

UNIVERSITY OF SOUTHERN QUEENSLAND, AUSTRALIA



**ANALYSIS AND CLASSIFICATION OF EEG SIGNALS**

A Dissertation Submitted by

Siuly

For the Award of

**Doctor of Philosophy**

July, 2012

## Abstract

Electroencephalography (EEG) is one of the most clinically and scientifically exploited signals recorded from humans. Hence, its measurement plays a prominent role in brain studies. In particular, the examination of EEG signals has been recognized as the most preponderant approach to the problem of extracting knowledge of the brain dynamics. EEG recordings are particularly important in the *diagnosis of epilepsy* and in *brain computer interface (BCI)*. In BCI systems, EEG signals help to restore sensory and motor functions in patients who have severe motor disabilities. Analysing EEG signals is very important both for supporting the diagnosis of brain diseases and for contributing to a better understanding of cognitive process.

Although EEG signals provide a great deal of information about the brain, research in classification and evaluation of these signals is limited. Even today the EEG is often examined manually by experts. Therefore, there is an ever-increasing need for developing automatic classification techniques to evaluate and diagnose neurological disorders. Classification techniques can help to differentiate EEG segments and to decide whether a person is healthy. A big challenge is for BCI systems to correctly and efficiently identify different EEG signals of different motor imagery (MI) tasks using appropriate classification algorithms to assist motor disabled patients in communication.

In this dissertation, we aim to develop methods for the analysis and classification of epileptic EEG signals and also for the identification of different categories of MI tasks based EEG signals in BCI's development.

In order to classify epileptic EEG signals, we propose two methods, simple sampling technique based least square support vector machine (SRS-LS-SVM) and clustering technique based least square support vector machine (CT-LS-SVM). The experimental results show that both algorithms perform well in the EEG signal classification and the CT-LS-SVM method takes much less execution time compared to the SRS-LS-SVM technique. The research findings also indicate that the proposed approaches are very efficient for classifying two categories of EEG signals. This

research can help to provide clinical information about patients who have epilepsy, neurological disorders, mental or physiological problems.

In BCI systems, if the MI tasks are reliably distinguished through identifying typical patterns in EEG data, motor disabled people could communicate with a device by composing sequences of these mental states. In this dissertation, for the identification of MI tasks in BCI applications, we developed three methods:

- (1) Cross-correlation based logistic regression (CC-LR).
- (2) Modified CC-LR with diverse feature sets.
- (3) Cross-correlation based least square support vector machine (CC-LS-SVM).

The experimental results have demonstrated the effectiveness of the methods for the identification of MI tasks. These techniques can assist clinical diagnoses and rehabilitation tasks.

Finally we investigated two issues for the MI classification:

- (1) Which algorithm performed better.
- (2) Which EEG data is more suitable for getting information about MI tasks.

Is it the motor area data or the all-channels data?

To answer these two questions, we considered the three algorithms: the CC-LS-SVM, the CC-LR and the cross-correlation based kernel logistic regression (CC-KLR). Based on the experimental results, we concluded that the CC-LS-SVM algorithm is the best algorithm for the MI tasks EEG signal classification, and the all-channels EEG data can provide better information than the motor area EEG data for the MI tasks classification. Furthermore, the CC-LS-SVM approach can correctly identify the discriminative MI tasks, demonstrating the algorithms superiority in the classification performance over other existing methods.

## Certification of Dissertation

I hereby declare that the work presented in this dissertation is my own and is, to the best of my knowledge and belief, original except as acknowledged in the text. This dissertation has not previously been submitted either in whole or in part for a degree at this or any other university.

Signed

(Siuly)

Date:

Signed

(Principal Supervisor)

(Associate Professor Yan Li)

Date:

Signed

(Associate Supervisor)

(Associate Professor Peng (Paul) Wen)

Date:

## Acknowledgements

First of all, I would like to thank almighty Allah, for His guidance and strength.

I would like to express my sincere gratitude and appreciation to my supervisor, Associate Professor Yan Li, for her continuous inspiration, support, guidance, encouragement, advice, patience, and individual feedback throughout the course of my PhD study. She was an approachable supervisor who motivated me to achieve my goals. I feel very grateful and blessed to have worked under her supervision. I would not imagine carrying out this research without her constant support and help. I would also like to express my sincere gratitude and appreciation to my associate supervisor Associate Professor Peng (Paul) Wen, for his continuous encouragement, support, valuable advice, and suggestions.

I gratefully acknowledge the University of Southern Queensland (USQ) for offering me the 2009 USQ Postgraduate Research Scholarship to perform this research successfully, without which, this work would not have been possible.

I would also like to thank Professor Mark Sutherland, Director of Centre for Systems Biology (CSBi) for granting me half-tuition fees support, and supporting me to attend conferences.

I would also like to thank to all of the staff in the Department of Mathematics and Computing for their co-operation. I would also like to express my gratitude and appreciation to my postgraduate colleagues in our research group for their discussions, comments and advice, especially during our group seminars. I would also like to thank all of my friends for their encouragement and supports.

I wish to express my appreciation to my honourable parents, loving brothers and sisters for their support. Last and most importantly, I wish to thank my husband, Dr. Enamul Kabir who stands beside me and encourages me constantly for every step of my work in this thesis. My thanks are also to my son, Shadman Enam Srijon for giving me happiness, joy and enduring love. I would like to dedicate this dissertation to my son, my beloved husband and my respectable parents.

# Contents

<b>Abstract</b>	i
<b>Certification of Dissertation</b>	iii
<b>Acknowledgements</b>	iv
<b>Contents</b>	v
<b>List of Figures</b>	x
<b>List of Tables</b>	xiv
<b>Publications Based on This Dissertation</b>	xviii
<b>1 Introduction</b>	<b>1</b>
1.1 Overview and Motivation of the Study	2
1.2 Problem Statements	3
1.3 Contributions of the Dissertation	4
1.4 Structure of the Dissertation	8
<b>2 Overview of EEG signal classification and its background knowledge</b>	<b>10</b>
2.1 Background knowledge related to EEG signals	10
2.1.1 Human Brain	11
2.1.1.1 Brain structures and their functions	11
2.1.1.2 Human brains' neurophysiology	12
2.1.2 Electroencephalography (EEG)	15
2.1.2.1 Nature or rhythms of the EEG signals	18
2.1.3 Epilepsy, epileptic seizures and their effects on EEG signals	20
2.1.4 Concept of Brain Computer Interfaces (BCIs)	21
2.1.4.1 Architecture of BCI systems	22
2.1.4.2 Application of BCI systems	24
2.2 Overview of EEG signal classification	25
2.2.1 Concept of classification	25
2.2.2 Types of classification	26

2.2.2.1	Supervised classification	26
2.2.2.2	Unsupervised classification	28
2.2.3	Structure of classification	29
2.2.4	Common used methods of the EEG signal classification	30
2.2.4.1	Methods used in the epileptic EEG classification	30
2.2.4.2	Methods used in the MI based EEG classification in BCI systems	32
2.3	Summary	35
<b>3</b>	<b>Two-stage Random Sampling with Least Square Support Machine</b>	<b>37</b>
3.1	Introduction	37
3.2	Related Work	39
3.3	Methodology	41
3.3.1	Random sample and sub-sample selection using SRS technique	42
3.3.2	Feature extraction from different sub-samples	43
3.3.3	Least square support vector machine (LS-SVM) for classification	44
3.3.4	Performance evaluation	46
3.4	Experimental Data	47
3.4.1	EEG epileptic data	47
3.4.2	Mental imagery tasks EEG data	49
3.4.3	Two-class synthetic data	50
3.5	Experimental Results and Discussions	50
3.5.1	Results for the EEG epileptic dataset	50
3.5.2	Results for the mental imagery tasks EEG dataset	55
3.5.3	Results for the two-class synthetic data	57
3.6	Conclusions	59
<b>4</b>	<b>Clustering Technique based Least Square Support Vector Machine</b>	<b>61</b>
4.1	Introduction	62

4.1.1	Objective of the research	63
4.2	Proposed Methodology	64
4.2.1	Clustering technique (CT) for feature extraction	64
4.2.1.1	Stage 1: Determination of clusters	65
4.2.1.2	Stage 2: Determination of sub-clusters	65
4.2.1.3	Stage 3: Statistical feature extraction	66
4.3	Experimental Data	66
4.4	Implementation of the proposed CT-LS-SVM algorithm	67
4.5	Experimental Results and Discussions	70
4.5.1	Classification results for the mental imagery tasks EEG data	71
4.5.2	Classification results for the EEG epileptic data	75
4.5.3	Classification results for the motor imagery EEG data	79
4.6	Conclusions	82
<b>5</b>	<b>Cross-correlation based Logistic Regression Method for Classification of MI Tasks</b>	<b>84</b>
5.1	Introduction	84
5.2	Theoretical Background	87
5.2.1	Cross-correlation technique	87
5.2.2	Logistic regression model	88
5.3	Proposed Methodology	90
5.3.1	Feature extraction using cross-correlation (CC) technique	90
5.3.2	MI tasks signal classification by logistic regression (LR)	92
5.3.3	Performance evaluation methods	92
5.3.3.1	$k$ -fold cross validation method	92
5.3.3.2	Classification accuracy	93
5.3.3.3	Confusion matrix	93
5.3.4	Experimental Data	93
5.4	Results and Discussions	94
5.4.1	Classification results for dataset IVa	94
5.4.2	Classification results for dataset IVb	100
5.5	Conclusions and Recommendations	103

<b>6</b>	<b>Modified CC-LR Algorithm with Diverse Feature Sets</b>	<b>104</b>
6.1	Background	104
6.2	Methodology	106
6.3	Experimental Evaluation and Discussions	109
6.3.1	Implementation of the CC technique for the feature extraction	109
6.3.2	MI classification results testing different features	114
6.3.3	A comparative study	120
6.4	Conclusions and Recommendations	122
<b>7</b>	<b>LS-SVM with Tuning Hyper Parameters: Improving Prospective Performance in the MI task Recognition</b>	<b>123</b>
7.1	Introduction	124
7.2	Review of the existing classification techniques	126
7.3	Proposed Method	128
7.3.1	Reference signal selection	129
7.3.2	Computation of a cross-correlation sequence	130
7.3.3	Statistical feature extraction	133
7.3.4	Classification	133
7.4	Performance Evaluation	135
7.5	Experiments and Results	136
7.5.1	Tuning the hyper parameters of the LS-SVM classifier	137
7.5.2	Variable selections in the logistic regression and kernel logistic regression classifiers	140
7.5.3	Performances on both datasets	141
7.5.4	Performance comparisons with the existing techniques	148
7.6	Conclusions	149
<b>8</b>	<b>An Investigation Study</b>	<b>152</b>
8.1	Background	152
8.2	Review of the existing research	154
8.3	Data and Implementation	156

8.4	Experiments and Results	159
8.4.1	Results for dataset IVa	159
8.4.2	Results for dataset IVb	165
8.5	Conclusions and Contributions	166
<b>9</b>	<b>Conclusions and Future Work</b>	<b>168</b>
9.1	Summary and Conclusions of the Dissertation	168
9.2	Future work	174
	<b>References</b>	<b>176</b>

## List of Figures

No	Title	Page
Figure 2.1	Anatomical areas of the brain	12
Figure 2.2	A simple structure of neuron	13
Figure 2.3	A single cortical pyramidal cell showing the current flow that contributes to the surface EEG during a net excitatory input	14
Figure 2.4	First recording of EEG signals made by Hans Berger	16
Figure 2.5	The international 10-20 electrode placement system	17
Figure 2.6	Example of different types of normal EEG rhythms	19
Figure 2.7	A general architecture of a BCI system	23
Figure 2.8	Mutually exclusive training and testing set	28
Figure 2.9	Process of signal classification in the biomedical engineering	29
Figure 3.1	Block diagram of the SRS-LS-SVM method for EEG signal classification	41
Figure 3.2	Random samples and sub-sample selection diagram using SRS technique	42
Figure 3.3	Example of five different sets of EEG signals taken from different subjects	48
Figure 3.4	Exemplary EEG signals for left hand movements (class 1), right hand movements (class 2) and word generation (class 3) taken from Subject 1	49
Figure 3.5	Exemplary mean feature points obtained by the SRS from healthy subjects with eyes open (Set A)	51
Figure 3.6	Exemplary mean feature points obtained by the SRS from epileptic patients during seizure-free intervals within the hippocampal formation of the opposite hemisphere of the brain (Set C)	51
Figure 3.7	Exemplary mean feature points obtained by the SRS from	52

	epileptic patients during seizure activity (Set E)	
Figure 3.8	ROC curve for healthy subjects with eye open (Set A) and epileptic patients during seizure activity (Set E) in the EEG epileptic data	55
Figure 3.9	Exemplary mean feature points obtained by the SRS from left hand movements (class 1) of Subject 1	56
Figure 4.1	Block diagram of the proposed methodology for EEG signal classification	64
Figure 4.2	Clustering technique diagram for obtaining different clusters, sub-clusters and statistical features	65
Figure 4.3	Comparison of execution time between the CT-LS-SVM and SRS-LS-SVM methods for the mental imagery tasks EEG data	74
Figure 4.4	Comparison of execution time between the CT-LS-SVM and SRS-LS-SVM methods for the EEG epileptic data	77
Figure 4.5	Comparison of execution time between the CT-LS-SVM and SRS-LS-SVM methods for the motor imagery EEG data	80
Figure 5.1	Example of a typical cross-correlogram	88
Figure 5.2	The typical signals of the RH and the RF MI tasks for each subject of dataset IVa	95
Figure 5.3	The typical cross-correlograms for the RH and the RF MI signals of each subject in dataset IVa	96
Figure 5.4	Classification plots for the 1-fold of subject 1 in dataset IVa	98
Figure 5.5	Performance comparisons of four other existing methods of the literature with the proposed CC-LR method	100
Figure 5.6	The typical signals and cross-correlograms for the RF and the LH MI signals of dataset IVb	101
Figure 5.7	Classification plot for the 1-fold of dataset IVb	102
Figure 6.1	Schematic diagram for the classification of the MI based EEG signal in BCIs	106
Figure 6.2	(a) Structure of the human brain (b) the international 10-	107

	20 electrode placement system	
Figure 6.3	The typical cross-correlograms for the RH and the RF MI signals of S1 of dataset IVa	111
Figure 6.4	The typical cross-correlograms for the RH and the RF MI signals of S2 of dataset IVa	111
Figure 6.5	The typical cross-correlograms for the RH and the RF MI signals of S3 of dataset IVa	112
Figure 6.6	The typical cross-correlograms for the RH and the RF MI signals of S4 of dataset IVa	112
Figure 6.7	The typical cross-correlograms for the RH and the RF MI signals of S5 of dataset IVa	113
Figure 6.8	The typical cross-correlograms for the RF and the LH MI signals of dataset IVb	114
Figure 6.9	Correct classification rate for the two features set, the four features set and the six features set in each of the three folds for: (a) S1 (b) S2 (c) S3 (d) S4 (e) S5 in dataset IVa. Error bars indicate standard deviation	117- 119
Figure 6.10	Correct classification rate for the two features set, the four features set and the six features set for each of the three folds in dataset IVb. Error bars indicate standard deviation	120
Figure 7.1	Fundamental structure of brain computer interface (BCI)	124
Figure 7.2	Block diagram of the proposed CC-LS-SVM technique for the MI EEG signal classification in BCIs development	129
Figure 7.3	Typical reference signals: (A) Dataset IVa and (B) Dataset IVb	129
Figure 7.4	Typical right hand and right foot MI signals and their respective cross-correlograms for subject <b>aa</b> in dataset IVa	131
Figure 7.5	Typical right foot and left hand MI signals and their respective cross-correlograms for dataset IVb	132
Figure 7.6	Partitioning design of the obtained feature vectors for the 10-fold cross-validation method	136
Figure 7.7 (A)	Process of the two-step grid search for optimizing the parameters $\gamma$ (gamma) and $\sigma^2$ (sig2) of the LS-SVM	138

	classifier in the 1-fold of subject <b>aa</b> of dataset IVa	
Figure 7.7 (B)	Process of the two-step grid search for optimizing the parameters $\gamma$ (gamma) and $\sigma^2$ (sig2) of the LS-SVM classifier in the 1-fold of dataset IVb	139
Figure 7.8	Comparisons of the individual classification accuracies among the logistic regression, kernel logistic regression and the LS-SVM for each of the 10-folds: (a) subject <b>aa</b> (b) subject <b>al</b> (c) subject <b>av</b> (d) subject <b>aw</b> (e) subject <b>ay</b> in dataset IVa	142- 143
Figure 7.9	Comparisons of the individual classification accuracies among the logistic regression, kernel logistic regression and the LS-SVM for each of the 10-folds in dataset IVb	144
Figure 8.1	Locations of electrodes for datasets IVa and IVb in BCI Competition III. 118 electrodes are shown labelled according to the extended international 10/20 system	157
Figure 8.2	Comparison of the performance between the motor area EEG and the all-channels EEG data for the CC-LS-SVM algorithm. The vertical lines show the standard errors of the test accuracies	161
Figure 8.3	Comparison of the performance between the motor area EEG and the all-channels EEG data for the CC-LR algorithm. The vertical lines show the standard error of the test accuracies	162
Figure 8.4	Comparison of the performance between the motor area EEG and the all-channels EEG data for the CC-KLR algorithm. The vertical lines show the standard error of the test accuracies	163
Figure 8.5	The comparison of the performance for the CC-LS-SVM, CC-LR and CC-KLR algorithms between the motor area data and the all-channels data. The vertical lines show the standard errors of the test accuracies	166

## List of Tables

No	Title	Page
Table 3.1	Experimental results and the area values under ROC curve for two-class pairs of the EEG signals for the EEG epileptic database	53
Table 3.2	Comparison of performance of our proposed method with two most recently reported methods for Set A and Set E of the EEG epileptic database	54
Table 3.3	Number of recorded values in four sessions from the mental imagery tasks EEG data (Data set V for BCI competition III)	56
Table 3.4	Experimental results for different pairs of two-class EEG signals for the mental imagery tasks EEG data (Data set V for BCI competition III)	57
Table 3.5	The original two-class synthetic data from Ripley (1996) and the extracted feature vectors by the SRS technique	58
Table 3.6	Confusion matrix for Ripley data (1996)	58
Table 3.7	Experimental results for the Ripley data (1996)	59
Table 4.1	The number of recorded values in four sessions from the mental imagery tasks EEG data	68
Table 4.2	The number of clusters and sub-clusters and time period for each cluster and sub-cluster of a class for a subject for the mental imagery tasks EEG data	68
Table 4.3	The number of clusters and sub-clusters and time period for each cluster and sub-cluster of a class for a subject for the motor imagery EEG data	69
Table 4.4	Performance comparison of the proposed CT-LS-SVM versus the SRS LS-SVM for different pairs of two-class EEG signals from the mental imagery tasks EEG data	72
Table 4.5	Classification accuracy of the proposed CT-LS-SVM	73

	method by the 10-fold cross validation for the mental imagery tasks EEG data	
Table 4.6	Performance comparison of different methods with the proposed CT-LS-SVM algorithm for the mental imagery tasks EEG data	74
Table 4.7	Performance comparison of the proposed CT-LS-SVM versus the SRS-LS-SVM method for different pairs of two-class EEG signals from the EEG epileptic data	76
Table 4.8	Classification accuracy of the proposed CT-LS-SVM method by the 10-fold cross validation for the EEG epileptic data	77
Table 4.9	The obtained performance with our proposed CT-LS-SVM method and other methods from the literature for healthy subjects with eyes open (Set A) and epileptic patients during seizure activity (Set E) of the EEG epileptic data	78
Table 4.10	Performance comparison between the CT-LS-SVM and SRS-LS-SVM for the motor imagery EEG data	79
Table 4.11	Classification accuracy of the CT-LS-SVM method by the 10-fold cross validation for the motor imagery EEG data	80
Table 4.12	Comparison of classification accuracy for the motor imagery EEG data with other EEG signal classification attempts	81
Table 4.13	Summary results of the proposed CT-LS-SVM approach and the SRS-LS-SVM method for the mental imagery tasks EEG data, the EEG epileptic data and the motor imagery EEG data.	82
Table 5.1	The information of original data for BCI Competition III, dataset IVa	95
Table 5.2	The 3-fold cross validation results by the CC-LR method on testing set for dataset IVa of BCI Competition III	97
Table 5.3	Confusion matrix for the 1-fold of subject 1 from dataset	98

	IVa	
Table 5.4	Performance comparison of the CC-LR algorithm with the R-CSP with aggregation and the CT-LS-SVM algorithms for dataset IVa, BCI III	99
Table 5.5	The 3-fold cross validation results by the proposed method on testing set for dataset IVb of BCI Competition III	101
Table 5.6	Confusion matrix for the 1-fold of dataset IVb	102
Table 6.1	Cross-validation results with the modified CC-LR method on testing set for dataset IVa	116
Table 6.2	Cross validation results by the proposed method on testing set for dataset IVb	119
Table 6.3	Comparison of the classification performance between our proposed algorithm and the most recent reported eight algorithms for dataset IVa in BCI Competition III	121
Table 7.1	Optimal values of the parameters $\gamma$ and $\sigma^2$ of the LS-SVM for dataset IVa	139
Table 7.2	Optimal values of the parameters $\gamma$ and $\sigma^2$ of the LS-SVM for dataset IVb	140
Table 7.3	Classification results by the 10-fold cross-validation method on testing set of dataset IVa	141
Table 7.4	Classification results by the 10-fold cross-validation method on testing set of dataset IVb	144
Table 7.5	Classification results of the three classifiers for the nine features and the six features for the reference signals, Fp1 and C3, in dataset IVa	146
Table 7.6	Classification results of the three classifiers for the nine features and the six features for the reference signals, Fp1 and C3, in dataset IVb	148
Table 7.7	Performance comparisons for dataset IVa	149
Table 8.1	Experimental results of the three algorithms reported in percentage (mean $\pm$ standard deviation) for dataset IVa	160
Table 8.2	The comparison of our three algorithms with two existing	163

	methods for the motor area data in dataset IVa	
Table 8.3	The comparison of our three algorithms with two existing methods for the all-channels data in dataset IVa	164
Table 8.4	Experimental results of the three algorithms reported in terms of the 3-fold cross validation accuracy (mean $\pm$ standard deviation) for dataset IVb	165

## Publications Based on This Dissertation

1. Siuly, Li, Y., (2012) 'Improving the separability of motor imagery EEG signals using a cross correlation-based least square support vector machine for brain computer interface', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, DOI (identifier) 10.1109/TNSRE.2012.2184838, in press. (ERA A\*, 2010)
2. Siuly, Li, Y. and Wen, P. (2011) '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. (ERA A\*, 2010)
3. Siuly, Li, Y. and Wen, P. (2011) 'Identification of Motor Imagery Tasks through CC-LR Algorithm in Brain Computer Interface', *International Journal of Bioinformatics Research and Applications* for publication, 2011 (in press). (ERA B, 2010)
4. Siuly, Li, Y. and Wen, P. (2011) '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. (ERA C, 2010)
5. Siuly, Li, Y. and Wen, P. (2010) 'Analysis and classification of EEG signals using a hybrid clustering technique', *The proceedings of the 2010 IEEE/ICME International Conference on Complex Medical Engineering (CME2010)*. pp. 34–39.
- 6 Siuly, Li, Y. and Wen, P. (2009) 'Classification of EEG signals using Sampling Techniques and Least Square Support Vector Machines', *The proceedings of Fourth International Conference on Rough Sets and Knowledge Technology (RSKT 2009)*, LNCS 5589 (2009), pp. 375-382.

7. Siuly, Li, Y., Wu, J. and Yang, J. (2011) 'Developing a Logistic Regression Model with Cross-Correlation for Motor Imagery Signal Recognition', *The proceedings of the 2011 IEEE IICME International Conference on Complex Medical Engineering May 22 - 25, Harbin, China*, pp. 502-507.
  
8. Siuly, Li, Y. and Wen, P., (2011) 'Comparisons between Motor Area EEG and all-Channels EEG for Two Algorithms in Motor Imagery Task Classification', submitted to the *International Journal of Intelligent Systems Technologies and Applications* for publication, February 2012. (ERA B, 2010)
  
9. Siuly, Li, Y. and Wen, P., (2012) 'Modified CC-LR algorithm with three diverse feature sets for motor imagery tasks classification in EEG based brain computer interface', submitted to the *Journal of Applied Mathematics and Computation*, February 2012. (ERA A, 2010)