

An Improved EEG Pattern Classification System Based on Dimensionality Reduction and Classifier Fusion

By

Akram Saleh Mousa AlSukker

Thesis submitted as a requirement for the degree of

Doctor of Philosophy

School of Electrical, Mechanical and Mechatronic Systems
Faculty of Engineering and Information Technology

University of Technology, Sydney (UTS)

July, 2012

Certificate

I, Akram Saleh Mousa AlSukker, declare that this thesis titled, An Improved EEG Pattern Classification System Based on Dimensionality Reduction and Classifier Fusion, and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has been previously submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Production Note:
Signature removed prior to publication.

Date:

6/7/2012

Acknowledgements

First and foremost, all my praise be to Allah the Most Gracious, the Most Merciful whose guidance constantly inspires me and has given me courage and strength to complete my PhD. I especially express my deepest gratitude to my supervisor, **Dr. Ahmed Al-Ani**, for his continual guidance, positive comments and tremendous support throughout my PhD candidature. His perpetual energy, enthusiasm and devotion to biomedical signal processing have motivated me to continue my research in the area. Special thanks go to **Dr. Rami Kushaba** for his collaboration, creative ideas and continuous support. He was my companion throughout my candidature and was of great help in my thesis writing. My associate supervisor **A/Prof. Adel Al-Jumaily** also deserves a special thank for his guidance, advice and assistance during the course of my PhD candidature. I wish to acknowledge the Department of Medical Informatics at the University of Technology, Graz, in Austria and the BCI competition web site for making the EEG datasets available for the research. I would also like to thank Ms Rosalind Elliott from Royal North Shore Hospital, Sydney, Australia for providing sleep study datasets. I would like to express my appreciation to Prof. Hung Nguyen for his support and for facilitating the use of laboratory equipment. My thanks are also due to A/Prof. Fazlul Huq of Discipline of Biomedical Science, The University of Sydney for kindly going through the thesis draft and for making useful suggestions towards improvement of language, its flow and coherence. I wish to thank my colleagues at UTS, in particular Mr. Thamer Hanoun for their company, friendship and support. My greatest heartfelt thanks go to my loving wife Zaynab Al-Eisawi for her constant support, encouragement and sacrifice throughout my candidature. I am also grateful to Allah for blessing us with two beautiful boys, Bassam and Yousef who have been a source of constant joy, love and affection in our heart and mind.

Dedication

I would like to dedicate this thesis to my mother and father, Noura and Saleh AlSukker. Throughout my lifetime, they always attempted to instill in me the desire to acquire new knowledge and their words of wisdom are a source of inspiration and strength.

Abstract

Analysis of brain electrical activities (Electroencephalography, EEG) presents a rich source of information that helps in the advancement of affordable and effective biomedical applications such as psychotropic drug research, sleep studies, seizure detection and brain computer interface (BCI). Interpretation and understanding of EEG signal will provide clinicians and physicians with useful information for disease diagnosis and monitoring biological activities. It will also help in creating a new way of communication through brain waves.

This thesis aims to investigate new algorithms for improving pattern recognition systems in two main EEG-based applications. The first application represents a simple Brain Computer Interface (BCI) based on imagined motor tasks, whilst the second one represents an automatic sleep scoring system in intensive care unit. BCI system in general aims to create a non-muscular link between brain and external devices, thus providing a new control scheme that can most benefit the extremely immobilised persons. This link is created by utilizing pattern recognition approach to interpret EEG into device commands. The commands can then be used to control wheelchairs, computers or any other equipment. The second application relates to creating an automatic scoring system through interpreting certain properties of several biomedical signals. Traditionally, sleep specialists record and analyse brain signal using electroencephalogram (EEG), muscle tone (EMG), eye movement (EOG), and other biomedical signals to detect five sleep stages: Rapid Eye Movement (REM), stage 1,... to stage 4. Acquired signals are then scored based on 30 seconds intervals that require manually inspecting one segment at a time for certain properties to interpret sleep stages. The process is time consuming and demands competence. It is thought that an automatic scoring system mimicking sleep expert rules will speed up the process and reduce the cost.

Practicality of any EEG-based system depends upon accuracy and speed. The more accurate and faster classification systems are, the better will be the chance to integrate them in wider range of applications. Thus, the performance of the previous systems is further enhanced using improved feature selection, projection and classification algorithms.

As processing EEG signals requires dealing with multi-dimensional data, there is a need to minimize the dimensionality in order to achieve acceptable performance with less computational cost. The first possible candidate for dimensionality reduction is employed using channel/feature selection approach. Four novel feature selection methods are de-

veloped utilizing genetic algorithms, ant colony, particle swarm and differential evolution optimization. The methods provide fast and accurate implementation in selecting the most informative features/channels that best represent mental tasks. Thus, computational burden of the classifier is kept as light as possible by removing irrelevant and highly redundant features.

As an alternative to dimensionality reduction approach, a novel feature projection method is also introduced. The method maps the original feature set into a small informative subset of features that can best discriminate between the different class. Unlike most existing methods based on discriminant analysis, the proposed method considers fuzzy nature of input measurements in discovering the local manifold structure. It is able to find a projection that can maximize the margin between data points from different classes at each local area while considering the fuzzy nature.

In classification phase, a number of improvements to traditional nearest neighbour classifier (k NN) are introduced. The improvements address k NN weighting scheme limitations. The traditional k NN does not take into account class distribution, importance of each feature, contribution of each neighbour, and the number of instances for each class. The proposed k NN variants are based on improved distance measure and weight optimization using differential evolution. Differential evolution optimizer is utilized to enhance k NN performance through optimizing the metric weights of features, neighbours and classes. Additionally, a Fuzzy k NN variant has also been developed to favour classification of certain classes. This variant may find use in medical examination. An alternative classifier fusion method is introduced that aims to create a set of diverse neural network ensemble. The diversity is enhanced by altering the target output of each network to create a certain amount of bias towards each class. This enables the construction of a set of neural network classifiers that complement each other.

All proposed feature reduction and classification algorithms have been tested and verified on different datasets including EEG (represented by BCIs and sleep scoring problem). The results obtained are very encouraging when comparing with their counterparts found in the literature.

Acronyms and Abbreviations

AAR: Adaptive Auto-Regressive
AASM: American Academy of Sleep Medicine
ACO: Ant Colony Optimization
ACOFUZZY: Fuzzy Ant Colony Optimization
ANOVA: Analysis of Variance
AR: Auto-Regressive
BCI: Brain computer Interface
BIC: Bayesian Information Criterion
BMI: Brain Machine Interface
BPSO: Binary Particle Swarm Optimization
BPSOMI: Mutual Information based Binary Particle Swarm Optimization
BSS: Blind Source Separation
CAT scan: Computerized Axial Tomography scan
*cb*NNE: Class biased Neural Network Ensemble
CSP: Common Spatial Pattern
CT scan: Computerized Tomography scan
*CWk*NN: Class Weighting *k* Nearest Neighbour
CWT: Continuous Wavelet Transform
DE: Differential Evolution
DEFS: Differential Evolution Feature Selection
DFT: Discrete Fourier Transform
DGA: Diverse Genetic Algorithm
DNF: Desired Number of Features
*DSk*NN: Dempster-Shafer *k* Nearest Neighbour
DWT: Discrete Wavelet Transform
EAs: Evolutionary Algorithms
ECG: electrocardiogram
ECoG: Electrocorticogram
EEG: Electroencephalogram
EMG: Electromyogram
EOG: electrooculogram
ERP: Event-Related Potential
ES: Exhaustive Search
EV: Evoked Potential
FCM: Fuzzy C-Means
FDR: Fisher Discrimination Rate
FFT: Fast Fourier Transform
*FISk*NN: Fuzzy Inference System *k* Nearest Neighbour

f k NN: Fuzzy k Nearest neighbour
FLDA: Fuzzy Linear Discriminant Analysis
fMRI: functional Magnetic Resonance Imaging
FT: Fourier Transform
GA: Genetic Algorithm
GARF: Relative Fitness scaling Genetic Algorithm
HGA: Hybrid Genetic Algorithm
IBPSO: Improved Binary Particle Swarm Optimization
ICA: Independent Component Analysis
ICU: Intensive Care Unit
 k NN: k Nearest Neighbour
LDA: Linear Discriminant Analysis
LFDA: Local Fisher Discriminant Analysis
LLE: Locally Linear Embedding
LMS: Least-Mean-Squares
LMV: Local Mean Vector
LPP: Locality Preserving Projections
LSDA: Locality Sensitive Discriminant Analysis
LVQ: Linear Vector Quantization
MEG: Magnetoencephalography
MI: Mutual Information
MIFE: Mutual Information Feature Evaluation
MIFS: Mutual Information Feature selection
 m k NN: Modified k Nearest Neighbour
ML: Machine Learning
MLP: Multi-Layer Perception
MRI: Magnetic Resonance Imaging
MSE: Mean Square Error
NCA: Neighbourhood Components Analysis
ND: Negative Distribution Factor
NIRS: Near-Infrared Spectroscopy
NN: Neural Network
NNE: Neural Network Ensemble
NoS: Number of Evaluated Subsets
NPE: Neighbourhood Preserving Embedding
NREM: Non Rypid Eye Movement
NW k NN: Neighbourhood Weighting k Nearest Neighbour
OLSFDA: Orthogonal Locality Sensitive Fuzzy Discriminant Analysis
PCA: Principal Component Analysis
PD: Positive Distribution Factor

PET: Positron Emission Tomography
PkNN: Pseudo k Nearest Neighbour
PSD: Power Spectral Density
PSG: Polysomnogram
PSO: Particle Swarm Optimization
PTA: Plus Take-Away
R&K: Rechtschaffen and Kales Rules
RBF: Radial Basis Function
REM: Rypid Eye Movment
RLS: Recursive-Least-Squares
ROC: Receiver Operating Characteristics curve
SA: Simulated Annealing
SBS: Sequential Backward Search
SCI: Spinal Cord Injuriy
SCP: Slow Cortical Potentials
sEMG: surface Electromyogram
SFFS: Sequential Forward Floating Search
SFS: Sequential Forward Search
SGA: Simple Genetic Algorithm
SOFM: Self-Organizing Feature Maps
SSVEP: Steady-State Visual Evoked Potentials
STFT: Short-Time Fourier Transform
SVD: Singular Value Decomposition
SVM: Support Vector Machine
SWS: Slow-Wave Sleep
TMS: Transcranial Magnetic Stimulation
TS: Tabu Search
TSP: Travel Saleman Problem
UCI: University of California, Irvine
uLDA: uncorrelated Linear Discriminant Analysis
WGSSE: Within the Generalized Group Sum of Squared Error
wkNN: Weighted k Nearest Neighbour
wmkNN: Weighted Modified k Nearest Neighbour
WPD: Wavelet Packet Decomposition
WPT: Wavelet Packet Transform
WT: Wavelet Transform
WTL: Win-Tie-Loss
WVD: Wigner-Ville Distribution

Contents

ABSTRACT	iv
ABBREVIATIONS	vi
LIST OF FIGURES	xiii
LIST OF TABLES	xvii
1 INTRODUCTION	1
1.1 Background about EEG-based Applications	3
1.2 Pattern Recognition using EEG Data	4
1.3 Research Objectives	8
1.4 Organization of the Thesis	9
1.5 Publications Arising Directly from the PhD Research	10
2 BACKGROUND AND LITERATURE REVIEW	12
2.1 Introduction	12
2.2 The Electroencephalogram, EEG	12
2.2.1 Brain anatomy	13
2.2.1.1 Nerve cells and action potential	14
2.2.2 Characteristics of EEG recording	16
2.2.3 EEG rhythms	19
2.2.4 Factors affecting EEG signal	20
2.2.5 Other instruments measuring the brain's activity	21
2.2.6 Advantages of EEG	22
2.2.7 EEG-based applications	23
2.2.7.1 Brain computer interface	23
2.2.7.2 Sleep scoring system	26

2.2.8	EEG Supporting Signals	28
2.2.8.1	Electrocardiography	28
2.2.8.2	Electromyogram	29
2.2.8.3	Electrooculography	29
2.3	Signal Analysis	30
2.3.1	General signal characteristics	31
2.3.2	Segmentation of EEG signals	31
2.3.3	Feature extraction from EEG	32
2.3.3.1	Time-domain feature extraction	34
2.3.3.2	Time-frequency domain feature extraction	36
2.3.3.3	Fourier transform	36
2.3.3.4	Auto-regressive linear model	38
2.3.3.5	Wavelet transform	39
2.3.4	Dimensionality reduction in EEG	43
2.3.4.1	The curse of dimensionality	44
2.3.4.2	Feature selection strategies	45
2.3.4.3	Stepwise search	46
2.3.4.4	Genetic algorithm	47
2.3.4.5	Simulated annealing	49
2.3.4.6	Tabu search	51
2.3.4.7	Ant colony optimization	52
2.3.4.8	Differential evolution	55
2.3.4.9	Particle swarm optimization	57
2.3.5	Feature projection	59
2.3.6	Classification	61
2.3.6.1	Classifier combination	66
2.3.6.2	Existing classifier fusion methods	67
2.4	Summary	69
3	BIO-INSPIRED FEATURE SELECTION IN EEG	70
3.1	Introduction	70
3.2	The Need for Feature Selection	71
3.3	Selecting Best Channels	71
3.3.1	GA representation in BCI problem	73
3.3.2	Initial channel selection based on literature	75
3.3.2.1	Improved channel selection through GA	76
3.3.2.2	Selecting channels versus features	77
3.3.2.3	Motor versus non-motor channels	78
3.4	Comparison of Search Algorithms Performance on BCI Data	79
3.4.1	Sequential search	80

3.4.2	Simulated annealing	80
3.4.3	Tabu search	82
3.4.4	Accuracy and computational complexity	83
3.5	Improved Genetic Algorithm Diversity	84
3.5.1	Proposed GA method	86
3.5.2	Experimental results on small datasets	88
3.5.3	Experimental results on highly redundant data	89
3.5.4	Experimental results on EEG datasets	92
3.5.5	Experimental results on large datasets	92
3.6	Enhanced Feature Selection Algorithm Using ACO	96
3.6.1	Local measure estimation in ACO	97
3.6.1.1	Mutual information estimation	97
3.6.1.2	Improved local measure estimation	98
3.6.2	Experimental results on highly redundant data	101
3.6.3	Experimental results on EEG datasets	102
3.7	A Novel Swarm Based Feature Selection Algorithm	106
3.7.1	Particle swarm optimization	106
3.7.2	The proposed binary PSO algorithm	107
3.7.3	Experimental results	111
3.8	Selection of Relevant Feature Subsets using Differential Evolution	116
3.8.1	Selection of subsets using an upper subset size limit	120
3.8.2	Experimental results	121
3.9	Comparison of Proposed Methods	132
3.10	Summary	134
4	DIMENSIONALITY REDUCTION THROUGH FEATURE PROJECTION IN EEG	135
4.1	Introduction	135
4.2	The Need for Feature Projection	136
4.3	Principal Component Analysis	137
4.4	Linear Discriminant Analysis	138
4.4.1	Fuzzy LDA	140
4.4.2	Other projection variants	140
4.5	Orthogonal Locality Sensitive Fuzzy Discriminant Analysis	141
4.6	Experimental Results	147
4.6.1	Experiment on sleep data	148
4.6.2	Additional experiments	154
4.7	Conclusion	161
5	CLASSIFICATION AND CLASSIFIER FUSION	162
5.1	Introduction	162

5.2	The k NN Classifier Variants	163
5.2.1	Weighted k NN	164
5.2.2	Pseudo nearest neighbour	164
5.2.3	Class-weighted k NN	164
5.2.4	Fuzzy k NN	165
5.2.5	Neighbourhood components analysis	166
5.2.6	Other k NN variants	166
5.3	General Description of the Proposed Algorithms	167
5.4	Improving k NN Classification	168
5.4.1	Optimizing k NN using weighted modified measure	168
5.4.1.1	Favouring a specific class	169
5.4.1.2	Experimental results on UCI datasets	171
5.4.1.3	Experimental results on EEG datasets	175
5.4.2	Optimizing k NN using differential evolution	176
5.4.2.1	Experimental results	178
5.4.2.2	Experimental results on EEG dataset	183
5.5	Enhanced Neural Network Ensemble	184
5.5.1	The proposed method	186
5.5.2	Experiments	187
5.6	Summary	191
6	SUMMARY AND FUTURE RESEARCH	192
6.1	Thesis Summary	192
6.2	Recommendation for Future Research	194
A	DATASETS DESCRIPTION	196
A.1	EEG Datasets Description	196
A.1.1	EEG1-dataset	196
A.1.1.1	Feature extraction	198
A.1.2	EEG2-dataset	199
A.1.2.1	Feature extraction	199
A.1.3	EEG3-dataset	199
A.1.3.1	Feature extraction	200
A.1.4	Sleep scoring dataset	200
A.2	Ethical Approvals	203
	Bibliography	204

List of Figures

1.1	BCI pattern recognition system.	4
1.2	10-20 International EEG system.	5
2.1	Human brain.	13
2.2	The lobes of a cerebral hemisphere.	15
2.3	Brain functions.	15
2.4	Neuron.	16
2.5	Resting and action potentials.	17
2.6	EEG recording	17
2.7	EEG rhythms.	18
2.8	Sleep cycle	26
2.9	Typical ECG waveform.	29
2.10	EMG recording.	29
2.11	EOG recording.	30
2.12	Pattern recognition.	31
2.13	Energy of a signal.	35
2.14	Daubechies mother wavelet.	40
2.15	Wavelet: a) Position b) Scale.	41
2.16	Level 3 decomposition using wavelets transform.	42
2.17	Level 3 decomposition using wavelet packets transform.	42
2.18	Crossover.	49
2.19	Mutation.	49
2.20	ICA in EEG.	60
2.21	Orthogonal projection in PCA.	60
2.22	SVM presentation in two-dimensional space.	62
2.23	Single neuron model.	64
2.24	Approaches to building classifier ensembles.	67
3.1	Increase of search spaces size by choosing more channels.	72

3.2	Roulette wheel selection.	74
3.3	Classification accuracy of the selected channels for the training and test sets.	77
3.4	Classification accuracy of the selected channel and feature/channel subset.	78
3.5	Position of Motor and Non-Motor EEG electrodes.	79
3.6	Classification accuracy of selected features from Motor and Non- Motor channels.	80
3.7	Sequential search classification accuracy of selected channels/features.	81
3.8	SA classification accuracy of selected channels/features.	82
3.9	TS classification accuracy of selected channels/features.	82
3.10	Comparison between different methods in all channels.	83
3.11	DGA flow chart.	87
3.12	Convergence of different methods when selecting 5 features from Madelon.	91
3.13	Classification performance of EEG1a-dataset.	93
3.14	Classification performance for different datasets.	95
3.15	Classification performance for different datasets.	96
3.16	Membership function: a) Fuzzy b) Crisp.	99
3.17	Classification accuracy for Madelon dataset.	102
3.18	Convergence results on EEG1-datasets.	103
3.19	Convergence results on EEG2-dataset.	104
3.20	BPSOMI forming a reduced feature space.	108
3.21	BPSOMI forming a reduced feature space.	110
3.22	A flowchart of the proposed BPSOMI algorithm.	111
3.23	Madelon results using BPSOMI.	112
3.24	EEG1a-dataset results using BPSOMI.	113
3.25	Convergence results on EEG1-datasets.	114
3.26	EEG2-dataset convergence results using BPSOMI.	115
3.27	Wheels constructed to select subsets of size 4	117
3.28	Flowchart of the proposed DEFS _W algorithm.	119
3.29	Wheels constructed to select subsets of size 6 or less.	121
3.30	Performance of constrained/non-constrained selection methods on Madelon dataset	123
3.31	EEG1a-dataset results using selected number of features.	124
3.32	Convergence results on EEG1b-datasets.	125
3.33	Average classification accuracies for each subset size across datasets 1	127
3.34	Average classification accuracies for each subset size across datasets 2	128
3.35	Convergence of different feature selection methods on different datasets.	129
3.36	Size of the best subset among all population member for each iteration.	130
3.37	Maximum, minimum and best subset size of the DEFS _W algorithm on 9 Tumor dataset.	131

3.38	Average k NN classification accuracies achieved by different feature selection methods	132
4.1	LDA: Two-dimensional, two-class data projected on \mathbf{w}	139
4.2	Average classification accuracy achieved using different projection and classification methods	150
4.3	Average classification accuracy achieved using different fuzzy discriminant analysis methods	153
4.4	Average class-wise accuracy across ten subjects using different methods. . .	154
4.5	k NN classification accuracy using different projection and classification methods on UCI datasets.	156
4.6	LDA classification accuracy using different projection and classification methods on UCI datasets.	157
4.7	LIBLINEAR classification accuracy using different projection and classification methods on UCI datasets.	158
4.8	k NN classification accuracy using different projection and classification methods on EEG datasets.	159
4.9	LDA classification accuracy using different projection and classification methods on EEG datasets.	160
4.10	LIBLINEAR classification accuracy using different projection and classification methods on EEG datasets.	161
5.1	Memberships of the inputs, Q_j and F_{av} , and output δ of the FIS.	171
5.2	ROC curves for different classifiers.	171
5.3	Rules of the fuzzy inference system.	171
5.4	ROC curves of the different methods.	175
5.5	Average class-wise accuracy for sleep datasets using different k NN variants. .	176
5.6	Generating synthetic data using MATLAB.	179
5.7	Scenario 1: Synthetic data	179
5.8	Scenario 2: Synthetic data	180
5.9	Results on balanced synthetic data with same class distribution using different k NN variants.	181
5.10	Results on unbalanced synthetic data with different class distribution using different k NN variants.	183
5.11	Average class-wise accuracy of different k NN variants for unbalanced EEG datasets.	184
A.1	Recording paradigm for EEG1-dataset.	197
A.2	Electrode placement on EEG1-dataset.	197
A.3	Wavelet packet decomposition of the 56 EEG channels.	198
A.4	Different classes of hand movements that the user imagined.	199

A.5 Placement of electrodes for polysomnography 202

List of Tables

2.1	Machine learning tasks.	31
3.1	Performance of the selected subsets.	76
3.2	Compromised classification accuracy versus time for channel search.	84
3.3	Compromised classification accuracy versus time for feature search.	85
3.4	Description of the datasets employed.	89
3.5	Convergence speed of GA variants.	90
3.6	Search space size and time saving.	90
3.7	Gene dataset description.	94
3.8	Statistical significance test results of ACOFUZZY	102
3.9	Statistical significance test results on EEG1-Dataset.	104
3.10	Statistical significance test results on EEG2-Dataset.	105
3.11	Statistical significance test results of BPSOMI.	114
3.12	Statistical significance test results of DEFS _W	125
3.13	Description of the datasets employed.	125
3.14	Statistical significance test results of DEFS	129
3.15	Comparison of proposed methods	133
4.1	Geometric mean error rates.	151
4.2	One-way analysis of variance test results.	152
4.3	Geometric mean error rates of OLSFDA	153
4.4	One-way analysis of variance test results of OLSFDA	153
5.1	UCI dataset description.	172
5.2	Average class-wise accuracy for a selected UCI dataset	173
5.3	Class-wise classification accuracy for different Fav values.	174
5.4	Area under the curve of the different methods.	175
5.5	Results on selected UCI datasets.	182
5.6	Performance of the different fusion methods	189

5.7	Performance of the different fusion methods on the sleep scoring data. . . .	190
5.8	One-way analysis of variance test results on the sleep scoring data.	191
A.1	Number of sleep stage epochs for each subject (judged by expert).	202