OP
435
S56
3009

SIGNAL ANALYSIS OF SLEEP ELECTROOCULOGRAPH

Peyman Shokrollahi, B.Sc.
Shiraz University, Iran, 2001

A thesis
presented to Ryerson University
in partial fulfillment of
Master of Applied Science
in the program of
Electrical and Computer Engineering

Toronto, Ontario Canada, 2009

©Peyman Shokrollahi, 2009

PROPERTY OF
RYERSON UNIVERSITY LIBRARY

Author's Declaration

I hereby declare that I am the sole author of this thesis.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Signature

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individual for the purpose of scholarly research.

Signature

Abstract

Signal Analysis of Sleep Electrooculogram

©Peyman Shokrollahi, 2009

**Master of Applied Science
Electrical and Computer Engineering
Ryerson University**

Measures of sleep physiology, not obvious to the human eye, may provide important clues to disease states, and responses to therapy. A significant amount of eye movement data is not attended to clinically in routine sleep studies because these data are too long, about six to eight hours in duration, and they are also mixed with many unknown artifacts usually produced from EEG signals or other activities. This research describes how eye movements were different in depressed patients who used antidepressant medications, compared to those who did not.

The goal is to track antidepressant medications' effects on sleep eye movements. Clinically used SSRIs such as Prozac (Fluoxetine), Celexa (Citalopram), Zoloft (Sertraline), the SNRI Effexor (Venlafaxine) have been considered in this study to assess the possible connections between eye movements recorded during sleep and serotonin activities. The novelty of this research is in the assessment of sleep eye movements, in order to track the antidepressant medications' effects on the brain through EOG channels. EOG analysis is valuable because it is a noninvasive method, and the following research is looking for findings that are invisible to the eyes of professional clinicians.

This thesis focuses on quantifying sleep eye movements, with two techniques: autoregressive modeling and wavelet analysis. The eye movement detection software (EMDS) with more than 1500 lines was developed for detecting sleep eye movements. AR coefficients were derived from the sleep eye movements of the patients who were exposed to antidepressant medications, and those who were not, and then they are classified by means of linear discriminant analysis. Also for wavelet analysis, discrete wavelet coefficients have been used for classifying sleep eye movements of the patients who were exposed to medication and those who were not.

Acknowledgment

I would like to acknowledge my parents and my family. I would not have been able to pursue my goals without their unyielding support and encouragement. I would particularly like to thank my father, Hadi Shokrollahi, my mother, Shohreh Kamali, and both my sisters, Elnaz and Mehrnaz, and especially my father for his comments and suggestions on editing this thesis.

I am deeply grateful to Dr. Sridhar Krishnan and Dr. Kristiina McConville, professors who have always been my true mentors and role models. Without their help and encouragement this Masters' project could never have been undertaken and completed. I am also grateful to Dr. Brian J. Murray for his clinical advice. He outlined the goal of project and provided the data.

I would like to thank the rock of my life for the last two years, my friend, and my love, Parisa Gavahi. She also helped me with plotting the mother wavelet figures of Chapter 2.

I am forever grateful to my mentor, Master Hassan Yousefzamani, who always inspired me with love, devotion, and generosity.

I would like to thank my fellow colleague from the Signal Analysis Research (SAR) group, Dr. Karthi Umapathi, for his enlightening and informative discussions with me about my research and Ms. Shillan Jabbar for plotting the figures of eyes and facial electrode placement of Chapter 1. I would also like to thank Dr. Robert Roseberry, Mr. Christopher Brierley, and Mr. Blake B. Walker for their suggestions on improving the style of this thesis.

I would also like to thank the many staff members of Ryerson University and Sunnybrook Health Sciences Centre, for providing me an enhanced educational and clinical experience. In particular I shall never forget the assistance provided by Dr. Mark I. Boulos and Mr. Dana Jewell.

Dedication

To my father, Hadi, my mother, Shohreh, both my sisters, Elnaz, Mehrnaz, my love, Parisa, and my musical mentor, Master Hassan Yousefzamani, for their love, support and encouragement.

Contents

1	Introduction	1
1.1	Sleep	2
1.1.1	Sleep Recording	3
1.1.2	Sleep Stages	4
1.1.3	EOG	5
1.2	Neurotransmitter Function in Controlling Mood	7
1.3	Depression and Risk Factors in Sleep Deprivation	9
1.4	Antidepressant Medications	10
1.4.1	Antidepressant Medication Development	10
1.4.2	SSRIs and SNRIs Mechanism of action	11
1.5	Eye Movement	13
1.6	Motivation	17
1.7	Contribution	17
1.8	Outline of the Thesis	18
2	Autoregressive Modelling and Wavelet Analysis Theories	20
2.1	The Uncertainty Principle and Heisenberg Rectangles	21
2.1.1	Scalogram	22
2.2	Wavelet Analysis	23
2.2.1	Signal Classes and Signal Space Analysis	24
2.2.2	Scaling Function Analysis	25
2.2.3	The Wavelet Function and Wavelet Vector Spaces	28
2.2.4	Continuous Wavelet Transform	29
2.2.5	Discrete Wavelet Transform	30
2.2.6	Log-Gabor Wavelet	32
2.3	All-Pole and Autoregressive Modelling	33
2.3.1	Burg Algorithm	35
2.4	Pattern Classification	37
2.4.1	Linear Discriminant Analysis	38
2.5	Receiver Operating Characteristic (<i>ROC</i>) Curve	44

3	Sleep Eye Movement Detection and Autoregressive Modelling	52
3.1	Eye Movement Detection	53
3.1.1	Eye Movement Detection, Algorithm 1	55
3.1.2	Eye Movement Detection, Algorithm 2	59
3.2	Autoregressive (AR) Modelling	63
3.2.1	EOG Segmentation	64
3.2.2	Model Order Optimization	68
3.3	Classification Results	73
3.3.1	AR Classification of Patients Exposed/Unexposed to Prozac	73
3.3.2	AR Classification of Patients Exposed/Unexposed to Celexa and Effexor	77
3.3.3	Receiver Operating Characteristic (<i>ROC</i>) Curve, Method I	82
3.3.4	Receiver Operating Characteristic (<i>ROC</i>) Curve, Method II	84
4	Time, Frequency, and Wavelet Analysis for Quantifying Sleep Eye Move- ments	87
4.1	Time-Domain Analysis	88
4.1.1	Non-stationarity of the EOG Signals and Its Solution	96
4.2	Frequency-Domain Analysis	99
4.3	Time and Frequency Analysis	99
4.4	Scalogram Analysis	101
4.4.1	Discussion	108
4.5	Slow Eye Movement Frequency Analysis Using Discrete Wavelet Transform	110
4.5.1	SEM Analysis with Consecutive Fixed Segmentation	110
4.5.2	SEM Analysis using EMDS	117
4.6	Feature Extraction Using DWT	119
4.6.1	Feature Extraction Using DWT Matrix, without Grouping	119
4.6.2	Feature Extraction Using DWT Matrix, by Grouping	126
4.6.3	Discussion	130
5	Conclusions	136
5.1	The Contributions	137
5.1.1	Clinical Contributions	138
5.1.2	Technical Contributions	138
5.2	Future Work	146
	Bibliography	148
	A Fourier Analysis	164
	B Gabor Function	167

List of Figures

1.1	Polysomnography Including EEG, EOG, EMG and Other Channels	3
1.2	The Polarity of the Eye	6
1.3	Three channels of EOG recordings; LOC, ROC, and VOC with Mastoid Channels A1 and A2.	7
1.4	The Mastoid channel, A2.	8
1.5	The Mastoid channel, A1.	8
1.6	The Procedure of the Communication between two Neurons by Neurotransmitters	11
1.7	The Leftward Eye Movement	13
1.8	The Rightward Eye Movement	13
1.9	The Upward Eye Movement	14
1.10	The Downward Eye Movement	14
1.11	The EOG Channels for Leftward Eye Movement (x-axis is time, samples, and y-axis is voltage, volts)	15
1.12	The EOG Channels for Rightward Eye Movement (x-axis is time, samples, and y-axis is voltage, volts)	15
1.13	The EOG Channels for Upward Eye Movement (x-axis is time, samples, and y-axis is voltage, volts)	16
1.14	The EOG channels for Downward Eye Movement (x-axis is time, samples, and y-axis is voltage, volts)	16
1.15	Organization of the thesis	19
2.1	Heisenberg Box	23
2.2	The Scaling and Wavelet Space	27
2.3	Coiflet Mother Wavelet of order 3 (x-axis is time, samples, and y-axis is amplitude, volts)	30
2.4	Daubechies Mother Wavelet of Order 4	31
2.5	Sub-band decompositions of discrete wavelet transform. $g[.]$ is the high pass filter. $h[.]$ is the low pass filter.	32
2.6	Two-dimensional feature vectors of two classes C_1 and C_2 . The vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ are the prototypes of the two classes. The solid line that represents the decision boundary is perpendicular to the straight dashed line, joining the two classes of prototypes.	38

2.7	A simple linear classifier with n input units and $n + 1$ weight vector	39
2.8	The hyperplane H which separates the feature space into R_1 and R_2	41
2.9	Two probability density functions of a quantity on which decisions are based with a possible threshold	47
2.10	Receiver Operating Characteristic (<i>ROC</i>) curve shows three possible points .	49
2.11	The comparison between the <i>ROC</i> curves of two tests	50
3.1	Methodology Used for the Sleep Eye Movement Classification of Patients Exposed/Unexposed to Antidepressant Medication	52
3.2	Examples of Rightward and Leftward Eye Movements	54
3.3	Examples of Upward and Downward Eye Movements	54
3.4	The Block Diagram of Eye Movement Detection Procedure	55
3.5	A Typical Hypnogram	56
3.6	The Verification Plot of a Leftward Eye Movement	61
3.7	The Verification Plot of a Rightward Eye Movement	61
3.8	The Verification Plot of an Upward Eye Movement	62
3.9	The Verification Plot of a Downward Eye Movement	62
3.10	The time duration of the sleep eye movements in samples (patient A1, Stage 3, y-axis unit is in volt)	65
3.11	The time duration of the sleep eye movements in samples (patient A1, Stage R, y-axis unit is in volt)	66
3.12	The time duration of the sleep eye movements in samples (patient C3, Wakefulness, y-axis unit is in volt)	67
3.13	The error plot (patient A1, Stage 1, leftward eye movement)	69
3.14	The error plot (patient A1, Stage2, rightward eye movement)	70
3.15	The AIC curves	71
3.16	The AIC curve for a patient who was exposed to Celexa derived from leftward movements, LOC, with segmentation of 400 samples	72
3.17	The AIC curve for a patient who exposed to Celexa derived from leftward movements, Stage 1, LOC, with segmentation of 500 samples	72
3.18	The classification results of LOC leftward AR coefficients	75
3.19	The classification results of ROC leftward AR coefficients	76
3.20	The classification results of LOC rightward AR coefficients	76
3.21	The classification results of ROC rightward AR coefficients	77
3.22	Distribution of three AR coefficients	78
3.23	Group 1 leftward eye movement AR coefficient accuracies	79
3.24	Group 1 rightward eye movement AR coefficient accuracies	79
3.25	Group 2 leftward eye movement AR coefficient accuracies	80
3.26	Group 2 rightward eye movement AR coefficient accuracies	80
3.27	ROC Curve with Method I	85
3.28	Receiver Operating Characteristic (<i>ROC</i>) Curve with Method II	86

4.1	The crosscorrelation of LOC and Mastoid signal (A1) derived from a patient N4 who was not exposed to antidepressant medication in Stage REM (x-axis is time in samples and y-axis is the crosscorrelation value).	89
4.2	The crosscorrelation of LOC and Mastoid signal (A2) derived from a patient N4 who was not exposed to antidepressant medication in Stage REM (x-axis is time in samples and y-axis is the crosscorrelation value).	90
4.3	The crosscorrelation of ROC and Mastoid signal (A1) derived from a patient N4 who was not exposed to antidepressant medication in Stage REM (x-axis is time in samples and y-axis is the crosscorrelation value).	90
4.4	The crosscorrelation of ROC and Mastoid signal (A2) derived from a patient N4 who was not exposed to antidepressant medication in Stage REM (x-axis is time in samples and y-axis is the crosscorrelation value).	91
4.5	The maximum crosscorrelation of the LOC, ROC, and VOC compared with Mastoid signals (A1 and A2) for five patients who were not exposed to antidepressant medication (N) and five patients who were exposed to Prozac. .	92
4.6	The maximum crosscorrelation of the LOC compared with Mastoid signals (A1 and A2) for five patients who were not exposed to antidepressant medication (N) and five patients who were exposed to Prozac, in different sleep stages. .	93
4.7	The mean value and variance of LOC, ROC, VOC, and Mastoids (A1 and A2) for Normal subjects (five patients who were not exposed to antidepressant medication) and Abnormal subjects (five patients who were exposed to Prozac medication)	94
4.8	The leaning curve of the RLS algorithm for an EOG signal (x-axis is time in samples and y-axis is error).	98
4.9	The EOG signal for a patient exposed to Prozac medication has been plotted in time-domain and in frequency-domain. The spectrum of the signal from 0 to 5 Hz is shown in the middle one. The bottom figure shows a complete spectrum of the signal. (Maximum frequency = 64Hz; the y-axis unit is in volt)	100
4.10	The amplitude scalogram of 5-second EOG signal (the sample is the unit of x-axis).	101
4.11	The amplitude scalogram of five second EOG signal. The energy distribution measurement has been illustrated in the right column of the graph starting from blue (0) to red dark (0.016)	102
4.12	The leave-one-out method classification accuracy results compared among stage energies of three unexposed patients derived from the scalogram. . . .	103
4.13	The leave-one-out method classification accuracy results compared among three stage energies of five unmedicated patients derived from the scalogram.	104
4.14	The leave-one-out method classification accuracy results compared among three stage energies of five patients exposed to Effexor medication, derived from the scalogram.	106

4.15	The leave-one-out method classification accuracy results compared among three stage energies of five patients exposed to Celexa medication.	107
4.16	The leave-one-out method classification accuracy results compared among three stage energies of three patients exposed to Prozac medication.	108
4.17	The mean value of accuracy results for patients exposed (E, L, and P) and unexposed (C) to medication.	109
4.18	The mean value of accuracy results for patients exposed and unexposed to medication (without patient C3).	110
4.19	The block diagram of the EOG wavelet decomposition with Daubechies of order 4 into 10 levels for classifying those who were exposed to antidepressant medication and those who were not.	111
4.20	An example of a SEM, detected between the marked points in LOC, ROC, and ROC – LOC.	112
4.21	The SEM signal and its 10-level decomposition. The signal duration is 10 seconds. The signal decomposed with Daubechies of order 4. The signal is ROC minus LOC of Stage 1, belonging to patient C1, who is not exposed to antidepressant medication. (D_1 is in the range of 10^{-5} , and D_2 and D_3 are in the range of 10^{-3})	113
4.22	Levels 8 – 10 details classification results derived from slow movement detection method	115
4.23	Level 8 to level 10 Details classification accuracy results, derived from slow movement detection method using the EMDS.	117
4.24	Details (D_4 to D_8) and approximation (A_8) of leave-one-out method accuracies for different features in which the classification performed over Stage 1, LOC, leftward movements in patients exposed/unexposed to antidepressant medication.	123
4.25	Details (D_4 to D_8) and approximation (A_8) the regular method accuracies for different features in which the classification performed over Stage 1, LOC, leftward movements in Normal/Abnormal set of data.	124
4.26	The leave-one-out method accuracies for different features in which the classification performed for details (D_4 to D_8) and approximation (A_8) of Stage 1, LOC, leftward movements in Normal/Abnormal dataset.	126
4.27	The leave-one-out method accuracies for different features in which the classification performed for details (D_4 to D_8) and approximation (A_8) of Stage 2, LOC, leftward movements in Normal/Abnormal dataset.	127
4.28	The leave-one-out method accuracies for different features in which the classification performed for details (D_4 to D_8) and approximation (A_8) of Stage 3, LOC, leftward movements in Normal/Abnormal dataset.	127
4.29	The leave-one-out method accuracies for different features in which the classification performed for details (D_4 to D_8) and approximation (A_8) Stage REM, LOC, leftward movements in Normal/Abnormal dataset.	128

4.30	The regular method accuracies for different features in which the classification performed for details (D4 to D8) and approximation (A8) Stage 1, LOC, leftward movements in Normal/Abnormal dataset.	128
4.31	The regular method accuracies for different features in which the classification performed for details (D4 to D8) and approximation (A8) Stage 2, LOC, leftward movements in Normal/Abnormal dataset of data.	129
4.32	The regular method accuracies for different features in which the classification performed for details (D4 to D8) and approximation (A8) Stage 3, LOC, leftward movements in Normal/Abnormal dataset.	129
4.33	The regular method accuracies for different features in which the classification performed for details (D4 to D8) and approximation (A8) Stage REM, LOC, leftward movements in Normal/Abnormal dataset.	130
A.1	The signal compounds of two sinusoidal waveforms with frequencies of 20 Hz and 40 Hz with in Time and Frequency domain.	165
A.2	The signal compounds of two sinusoidal waveforms with frequencies of 20 Hz and 40 Hz with in Time and Frequency domain.	166
B.1	Gabor Filter Transfer function.	168

List of Tables

1.1	SSRI Medications	12
1.2	SNRI Medications	12
2.1	Decision Matrix	46
2.2	Definitions of the various Decision Performance	47
2.3	Definitions of the various Decision Performance	47
3.1	Eye Movement Decision Table	55
3.2	An example of the quantification of detected eye movements with different threshold and window size	59
3.3	Eye Movement Duration Examples	63
3.4	The classification results of LOC leftward AR coefficients	74
3.5	The classification results of ROC leftward AR coefficients	74
3.6	The classification results of LOC rightward AR coefficients	75
3.7	The classification results of ROC rightward AR coefficients	75
3.8	The leave-one-out method accuracies (%) of the classification of Group 1 leftward eye movements' AR coefficients	81
3.9	The leave-one-out method accuracies (%) of the classification of Group 1 rightward eye movements' AR coefficients	81
3.10	The leave-one-out method accuracies (%) of the classification of Group 2 leftward eye movements' AR coefficients	81
3.11	The leave-one-out method accuracies (%) of the classification of Group 2 rightward eye movements' AR coefficients	81
4.1	The Crosscorrelation Comparison	91
4.2	The Crosscorrelation Comparison of LOC and Mastoid Channels	95
4.3	The Crosscorrelation Comparison of ROC and Mastoid Channels	95
4.4	The Crosscorrelation Comparison of VOC and Mastoid Channels	96
4.5	The Crosscorrelation of Mastoid Signals with EOG for Normal Patients (those who were not exposed to Prozac medication)	96
4.6	The Crosscorrelation of Mastoid Signals with EOG for Abnormal Patients (those who were exposed to Prozac medication)	97
4.7	Mean Value Comparison	97
4.8	Variance Value Comparison	97

4.9	The classification accuracy results of all stage comparisons of maximum energies derived by scalogram analysis for patients who were unexposed to antidepressant medication	104
4.10	The accuracy results of maximum energy comparisons for Stage 3 with other stages for patients unexposed to antidepressant medication.	105
4.11	The accuracy results of maximum energy comparisons for Stage 3 with other stages, for patients who were exposed to Effexor medication.	106
4.12	The accuracy results of maximum energy comparisons for Stage 3 with other stages, for patients who were exposed to Celexa medication.	107
4.13	The accuracy results of maximum energy comparisons for Stage 3 with other stages, for patients who were exposed to Prozac medication.	108
4.14	The mean values of the accuracy results for patients unexposed (C) and exposed (E, L, and P) to medication (C3 is a patient who was not exposed to antidepressant medication).	109
4.15	The range of frequencies in each decomposition level	114
4.16	The leave-one-out method and the regular method accuracies for D8–D10 classification of SEMs detection method.	115
4.17	The leave-one-out method and the regular method accuracies for A10 and D8-D10 classification of SEMs detection method	116
4.18	The leave-one-out method and the regular method accuracies for D8–D10 classification of SEMs detection method	118
4.19	The range of frequencies in each decomposition level	121
4.20	The accuracy results of the leave-one-out method for four features performed over all entire matrix of wavelet coefficients (Abs=Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)	125
4.21	The accuracy results of the regular method for four features performed over all entire matrix of wavelet coefficients (Abs=Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)	125
4.22	The accuracy results of the leave-one-out method for four features performed over a group of five in a matrix of wavelet coefficients. The results were derived from Normal-Abnormal set, Stage 1. (Abs=Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)	131
4.23	The accuracy results of the leave-one-out method for four features performed over a group of five in a matrix of wavelet coefficients. The results were derived from Normal-Abnormal set, Stage 2. (Abs =Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)	131
4.24	The accuracy results of the leave-one-out method for four features performed over group of five in a matrix of wavelet coefficients. The results were derived from Norma-Abnormal set, Stage 3. (Abs=Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)	131

4.25	The accuracy results of the leave-one-out method for four features performed over a group of five in a matrix of wavelet coefficients. The results were derived from Normal-Abnormal set, Stage REM. (Abs =Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)	132
4.26	The accuracy results of the regular method for four features performed over a group of five in a matrix of wavelet coefficients. The results were derived from Normal-Abnormal set, Stage 1. (Abs =Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)	132
4.27	The accuracy results of the regular method for four features performed over a group of five in a matrix of wavelet coefficients. The results were derived from Normal-Abnormal set, Stage 2. (Abs =Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)	132
4.28	The accuracy results of the regular method for four features performed over a group of five in a matrix of wavelet coefficients. The results were derived from Normal-Abnormal set, Stage 3. (Abs =Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)	133
4.29	The accuracy results of the regular method for four features performed over a group of five in a matrix of wavelet coefficients. The results were derived from Normal-Abnormal set, Stage REM. (Abs =Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)	133

List of Acronyms

A - Abnormal; patient who has been exposed to depression medication
Abs - Absolute
AIC - Akaike Information Criterion
AUC - Area Under the Curve
Avg - Average
AR - Autoregressive
C - Control; patient who has not been exposed to depression medication
CWT - Continuous Wavelet Transform
D - Daubechies
Decom - Decompose
DWT - Discrete Wavelet Transform
E - Effexor; patient who has been exposed to Effexor medication
ECG - Electrocardiogram
EEG - Electroencephalogram
EMDS - Eye Movement Detection Software
EMG - Electromyogram
EOG - Electrooculogram
FT - Fourier Transform
FFT - Fast Fourier Transform
FN - False Negative
FNF - False Negative Fraction
FP - False Positive
FPF - False Positive Fraction
Hz - Hertz
IDWT - Inverse Discrete Wavelet Transform
IFT - Inverse Fourier Transform
L - Celexa; patient who has been exposed to Celexa medication
LDA - Linear Discriminant Analysis
LOC - Left Oculogram
MAOI - Monoamine Oxidase Inhibitor
MRA - Multiresolution Analysis
N - Normal; patient who has not been exposed to depression medication
NREM - Non-Rapid Eye Movement
P - Prozac; patient who has been exposed to Prozac medication
R - REM; Rapid Eye Movement; one of the sleep stages
REM - Rapid Eye Movement
RLS - Recursive Least Square
ROC - Right Oculogram
ROC - Receive Operating Characteristics
S - Stage; sleep stage

SEM - Slow Eye Movement
SNRI - Serotonin Norepinephrine Reuptake Inhibitor
SSRI - Selective Serotonin Reuptake Inhibitor
Std - Standard Deviation
STFT - Short Time Fourier Transform
TCA - Tricyclic Antidepressant
TN - True Negative
TNF - True Negative Fraction
TP - True Positive
TPF - True Positive Fraction
Val - Value
VOC - Vertical Oculogram
WT - Wavelet Transform

Chapter 1

Introduction

HUMAN beings spend about one-third of their lives sleeping, a necessary part of existence which includes a variety of unique physiological states. The suspension of consciousness and the experience of dreaming have been discussed among physicians, philosophers, and poets for centuries. Furthermore, in the last three decades, academic studies on sleep have found a large number of connections between sleep and physical and mental well-being. In this regard, neurologists and psychiatrists are more interested in sleep studies than other scientists because of the invaluable physiological information that exists in sleep data recordings, including electroencephalographic (EEG), electrooculographic (EOG), and electromyographic (EMG) recordings, captured in routine sleep studies. In order to have a better understanding of sleep data recordings, doctors and sleep experts receive extensive training to be sufficiently qualified in sleep data analysis. However, there is a wealth of physiological information that is not immediately apparent and is not readily available.

Signal acquisition technology is able to gather an extensive database of sleep recordings. EEG, EOG, ECG, EMG, and other channels record the electrical activity of the brain, eyes, heart, muscles, and other organs during sleep. These types of two-dimensional signals, representing voltage variations over time, show the physiological activity of different organs. Furthermore, each channel includes extensive information, and the study of each channel needs broad research. Thus, among all channels, EOG signals, which represent eye movement activity, have been chosen for this research.

During recent years, the medical industry has better understanding of sleep disorder problems. The medical industry has helped to alleviate sleep disturbances, reduce their side effects, and enhance medications to increase the duration of sleep. These improvements are reflected in the frequent publications about sleep medications, as well the availability of medications that may have effects on sleep. Antidepressant medications are among those which may have extensive effects. This research concerns antidepressant medications' effects on eye movements recorded in EOG signals during sleep.

This chapter is divided into following sections: sleep, neurotransmitter function in controlling mood, depression and risk factors in sleep deprivation; antidepressant medications, eye movement, motivation, contribution, and outline of the thesis.

1.1 Sleep

Sleep is known to be governed by the diencephalon, which is the area in the centre of the brain just above the brain stem. Sleep is characterized by cyclic, "reversible loss of consciousness, reduced sensory and motor functions linking the brain with the environment, internally generated rhythmicity and a restorative quality that cannot be achieved [otherwise]" [1].

In 1868, Greisinger discovered eye movements during sleep. Freud's suggestion, acting out of dreams, predicted the discovery of muscle paralysis during Rapid Eye Movement sleep. The report by Aserinsky and Kleitman explained the existence of "regularly occurring periods of eye motility during sleep" that were associated with dreaming. They described the occurrence of the repetitive cycles which are called rapid eye movements (REM) and non-rapid eye movement (NREM). In 1953, it was discovered that REM sleep has a cyclic pattern in sleep. Before this investigation, some investigators had suspected that two states of sleep existed [1]. These studies led to a solution for many sleep problems.

More than 50 million Americans suffer from some form of a sleep disorder. The disorder of sleep and wakefulness hinders daily functions, and affects health and longevity [2]. Sleep research is one of the topics in research by which the quality of human life can be improved. In this regard, a great deal of data analysis in sleep-related materials exists, such as time-

frequency analysis of sleep [3], sleep time duration [4], sleep apnea analysis [5] [6] [7], sleep EEG analysis [8] [9] [10], and sleep EEG-ECG analysis [11].

1.1.1 Sleep Recording

Sleep recordings are categorized into NREM and REM sleep state. Each state has unique physiological characteristics including: variations in brain wave patterns, eye movements, and muscle tone. Polysomnography (a sleep study) captures this information with EEG, EOG, and EMG recordings. While sleep clinicians use this information in clinical management, there is a wealth of information that is not apparent to casual visual inspection [12]. Figure 1.1 represents a sample of polysomnography including EEG, EOG, EMG, ECG and other channels.

Figure 1.1: Polysomnography Including EEG, EOG, EMG and Other Channels

Polysomnography is also used to assess normal and abnormal sleep patterns, sleep stages,

and wakefulness. Furthermore, characteristics of recorded data include much unknown information. Quantitative methods such as Fast Fourier Transform (FFT), wavelet transforms, and digital filters are also used for further assessment to characterize these biological signals. As a rule, a bio-signal changes in a very complicated manner, and it is difficult, even for specialists, to extract useful information directly from raw data. Therefore, a computer assisted method was required for further assessment of these data [13].

1.1.2 Sleep Stages

Sleep stages are defined by using EEG, EOG, and EMG features. In the 1960s, the need for standardized quantitative assessment of sleep to facilitate research and clinical evaluation led to the publication of the AASM manual for the scoring of sleep and associated events [2] that is still used today. Although sleep study is categorized as NREM and REM sleep, NREM sleep is classified into further categories, Stages 1-4, by the help of different polysomnographic analyses considering that each state has unique physiological characteristics including variations in brain wave patterns, eye movements, and muscle tone.

Most Prominent Features of The Stages

Stage categories for scoring epochs are as follows:

- Wakefulness
 - Recognized by existing low voltage and high frequency signals in EEG channels;
- Movement Time
 - Recognized by scoring epoch during which the EEG signal is obscured by movements of the subject;
- Stage 1
 - Recognized by a relatively higher voltage and lower frequency than wakefulness in EEG signals with slow eye movements;

- Stage 2
 - Recognized by a relatively higher voltage and lower frequency than Stage 1 in EEG signals with the presence of a special waveform in EEG called the K-complex and sleep spindles;
- Stages 3 and 4
 - Recognized by moderate amounts of high amplitude with slow wave activity, known as deep sleep;
- REM
 - Recognized by the presence of rapid eye movements in EOG channels with low voltage and high frequency waveforms in EEG;

Sleep scoring is one of the topics in which much research has been conducted. The following publications are important in this regard for automatic sleep scoring [14] [15] [16], Stage 2 scoring [17], scoring through non-invasive methods [18] [19], the classification of sleep stages [20] [21] [22], and sleep monitoring [23]. Furthermore, due to the importance of sleep data acquisition, one of the main topics in sleep research is instrumentation and sensor development. Also, various signal processing techniques such as filtering analysis [24], wavelet analysis [25], multifractal analysis [26], and autoregressive modelling [27] have been used in sleep literature. Most of the techniques mentioned have been performed on EEG signals.

1.1.3 EOG

EOG channels of polysomnography in which the signals were recorded for this study were considered. The EOG records the potential difference between the cornea and the retina. The eyeball is a dipole, with a positive pole at the cornea and a negative pole at the retina. Figure 1.2 represents the polarity of the eye.

Figure 1.2: The Polarity of the Eye

By the rotation of the eye globe, the cornea moves with respect to the electrodes placed near the eyes. If the eye (cornea) moves toward an electrode, it will induce a positive potential in the electrode and if the eye moves away, it will induce negative potential in the electrode. Therefore, for recording the horizontal and vertical eye movements the electrodes are placed diagonally [13].

Recordings of eye movements from three EOG and two reference channels with the routine polysomnographic recording were used in this work. The sampling rate is 128 samples per second. The AC channels are as follows:

- **LOC:** a signal that is recorded by the electrode placed 1cm below and 1cm left of the left eye of the subject;
- **ROC:** a signal that is recorded by the electrode placed 1cm above and 1cm to the right of the right eye of the subject;
- **VOC:** a signal that is recorded by an electrode placed 1cm below the middle of the right eye of the subject;
- **Mastoid Signals (A1, A2):** signals recorded by an electrode placed on a bony surface

and used as a reference.

To remove artifacts from EOG signals, processing and analysis are performed on signal differences between an EOG signal and references as follows: LOC–A2, ROC–A1, and VOC–A1, which are referred to as EOG signals in this research. Sleep staging was provided, using the routine scoring methods [2]. Figure 1.3, 1.4, and 1.5 demonstrate the placements of the electrodes on the subject. EOG signals recorded with this configuration at the sleep lab of Sunnybrook Health Science Centre were used in this research.

Figure 1.3: Three channels of EOG recordings; LOC, ROC, and VOC with Mastoid Channels A1 and A2.

1.2 Neurotransmitter Function in Controlling Mood

Figure 1.6 shows how an action potential moves from one neuron to another, and how two neurons communicate with each other through neurotransmitters. A neurotransmitter molecule communicates (information) through synaptic cells. It exchanges the signals from a presynaptic neuron to the postsynaptic neuron. Transmitters are packed in vesicles docking at the presynaptic terminals. At arrival, the electric impulse, called action potential,

Figure 1.4: The Mastoid channel, A2.

Figure 1.5: The Mastoid channel, A1.

dumps the neurotransmitter in the docked vesicles into the synaptic cleft, the gap between presynaptic and postsynaptic neuron. The receptors on the postsynaptic neuron receive the neurotransmitters [28].

Neurotransmitters play a great role in controlling the mood regularities. The irregularities in their transmission can cause several diseases in humans, such as depression.

1.3 Depression and Risk Factors in Sleep Deprivation

Depression is linked to lower levels of neurotransmitters, especially serotonin. Many organizations and educational programs have attempted to increase public awareness about depression. The US Depression Awareness Recognition and Treatment Program, the UK Defeat Depression Campaign, the Royal College of Psychiatrists, and the Royal College of General Practitioners-Health and Work Training are among these programs and organizations [29] [30]. Psychiatric depression is considered to be one of the more common and treatable mental disorders [31]. The statistics show that the the average lifetime risk of depression is between 4% to 12% for males and 12% to 26% for females [32] [33] [34]. Depression is more common among the elderly population, due to additional risk factors such as physical disability, lower income, cognitive impairment, and lack of social support [35] [34]. Furthermore, depression in elderly people is harder to manage if there is concurrent physical illness. Also, for many patients, it is required to maintain their long-term, even lifetime, antidepressant maintenance therapy [36]. Depression treatments with fewer side effects are required, and these effects are minimized when patients are well informed [34].

Most patients with depression have sleep disturbances that may be accompanied by excessive sleeplessness, known as insomnia. The characteristics of sleep disturbance in patients with depression includes short REM sleep latency on all-night polysomnography and an increased REM density. Depression can be treated in many ways; one of them is antidepressant medication such as Selective Serotonin Reuptake Inhibitors (SSRIs) and Serotonin Norepinephrine Reuptake Inhibitors (SNRIs), which will be discussed in the next section in detail [37].

1.4 Antidepressant Medications

Sleep disorders often require assessment in a sleep laboratory. They frequently coexist with other medical and psychiatric conditions (e.g. cardiovascular disease, depression, or diabetes) as well as different medication exposures (i.e.: harmful effects). Many sleep state changes have been seen in clinical depression (i.e reduced activity) and with respect to medication exposures [38]. While REM sleep eye movements have been of historical interest to clinicians, some interesting eye movements are noted in NREM sleep after exposure to antidepressant medications such as Prozac [39] [40]. A way to quantify sleep eye movement changes may help detect patterns of physiology and response to treatment that have not been apparent to clinicians so far. The main group of medications that have been used in this research are the SSRIs and SNRIs.

SSRIs are referred to as a pharmacologically homogeneous group. They represent only differences in serotonin reuptake inhibitory potency. On the other hand, they have different effects on neurotransmitter receptors and pharmacokinetic profiles, which is related to the rate of uptaking a drug [41].

1.4.1 Antidepressant Medication Development

In late 1950s, the first effective antidepressant medications, monoamine oxidase inhibitors (MAOIs) and the tricyclic antidepressants (TCAs), were launched to the pharmacies. However, these agents have been limited in clinical applications, due to their different side effects and “premature discontinuation of therapy” [34] [42]. Selective serotonin re-uptake inhibitors (SSRIs) is the new class of antidepressant medications which has become available in the past few years. They are as effective as TCAs with fewer side effects. Furthermore, it should be mentioned that they are not “perfect” antidepressants; they have side effects, they do not work immediately, and they do not work for all patients. However, they enhance the pharmacotherapy of depression significantly [34].

1.4.2 SSRIs and SNRIs Mechanism of action

Although SSRIs do not bind to any specific neuroreceptor system, they produce antidepressant effects through the selective blockade of serotonin re-uptake [43]. Re-uptake of serotonin, after its release into the cleft, stops the neuron activity (see Figure 1.6).

Figure 1.6: The Procedure of the Communication between two Neurons by Neurotransmitters

A serotonin re-uptake carrier is responsible for its re-uptake through the cellular membrane. Inhibition of the serotonin uptake carrier makes the serotonin concentration increase in the synapse [44].

By increasing the concentration of serotonin in the synaptic cleft through inactivation of presynaptic autoreceptors, serotonergic (related to neurotransmitter serotonin) turnover is reduced. On the other hand, by increasing the amount of synaptic serotonin, presynaptic serotonin autoreceptors become desensitized. This action is accompanied by increased serotonin release and enhancement of serotonergic neurotransmission. However, this outcome does not occur immediately, and takes about 14 days after starting a treatment [45]. This observation is pertinent, because of the “delayed onset of antidepressant efficacy” in most cases [34].

SNRIs comprise a new class of antidepressant medications. SNRIs are developed from SSRIs to affect not only serotonin, but also norepinephrine and other neurotransmitter systems. These medications have similar functions to SSRIs in which they inhibit the reuptake of neurotransmitters, which have an effect on mood [46].

Tables 1.1 and 1.2 show the generic and trade names of SSRI's and SNRI's antidepressant medications.

Table 1.1: SSRI Medications	
Trade Name	Generic Name
Prozac	Fluoxetine
Celexa	Citalopram
Zoloft	Sertraline
Paxil	Paroxetine

Table 1.2: SNRI Medications	
Trade Name	Generic Name
Effexor	Venlafaxine

By reviewing some of the advantages and disadvantages of SSRI medications, one can understand that the SSRIs are used for depression treatment, and they were developed to be as effective as reference medication, TCAs. One who is interested in this development can study Fluvoxamine(with trade name of Luvox) [47], Paroxetine(with trade name of Paxil) [48], Citalopram (with trade name of Celexa) [49], and Sertraline(with trade name of Zoloft) [50]. Also, advantages and disadvantages of SSRIs are discussed in [34]. Further, the SSRIs are effective in the relief of sleep disturbance, within the depression context [34] [51]. The SSRIs have other advantages and disadvantages which were not of concern to this research. For example, they potentiate the effects of alcohol or some other drugs. One may refer to [52] [53] for further information.

1.5 Eye Movement

The study of eye movement is a source of information to both scientists and clinicians. Neurobiologists are interested in the study of eye movement, because the inspection of eye movements provides an opportunity to understand brain functionality. Ophthalmologists and neurologists are also interested in eye movement study because they can find clues to the localization of disease in the brain. Over the past thirty years, eye movements have been considered as a source of insight into different brain disorders [54]. Furthermore, it should be considered that unconscious eye movements during sleep are also a good source of information for clinicians, in order to track disease states and responses to therapy.

This section talks about the basic elements of eye movements which are used in this thesis. Also, the publications which are mostly about eye movement analysis during wakefulness are considered.

Eye movements are categorized into four types of movements: leftward, rightward, upward and downward as shown in Figures 1.7, 1.8, 1.9, and 1.10.

Figure 1.7: The Leftward Eye Movement

Figure 1.8: The Rightward Eye Movement

Figure 1.9: The Upward Eye Movement

Figure 1.10: The Downward Eye Movement

Figures 1.11, 1.12, 1.13, and 1.14 represent the configuration of EOG channels for different types of movements.

Furthermore, eye movement measurement during wakefulness is one of the most important topics in many research areas. Different types of analysis have been applied to EOG signals, such as simple nonlinear deterministic processes [55], different types of algorithms, [56] [57], filtering [58], modelling of eye movements [59] [60]. Modelling of eye movements is categorized into special eye movements such as saccadic eye movements [61] [62] or vergence eye movements [63]. Machine learning [64] [65], analysis of eye movement dysfunctions, such as saccadic [66] [67] [68], vigilance [69], or quick phase [70], and human brain functionality [71] [72] [73] are important topics in the analysis of wakefulness eye movements. There are many publications on eye movement tracking [74] [75] [76] and removing eye movements from EEG as artifacts [77].

EOG is also important in human computer interactions [78] [79] [80]. Further, EOG is used directly in control systems [81]. Some of the methods are used for processing visual motion in animals [82] or for analysis of human eye [83]. Different kinds of control and

Figure 1.11: The EOG Channels for Leftward Eye Movement (x-axis is time, samples, and y-axis is voltage, volts)

Figure 1.12: The EOG Channels for Rightward Eye Movement (x-axis is time, samples, and y-axis is voltage, volts)

Figure 1.13: The EOG Channels for Upward Eye Movement (x-axis is time, samples, and y-axis is voltage, volts)

Figure 1.14: The EOG channels for Downward Eye Movement (x-axis is time, samples, and y-axis is voltage, volts)

instrumentation systems are also developed on the basis of eye movements [84] [85] [86]; one who is interested in this field may refer to [87] [88] [89].

1.6 Motivation

A significant amount of eye movement data is not attended to clinically in routine sleep studies because these data are too long, about six to eight hours in duration, and they are also mixed with many unknown artifacts usually produced from EEG signals or other activities. Measures of sleep physiology, not obvious to the human eye, may provide important clues to disease states, and responses to therapy. For example, it has been noted that eye movements change significantly in the patients exposed to antidepressant medications. This research describes how eye movements were different in depressed patients who used antidepressant medications, compared to those who did not.

The goal is to track antidepressant medications' effects on sleep eye movements. Clinically used SSRIs such as Prozac, Celexa, Zoloft, the SNRI Effexor have been considered in this study to assess the possible connections between eye movements recorded during sleep and serotonin activities. The novelty of this research is in the assessment of sleep eye movements, in order to track the antidepressant medications' effects on the brain through EOG channels. EOG analysis is valuable because it is a noninvasive method, and the following research is looking for findings that are invisible to the eyes of professional clinicians.

Discovering an association between the two classes of patients (exposed and unexposed to medication) is an important step in biomedical signal analysis applied to medication effects. However, why and how these two classes are related in sleep eye movements should not be limited to this research; they need further research in the future.

1.7 Contribution

The proposed contributions of this research are summarized as bellow:

- Eye Movement Detection Software (EMDS) (Chapter 3)

- Band Pass Filter (0.3-20 Hz)
- Verification Plots
- Returning Eye Movement Time Durations
- Returning Other Eye Movement Specifications Such as Time Index and Amplitude of Detected Movements
- Time Domain Analysis (Chapter 3)
 - Autoregressive Modelling
- Time-Frequency Domain Analysis (Chapter 4)
 - Wavelet Analysis

1.8 Outline of the Thesis

This thesis focuses on quantifying sleep eye movements, with different techniques, in order to extract information. Quantifying sleep eye movements that can measure the sleep physiology not obvious to the human eyes may provide important clues to disease states and responses to therapy.

To this end, Chapter 2 introduces the theories of applied techniques, namely autoregressive (AR) modelling and wavelet analysis. Chapter 2 also gives the reader some details in mathematics, which support the implemented methods of this research.

Chapter 3, focuses on quantifying sleep eye movements with AR modelling. In that chapter, the eye movement detection software (EMDS), which is developed for detecting sleep eye movements, first will be described, then Chapter 3 presents how AR coefficients, derived from the sleep eye movements of the patients who were exposed to antidepressant medications, and those who were not, are classified by means of linear discriminant analysis. To evaluate this method, a Receiver Operating Characteristic (*ROC*) curve has been developed.

Chapter 4, discusses the time, frequency, and time-frequency analysis of sleep eye movements. In this chapter, discrete wavelet coefficients have been used for classifying sleep eye

movements of the patients who were exposed to medication and those who were not. Chapters 3 and 4 discuss how the eye movement waveforms are different between patients who were exposed to antidepressant medication and those who were not.

Chapter 5, which is the last chapter, presents the conclusions and suggests future directions inspired by the presented research.

In summary, this work tracks the antidepressant medication effects on sleep eye movements by means of the classification of AR and wavelet coefficients. Figure 1.15 represents the block diagram of the organization of this thesis.

Figure 1.15: Organization of the thesis

Chapter 2

Autoregressive Modelling and Wavelet Analysis Theories

IN this chapter the theories behind the methods used in sleep eye movement analysis are discussed. Autoregressive (AR) modelling and wavelet analysis are the main mathematical techniques used in this research.

Signal modelling is an important method which can be used in a variety of fields, including biomedical engineering. Modelling of a biomedical signal is important for distinguishing abnormalities in the waveform of the signal. In some areas such as heart signals (electrocardiogram [ECG]), the waveform has a known shape; however, in other areas such as sleep eye movements, it does not. There is tremendous amount of information buried in electrooculogram (EOG) that is unknown to clinicians. Any new information that can be extracted from this signal can help neurologists and psychiatrists in their diagnostic or therapeutic analyses. Therefore, possible methods for modelling of EOG signals were considered in this research. If it is possible to model a signal accurately with a small number of parameters k , where k is much smaller than the length of the signal, the signal singularities can be characterized by them. These singularities may be used for the detection of abnormalities. For example, if one can find the special pattern for normal sleep eye movements similar to what has been found for the normal waveform in the ECG, then the gates will open for sleep disorder therapies.

Signal modelling is also used in signal prediction and signal interpolation. In both of these

problems, a signal $x(n)$ is known over an interval of time and the goal is to predict $x(n)$ for the next sample. In the prediction problem, as an example, $x(n)$ is known over the duration of N samples and the goal is to predict the $x(N + 1)$ value. In the interpolation problem, one or more consecutive values may be distorted and the goal is to accurately predict these distorted samples. For example, in some sleep disorder patients the ECG signals distorted the EOG signals. In these cases, the need for a prediction method for recovering the signals is significant. One of the famous modelling methods is autoregressive (AR) modelling, which will be discussed in detail in this chapter.

On the other hand, a wavelet transform can help transform the EOG signal from the time-domain into time and frequency space. This transformation can help to extract more information about normalities and abnormalities in EOG signals. Wavelet analysis may be a good method for analyzing non-stationary signals such as EOG. Wavelet analysis can represent high and low frequency information with different spatial resolution in a computationally efficient way. Wavelet analysis, which was used first in seismic geology, is developed in other fields such as physics, mathematics, and electrical engineering. Consequently, this method has been developed for many applications in medical and industrial fields such as FBI Fingerprint Compression [90], data denoising [91], and data mining [92].

In this research, AR modelling and wavelet analysis have been used for extraction and classification of sleep eye movements of those who were exposed to antidepressant medication and those who were not. This chapter discusses the theories that support the methods used in this research. This chapter is divided into the following six sections: the uncertainty principle, wavelet analysis, all-pole and autoregressive modelling, pattern classification, and receiver operating characteristic *ROC* curve.

2.1 The Uncertainty Principle and Heisenberg Rectangles

The uncertainty principle states that “the energy spread of a function and its Fourier transform [see Appendix A] cannot be simultaneously arbitrarily small [93].” In 1946, Gabor

defined elementary time-frequency atoms. Time-frequency atoms are the waveforms with minimal energy spread in the time-frequency plane. He proposed to decompose signals over these atoms to measure the time-frequency energy spread, or in a simple word, “information” [93].

Gabor atoms are constructed by a translated window g in the time and frequency domain. Equation 2.1 defines this window.

$$g_{\mu,\gamma}(t) = g(t - u) \exp(i\gamma t) \quad (2.1)$$

The energy of $g_{\mu,\gamma}$ is spread over an interval of σ_t in the neighborhood of μ and γ (mean value of the Gaussian distribution over the time and frequency axes respectively, see Figure 2.1), where the σ_t is measured by the standard deviation of $|g|^2$.

On the other hand, σ_ω measures the domain where the Fourier transform of $g_{\mu,\gamma}$ is non-negligible. The Fourier transform of g is $\hat{g}$ that is derived from Equation 2.2:

$$\hat{g}_{\mu,\gamma}(\omega) = \hat{g}(\omega - \gamma) \exp(-i\mu(\omega - \gamma)) \quad (2.2)$$

The time-frequency localization of an atom, called the Heisenberg Box, is represented in Figure 2.1. The Heisenberg Box is a rectangle with time width of σ_t and frequency height of σ_ω .

Although the length and width of the Heisenberg Box may vary, the area of the box is derived from the uncertainty principle by the inequality of Equation 2.3. This area is minimum if g is a Gaussian distribution. In this case the atoms are called *Gabor Functions*.

$$\sigma_t \sigma_\omega \geq \frac{1}{2} \quad (2.3)$$

2.1.1 Scalogram

Wavelet analysis has been used in measuring the energy of a signal in a time-frequency neighborhood of f_i and t_i located in a resolution box, the Heisenberg Box. This wavelet energy measuring procedure is called a scalogram [94].

Figure 2.1: Heisenberg Box

2.2 Wavelet Analysis

Wavelets are powerful tools that can be used to analyze the time-frequency behaviour of a signal with finite energy. Wavelets decompose the signal into the time and frequency domains. The wavelet decomposition can have the advantage of better performance in time varying signals, and/or on signals with singular characters, such as EOG and EEG, than other linear time-frequency transforms such as Short Time Fourier. This is due to decomposing the signals into different scales. Wavelet analysis has been applied in many disciplines such as data compression, image analysis, and signal processing [94].

Wavelet analysis reveals aspects such as trends, breakdown points, discontinuities in higher derivatives, and self-similarity, while other signal analysis techniques may not be able to do so. Furthermore, wavelet analysis is often used for de-noising a signal without much

degradation.

Wavelet transform is a linear transform. In other words, the wavelet transforms the time-frequency information of a signal, with a family of waveforms, into the time and frequency domain without missing any information [94]. Thus, a signal or function $f(t)$ can be analyzed, only if it can be expressed as a linear decomposition in the form such as Equation 2.4.

$$f(t) = \sum_l a_l \psi_l(t) \quad (2.4)$$

where l is an integer index for the finite or infinite sum, a_l is the set of real-valued expansion coefficients, and $\psi_l(t)$ is a set of real-valued functions of t , known as the expansion set. If the expansion of Equation 2.4 is unique, then the expansion set will be called the *basis* [95].

2.2.1 Signal Classes and Signal Space Analysis

There are three types of signals; $L^2(\mathfrak{R})$, $L^1(\mathfrak{R})$, and *Distributions*.

A space that is particularly important in signal processing is $L^2(\mathfrak{R})$. This is the space of all functions, $f(t)$, with finite energy, i.e., it contains all functions with finite square integrals: $f \in L^2 \Rightarrow \int |f(t)|^2 dt = E < \infty$ [95].

$L^1(\mathfrak{R})$ is the next most basic class of signals. It contains all functions with finite absolute value integrals: $f \in L^1 \Rightarrow \int |f(t)|^1 dt = E < \infty$ [95].

A more general class of signals are Distributions. These signals, or functions, are not defined by having “values.” They are defined by the value of their inner product with a normal function. An example of this family is the Dirac delta function, $\delta(t)$, which is defined as: $f(T) = \int f(t)\delta(t - T)dt$ [95].

A *function space* is a linear vector space, with finite or infinite dimensions, where the vectors are functions, and the scalars are real numbers. The inner product of $f(t)$ and $g(t)$ is a scalar a , obtained from Equation 2.5 [95].

$$a = \langle f(t), g(t) \rangle = \int f^*(t)g(t)dt \quad (2.5)$$

where $f^*(t)$ is the conjugate of $f(t)$. The *Norm* of the *length* of a vector is defined by Equation 2.7 [95].

$$\|f\| = \sqrt{|\langle f, f \rangle|} \quad (2.6)$$

If the inner product of two signals (vectors) is zero, those signals are called *orthogonal*. If the basis is an orthogonal, then:

$$\langle \psi_k(t), \psi_l(t) \rangle = \int \psi_k(t) \psi_l(t) dt = 0 \quad k \neq l \quad (2.7)$$

Coefficients of Equation 2.4 are derived from Equation 2.8, by the inner product where: [95]

$$a_k = \langle f(t), \psi_k(t) \rangle = \int f(t) \psi_k(t) dt \quad (2.8)$$

A simple *scaling* and *translation* can generate a single scaling function or a wavelet (Equation 2.9) [95].

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k) \quad j, k \in \mathbb{Z} \quad (2.9)$$

where $\mathbb{Z}$ is the set of all integers, and the factor $2^{j/2}$ maintains the norm value of $\psi_{j,k}(t)$, as a constant independent of scale j [95].

2.2.2 Scaling Function Analysis

A set of scaling functions is defined in a simpler form such as Equation 2.10, in terms of translation of a basic scaling function

$$\varphi_k(t) = \varphi(t - k) \quad k \in \mathbb{Z} \quad \varphi \in L^2 \quad (2.10)$$

Suppose the subspace of $L^2(\mathbb{R})$, which is called ν_0 , spanned by $\varphi_k(t)$ is defined in Equation 2.11.

$$\nu_0 = \overline{\text{Span}_k \{\varphi_k(t)\}} \quad (2.11)$$

where integer k changes from $-\infty$ to $+\infty$; and the closure which is denoted by over-bar means $f(t) = \sum_k a_k \varphi_k(t)$ for any $f(t) \in \nu_0$ [95].

The size of the subset can be spanned by changing the time scale of the scaling functions. Equation 2.12 represents such a change [95].

$$\varphi_{j,k}(t) = 2^{j/2} \varphi(2^j t - k) \quad (2.12)$$

Functions represented by Equation 2.12 are the subset of bigger spaces than ν_0 , known as space ν_j defined in Equation 2.13.

$$\nu_j = \overline{\text{Span}_k\{\varphi_k(2^j t)\}} = \overline{\text{Span}_k\{\varphi_{j,k}(t)\}} \quad (2.13)$$

in which $k \in \mathbb{Z}$. This means $f(t) = \sum_k a_k \varphi(2^j t + k)$ for $f(t) \in \nu_j$ [95].

In conclusion, by increasing j (if $j > 0$), the scales become finer since $\varphi_{j,k}(t)$ becomes narrower; therefore, it is translated into smaller steps. If $j < 0$, by increasing the absolute value of j , $\varphi_{j,k}(t)$ becomes wider and, only coarse information can be represented by wider scaling functions [95]. This concept was originally derived from map analysis. In a coarse scale, the map includes wider information without presenting the details; however, a map with the same size and a finer scale has narrower information but presents more details.

Multiresolution Analysis in Subspaces

For formulating the basic requirement of multiresolution analysis (MRA), the following relationship in Equation 2.14 should be considered, because of the intuitive ideas of scale.

$$\dots \subset \nu_{-3} \subset \nu_{-2} \subset \nu_{-1} \subset \nu_0 \subset \nu_1 \subset \nu_2 \subset \nu_3 \subset \dots \subset L^2 \quad (2.14)$$

where $\nu_{-\infty} = \{\emptyset\}$ and $\nu_{+\infty} = L^2$. This equation demonstrates that each coarse space is a subset of a finer scale, which means the information at finer scale also exists at the more coarse scale.

To satisfy the natural scaling condition, the following relation is expected:

$$f(t) \in \nu_j \Leftrightarrow f(2t) \in \nu_{j+1} \quad (2.15)$$

Equation 2.15 shows that elements in a space are simply scaled for the next space. Figure 2.2 represents the condition of Equation 2.15, which means if $\varphi(t) \in \nu_0$, then $\varphi(t) \in \nu_1$, and the space is spanned by $\varphi(2t)$. Therefore, $\varphi(t)$ can be expressed in terms of $\varphi(2t)$ in the form of Equation 2.16. This equation is the basic equation of multiresolution formulation [95].

Figure 2.2: The Scaling and Wavelet Space

$$\varphi(t) = \sum_n h(n) \sqrt{2} \varphi(2t - n) \quad n \in \mathbb{Z} \quad (2.16)$$

In Equation 2.16, $h(n)$ are real or complex coefficients, known as scaling function coefficients, scaling filters, or scaling factors. Also, $\sqrt{2}$ keeps the norm of the scaling function. Equation 2.16 is called the refinement equation, the dilation equation, or the multiresolution analysis equation (MRA) [95]. In this research, the EOG signals were decomposed into different scales, and the resulting coefficients were used for classification of patients who were exposed to antidepressant medications and those who were not.

2.2.3 The Wavelet Function and Wavelet Vector Spaces

The important features of a signal can be described by defining a set of functions, $\psi_{j,k}(t)$, that span the differences between the spaces by various scales of the scaling functions. These functions, $\psi_{j,k}(t)$, are slightly different from scaling functions $\varphi_{j,k}(t)$, and they are known as wavelets.

The scaling functions are orthogonal to wavelets. The advantages of this orthogonality are in the simplicity of the calculation of expansion coefficients. This means that all members of ν_j and ω_j are orthogonal, which can be described characteristically in the mathematical form of Equation 2.17:

$$\langle \varphi_{j,k}(t), \psi_{j,l}(t) \rangle = \int \varphi_{j,k}(t) \psi_{j,l}(t) dt = 0 \quad j, k, l \in \mathbb{Z} \quad (2.17)$$

The wavelet spanned subspace, ω_0 , is defined in Equation 2.18, where

$$\nu_1 = \nu_0 \oplus \omega_0 \quad (2.18)$$

The general form of Equation 2.18 can be described in Equation 2.19 [95].

$$L^2 = \nu_0 \oplus \omega_0 \oplus \omega_1 \oplus \dots \quad (2.19)$$

Figure 2.2 represents the nesting of the scaling function spaces. One can understand from this figure that all of the wavelet spaces, except the zero element, are the disjoint differences. This means: wavelet spaces are orthogonal complements ($\nu_0 \perp \omega_0 \perp \omega_1 \perp \dots$) [95].

Any function, $g(t) \in L^2(\mathbb{R})$, can be written in the form of Equation 2.20 as a series of expansions in terms of a scaling function (first term of the right side of the equation) and wavelets (second term of the right side of the equation) [95].

$$g(t) = \sum_{k=-\infty}^{\infty} c(k) \varphi_k(t) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} d(j, k) \psi_{j,k}(t) \quad (2.20)$$

The coefficients of Equation 2.20 are derived from Equation 2.21 and Equation 2.22 [95].

$$c(k) = c_0(k) = \langle g(t), \varphi_k(t) \rangle = \int g(t) \varphi_k(t) dt \quad (2.21)$$

$$d_j(k) = d(j, k) = \langle g(t), \psi_{j,k}(t) \rangle = \int g(t) \psi_{j,k}(t) dt \quad (2.22)$$

2.2.4 Continuous Wavelet Transform

By definition, the continuous wavelet transform (CWT), $Wf(a, b)$, of a finite-energy signal $f(t)$ is the correlation of $f(t)$ with a mother wavelet.

It is not a weakness for a wavelet that it produces a time-scale and not a time-frequency view of a signal. A time-scale introduces a different way of viewing data, which is a natural way of viewing data from a large number of natural phenomena. For example, the lunar landscape has a ragged surface due to bombardment by different sizes of meteorites. If this surface is imagined as a one-dimensional signal, then it can be interpreted that the signal has the components of different scales; large features are made by large meteorites and fine features by small ones. In most natural signal analyses such as modelling the natural structures of clouds, mountain landscapes, coastlines, blood vessels, and tree branches, thinking in terms of scale makes more sense than thinking in terms of frequency, because the repetitive shapes are different in terms of scales. Also, there is an inverse relationship between scale and frequency: mathematically, scale is one over frequency.

The CWT is defined by Equation 2.23, and the Inverse Continuous Wavelet Transform (ICWT) is defined in Equation 2.24. In CWT, $f(t)$ is a continuous function over the variable t ; thus, $F(a, b)$ is a continuous function over the variables a and b .

$$F(a, b) = \int f(t) w\left(\frac{t-a}{b}\right) dt \quad (2.23)$$

$$f(t) = \int \int F(a, b) w\left(\frac{t-a}{b}\right) da db \quad (2.24)$$

where $w(t)$ is the basic wavelet and $a, b \in \mathbb{R}$ are real continuous variables. Daubechies discussed conditions for the wavelet $w(t)$ to support the ICWT [96].

Figure 2.3 represents the Coiflet mother wavelet of order 3.

Figure 2.3: Coiflet Mother Wavelet of order 3 (x-axis is time, samples, and y-axis is amplitude, volts)

Figure 2.4 represents the Daubechies mother wavelet of order 4. Daubechies wavelets are orthogonal wavelets with a different order of smoothness.

2.2.5 Discrete Wavelet Transform

Figure 2.5 represents the procedure of multiresolution decomposition of a signal $x[n]$. Each stage consists of two digital filters, high pass and low pass, and two down samplers, by 2. In other words, $g[\cdot]$ is the high pass filter in nature, and $h[\cdot]$ is the low pass filter in nature. After passing the signal through the high pass filter, the downsampled result by 2 is called “detail”. After passing the signal through low pass, the downsampled result by 2 is called “approximation” [97]. The approximation part will go through further decompositions, as

Figure 2.4: Daubechies Mother Wavelet of Order 4

represented in Figure 2.5.

The discrete wavelet transform (DWT) concept can be described in mathematical form. In this regard, more general statements of Equation 2.20 are considered in Equation 2.25 and 2.26.

$$g(t) = \sum_k c_{j_0}(k) 2^{j_0/2} \varphi(2^{j_0/2}t - k) + \sum_k \sum_{j=j_0}^{\infty} d_j(k) 2^{j/2} \varphi(2^{j/2}t - k) \quad (2.25)$$

$$g(t) = \sum_k c_{j_0}(k) \varphi_{j_0,k}(t) + \sum_k \sum_{j=j_0}^{\infty} d_j(k) \varphi_{j,k}(t) \quad (2.26)$$

where j_0 determines the first scaling function space, since $L^2 = \nu_{j_0} \oplus \omega_{j_0} \oplus \omega_{j_0+1} \oplus \dots$. The coefficients of Equation 2.25 and 2.26 are called the DWT of the signal $g(t)$. By having certain conditions, the original signal can be described completely by these wavelet coefficients. Also, these coefficients can be used for analysis, description, approximation, and filtering, just the same as Fourier series coefficients.

Figure 2.5: Sub-band decompositions of discrete wavelet transform. $g[\cdot]$ is the high pass filter. $h[\cdot]$ is the low pass filter.

The wavelet coefficients can be derived by Equations 2.27 and 2.28, where as the higher the scale, the higher the resolution.

$$c_j(k) = \langle g(t), \varphi_{j,k}(t) \rangle = \int g(t) \varphi_{j,k}(t) dt \quad (2.27)$$

$$d_j(k) = \langle g(t), \psi_{j,k}(t) \rangle = \int g(t) \psi_{j,k}(t) dt \quad (2.28)$$

2.2.6 Log-Gabor Wavelet

In 1987, the solution to the Gabor wavelet (see Appendix B) was proposed by Field [98]. He discovered that natural images are better when passed through Gaussian filters and presented in logarithmic frequency scales. Equation 2.29 shows the transfer function of the log-Gabor function [98].

$$G(\omega) = \exp(-\log(\omega/\omega_0)^2)/(2(\log(k/\omega_0))^2) \quad (2.29)$$

where ω_0 is the centre frequency of the filter and k/ω_0 should be a constant while ω_0 is varying. The log-Gabor wavelet has two characteristics. It has no DC value and it has an extended tail at high frequencies. Furthermore, the second characteristic has been used

in filtering of the natural images where their amplitude spectra decrease at approximately $1/\omega$ [98].

2.3 All-Pole and Autoregressive Modelling

All-pole models are popular for several reasons. The physical process by which a signal is generated will result in an all-pole (autoregressive) signal in some applications. Furthermore, in other applications for which it may not be possible to fit an all-pole model for the signal, an all-pole model can be found to approximate the process, because in many different types of applications all-pole models have been found to provide a sufficiently accurate representation. Also, all-pole models are special structures that are found in all-pole equations, which make algorithms fast and efficient for finding the all-pole parameters [99].

Signal $x(n)$ has a real or complex value, and it is equal to zero for $n < 0$. Equation 2.30 represents the form of all-pole modelling of $x(n)$.

$$H(z) = \frac{b(0)}{A_p(z)} = \frac{b(0)}{1 + \sum_{k=1}^p a_p(k)z^{-k}} \quad (2.30)$$

The coefficients of all-pole model $a_p(k)$, can be derived by different methods. For example, with Prony's method, the coefficients can be found by minimizing the error given in Equation 2.31 [99].

$$\varepsilon_p = \sum_{n=0}^{\infty} |e(n)|^2 \quad (2.31)$$

where

$$e(n) = x(n) + \sum_{k=1}^p a_p(k)x(n-k) \quad (2.32)$$

Therefore, normal equations which are used for evaluating the all-pole model coefficients by Prony's method are derived from Equation 2.33 and 2.34 [99].

$$\sum_{l=1}^p a_p(l)r_x(k-l) = -r_x(k) \quad k = 1, 2, \dots, p \quad (2.33)$$

$$r_x(k) = \sum_{n=0}^{\infty} x(n)x^*(n-k) \quad (2.34)$$

Also, Equation 2.35 represents the initial value for solving the Normal equations, and is derived by satisfying the energy constraint [99]:

$$b(0) = \sqrt{\varepsilon_p} \quad (2.35)$$

By the above assumption, Equation 2.35, the minimum error is evaluated by Equation 2.36 [99].

$$\varepsilon_p = r_x(0) + \sum_{k=1}^p a_p(k)r_x^*(k) \quad (2.36)$$

If a white noise $v(n)$ is filtered with a causal linear shift-invariant filter having a rational system function with an all-pole filter of the form of Equation 2.30, the filter is called an autoregressive (AR) process of order p , and it is referred to as an AR(p) process. In other words, AR(p) is a wide-sense stationary process which may be generated by filtering unit variance white noise, $v(n)$, with an all-pole filter of the form of Equation 2.30 [99].

One can accept that the autocorrelation sequence of an AR process satisfies the Yule-Walker equations of the form:

$$r_x(k) + \sum_{l=1}^p a_p(l)r_x(k-l) = |b(0)|^2\delta(k) \quad k \geq 0. \quad (2.37)$$

Therefore, if autocorrelation values $r_x(k)$ are given for $k = 1, 2, \dots, p$, then the AR coefficients can be derived. This method is known as the Yule-Walker method. By comparing Yule-Walker equations and normal equations of Prony's method for all-pole modelling, one can understand that the two sets of equations are identical. The only difference between them is how the autocorrelation $r_x(k)$ is defined in each equation; $r_x(k)$ is a deterministic

autocorrelation in all-pole modelling of Prony's method, and it is a statistical autocorrelation in the Yule-Walker method. Also $b(0)$ is evaluated by:

$$|b(0)|^2 = r_x(0) + \sum_{k=1}^p a_p(k)r_x(k). \quad (2.38)$$

Furthermore, the statistical autocorrelation can be estimated from a sample realization of the process. For example, if $x(n)$ is determined for $0 \leq n < N$, then $\hat{r}_x(0)$ can be estimated by using Equation 2.39, the autocorrelation definition:

$$\hat{r}_x(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)x(n-k) \quad (2.39)$$

Therefore, despite the differences between stochastic and deterministic all-pole modelling, the two methods are almost equivalent when the two autocorrelations are estimated.

2.3.1 Burg Algorithm

In several approaches, to estimate the optimal all-pole model coefficients $a_p(k)$, an error is defined, and an attempt is made to minimize it. One of the methods for parameterizing the all-pole filter is to use lattice filter. In this method the reflection coefficient is used to minimize the errors. One of these lattice filter approaches is Burg's method.

The Burg algorithm has been used in this work to estimate AR coefficients. In the 1960s, he developed a method, called the *maximum entropy method*, for spectrum estimation. A part of this method involves finding an all-pole model for the data. He also demonstrated how reflection coefficients can be computed sequentially, by minimizing the sum of the forward and backward predication errors [99].

$$\varepsilon_j^B = \varepsilon_j^+ + \varepsilon_j^- = \sum_{n=j}^N |e_j^+(n)|^2 + \sum_{n=j}^N |e_j^-(n)|^2 \quad (2.40)$$

By refereing to Equation 2.40, the *blending* of forward and backward prediction errors, used in Burg's method, shows an equal emphasis on ε_j^+ and ε_j^- . In addition, what makes the Burg error, ε_j^B , valuable, is that sequentially minimizing ε_j^B bounds the reflection coefficients

by one in magnitude ($|\Gamma_j^B| \leq 1$); thus, the model is stable. The reflection coefficients are derived from Equation 2.41.

$$\Gamma_j^B = -\frac{2 \sum_{n=j}^N e_{j-1}^+(n)[e_{j-1}^-(n-1)]^*}{\sum_{n=j}^N \{|e_{j-1}^+(n)|^2 + |e_{j-1}^-(n-1)|^2\}} = -\frac{2\langle e_{j-1}^+, e_{j-1}^- \rangle}{\|e_{j-1}^+\|^2 + \|e_{j-1}^-\|^2} \quad (2.41)$$

The Burg error can be written in general form (Equation 2.42) with the initial value of Equation 2.43:

$$\varepsilon_j^B = \{\varepsilon_{j-1}^B - |e_{j-1}^+(j-1)|^2 + |e_{j-1}^-(N)|^2\}[1 - |\Gamma_j^B|^2] \quad (2.42)$$

$$\varepsilon_0^B = \sum_{n=0}^N \{|e_0^+(n)|^2 + |e_0^-(n)|^2\} = 2 \sum_{n=0}^N |x(n)|^2 \quad (2.43)$$

It can be assumed that,

$$D_j = \varepsilon_{j-1}^B - |e_{j-1}^+(j-1)|^2 + |e_{j-1}^-(N)|^2 \quad (2.44)$$

then the desired recursion can be derived by Equation 2.45

$$D_{j+1} = D_j[1 - |\Gamma_j^B|^2] - |e_j^+(j)|^2 - |e_j^-(N)|^2 \quad (2.45)$$

The Burg algorithm is as follows:

1. Initialize the recursion

- $e_0^+(n) = e_0^-(n) = x(n)$
- $D_1 = 2 \sum_{n=1}^N \{|x(n)|^2 - |x(n-1)|^2\}$

2. For $j = 1$ to p

- $\Gamma_j^B = -\frac{2}{D_j} \sum_{n=j}^N e_{j-1}^+(n)[e_{j-1}^-(n-1)]^*$
- For $n = j$ to N

- $e_j^+(n) = e_{j-1}^+(n) + \Gamma_j^B e_{j-1}^-(n-1)$
- $e_j^-(n) = e_{j-1}^-(n-1) + (\Gamma_j^B)^* e_{j-1}^+(n)$
- $D_{j+1} = D_j(1 - |\Gamma_j^B|^2) - |e_j^+(j)|^2 - |e_j^-(N)|^2$
- $\varepsilon_j^B = D_j[1 - |\Gamma_j^B|^2]$

2.4 Pattern Classification

In this section, the theories supporting the classification procedures in this research are discussed. Rangayyan defines pattern recognition or classification as: “categorizing of input data into identifiable classes via the extraction of significant features or attributes of the data from a background of irrelevant detail” [100]. In biomedical signal analysis, one can extract quantitative features from the signal. Each real signal may consist of infinite features. For evaluating the best classifier, the choosing of features is also important. Nevertheless, each signal can be represented by a feature vector $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$, which is also called the measured vector, or the pattern vector in which x_i represents the measurement of a feature as a real value. Therefore, the end of vector $\mathbf{x}$ represents a point in an n-dimensional Euclidean space [100]. Figure 2.6 represents the vectors of similar objects that are formed as clusters.

Appropriate design of preprocessing and the feature extraction procedure are the keys to having good classification results. In the pattern classification problems, one is eager to find the optimal decision boundaries to separate the data into pattern classes based on the feature vectors [100].

In the literature of classification, *training set* refers to a given set of feature vectors of known categorization with which the classifier is designed. Also, *testing set* refers to new feature vectors of known classes, which test the performance of the classifier. The approach of using a training set and then a testing set is known as *supervised* pattern classification; the training set is used to develop the mathematical function that can characterize the separation between the classes, and the testing set is used for testing the classification performance by using the provided mathematical function [100].

Figure 2.6: Two-dimensional feature vectors of two classes C_1 and C_2 . The vectors z_1 and z_2 are the prototypes of the two classes. The solid line that represents the decision boundary is perpendicular to the straight dashed line, joining the two classes of prototypes.

2.4.1 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) was introduced by Fisher in 1936. Fisher used the LDA analysis for taxonomic analysis in [101]. In 1968, Altman introduced LDA in the failure prediction literature. LDA has been used in many applications, especially in two-case classifications for failure prediction in manufacturing, and for management or normal/abnormal classification in medical applications such as cancer detection or heart disease.

In this research, LDA has been applied in two-group cases, known as binary cases. In a binary case, LDA is based on the linear discriminant function of the independent variables, $(x_1, x_2, x_3, \dots, x_n)$, in Equation 2.46. Figure 2.7 represents Equation 2.46.

$$d(\mathbf{x}) = w_1x_1 + w_2x_2 + \dots + w_nx_n + w_{n+1} = \mathbf{w}^T\mathbf{x} \quad (2.46)$$

where $d(\mathbf{x})$ is the score of the function, w_{n+1} is a constant, known as *bias* or *threshold weight*, and $(w_1, w_2, w_3, \dots, w_n)$ are discriminant coefficients called *weight vectors*. To maximize the

Figure 2.7: A simple linear classifier with n input units and $n + 1$ weight vector

distance between the group means of $d(\mathbf{x})$, the parameters of the function are traditionally estimated with the ordinary least-squares method [102] [103].

Also, $d(\mathbf{x})$ is called a linear discriminant or decision function, where the following equation exists: $\mathbf{x} = (x_1, x_2, \dots, x_n, 1)^T$, also known as the feature vector, augmented by an additional entry equal to unity. The equation of $\mathbf{w} = (w_1, w_2, \dots, w_n, w_{n+1})^T$ is also known as the augmented weight vector [100].

A two-class pattern classification can be described by Equation 2.47.

$$d(\mathbf{x}) = \mathbf{w}^T \mathbf{x} \begin{cases} > 0 & \text{if } \mathbf{x} \in C_1 \\ \leq 0 & \text{if } \mathbf{x} \in C_2 \end{cases} \quad (2.47)$$

where C_1 and C_2 are the two classes. Here, $d(\mathbf{x})$ is the discriminant function separating the two classes, as represented in Figure 2.6 [100].

In the two-case method (i.e two-class method), each of the two classes is separable from the other by a single decision surface. In other words, if $d_i(\mathbf{x}) > 0$, then $\mathbf{x} \in C_i$ and $i = 1$ or $i = 2$.

Patterns which are separable by linear decision functions are said to be linearly separable.

It should also be added that in other cases, such as in an infinite variety of complex decision boundaries, it will be described by generalized nonlinear decision functions of the feature vectors, as has been shown in Equation 2.48:

$$d(\mathbf{x}) = w_1 f_1(\mathbf{x}) + w_2 f_2(\mathbf{x}) + \dots + w_k f_k(\mathbf{x}) + w_{k+1} = \sum_{i=1}^{k+1} w_i f_i(\mathbf{x}) \quad (2.48)$$

$\{f_i(\mathbf{x})\}$ where $(i = 1, 2, \dots, k)$ are real single-valued functions of $\mathbf{x}$ and $f_{k+1}(\mathbf{x}) = 1$ [100].

In some cases $f_i(\mathbf{x})$ is nonlinear in the n -dimensional space of $\mathbf{x}$. In these cases, the feature vector is transformed to: $\mathbf{x}^\dagger = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_k(\mathbf{x}), 1)^T$, and $d(\mathbf{x})$ is defined linearly as $d(\mathbf{x}) = \mathbf{w}^T \mathbf{x}^\dagger$, where $\mathbf{w} = (w_1, w_2, \dots, w_k, w_{k+1})^T$ [100].

Hyperplane as the Decision Surface

In general, the vector of x has usually more than two dimensions, so when $d(\mathbf{x})$ is linear, the decision surface that separates points assigned to each class, is a *Hyperplane*, H . In this case, if $\mathbf{x}_1$ and $\mathbf{x}_2$ are both on the decision surface, then Equation 2.49 and 2.50 will be true.

$$d_1(\mathbf{x}) = \mathbf{w}^T \mathbf{x}_1 = d_2(\mathbf{x}) = \mathbf{w}^T \mathbf{x}_2 \quad (2.49)$$

$$\mathbf{w}^T (\mathbf{x}_1 - \mathbf{x}_2) = 0 \quad (2.50)$$

This means that $\mathbf{w}$ is normal to any vector lying in the hyperplane; in other words, it is perpendicular to the hyperplane. It can be assumed that hyperplane H divides the feature space into two parts: decision region R_1 for the class C_1 , and R_2 for the class C_2 [103].

The normal vector $\mathbf{w}$ points toward R_1 , because $d(\mathbf{x}) > 0$, if $\mathbf{x}$ is in R_1 . By convention, if $\mathbf{x}$ is in R_1 , it is said to be on the *positive* side of H , and if it is in R_2 , it is said to be on the *negative* side of H [103].

$$\mathbf{x} = \mathbf{x}_p + r \frac{\mathbf{w}}{\|\mathbf{w}\|} \quad (2.51)$$

In Equation 2.52, $d(\mathbf{x})$ measures the distance of $\mathbf{x}$ from the hyperplane [103]. In this equation, $\mathbf{x}_p$ is a normal projection of feature vector $\mathbf{x}$ onto hyperplane H , and r is the

distance between $\mathbf{x}$ and $\mathbf{x}_p$. (See Figure 2.8). Therefore, r is positive if $\mathbf{x}$ is on the positive side of H , and negative if it is on the negative side of H . Since $d(\mathbf{x}_p) = 0$, from discriminant function of $d(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0$, Equation 2.52 (which is equivalent to 2.53) is derived.

$$d(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0 = r \|\mathbf{w}\| \quad (2.52)$$

or

$$r = \frac{d(\mathbf{x})}{\|\mathbf{w}\|} \quad (2.53)$$

Therefore, the distance from the origin to the hyperplane of H is derived by $\frac{w_0}{\|\mathbf{w}\|}$, in which the origin is on the positive side of H if $w_0 > 0$, and the origin is on the negative side if $w_0 < 0$. The hyperplane passes through the origin, and so the following equation exists: $d(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$ if $w_0 = 0$. In this case, $d(\mathbf{x})$ is said to have *homogeneous* form. Figure 2.8 represents this algebraic result [103].

Figure 2.8: The hyperplane H which separates the feature space into R_1 and R_2 .

In other words, a linear discriminant function divides the feature space into separate regions by a hyperplane decision surface. The normal vector $\mathbf{w}$ determines the origin of

the surface. Also, the bias w_0 determines the location of the surface. Therefore, $d(\mathbf{x})$ is proportional to the signed distance from $\mathbf{x}$ to the hyperplane. If $\mathbf{x}$ is on the positive side, then $d(\mathbf{x}) > 0$ and if $\mathbf{x}$ is on the negative side, then $d(\mathbf{x}) < 0$ [103].

Distance Function

As considered in Figure 2.6, the prototypes $\mathbf{z}_1$ and $\mathbf{z}_2$ are typically computed as the average of feature vectors for each of the two classes. Equation 2.54 shows the Euclidean Distance between an arbitrary pattern vector $\mathbf{x}$ and the i^{th} prototype, where $i = 1$ or 2 .

$$D_i = \|\mathbf{x} - \mathbf{z}_i\| = \sqrt{(\mathbf{x} - \mathbf{z}_i)^T (\mathbf{x} - \mathbf{z}_i)}. \quad (2.54)$$

A simple way of classifying the vector $\mathbf{x}$ is to derive the length of this vector, and the smallest distance to any classes results in the fact that it belongs to that class. This fact is represented in Inequation 2.55.

$$\text{if } D_i < D_j \ \forall j \neq i, \text{ then } \mathbf{x} \in C_i \quad (2.55)$$

Equation 2.56 represents a relationship between distance functions and discriminant functions.

$$D_i^2 = \|\mathbf{x} - \mathbf{z}_i\|^2 = (\mathbf{x} - \mathbf{z}_i)^T (\mathbf{x} - \mathbf{z}_i) = \mathbf{x}^T \mathbf{x} - 2\mathbf{x}^T \mathbf{z}_i + \mathbf{z}_i^T \mathbf{z}_i = \mathbf{x}^T \mathbf{x} - 2(\mathbf{x}^T \mathbf{z}_i - \frac{1}{2} \mathbf{z}_i^T \mathbf{z}_i) \quad (2.56)$$

Minimizing the D_i means minimizing D_i^2 ; therefore, $(\mathbf{x}^T \mathbf{z}_i - \frac{1}{2} \mathbf{z}_i^T \mathbf{z}_i)$ should be maximized, so it is called the decision function. See Equation 2.57 below:

$$d_i(\mathbf{x}) = (\mathbf{x}^T \mathbf{z}_i - \frac{1}{2} \mathbf{z}_i^T \mathbf{z}_i), \quad i = 1, 2 \quad (2.57)$$

and see the decision rule, which is stated in Inequation 2.58.

$$\text{if } d_i < d_j \ \forall j \neq i, \text{ then } \mathbf{x} \in C_i \quad (2.58)$$

This is known as the linear discriminant function [100].

Leave-One-Out Method

When a limited number of sample vectors with known classifications are used, it is important to know how many samples may be used for training a classifier. One should consider that once the classifier has been designed, another independent set of samples of known classification must be used for testing the classification. In the case that there exist a large number of samples for classification, the set may be randomly split into two approximately equal parts. One is used for training and the other is used for testing. The random-splitting can occur several times in order to generate several classifiers, and one of these classifiers can be chosen as the best classifier on the basis of its performance [100].

The leave-one-out method is a suitable method for estimating the classification accuracy of a pattern classification, especially when the number of available samples is small. In this method, one of the samples is excluded as the testing set, and the remaining samples are used as the training set. This procedure is continued by using other samples. Therefore, in N samples classification, with this method, $N - 1$ samples are used as trainers to design the classifier, and one sample is used for testing the classification each time; and this procedure is repeated N times. Finally, an average classification accuracy is evaluated using all the test results [100].

In a simple case consideration, it is assumed that the sample set covariances of the two classes are equal. For example, there are two sample sets: $S_1 = \{x_1^{(1)}, \dots, x_{N_1}^{(1)}\}$ from the class of C_1 , and $S_2 = \{x_1^{(2)}, \dots, x_{N_2}^{(2)}\}$ from the class of C_2 . N_1 and N_2 are the numbers of the samples in the sets of S_1 and S_2 respectively. Also, it is assumed that prior probabilities of the two cases are equal. For the simplification of this example, the Bayes classifier and the assumption that x has a multivariate Gaussian PDF distribution have been used. Therefore, x is assigned to class C_1 if it satisfies Equation 2.59:

$$(x - m_1)^T(x - m_1) - (x - m_2)^T(x - m_2) > \theta \quad (2.59)$$

where θ is a threshold. The sample mean, $\tilde{m}_i$, is given by Equation 2.60.

$$\tilde{m}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_j^{(i)} \quad (2.60)$$

In the leave-one-out method, one sample, $x_k^{(i)}$, is excluded from the training set; then after designing the classifier it is used as the sample test. In this case, the mean for class C_i , without $x_k^{(i)}$, is $\tilde{m}_{ik}$, which is computed in Equation 2.61.

$$\tilde{m}_{ik} = \frac{1}{N_i - 1} \left[\sum_{j=1}^{N_i} x_j^{(i)} - x_k^{(i)} \right] \quad (2.61)$$

Therefore, the distance of $x_k^{(i)}$ from the approximated mean is evaluated by Equation 2.62. Also, the decision rule for testing a sample, $x_k^{(1)}$, from C_1 , is evaluated by Equation 2.63.

$$x_k^{(i)} - \tilde{m}_{ik} = \frac{N_i}{N_i - 1} (x_k^{(i)} - \tilde{m}_i) \quad (2.62)$$

$$\begin{aligned} (x_k^{(1)} - \tilde{m}_{1k})^T (x_k^{(1)} - \tilde{m}_{1k}) - (x_k^{(1)} - \tilde{m}_{12})^T (x_k^{(1)} - \tilde{m}_{12}) = \\ = \left(\frac{N_1}{N_1 - 1} \right)^2 (x_k^{(1)} - \tilde{m}_1)^T (x_k^{(1)} - \tilde{m}_1) - (x_k^{(1)} - \tilde{m}_2)^T (x_k^{(1)} - \tilde{m}_2) > \theta \end{aligned} \quad (2.63)$$

One should note that only $\tilde{m}_1$ is changed when $x_k^{(1)}$ is tested, and $\tilde{m}_2$ is not. On the other hand, when a sample $x_k^{(2)}$ from C_2 is tested, the decision rule has been derived from Equation 2.64.

$$\begin{aligned} (x_k^{(2)} - \tilde{m}_1)^T (x_k^{(2)} - \tilde{m}_1) - (x_k^{(2)} - \tilde{m}_{2k})^T (x_k^{(2)} - \tilde{m}_{2k}) = \\ = ((x_k^{(2)} - \tilde{m}_1)^T (x_k^{(2)} - \tilde{m}_1) - \left(\frac{N_2}{N_2 - 1} \right)^2 (x_k^{(2)} - \tilde{m}_2)^T (x_k^{(2)} - \tilde{m}_2)) < \theta \end{aligned} \quad (2.64)$$

The leave-one-out method provides an almost unbiased estimate of the classification accuracy for a given classification method [100].

2.5 Receiver Operating Characteristic (ROC) Curve

Any assessment of diagnostic performance seems to require some comparison of diagnostic decisions with the "truth." The simplest measure of diagnostic decision quality is the fraction of cases for which the physician is correct, often is called accuracy. Sensitivity and specificity represent two kinds of accuracy, the first for actually positive cases and the second for

actually negative cases, from which accuracy is derived. Equations 2.65, 2.66, and 2.67 mathematically represent them.

$$\text{Sensitivity} = \frac{[\text{Number of True Positive (TP) Decisions}]}{[\text{Number of Actually Positive Cases}]} \quad (2.65)$$

$$\text{Specificity} = \frac{[\text{Number of True Negative (TN) Decisions}]}{[\text{Number of Actually Negative Cases}]} \quad (2.66)$$

$$\text{Accuracy} = \text{Sensitivity} \times P(D^+) + \text{Specificity} \times P(D^-) \quad (2.67)$$

$P(D^+)$ is the prior probability of the actual presence of the disease in a case from the population studied, and $P(D^-) = 1 - P(D^+)$ represents the prior probability that disease is actually absent in a case from the studied population [104].

It should be considered that sensitivity is also called True Positive Fraction (TPF), and specificity is also called True Negative Fraction (TNF). Medically, true positive means that the test indicates the presence of disease, when it is in fact present. False negative means that the test indicates the presence of disease, when it is, in fact, not the case. Accordingly, two new terms can be suggested by False Positive Fraction (FPF) and False Negative Fraction (FNF), described in Equations 2.68, 2.69.

$$\text{FPF} = \frac{[\text{Number of False Positive (FP) Decisions}]}{[\text{Number of Actually Negative Cases}]} \quad (2.68)$$

$$\text{FNF} = \frac{[\text{Number of False Negative (FN) Decisions}]}{[\text{Number of Actually Positive Cases}]} \quad (2.69)$$

For either actual state, positive or negative, the summation of the number of correct and false decisions is equal to the number of cases with that actual state. On the other hand,

$$\text{TPF} + \text{FNF} = 1$$

and

$$\text{TNF} + \text{FPF} = 1$$

The results of the decision table can be presented in a table such as Table 2.1 [104].

Table 2.1: Decision Matrix			
Actual State	Test Result Positive (T+)	Test Result Negative (T-)	Total Actual States
Positive (D^+)	TP	FN	Actually Positive
Negative (D^-)	FP	TN	Actually Negative
Total Test Results	Positive Decisions	Negative Decisions	

The following results can be derived from the Table 2.1 [104]:

$$TPF = TP / \text{Actually Positive}$$

$$FPF = FP / \text{Actually Negative}$$

$$FNF = 1 - TPF$$

$$TNF = 1 - FPF$$

$$P(D^+) = \text{Actually Positive} / \text{Total Cases}$$

$$P(D^-) = 1 - P(D^+)$$

$$\text{Accuracy} = TPF \times P(D^+) + TNF \times P(D^-)$$

One can also describe FPF as the conditional probability $P(T + | D+)$ which means “[t]he probability of a positive test, given the absence of disease.” Similarly, TPF equals $P(T + | D+)$, FNF equals $P(T - | D+)$, and TNF equals $P(T - | D-)$ [104].

The definitions are summarized in Table 2.3 [104].

Figure 2.9 represents the overlap of the distribution of result values in actually positive and actually negative result values. It is considered that one cannot find a single threshold to separate the population clearly. As one can see, TPF, FPF, FNF, and TNF change by changing the threshold value. Therefore, a threshold value should be chosen in such a way to compensate for the gains and losses, and to make a balance between the decision fractions [104].

Table 2.2: Definitions of the various Decision Performance

Definition
$TPF = \text{Sensitivity} = P(T+ D+)$
$FPF = 1 - \text{Specificity} = P(T+ D-)$
$TNF = \text{Specificity} = P(T- D-)$
$FNF = 1 - \text{Sensitivity} = P(T- D+)$
Disease Prevalence = $P(D+)$

Table 2.3: Definitions of the various Decision Performance

Definition
$TPF = \text{Sensitivity} = P(T+ D+)$
$FPF = 1 - \text{Specificity} = P(T+ D-)$
$TNF = \text{Specificity} = P(T- D-)$
$FNF = 1 - \text{Sensitivity} = P(T- D+)$
Disease Prevalence = $P(D+)$

Figure 2.9: Two probability density functions of a quantity on which decisions are based with a possible threshold

The confidence threshold depends on factors such as the estimate of prior probability and the assessment of the consequences of correct and incorrect decisions. Furthermore, in most cases this threshold is not fixed, and it varies, which causes the variation in decision fractions [104].

Although change in the decision threshold is caused by the variation of decision of all four factors, only TPF and FPF can be used to determine all of four of the decision factors (Table 2.1). Therefore, these two fractions can be tracked to evaluate a method by changing the decision threshold [104].

By referring to Figure 2.9, one can see that raising the decision threshold must increase both FPF and TPF. Therefore, FPF and TPF increase or decrease simultaneously as the decision threshold is changed, no matter what the forms of the two distributions are [104].

The TPF and FPF values can be plotted as “y” and “x”, coordinate values, while changing the decision threshold such as in Figure 2.10. The axis of this graph stretches from zero to one, because they are the limits of possible FPF and TPF (probabilities). It is considered that all points of this graph represent all possible combinations of TPF and FPF. This curve is called the Receiver Operating Characteristic (*ROC*) curve. Since it describes the inherent detection characteristics of the test, and since the receiver of the test information can operate at any point on the curve corresponding to the appropriate decision threshold, the receiver operating characteristic *ROC* curve is known for evaluating a diagnostic test [104].

Three possible operating points are shown in Figure 2.10: a strict threshold represents a case called positive only if judged almost definitely positive; a moderate threshold; and a lax threshold. Lax represents a case called positive in any suspicion of disease [104].

To understand what the curve means, it is considered that if disease prevalence is low, FPF must be kept small by choosing an appropriate threshold value; otherwise, all positive decisions will be considered false positive decisions, which may cause many unnecessary examinations or treatments of the patients. Also, FPF should be kept small for cases in which the consequences of a false positive decision are very dangerous, perhaps due to a high-risk, unnecessary surgery. In both cases, the decision maker should operate on the

Figure 2.10: Receiver Operating Characteristic (*ROC*) curve shows three possible points

lower left part of the receiver operating characteristic *ROC*. On the other hand, if the same test with the same receiver operating characteristic (*ROC*) curve is applied to a population with high disease prevalence, the decision threshold should be chosen higher on the curve for accepting higher FPF in order to keep FNF low and TPF high [104].

Figure 2.11 represents the comparison between test A and test B. Test A is superior to test B, because for the same FPF, test A has a higher TPF than test B, and for the same TPF, test A has a lower FPF than test B. Actually, the test performance is checked by the area under the receiver operating characteristic (*ROC*) curve. The greater the area under the curve, the better the performance of the test [104].

Figure 2.11: The comparison between the ROC curves of two tests

Receiver Operating Characteristics (*ROC*) curve is a common tool for measuring the effectiveness of medical modalities. Receiver Operating Characteristic (*ROC*) is derived from a standard detection theory, based on sensitivity and specificity definition. The Receiver

Operating Characteristic (*ROC*) curve is plotted based on True Positive versus False Negative decisions [105]. Also, one should know that in the detection of an illness in patients, it is often better to optimize the area under the Receiver Operating Characteristic (*ROC*) curve, rather than minimizing the error of classification [106]. Therefore, the area under the curve can be considered as one of the most important criteria in the evaluation of a method.

Chapter 3

Sleep Eye Movement Detection and Autoregressive Modelling

ELECTROOCULOGRAM (EOG) signals recorded during sleep have been considered by researchers for analysis recently. This research attempts to show that eye movements change significantly in patients exposed to antidepressant medications. This chapter describes the proposed methods of this research that how eye movements are detected, autoregressive (AR) coefficients of eye movements are derived, and the LDA classifications are used to represent the difference in the sleep eye movements of depressed patients who used antidepressant medications, and those who did not. Figure 3.1 represents the block diagram of the methodology used in this chapter. Each part of this block diagram is discussed in detail.

Figure 3.1: Methodology Used for the Sleep Eye Movement Classification of Patients Exposed/Unexposed to Antidepressant Medication

In this chapter, two groups of five patients who were exposed to Prozac (one group called 'A', the other called 'P'), one group of five patients who were exposed to Celexa (called 'L'), one group of five patients who were exposed to Effexor (called 'E'), and two groups

of five patients who were not exposed to antidepressant medication (one group called 'N', the other called 'C') were considered. Due to confidentiality of the patient information, the other information such as name, gender, and age of the patients were not released through the sleep lab by the time of this research.

3.1 Eye Movement Detection

Sleep eye movements are analyzed by expert clinicians. The long period of sleep recording, which is about six to eight hours, makes the precise analysis of eye movements impossible even for qualified experts. In order to analyze the sleep eye movements, eye movement detection software is required for quantification. On the other hand, a computer-assisted method is able to extract more information out of recorded data, and provides the time locations of detected movements for further assessments.

In order to develop a methodology, manual eye movement detection by clinicians should be considered. Figures 3.2 and 3.3 represent the eye movements' signals when the subject looks toward four different directions; leftward, rightward, upward, and downward. The sampling rate in these figures is 128 samples per second, and the y-axes are voltages with the unit of volt. Although the ROC signals in upward and downward movement are not recorded well in the up and down examples (see Figures 3.3), the four different types of the eye movements (leftward, rightward, upward, and downward) follow the decision rules of Table 3.1. On the basis of the eye movement decision table, if LOC-A2 (called loc, in lowercase) has a positive peak value, represented by (+), and ROC-A1 (called roc) has a negative peak, represented by (-), and VOC-A1 (voc) has a negative peak, also represented by (-), then the movement is classified as rightward movement. The other cases can be derived from Table 3.1.

The eye movement detection procedure block diagram is plotted in Figure 3.4. The following sections describe each block of this procedure in detail.

Figure 3.2: Examples of Rightward and Leftward Eye Movements

Figure 3.3: Examples of Upward and Downward Eye Movements

Table 3.1: Eye Movement Decision Table

EOG Signal	Right	Left	Down	Up
LOC - A2	+	-	-	+
ROC - A1	-	+	+	-
VOC - A1	-	+	-	+

Figure 3.4: The Block Diagram of Eye Movement Detection Procedure

3.1.1 Eye Movement Detection, Algorithm 1

The block diagram of Figure 3.4 represents Algorithm 1, which is used for eye movement detection. This algorithm includes sleep stage segmentation, bandpass filter, time segmentation, DC and above 3 standard deviation (SD) elimination, normalization, threshold and windowing, and Boolean classification. The inputs of Algorithm 1 are EOG signals including loc, roc, and voc. The outputs of this algorithm are detected movements in sleep EOG signals. In the following section, the specifications of this algorithm are discussed.

Sleep Stage Segmentation

In the first stage of the detection procedure, the EOG signals of each subject are divided into six sleep stages (wakefulness, Stages 1 – 4, and REM sleep) by following the clinical procedures in eye movement detections. Sleep stage segmentation is done by means of provided hypnograms.

The provided EOG signals, from the sleep laboratory at the Sunnybrook Health Sciences

Centre, include the six to eight hours of EOG signals (LOC, ROC, and VOC), two mastoid signals, and a hypnogram (see Figure 3.5). Also for more information about EOG signals, one can refer to Section 1.1.3.

Sleep stage information over time for each subject, called a *hypnogram*, was provided with full polysomnographic information by clinicians [2]. Figure 3.5 represents about 7.5 hours of ROC and LOC data with a hypnogram. In the hypnogram part of Figure 3.5, wakefulness is represented by zero, Stage REM by 5, Stage 1 by 1, Stage 2 by 2, and so forth. Although Figure 3.5 was provided for the time analysis of EOG signals in the first stage of this study, it was used thereafter for EOG segmentation.

Figure 3.5: A Typical Hypnogram

Lowpass Filter

The EOG signal is recorded with the sampling rate of 128 samples per second. The range of frequency of the EOG signals which are processed in this research extends from zero to 64Hz . This is because of the fact that these signals have been recorded at a sampling rate of 128 samples per second, which is twice the maximum frequency, 64Hz , pursuant to the Nyquist sampling theorem [107]. However, by the frequency analysis, which was done in Chapter 4, one can see that most of the frequency components are below 20Hz . For simplicity in Algorithm 1, the lowpass Finite Impulse Response (FIR) filter was used. Also, the range of the filter was designed for the cutoff frequency 20Hz in order not to lose much information. Different orders of the FIR filter were used, but the final filter with an order of 200 was chosen. Although the order of this filter may be high for hardware implementation, the purpose is to use a more ideal filter for cleaning the signal rather than a practical hardware filter.

Time Segmentation

Expert sleep clinicians analyze the eight-hour sleep signals in 30-second epochs. The full polysomnographic information, including EOG, EMG, EEG, and other channels are analyzed by clinicians in order to monitor the details of sleep onset of a patient. It should be mentioned that other durations, such as five-second and ten-second epochs can also be used. Furthermore, the 30-second epochs are more common and are used in Algorithm 1 for sleep eye movement detection.

DC and Above Three Standard Deviation Elimination

The technique that clinicians use to detect eye movement is based on the detection of antiphase waveforms in EOG channels, regardless of DC fluctuations. Therefore, the mean value of each 30-second epoch was evaluated. Thereafter, the difference between the signal and mean value in the epoch was used as the DC-eliminated signal. In addition, the values above three SD are removed by evaluating the SD within the epoch and setting the values

above three SD to zero.

Normalization

The 30-second epoch DC-eliminated and above-3-SD-eliminated signal is used for normalization. EMDS detects the maximum value for each processed signal. Then, all the values of the epoch were divided by the maximum value. Thus, the signal was normalized for comparison. This method is used by clinicians for sleep eye movement detection. The normalization is used by clinicians to place noise below the threshold value, so that it will not be interpreted as a movement. It should be considered that a great number of antiphase waveforms exist in the noise level in the actual EOG signals, which are not eye movements; or they may be movements but not distinguished as eye movements.

Thresholding and Windowing

A certain threshold was used for eye movement detection. The values above this threshold in EOG signals are assumed eye movements. The threshold value was evaluated by the experiment. Furthermore, the software should detect a movement if there is a combination of peak values above the threshold in different EOG channels, according to Table 3.1. The time indices of these maxims and minims are not equal. If the software decides only on the maximum and minimum, which has the same time index, then that may cause the loss of eye movement detection. The peak values of LOC, ROC, and VOC are compared in a defined window.

There is no golden standard for the threshold and window size values. Furthermore, if the threshold value is high, a large number of movements are missed in the detection procedure; and if the threshold is low, an antiphase noise will be detected as an eye movement. On the other hand, if the window size is too small, it causes the loss of eye movement detection; and if the window size is too big, it causes an increase in the number of detections, due to existing maximum and minimums within one particular window.

The threshold value of 0.2 and a window size of 15 samples, experimentally, were chosen. Table 3.2 represents the number of detected eye movements with different thresholds and

window sizes. The detection program in this particular evaluation runs over a segment of 2000 samples. The threshold of 0.1 detects noises as eye movements. Furthermore, the window size of 15 samples seems to be more reasonable in the detection process.

Table 3.2: An example of the quantification of detected eye movements with different threshold and window size

Threshold Window Size Eye Movement Type	0.1 15	0.1 30	0.2 15	0.2 30
Downward	62	66	23	38
Leftward	76	76	46	47
Upward	61	67	26	32
Rightward	82	85	49	48

Boolean Classification

Table 3.1 was used by clinicians to detect the eye movements. On the basis of the eye movement decision table, if loc (LOC-A2) has a negative peak value, represented by (−), roc has a positive peak, represented by (+), and voc has a positive peak, represented by (+), then the eye movement is classified as leftward movement. The other types of movements can be detected by the peak status mentioned in Table 3.1.

3.1.2 Eye Movement Detection, Algorithm 2

Due to the high complexity of Algorithm 1, the processing of 6 to 8 hours of EOG signals took a long time, usually over an hour. Therefore, the modification of Algorithm 1 to Algorithm 2, for reducing the complexity, was necessary. In this section, the changes in Algorithm 1 that led to the development of Algorithm 2, are discussed.

Using Vectors

Algorithm 1 produced objects from the segments of each EOG signal, that were higher than the threshold value. These objects were compared according to Table 3.1 for eye movement detection. The process of making objects for all parts of the signal above the threshold had

increased the complexity of Algorithm 1. Also, if there was no match between the objects, according to Table 3.1, the EMDS software would be crashed and locked in infinite loop. Furthermore, making vectors, instead of objects from the parts of each EOG signal, decreases the complexity. Therefore, vectors were used instead of objects in Algorithm 2. With this change, the complexity of the program was reduced and the detection procedure became faster.

Band Pass Filter

The other modification for making Algorithm 2 was using the band pass filter instead of the low pass filter in Algorithm 1. The 2007 *AASM Manual of the Scoring of Sleep and Associated Events* reports of the use of a bandpass filter of $0.3 - 35\text{Hz}$. However, the data, which were used in this research, were recorded with a bandpass filter of $0.3 - 64\text{Hz}$. As will be discussed in Chapter 4, the main frequency components of EOG are in the range of $0.3 - 20\text{Hz}$. The FIR filter with a low cutoff frequency 0.3Hz and a high cutoff frequency 20Hz , with an order of 200, was provided for filtering the signal.

Verification Plots

Although some of the detected movements were plotted for verification in Algorithm 1, Algorithm 2 has the verification plots of all detected movements. This feature can be used by clinicians for the verification of eye movement detection. Figures 3.6, 3.7, 3.8, and 3.9 represent examples of verification plots for each type of movement: leftward, rightward, upward, and downward. The left column of figures shows the signal after processing and the right one shows the original signals for each EOG channel. The unit of x-axis is sample. Further, y-axis is the amplitude of the signal in volts which is normalized in the left column of figures.

Returning Eye Movement Time Duration

One of the features of Algorithm 2 is the evaluation of sleep eye movement durations. This evaluation is one of the clinical concerns in sleep studies. Although one can create many

Figure 3.6: The Verification Plot of a Leftward Eye Movement

Figure 3.7: The Verification Plot of a Rightward Eye Movement

increased the complexity of Algorithm 1. Also, if there was no match between the objects, according to Table 3.1, the EMDS software would be crashed and locked in infinite loop. Furthermore, making vectors, if the number of objects is large, increases the complexity. Therefore, when the number of objects is large, the complexity of the detection procedure became faster.

Band Pass Filter

The other modification for making Algorithm 2 faster is using the band pass filter instead of the low pass filter in Algorithm 1. The band pass filter is used in the Scoring of Sleep and Associated Events reports of the Sleep and Associated Events software. The band pass filter is 0.3 – 35 Hz. However, the data, which were used in this research, were recorded with a bandpass filter of 0.3 – 64 Hz. As will be discussed in the next section, the band pass filter is used in the range of 0.3 – 20 Hz. The FIR filter with a low cutoff frequency 0.3 Hz and a high cutoff frequency 20 Hz, with an order of 200, was provided for filtering the signal.

Figure 3.8: The Verification Plot of an Upward Eye Movement

Verification Plots

Although some of the detected eye movements are not correct, the verification in Algorithm 1, Algorithm 2 has the verification feature. This feature can be used by clinicians for the verification of the detected eye movements. Figures 3.6, 3.7, 3.8, and 3.9 represent examples of verification plots for leftward, rightward, upward, and downward eye movements. The left column shows the original signal after processing and the right one shows the original signal. The unit of x-axis is sample. Farther, y-axis is the amplitude. The unit of y-axis is normalized in the left column of figures.

Returning Eye Movement Time Duration

One of the features of the EMDS software is the returning eye movement time duration. This evaluation is one of the clinical concerns in sleep studies. Although one can create many

Figure 3.9: The Verification Plot of a Downward Eye Movement

classes of time duration for sleep eye movements, clinically, the eye movements are categorized into two main categories: slow eye movements, and fast eye movements.

Algorithm 2 is able to derive the duration of an eye movement. The duration of an eye movement was derived for the detected eye movement. A detected eye movement, which is a signal passed through the filter, with its mean and the values above 3 SD removed, is known as a post-processed signal. Therefore, the zero crossing can be used to derive the duration of the detected movement. The AASM manual explains that the eye movements below 500ms in duration are considered fast movements; and above this time are considered slow movements [2]. Table 3.3 represents the duration examples of the detected eye movements.

Table 3.3: Eye Movement Duration Examples

21	68	27	7	141
15	48	37	57	64
123	21	86	53	5
17	1	7	32	92

Returning Other Eye Movement Specifications

This software also returns the other specifications of detected eye movements such as a time index, and maximum amplitude of the detected movements for each EOG channel.

3.2 Autoregressive (AR) Modelling

After the detection of eye movements, the Burg method was applied to derive the AR coefficients of each movement type (leftward, rightward, upward, and downward). After detecting the movements, the AR coefficients were derived using the Burg method. AR coefficients represent the waveforms of the signals. Therefore, AR coefficients can be used as features for the classification of the sleep eye movement signals for those patients who were exposed to medication and those who were not. The AR coefficients of all ten subjects (five patients who were exposed to Prozac medication and five patients who were not) were grouped on the basis of the sleep stages and the type of movements. This section addresses

how the AR modelling can be used for EOG analysis and how the optimum AR model order was derived for this purpose.

3.2.1 EOG Segmentation

It should be considered that EOG signals are non-stationary signals, because their statistical characteristics change over time. On the other hand, AR coefficients can be used for stationary signals. To solve the non-stationary problem of EOGs, in order to use them in AR modelling, the fixed segmentation, which is the small window of fixed duration, was used. This is the simplest method to break an EOG signal into quasi-stationary segments. The non-stationarity of sleep EOGs is discussed in Section 4.1.1 in more detail.

The time duration of sleep eye movements lasts from 30 samples, with a sampling rate of 128 samples per second, and to more than 700 samples in a few cases. The time duration evaluation, which was achieved by inspection, was done over the randomly chosen patients from each group ('A' who were exposed to Prozac, 'E' who were exposed to Effexor, 'L' who were exposed to Celexa, 'C' (Controls) who were not exposed to any antidepressant medication). Random epochs from all the stages of each patient were considered for the evaluation of sleep eye movement durations. Accordingly, a duration of 300 samples was chosen for segmenting the signal because most of the eye movements are in the range of 250–500 samples. Figure 3.10 represents examples of the sleep eye movement time durations from patient A1 in Stage 3. Figure 3.11 shows another sample of the same patient in Stage REM. Figure 3.12 shows the examples of the sleep eye movement time duration from patient C3 (from patients who were not exposed to any antidepressant medication, known as the Control group) in wakefulness. It should be considered that the segment including an eye movement is between the two marked datapoints. In order to derive the AR model of the EOG signals, the EOGs were segmented into 301 samples; 150 samples before and 150 samples after the detected movements as well as the detected sample, in order to be stationary or quasi-stationary.

Figure 3.10: The time duration of the sleep eye movements in samples (patient A1, Stage 3, y-axis unit is in volt)

how the AR modelling can be used for EOG analysis and how the optimum AR model order was derived for this purpose.

3.2.1 EOG Segmentation

It should be considered that EOG signals are non-stationary signals, because their statistical characteristics change over time. On the other hand, AR coefficients can be used for stationary signals. To solve the non-stationary problem of EOGs, in order to use them in AR modelling, the fixed sleep eye movement time duration was used.

Figure 3.11: The time duration of the sleep eye movements in samples (patient A1, Stage R, y-axis unit is in volt)

Control group) in wakefulness. It should be considered that the segment including an eye movement is between the two marked datatoints. In order to derive the AR model of the EOG signals, the EOGs were segmented into 301 samples: 150 samples before and 150 samples after the detected movements as well as the detected sample, in order to be stationary or quasi-stationary.

Figure 3.12: The time duration of the sleep eye movements in samples (patient C3, Wakefulness, y-axis unit is in volt)

3.2.2 Model Order Optimization

Two methods were used for evaluating the AR model order. One method of evaluating model order is to use the error plots and the other is to use the Akaike Information Criterion (AIC) method [108]. Both of them are discussed in this part.

Error Plots

The AR model order of detected eye movements was derived by inspection through minimum mean-square values of the prediction error. The optimal AR model was determined to be 27.

Bonnet modelled the EOG signal as an autoregressive of order 2 and estimated the parameters with a Kalman filter [109]; however, this model order may not be accurate for EOG analysis. Although the low model order may reduce the complexity of the procedure, it increases the error. The inspection over the minimum mean-square values was derived from the 175 plots. Figures 3.13 and 3.14 represent the mean-square error derived from AR modelling. In these plots, the mean square error was derived for different model orders, 1 to 50. The value of 27 was evaluated by averaging over the different chosen model orders.

AIC Criteria

The Akaike Information Criterion (AIC) [14] is a common method used for selecting an appropriate AR model order based on the input process. This method, which was discussed in Chapter 2 in detail, was used for evaluating the optimum AR model order. It is assumed that input process has Gaussian statistics. The AIC formula for an AR process is represented in Equation 3.1.

$$AIC = \ln(\varepsilon) + \frac{2p}{N} \quad (3.1)$$

where ε is the modelling error, p is the AR model order, and N is the number of data samples. The second term in Equation 3.1 is the penalty for use of extra AR coefficients. In other words, although increasing the model order leads to having a higher accuracy value in modelling the signal, it increases the complexity, which could be an important factor in

analyzing long data, such as 6 – 8 hours of sleep EOG signals. The minimum of the AIC curve corresponds to the optimum model order.

Figure 3.15 represents the AIC values versus AR model orders (1 to 100). Figure 3.15 specifies the AIC curves for ROC and LOC signals of leftward and rightward detected movements. In order to plot these curves, the AIC values were derived for model order 2 to 100 in steps of 2 ($p = 2 : 2 : 100$), separately for rightward and leftward movements. The length of the signal (N) was 300 samples (at a sampling rate of 123 samples per second) for the AIC evaluation. The derived AIC values of the rightward and leftward eye movements of

Figure 3.13: The error plot (patient A1, Stage 1, leftward eye movement)

Other segmentation values (N) such as 400 and 500 samples were tested and the same results were derived. Figures 3.16 and 3.17 represent the AIC curves for the segmentation of 400 and 500 samples respectively. Although the optimum model order is in a range of 27 – 37, which corresponds to the different minimum values of the AIC curve derived from the different signals, the model order of 27 is still a good choice for the EOG modelling.

3.2.2 Model Order Optimization

Two methods were used for evaluating the AR model order. One method of evaluating model order is to use the error plots and the other is to use the Akaike Information Criterion (AIC) method [108]. Both of them are discussed in this part.

Error Plots

The AR model order of detected eye movements was derived by inspection through minimum mean-square value of prediction error. The AR model order was determined to be

Figure 3.14: The error plot (patient A1, Stage2, rightward eye movement)

$$AIC = \ln(\epsilon) + \frac{2p}{N} \quad (3.1)$$

where ϵ is the modelling error, p is the AR model order, and N is the number of data samples. The second term in Equation 3.1 is the penalty for use of extra AR coefficients. In other words, although increasing the model order leads to having a higher accuracy value in modelling the signal, it increases the complexity, which could be an important factor in

analyzing long data, such as 6 – 8 hours of sleep EOG signals. The minimum of the AIC curve corresponds to the optimum model order.

Figure 3.15 represents the AIC values versus AR model orders (1 to 100). Figure 3.15 specifies the AIC curves for ROC and LOC signals of leftward and rightward detected movements. In order to plot these curves, the AIC values were derived for model order 2 to 100 in steps of 2 ($p = 2 : 2 : 100$), separately for rightward and leftward movements. The length of the signal (N) was 300 samples (at a sampling rate of 128 samples per second) for the AIC evaluation. The derived AIC values of the rightward and leftward eye movements of the ROC and LOC signals were then averaged to derive the final curves in Figure 3.15.

Figure 3.15: The AIC curves

Other segmentation values (N) such as 400 and 500 samples were tested and the same results were derived. Figures 3.16 and 3.17 represent the AIC curves for the segmentation of 400 and 500 samples respectively. Although the optimum model order is in a range of 27 – 37, which corresponds to the different minimum values of the AIC curve derived from the different signals, the model order of 27 is still a good choice for the EOG modelling.

Figure 3.16: The AIC curve for a patient who was exposed to Celexa derived from leftward movements, LOC, with segmentation of 400 samples

Figure 3.17: The AIC curve for a patient who exposed to Celexa derived from leftward movements, Stage 1, LOC, with segmentation of 500 samples

3.3 Classification Results

The goal of this research is to present the differences in the waveform of the sleep eye movement signals in response to antidepressant medication, which is not evident to clinicians in routine sleep studies. Measures of sleep physiology not obvious to the human eye may provide important clues to disease states and response to therapy.

After deriving the AR coefficients of the eye movements recorded during sleep, Linear Discriminant Analysis (LDA) was used to represent the difference between the eye movement waveforms of the signal of the patients who were exposed to antidepressant medications and those who were not.

In this section, LDA was used for two major comparisons. One comparison was done between the sleep eye movements of a group of five depressed patients, taking Prozac, with the generic name of Fluoxetine, and a group of five depressed patients taking no antidepressant medication. Another comparison was made between two other groups of five patients who were exposed to antidepressant medications, Celexa, with the generic name of Citalopram, and Effexor, with the generic name of Venlafaxine, with five patients who were not exposed to antidepressant medications (Controls).

3.3.1 AR Classification of Patients Exposed/Unexposed to Prozac

This section describes a classification of AR coefficients derived from a group of five patients who used Prozac medication, and a group of five patients who did not. AR coefficients of eye movements recorded during sleep were derived and classified by means of discriminant analysis. The methodology used for this classification included eye movement detection and the evaluation of AR coefficients with the fixed segments. The optimized model order of 27 was derived by minimum square error inspection. AR coefficients were then derived on the basis of this optimum value. The results were classified with a linear discriminant function. The overall average accuracy was 82.8%. An accuracy result of a classification represents the portion of the correct assigned number of cases into their classes to the all number of cases. Also, this accuracy represents the differences between the waveforms of the sleep eye

movements of the patients who were exposed to Prozac medication and those who were not.

In other words, after evaluating the AR coefficients, the AR coefficients of all ten patients, five patients who were not exposed to antidepressant medication ($N1$ to $N5$), and five patients who were exposed to Prozac medication ($A1$ to $A5$), were grouped on the basis of the sleep stages and the type of movements.

LDA was applied to classify AR coefficients into normal and abnormal. This research is only focused on the results of left and right movement classification for all the sleep stages. The sensitivity, specificity, and efficiency are shown in Tables 3.4, 3.5, 3.6, 3.7, and Figures 3.18, 3.19, 3.20, 3.21. In deriving accuracy, $P(D^+)$ and $P(D^-)$ are both approximately 0.5 because the numbers of comparisons in both groups are approximately equal.

Table 3.4: The classification results of LOC leftward AR coefficients

Stages	Sensitivity (%)	Specificity (%)	Accuracy (%)
Stage 1	65.8	66.1	67.3
Stage 2	74.0	66.0	70.0
Stage 3	95.1	94.9	95.0
Stage 4	99.0	96.0	97.5
Stage REM	79.5	84.7	82.1
Wakefulness	74.6	70.4	72.5

Table 3.5: The classification results of ROC leftward AR coefficients

Stages	Sensitivity (%)	Specificity (%)	Accuracy (%)
Stage 1	67.8	77.5	72.7
Stage 2	87.6	79.9	83.8
Stage 3	93.5	93.1	93.3
Stage 4	96.6	89.1	92.9
Stage REM	74.6	95.0	84.8
Wakefulness	79.3	72.6	76.0

Table 3.6: The classification results of LOC rightward AR coefficients

Stages	Sensitivity (%)	Specificity (%)	Accuracy (%)
Stage 1	75.8	76.2	76.0
Stage 2	71.9	84.5	78.7
Stage 3	96.5	93.8	95.2
Stage 4	100	98.2	99.1
Stage REM	77.9	71.0	74.5
Wakefulness	72.9	68.6	70.8

Table 3.7: The classification results of ROC rightward AR coefficients

Stages	Sensitivity (%)	Specificity (%)	Accuracy (%)
Stage 1	77.7	72.1	74.9
Stage 2	72.0	84.8	78.4
Stage 3	96.5	88.5	92.5
Stage 4	97.2	96.4	96.8
Stage REM	77.0	97.5	87.3
Wakefulness	78.4	73.4	75.9

Figure 3.18: The classification results of LOC leftward AR coefficients

Figure 3.19: The classification results of ROC leftward AR coefficients

Figure 3.20: The classification results of LOC rightward AR coefficients

Figure 3.21: The classification results of ROC rightward AR coefficients

3.3.2 AR Classification of Patients Exposed/Unexposed to Celexa and Effexor

This section describes the method used to demonstrate that the eye movements were different in patients who used antidepressant medications, compared to those who did not. Groups 1 and 2 included five patients who used Celexa and Effexor, respectively, compared to five patients not taking any antidepressant. This method included eye movement detection with eye movement detection software, and evaluation of AR coefficients with fixed segments. The AIC method was used for determination of an appropriate model order of 27. AR coefficients were derived on the basis of this optimum value and were then classified with a linear discriminant function. These coefficients represent the waveform of the sleep eye movements of all three groups. Figure 3.22 represents the distribution of three AR coefficients (out of 27 AR coefficients) that were used for linear discriminant classification. The overall average of the regular method accuracies were 76.4%, and 78.7% for Groups 1 and 2 respectively. The overall average of the leave-one-out method accuracies were 75.5% and 77.5% for Groups 1 and 2.

Figure 3.22: Distribution of three AR coefficients

Figures 3.23, 3.24, 3.25, and 3.26 are the leave-one-out method accuracy results of Groups 1 and 2 classifications. Figures 3.23, and 3.24 show the accuracy results of the classification of the AR coefficients of leftward and rightward eye movements for Group 1. In these cases, the patients did not experience Stage 4 sleep. The lower accuracy in Stage wakefulness and REM may be due to the fact that there are more fast movements in Stage wakefulness and REM which may be more easily corrupted by artifacts.

Figures 3.25 and 3.26 show the leave-one-out accuracies for Group 2. It was possible to obtain these better results for Stage 3 and Stage 4 because of better detection of eye movements in these stages. The overall average of the regular method accuracies were in the range of 72% to 80%. The overall average of the leave-one-out accuracies were in the range of 71% to 80%. Tables 3.8, 3.9, 3.10, and 3.11 show the accuracy results of Figures 3.23, 3.24, 3.25, and 3.26 in detail.

Table 3.8: The leave-one-out method accuracies (%) of the classification of Group 1 leftward eye movements' AR coefficients

Figure 3.23: Group 1 leftward eye movement AR coefficient accuracies

Table 3.10: The leave-one-out method accuracies (%) of the classification of Group 2 leftward eye movements' AR coefficients

Figure 3.24: Group 1 rightward eye movement AR coefficient accuracies

Figure 3.25: Group 2 leftward eye movement AR coefficient accuracies

Figure 3.26: Group 2 rightward eye movement AR coefficient accuracies

Table 3.8: The leave-one-out method accuracies (%) of the classification of Group 1 leftward eye movements' AR coefficients

Stages	LOC	ROC
Stage 1	76.0	80.1
Stage 2	69.4	80.2
Stage 3	78.8	80.0
Stage REM	64.4	72.4
Wakefulness	68.9	75.8

Table 3.9: The leave-one-out method accuracies (%) of the classification of Group 1 rightward eye movements' AR coefficients

Stages	LOC	ROC
Stage 1	76.5	79.6
Stage 2	69.4	82.2
Stage 3	86.6	88.1
Stage REM	65.4	71.1
Wakefulness	72.9	76.9

Table 3.10: The leave-one-out method accuracies (%) of the classification of Group 2 leftward eye movements' AR coefficients

Stages	LOC	ROC
Stage 1	71.3	73.7
Stage 2	79.1	75.7
Stage 3	89.3	86.8
Stage 4	87.6	77.2
Stage REM	73.0	76.6
Wakefulness	65.2	69.2

Table 3.11: The leave-one-out method accuracies (%) of the classification of Group 2 rightward eye movements' AR coefficients

Stages	LOC	ROC
Stage 1	72.3	73.1
Stage 2	79.9	81.6
Stage 3	91.0	85.8
Stage 4	85.4	88.0
Stage REM	68.5	73.9
Wakefulness	66.3	69.7

3.3.3 Receiver Operating Characteristic (*ROC*) Curve, Method I

One of the criteria that can be used in evaluating a method is Receiver Operating Characteristic (*ROC*). The theory of the Receiver Operating Characteristics (*ROC*) curve was described in Chapter 2. Here, a method which was used for evaluating the Receiver Operating Characteristic (*ROC*) curve of the AR modelling is described.

Suppose there are two matrices of AR coefficients N and A . Both matrices comprise AR coefficients derived from fixed segments (301 samples), including detected movements (150 samples before and 150 samples after a detected movement). Matrix N was derived from normal patients and Matrix A from abnormal patients. Each row of both matrices consists of 27 AR coefficients derived from a detected movement.

At the first step, A and N are normalized over each row of the matrix. Thereafter, the minimum and maximum of matrix N and A were evaluated for columns of normalized matrices. Therefore, each column of A or N has a minimum, a maximum, and a range (the distance between the minimum and the maximum). There are six possibilities in comparisons between the range of each column of A and N :

1. The range of A is within the range of N ;
2. The range of N is within the range of A ;
3. The range of N has overlap with the range of A , and the range of N is higher than A ;
4. The range of N has overlap with the range of A , and the range of N is lower than A ;
5. The range of N does not overlap with the range of A , and the range of N is higher than A ;
6. The range of N does not overlap with the range of A , and the range of N is lower than A .

Also, the mean value of normalized matrices (A and N) are derived for each column. Then, the size of two matrices is adjusted, if they are not the same size, by cutting the larger

matrix. Suppose the k th row of matrix N is considered, each element of this row can be referred to as $value(i, k)$. Also, the distance between a higher maximum and lower minimum of each column in A and N are derived. The step size is derived by dividing the distance by 100. As one can see, the step size will differ from column to column corresponding to the range of the distance.

If one considers the k th column of A and N , the threshold value corresponding to this column starts from the minimum of the lower range and increases to the maximum of the higher range incrementally with step size.

The decision vector is made by comparing $value(i, k)$ and its corresponding threshold for each row of A and N . Cases 1 and 2 are laid aside because the normal range is within the abnormal, or vice versa, and one cannot distinguish normal from abnormal. Other cases should be as follows:

- In case 3: if $value(i, k) < threshold$, then $decision_vector(1, k) = -1$;
- In case 4: if $value(i, k) < threshold$, then $decision_vector(1, k) = +1$;
- In case 5: if $value(i, k) < threshold$, then $decision_vector(1, k) = -1$;
- In case 6: if $value(i, k) < threshold$, then $decision_vector(1, k) = +1$.
- Otherwise $decision_vector(1, k) = 0$

Therefore, in a decision vector -1 represents the abnormal and $+1$ represents the normal. By averaging the decision vector, one can determine whether the case (derived from each row of A and N) is normal or abnormal. The decision factors, True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), are derived using the following decision rules:

- If it is picked from N and considered normal, add one to TP counter;
- If it is picked from N and considered abnormal, add one to FN counter;
- If it is picked from A and considered abnormal, add one to TN counter;

- If it is picked from A and considered normal, add one to FP counter.

By evaluating the decision factors, the sensitivity and specificity are evaluated on the basis of Equations 3.2, and 3.3.

$$Sensitivity = \frac{TP}{TP + FN} \quad (3.2)$$

$$Specificity = \frac{TN}{TN + FP} \quad (3.3)$$

Sensitivity is plotted versus $(1 - Specificity)$ for different thresholds and the Receiver Operating Characteristic (ROC) plot is derived. Figure 3.27 represents the Receiver Operating Characteristic (ROC) curve of the classification of the AR coefficients belonging to the patients who were exposed to Effexor and those who were not exposed to any antidepressant medication (Controls). The reason that this curve does not end on $(1, 1)$ is that the decision vector is used. If there were only one case, the decision factors (TN, TP, FP, and FN) would have a consistent accumulation pattern. In this case, the average of some of the cases become zero, which leads to removal of the case from Receiver Operating Characteristic (ROC) evaluation. Therefore, the Receiver Operating Characteristic (ROC) curve may not reach $(1, 1)$.

3.3.4 Receiver Operating Characteristic (ROC) Curve, Method II

In this method, the Receiver Operating Characteristic (ROC) curve was derived by computing the performance curve for classifier output. In this method, one EOG channel (for example LOC) for one type of movement such as leftward was chosen for the exposed and unexposed classification. The performance curve returns the TP and FP as outputs for plotting the classification performance. The area under the curve (AUC) is equal to the probability of discrimination; that is, the ability of the test to correctly differentiate those who were exposed to antidepressant medication from those who were not. The performance of a classification is measured by AUC. An area of 1.00 represents a perfect classification and 0.50 represents a worthless one. The AUC ranges are ranked as follows,

Figure 3.27: ROC Curve with Method I

- 0.90 – 1.00 = Excellent;
- 0.80 – 0.90 = Good;
- 0.60 – 0.70 = Poor;
- 0.50 – 0.60 = Fail.

The AUC of Figure 3.28 is 82.17%. In conclusion, the Receiver Operating Characteristic (*ROC*) curve has evaluated the autoregressive modelling method with larger than 80% area under the curve, which is in the *Good* range.

Figure 3.28: Receiver Operating Characteristic (*ROC*) Curve with Method II

Chapter 4

Time, Frequency, and Wavelet Analysis for Quantifying Sleep Eye Movements

As discussed in Chapter 1, a significant amount of data is not evident clinically in sleep studies. Measures of sleep physiology, not immediately apparent, may provide important clues to diseases and responses to therapy. As stated, it has been noted that eye movements change significantly in the patients exposed to antidepressant medications. This chapter describes how eye movements were different in depressed patients who used antidepressant medications, compared to those who did not, by means of wavelet analysis.

Electrooculogram (EOG) provides a significant amount of data which can be studied by dividing it into different frequency sub-bands for analysis. The same sub-bands can then be compared with the same group of patients, in different sleep stages, or between different groups of patients. Specifically in this research, the focus is on the comparison between patients who were exposed to antidepressant medication and those who were not. In this chapter, the following EOG analyses, which are proposed by this research, will be discussed:

- Time-Domain Analysis;
- Frequency-Domain Analysis;
- Time and Frequency Analysis;

- Scalogram Analysis;
- Slow Eye Movement Frequency Analysis Using Discrete Wavelet Transform;
- Feature Extraction Using Discrete Wavelet Transform.

In this chapter, two groups of five patients who were exposed to Prozac (one group called 'A', the other called 'P'), one group of five patients who were exposed to Celexa (called 'L'), one group of five patients who were exposed to Effexor (called 'E'), one group of five patients who were not exposed to antidepressant medication (called 'N'), and another group of eight patients who were not exposed to antidepressant medication (called 'C') were considered. Due to confidentiality of the patient information, there was no access to more information of patients such as name, gender, and age.

4.1 Time-Domain Analysis

In the early stage of this study, the simple method time-domain analysis of the eye movements has been used for a better understanding of EOG signals rather than analyzing the signals with advanced tools such as wavelet or Short Time Fourier Transform analysis. In the beginning, the attempt was to understand why the reference signals (mastoid signals, A1 and A2) were used. The mastoid signals represent artifacts. As was discussed in the previous chapters, the differences between LOC, ROC, and VOC and the mastoid signals (LOC-A2, ROC-A1, and VOC-A1) were used in order to cancel the artifacts of the EOG signals. However, the amplitudes of the reference signals are subtracted from those of the EOG signals in such a way that the location of the electrode for each reference signal is at the greatest distance from the location of the EOG. For measuring the similarities between mastoid channels and EOG channels, the crosscorrelation between one of the EOG signals and one of the mastoid signals has been derived. Figures 4.1, 4.2, 4.3, and 4.4 represent the crosscorrelation of the LOC, and ROC signals with reference signals. As can be seen in the figures, the crosscorrelation is maximum at the time of zero. This confirms the fact that

subtracting these channels without any time shifting results in a cleaner EOG signal, than with a time shifting.

Figure 4.1: The crosscorrelation of LOC and Mastoid signal (A1) derived from a patient N4 who was not exposed to antidepressant medication in Stage REM (x-axis is time in samples and y-axis is the crosscorrelation value).

Further, the comparisons between the maximum crosscorrelation of the LOC, ROC, and VOC with one of the mastoid signals for those who were not exposed to antidepressant medication (called Normal patients) and those who were exposed to Prozac (called Abnormal patients) have been derived. The results are illustrated in Figure 4.5. The comparisons between the maximum crosscorrelation of the LOC with mastoid signals in different sleep stages for Normal and Abnormal patients have been derived. The results are illustrated in Figure 4.6.

Also, the comparison between the maximum crosscorrelation of the ROC and VOC with mastoid signals for the same patients at different sleep stages was derived. The simple method time-domain analysis continued with statistical analyses such as mean and variance. The mean and variance were evaluated for LOC, ROC, VOC, and mastoid signals. The results are plotted in Figure 4.7. (which is an example of all EOG channels) are shown in

Figure 4.2: The crosscorrelation of LOC and Mastoid signal (A2) derived from a patient N4 who was not exposed to antidepressant medication in Stage REM (x-axis is time in samples and y-axis is the crosscorrelation value).

Figure 4.3: The crosscorrelation of ROC and Mastoid signal (A1) derived from a patient N4 who was not exposed to antidepressant medication in Stage REM (x-axis is time in samples and y-axis is the crosscorrelation value).

Figure 4.4: The crosscorrelation of ROC and Mastoid signal (A2) derived from a patient N4 who was not exposed to antidepressant medication in Stage REM (x-axis is time in samples and y-axis is the crosscorrelation value).

In conclusion, the following Tables 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, were derived from the comparing of the crosscorrelation of EOG and mastoid channels. The compared maximum crosscorrelations, which are shown in Figure 4.5, were counted, and the more frequent of the higher maximum crosscorrelations of each pair in any group of five patients were determined. If the maximum crosscorrelation is more frequent for LOC–A1 among the five patients than LOC–A2, it will then be shown in the Table as LOC–A1 > LOC–A2.

Table 4.1: The Crosscorrelation Comparison

Crosscorrelation	
Normal	LOC–A1 > LOC–A2
Abnormal	LOC–A1 < LOC–A2
Normal	ROC–A1 < ROC–A2
Abnormal	ROC–A1 < ROC–A2
Normal	VOC–A1 < VOC–A2
Abnormal	VOC–A1 $\ll$ VOC–A2

The results from Figure 4.6 (which is an example of all EOG channels) are shown in (A2) for five patients who were not exposed to antidepressant medication (N) and five patients who were exposed to Prozac, in different sleep stages.

Figure 4.5: The maximum crosscorrelation of the LOC, ROC, and VOC compared with Mastoid signals (A1 and A2) for five patients who were not exposed to antidepressant medication (N) and five patients who were exposed to Prozac.

Figure 4.6: The maximum crosscorrelation of the LOC compared with Mastoid signals (A1 and A2) for five patients who were not exposed to antidepressant medication (N) and five patients who were exposed to Prozac, in different sleep stages.

Figure 4.7: The mean value and variance of LOC, ROC, VOC, and Mastoids (A1 and A2) for Normal subjects (five patients who were not exposed to antidepressant medication) and Abnormal subjects (five patients who were exposed to Prozac medication)

Tables 4.2, 4.3, 4.4 where the LOC, ROC and VOC with the reference signals were compared for normal and abnormal patients in different sleep stages. In those tables, the more frequent maximum crosscorrelations were determined, and if the the maximum crosscorrelation of an A1 and EOG signal was more frequent than that of an A2 and EOG signal, then it is shown that $A1 > A2$ (See Table 4.2).

Table 4.2: The Crosscorrelation Comparison of LOC and Mastoid Channels

LOC		
Stage	Normal	Abnormal
St1	$A1 < A2$	$A1 < A2$
St2	$A1 > A2$	$A1 < A2$
St3	$A1 > A2$	$A1 > A2$
St4	$A1 > A2$	$A1 > A2^a$
REM	$A1 > A2$	$A1 > A2^b$

^aThere is only one result in the chart.

^bFor all the results in the chart

Table 4.3: The Crosscorrelation Comparison of ROC and Mastoid Channels

ROC		
Stage	Normal	Abnormal
St1	$A1 < A2^a$	$A1 < A2$
St2	$A1 < A2^a$	$A1 < A2$
St3	$A1 < A2$	$A1 < A2$
St4	$A1 < A2$	$A1 < A2^b$
REM	$A1 < A2$	$A1 < A2$

^aFor all the results in the chart

^aFor all the results in the chart

^bThere is only one result in the chart.

In Tables 4.5 and 4.6, the maximum crosscorrelation of EOG signals with each reference signal was compared. The more frequent maximum crosscorrelations among the patients were determined. If the maximum crosscorrelation of each EOG signal is more frequent than that of another EOG signal, then that signal is shown as greater than the other EOG; e.g. $LOC > ROC > VOC$. For greater clarity, the crosscorrelations between the normal and

Table 4.4: The Crosscorrelation Comparison of VOC and Mastoid Channels

VOC		
Stage	Normal	Abnormal
St1	$A1 < A2$	$A1 < A2^a$
St2	$A1 < A2$	$A1 < A2$
St3	$A1 > A2$	$A1 < A2$
St4	$A1 > A2$	$A1 < A2^b$
REM	$A1 < A2$	$A1 < A2^a$

^aFor all the results in the chart

^bThere is only one result in the chart.

^aFor all the results in the chart

abnormal patients in different sleep stages were compared.

Table 4.5: The Crosscorrelation of Mastoid Signals with EOG for Normal Patients (those who were not exposed to Prozac medication)

Crosscorrelation		
Stage	Normal/A1	Normal/A2
St1	$VOC > LOC > ROC$	$VOC > ROC > LOC$
St2	$VOC > LOC > ROC$	$VOC > ROC > LOC$
St3	$LOC > VOC > ROC$	$VOC = LOC > ROC$
St4	$LOC > VOC > ROC$	$VOC > LOC > ROC$
REM	$VOC > LOC > ROC$	$VOC > ROC > LOC$

In Tables 4.7 and 4.8, the mean value and the variance of the EOG signals and reference signals are compared.

4.1.1 Non-stationarity of the EOG Signals and Its Solution

As was stated, the EOG signals are non-stationary. The non-stationarity of the signals was investigated by applying them to an adaptive filter such as the Recursive Least Square (RLS) algorithm. Figure 4.8 illustrates the learning curve of the RLS algorithm. Theoretically, the learning curve represents the output error of the adaptive filter. If the signal is stationary, the learning curve converges to a constant value which is the noise variance. However, Figure 4.8 shows that the learning curve converges to multiple values; thus, the EOG signal can not

Table 4.6: The Crosscorrelation of Mastoid Signals with EOG for Abnormal Patients (those who were exposed to Prozac medication)

Crosscorrelation		
Stage	Abnormal/A1	Abnormal/A2
St1	LOC > ROC > VOC	VOC > LOC > ROC
St2	LOC > ROC > VOC	VOC > LOC > ROC
St3	LOC > VOC > LOC	VOC = LOC = ROC
St4	LOC > VOC > ROC ^a	LOC > VOC > ROC
REM	LOC > ROC > VOC	VOC > ROC > LOC

^aThere is only one result in the chart.

Table 4.7: Mean Value Comparison

Mean Value	
Normal	A1 > A2
Abnormal	A1 < A2
Normal	ROC > LOC & VOC
Abnormal	ROC & VOC > LOC

Table 4.8: Variance Value Comparison

Variance	
Normal	A1 = A2
Abnormal	A1 > A2
Normal	VOC > ROC > LOC
Abnormal	LOC > ROC > VOC

be a stationary signal; therefore, it must be considered as a non-stationary signal. As was discussed in Chapter 2, solving the non-stationary problem of EOG signals can be achieved by signal segmentation. In that chapter, the method of fixed segmentation in which a non-stationary signal is divided into stationary or approximately stationary parts was explained; the signal should be segmented to fixed small parts whose lengths are much smaller than the length of the original signal. By this method, the produced small segments are stationary or quasi-stationary. This method was used to solve the non-stationarity of EOG signals in this research.

Figure 4.8: The learning curve of the RLS algorithm for an EOG signal (x-axis is time in samples and y-axis is error).

4.2 Frequency-Domain Analysis

The AASM Manual for the scoring of sleep and associated events recommends using the bandpass filter (0.3 to 35Hz) for EOG signals [2]. However, as one can see in Figure 4.9, the range of frequency of the EOG signals which are processed in this research extends from zero to 64Hz. This is because of the fact that these signals have been recorded with a sampling rate of 128 samples per second, which is twice the maximum frequency, 64Hz, pursuant to the Nyquist sampling theorem [107].

For instance, one of the EOG signals, from a patient exposed to Prozac medication, has been plotted in time-domain and frequency-domain, in Figure 4.9. The top one is the time-domain representation of the signal. The middle figure represents the spectrum of the signal below the frequency of 5Hz; the first few frequency samples have been removed for better representation of the signal, due to having large peak values at DC. The bottom figure shows the complete spectrum of the signal, in which the maximum frequency is 64Hz. In addition, Figure 4.9 represents a raw signal, without any processing. By considering the different spectral plots, it is obvious that the main frequency information exists below 20Hz.

4.3 Time and Frequency Analysis

Frequency information in particular time is extracted by means of wavelet analysis [94]; thus, one is able to assess the transient information in EOG signals. This information can be used, first to understand the nature of the EOG signal, and second to compare signals of patients exposed to the medication with signals of those not exposed to the medication in order to distinguish the effects of antidepressant medication on eye movement signals. In this regard, wavelet transforms were applied to EOG signals in three methods: scalogram analysis, slow eye movement analysis, and feature extraction.

Figure 4.9: The EOG signal for a patient exposed to Prozac medication has been plotted in time-domain and in frequency-domain. The spectrum of the signal from 0 to 5 Hz is shown in the middle one. The bottom figure shows a complete spectrum of the signal. (Maximum frequency = 64Hz; the y-axis unit is in volt)

4.4 Scalogram Analysis

The wavelet energy measuring procedure is called scalogram analysis, which measures the energy of a signal in a time-frequency neighborhood of f_i and t_i located in a resolution box, Heisenberg Box [94].

The scalogram of the signal was considered. The scalogram plots the power of each time-frequency box for different scales. Figure 4.10 represents five seconds of LOC belonging to patient C1, who was not exposed to the antidepressant medication (Stage 1, samples 10000:10640), and Figure 4.11 shows a example of an amplitude scalogram.

Figure 4.10: The amplitude scalogram of 5-second EOG signal (the sample is the unit of x-axis).

The scalogram analysis was done with the the help of Log-Gabor filters, provided by Zhiatao [98]. Gabor filters are famous for their application in the localization of spatial and frequency information. Due to the Gabor filter's limitation in maximum bandwidth and because of not being optimal for broad spectral information, the Log-Gabor function was proposed. On the other hand, Log-Gabor filters can have an arbitrary bandwidth characteristic; therefore, they are used to derive the amplitude of the scalogram.

Figure 4.11: The amplitude scalogram of five second EOG signal. The energy distribution measurement has been illustrated in the right column of the graph starting from blue (0) to red dark (0.016)

The EOG signal was segmented to fixed 5-second epochs without respect to eye movements. Thereafter, the scalogram amplitude matrix of these fixed segments was evaluated. As was discussed theoretically in Chapter 2, the elements of the rows of the amplitude matrix represent the power of each time-frequency box (Heisenberg rectangle) for every scale. To find out the maximum energy for every scale, first the mean value of each row of the amplitude scalogram matrix was derived; and second, the maximum mean value was chosen. Consequently, the maximum power and its corresponding scales were found with the help of these mean values for the whole scalogram amplitude matrix.

The scalogram amplitude matrices for all fixed 5-second epochs in every sleep stage were studied. The most frequent scale corresponding to maximum power was chosen for each stage; then the maximum power values of this scale were extracted from all the 5-second epochs in a sleep stage. These maximum power values of each pair of the sleep stages for a patient were determined; they were compared with the help of a classifier. A linear discriminant classifier was used in this analysis, in which the leave-one-out method is used

to determine the accuracy of the classification. For example, the maximum power values of Stage 1, S1, were compared to those of the Stage 2, S2, and classified. The accuracy percentages of the classification of different sleep stages then were plotted in Figure 4.12. The power results were used to compare the energy between stages of three patients who were not exposed to antidepressant medications.

Figure 4.12: The leave-one-out method classification accuracy results compared among stage energies of three unexposed patients derived from the scalogram.

The accuracy value of about 50% in patient C3 shows that there is not much difference between the power results in different stages for this patient. However, the accuracy values of about 70% of two other patients for power results in contrasting Stage 3 with Stages 1, 2, and REM are to encourage one to check this finding for other patients. Table 4.9 presents the precise accuracy results of Figure 4.12.

Five other patients who were not exposed to antidepressant medications, (C4, C5, C6, C8, and C9), were considered for the same comparison, and the leave-one-out accuracy results are displayed in Figure 4.13.

The approximately 70% accuracy result in classifying Stages 2 and 3 reveal the fact that

Table 4.9: The classification accuracy results of all stage comparisons of maximum energies derived by scalogram analysis for patients who were unexposed to antidepressant medication

Stages	C1 (%)	C2(%)	C3(%)
S1_S2	55.8	58.4	54.4
S1_S3	80.5	87.4	45.9
S1_SR	59.5	54.5	49.3
S2_SR	64.8	50.1	55.9
S3_SR	72.7	75.7	59.4
S3_S2	86.1	72.6	61.2

Figure 4.13: The leave-one-out method classification accuracy results compared among three stage energies of five unmedicated patients derived from the scalogram.

these two stages are different in terms of maximum energy for the group of patients who were not exposed to antidepressant medication. The precise results of Figure 4.13 are shown in Table 4.10. In that table, the leave-one-out classification accuracy results have been shown, and Stage 3 is compared with other stages (1, 2, and REM) on maximum energies, derived by scalogram analysis from patients who were unexposed to antidepressant medication.

Table 4.10: The accuracy results of maximum energy comparisons for Stage 3 with other stages for patients unexposed to antidepressant medication.

Stages	C4 (%)	C5(%)	C6(%)	C8(%)	C9(%)
S3_S1	71.1	74.1	75.1	60.8	52.6
S3_S2	80.8	68.2	74.7	81	69.5
S3_SR	64.1	46.1	78	76.5	68.5

The study continued to find out whether such a difference in maximum energy between Stage 3 and other stages for the groups exposed to medication existed. In this regard, three groups of patients were considered; the patients were exposed to three types of antidepressant medications: Effexor, Celexa, or Prozac.

First, the comparison method to find out the energy difference between Stage 3 and other stages (Stages 1, 2, and REM) by this method was applied to those who were exposed to Effexor medication. The results given in Figure 4.14 show that the energy difference is lower than for those patients who were not exposed to any antidepressant medication. Once more, it should be emphasized that this method shows the difference between maximum energy, derived by scalogram, and it does not show which group has more or less energy; it just shows that they are different. The reason for this comparison is to find out the invisible differences between those who were exposed to antidepressant medication and those who were not.

The patients of this comparison were chosen randomly from the group exposed to Effexor medication.

The precise results of Figure 4.14 are displayed in Table 4.11, where the leave-one-out classification accuracy results are shown.

Figure 4.15 shows the results of the leave-one-out classification of Stage 3 and other

Figure 4.14: The leave-one-out method classification accuracy results compared among three stage energies of five patients exposed to Effexor medication, derived from the scalogram.

Table 4.11: The accuracy results of maximum energy comparisons for Stage 3 with other stages, for patients who were exposed to Effexor medication.

Stages	E2 (%)	E3(%)	E4(%)	E5(%)	E7(%)
S3_S1	33.3	64.1	47.9	57.2	59.8
S3_S2	68.8	86.3	62.7	43.6	67.6
S3_SR	54.5	66.9	48.4	48	43.6

stage (1, 2, and REM) maximum energies derived from the scalogram for patients who were exposed to the antidepressant Celexa.

Figure 4.15: The leave-one-out method classification accuracy results compared among three stage energies of five patients exposed to Celexa medication.

It can be considered, the patients who were exposed to Celexa do not have as low accuracy results as those who were exposed to Effexor. It means that there exist differences between the Stage 3 and other stage (1, 2, and REM) energies derived from the scalograms. The precise results of Figure 4.15 are presented in Table 4.12. In this table, the leave-one-out classification accuracy results are shown.

Table 4.12: The accuracy results of maximum energy comparisons for Stage 3 with other stages, for patients who were exposed to Celexa medication.

Stages	L1 (%)	L3(%)	L4(%)	L7(%)	L8(%)
S3_S1	75.3	44.2	45.8	47.3	70
S3_S2	86.8	59.6	39.7	81.2	69.7
S3_SR	59.1	74.1	45.9	41.2	41.8

The same method was applied to patients who were exposed to the antidepressant Prozac and the results are shown in Figure 4.16 and Table 4.13. It is obvious that these results are

close those of the patients who were exposed to Celexa; i.e., there exist differences between the Stage 3 and other stage (1, 2, and REM) energies derived from the scalograms.

Figure 4.16: The leave-one-out method classification accuracy results compared among three stage energies of three patients exposed to Prozac medication.

Table 4.13: The accuracy results of maximum energy comparisons for Stage 3 with other stages, for patients who were exposed to Prozac medication.

Stages	P3 (%)	P4(%)	P5(%)
S3_S1	75.8	61.9	43.3
S3_S2	85.1	71.7	69.7
S3_SR	62.5	64.4	42.2

4.4.1 Discussion

In conclusion, by deriving the mean value of the difference between energies represented by the accuracies for each group of patients, about 70% of the accuracies Stage 3 are different from those of Stage 2 for all groups. The comparison data are displayed in Figures 4.17, and 4.18, and in Table 4.14. In addition, Table 4.14 shows the higher differences for patients who were unexposed to medication as compared to those who were exposed to any

antidepressant medication. In comparison of Stage 3 with Stages 1 and R (REM), the difference was higher for those who were not exposed to the medication than for those who were exposed. As was discussed, patient C3 in the group unexposed to medication is possibly the exception in this group, perhaps due to the medical background, data recording, or other medical or technical problems. The results were reevaluated without C3 as well. The classification results of unexposed patients were increased by ignoring this patient's data.

Figure 4.17: The mean value of accuracy results for patients exposed (E, L, and P) and unexposed (C) to medication.

Table 4.14: The mean values of the accuracy results for patients unexposed (C) and exposed (E, L, and P) to medication (C3 is a patient who was not exposed to antidepressant medication).

Stages	C (%)	C without C3 (%)	E(%)	L(%)	P(%)
S3_S1	68.44	71.66	52.46	56.52	60.33
S3_S2	74.26	76.13	65.76	67.40	75.50
S3_SR	67.62	68.80	52.28	52.42	56.37

Figure 4.18: The mean value of accuracy results for patients exposed and unexposed to medication (without patient C3).

4.5 Slow Eye Movement Frequency Analysis Using Discrete Wavelet Transform

The EOG is a useful indicator of sleep onset. The occurrence of slow eye movements in the EOG is a sign of sleepiness. Slow Eye Movement (SEM) activity starts in pre-sleep wakefulness and linearly increases in this stage. It continues in Stage 1 and declines progressively in the first minutes of Stage 2 [110].

In this research, the detection of SEMs for these comparisons are not of interest. Pre-sleep wakefulness and Stage 1 EOG signals are considered in order to find singularities that were demonstrated in the differences between the groups who were exposed to antidepressant medication and those who were not.

4.5.1 SEM Analysis with Consecutive Fixed Segmentation

The SEM activity analysis is based on the wavelet multiresolution analysis of sleep EOG signals. The fixed segmentation was used without respect to eye movements. In the method

proposed by Magosso, the EOG activities during the wake-sleep transition were analyzed to detect SEM by wavelet [111]. SEM is one of the characteristics of the wake-sleep transition; therefore, it is suitable to be studied for enhancing the methods that prevent drowsiness during driving and for the diagnosing of several pathological conditions [111]. By following Magosso's method, the difference between the two EOG channels $[(ROC-A1) - (LOC-A2)]$ was decomposed by a discrete wavelet decomposer. A mother wavelet, Daubechies of order 4, proposed by Magosso, was used in this method because it resembles SEM [111]. The decomposition was derived for ten levels. The detail and approximation coefficients derived from each level of decomposition were used for classification. The wavelet theory which support this method is explained in detail in Chapter 2. In this classification, linear discriminant analysis was applied for classifying those who were exposed to the medication and those who were not. The proposed method is presented in the block diagram of Figure 4.19.

Figure 4.19: The block diagram of the EOG wavelet decomposition with Daubechies of order 4 into 10 levels for classifying those who were exposed to antidepressant medication and those who were not.

Figure 4.20 represents a SEM activity. This is detected in Stage 1 of patient C1 (from the Control group who did not use the antidepressant medication). The LOC, ROC, and the difference between LOC and ROC are plotted.

Eye movement is detected as SEM by the following clinical criteria:

- Slow sinusoidal excursion (0.2-0.6 Hz) lasts more than one second [111];
- Its amplitude is between 20 to 200 μV [111];

Figure 4.20: An example of a SEM, detected between the marked points in LOC, ROC, and ROC - LOC.

- Antiphase deflection exists between two horizontal LOC and ROC channels [112];
- Onsets of eye movements occur within 300 ms of one another [112];
- Movements begin and end near zero velocity [112];
- Artifacts, such as blinks or EEG/EMG artifacts, are absent.

These criteria are applied by clinicians to detect SEM visually [113] [114]. Figure 4.21 represents a sample of SEM activity and its ten level decompositions. The detail and approximation coefficients can be used for composing the signals by Inverse Discrete Wavelet Transform (IDWT), for which the evaluation is not a concern of this research. The EOG signals were sampled with the rate of 128 samples per second, which can result in the frequency range for each level of decomposition, as shown in Table 4.15.

Figure 4.21: The SEM signal and its 10-level decomposition. The signal duration is 10 seconds. The signal decomposed with Daubechies of order 4. The signal is ROC minus LOC of Stage 1, belonging to patient C1, who is not exposed to antidepressant medication. (D1 is in the range of 10^{-5} , and D2 and D3 are in the range of 10^{-3})

Table 4.15: The range of frequencies in each decomposition level

Decomposition Level	Frequency Range (Hz)
Signal	0–64
D1	32–64
D2	16–32
D3	8–16
D4	4–8
D5	2–4
D6	1–2
D7	0.5–1
D8	0.25–0.5
D9	0.125–0.25
D10	0.0625–0.0125
A10	0–0.0625

To capture the frequency between $0.2 - 0.6\text{Hz}$, details below scale seven were processed. Ten-level decomposition is enough because there would be no information at lower frequencies. Wavelet decomposition were applied to the difference of ROC and LOC since the difference signal had higher values than the original EOG signal in the presence of conjugate eye synchrony. The antiphase waveforms of ROC and LOC during the presence of conjugate eye movements are valuable criteria which are used by evaluating the difference of the left and right EOG signals. As one can see, the details of D8–D10 have the maximum value in SEM detection.

The objective was to find out whether the details of slow movements (D8–D10) are good features to classify the EOG signals belonging to those who were exposed to the medication, and those who were not. In this regard, the levels 8 – 10 Detail coefficients were extracted and used as features to classify the patients, i.e. those exposed to the medications (Effexor, E, Celexa, L, or Prozac, P), and those unexposed (Control, C). Figure 4.22 represents the regular and leave-one-out results of this classification. This figure shows that these details cannot be used directly for this classification and that they need some manipulation. Table 4.22 represents the accuracy results.

In the next step, the normalized value of the coefficients of D8–D10 and A10 (Approx-

Figure 4.22: Levels 8 – 10 details classification results derived from slow movement detection method

Table 4.16: The leave-one-out method and the regular method accuracies for D8–D10 classification of SEMs detection method.

Groups	Leave One Out Method (%)	Regular Method (%)
Effexor Control Stage 1	51.6	53.0
Effexor Control Stage W	52.2	53.3
Prozac Control Stage 1	50.4	52.4
Prozac Control Stage W	50.3	51.8
Celexa Control Stage 1	50.9	52.4
Celexa Control Stage W	50.4	51.5

imation) were derived for classification. In other words, each vector of the Approximation and Details is first derived from each 30-second fixed segment, then normalized with its maximum absolute value. After deriving Approximation and Detail matrices for all patients of all epochs (30 seconds), the LDA classifier was applied to those patients who were exposed to the antidepressant medication and to those who were not. Table 4.17 represents the results.

Table 4.17: The leave-one-out method and the regular method accuracies for A10 and D8-D10 classification of SEMs detection method

Groups	Leave One Out Method (%)	Regular Method (%)
Prozac Control Stage 1	49.8	61.4
Prozac Control Stage W	55.5	63.4
Effexor Control Stage 1	49.6	57.2
Celexa Control Stage 1	49.4	59.2
Zoloft Control Stage 1	49.3	57.7

Furthermore, the Zoloft group (which consists of five patients who were exposed to Zoloft medication) and Control group (which consists of five patients who were not exposed to any antidepressant medication) were classified with this method, but instead of A10 and D8–D10 coefficients, A10 and D3–D10 were considered for classification. This did not significantly improve the accuracy: The leave-one-out accuracy was 51.3% and the regular method accuracy was 56.5%.

Also, another set of data, called Normal and Abnormal, were checked. This group includes patients with sleep apnea, where Abnormal patients were exposed to Prozac medication and Normal ones were not. The leave-one-out accuracy was 56.8% and the regular method accuracy was 63.9%.

Discussion

The DWT coefficients in Stage 1 do not show any remarkable difference between the classified groups of patients. This could be due to the large number of noise level signals that were compared for both groups, exposed and unexposed to the medication, in the classification. On the other hand, the result may imply that the waveforms of the signal in terms of SEMs

have not been affected by this type of medication. In other words, the detail coefficients corresponding to SEM (D8–D10) does not show any difference, which means that the waveform amplitudes of classified groups with fixed segmentation are not much different.

4.5.2 SEM Analysis using EMDS

Instead of consecutive fixed segmentation, Eye Movement Detection Software (EMDS) was used to detect eye movements, and then the classification was made with the same procedure as in the previous subsection. The size of the considered segment is 501 samples, 250 samples before and 250 samples after detected movement, as well as the detected sample. Daubechies of order 4 with 10 level decomposition was used and only D8–D10 was classified to show the difference of the SEM waveforms of patients exposed and unexposed to medication. Figure 4.23 and Table 4.18 represent the accuracy results of this comparison.

Figure 4.23: Level 8 to level 10 Details classification accuracy results, derived from slow movement detection method using the EMDS.

Table 4.18: The leave-one-out method and the regular method accuracies for D8–D10 classification of SEMs detection method

Groups	Leave One Out Method (%)	Regular Method (%)
Effexor Control Stage 1	59.8	60.2
Effexor Control Stage W	56.9	57.4
Prozac Control Stage 1	61.5	61.9
Prozac Control Stage W	69.4	69.5
Celexa Control Stage 1	58.6	58.8
Celexa Control Stage W	55.0	55.4

Discussion

There was an improvement of 10% in the accuracies of the classification as a result of taking into consideration the eye movements. In the previous case, one of the possible causes of low accuracies was due to the large number of segments, with no eye movement. By using EMDS, one can see that the results improve. Especially in the case of Prozac and Control patients, the accuracy increases to approximately 70%. The accuracy resulted from the classification represents the accuracy of the correctly assigned number of the cases to their classes. In this case, this accuracy of approximately 70% indicates that the difference between the waveforms of detected movements of the sleep eye movements of the patients who were exposed to antidepressant medication and those who were not is remarkable. In other words, the details (D8-D10) derived from the correlation between the mother wavelet and the detected movements are classified. The accuracy result of this classification represents that how the classifier has been successful in differentiating the detail coefficients. If the accuracy were 100%, it would mean that the difference between the coefficients was much, so the detail coefficients of one of the classes are very similar to mother wavelet and the other are not. Therefore, the accuracy of the classification shows the difference between the waveforms of the detected movements for the patients who were exposed to medication and those who were not. This also shows that sleep eye movements, which are unconscious eye movements controlled by the brain, have been affected largely by the exposure to Prozac. Therefore, Serotonin controlled by Prozac affects eye movements by changing the waveforms of the

signals in low frequencies. Using the D8–D10 decompositions represent the low frequency sub-bands in the signals.

4.6 Feature Extraction Using DWT

As was discussed in Chapter 2, the Wavelet Transform (WT) is used for addressing the non-stationary signal problems, describing a time-domain signal in terms of coefficients. These coefficients show the correlation between the mother wavelet and the signal for each scale. In other words, a signal will be described in terms of building blocks that are the family of functions, derived from a single generating function called a mother wavelet, by translating and dilating operation. One of the advantages of WT is that it works on varying window sizes. This property may allow the analysis to have a better performance than other transforms such as Fourier Transform (FT) or Short Time Fourier Transform (STFT). WT uses varying sizes of broad windows for low and high frequencies. Thus, the signal can be described in an optimal time-frequency resolution in all frequency ranges. This means that WT can detect the transient information of the non-stationary signals. The specification of time and frequency localization is one of the main features of WT. The WT of a signal is the decomposition of the signal over the translation and dilation of mother wavelets [97].

WT can be used to extract features from EOG, by performing a spectral analysis of the signal. These features characterize the EOG waveforms. Therefore, they can be used to reflect the antidepressant medication effects that are not apparent to clinicians [97].

4.6.1 Feature Extraction Using DWT Matrix, without Grouping

As was discussed in Chapter 1, the EOG signal is considered to be useful in deriving the information of sleep eye movements. The analysis of EOG signals are important for neurologists and psychologists in order to assess the brain functionality. In this research, the attempt is to track the effects of antidepressant medication (SSRI and SNRI) on sleep eye movements. Since eye movement recording during sleep is a non-invasive method, it would be more worthy if one could find the relationship between the EOG and Serotonin re-uptaking

behaviour. Thus, EOG can extract valuable information, for neurologists and psychiatrists.

A simple way to display the differences between those who were exposed to medication and those who were not, is to conduct a spectral analysis. The WT provides general techniques for many tasks, such as biomedical signal analysis. One Technique is the ability to compute and manipulate data in features. with features being defined as manipulated data in compressed parameters. Therefore, EOG signals can be compressed into the parameters, and be characterized in order to study the behaviour of the EOG signals. These features use the less of parameters to represent EOG signals, and they are important for diagnostic and therapeutic purposes. As described in Chapter 2, the WT can be thought of as an extension of the classic Fourier transform which works on a multi-scale basis, rather than working on a single scale (time-domain or frequency-domain). This multi-scale feature of WT decomposes a signal into a number of scales; each scale represents singularities of the signal in a particular sub-band [97]. The procedure of multiresolution decomposition used in WT was described in Chapter 2.

The EOG signals can be decomposed to different structures occurring at different time-scales. In other word, it is a superposition of different time-scale structures. One of the purposes of wavelet analysis is to decompose the signals to different levels of time-scale structures for further analysis. This method is very suitable for non-stationary signals such as EOG because it is able to localize time and frequency precisely. This time frequency localization property of wavelets is known as *compact support*, which is one of the most attractive features in WT. Having such a precise time frequency localization characteristic is due to the fact that WT uses a variety of window sizes, instead of fixed windows.

The selection of an appropriate mother wavelet and the number of decomposition levels are very important in wavelet analysis. In this case, the mother wavelet of Coiflet of order 3, which was proposed by Ayush Bhandari, was used [115]. The number of decomposition levels was chosen as eight levels. Table 4.19 shows the frequency range of each level of decomposition. According to this table, the information under $0.25Hz$ exists in A8. On the other hand, due to the use of a lowpass filter with a cutoff frequency of $0.3Hz$ during

clinic recordings, further decompositions may not be useful. Also, it is not necessary to evaluate the lower frequencies because dominant frequencies do not exist below 8-level of decompositions [115].

Table 4.19: The range of frequencies in each decomposition level

Decomposition Level	Frequency Range (Hz)
Signal	0 – 64
D1	32 – 64
D2	16 – 32
D3	8 – 16
D4	4 – 8
D5	2 – 4
D6	1 – 2
D7	0.5 – 1
D8	0.25 – 0.5
A8	0 – 0.25

The DWT provides a compact representation of the energy distribution of EOG signals in time and frequency domain. Therefore, the computed wavelet coefficients, DWTs, are considered feature vectors which represent the signals.

EMDS was used to detect the movement. Among the four types of movements, leftward, rightward, upward, and downward, the leftward movement was considered in this section. Also among the three EOG signals, LOC was chosen for this analysis. In other words, LOC signal parts containing leftward movements were analyzed. There is no restriction on the selection of any types of movement or any EOG channel, and the method can be expanded to other movements and channels as well. In this section, because of the computational time and complexity of EOG signals, only a part of the data was considered.

The fixed segments of 501 samples containing eye movements were chosen (250 samples before and 250 samples after the detected movement, as well as the detected sample). These segments were decomposed to 8-levels by the Coiflet of order 3 as their mother wavelet. The detail wavelet coefficients (D1, D2, D3, D4, D5, D6, D7, and D8) at the first, second, third, fourth, fifth, sixth, seventh, and eighth levels have the 259, 138, 77, 47, 32, 24, 20,

and 18 coefficients respectively, and the approximation (A8) has 18 coefficients. By this way, 633 wavelet coefficients were derived. Although D1, D2, are not in the range of the desired frequency (0 – 15Hz), the dimensionality of the extracted feature vector is still big. In order to reduce the dimensionality of the feature vectors, statistics, over the set of wavelet coefficients, were derived. The statistical features that were used are as follows:

1. Mean of the absolute values of the coefficients in each sub-band:

$$\overline{|X|} = \frac{1}{N} \sum_{i=1}^N |X(i)| \quad (4.1)$$

where X is the vector of DWT coefficients of each sub-band, i is the index of detected movements, and N is the total number of detected movements;

2. Average power of the wavelet coefficients in each sub-band:

$$\overline{P} = (\sum |X|^2) / (\text{length}(X)) \quad (4.2)$$

where X is the vector of coefficients and $\overline{P}$ is an average power vector;

3. Standard deviation of the coefficients in each sub-band:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (X_i - \overline{X})^2} \quad (4.3)$$

where X_i is the vector of coefficients in a sub-band, $\overline{X}$ is a mean value of that sub-band, N is the total number of detected movements;

4. Ratio of absolute mean values of adjacent sub-band:

$$\text{Ratio} = \frac{(\sqrt{\sum |\overline{X}|^2})_{i+1}}{(\sqrt{\sum |\overline{X}|^2})_i} \quad (4.4)$$

where $\overline{X}$ is the mean value of the DWT matrix of a sub-band, and i is the index of the sub-band.

Furthermore, the frequency distribution are represented by features 1 and 2 in Equation 4.1 and 4.2. The amount of changes in frequency distribution are represented by feature 3 and 4 in Equation 4.3 and 4.4 .

After deriving the detected movements for each patient and conducting the segmentation, the wavelet coefficients of segmented vectors was computed. The mean absolute value, average power, standard deviation, and the ratio of absolute mean values of adjacent sub-bands were evaluated for each decomposition level of each patient. The final feature result provides a vector for each patient. Therefore, five vectors were assigned to five members of the group who were exposed to antidepressant medication and five vectors were assigned to five patients who were not exposed to antidepressant medication. These vectors were used to classify the patients by the linear discriminant analysis (LDA). Figure 4.24 displays the leave-one-out method accuracy results of LDA classification.

Figure 4.24: Details (D4 to D8) and approximation (A8) of leave-one-out method accuracies for different features in which the classification performed over Stage 1, LOC, leftward movements in patients exposed/unexposed to antidepressant medication.

Figure 4.25 displays the results of the regular method accuracies of different features. As described in Chapter 2 and by looking at this figure, it can be understood that the leave-

one-out method results are more trustworthy than the regular method. However, the low accuracies in the leave-one-out method implies that there is not much difference between the selected features for patients who were exposed to medication and those who were not. In other words, the information might be lost in deriving average values for most of the features.

Figure 4.25: Details (D4 to D8) and approximation (A8) the regular method accuracies for different features in which the classification performed over Stage 1, LOC, leftward movements in Normal/Abnormal set of data.

However, the main reason for the poor results could be due to the low number of comparisons. Only five feature vectors of the patients who were exposed to antidepressant medications were compared with five feature vectors of those who were not. Also, making the average of all values to derive the feature vector might have caused inaccurate results. This means the information about the differences between the patients who were exposed and those who were not exposed to antidepressant medication may be removed by averaging or by evaluating the standard deviation for the entire matrix.

Table 4.20: The accuracy results of the leave-one-out method for four features performed over all entire matrix of wavelet coefficients (Abs=Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)

Level of Decom.	Mean of Abs. Val. (%)	Avg. Power (%)	Std (%)	Ratio (%)	Entire Features (%)
A8	40	40	40	70	60
D8	80	30	70	70	50
D7	60	30	30	60	30
D6	30	40	50	70	40
D5	50	40	60	70	50
D4	60	40	40	60	70

Table 4.21: The accuracy results of the regular method for four features performed over all entire matrix of wavelet coefficients (Abs=Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)

Level of Decom.	Mean of Abs. Val. (%)	Avg. Power (%)	Std (%)	Ratio (%)	Entire Features (%)
A8	90	90	80	70	90
D8	100	100	100	70	100
D7	90	100	100	70	100
D6	100	100	90	70	100
D5	100	100	100	70	100
D4	100	90	90	60	100

4.6.2 Feature Extraction Using DWT Matrix, by Grouping

To solve the low accuracy problem of the feature extraction from the matrix, as a whole (see Section 4.6.1), the features are derived from the matrix by grouping the data. After deriving the matrix of coefficients, instead of evaluating the features directly from the entire matrix, they are evaluated from each group of five rows of the matrix. The five leftward eye movement wavelet coefficients that comprise a part of this matrix, derived from LOC, were used as a group in feature extraction. By this method, the mean of the absolute values, average power, standard deviation, and ratio of absolute mean values of adjacent sub-bands, are evaluated for each nonconsecutive 25 seconds (because the detected eye movements are not in consecutive time) from their corresponding wavelet coefficients. This corresponds to wavelet coefficients derived from five movements that each of which takes five seconds. Therefore, each derived feature matrix corresponds to one level of decomposition, and includes all four features. Figures 4.26, 4.27, 4.28, 4.29, 4.30, 4.31, 4.32, and 4.33 display the results.

Figure 4.26: The leave-one-out method accuracies for different features in which the classification performed for details (D4 to D8) and approximation (A8) of Stage 1, LOC, leftward movements in Normal/Abnormal dataset.

Figure 4.27: The leave-one-out method accuracies for different features in which the classification performed for details (D4 to D8) and approximation (A8) of Stage 2, LOC, leftward movements in Normal/Abnormal dataset.

Figure 4.28: The leave-one-out method accuracies for different features in which the classification performed for details (D4 to D8) and approximation (A8) of Stage 3, LOC, leftward movements in Normal/Abnormal dataset.

4.2.2 Feature Extraction Using DWT Matrix, by Grouping

Figure 4.29: The leave-one-out method accuracies for different features in which the classification performed for details (D4 to D8) and approximation (A8) Stage REM, LOC, leftward movements in Normal/Abnormal dataset.

Figure 4.30: The regular method accuracies for different features in which the classification performed for details (D4 to D8) and approximation (A8) Stage 1, LOC, leftward movements in Normal/Abnormal dataset.

Figure 4.31: The regular method accuracies for different features in which the classification performed for details (D4 to D8) and approximation (A8) Stage 2, LOC, leftward movements in Normal/Abnormal dataset of data.

Figure 4.32: The regular method accuracies for different features in which the classification performed for details (D4 to D8) and approximation (A8) Stage 3, LOC, leftward movements in Normal/Abnormal dataset.

Figure 4.33: The regular method accuracies for different features in which the classification performed for details (D4 to D8) and approximation (A8) Stage REM, LOC, leftward movements in Normal/Abnormal dataset.

Tables 4.22, 4.23, 4.24, 4.25, 4.26, 4.27, 4.28, and 4.29 contain the accuracy values of leave-one-out and the regular method classification.

4.6.3 Discussion

The leave-one-out method has fewer numeric values than the regular method, but it provides more reliable results. As this research is interested in extracting the result from the differences between the patients, the accuracy rate, which is in the range of 65% to 75%, seems to be a good indication of those differences (see Figures 4.26, 4.27, 4.28, and 4.29). By looking at the leave-one-out results, one can understand that D8, which contains the 0.25Hz to 0.5Hz frequencies, has the maximum accuracy for Stage 1 (see Figure 4.26). This means that in this frequency sub-band the differences between the patients are most clearly revealed, in regards to a mean absolute value feature. There is a maximum difference in the mean absolute value features. S1 with 71% accuracy (see Table 4.22), S3 with almost 69% (see Table 4.24), S2 with 65% (see Table 4.23) show that the waveform of the signal for

Table 4.22: The accuracy results of the leave-one-out method for four features performed over a group of five in a matrix of wavelet coefficients. The results were derived from Normal-Abnormal set, Stage 1. (Abs=Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)

Level of Decom.	Mean of Abs. Val. (%)	Avg. Power (%)	Std (%)	Ratio (%)	Entire Features (%)
A8	62	57	65	57	67
D8	71	57	68	65	75
D7	69	54	67	52	72
D6	68	51	63	60	68
D5	65	49	63	24	68

Table 4.23: The accuracy results of the leave-one-out method for four features performed over a group of five in a matrix of wavelet coefficients. The results were derived from Normal-Abnormal set, Stage 2. (Abs =Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)

Level of Decom.	Mean of Abs. Val. (%)	Avg. Power (%)	Std (%)	Ratio (%)	Entire Features (%)
A8	63	78	53	64	72
D8	65	75	57	51	74
D7	55	71	54	57	70
D6	59	73	55	52	70
D5	54	73	54	56	73

Table 4.24: The accuracy results of the leave-one-out method for four features performed over group of five in a matrix of wavelet coefficients. The results were derived from Normal-Abnormal set, Stage 3. (Abs=Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)

Level of Decom.	Mean of Abs. Val. (%)	Avg. Power (%)	Std (%)	Ratio (%)	Entire Features (%)
A8	50	67	53	10	77
D8	69	75	63	53	71
D7	55	77	51	57	67
D6	57	63	59	61	71
D5	40	67	48	57	77

Table 4.25: The accuracy results of the leave-one-out method for four features performed over a group of five in a matrix of wavelet coefficients. The results were derived from Normal-Abnormal set, Stage REM. (Abs =Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)

Level of Decom.	Mean of Abs. Val. (%)	Avg. Power (%)	Std (%)	Ratio (%)	Entire Features (%)
A8	50	67	39	51	59
D8	51	63	42	51	70
D7	51	64	54	54	71
D6	50	65	51	54	65
D5	48	62	52	60	67

Table 4.26: The accuracy results of the regular method for four features performed over a group of five in a matrix of wavelet coefficients. The results were derived from Normal-Abnormal set, Stage 1. (Abs =Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)

Level of Decom.	Mean of Abs. Val. (%)	Avg. Power (%)	Std (%)	Ratio (%)	Entire Features (%)
A8	64	58	68	57	70
D8	75	60	73	65	82
D7	75	60	73	53	83
D6	73	59	71	60	76
D5	72	62	74	50	73

Table 4.27: The accuracy results of the regular method for four features performed over a group of five in a matrix of wavelet coefficients. The results were derived from Normal-Abnormal set, Stage 2. (Abs =Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)

Level of Decom.	Mean of Abs. Val. (%)	Avg. Power (%)	Std (%)	Ratio (%)	Entire Features (%)
A8	64	79	56	64	74
D8	68	77	61	51	81
D7	59	75	59	57	77
D6	64	74	62	52	82
D5	54	73	55	56	73

Table 4.28: The accuracy results of the regular method for four features performed over a group of five in a matrix of wavelet coefficients. The results were derived from Normal-Abnormal set, Stage 3. (Abs =Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)

Level of Decom.	Mean of Abs. Val. (%)	Avg. Power (%)	Std (%)	Ratio (%)	Entire Features (%)
A8	71	71	65	48	97
D8	95	93	88	55	100
D7	87	93	85	59	100
D6	89	79	89	61	100
D5	81	89	87	57	100

Table 4.29: The accuracy results of the regular method for four features performed over a group of five in a matrix of wavelet coefficients. The results were derived from Normal-Abnormal set, Stage REM. (Abs =Absolute, Avg=Average, Decom=Decompose, Std=Standard Deviation, Val=Value)

Level of Decom.	Mean of Abs. Val. (%)	Avg. Power (%)	Std (%)	Ratio (%)	Entire Features (%)
A8	62	72	54	51	67
D8	62	71	59	51	85
D7	63	71	67	54	88
D6	68	74	67	54	86
D5	70	73	65	60	82

those who were exposed to medication is different from that of those who were not. This fact simply represents the idea of detection of the antidepressant medication effects by EOG analysis. In other words, EOG can detect SSRIs effects, and this method is able to detect the SSRI behaviour through EOG signals. However, mean absolute values in Stage REM show accuracy is too low (50%) to differentiate exposed and unexposed patients (see Figure 4.29).

In Stage 1 (see Figure 4.26), the means of absolute values show about 70% accuracy in the classification for D8, D7, and D6. This means that there are differences in the frequency range of $0.25 - 2Hz$ between the patients who were exposed and unexposed to antidepressant medication. Standard deviations, which in this case show the range of deviation in each sub-band, represent maximum differences in the range of $0.25 - 1Hz$ (D8 and D7). Although, about 60% accuracies in classification do not represent significant differences, the average 65% accuracy in this case represents a kind of existence of differences between deviations in the DWT coefficients. This accuracy rate represents the difference between the waveform of the signal affected by medication exposure and was not initially evident in the signal.

In Stage 1 (see Figure 4.26), the “all features” have 72%–75% accuracy in differentiating between exposed and unexposed patients. Furthermore, when all features, (mean of absolute, average power, standard deviation, and ratio) were used for classification, the accuracy percentage was increased as expected.

In Stage 2 (see Figure 4.27), the classification accuracies of the “average power” feature are about 74% on average. Also, the “all features” shows the accuracy of above 70% in this stage.

In Stage 3, the “average power” shows 75% accuracy with the leave-one-out method in D8 (frequency range of $0.25 - 0.5Hz$), and 77% in D7 (frequency range of $0.5 - 1Hz$). In other frequency sub-bands the derived the leave-one-out method accuracies are 67% and 63%, which show a difference in the power of the DWT coefficients (see Table 4.24).

In Stage 3, the “all features” calculation has a maximum accuracy of 77% in the leave-one-out method, both in A8 (frequency range of $0 - 0.25Hz$) and in D5 (frequency range of $2 - 4Hz$). This means that if all four features are considered, the accuracy rate improves to 70%

and above perhaps due to the fact that if the feature vector dimensionality increases, then the classifier performance improves. This means that the difference between the waveform representatives (DWT coefficients), for the patients who were exposed and unexposed the antidepressant medication are distinguished better with an accuracy rate of about 70% and above (see Figure 4.28).

In Stage REM, there are not many differences between the features. "Mean of absolute value", "standard deviation", and "ratio" for patients who were exposed and unexposed to medication show an accuracy rate of the leave-one-out method at about 50%, which indicates that the difference is minimal (see Figure 4.29). However, "average power" has an accuracy of about 64.2% on average. Also, the accuracy of "all (entire) features" is 71% for D7 (frequency range of $0.5 - 1\text{Hz}$) and 70% for D8 (frequency range of $0.25 - 0.5\text{Hz}$). D5 (frequency range of $2 - 4\text{Hz}$) has an accuracy of 67% and D6 (frequency range of $1 - 2\text{Hz}$) has the least accuracy rate of 65% (see Table 4.25).

In conclusion, the "mean of absolute value" is a good criterion only for Stage 1, and "Average Power" is good for all stages except for Stage 1. Although "Standard Deviation" and "Ratio" do not achieve positive results, both are significant in "all feature" accuracies; i.e., by removing each of these two features, the accuracies will decrease.

Further assessment of depression and related medicative states can be conducted by using neural networks, adaptive segmentation, and adaptive mother wavelet selection. Neural networks can improve the classification results by implementing linear discriminants in a space where the inputs of classification are nonlinear; i.e., the nonlinearity of the input of the classification can be resolved by using neural networks. Adaptive filters such as Kalman filter, Hysteresis filter [109], or fast least square algorithms [116] can be used for EOG segmentation. Adaptive mother wavelet selection also can be used to detect the different waveforms of sleep eye movements. By this method, one may be able to distinguish the pathological waveforms from non-pathological waveforms in the EOG signals.

Chapter 5

Conclusions

SLEEP is a necessary part of human existence. Sleep disorders have a great impact on the well-being of human life. More than 50 million Americans suffer from sleep disorders. These disorders hinder human daily function and affect health and longevity [2]. In the last three decades, academic studies on sleep have found a large number of connections between sleep and physical and mental well-being. The physiological information recorded from patients during sleep is invaluable for scientists, especially for neurologists and psychiatrists. While sleep clinicians use sleep information in clinical management, there is a wealth of information that is not apparent on casual visual inspection, even for experts [12]. However, more invaluable physiological information can be achieved through mathematical and technical methods. In other words, further processing will help clinicians to utilize invaluable information that is not readily available to them in routine sleep studies. Measures of sleep physiology not obvious to the human eye may provide important clues to disease states and its response to therapies.

Sleep disorders frequently coexist with other medical and psychiatric conditions (e.g., cardiovascular disease, depression, or diabetes), and they are affected by different medication exposure; i.e., harmful effects. Furthermore, many sleep state changes have been observed in clinical depression studies (reduced activity), with respect to medication exposure [38]. Thus, it becomes evident that medications are significant in sleep studies.

Most patients with depression experience sleep disturbance that may be accompanied

by excessive sleeplessness or insomnia. The characteristics of sleep disturbance in patients with depression can be useful for treatment. Furthermore, many methods are used to treat these patients nowadays. Exposure to antidepressant medications such as selective serotonin re-uptake inhibitors (SSRIs) and serotonin norepinephrine re-uptake inhibitors (SNRIs), has recently become a common method [37].

This thesis has described how sleep eye movements were different in depressed patients who used antidepressant medications compared to those who did not. This study attempts to present the antidepressant medications' effects on sleep eye movements. Clinically used SSRIs such as Prozac (with the generic name of Fluoxetine), Celexa (with the generic name of Citalopram), and Zoloft (with the generic name of Sertraline), and SNRIs such as Effexor (with the generic name of Venlafaxine) were considered in this study to assess the possible connections between eye movements recorded during sleep and serotonin effects. The novelty of this research is in the assessment of sleep eye movements in order to track the antidepressant medications' effects on the brain, through the EOG channels. EOG analysis is valuable because it is a noninvasive method of study, and the research at hand is looking for findings that have been invisible to the eyes of professional clinicians. This research shows that there is a remarkable difference between the waveforms of sleep eye movements for the patients who were exposed to medication and those who were not. This difference is not evident in routine sleep studies because the EOG signals are too long.

5.1 The Contributions

To quantify and assess the effects of antidepressant medication on sleep eye movements, the following contributions were made:

- The eye movement detection software (EMDS) was written for detecting sleep eye movements through EOG signals.
- Autoregressive (AR) modelling and wavelet analysis were used for classifying sleep eye movements of those who were exposed to antidepressant medication and those who

were not.

The contributions of this research can be summarized as two main categories, technical and clinical.

5.1.1 Clinical Contributions

The clinical contribution of this research was the development of Eye Movement Detection Software (EMDS). This software was used to quantify sleep eye movements, and it has the following features:

- Band Pass Filter: An FRI filter in the frequency range of $0.3 - 20Hz$ with the order of 200 was used to remove artifacts;
- Verification Plot: All detected movements are plotted in pre-process and post-process figures, including all EOG channels for verification that can be used by clinicians;
- Returning Eye Movement Time Duration: The time durations of detected eye movements are returned. These time durations, which are categorized into two main categories: slow eye movements and fast eye movements, are the clinical concerns in sleep studies;
- Returning Other Eye Movement Specifications: This software also returns the other specifications of detected eye movements such as time index and maximum amplitude of the detected movements for each EOG channel.

The EMDS and the named features of this software can be used for further analysis such as quantification and classification of detected movements in clinical studies.

5.1.2 Technical Contributions

The technical contribution of this research is summarized in two sections: AR modelling and wavelet analysis of EOG signal. In both methods, linear discriminant classifiers were used to evaluate differences between the waveforms of sleep eye movements for the patients who were exposed to antidepressant medication and of those who were not.

Autoregressive Analysis

The methodology, which was used for the classification of the sleep eye movement of the patients who were exposed to the Prozac (antidepressant) medication, and of those who were not, was to derive AR modelling. The AR coefficients of fixed segments, including the detected eye movements, were derived. The signal was segmented because EOG is a non-stationary signal, and segmentation makes each segment become stationary or quasi-stationary. The optimum model order was 27 derived by the minimum square error inspection. The results were classified with a linear discriminant function. Then, the classification overall average accuracy rate was determined to be 82.8% [12].

Thereafter, the methodology was developed by using the improved eye movement detection software and deriving the optimum AR model order by employing the Akaike Information Criterion (AIC) method [108]. Other groups of patients were tested. Groups 1 and 2, each included five patients who used Celexa and Effexor, respectively, compared to five patients not taking any antidepressants. AR coefficients of sleep eye movements belonging to each group were derived and classified using linear discriminant analysis. The model order was also 27 evaluated by AIC method. The overall averages of the regular method accuracies were 76.4% and 78.7% for groups 1 and 2, respectively. The overall averages of the leave-one-out method accuracies were 75.5% and 77.5% for the same groups [117].

Finally, the Receiver Operating Characteristic (*ROC*) curve was evaluated for the autoregressive modelling method. This method evaluates the performance of a method by measuring the area under the Receiver Operating Characteristic (*ROC*) curve. The area under the curve (AUC) is equal to the probability of discrimination; that is the ability of the test to correctly distinguish those who were exposed to antidepressant medication from those who were not. An area of 1.00 represents a perfect classification and 0.50 represents a worthless one. The AUC ranges are ranked as follows [118]:

- 0.90 – 1.00 = Excellent;
- 0.70 – 0.90 = Good;

- $0.60 - 0.70 = \text{Poor};$

- $0.50 - 0.60 = \text{Fail}.$

As was shown in Subsection 3.3.4, 82.17% derived for AUC represents a “good” evaluation for this autoregressive method.

AR coefficients represent the waveforms of sleep eye movements. Therefore, the classification of AR coefficients of patients who were exposed to medication and of those who were not represents the differences between their waveforms. In this research, the achieved accuracies, which are above 70% in average, demonstrate the differences between the waveforms of sleep eye movements of the patients who were exposed to antidepressant medication and those who were not. The results demonstrate that eye movements can be quantified and characterized with this approach. This method will allow the development of new metrics that may assist in disease classification and response to treatment, in a variety of neuropsychiatric conditions.

The following discussion points should be considered,

1. In earlier studies, the AR model order of 2 was used for EOG analysis [109]. However, the model order of 2 may not be accurate enough for this research. Therefore, the optimum model order of 27 was derived using two methods: the minimum square error inspection, and the AIC method. The complexity of the AIC method was less than the complexity of the minimum square error inspection.
2. All-pole models are the special structures that make algorithms fast and efficient for finding the all-pole parameters [99]. Among different types of algorithms for all-pole modelling, the Burg method was chosen because it always produces a stable model and it handles short data records well [119]. However, the Burg method does not perform well for noisy data [119].
3. LDA classification is recommended for most clinical applications [120]. LDA classification is simpler than other classifiers such as neural networks. Although the use of more

advanced classifiers, such as neural networks, may increase the accuracy rates, the accuracy results from LDA classifications were above 70% in most of the cases, which was satisfactory. However, simpler classifications such as k-mean clustering were tried in this research and the results were not satisfactory.

Wavelet Analysis

The scalogram analysis, Section 4.4, explained that the EOG signal was segmented to fixed 5-second epochs, with no respect to the inclusion of eye movements. Thereafter, the scalogram amplitude matrix of these fixed segments was evaluated. As was discussed theoretically in Chapter 2, the elements of the rows of the amplitude matrix represent the energy of each time-frequency box (Heisenberg rectangle) for each scale. To find out the maximum energy for each scale, the mean value of each row of the amplitude scalogram matrix, which corresponds to the energy in that scale, was derived. Consequently, the maximum energy and its corresponding scales were determined with the help of these mean values, for each scale. In the results section, we concluded that the energy difference, derived from the scalogram, has an accuracy rate of approximately 70% in comparing Stages 2 and 3. The comparison data is presented in Figures 4.17 and 4.18, and in Table 4.14. Also, Table 4.14 in Chapter 4 clearly shows the greater differences for patients who were unexposed to medication as compared to those who were exposed to any medication. In comparison between Stage 3 with Stages 1 and REM, the difference was greater for those who were not exposed to medication than for those who were. As discussed, patient C3, who is in the group of patients unexposed to antidepressant medication, is possibly the anomaly in this group, perhaps due to medical background, data recording, or other medical or technical problems. The results were reevaluated without C3 as well. The classification accuracy increased when omitting this patient's data.

In Section 4.5, two methods of using consecutive fixed segmentations and fixed segmentations that include the detected movements were described. First, consecutive fixed segmentations were used without respect to eye movements. The first algorithm was based on

wavelet multiresolution analysis. The difference between the two EOG channels $[(\text{ROC}-\text{A1}) - (\text{LOC}-\text{A2})]$ was given to the wavelet decomposer; Mother wavelet Daubechies of order 4 was used in this regard. It was derived from the literature on sleep Slow Eye Movement (SEM) detections [111]. It was concluded that the Discrete Wavelet Transform (DWT) coefficients in Stage 1 may not represent any differences between the classified groups with this method or they may have no differences at all. In other words, the Detail coefficients corresponding to SEM (D8-D10) do not show any difference, which means that the waveform amplitudes of classified groups with consecutive fixed segmentation (10-second or 30-second) do not show much difference. Further, the reason can be due to considering a large number of segments in each group that only contain noise. In fact, the accuracy decreases if a large number of noisy segments are classified.

Second, instead of fixed segmentation, EMDS was used to detect the eye movements, and then the classification was made. The size of the considered segment was about 501 samples, including 250 samples before and 250 samples after the detected movement. Mother wavelet Daubechies of order 4 with 10-level decompositions was used, and only D8–D10 was classified to show the differences between SEM waveforms of patients exposed and unexposed to medication. Figure 4.23 and Table 4.18 represent the accuracy results of this comparison. The decomposition was derived for 10 levels. The details and approximations of each level were sent to a linear discriminant classifier for classifying those who were exposed to medication and those who were not. The block diagram of the method was presented in Figure 4.19.

In this diagram, it was concluded that there is an improvement of 10% in the results if the fixed segmentations including the eye movements are used instead of consecutive fixed segmentation without respect to eye movements. In the case of consecutive fixed segmentations, one of the possible causes for the low accuracies was the consideration of a large number of segments without any eye movement, i.e., they are only noises. However, the accuracy results were improved by only considering the segments that include eye movements. The Eye Movement Detection Software (EMDS) was used in this regard. Especially, in

the case of Prozac and Control classification, the difference between the waveform of those signals has grown up to 70% in accuracy. This shows that SEM, which are unconscious eye movements controlled by the brain, were affected, probably, by the exposure to Prozac. Because of using D8–D10 decompositions in the analysis, it can be concluded that Prozac has an effect on eye movements by changing the waveform of the signal at low frequencies.

In Section 4.6, two methods using four statistical features were discussed:

1. Mean of the absolute values of the coefficients in each sub-band;
2. Average power of the wavelet coefficients in each sub-band;
3. Standard deviation of the coefficients in each sub-band;
4. Ratio of absolute mean values of adjacent sub-bands.

In the first method, the features were derived using the DWT matrices without grouping (see Section 4.6.1). The accuracies were low (see Figures 4.24) because a low number of comparisons in the classification were made; i.e., five feature vectors of the five patients who were exposed to antidepressant medication and five feature vectors of five patients who were not exposed.

In the second method, instead of evaluating the four above features from the matrix, as a whole (see Section 4.6.1), they were evaluated from the matrix by grouping the data. Features were evaluated by grouping each of the five rows of the wavelet coefficients matrix, where the coefficients in the matrix correspond to leftward eye movement wavelet coefficients. With this method, the mean of the absolute values, average power, standard deviation, and ratio of absolute mean values of adjacent sub-bands were evaluated for each inconsecutive 25-second period (i.e., detected movements have different time indices), from their corresponding wavelet coefficients. This corresponds to wavelet coefficients derived from five movements each of which takes five seconds. Thus, each matrix of features corresponds to one level of decomposition and includes all four statistical features. With this method the following conclusions were drawn:

- The leave-one-out method has fewer numeric values than the regular method, and it provides more reliable results. As this research is interested in extracting the differences that characterize the patients, the accuracy rate which has increased from 65% to 75% seems to be a good indication of those differences. By looking at the leave-one-out results (see the figures of Section 4.6.2), one can understand that D8, which contains the $0.25Hz$ to $0.5Hz$ frequencies, has the highest accuracy for Stage 1 (S1) (see Figure 4.26). This means that in this frequency sub-band the differences of the patients are most clearly revealed, with regards to a “mean absolute value” feature. The leave-one-out accuracy was maximum for D8 in the “mean of absolute value” feature of S1, S3, and S2. S1 with a 71% difference (see Table 4.22), S3 with almost 69% (see Table 4.24), and S2 with 65% (see Table 4.23) show that the waveforms of the signals for patients who were exposed to medication are different from those of patients who were not. The differentiation between the patients simply implies the idea of detection of antidepressant medications’ effects by EOG analysis. In other words, EOG can detect SSRI effects, and this method is capable of detecting SSRI behaviour through the EOG signals. However, the mean absolute values in Stage REM do not show enough accuracy (they are about 50%) to differentiate patients who were exposed to antidepressant medication from those who were not (see Figure 4.29).
- In Stage 1 (see Figure 4.26), the means of absolute values show about 70% accuracy in the classification for D8, D7, and D6. This means that there are differences in the frequency range of $0.25 - 2Hz$ between the patients who were exposed and unexposed to antidepressant medication. Standard deviations, which in this case show the range of deviation in each sub-band, represent maximum differences in the range of $0.25 - 1Hz$ (D8 and D7). Although, about 60% accuracies in classification do not represent a significant difference, the average 65% accuracy in this case represents a kind of existence of differences between deviations in the DWT coefficients. This accuracy rate represents the difference between the waveform of the signal affected by medication exposure, which was not apparent in the signal at first.

- In Stage 1, the “all features” have 72% – 75% accuracy in differentiating between the patients exposed and unexposed to medication. Furthermore, when all features (mean of absolute, average power, standard deviation, and ratio) were used for classification, the accuracy rate increased, perhaps due to the fact that if the dimension of the feature vector increases, the classifier performance improves (see Figure 4.26).
- In Stage 2, the classification accuracy of the “average power” feature was about 74% on average. Also, an “all features” calculation gave an accuracy of above 70% in this stage (see Figure 4.27).
- In Stage 3, “average power” showed 75% accuracy with the leave-one-out method in D8 (frequency range of 0.25 – 0.5Hz), and 77% in D7 (frequency range of 0.5 – 1Hz). In other frequency sub-bands the derived leave-one-out method accuracies were 67% and 63%, which show a difference in the power of the DWT coefficients (see Figure 4.28).
- In Stage 3, “all features” had a maximum accuracy of 77% in the leave-one-out method, both in A8 (frequency range of 0 – 0.25Hz) and in D5 (frequency range of 2 – 4Hz). This means that if all four features are considered, the accuracy rate will improve to 70% and above. Further, this means that the differences between the waveform representatives (DWT coefficients) for the patients exposed and unexposed to medication are distinguished much better, with an accuracy rate of about 70% and above (see Figure 4.28).
- In Stage REM, the “mean of absolute value”, “standard deviation”, and “ratio” features did not show many differences between patients exposed and unexposed to medication because the leave-one-out method showed an accuracy rate of about 50%, which indicates that the difference is minimal (see Figure 4.29). However, “average power” had an accuracy of about 64.2% on average. Also, the accuracy of (all [entire] features) was 71% for D7 (frequency range of 0.5 – 1Hz) and 70% for D8 (frequency range of 0.25 – 0.5Hz). D5 (frequency range of 2 – 4Hz) had an accuracy of 67% and D6 (frequency range of 1 – 2Hz) had the least accuracy, at 65% (see Table 4.25).

According to the above results, one can conclude that the “mean of absolute value” is a good criterion only for Stage 1, and “Average Power” is good for all stages except for Stage 1. Although “Standard Deviation” and “Ratio” do not achieve positive results, both are significant in “all feature” accuracies; i.e., by removing each of these two features, the accuracies will decrease.

Comparison between Autoregressive Analysis and Wavelet Analysis

AR analysis is an established method for modeling the physical process such as sleep eye movements. Also the Burg algorithm that was used for deriving AR coefficients is fast and efficient. The higher accuracy was derived by AR modelling analysis than wavelet analysis. However, wavelet analysis can detect the singularities of the sleep eye movements more efficient than AR modelling. Wavelet analysis can assess the transient information and extract the frequency information in particular time. Wavelet analysis map the time domain signal information recorded on polysomnograms to time-frequency domain. Therefore, it adds one dimension to the signal, so it has more ability than AR modelling to extract new information useful to clinical applications. Also, in this research wavelet analysis was faster than AR modelling.

5.2 Future Work

The future work can be summarized as follows:

EMDS, which was used to detect the eye movements, can be developed further for better eye movement detection. This software can evaluate more accurate time durations of sleep eye movements. Among the four types of movements, leftward, rightward, upward, and downward, the horizontal movements were considered in this research; that can be extended to verticals. Time durations also can be categorized into more classes than the two classes of slow and fast movements used in this study.

Further, by using autoregressive modelling (AR), instead of AR coefficients, pole analysis can be used for broader analysis.

In the feature selection part of wavelet analysis, among the three EOG signals, LOC was chosen for further analysis. In other words, LOC signal parts, containing leftward movements, were analyzed. There is no restriction in selection of any types of movements, or any EOG channel, and the method can be expanded to the others in future work. In this section, because of the computational time and complexity of EOG signals, the data were partially considered. Thus, the method (explained in Section 4.6) can be expanded to the data recording for the rightward, upward, and downward movements, and to ROC and VOC channels.

In future work, the filtering of EOG signals can be improved as well. Adaptive filters may be used in this regard. Adaptive filtering can be used for cancelling other channels such as EEG, ECG, or abdomen artifacts. Also in the wavelet analysis method, adaptive windowing can be used for different detected eye movements. Adaptive mother wavelets can also be used for different waveforms of eye movements by detecting the normal and abnormal waveforms of eye movements in categorizing the waveforms, and confirming the results with clinicians. By this method, the pathological waveforms can be distinguished from non-pathological waveforms in the EOG signals.

Further assessment of depression and related medicative states can be conducted by using neural networks, adaptive segmentation, and adaptive mother wavelet selection. Neural networks can improve the classification results by implementing linear discriminants in a space where the inputs of the classification are nonlinear; i.e., the nonlinearity of the input of the classification can be resolved by using neural networks. Adaptive filters such as the Kalman filter, Hysteresis filter [109], or fast least square algorithms [116] can be used for EOG segmentation.

Discovering an association between the two classes of patients; exposed and unexposed to medication, is an important step in biomedical signal analysis of medication. Furthermore, why and how these two classes are related to sleep eye movements should not be limited to this research, and they need independent research in the future.

Bibliography

- [1] M. S. Aldrich, *Sleep Medicine*. Oxford University press, USA, pp. 3–7, 1999.
- [2] C. Iber, S. Israel, A. L. Chesson, and S. F. Quan, *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications.*, American Academy of Sleep Medicine Westchester, IL, pp. 19–19, 2007.
- [3] T. Tohara, M. Katayama, A. Takajyo, K. Inoue, S. Shirakawa, M. Kitado, T. Takahashi, and Y. Nishimura, *Time frequency analysis of biological signals during sleep*. SICE, 2007 Annual Conference, pp. 1925–1929, 2007.
- [4] J. Merilahti, A. Saarinen, J. Parkka, K. Antila, E. Mattila, and I. Korhonen, *Long-Term Subjective and Objective Sleep Analysis of Total Sleep Time and Sleep Quality in Real Life Settings*. Engineering in Medicine and Biology Society, EMBS 2007, 29th Annual International Conference of the IEEE, pp. 5202–5205, 2007.
- [5] T. Penzel, J. W. Kantelhardt, L. Grote, J. H. Peter, and A. Bunde, *Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea*. IEEE Transactions on Biomedical Engineering, vol. 50, no. 10, pp. 1143–1151, 2003.
- [6] C. S. P. Zhi-De Deng, N. M. Arzeno, and E. S. Katz, *Heart Rate Variability in Pediatric Obstructive Sleep Apnea*. Engineering in Medicine and Biology Society, EMBS '06, 28th Annual International Conference of the IEEE, pp. 3565–3568, 2006.
- [7] H. Nazeran, R. Krishnam, S. Chatlapalli, Y. Pamula, E. Haltiwanger, and S. Cabrera,

- Nonlinear Dynamics Analysis of Heart Rate Variability Signals to Detect Sleep Disordered Breathing in Children*. Engineering in Medicine and Biology Society, EMBS'06, 28th Annual International Conference of the IEEE, pp. 3873–3878, 2006.
- [8] U. Malinowska, P. J. Durka, K. J. Blinowska, W. Szelenberger, and A. Wakarow, *Micro- and macrostructure of sleep EEG*. IEEE Engineering in Medicine and Biology Magazine, vol. 25, no. 4, pp. 26–31, 2006.
- [9] E. M. Ventouras, I. Alevizos, P. Y. Ktonas, H. Tsekou, T. Paparrigopoulos, I. Kalatzis, C. R. Soldatos, and G. Nikiforidis, *Independent Components of Sleep Spindles*. Engineering in Medicine and Biology Society, EMBS 2007, 29th Annual International Conference of the IEEE, pp. 4002–4005, 2007.
- [10] U. R. Abeyratne, V. Swarnkar, and S. I. Rathnayake, *Sleep-Stage and Event Dependency of Brain Asynchrony as manifested through surface EEG*. Engineering in Medicine and Biology Society, EMBS 2007, 29th Annual International Conference of the IEEE, pp. 709–712, 2007.
- [11] A. H. Khandoker, C. K. Karmakar, and M. Palaniswami, *Analysis of coherence between sleep EEG and ECG signals during and after obstructive sleep apnea events*. Engineering in Medicine and Biology Society, EMBS 2008, 30th Annual International Conference of the IEEE, pp. 3876–3879, 2008.
- [12] P. Shokrollahi, S. Krishnan, K. Umapathy, K. McConville, M. I. Boulos, D. Jewell, and B. J. Murray, *A method for quantifying sleep eye movements that reflects medication effects*. In Proc., World Congress in Medical Physics and Biomedical Engineering, The International Federation for Medical and Biological Engineering, Munich, Germany, September 2009.
- [13] M. Hanaoka, M. Kobayashi, and H. Yamazaki, *Automated sleep stage scoring by decision tree learning*. Engineering in Medicine and Biology Society, 2001, Proceedings of the 23rd Annual International Conference of the IEEE, vol. 2 2001.

- [14] S. J. Redmond, and C. Heneghan, *Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea*. IEEE Transactions on Biomedical Engineering, vol. 53, no. 3, pp. 485–496, 2006.
- [15] J. Virkkala, J. Hasan, R. Velin, S. L. Himanen, A. Varri, and E. J. W. Van Someren, *Automatic sleep detection using activity and facial electrodes*. Engineering in Medicine and Biology Society, 2008, EMBS 2008, 30th Annual International Conference of the IEEE, pp. 1639–1642, 2008.
- [16] S. Redmond, and C. Heneghan, *Electrocardiogram-based automatic sleep staging in sleep disordered breathing*. Computers in Cardiology, 2003, pp. 609–612, 2003.
- [17] D. Gorur, U. Halici, H. Aydin, G. Ongun, F. Ozgen, and K. Leblebicioglu, *Sleep spindles detection using short time Fourier transform and neural networks*. Proceedings of the 2002 International Joint Conference on Neural Networks, 2002, IJCNN'02, vol. 2, 2002.
- [18] T. Watanabe, and K. Watanabe, *Noncontact method for sleep stage estimation*. IEEE transactions on biomedical engineering, vol. 51, no. 10, pp. 1735–1748, 2004.
- [19] E. Komatsu, Y. Kurihara, and K. Watanabe, *Sleep Stage Estimation by Non-invasive Bio-measurement*. SICE-ICASE, 2006, International Joint Conference, pp. 1494–1499, 2006.
- [20] A. Lewicke, E. Sazonov, M. J. Corwin, M. Neuman, and S. Schuckers, *Sleep Versus Wake Classification From Heart Rate Variability Using Computational Intelligence: Consideration of Rejection in Classification Models*. IEEE Transactions on Biomedical Engineering, vol. 55, no. 1, pp. 108–118, 2008.
- [21] I. Zhovna, and I. D. Shallow, *Automatic detection and classification of sleep stages by multichannel EEG signal modeling*. Engineering in Medicine and Biology Society, EMBS 2008, 30th Annual International Conference of the IEEE, pp. 2665–2668, 2008.

- [22] E. Estrada, P. Nava, H. Nazeran, K. Behbehani, J. Burk, and E. Lucas, *Itakura Distance: A Useful Similarity Measure between EEG and EOG Signals in Computer-aided Classification of Sleep Stages*. 27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005, IEEE-EMBS 2005, pp. 1189–1192, 2005.
- [23] S. Devot, A. M. Bianchi, E. Naujokat, M. O. Mendez, A. Brauers, and S. Cerutti, *Sleep Monitoring Through a Textile Recording System*. Engineering in Medicine and Biology Society, EMBS 2007, 29th Annual International Conference of the IEEE, pp. 2560–2563, 2007.
- [24] S. Romero, M. A. Mananas, S. Clos, S. Gimenez, and M. J. Barbanoj, *Reduction of EEG artifacts by ICA in different sleep stages*. Engineering in Medicine and Biology Society, Proceedings of the 25th Annual International Conference of the IEEE, vol. 3, 2003.
- [25] D. Matsiki, and X. Deligianni, and E. Vlachogianni-Daskalopoulou, and L. J. Hadjileontiadis, *Wavelet-based Analysis of Nocturnal Snoring in Apneic Patients Undergoing Polysomnography*. Engineering in Medicine and Biology Society, EMBS 2007, 29th Annual International Conference of the IEEE, pp. 1912–1915, 2007.
- [26] I. H. Song, Y. S. Ji, B. K. Cho, J. H. Ku, Y. J. Chee, J. S. Lee, M. Lee, I. Y. Kim, and S. I. Kim, *Multifractal Analysis of Sleep EEG Dynamics in Humans*. 3rd International IEEE/EMBS Conference on Neural Engineering, 2007. CNE'07, pp. 546–549, 2007.
- [27] M. O. Mendez, O. P. Villantieri, A. M. Bianchi, and S. Cerutti, *Sleep Analysis for Wearable Devices Applying Autoregressive Parametric Models*. Engineering in Medicine and Biology Society, 2005, IEEE-EMBS 2005, 27th Annual International Conference of the IEEE, pp. 7353–7356, 2005.
- [28] Jokerwe, *Neural synapse; Acetylcholine*. [http : //www.youtube.com/watch](http://www.youtube.com/watch), October, 2007.

- [29] E. S. Paykel, and R. G. Priest, *Recognition and management of depression in general practice: consensus statement*. Br Med Assoc, British Medical Journal, vol. 305, no. 6863, pp. 1198, 1992.
- [30] D. A. Regier, R. M. Hirschfeld, F. K. Goodwin, J. D. Burke Jr, J. B. Lazar, and L. L. Judd, *The NIMH Depression Awareness, Recognition, and Treatment Program: structure, aims, and scientific basis*. Am Psychiatric Assoc, American Journal of Psychiatry, vol. 145, no. 11, pp. 1351–1357, 1988.
- [31] E. S. Paykel, P. Freeling, and J. A. Hollyman, *Are tricyclic antidepressants useful for mild depression? A placebo controlled trial*. Pharmacopsychiatry, vol. 21, no. 1, pp. 15, 1988.
- [32] M. M. Weissman, and J. K. Myers, *Affective disorders in a US urban community: the use of research diagnostic criteria in an epidemiological survey*. Archives of General Psychiatry, vol. 35, no. 11, pp. 1304–1311, 1978.
- [33] T. Helgason, *Epidemiological investigations concerning affective disorders*. Origin, Preventions, and Treatment of Affective Disorders, Schou M and Stromgren E (eds), Academic Press, London, UK, pp. 241–255 1979.
- [34] R. Lane, D. Baldwin, and S. Preskorn, *The SSRIs: advantages, disadvantages and differences*. Oxford, UK: Oxford University Press, Journal of Psychopharmacology, vol. 9, no. 2, pp. 163–178, 1995.
- [35] D. Blazer, B. Burchett, C. Service, and L. K. George, *The association of age and depression among the elderly: an epidemiologic exploration*. journal=Journal of gerontology, vol. 46, no. 6, pp. M210, 1991.
- [36] J. F. Greden *Antidepressant maintenance medications: when to discontinue and how to stop*. The Journal of clinical psychiatry, vol. 54, pp. 39, 1993.

- [37] M. J. Thorpy, and J. Yager, *The encyclopedia of sleep and sleep disorders*. Facts on File New York, 2001.
- [38] B. Saletu, R. Frey, M. Krupka, P. Anderer, J. Grünberger, and W. R. See, *Sleep laboratory studies on the single-dose effects of serotonin reuptake inhibitors paroxetine and fluoxetine on human sleep and awakening qualities*. Sleep, vol. 14, no. 5, pp. 439, 1991.
- [39] C. H. Schenck, M. W. Mahowald, S. W. KIM, K. A. O'Connor, and T. D. Hurwitz, *Prominent eye movements during NREM sleep and REM sleep behavior disorder associated with fluoxetine treatment of depression and obsessive-compulsive disorder*. American Academy of Sleep Medicine, Sleep(New York, NY), vol. 15, no. 3, pp. 226–235, 1992.
- [40] R. Armitage, M. Trivedi, and A. J. Rush, *Fluoxetine and oculomotor activity during sleep in depressed patients*. Elsevier Science, Neuropsychopharmacology(New York, NY), vol. 12, no. 2, pp. 159–165, 1995.
- [41] C. Sanchez, and E. Meier, *Behavioral profiles of SSRIs in animal models of depression, anxiety and aggression*. Springer, Psychopharmacology, vol. 129, no. 3, pp. 197–205, 1997.
- [42] D. Johnson, *Non-compliance with antidepressant therapy-an underestimated problem*. Intern Med, vol. 11, pp. 14–17, 1986.
- [43] D. T. Wong, F. P. Bymaster, L. R. Reid, and P. G. Threlkeld, *Fluoxetine and two other serotonin uptake inhibitors without affinity for neuronal receptors*. Biochemical pharmacology, vol. 32, no. 7, pp. 1287, 1983.
- [44] R. W. Fuller, and D. T. Wong, *Serotonin uptake and serotonin uptake inhibition*. Blackwell Publishing Ltd, Annals of the New York Academy of Sciences, The Neuropharmacology of Serotonin, vol. 600, no. 1, pp. 68–80, 1990.

- [45] P. Blier, C. De Montigny, and Y. Chaput, *A role for the serotonin system in the mechanism of action of antidepressant treatments: preclinical evidence*. The Journal of clinical psychiatry, vol. 51, pp. 14, 1990.
- [46] N. Rashmi and M. Dombeck, *Bipolar disorder treatment SSRI and SNRI antidepressants*. [http : //www.mentalhelp.net/poc/view_doc.php](http://www.mentalhelp.net/poc/view_doc.php), 2007.
- [47] P. Benfield, and A. Ward, *Fluvoxamine: a review of its pharmacodynamic and pharmacokinetic properties, and therapeutic efficacy in depressive illness.*, Adis International, Drugs(Basel) vol. 32, no. 4, pp. 313–334, 1986.
- [48] K. L. Dechant, and S. P. Clissold, *Paroxetine. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic potential in depressive illness*. Drugs, vol. 41, no. 2, pp. 225, 1991.
- [49] R. J. Milne, and K. L. Goa, *Citalopram. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic potential in depressive illness*. Drugs, vol. 41, no. 3, pp. 450, 1991.
- [50] D. Murdoch, and D. McTavish, *Sertraline: a review of its pharmacodynamic and pharmacokinetic properties, and therapeutic potential in depression and obsessive-compulsive disorder*. Adis International, Drugs(Basel), vol. 44, no. 4, pp. 604–624, 1992.
- [51] Y. D. Lapierre, *Controlling Acute Episodes of Depression*. International clinical psychopharmacology, vol. 6, pp. 23, 1991.
- [52] M. Peet, *Induction of mania with selective serotonin re-uptake inhibitors and tricyclic antidepressants*. RCP, The British Journal of Psychiatry, vol. 164, no. 4, pp. 549–550, 1994.
- [53] W. E. Bunney, *Psychopharmacology of the switch process in affective illness*. Raven Press, Psychopharmacology: A generation of progress, pp. 1249–1259, 1978.

- [54] R. J. Leigh, D. S. Zee, J. A. Sharpe, O. Toronto, J. D. Geyer, T. A. Payne, P. R. Carney, and M. S. Aldrich, *Neurology of Eye Movements, Contemporary Neurology Series*. Oxford Univestity Press, Fourth Edition, 2006.
- [55] J. Trammell, and P. Ktonas, *A simple nonlinear deterministic process may generate the timing of rapid eye movements during human REM sleep*. First International IEEE EMBS Conference on Neural Engineering, 2003, Conference Proceedings, pp. 324–327, 2003.
- [56] T. Yagi, K. Gouhara, and Y. Uchikawa, *An algorithm of eye movement in selective fixation*. IEEE International Conference on Neural Networks, 1993, pp. 761–765, 1993.
- [57] A. T. Bahill, and J. D. McDonald, *Frequency limitations and optimal step size for the two-point central difference derivative algorithm with applications to human eye movement data*. IEEE Transactions on Biomedical Engineering, pp. 191–194, 1983.
- [58] H. Sakaino, N. Tetsutani, F. Kishino, *A design of optimal filtering for a gaze-controller based on therelationship between image-features and eye movements*. IEEE International Conference on Systems, Man and Cybernetics, 1992, pp. 340–345, 1992.
- [59] J. J. M. Kierkels, G. J. M. van Boxtel, and L. L. M. Vogten, *A model-based objective evaluation of eye movement correction in EEG recordings*. IEEE Transactions on Biomedical Engineering, vol. 53, no. 2, pp. 246–253, 2006.
- [60] S. Ojima, and S. Yano, *Eye movement model with neural oscillators*. IEEE International Conference on Neural Networks, 1995, Proceedings, vol. 5, 1995.
- [61] L. Manfredi, E. S. Maini, P. Dario, C. Laschi, B. Girard, N. Tabareau, and A. Berthoz, *Implementation of a neurophysiological model of saccadic eye movements on an anthropomorphic robotic head*. International Conference on Humanoid Robots, 2006, 6th IEEE-RAS, pp. 438–443, 2006.

- [62] N. Ikawa, T. Kurata, and S. Kato, *A neural network model of programming for the saccade goal*. Neural Networks, 1993. IJCNN'93-Nagoya, Proceedings of 1993 International Joint Conference, vol. 2, 1993.
- [63] J. L. Horng, and J. L. Semmlow, and G. K. Hung, and K. J. Ciuffreda, *Autoregressive analysis of dynamic vergence eye movement*. Bioengineering Conference, 1993, Proceedings of the 1993 IEEE Nineteenth Annual Northeast, pp. 69–71, 1993.
- [64] W. Kienzle, F. A. Wichmann, B. Scholkopf, and M. O. Franz, *Learning an interest operator from human eye movements*. Citeseer, IEEE Conference on Computer Vision and Pattern Recognition, vol. 17, pp. 22, 2006.
- [65] Y. Liang, M. L. Reyes, and J. D. Lee, *Real-time detection of driver cognitive distraction using support vector machines*. IEEE Transactions on Intelligent Transportation Systems, vol. 8, no. 2, pp. 340–350, 2007.
- [66] A. Ueno, Y. Ota, M. Takase, and H. Minamitani, *Parametric analysis of saccadic eye movement depending on vigilance states*. Engineering in Medicine and Biology Society, 1996, Bridging Disciplines for Biomedicine, Proceedings of the 18th Annual International Conference of the IEEE, vol. 5 1996.
- [67] A. T. Bahill, *How does the saccadic eye movement controller adapt for pathological states?* 1977 IEEE Conference on Decision and Control including the 16th Symposium on Adaptive Processes and A Special Symposium on Fuzzy Set Theory and Applications, vol. 16, 1977.
- [68] J./ D./ Enderle, and J./ W./ Wolfe, *Time-optimal control of saccadic eye movements*. IEEE Transactions on Biomedical Engineering, pp. 43–55, 1987.
- [69] A. Ueno, Y. Ota, M. Takase, and H. Minamitani, *Relationship between vigilance levels and characteristics of saccadic eye movement*. IEEE 17th Annual Conference Engineering in Medicine and Biology Society, 1995, vol. 2, 1995.

- [70] A. Iijima, A. Ueno, M. Nishimura, N. Ishikawa, H. Hosaka, and H. Minamitani, *Image analysis of quick phase eye movement in nystagmus*. Engineering in Medicine and Biology, 1999. 21st Annual Conf. and the 1999 Annual Fall Meeting of the Biomedical Engineering Soc. BMES/EMBS Conference, 1999, Proceedings of the First Joint, vol. 2, 1999.
- [71] H. Takagi, *Development of an eye-movement enhanced translation support system*. Computer Human Interaction, 1998, Proceedings, 3rd Asia Pacific, pp. 114–119, 1998.
- [72] A. N. Kumar, R. J. Leigh, and S. Ramat, *The brainstem switch for gaze shifts in humans*. Engineering in Medicine and Biology Society, 2001, Proceedings of the 23rd Annual International Conference of the IEEE, vol. 1, 2001.
- [73] T. Kurita, H. Shimai, Y. Baba, T. Mishima, M. Tanaka, S. Akaho, and S. Umeyama, *Gaze control on virtual active vision system with binocularfish-eye lenses*. 2000 IEEE International Conference on Systems, Man, and Cybernetics, vol. 3, 2000.
- [74] A. Iijima, M. Haida, N. Ishikawa, H. Minamitani, and Y. Shinohara, *Development of an eye information analysis system with a small display for evaluation of eye tracking functions*. Engineering in Medicine and Biology, 2002, 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002, Proceedings of the Second Joint, vol. 3, 2002.
- [75] S. R. Muir, M. R. MacAskill, D. Herron, H. Goelz, T. J. Anderson, and R. D. Jones, *EMMA-an eye movement measurement and analysis system*. Australasian College of Physical Scientists, Australasian Physical and Engineering Sciences in Medicine, vol. 26, no. 1, pp. 18–24, 2003.
- [76] R. S. Remmel, *Use of an Electromagnetic Eye Movement Monitor for Easy Measurement of Arm Movements*. IEEE Transactions on Biomedical Engineering, vol. 53, no. 11, pp. 2356–2361, 2006.

- [77] H. Hallez, A. Vergult, R. Phlypo, P. Van Hese, W. De Clercq, Y. D'Asseler, R. Van de Walle, B. Vanrumste, W. Van Paesschen, S. Van Huffel, and I. Lemahieu, *Muscle and eye movement artifact removal prior to EEG source localization*. Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS06), pp. 1002–1005, 2006.
- [78] T. Miyoshi, and A. Murata, *Input device using eye tracker in human-computer interaction*. 10th IEEE International Workshop on Robot and Human Interactive Communication, 2001, Proceedings, pp. 580–585, 2001.
- [79] A. Funase, T. Yagi, Y. Kuno, and Y. Uchikawa, *Prediction of eye movements from EEG*. Neural Information Processing, 1999, Proceedings, ICONIP'99, 6th International Conference, vol. 3, 1999.
- [80] Y. Lin, and W. J. Zhang, *Combining eye movement and hand movement measures for evaluating human-machine interfaces*. IEEE International Conference on Systems, Man and Cybernetics, 2003, vol. 2, 2003.
- [81] A. R. Kherlopian, J. P. Gerrein, M. Yue, K. E. Kim, J. W. Kim, M. Sukumaran, and P. Sajda, *Electrooculogram based system for computer control using a multiple feature classification model*. Engineering in Medicine and Biology Society, 2006, EMBS'06, 28th Annual International Conference of the IEEE, pp. 1295–1298, 2006.
- [82] F. Mura, and I. Shimoyama, *Visual guidance of a small mobile robot using active, biologically-inspired, eye movements*. 1998 IEEE International Conference on Robotics and Automation, 1998 Proceedings, vol. 3, 1998.
- [83] S. S. Hacisalihzade, J. S. Allen, and L. W. Stark, *Automatic analysis of eye movements with a PC*. Engineering in Medicine and Biology Society, 1989, Images of the Twenty-First Century, Proceedings of the Annual International Conference of the IEEE Engineering, pp. 652–653, 1989.

- [84] A. Daftari, T. L. Alvarez, M. L. Kung, and J. L. Semmlow, *Neuro-control in divergence eye movements*. Bioengineering Conference, 2003 IEEE 29th Annual, Proceedings, pp. 11–12, 2003.
- [85] S. Merchant, and Y. Kwon, and T. Schnell, and T. Etherington, and T. Vogl, *Evaluation of synthetic vision information system (SVIS) displays based on pilot performance*. Digital Avionics Systems, 2001, DASC, The 20th Conference, vol. 1, 2001.
- [86] H. Wakamatsu, M. Kuwano, and H. Suda, *Realization of physiological eye movements by automatic selection of control laws using artificial neural network*. IEE, IEE Conference Publication, pp. 113–113, 1993.
- [87] A. Takanishi, T. Matsuno, and I. Kato, *Development of an anthropomorphic head-eye robot with twoeyes-coordinated head-eye motion and pursuing motion in the depthdirection*. Intelligent Robots and Systems, 1997, IROS'97, Proceedings of the 1997 IEEE/RSJ International Conference, vol. 3, 1997.
- [88] X. Zhang, and Y. Sato, *Cooperative movements of binocular motor system*. IEEE International Conference on Automation Science and Engineering, 2008, CASE 2008, pp. 321–327, 2008.
- [89] K. Schreiber, and T. Haslwanter, *Improving calibration of 3-D video oculography systems*. IEEE Transactions on Biomedical Engineering, vol. 51, no. 4, pp. 676–679, 2004.
- [90] J. N. Bradley, C. M. Brislawn, and T. Hopper, *FBI wavelet/scalar quantization standard for gray-scale fingerprint image compression*. Proceedings of SPIE, vol.293, no. 1993, 1961.
- [91] V. J. Barclay, R. F. Bonner, and I. P. Hamilton, *Application of wavelet transforms to experimental spectra: smoothing, denoising, and data set compression*. Anal. Chem, vol. 69, no. 1, pp. 78–90, 1997.

- [92] J. Han, and M. Kamber, *Data mining: concepts and techniques*. Morgan Kaufmann, 2006.
- [93] S. Mallat, *A wavelet tour of signal processing*. Academic press, 1999.
- [94] N. D. Kelley, R. M. Osgood, J. T. Bialasiewicz, and A. Jakubowski, *Using time-frequency and wavelet analysis to assess turbulence/rotor interactions*. Prepared for the Collection of Technical Papers, 19th ASME Wind Energy Symposium, 38th Aerospace Sciences Meeting and Exhibit, Reno, NV, Jan. 10-13, 2000, vol. 19, pp. 10–13, 2000.
- [95] C. S. Burrus, R. A. Gopinath, H. Guo, J. E. Odegard, and I. W. Selesnick, *Introduction to wavelets and wavelet transforms: a primer*. Prentice Hall Upper Saddle River, NJ, 1998.
- [96] I. Daubechies, *Ten lectures on wavelets*. Society for Industrial Mathematics, 1992.
- [97] İ. Güler, and E. D. Übeyli?, *ECG beat classifier designed by combined neural network model*. Elsevier, Pattern Recognition, vol. 38, no. 2, pp. 199–208, 2005.
- [98] X. Zhitao G. Chengming Y. Ming, and L. Qiang, *Research on log Gabor wavelet and its application in image edge detection*. 6th International Conference on Signal Processing, 2002, vol. 1, 2002.
- [99] M. H. Hayes, *Statistical digital signal processing and modeling*. John Wiley & Sons, Inc. New York, NY, USA, 1996.
- [100] R. M. Rangayyan, *Biomedical signal analysis*. Wiley-Interscience New York, NY, 2002.
- [101] R. Fisher, *The use of multiple measurements in taxonomic problems*. Ann. Eugenics, vol. 7, pp. 179–188, 1936.
- [102] E. K. Laitinen, *Classification accuracy and correlation: LDA in failure prediction*. Elsevier, European Journal of Operational Research, vol. 183, no. 1, pp. 210–225, 2007.

- [103] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. Wiley New York, 2001.
- [104] C. E. Metz, *Basic principles of ROC analysis*. Seminars in nuclear medicine, vol. 8, no. 4, pp. 283, 1978.
- [105] S. Wang, C. I. Chang, S. C. Yang, G. C. Hsu, H. H. Hsu, P. C. Chung, S. M. Guo, and S. K. Lee, *3D ROC analysis for medical imaging diagnosis*. 27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005, IEEE-EMBS 2005, pp. 7545–7548, 2005.
- [106] D. M. J. Tax, and R. P. W. Duin, *Linear model combining by optimizing the Area under the ROC curve*. 18th International Conference on Pattern Recognition, ICPR 2006, vol. 4, 2006.
- [107] J. G. Proakis, and D. G. Manolakis, *Digital signal processing: principles, algorithms, and applications*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, pp. 30–30, 1996.
- [108] H. Akaike, *A new look at the statistical model identification*. IEEE transactions on automatic control, vol. 19, no. 6, pp. 716–723, 1974.
- [109] P. Bonnet, V. Buzenac, P. Baylou, M. Najim, and J. Paty *EOG Segmentation Using Kalman And Hysteresis Filters*. Engineering in Medicine and Biology Society, 1992, Proceedings of the Annual International Conference of the IEEE, vol. 6, 1992.
- [110] L. De Gennaro, M. Ferrara, F. Ferlazzo, and M. Bertini, *Slow eye movements and EEG power spectra during wake-sleep transition*. Elsevier, Clinical Neurophysiology, vol. 111, no. 12, pp. 2107–2115, 2000.
- [111] E. Magosso, F. Provini, P. Montagna, and M. Ursino, *A wavelet based method for automatic detection of slow eye movements: A pilot study*. Elsevier, Medical Engineering and Physics, vol. 28, no. 9, pp. 860–875, 2006.

- [112] H. S. Porte, *Slow horizontal eye movement at human sleep onset*. Blackwell Science Ltd, Journal of sleep research, vol. 13, no. 3, pp. 239–249, 2004.
- [113] L. Torsvall, and T. Åkerstedt, *Extreme sleepiness: quantification of EOG and spectral EEG parameters*. Informa Healthcare, International Journal of Neuroscience, vol. 38, no. 3, pp. 435–441, 1988.
- [114] L. D. Gennaro, A. Devoto, F. Lucidi, and C. Violani, *Oculomotor changes are associated to daytime sleepiness in the multiple sleep latency test*. Blackwell Science Ltd, Journal of sleep research, vol. 14, no. 2, pp. 107–112, 2005.
- [115] A. Bhandari, V. Khare, M. Trikha, and S. Anand, *Wavelet based Novel Technique for Signal Conditioning of Electro-Oculogram Signals*. Annual India Conference, 2006, pp. 1–6, 2006.
- [116] V. Buzenac, R. Settineri, M. Najim, and J. Paty, *EOG segmentation using fast algorithms*. 1993 IEEE International Symposium on Circuits and Systems, 1993, ISCAS'93, pp. 826–829, 1993.
- [117] P. Shokrollahi, S. Krishnan, K. Umapathy, K. McConville, M. I. Boulos, D. Jewell, and B. J. Murray, *Computer-assisted method for quantifying sleep eye movements that reflects medication effects*. In Proc., The 31st International Conference of the IEEE Engineering in Medicine and Biology Society, (EMBC 2009), Minneapolis, Minnesota, USA, September 2009.
- [118] T. G. Tape *The Area Under an ROC Curve*. University of Nebraska Medical Centre, <http://gim.unmc.edu/dxtests/roc3.htm>.
- [119] J. Vanherzeele, P. Guillaume, and S. Vanlanduit, *Improved Fourier analysis using parametric frequency-domain transfer-function estimators*. Elsevier, Mechanical Systems and Signal Processing, vol. 21, no. 4, pp. 1704–1716, 2007.

- [120] A. Nijssen, K. Maquelin, L. F. Santos, P. J. Caspers, T. C. B. Schut, J. C. den Hollander, M. H. A. Neumann, and G. J. Puppels, *Discriminating basal cell carcinoma from perilesional skin using high wave-number Raman spectroscopy*. Journal of Biomedical Optics, vol. 12, pp. 034004, 2007.

Appendix A

Fourier Analysis

The Fourier Transform (FT) is applied to linear time-invariant signals. Sinusoidal waves, $e^{i\omega t}$ are eigenvectors of this linear time-invariant process. This process is specified by the eigenvalues $\hat{h}(\omega)$ in which a signal, f , can be decomposed to a sum of sinusoidal eigenvectors $\{e^{i\omega t}\}_{\omega \in \mathbb{R}}$. Equation A.1 represents the Fourier transform of f [93].

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \quad (\text{A.1})$$

If f has finite energy, the theory of Fourier integrals represents the derivation of f from $\hat{f}$ by Equation A.2, which is known as Inverse Fourier Transform (IFT).

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega)e^{i\omega t} d\omega \quad (\text{A.2})$$

Since FT is known as a linear time invariant operator, it is used in stationary signal processing. However, in the case of non-stationary signal processing, the Fourier analysis may not be a good method, since it cannot detect the transient phenomena in the signal, such as detection of a word pronounced at a particular time in speech processing [93].

Figure A.1 represents the signal which consists of two sinusoidal waveforms with frequencies of 20 Hz and 40 Hz. The second plot, in Figure A.1, represents the signal in the frequency-domain, which is called FFT.

Figure A.2 represents the signal that also consists of 20 and 40 HZ sinusoidal waveforms and its FFT plot. Figure A.2 has 20 Hz and 40 Hz waveforms in reverse, with respect to

Figure A.1: The signal compounds of two sinusoidal waveforms with frequencies of 20 Hz and 40 Hz with in Time and Frequency domain.

Figure A.1. As represented, Figure A.1 is different from Figure A.2 in the time-domain, but their FFT is the same. This example clearly shows that the Fourier transform cannot represent the time aspect of the signal. Equation A.1 derived the Fourier coefficient by correlating f with $e^{i\omega t}$, $\hat{f}(\omega)$ strongly depends on the values of $f(t)$ for all time, t , because $e^{i\omega t}$ covers the whole real values $\mathbb{R}$ [93]. In other words, the signal in the time-domain cannot be recovered only from Fourier information in the frequency-domain.

frequency-domain. The transfer function is the summation of two Gaussian functions located, symmetrically, around the centre frequency. The two Gaussian functions that make the Gabor overlap at the origin if the standard deviation of these Gaussian functions are larger than one third of centre frequency, and they will result an unwanted DC component in the filter [36].

Where the centre frequency is equal to three standard deviations, the bandwidth is about one octave. In other words, the cutoff frequencies fall approximately on 2x and 4x (approximately from 0.8 to 1.5 on frequency axis of Figure B.1), which are equal to the one

Figure A.2: The signal compounds of two sinusoidal waveforms with frequencies of 20 Hz and 40 Hz with in Time and Frequency domain.

processing. However, in the case of non-stationary signal processing, the Fourier analysis may not be a good method, since it cannot detect the transient phenomena in the signal, such as detection of a word pronounced at a particular time in speech processing [93].

Figure A.1 represents the signal which consists of two sinusoidal waveforms with frequencies of 20 Hz and 40 Hz. The second plot, in Figure A.1, represents the signal in the frequency-domain, which is called FFT.

Figure A.2 represents the signal that also consists of 20 and 40 HZ sinusoidal waveforms and its FFT plot. Figure A.2 has 20 Hz and 40 Hz waveforms in reverse, with respect to

Appendix B

Gabor Function

The Gabor function is a Gaussian function modulated from a sine and a cosine wave. The main application of Gabor wavelet is in feature detection. Gabor filters, traditionally, were used in Short Fourier Transform analysis to locate the time and frequency information of the signal. Furthermore, Gabor filters have some advantages and disadvantages. The Gabor function is the unique function for presenting the lower bound of the uncertainty principle, and it has a good directional selectivity. On the other hand, the maximum bandwidths are limited, and they are not optimal to localize the time information for signals with broad spectral information [98].

The Gabor Function design with an arbitrary wide bandwidth is impossible. The reason, in fact, is due to the small DC components that are in the even-symmetric Gabor Function. Figure B.1 represents a high bandwidth, even-symmetric Gabor filter in the frequency-domain. The transfer function is the summation of two Gaussian functions located, symmetrically, around the centre frequency. The two Gaussian functions that make the Gabor overlap at the origin if the standard deviation of these Gaussian functions are larger than one third of centre frequency, and they will result an unwanted DC component in the filter [98].

Where the centre frequency is equal to three standard deviations, the bandwidth is about one octave. In other words, the cutoff frequencies fall approximately on 2σ and 4σ (approximately from 0.5 to 1.5 on frequency axis of Figure B.1), which are equal to the one

Figure B.1: Gabor Filter Transfer function.

octave bandwidth [98].

Publications

Journals

- Peyman Shokrollahi, Sridhar Krishnan, Karthikeyan Umapathy, Kristiina McConville, Mark I. Boulos, Dana Jewell, and Brian J. Murray, Signal Analysis of Sleep Electrooculogram. *Journal of Medical Engineering and Physics*.
- Karthikeyan Umapathy, Peyman Shokrollahi, Kumaraswamy Nanthakumar, and Srihar Krishnan, (Under review). Instantaneous features in biomedical signal processing. Submitted to *Annals of Biomedical Engineering: The Journal of the Biomedical Engineering Society*.

Conferences

- Peyman Shokrollahi, Sridhar Krishnan, Karthikeyan Umapathy, Kristiina McConville, Mark I. Boulos, Dana Jewell, and Brian J. Murray (Accepted). A method for quantifying sleep eye movements that reflects medication effects. *International Federation for Medical and Biological Engineering*, Munich, Germany.
- Peyman Shokrollahi, Sridhar Krishnan, Karthikeyan Umapathy, Kristiina McConville, Mark I. Boulos, Dana Jewell, and Brian J. Murray (Accepted). Computer-assisted method for quantifying sleep eye movements that reflects medication effects. *The 31st International Conference of the IEEE Engineering in Medicine and Biology Society*, Minneapolis, Minnesota, USA.
- Peyman Shokrollahi, Sridhar Krishnan, Karthikeyan Umapathy, Kristiina McConville, Mark I. Boulos, Dana Jewell, and Brian J. Murray (April 2009, Accepted). Quantifying sleep eye movements (Poster presentation). *Innovation Symposium*, Ryerson University, Toronto.