

Variant Maps on Normal and Abnormal ECG Data Sequences

Yan Ji¹, Jeffrey Zheng^{1,2*}, Yinfu Xie³ and Tao Shou³

¹School of Software, Yunnan University, Kunming 650091, China

²Key Lab of Yunnan Software Engineering, Kunming 650091, China

³The First People's Hospital of Yunnan Province, Kunming 650031, China

*Corresponding author: Jeffrey Zheng, School of Software, Yunnan University, Kunming, China 650091, Tel: +86-1310883909; E-mail: conjugatologic@yahoo.com

Received date: July 30, 2016; Accepted date: September 05, 2016; Published date: September 12, 2016

Copyright: © 2016 Yan J, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Abstract

ECG data sequences are classical and most reliable clinical data for patients to provide complex physiological and pathological information diagnosing various heart diseases. Extracting dynamic information from ECG signal time sequences, Poincare maps have developed as classical assistant tools using two dimension maps as important basis for medical doctors to diagnose multiple cardiovascular diseases. Since simulation systems of human heart could be extremely complicated on chaos behaviors, Poincare maps based on paired measures have some limitations to excavate ECG data sequences on special physiological and pathological information.

In this paper, we propose a new measuring model based on multi-dimensional measurements using variant maps to handle ECG data sequences in refined visual representations. System architecture of this model and their core components are discussed. Under this construction, normal and abnormal ECG data sequences can be represented as variant maps. Sample results are illustrated as a set of two dimensional variant maps for selected ECG data sequences.

Keywords: Heart diseases; ECG data sequences; Variant maps; Visualization; Probability measurement

Introduction

Advanced health informatics is listed as one of the fourteen engineering grand challenges for the 21st century [1]. Activities in this field include acquiring, managing, and using biomedical information. It is necessary to establish health assurance systems from personal to global levels, to enhance the quality and efficiency of medical care and the response to widespread public health emergencies [1].

Most people are well known that the Electrocardiogram (ECG) are represented as irregular curves on long grid papers to record electrical activities of a heart. However beside physician experts, a few people can really understand possible meanings of complicated ECG curves. From a measurement viewpoint, ECG signals are physical indices as heart activities in cycles to indicate various electrical signals. There are various states before contraction, pre excites on the heart, and then spreading to the patient's body, identifying electrical levels on the body to produce potential differences. ECG devices record these signals in potential differences, such graphic patterns are called electrocardiogram [1].

Due to noninvasive nature and intrinsic information on heart activities [1], ECG signals are the most common signals of heart diseases widely used in general medical practice over the world of health environment. ECG can examine the nature of the diagnosis of patient's heart with arrhythmia or rhythmia properties. The collection of ECG data obtained from volunteers is useful with huge amount of data sequences. In modern world, heart diseases are already occurring as the largest portion of daily health diseases on the transmission

system in human society [2]. The analysis of ECG data sequences plays an essential role to diagnose various heart diseases in clinic practices in the field of health information to be giant projects in advanced medical practices over the world.

At present, in the direction of medical diagnosis on ECG signals, a useful processing model is based on Poincare maps invented by the eminent French scientist in 100 years ago that are widely applied in modern chaos theory and complex systems. Based on a set of paired ECG data on selected sequences, this type of maps can be generated as visual maps in the identification of different types of cardiac arrhythmias [2]. Using Poincare maps, ECG signals are transformed as scattered point diagram based on a nonlinear analysis method as the foundation. It is feasible for computer software to convert longer ECG data sequences into simple graphical representation as 2D maps [3].

Applying modern probability and statistical analysis tools, this type of processing models could be useful to make feature extractions mining implicit information in a large set of sample ECG data sequences. It is difficult for classical ECG technologies to identify a particular class of heart diseases with nonlinear physiological or pathological information. Due to specific types of diseases may be observed as a family of similar curves with the same variation on measurements from data sequences, their spectrum in feature maps may be identified as normal rhythm and arrhythmia ECG signals.

To explore efficient technology analysis ECG signals using advanced methodologies, a new model based on variant maps [2-10] are applied to ECG data sequences.

In consequent section, variant maps on ECG data sequences to identify normal/abnormal ECG data sequences will be described and different visual results will be analyzed.

System Architecture

In this section, architecture of a variant map system and its core components are discussed to process ECG data sequences generating variant maps with the use of diagrams. The refined definitions and equations of this system are described.

Architecture

The Architecture of the ECG Variant Map system is composed of three components: Transformation Component TC, Measure Component MC, and Visualization Component VC shown in Figure 1. Each component is composed of one ~ four modules respectively discussed in the next subsection.

A list of parameters can be described as follows:

$$\left\{ \begin{array}{l} E: \text{Multiple value ECG data sequences} \\ W: \text{Window size} \\ R: \text{The steady state interval value, } 0 < R < 1 \\ M: \text{Segment length} \end{array} \right.$$

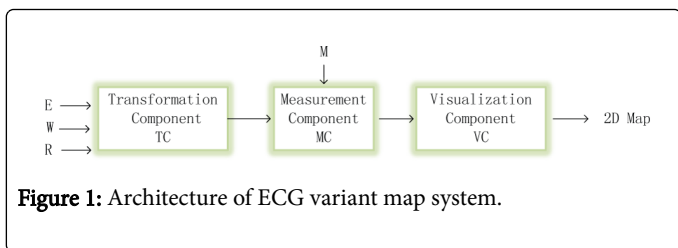


Figure 1: Architecture of ECG variant map system.

Under this construction, the component of TC processes a selected ECG data sequence E and two control parameters $\{W, R\}$ as inputs: E an ECG data sequence with N elements, W pre-set Window-Size value, and R the steady state interval value. Undertaken relevant computations and processing, the output of TC component is composed of a Pseudo-DNA data sequence that will be provided as an input data sequence for the MC. Inputs from TC with a segment value M as a control parameter are processed in the MC. The output of the MC is organized as four vectors of probability measurements. Selecting two vectors from the four probability vectors, they become the input data of the VC component, the output of the VC is composed of a pair of position values created from each M values from the two selected vectors. The pair of positions in the VC determines a projected position as a visual point. After all elements on the selected ECG data sequence are processed, multiple segments are transformed as a set of multiple paired values to generate relevant graphs to indicate their distribution properties on 2D variant maps respectively.

Core modules

It is essential to describe core modules in each component. Three components are described as follows.

The TC component

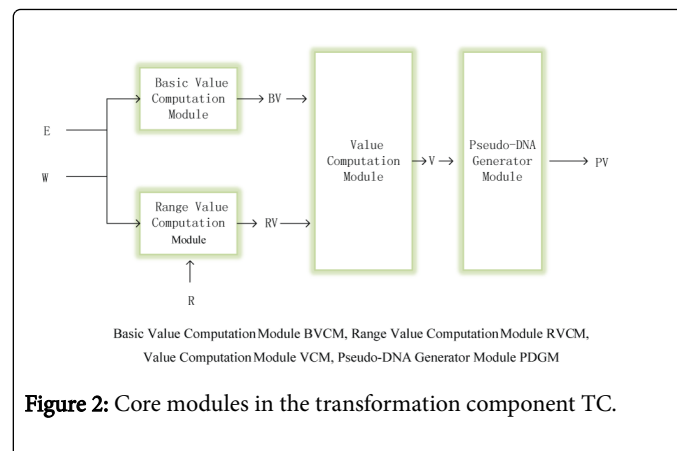


Figure 2: Core modules in the transformation component TC.

The TC component is composed of four modules: Basic Value Computation Module BVC, Range Value Computation Module RVC, Value Computation Module V and Pseudo-DNA Generator Module PDGM shown in Figure 2. The output of the TC component provides its output as the input of the MC component.

Main I/O parameters of the TC component are organized in three groups (Input, Intermediate and Output):

Input group:

E ECG signals with N elements;

W pre-set Window-Size value;

R the steady state interval value;

Intermediate output/input group

$BV: BV = \{bv_0, bv_1, bv_2, \dots, bv_i, \dots, bv_N\}$, Basic value sequence; $RV:$

$RV = \{rv_0, rv_1, rv_2, \dots, rv_i, \dots, rv_N\}$, Range value sequence; $V:$

$V = \{v_0, v_1, v_2, \dots, v_i, \dots, v_N\}$, Value sequence;

Output group:

$PV: PV = \{pv_0, pv_1, pv_2, \dots, pv_i, \dots, pv_N\}$, Pseudo-DNA sequence.

The inputs of BVC are E, W and the output is BV . The inputs of RVC are E, W, R and the output is RV . The inputs of V are E, W, R, BV, RV and its output is V . This vector provides the input of the PDGM, and the PDGM's output is PV to be transformed as a Pseudo-DNA sequence.

The MC component

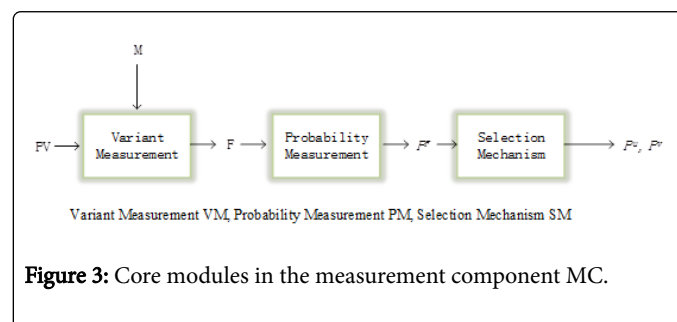


Figure 3: Core modules in the measurement component MC.

The MC component is composed of three modules: Variant Measurement VM, Probability Measurement PM, and Selection Mechanism SM shown in Figure 3. The inputs of VM component are composed of a Pseudo-DNA sequence PV and a segment length M, its output is a sequence composed by four symbols: {⊥, +, −, T}. The input of the PM component is the output of VM component, and the output is composed of probability measurement sequences: four sets of probability measurements. The input of SM component is the four sets of probability measurements and the output is selected from two sets of probability measurements (P^F).

The MC component is shown as Figure 3 and its I/O parameters are listed as follows.

Input group: PV: Pseudo-DNA sequence composed by four bases: {A, C, G, T};

M: segment length;

Intermediate output/input group: N: length of ECG sequence;

0-1 sequence: sequences obtained by respectively map {A, C, G, T};

F: multiple sequences are composed by four symbol: $F \in \{\perp, +, -, T\}; 1 \leq i \leq N/M; M^F$: Four possible measures: $M^\perp, M^+, M^-, M^T; F \in \{\perp, +, -, T\}; P^F$: probability measure sequences of four symbols: $F \in \{\perp, +, -, T\}$; Output group: P^u, P^v : probability measure sequences of two symbols: $u \in \{\perp, +\}, v \in \{-, T\}$.

The inputs of MC component are a Pseudo-DNA sequence PV, M and intermediate input parameters. Final output is probability measure sequences of two symbols. In this paper, we select a paired of probability measure sequences on two symbols and respectively.

The VC component

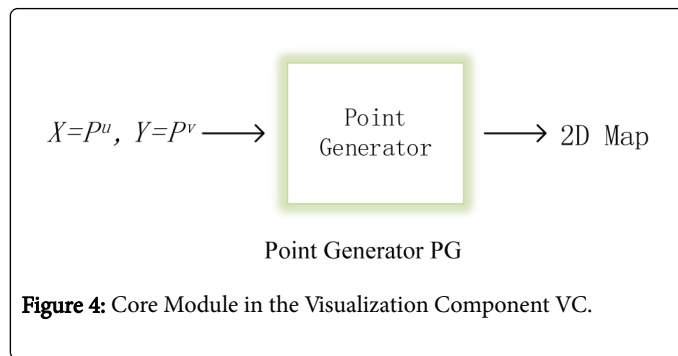


Figure 4: Core Module in the Visualization Component VC.

This VC component is composed of one module only: Point Generator PG. The input of PG is a set of paired measures X and Y; $X = P^u = \{P^u_0, P^u_1, P^u_2, \dots, P^u_i, \dots, P^u_{N/M-1}\}$, $Y = P^v = \{P^v_0, P^v_1, P^v_2, \dots, P^v_i, \dots, P^v_{N/M-1}\}$, $u = \perp, v = T$. Final output is 2D variant maps of selected ECG data sequences.

The VC component is shown in Figure 4 and its I/O parameters are listed as follows.

Input group $X = P^u, Y = P^v$, a set of paired probability measure sequences;

Output group: 2D map This part uses paired values to repeat process on rest segments generated from the selected ECG data sequence.

Detailed Description

Section 2 provides system description on ECG variant map system, it is necessary to make explanation for its further details.

Parameter description

$E = \{e_0, e_1, e_2, \dots, e_s, \dots, e_{N-1}\}$: Selected ECG data sequence with length N. $D = \{d_0, d_1, d_2, \dots, d_i, \dots, d_{N-1}\}$: Generated Pseudo-DNA sequence PV with length N, $d_i \in \{A, C, G, T\}$. $X^d = \{x_0, x_1, x_2, \dots, x_i, \dots, x_{N-1}\}$: 0-1 sequences with length N, $d \in \{A, C, G, T\}$. $X^d_i = \{x_{(i,0)}, x_{(i,1)}, x_{(i,2)}, \dots, x_{(i,j)}, \dots, x_{(i,M-1)}\}$: 0-1 sequence of every segment $0 \leq i \leq N/M$ and length of every segment $0 \leq j < M, d \in \{A, C, G, T\}$.

Transformation component

BVC module is used to compute the basic value of the selected ECG data sequence. Equation of computation BV is listed as follows:

$$bv_i = \frac{1}{W} \sum_{s=i}^{W+i-1} e_s, \quad (0 \leq i \leq N-W; 0 \leq s < N)$$

RVC module is used to compute the range value of the selected ECG data sequence. Equation of computation RV is listed as follows:

$$rv_j = \frac{R}{2} (\max\{e_s, e_{s+1}, e_{s+2}, \dots, e_{s+W-1}\} - \min\{e_s, e_{s+1}, e_{s+2}, \dots, e_{s+W-1}\})$$

$(0 \leq s \leq N-W; s = j)$

$\max\{e_s, e_{s+1}, e_{s+2}, \dots, e_{s+W-1}\}$: Max value in Window-Size.
 $\min\{e_s, e_{s+1}, e_{s+2}, \dots, e_{s+W-1}\}$: Min value in Window-Size.

VC module is used to compute final value of the selected ECG data sequence. Equation of computation V is listed as follows:

$$v_i = \frac{2 * (e_i - bv_i)}{R * rv_i}, \quad (0 \leq i \leq N-W)$$

According to the following method, a Pseudo-DNA sequence is produced:

$$d = \begin{cases} A, & v_i \geq R; \\ G, & 0 \leq v_i < R; \\ C, & -R \leq v_i < 0; \\ T, & v_i < -R. \end{cases}$$

Measurement component

A Pseudo-DNA sequence is converted into four 0-1 sequences. Using the symbol to show $Map_A, Map_G, Map_C, Map_T$ Detailed method [1-10] is described as follows:

$$X^d = \begin{cases} 0, & d_i \neq d; \\ 1, & d_i = d. \end{cases} (d \in \{A, C, G, T\}, 0 \leq i < N)$$

Example: A Pseudo-DNA sequence is AGCTAAAGGGTTCGCTACGCGGCTA. Then four distinct sequences can be generated:

$$Map_A: X_A=1000111000000000100000001;$$

$$Map_G: X_G=0100000111000100001011000;$$

$$Map_C: X_C=0010000000001010010100100;$$

$$Map_T: X_T=0001000000110001000000010.$$

Each 0-1 sequence can be segmented by length M into multiple segments and using X_i^F to show 0-1 sequence of the i-th segment. Example: If a 0-1 sequence of Map_C is 0010000000001010010100100 and length M=5, then this sequence can be arranged as following segments: $X_0^C=00100$ $X_1^C=00000$ $X_2^C=00101$ $X_3^C=00101$ $X_4^C=00100$

0-1 sequences are converted into variant sequences through method [1] as follows:

$$F = \begin{cases} \perp, & x_{i,j} = 0 \text{ and } x_{i+1,j} = 0; \\ +, & x_{i,j} = 0 \text{ and } x_{i+1,j} = 1; \\ -, & x_{i,j} = 1 \text{ and } x_{i+1,j} = 0; \\ \top, & x_{i,j} = 1 \text{ and } x_{i+1,j} = 1; \end{cases} (0 \leq i \leq M-2)$$

0-1 sequences are converted into variant sequences of being composed variant symbol $\{\perp, +, -, \top\}$ through VM component. Example: if $X_0^C=00100$ $X_1^C=00000$ $X_2^C=00101$ $X_3^C=00101$ $X_4^C=00100$, then $f_0 = \perp \perp - \perp \perp$ $f_1 = \perp \perp + \perp +$ $f_2 = \perp \perp \top \perp \top$ $f_3 = \perp \perp \top \perp -$. If using M^F represent the number of variant symbol F of every segment variant sequence and $f_0 = \perp \perp - \perp \perp$ $f_1 = \perp \perp + \perp +$ $f_2 = \perp \perp \top \perp \top$ $f_3 = \perp \perp \top \perp -$, then $M^\perp = \{4, 3, 3, 3\}$, $M^+ = \{0, 2, 0, 0\}$, $M^- = \{1, 0, 0, 1\}$, $M^\top = \{0, 0, 2, 1\}$. $M = M^\perp + M^+ + M^- + M^\top = 5$.

The equation of computation probability measures on each segmented variant sequence is listed as follows:

$$P^F = \frac{M^F}{M}, F \in \{\perp, +, -, \top\}$$

Visualization component

Using this set of measurements, projective functions can be established to select a pair of values to transform an ECG data sequence into a 2D map as follows.

Let (x_i, y_i) be the i-th pair of values, $x_i = P_i^u$ and $y_i = P_i^v$, ($u \in \{\perp, +\}$, $v \in \{-, \top\}$; $0 \leq i < N/M$)

Then each pair of values locates a specific position on a 2D map for the selected ECG data sequence.

Each (x_i, y_i) determines a specific position on 2D plane. A series of projections are repeated on relevant segments and each segment corresponding to a projective point on the 2D map. This makes all processed measurements be their projective points, and finally to generate a 2D maps for the selected ECG data sequence. Each segment on the ECG data sequence generates a specific point on a 2D map; it is essential to recursively process all relevant segments using the VM.

Results of ECG Variant Maps

Applying this process based on variant model and visualization method, normal and abnormal ECG data sequences can be distinguished.

Sample ECG data information

In this paper, all ECG data sequences are provided by the First People's Hospital of Yunnan Province. A total amount of ECG data has 500 MB to collect from 220 thousand records. All records are diagnosed and classified by ECG experts. Among this data set, normal ECG data has about 138 MB and abnormal ECG data has about 362 MB. The data format is briefly shown in Figure 5.

Figure 5: ECG data format.

It is always difficult to imagine map results from control parameters directly. A list of controlled effects will be illustrated in this section. Let readers easier look at different visual effects via sample results of ECG variant maps under various controlled parameters.

Normal and abnormal ECG samples

Applying normal and abnormal ECG data sequences, different conditions illustrate their spatial distributions in a controllable environment. Sample 1D maps are shown in Figure 6.

ECG variant maps based on Pseudo-DNA

Four maps of on bases of Pseudo-DNA generated by normal and abnormal ECG data sequence are shown in Figure 7.

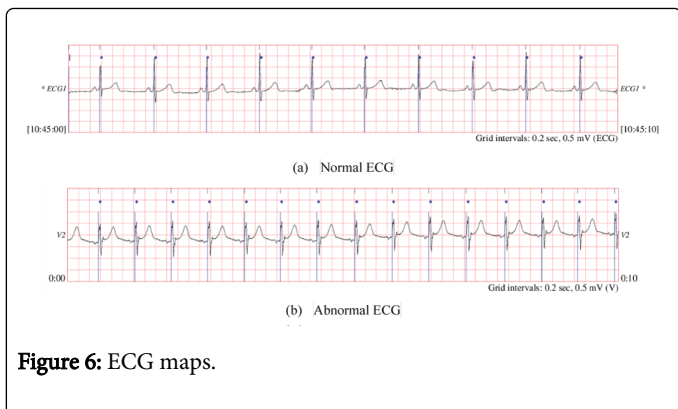


Figure 6: ECG maps.

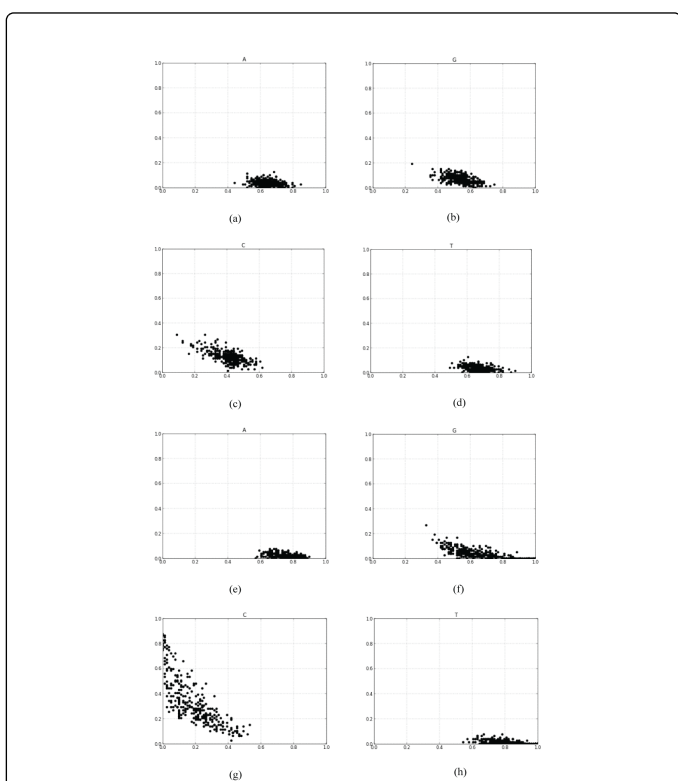


Figure 7: The ECG Variant Maps of ECG data sequences; (a-d) normal results on (a) Map_A , (b) Map_G , (c) Map_C , (d) Map_T ; (e-h) abnormal results on (e) Map_A , (f) Map_G , (g) Map_C , (h) Map_T respectively.

In Figure 7, eight ECG variant maps of ECG data sequences are listed under the condition $M=79$, $N=25561$, $W=30$, $R=0.95$. Where maps (a-d) are shown the results of normal ECG sample on Map_A , Map_G , Map_C , Map_T respectively. Maps (e-h) are shown the results of abnormal ECG sample on Map_A , Map_G , Map_C , Map_T respectively.

Sample results of ECG variant maps on different parameters

In Figure 8, six 2D maps are illustrated in the range of $M=70-100$, $N=25561$, $R=0.95$ for comparison, its scatter points are more compact with M value. Scatter points are mostly distributed between 0.2 to 0.6 region on X-axis.

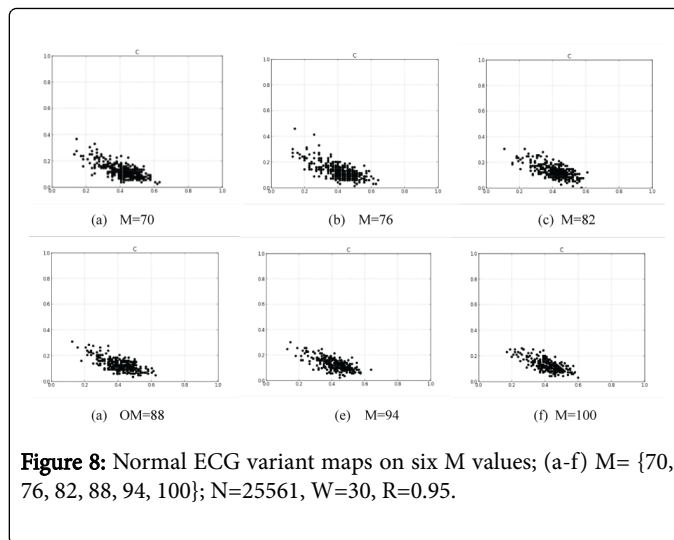


Figure 8: Normal ECG variant maps on six M values; (a-f) $M= \{70, 76, 82, 88, 94, 100\}$; $N=25561$, $W=30$, $R=0.95$.

In Figure 9, six 2D maps are illustrated in the range of $M=70-100$, $N=25561$, $R=0.95$ for comparison, scatter points illustrate arc-shaped distributions significantly different from normal cases.

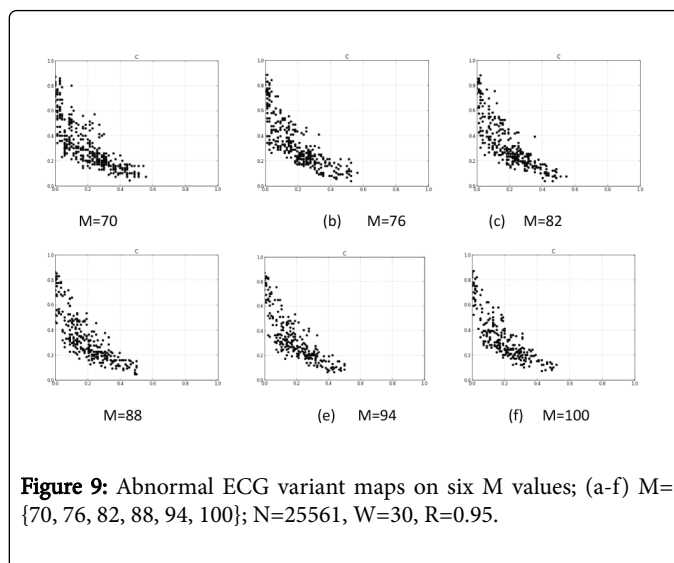


Figure 9: Abnormal ECG variant maps on six M values; (a-f) $M= \{70, 76, 82, 88, 94, 100\}$; $N=25561$, $W=30$, $R=0.95$.

Result's Analysis

Figure 7 contains eight maps (a-h). Four of them (a-d) are generated from a normal ECG data sequence and other four maps (e-h) are generated from an abnormal ECG data sequence on the parameters $M=79$, $N=25561$, $W=30$, $R=0.95$. Visual differences can be identified between normal and abnormal ECG maps. Among the eight maps (a-h), map (g) is showing a curved distribution significantly different from other maps (a-f) and (h) in Figure 6. So a Map_C as a selected projection from a Pseudo-DNA sequence may provide a better distinction between normal and abnormal ECG maps.

To illustrate visual effects of 2D maps under different control parameters, six 2D maps on normal ECG cases are generated in the region of $W=30$, $R=0.95$ and $M=70-100$, different distributions can be observed, and their visual distributions are shown in Figure 8. It is interesting to observe different maps when parameter M changed. A

larger M value makes a tighter distribution. Six maps have a main distribution range on 0.2-0.6 on X-axis.

Six 2D maps on abnormal ECG cases are selected in the range of $M=70-100$, $N=25561$, $W=30$, $R=0.95$ for comparison in Figure 9, Six maps have a main distribution range on 0.0-0.5 on X-axis. Significant differences can be identified compared with maps of normal ECG sample shown in Figure 8.

Conclusion

This paper proposes the ECG variant map system to transform ECG data sequences as variant ECG maps. Two ECG data sequences are selected as samples. Each ECG data sequence generates a pseudo-DNA sequence PV and four projective maps on {A, G, C, T} respectively. This system uses variant measurements and visualization method, to transform selected ECG data sequences as maps.

Using this type of multiple maps, it is convenient for multiple samples to make relevant processing and comparison. Each selected ECG data sequence has been pre-distinguished as either normal or abnormal case. Their ECG variant maps can be used to illustrate specific visual features to distinguish normal or abnormal ECG sample. The listed normal/abnormal ECG samples contain significant characteristics in their scatter point maps. This method may provide assistant information on further heart diseases diagnosis in exploring future medical ECG applications.

Acknowledgement

Thanks to the school of software Yunnan University, to the key laboratory of Yunnan software engineering for excellent working

environment. Financial supports to this project are provided by National Science Foundation of China (61362014) and the Yunnan Advanced Overseas Scholar Project (W8110305).

References

1. National Academy of Engineering, Grand challenges for engineering.
2. Pandey S, Voorsluys W, Niu S, (2012) An autonomic cloud environment for hosting ECG data analysis services [J]. *Future Generation Computer Systems* 28: 147-154.
3. Khair M (2012) A leadless wireless ECG measurement system and method for measuring of bio- potential electric activity of the heart. US 8838218 B2.
4. Kawamoto K, Houlihan C, Balas A, Lobach D (2005) Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *British Medical Journal* 330: 765-768.
5. Roob SJ, Meintzer AB (2014) Non-invasive fetal ECG analysis. *Peritoneal Dialysis International* 35: 1521-1536.
6. Li F, Xiang J, Lorenz Plot RR (2014) Beijing, People's Press edition ISBN 978-7-117-18767-1.
7. Sitzman DA, Johnson TL, Treacy ST (2014) System and method for signal quality indication and false alarm reduction in ECG monitoring systems: US Patent, US 8805482 B2.
8. Zheng J, Zheng C (2010) A framework to express variant and invariant functional spaces for binary logic. *Front Electr Electron Eng* 5: 163-172.
9. Zheng J, Zhang W, Luo J, Zhou W, Shen R (2013) Variant Map System to Simulate Complex Properties of DNA Interactions Using Binary Sequences. *Advances in Pure Mathematics* 3: 5-24.
10. Zheng J, Zheng C, Kunii TL (2013) Interactive Maps on Variant Phase Space, Emerging Application of Cellular Automata. *InTech Press* 6: 113-196.