read
	0.750	0.333	0.750	0.750	0.750	0.813	Write
	0.714	0.298	0.714	0.714	0.714	0.813	Average
Lazy K-Star (With 3 CBFS Features)	0.833	0.250	0.714	0.714	0.714	0.896	Read
	0.750	0.167	0.857	0.857	0.857	0.896	Write
	0.786	**0.202**	**0.796**	**0.786**	**0.787**	**0.896**	Average
AdaBoostM1	0.833	0.125	0.833	0.833	0.833	0.750	Read
(With all Features)	0.875	0.167	0.875	0.875	0.875	0.750	Write
	0.857	0.149	0.857	0.857	0.857	0.750	Average
AdaBoostM1 (With 3 CBFS Features)	0.667	0.000	1.000	1.000	1.000	0.667	Read
	1.000	0.333	0.800	0.800	0.800	0.667	Write
	0.857	**0.190**	**0.886**	**0.857**	**0.851**	**0.667**	Average
Random Tree (With all Features)	0.833	0.000	1.000	1.000	1.000	0.917	Read
	1.000	0.167	0.889	0.889	0.889	0.917	Write
	0.929	0.095	0.937	0.929	0.927	0.917	Average
Random Tree (With 3 CBFS Features)	0.667	0.375	0.571	0.571	0.571	0.646	Read
	0.625	0.333	0.714	0.714	0.714	0.646	Write
	0.643	**0.351**	**0.653**	**0.653**	**0.645**	**0.646**	Average

We use the default settings and apply as possible conventional classifiers to EOG datasets before and after EBFS feature selection, and obtain the results of percentage split (66%- for training data) in Table 4. In the table we report performance by TP Rate, FP Rate, Precision, Recall, F-Measure and ROC Area. The dataset is divided into training set with two-third of original and rest one-third for testing instances. We focus CBFS feature selection algorithm in order to improve the classification. The detailed accuracy by class results are shown in Table 4 with an emphasis on the difference before and after feature selection. This shows after feature selection class accuracy increase by minimum number of features (3 out of 13 features). The evaluation of a feature subset becomes simpler than that of a full set.

The Bayes theorem based classifiers Naive Bayes and Bayes Net showed a noticeable difference in accuracy. The Precision(0.844), Recall(0.786) and F- Measure(0.767) of BayesNet is little more than that of the other method NaiveBayes. The ROC area(0.750) indicates that it performs as usual as the other methods such as support vector machine, neural network and tree based method.

By using CBFS features the SVM based classifier and Neural Network based classifier performance is increased by precision as 84.4%, recall as 78.6%, and has F-Measure of 76.7% with ROC Area of 75%. The lazy learner with CBFS classified the task with improved ROC area 89.6%, F-Measure of 78.7% but the precision value 79.6% which is lower than that of other classifiers.

Bagging and Boosting (after CBFS) performance evaluation by AdaBoostM1 shows the maximum precision of 88.6% and recall of 85.7%. But the ROC area for which is low (66.7%).

Though all the classifier performance is increased by CBFS feature selection algorithm is increased in terms of precision, recall and accuracy the Decision Tree based classifier performance is decreased by using CBFS features. The features set with all features classifies reading and writing on an average of TP as 92.9% , precision as 93.7%,recall as 92.9% , F-Measure as 92.7% and ROC area as 91.7% which is the best among all other classifiers performance. When CBFS features are used with decision tree based classifier (Random Tree) performance is worsen (precision 65.3%,recall 65.3%, F-measure 64.5% and ROC Area 64.6%).

CONCLUSION AND FUTURE WORK

By using CBFS we improved the feature extraction (in EOG signals) stage by selecting efficient subset of features. We extract 3 (Mean, Trim mean, Median) clearly separable statistical features from a database of EOG signals. The classification result Table 5 shows that decision tree based classifier with CBFS features in EOG signal is not suitable. All the features may be considered for classification using the Decision Tree classifier for better performance. The reduced set of CBFS features with other classifiers alone we classified EOG signals effectively and efficiently for reading and writing activity. By this we recommend good features for recognizing reading and writing in EOG signals. We analyse the recognition accuracy as a function of the feature subset size using conventional classifiers. CBFS CScore evaluates each statistical feature based on degree of condensation of samples to the centroid of the classes such as read and write, and reduces the validation errors. CBFS features took minimum time for building the model with compared to all features. The problem of over fitting exists between the classes. This work can be extended in various directions. We plan to explore a line of research that focuses on comparison of different feature selection methods with different set of features from EOG signals of various activities.

REFERENCES

[1] A. Bulling, J.A. Ward, H. Gellersen and G. Troster, "Robust Recognition of Reading Activity in Transit using Wearable ElectroOculoGraphy", in proceedings of 6th International Conference on Pervasive Computing, Springer, 2008.

[2] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics*, 2007.

[3] A. Bulling, D. Roggen and G. Troster, "Wearable EOG goggles: Seamless sensing and context-awareness in everyday environments", Journal of Ambient Intelligence and Smart Environments, 2009.

[4] M. A. Tinati and B. Mozaffary, "A wavelet packets approach to electrocardiograph baseline drift cancellation," International Journal of Biomedical Imaging, 2006.

[5] D. L. Donoho, "De-noising by soft thresholding," *IEEE Transactions on Information Theory*, 1995.

[6] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in Proceedings of Symposium on Eye Tracking Research & Applications (ETRA '00), 2000.

[7] H. Peng, "mRMR Feature Selection Toolbox for MATLAB," February 2008; http://research.janelia.org/peng/proj/mRMR/.

[8] Muhsin Ozdemir, Ph.D Dissertation, Evolutionary Computing For Feature Selection And Predictive Data Mining, Rensselaer Polytechnic InstituteTroy, New York,2002

[9] Bulling A, Ward JA, Gellersen H, Troster G. "Eye Movement analysis for activity recognition using electrooculography" *IEEE Trans Pattern Anal Mach Intell* 2011.

[10] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and minredundancy," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005.

[11] Bing Long, Shulin Tian & Houjun Wang,"Feature Vector Selection Method Using Mahalanobis Distance for Diagnostics of Analog Circuits Based on LS-SVM," Springer Science, LLC 2012.

[12] Xuejun Dong, Paul Nyren, Bob Patton, Anne Nyren, Jim Richardson, and Thomas Maresca," Wavelets for Agriculture and Biology: A Tutorial with Applications and Outlook", BioScience, May 2008.

[13] S. S. Hacisalihzade, L.W. Stark, and J. S. Allen, "Visual perception and sequences of eye movement fixations: a stochastic modeling approach," *IEEE Transactions on Systems Man and Cybernetics*, 1992.

[14] A. Bulling, D. Roggen, and G. Tr¨oster, "It's in Your Eyes – Towards Context-Awareness and Mobile HCI UsingWearable EOG Goggles," in Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp '08), 2008.

[15] M. Dash, H. Liu, H. Motoda, Consistency based feature selection, in: Proceedings of Fourth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Kyoto, Japan, 2000.

[16] F. T. Keat, S. Ranganath, and Y. V. Venkatesh, "Eye gaze based reading detection," in *Proceedings of IEEE Conference on Convergent* Technologies *for the Asia-Pacific Region*, vol. 2, 2003, pp. 825–828.

[17] A.L. Blum, P. Langley, Selection of relevant features and examples in machine learning, Artificial Intelligence 97 (1997).

[18] Jain & Zongker," Feature Selection: Evaluation, Application,and Small Sample Performance", *IEEE Trans. PAM,* 1997.

[19] Y. Yamagishi K, Hori J, Miyakawa M. Development of EOG-based communication system controlled by eight-directional eye movements. *IEEE Trans Med Biol Eng* 2006.

[20] S. Aungsakul, A. Phinyomark, P. Phukpattaranont, C. Limsakul, Evaluating Feature Extraction Methods of Electrooculography (EOG) Signal for Human-Computer Interface, Procedia Engineering 32 (2012) 246 – 252.

[21] V.N. Vapnik, "*Statistical Learning Theory"*, Wiley-Interscience Publications, 1998.

[22] Haun Liu, H. Motoda, "Feature Selection for Knowledge Discovery and Data Mining", Kluwer Academic publishers,1998.

[23] Monica M. Chamay, Se-do Oh, and Young-Jin Kim, "Development of Dimension Reduction Algorithm in Diagnostic System with Artificial Neural Network" Proceedings of the International MultiConference of Engineers and Computer Scientists 2013, Hong Kong.

[24] J. Han and M. Kamber. *Data Mining: Concepts and Techniques.*Elsevier (2006).

[25] M. M. Hayhoe and D. H. Ballard, "Eye movements in natural behavior," *Trends in Cognitive Sciences*, 2005.

[26] Y. Huang, D. Xu and F Nie, *Semi-Supervised Dimension Reduction using Trace Ratio Criterion.* IEEE transactions on Neural Networks and Learning Systems. 2012.

[27] Ishan Ranjan, Spectral feature extraction of biomedical signal, Concordia University,1991.

[28] Z. Lv, X. Wu, M. Li, A Research on EOG Feature ParametersExtraction Based on Linear Predictive Coding Model, *International Conference IEEE, Bioinformatics and Biomedical Engineering*, USA (2009).

[29] P. Lall, *Decorrelated feuature space and Neural nets based framework for failure modes clustering in Electronics subjected to mechanical shock.* IEEE Transactions on Reliability, 2012.

[30] Mehran Sahami, Ph.D. Dissertation Using Machine Learning to improve information access, Standford University (1999).

[31] Pablo A. Estevez, Michel Tesmer, Claudio A. Perez, "Normalized Mutual Information Feature Selection," *IEEE Transactions on Neural Networks, February 2009.*

[32] Gavin Brown, Adam Pocock, Ming-Jie Zhao, Mikel Lujan, "Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection," Journal of Machine Learning Research (2012).

[33] H. Cheng, Z. Qin, C. Feng, Y. Wang, and F. Li. Conditional mutual information-based feature selection analyzing for synergy and redundancy. *Electronics and Telecommunications Research Institure (ETRI) Journal,* 2011.

[34] G. Brown. A new perspective for information theoretic feature selection. *In International conference on Artificial Intelligence and Statistics, volume* 5, 2009.

[35] D. D. Salvucci and J. R. Anderson, "Automated eye-movement protocol analysis," *Human-Computer Interaction*, vol. 16, 2001.

[36] K. S. Balagani and V. V. Phoha. On the feature selection criterion based on an approximation of multidimensional mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 32(7),2010.

[37] B. Guo and M.S. Nixon. Gait feature subset selection by mutual information. *IEEE Trans Systems, Man and Cybernetics,* January 2009.

[38] A. Jakulin. *Machine Learning Based on Attribute Interactions*. PhD thesis, University of Ljubljana, Slovenia, 2005.

[39] Oh S (2011) A New Feature Evaluation Method Based on category Overlap. Comput Biol Med.

[40] H. Liu, H. Motoda. Computational Methods of Feature Selection. Chapman & Hall CRC, 2008.

[41] Ruo-Fei Du, Ren-Jie Liu, Tian-Xiang Wu and Bao-Liang Lu,"Online Vigilance Analysis Combining Video and Electrooculography Features", in proceedings of 19th International Conference on Neural Information Processing (iconip'12),2012.