MORPHOLOGICAL CHARACTERIZATION OF ECG SIGNAL ABNORMALITIES: A NEW APPROACH

Mohamed O. Ahmed Omar[1,3], Nahed H. Solouma[2], Yasser M. Kadah[3]

[1] Misr University for Science and Technology, 6th October City, Egypt

[2] National Institute of Laser Enhanced Sciences, Cairo University, Giza, Egypt

[3] Biomedical & System Department, Faculty of Engineering, Cairo University, Giza, Egypt

E-mail: eng_mohamedomar@hotmail.com

Abstract- **Examining the ECG signal carefully may lead to an accurate diagnosing for the heart abnormalities. Several techniques have been conducted in the area of arrhythmia detection using classical methods of classification. In the last two decades, there has been an increasing interest in applying techniques from the domains of nonlinear analysis and chaos theory in the characterization of the ECG signal. In this work, we propose a new characterization method based on the morphological analysis of the ECG signal. The data set used were taken from the world-famous MIT-BIH Arrhythmia database, as well as several other international ECG signal databases. Most of the heart abnormalities were detected using this novel method with a high degree of confidence.**

I. INTODUCTION

Cardiac arrhythmias are alterations of cardiac rhythm that disrupt the normal synchronized contraction sequence of the heart and reduce pumping efficiency. Causes include rate variations of the cardiac pacemaker, ectopic pacemaker sites, and abnormal propagation of pacing impulses through the specialized cardiac conduction system [1].

Types and frequency of occurrence of arrhythmias provide an important indication of the electrical stability of the heart. In particular, certain ventricular arrhythmias are thought to indicate susceptibility to life-threatening conditions. Since arrhythmias can be suppressed by anti-arrhythmic drugs, early recognition is important [1].

Conventional methods of monitoring and diagnosing arrhythmia rely on detecting the presence of particular signal features by a human observer. Due to the large number of patients in intensive care units and the need for continuous observation of such conditions, several techniques for automated arrhythmia detection have proliferated since the early 1960's to attempt to solve this problem and many are used clinically [2]. Such techniques work by transforming the mostly qualitative diagnostic criteria into a more objective quantitative signal feature classification problem. Classical techniques have been used to address this problem such as the analysis of electrocardiogram (ECG) signals for arrhythmia detection using the autocorrelation function [3], using frequency domain features [4], using time frequency analysis [5], and wavelet transform [6], [7]. Other techniques used adaptive filtering [8], sequential hypothesis testing [9], [10].

In this work, we propose a comprehensive approach to the problem of diagnosing heart abnormalities from ECG signals. This approach classifies the ECG signal as being normal or abnormal and then subclassifies the abnormal cases according to its diseases by extracting new features from ECG signals. Feature extraction is based on the morphology of the ECG signal.

The results of this work are expected to outline a useful diagnostic tool that can be implemented in modern cardiac monitors to assist physicians reach more accurate diagnostic rates.

This paper is classified as follows: The first section, is an introduction and a summary for the pervious work. The second section, describe the materials used in this work and the methods applied to detect different heart abnormalities. In the third section, we provide the results of this work and finally we discuss the result and give a conclusion in the last section.

II. METHODOLOGY

2.1 *Data Acquisition*

In this study, a total of more than two hundreds ECG signals were acquired from the world-famous MIT-BIH Arrhythmia database. MIT-BIH Arrhythmia database consists roughly 109,000 beats that have been manually annotated by at least two cardiologists working independently. Each signal file contains two signals sampled at 360 Hz. Modified lead II (MLII) has been provided as one of the two channels ECG recordings [11].

After acquiring the data from MIT-BIH arrhythmia data base, we transformed it into text files for further analysis. We used the Matlab 7.0 Mathwork,Inc. to build the proposed system, because it has robust tool boxes that can help in this work. From the Matlab tool boxes we used the signal processing tool box, the multivariate statistics tool box and others.

2.2 *Data Preprocessing*

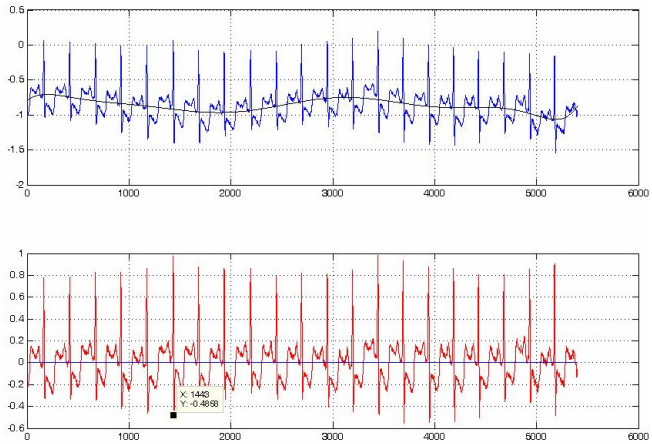
Two problems were found as we started dealing with data : (1) the presence of noise which distort the signal and (2) the base line is incorrect which has a wave form that resembles sea waves in most cases.

2.2.1 *Baseline correction:*

we used a polynomial curve fitting algorithm to correct the baseline. The whole signal was fitted and then the original signal was subtracted from this polynomial to set the correct baseline of the signal to zero as shown in fig (1).

The polynomial equation:

$$P(x) = P_1 x^n + P_2 x^{n-1} + \cdots\cdots + P_n x + P_{n+1}$$



Fig(1) Baseline correction after using polynomial fitting.

2.2.2 *Inflection point detection:*

All waves P, T, U, Q, R, S have inflection point throughout the whole signal as shown in fig (2). These points are known as a peak or a valley.

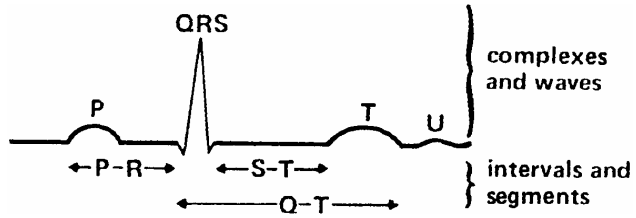


Fig (2) ECG signal

In ordinary math, the first differentiation gets the inflection points and the second gets if it is either a peak or a valley. However, since fixed equation can be fitted to the signal, mathematics failed when numbers were substituted by zero. To solve this problem, the signal points were grabbed and sorted to be a peak or a valley as the edges rise or fall.

Automatic threshold:

Threshold was used to help initially extracting the **R** wave from the upper side of the ECG signal or the **S** wave from the opposite side, from which other features could be extracted. After determining inflection points, it is preferred to build the automatic threshold on the upper side of the inflection point (peaks) and when necessary at the opposite side (valley). This process is done by rearranging the peaks or valley of the signal in descending order and locating the threshold line between the highest difference between them.

Peak Separation :

Peak is determined by detecting the change in the wave from positive to negative.

Valley Separation:

Valley is determined by detecting the change in the ECG signal from negative to positive.

2.3 *Feature Extraction*

From pervious procedure we could now, extract **R** , **S** waves and determine the **R**- **R** intervals, and hence we can get an initial estimate for the heart rate.

2.3.1 *R wave detection*

After baseline correction and building an automatic threshold, we could detect the **R** wave as the highest value in the ECG signal by separating the points above the threshold from the rest of the signal. This concept is illustrated in figure(3).

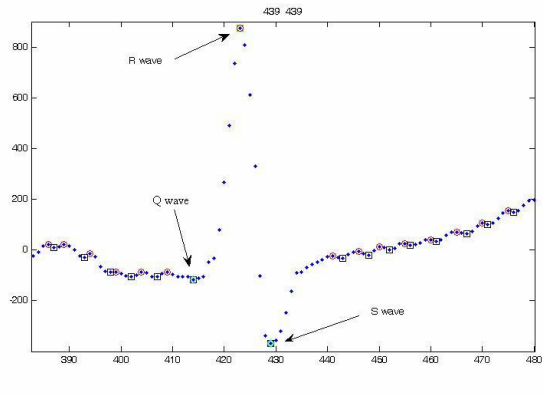


Fig (3) Q, R ,S waves detection.

2.3.2 *Detecting the **R-R** interval:*

The R-R interval is the measured period between two consecutive R waves.

2.3.3 *Heart Rate Determination*

Heart rate is the number of beats during a period of time. The following equation describe how could we measur the heart rate:

$$HR = \frac{\text{No. of R wave}}{\text{Time between (1st \& last) R wave}}$$

2.3.4 *Q and S waves Detection*

The **S** and **Q** waves were detected by dividing the **R-R** interval into four equal groups. By studying the inflection points (peaks and valleys) which were detected previously, we can extract the Q wave as the valley directly before the R wave and the S wave as the biggest valley directly after the R wave as shown in figure (3).

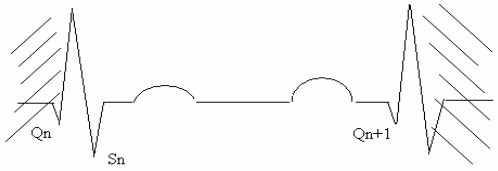
2.3.5 *QRS complex detection*

By knowing the **R**, **Q** and **S** waves, it is easy to determine the QRS complex because the three values are the three following inflection points. QRS interval was the result from subtracting the S and Q waves.

2.3.6 *Detecting the T wave*

2.3.6.1 *Smoothing period between Q and S waves*

knowing that P, T and U waves located between Q and S waves, The period between them must therefore be cut in order to detect the pervious waves. The first Q wave however, must be neglected if it appeared before the R wave because there may be no S wave with it and this period started from S_n and ended at Q_{n+1} as shown in figure(4).



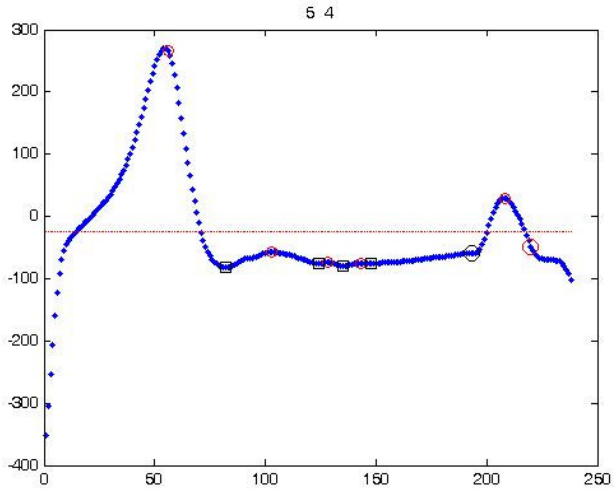


Fig (4) Signal cutting between S & Q waves (QS period)

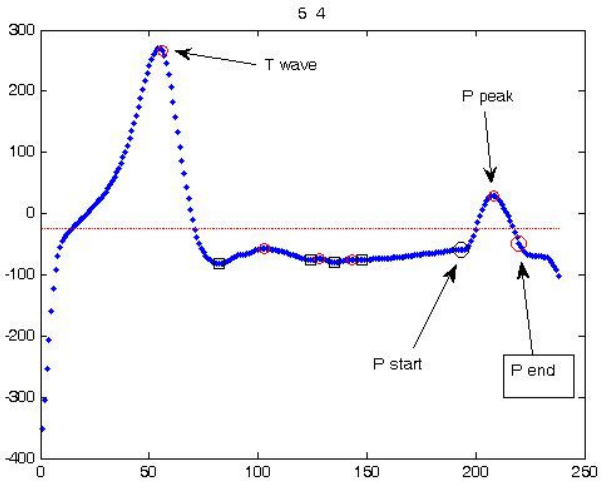


Fig (5) shows P and T waves detection.

2.3.6.2 *Smoothing period between Q & S waves*

From the observation when dealing with ECG signals, the most distortion period is between Q & S waves. It has much more inflection points which must be filtered to get the required inflection points that referred to P & T waves. After many trials, smoothing was found to be enough to do this job.

T detection, first smoothing take place to avoid taking wrong peaks or valleys and it was proven experimentally on many diseases that smoothing can be stopped after 15 peaks and valley totally, then by assuming it is the highest peak or valley in the first half of the regions directly after S wave. As shown in figure (5)

2.3.7 *P wave detection*

After further smoothing till number of peaks and valleys totally reaches 6, also experimentally then by assuming it is the highest peaks in the second half of the regions directly before Q wave. Shown in figure (5)

2.3.8 *ST detection*

After getting S and T waves the points between them are sorted positively and negatively to their places from the baseline to determine which of them has more points than the other. Shown in figure (6)

2.3.9 *QT detection*

QT interval was the result from subtracting Q and T waves as shown in figure (7).

2.3.10 *PR interval*

PR interval measured from beginning of P wave to the beginning of QRS Complex as shown in figure (7).

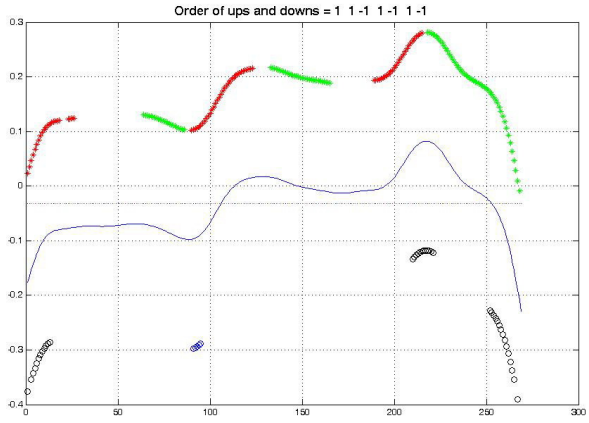


Fig (6) ST segment detection

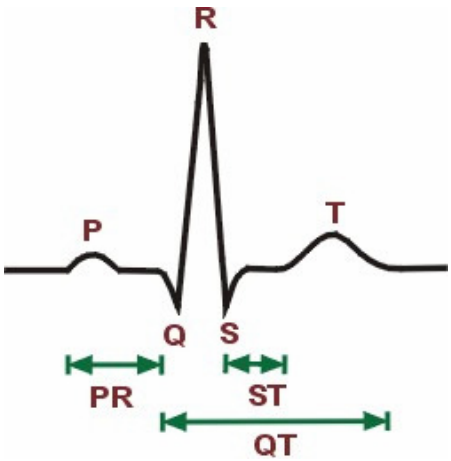


Fig (7) ECG signal intervals.

2.4 *Classification Techniques*

The output stage is a classifier to assign a pattern to a certain class, or descriptor for the pattern. Generally speaking, the data set is usually divided into a training set and a test set. Some features are to be extracted from every input (training or test) to represent the pattern. The subject of pattern recognition is subdivided into: the statistical pattern recognition, and the neural network approach. [12].

Statistical Pattern Recognition

There exist many statistical classification methods such as the supervised and unsupervised learning methods. The unsupervised learning method, known as clustering, attempts to develop a representation for the given sample data. This method is used when the data set is unlabeled, i.e. the different classes included in the data are not known, such as k-means classifier. Using the unsupervised learning, subsets of these data may be formed into natural groupings or 'clusters', where each cluster most likely corresponds to an underlying pattern class [12]. Though, we already had labeled data we used with this classifier to make sure the data sets are valid. The non-parametric methods of classification is the supervised learning such as Euclidian distance classifier, and the nearest neighbor (KNN) classifier were used in our work.

III. RESULT

We applied T-test hypothesis on all features in discriminating different ECG signal morphologies. This test enable us to test the significance of each feature and choose the most discriminate one before introducing it to the different classification algorithms we developed.

We tested totally 12 features extracted from ECG signals in 61 ECG samples that represent normal and abnormal waves, after the test we performed about 84 tests and had 68 significant (81%), and 16 insignificant (19%). From the visual inspection of the extracting features and wave morphology we have the results shown in table (1).

Table (1), statistical results shows the accuracy of all features extracting

ECG	Features											
	Ph	Pw	Q	R	S	QRS	T	ST	RR	HR	QT	PR
stdb	98%	76%	100%	100%	91%	90%	71%	67%	100%	100%	67%	76%
svdb	79%	80%	100%	100%	100%	98%	92%	79%	100%	100%	98%	80%
itdb	83%	88%	100%	100%	100%	100%	100%	100%	100%	100%	100%	88%
vfdb	38%	34%	81%	87%	78%	78%	41%	84%	79%	79%	49%	34%
cudb	90%	86%	98%	98%	88%	88%	83%	83%	99%	99%	83%	86%
cdb	88%	55%	100%	100%	99%	100%	77%	91%	100%	100%	77%	55%
nstdb	88%	56%	100%	100%	100%	100%	100%	98%	100%	100%	123%	56%
Normal	100%	74%	100%	100%	100%	100%	100%	100%	100%	100%	100%	74%
Total accuracy %	81%	65%	98%	99%	95%	95%	82%	88%	98%	98	83%	65%

We have applied Euclidian distance classifier. At first we tried to use the average of all features (significant and insignificant)in this classifier, as we had much more samples to be classified. Second trials was using only significant features. Third trials was using one sample from the patients representing all ECG signals at constant duration. Forth trials was using three samples from the patient representing all ECG signal. The result we got in classifying the 61 ECG samples is shown in table (2).

Table (2), Euclidian distance classifier results

Trail	Nor.	Stdb	Svdb	Itdb	Vfdb	Cudb	Cdb	Nstdb	Total
1	100%	33%	38%	100%	0%	67%	0%	100%	51.7%
2	100%	44%	88%	100%	20%	67%	0%	100%	62%
3	100%	67%	75%	100%	20%	67%	20%	100%	67.2%
4	100%	78%	88%	100%	40%	67%	30%	100%	74.1%

Supervised pattern recognition used also by nearest neighbor classifier (KNN classifier), gave the result shown in table (3).

First trials was using one sample from the patients representing all ECG signals at constant duration. Second trials was using three sample from the patient representing all ECG signals at constant duration.

Table (3), KNN classifier results.

Trail	Nor.	Stdb	Svdb	Itdb	Vfdb	Cudb	Cdb	Nstdb	Total
1	100%	99%	100%	100%	0%	34%	0%	100%	66%
2	100%	99%	100%	100%	10%	16%	20%	100%	67%

From classification results shown in tables (1) and (2) we can increase the sensitivity of these classifiers when we using more sample.

IV. DISCUSSION AND CONCLUSION

In this work, we have characterized the features extracted from the ECG signal based on their morphology. That is, we have established a new approach to classify between normal and abnormal signals. The ECG signal used in this approach was taken from MIT-BIH Arrhythmia database. Furthermore, we have used the Euclidian distance classifier and the nearest neighbor (KNN) classifier in order to classify the differences between the signals.

The proposed approach has yielded reasonable classification accuracy between the normal ECG and the abnormal ECG. Our approach has yielded low classification accuracy in some abnormal ECG signals. Therefore, it required using more ECG samples in order to enhance the performance of our approach for the similar signal.

REFERENCES

[1] D. A. Coast, R. M. Stern, G. G. Vano and S. A. Biller, "An approach to cardiac arrhythmia analysis using Hidden Markov Model, " *IEEE Trans. Biomed. Eng.* ,vol. 37, no. 9, Sep. 1990.

[2] C. L. Feldman, "Computer detection of cardiac arrhythmias: Historical review, " *Amer. Rev. Diag.*, vol. 2, no. 5, pp. 138-145, 1983.

[3] S. G. Guillén, M. T. Arredondo, G. Martin, and J. M. F. Corral, "Ventricular fibrillation detection by autocorrelation function peak analysis," *J. Electrocardiol.*, vol. suppl. 22, pp. 253–262, 1989.

[4] K. Minami, H. Nakajima, and T. Toyoshima, “Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network,” *IEEE Trans. Biomed. Eng.*, vol. 46, pp. 179–185, Feb. 1999.
[5] V. X. Afonso and W. J. Tompkins, “Detecting ventricular fibrillation, selecting the appropriate time-frequency analysis tool for the application,” *IEEE Eng. Med. Biol. Mag.*, pp. 152–159, Mar. 1995.
[6] L. Khadra, A. S. Al-Fahoum, and H. Al-Nashash, “Detection of life threatening cardiac arrhythmias using the wavelet transformation,” *Med. Biolo. Eng. Comput.*, vol. 35, pp. 626–632, Nov. 1997.
[7] A. S. Al-Fahoum and I. Howitt, “Combined wavelet transformation and radial basis neural networks for classifying life-threatening cardiac arrhythmias,” *Med. Biolo. Eng. Comput.*, vol. 37, pp. 566–573, Sept. 1999.
[8] N. V. Thakor and Y. Zhu, “Applications of adaptive filtering to ECG analysis: Noise cancellation and arrhythmia detection,” *IEEE Trans. Biomed. Eng.*, vol. 38, pp. 785–794, Aug. 1991.
[9] N. V. Thakor, Y. Zhu, and K. Pan, “Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm,” *IEEE Trans. Biomed. Eng.*, vol. 37, pp. 837–843, Sept. 1990.
[10] N. V. Thakor, A. Natarajan, and G. F. Tomaselli, “Multiway sequential hypothesis testing for tachyarrhythmia discrimination,” *IEEE Trans. Biomed. Eng.*, vol. 41, pp. 480–487, May 1994.
[11] MIT-BIH Arrhythmia database directory, 3rd Edition, Harvard-MIT Division of Health Science and Technology, July, 1992.
[12] R. J. Schalkoff, "Pattern Recognition: Statistical, Structural, and Neural Approaches", JOHN WILEY & SONS,INC., 1992.