Automated Risk Identification of Myocardial Infarction Using Relative Frequency Band Coefficient (RFBC) Features from ECG

Gohel Bakul* and U.S. Tiwary

Indian Institute of Information Technology, Allahabad, 211012, India

Abstract: Various structural and functional changes associated with ischemic (myocardial infarcted) heart cause amplitude and spectral changes in signals obtained at different leads of ECG. In order to capture these changes, Relative Frequency Band Coefficient (RFBC) features from 12-lead ECG have been proposed and used for automated identification of myocardial infarction risk. RFBC features reduces the effect of subject variabilty in body composition on the amplitude dependent features. The proposed method is evaluated on ECG data from PTB diagnostic database using support vector machine as classifier. The promising result suggests that the proposed RFBC features may be used in the screening and clinical decision support system for myocardial infarction.

1. INTRODUCTION

In recent years, computer intervention in medicine has been explored intensively. The computer intervention provides advantages like (1) automated analysis (2) quantitative measurement and quantitative complex analysis (3) availability of algorithms that find out hidden pattern or knowledge that could not be made out simply by looking at data by human or an expert and (4) avoidance of subjective variation. Thus many automated methods and algorithms have been developed and used in ECG analysis to evaluate cardiac disorders. One of the important cardiac disorders, Myocardial Infarction (MI) is the leading cause of death and disability in the world. Lack of blood supply (ischemia) to heart tissue, due to the blockade in the coronary artery cause myocardial infarction. It is a life threatening condition and early accurate diagnosis of MI or predicting risk of MI is very important. Electrocardiogram (ECG) is the widely used cheap diagnostic tool to investigate various cardiac abnormalities and it is a first line investigation in MI. In ECG, ST-segment changes and T wave inversion criteria are mainly used for acute MI detection, but they have low sensitivity range from 10 % to 50 % as mentioned) in different studies [1-4]. Another drawback of this criterion is its transient nature, which means it is only present in ECG at the time of ischemic attack. So finding the features that give information about MI or about risk of MI during emergency is quite useful. Lots of efforts have been made to increase the ECG sensitivity in the diagnosis of MI that include; Exercise Tolerance Test (ETT), HR-ST hysteresis, body surface potential map (BSPM), different QRS scoring system, ST segment changes [1, 5-7] and different machine learning based approaches mainly using ST segment information [5, 6]. These different approaches have their own advantages and limitations in terms of accuracy, convenience, automation, dataset availability, etc. In this work, our goal is the automatic extraction of new features from 12-lead ECG that should accurately discriminate between myocardial Infarction (MI) group and Healthy Control (HC) group with higher sensitivity.

Recording fluctuations in electrical potential of myocardial fibers (heart muscle) on body surface is called the Electrocardiogram (ECG) (Fig. **1**). Vector that represents magnitude and direction of electric field generated through heart is called "cardiac vector" [7] (Fig. **2**). In the ECG, amplitude recorded on a given lead depends on the projection of "cardiac vector" on that lead at that time. Hypertrophic and atrophic structural changes as well as electrical conduction of the heart are frequently associated with ischemic heart. These changes affect the magnitude and direction of normal cardiac vector [7] causing alteration in the amplitude and morphology of ECG waves in different leads. These distort the patterns in ECG waves on different leads, which may help to identify ischemic heart from healthy heart. But it is inappropriate to use the absolute amplitude changes or its dependent feature directly. In this paper, we will show that proposed features reduce the effect of body composition without using any specific body composition information.

In our previous study [8] we had proposed relative amplitude features of R and S peaks between two leads. In this paper, we extend this approach to spectral domain and RFBC features are proposed. Details of these features are described in subsequent section. Finally, RFBC features are evaluated on ECG data of MI and healthy control group from PTB diagnostic database using support vector machine.

2. METHODOLOGY

2.1. Relative Amplitude and Body Composition

The amplitude of ECG signal is affected by body impedance which depends on body composition. Body

*Address correspondence to this author at the Indian Institute of Information Technology, Allahabad (IIITA), Allahabad, 211012, India; Tel/Fax: 91-9696481302; Email: bcgohel@iiita.ac.in
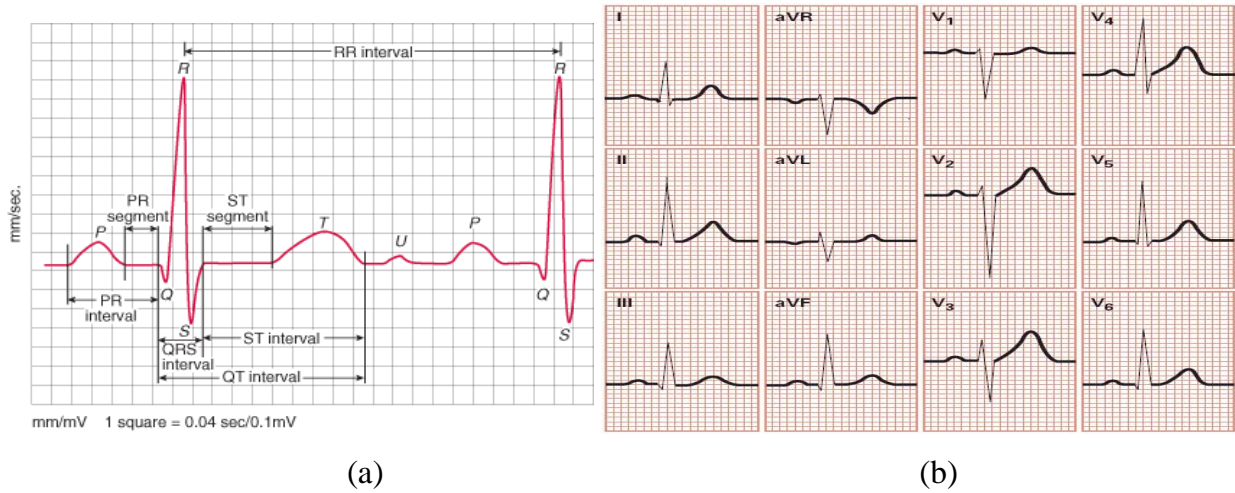
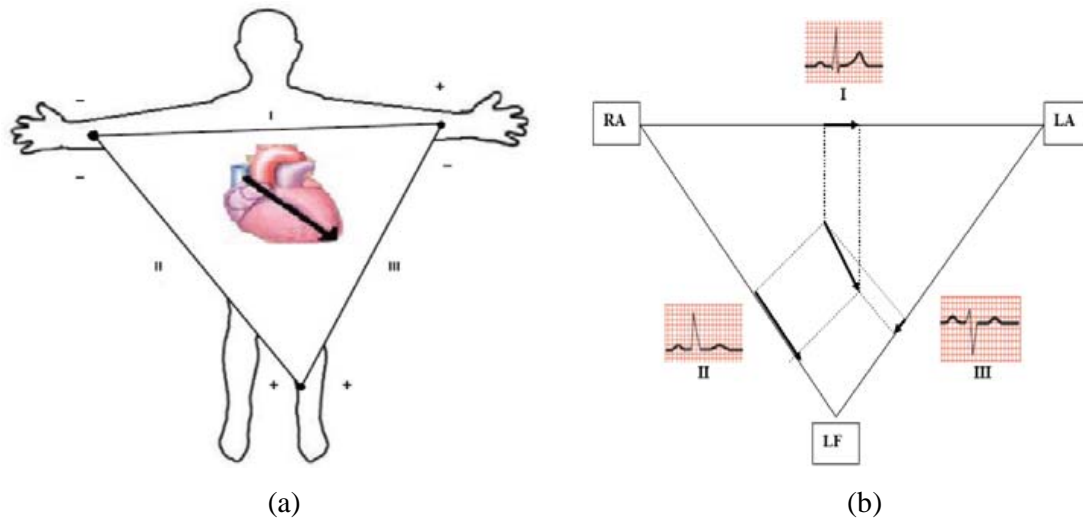**Fig. (1).** (**a**) Normal waveforms of ECG (**b**) 12-lead ECG signal.

Fig. (2). (**a**) Cardiac vector (black arrow in centre) (**b**), electrode connectivity of I, II, III lead and generation of ECG wave in different lead(s) through projection of cardiac vector.

composition variability across different persons gives undesirable effect on ECG amplitude and its dependent features. To reduce the body composition effect on ECG amplitude, different methods have been proposed in literature, which are mainly based on correlation of ECG amplitude with parameters that provide information regarding body composition, like age, sex, Body Mass Index (BMI), fat-free mass derived from bioelectric impedance method, skin fold thickness etc. [9-11]. Methods that use age, sex and BMI information are not very effective. The estimation of fat-free mass is a tedious job and requires a different instrument. In this paper, the relative amplitude based approach is adopted which does not require any body composition information.

Body mass (mainly body fat), distance between the electrodes or distance of electrodes from the heart characterizes the body impedance [9, 12]. Let us assume the body as structurally proportional in all directions from the heart for each subject. So the body mass is approximately proportional to the distance between electrodes and distance of electrodes from heart for each subject. Thus the body

impedance (R_i) exert on different leads are proportional to each other. So

$$\frac{R_i}{R_j} = k_{ij}, i \neq j \tag{1}$$

Where i, j are different leads of ECG and k_{ij} can be considered as constant for selected pair of leads (i, j).

It should be noted that proportion of body parts or composition (anthropometric measures) varies from one person to another and hence k_{ij} also varies. But the amount of variation is very small compared to absolute body composition variations. Thus the relative coefficients reduce the body composition effect on ECG considerably.

Voltage (V) of ECG signal depends on electrical current through heart (C) and body impendence (R). So

$$V = CR \tag{2}$$

$$\frac{V_i}{V_j} = \frac{C_i R_i}{C_j R_j} \tag{3}$$

$$\frac{V_i}{V_j} = \frac{k_{ij}C_i}{C_j} \tag{4}$$

As k_{ij} is constant for given pair of leads (i, j), from eqn. 4 the relative amplitude between two leads do not depend on body resistance (R) It is also true for frequency domain as

$$F(k \bullet x) = k \bullet F(x)$$

Where F represents the Fourier transform.

Furthermore, to account for the variability (person to person and male to female) in distribution of fat between limb and chest area, we treated limb leads and chest leads separately.

2.2. Representation of ECG Signal in Terms of RFBC Features

In our previous paper we had defined the relative amplitude features [8] as follows:

It is the relative amplitude between the R or S peak (whichever is maximum) of one lead and R or S peak (whichever is maximum) of another lead of ECG signal (eqn. 5 and Fig. **3**). Relative amplitude features (eqn. 6) [8] are calculated for different combination of limb leads (I,II, III, aVR, aVL, AVF) and different combination of chest leads $(V_1, V_2, V_3, V_4, V_5, V_6)$ as follow

$$A_{ij} = \frac{(a_i - a_j)}{(a_i + a_j)} \quad i \neq j \tag{5}$$

$$a_i = argmax\{|R_i|, |S_i|\}$$

$$a_j = argmax\{|R_j|, |S_j|\}$$

here $i, j = 1$ to 6 for limb leads and $i, j = 7$ to 12 for chest leads

$$V_{ij} = [A_{ij} \, d_i \, d_j] \text{ where } d = \begin{cases} 1 \text{ if } R > S \\ -1 \text{ if } R < S \end{cases} \tag{6}$$

The relative amplitude only considers the R and S peaks and the changes associated with other ECG waveforms are missed. Further, measuring amplitude of various characteristic ECG waveforms and the signal segmentation are difficult and error prone. So we have extended the relative amplitude approach to spectral domain, as the

different waveforms of ECG have different frequency range and cover more information apart from R and S peaks. It can also capture the changes in spectral properties of ECG waveforms. So we have proposed the relative band frequency coefficient (RFBC) features. It is the relative amplitude of different frequency bands between two leads and can be calculated as described below.

Standard-12 leads ECG signal (x) from each subject was used. To remove low frequency baseline drift, ECG signal x_i (where $i = 1$ *to* 12 *leads*) was passed through high pass forward-backward FIR filter with 2Hz cut off frequency (Fig. **4b**). ECG waves like R and S waves as well as T and inverted T waves have opposite directions but same frequency spectra. To separate them ECG signal x_i was divided into positive peak signal x_i^p (above the baseline) and negative peak signal x_i^n (below the baseline) (Fig. **4c** and eqn. 7).

$$x_i^p = \begin{cases} x_i, & x_i > 0 \\ 0, & x_i < 0 \end{cases} \text{ and } x_i^n = \begin{cases} 0, & x_i > 0 \\ x_i, & x_i < 0 \end{cases} \tag{7}$$

Fourier spectrum X_i^p and X_i^n was calculated by taking fast discrete fourier transform of x_i^p and x_i^n respectively (eqn. 8). As most of the energy of ECG waveforms is concentrated in frequency up to 40 Hz only, the frequency spectrum up to 40 Hz has been considered in the analysis (Fig. **4d**). This also removes high frequency noise and power line noise (50 Hz), Frequency spectrum was broken into S equally spaced bins (frequency bands, sized (40-2)/S Hz)) and the cumulative amplitude of each bin b_m was calculated (Fig. **4e** and eqn. 9)

$$X_i^p[k] = \sum_{l=0}^{N-1} x_i^p[l] . e^{-\frac{j2\pi lk}{N}} \tag{8}$$

Where $k = 0$ to N-1 and N is total no. of samples

$$W_i^p[b_m] = \sum_{k \in b_m} X_i^p[k] \tag{9}$$

Where $b_m = 1$ to S bin

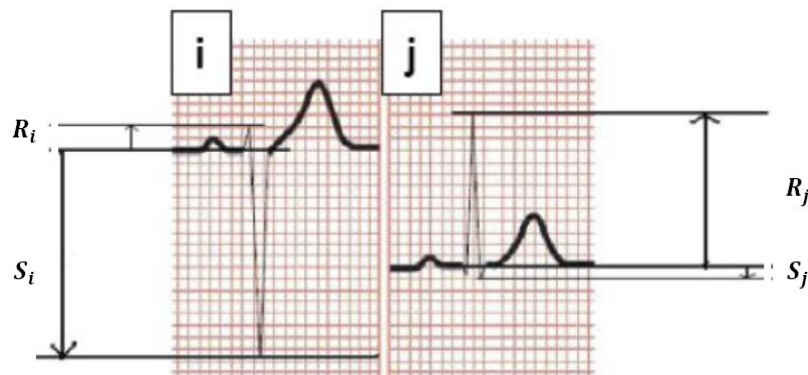RFBC features were calculated for each bin for each pair of leads (eqn. 10).

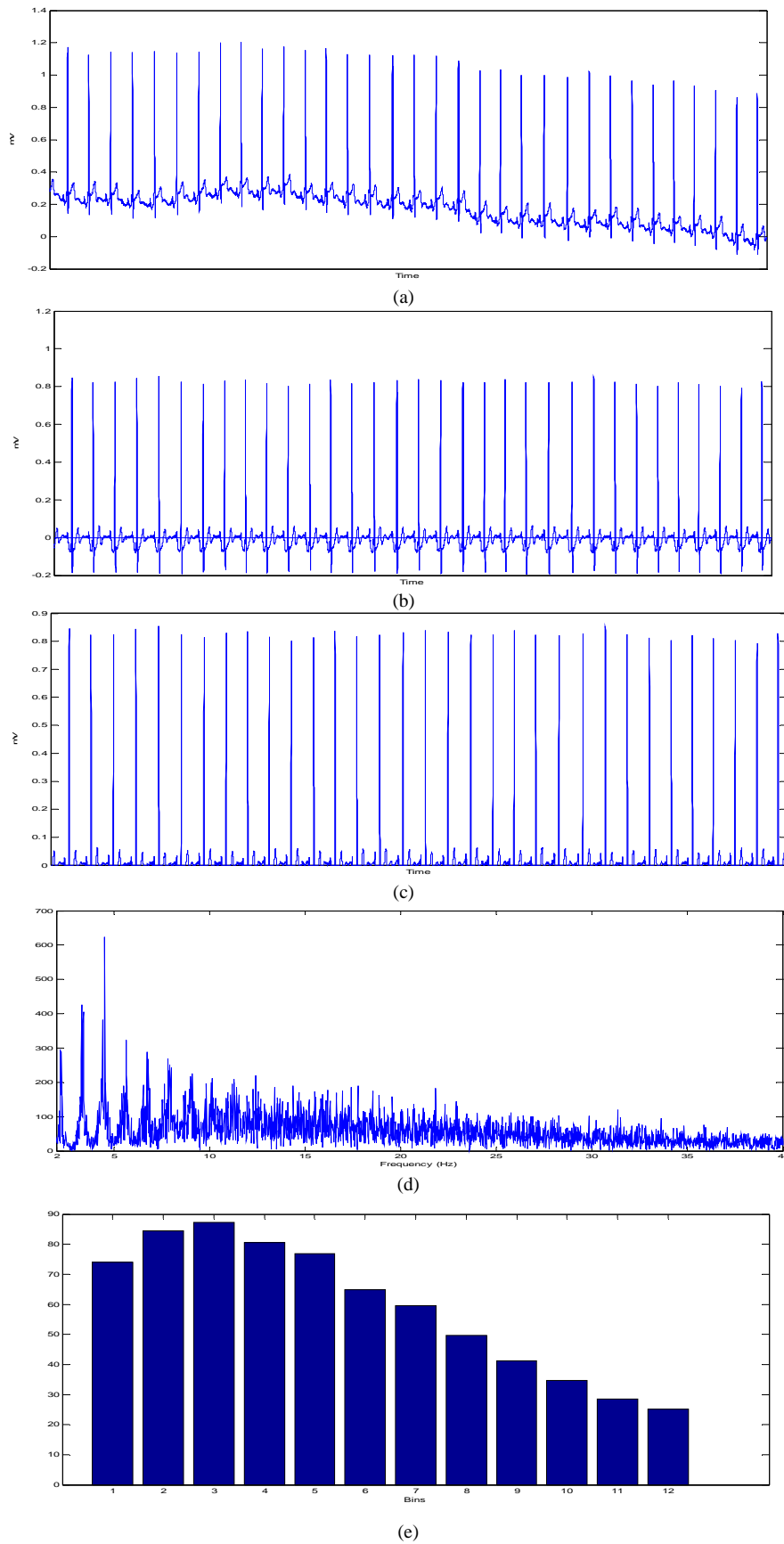**Fig. (3).** R and S peak in two different ECG lead.

Fig. (4). (**a**) ECG signal (**b**) baseline corrected signal (**c**) positive peak signal x_i^p (**d**) DFT spectrum of x_i^p (2 Hz to 40 Hz) (**e**) value of 12 equally spaced bins of DFT spectrum of x_i^p.

$$RFBC[b_m]_{ij}^p = \frac{W_i^p[b_m] - W_j^p[b_m]}{W_i^p[b_m] + W_j^p[b_m]} \tag{10}$$

here, for limb leads (L) $i = 1$, $j = 2$ to 6 (5 pair(i, j))

For chest leads (C) $i = 7$, $j = 8$ to 12 (5 pair(i, j))

Calculation of RFBC features of x_i^n is also done in the same way. For one subject, total number of RFBC features with S number of bins is $S*2*10$. We have taken $S = 12$ in this analysis.

3. CLASSIFICATION BASED ON RFBC FEATURES

Various machine learning techniques have been intensively employed in clinical decision support system particularly artificial neural network [5, 6] and support vector machine [13]. SVM is preferred over ANN as it yields global and unique solution, have a simple geometrical interpretation, use structural risk minimization, and it is less prone to over fitting etc. Thus we have used SVM for classification.

SVM is the novel classification technique based on VC dimension theory [13, 14]. The central idea of SVM is to find hyperplanes (H$_1$, H$_2$) with largest possible margin M that separates the data points $X = \{x_1, x_2 \ldots x_i\} \in R^d$ in two classes $y_i \in \{1, -1\}$ by finding a weight vector $w \in R^d$ and offset b. The hyperplanes (H$_1$, H$_2$) are pushed apart until they meet corresponding class of data points which are called support vectors. Since

$$M = \frac{2}{\|w\|} \tag{11}$$

$$((x, w) + b) \geq +1 \text{ for } y_i = +1 \tag{12}$$

$$((x, w) + b) \leq -1 \text{ for } y_i = -1 \tag{13}$$

Maximization of margin M requires minimization of $\|w\|$. So

$$\min_{w \in R^d} (\mathrm{w}, \mathrm{w})$$

subject to $y_i((x, w) + b) - 1 \geq 0 \ \forall_i$

One of the variants of this algorithm is soft margin SVM which allows minimum amount of misclassification or error for dataset that are not linearly separable by hyper planes and lead to following optimization problem

$$\min_{w \in R^d} (\mathrm{w}, \mathrm{w}) + C \sum_{i=1}^{n} \xi_i^2 \tag{14}$$

subject to $y_i((x, w) + b) \geq 1 - \xi_i$ and $\xi_i \geq 0 \ \forall_i$

Where parameter ξ are called slack variables that represent total training error. The tradeoff between training error and maximal margin M is controlled by the regulation parameter C. If C is set too large, classifier is over fitted and if it is set too small, classifier is under fitted, which means large training error. So proper value of C is important for the performance of the classifier.

In the non-linear SVM, first the input vector x_i is mapped to higher dimension feature space F using mapping function $\varphi(x)$ and then a linear classification is performed in this feature space F. In practice kernel function $K(x_i, x_j)$ is used, as it offers implicit mapping into inner product feature space (reproducing kernel Hilbert spaces), $K(x_i, x_j) = (\varphi(x_i), \varphi(x_j)) = (x_i, x_j)^D$. This means explicit mapping $\varphi(x)$ is not required, as the inner product provides sufficient mapping. Mostly used kernels are polynomial, gaussian radial basis function (RBF), exponential radial basis function, sigmoid kernel (*tanh*) etc. In this study, a sigmoid kernel was used.

4. RESULTS AND DISCUSSION

ECG signal of healthy (HC) subjects and subjects with MI was downloaded from online PTB diagnostic database [15]. This dataset has 12 standard leads for each subject, sampled at 1000 Hz sampling frequency. RFBC features were calculated on 52 HC and 104 MI subjects and classification was done using support vector machine with one leave out cross validation. Value of S is selected as 12, providing 3.16 Hz wide frequency band bin. By manually optimizing the parameters and kernel of SVM the following results (Table **1**) are obtained.

To compare the accuracy of RFBC features, we have also tabulated our previous results using relative amplitude features [8]. The relative amplitude features yields 71.07 % and 74.35% accuracy with Back-propagation neural network and Support vector machine based classifications respectively, while RFBC yields accuracy of 85.23 %, sensitivity 85.57 % and Specificity: 83.97 % with support vector machine (Table **1**). Thus RFBC features outperform the relative amplitude features. Increase in accuracy with RFBC is due to the fact that it captures information regarding all characteristic waves in terms of frequency spectrum while the relative amplitude features capture information based on only R and S. However, it is worth mentioning that the measurement of classification accuracy of acute MI is only possible when ECG data is labeled simultaneously with confirmatory test for acute MI e.g. blood troponin level. The PTB diagnostic database is not labeled with confirmatory test. Further, features may represent chronic structural changes associated with ischemic heart instead of ischemic event directly. So the results mentioned above may not represent sensitivity and specificity of acute MI in true sense. It is better to infer that the proposed features represent the risk of having MI rather than true sensitivity and specificity of acute ischemic event.

Different ECG features e.g. prominent Q-wave and ventricular hypertrophy changes which are used presently to identify risk of MI have low sensitivity [3, 9, 16]. Prominent Q-wave in ECG also denotes old myocardial infarction. Methods representing ventricular hypertrophic (VH) changes have low sensitivity because of the effect of body composition variation across the subjects [9, 16]. Exercise Tolerance Test [1] is also useful in identifying risk of MI but cannot be performed in emergency. Hence our proposed features, as it has high accuracy, may be beneficial in risk identification of MI at the time of emergency as well as for regular screening of MI.

Table 1.   Result of MI and HC ECG Signal Classification Based on Relative Amplitude Features and RFBC Features

Feature	Classifier	Validation Approach	Accuracy %
Relative Amplitude [16]	Back-Propagation Neural Network (3 layer - 30:6:1)	3:2 train-test ratio ($200 \times$ randomly chosen datasets)	Average : 71.07 % Best case: 81.81 %
Relative Amplitude	SVM Kernel: Sigmoid $C = 100$	One-leave out cross validation	Accuracy: 74.35 % Sensitivity: 69.23 % Specificity: 84.61 %
RFBC	SVM Kernel: Sigmoid $C = 100$	One-leave out cross validation	Accuracy: 85.23 % Sensitivity: 85.57 % Specificity: 83.97 %

Furthermore, this method also reduces the body composition effect on features without using any body specific information like age, sex, body weight, BMI and body impedance. Thus this method may be helpful in the diagnosis of ventricular hypertrophies through ECG.

5. CONCLUSION AND FUTURE WORK

Here, we have presented the new features, RFBC from 12-lead ECG that discriminate MI subject from healthy subject automatically with good accuracy. The main characteristic of this method is the reduced effect of inter-subject body composition variation. It might be used for the identification of the risk of MI or in clinical decision support system for MI effectively. In future, we would like to evaluate the proposed features in identification of ventricular hypertrophy.

REFERENCES

[1]   N. Herring and D.J. Paterson, "ECG diagnosis of acute ischemia and infarction: past, present and future", *QJM*, vol. 99, no. 4, pp. 219-230, 2006.

[2]   P. H. Ginn and Barbara, "How accurate is the use of ECGs in the diagnosis of myocardial infarct?", *J. Fam. Pract.,* vol. 55, no. 6, June, 2006.

[3]   F.M. Asch, Shah S, C. Rattin, S. Swaminathan, A. Fuisz and J. Lindsay, "Lack of sensitivity of the electrocardiogram for detection of old myocardial infarction: a cardiac magnetic resonance imaging study", *Am. Heart J.,* vol. 152, no. 4, pp. 742-748, Oct. 2006.

[4]   J. Zimmerman, R. Fromm, D. Meyer, A. Boudreaux, C. Wun, R. Smalling, B. Davis, G. Habib and R. Roberts, "Diagnostic Marker Cooperative Study for the Diagnosis of Myocardial Infarction", *Circulation,* vol. 99, pp. 1671-1677, 1999.

[5]   H. Haraldsson, L. Edenbrandt and M. Ohlsson "Detecting acute myocardial infarction in the 12-lead ECG using Hermite expansions and neural networks", *Artif. Intell. Med.*, vol. 32, pp. 127-136, 2004.

[6]   B. Hedén, H. Öhlin, R. Rittner and L. Edenbrandt "Acute myocardial infarction detected in the 12-lead ECG by artificial neural networks", *Circulation*, vol. 96, pp. 1798-1802, 1997.

[7]   G. Cliford, F. Azuaje and P. McSharry. *Advanced Methods and Tools For ECG Data Analysis,* Artech House Publishers, Oct. 2006.

[8]   B. Gohel, U.S.Tiwary and T.Lahiri, "Relative Amplitude based Features of characteristic ECG-Peaks for Identification of Coronary Artery Disease", *in proceeding of International Conference on Intelligent Human Computer Interaction, Jan 20-23, 2009 Allahabad, India.* Springer, 2009.

[9]   O. Tochikubo, E. Miyajima, T. Shigemasa and M. Ishii "Relation between body fat–corrected ECG voltage and ambulatory blood pressure in patients with essential hypertension", *Hypertension,* vol. 33, pp. 1159-1163, 1999.

[10]   P. M. Okin, S. Jern, R. B. Devereux, S. E. Kjeldsen and B. Dahlöf, "Effect of Obesity on Electrocardiographic left ventricular hypertrophy in hypertensive patients", *Hypertension*, vol. 35, pp. 13-18, 2000.

[11]   H. C. Lukaski, W. W. Bolonchuk, C. B. Hall and W. A. Siders "Validation of tetrapolar bioelectrical impedance method to assess human body composition", *J. Appl. Physiol.,* vol. 60, pp. 1327-1332, 1986.

[12]   F. Morris, J. Edhouse, W. J. Brady and J. Camm, *ABC of clinical ectrocardiography,* BMJ book, 2003.

[13]   D. Confortia and R. Guido, "Kernel-based Support Vector Machine classifiers for early detection of myocardial infarction", *Optim. Methods Softw.*, vol. 20, pp. 401-413, April 2005.

[14]   S. Sanei and J.A. Chambers, *EEG signal processing,* John Wiley & Sons Ltd, 2007.

[15]   A. L. Goldberger, A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. Peng and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signal", *Circulation*, vol. 101, no. 23, pp. 215-220, June 2000.

[16]   D. Levy, S. B. Labib, K. M. Anderson, J. C. Christiansen, W. B. Kannel and W. P. Castelli, "Determinants of sensitivity and specificity of electrocardiographic criteria for left ventricular hypertrophy", *Circulation,* vol. 81, pp. 815-820, 1990.