ECG Analysis Using DWT and Wavelet Coefficient to Reduce the Feature and SVM-ICP for Classification and Matching

Janan A. Mahdi
Department of Control and System Engineering,
University of Technology,
Baghdad, Iraq.
jananmahl@yahoo.com,
60011@uotechnology.edu.iq

Received on: 04/10/2017
Accepted on: 01/03/2018
Published online: 25/8/2018

Abstract - *The Electrocardiogram (ECG) considered as one of the important issue in the medical field (hospitals and clinics), which is used to represent the health of a heart. Increasing patients of heart has supposed to design an automatic computerization technique to classify various abnormalities of the heart activities; to reduce the analysis time and detection mistakes. This research focusing on achieve high performance of classifying abnormal ECG by applying different methods. The first method is Discrete Wavelet Transform (DWT) with 4-level to transform the ECG signal and extract the feature extraction and Wavelet Energy (WE) during feature extraction as feature vector. In classification phase has used Support Vector Machine (SVM) to train datasets and classify the test samples, in matching phase, find closest vector of test to the training datasets method has used by applying Iterative Closest Point (ICP).*

Keywords: ECG, DWT, Wavelet Energy, SVM, ICP

How to cite this article: J.A. Mahdi, "ECG Analysis Using DWT and Wavelet Coefficient to Reduce the Feature and SVM-ICP for Classification and Matching," *Engineering and Technology Journal*, Vol. 36, Part A, No. 8, pp. ٩٢٥-٩٢٩, 2018.

1. Introduction

The basic of electrocardiography (ECG) is came from electrical activities of the heart rhythms recorded over time. ECG signal has observed, investigated, and analyzed by specialist in cardiology [1]. The observation method of analysis ECG leads sometimes to mistakes and takes time to recognition and classification the tested sample, for these reasons a computerized an ECG signals been designed, in addition to that, the early analysis of cardiac arrhythmia helps to decide the right anti-arrhythmic drugs [2]. ECG datasets have collected from Physionet which designed a web-service to download ECG data in many format and one of these format is .mat where has been depended in our work. Also, includes more than 50 collections of cardiopulmonary, neural, and other biomedical signals in several of cases such as sudden cardiac death, congestive heart failure, epilepsy, gait disorders, sleep apnea, and aging. These data has collected from a wide range of studies, as developed and contributed by members of the research community [3]. Nurul et al. [4]; they have proposed second order dynamic system (SODS) technique in feature extraction of ECG signals. Also, using Hybrid-recurrent Network (HRN). Khalil and Khalil [5]; they have been proposed a novel technique for ECG classification based on Legendre polynomial, simple logistic and KNN has applied to classify the ECG for normal and abnormal. The total classification accuracy is 90.0%. Czarina and Jastine [6], in this proposed work a comparative study focused on ECG classification between ANFIS and SVM to classifying normal and abnormal heartbeats. The performance of DWT-ANFIS and DWT-SVM in term of total classification accuracy is 90.0%. Aljafar et al. [7]; in this paper, they present a method to classification of multi-lead ECG signals for normal or abnormal using CSP as the feature extractor. The work consists of two main stages: CSP-based feature extraction and classification. There are three classifiers were employed in the classification stage: LDA, NB, and SVM. The total classification accuracy is 88.2%.

2. Proposed Work

The main phases of proposed work are illustrated in Figure 1; these phases of proposed work will be discussed in details in below:

I. Collected data

ECG recorded data has been collected from Physio Net web service, and the number of database of ECG is 48 recordings of 30-min, and there is 23 recordings for clinically arrhythmia abnormal cases [6]. In our study, have collect 6-cases/heart diseases, 20 samples for each case, and the duration of recording of each samples has been trim in 1min. Figures 2 shown the samples for each case.

https://doi.org/10.30684/etj.36.8A.13
2412-0758/University of Technology-Iraq, Baghdad, Iraq

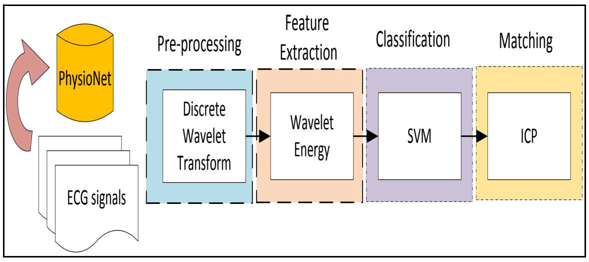

Figure 1: main steps of proposed work

II. Preprocessing using DWT

After collecting data of ECG, applied Discrete Wavelet Transform to convert an ECG signal from time domain to frequency domain [8]. The type of DWT selected in this work is DB4 mother wavelet with 4-level. The structure of DWT of proposal work shown in Figure 3. Where x is the ECG signal, cAs is the low-pass for each level of DWT and cDs is the high-pass filter for each level of DWT, C is wavelet decomposition vector, and L is bookkeeping vector, and Len is the length of each coefficient. The total length of all details coefficients equal to the length of original signal, which is (7200) samples in minute. Below a DWT algorithm represented in pseudo code form.

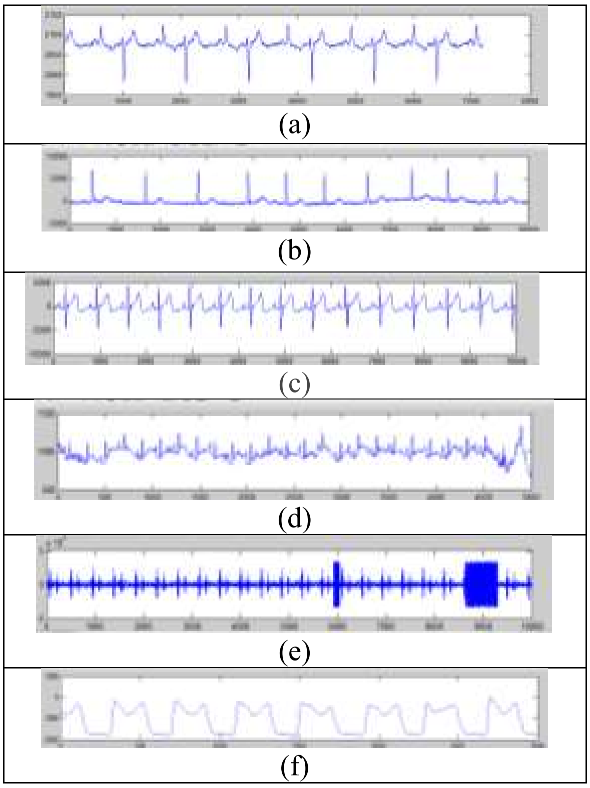

Figure 2: Abnormal ECG signals, (a) AAMI EC13, (b) Termination of Atrial Fibrillation, (c) coronary artery disease, (d) fetal scalp electrograms and uterine muscular activity, (e) Fetal PCG, (f) Gait in Neurodegenerative

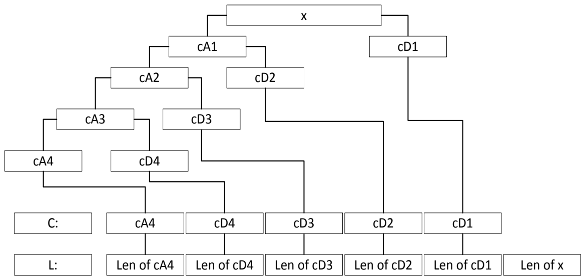

Figure 3: DWT 4-level structure

Wavelet 1-DWT Algorithm
Input: x signal, **n** number of levels **Output: D** details coefficient, **L** approximation coefficient

for each level
[Lowpass, Highpass]=convolving(x); *//filter the input signal to low and high pass*
D_n(details coeff.) = Highpass; // the details coefficient is equaled to high pass filter
L_n(approximation coeff.)= Lowpass // the approximation coefficient is equaled to low pass filter
return D and L
end

We have taken one sample for each class; six classes; and the result described in Table 1.

III. Feature extraction using wavelet energy

After we have get the resulted vector of wavelet decomposition "C" for each sample, another method applied to find the wavelet energy as defined in equation (1) [9]:

$$E_j = \sum_{k=1}^{n} d_{j,k}^2 \tag{1}$$

Where d is wavelet coefficients for each level (four levels), n number of the wavelet levels.
The algorithm of Wavelet Energy has represented in pseudocode below.

Wavelet Energy Algorithm
Input: d wavelet coefficient, **n** number of levels **Output: E** wavelet energy
for each level **E_j** = sum(d^2); *// sum the squared wavelet coefficient at current level* **return E** end

The result of finding wavelet energy for each DWT decomposition vector (C) in Table 1 is mentioned in Table 2.

Table 1: DWT results for six-samples, one-sample from each class of ECG

ECG Sample	C			
Sample 1 of class 1	906	1805	3603	3000
Sample 1 of class 2	381	755	1503	3000
Sample 1 of class 3	318	630	1253	3000
Sample 1 of class 4	1256	2505	5003	3000
Sample 1 of class 5	631	1255	2503	3000
Sample 1 of class 6	131	255	503	3000

Table 2: The result of finding wavelet energy for each DWT decomposition vector (C)

DWT decomposition vector					Wavelet Energy			
906	906	1805	3603	3000	0.0565	0.0634	0.1205	0.8963
381	381	755	1503	3000	0.0003	0.0006	0.0055	0.0301
318	318	630	1253	3000	0.0054	0.0525	0.2819	4.3379
1256	1256	2505	5003	3000	16.1769	46.2187	28.5176	6.2081
631	631	1255	2503	3000	0.0095	0.0600	0.1596	0.0621
131	131	255	503	3000	0.01	0.0028	0.0260	0.0702

IV. Heart diseases classification using SVM

The last phase of proposed work is post-processing divided into two steps; classification and matching. The classification method used to classify the ECG recorded classes is SVM, which is considered as learning method with high efficient and has precision result. The principle of SVM is find the maximum margin between two classes of data [10]. The input data d_{ij} to the SVM has separated into two classes [11]:

$$Ds = \{(d_{1i}l_{1i}, i = 1, \dots, N_1)\}, \{(d_{2j}l_{2j}, j = 1, \dots, N_2)\} \quad (2)$$

Where Ds is the datasets contains of two classes, d_{1i} represents the first class, and d_{2j} represents the second class. l_{1i} defined the labels for first class with labels (0) and l_{2j} define the labels for second class with labels (1). The output data represented by linear classifier for two classes as defined in equation (3) [12].

$$c(n) = sign[w^T d + b] \quad (3)$$

Where w is the weight, b is the bias,

The $c(n)$ resulted the label for each class and is declared in equation (4) [13]:

$$\begin{cases} w^T d + b \geq +1, & if\ c(n) = +1 \\ w^T d + b \leq -1, & if\ c(n) = -1 \end{cases} \quad (4)$$

The number of trained samples can be defined in equation (5).

$$trained_{samples} = trs * Nclass \quad (5)$$

Where trs is the number of selected samples, which is (17), samples, Nclass is number of classes that is (6 classes/cases). Then, the total 102 samples. For that, there is a need to specify (6) labels for each class to be defined for classifier as equation (2). The results for two classes during test phase are shown in Figures 3.

SVM classifier
Input: d data, **l** labels, **n** length of data for each class. **Output: c** linear classifier
For 0 to n -1 **Ds** = [d1i,l1i,i=1,…,N1],[d2j,l2j,j=1,…,N2]; // $d_{n(i,j)}$ represents two classes with labels is the weight, b the bias. If c(n)==+1 w^td+b>=+1 // label for class 1 If c(n) ==-1 w^td+b<=-1 // label for class 2 **return c** end

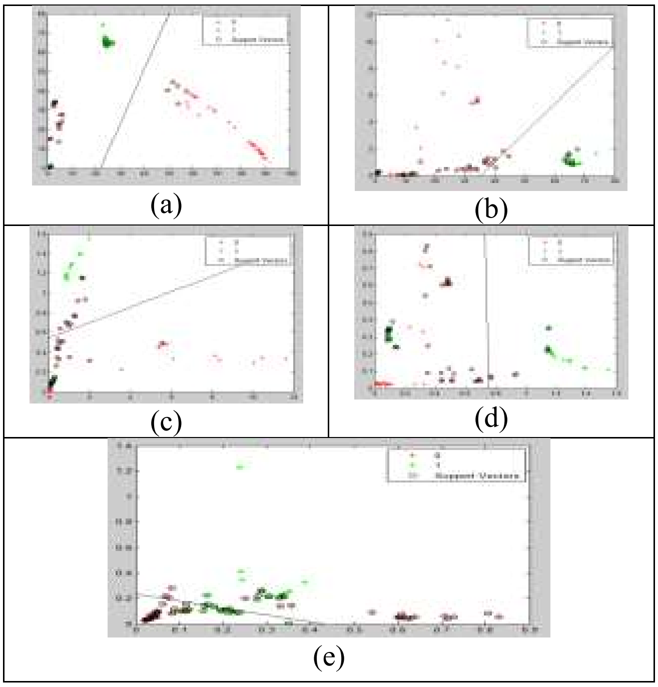

Figure 3: SVM results during training phase, (a) first sample of classes (1 and 2), (b) first sample of

classes (2 and 3), (c) first sample of classes (3 and 4), (d) first sample of classes(4 and 5), (a) first sample of classes (5 and 6).

In figure above, the label (0) represents first sample of current class, and label (1) represents the first sample of next class.

V. Matching process using ICP

The classifier of SVM retunes the 'label', which represents the test sample corresponding class in training model. To find the best sample of training closest to test sample in specific class resulted by SVM we have used Iterative Closest Point (ICP). The implementation of ICP is to select one of best closest feature for test sample from training datasets. The procedure of ICP algorithm is explained [14]:

- There are two sets of points, model (M) and data points (D), and find the corresponding each point of data to all points of model:

$$D(M) = min \, \|M - D\|^2 \qquad (6)$$

- Find the minimize error by using rotation and translation matrix:

$$(R, t) = \min \sum_{m \in M} d^2(Rm + t, D) \qquad (7)$$

Both of transformations (rotation and translation) matrixes defined in equations (8 and 9) [15].

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \qquad (8)$$

$$T = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \qquad (9)$$

ICP algorithm
Input: two set points; model **(M)** and data **(D)** **Output: ME** (minimum error)

for each point of D
  for each point of M
      ME= find minimum result = $\|\mathbf{M} - \mathbf{D}\|^2$
      return ME
  end
end

3. Performance Measurement for Results

The quality measurement metrics used during measure the performance and quality of results for design of proposed work are MSE and PSNR. The performance metrics used to evaluates the results of proposed work are defined as total true positive (TP), total true negative (TN), total false positive (FP), total false negative (FN), sensitivity, specificity, and accuracy.

Table 3: declares the experiments results of best-matched tested samples corresponded with training samples by using ICP

Class	Feature of training sample				Feature of test sample				MSE	PSNR
1	0.3496	0.0075	0.0179	0.0061	0.3496	0.0074	0.0180	0.0059		
1	0.3499	0.0077	0.0181	0.0055	0.3497	0.0075	0.0183	0.0058		
1	0.3483	0.0052	0.0193	0.0042	0.3486	0.0060	0.0199	0.0060		
2	0.2368	0.0910	0.0882	0.0903	0.3253	0.1973	0.4218	1.0015		
2	0.1796	0.1385	0.1331	0.1487	0.1788	0.1495	0.1354	0.1685		
2	0.1654	0.1552	0.1514	0.2354	0.1662	0.1608	0.1576	0.2431		
3	0.3269	0.2025	0.4330	1.0079	0.3870	0.2472	0.5138	1.2780		
3	0.3017	0.2148	0.4277	1.1191	0.3603	0.2211	0.4536	1.0074		
3	0.2776	0.1897	0.4053	0.9924	0.2867	0.1343	0.5286	1.7706		
4	0.0203	0.0287	0.0301	0.0460	0.0216	0.0281	0.0363	0.0664	0.562	54.664
4	0.0222	0.0287	0.0364	0.0667	0.0271	0.0251	0.0132	0.0056	3	6
4	0.0245	0.0295	0.0361	0.0631	0.0302	0.0241	0.0233	0.0254		
5	0.7260	0.0558	0.5406	2.2823	0.7501	0.0655	0.1776	0.0304		
5	0.3549	0.1439	0.6733	2.4048	0.3253	0.1973	0.4218	1.0015		
5	0.5407	0.0907	0.5171	1.9639	0.6145	0.0806	0.5071	2.0384		
6	0.0882	0.1213	0.1086	0.1489	0.0782	0.1954	0.0423	0.1119		
6	0.0432	0.0510	0.0283	0.0242	0.0441	0.0463	0.0179	0.0192		
6	0.0470	0.0552	0.0145	0.0314	0.0441	0.0463	0.0179	0.0192		

Table 4: shows performance metrics results of proposed work

TP	TN	FP	FN	Sensitivity	Specificity	Accuracy
796	130	11	14	0.9827	0.9219	0.9645

1.
2.

4. Conclusion

The main goal of this work is to achieve a high accuracy of classification the abnormal ECG recorded signal into (6) classes/diseases. We have used methods applied during three phases (preprocessing, feature extraction, and post-processing). In the preprocessing phase, DWT has applied to transform the signal of ECG to frequency domain and decrease the length of signal by selecting number of Wavelet transform level. In feature extraction phase, we have calculate the coefficient (wavelet energy) of each level to reduce the length of ECG data. In that last phase of proposed work (post-processing) there are two methods used, SVM for classification and ICP for matching. The achieved accuracy using these proposed methods is (96.45) and the quality metrics to measure the performance of work are MSE (0.562) and PSNR (54.66). In contrast to other researches, in [5] they have achieved accuracy 90.00% and the classification of ECG based on Legendre polynomial, simple logistic and KNN has applied to classify the ECG for normal and abnormal. In [6], they have achieved classification accuracy 90.0%, based on ANFIS and SVM to classifying normal and abnormal heart beats. Other research [7], the accuracy achieved in this research is 88.2% base on LDA, NB, and SVM.

References

[1] S. Nady, M. Massoud and R. Gharieb. "Combining Continuous Wavelet Transform and Teager-Kaiser Energy Operator for ECG Arrhythmia Detection," IEEE, 8th Cairo International Biomedical Engineering Conference (CIBEC), 76-79, 2016.

[2] M. Shivajirao, L. Sanjay and A. Ashok. 2001. Arrhythmia Disease Classification using Artificial Neural Network Model. IEEE, International Conference on Computational Intelligence and Computing Research. 1-4, 2001

[3] B. George, S. Moody, G. Roger, S. Mark and L. Ary, "PhysioNet: Physiologic Signals, Time Series and Related Open Source Software for Basic," Clinical, and Applied Research. 33rd Annual International Conference of the IEEE EMBS Boston, Massachusetts USA. 8327-8330, 2011.

[4] A. Nurul, A. Mohd and M. Norlaili, "ECG Features Extraction using Second-Order Dynamic System and Regeneration using Hybrid Recurrent Network," IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2016.

[5] I. Khalil and A. Khalil, "Classification of ECG Signals Using Legendre Moments," International Journal of Bioinformatics and Biomedical Engineering; AIS, 2015.

[6] M. Czarina and P. Jastine, "Comparative Study between DWT-ANFIS and DWT-SVM in ECG Classification," IEEE, Region 10 Conference (TENCON). 2052-2055, 2016.

[7] L. Aljafar, N. Turky, R. Alotaiby, R. Al-Yami1, A. Saleh and A.J. Zouhair, "Classification of ECG Signals of Normal and Abnormal Subjects Using Common Spatial Pattern," International Conference on Electronic Devices, Systems and Applications (ICEDSA); IEEE, 2016.

[8] K. Jittiphong and N. Atthapol, "Selection of Proper Input Pattern in Fuzzy Logic Algorithm for Classifying the Fault Type in Underground Distribution System," IEEE Region 10 Conference (TENCON). 2650-2655, 2016.

[9] Y. Cao, G. Jing and Y. Guang, "Study on Speech Endpoint Detection Algorithm Based on Wavelet Energy Entropy," Chinese Control and Decision Conference (CCDC). 3965 - 3969, 2016.

[10] A. Shigeo, "Three-Layer Neural. Abe, S. In: Advances in Pattern Recognition Networks (eds)," Support Vector Machine for Pattern Classification. Springer press. 355, 2005.

[11] M. Neha and G. Suendra, "Maximal Margin Multi-classifier based on SVM Hyperparameter Turning," IEEE, Communication and Control. International Conference on Computer, Communication and Control (IC4). 1-5, 2015.

[12] S. Taeshik, S. Jungtaek and M. Jongsubm, "SVM Approach with a Genetic Algorithm for Network Intrusion Detection," In Classification (eds). Computer and Information Sciences-ISCIS 2005. 20th International Symposium Istanbul Turkey, Springer press. 229, 2005.

[13] A. Johan and V. Tony, "Maximal margin classification and linear SVMs. In: Linear SVM classifier: separable case (eds). Least Squares Support Vector Machines," World Scientific; 31, 2005.

[14] M. Alexander, M. Michael and K. Ron, "In the Rigid Kingdom. Bronstein," A. In: Iterative Closest Point Algorithm (eds). Numerical Geometry of Non-Rigid Shapes. Springer Sci. 125, 2008.

[15] L. Yeunsook, M. Mounir and T. Jeong, "3D Matching Applied ICP Algorithm for 3D Facial Avatar Modeling," Lee, Y. In: 3D Data Extraction Using Stereo Camera (eds). Aging Friendly Technology for Health and Independence. 8th International Conference on Smart Homes and Health Telematics, ICOST. 222, 2010.

Author biography

Janan Atiyah Mahdi: She is working in as a lecture Control and Systems Engineering Dept., University of Technology, Baghdad, Iraq. She is published

many papers in the field of Encryption Algorithm