Scientific Computing Department
Faculty of Computer and Information Sciences
Ain Shams University

Classification of Electrocardiogram (ECG) signals for Diagnosis of Heart diseases

Thesis submitted as a partial fulfillment of the requirements for the degree of Master of Science in Computer and Information Sciences

By

Hadeer Hussein Ibrahim El-Saadawy
Teaching Assistant at Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University

Under Supervision of

Prof. Dr. Mohamed Fahmy Tolba
Professor of Scientific Computing,
Faculty of Computer and Information Sciences,
Ain Shams University

Prof. Dr. Howida Abdel-Fattah Shedeed
Professor of Scientific Computing
Vice dean of graduate studies,
Faculty of Computer and Information Sciences,
Ain Shams University

Dr. Manal Mohsen Tantawi
Assistant Professor at Scientific Computing Department,
Faculty of Computer and Information Sciences,
Ain Shams University

January – 2018
Cairo

Acknowledgment

First of all I thank Allah, the most merciful and gracious, who gave me the knowledge, patience and strength to complete this thesis, and blessed me with his inspired gifts to overcome the obstacles I encountered.

I would like to express my deep gratitude to my supervisors who I'm lucky to work under their supervision; Prof. Dr. Mohamed Tolba for his usual support, patience, encouragement and guidance, Prof. Dr. Howida Shedeed for her usual support, motivation and guidance and Dr. Manal Tantawi the one who I am lucky to have by my side not only a supervisor but an elder sister too, I extend my utmost gratitude and appreciation for your technical and scientific help, continuous supportive guidance in both technical and non-technical issues and for always believing in me. I am deeply thankful.

I would like to thank the world best gift, the most supportive family. I would like to thank Mum and Dad who have devoted themselves to support me in my whole life, not just this work for their endless passionate support and encouragement and the sleepless nights they spent to make it easier for me. And my sisters Mahitab, Alaa and Esraa for always being by my side in the downs and ups. Thanks my sisters for your usual moral support and special thanks for Alaa for her technical support.

My family, thanks for being the shoulder I can always depend on and for constantly pushing me to become the person I want to become and create the life I want for myself. This thesis dedicated to you, to make you proud. Without you, everything is nothing.

I would also like to thank the world best friend Yasmin Khaled for always being by my side. Thanks for your constant encouragement in the most difficult times, for accepting me through the tough time. I am really grateful to have you in my life.

Also, I want to thank my friends Eman Hamdi, Ghada Hamed, Alaa Atef, Alaa Salah, Doaa Mohsen and Aya Zaki for their motivational encouragement and support, without you it would have been much harder.

Last but not least, I would like to thank all my professors, colleagues and students who kept on encouraging me. Thank you for being in my life.

Abstract

The Electrocardiogram (ECG) has been introduced for decades as a powerful tool for diagnosing heart diseases. Hence, the automation of analyzing a rich source of information like ECG for diagnostic purposes is very crucial, since it helps 24-hour monitoring and instant discovering of cardiac disorders which need rapid medical aid in clinical situations.

Cardiac arrhythmias mean abnormal activities in the heart upon certain conditions and mainly consist of two types. One of them is life threatening and can cause death. On the other hand, the other type is cardiac arrhythmia which is our interest in this study. It needs attention to avoid deterioration, but it is not critical as life threatening as the first one. Thus, ECG heartbeats should be continuously examined and classified.

This thesis proposes an automatic reliable two-stage hybrid hierarchical method for ECG heartbeat classification. The heartbeats are segmented dynamically to avoid the consequences of the heart rate variability. Discrete Wavelet Transform (DWT) is utilized to extract morphological features that describe the segmented heartbeat. The extracted features are then reduced by using Principal Component Analysis (PCA). Subsequently, the resulted features along with four RR features are fed into a Support Vector Machine (SVM) to classify five categories (first stage). Thereafter, the heartbeats are further classified to one of the classes belonging to the assigned category (second stage). Two different strategies for classification have been investigated: One versus All and One versus One. The proposed method has been applied on data from lead 1 and lead 2. A new fusion step is introduced, where a stacked generalization algorithm is applied and different types of classifiers have been examined. Experiments have been carried out using MIT-BIH database. The best overall and average accuracies obtained by the first stage are 98.40% and 97.50% respectively. For the second stage, 94.94% and 93.19% are the best overall and average accuracies obtained respectively. The best results are achieved using SVM with One versus One classification strategy for both stages and decision trees classifier for the fusion step.

Table of Contents

Acknowledgment II
Abstract III
Table of Contents V
List of Figures VII
List of Tables X
List of Abbreviations XIV
Chapter 1. Introduction 2
1.1 Thesis Motivations 2
1.2 Thesis Objectives 3
1.3 Thesis Achievements 4
1.4 Thesis Organization 5
Chapter 2. Medical Background 7
2.1 Heart Anatomy 7
2.2 Blood Flow 8
2.3 Heart Electrical Activity 10
2.4 Electrocardiogram (ECG) 11
2.5 Heart Diseases 13
Chapter 3. Related Work 27
3.1 General Overview 27
3.2 Related Studies 29
3.3 Comparative Analysis 34
Chapter 4. Proposed Method – A Hybrid Hierarchical Method 38
4.1 Dataset 39
4.2 Pre-processing 40
4.2.1 Filtering 40
4.2.2 QRS Detection 41
4.2.3 Beats Segmentation 42
4.3 Features Extraction 44
4.3.1 Morphological Features 44
a) Continuous Wavelet Transform (CWT) 45
b) Discrete Wavelet Transform (DWT) 47
c) Wavelet Families 49
4.3.2 Dynamic Features 50
4.4 Features Reduction 51
4.5 Classification 52

4.5.1 Probabilistic Neural Networks (PNN) 53
4.5.2 Random Forest Trees 54
4.5.3 Softmax Regression 55
4.5.4 Decision Trees 56
4.5.5 Support Vector Machine (SVM) 57
a) One versus All SVM 59
b) One versus One SVM 59
4.6 Lead1 and Lead2 Fusion 60
4.6.1 First Stage Fusion 60
a) Rejection Method 60
b) Stacked Generalization Method 61
4.6.2 Second Stage Fusion 61
Chapter 5. Experimental Results 64
5.1 Lead1 Results 65
5.1.1 SVM Classifier 65
5.1.2 PNN Classifier 67
5.1.3 Softmax Regression Classifier 70
5.1.4 Random Forest Trees Classifier 72
5.2 Lead2 Results 75
5.2.1 SVM Classifier 75
5.2.2 PNN Classifier 77
5.2.3 Softmax Regression Classifier 80
5.2.4 Random Forest Trees Classifier 82
5.3 Analysis of Lead1 and Lead2 Results 85
5.4 Different Classification Strategies With and Without RR Features Results 87
5.5 Results of Leads Fusion for First Stage 89
5.6 Results of Second Stage and Leads Fusion 91
5.7 Analysis of First and Second Stages Results 92
Chapter 6. Conclusion and Future Work 97
6.1 Conclusion 97
6.2 Future Work 98
List of Publications 99
References 101

List of Figures

Figure 2. 1 Heart Anatomy [20] 7
Figure 2. 2 The blood flow in and out of the heart [20] 8
Figure 2. 3 The Heart Electrical Activity [24]... 10
Figure 2. 4 An ECG trace and a magnified labeled heartbeat [24] 11
Figure 2. 5 The interpretation for the three complexes of the ECG heartbeat [24].. 12
Figure 2. 6 Normal Beat captured from MIT-BIH dataset for a) lead 1 and b) lead 2 13
Figure 2. 7 Left Bundle Branch Block Beat captured from MIT-BIH dataset for a) lead 1 and b) lead 2 14
Figure 2. 8 Right Bundle Branch Block Beat Captured from MIT-BIH dataset for a) lead 1 and b) lead 2 15
Figure 2. 9 Atrial Escape Beat captured from MIT-BIH dataset for a) lead 1 and b) lead 2... 16
Figure 2. 10 Nodal (Junctional) Escape Beat captured from MIT-BIH dataset for a) lead 1 and b) lead 2 16
Figure 2. 11 Atrial Premature Contraction Beat captured from MIT- BIH dataset for a) lead 1 and b) lead 2 17
Figure 2. 12 Aberrated Atrial Premature Beat captured from MIT-BIH dataset for a) lead 1 and b) lead 2 18
Figure 2. 13 Blocked Atrial Premature Beat captured from MIT-BIH dataset for a) lead 1 and b) lead 2 19
Figure 2. 14 Nodal (Junctional) Premature Beat captured from MIT-BIH dataset for a) lead 1 and b) lead 2 20

Figure 2. 15 Premature Ventricular Contraction Beat captured from MIT-BIH dataset for a) lead 1 and b) lead 2 21
Figure 2. 16 Ventricular Escape Beat captured from MIT-BIH dataset for a) lead 1 and b) lead 2 21
Figure 2. 17 Ventricular Flutter Wave Beat captured from MIT-BIH dataset for a) lead 1 and b) lead 2 22
Figure 2. 18 Fusion of Ventricular and Normal Beat captured from MIT-BIH dataset for a) lead 1 and b) lead 2 23
Figure 2. 19 Fusion of Paced and Normal Beat captured from MIT-BIH dataset for a) lead 1 and b) lead 2 24
Figure 2. 20 Unclassifiable Beat captured from MIT-BIH dataset for a) lead 1 and b) lead 2 24
Figure 2. 21 The five main categories according to the ANSI/AAMI EC57: 1998 standard. 25
Figure 3. 1 Main Steps for ECG based heartbeat classification process 27
Figure 4. 1 The main steps of the proposed hierarchical method 38
Figure 4. 2 Normal beat a) before b) after preprocessing 41
Figure 4. 3 A segmented heartbeat using Equation (4. 2) with different values for the parameters “c” and “σ” 44
Figure 4. 4 Frequency-time domain as represented by CWT 47
Figure 4. 5 Discrete Wavelet Decomposition [2] 48
Figure 4. 6 Haar Wavelet 49
Figure 4. 7 Daubechies wavelets Family 49
Figure 4. 8 Biorthogonal Wavelets Family 50
Figure 4. 9 The proposed hierarchical classification stages 53
Figure 4. 10 PNN Architecture [69] 54
Figure 4. 11 Softmax Regression Achitecture [72] 56

Figure 4. 12 Mapping the input data points into a high dimension features space [75] 58
Figure 4. 13 Building hyperplanes to separate between the data points and constructing the hyperplane that maximizes the margin [75] 58
Figure 4. 14 One versus One SVM Classification strategy for first stage ... 59
Figure 4. 15 The Rejection method 60
Figure 5. 1 The discrete wavelet decomposition levels for both a) Lead1 and b) Lead2 86
Figure 5. 2 The final two-stage hierarchical method.................................... 93

List of Tables

Table 3. 1 A brief survey on the key studies for Heart beat Classification.. 34
Table 4. 1 The training and testing percentages used in the experiments [1] ...39
Table 4. 2 The mapping to the five main categories according to the ANSI/AAMI EC57: 1998 standard [1] ... 40
Table 4. 3 The number of beats used in testing the second level classification experiments ... 62
Table 5. 1 Lead1 using Fixed Segmentation Strategy and SVM Classifier . 65
Table 5. 2 Lead1 using 1st Dynamic Segmentation Strategy and SVM Classifier ... 66
Table 5. 3 Lead1 using 2nd Dynamic Segmentation Strategy and SVM Classifier ... 66
Table 5. 4 Lead1 using 3rd Dynamic Segmentation Strategy and SVM Classifier ... 67
Table 5. 5 Lead1 using Fixed Segmentation Strategy and PNN Classifier.. 68
Table 5. 6 Lead1 using 1st Dynamic Segmentation Strategy and PNN Classifier ... 68
Table 5. 7 Lead1 using 2nd Dynamic Segmentation Strategy and PNN Classifier ... 69
Table 5. 8 Lead1 using 3rd Dynamic Segmentation Strategy and PNN Classifier ... 69
Table 5. 9 Lead1 using Fixed Segmentation Strategy and Softmax Regression Classifier ... 70
Table 5. 10 Lead1 using 1st Dynamic Segmentation Strategy and Softmax Regression Classifier ... 71

Table 5. 11 Lead1 using 2nd Dynamic Segmentation Strategy and Softmax Regression Classifier 71
Table 5. 12 Lead1 using 3rd Dynamic Segmentation Strategy and Softmax Regression Classifier 72
Table 5. 13 Lead1 using Fixed Segmentation Strategy and Random Forest Trees Classifier 73
Table 5. 14 Lead1 using 1st Dynamic Segmentation Strategy and Random Forest Trees Classifier 73
Table 5. 15 Lead1 using 2nd Dynamic Segmentation Strategy and Random Forest Trees Classifier 74
Table 5. 16 Lead1 using 3rd Dynamic Segmentation Strategy and Random Forest Trees Classifier 74
Table 5. 17 Lead2 using Fixed Segmentation Strategy and SVM Classifier 75
Table 5. 18 Lead2 using 1st Dynamic Segmentation Strategy and SVM Classifier 76
Table 5. 19 Lead2 using 2nd Dynamic Segmentation Strategy and SVM Classifier 76
Table 5. 20 Lead2 using 3rd Dynamic Segmentation Strategy and SVM Classifier 77
Table 5. 21 Lead2 using Fixed Segmentation Strategy and PNN Classifier 78
Table 5. 22 Lead2 using 1st Dynamic Segmentation Strategy and PNN Classifier 78
Table 5. 23 Lead2 using 2nd Dynamic Segmentation Strategy and PNN Classifier 79
Table 5. 24 Lead2 using 3rd Dynamic Segmentation Strategy and PNN Classifier 79

Table 5. 25 Lead2 using Fixed Segmentation Strategy and Softmax Regression Classifier 80
Table 5. 26 Lead2 using 1st Dynamic Segmentation Strategy and Softmax Regression Classifier 81
Table 5. 27 Lead2 using 2nd Dynamic Segmentation Strategy and Softmax Regression Classifier 81
Table 5. 28 Lead2 using 3rd Dynamic Segmentation Strategy and Softmax Regression Classifier 82
Table 5. 29 Lead2 using Fixed Segmentation Strategy and Random Forest Trees Classifier 83
Table 5. 30 Lead2 using 1st Dynamic Segmentation Strategy and Random Forest Trees Classifier 83
Table 5. 31 Lead2 using 2nd Dynamic Segmentation Strategy and Random Forest Trees Classifier 84
Table 5. 32 Lead2 using 3rd Dynamic Segmentation Strategy and Random Forest Trees Classifier 84
Table 5. 33 Results using SVM classifier for both lead1 and lead2, db8 mother wavelet for lead1 and bior2.4 mother wavelet for lead2................. 87
Table 5. 34 Lead 1 and Lead 2 classification ten trials average results using One versus All SVM classifier 88
Table 5. 35 Lead 1 and Lead 2 classification ten trials average results using One versus One SVM classifier without and with RR features 89
Table 5. 36 Fusion classification ten trials average results using rejection method and stacked generalization with different classification methodologies in the second layer... 90
Table 5. 37 Average accuracy for classifying the classes of each category separately 91

Table 5. 38 Classification ten trials average results for category N, S and Q for Lead2 and category V for Lead1 using One Versus One SVM methodology 92
Table 5. 39 Comparison with the existing studies....................................... 94

List of Abbreviations

Abbreviation	**Stands for**
AE	**A**trial **E**scape
ANN	**A**rtificial **N**eural **N**etwork
AP	**A**berrated Atrial **P**remature
APC	**A**trial **P**remature **C**ontraction
AV	**A**trio**V**entricular
BAP	**B**locked **A**trial **P**remature
DWT	**D**iscrete **W**avelet **T**ransform
ECG	**E**lectro**C**ardio**G**ram
FPN	**F**usion of **P**aced and **N**ormal
ICA	**I**ndependent **C**omponent **A**nalysis
KPCA	**K**ernel **P**rincipal **C**omponent **A**nalysis
LBBB	**L**eft **B**undle **B**ranch **B**lock
LDA	**L**inear **D**iscriminant **A**nalysis
NE	**N**odal (Junctional) **E**scape
NN	**N**eural **N**etwork
NOR	**NOR**mal
NP	**N**odal (Junctional) **P**remature
PCA	**P**rincipal **C**omponent **A**nalysis
PNN	**P**robabilistic **N**eural **N**etwork
PVC	**P**remature **V**entricular **C**ontraction
RBBB	**R**ight **B**undle **B**ranch **B**lock
SA	**S**ino**A**trial
SVM	**S**upport **V**ector **M**achine

UN	**UN**classifiable
VE	**V**entricular **E**scape
VF	**V**entricular **F**lutter Wave
VFN	**F**usion of **V**entricular and **N**ormal
WT	**W**avelet **T**ransform