IDENTIFICATION AND CLASSIFICATION OF ECG ABNORMALITIES USING RECURRENCE QUANTIFICATION ANALYSIS

Walid El-Atabany, Abou-Bakr M. Youssef, and Yasser M. Kadah
Biomedical Engineering Department, Cairo University, Giza, Egypt

Abstract–**Cardiac arrhythmias disrupt the normal synchronized contraction sequence of the heart and reduce the pumping efficiency. Therefore, the detection and classification of such arrhythmias is essential. In this paper, four types of ventricular arrhythmias (PVC, VB, VT and VF) were considered; a number of features were extracted from ECG signal using a nonlinear dynamical signal analysis technique, recurrence quantification analysis (RQA). Then three statistical classifiers are used in the classification, results confirmed the robustness of the new techniques and demonstrate its value as a new diagnostic tool.**

Keywords – ***ECG, arrhythmia analysis, nonlinear dynamics***

I. INTRODUCTION

It has been reported that cardiac arrhythmias account for over 50% of all deaths due to heart disease in the United States [1]. Prevention of such cardiac deaths requires prompt and accurate identification of the arrhythmias from the electrocardiogram (ECG) recordings. In addition, arrhythmia classification is also an important task in interpretive systems, which is intended to provide a diagnostic classification of ECG [2]. It is of paramount importance to differentiate potentially lethal arrhythmias such as ventricular fibrillation (VF) and ventricular tachycardia (VT), from more benign problems as manifested in superventricular tachycardia (SVT)[3]. Thus, the development of accurate noninvasive techniques for identifying patients at risk of lethal arrhythmias is essential to reducing mortality from cardiac deaths.

Various detection algorithms have been reported, which can be classified as linear techniques such as sequential hypothesis testing[4], Autoregressive Modeling of ECG signal[5], frequency domain features[6], wavelet analysis[7],[8], And nonlinear techniques, which uses the concept of chaotic systems to describe and extract some features from the ECG signal. All these methods exhibit advantages and disadvantages, some being too difficult to implement and compute for AEDs and ICDs, and some having low specificity in differentiating between various types of arrhythmias. In this paper we proposed a novel technique to be used in arrhythmia classification, which take into consideration the problem of characterizing the nonlinear dynamics (chaos theory) of the ECG signal and its variation with different arrhythmia types, where a nonlinear dynamical signal analysis technique, recurrence quantification analysis (RQA), was applied to the ECG signal. It is an extension of a graphical method called recurrence plot analysis; Recurrence plots are most usefully described using a set of features collectively known as recurrence quantification analysis (RQA). The proposed implementations were used to compute these features for a large number of independent ECG signals belonging to five different ECG signal types from the MIT-BIH Arrhythmia Database [9]. The results are studied to detect statistically significant differences among different arrhythmia types. Finally, statistical classification techniques are used to assess the diagnostic effectiveness.

II. METHODS

Eckmann have introduced a tool, which can visualize the recurrence of states $\vec{x}_i$ in a phase space [10]. Usually, a phase space does not have a dimension (two or three), which allows it to be pictured. Higher dimensional phase spaces can only be visualized by projection into the two or three dimensional sub-spaces. However, Eckmann's tools enable us to investigate the *m*-dimensional phase space trajectory through a two-dimensional representation of its recurrences (Fig. 1). Such recurrence of a state at time *i* at a different time *j* is pictured within a two-dimensional squared matrix with black and white dots, where black dots mark a recurrence, and both axes are time axes. This representation is called *recurrence plot (RP).* Such an RP can be mathematically expressed as

$$R_{i,j}^{m,\varepsilon} = \left(\varepsilon - \left\|\vec{x}_i - \vec{x}_j\right\|\right), \quad \vec{x}_i \in R^m, \; i, j = 1.....N , \qquad (1)$$

Where N is the number of considered states x_i;ε is a threshold distance, ‖ . ‖ a norm and (.) the Heaviside function. The phase space vectors for one-dimensional time series u_i from observations can be reconstructed by using the Taken's time delay method, $\vec{x}_i = (u_i, u_{i+\tau}, ..., u_{i+(m-1)\tau})$ [11]. The dimension m can be estimated with the method of false nearest neighbors[11][12]. The cutoff distance ε defines a sphere centered at $\vec{x}_i$. If $\vec{x}_j$ falls within this sphere, the state will be close to $\vec{x}_i$ and thus $R_{i,j} = 1$. ε Can be either constant for all $\vec{x}_i$ [13] or they can vary in such a way that the sphere contains a predefined number of close states [14]. In this paper a fixed ε is used, resulting in a symmetric RP.

Zbilut and Webber have recently developed the recurrence quantification analysis (RQA) to quantify an RP [15-17]. They extracted some measures from the RP based on the recurrence point density and the diagonal structures in the recurrence plot, which are; the *recurrence rate* (RR), the *determinism* (DET), the *maximal length of diagonal structures* (Lmax), the *average length of the diagonal structures* (L), and the *entropy* (ENT). Gao has recently

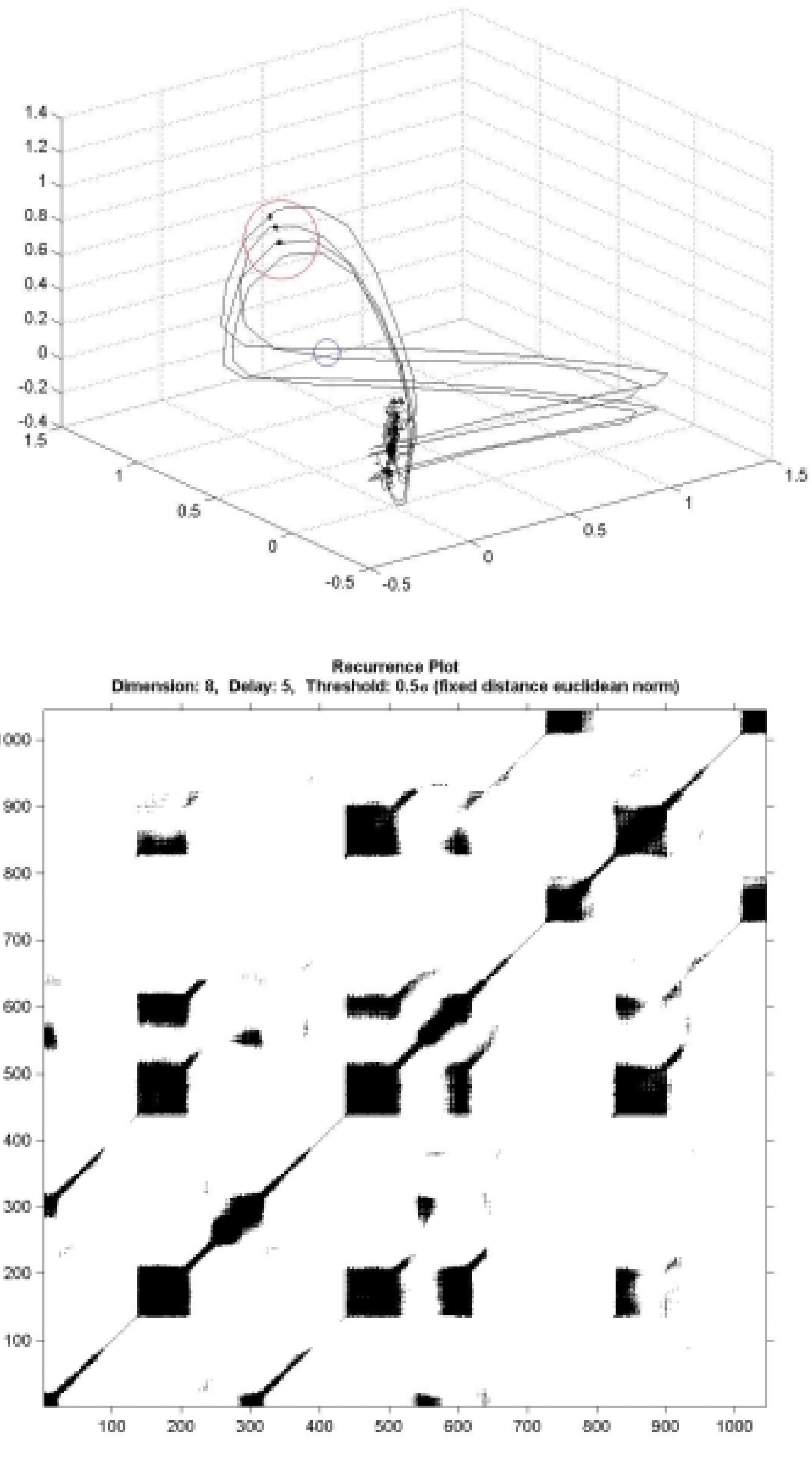

Figure 1 (a) Segment of the phase space trajectory of the ECG signal and (b) its corresponding recurrence plot. A point of the trajectory at *j* which falls into the neighborhood (large circle in (a)) of a given point at *i* is considered as a recurrence point (black points on the trajectory inside the large circle). This is marked with black point in the RP at the location (*i*, *j*). A point outside the neighborhood (small circle in (a)) causes a white point in the RP.

introduced another measures based on the vertical structures in the RP, which are; the *laminarity* (LAM), the *average length of vertical structures* (TT), and the *maximal length of vertical structures* (Vmax).

III. RESULTS AND DISCUSSION

The ECG signals used in this paper were obtained from the MIT-BIH arrhythmia database [9]; the data set was composed of five different types including normal sinus rhythm (NSR), premature ventricular couplet (PVC), ventricular bigeminy (VB), ventricular tachycardia (VT), and ventricular fibrillation (VF). The data set was divided into learning and testing data set, 52 independent signals for the learning set of each type and 26 independent signals for the test set of each type with each signal length 3 sec. All the signals were resampled at 360 samples/s.

The *m*-dimensional phase space trajectory of the ECG signals was reconstructed using the delay time embedding method, where the embedding dimension *m* was calculated using the false nearest neighborhood (*FNN*) algorithm ($m = 8$), and the delay time *l* was calculated from the first minimum of the mutual information function ($l = 6$) [18] using the *TSTOOL* package [19]. Then the recurrence plot (RP) was reconstructed with $\varepsilon = 0.7$, and the eight features (RR, DET, L, Lmax, ENT, LAM, TT, Vmax) of the recurrence quantification analysis (RQA) were extracted using the "Cross Recurrence Plot Toolbox" [20] to form the features vectors. I used the significance test in this paper to assess the use of the parameters extracted from the new technique for discriminating between different ECG signal types. Results of significance test for each feature was shown in tables I-VIII

TABLE I. P-values of t-test for RR.

Type	PVC	VT	VB	VF
NR	<1.0e-16	<1.0e-16	**0.2721**	<1.0e-16
PVC		<1.0e-16	<1.0e-16	<1.0e-16
VT			0.0033	<1.0e-16
VB				<1.0e-16

TABLE II. P-values of t-test for DET.

Type	PVC	VT	VB	VF
NR	<1.0e-16	**0.2069**	<1.0e-16	<1.0e-16
PVC		0.0098	**0.6024**	<1.0e-16
VT			0.0205	<1.0e-16
VB				<1.0e-16

TABLE III. P-values of t-test for L.

Type	PVC	VT	VB	VF
NR	<1.0e-16	<1.0e-16	0.0006	**0.3590**
PVC		0.0001	**0.2917**	<1.0e-16
VT			<1.0e-16	<1.0e-16
VB				0.0001

TABLE IV. P-values of t-test for Lmax.

Type	PVC	VT	VB	VF
NR	<1.0e-16	<1.0e-16	<1.0e-16	<1.0e-16
PVC		<1.0e-16	<1.0e-16	<1.0e-16
VT			0.0045	0.0138
VB				<1.0e-16

TABLE V. P-values of t-test for ENT.

Type	PVC	VT	VB	VF
NR	**0.5277**	<1.0e-16	<1.0e-16	<1.0e-16
PVC		<1.0e-16	<1.0e-16	<1.0e-16
VT			<1.0e-16	<1.0e-16
VB				<1.0e-16

TABLE VI. P-values of t-test for LAM.

Type	PVC	VT	VB	VF
NR	<1.0e-16	<1.0e-16	<1.0e-16	0.0011
PVC		<1.0e-16	0.0065	**0.5761**
VT			<1.0e-16	<1.0e-16
VB				0.0460

TABLE VII. P-values of t-test for TT.

Type	PVC	VT	VB	VF
NR	<1.0e-16	0.0001	0.0016	<1.0e-16
PVC		0.0002	0.0005	<1.0e-16
VT			**0.8908**	<1.0e-16
VB				<1.0e-16

TABLE VIII. P-values of t-test for Vmax.

Type	PVC	VT	VB	VF
NR	<1.0e-16	**0.9383**	0.0035	<1.0e-16
PVC		<1.0e-16	0.0004	<1.0e-16
VT			0.0188	<1.0e-16
VB				<1.0e-16

Results of the significant test of the RQA features confirm that normal ECG signals can be statistically differentiated from abnormal by using the extracted features, where various features exhibit very low P-values (i.e. significantly difference) when used to distinguish between some types of arrhythmias and large P-values (i.e. no statistically significant difference) for other types. So we can merge all of these features in one vector to be used in the detection and classification process of different arrhythmias types, three statistical classifiers are used in this paper; minimum distance classifier, Bayes minimum-error classifier, and voting k-nearest neighbor (k-NN) classifier [21].

In the classification process we first tried to use each feature separately but the result was not be good, but when we combined all the features extracted from the proposed technique the results showed the robustness of the new technique in the detection of abnormality of the ECG signal. Classification results for only normal versus abnormal ECG shown in Table IX, and the results of applying the three statistical classifiers to classify the 5 different ECG types are listed in Table X.

TABLE IX. RESULTS OF THE THREE CLASSIFIERS FOR THE DETECTION PROBLEM

Classifier	Specificity	Sensitivity
Min. distance	88.4615	75.9615
Bayes	96.1538	96.1538
K-NN (K=2)	80.9524	95.8333

TABLE X. RESULTS OF THE THREE CLASSIFIERS FOR THE CLASSIFICATION PROBLEM

	Min. Distance	Bayes	K-NN (K=2)
Specificity	88.4615	96.1538	85.0000
Sensitivity for PVC	84.6154	88.4615	100.000
Sensitivity for VT	19.2308	73.0769	68.4211
Sensitivity for VB	23.0769	84.6154	73.3333
Sensitivity for VF	61.5385	88.4615	90.4762

Results of the detection and classification process showed that Bayes minimum-error classifier seems to provide the best results which means that the clusters follow a Gaussian distribution, followed by the k-nearest neighbor classifier (the best at k=2). And the results of minimum distance classifier were rather poor which means that the classes are not linearly separable.

IV. CONCLUSIONS

Applying nonlinear signal processing techniques to signals like ECG provides very useful information for detection of cardiac abnormalities. The proposed recurrence quantification analysis (RQA) technique have been shown to be effective for the classification of cardiac arrhythmias in critically ill patients as shown in the results of a large data set of actual ECG signals from five different classes, which indicate the value of such techniques in the diagnosis of heart disease in intensive care units (ICU).

REFERENCES

[1] S. C. Vlay,“ Sudden cardiac death: An overview,” in *Manual of Cardiac Arrhythmias*, S. C. Vlay, Ed. Boston, MA: Little, Brown and Company, 1988.

[2] V. X. Afonso, W. J. Tompkins, T. Q. Nguyen, and S. Luo, “ECG beat detection using filter banks,” *IEEE Trans. Biomed. Eng*., vol. 46, pp.192– 202, Feb. 1999.

[3] S. W. Chen and P. M. Clarkson, “Detection of cardiac arrhythmias using a damped exponential modeling algorithm,” in *IEEE Proc. 1996 Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP 96)*, vol. 3, pp.1776– 1779, May 1996

[4] N. V. Thakor, Y. S. Zhu, and K. Y. Pan, “Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm,” *IEEE Trans. Biomed. Eng*., vol. 37, pp.837– 843, Sept. 1990.

[5] D. Ge, N. Srinivasan, and S. M Krishnan. “Cardiac arrhythmia classification using autoregressive modeling,” *BioMed*, Nov 2002

[6] K. Minami, H. Nakajima, and T. Toyoshima, “Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network,” *IEEE Trans. Biomed. Eng*, vol. 46, no. 2, pp. 179-185, February 1999.

[7] L. Khadra, A.S. Al-Fahoum, and H. Al-Nashash, “Detection of life-threatening cardiac arrhythmias using the wavelet transformation,” *Medical & Biological Engineering & Computing*, vol. 35, pp 626-632, November 1997.

[8] A.S. Al-Fahoum, I. Howitt, “Combined wavelet transformation and radial basis neural networks for classifying life-threatening cardiac arrhythmias,” *Medical & Biological Engineering & Computing*, vol. 37, pp 566-573, September 1999.

[9] The MIT-BIH Arrhythmia Database, *3rd ed*., Harvard-MIT Division of Health Sciences and Technology, May 1997.

[10] J. P. Eckmann, S. O. Kamphorst, and D. Ruelle, “Recurrence Plots of Dynamical Systems,” *Europhysics Letters*, vol. 5, pp. 973-977, 1987.

[11] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, England, 1997.

[12] J. H. Argyris, G. Faust, and M. Haase, *An Exploration of Chaos*, North-Holland, Amsterdam, 1994.

[13] M. Koebbe, and G. Mayer-Kress, “Use of Recurrence plots in the Analysis of Time-Series Data.” In: Casdagli, M., Eubank, S. (Eds.), *Proceedings of SFI Studies in the Science of Complexity*, Vol. XXI. Addison-Wesely, Redwood City. pp. 361-378, 1992.

[14] J. P. Eckmann, S.O. Kamphorst, D. Ruelle, “Recurrence Plots of Dynamical Systems,” *Europhysics Letters, vol.* 5, 973-977,1987.

[15] C. L. Webber, Jr., and J. P. Zbilut, “Dynamical assessment of physiological systems and states using recurrence plot strategies,” *Journal of Applied Physiology,* vol. 76, no. 2, pp. 965-973, 1994.

[16] L. L. Trulla, A. Giuliani, J. P. Zbilut, and C. L. Webber, Jr., “Recurrence quantification analysis of the logistic equation with transients.” *Physics Letters A* , vol. 223, no. 4, pp. 255-260, 1996.

[17] J. P. Zbilut, V. L. Webber, Jr., “Embeddings and delays as derived from quantification of recurrence plots,” *Physics Letters A* vol. 171, no. 3-4, pp. 199-203, 1992.

[18] A. M. Fraser, H. L. Swinney, “Independent coordinates for strange attractors from mutual information.” *Physical Review A* , vol. 33, no. 2, pp. 1134-1140, 1986.

[19] TSTOOL package Ver.1.1 web site “http://www.physik3.gwdg.de/tstool”

[20] N. Marwan, University of Potsdam, Germany, “Cross Recurrence Plot Toolbox for Matlab”, Version 4, Release 12,

[21] Y.M. Kadah, A.A. Farag, J.M. Zurada, A.M. Badawi, and A.M. Youssef, “Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images,” *IEEE Trans. Medical Imaging*, vol. 15, no. 4, pp. 466-478, August 1996.