Feature Extraction and Classification of ECG Signal using Independent Component Analysis

Anju G. Nair, Paul Thomas

Mar Baselios Engineering College, Trivandrum, Kerala, India

Abstract— Electrocardiogram (ECG) is the electrical trace of the mechanical and electrical activities of the heart. It provides information on the rhythm and functioning of the heart. ECG feature extraction and classification is the core of diagnosing cardiac disorders. Many techniques are reported for ECG feature extraction and classification. These techniques include Wavelet transforms, Principal Component Analysis (PCA), Artificial Neural Network (ANN), Self Organizing Map (SOM) and Support vector machine (SVM). Though wavelet transforms are powerful tool for ECG feature extraction, they are limited by the choice of the wavelet. Wavelet transform, along with ANN and a training of 20%, has reported a performance of 90% with regard to specificity, sensitivity and accuracy. The Performance levels obtained with other techniques are lesser. In this paper a novel scheme for ECG feature extraction and classification is proposed, based on ICA (Independent Component Analysis) and SVM. ICA is a statistical technique which linearly transforms the observed random data into components which are independent from each other. SVM binary classifiers are sample based statistical learning algorithms that construct a maximum margin separating hyper plane. A higher level of performance is demonstrated for all three performance parameters, with accuracy more than 95%. Standard ECG signals available in MTB data base is used in this study.

Keywords— *Artifical Neural Network (ANN), Independent Component Analysis (ICA), Support Vector Machine(SVM).*

I. INTRODUCTION

ECG contains a plethora of information on the normal and pathological physiology of the heart. ECG is clinically used to ascertain the health of the heart. It is difficult for a physician to visually analyze an ECG and arrive at conclusion. This calls for automated techniques for the analysis and classification of ECG signal.

The most widely observed cardiac disorder is arrhythmias, the most lethal arrhythmia being Ventricular fibrillation (VF). Ventricular fibrillation is the result of uncoordinated contraction of ventricular muscles, making the ventricles fever instead of contracting.

There are different techniques to extract features from ECG. They include wavelet transform, spectral analysis, PCA and various other complexity measurement schemes [10]. Performance achieved with wavelet transform is limited by the choice of the wavelet. PCA can bring out beat to beat changes in ECG features. Techniques for classification of ECG features include ANN, SOM and SVM. ANN and SVM are machine learning techniques, where the weights get refined with learning. Figure1 shows the entire architecture of feature extraction and classification.

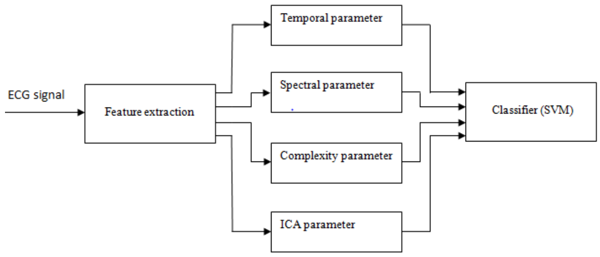

Fig.1: Block diagram for feature extraction and classification.

The present study aimed at building a high performance ECG classifier with ICA used for feature extraction and SVM used for classification. ECG signals from MIT database is used in this study. This study on ECG signals with VF, has demonstrated that ICA with SVM is a better classifier for ECG signals.

II. ECG FEATURE EXTRACTION

ECG signal is first pre processed using a moving average filter of order 5 to remove high frequency artifacts such as those from muscles and EMI.

This study is based on eight features extracted from ECG. 1) Exponent Crossing Count (ECC), 2)Threshold Crossing Interval (TCI), 3)Threshold crossing sample count (TCSC), 4) ECG Temporal area (TA), 5) ECG Spectral area (SA), 6) Sample entropy (SE), 7) Median frequency (MF) and 8) features from ICA. First three features are extracted from an eight seconds stretch of ECG.

A. *Exponent crossing count (ECC)*

ECC is extracted using Standard Exponential (SEA) algorithm. The sample with maximum amplitude is identified in the ECG signal. This maximum amplitude is M and the corresponding time instant is t_m. From this maximum point, exponentially decreasing curves are fitted to both sides of the sample as

$$As(t) = Mexp\left(-\frac{|t-t_m|}{\tau}\right) \quad (1)$$

τ is a time constant, set to 3s in this study.

ECC, the number of crossings per minute, is

$$N = 7.5 * n \quad (2)$$

n is the number of crossings of the curves with the ECG.

A modification to SEA is proposed [1] with exponentially decreasing curves fitted to both sides of each QRS peak. Average of the individual counts is used to compute ECC. The modification to STE is termed as modified exponent (MEA).

A. Threhold Crossing Count (TCC)

TCI [1] is the average threshold crossing interval. The 8s original is divided into 8 blocks of 1s each. The threshold value in a block is set at 20% of maximum sample value in that block. This is illustrated in fig 2.

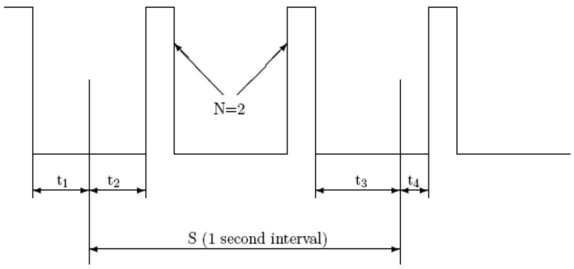

Fig.2. TCI illustration

$$TCI = \frac{1000}{(N-1)+\frac{t_2}{t_1+t_2}+\frac{t_3}{t_3+t_4})}[ms] \quad (3)$$

Here, N is the sampling rate and t_1, t_2, t_3, t_4 are shown in fig 2.

B. Threhold Crossing Sample Count (TCSC)

TCSC [1] is the number of threshold crossings of the ECG samples in the window. The value of threshold is suitably taken based on empirical studies. The algorithm is suitable for VF detection and SR.

C. Mean Absolute Value (MAV)

MAV is the mean of absolute values of ECG samples in the window.

D. VF Algorithm

First, compute mean frequency of the ECG signal and then apply narrow band elimination filter to this region.

E. Spectral bands

A Fast Fourier Transform (FFT) is carried on ECG signal to get the different frequency bands [1].

F. ECG Temporal Area

Here the intersection area of an ECG signal and its delayed version is computed [5].The signal and its delayed version (by 0.5s) plotted on orthogonal axis and checked for the overlapping area. This done on a 40×40 grids to ease computations .The measure is given by the relation:

$$d = \frac{number\ of\ visited\ boxes}{number\ of\ all\ boxes} \quad (5)$$

G. ECG Spectral Area

Here the intersection area of an ECG signal and its Hilbert transform version is computed [6].The signal and its Hilbert transform version plotted on orthogonal axis and checked for the overlapping area. This done on a 40×40 grids to ease computations .The measure is given by the relation:

$$d = \frac{number\ of\ visited\ boxes}{number\ of\ all\ boxes} \quad (6)$$

H. Median Frequency (MF)

MF [7] is the central frequency of the ECG.

I. Features from ICA

ICA provides a linear representation of non gaussian data, in which the original vector is a linear combination of their basis vector. ICA has an advantage that it uses higher order statistics to reduce higher order dependencies rather than second order statistics in PCA.

The first step in ICA is to make x a zero mean variable, by subtracting the expected value from x. The columns of the transformation matrix of a signal are its basis function.

FAST ICA learning algorithm can be used to generate this basis function. The algorithm as follows:

1. Choose an initial weight w
2. The weight is updated using the eqn:

$$w = w+/||w^+|| \quad (7)$$

Where, w+ = $E\{xg(w^T\ x)\}-E\{g(w^T\ x)\}w$. Repeat the above steps until the weight is updated correctly.

III. SVM CLASSSIFIER

SVM is a binary classifier and it is used to solve regression tasks. SVM is a linear extension of perceptron based learning technique .The SVM has the advantage that it minimizes the empirical classification error and maximises the geometric margin. In the case of non linear relationship modeling, they build a boundary that separate data using a linear hyper plane.

A. Defining the hyperplane

We use the pre-classified signals for training the SVM. These signals belong to two classes with VF and without VF. The features extracted from these signals are used to define the hyperplane separating both the sets. The criterion for this classification is maximum Euclidian distance from hyperplane.

B. Training Algorithm

First set the training data chosen after training set selection procedure. Generate the training pairs based on attribute extracted from the signal. SVM is trained according to training pairs using the function svm.train in MATLAB.

IV. PERFORMANCE EVALUATION

The performances of classification are specified in terms of sensitivity (SE) and specificity (SP). SE is the percentage of correct detection of arrthymias condition, in this case VF.

$$SE = \frac{TP}{TP+FN} \tag{16}$$

SP is the percentage of correct detection of Non disease condition.

$$SP = \frac{TN}{TN+FP} \tag{17}$$

Where TP represents the number of true-positive decisions, FN the number of false-negative decisions, TN the number of true negative decisions, and FP the number of false-positive decisions.

V. RESULTS

The ECG signal from MIT database is preprocessed and the signal is divided into a set of non overlapping samples of length 8s. Fig.3 shows the preprocessed ECG signal by taking the sampling frequency Fs=250.

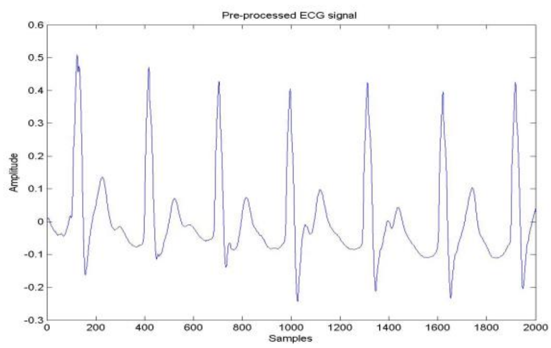

Fig.3.Preprocessed ECG signal

Typical features extracted from a signal with VF are shown in TABLE I and feature extracted from a signal with Non-VF are shown in TABLE II.

TABLE I. ANALYSIS OF ECG PARAMETER FOR VF CASE

Parameter	Score
TCSC	9.500
TCI	1.460
STE	2.125
MAV	0.239
MEA	0.025
M	0.866
A2	0.803
FM	2.682
VF	0.512
SE	0.394
TA	0.486
SA	0.502

TABLE II. ANALYSIS OF ECG PARAMETER FOR Non-VF CASE

Parameter	Score
TCSC	4.333
TCI	1.994
STE	0.875
MAV	0.103
MEA	0.000
M	1.707
A2	0.483
FM	1.707
VF	0.609
SE	0.273
TA	0.260
SA	0.340

Classification of the extracted features is tried in three different methods. Method1 temporal, spectral and complexity parameters are used. In method2 the classification is based on feature extracted using ICA technique. Method3 is a combination of method1 and method2. The results obtained in the following are tabulated in TABLE III below.

TABLE III. COMPARISON OF PARAMATER PERFORMANCE

Features trained	Sensitivity (%)	Specificity (%)	Accuracy (%)
Method1	100.00	83.33	90.00
Method2	75.00	100.00	90.00
Method3	100.00	98.00	99.20

The measure of SE, SP and accuracy was calculated for above three methods. Method1 shows that SE, SP and accuracy in the range above 80%. Method2 shows improved SP and accuracy but SE lacks below 80%. Method3 shows improved performance compared to method1 and method2.

VI. CONCLUSIONS

In this paper, we address ECG feature extraction and classification system using a combination of ICA and SVM. ICA is a technique in which the observed data are linearly transformed into components that are linearly independent. SVM classifiers have been extensively used in ECG signal for better classification of VF. To conclude, the obtained results showed that using ICA along with temporal, spectral and complexity parameters, we can able to attain accuracy about 99.2%. As a future scope we can do this for feature extraction and classification for more number of arrthymias.

ACKNOWLEDGMENT

The authors acknowledge the support of the Mar Baselios College Of Engineering And Technology In This Research.

REFERENCES

[1] Felipe Alonso-Atienza, Eduardo Morgado, Lorena Fern´andez-Mart´ınez, Arcadi Garc´ıa-Alberola and Jose´ Luis Rojo-A´lvarez, "Detection of Life-Threatening Arrhythmias Using Feature Selection and Support Vector Machines", *IEEE Trans. Biomed. Eng.*, vol.61, no. 3, Mar. 2014.

[2] S. Kuo and R. Dillman, "Computer detection of ventricular fibrillation,"in *Proc. Comput. Cardiol.*, 1978, pp. 2747–2750.

[3] S. Barro, R. Ruiz, D. Cabello, and J. Mira, "Algorithmic sequential decision making in the frequency domain for life threatening centricular arrhythmias and aimitative artifacts: A diagnostic system," *J. Biomed. Eng.*, vol. 11, no. 4, pp. 320–328, Jul. 1989.

[4] H. Li, W. Han, C. Hu, andM.-H.Meng, "Detecting ventricular fibrillation by fast algorithm of dynamic sample entropy," in *Proc. IEEE Int. Conf. Robot. Biomimet.*, Dec. 2009, pp. 1105–1110.

[5] A. Amann, R. Tratnig, and K. Unterkofler, "Detecting ventricular fibrillation by time-delay methods," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 1, pp. 174 –177, Jan. 2007.

[6] A. Amann, R. Tratnig, and K. Unterkofler, "A new ventricular fibrillation detection algorithm for automated external defibrillators," in *Proc Comput. Cardiol.*, Sep. 2005, pp. 559–562.

[7] R. Dzwonczyk, C. Brown, and H.Werman, "The median frequency of the ECG during ventricular fibrillation: Its use in an algorithm for estimating the duration of cardiac arrest," *IEEE Trans. Biomed. Eng.*, vol. 37, no. 6, pp. 640–646, Jun. 1990.

[8] Hyv¨arinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.

[9] Y. W. Chen and C. J. Lin, "Combining SVMs with various feature selection strategies," in *Feature extraction: Foundations and Applications*, I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Eds.Secaucus, NJ,USA: Springer-Verlag, 2006.

[10] Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 75–81, Jan. 1976.

Anju G. Nair received his B.Tech Degree in Applied Electronics and Instrumentation Engineering from Marian Engineering College, Trivandrum affiliated to kerala University in 2013, and is presently pursuing M.Tech (Signal Processing) degree from Mar Baselios Engineering College, trivandrum. His areas of interest are DSP Design and Testing of IC, Biomedical Signal Processing etc.