

AN OPTIMIZATION METHOD BASED ON GENETIC ALGORITHM FOR  
HEART RATE VARIABILITY ANALYSIS IN THE PREDICTION OF THE  
ONSET OF CARDIAC ARRHYTHMIA

✓

BOON KHANG HUA

UNIVERSITI TEKNOLOGI MALAYSIA

AN OPTIMIZATION METHOD BASED ON GENETIC ALGORITHM FOR  
HEART RATE VARIABILITY ANALYSIS IN THE PREDICTION OF THE  
ONSET OF CARDIAC ARRHYTHMIA

BOON KHANG HUA

A thesis submitted in fulfilment of the  
requirements for the award of the degree of  
Doctor of Philosophy (Electrical Engineering)

Faculty of Electrical Engineering  
Universiti Teknologi Malaysia

JUNE 2017

Dedicated to my beloved family.

## ACKNOWLEDGEMENT

The past four years has been one of the most challenging yet exciting journey for me, and I feel very grateful to be able to experience it first hand, while shaping myself to become a better person.

First and foremost, I would like to express my deepest gratitude to my dear supervisor, Prof. Dr. Mohamed Khalil Mohd. Hani for his encouragement, criticism, and love during my Ph.D. journey. His great enthusiasm for the research and philosophy of life have enlightened me in a significant way. I always look up to him not just being a mentor, but a role model, and a father.

My sincerest appreciation also goes to my co-supervisor DR. Malarvili A/P Bala Krishnan for her guidance towards my research work. I would also like to convey my gratitude to Assoc. Prof. Dr. Muhammad Nadzir Marsono and Dr. Usman Ullah Sheikh for their comments and criticism during my Ph.D. proposal defense.

It was also a great privilege to work closely with my fellow seniors Sia Chen Wei and Lee Yee Hui for their technical supports and advices. Not to forget the members of the VeCAD Research Laboratory: Han Chien, Hui Ru, Ikmal, Jeevan, Jia Wei, Omid, Vidya, and Yin Zhen. It would be a dull and lonely journey without your company.

Most importantly, I would like to thank my family, especially my dear parents for always being there for me, and for their love and patience. Thank you for the boundless supports to pursue my dream. Without them, this Ph.D. journey would have been impossible. Thank you.

## ABSTRACT

Heart rate variability (HRV) is one of the common biological markers for developing a diagnostic system of cardiovascular disease. HRV analysis is used to extract statistical, geometrical, spectral and non-linear features in such diagnostic system. The diagnostic accuracy can be maximized by applying a feature selection step that selects an optimal feature subset from the extracted features. However, there are shortcomings in using only the feature selection for optimizing a diagnostic system that is based on HRV analysis. One of the main limitations is that the parameters of HRV feature extraction algorithms are not optimized for maximal performance. In addition, the feature selection process does not consider the feature cost and misclassification error of the selected optimal feature subset. Therefore, this thesis proposes a multi-objective optimization method that is based on the non-dominated sorting genetic algorithm to overcome these shortcomings in a cardiac arrhythmia prediction system. It optimizes the HRV feature extraction parameters, selects the best feature subset, and tunes the classifier parameters simultaneously for maximum prediction performance. The proposed optimization algorithm is applied in two cardiac arrhythmia cases, namely the prediction of the onsets of paroxysmal atrial fibrillation (PAF) and ventricular tachyarrhythmia (VTA). In the proposed approach, trade-off between multiple optimization objectives that contradict to each other are also analyzed. The optimization objectives include the feature count, measurement cost, prediction sensitivity, specificity and accuracy rate. The following results prove the effectiveness of the proposed optimization algorithm in the two arrhythmia cases. Firstly, the PAF onset prediction achieves an accuracy rate of 89.6%, which significantly outperforms most of the previous works. This accuracy rate is achieved even with the HRV signal length being reduced from the typical 30 minutes to just 5 minutes (a reduction of 83%). In the case of VTA onset prediction, the accuracy rate of 78.15% is achieved with 5-minute signal length. This result outperforms previous works. Another significant result is the sensitivity rate improvement with the trade-off of lower specificity and accuracy rate for both PAF and VTA onset predictions. For instance, the sensitivity rate of the VTA onset prediction system improved from 81.48% to 92.59% while the accuracy rate reduced from 78.15% to 72.59%.

## ABSTRAK

Kebolehubahan kadar jantung (HRV) adalah salah satu penanda biologi yang popular untuk membangunkan sistem diagnostik penyakit kardiovaskular. Analisis HRV digunakan untuk mengekstrakkan ciri-ciri statistik, geometri, spektrum dan tidak-linear dalam sistem diagnostik. Ketepatan diagnostik boleh dimaksimumkan dengan menggunakan kaedah pemilihan ciri yang menentukan subset ciri-ciri optimum daripada ciri-ciri yang telah diekstrak. Walau bagaimanapun, terdapat beberapa kelemahan jika hanya menggunakan pemilihan ciri untuk mengoptimumkan sistem diagnostik yang berasaskan analisis HRV. Salah satu kelemahan utama ialah parameter-parameter dalam algoritma pengekstrakan ciri-ciri HRV tidak dioptimumkan untuk mencapai prestasi yang maksimum. Selain itu, proses pemilihan ciri tidak mengambil kira kos ciri dan ralat salah pengelasan dalam subset ciri optimum yang dipilih. Oleh yang demikian, tesis ini mencadangkan kaedah pengoptimuman pelbagai objektif berasaskan algoritma genetik isihan bukan dominan untuk mengatasi kekurangan tersebut dalam sistem ramalan aritmia jantung. Ia mengoptimumkan parameter pengekstrakan ciri HRV, memilih subset ciri yang terbaik dan menala parameter pengelas dengan serentak untuk prestasi ramalan yang maksima. Algoritma pengoptimuman yang dicadangkan digunakan dalam dua kes jantung aritmia, iaitu ramalan permulaan fibrilasi atrial paroksismal (PAF) dan takiaritmia ventrikel (VTA). Dalam pendekatan yang dicadangkan, tukar-ganti antara pelbagai objektif pengoptimuman yang bercanggah antara satu sama lain juga dianalisis. Objektif pengoptimuman merangkumi bilangan ciri, kos pengukuran, kadar ramalan sensitiviti, spesifisiti, dan ketepatan. Hasil berikut membuktikan keberkesanan algoritma pengoptimuman yang dicadangkan dalam dua kes aritmia. Pertama, ramalan permulaan PAF mencapai kadar ketepatan 89.6%, dan kadar ini lebih tinggi daripada kebanyakan kerja sedia ada. Kadar ketepatan ini dicapai walaupun panjang isyarat HRV dikurangkan dari tempoh tipikal 30 minit kepada hanya 5 minit (pengurangan sebanyak 83%). Dalam kes ramalan permulaan VTA, kadar ketepatan sebanyak 78.15% telah dicapai dengan isyarat HRV sepanjang 5 minit. Keputusan ini adalah lebih baik daripada kerja sedia ada. Satu lagi hasil yang penting ialah peningkatan kadar sensitiviti dengan pengurangan kadar spesifisiti dan ketepatan untuk kedua-dua ramalan permulaan PAF dan VTA. Contohnya, kadar sensitiviti dalam ramalan VTA dipertingkatkan dari 81.48% kepada 92.59% sementara kadar ketepatan berkurang dari 81.48% kepada 72.59%.

## TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	<b>DECLARATION</b>	ii
	<b>DEDICATION</b>	iii
	<b>ACKNOWLEDGEMENT</b>	iv
	<b>ABSTRACT</b>	v
	<b>ABSTRAK</b>	vi
	<b>TABLE OF CONTENTS</b>	vii
	<b>LIST OF TABLES</b>	xiii
	<b>LIST OF FIGURES</b>	xvi
	<b>LIST OF ABBREVIATIONS</b>	xx
	<b>LIST OF APPENDICES</b>	xxii
<b>1</b>	<b>INTRODUCTION</b>	1
	1.1 Heart Rate Variability Analysis - Overview	1
	1.2 Cardiac Arrhythmia	2
	1.3 Problem Statement	3
	1.4 Objective	6
	1.5 Scope of Work	6
	1.6 Contributions	7
	1.7 Thesis Organization	9
<b>2</b>	<b>LITERATURE REVIEW</b>	10
	2.1 Introduction	10
	2.2 Previous Works for Arrhythmia	10
	2.3 Onset Prediction of Paroxysmal Atrial Fibrillation (PAF)	12
	2.3.1 Research Issue - The requirement of Long Duration in HRV Signal	15
	2.4 Onset Prediction of Ventricular Tachyarrhythmia (VTA)	16

2.5	Feature Selection Algorithms	19
2.5.1	Overview	19
2.5.2	Feature Selection based on Genetic Algorithm (GA)	21
2.6	Optimization Issue in Parameter Values and Settings of HRV Analysis	22
2.7	Feature Cost	25
2.8	Misclassification Error and Cost in Biomedical Application	27
2.9	Hybrid Feature Selection Model based on GA	30
2.10	Duplicate Chromosomes Issue in GA	32
2.11	Multi-objective Optimization Issue	34
<b>3</b>	<b>BACKGROUND AND THEORY</b>	<b>36</b>
3.1	Heart Rate Variability (HRV)	36
3.2	HRV Pre-processing	38
3.2.1	Heart Rate Correction	38
3.2.2	Interpolation	38
3.2.3	Detrending of the HRV	39
3.3	HRV Analysis	41
3.3.1	Time Domain Analysis	41
3.3.2	Frequency Domain Analysis	42
3.3.2.1	Fast Fourier Transform	43
3.3.2.2	AR Model	45
3.3.2.3	Higher Order Spectral (HOS) Analysis	48
3.3.2.4	Smoothing Window Function	51
3.3.3	Non-linear Analysis	54
3.3.3.1	Sample Entropy	54
3.3.3.2	Poincare Plot	55
3.4	Support Vector Machine (SVM)	56
3.5	Genetic Algorithm	57
3.5.1	Theory	57
3.5.2	GA based Feature Selection Algorithm	59
3.6	Non-dominated Sorting Genetic Algorithm III	62
3.6.1	Basic Definition of Pareto Dominance Concept	62
3.6.2	Non-dominated Sorting Genetic Algorithm III	63



	3.6.2.1	Non-dominated Sorting	64
	3.6.2.2	Generate Reference Point	64
	3.6.2.3	Adaptive Normalization	66
	3.6.2.4	Association Operation	67
	3.6.2.5	Niche Preservation Operation	68
<b>4</b>		<b>RESEARCH METHODOLOGY</b>	<b>70</b>
	4.1	Research Approach	70
	4.2	Software Libraries and Tools	71
	4.2.1	LibSVM Library	71
	4.2.2	POSIX Threads (Pthreads) Library	72
	4.2.3	Normal Random Number Generators (NORMAL) Library	73
	4.2.4	Matrix Laboratory (MATLAB)	74
	4.3	Test Dataset	75
	4.3.1	Atrial Fibrillation Prediction Database (AFPDB)	75
	4.3.2	Spontaneous Ventricular Tachyarrhythmia Database (SVTAB)	76
	4.4	Performance Metrics	77
<b>5</b>		<b>MODELLING AND ALGORITHM DEVELOPMENT IN PROPOSED OPTIMIZATION METHOD FOR AR- RHYTHMIA PREDICTION</b>	<b>79</b>
	5.1	Baseline Model	79
	5.2	Proposed Optimization Algorithm	81
	5.3	Chromosome Design in Proposed Optimization Algorithm	84
	5.4	Fitness Evaluation Stage in Proposed Optimization Algorithm	87
	5.4.1	HRV Feature Extraction	87
	5.4.2	Statistical Performance Evaluation	87
	5.4.2.1	uFilter method	88
	5.4.2.2	Min-Redundancy and Max- Relevance Criterion based Mutual Information (mRMR- MI) Method	91

	5.4.2.3	Discretization Process for MI Computation	92
	5.4.3	Hybrid Feature Selection Model	95
	5.4.3.1	Local Search Operation (LSO)	95
	5.4.3.2	Duplication Handling Algorithm	97
	5.4.4	Prediction Performance Evaluation	101
	5.4.5	Feature Cost Evaluation	103
	5.4.5.1	Feature Count	103
	5.4.5.2	Measurement Cost	104
	5.4.6	Parallel Implementation using POSIX Thread	106
5.5		Non-dominated Sorting	107
	5.5.1	Fast-non-dominated Sorting (FNDS)	107
	5.5.2	Fast-non-dominated Sorting (FNDS) without Statistical Fitness Value	109
5.6		Adaptive Normalization Operation	109
5.7		Genetic Operator	110
5.8		Random Number Generator and Seed Number	111
<b>6</b>		<b>EXPERIMENTAL WORKS AND RESULTS</b>	<b>112</b>
	6.1	Experimental Work: Tuning the Parameters and Settings in HRV Feature Extraction (FE)	112
	6.1.1	Experimental Setup	113
	6.1.2	Analysis for uFilter Method	114
	6.1.3	Analysis for mRMR-MI method	118
	6.1.4	Summary	120
	6.2	Experimental Work: Duplicate Chromosome Issue in Proposed Optimization Algorithm Method	121
	6.2.1	Experimental Setup	121
	6.2.2	Severity of the Duplicate Chromosomes	122
	6.2.3	Quality of the Solution	124
	6.2.4	Summary	126
	6.3	Experimental Work: Optimization Without the Measurement Cost Minimization	126
	6.3.1	Experimental Setup	127
	6.3.2	Trade-off Analysis: Accuracy, Sensitivity, Specificity and Feature Count	128
	6.3.3	Discussion for the Relationship	132

6.4	Experimental Work: Optimization With the Measurement Cost Minimization	134
6.4.1	Experimental Setup	135
6.4.2	Capability in Searching the Solutions With Low Measurement Cost	135
6.4.3	Performance Comparison between the Solutions With and Without the Consideration of the Measurement Cost Minimization	137
6.4.3.1	Trade-off between Sensitivity and Specificity Rate	139
6.4.3.2	Comparison between Feature Ranking Method and Feature Extraction Time Method.	139
6.4.3.3	Trade-off between Prediction Performance and Measurement Cost.	140
6.4.3.4	Performance Comparison for the Solutions With Highest Accuracy.	141
6.4.3.5	Comparison between uFilter and mRMR-MI Method.	142
6.4.4	Summary	142
6.5	Experimental Work: Proposed Optimization Algorithm and Typical GA based Feature Selection Process	143
6.5.1	Experimental Setup	143
6.5.2	Performance Comparison	145
6.5.3	Summary	146
6.6	Benchmarking Results	146
6.6.1	PAF Onset Prediction	146
6.6.2	VTA Onset Prediction	149
<b>7</b>	<b>CONCLUSION</b>	<b>152</b>
7.1	Concluding Remarks	152
7.2	Contributions	153
7.3	Future Work	155

**REFERENCES**

157

Appendices A – D

170 – 182

## LIST OF TABLES

TABLE NO.	TITLE	PAGE
2.1	Overview of clinical research problems applying HRV analysis.	10
2.2	Arrhythmia related research using ECG signal.	11
2.3	Previous works for PAF onset prediction based on PAC method that use 30 minute ECG signal duration.	13
2.4	Previous works for PAF onset prediction based on HRV analysis that use 30 minute signal duration.	14
2.5	Previous works for PAF onset prediction based on HRV analysis that use less than 10 minute signal duration.	14
2.6	Previous works for VTA onset prediction based on the HRV analysis.	18
2.7	Summary of feature selection techniques in the previous works that use HRV analysis.	20
2.8	Strengths and weaknesses when GA is applied in problems based on HRV analysis.	21
2.9	Example of previous works using HRV analysis that use recommended parameter values from other research.	23
2.10	Misclassification error of the classification algorithm in medical application.	28
2.11	Fitness function of GA based feature selector that takes into account misclassification error penalization for non-HRV previous works.	29
2.12	Classification performance of optimized algorithm in the work of Yu and Lee.	30
5.1	List of the HRV features.	81
5.2	Definition of the chromosome representation.	85
5.3	Joint probability table that is formed by using feature $f_i$ and $f_j$ .	93
5.4	Ranking of the feature extraction algorithm.	105

6.1	Four experiments according to different statistical fitness functions and predicted arrhythmia type.	127
6.2	Pareto optimal solutions that belong to “Group1” in Figure 6.12.	130
6.3	Pareto optimal solutions that belong to “Group2” in Figure 6.12.	130
6.4	Comparison of the average measurement cost and average accuracy of Pareto optimal solutions given by the proposed optimization algorithm that use different measurement cost minimization methods, statistical fitness functions, and arrhythmia prediction.	136
6.5	Performance comparison of the Pareto optimal solutions before and after measurement cost minimization.	138
6.6	Comparison of the solutions with highest accuracy given by the proposed optimization algorithm with and without measurement cost minimization respectively for different arrhythmia prediction (PAF or VTA) and statistical fitness function (uFilter or mRMR-MI).	141
6.7	Performance comparison between typical GA based feature selection algorithm and proposed optimization algorithm.	145
6.8	Benchmarking results of PAF onset prediction using less than 10 minute signal duration.	146
6.9	Benchmarking results of PAF onset prediction using 30 minute signal duration.	147
6.10	Solutions that have accuracy above 80% with incremental sensitivity rate (with the exception of solution 4).	148
6.11	Benchmarking results of VTA onset prediction.	149
6.12	Solutions that have accuracy above 70% with incremental sensitivity rate (with the exception of solution 10).	151
B.1	Pareto optimal solutions that belong to “Group1” in Figure B.1.	173
B.2	Pareto optimal solutions that belong to “Group2” in Figure B.1.	173
B.3	Pareto optimal solutions that belong to “Group1” in Figure B.2.	175
B.4	Pareto optimal solutions that belong to “Group2” in Figure B.2.	175
B.5	Pareto optimal solutions that belong to “Group1” in Figure B.3.	177

B.6	Pareto optimal solutions that belong to “Group2” in Figure B.3.	178
C.1	Pre-defined parameter values and settings of HRV feature extraction algorithms.	181
D.1	Statistical Features For PAF and VTA Prediction Cases.	182

## LIST OF FIGURES

<b>FIGURE NO.</b>	<b>TITLE</b>	<b>PAGE</b>
1.1	Simplified block diagram of HRV based classification system.	4
1.2	Diagnostic system model in the HRV based research.	7
2.1	Feature selection model with GA.	22
2.2	Optimization model for the CBIR system.	24
2.3	Feature extraction parameters that can be optimized.	25
2.4	Flow chart for the hybrid feature selection model in the simple GA.	31
2.5	General framework of GA that handles duplicate chromosome [88].	33
3.1	Schematic of Electrocardiogram	37
3.2	RR tachogram of a human.	37
3.3	Power spectrum density (PSD) of the HRV data	43
3.4	The regions in the estimated bispectrum	50
3.5	Simple GA.	58
3.6	Difference between the common GA based feature selection algorithm and the algorithm proposed by Huang and Wang.	59
3.7	The flow chart of the GA based feature selection by Huang and Wang.	60
3.8	The chromosome design proposed by Huang and Wang.	61
4.1	35 minutes of ECG signal and the corresponding HRV data for PAF onset prediction.	76
4.2	15 minutes of HRV data for VTA onset prediction.	77
5.1	Interaction between HRV based arrhythmia prediction system, and the proposed optimization algorithm.	80
5.2	Proposed Chromosome Representation	84
5.3	Flow chart for local search operation (LSO)	97
5.4	Flow chart of the duplication handling algorithm	98
5.5	Prediction performance evaluation process of a chromosome	102
5.6	Simple block diagram for parallel implementation on a single CPU with 4 physical cores and 8 logical cores.	106



6.1	Flow chart of the simple GA in this experiment.	113
6.2	Chromosome representation for the simple GA.	114
6.3	Number of features that pass the uFilter test at various significance levels before and after the HRV FE parameters optimization: a) VTA onset prediction, and b) PAF onset prediction.	116
6.4	Sum of uFilter statistic values of the feature sets before and after the HRV FE parameters optimization for: (a) VTA onset prediction, and (b) PAF onset prediction.	116
6.5	Prediction accuracy improvement of HRV feature sets before and after the HRV FE parameters and settings optimization for: (a) VTA onset prediction, (b) PAF onset prediction.	117
6.6	Sum of the mRMR-MI value for top ranked 5, 10, 15 and 20 features before and after the optimization: (a) VTA onset prediction, and (b) PAF onset prediction.	119
6.7	Number of features that can pass the uFilter test at various statistical significance levels before and after the optimization for: (a) VTA onset prediction, and (b) PAF onset prediction.	120
6.8	Performance accuracy improvement of HRV feature sets before and after the optimization for: (a) VTA onset prediction, (b) PAF onset prediction.	120
6.9	Number of duplicate chromosomes in the population of the proposed optimization algorithm without enabling the proposed duplication handling process for VTA onset prediction: a) population P, and b) population Q.	123
6.10	Comparison of the best solution obtained from proposed optimization algorithm with and without the duplication handling process for VTA onset prediction: a) Accuracy rate, and b) Feature count.	124
6.11	Comparison of the best solution obtained from proposed optimization algorithm with and without the duplication handling process for PAF onset prediction: a) Accuracy rate, and b) Feature count.	125
6.12	3D scatter plot for the Pareto optimal solutions in experiment 1: accuracy versus sensitivity versus specificity ratio, and feature count (color bar).	129
6.13	2D view of the 3D scatter plot in Figure 6.12 for experiment 1: prediction accuracy vs sensitivity.	131

6.14	Comparison of the average feature count between “Group1” and “Group2” in 4 experiments.	133
6.15	Distribution of the Pareto optimal solutions before and after considering the measurement cost minimization: a) and b) shows numerical feature ranking; c) and d) shows time (us).	137
B.1	3D scatter plot for the Pareto optimal solutions in experiment 2: accuracy versus sensitivity versus specificity ratio, and feature count (color bar).	171
B.1	3D scatter plot for the Pareto optimal solutions in experiment 2: accuracy versus sensitivity versus specificity ratio, and feature count (color bar).	172
B.2	3D scatter plot for the Pareto optimal solutions in experiment 3: accuracy versus sensitivity versus specificity ratio, and feature count (color bar).	174
B.2	3D scatter plot for the Pareto optimal solutions in experiment 3: accuracy versus sensitivity versus specificity ratio, and feature count (color bar).	175
B.3	3D scatter plot for the Pareto optimal solutions in experiment 4: accuracy versus sensitivity versus specificity ratio, and feature count (color bar).	176
B.3	3D scatter plot for the Pareto optimal solutions in experiment 4: accuracy versus sensitivity versus specificity ratio, and feature count (color bar).	177
B.4	Distribution of the Pareto optimal solutions from the experiment denoted by the row (“Feature ranking method” + “uFilter” + “VTA”) in Table 6.4.	179
B.5	Distribution of the Pareto optimal solution from the experiment denoted by the row “Feature extraction time method” + “uFilter” + “VTA” in Table 6.4.	179
B.6	Distribution of the Pareto optimal solutions from the experiment denoted by the row (“Feature ranking method” + “mRMR-MI + “PAF”) in Table 6.4.	179
B.7	Distribution of the Pareto optimal solutions from the experiment denoted by the row (“Feature extraction time method” + “mRMR-MI + “PAF”) in Table 6.4.	180
B.8	Distribution of the Pareto optimal solutions from the experiment denoted by the row (“Feature ranking method” + “mRMR-MI + “VTA”) in Table 6.4.	180

B.9	Distribution of the Pareto optimal solutions from the experiment denoted by the row (“Feature extraction time method” + “mRMR-MI + “VTA”) in Table 6.4.	180
-----	---	-----

**LIST OF ABBREVIATIONS**

2D	-	2 Dimensional
3D	-	3 Dimensional
ACC	-	Accuracy
AF	-	Atrial Fibrillation
AFPDB	-	Atrial Fibrillation Prediction Database
ANN	-	Artificial Neural Network
ANOVA	-	Analysis Of Variance
ANS	-	Autonomic Nervous System
API	-	Application Programming Interface
AR	-	Autoregressive
CBIR	-	Content-based Image Retrieval
CHF	-	Congestive Heart Failure
CPU	-	Central Processing Unit
DFT	-	Discrete Fourier Transform
DIT	-	Decimation-in-time
ECG	-	Electrocardiogram
FE	-	Feature Extraction
FFT	-	Fast Fourier Transform
FN	-	False Negative
FP	-	False Positive
GA	-	Genetic Algorithm
GSA	-	Gravitational Search Algorithm
HF	-	High Frequency
HOS	-	Higher Order Spectral
HR	-	Heart Rate
HRV	-	Heart Rate Variability
ICD	-	Implantable Cardioverter Defibrillator

IEEE	-	Institute of Electrical and Electronics Engineers
LF	-	Low Frequency
LSO	-	Local Search Operation
MATLAB	-	Matrix Laboratory
MI	-	Mutual Information
MOP	-	multi-objective problem
mRMR	-	Min-Redundancy and Max-Relevance
NF	-	Number of Feature
NORMAL	-	Normal Random Number Generators
NSGA-III	-	Non-dominated Sorting Genetic Algorithm III
OS	-	Operating System
PACs	-	Premature Atrial Contractions
PAF	-	Paroxysmal Atrial Fibrillation
PC	-	Personal Computer
PNS	-	Parasympathetic Nervous System
POSIX	-	Portable Operating System Interface
PSD	-	Power Spectrum Density
PVCs	-	Premature Ventricular Contractions
SA	-	Sinus Atrial
SCD	-	Sudden Cardiac Death
SD	-	Standard Deviation
SEN	-	Sensitivity
SNS	-	Sympathetic Nervous System
SPA	-	Smoothing Prior Approach
SPE	-	Specificity
SVM	-	Support Vector Machine
SVTAB	-	Spontaneous Ventricular Tachyarrhythmia Database
UNIX	-	Uniplexed Information Computing System
VF	-	ventricular Fibrillation
VT	-	Ventricular Tachycardia
VTA	-	Ventricular Tachyarrhythmia
WFDB	-	Waveform Database

**LIST OF APPENDICES**

<b>APPENDIX</b>	<b>TITLE</b>	<b>PAGE</b>
A	List of Publication	170
B	Supplementary Results	171
C	Parameter Values and Settings of HRV Feature Extraction Algorithm	181
D	List of Statistical Significant Features	182

## **CHAPTER 1**

### **INTRODUCTION**

Over the last few decades, there has been a widespread interest in the study of variations in the beat-to-beat timing of the heart, known as heart rate variability (HRV). The analysis of HRV has been used to develop the automated classification algorithm that is able to diagnose various types of cardiovascular diseases. This chapter introduces the overview of HRV analysis, defines research problems, objectives as well as research contribution and novelty of the work.

#### **1.1 Heart Rate Variability Analysis - Overview**

Heart rate (HR) represents the number of contractions per minute that occurs in the heart. The heart rate variation (HRV) occurs because of the rising and falling of heart rate that are affected by various factors such as human activity or cardiovascular related diseases. The irregularity of time interval in human heartbeats was first noted in the early 1600s. However, its physiological importance was only appreciated in 1965 when Hon and Lee [1] found that the changes in pattern of HR pattern preceded the fetal distress before changes in the baseline heart rate (average heart rate over 60 seconds). In late 1980, clinical importance of HRV became apparent when it was confirmed that the HRV was a strong and independent predictor of mortality following an acute myocardial infarction [2].

Today, there are active research interests in using the HRV as a biological marker to diagnosis various cardiovascular related diseases such as arrhythmia, diabetes and heart failure. Such interests arise because researchers have proved that the HRV is one of the most promising marker to assess the autonomic nervous system (ANS) activity [3], which can be correlated to the cardiovascular disease.

ANS is peripheral nervous system of the human body that is able to influence the activity of the internal organs below the level of consciousness. It affects digestion, heart rate, respiratory rate, salivation, perspiration pupillary dilation sexual arousal, urination and etc. There are two subsystems in ANS: parasympathetic nervous system (PNS) and sympathetic nervous system (SNS). Both subsystems operate independently in some responses and work co-operatively in others [4]. Researchers have found that the increased SNS activity or decreased PNS activity is associated with heart rate acceleration and vice versa. The relationship between ANS and HRV has encouraged the development of HRV markers.

The HRV markers can be obtained through HRV analysis [3]. The HRV analysis involves the use of different techniques or algorithms to evaluate the variation of the heart rate mathematically. These techniques can be divided into three categories: time domain, frequency domain, and non-linear analysis. Based on these techniques, various features are extracted and studied for the application in various medical research problems. The main objective is to use the HRV features to diagnose the patient, to detect occurrence of the disease event in real-time and even predict such event before it happens.

## **1.2 Cardiac Arrhythmia**

Among the HRV based research problems, the interest of this thesis is focus on improving the algorithm that uses the HRV analysis to predict the onset of the cardiac arrhythmia [3, 5]. Cardiac arrhythmia is a condition in which the electrical impulses that regulate dilation and contraction of heart do not function properly. This will cause the abnormal heart rhythm in which the heart beat too fast, too slow or irregularly (erratically). When this happens, the heart may not pump enough blood to the body and consequently damage the brain, heart and other organs.

Arrhythmia needs to be treated by specialist physician. There are many types of arrhythmia and can be classified according to the heart rate and mechanism. Arrhythmia is called tachycardia when heart beats too fast (over 100 beats per minute) and called bradycardia when heart beats that is too slow (less than 60 beats per minute). Other examples of arrhythmia include premature atrial contractions (PACs), Atrial Fibrillation (AF), ventricular fibrillation (VF), premature ventricular contractions (PVCs), heart block (First, Second and Third Degree) and more.



Arrhythmia can also be divided into life threatening and non-life-threatening arrhythmia. Occurrence of life-threatening arrhythmia such as ventricular tachyarrhythmia (VTA) can cause immediate death to a patient. Non-life threatening arrhythmia has no immediate threat to health and life of patient. However, it has long term negative impact on the health of patient. One of the more common non-life threatening arrhythmia is Atrial Fibrillation (AF). The AF increases the risk in mortality rate, stroke, heart failure and also leads to impaired cognitive function [6]. There are three types of AF, namely paroxysmal AF (PAF), persistent AF, and chronic AF. Patients often start with PAF and slowly evolving to persistent AF and chronic AF.

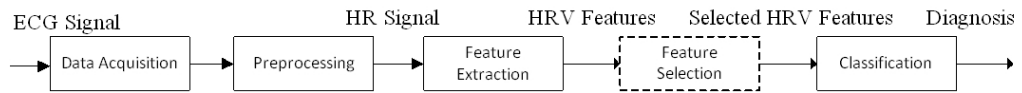
One of the popular equipment that can treat the arrhythmia is Implantable cardioverter defibrillator (ICD) [7]. It has small form factor (2cm-3cm) and is battery-powered device. It is implanted in the human body through clinical surgery. Its primary function is to detect the arrhythmia and restore it back to normal rhythm by using an electrical pulse to shock the heart muscle. It is used to treat prevalence arrhythmia such as atrial fibrillation and ventricular tachyarrhythmia.

### 1.3 Problem Statement

Two major research problems are considered in this thesis. The first issue is about the optimization of the HRV based prediction methods, and the second is about overcoming weaknesses of existing arrhythmia prediction methods.

The block diagram in Figure 1.1 shows the overview of the classification system that uses the HRV analysis for diagnosis. It shows stages of data acquisition, pre-processing, HRV feature extraction, and a supervised classifier. Initially, electrocardiogram (ECG) signal is acquired and fed to the pre-processing stage. During the pre-processing, QRS complexes of the ECG signal are detected for HRV quantification. The quantified HRV are also corrected and HRV sequences are resampled to certain frequency. Then, different HRV features are extracted in feature extraction stage. Optionally, the feature selection process is applied to select the best HRV features that can lead to high prediction performance. Finally, the supervised classification model is trained with extracted HRV features for disease diagnosis.

There are several research gaps regarding the feature selection algorithms, which have been used to optimize the performance of the HRV based classification



**Figure 1.1:** Simplified block diagram of HRV based classification system.

system.

The first outstanding issue in the optimization process is the parameter values and settings in both HRV pre-processing and feature extraction stages are not optimized (tuned) for maximum classification performance. Most previous works based on HRV analysis simply adopt the pre-defined parameter values and settings from other works. To maximize the discrimination capability of the extracted features, parameters of feature extraction algorithm should be tuned for different application and database [8]. This issue arises mainly because of the lack of automatic methods to simultaneously tune different algorithms except either by grid search or trial and error. For example, it is difficult to optimize both pre-processing and feature extraction stages at the same time. Therefore, an optimization algorithm that is able to simultaneously tune the parameters and settings in both stages for maximum classification performance is required.

Another research issue is that feature cost is not minimized explicitly during feature selection process. In previous works [9, 10], the main objective of the feature selection is to optimize (select) a HRV feature subset with respect to a single objective only. Therefore, the feature cost may be reduced but not minimized. Feature cost can be divided into two types: feature count (number of features in the selected subset) and measurement cost (time taken to extract all features in the selected subset). Both costs are important factors that affect the complexity and implementation of the classification algorithm, especially when it operates in real-time. Minimization of the measurement cost can reduce the feature extraction time in HRV feature extraction stage while minimizing the feature count can reduce complexity of the trained supervised classifier.

The existing HRV based works typically does not take into account the misclassification cost. The misclassification cost can affect both classification sensitivity and specificity rate. In medical applications, sensitivity rate is more important than specificity rate and accuracy rate because it is the success rate in recognizing the patient with disease. The failure to diagnose a patient has larger negative consequences than a failure to diagnose healthy person. Therefore, it is desired that the HRV based classification algorithms have higher sensitivity rate while

maintaining the good accuracy rate.

Genetic algorithm (GA) is usually adopted to solve multi-objective optimization problem. However, the simple GA, which is employed in previous related works, is not suitable for this work because multiple-objective optimization is required. One of the reason is that the multiple fitness functions need to be linearly combined with different weights in simple GA. Trial and error is required to tune the weights values in order to obtain a solution with desired performance. It is further complicated by the need to combine different fitness functions with difference units. Moreover, trade-off analysis between multiple fitness functions is required during the solution selection in multiple objective optimization.

In previous works based on HRV analysis, there is a shortcoming in the feature selection model of the GA: the possibility to select a feature that is affected by noisy data. This issue is well known in feature selection methods based on heuristic search such as GA. In non-HRV research works, hybrid simple GAs have been proposed to reduce this possibility. Under this model, the statistical significance test or mutual information (a types of correlation measures) of the feature is examined before it is selected to form the final feature subset. This hybrid feature selection model is integrated into the proposed optimization algorithm.

In the case of PAF onset prediction, most of the existing HRV based methods require a 30 minutes signal duration for feature extraction in order to achieve acceptable prediction accuracy levels (80% and above). Although there have been some research [11, 12, 13] that investigated the use of shorter HRV signal duration for PAF onset prediction, their prediction accuracy rates were significantly lower than 80%, as achieved by other previous works that employed 30 minutes signal [14, 5, 15, 16]. Long time duration of signal for feature extraction process is not suitable when the prediction system is implemented in implantable cardioverter defibrillator (ICD). Using the ICD to restore the PAF back to normal rhythm [17] through electrical pulse is one of the main treatments for PAF patient. A PAF onset predictor enable ICD to continuously predict the PAF onset, and possibly allowing it to be terminated early.

In recent years, much research [7, 18, 19] have shown interest in addressing the power consumption issue of the battery powered device such as ICD or similar devices, which use the HRV analysis for real time disease diagnosis. In the case of PAF onset prediction methods, the main concern is that long duration of signal and

compute-intensive HRV analysis may burden the ICD battery life, and consequently shortening its operation time. It can lead to more frequent surgery processes for battery replacement, which can affect the health of the patient [7]. (Generally, the ICD device is expected to operate for more than 5 years after it is implanted in the human body). Therefore, one of the research problem in this thesis is to reduce the required HRV signal length from 30 minutes to 5 minutes, while achieving acceptable prediction accuracy rate. The idea of the signal length reduction is also inspired by Chesnokov et al [20], who have suggested one of their future work is to examine the use of 5-15 minutes HRV signal for PAF onset prediction. Furthermore, HRV signal length of 5 minutes is the minimum length that can produce reliable HRV spectral features [4, 21].

#### **1.4 Objective**

The goal of this research is to propose a multi-objective GA based optimization algorithm for application in cardiac arrhythmia prediction system that requires shorter signal length for HRV analysis. In detail, the objectives are:

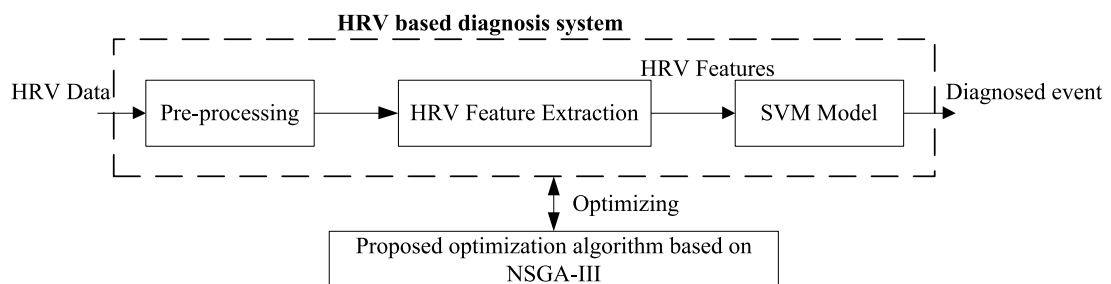
1. To propose an optimization method, based on the non-dominated sorting genetic algorithm III (NSGA-III), that can simultaneously optimize the parameters and settings in stages of HRV feature extraction, feature selection, and classifier of a cardiac arrhythmia prediction system.
2. To analyze and demonstrate the effectiveness of the proposed optimization algorithm by applying it in the prediction of the onset of cardiac arrhythmia in two cases: (a) paroxysmal atrial fibrillation (PAF) using HRV signal duration shorter than the typical 30 minutes, and (b) ventricular tachyarrhythmia (VTA).

#### **1.5 Scope of Work**

- HRV is the main signal in this work. It is converted from ECG signal or any other type of source.
- This work is limited to two cardiac arrhythmia prediction problems: PAF and VTA onset prediction.
- HRV feature extraction stage applies time domain, frequency domain, and non-linear analyses. The classifier is based on the support vector machine (SVM)

model. In this work, with some modification, the SVM is implemented by using the open source library called LIBSVM.

- No new test datasets are created for the experimental work in this thesis. Instead, standard databases are used for fair benchmarking and analysis. They are Atrial Fibrillation Prediction Database (AFPDB) and Spontaneous Ventricular Tachyarrhythmia Database (SVTAB).
- All algorithms are implemented in C/C++, and MATLAB is used for the analysis of experimental works.
- During the optimization, the trade-off between multiple optimization objectives that contradict to each other are also analyzed. The optimization objectives include the feature count, measurement cost, prediction sensitivity, specificity and accuracy rate.
- This research is confined to solving the HRV based arrhythmia prediction problem, although the proposed optimization method can be used in any HRV based research problem that uses the diagnostic system that is based on the model shown in Figure 1.2.



**Figure 1.2:** Diagnostic system model in the HRV based research.

## 1.6 Contributions

The contributions of the thesis are

- A multi-objective optimization algorithm is proposed to improve the performance of the HRV based prediction system. It is based on non-dominated sort genetic algorithm III (NSGA-III), which is a state-of-the-art optimization algorithm that employs the Pareto optimal concept. To our knowledge, it is the first attempt to use NSGA-III in the optimization of the HRV based prediction

system. Various modifications are proposed for the NSGA-III in order to tackle the various optimization goals.

- By applying the proposed optimization algorithm, the shortcomings in two arrhythmia prediction problems are overcome:
  - In the **PAF** onset prediction, the required length of HRV signal for the prediction is reduced by 83% from 30 minutes to 5 minutes, while improving the accuracy rate to 89.62%. Even with stricter performance evaluation approach, this accuracy rate is still higher than the previous works, which cannot achieve the accuracy above 70% when using less than 10 minutes HRV signal for prediction.
  - In the **VTA** onset prediction, the prediction system achieves prediction accuracy of 78.15%. It outperforms all previous works even with the application of stricter performance evaluation.
- The improvements of the proposed optimization algorithm over existing methods are:
  - The optimization process is extended from feature selection only to whole HRV based prediction system simultaneously, which include HRV pre-processing, feature extraction and SVM classifier.
  - The **hybrid feature selection model** is used to examine the quality of the features before they are used to form the feature subset. It increases the confidence level of the selected feature subset by reducing the possibility to select the features that are affected by noisy data.
  - A novel **duplication handling algorithm** that suits the hybrid selection model is proposed to handle the duplicate chromosome issue when adapting the NSGA-III. The end result is that the computation cost of the optimization process is reduced by partially avoiding the redundant evaluation of the duplicates, while the accuracy of the selected feature subset is improved.
  - The **feature costs**, which are represented by feature count and measurement cost, are minimized during the feature selection.
  - The optimization of the **misclassification cost** is taken into account. The experimental results prove that the sensitivity rate of the prediction system can be improved but at the expense of reduced specificity and accuracy rate. It should be noted that the reduced accuracy rate is still acceptable when compared to previous works.
  - The optimization concept in the NSGA-III allows the HRV based

system designer to analyze the trade-off between multiple optimization objectives that contradict to each other. Furthermore, trial and error is not required to tune the weight coefficients because different objectives are not linearly combined into single composite objective.

## **1.7 Thesis Organization**

The rest of the thesis is organized as follows. Chapter 2 summarizes the literature reviews and comparisons of related previous works to clarify the research rationale. Chapter 3 provides the fundamental background knowledge regarding the research. Chapter 4 presents research methodology. Chapter 5 presents the detail description of the HRV based arrhythmia prediction system and the proposed optimization algorithm. Chapter 6 presents the analysis and discussion about the verification and benchmarking of the algorithm. In the last chapter, the contribution of the research work is summarized and the potential future works are stated.

## REFERENCES

1. Horn, E. and Lee, S. Electronic evaluations of the fetal heart rate patterns preceding fetal death: further observation. *Am J Obstet Gynecol*, 1995. 87: 824–826.
2. Bigger, J. T., Fleiss, J. L., Steinman, R. C., Rolnitzky, L. M., Kleiger, R. E. and Rottman, J. N. Frequency domain measures of heart period variability and mortality after myocardial infarction. *Circulation*, 1992. 85(1): 164–171.
3. Rajendra Acharya, U., Paul Joseph, K., Kannathal, N., Lim, C. M. and Suri, J. S. Heart rate variability: a review. *Med Biol Eng Comput*, 2006. 44(12): 1031–51.
4. Malik, M., Camm, A., Bigger, J., Breithardt, G., Cerutti, S., Cohen, R., Coumel, P., Fallen, E., Kennedy, H., Kleiger, R. *et al.* Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use. *European Heart Journal*, 1996. 93(5): 1043–1065.
5. Mohebbi, M. and Ghassemian, H. Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal. *Comput Methods Programs Biomed*, 2012. 105(1): 40–9.
6. Manolis, A. J., Rosei, E. A., Coca, A., Cifkova, R., Erdine, S. E., Kjeldsen, S., Lip, G. Y., Narkiewicz, K., Parati, G., Redon, J. *et al.* Hypertension and atrial fibrillation: diagnostic approach, prevention and treatment. *Journal of hypertension*, 2012. 30(2): 239–252.
7. Kim, K., Cho, U., Jung, Y. and Kim, J. Design and implementation of biomedical SoC for implantable cardioverter defibrillators. *Solid-State Circuits Conference, 2007. ASSCC'07. IEEE Asian. IEEE.* 2007. 248–251.
8. Rashedi, E., Nezamabadi-Pour, H. and Saryazdi, S. A simultaneous feature adaptation and feature selection method for content-based image retrieval systems. *Knowledge-Based Systems*, 2013. 39: 85–94.
9. Yu, S.-N. and Lee, M.-Y. Bispectral analysis and genetic algorithm for congestive heart failure recognition based on heart rate variability. *Computers*



- in biology and medicine*, 2012. 42(8): 816–825.
10. Lucena, F., Barros, A. K. and Ohnishi, N. The performance of short-term heart rate variability in the detection of congestive heart failure. *BioMed Research International*, 2016. 2016.
  11. Narin, A., Isler, Y. and OMalleyMalleyzer, M. Early prediction of Paroxysmal Atrial Fibrillation using frequency domain measures of heart rate variability. *2016 Medical Technologies National Congress (TIPTEKNO)*. 2016. 1–4. doi:10.1109/TIPTEKNO.2016.7863110.
  12. Hickey, B. and Heneghan, C. Screening for paroxysmal atrial fibrillation using atrial premature contractions and spectral measures. *Computers in Cardiology, 2002*. IEEE. 2002. 217–220.
  13. Yang, A. and Yin, H. Prediction of paroxysmal atrial fibrillation by footprint analysis. *Computers in Cardiology 2001*. IEEE. 2001. 401–404.
  14. Costin, H., Rotariu, C. and Pășărică, A. Atrial fibrillation onset prediction using variability of ECG signals. *2013 8th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*. IEEE. 2013. 1–4.
  15. Chesnokov, Y. V. Complexity and spectral analysis of the heart rate variability dynamics for distant prediction of paroxysmal atrial fibrillation with artificial intelligence methods. *Artificial intelligence in medicine*, 2008. 43(2): 151–165.
  16. Tran, T., McNamers, J., Aboy, M. and Goldstein, B. Prediction of paroxysmal atrial fibrillation by analysis of atrial premature complexes. *IEEE Transactions on Biomedical Engineering*, 2004. 51(4): 561–569.
  17. Prystowsky, E. N. Management of atrial fibrillation: therapeutic options and clinical decisions. *The American journal of cardiology*, 2000. 85(10): 3–11.
  18. Massagram, W., Hafner, N., Chen, M., Macchiarulo, L., Lubecke, V. M. and Boric-Lubecke, O. Digital heart-rate variability parameter monitoring and assessment ASIC. *IEEE transactions on biomedical circuits and systems*, 2010. 4(1): 19–26.
  19. Fang, W.-C., Huang, H.-C. and Tseng, S.-Y. Design of heart rate variability processor for portable 3-lead ECG monitoring system-on-chip. *Expert Systems with Applications*, 2013. 40(5): 1491–1504.
  20. Chesnokov, Y., Holden, A. and Zhang, H. Distant prediction of paroxysmal atrial fibrillation using HRV data analysis. *Computers in Cardiology*, 2007. 34: 455–458.

21. Clifford, G. D. and Tarassenko, L. Quantifying errors in spectral estimates of HRV due to beat replacement and resampling. *IEEE transactions on biomedical engineering*, 2005. 52(4): 630–638.
22. Acharya, R., Kumar, A., Bhat, P., Lim, C., Kannathal, N., Krishnan, S. *et al.* Classification of cardiac abnormalities using heart rate signals. *Medical and Biological Engineering and Computing*, 2004. 42(3): 288–293.
23. Asl, B. M., Setarehdan, S. K. and Mohebbi, M. Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. *Artificial intelligence in medicine*, 2008. 44(1): 51–64.
24. Rouhani, M. and Soleymani, R. Neural Networks based Diagnosis of heart arrhythmias using chaotic and nonlinear features of HRV signals. *Computer Science and Information Technology-Spring Conference, 2009. IACSITSC'09. International Association of. IEEE. 2009. 545–549.*
25. Al-Angari, H. M. and Sahakian, A. V. Use of sample entropy approach to study heart rate variability in obstructive sleep apnea syndrome. *IEEE Transactions on Biomedical Engineering*, 2007. 54(10): 1900–1904.
26. Ebrahimzadeh, E., Pooyan, M. *et al.* Early detection of sudden cardiac death by using classical linear techniques and time-frequency methods on electrocardiogram signals. *Journal of Biomedical Science and Engineering*, 2011. 4(11): 699.
27. Norman, G., Karelina, K., Morris, J., Zhang, N., Berntson, G. and DeVries, C. Heart rate variability predicts cell death and inflammatory responses to global cerebral ischemia. *Frontiers in physiology*, 2012. 3: 131.
28. Penzel, T., Kantelhardt, J. W., Grote, L., Peter, J.-H. and Bunde, A. Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea. *IEEE Transactions on biomedical engineering*, 2003. 50(10): 1143–1151.
29. Ebrahimi, F., Setarehdan, S.-K., Ayala-Moyeda, J. and Nazeran, H. Automatic sleep staging using empirical mode decomposition, discrete wavelet transform, time-domain, and nonlinear dynamics features of heart rate variability signals. *Computer methods and programs in biomedicine*, 2013. 112(1): 47–57.
30. Lake, D. E., Richman, J. S., Griffin, M. P. and Moorman, J. R. Sample entropy analysis of neonatal heart rate variability. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 2002. 283(3): R789–R797.

31. Thayer, J. F., Yamamoto, S. S. and Brosschot, J. F. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. *International journal of cardiology*, 2010. 141(2): 122–131.
32. Seyd, A., Joseph, P. K. and Jacob, J. Automated diagnosis of diabetes using heart rate variability signals. *Journal of medical systems*, 2012. 36(3): 1935–1941.
33. Pontet, J., Contreras, P., Curbelo, A., Medina, J., Noveri, S., Bentancourt, S. and Migliaro, E. R. Heart rate variability as early marker of multiple organ dysfunction syndrome in septic patients. *Journal of critical care*, 2003. 18(3): 156–163.
34. Ince, T., Kiranyaz, S. and Gabbouj, M. A generic and robust system for automated patient-specific classification of ECG signals. *IEEE Transactions on Biomedical Engineering*, 2009. 56(5): 1415–1426.
35. Ye, C., Kumar, B. V. and Coimbra, M. T. Heartbeat classification using morphological and dynamic features of ECG signals. *IEEE Transactions on Biomedical Engineering*, 2012. 59(10): 2930–2941.
36. Jiang, W. and Kong, S. G. Block-based neural networks for personalized ECG signal classification. *IEEE Transactions on Neural Networks*, 2007. 18(6): 1750–1761.
37. Zong, W., Mukkamala, R. and Mark, R. A methodology for predicting paroxysmal atrial fibrillation based on ECG arrhythmia feature analysis. *Computers in Cardiology 2001*. IEEE. 2001. 125–128.
38. Alcaraz, R. and Rieta, J. J. A review on sample entropy applications for the non-invasive analysis of atrial fibrillation electrocardiograms. *Biomedical Signal Processing and Control*, 2010. 5(1): 1–14.
39. Alcaraz, R., Abásolo, D., Hornero, R. and Rieta, J. J. Optimal parameters study for sample entropy-based atrial fibrillation organization analysis. *Computer methods and programs in biomedicine*, 2010. 99(1): 124–132.
40. Lynn, K. and Chiang, H. A two-stage solution algorithm for paroxysmal atrial fibrillation prediction. *Computers in Cardiology 2001*. IEEE. 2001. 405–407.
41. Thong, T. and Raitt, M. H. Predicting imminent episodes of ventricular tachyarrhythmia using heart rate. *Pacing and clinical electrophysiology*, 2007. 30(7): 874–884.
42. Joo, S., Choi, K.-J. and Huh, S.-J. Prediction of spontaneous ventricular

- tachyarrhythmia by an artificial neural network using parameters gleaned from short-term heart rate variability. *Expert Systems with Applications*, 2012. 39(3): 3862–3866.
43. Rozen, G., Kobo, R., Beinart, R., Feldman, S., Sapunar, M., Luria, D., Eldar, M., Levitan, J. and Glikson, M. Multipole analysis of heart rate variability as a predictor of imminent ventricular arrhythmias in ICD patients. *Pacing and Clinical Electrophysiology*, 2013. 36(11): 1342–1347.
  44. Wollmann, C., Gradaus, R., Böcker, D., Fetsch, T., Hintringer, F., Hoh, G., Hatala, R., Podczeck-Schweighofer, A., Kreutzer, U., Kamaryt, P. *et al.* Variations of heart rate variability parameters prior to the onset of ventricular tachyarrhythmia and sinus tachycardia in ICD patients. Results from the heart rate variability analysis with automated ICDs (HAWAI) registry. *Physiological measurement*, 2015. 36(5): 1047.
  45. Ebrahimzadeh, E., Pooyan, M. and Bijar, A. A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analyses from HRV signals. *PloS one*, 2014. 9(2): e81896.
  46. Camm, A. J., Kirchhof, P., Lip, G. Y., Schotten, U., Savelieva, I., Ernst, S., Van Gelder, I. C., Al-Attar, N., Hindricks, G., Prendergast, B. *et al.* Guidelines for the management of atrial fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC). *European heart journal*, 2010. 31(19): 2369–429.
  47. Du, X., Ninomiya, T., de Galan, B., Abadir, E., Chalmers, J., Pillai, A., Woodward, M., Cooper, M., Harrap, S., Hamet, P. *et al.* Risks of cardiovascular events and effects of routine blood pressure lowering among patients with type 2 diabetes and atrial fibrillation: results of the ADVANCE study. *European heart journal*, 2009. 30(9): 1128–1135.
  48. Al-Khatib, S. M., Wilkinson, W. E., Sanders, L. L., McCarthy, E. A. and Pritchett, E. L. Observations on the transition from intermittent to permanent atrial fibrillation. *American heart journal*, 2000. 140(1): 142–145.
  49. Prakash, A., Saksena, S., Hill, M., Krol, R. B., Munsif, A. N., Giorgberidze, I., Mathew, P. and Mehra, R. Acute effects of dual-site right atrial pacing in patients with spontaneous and inducible atrial flutter and fibrillation. *Journal of the American College of Cardiology*, 1997. 29(5): 1007–1014.
  50. Zipes, D. P. and Wellens, H. J. Sudden cardiac death. *Circulation*, 1998. 98(21): 2334–2351.
  51. de Luna, A. B., Coumel, P. and Leclercq, J. F. Ambulatory sudden cardiac

- death: mechanisms of production of fatal arrhythmia on the basis of data from 157 cases. *American heart journal*, 1989. 117(1): 151–159.
52. Tereshchenko, L. G., Fetcs, B. J., Domitrovich, P. P., Lindsay, B. D. and Berger, R. D. Prediction of ventricular tachyarrhythmias by intracardiac repolarization variability analysis. *Circulation: Arrhythmia and Electrophysiology*, 2009. 2(3): 276–284.
  53. Reed, M., Robertson, C. and Addison, P. Heart rate variability measurements and the prediction of ventricular arrhythmias. *An International Journal of Medicine*, 2005. 98(2): 87–95.
  54. Watanabe, M. A. Heart rate turbulence slope reduction in imminent ventricular tachyarrhythmia and its implications. *Journal of cardiovascular electrophysiology*, 2006. 17(7): 735–740.
  55. Bilgin, S., Çolak, O. H., Polat, O. and Koklukaya, E. Estimation and evaluation of sub-bands on LF and HF base-bands in HRV for Ventricular Tachyarrhythmia patients. *Expert Systems with Applications*, 2009. 36(6): 10078–10084.
  56. Saeys, Y., Inza, I. and Larrañaga, P. A review of feature selection techniques in bioinformatics. *bioinformatics*, 2007. 23(19): 2507–2517.
  57. Narin, A., Isler, Y. and Ozer, M. Investigating the performance improvement of HRV Indices in CHF using feature selection methods based on backward elimination and statistical significance. *Computers in biology and medicine*, 2014. 45: 72–79.
  58. Bsoul, M., Minn, H. and Tamil, L. Apnea MedAssist: real-time sleep apnea monitor using single-lead ECG. *IEEE Transactions on Information Technology in Biomedicine*, 2011. 15(3): 416–427.
  59. De Chazal, P., Heneghan, C., Sheridan, E., Reilly, R., Nolan, P. and O'Malley, M. Automatic classification of sleep apnea epochs using the electrocardiogram. *Computers in Cardiology 2000*. IEEE. 2000. 745–748.
  60. Xie, B. and Minn, H. Real-time sleep apnea detection by classifier combination. *IEEE Transactions on Information Technology in Biomedicine*, 2012. 16(3): 469–477.
  61. Babaeizadeh, S., White, D. P., Pittman, S. D. and Zhou, S. H. Automatic detection and quantification of sleep apnea using heart rate variability. *Journal of electrocardiology*, 2010. 43(6): 535–541.
  62. De Chazal, P., Heneghan, C., Sheridan, E., Reilly, R., Nolan, P. and O'Malley,

- M. Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. *IEEE Transactions on Biomedical Engineering*, 2003. 50(6): 686–696.
63. Mendez, M. O., Bianchi, A. M., Matteucci, M., Cerutti, S. and Penzel, T. Sleep apnea screening by autoregressive models from a single ECG lead. *IEEE transactions on biomedical engineering*, 2009. 56(12): 2838–2850.
64. İşler, Y. and Kuntalp, M. Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure. *Computers in Biology and Medicine*, 2007. 37(10): 1502–1510.
65. Akhter, N., Dabhade, S., Bansod, N. and Kale, K. Feature Selection for Heart Rate Variability Based Biometric Recognition Using Genetic Algorithm. In: *Intelligent Systems Technologies and Applications*. Springer. 91–101. 2016.
66. Jelinek, H. F., Abawajy, J. H., Cornforth, D. J., Kowalczyk, A., Negnevitsky, M., Chowdhury, M. U., Krones, R. and Kelarev, A. V. Multi-layer attribute selection and classification algorithm for the diagnosis of cardiac autonomic neuropathy based on HRV attributes. *AIMS Medical Science*, 2015. 2(4): 396–409.
67. Alvarez, D., Hornero, R., Marcos, J. V. and del Campo, F. Feature selection from nocturnal oximetry using genetic algorithms to assist in obstructive sleep apnoea diagnosis. *Medical engineering & physics*, 2012. 34(8): 1049–1057.
68. Ocak, H. A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being. *Journal of medical systems*, 2013. 37(2): 1–9.
69. Boardman, A., Schlindwein, F. S., Rocha, A. P. and Leite, A. A study on the optimum order of autoregressive models for heart rate variability. *Physiological Measurement*, 2002. 23(2): 325.
70. Moody, G., Goldberger, A., McClennen, S. and Swiryn, S. Predicting the onset of paroxysmal atrial fibrillation: The Computers in Cardiology Challenge 2001. *Computers in Cardiology 2001*. IEEE. 2001. 113–116.
71. Corthout, J., Van Huffel, S., Mendez, M., Bianchi, A., Penzel, T. and Cerutti, S. Automatic screening of obstructive sleep apnea from the ECG based on empirical mode decomposition and wavelet analysis. *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2008. 3608–3611.
72. Tuzcu, V., Nas, S., Börklü, T. and Ugur, A. Decrease in the heart rate

- complexity prior to the onset of atrial fibrillation. *Europace*, 2006. 8(6): 398–402.
73. Paclik, P., Duin, R. P., van Kempen, G. M. and Kohlus, R. On feature selection with measurement cost and grouped features. *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Springer, 2002. 461–469.
74. Weiss, Y., Elovici, Y. and Rokach, L. The CASH algorithm-cost-sensitive attribute selection using histograms. *Information Sciences*, 2011. 222: 247–268.
75. Huang, C.-L. and Wang, C.-J. A GA-based feature selection and parameters optimization for support vector machines. *Expert Systems with Applications*, 2006. 31(2): 231–240.
76. Zhao, M., Fu, C., Ji, L., Tang, K. and Zhou, M. Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes. *Expert Systems with Applications*, 2011. 38(5): 5197–5204.
77. Wians, F. H. Clinical laboratory tests: which, why, and what do the results mean? *Laboratory Medicine*, 2009. 40(2): 105–113.
78. Pendharkar, P. C. A maximum-margin genetic algorithm for misclassification cost minimizing feature selection problem. *Expert Systems with Applications*, 2013. 40(10): 3918–3925.
79. Chen, N., Ribeiro, B., Vieira, A. S., Duarte, J. and Neves, J. C. A genetic algorithm-based approach to cost-sensitive bankruptcy prediction. *Expert Systems with Applications*, 2011. 38(10): 12939–12945.
80. West, D. Neural network credit scoring models. *Computers & Operations Research*, 2000. 27(11): 1131–1152.
81. Oreski, S. and Oreski, G. Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert systems with applications*, 2014. 41(4): 2052–2064.
82. El-Mihoub, T. A., Hopgood, A. A., Nolle, L. and Battersby, A. Hybrid Genetic Algorithms: A Review. *Engineering Letters*, 2006. 13(2): 124–137.
83. Huang, J., Cai, Y. and Xu, X. A hybrid genetic algorithm for feature selection wrapper based on mutual information. *Pattern Recognition Letters*, 2007. 28(13): 1825 – 1844.

84. Huang, J. and Rong, P. *A Hybrid Genetic Algorithm for Feature Selection Based on Mutual Information*, Boston, MA: Springer US. 2009, 125–152.
85. Hanchuan, P., Fuhui, L. and Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005. 27(8): 1226–1238.
86. Perez, N. P., Guevara Lopez, M. A., Silva, A. and Ramos, I. Improving the Mann-Whitney statistical test for feature selection: An approach in breast cancer diagnosis on mammography. *Artificial Intelligence in Medicine*, 2015. 63(1): 19–31.
87. Ronald, S. Duplicate genotypes in a genetic algorithm. *Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on.* IEEE. 1998. 793–798.
88. Lou, Y. and Yuen, S. Y. Non-revisiting genetic algorithm with adaptive mutation using constant memory. *Memetic Computing*, 2016: 1–22.
89. Saroj and Devraj. *A Non-revisiting Genetic Algorithm with Adaptive Mutation for Function Optimization*, Berlin, Heidelberg: Springer Berlin Heidelberg. 2012, 288–297.
90. Ratnoo, S. and Kamboj, D. A non-revisiting Genetic Algorithm with adaptive mutation for Function Optimization. *International Journal of Advanced Research in Computer Science*, 2011. 2(6).
91. Chow, C. K. and Yuen, S. Y. Continuous non-revisiting genetic algorithm with overlapped search sub-region. *IEEE Congress on Evolutionary Computation.* IEEE. 2012. 1–8.
92. Chow, C. K. and Yuen, S. Y. Continuous non-revisiting genetic algorithm with random search space re-partitioning and one-gene-flip mutation. *IEEE Congress on Evolutionary Computation.* IEEE. 2010. 1–8.
93. Yuen, S. Y. and Chow, C. K. A genetic algorithm that adaptively mutates and never revisits. *IEEE transactions on evolutionary computation*, 2009. 13(2): 454–472.
94. Yuen, S. Y. and Chow, C. K. A non-revisiting genetic algorithm. *IEEE Congress on Evolutionary Computation.* IEEE. 2007. 4583–4590.
95. Hajela, P. and Lin, C.-Y. Genetic search strategies in multicriterion optimal design. *Structural optimization*, 1992. 4(2): 99–107.



96. Murata, T. and Ishibuchi, H. MOGA: multi-objective genetic algorithms. *Evolutionary Computation, 1995., IEEE International Conference on.* IEEE. 1995, vol. 1. 289.
97. Konak, A., Coit, D. W. and Smith, A. E. Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety*, 2006. 91(9): 992–1007.
98. Deb, K. and Jain, H. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints. *IEEE Transactions on Evolutionary Computation*, 2014. 18(4): 577–601.
99. Bi, X. and Wang, C. An improved NSGA-III algorithm based on objective space decomposition for many-objective optimization. *Soft Computing*, 2016: 1–28.
100. Gacek, A. and Pedrycz, W. *ECG signal processing, classification and interpretation: a comprehensive framework of computational intelligence.* Springer Science & Business Media. 2011.
101. McNames, J., Thong, T. and Aboy, M. Impulse rejection filter for artifact removal in spectral analysis of biomedical signals. *Engineering in Medicine and Biology Society, 26th Annual International Conference of the IEEE.* 2004, vol. 1. 145–148.
102. Jang, D.-G., Hahn, M., Jang, J.-K., Farooq, U. and Park, S.-H. A comparison of interpolation techniques for RR interval fitting in AR spectrum estimation. *2012 IEEE Biomedical Circuits and Systems Conference (BioCAS).* IEEE. 2012. 352–355.
103. Tarvainen, M. P., Ranta-aho, P. O. and Karjalainen, P. A. An advanced detrending method with application to HRV analysis. *IEEE Transactions on Biomedical Engineering*, 2002. 49(2): 172–175.
104. Zhang, F., Chen, S., Zhang, H., Zhang, X. and Li, G. Bioelectric signal detrending using smoothness prior approach. *Medical Engineering & Physics*, 2014. 36(8): 1007 – 1013.
105. Harris, F. J. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 1978. 66(1): 51–83.
106. Cooley, J. W., Lewis, P. A. and Welch, P. D. The fast Fourier transform and its applications. *IEEE Transactions on Education*, 1969. 12(1): 27–34.
107. Bos, R., de Waele, S. and Broersen, P. M. Autoregressive spectral estimation

- by application of the Burg algorithm to irregularly sampled data. *IEEE transactions on instrumentation and measurement*, 2002. 51(6): 1289–1294.
108. De Waele, S. and Broersen, P. M. The Burg algorithm for segments. *IEEE Transactions on Signal Processing*, 2000. 48(10): 2876–2880.
  109. Stoica, P. and Moses, R. L. *Introduction to spectral analysis*. vol. 1. Prentice hall Upper Saddle River. 1997.
  110. Zhou, S.-M., Gan, J. Q. and Sepulveda, F. Classifying mental tasks based on features of higher-order statistics from EEG signals in brain computer interface. *Information Sciences*, 2008. 178(6): 1629–1640.
  111. Pinhas, I., Toledo, E., Aravot, D. and Akselrod, S. Bicoherence analysis of new cardiovascular spectral components observed in heart-transplant patients: statistical approach for bicoherence thresholding. *IEEE Transactions on Biomedical Engineering*, 2004. 51(10): 1774–1783.
  112. Nikias, C. L. and Raghuveer, M. R. Bispectrum estimation: A digital signal processing framework. *Proceedings of the IEEE*, 1987. 75(7): 869–891.
  113. Chua, K. C., Chandran, V., Acharya, U. R. and Lim, C. M. Cardiac state diagnosis using higher order spectra of heart rate variability. *J Med Eng Technol*, 2008. 32(2): 145–55.
  114. Akay, M. *Nonlinear Biomedical Signal Processing Vol. II: Dynamic Analysis and Modeling*. Wiley-IEEE press. 2000.
  115. Richman, J. S. and Moorman, J. R. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, 2000. 278(6): H2039–H2049.
  116. Pincus, S. M. and Goldberger, A. L. Physiological time-series analysis: what does regularity quantify? *American Journal of Physiology - Heart and Circulatory Physiology*, 1994. 266(4): H1643–H1656.
  117. Woo, M. A., Stevenson, W. G., Moser, D. K., Trelease, R. B. and Harper, R. M. Patterns of beat-to-beat heart rate variability in advanced heart failure. *American Heart Journal*, 1992. 123(3): 704 – 710.
  118. Cortes, C. and Vapnik, V. Support-Vector Networks. *Machine Learning*, 1995. 20(3): 273–297.
  119. Holland, J. H. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. U Michigan Press. 1975.
  120. Das, I. and Dennis, J. *Normal-Boundary Intersection: An Alternate*

- Method for Generating Pareto Optimal Points in Multicriteria Optimization Problems*. Technical report. DTIC Document. 1996.
121. Wierzbicki, A. P. The use of reference objectives in multiobjective optimization. In: *Multiple criteria decision making theory and application*. Springer. 468–486. 1980.
  122. Chang, C.-C. and Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2011. 2(3): 1–27.
  123. Butenhof, D. R. *Programming with POSIX threads*. Addison-Wesley Professional. 1997.
  124. Burkardt, J. Normal Random Number Generators (NORMAL) Library. [https://people.sc.fsu.edu/~jburkardt/cpp\\_src/normal/normal.html/](https://people.sc.fsu.edu/~jburkardt/cpp_src/normal/normal.html/), 2012.
  125. MATLAB. *version 8.4.0.150421 (R2014b)*. Natick, Massachusetts: The MathWorks Inc. 2014.
  126. Silva, I. and Moody, G. B. An open-source toolbox for analysing and processing physionet databases in matlab and octave. *Journal of open research software*, 2014. 2(1).
  127. Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K. and Stanley, H. E. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*, 2000. 101(23): e215–e220.
  128. <http://physionet.org/physiobank/database/mvtdb/>. URL <http://physionet.org/physiobank/database/mvtdb/>.
  129. Gibbons, J. D. and Chakraborti, S. *Nonparametric statistical inference*. Springer. 2011.
  130. Ge, H. and Hu, T. Genetic Algorithm for Feature Selection with Mutual Information. *Computational Intelligence and Design (ISCID), 2014 Seventh International Symposium on*. vol. 1. 116–119.
  131. Irani, K. B. Multi-interval discretization of continuous-valued attributes for classification learning. 1993.
  132. Bengio, Y. and Grandvalet, Y. No Unbiased Estimator of the Variance of K-Fold Cross-Validation. *J. Mach. Learn. Res.*, 2004. 5: 1089–1105. ISSN 1532-4435.
  133. Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T. A fast and

- elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 2002. 6(2): 182–197.
134. Yuan, Y., Xu, H. and Wang, B. An Improved NSGA-III Procedure for Evolutionary Many-objective Optimization. *In Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation*. ACM. 661–668.