BIRZEIT UNIVERSITY
Birzeit, Palestine

BIRZEIT UNIVERSITY- FUCULTY OF INFORMATION TECHNOLOGY

SCIENTIFIC COMPUTING DEPARTMENT

Classification of Electrocardiogram (ECG) Arrhythmias Based on Neuro-Fuzzy System

By

Sabrin Hamoudeh

Superviser

Dr. Wasel Ghanem

This thesis was submitted to the Scientific Computing Department in partial fulfillment of the requirements for the degree of M.Sc. in Scientific Computing

BIRZEIT

-April 2010-

Classification of Electrocardiogram (ECG) Arrhythmias Based on Neuro-Fuzzy System

By

Sabrin Hamoudeh

This thesis was defended successfully on …………………… and approved by:

Committee members	**Signature**
1. …………………………………………………………………	………………………………………
2. …………………………………………………………………	………………………………………
3. …………………………………………………………………	………………………………………
4. …………………………………………………………………	………………………………………

To

MY COMING FAMILY

TABLE OF CONTENT

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Meaning
Neuro-Fuzzy	**N**eural **N**etwork & **F**uzzy logic
ANN	**A**rtificial **N**eural **N**etwork
PCA	**P**rincipal **C**omponent **A**nalysis
DWT	**D**iscrete **W**avelet **T**ransform
CPU	**C**entral **P**rocessing **U**nit
ECG	**E**lectro**C**ardio**G**ram
VPC	**V**entricular **P**remature **C**ycle
MLPN	**M**ulti **L**ayer **P**erceptron **N**etwork
EKF	**E**xtended **K**alman **F**ilter
FIS	**F**uzzy **I**nference **S**ystem
ANFIS	**A**daptive **N**euro-**F**uzzy **I**nference **S**ystem
DCT	**D**iscrete **C**osine **T**ransform
FLC	**F**uzzy **L**ogic **C**ontroller
UNANR	**U**nbiased & **N**ormalized **A**daptive **N**oise **R**eduction
HOS	**H**igher **O**rder **S**tatistics
SAN	**S**ino **A**trial **N**ode
RBBB	**R**ight **B**undle **B**ranch **B**lock
LBBB	**L**eft **B**undle **B**ranch **B**lock
P	**P**aced beat
N	**N**ormal beat
FT	**F**ourier **T**ransform
WT	**W**avelet **T**ransform
CWT	**C**ontinuous **WT**
P	**P**roportional
I	**I**ntegral
D	**D**erivative
Neg	**N**egative
Pos	**P**ositive
NB	**N**egative **B**ig
NM	**N**egative **M**edium
PM	**P**ositive **M**edium

PB	**P**ositive **B**ig
BOA	**B**isector **O**f **A**rea
MOM	**M**ean **o**f **M**axima
EMG	**E**lectro**M**yo**G**raphy
IIR	**I**nfinite **I**mpulse **R**esponse
LMS	**L**east **M**ean **S**quare
FIR	**F**inite **I**mpulse **R**esponse
FFT	**F**ast **F**ourier **T**ransform
MIT	**M**assachusetts **I**nstitute of **T**echnology
BIH	**B**oston's **B**eth **I**srael **H**ospital
FD1	**F**irst **D**erevative algorithm

كلمة شكر

لا بد لنا ونحن نخطو خطواتنا الأخيرة في الحياة الجامعية من وقفة نعود إلى أعوام قضيناها في رحاب الجامعة مع أساتذتنا الكرام الذين قدموا لنا الكثير باذلين بذلك جهودا كبيرة في بناء جيل الغد

لتبعث الأمة من جديد...

وقبل أن نمضي نقدم أسمى آيات الشكر والامتنان والتقدير والمحبة إلى الذين حملوا أقدس رسالة في الحياة....

إلى الذين مهدوا لنا طريق العلم والمعرفة....

إلى جميع أساتذتنا الأفاضل..........

واخص بالتقدير والشكر:

الدكتور واصل غانم

الذي تفضل بالإشراف على هذا البحث فجزاه الله عنا كل خير

وكذلك أعضاء لجنة النقاش :

الدكتور وصفي الكفري

الدكتور رشيد الجيوسي

إلى من زرعوا التفاؤل في دربنا وقدموا لنا المساعدات والتسهيلات , ربما دون أن يشعروا بدورهم بذلك فلهم منا كل الشكر.

أما الشكر الذي من النوع الخاص فنحن نتوجه به أيضا إلى كل من لم يقف إلى جانبنا, ومن وقف في طرقنا وعرقل مسيرة بحثنا وزرع الشوك في طريقه , فلولا وجودهم لما أحسسنا بمتعة البحث, ولا حلاوة المنافسة الايجابية, ولولاهم لما وصلنا إلى ما وصلنا إليه, ولا اكتشفنا مكنون ذاتنا فلهم منا كل الشكر.

الاهداء

يا من افتقدك منذ الصغر

يا من علمني النجاح والصبر

يا من افتقده في مواجهة الصعاب

ولم تمهله الدنيا لأرتوي من حنانه

يا من أودعتني لله أهديك هذا البحث...

أبي

إلى من تتسابق الكلمات لتخرج معبرة عن مكنون ذاتها

من علمتني وعانت الصعاب لأصل إلى ما أنا فيه

إلى من تشفي بلمستها أنين السنين وعثرات الزمان...

إلى أمي الغالية

إلى من آثروني على أنفسهم

إلى من علموني علم الحياة

إلى الضحكات التي عاركت ضحكتي

إلى الدموع التي واست دمعتي

إلى الخطى التي شدت خطوتي ..

إلى أخي وأخواتي

إلى القلب الذي لون الحياة حولي

إلى الروح التي سكنت روحي

إلى من ينبض بنبضي واكبر بتواجده على جدران قلبي

وعندما تكسوني الهموم أسبح في بحر حنانه ليخفف من آلامي

إلى رفيق دربي ونبراس الأمان في رحلتي نحو الحياة ...

إلى خطيبي "محمد"

إلى كل من أضاء بعلمه عقل غيره

أو هدى بالجواب الصحيح حيرة سائليه

فاظهر بسماحته تواضع العلماء

وبرحابته سماح العارفين...

إلى أساتذتي الفاضلين

إلى كل العيون التي آزرت تجربتي

إلى كل نفس زرعت في طريقي وردة...

إلى زميلاتي وزملائي

إلى من أعطوني الثقة بتواجدهم هنا

إلى من أرى بعيونهم أني قادرة على التغيير والتجديد

إليكم جميعا اهدي هذا البحث

المستخلص

تهدف هذه الرسالة إلى تصنيف إشارات القلب (ECG) باستخدام الشبكة العصبية الصناعية والمنطق الضبابي. حيث سيتم استخدام الشبكة العصبية المغذية للأمام متعددة الطبقات لتصنيف هذه الإشارة فيما إذا كانت طبيعية أم لا. و سيتم تعليم هذه الشبكة باستخدام طريقة الانتشار الخلفي والمنطق الضبابي والذي بدوره سيقوم بتسريع معدل التقارب لمعدل التعلم (Learning rate of convergence) من خلال التحكم بمعدل التعلم بواسطة (if-then rules) مقارنة بالطرق والأساليب المستخدمة سابقا. من ناحية أخرى, هناك طريقتين سيتم استخدامهما في عملية استخراج النقاط المميزة في إشارة القلب وهما : تحليل العناصر الأساسية(PCA) وتحويل الموجه المنفصل(DWT), بحيث ستكون معاملات هذه الطرق هي المدخلات للشبكة العصبية بالإضافة إلى نظرية شانون (Shannon Entropy) التي ستحدد العدد المثالي لهذه المعاملات لتعليم الشبكة. ونتيجة لذلك, فقد أثبتت النتائج بان تعديل طريقة التعلم "الانتشار الخلفي" لتحتوي (if-then rules) أكثر فاعلية من الطريقة التقليدية "الانتشار الخلفي" في تسريع عملية تعليم الشبكة العصبية.

Abstract

Classification of ECG beats using Neuro-Fuzzy approach is the aim of this work. In other words, feed-forward artificial neural network (ANN) is trained by the back-propagation algorithm in combination with fuzzy logic rules that controls the learning rate by if-then rules. Therefore, faster rate of convergence is achieved by this new designation in comparison with those in literature which they used the traditional back-propagation. On the other hand, Principal Component Analysis (PCA) and Discrete Wavelet Transform (DWT) are two powerful techniques that they used for feature extraction. In addition, Shannon entropy criterion is used for determining the optimal number of coefficients for the above techniques; therefore, dimensionality reduction is achieved. The performance parameters of the system are the CPU-time and the percentage average accuracy. As a result, three classifiers are proposed from this combination: PCA-ANN and DWT-ANN classifiers that they trained by the traditional back-propagation algorithm and DWT-Neuro-Fuzzy classifier with percentage average accuracy 96.83%, 98.59% and 98.68% and CPU-time 9.69, 7.75, and 7.42 seconds respectively.

1 Introduction

Everywhere around us are signals that need to be analyzed; Human speech, medical images, music and many other types of signals. One of the most important signals in biomedical field is the Electrocardiogram signal (ECG).

Electrocardiograms (ECG) are the signals that originate from the activity of the human heart; it consists of waves and complexes with segments and intervals arranged in a specific way. Any abnormalities in these signals lead to what is known arrhythmias. Various heart diseases can be easily diagnosed by an experienced cardiologist by looking at the ECG waveforms printout. But, Due to large number of patients in the intensive care units that they need continuous observations, several computerized ECG diagnosis tools have been developed. These tools are one application of pattern recognition which has generally three main stages: Preprocessing stage consists of techniques that segment and de-noise the ECG signals and this is aimed to best feature extraction and high degree of recognition accuracy. Then, Feature extraction stage which consists of techniques for Extracting crucial information from ECG patterns to be classified. The final stage is the post-processing stage in which the ECG patterns can be categorized into different classes or categories.

This research is aimed to development new computerized ECG tool based on hybrid system of neural network and fuzzy logic (Neuro-Fuzzy approach) for ECG signals that analyze and diagnose them.

1.1 Neuro-Fuzzy Literature Survey

In literature, many ECG pattern recognition systems based on Neuro-Fuzzy approach were developed. These systems formed from different combination of various pre-processing and feature extraction techniques.

In the past years, many ECG classifiers based on statistical methods, clustering methods, expert systems and marcov models were developed. These methods had drawbacks in their sensitivity to noise and failure to deal with ambiguous patterns, therefore, the Artificial Neural Network (ANN) was proposed to solve these problems. After that, a comparative study between ANN and traditional techniques yield to a result that the best ECG classification performance was obtained from the hybrid system of neural network and fuzzy logic (neuro-fuzzy approach) [1]. The proposed system was composed of feed-forward ANN trained with back-propagation record 94% accuracy.

Another traditional designation of ECG neuro-fuzzy system was proposed in [2]. It based on two feature extraction methods: the shift invariant method to detect the R-peaks and the transition model state to detect the ventricular premature cycle (VPC) based on the duration of R-R interval, and classification phase consists of: multi layer perceptron network (MLPN) trained with extended kalman filter (EKF) grouping the inputs to be inputted to the fuzzy inference system (FIS). As a result, it has 93.27% average classification rate, but it needs to modify the membership functions and rules of the fuzzy inference system to achieve best classification rate.

Therefore, new proposed systems composed of neuro-fuzzy designation called adaptive neuro-fuzzy inference system (ANFIS) were developed. Each of them has different feature extraction techniques such as wavelet transform [3], Largest Lyapunov exponent and spectral entropy [4], Lyapunov exponents[5], principal component analysis (PCA) and discrete cosine transform (DCT)[6].

At the same time, new approach of neuro-fuzzy classifier was proposed in Wagner et.al [7]. Three layer neural networks was used to train the fuzzy inference system, the first result of classification of this proposed system is not acceptable for use in real time applications. On the other hand, the classification accuracy will be more significant and accurate if better features are used.

Another works based on neuro-fuzzy systems but with different approach were developed. These systems consist of fuzzy c-mean clustering techniques in combination with neural networks but with different feature extraction techniques in Engin et.al [8], Ceylan et.al [9] [10], and Froese et.al[11].

1.2 Thesis Objectives

The slow rate of convergence of back-propagation algorithm is the major drawback of the above methods that were developed in literature, especially if on-line learning is required. Therefore, fuzzy logic controller (FLC) is suggested in [12] to accelerate the rate of convergence by controlling the learning parameter of back-propagation algorithm with fuzzy if-then rules. To achieve this objective, new hybrid system of artificial neural network (ANN) and fuzzy logic will be adopted in this research. That the ANN is trained with back-propagation algorithm with momentum in combination with fuzzy logic rules. The general block diagram of the proposed system in this thesis is shown in figure (1).

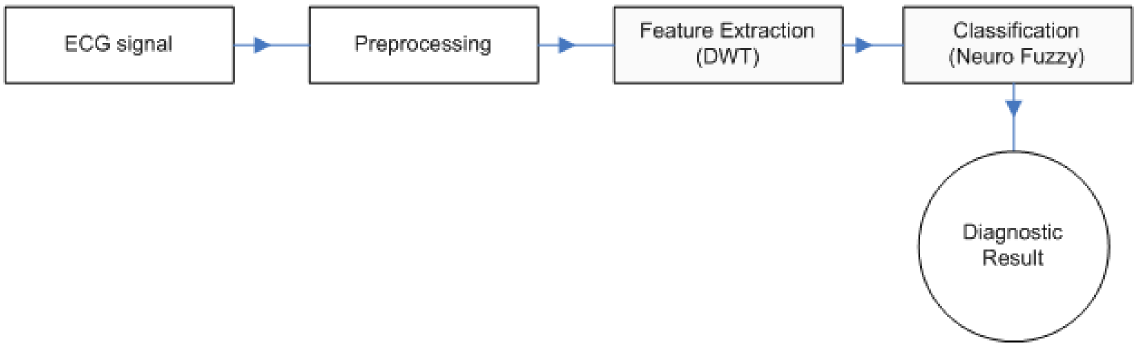


Figure 1: General block diagram of proposed system in this thesis.

On the other hand, this thesis is also focused on proving that the entropy criterion can efficiently find the optimal number of coefficients of PCA and DWT that is needed to train the ANN and satisfy high degree of classification rate.

Therefore, the main objective of this thesis is Development of computerized ECG diagnostic tool using neuro-fuzzy approach based on the following methodology:

- ECG preprocessing consists of segmentation and noise removal using UNANR model.
- ECG feature extraction using two feature extraction techniques: PCA and DWT.
- ECG Classification using neural network trained with fuzzy controlled back-propagation.

The main contribution in this thesis can be summarized as following:

- Proving that the entropy criterion efficiently finds the optimal number of DWT and PCA which achieve higher degree of accuracy.
- Proving that the fuzzy logic controller efficiently controls the learning rate of back-propagation algorithm with momentum with faster rate of convergence and high degree of accuracy.

1.3 Thesis Outline

This thesis is divided into seven chapters that organized as follows:

The first chapter is an introduction that provides summary on the ECG computerized system, literature review on ECG neuro-fuzzy systems, and aims and objective that will be achieved from this work. In the next chapter, an overview on the physiological basis of the heart, its functions, leads, and abnormalities is given. In addition to the ECG beats to be categorized. Chapter three gives an explanation of the ECG noise removal techniques is given. It provides theoretical and experimental results for using unbiased and normalized noise reduction (UNANR) model in ECG de-noising. Chapter four presents basic theory for discrete wavelet transform (DWT) and principal component analysis (PCA) as feature extraction techniques, neural network and fuzzy logic controllers (FLC). Research methodology is presented in Chapter five. It explains the experimental procedures to achieve the thesis objectives. In Chapter six, the results of experiments is discussed and compared to the literature. The last chapter presents the conclusion of the work. In addition to suggestions for future works.

2 The physiological basis of the Electrocardiogram (ECG)

Understanding the physiological basis of the ECG before any signal processing is important to review measurement conventions of the standard ECG [13]. This chapter shows the physiological concepts of normal ECG and its abnormality. It explains the heart physiology, ECG waveform, and the electrocardiogram arrhythmias.

2.1 Heart physiology

It is well known that one sort of tissue that form the heart is the muscle (myocardium) that is rhythmically contract and circulates the blood throughout the body during cycle called the cardiac cycle. This cycle is controlled by the conduction system in the heart as shown in figure (2).

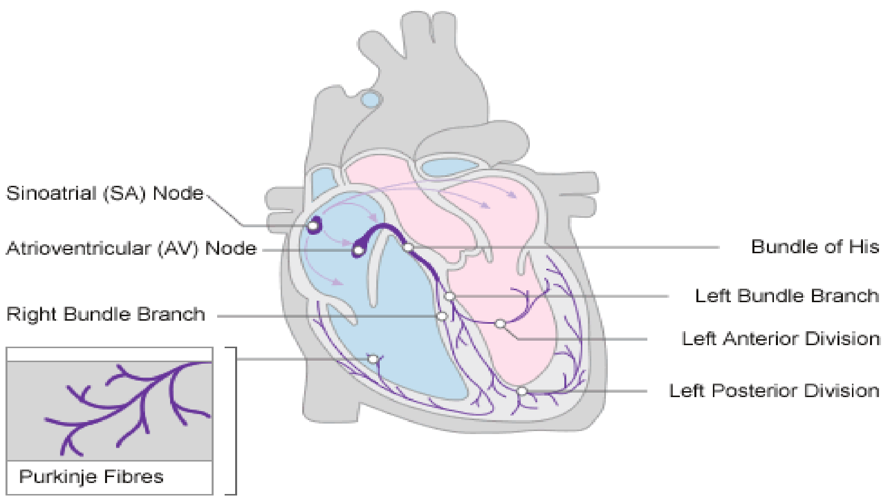


Figure 2: The conduction system of the heart [14].

The normal heartbeat begins as an Electrical impulse originates in the sinoatrial node (SAN) and rapidly spread over the atria and causes them to contract (depolarized). After 70msec, these impulses are transmitted through the Bundle of His and Purkinje network to the ventricles which they start to contract (depolarized). After the depolarization process for the atria and the ventricles, the cells in myocardium begin entering the refractory period and re-polarizing. This result in potential difference on the surface of the skin can be measured at selectively electrodes (leads) placed on the skin. The measured signal is known as (electrocardiogram) ECG signal which indicate the overall rhythm of the heart, it consists of significant waves distinguish the normal and abnormal heart beat. Abroad number of factors affects the ECG, this cause irregular in the rhythms and leads to what is known arrhythmia [15].

2.1.1 ECG Deflections and leads

The ECG waveform corresponding to a single normal heartbeat consists of significant waves or deflections [16], these waves as shown in figure (3) are:

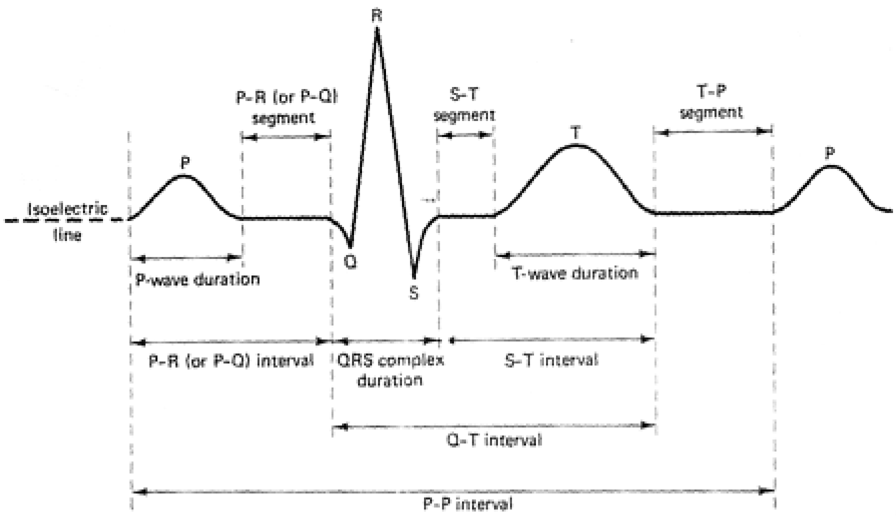


Figure 3: ECG waveform [17].

- P wave: represents the depolarization of the tow atria. Normally, this wave has amplitude ranging from 0 to 0.3mv.
- QRS complex: represents the depolarization of the two ventricles.
- T waves: represents the re-polarization of the ventricles.

In order to record the electrical impulses of the heart, electrodes are placed on the skin. The signal that goes between two electrodes is known as lead [18]. Basically, there are 12 leads divided into two groups:

1. Bipolar leads: These types of leads have one positive and one negative pole. This leads are called the limb leads since the electrodes that form these signals are placed on the limbs.
2. Unipolar leads: These types of leads have just one positive pole and the negative pole is composites from the signals from other electrodes.

2.1.2 ECG arrhythmias

When the conduction problems have repeated morphological changes or abnormalities, then the ECG degenerates into unrecognizable pattern, on other words, an arrhythmia is occurred [13]. There are three general approaches to arrhythmia analysis explained as following:

- Performing QRS detection and beat classification, labeling an arrhythmia as series of beats of a particular type.
- Analyzing a section of ECG that have several beat intervals, and calculate statistics such as variance on which the arrhythmia classification is performed.
- Constructing model for different rhythms, then compare the observed signal to this model.

The first approach is adopted in this research. Therefore, four types of beats will be used in this work, and they are:

- Right bundle branch block(RBBB)
- left bundle branch block(LBBB)
- Paced beats(P)

- Normal beats(N)

3 ECG Pre-Processing

Unfortunately, ECG signals contaminated with noises and artifacts exist within interest frequency bands and have the same ECG morphology. So, preprocessing techniques have to be used to clean the signal to get better processing results. But before going into the details of these techniques, it is very important to know the characteristic points of those noises.

3.1 Noise Sources

ECG is not noise free signals, the common sources and brief descriptions for these noises are [19]:

- Power line interference: Power line interference consists of 50 Hz or 60 Hz and its harmonics, and its amplitude is up to 50 % of the peak to peak ECG amplitude.
- Electrode contact noise: Electrode contact noise is transient noise caused by loss of contacts between the electrode and the skin.
- Patient-electrode artifacts: Patient-electrode artifacts are transient baseline changes caused by the variation in the impedance between the electrode and the skin.
- Electromyography (EMG) noise: EMG noise is caused by muscle contraction. It has 10% amplitude variation of peak to peak ECG signal with 50ms duration.
- Baseline drift: Baseline drift or wander is caused from respiration. It has 15% of peak to peak ECG amplitude variation, and it is at low frequencies from 0.15 to 0.3 Hz.
- Electrosurgical noise: Electrosurgical noise is caused by electrical surgery instruments. It is very dangerous noise sources since it destroys the ECG signal. It represented by large amplitude reach to 200% of peak to peak ECG signal at frequencies between 100 *kHz* and 1 *MHz*.

3.2 De-noising techniques

In literature, both linear and nonlinear methods are used for ECG de-noising. Basic survey on most frequently used techniques for ECG de-noising in [20] will be discussed briefly as follows:

- Moving average filter

The main idea of this filter is doing signal smoothing. It is useful apply it several times in cascade to achieve the desired amount of smoothing by choosing the appropriate window size at each time [20]. The input to this filter is the ECG signal then the output of this filter response is subtracted from the ECG signal to get estimated ECG signal components.

- IIR comb filter

Infinite impulse response (IIR) comb filter is a digital notching filtering with order *n* and bandwidth (*bw*) at *-3dB*, the difference equation and the transfer function of this filter is given by equation (1) and equation (2), respectively [21]:

$$\boldsymbol{y(m) = bx(m) - bx(m-n) + ay(m-n)} \qquad (1)$$

$$\boldsymbol{H(z) = b * \frac{1-z^{-n}}{1-az^{-n}}} \qquad (2)$$

Where *b* & *a* are the filter coefficients and *n* is the filter order.

The filter parameters are:

- ω_o: is the frequency to be removed from the signal

-q: The quality factor (Q factor) is equal to ω_o / bw.

-f_S: signal sampling frequency, which is it equal to 360 Hz for ECG signal.

-f_o : signal frequency to be removed, which is it equal to 60 Hz for ECG power line interference.

-The order of filter is f_s/f_o.

This filter rejects the frequency component which is equal 60 Hz and its harmonics from the ECG signal. The ECG signal is the input to this filter, and the output of it is ECG signal that is free from power line interference (60Hz and its harmonics).

- Adaptive filtering for noise reduction

Some criterion such as estimated mean squared error and correlation is required in adaptive filtering techniques responsible on tuning the parameters which used for the processing of signals [22]. Figure (4) is the general block diagram of filter as noise canceller.

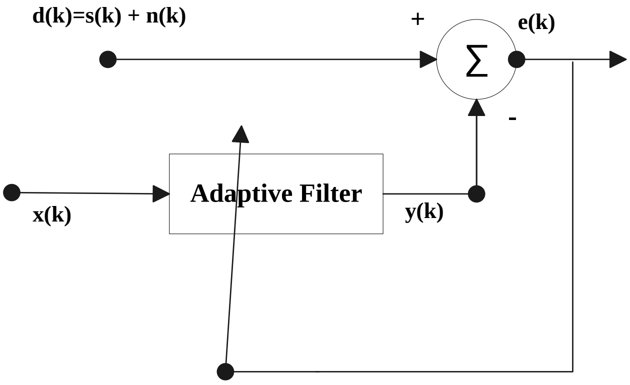


Figure 4: adaptive filter as noise canceller [22].

- *k*: is the iteration number.
- *x(k)*: is the input signal.
- *d(k)*: is the desired signal.
- *y(k)*: is the adaptive filter output.
- *e(k)*: is the error signal and it is equal to *d(k)-y(k).*
- *s(k):* is the desired signal and *n(k):* is the noise.

Once the error is obtained, an objective function is formed to be used by the adaptation algorithm in order to determine appropriate updating of the filter coefficients until the adaptive filter output signal is matching the desired signal as possible.

One of models of adaptive filtering is (Least Mean Square) LMS and the other is UNANR (Unbiased and Normalized Adaptive Noise Reduction) will be discussed as follows:

- **LMS model**

The most commonly used type of adaptive filters is the stochastic gradients approach in which the cost function is defined as mean-squared error and the steepest descent method is used to find the minimum of the error [23].

One of the most popular stochastic gradient algorithms is least mean square algorithm (LMS) which was devised by Widrow and Hoff in 1959[23].

The general block diagram of LMS interference cancellation is shown in figure (5).

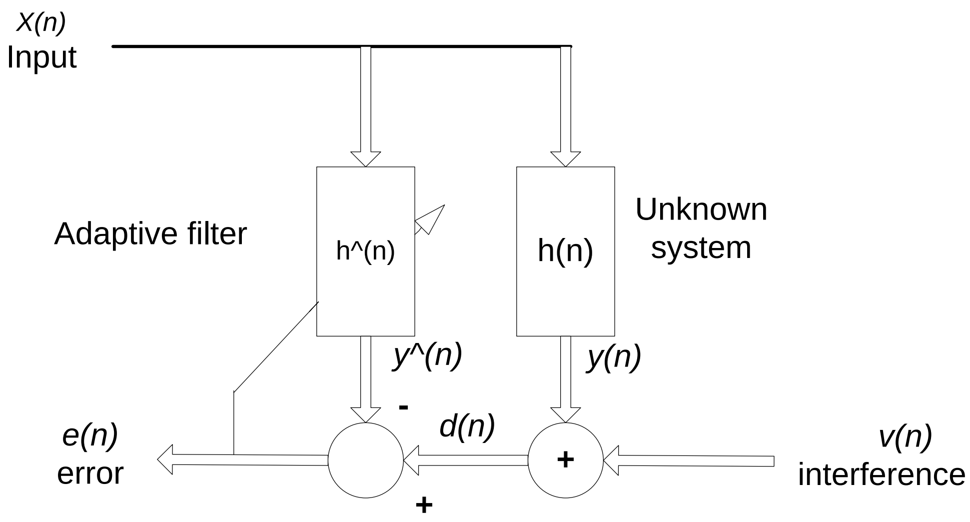


Figure 5: The general block diagram of LMS filter as noise canceller [24].

The response of the filter $y^{\wedge}(n)$ is calculated according to equation(3).

$$\boldsymbol{y}^{\wedge}(\boldsymbol{n}) = \boldsymbol{h}^{\wedge\boldsymbol{T}}(\boldsymbol{n}).\boldsymbol{x}(\boldsymbol{n}) \qquad (3)$$

Where:

-*n*: is the time.

-*p*: is the filter order.

- $x(n) = [x(n)\ x(n-1) \dots x(n-p+1)]^T$ is the ECG signal.

- $h^{\wedge}(n) = [h_0^{\wedge}(n)\ \ h_1^{\wedge}(n)\ \ ...\ \ h_{p-1}^{\wedge}(n)]^T$, $h(n) \in \mathbb{C}^p$ is the LMS filter coefficients at time n.

The LMS model modifies its coefficients that get convolved with the ECG input signal in the adaptation process, so as to estimate the noise in the ECG signal. The output of the LMS system subtracts the LMS-model response from the desired signal *d(n)* at each time instant (*n*) in order to estimate the error, *e(n).*

Now the output *y(n)* is calculated by equation(4).

$$\boldsymbol{y}(\boldsymbol{n}) = \boldsymbol{h}^{\boldsymbol{T}}(\boldsymbol{n}).\boldsymbol{x}(\boldsymbol{n}) \qquad (4)$$

Where:

- *y(n):* ECG signal component.

And the desired output *d(n)* is calculated by equation (5).

$$\boldsymbol{d}(\boldsymbol{n}) = \boldsymbol{y}(\boldsymbol{n}) + \boldsymbol{v}(\boldsymbol{n}) \qquad (5)$$

Where:

-*v(n):* is the additive white noise.

-*d(n):* is the desired ECG signal that contaminated with the additive white noise.

Now, the error *e(n)* is defined by equation(6)

$$\boldsymbol{e}(\boldsymbol{n}) = \boldsymbol{d}(\boldsymbol{n}) - \boldsymbol{y}^{\wedge}(\boldsymbol{n}) = \boldsymbol{d}(\boldsymbol{n}) - \boldsymbol{h}^{\wedge\boldsymbol{T}}(\boldsymbol{n}).\boldsymbol{x}(\boldsymbol{n}) \qquad (6)$$

The LMS coefficient adaptation process aims to find the filter weights *h(n)* which minimize the cost function *j(n)* according to steepest descent. To do this, we start with defining the cost function *j(n)* by equation (7).

$$\boldsymbol{j}(\boldsymbol{n}) = \boldsymbol{E}\{\boldsymbol{e}^{\boldsymbol{2}}(\boldsymbol{n})\} \qquad (7)$$

Where:

-*E{.}:*is the expected value.

This cost function is the mean squared error, and it is minimized by the LMS. Applying the steepest descent means to take the partial derivative with respect to the filter coefficient, this is given by equation (8).

$$\nabla_{h^{\wedge T}} j(n) = \nabla_{h^{\wedge T}} E\{e(n)e^{*}(n)\} = 2E\{\nabla_{h^{\wedge T}}(e(n)e^{*}(n))\} \quad (8)$$

$$\nabla_{h^{\wedge T}}\, e(n) = \nabla_{h^{\wedge T}}(d(n) - h^{\wedge T}(n).x(n)) = -x(n) \quad (9)$$

Substitute equation (9) into equation (8) we get the following equation:

$$\nabla\, j(n) = -2E\{x(n)e^{*}(n)\} \quad (10)$$

By substituting equation (10) into the standard steepest descent algorithm, the LMS adaptation rule is obtained by equation (11).

$$h^{\wedge}(n+1) = h^{\wedge}(n) - \frac{\mu}{2}\nabla j(n) = h^{\wedge}(n) + \mu\, E\{x(n)e^{*}(n)\} \quad (11)$$

Where:

-μ: is the learning rate.

Generally, the expectation value is not computed in equation (11), the instantaneous estimate of that expectation is used, that equation (11) has become:

$$h^{\wedge}(n+1) = h^{\wedge}(n) - \frac{\mu}{2}\nabla j(n) = h^{\wedge}(n) + \mu\, x(n)e^{*}(n) \quad (12)$$

LMS algorithm can be summarized as follows:

Step1: initialization. Initialize the LMS weights.

Step2: error calculation by equation (6).

Step 3: adaptation process by equation (12).

-UNANR model

The unbiased and normalized noise reduction (UNANR) has been proposed by Wu et. al. [25]. It is transversal, linear, finite impulse response filter. Figure (6) provides the UNANR model with order M.

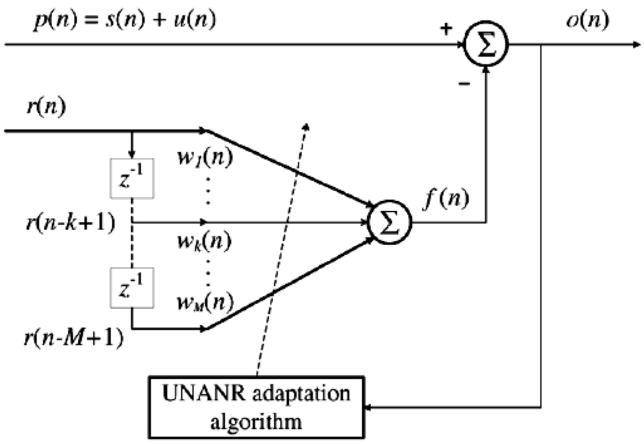


Figure 6: UNANR model with order M [25].

The response of the filter *(f (n))* is calculated according to equation (13).

$$\boldsymbol{f(n) = \sum_{m=1}^{M} w_m(n)\, r(n-m+1)} \qquad (13)$$

-*M*: UNANR-model order.

- $w_m(n)$: UNANR coefficients

- $r(n-m+1)$: The reference input noise at the present (*m=1*) and preceding *m-1*.

- $(1 < m \le M)$: input samples.

The UNANR coefficients are normalized to provide unit gain at DC such that:

$$\boldsymbol{\sum_{m=1}^{M} w_{m(n)} = 1} \qquad (14)$$

The UNANR model modifies its coefficients that get convolved with the reference input in the adaptation process, so as to estimate the noise in the ECG signal. The output of the UNANR system subtracts the UNANR-model response from the primary input at each time instant (*n*) in order to estimate the ECG signal component, $s^{\wedge}$, and this is calculated by equation(15).

$$\boldsymbol{s^{\wedge} = o(n) = p(n) - f(n)} \qquad (15)$$

Where:

- $s^{\wedge}$: ECG signal component.

- $f(n)$: The response of the UNANR model.

- $p(n)$: Primary input.

Be aware that the primary input includes the desired ECG component and the additive white noise as shown in equation (16).

$$\boldsymbol{p(n) = s(n) + u(n)} \qquad (16)$$

Where:

$s(n)$: Desired ECG component

$u(n)$: Additive white noise.

Then, both sides of equation (15) are squared, and then the formula in equation (17) is obtained.

$$\boldsymbol{s^{\wedge 2} = p^2\,(n) + f^2(n) - 2p(n)f(n)} \qquad (17)$$

After that equation (16) is substituted in equation (17), and then equation (18) and equation (19) are obtained.

$$\boldsymbol{s^{\wedge 2} = [s\,(n) + u(n)]^2 + f^2(n) - 2[s\,(n) + u(n)]f(n)} \qquad (18)$$

$$\boldsymbol{s^{\wedge 2} = s^2(n) + 2s(n)u(n) + u^2(n) + f^2(n) - 2[s\,(n) + u(n)]f(n)} \qquad (19)$$

The UNANR coefficient adaptation process aims to minimize the instantaneous error $\epsilon\ (n)$ between the estimated signal power $s^{\wedge 2}(n)$ and the desired signal power $s^2(n)$ to get equation (20).

$$\in \boldsymbol{(n) = s^{\wedge 2} - s^2(n) = 2s(n)u(n) + u^2(n) + f^2(n) - 2[s\,(n) + u(n)]f(n)} \qquad (20)$$

To achieve the above goal, the UNANR coefficients should be optimized according to the steepest-descent algorithm. The convergence in the multidimensional coefficient space follows a search path provided by the negative gradient direction given by the following equations:

$$-\nabla_{w_k} \in (n) = -\frac{\partial f^2(n)}{\partial w_k} + 2\ \frac{\partial [s(n)+u(n)]\, f(n)}{\partial w_k} \qquad (21)$$

$$-\boldsymbol{\nabla}_{\boldsymbol{w_k}} \in (\boldsymbol{n}) = -\mathbf{2}\, \boldsymbol{r}(\boldsymbol{n}-\boldsymbol{k}+\mathbf{1}) \sum_{\boldsymbol{m}=\mathbf{1}}^{\boldsymbol{M}} \boldsymbol{w_m}(\boldsymbol{n})\boldsymbol{r}(\boldsymbol{n}-\boldsymbol{m}+\mathbf{1}) - \mathbf{2}\boldsymbol{p}(\boldsymbol{n})\boldsymbol{r}(\boldsymbol{n}-\boldsymbol{k}+\mathbf{1}) \quad (22)$$

$$-\boldsymbol{\nabla}_{\boldsymbol{w_k}} \in (\boldsymbol{n}) = -\mathbf{2}\, \boldsymbol{r}(\boldsymbol{n}-\boldsymbol{k}+\mathbf{1}) * \left[\sum_{\boldsymbol{m}=\mathbf{1}}^{\boldsymbol{M}} \boldsymbol{w_m}(\boldsymbol{n})\boldsymbol{r}(\boldsymbol{n}-\boldsymbol{m}+\mathbf{1}) - \boldsymbol{p}(\boldsymbol{n})\right] \quad (23)$$

By substituting equation (14) and equation (23) into the standard steepest descent algorithm then the UNANR adaptation rule is obtained by equation (24).

$$\boldsymbol{w_k}(\boldsymbol{n}+\mathbf{1}) = \boldsymbol{w_k}(\boldsymbol{n}) - \boldsymbol{\eta}\boldsymbol{\nabla}_{\boldsymbol{w_k}} \in (\boldsymbol{n}) \qquad (24)$$

Now, by substituting equation (23) into equation (24), then the UNANR adaptation rule is obtained by equation (26).

$$\boldsymbol{w_k}(\boldsymbol{n}+\mathbf{1}) = \boldsymbol{w_k}(\boldsymbol{n}) - \mathbf{2}\boldsymbol{\eta}\boldsymbol{r}(\boldsymbol{n}-\boldsymbol{k}+\mathbf{1}) * \left[\sum_{\boldsymbol{m}=\mathbf{1}}^{\boldsymbol{M}} \boldsymbol{w_m}(\boldsymbol{n})\boldsymbol{r}(\boldsymbol{n}-\boldsymbol{m}+\mathbf{1}) - \boldsymbol{p}(\boldsymbol{n})\right] \quad (25)$$

$$\boldsymbol{w_k}(\boldsymbol{n}+\mathbf{1}) = \boldsymbol{w_k}(\boldsymbol{n}) + \mathbf{2}\boldsymbol{\eta}\boldsymbol{r}(\boldsymbol{n}-\boldsymbol{k}+\mathbf{1}) * \sum_{\boldsymbol{m}=\mathbf{1}}^{\boldsymbol{M}} \boldsymbol{w_m}(\boldsymbol{n})\,[\boldsymbol{p}(\boldsymbol{n}) - \boldsymbol{r}(\boldsymbol{n}-\boldsymbol{m}+\mathbf{1})] \quad (26)$$

Where:

-η : UNANR adaptive filter Learning rate.

UNANR algorithm can be summarized as follows [25]:

Step 1:*Initialization*. Initialize the coefficients to have uniformly distributed random values with zero-mean and unit variance then normalize them to have unit sum.

Step 2: *Activation*. At time instant *n*, activate the UNANR model with noise reference *r(n)* and estimated values of the coefficients $\hat{w}_k(n)$Then calculate the response of filter *f(n)* by equation(27).

$$\boldsymbol{f}(\boldsymbol{n}) = \sum_{\boldsymbol{m}=\mathbf{1}}^{\boldsymbol{M}} \hat{\boldsymbol{w}}_{\boldsymbol{m}}\ (\boldsymbol{n})\boldsymbol{r}(\boldsymbol{n}-\boldsymbol{m}+\mathbf{1}) \qquad (27)$$

Step 3: *Adaptation of coefficients*. Update the coefficients for the next time instant *n* + 1 according to equation (28).

$$\mathbf{w}_{\mathbf{k}}^{\wedge}(\mathbf{n}+\mathbf{1}) = \frac{w_k(n+1)}{\sum_{m=1}^{M} w_k(n+1)} \qquad (28)$$

Step 4: *Continuation*. Increment time instant *n* by one and go back to Step 2.

4 Theoretical background of techniques

This chapter explains the basic theoretical concepts of techniques that are used in ECG feature extraction, neural network and fuzzy logic controller classification.

4.1 Feature extraction

Extracting crucial information from patterns called feature extraction procedure [26]. In literature; many techniques have been proposed as feature extraction methods. The application is one of the most important factors in choosing the suitable extraction method. Two powerful feature extraction techniques: discrete wavelet transform (DWT) and principal component analysis (PCA) will be discussed below.

4.1.1 Discrete Wavelet Transform (DWT)

In practice, most of the signals are time-domain in their raw format. In many cases, mathematical transformations are required to obtain further information hidden in frequency content of the signal. The most popular transformations are Fourier transform and Wavelet Transform. The nature of processed signal determines the transformation type used [27].

1. **Fourier Transform (FT)**

Fourier transform gives the frequency spectrum of the signal. This is useful in studying the frequency content of it. The signal in this transform is translated into a function in frequency domain; the coefficients of this function represent the sine and cosine function contribution at each frequency [28]. Unfortunately, frequency domain representation is the main disadvantage of this transform if the event time occurrence is required or the nature of signal in hand is non-stationary (such as biological signals). In this case, time-frequency representation by WT is developed to overcome this problem.

2. Wavelet Transform(WT)

Wavelet transform provides time-frequency representation that the time and frequency information is provided simultaneously. The idea behind the WT is to decompose the signal into a set of basis functions called wavelets. When those wavelets are discretely sampled, then the WT is called discrete WT (DWT).

The DWT of a signal $x \in R^n$, where x in this thesis is the ECG signal, is calculated by passing it through low and high pass filters [29]. First the ECG signal is passed through low pass filter with impulse response h resulting in a convolution of two as illustrated in equation (29).

$$\boldsymbol{y}[\boldsymbol{n}] = (\boldsymbol{x} * \boldsymbol{h})[\boldsymbol{n}] = \sum_{\boldsymbol{k}=-\infty}^{\infty} \boldsymbol{x}[\boldsymbol{k}]\boldsymbol{h}[\boldsymbol{n}-\boldsymbol{k}] \tag{29}$$

Where:

-x: is the ECG signal which is $[x_0, x_1, \ldots., x_n]$.

-n: is the ECG signal sample.

-h: is the low pass filter.

The signal is also decomposed simultaneously using the high pass filter g, which is very related to the low pass filter and they are known as a quadrature mirror filter. The outputs of the low and high pass filters are called the approximation and detail coefficients respectively. Those outputs are then down sampled by two and given the following final outputs in equation (30) and equation (31).

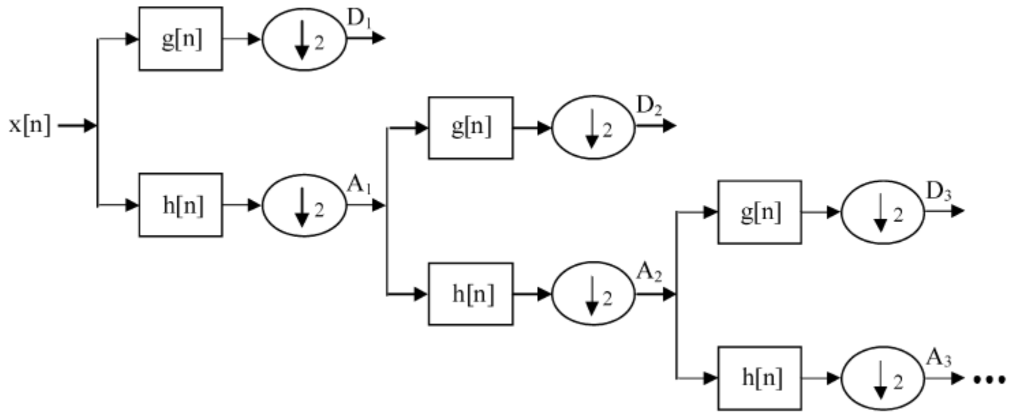
$$\boldsymbol{y}_{\boldsymbol{low}}[\boldsymbol{n}] = \sum_{\boldsymbol{k}=-\infty}^{\infty} \boldsymbol{x}[\boldsymbol{k}]\boldsymbol{g}[\boldsymbol{2n}-\boldsymbol{k}] \tag{30}$$

$$\boldsymbol{y}_{\boldsymbol{high}}[\boldsymbol{n}] = \sum_{\boldsymbol{k}=-\infty}^{\infty} \boldsymbol{x}[\boldsymbol{k}]\boldsymbol{h}[\boldsymbol{2n}+\boldsymbol{1}-\boldsymbol{k}] \tag{31}$$

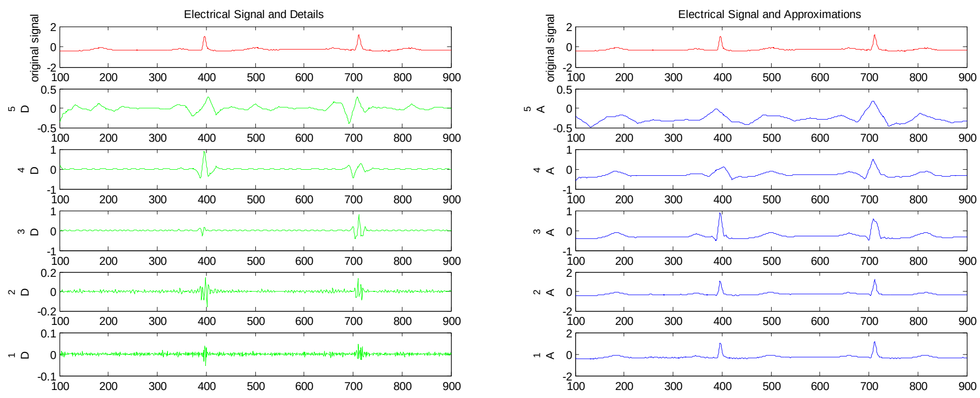
This decomposition has halved the time resolution since half of each filter output characterizes the signal. On the other hand the frequency resolution has been doubled since each filter output has half the frequency band of the input signal, which is the ECG signal.

Then, the approximation is decomposed and the process is repeated as shown in figure (7-a).

But, the decomposed signal is reached into a level beyond which it can't be decomposed, so the maximum number of decomposition levels is equal to Log_2N, where N is the number of samples in the signal [24]. Thus, the DWT detail coefficients are forming the ECG feature vector.



(a)



(b)

Figure 7: a) two-channel filter bank implementation of DWT applied to the data vector *x*. The blocks g [.] and h[.] represent the high pass filter and the low pass filter respectively. The implementation can be carried out in more resolution by successively splitting the low pass filter [30].b) ECG signal approximation and detail coefficients at levels 1, 2, 3, 4&5.

4.1.2 Principal Component Analysis (PCA)

One of the most common feature extraction algorithms is principal component analysis (PCA). It is linear technique aims to dimensionality reduction by solving Eigen value problem. Dimensionality reduction is achieved by transformation process from higher *m* dimensional vector data into lower *n* dimensional vector data, then projecting the original vector data into the new orthogonal vectors [31].

In other words, PCA is a technique for finding the eigenvectors and eigenvalues for the covariance matrix R_{xx} of $x(k) \in \mathbb{R}^m$ in equation (32) and this is equivalent to Karhunen-Loeve transformation[26].

$$\boldsymbol{R_{xx}} = \boldsymbol{E}\{\boldsymbol{x}(\boldsymbol{k})\boldsymbol{x}^{\boldsymbol{T}}(\boldsymbol{k})\} = \boldsymbol{V\Lambda V} \in \mathbb{R}^{\boldsymbol{mxm}} \quad (32)$$

Where

-$E\{.\}$: is the Expected value.

-$x(k)$: is the ECG signal segments.

- $\Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \dots, \lambda_m\}$ is the diagonal matrix containing the eigenvalues.

-$V = [v_1, v_2, \dots, v_m] \in \mathbb{R}^{mxm}$ is the corresponding unitary matrix consisting of the unit length eigenvectors referred to as principal eigenvectors.

The Karhunen-Loeve-transform calculates the linear transformation of an input vector by equation (33).

$$\boldsymbol{y_p} = \boldsymbol{V_S^T}\,\boldsymbol{x} \quad (33)$$

Where:

-$x = [x_1(k), x_2(k), \dots\dots x_m(k)]^T$ is the zero mean ECG input vectors.

-$y_p = [y_1(k), y_2(k), \dots\dots y_n(k)]^T$ is the output vector of ECG principal components (*PCs*).

-$V_S = [v_1(k), v_2(k), \ldots\ldots v_n(k)]^T \in \mathbb{R}^{mxn}$ is the set of signal subspace eigenvectors with orthogonal vector $v_i = [v_{i1}, v_{i2}, \ldots\ldots v_{in}]^T$.

-$v_i = (i = 1,2,..,n)$ are the eigenvectors of the covariance matrix, while the variance of the *PCs* y_i is the corresponding principal eigenvalues.

Therefore, the basic problem we try to solve is the standard eigenvalue problem in equation (34).

$$\boldsymbol{R_{xx} v_i = \lambda_i \, v_i} \tag{34}$$

In the standard PCA algorithm, extracting the principal components is as follows:

- Centering the observations (the observations in this research are the ECG beats) by subtracting the mean from each, such that

$$\boldsymbol{\sum_{i=1}^{m} x_i = 0} \tag{35}$$

Where: $x_i \in R^N$, $i=1 \ldots m$ is the observations.

- Calculating the diagonalizing covariance matrix C by equation (36).

$$\boldsymbol{C = \frac{1}{m} \sum_{j=1}^{m} X_j X_j^T} \tag{36}$$

- Solving the Eigen value problem in equation(9) to find the nonzero eigenvector v and positive eigenvalues $\lambda \geq 0$

$$\boldsymbol{\lambda v = C v} \tag{37}$$

- Choosing specific number of principal components according to threshold, and forming the ECG feature vector. Entropy criteria discussed in the next section is used to determine the value of this threshold.
- Forming the new observation set by projecting them to the feature vector.

4.1.3 Information theory

Information theory was developed in 1948 by Claude Shannon in his seminal work "a mathematical theory of communication" [32]. Mathematically, the information (I) of observing the symbol (a) and has the probability (p), is measured by equation (38).

$$\boldsymbol{I = \log(1|p)} \tag{38}$$

The average information is called the entropy [33]. It is a measurement of uncertainty in any random variable. The most common entropy criterion is Shannon entropy. If P is probability distribution have the set of probabilities $\{p_1, ..., p_n\}$, the entropy, H, of the distribution P is defined by equation(39).

$$\boldsymbol{H(P) = \sum_{i=1}^{n} p_i * \log(1|p_i)} \tag{39}$$

And this is equal to the expected value of the self information (the information from one symbol). The maximum entropy is obtained when there is equi-probable distribution.

To calculate this entropy for the ECG signal, the probability is needed. Therefore, a histogram for ECG signal has to be calculated to estimate this probability then the entropy is calculated as follows:

- Histogram of ECG signal $h(v)$ is calculated

  Where:

  -v: is the ECG signal.

  -$h\,(.)$: is the histogram.

- then the probability p_i is estimated according to the histogram by the following equation:

$$\boldsymbol{p_i = \frac{h(v)}{\sum_i h(v)}} \tag{40}$$

- then the entropy for ECG signal is calculated according to the following equation:

$$\boldsymbol{Entropy = -\sum_i p_i \log p_i} \tag{41}$$

4.2 Classification

Classification problems can be defined as finding the convenient way mapping the input space to output space as shown in figure(8), this convenient way can be occupied by different techniques or the combination between them, some of these techniques are:

- Neural network
- Fuzzy logic

Figure 8: A classification problem.

4.2.1 Artificial Neural Network (ANN)

Artificial neural network is the mathematical model of biological neural network [34]. It is learned by example to do specific thing.

4.2.1.1 Historical background

The evolution of the artificial neural networks has been beginning since the 1950s. The period 1950-1960 was the golden time of the artificial neural network, the first success of neuro-computing was achieved. The period 1967-1982 the research in this field was died then resurrect again in 1983. Now, the research on neural network field has become a hot subject [35].

4.2.1.2 The structure of ANN

Artificial neural network is composed of collection of nodes or units called neurons connected by directed links. The strength and the sign of each link is determined by numeric weight associated to it [36].

- **simple neuron**

The basic unit in artificial neural network is called a neuron. The general block diagram of neuron is shown in figure (9).

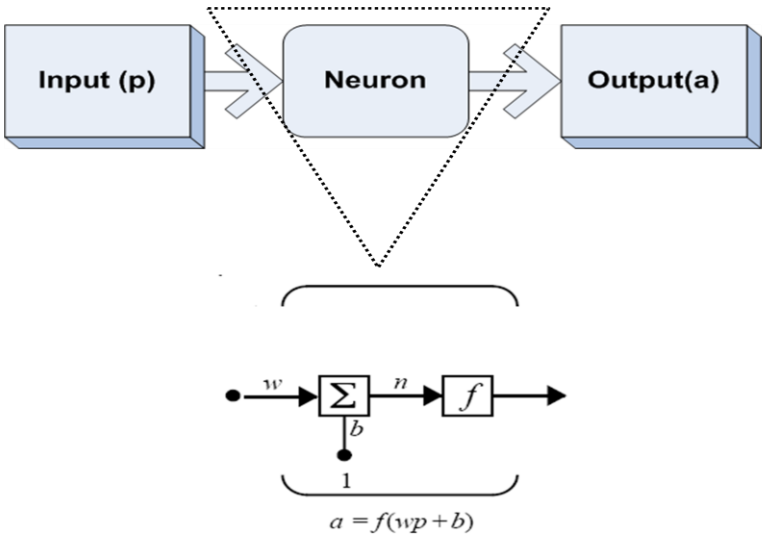


Figure 9: The general block diagram of neuron.

The input $\boldsymbol{p}$ (vector or scalar) is multiplied by the weight $\boldsymbol{w}$ (vector or scalar according to the input $\boldsymbol{p}$) then added to the bias to form the value *n* according to the formula in equation (42).

$$\boldsymbol{n = wp + b} \tag{42}$$

Then it applies an activation or transfer function *f(.)* to the value *n* to obtain the output *a* by equation(43).

$$\boldsymbol{a = f(n) = f(wp + b)} \tag{43}$$

- **Transfer functions**

There are many types of transfer functions the neural network can use. The most famous types are:

1. Hard- limit: the value of *y- axis* is *0* or *1* so this transfer function limit the output of the neuron to either *0* if the net input argument *n* is less than *0*, or *1* if the net input argument *n* is greater than or equal to zero. This transfer function is commonly used in the perceptron neural network. Figure (10) shows the general block diagram of hard-limit transfer function.

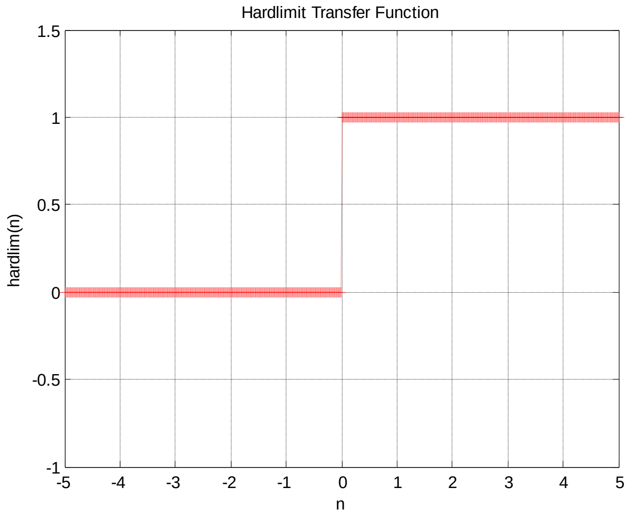


Figure 10: hard-limit transfer function.

2. Linear transfer function. The general block diagram of linear transfer function is shown in figure (11).

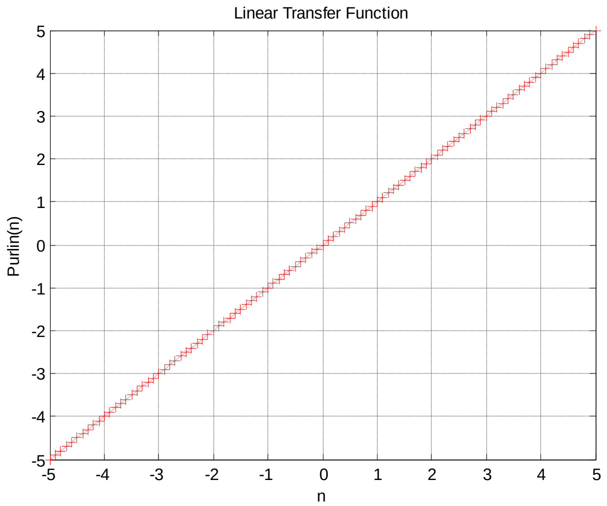


Figure 11: linear transfer function.

3. log-sigmoid transfer function:

This transfer function takes any value of input from plus to minus infinity and squashes the output value between *0* and *1*. It is used in the back propagation algorithm since it is differentiable function. The general block diagram is shown in figure (12).

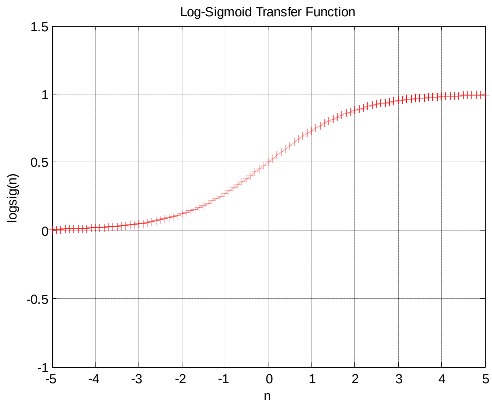


Figure 12: log-sigmoid transfer function.

- **Layer**

The collection of neurons that operates on parallel is called layer in the network. Generally, any simple neural network is consists of three layers: the first one is the input layer, the second one is the hidden, and the third one is the output layer. The general block diagram of layer is shown in figure (13).

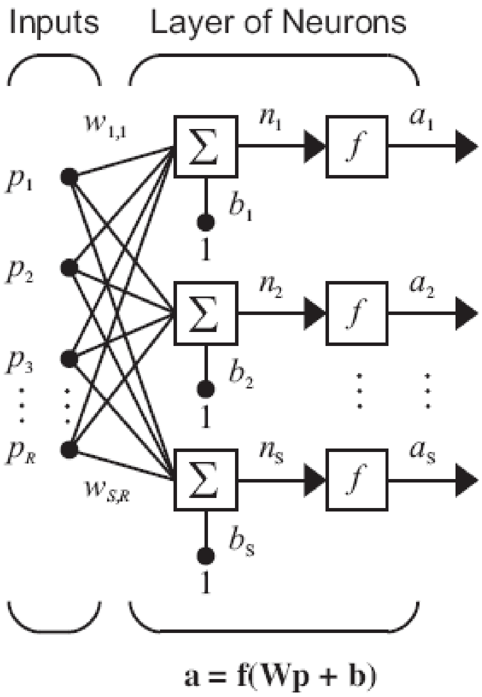


Figure 13: the general block diagram of layer [37].

4.2.1.3 Multilayer Feed- forward neural networks

One of the most popular neural networks type is multilayer feed-forward which have one or more layers of hidden units. Figure (14) shows the general block diagram of feed forward network. In addition, it has no feedback or delays connections; in this case it is not necessary to take inconsideration whether or not the inputs occur in a particular time sequence. So the output is calculated directly from the inputs through feed forward connections. It is well known that this network learn by example through different algorithms called training algorithms.

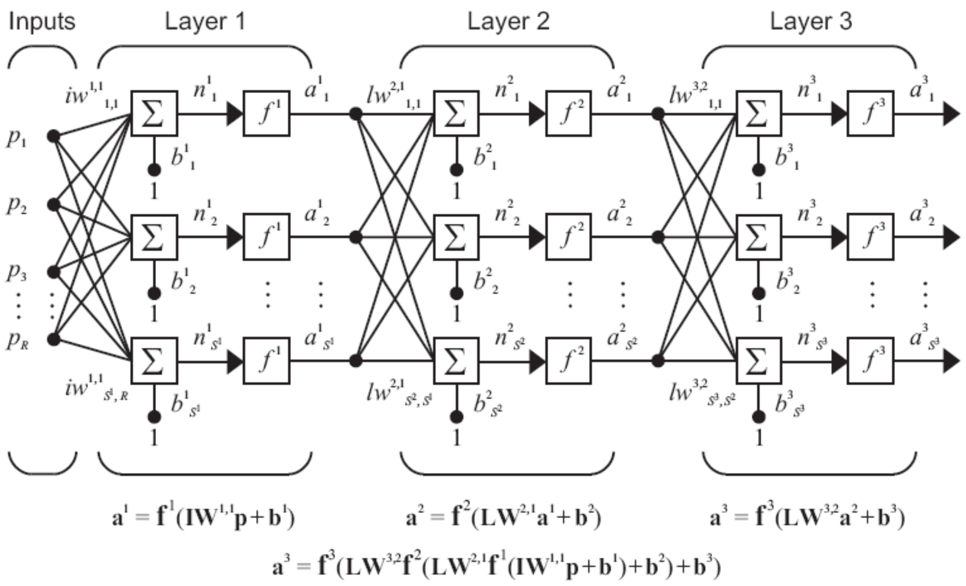


Figure 14: the general block diagram of feed forward network [37].

- **Training algorithm**

There are two main modes any neural networks have: feed-forward and learning. In Feed-forward mode, the output from the output neurons is required; this consists of presenting the input patterns into the input neurons then passing them through the network. But, in supervised learning mode, the objective is to changing the network parameters such as weights in order to make the actual output very close to the desired or target patterns. Back-propagation is very popular and most general supervised algorithm used to train the multilayer feed-forward neural network, it is an extension of LMS algorithm for linear systems [38].

The implementation of Back-propagation algorithm is as follows [39]:

Once the output from each neuron is obtained, the error signal can be calculated. At each iteration (*n*) the error signal $e_j(n)$ at the output of neuron *j* is calculated by equation(44).

$$\boldsymbol{e_j\,(n) = d_j(n) - y_j(n)} \qquad (44)$$

Where:

- $d_j(n)$: the computed output from neuron *j* at iteration *n*.

- y_j *(n)*:the desired or target at the output of neuron *j* at iteration *n*.

Minimizing the error signal e_j *(n)* is the objective of back-propagation algorithm which means that the actual output is very close to the desired response.

Then, the total error energy ξ *(n)* is calculated by equation (45).

$$\boldsymbol{\xi(n) = \frac{1}{2} \sum_{j\epsilon c} e_j^2 (n)} \qquad (45)$$

Where:

- *C*: is the number of neurons in the output layer.
- $1/2*(ej_{(n)})^2$: is the instantaneous error energy at neuron *j*.

If there is total number of patterns *N*, then the average total error energy for each pattern is calculated by equation (46).

$$\boldsymbol{\xi_{av} = \frac{1}{N} \sum_{n=1}^{N} \xi(n)} \qquad (46)$$

ξ_{av} is the cost function to be minimized. To do this minimization, the neural network free parameters (the weights and biases) are adjusted or updating by simple gradient descent procedure as in equation (47).

$$\boldsymbol{w_{ji}(n+1) = w_{ji}(n) - \eta \frac{\partial \xi}{\partial w_{ji}}} \qquad (47)$$

Where:

-w_{ji} : is the weight associated to neuron *i* and *j*.

- η: is the learning rate.

Slow rate of convergence is one of drawback of standard back propagation algorithm; the other is the local minima problem [40]. So momentum term is added to equation (47) to be [46]:

$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji}(n) + \alpha \Delta w_{ji}(n-1) \qquad (48)$$

Where:

-α: is the momentum constant, its value between [*0*-*1*].

-Δw_{ji} : is the update weight value associated to neuron *i* and *j*.

Equation (48) describes the back-propagation with momentum algorithm.

4.2.2 Fuzzy systems

Fuzzy logic is the best and fast way for mapping the input space to the output space [41]. The basis of fuzzy logic is the fuzzy set theory which was introduced by Zadeh in 1960s [42]. The elements of this set are associated with fuzzy membership value between *0* (completely false) and *1* (completely true) in contrast to the two-valued logic sets (Boolean logic) which allow just two values to its elements ; the true value(*1*) or the false (*0*) value. There is a strong relationship between fuzzy logic and fuzzy subset theory, a fuzzy set *A* in a set *X* is characterized by its membership function

$$\mu_A : X \rightarrow [0, 1] \qquad (49)$$

Where:

$\mu_A(x)$ is the degree of membership of element *x* in fuzzy set *A* for each $x \in X$.

It is clear that *A* is completely determined by the set of tuples as follows:

$$A = \{(x, \mu_A(x)) | x \in X\} \qquad (50)$$

That if $X = \{x_1, x_2, \ldots \ldots, x_n\}$ then the fuzzy set *A* is equal to:

$$A = \mu_1/x_1 + \; ..+\mu_n/x_n \qquad (51)$$

Where:

- μ_i/x_i , $i=1,\ldots,n$ is the grade of membership of x_i in *A*, and the plus sign is the union operation.

On the other hand, there are many operations can be done on fuzzy sets. If there another fuzzy set called *B* in *X*, then the following operations are defined as follows:

- intersection operation, the intersection of *A* and *B* is defined as:

$$(\boldsymbol{A} \cap \boldsymbol{B})(\boldsymbol{t}) = \mathbf{min}\{\boldsymbol{A}(\boldsymbol{t}), \boldsymbol{B}(\boldsymbol{t})\}, \forall \boldsymbol{t} \in \boldsymbol{X} \tag{52}$$

  And by logical sense, this is AND operator.

- Union operation, the union of *A* and *B* is defined as:

$$(\boldsymbol{A} \cup \boldsymbol{B})(\boldsymbol{t}) = \mathbf{max}\{\boldsymbol{A}(\boldsymbol{t}), \boldsymbol{B}(\boldsymbol{t})\}, \forall \boldsymbol{t} \in \boldsymbol{X} \tag{53}$$

  And by logical sense, this is OR operator.

Every fuzzy set is called linguistic variable, and has different values, each has specified range of membership function. The membership function of all of values is combined to form the linguistic variable membership function. Let the fuzzy set *A* be the speed variable, then this linguistic variable could have the values: low, moderate, and fast. Therefore, fuzzy rule base are performed for those fuzzy sets by *if -then* rules. Let C be another fuzzy set, then the following rules can be defined as follows:

If A is (value) AND B is (value) Then C is (value)

As a result, fuzzy logic is more adequate in describing the human reasoning so it is applied in rule base controllers [43].

4.2.2.1 Introduction to the fuzzy logic controllers

Conventional controllers are derived from control theory techniques were developed in the past decades to control dynamical systems, the behavior of which described by mathematical models of open loop process [44].

The feedback controller is needed to guarantee the desired response of the output. Figure (15-a) is the general block diagram of feedback control system, where [42]:

- y^* : is the reference input.

- y: is the output of system .

- u: is the control action(the output of the controller).

- e: is the error between y and y^*.

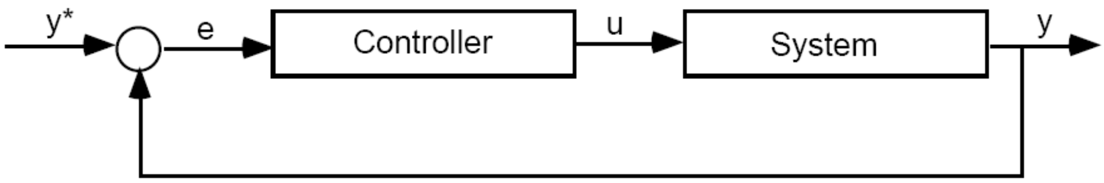
The control action u can be provided by the general form of discrete-time control law in equation (54).

$$\mathbf{u(k) = f(e(k), e(k-1), \ldots, e(k-\tau), u(k-1), \ldots\ldots, u(k-\tau)} \quad (54)$$

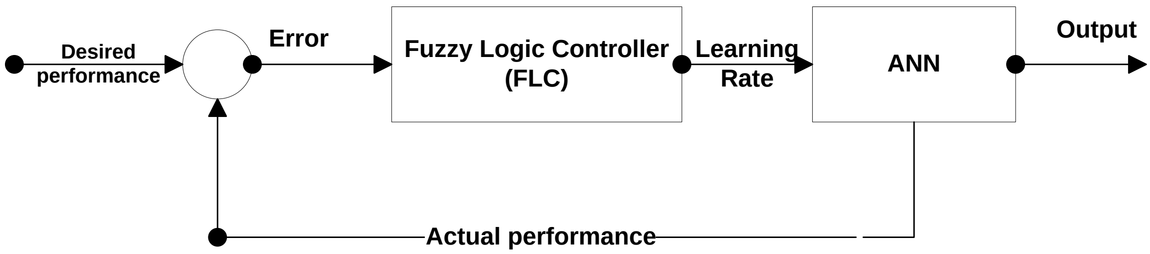
Where:

-τ: is the order of the controller.

- f: is nonlinear function.



(a)



(b)

Figure 15: (a) the general block of feedback controller. (b) The general block diagram of fuzzy logic controller (FLC) [42].

There are different control algorithms: proportional (P), Integral (I), and Derivative (D). In addition to their combination can be derived from the control law in equation (54) [42].

In the same manner, the idea of formulating the control algorithm by logical rules based on fuzzy algorithms was introduced by L.A.Zadeh [42].those controllers called fuzzy logic controllers (FLC).

Figure (15-b) shows the general block diagram of fuzzy logic controller (FLC) with ANN. After applying the ECG feature vector to the ANN, then back-propagation algorithm calculate the error, and then find the new value of the learning rate by if-then rules determined by the FLC. After that, this value is used by equation (48) to calculate the update value of the back-propagation algorithm weights according to the error and change in error.

4.2.2.2 Structure of a fuzzy logic controller

Fuzzy controllers have specific component characteristics. In figure (16) the general block diagram of fuzzy controller, diagram block by block is explained as follows [45]:

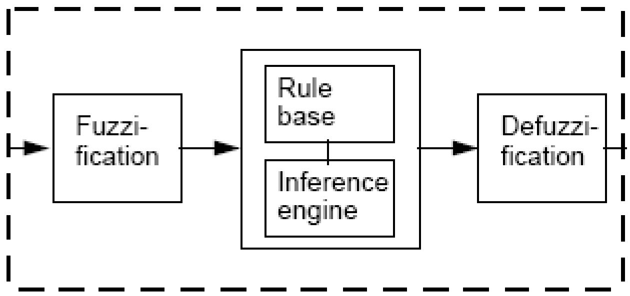


Figure 16: blocks of fuzzy controller [45].

1. Fuzzification: in this step, each input data is matched to degrees of membership by lock up in one or several membership functions.
2. Rule base: it contains many rules in the if-then format. These rules may have several linguistic variables in the condition and the consequent parts. In general, the controller needs three linguistic input variables: the error, the change in error and the accumulated error. There are three different types of rules format :

- End-user format: the rules in the controller in if-then format, as shown below:

  1) If error is Neg and change in error is Neg then output is NB.

  2) If error is Neg and change in error is Zero then output is NM.

  3) If error is Neg and change in error is Pos then output is Zero.

  4) If error is Zero and change in error is Neg then output is NM.

  5) If error is Zero and change in error is Zero then output is Zero.

  6) If error is Zero and change in error is Pos then output is PM.

  7) If error is Pos and change in error is Neg then output is Zero.

  8) If error is Pos and change in error is Zero then output is PM.

  9) If error is Pos and change in error is Pos then output is PB.

- Relational format: the same set of rule can be written as shown in figure (17). The rightmost column is the output and the leftmost columns are the inputs, on the other hand the top row is the names of variables and each other rows are a rule. This format have either of two connectives the (AND) or (OR) logical, and it is more suitable for relational database.

Error	Change in error	Output
Neg	*Pos*	*Zero*
Neg	*Zero*	*NM*
Neg	*Neg*	*NB*
Zero	*Pos*	*PM*
Zero	*Zero*	*Zero*
Zero	*Neg*	*NM*
Pos	*Pos*	*PB*
Pos	*Zero*	*PM*
Zero	*Neg*	*Zero*

Figure 17 : relational rule base format [45].

- Tabular linguistic format: the third form is as shown in figure (18). The input variables along the axes and the output variable inside the table.

		Change in error		
		Neg	Zero	Pos
Error	Neg	NB	NM	Zero
	Zero	NM	Zero	PM
	Pos	Zero	PM	PB

Figure 18: Tabular linguistic rule base format [45].

Be aware in choosing the membership functions into two recommendations:

- Start with triangular sets. All the input and output membership functions should be symmetrical triangles of the same width, and ramps for the leftmost and rightmost.
- The overlap should be at least 50%. This means that initially each variable should be member at least in two sets.

3. Inference engine: in this step aggregation process is applied to find the firing strength of each rule, and then find the resulting fuzzy set.
4. Defuzzification: the resulting fuzzy set is converted into number to act as a control signal sent the process, there are several deffuzification methods, as:

- Center of gravity (COG) : it is the most frequently used method, the crisp output is calculated by equation(55).

$$\boldsymbol{u} = \frac{\sum_i \mu(x_i\ )x_i}{\sum_i \mu(x_i\ )} \qquad (55)$$

Where:

-x_i: is the element.

-μ: is the membership value.

- Bisector of area (BOA).
- Mean of maxima (MOM).

5 Research methodology

ECG signal diagnostic system involves signal preprocessing including segmentation and noise removal, features extraction and artificial neural network (ANN) techniques. The output is used for ECG signal classification that gives indication of the patient heart condition.

The experimental procedure is as follows:

1) ECG preprocessing including segmentation and noise removal.
2) Feature extraction selection technique.
3) Neuro-fuzzy network for classification.

5.1 Experimental Tools

- **The Matlab Environment**

One of the most powerful and easy environment for technical computations is matlab. It has greatest features and capabilities over other traditional means of numerical computing. It includes a family of applications called toolboxes and it has flexibility in accepting new toolboxes. As a result, matlab is widely used in research fields, such as biomedical engineering research. The toolboxes used in this research are:

- Signal processing toolbox: this toolbox is a collection of tools built on matlab computing environment, it provides wide range of operations in signal processing field. Some of The functions provided by this toolbox is [46]:

  - Design and analysis Analog and digital filters.
  - FIR and IIR digital filter design.
  - Different transforms such as WT and FFT
  - Linear prediction

- Parametric modeling.

- Neural network toolbox: this toolbox provides tools for design, implementation, and visualization and simulation different types of neural networks. The main features of this toolbox is:

  - Supports different types of supervised and unsupervised network architecture.

  - Supports many training and learning functions.

  - Open number of design neural network nodes and layers.

- Fuzzy logic toolbox: this toolbox provides tools for design, implementation, and visualization and simulation fuzzy inference system. The main feature of this toolbox is in creating and editing different fuzzy rules and membership functions.

- **MIT/BIH database**

Since 1975, the laboratories at MIT and Boston's Beth Israel Hospital have supported the research into arrhythmia analysis. One of the major products of this effort is the MIT/BIH arrhythmia database. This database contains 48 record over 30 minutes obtained from 47 subjects categorizes as [47]:

- 23 recordings were chosen at random from a set of 4000, 24-hour ambulatory ECG recordings (numbered from 100 to 124).
- 25 recordings were selected from the same set to include variety of rare but clinically important phenomena (numbered from 200 to 234).

Each of these recordings has one or more sets of annotation files, these annotations are labels to specific locations within a recording and they indicate to the time of occurrence and types of each individual heart beat. There were many programs were developed to read these files. WFDB [48] library is one of those programs which used in this research to read the recordings annotation files. Recordings number 100, 101, 102, 107, 111, 124 and 212 which contains the four type of beats as shown in table (1) will be used in this research to form the output of ECG recognition system

Class name	Beat Symbol	Beat name
NORMAL	N	Normal beat
LBBB	L	Left bundle branch block beat
RBBB	R	Right bundle branch block beat
PACE	P	Paced beat

Table 1: The four output beats type of the ECG diagnostic system.

5.2 ECG segmentation

Extraction specific type of beats from each recording is the objective of this section. Typical ECG signal can be decomposed into three different types of groups that have the basic ECG elements as shown in figure (19) [49].

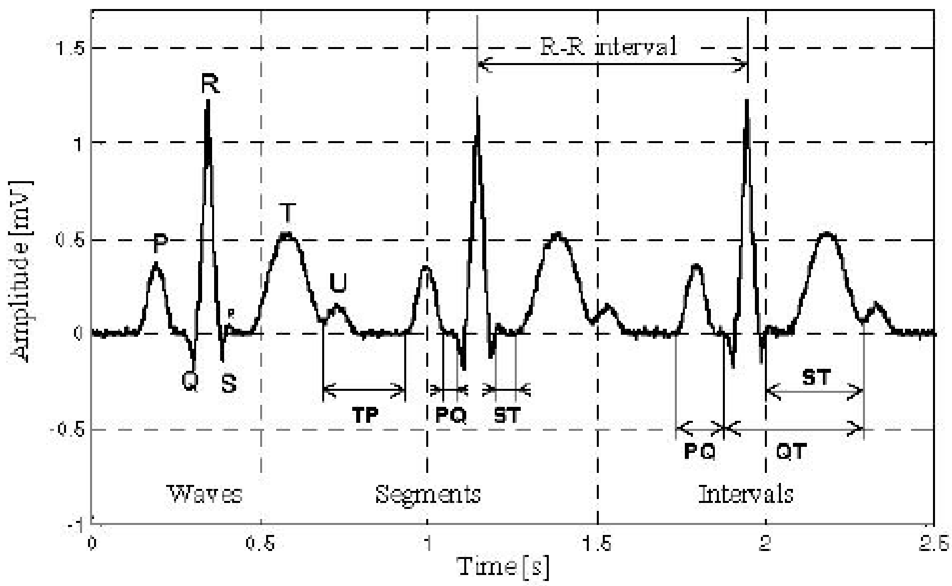


Figure 19: standard ECG signal [49].

Segmentation is based on onset of QRS detection in combination with WFDB library as follows:

1. Onset of QRS detection algorithm:

It had been proved in [50] that the (**F**irst **D**erivative only) FD1 algorithm is give the best performance in QRS detection algorithm. So FD1 algorithm will be adopted.

If *x (n)* is the signal and n is the number of signal sample, then the first derivative algorithm is in two steps:

- Find the derivative *y(n)* of the signal by equation (56).

$$\boldsymbol{y(n) = -2x(n-2) - x(n-1) + x(n+1) + 2x(n+2)} \qquad (56)$$

- Then, Finding the threshold according to equation (57).

$$\boldsymbol{slope_{threshold} = 0.7\max(data)} \qquad (57)$$

Then search for the points which exceed this threshold. So the first point that exceeds this threshold is taken as the onset of QRS complex. The FD1 steps are shown in figure (20-c) detect three onsets of QRS this is the same number of QRS in the figure (20-a).

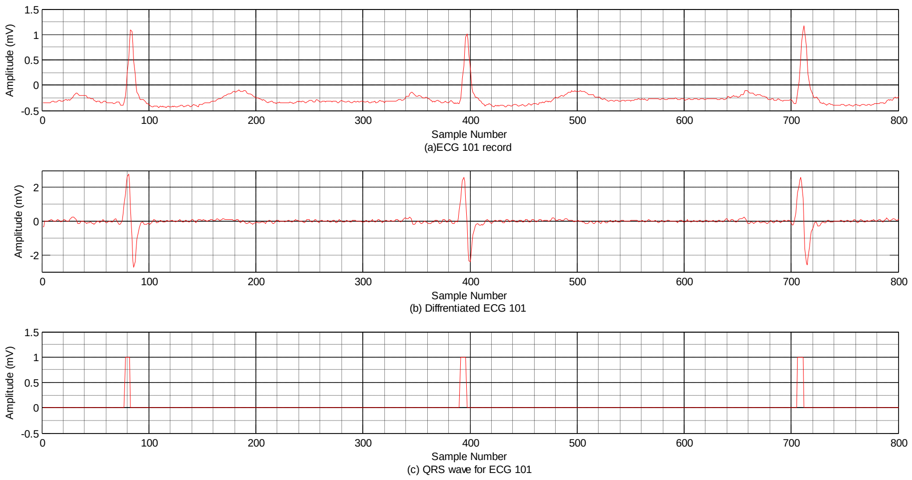


Figure 20: FD1 algorithm steps. a) ECG record 101. b) Differentiated ECG 101. c) QRS wave for ECG 101.

2. Once the FD1 algorithm is applied, then the onset of the QRS complex is known. Then we find the R peak from the WFDB Software package, for example: the onset sample position of the first QRS complex by FD1 is equal to 78 and from the

second column in figure (21) the Rpeak for the first QRS complex is equal to 83 so 78 + 5 is the position of the sample Rpeak, and so on .after that we extract 100(we trim the last 20 samples in the codes to be 180sample in each segment only) samples in either sides of R peak.

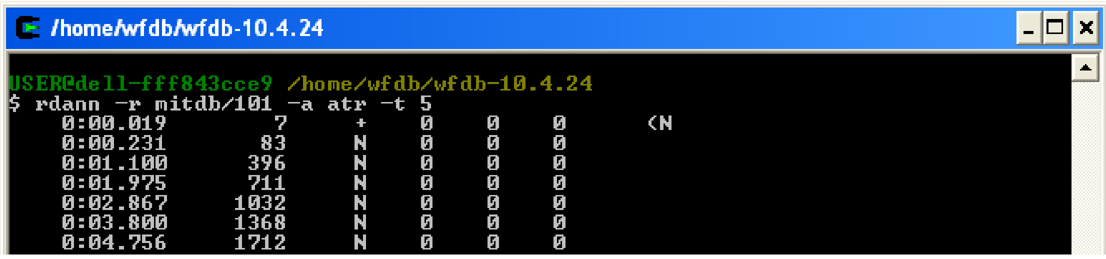


Figure 21: WFDB for the first five seconds of ECG record 101, the second column is the position of R peaks.

3. Mean and variance normalization:

Normalization process for each beat is applied to have zero mean and unit variance.

The above procedure is for the normal and the left beats but the other ones the value of the R peaks from the WFDB library are taken then the segmentation process is continued as in the normal and left beat.

5.3 ECG De-noising

Reducing the noise from ECG signals is the objective of this section. The de-noising process is divided into three main areas as mentioned in [25] as shown in figure (22):

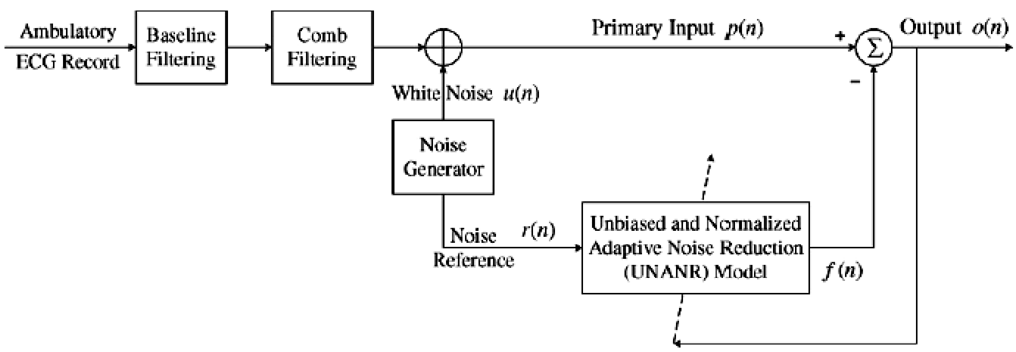


Figure 22: overview of the de-noising process as mentioned in [25].

5.3.1 Setup

Record 101 with 21600 samples from MIT/BIH arrhythmia database is used as input to two stage moving average filter for baseline drift removal. The baseline free signal is then used as input to IIR comb filter for power line interference removal purposes. Then, the output of this comb filter is contaminated with 5dB noise to formulate the primary input to the UNANR filter while the 5dB is used as reference input. The output should be noise free ECG signal.

5.3.2 Procedure and Results

There are four stages that the ECG de-noising process undergoes as following:

- **Baseline drift removal**

Two stages moving average filter is used for baseline removal with window length of 1/3 (7200sample) and 1/2 (10800 sample) the number of samples of the input signal in the first and second stage respectively. Then the output is subtracted from the ECG_101 signal. The output of two stage moving average filter is as shown in the figure (23).

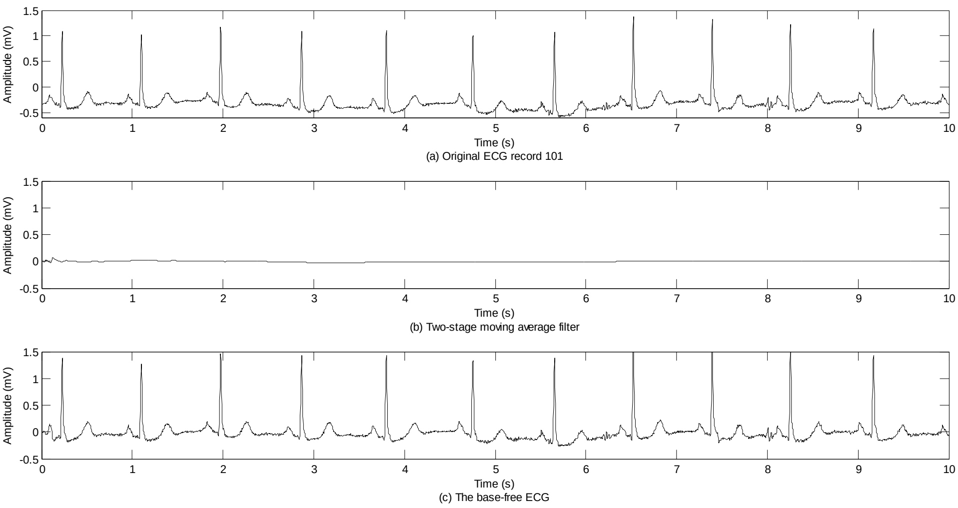


Figure 23: The output of two stage moving average filter [25].

- **Power line interference removal**

IIR comb filter is used for power line interference removal (60Hz and its harmonics). The parameters of the IIR comb filter are as follows:

- The coefficients of the IIR comb filter is specified to 4-decimated-digit word length so the transfer function of the filter is given by equation (58).

$$\mathbf{H(z) = 0.9502\ \frac{1-z^{-6}}{1+z^{-1}\ -0.9004\,z^{-6}}} \qquad (58)$$

- $\omega_o = 2\pi(60)$ is the frequency to be removed from the signal
- $bw = (f_o/(f_s/2))/q$
- $q=30$The quality factor (Q factor) is equal to $\omega o/\,bw$.
- $f_s=360Hz$ signal sampling frequency.
- $f_o = 60Hz$ signal frequency to be removed.
- $n=6$ The order of filter is f_s/f_o

Figure (24) shows the frequency response of the filter. The baseline free ECG signal is the input to this filter. The output should be power line interference free ECG signal as shown in figure (25).

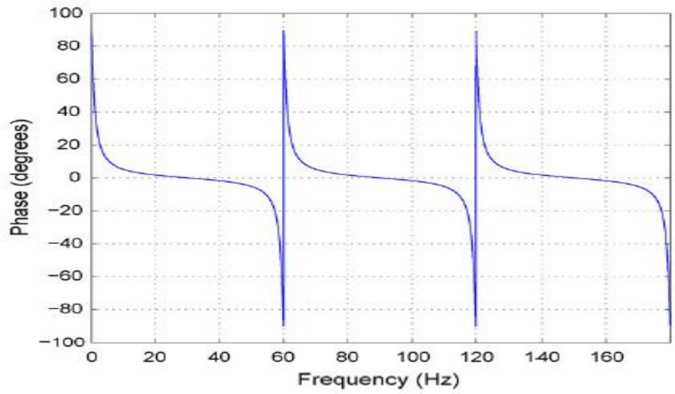


Figure 24: The IIR comb filter frequency response [25].

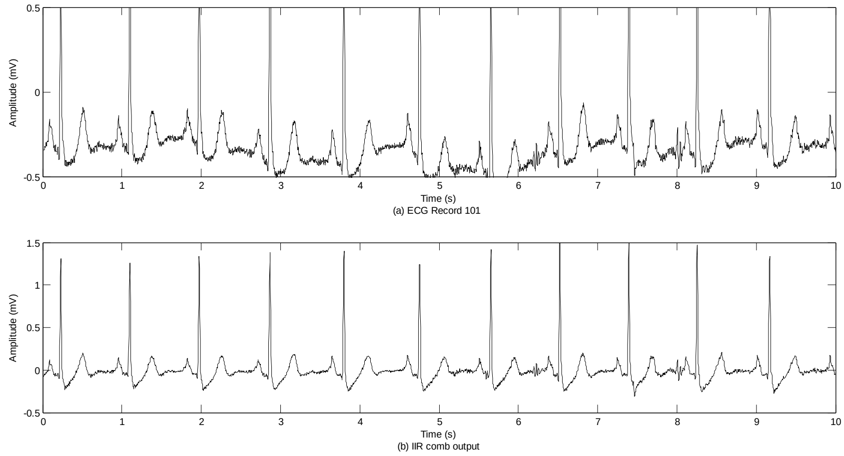


Figure 25: a) ECG record 101. b) IIR comb filter output.

- **Noise generator**

Additive white noise that is contaminated with the input signal(IIR comb free signal) and the noise reference as depicted in figure (22) is generated from the noise generator with an assigned (signal to noise ratio) SNR. In this experiment many SNR values is used and SNR improvement according to the following formula

SNRimprovement=inputSNR-10$\log_{10}$(signalpower/Noise power).

- **UNANR model**

The output of IIR comb filter is input to the UNANR filter. The output of UNANR filter and 5-dB noise is shown in figure (26).

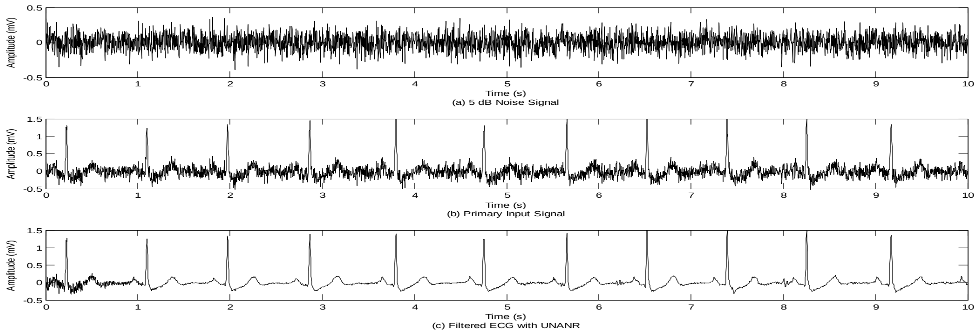


Figure 26: the output of UNANR filters: a) 5-dB noise. b) The IIR comb filter contaminated with 5-dB noise. c) The output of UNANR filters (clean ECG).

- **LMS model**

The LMS model is used instead of UNANR model in figure (22). The output of IIR comb filter is input to the LMS filter. The output of LMS filter and 5-dB noise shown in figure (27).

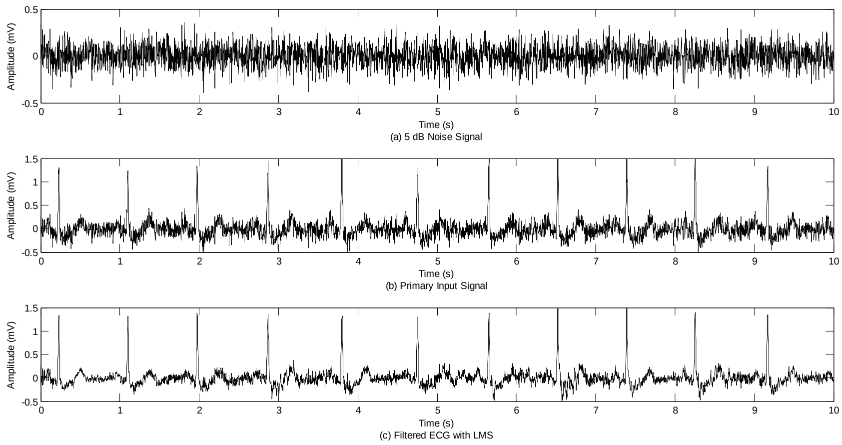


Figure 27: the output of LMS filters: a) 5-dB noise. b) The IIR comb filter contaminated with 5-dB noise. c) The output of LMS filters (clean ECG).

- **LMS and UNANR comparison**

In UNANR model the ECG signal is more noise free than LMS as shown in figure (28).

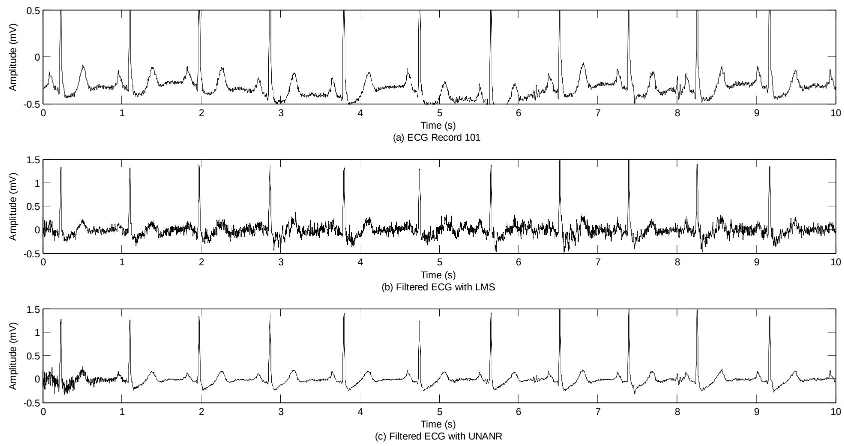


Figure 28: LMS and UNANR output: a) ECG record 101. b) Filtered ECG with LMS. c) Filtered ECG with UNANR [25].

5.4 ECG feature extraction

Three main objectives are expected from this section:

- Choosing the best feature extraction technique.
- Proving the relationship of each feature extraction technique to the entropy.
- Know the cpu- time of the back-propagation with momentum algorithm.

Two feature extraction techniques are used for feature extraction: principal component analysis (PCA) and discrete wavelet transform (DWT). The coefficients of each technique is formed the ECG feature vector. To choose the optimal number of these coefficients Shannon entropy criterion is used. The percentage accuracy of artificial neural network (ANN) is the measure of which technique will be used in neuro-fuzzy classifier. The general scheme of ANN classifier is shown in figure (29).

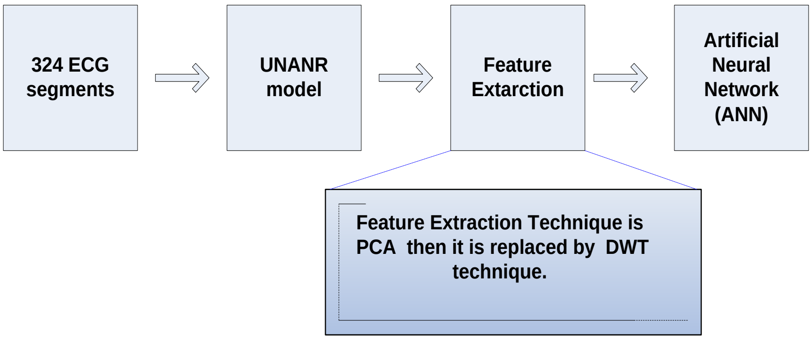


Figure 29: the general scheme of ANN classifier used in feature extraction phase.

The other blocks in figure (29) are discussed as follows:

1. Feature Extraction block:

   This block is divided into two parts: PCA algorithm is used as feature extraction technique in the first part. Therefore, PCA-ANN classifier is proposed. Then PCA technique is replaced in part two with DWT algorithm as feature extraction technique to form the DWT-ANN classifier.

<u>Pert one: PCA-ANN</u>

PCA-ANN general scheme is shown in figure (30). The output of the UNANR model is the input to the PCA algorithm to form The ECG feature vector by the principal components (PCA coefficients) to be the input to the ANN. These principal components are the eigenvectors corresponding to the eigenvalues more than specific threshold. This threshold value determines the ECG feature vector dimension, hence the number of neurons in the ANN input layer, and in addition, it is indication to the optimum number of PCA coefficients.

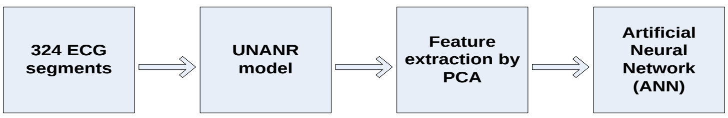


Figure 30: general scheme for PCA-ANN classifier.

The significant point in PCA algorithm is the threshold value. This value is determined by the following two ways:

- Shannon entropy : Shannon entropy is applied by the following hypothesis: if the entropy after the PCA for each signal is calculated, then the mean of the entropy is obtained and divided by the number of samples in the signal and the percentage of the result is taken, this percentage entropy is equal to the PCA threshold value that can be obtained experimentally and give the best percentage accuracy results which means optimal number of PCA coefficients.
- Experimentally (by trial and error): To prove the hypothesis mentioned above, the trial and error way is used. At each time, the PCA thresholds are changed, and the percentage accuracy of training, testing, and validation and percentage entropy is calculated. Then the threshold for best accuracy is choosing.

<u>Part two: DWT-ANN</u>

DWT-ANN general scheme is shown in figure (31). The output of the UNANR model is the input to the DWT algorithm to form The ECG feature vector by the detail coefficients to be the input to the ANN.

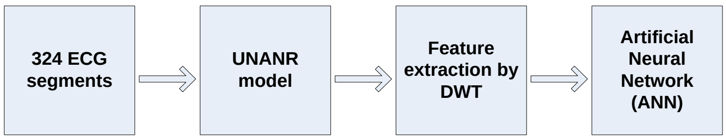


Figure 31: general scheme for DWT-ANN classifier.

Choosing the optimal number of Wavelet coefficients and the decomposition level are the most critical points in designing the DWT algorithm. The maximum number of decomposition level is calculated by the formula $\log_2 N$, thus N is equal to 180 samples for each beat. Each level has specific number of coefficients. The question here, at what level the maximum percentage accuracy by ANN is recorded? What is the enough number of coefficients is needed to train the ANN?

These questions are answered by two ways:

- By the following hypothesis:

  The decomposition level of DWT is equal to the level where the entropy of all signals is close to each other but not equal zero.

- By trial and error(Experimentally):

  To prove the hypothesis mentioned above, the trial and error way is used. At each time, The DWT decomposition levels are changed, and the percentage accuracy for training, testing and validation is calculated. Then the level of the best percentage accuracy is choosing.

2. ANN for ECG classification block: Developing a classifier involves choosing appropriate classifier model then determines the training algorithm to train and then test the ECG input signal to be classified into different categories.

   Multi-layer Feed-forward ANN is the classifier model will be used in this thesis. It will be trained by the back-propagation with momentum algorithm. This algorithm is gradient descent algorithm designed to minimize the cost function which is defined as the mean square error between the actual output of the feed-forward ANN and the desired output. For that, it requires continues differentiable transfer function. On the other hand, this algorithm has adaptive learning rate.

   The general guidelines for training process are:

   - Defining the training patterns.
   - Create the network object.
   - Train the network
   - Simulate the trained network to new inputs.

The first step in designing the feed-forward ANN is determining the categories in which the ECG signals will be classified; the most common categories were selected from the MIT/BIH database. The selected ECG signals which will be used in this research are divided into four categories as shown in table (1). To achieve clustering the ECG input patterns into four categories, a target vector formed from two Boolean values for each category or class is defined. Thus, there are two neurons in the ANN output layer. The categories or output target vector is defined in table (2).

beat name	Beat Symbol	Output vector
NORMAL	N	01
LBBB	L	10
RBBB	R	00
PACE	P	11

Table 2: the two Boolean value of the output target vector

After that, the input patterns from the segmentation process are collected in one input matrix and its corresponding target matrix is formed with combination of 1s and 0s. This is required to chick the performance of the network. The percentage accuracy is will be used in this research to chick the performance of the network. Three ways cross validation [51] technique is then used to divide these patterns randomly into three sets after feature extraction phase as following:

- Training set (50% of the input patterns): to train the neural network.

- Testing set (25% of the input patterns): to test the performance of the trained neural network.

- Validation set (25% of the input patterns): to avoid over-fitting phenomena, that the training phase is stopped if the validation error reaches specified limit.

On the other hand, this network can have more than one hidden layer with differentiable transfer functions such as log sigmoid and linear. Each neuron has its Wight that adapted after each iteration by the back-propagation with momentum training algorithm. In this research, the tangent-sigmoid and linear transfer functions are used in one hidden layer with 60 neurons and output layer with two neurons respectively, with randomly initial weights and zero biases. In addition to the following parameters:

- Learning rate= 0.05
- Momentum constant=0.01
- Performance goal=0.0001
- Maximum number of epochs=10000

5.5 Neuro-Fuzzy classifier

The objective of this section is to improve the time of convergence of the back-propagation training algorithm. The CPU-time is the performance measurement for that. The general scheme of this part is shown in figure (32). The feature extraction technique is the DWT .The last block will be designed in this section. In this block the Fuzzy logic rules are used to train the artificial neural network (ANN) this is what is called Neuro-Fuzzy classifier which is the main aim of this thesis.

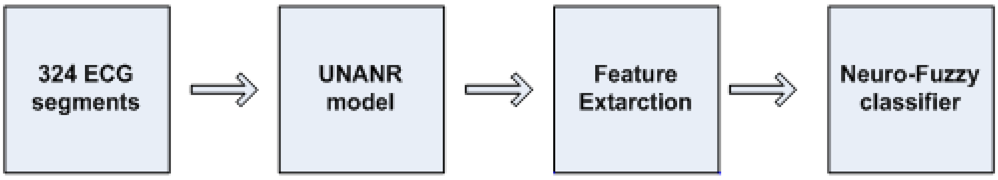


Figure 32: the general scheme of new hybrid classifier of neural network and fuzzy logic (Neuro-Fuzzy) which is used in this experiment.

As mentioned above, there are two free parameters for the back-propagation with momentum training algorithm: the learning rate and momentum constant. Till now, there are no general guidelines for choosing the values of these parameters; it is up to the user to come up with values that give fast convergence. As a result, Fuzzy control of learning rate is suggested here to achieve fast rate of convergence.

Back propagation with momentum algorithm is modified in this experiment to include fuzzy if-then rules. This is done by considering the linguistic variables Error(E) categorized as low, medium and high, and Change in error(ΔE) categorized as positive, negative and zero, As fuzzy input variables in the if condition, In addition to the Change in learning rate(Δη) which is categorized as zero, negative small and positive small, as fuzzy output variable. The membership function for each of these variables is defined in figure (33), figure (34) and figure (35). During each iteration *(n)*, the back-propagation with momentum algorithm has:

- E_n: the error at iteration n.
- ΔE_n: the change of error which is equal to $(E_n\text{-}E_{n-1})$.

- $\Delta\eta_n$: change of learning rate at iteration n that indicate the amount by which learning rate is updated.

The learning rate parameter of the back-propagation with momentum algorithm is incremented by the value $\Delta\eta_n$ which is specified by the fuzzy rules. The rules chosen for

ECG problem are shown in tabular format in table (3). From this table, there are 9 rules used in this experiment, some of them are:

```
1. If (error is Low) or (changeOfError is negative) then (deltalearningrate is positive small) (1)
2. If (error is Low) or (changeOfError is zero) then (deltalearningrate is negative small) (1)
3. If (error is Low) or (changeOfError is positive) then (deltalearningrate is zero) (1)
```

By this algorithm, the fuzzy rules are used to train the neural network and the training, testing and validation accuracy is obtained.

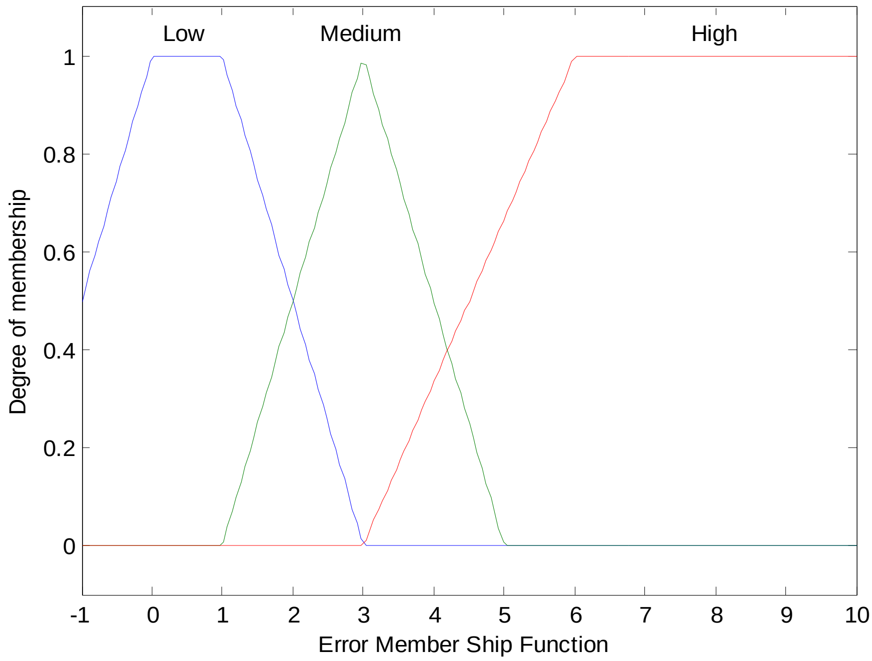


Figure 33: Error variable membership function.

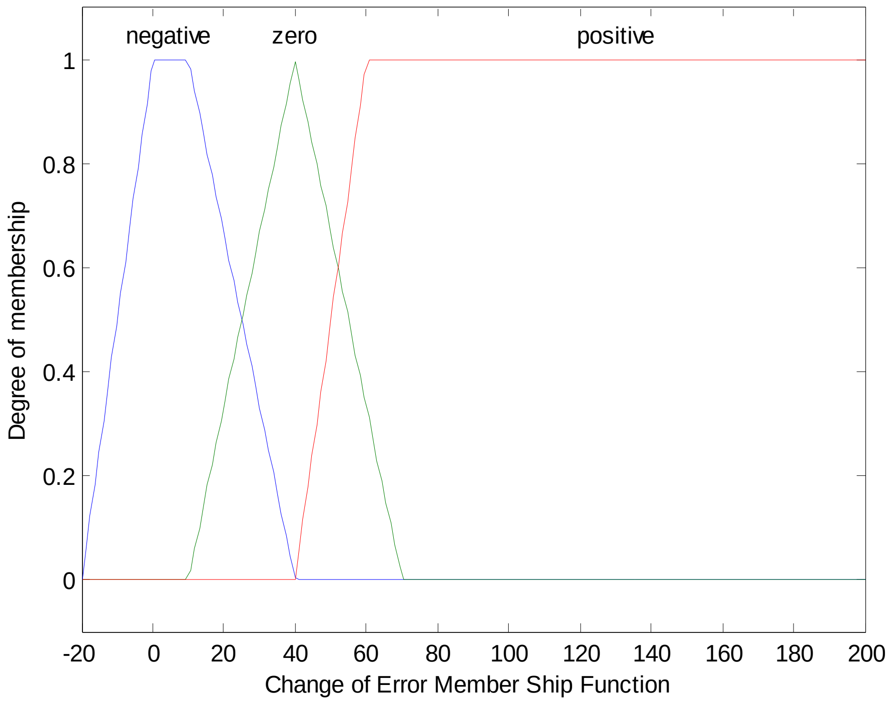


Figure 34: Change of error membership function.

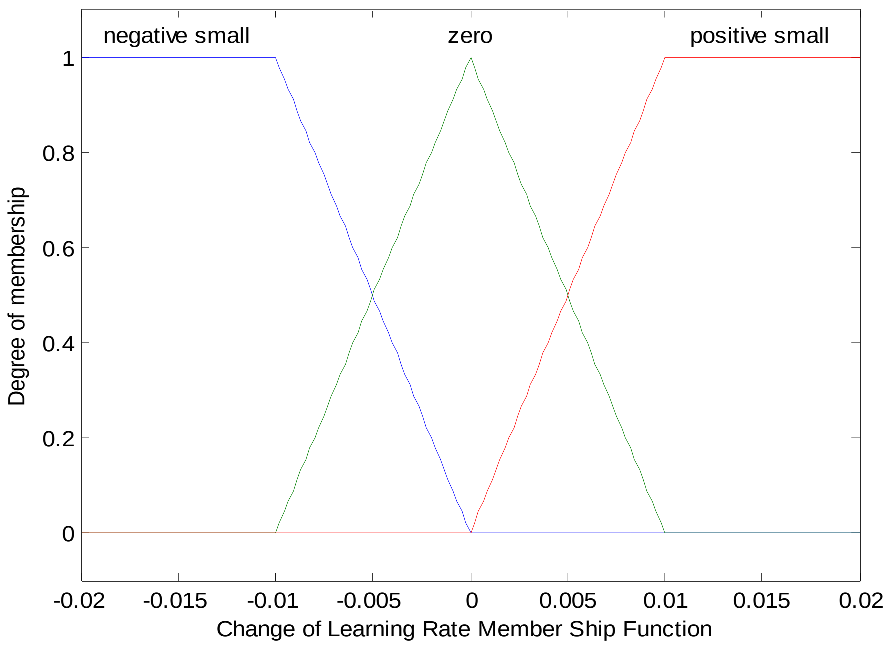


Figure 35: Change of learning rate membership function.

Error		Change of Error		
		Negative	Zero	Positive
	Low	Positive Small	Negative Small	Zero
	Medium	Positive Small	Positive Small	Negative Small
	High	Zero	Negative Small	Negative Small

Table 3: The tabular format of the ECG application Rules.

6 Results and discussion

In this chapter, the results of ECG de-noising, segmentation, feature extraction, and classification will be discussed.

6.1 Results and discussion of ECG segmentation

Since the ECG signals in the MIT/BIH database is annotated as heartbeats, so extraction specific type of beats called segments is required to form the input patterns to the ECG diagnostic system. The beats according to specific type of four beats are extracted from its R peak position with ± 100 samples as shown in figure (37). Figure (38) shows 48 segments of ECG record 101. Then normalization process is applied to each segment to have zero mean and unit variance in order to get better recognition results as shown in figure (39).

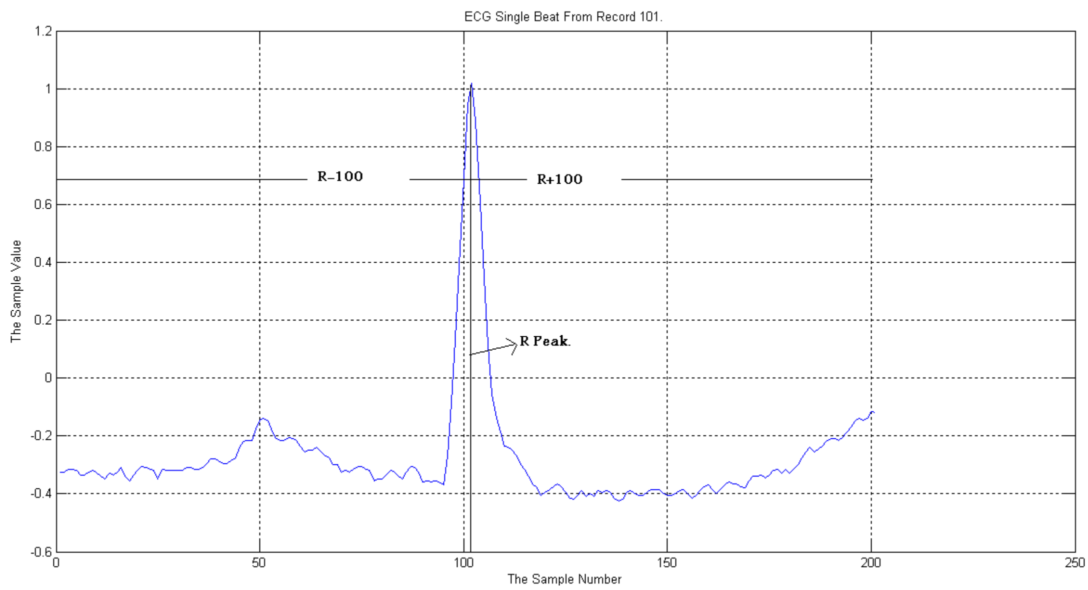


Figure 36: Extraction process of single heart beat.

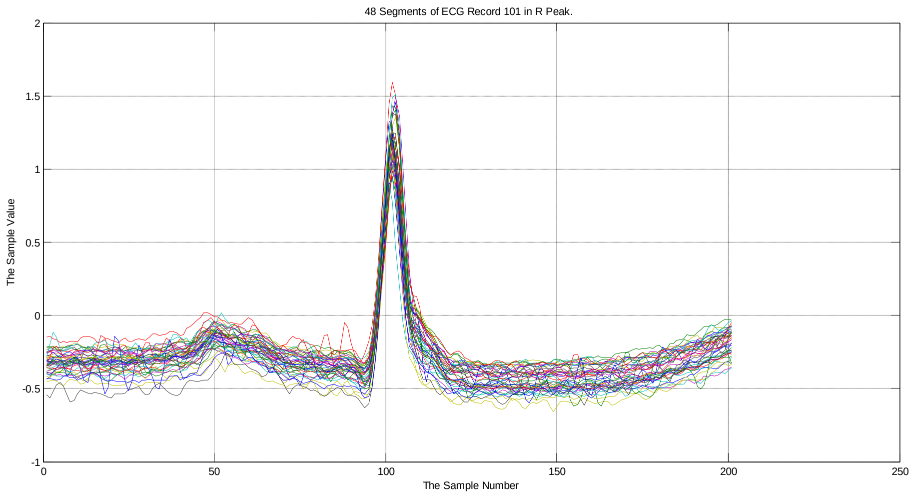


Figure 37: ECG record 101 (48) segments.

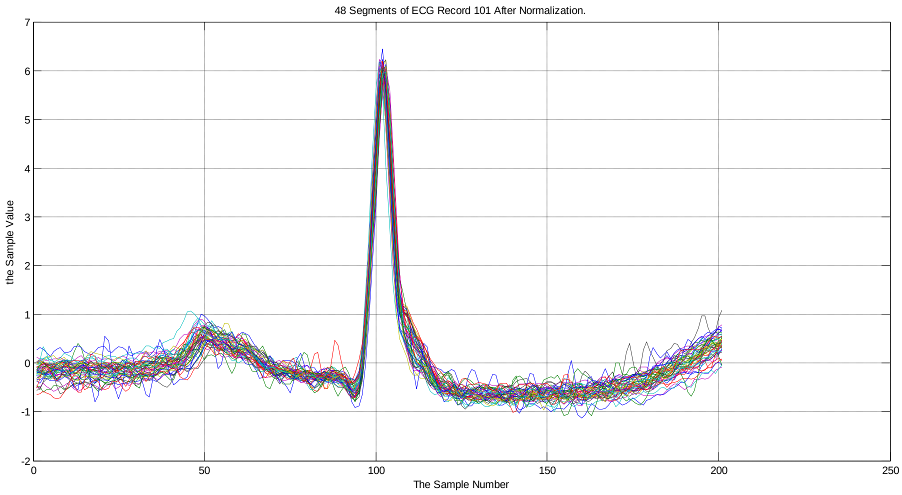


Figure 38: ECG record 101 (48) segments after normalization.

As a result of segmentation process, table (4) is formed. This table displays the number of each segment according to each beat type inside the table, in addition to the record number from where these segments are taken. Therefore, the input patterns are 324 segments for four types of beat, each has180 samples.

Record Number	Beat Type			
	RBBB(R)	NORMAL(N)	LBBB(L)	PACE(P)
100	-	24	-	-
101	-	66	-	-
102	-	-	-	28
107	-	-	-	50
111	-	-	60	-
124	34	-	-	-
212	60	-	-	-
Total # of segments	96	90	60	78

Table 4: The number of each beat type segments.

These input patterns are entered to the UNANR model discussed in chapter four for noise removal, and then it is now ready for feature extraction process.

6.2 Results and discussion of ECG de-noising

Smoothing process from figure (23-b) extract the low frequency bands (0-5Hz approximately) from the original ECG signal, this is the same frequency bands baseline drift occupied. So when the original ECG signal subtracted from this two stage moving average filter the output will be baseline free ECG signal and the signal is placed on the isoelectric of the ECG recording as shown in figure (23-c).

The power line interference is successfully eliminated from ECG signals by the IIR comb filter as shown in figure (25-b), since from figure (24) the selected comb filter parameters reject the frequencies of 60Hz and its harmonics, those frequency values are the same values the power line interference occupy. Other type of noises is eliminated by the UNANR model as shown from figure (28-c), in addition, the P-QRST is obvious and clear, but these things is not achieved from the LMS model as shown in figure (28-b). So, for accurate ECG morphology detection the UNANR model is used.

6.3 Results and discussion of ECG feature extraction

The main objective of feature extraction phase is to form ECG feature vector with optimum number of features and higher classification rate. This phase is divided into two parts as following:

- **Part one: PCA_ANN classifier**

Experimentally, when the PCA thresholds are changed, the accuracy and percentage entropy is calculated and their values as shown in table (5). From table (5) the highest percentage testing accuracy is equal to 97.87% but the validation is worse as shown in figure (39), so the percentage average (validation, testing, and training) accuracy is calculated and the highest average is equal to 95.57% $\rightarrow$ 52 coefficient $\rightarrow$PCA threshold ≥ 5.

Now, it can be noticed from the value of entropy in table(5), that all of them is greater than 5 which is equal to the PCA threshold obtained experimentally, figure (40) shows this graphically.

PCA_Threshold	Percentage Entropy (%)	PCA Coefficient numbers	Percentage Training Accuracy (%)	Percentage Testing Accuracy (%)	Percentage Validation Accuracy (%)	Percentage Average Accuracy (%)
1	5.1469	69	94.89	91.57	86.8	91.08
1.5	5.1489	64	95.96	97.87	85.77	93.2
2	5.1499	63	95.45	93.7	91.28	84.68
2.5	5.1502	59	96.25	94.74	89.03	93.34
3	5.1504	57	96.79	94.74	92.34	94.62
3.5	5.1507	54	96.79	93.7	91.25	93.9
4	5.1507	54	97.34	95.78	91.25	94.79
4.5	5.1509	52	96.51	96.82	89.08	94.14
5	5.1509	52	97.87	97.59	91.25	95.57
5.5	5.1512	50	95.98	95.78	87.99	93.25
6	5.1512	50	95.98	95.78	87.99	93.25
6.5	5.1513	49	96.79	94.74	90.16	93.89
7	5.1513	49	96.79	94.74	90.16	93.89
7.5	5.1513	49	96.79	94.74	90.16	93.89
8	5.1514	48	96.78	95.78	90.11	94.22
8.5	5.1516	47	96.5	97.87	86.8	93.72
9	5.1516	47	96.5	97.87	86.8	93.72
9.5	5.1516	47	95.99	94.74	87.99	92.91

Table 5: Numerical Results for Different PCA Thresholds.

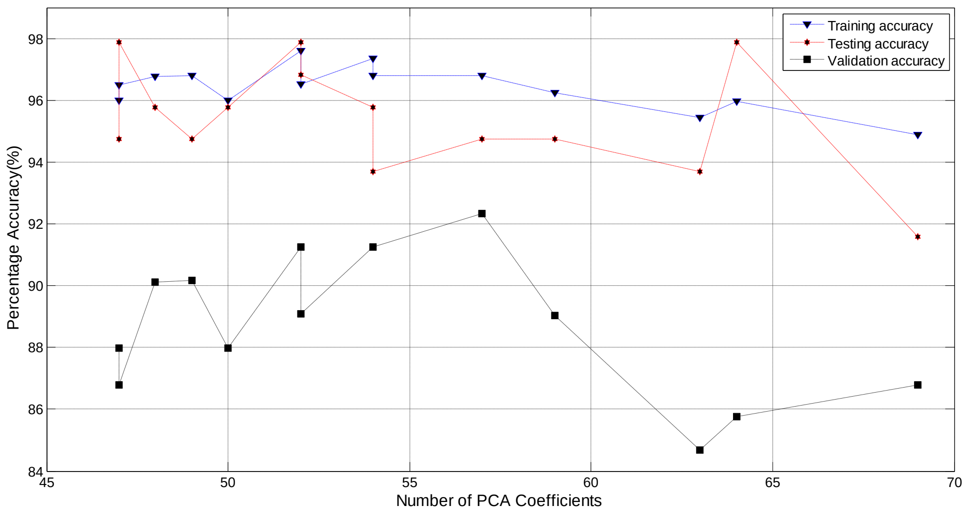


Figure 39: The Percentage Accuracy for Training, Testing, and Validation.

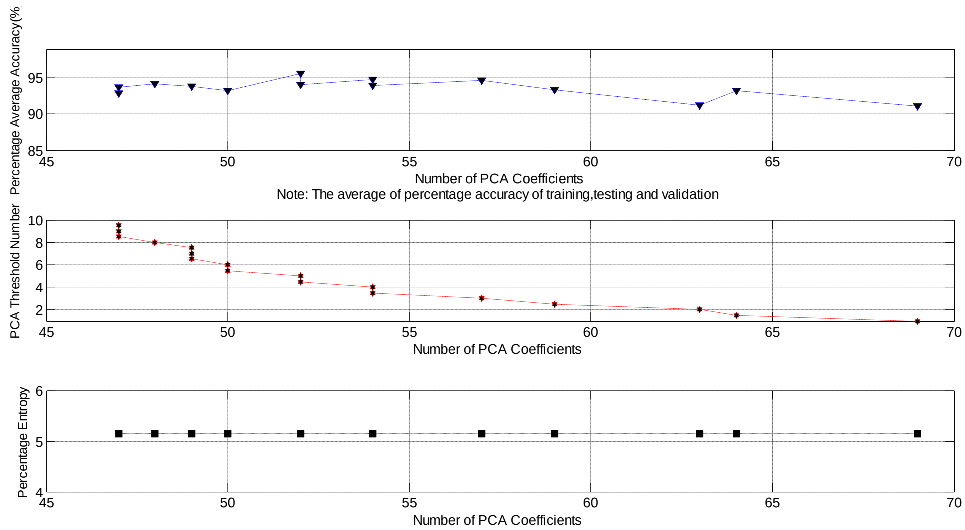


Figure 40: The Relation between the Percentage Entropy and PCA Threshold, PCA Coefficients Number, and the Percentage Accuracy.

Based on experimental results, it had been proved that the PCA threshold obtained from the experimental results is equal to the calculated percentage entropy value which is mentioned in the hypothesis. As a result, the ECG feature vector is formed from 52 PCA coefficients whose are the eigenvectors corresponding to the Eigenvalues more than 5 (PCA threshold). The architecture of PCA-ANN (52:60:2) classifier is shown in figure (41). The final accuracy results obtained by this architecture are shown in table (6).

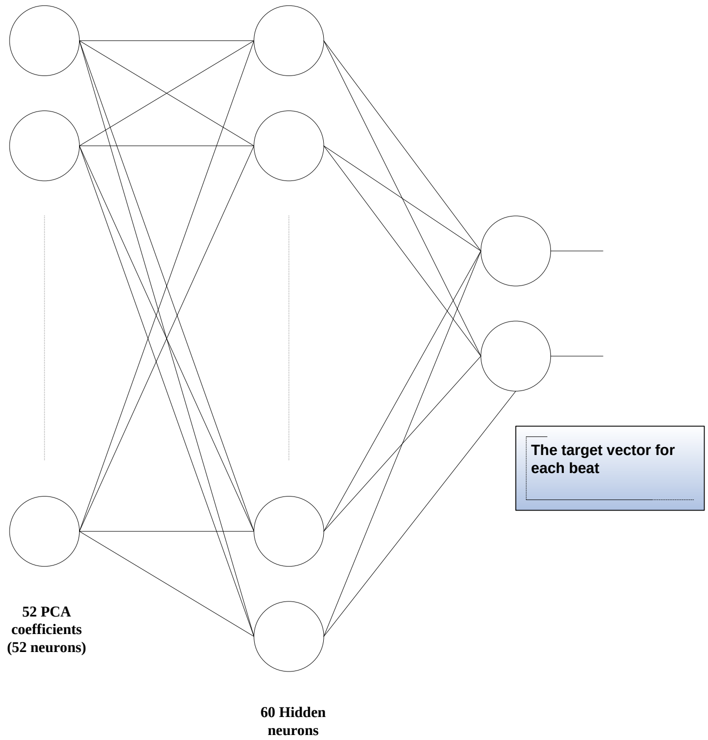


Figure 41: The architecture of PCA-ANN classifier.

Class Name	Training Accuracy	Testing Accuracy	Validation Accuracy
RBBB	90.42	91.66	69.56
NORMAL	98.88	95.65	95.45
LBBB	100	100	100
PACE	100	100	100
Average Accuracy	97.33	96.83	91.25

Table 6: The Final Percentage Accuracy Results for Different beats Type.

- **Part two: DWT-ANN**

Experimentally, when the DWT decomposition levels are changed, the percentage accuracy is calculated and their values as shown in table (7). From table (7) the highest percentage testing accuracy is equal to 98.68% → 4 coefficient →DWT decomposition level = 7.

Now, it can be noticed from the value of entropy in table(7), that all of them is close to each other at level 7 but not equal zero as shown in figure (42) which is equal to the DWT decomposition level obtained experimentally, figure (43) and figure (44) shows this graphically.

Wavelet Decomposition Level	Wavelet Coefficient numbers	Percentage Training Accuracy (%)	Percentage Testing Accuracy (%)	Percentage Validation Accuracy (%)
2	47	82.71	71.1	82.4
3	25	85.78	83.28	85.77
4	14	97.6	96.82	92.39
5	8	99.14	96.65	98.91
6	5	99.12	97.59	97.72
7	4	98.61	98.68	96.69
8	3	86.86	86.72	80.23

Table 7: Numerical Results for Different DWT Decomposition Level.

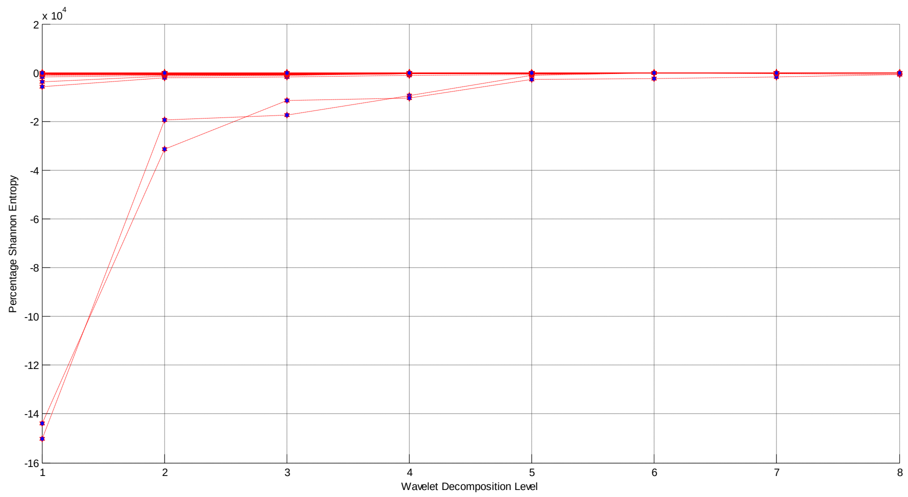


Figure 42: The value of entropy values at each level.

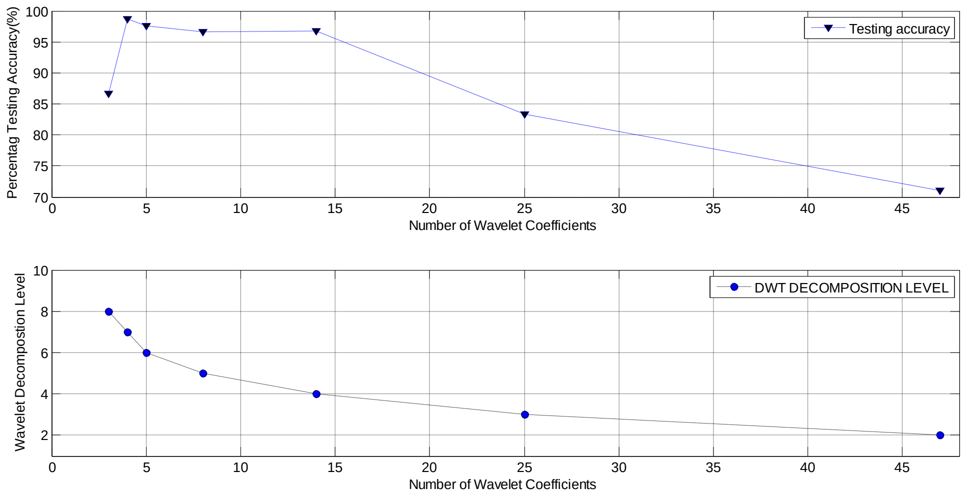


Figure 43: The Relation between the Percentage testing accuracy and DWT decomposition level.

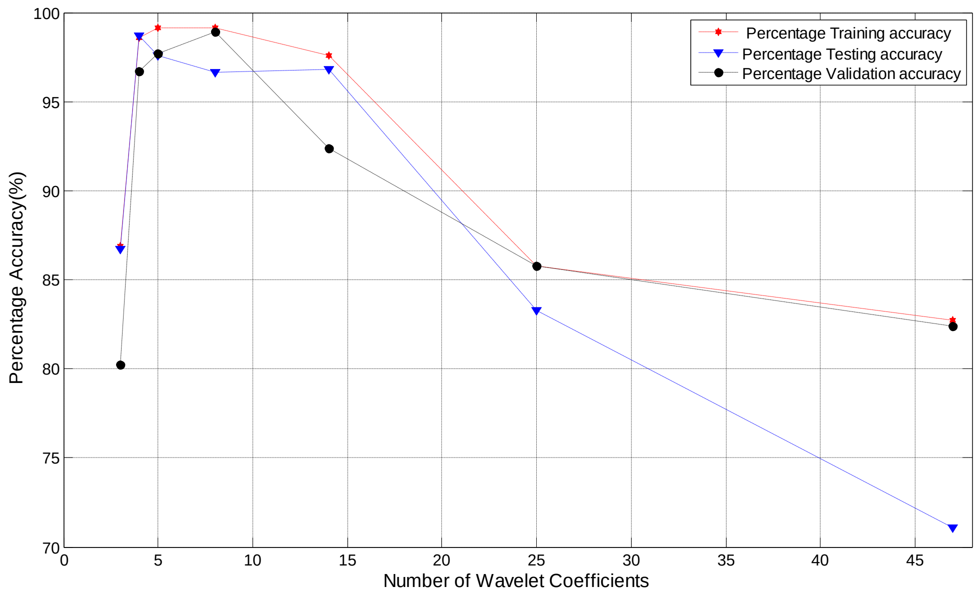


Figure 44: The Percentage Accuracy for Training, Testing, and Validation.

Based on experimental results, it had been proved that the DWT level obtained from the experimental results is equal to the level where the entropy of all signals is close to each other but not equal zero. That there is no information can obtain from the level where the entropy is equal to zero. As a result, the ECG feature vector is formed from the DWT detail coefficients from level 7 which is equal to 4 coefficients. On the other hand, in [11] the DWT till now depends on trial and error to find the optimum number of coefficients, but by this experiment, the entropy capable to determine this optimum number. The architecture of DWT-ANN (4:60:2) classifier is shown in figure (45). The final accuracy results obtained from this architecture is shown in table (8).

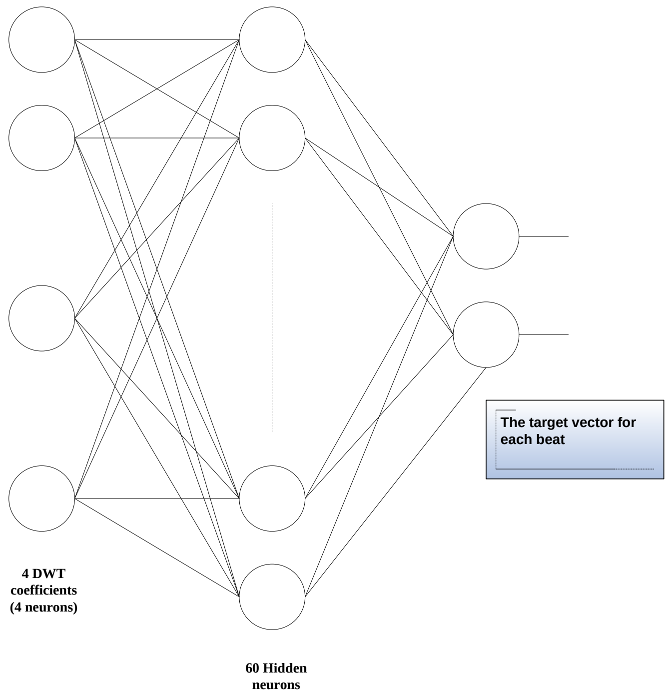


Figure 45: The architecture of DWT-ANN classifier.

Class Name	Training Accuracy	Testing Accuracy	Validation Accuracy
RBBB	95.74	100	91.30
NORMAL	100	100	95.45
LBBB	100	100	100
PACE	98.71	94.73	100
Average Accuracy	98.61	98.59	96.69

Table 8: The Final Percentage Accuracy Results for Different beats Type.

In summary, Based on the experimental results from table(6) and table(8), the DWT gives best ECG classification accuracy with 4 coefficients, due to the similarities between the mother wavelet (Db2) and the ECG signal which they convolved to each other, dB2 is shown in figure (46). Consequently DWT is considered the feature extraction in the neuoro-fuzzy classifier with the same implementation.

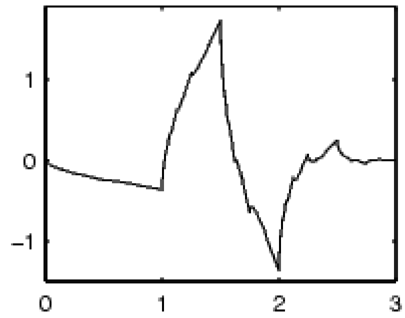


Figure 46: dB2 shape [52].

6.4 Results and discussion of neuro-fuzzy classification

DWT is used as feature extraction technique to form the ECG feature vector of 4 coefficients (4 neurons), so the architecture of this classifier (DWT-Neuro-Fuzzy) the same as DWT-ANN classifier mentioned in section (6.4). But the training algorithm is modified to have the rules mentioned in section (5.4), this rules accelerate the convergence of learning rate as shown in table (9). The performance measurement of that is the CPU time required to train the network. On the other hand the ECG classification results as shown in table (10) is better than the other two classifiers which they trained by the back-propagation with momentum algorithm.

Classifier Name	PCA-ANN	DWT-ANN	DWT-Neuro-Fuzzy
CPU-time(sec)	9.69	7.75	7.42

Table 9: performance measure of each proposed classifiers.

Class Name	Training Accuracy	Testing Accuracy	Validation Accuracy
RBBB	100	100	100
NORMAL	98.88	100	90.91
LBBB	100	100	100
PACE	98.71	94.73	100
Average Accuracy	99.4	98.68	97.72

Table 10: The Final Percentage Accuracy Results for Different beats Type recorded by (DWT-Neuro-Fuzzy).

6.5 Comparison study

Table (11) shows a comparison study with the more related work referenced in [11] in which the ECG beats were taken from the MIT/BIH database. Generally, the average testing accuracy recorded by the proposed classifiers in this research is higher than those in [11] which they have approximately very similar implementation parameters that the ANN is trained by the coefficients of DWT or PCA. On the other hand, the number of coefficients that is used to train the ANN for the proposed systems in this thesis is less than those systems which proposed in [11], which means less system complexity. But, the proposed DWT-Neuro-Fuzzy classifier is the best among those in [11] and the other two proposed systems in this work, that it has highest percentage accuracy and less system complexity since it has feature vector composed from 4 coefficients, thus, ANN has 4 input neurons in the input layer, and this is can be observed in table (11).

Classifier name	Training algorithm	Feature extraction technique	Number of beats	Number of coefficients	Average testing accuracy (%)
Back-prop.ANN[11]	Back-propagation	DWT	2	11	96.5
GA-ANN[11]	Genetic algorithm(GA)	DWT	2	8	95.9
Proposed PCA-ANN	Back-propagation	PCA	4	52	96.83
Proposed DWT-ANN	Back-propagation	DWT	4	4	98.59
Proposed DWT-Neuro-Fuzzy	Fuzzy controlled back-propagation	DWT	4	4	98.68

Table 11: A comparison study between the proposed systems in this research and the most related classifiers in [11].

7 Conclusion and future work

Hybrid system of artificial neural network (ANN) and fuzzy logic (Neuro-Fuzzy system) has been developed in this thesis, in which four types of heart beats have been categorized. The most important point in ECG diagnostic is to extract efficient features and to find suitable structure of algorithm and techniques of the classifier. Therefore, two feature extraction techniques: PCA & DWT were tested in this work according to the highest percentage accuracy of two proposed classifiers PCA-ANN and DWT-ANN, in which the ECG feature vector is formed from the coefficients of these techniques. Thus, the optimal number of those coefficients was chosen effectively according to the entropy criterion approach as explained in chapter five instead of trial and error approach used in literature. On the other hand, ANN trained with Fuzzy-back-propagation algorithm with momentum is used as final classifier, that Fuzzy logic rules play an important role in accelerating the rate of convergence of the traditional back-propagation algorithm with momentum by controlling the learning rate.

Conclusions based on experimental results for different experiments can be presented as follows:

- In preprocessing phase, UNANR model provided best signal to noise ratio than LMS model. Therefore, the UNANR model was used in this work for ECG de-noising. In addition, the moving average and the comb IIR filters are successfully remove the baseline wandering and power line interference noises, respectively.

- In feature extraction phase, two feature extraction techniques were tested to form the ECG feature vector: principal component analysis (PCA) and discrete wavelet transform (DWT), which composed of the coefficients of these techniques. The most important point here, is how efficiently find the optimal number of those coefficients that is sufficiently to train the ANN. In literature, trial & error is the most successfully approach used to determine the optimal number of PCA and DWT coefficients. In this work, Shannon entropy criterion approach discussed in chapter five was successfully used to determine the optimal number of coefficients which is equal 4 for DWT and 52 for PCA instead of traditional trial and error approach used in literature, and this according to the following two hypothesis:

1. If the entropy after the PCA for each signal is calculated, then the mean of the entropy is obtained and divided by the number of samples in the signal and the percentage of the result is taken, this percentage entropy is equal to the PCA threshold value that can be obtained experimentally and give the best percentage accuracy results which means optimal number of PCA coefficients.
2. The decomposition level of DWT is equal to the level where the entropy of all signals is close to each other but not equal zero.

- The best recognition accuracy was obtained from the proposed DWT-ANN classifier compared to the proposed PCA-ANN. Therefore; ECG feature vector with 4 DWT detail coefficients trained the ANN successfully with higher average accuracy and less CPU time compared with PCA technique. Consequently, DWT was adopted as feature extraction technique for the neuro-fuzzy classifier for its effectiveness and system complexity reduction.
- In classification phase, feed-forward ANN trained with fuzzy back-propagation algorithm was used as final classifier (DWT-Neuro-Fuzzy classifier). Therefore, Fuzzy logic controller has the capability to speed up the convergence of the back-propagation algorithm by controlling the learning rate with fuzzy if-then rules.

- DWT-Neuro-Fuzzy classifier recorded the highest average recognition rate and less CPU-time among other proposed classifiers: PCA-ANN and DWT-ANN.
- DWT-Neuro-Fuzzy classifier recorded the highest average testing accuracy compared to those in literature.

7.1 Future Work

Due to the significance of Computerized ECG diagnostic system in clinical practice today, many researches in this field are developed. Therefore, many ideas can be proposed in future, some of them can be summarizes as follows:

- In preprocessing phase, other techniques could be tested to improve the de-noising process; one suggestion of technique is the PCA in combination with DWT.
- In the feature extraction phase, other techniques could be tested such as the Kernel-PCA which is nonlinear technique.
- In the classification phase, UNANR can be used to train the ANN.

8 Bibliography

[1] R.Silipo, G.Bortolan," *Neural and Traditional Techniques in Diagnostic ECG Classification,"* University of Florence, Italy.2000.

[2] Y.P.Meau, F.Ibrahim, S.A.L.Narainasamy, R.Omar, *" Intellegent Classification of Electrocardiogram (ECG) Signal Using Extended Kalman Filter (EKF) Based Neuro-Fuzzy System,"* Elsiveir, Science direct, 2006.

[3] I.Guler, E.D.Ubeyli," *Application of Adaptive NeuroFfuzzy Inference System for Detection of Electrocardigraphic Changes in Patients with Partial Epilepsy Using Feature xtraction,"* Elsiveir, Science direct, 2004.

[4] P.K.Sadasivan, et al, *" Cardiac State Diagnosis Using Adaptive Neuro-Fuzzy Technique,"* Elsiveir, Science direct, 2006.

[5] E.D.Ubeyli," *Adaptive Neuro-Fuzzy Inference System for Classification of ECG Signals Using Lyapunov Exponents,"* Elsevier Ireland Ltd.2008

[6] V-E.Neagoe, I-F.Iatan, S.Grunwald ,*"A Neuro-Fuzzy Approach to Classification of ECG Signals for Ischemic Heart Disease Diagnoses,"* National Center for Biotechnology Information , U.S. National Library of Medicine, 2003. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1480049 /

[7] G.V.Wagner, S.Hey, W.Stork, K.-D.M-Glasser, *" Classification of ECG Signals Using A Neuro-Fuzzy Approach,"* , Germany.2001

[8] M.Engin, *" ECG Beat Classification Using Neuro-Fuzzy Network,"* Elsiveir, Science direct, 2004.

[9] R.Ceylan, Y.Ozbay, B.Karlik, *"A Fuzzy Clustering Neural Network Architecture for Classification of ECG Arrythmias,"* Elsiveir, Science direct, 2005.

[10] R.Ceylan, Y.Ozbay, B.Karlik, *" A Novel Approach for Classification of ECG Arrhythmias: Type-2 fuzzy clustering neural network,"* Elsiveir, Science direct, 2008.

[11] T.Froese, S.Hadjiloucas, R.K.H.Galvao, V.M.Becerra, C.J.Coelho," *Comparison of Extra Systolic ECG Signal Classifiers Using Discrete Wavelet Transforms,"* ELSEVIER,Science direct,2006.

[12]P.Arabshahi, J.J.Choi, R.J.M arks II, T.P.Candell,*" Fuzzy Control of BackPropagation*," Boeing computer services, Seattle, IEEE, 1992.

[13] G.D.Clifford, F.Azuaje, and R.G.McSharry, "*Advanced Methods and Tools for ECG Data Analysis*", Boston/London: Artech house,INC., 2006.

[14] The university of Nottingham, *"cardiology teaching package*," cardiac conduction system, Division of Nursing, University of Nottingham. Available: http://www.nottingham.ac.uk/nursing/practice/resources/cardiology/function/conduction.php . 2010

[15] Royal scientific society, Japan international cooperation agency, "*Basic cardiovascular anatomy and ohysiology*". 1997.

[16]R.Richter," Basic Electrocardiography,"2005. [online]. Available: http://www.sh.lsuhsc.edu/fammed/OutpatientManual/BasicECG.htm, [Accessed: Feb.6, 2010].

[17] Yale university, "Electrocardiography", BME 355 lab, 2007. Available. http://www.google.com/imgres?imgurl=http://noodle.med.yale.edu/~staib/bme355/ecg/ecg.gif&imgrefurl=http://noodle.med.yale.edu/~staib/bme355/ecg/prep.htm&h=249&w=383&sz=7&tbnid=vHWCtknhtjvAGM:&tbnh=80&tbnw=123&prev=/images%3Fq%3Dnormal%2BECG%2BDIAGRAM&hl=ar&usg=__CMOgaNSqg6SXbCLGmjyodvCt9f4=&ei=ohE-SonROIeE-QagyOGtBQ&sa=X&oi=image_result&resnum=1&ct=image

. 2009.

[18] Wikipedia, “Electrocardiography," free encyclopedia, 2009. Available. http://en.wikipedia.org/wiki/Electrocardiogram. 2009.

[19] G.M.Friesen, T.C.Jannett, M.A.Jadallah, S.L.Yates, S.R.Quint, and H.R.Nagle, "*A Comparison of the Noise Sensitivity of Nine QRS Detection Algorithms,"* IEEE Trans. Biomed. Eng. Vol.37, pp. 85 – 98, Jan.1990.

[20] D.Novak, *"Processing of ECG Signal Using Wavelet,"* M.S.thesis, Czech Technical University, Prague. 2003

[21] Filter Design toolbox, mathworks Inc, http:// www.mathworks.com

[22] Vijaykumar V R, and Kanagsapabathy P. *"Modified Adaptive Filtering Algorithm for Noise Cancellation in Speech Signals"*. 2007.

[23] Haykin S, and Widrow B." *Least-Mean-Square Adaptive Filters"*. Wiley, 2003.

[24] Wikipedia, the free encyclopedia, *"* Least mean squares filter,*"* Least mean squares filter, 2010.Available: http://en.wikipedia.org/wiki/Least_mean_squares_filter. [May,2010].

[25] Wu YF, et al. *" Filtering electrocardiographic signals using an unbiased and normalized adaptive noise reduction system,* " Med Eng Phys 2008.

[26] A.Cichocki, S.Amari, *" Adaptive blind signal and image processing : learning algorithms and applications,"* Chichester, England: Wiley ،2003.PP.87-90.

[27] R.Polikar, " *The Wavelet Tutorial Part I,"* The Wavelet Tutorial Part I by ROBI POLIKAR, june 5,1996. [Online]. Available: http://users.rowan.edu/~polikar/WAVELETS/WTpart1.html.

[28] A.Graps, *"An Introduction to Wavelets,"* IEEE Com. Science and Eng.Vol.2, no.2, 1995.

[29] Wikipedia, the free encyclopedia, *"Discrete wavelet transform,"* Discrete wavelet transform, 2010.Available: http://en.wikipedia.org/wiki/Discrete_wavelet_transform. [May,2010].

[30] D.Cvetkovic, E.D.Ubeyli. I.Cosic, " *Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: A pilot study,"* Elsevier. [Online]. Available at Sciencedirect. 2007.

[31] G.Chandroth, A J C.Sharkey, N E.Sharkey, *"Vibration signatures, wavelets and principal components analysis in diesel engine diagnostics,"* University of Sheffield,UK.

[32] T.Carter, "*An Introduction to Information Theory and Entropy*," Santa Fe, 2007.

[33] Wikipedia," *Information Theory,"* information theory-wikipedia, the free encyclopedia, Wikipedia, 31 january 2010. [Online]. Available: http://en.wikipedia.org/wiki/Information_theory.

[34] E.S.Gopi," *Algorithm Collections for Digital Signal Processing Applications using Matlab,"* Dordrecht, the Netherlands: Springer.

[35] K.J.Ray Liu, "*Pattern Recognition and Image Preprocessing*, "university of Maryland, Maryland, second edition, Marcel Dekker, 2002.

[36] S.J.Russell, P.Norvig," *Artificial Intelligence A Modern Approach*," second edition, Upper Saddle River, New Jersey: Pearson Education, Inc, 2003.

[37] H.Demuth, M.Beale, M.Hagan," *Neural Network Toolbox™ 6 User's Guide,"* MathWorks, Inc., 1992-2009.

[38] R.O.Duda, P.E.Hart, D.G.Stork," *Pattern Classification,"* second edition, 2000.

[39] Z.Zainuddin, N.Mahat, Y.Abu Hassan, "*Improving the Convergence of the Back propagation Algorithm Using Local Adaptive Techniques,"* World Academy of Science, Engineering and Technology . 2005.

[40] V.V.Phansalkar, P.S.Sastry," *Analysis of the Back-Propagation Algorithm with Momentum,"* IEEE transactions on neural network, VOL.5.NO, 3, May 1994.

[41] N.Ahmed, *" Fuzzy Logic Control Using Matlab Part I,"* Khwarzimic Scinence Society at the University of Lahore, 2003.

[42] R.Fuller," *Neural Fuzzy System,"* 1995.

[43] J.Jantzen, *"Tutorial on Fuzzy Logic,"* Technical University of Denmark, Department of Automation, Bldg 326, DK-2800 Lyngby, DENMARK.Tech. Report no 98-E 868, 19 Aug 1998 (logic).

[44] P.Antasaklis," *Defining Intelligent Control*," IEEE Control System Society, 1993. Available: http://www.nd.edu/~pantsakl/elecpubs/isis-94-001.pdf

[45] J.Jantzen, *"Design of Fuzzy Controllers,"* Technical University of Denmark, Department of Automation, Bldg 326, DK-2800 Lyngby, DENMARK.Tech. Report no 98-E 864 (design), 19 Aug 1998.

[46] Signal processing toolbox, mathworks Inc, http://www.mathworks.com

[47] MIT/BIH Arrhythmia Database, Harvard University and Massachusetts Institute of Technology Division of Health Sciences and Technology, Cambridge, MA. Available: http://www.physionet.org.

[48] physionet, " WFDB *quick start for MS-Windows 95/98/NT/2000/ME/XP/Vista/7," WFDB quick start for MS-Windows 95/98/NT/2000/ME/XP/Vista/7,"* 13 October, 2009, [Online]. Available: http://www.physionet.org/physiotools/wfdb-windows-quick-start.shtml

[49] A.Josko, "*Discrete Wavelet Transform In Automatic ECG Signal Analysis*," Technology Conference-IMTC, Warsaw, Poland. May 1-3, 2007.

[50] G.Q.Gao, "Computerized Detection and Classification of Five Cardiac Conditions," M.S.thesis, Auckland university, New Zealand, 2003.

[51] R.Gutierrez, *"lecture 13: validation,"* Wright State University. Available: http://users.soe.ucsc.edu/~manfred/opubs/crossval.pdf .

[52] Wavelet toolbox, mathworks Inc, http://www.mathworks.com