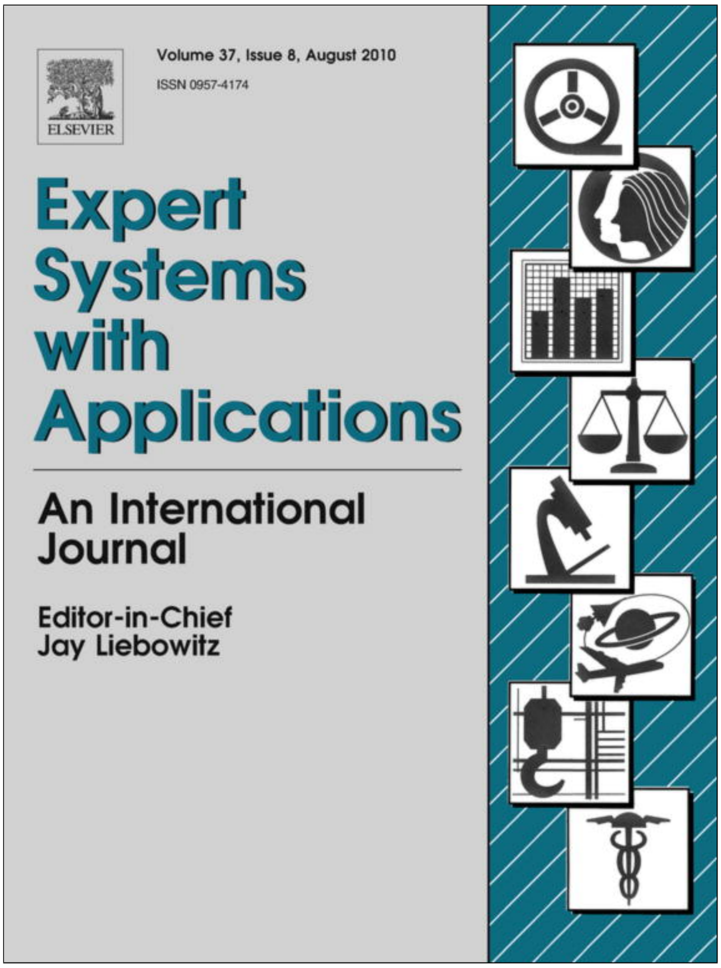
Volume 37, Issue 8, August 2010
ISSN 0957-4174
ELSEVIER
Expert Systems with Applications
An International Journal
Editor-in-Chief
Jay Liebowitz

Expert Systems with Applications 37 (2010) 5751–5757

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification

Hamid Khorrami *, Majid Moavenian

Department of Mechanical Engineering, Ferdowsi University of Mashhad, Iran

ARTICLE INFO

Keywords:
MLP
SVM
ECG
DWT
CWT
DCT

ABSTRACT

In this study we have proposed and compared use of CWT (Continues Wavelet Transform) with two powerful data transformation techniques DWT (Discrete Wavelet Transform), and DCT (Discrete Cosine Transform) which have already been in use, in order to improve the capability of two pattern classifiers in ECG arrhythmias classification. The classifiers under examination are MLP (Multi-Layered Perceptron, a conventional neural network) and SVM (Support Vector Machine). The training or learning algorithms used in MLP and SVM are BackPropagation (BP) and Kernel–Adatron (K–A), respectively. The ECG signals taken from MIT-BIH arrhythmia database are used to classify four different arrhythmias together with normal ECG. The output of MLP and SVM classifiers in terms of training performance, testing performance or generalization ability and training time are compared. MLP and SVM training and testing stages have been carried out twice. At first, only one lead (II) is used, and then a second ECG lead (V1) has been added to the training and testing datasets. Three feature extraction techniques are applied separately to datasets before classification. The results show that selection of the best feature extraction method will depend on the substantial value considered for training time, training and testing performance. This is stated because when applying MLP or SVM, addition of CWT and DCT will show the advantage only when training performance and testing performance are important, respectively. Generally speaking only testing performance with single lead for MLP shows superiority over SVM.

1. Introduction

The electrocardiogram (ECG) is the electrical activity signal of the heart which is very important in heart disease diagnosing because every arrhythmia in ECG signals can be relevant to a heart disease. The main problem in heart disease diagnose using ECG is that the normal ECG may differ for each person and sometimes one disease has dissimilar signs on different patient's ECG signals. Also, two distinct diseases may have approximately identical effects on normal ECG signals. These problems complicate the heart disease diagnose. So, utilization of pattern classifier techniques can improve the new patients ECG arrhythmia diagnoses.

ANN (artificial neural network) is a conventional classifier used for ECG arrhythmias classification. MLP is introduced to be able to recognize and classify ECG signals more accurately than other ANN methods. However, MLP with (BP) training algorithm suffers from slow convergence to local and global minima and from random settings of weights, initial values (Özbay, Ceylan, & Karlik, 2006). Improvement of ANN's performance have been the subject of new researches on ECG arrhythmias classification by application of various feature extraction techniques. Discrete wavelet transform is used to improve the performance of MLP with (BP) training algorithm and also compared with other feature extraction and data reduction methods (Ceylan & Ozbay, 2007). Addition of DWT has improved the accuracy of MLP neural network (Froese, Hadjiloucas, Galvão, Becerra, & José Coelho, 2006). Also an ECG beat classification system based on DWT and probabilistic neural network (PNN) is proposed to discriminate six ECG beat types (Yu & Chen, 2007). The ECG recordings were processed using CWT and DWT in an effort to predict the maintenance of sinus rhythm after cardioversion in patients with persistent atrial fibrillation (Cervigón, Sánchez, Castells, Blas, & Millet, 2007). Use of DWT in analysis of electrocardiographic changes in partial epileptic patient (Ubeyli, 2008).

Survey on SVM classifiers shows that numerous research works have been carried out. It is pointed out that SVM classifiers do not trap in local minima points and need less training input therefore they are faster than ANN (Abe, 2005). Classification of SVM for ECG signals assisted by feature extraction methods carried out by Acir (2006). He introduced a novel SVM classification

* Corresponding author. Tel.: +98 9155515832; fax: +98 5118763304.
E-mail addresses: khorrami.hamid@gmail.com, hamid_322@yahoo.com (H. Khorrami), moaven@ferdowsi.um.ac.ir (M. Moavenian).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.02.033

based on a perturbation method for ECG arrhythmias classification and improved SVM performance by employing DCT and DWT. It's experimental results show that DCT–SVM structure has better accuracy than DWT–SVM structure. Heart valve disease diagnosis is carried out using SVM and ANN (artificial neural network) together with DWT (Comak, Arslan, & Turkoglu, 2007). A novel classifier proposed by the authors is SVM classifier with (K–A) training algorithm.

In this paper we have constructed eight different structures for ECG signals classification. These structures are MLP, DWT-MLP, CWT-MLP, DCT-MLP, SVM, DWT-SVM, CWT-SVM and DCT-SVM, (Fig. 1). For training and testing structures, first a one lead ECG signal (II) was used which contained four different arrhythmias accompanied by normal ECG signal. Then two lead ECG signals (II and V1) were used as the second type of ECG dataset for MLP and SVM training.

2. Materials and preprocessing

The ECG signals for training and testing datasets are taken from MIT-BIH arrhythmia database (mitdb). This contains two lead ECG signals of 48 patients. The selected Arrhythmias are LBBB (Left Bundle Branch Block), RBBB (Right Bundle Branch Block), PAB (Premature Atrial Beat) and PVB (Premature Ventricular Beat). Ninety beats were chosen for each arrhythmia and normal ECG divided into three groups of training, testing and validation (see Table 1). Each ECG beat is a matrix (334×1) when one ECG lead is used (II) and a matrix (668×1) when two ECG leads are used (II andV1). Every ECG signal has five distinct points (P, Q, R, S and T) used for the interpretation of the ECG (Fig. 2). Every R–R interval duration was considered as a beat in the study. Because no-fixed ECG base line exists for each individual patient every beat was located in zero to one vertical scale for better arrhythmias classification.

2.1. Multi-Layered Perceptron

In our study, a three-layered feed-forward neural network was trained, using (BP) algorithm. The (BP) training algorithm with generalized delta learning rule is an iterative gradient algorithm designed to minimize the mean square error between the actual output of a multilayered feed-forward neural network and a desired output. Each layer is fully connected to the previous layer, and there exist no any other connection.

2.2. Backpropagation algorithm (summary)

Given a finite length input patterns $x_1(k), x_2(k), \ldots, x_n(k) \in \Re$, $(1 \leqslant k \leqslant K)$ and the desired patterns $x_1(k), x_2(k), \ldots, x_m(k) \in \Re$,

Step 1: Select the total number of layers M, the number $n_i (i = 1,2,\ldots,M-1)$ of the neurons in each hidden layer, and an error tolerance parameter $\varepsilon > 0$.
Step 2: Randomly select the initial values of the weight vectors $w_{aj}^{(i)}$ for $i = 1,2,\ldots,n_i$.
Step 3: Initialization:

$$w_{aj}^{(i)} \leftarrow w_{aj}^{(i)}(0), \quad E \leftarrow 0, \quad k \leftarrow 1$$

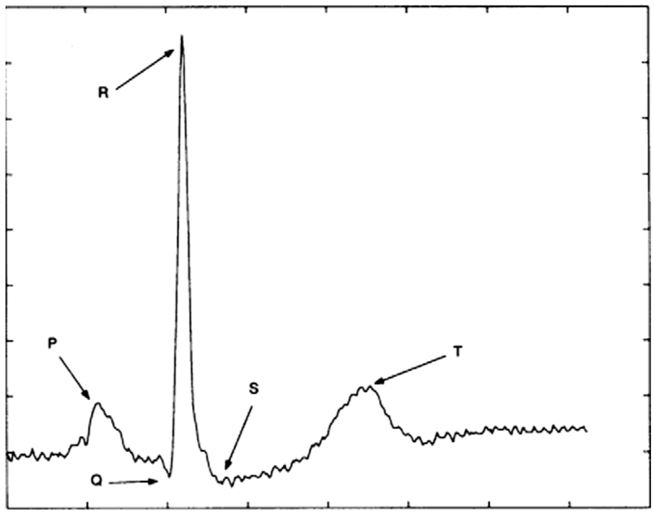

Fig. 2. Standard ECG beat.

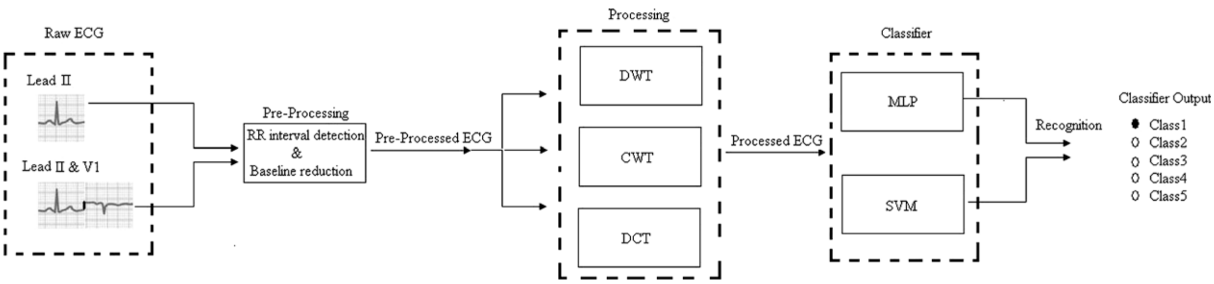

Fig. 1. Schematic of structures.

Table 1
Number of training, testing and validation data in first and second types of datasets.

		LBBB	RBBB	Normal	PVC	PAB	Total
Number of training data beats		50	50	50	50	50	250
Number of validation data beats	For MLP	30	30	30	30	30	150
	For SVM	0	0	0	0	0	0
Number of testing data beats		10	10	10	10	10	50
MIT-BIH data file		111-207-214-109	118-207-212-231	101-105-209-234	107-108-109-119-200-203-207-223-233	118-200-201-202-207-209-	

Step 4: Calculate the neural outputs

$$\begin{cases} s_j^{(i)} = (w_{aj}^{(i)})^T x_a^{(i-1)} \\ x_j^{(i)} = \sigma\left(s_j^{(i)}\right) \end{cases} \tag{1}$$

For $i = 1,2,\ldots,M$ and $j = 1,2,\ldots,n_i$.

Step 5: Calculate the output error

$$e_j = d_j - x_j^{(M)} \tag{2}$$

for $j = 1,2,\ldots,m$.

Step 6: Calculate the output deltas

$$\delta_j^{(M)} = e_j \sigma'\left(s_j^{(M)}\right) \tag{3}$$

Step 7: Recursively calculate the propagation errors of the hidden neurons

$$e_j^{(i)} = \sum_{l=1}^{n_i+1} \delta_l^{(i+1)} w_{lj}^{(i+1)} \tag{4}$$

From the layer $M-1$, $M-2$, ..., to layer 1.

Step 8: Recursively calculate the hidden neural delta values:

$$\delta_j^{(i)} = e_j \sigma'\left(s_j^{(i)}\right) \tag{5}$$

Step 9: Update weight vectors

$$w_{aj}^{(i)} = w_{aj}^{(i)} + \eta \delta_j^{(i)} x_a^{(i-1)} \tag{6}$$

Step 10: calculate the error function

$$E = E + \frac{1}{k}\sum_{j=1}^{m} e_j^2 \tag{7}$$

Step 11: if $k = K$ then go to step 12; otherwise, $k \leftarrow k+1$ and go to step 4.
Step 12: if $E \leqslant \varepsilon$ then go to step 13; otherwise go to step 3.
Step 13: learning is completed. Output the weights (Gupta, Jin, & Homma, 2003).

After completing the training procedure of the neural network, the weights of MLP are frozen and MLP is made ready for testing stage. MATLAB software is employed to run structures using MLP with BP algorithm.

2.3. Support vector machines

A special form of ANNs are SVMs, introduced by Boser in 1992. The SVM performs classification by non-linearly mapping their n-dimensional input into a high dimensional feature space. In this high dimensional feature space a linear classifier is constructed. Doing the explicit mapping would be computationally unreasonable, and the algorithm avoids that by introducing the kernel, which is possible since the algorithm only uses the scalar product of the inputs. From this the classification problem is translated into a convex quadratic optimization problem, which due to its convexity has a unique solution.

The simplest version of a SVM is the so-called Maximal Margin Classifier. It is applicable only when data are linearly separable. It is a good start for understanding the basic ideas behind more sophisticated SVMs. Consider a linearly separable dataset $\{(X_i, d_i)\}$, where X_i is the input pattern for the i:th example and d_i is the corresponding desired output $\{-1, 1\}$. The assumption, "the dataset is linearly separable", means there exist a hyperplane working as the decision surface. We can write:

$$\begin{aligned} &W^T X_i + b \geqslant 0, \text{ then } d_i = +1 \\ &W^T X_i + b \leqslant 0, \text{ then } d_i = -1 \end{aligned} \tag{8}$$

where $W^T X + b$ is the output function. The distance from the hyperplane to the closest point is called the geometric margin. The idea is, to have a good machine, so the geometric margin needs to be maximized. To get that, we first introduce the marginal function $W^T X + b$. Because the dataset is linearly separable we can rewrite Eq. (8) as follow:

$$\begin{aligned} &W^T X^+ + b = +1 \\ &W^T X^- + b = -1 \end{aligned} \tag{9}$$

where $X^+(X^-)$ is the closest data point on the positive (negative) side of the hyperplane. Now it is straight forward to compute the geometric margin

$$\begin{aligned} \gamma &= \frac{1}{2}\left(\frac{W^T X^+ + b}{|w|} - \frac{W^T X^- + b}{|w|}\right) \\ &= \frac{1}{2|w|}(W^T X^+ + b - W^T X^- - b) = \frac{1}{2|w|}(1-(-1)) = \frac{1}{|w|} \end{aligned} \tag{10}$$

Hence, equivalent to maximize the geometric margin is fixing the functional margin to one and minimizing the norm of the weight vector,$|w|$. This can be formulated as a quadratic (ww^T) problem with inequality constraints $d_i(w^T x_i + b) \geqslant 1$.

$$\begin{aligned} &\min : \frac{1}{2} W^T W \ \text{(quadratic-problem)} \\ &\text{subject to} : d_i\left(w^T x_i + b\right) \geqslant 1 \end{aligned} \tag{11}$$

By the use of Lagrange multipliers $\alpha_i \geqslant 0$ the original problem is transformed into the dual problem. From the Kuhan–Tuker theory we have the following condition

$$\alpha_i\left[d_i\left(W^T x_i + b\right) - 1\right] = 0 \tag{12}$$

which means only the points with functional margin unity contributes to the output function. These points are called the Support Vectors. Since they are supporting, the separating hyperplane. For more information about SVM classifying, non-separable datasets and classifying more than two classes, see (Abe, 2005).

2.4. Kernel–Adatron algorithm (summary)

Support vector machines work by mapping training data for classification tasks into a high dimensional feature space. In the feature space they then find a maximal margin hyperplane which separates the data. This hyperplane is usually found using a quadratic programming routine which is computationally intensive and non trivial to implement. In this section we briefly explain the (K–A) algorithm for SVM classification. The algorithm is simple and can find rapid solution for SVM classification with an exponentially fast rate of convergence (in the number of iterations) towards the optimal solution as follows:

Step 1: Initialize Lagrangian parameters $\alpha_i = 1$.
Step 2: Starting from pattern $i = 1$, for labeled points $\{(x_i, y_i)\}$ calculates

$$z_i = y_i \sum_{j=1}^{p} \alpha_i y_i K(x_i, x_j) \tag{13}$$

Step 3: For all patterns i calculate

$$\gamma_i = y_i z_i \tag{14}$$

and execute steps 4–5 below.

Step 4: Let

$$\delta\alpha^i = \eta(1 - \gamma^i) \quad (15)$$

Be the proposed change to the multipliers α^i.

Step 5.1: If $(\alpha^i + \delta\alpha^i) \leqslant 0$ then the proposed change to the multipliers would result in a negative α^i. Consequently to avoid this problem we set $\alpha^i = 0$.

Step 5.2: If $(\alpha^i + \delta\alpha^i) > 0$ then the multipliers are updated through the addition of the $\delta\alpha^i$i.e. $\alpha^i \leftarrow \alpha^i + \delta\alpha^i$.

Step 6: Calculate the bias b from

$$b = \frac{1}{2}(\min(z_i^+) + \max(z_i^-)) \quad (16)$$

where z_i^+ are those patterns i with class label +1 and z_i^- are those with class label -1.

Step 7: If a maximum number of presentations of the pattern set has been exceeded then stop, otherwise return to step 2. The kernel $K(x, x')$ can be any function satisfying Mercer's condition; in particular it is possible to use RBF or polynomial kernels (Abe, 2005). Some conventional kernels are introduced in Table 2.

3. Wavelet transform

In 1982, Jean Morlet, a French geophysical engineer, discovered the idea of the wavelet transform. Morlet first introduced the idea of wavelets as a family of functions constructed from translations and dilatations of a single function called the "mother wavelet". Many of researchers (Grossmann, Meyer, Mallat, Daubechies, etc.) developed and enhanced this new signal-processing tool to make it the most efficient and used technique in the structural health-monitoring field. The wavelet transform can be thought of as an extension of classic Fourier transform, except that, instead of working on single scale (time or frequency), it works on a multi-scale basis. The wavelet transform can be classified as continuous or discrete.

3.1. Continues wavelet transform

The wavelet analysis has been introduced as a windowing technique with variable-sized regions. Wavelet decomposition introduces the notion of scale as an alternative to frequency, and maps a signal into a time-scale plane as shown in the (Fig. 3). This is equivalent to the time-frequency plane used in the STFT (short time Fourier transform). Each scale in the time-scale plane corresponds to a certain range of frequencies in the time-frequency plane. The term wavelet means a small wave. A wavelet is a waveform of limited duration. Wavelets are localized waves that extend for a finite time duration compare to sine waves which extend from minus to plus infinity. The comparison with the Fourier analysis is now clear. The wavelet analysis is the decomposition of a signal into shifted and scaled versions of the original wavelet whereas the Fourier analysis is the decomposition of a signal into sine waves of different frequencies.

Mathematically, the continuous wavelet transform of a function $f(t)$ is defined as the integral transform of $f(t)$ with a family of wavelet functions, $\psi_{a,b}(t)$:

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) * \psi\left(\frac{t-b}{a}\right) dt \quad (17)$$
$$a \in R^+ - \{0\}, \ b \in R$$

In other words, the continuous wavelet transform (CWT) is defined as the sum of the signal multiplied by scaled and shifted versions of the wavelet function ψ:

$$CWT(scale, position) = \int_{-\infty}^{+\infty} f(t) * \psi(scale, position, t) dt \quad (18)$$

The function $\psi(t)$ is commonly called the mother wavelet and the family of functions $\psi_{a,b}(t)$ is called daughter wavelets. The daughter wavelets are derived from scaling and shifting the mother wavelet. The scale factor a represents the scaling of the function $\psi(t)$, and the shift factor b represents the temporal translation of the function. The results of the CWT are number of special wavelet coefficients located in matrix C (function of scale and position). For more information see Mallet (1999). It is important to know that determination of CWT scale parameter and mother wavelet are very significant in ECG feature extraction.

3.2. Discrete wavelet transform

Calculating wavelet coefficients at every possible scale is a fair amount of work, and it generates an awful lot of data. What if we choose only a subset of scales and positions at which to make our calculations? It turns out, rather remarkably, that if we choose scales and positions based on powers of two, so-called *dyadic* scales and positions. We obtain such an analysis from the discrete wavelet transform (DWT). DWT works like a bandpass filter and we can do DWT for a signal several levels. Each level decompose the input signal into approximations (low frequency part of initial signal) and details (high frequency part of initial signal). The next level of DWT is done upon approximations. It is also essential to notice that determination of DWT level and mother wavelet are very important in ECG feature extraction.

Table 2
Some conventional kernels.

Kernel function	Type of classifier
$K(x, x_i) = \exp(-\gamma\|x - x_i\|^2)$	Gaussian radial basis function (RBF)
$K(x, x_i) = (x^T x_i + 1)^d$	Polynomial of degree d
$K(x, x_i) = \tanh(x^T x_i - \theta)$	Multi-Layer Perceptron

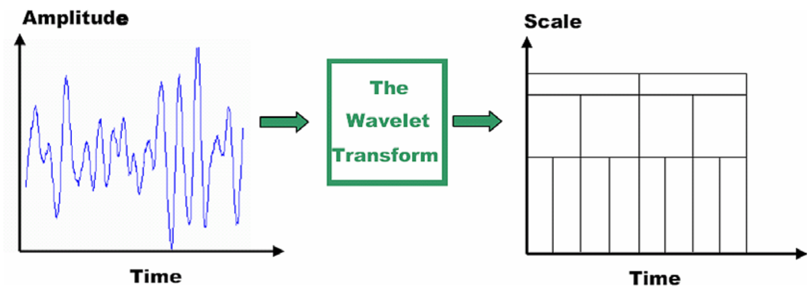

Fig. 3. Wavelet transform domain.

$$DWT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) * \psi\left(\frac{t-b}{a}\right) dt$$
$$a = 2^j, \quad b = k2^j, \quad (k,j) \in Z^2 \tag{19}$$

In this study the best results of ECG signals classification were obtained for DWT–MLP and CWT–MLP structures by examining different already most used mother wavelets to select the best for DWT and CWT techniques, Also, optimal scale parameter in CWT and optimal decomposition level parameter in DWT were specified empirically. Further we compared DWT–MLP and CWT–MLP also DWT–SVM and CWT–SVM structural performances, by employing the selected optimal parameters found for DWT and CWT.

4. Discrete cosine transform

One of the main reasons for employing DCT in ECG signals classification is the ability to compress signals. Compression means that you can restore signal information in a restrict number of DCT coefficients. The most common DCT definition of a one-dimetional (1D) sequence of length N is

$$C(u) = \alpha(u) \sum_{x=0}^{N-1} f(x) \cos\left[\frac{\pi(2x+1)u}{2N}\right] \tag{20}$$

for $u = 0, 1, 2, \ldots, N-1$. Similarly, the inverse transformation is defined as

$$f(x) = \sum_{u=0}^{N-1} \alpha(u) C(u) \cos\left[\frac{\pi(2x+1)u}{2N}\right] \tag{21}$$

for $x = 0, 1, 2, \ldots, N-1$. In both Eqs. (20) and (21) $\alpha(u)$ is defined as

$$\alpha(u) = \sqrt{1/N}, u = 0$$
$$\&$$
$$\alpha(u) = \sqrt{2/N}, u \neq 0 \tag{22}$$

It is clear from (20) that for $C(u=0) = \sqrt{1/N}\sum_{x=0}^{N-1} f(x), u = 0$. Thus, the first transform coefficient is the average value of the sample sequence. To fix ideas, ignore the $f(x)$ and $\alpha(u)$ component in (20). The plot of $\sum_{x=0}^{N-1} \cos\left[\frac{\pi(2x+1)u}{2N}\right]$ for $N = 8$ and varying values of u is shown in (Fig. 4). In accordance with our previous observation, the first the top-left waveform ($u = 0$) renders a constant (DC) value, whereas, all other waveforms ($u = 1, 2, \ldots, 7$) give waveforms at progressively increasing frequencies (Pennebaker & Mitchell, 1993).

If the input sequence has more than N sample points then it can be divided into sub-sequences of length N and DCT can be applied to these chunks independently. Here, a very important point to note is that in each such computation the values of the basis function points will not change. Only the values of $f(x)$ will change in each sub-sequence. This is a very important property, since it shows that the basis functions can be pre-computed offline and then multiplied with the sub-sequences. This reduces the number of mathematical operations (i.e., multiplications and additions) thereby rendering computation efficiency. MATLAB software is used to calculated DCT coefficients in this paper.

5. Structure

In this study, eight different structures were formed for classification of ECG arrhythmias. In structures one to four MLP classifier with (BP) training algorithm has been trained and tested. CWT, DWT and DCT are added to form structures two to four, respectively. Training and testing performance (Tr.P and Te.P) plus training time (Tr.T) are selected as base for evaluation of these structures to candidate the best. See Table 3. In training MLP classifier, usually we try to maximize classification performance of the training data. But if the classifier is too fit for the training data, the classification ability for test data, i.e., the generalization ability is degraded. This phenomenon is called *overfitting*. To solve the *overfitting* problem during learning of MLP we had a random selection

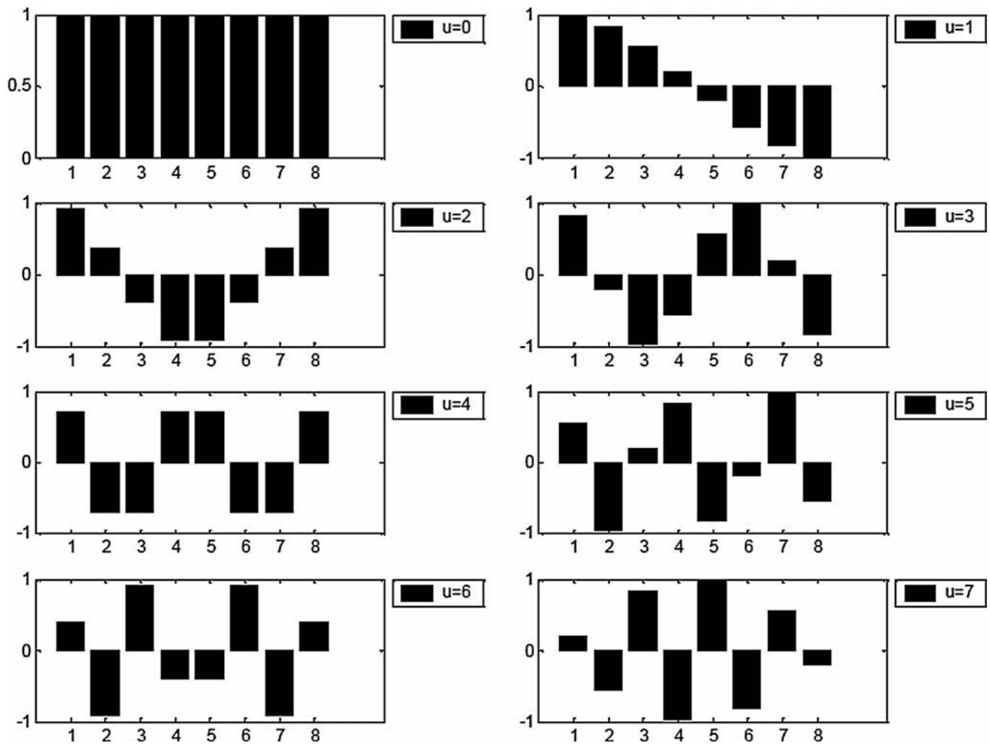

Fig. 4. Eight DCT components.

Table 3
Tr.P, Te.P and Tr.T for structures one to four.

Structure	Dataset type	Training performance	Training time (min:s)	Number of iterations	Testing performance
MLP	1	0.0406339	03:28	5000	0.0516
CWT-MLP	1	0.0182962	03:50	5000	0.0552
DWT-MLP	1	0.0349334	02:47	5000	0.0438
DCT-MLP	1	0.0572522	03:03	5000	0.0668
MLP	2	0.0213711	05:54	5000	0.1305
CWT-MLP	2	0.0056344	04:58	5000	0.1048
DWT-MLP	2	0.0228386	03:03	5000	0.1422
DCT-MLP	2	0.0293947	03:18	5000	0.0865

of ECG beats validation data which is used to prevent *overfitting* in training stage.

Learning or training of MLP and SVM has been done with two types of datasets according to Table 1. The first type of dataset is obtained from one lead ECG signals (II) and the second one from two lead ECG signals (II and V1). In these structures MLP classifier with three layers containing 34, 50 and 5 neurons, respectively, gives the best performance results in training and testing stages.

In the next for structures we have utilized SVM classifier with RBF kernel function used by (K–A) algorithm. RBF mapes datasets into a high-dimensional feature space. Since five classes (arrhythmias) are involved in classification and they are more than two, one-against-all method is used for SVM see (Abe, 2005). Also, in (K–A) algorithm SVMs are motivated by the concept of training and using only those inputs that are near to the decision surface. These inputs provide the main information for classification. CWT, DWT and DCT are added to the SVM structure in order to improve the results. See Table 4.

6. Calculation of training and test performances

Training and test performances are calculated and presented in Tables 3 and 4 using Eq. (23)

$$MSE = \left(\frac{\sum_{j=1}^{P} \sum_{i=1}^{N} (d_{ij} - y_{ij})^2}{NP} \right) \tag{23}$$

where P = number of sample points in each beat, N = number of beats in input matrix, d_{ij} = desired output of classifier for jth sample point and ith beat, y_{ij} = real output of classifier for jth sample point and ith beat. MSE = mean square error.

7. Organization and manipulation of the results

The results of experiments carried out in this research are presented in Tables 3 and 4. Table 3, only deals with application of MLP structure, when feature extractions are active with single lead. Table 4 is designed so that while it is dealing with application of SVM it compares the results obtained by SVM and MLP together. This is done by dividing each row of Table 4 related to special structure in two rows (In the first row Tr.P and Te.P of Table 3 for each row is kept constant and Tr.T and number of iterations are found when SVM is active. In the second row Tr.T of Table 3 for each row is kept constant then Tr.P, number of iterations and Te.P are found when SVM is active.

8. Conclusion

Classification of ECG arrhythmias taken from different and numerous patients are common tools for prediction of the existence of arrhythmia in ECG signal. In this research, a comparison of two classifiers, MLP and SVM, using three feature extraction methods is done. Eight constructed structures explained in section one have been under consideration with respect to Tr.P, Te.P and Tr.T.

Analysis of the results shown in Table 3, shows that, when feature extractions are active with single lead, Tr.P is reduced for CWT and DWT and Tr.T is reduced for DWT and DCT. It should be noted that Te.P is only reduced for DWT. When the second lead is added the Tr.P has improved nearly two times for all feature extraction methods where as Tr.T increases no more than 30%, while Te.P shows an increase of at least 2.5 times except for DCT.

Table 4
Tr.P, Te.P and Tr.T for structures five to eight.

Structure	Dataset type	Training performance	Training time (min:s)	Number of iterations	Testing performance
SVM	1	0.0406339[a] 0.0100045	00:25 03:28[a]	39 334	0.0516[a] 0.0651
CWT-SVM	1	0.0182962[a] 0.01267129	00:28 03:50[a]	56 306	0.0552[a] 0.0984
DWT-SVM	1	0.0349334[a] 0.0095859	00:09 02:47[a]	49 389	0.0438[a] 0.0680
DCT-SVM	1	0.0572522[a] 0.015230	00:05 03:03[a]	29 544	0.0668[a] 0.0555
SVM	2	0.0213711[a] 0.0091081	01:52 05:54[a]	72 325	0.1305[a] 0.1086
CWT-SVM	2	0.00563444[a] 0.011462188	– 04:58[a]	– 306	0.1048[a] 0.1189
DWT-SVM	2	0.0228386[a] 0.010039938	00:42 03:03[a]	62 259	0.1422[a] 0.1110
DCT-SVM	2	0.0293947[a] 0.0099505	00:19 03:18[a]	59 488	0.0865* 0.1031

[a] Indicates the reference value used from Table 3 in Table 4.

Similar analysis of Table 4 shows that for SVM structure, when feature extractions are active with single lead, Tr.P is reduced not considerably for all feature extraction methods but Tr.T shows a noticeable reduction down to one fifth for DWT and DCT but Te.P is only increased for CWT. When the second lead is added there are not noticeable changes in Tr.P but Tr.T and Te.P are showing noticeable changes.

Comparison of MLP and SVM based on the results reported in Tables 3 and 4 shows that SVM classifier with (K–A) training algorithm has improved Tr.P at least four times. Also the time spent to reach the same Tr.P as for MLP, is reduced 3:03 min:s.

Implementation of the selected and proposed classification structures shows that selection of the best feature extraction method will depend on the substantial value considered for training time, training and testing performance.

References

Abe, S. (2005). *Support vector machines for pattern classification*. London, NJ: Springer-Verlag.

Acir, N. (2006). A support vector machine classifier algorithm based on perturbation method and its application to ECG beat recognition systems. *Expert Systems with Applications, 31*, 150–158.

Cervigón, R., Sánchez, C., Castells, F., Blas, J. M., & Millet, J. (2007). Wavelet analysis of electrocardiograms to characterize recurrent atrial fibrillation. *Journal of the Franklin Institute, 344*, 196–211.

Ceylan, R., & Ozbay, Y. (2007). Comparison of FCM, PCA and WT techniques for classification ECG arrhythmias using artificial neural network. *Expert Systems with Applications, 33*, 286–295.

Comak, E., Arslan, A., & Turkoglu, I. (2007). A decision support system based on support vector machines for diagnosing of the heart valve diseases. *Computers in Biology and Medicine, 37*, 21–27.

Froese, T., Hadjiloucas, S., Galvão, K. H. R., Becerra, V. M., & José Coelho, C. (2006). Comparison of extrasystolic ECG signal classifiers using discrete wavelet transforms. *Pattern Recognition Letters, 27*, 393–407.

Gupta, M., Jin, L., & Homma, N. (2003). *Static and dynamic neural networks from fundamentals to advanced theory*. NJ: John Wiley and Sons.

Mallet, S. (1999). *A wavelet tour of signal processing*. NJ: Academic Press.

MIT-BIH Arrhythmia Database, <http://www.physionet.org./physiobank/database/mitdb/>.

Özbay, Y., Ceylan, R., & Karlik, B. (2006). A fuzzy clustering neural network architecture for classification of ECG arrhythmias. *Computers in Biology and Medicine, 36*, 376–388.

Pennebaker, W. B., & Mitchell, J. L. (1993). *JPEG – Still image data compression standard*. NJ: International Thomson Publishing.

Ubeyli, E. (2008). Support vector machines for detection of electrocardiographic changes in partial epileptic patients. *Engineering Applications of Artificial Intelligence, 21*, 1196–1203.

Yu, S., & Chen, Y. (2007). Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network. *Pattern Recognition Letters, 28*, 1142–1150.