ECG ANALYSIS FOR ARRHYTHMIA DETECTION AND CLASSIFICATION

BY

HAMZA BAALI

A dissertation submitted in fulfilment of the requirement for the degree of Doctor of Philosophy (Mechatronics Engineering)

Kulliyyah of Engineering
International Islamic University Malaysia

AUGUST 2014

ABSTRACT

Though various techniques have been suggested for the analysis of ECG signals, interpretation of these signals, especially as they affect human health, has posed some difficulties. Consequently, the best way of interpreting these physiological signals by electric measurements from the body surface in terms of cardiac electric activity remains an active research topic till today. This research tackles three problems related to ECG analysis namely, parametric modeling, period normalization (interpolation) and classification of arrhythmias. In order to model the signal, each heartbeat is first mapped into a new domain where the transform coefficients vector would be sparse. The coefficients vector is then approximated to a sum of damped sinusoids. The transform matrix is generated based on the combination of linear prediction (LP) and the singular values decomposition (SVD) of the LPC filter impulse response matrix. This approach leads to relatively satisfactory compression ratio (*CR*) as compared to existing techniques. Though parametric modeling of ECG signals has a central role in real time transmission and classification of heart abnormalities (arrhythmias), the compression ratios achieved are not suitable for storage purpose. Therefore, 2D ECG compression schemes are adopted where the beats of differing periods should be equalized to the same period length and then arranged in an image matrix before the application of image compression algorithm. Limitations of the existing techniques for ECG period equalization are highlighted and a new frequency domain approach for period normalization has been developed. The proposed approach is signal dependent and able to adapt to the signal characteristics. An analytical model to generate basis functions has also been developed. The merits of the proposed technique are appreciated when compared to other techniques commonly used in the literature. Finally, an algorithm for arrhythmia classification that conforms to the recommended practice of the Association for the Advancement of Medical Instrumentation (AAMI) is presented. Three inter-patient classification scenarios have been considered namely, detection of ventricular ectopic beats (VEBs), detection of supraventricular ectopic beats (SVEBs) and the multiclass recommended taxonomy.A novel set of features extraction via the application of orthogonal transformation of the ECG signal has been developed. These features in conjunction with some commonly used features are fed into the Regularized Least Squares Classifier (RLSC) with linear kernel. The proposed classification scheme shows good separation capability between the classes of ECG arrhythmias as it has achieved a Balanced Classification Rate (*BCR*) of 83.9 % for the multiclass scenario which is comparable to the state-of-the-art performance of automatic arrhythmia classification algorithms.

ملخص البحث

على الرغم من اقتراح تقنيات مختلفة لتحليل إشارات تخطيط القلب، لا يزال تفسير هذه الإشارات ، خاصة وأنها تؤثر على صحة الإنسان ، يطرح بعض الصعوبات. وبالتالي ، فإن أفضل طريقة لتفسير هذه الإشارات الفيزيولوجية االمقاسة على سطح الجسم لمعرفة النشاط الكهربائي في القلب لا يزال يشكل موضوع بحث نشط حتى اليوم.هذا البحث يتناول ثلاث مشاكل تتعلق بتحليل تخطيط القلب وهي التمثيل الرياضي للاشارة ، الاستيفاء، وتصنيف عدم انتظام ضربات القلب . من أجل التمثيل الرياضي للاشارة ، يتم اولا تمثيل كل نبضة في مجال جديد حيث تكون اغلب معاملات الشعاع في المجال الجديد صغيرة.بعدها يتم تقريب الشعاع الناتج عن التحويلة الى مجموع منحنيات جيبية متخامدة. يتم إنشاء مصفوفة التحويل باستعمال التنبؤ الخطي (LP) و (SVD) للفلترLPC. هذه الطريقة ادت إلى نسبة ضغط مرضية نسبيا (CR) بالمقارنة مع التقنيات الحالية. على الرغم من ان التمثيل الرياضي لاشارة القلب له دور مركزي في إرسال هذه الاشارة في الوقت االمحدد، و تصنيف تشوهات القلب (عدم انتظام ضربات القلب)، تبقى نسب الضغط المحققة ليست مناسبة لأغراض التخزين. لذلك، يتم اعتماد مخططات ضغط 2D ECG حيث ينبغي أن تساوى فترات الدقات المختلفة ومن ثم ترتيبها في مصفوفة قبل تطبيق خوارزمية ضغط الصورة . تم تسليط الضوء على نقائص التقنيات الحالية لتسوية فترات دقات القلب و تطوير طريقة جديدة اهذا الغرض. الطريقة المقترحة قادرة على التكيف مع خصائص الإشارة. كما تم تطوير نموذج تحليلي لتوليد اشعة مصفوفة التحويل. وتمت دراسة مزايا التقنية المقترحة بالمقارنة مع التقنيات الأخرى المستخدمة عادة في االدراسات السابقة. أخيرا، تم تقديم خوارزمية لتصنيف عدم انتظام ضربات القلب يتوافق مع الممارسات الموصى بها من جمعية النهوض الأجهزة الطبية (AAMI). وقد اعتبرنا خلال الدراسة ثلاثة سيناريوهات تصنيف حالات (VEBs)، والكشف عن (SVEBs) بالاضافة الى التصنيف المتعدد للاضطرابات. وقد تم تطوير مجموعة جديدة من الميزات عبر تطبيق التحويل المتعامدة لإشارة تخطيط القلب. وتم تغذية هذه الميزات جنبا إلى جنب مع بعض الميزات االمستعملة عادة استخدامها في مصنف (RLSC) مع نواة الخطي. بين مخطط التصنيف المقترح القدرة فصل جيدة بمعدل (BCR) من 83.9٪ لسيناريو multiclass وهو مشابه لأداء خوارزميات التصنيف عدم انتظام ضربات القلب المستعملة .

APPROVAL PAGE

The thesis of Hamza Baali has been approved by the followings:

Momoh J.E Salami
Supervisor

Othman.O.Khalifa
Internal Examiner

Nahrul Khair Bin Alang Rashid
Internal Examiner

Yussof Mashor
External Examiner

Md. Yousuf Ali
Chairman

DECLARATION

I hereby declare that this dissertation is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institution.

Hamza Baali

Signature …………………...……………… Date ……………..……

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

ECG ANALYSIS FOR ARRHYTHMIA DETECTION AND CLASSIFICATION

Affirmed by: Hamza Baali

------------------------------ ---------------------------
Signature Date

ACKNOWLEDGEMENTS

All praise is due unto Allah who has made it possible for one to attain this level of academic pursuit. We praise Him for endowing us with the time, ability and resources for the accomplishment of this goal.

With due respect and a sense of appreciation, I acknowledge the support, contributions and encouragement of my supervisors Prof. Dr. Momoh and Prof. Dr. Rini towards the completion of my work; and particularly, the financial support of Prof. Dr. Rini is highly appreciated. I pray Allah to reward them abundantly.

May the blessings of Allah continue to shower on my parents Mr. & Mrs. Baali who bear the burden of responsibility for my social and academic wellbeing both in times of ease and adversity. May Allah grant them the opportunity to reap the fruits of their labour. I do also acknowledge the moral support of my brothers and sisters; and other relations of mine who have kindly contributed to the success of this venture.

I offered my unreserved gratitude to my friends who have provided a great deal of moral support and guidance in the course of my studies. May Allah reward them accordingly.

Finally, I pray Allah to reward all those who have contributed in one way or the other to the successful completion of my programme.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACC	Accuracy
AAMI	Association Advancement of Medical Instrumentation
AZTEC	Amplitude Zone Time Epoch Coding
AR	Autoregressive
ARX	Autoregressive model with Exogeneous Input
BCR	Balanced Classification Rate
CVD	Cardiovascular Disease
CELP	Code Excited Linear Prediction
CR	Compression Ratio
CORTES	Coordinate Reduction Time Encoding System
CT	Cosine Transform
DC	Direct Current
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DWPT	Discrete Wavelet Packet Transform
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
EBR	Energy Based Ratio
EDS	Exponentially Damped Sinusoids
FIR	Finite Impulse Response
F	Fusion beats
HBF	Hermite Basis Functions
HALF	High Amplitude Low Frequency

HOS	High Order Statistics
IIR	IIR Infinite Impulse Response
ICU	Intensive Care Units
IDCT	Inverse Discrete Cosine Transform
KLT	Karhunen-Loève Transform
LM	Levenberg-Marquardt
LDA	Linear Discriminant Analysis
LPCs	Linear Prediction Coefficients
LAHF	Low Amplitude High Frequency
ML	Machine Learning
MSE	Mean Squared Error
MA	Moving Average
ARMAX	Moving Average with Exogenous Variable
MRA	Multi-Resolution Analysis
N	Normal beats
NN	Neural Networks
OVA	One-Versus-All
PDWT	Packet Discrete Wavelet Transform
PRD	Percent Root mean square Difference
+P	Positive Predictivity
PVC	Premature Ventricular Contraction
RLS	Regularized Least Squares
RLSC	Regularized Least Squares Classifier
SRC	Sampling Rate Conversion
SAPA	Scan Along Polygonal Approximation

Se	Sensitivity
SA	Sinoatrial
SA node	Sinoatrial node
SNR	SNR Signal to Noise Ratio
SVEBs	Supraventricular Ectopic Beats
Sp	Specificity
SVM	Support Vector Machine
TP	Turning Point
2D	Two-Dimensional
VEBs	Ventricular Ectopic Beats
VE	Ventricular Escape
WT	Walsh Transform
WHO	World Health Organization
2D DCT	2D Discrete Cosine Transform

LIST OF SYMBOLS

V_1, V_{II}, V_{III}	Lead voltages
ϕ_L	Potential of the left arm
ϕ_R	Potential of the right arm
ϕ_F	Potential of the left foot
aVL, aVR and aVF	Unipolar leads
f_s	Sampling frequency
f_0	Fundamental frequency
b_0	Gain
r	Notch filter width
ω_0	Oscillation frequency
z^n	Complex exponential sequence
y(n)	ECG signal (zero state response)
ε	Threshold
s(n)	ECG signal (Full response)
$\mu(n)$	Zero input response
$\mathbf{w}$	Transform ECG vector using KLT
$\boldsymbol{\varphi}_k$	Basis vector
$\boldsymbol{\Phi}$	KLT matrix
N	ECG heartbeat length
$\hat{\boldsymbol{y}}$	Approximated ECG
$\mathbf{C}$	Covariance matrix
λ_k	K^{th} eigenvalue

$\mathbf{\Lambda}$	Eigenvalues matrix
$\boldsymbol{J}_m$	Mean square error
$\mathbf{u}$	Vector of mean value
$\boldsymbol{U}$	Left singular values matrix
$\boldsymbol{V}$	Right singular values matrix
$\boldsymbol{D}$	Singular values matrix
σ_i	i^{th} singular values
ψ_ℓ	Hermite basis function
H_l	Hermite polynomials
c_{lm}	Expansion coefficient
E	Sum of squared error
H	Filter transfer function
q	Shift operator
$\psi_{j,k}$	Wavelet function
$\varphi_{j,k}$	Scaling function
$d_{j,k}$	Wavelet coefficient
a_j^n	Approximation coefficient
$\mathbb{Z}$	Complex numbers space
$\mathbb{R}$	Real numbers
$L^2(\mathbb{R})$	Hilbert space
V_j	Approximation space
W_j	Wavelet space
a_i	i^{th} LPC coefficient
E	Sum of squared error
R	Correlation

e	Residual error
$A(z)$	Analysis filter
$H(z)$	Synthesis filter
$\boldsymbol{Y}$	ECG vector
$\boldsymbol{H}$	Impulse response matrix
$\boldsymbol{e}$	Residual error vector
$\boldsymbol{\theta}$	Transformed ECG vector
$\boldsymbol{\zeta}$	Transformed Residual error vector
$\hat{\theta}(n)$	Estimated ECG
A_m	Amplitude of the complex exponential
σ_m	Damping factor in seconds^{-1}
f_m	Oscillation frequency in hertz
φ_m	Initial phase in radians
T	Sampling time in seconds
$F(z)$	Auxiliary polynomial
$\boldsymbol{\theta}^{nz}$	Normalized transformed ECG
$\boldsymbol{Y}^{nz}$	Normalized ECG vector
N^*	Normalization period
$\boldsymbol{H}^{nz}$	Normalized impulse response matrix
$\boldsymbol{U}^{nz}$	Normalization Left singular vector matrix
$\boldsymbol{B}$	Inverse of $\boldsymbol{H}$
$\boldsymbol{E}_{re}$	Residual error energy
$skew$	Third order central moment
$Kurt$	Fourth order central moment
x_i	i^{th} feature element

$\hat{x}_i$	Normalized ith feature element
V	loss function
K	Kernel function
f^*	Regularization solution
$\|f\|_K^2$	Norm in the Hilbert space
$\mathbf{I}$	Identity Matrix
$\boldsymbol{c}$	Closed form solution
$\mathbf{x_t}$	Testing point

CHAPTER ONE

INTRODUCTION

1.1 OVERVIEW

Cardiovascular Diseases (CVDs) are the leading causes of death in the world, where more than 80% of these cases are found in developing countries (Goldberger et al., 2000). This leading position will last for the next thirty years as forecasted by the World Health Organization (WHO) (Organization, 2012). In terms of numbers, CVDs claimed the lives of about 17.3 million of the world population (i.e., 30% of the global deaths) in 2008. In addition, the estimated economical cost of heart related diseases in the United States only was about 316.4 us $ billion in 2010. This cost covers health care services, medications and decrease in productivity (Frieden, 2010). For accurate and early-on assessing of different cardiac diseases Electrocardiogram (ECG) is a crucial non-invasive diagnostic tool. Abnormalities in both electrical generation and conduction at different levels in the heart are reflected on the surface ECG as deviations from the normal heart rhythm. The term arrhythmia is used to refer to these deviations (Clifford et al., 2006).

In general, the main challenge in developing countries is due to an inadequate number of physicians who are able to read and analyze ECG signal particularly in rural areas. In developed countries, on the other hand, the increasing number of patients in Intensive Care Units (ICU) and the large amount of data recorded by the Holter monitors make it almost unfeasible for the physicians to manually analyze all the acquired data (Goldberger et al., 2000). One of the engineering solutions for the mentioned problem is applying machine learning techniques.

Machine Learning (ML), provides an automatic and low cost analysis of ECG data, which can assist the human being. Such analysis can reveal hidden information that is crucial for final decision. The fast development in ML algorithms and the possibility of extraction of discriminative and stable features using signal processing techniques give rooms to improve on the current state-of-the-art. Consequently, the possibility to save many lives if heart abnormalities are detected early-on and accurately is garanteed.

Generally, learning algorithms can be grouped into two main categories. In the first, supervised learning, it is assumed that a set of training data is available, and the classifier is designed by exploiting this apriori known information (Chazal et al., 2004; Clifford et al., 2006; Kampouraki et al., 2009; Minami et al., 1999; Osowski et al., 2004). In the second category, training data of known class labels is not available. In this type of problems, a set of feature vectors is given and the goal is to unravel the underlying similarities and cluster (group) “similar” vectors together. This is known as unsupervised pattern recognition, unsupervised learning or clustering (Francisco et al., 2007; Khawaja, 2006; Sotelo, 2010). With the increase of the number of abnormalities the clustering task becomes more challenging with unsupervised learning methods.

One of the most promising algorithms for supervised learning is the Regularized Least Squares Classifier (RLSC). This algorithm has shown to perform as accurately as Support Vector Machine (SVM) with some advantages in terms of reduced computational complexity and memory requirements when applied with linear kernel (Rifkin, 2002).

Parametric modeling of the ECG signal serves to reduce the size of the data for real-time transmission and to provide features for signals classification. However, by applying parametric modeling techniques only intersamples (intrabeat) redundancy is exploited. Interbeats redundancy manifested by the quasi-periodicity of the ECG signal can be exploited by adopting Two-Dimensional (2D) image compression algorithms which are more suitable for storage purposes.

1.2 PROBLEM STATEMENT AND ITS SIGNIFICANCE

Usually a large amount of ECG data is recorded from each patient (about 100,000 heartbeats daily), hence there is a need to store and retrieve these data efficiently for further consultation. Furthermore, some special applications of telemedicine, where consultation between medical specialists in different locations is conducted in real-time, need a compact representation of the ECG signal. This compact representation of the signal helps to reduce the time and the cost of transmission through telecommunication networks. Examples of these applications include transmissions initiated from an ambulance or a patient's home to the hospital for early diagnosis.

The ability of signal processing techniques to detect and classify automatically and rapidly the large amount of data generated by the Holter monitor at low cost when compared to manual analysis has brought the interest of many researchers in the last decades to develop new algorithms for automatic ECG monitoring.

Each of these research problems are discussed more specifically as follows:

a) Lack of efficient modeling techniques for ECG signals. Existing techniques fail to bring acceptable signal reconstruction for clinical evaluation in many cases due to the fact that they are based on symmetry assumptions or due to the large variety in the morphology of the ECG within and across patients (Osowski and Linh, 2001). The assumptions of the symmetry of the ECG waves are suitable to model normal and some pathological rhythms but they do not hold for abnormal rhythms which are clinically more important.

b) The best way for ECG period normalization still posses some challenges. Due to the capability of 2D ECG compression algorithms to exploit further the redundancy in the signal when compared to 1D compression algorithms and to yield better compression ratios, an extensive research effort has been devoted to their development in recent years. In 2D ECG compression, the beats of differing periods should be equalized to the same period length using different techniques. These techniques can be classified into two main categories, namely signal extension and period normalization techniques. Unlike signal extension, period normalization techniques are lossy. However, the latter produce a lower Percent Root mean square Difference (PRD) as compared to the former when used with 2D ECG coder (Chou et al., 2006). This outperformance is justified by the fact that period normalization provides higher inter-beat correlations compared to signal extension. The main problem of the widely used period normalization technique for ECG signals introduced by Wei et al is that it cannot process extremely irregular ECG very well (Wei et al., 2001). This problem has been observed and documented by (Chou et al., 2006). Sampling Rate Conversion (SRC) in the frequency domain using sinusoidal transforms (Discrete Fourier Transform (DFT) or Discrete Cosine Transform (DCT)), has not received sufficient attention from the research community in the past and have not been considered for ECG signals period normalization. However, recently Bi and Mitra have shown the merits of this approach in terms of lower computational complexity