Applied Soft Computing 11 (2011) 2676–2686

Contents lists available at ScienceDirect

Applied Soft Computing

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

Control chart pattern recognition using a novel hybrid intelligent method

Vahid Ranaee*, Ata Ebrahimzadeh

Faculty of Electrical and Computer Engineering, Babol University of Technology, Babol, Iran

ARTICLE INFO

Article history:
Received 6 October 2009
Received in revised form 28 August 2010
Accepted 27 October 2010
Available online 4 November 2010

Keywords:
Control chart pattern recognition
Multi-resolution wavelet transform
Support vector machine
Genetic algorithm
Hybrid intelligent method

ABSTRACT

Control chart patterns (CCPs) are important statistical process control tools for determining whether a process is run in its intended mode or in the presence of unnatural patterns. Automatic recognition of abnormal patterns in control charts has seen increasing demands nowadays in the manufacturing processes. This paper presents a novel hybrid intelligent method (HIM) for recognition of common types of CCP. The proposed method includes three main modules: a feature extraction module, a classifier module and an optimization module. In the feature extraction module, the multi-resolution wavelets (MRW) are proposed as the effective features for representation of CCPs. These features are novel in this area. In the classifier module, because of the promising generalization capability of support vector machines, a multi-class SVM (SVM) based classifier is proposed. In support vector machine training, the hyper-parameters have very important roles for its recognition accuracy. Therefore, in the optimization module, an efficient genetic algorithm is proposed for selecting of appropriate parameters of the classifier. Simulation results confirm that the proposed system outperforms other methods and shows high recognition accuracy about 99.37%.

1. Introduction

Control chart patterns (CCPs) are important statistical process control tools for determining whether a process is run in its intended mode or in the presence of unnatural patterns. CCPs can exhibit six types of pattern: normal (NR), cyclic (CC), increasing trend (IT), decreasing trend (DT), upward shift (US) and downward shift (DS) [1]. Except for normal patterns, all other patterns indicate that the process being monitored is not functioning correctly and requires adjustment. Fig. 1 shows six pattern types of control chart.

Over the years, numerous supplementary rules known as zone tests or run tests [2] have been proposed to analyze control charts. Interpretation of the process data still remains difficult because it involves pattern recognition tasks. It often relies on the skill and experience of the quality control personnel to identify the existence of an unnatural pattern in the process.

An efficient automated control chart pattern (CCP) recognition system can compensate this gap and ensure consistent and unbiased interpretation of CCPs leading to lesser number of false alarms and better implementation of control charts. Aiming this, several approaches have been proposed for CCP recognition. Some of the researchers used the expert systems [2], fuzzy-clustering method [3] and decision tree (DT) based classifiers [4]. Several papers have used the artificial neural networks (ANN) for recognition of CCPs [5–15]. ANN can be simply classified into two main categories, including supervised ANN and unsupervised ANN. Literature review shows that the techniques, which use MLP neural networks as the classifier, have high performances. The advantage with a neural network is that it does not require the provision of explicit rules or templates. Most the existing techniques used the unprocessed data as the inputs of CCPs recognition system. The use of unprocessed CCP data has further many problems such as the amount of data to be processed is large. On the other hand, a feature-based approach is more flexible to deal with a complex process problem, especially when no prior information is available. If the features represent the characteristic of patterns explicitly and if their components are reproducible with the process conditions, the classifier recognition accuracy will increase [15]. Further, if the feature is amenable to reasoning, it will help in understanding how a particular decision was made and thus makes the recognition process a transparent process. Features could be obtained in various forms, including principal component analysis shape features [11,13], correlation between the input and various reference vectors [16], and statistical correlation coefficients [17].

Based on the published papers, there exist some important issues in the design of automatic CCPs recognition system which if suitably addressed, lead to the development of more efficient recognizers. One of these issues is the extraction of the features. In this paper for obtaining the compact set of features which capture the prominent characteristics of the CCPs in a relatively small number of the components, the multi-resolution wavelets analysis is pro-

* Corresponding author.
E-mail address: vranaei@yahoo.com (V. Ranaee).

1568-4946/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2010.10.016

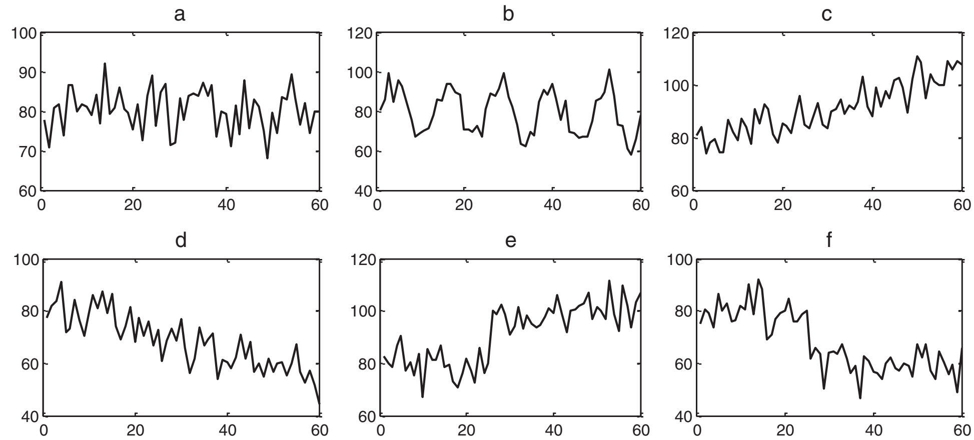

Fig. 1. Six common types of CCPs: (a) normal, (b) cyclic, (c) upward trend, (d) downward trend, (e) upward shift and (f) downward Shift.

posed. Another issue is related to the choice of the classification approach to be adopted. Literature review shows the systems that use artificial neural networks (ANNs) as the classifiers have high performances. However, with regard to effectiveness of ANNs, they suffer from some drawbacks. For example such as the requirement for a large amount of training data, 'over-fitting', slow convergence velocity and relapsing into a local extremum easily [18]. The practicability of ANNs is limited due to these weaknesses. Support vector machines (SVMs), based on statistic all earning theory, are gaining applications in the area of pattern recognition because of their excellent generalization capability [19]. Using SVMs is the method that is receiving increasing attention, with remarkable results recently [19]. The main difference between ANNs and SVMs is the principle of risk minimization. An ANN implements empirical risk minimization to minimize the error on the training data, whereas an SVM implements the principle of structural risk minimization in place of experiential risk minimization, which makes it have excellent generalization ability in the situation when there is a small sample. The largest problems encountered in setting up the SVM model are how to select the kernel function and its parameter values. The parameters that should be optimized include the penalty parameter (C) and the kernel function parameters such as the value of gamma (γ) for the radial basis function (RBF) kernel.

Turning back to CCPs recognition systems, it can be found that the selection of the best free parameters of the adopted classifier is generally done empirically. On the other hand using of SVM has some difficulties, which are how to select the optimal kernel function type, most appropriate the hyper-parameters values for SVM training and testing stages. Therefore in this study, we used an efficient optimizer for finding the optimum values of hyper-parameters, i.e., the kernel parameter and classifier parameters.

The rest of paper is organized as follows. Section 2 explains the feature extraction. Section 3 presents the classifier. Section 4 presents the optimization and the proposed model. Section 5, shows some simulation results and finally Section 6 concludes the paper.

2. Feature extraction

Feature extraction plays an important role for CCPs recognition problem. On the other hand control charts patterns are non-stationary signals which have highly complex time-frequency characteristics. For example, trend patterns are typically lower in frequency for a longer period of time while shift patterns have high frequency content for a short time.

Wavelet transformation retains the time variable information in the signal. The transformation is, instead, on scale and frequency. It allows frequency analysis and statistical analysis, with time captured in the multiple levels of decomposition [20]. However, proper choices of wavelet family, order and decomposition level are needed to retain the signal characteristics. Based on these observations, it was decided to try the multi-resolution wavelet analysis for recognition of CCPs. The multi-resolution wavelet analysis (MRWA) provides a collection of the mathematical theory to denote a function by means of projection onto a nested sequence of approximation spaces, where the wavelet coefficients will be applied as the parameter that determines where the data distribution can be coarsened or refined. Multi-resolution analysis was developed by Mallat [20] as an efficient and practical filtering algorithm. It was created as a theoretical basis to denote signals that decompose in finer and finer detail.

The first stage of decomposition will give the first level approximation which if decomposed will give the second level approximation and so on. Detail analysis is applied with a contracted, high frequency version of the mother wavelet, while approximation analysis is applied with a dilated, low frequency version of the same wavelet.

Denoting the wavelet coefficients with $h(i,j)$ and introducing the scaling function $\phi(t)$ as

$$\phi(t) = \sum_{j,k} h(j,k)\psi_{j,k}(t) \tag{1}$$

it will be proved that

$$f(t) = \sum_{n} c_j(n)\phi_{j,n} + \sum_{n} d_j(n)\psi_{j,n} \tag{2}$$

The first sum in $f(t)$ is the approximation and the second one is the details loosed during it. The approximation coefficients c_j and the details coefficients d_j for each level of decomposition can be found based on the coefficients derived from the precedent level as:

$$c_{j-1}(n) = \sum_{k} h(k-2n)c_j(k) \tag{3}$$

$$d_{j-1}(n) = \sum_{k} g(k-2n)c_j(k) \tag{4}$$

$$\text{With}\, g(n) = (-1)^n h(1-n) \tag{5}$$

As Fig. 2 shown, the original signal S is decomposed into four level by one-dimensional wavelet.

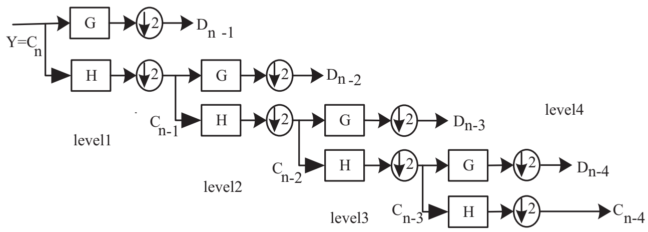

Fig. 2. Multi resolution decomposition of signal Y.

With G being a low-pass filter and H being a high-pass filter according to:

$$G\{a_n\} = \sum_k g(n-2k)a_k \tag{6}$$

$$H\{a_n\} = \sum_k h(n-2k)a_k \tag{7}$$

$$\text{and } c_{j-1}(n) = H\{c_j\} = \sum_k h(k-2n)c_j(k) = H \times c_j \tag{8}$$

$$d_{j-1}(n) = G\{c_j\} = \sum_k g(k-2n)c_j(k) = G \times c_j \tag{9}$$

3. Support vector machine (SVM)

We have proposed a multi-class SVM based classifier (MCSVM) that has a hierarchical structure. SVMs were introduced on the foundation of statistical learning theory. The basic SVM deals with two-class problems; however, it can be developed for multi-class classification [21]. Following subsections brief describe the binary SVM and MCSVM.

3.1. Binary SVM (BSVM)

SVM performs classification tasks by constructing optimal separating hyper-planes (OSH). OSH maximizes the margin between the two nearest data points belonging to two separate classes. Suppose the training set, (x_i, y_i), $i = 1, 2, \ldots, l$, $x \in R^d$, $y \in \{-1, +1\}$ can be separated by the hyper-plane $w^T x + b = 0$, where $\vec{w}$ is the weight vector and b is bias. If this hyper-plane maximizes the margin, then the following inequality is valid for all input data:

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1, \quad \text{for all } \mathbf{x}_i \quad i = 1, 2, \ldots, l \tag{10}$$

The margin of the hyper-plane is $2/\left\|\vec{w}\right\|$. Thus, the problem is the maximizing of the margin by minimizing of $\|w\|^2$ subject to (1). This is a convex quadratic programming (QP) problem and Lagrange multipliers (a_i, $i = 1, \ldots, l$; $\alpha_i \geq 0$) are used to solve it:

$$L_P = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l}\alpha_i[y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1] \tag{11}$$

After minimizing L_P with respect to w and b, the optimal weights are given by:

$$\mathbf{w}^* = \sum_{i=1}^{l}\alpha_i^* y_i \mathbf{x}_i \tag{12}$$

The dual of the problem is given by [22]:

$$L_d = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j \tag{13}$$

To find the OSH, it must maximize L_d under the constraints of $\sum_{i=1}^{l}\alpha_i y_i = 0$. The Lagrange multipliers are only non-zero ($\alpha \succ 0$) when $y_i(\mathbf{w}^T\mathbf{x}_i + b) = 1$. Those training points, for which the equality in (1) holds, are called support vectors (SV) that can satisfy $\alpha_i \succ 0$. The optimal bias is given by:

$$b^* = y_i - \mathbf{w}^{*T}\mathbf{x}_i \tag{14}$$

for any support vector $\mathbf{x}_i$. The optimal decision function (ODF) is then given by:

$$f(\mathbf{x}) = sgn(\sum_{i=1}^{l} y_i\alpha_i^*\mathbf{x}^T\mathbf{x}_i + b^*) \tag{15}$$

where α_i^*'s are optimal Lagrange multipliers. For input data with a high noise level, SVM uses soft margins can be expressed as follows with the introduction of the non-negative slack variables ξ_i, $i = 1, \ldots, l$:

$$y_i(w^T x_i + b) \geq 1 - \xi_i \quad \text{for} \quad i = 1, 2, \ldots, l \tag{16}$$

To obtain the OSH, it should be minimizing the $\Phi = 1/2\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\xi_i^k$ subject to (16), where C is the penalty parameter, which controls the tradeoff between the complexity of the decision function and the number of training examples, misclassified.

In the nonlinearly separable cases, the SVM map the training points, nonlinearly, to a high dimensional feature space using kernel function $K(\vec{x}_i, \vec{x}_j)$, where linear separation may be possible. There are several Kernel functions:

Linear:

$$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i.\vec{x}_j \tag{17}$$

Gaussian radial basis function (GRBF):

$$K(x_i, x_j) = \exp\frac{-\left\|x_i - x_j\right\|^2}{2\sigma^2} \tag{18}$$

Polynomial:

$$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i.\vec{x}_j + 1)^d \tag{19}$$

Sigmoid:

$$K(\vec{x}_i, \vec{x}_j) = \tanh(\gamma\vec{x}_i.\vec{x}_j + \eta) \tag{20}$$

where σ, d, γ and η are the parameters of the kernel functions. After a kernel function is selected, the QP problem will become:

$$L_d = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{21}$$

the α_i^* is derived by:

$$\alpha_i^* = arg\max_{\alpha} L_d$$
$$0 \leq \alpha_i \leq C; \quad i = 1, 2, \ldots, l; \quad \sum_{j=1}^{l}\alpha_i y_i = 0 \tag{22}$$

After training, the following, the decision function, becomes:

$$f(\mathbf{x}) = sgn(\sum_{i=1}^{l} y_i\alpha_i^* K(\mathbf{x}, \mathbf{x}_i) + b^*) \tag{23}$$

The performance of SVM can be controlled through the term C and the kernel parameter which are called hyper-parameters. These parameters influence on the number of the support vectors and the maximization margin of the SVM. The suitable selection of parameters of SVM plays an important role on the classification

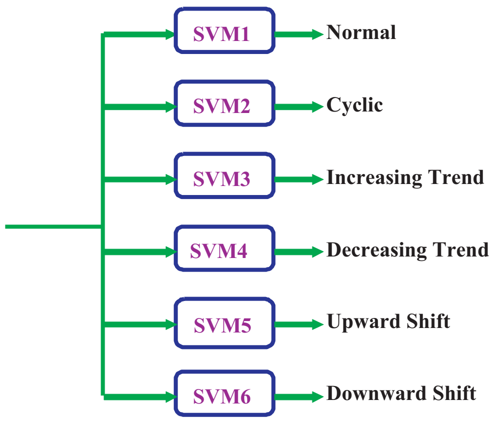

Fig. 3. Multi-class SVM-based classifier.

performance. In this paper, genetic algorithms are applied to select the parameters of SVM.

3.2. Multi-class SVM-based classifier (MCSVM)

There are two widely used methods to extend binary SVMs to multi-class problems [22]. One of them is called the one-against-all (OAA) method. Suppose we have a P-class pattern recognition problem, P independent SVMs are constructed and each of them is trained to separate one class of samples from all others. When testing the system after all the SVMs are trained, a sample is input to all the SVMs. Suppose this sample belongs to class P_1 ideally only the SVM trained to separate class $P1$ from the others can have a positive response. Another method is called one-against-one (OAO) method. For a P-class problem, $P(P-1)/2$ SVMs are constructed and each of them is trained to separate one class from another class. Again, the decision of a testing sample is based on the voting result of these SVMs. In this paper we have used the OAA method. The structure of this classifier is showed in Fig. 3.

4. Genetic algorithm and hybrid intelligent system

As mentioned in Section 1, in this study, we proposed a novel hybrid intelligent method (HIM) that conducts finding the optimum parameters of the classifier. This optimization has done by a genetic algorithm. Following subsections describe the GA and HIM.

4.1. Genetic algorithm

Genetic algorithm (GA) is a stochastic optimization algorithm, which adopts Darwin's theory of survival of the fittest [23]. Fig. 4 shows an overview of the genetic algorithm.

1. [Start] Generate random population of *n* chromosomes (suitable solutions for the problem)
2. [Fitness] Evaluate the fitness *f(x)* of each chromosome *x* in the population
3. [New population] Create a new population by repeating following steps until the new population is complete
1. [Selection] Select two parent chromosomes from a population according to their fitness (the better fitness, the bigger chance to be selected)
2. [Crossover] With a crossover probability cross over the parents to form new offspring (children). If no crossover was performed, offspring is the exact copy of parents.
3. [Mutation] With a mutation probability mutate new offspring at each locus (position in chromosome).
4. [Accepting] Place new offspring in the new population
4. [Replace] Use new generated population for a further run of the algorithm
5. [Test] If the end condition is satisfied, stop, and return the best solution in current population
6. [Loop] Go to step 2

Fig. 4. Outline of the basic genetic algorithm.

4.1.1. Chromosome representation, initial population, fitness function and selection

In GA, a candidate solution for a specific problem is called an individual or a chromosome and consists of a linear list of genes. How to encode a solution of the problem into a chromosome is a key issue for genetic algorithms. In this paper the real code representation is used. A chromosome of the proposed HIM includes the free parameter of C and σ of a SVM classifier. Then, each chromosome has two genes (segments). The free parameter of C is represented as natural numbers and the free parameter of σ is positive number. The evolutionary algorithms start with an arbitrarily initialized population. The initial population is being represented as $P = \{p_1, p_2, \ldots, p_{N_s}\}$, where N_s denotes the size of initial population.

Each individual is decided by an evaluating mechanism to obtain its fitness value. Based on this fitness value and undergoing genetic operators, a new population is generated iteratively with each successive population referred to as a generation. The fitness function is based on classification accuracy of classifier, which is as follows:

$$\text{Fitness function} = \frac{y_t}{y_t + y_f} \times 100 \tag{24}$$

where y_t and y_f denote the number of true and false classifications, respectively.

In a genetic algorithm, the selection of individuals to produce the successive generations plays a vital role. Two well known selection methods are used to select the population individuals for the mating pool: tournament method and roulette wheel selection (RWS) [23].

4.1.2. Genetic operators

Genetic operators are two basic types: mutation and crossover. Mutation alters one individual to produce a single new solution whereas crossover produces two new individuals from two existing individuals (parents). In this study, an arithmetic crossover [24] is applied. Crossover is performed to each pair of the mating pool with a crossover probability, P_c. This operator tends to enable to the evolutionary process to move toward promising regions of the search space. The mutation operation follows the crossover to determine whether a chromosome should mutate to the next generation or not with a mutation probability, P_m. In this study, uniform mutation method is applied in the presented model, which is as follows:

$$X_{old} = \{x_1, x_2, \ldots, x_n\}, \tag{25}$$

$$X_k|_{new} = L_k + r \times (U_k - L_k), \tag{26}$$

$$X_{new} = \{x_1, x_2, \ldots, X_k|_{new}, \ldots, x_n\} \tag{27}$$

where n denotes the number of parameters, r represents a random number in the range (0,1), and k is the mutation location. L_k and U_k are the lower and upper bounds in location k on the parameter, respectively. X_{old} represents the population before the mutation operation; and X_{new} represents the new population following the mutation operation.

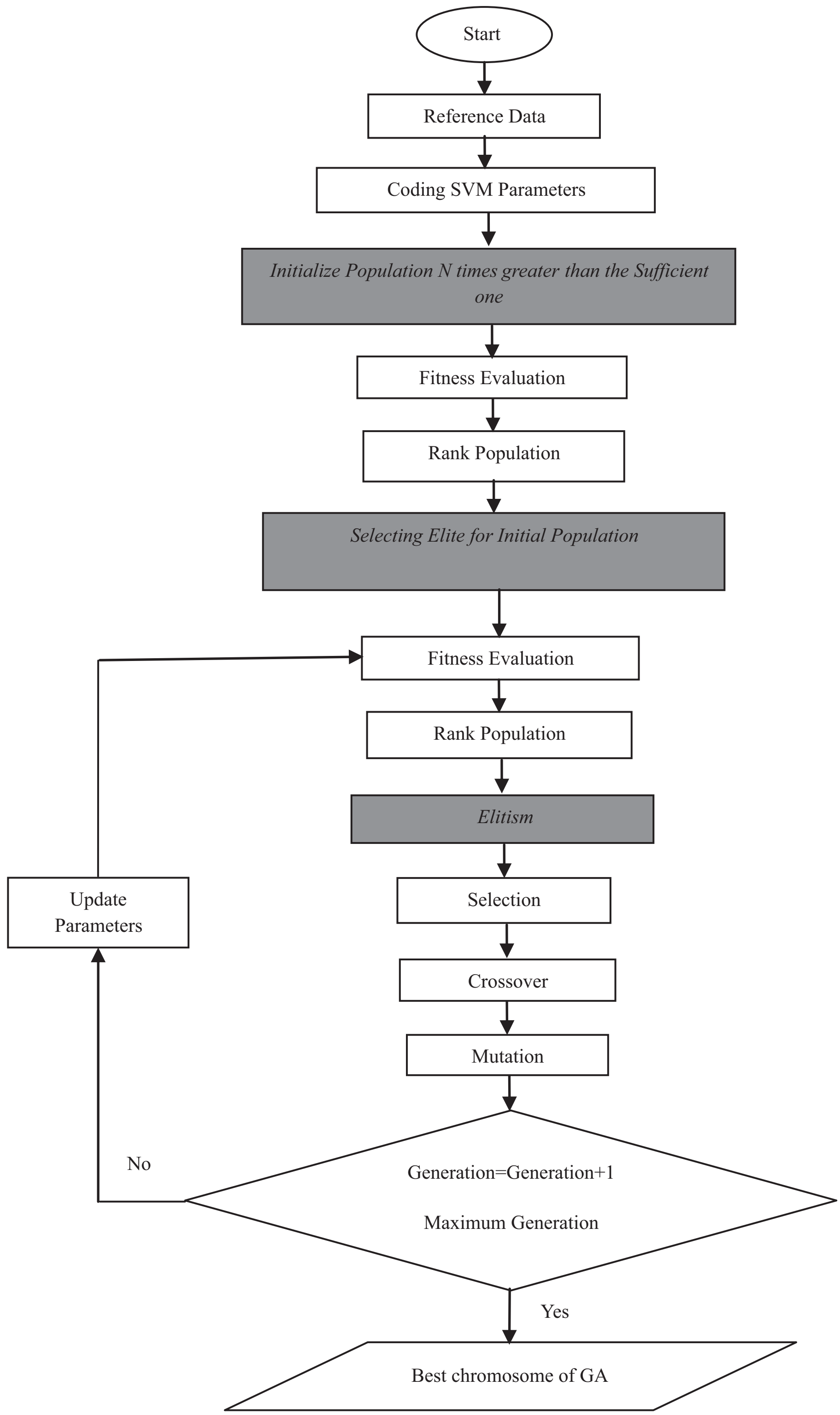

Fig. 5. Flowchart of the proposed algorithm.

4.2. Hybrid intelligent method (HIM)

HIM is a modified version of GA which properly find the optimized parameters. Fig. 6 shows the structure of the proposed hybrid intelligent method.

As it is shown in Fig. 5, the proposed algorithm is started with a random search, N times of elite initial population. Then, fitness function is calculated for each of the selected population. After ranking the population in the previous stage, modified elite initial population is obtained. It is also considered in the calculation of the fitness function, in the next step.

Furthermore, as a novel approach, we have implemented a type of elitism such that at each generation the best string seen up to that generation is preserved into the next population. This eliminates the loss of good chromosomes and keeps the diversity of population. In this method, generational replacement with probability P_r

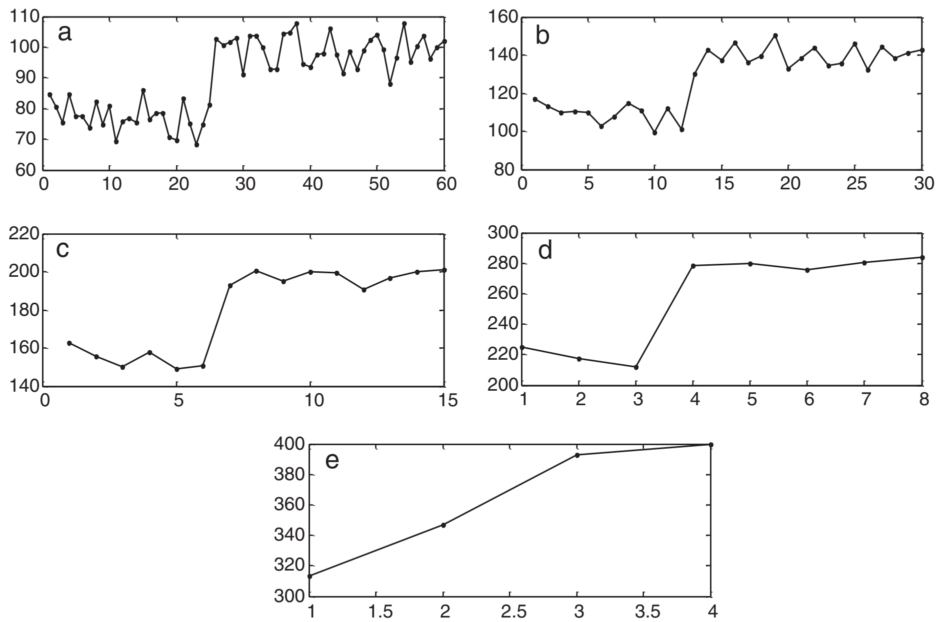

Fig. 6. (a) Original signal for a upward shift pattern of control chart patterns; (b) level 1 DWT approximated coefficients for the approximation of (a); (c) level 2 DWT approximated coefficients for the approximation of (a); (d) level 3 DWT approximated coefficients for the approximation of (a); (e) level 4 DWT approximated coefficients for the approximation of (a).

is applied, in which parents are replaced with children, while the $n_s \times P_r$ best chromosomes are maintained. A compression factor R_c is then applied in some generations to reduce the search space of parameters as:

$$U_j^{new} = (1 - R_c)C_{best,j} + R_c U_j^{old} \tag{28}$$

$$L_j^{new} = (1 - R_c)C_{best,j} + R_c L_j^{old} \tag{29}$$

where U_j and L_j are the upper and lower bounds of the j^{th} individual genes, respectively. And $C_{best,j}$ is the j^{th} first-ranked individual gene in the generation in which the domain compression operator is applied. Simulation results and efficiency of the proposed method will be discussed in the next section.

5. Simulation results

In this section we evaluate the performance of proposed recognizer. For this purpose we have used the practical and real world data [25]. This dataset contains 600 examples of control charts. For this study, we have used 20% of data for training the classifier and the rest for testing.

The feature extraction procedure is implemented based on the multi resolution wavelet decomposition of the CCPs to their time-frequency localized coefficients. Based on the extensive simulations it is found that, the SVM with GRBF kernel has better results than the other kernels such as linear and polynomial. So, in this study, for the SVM classifier, GRBF was adopted as kernel function.

We apply GA for finding the optimum parameters of SVMs. What can be grasped from the literature is that good GA performance requires the choice of a moderate population size, a high crossover probability, and a low mutation probability. Thus, Table 1 summarizes all HIM training parameters.

In this section, we have done several experiments for evaluating of the proposed method:

5.1. Experiment 1: classification in the whole original hyper-dimensional feature space

As mentioned earlier, in this experiment, we applied the SVM classifier directly on the entire original hyper-dimensional feature space, which is made up of 12 features. Table 2 shows performance comparison between the one-against-one (OAO) and one-against-all (OAA) methods of the SVM classifier with different kernels. We chose the best SVM classifier parameter values to maximize this prediction.

As reported in Table 2, percent of the OAO and OAA methods achieved with the SVM classifier based on the Gaussian kernel (SVM–RBF) on the test set were equal to 96.46% and 97.29%, respectively. These results were better than those achieved by the SVM-linear and the SVM-poly. Indeed, percent of the OAO (and OAA) methods were equal to 90.83% (91.67%) for the SVM-linear classifier and 93.75% (94.79%) for the SVM-poly classifier. As it can be seen from Table 3, separating increasing trend (IT) and the up shift (US) patterns as well as between the decreasing trend (DT) and the downward shift (DS) patterns is so difficult due to similarity between them. In addition, it provides reference classification accuracies in order to quantify the capability of the proposed GA–SVM (HIM) classification system to further improve these interesting results.

Table 1
HIM training parameters.

Parameter	Value
Number of generations	$N_s = 100$
Population size	$N_g = 40$
Gamma range	$\gamma \in (0, 1)$
Penalty range	$C \in (0 - 100)$
Selection type	Roulette wheel and tournament
Crossover type	Arithmetic
Mutation type	Uniform
Elitism	$P_r = 0.05$
Crossover probability	$P_c = 0.99$
Mutation probability	$P_m = 0.12$
Compression factor	$R_c = 0.99$

Table 2
Performances comparison of the one-against-one (OAO) method and one-against-all (OAA) methods with different kernels.

Method		Recognition accuracy (%)		NOR	CYC	IT	DT	US	DS
			NOR	98.75	1.25	0	0	0	0
			CYC	13.75	86.25	0	0	0	0
SVM-linear	OAO	90.83	IT	0	0	82.5	0	17.5	0
			DT	0	0	0	98.75	0	1.25
			US	0	1.25	8.75	0	90	0
			DS	0	0	0	12.5	0	87.5
			NOR	97.5	2.5	0	0	0	0
			CYC	16.25	81.25	0	2.5	0	0
SVM-linear	OAA	91.67	IT	0	0	96.25	0	3.75	0
			DT	0	0	0	100	0	0
			US	0	0	11.25	0	88.75	0
			DS	0	0	0	13.75	0	86.25
			NOR	100	0	0	0	0	0
			CYC	0	100	0	0	0	0
SVM-poly	OAO	*93.75*	IT	3.75	0	98.75	0	1.25	0
			DT	1.25	0	0	95	0	3.75
			US	0	7.5	1.25	0	91.25	0
			DS	0	15	0	7.5	0	77.5
			NOR	92.5	7.5	0	0	0	0
			CYC	17.5	82.5	0	0	0	0
SVM-poly	OAA	*94.79*	IT	0	0	98.75	0	1.25	0
			DT	0	0	0	100	0	0
			US	0	1.25	2.5	0	96.25	8.75
			DS	0	0	0	1.25	0	98.75
			NOR	100	0	0	0	0	0
			CYC	16.25	93.75	0	0	0	0
SVM-rbf	OAO	*96.46*	IT	0	0	98.75	0	1.25	0
			DT	0	0	0	100	0	0
			US	0	3.75	2.5	0	97.5	0
			DS	0	0	0	1.25	0	98.75
			NOR	100	0	0	0	0	0
			CYC	7.5	92.5	0	0	0	0
SVM-rbf	OAA	*97.29*	IT	0	0	98.75	0	1.25	0
			DT	0	0	0	98.75	0	1.25
			US	0	3.75	1.25	0	95	0
			DS	0	0	0	1.25	0	98.75
			NOR	100	0	0	0	0	0
			CYC	0	100	0	0	0	0
GA-SVM (HIM)	OAO	99.17	IT	0	0	98.75	0	1.25	0
			DT	0	0	0	100	0	0
			US	0	0	1.25	0	98.75	0
			DS	0	0	0	2.5	0	97.5
			NOR	100	0	0	0	0	0
			CYC	0	100	0	0	0	0
GA-SVM (HIM)	OAA	99.37	IT	0	0	98.75	0	1.25	0
			DT	0	0	0	100	0	0
			US	0	0	0	0	100	0
			DS	0	0	0	1.25	0	98.75

As already stated the features play a vital role in classification of digital signal types. In order to investigate the effectiveness of the selected features, we have used the features that have been introduced in some references. Table 3 shows this comparison. Other simulations setup is the same. Results imply that the proposed features have effective properties in control chart patterns representation.

Table 3
Comparison among the proposed features and some the features that have introduced in other references.

Ref. no.	Features	Total accuracy
[16]	Correlation between the input and various reference	93.94
[17]	Vectors statistical correlation coefficients	95.19
[14]	Shape features	96.28
Proposed features (SVM-rbf)(OAA)	Multi-resolution wavelets	97.29
Proposed method (GA-SVM)(OAA)	Multi-resolution wavelets	99.37

5.2. Experiment 2: comparison HIM with classical SVM

In the second experiment, we compare HIM and classical SVM with the Haar wavelet at various levels of wavelet decomposition which results shown in Table 4. In this experiment, MRW was performed up to level 5 on the CCPs. The approximated coefficients (see Fig. 6) are computed in various levels for each of the CCPs. Wavelets

Table 4
Performance comparisons of classical SVM and HIM with different levels of the wavelet decomposition.

Various levels of the wavelet decomposition	Number of features	Classical SVM Classification accuracy (%) Testing	GA based SVM (HIM) Classification accuracy (%) Testing
Level 1	30	96.88	98.96
Level 2	15	96.87	99.37
Level 3	8	93.12	95.83
Level 4	4	86.04	89.17
Level 5	2	82.08	81.46

Table 5
The best parameters of SVM classifier with the GRBF kernel function and estimation parameter for different runs.

Run	C	γ	Best fitness
1	332	0.000712	99.37
2	328	0.000749	99.37
3	230	0.000694	99.37
4	278	0.000739	99.37
5	153	0.000731	99.37
Mean	264.2	0.000725	99.37
	±111.2	±0.000031	0

are able to separate deterministic and stochastic components of a signal by capturing deterministic changes in a relatively small number of large coefficients. Classifier operates on a reduced set of coefficients which is obtained from the MRWA stage. Adequate recognition performance was provided with only an few feature using 'Haar' wavelet (see Fig. 6) as compared to an 60 features. This resulted to reduce the complexity of classifier in an input vector. Also, Fig. 7 shows the distribution of the six classes in the two-dimensional space planes when level of wavelet decomposition is 2.

This reduced set of coefficients is found to improve efficiency and classification accuracy. By comparing the results of HIM and classical SVM, it can be seen that the results of HIM are better than those of classical SVM for testing set. Fig. 8 shows the influences of the number of feature on the testing accuracy. For HIM, the classification accuracy increases with the increment of level of the wavelet decomposition and classification accuracy maximum (99.37%) is achieved when number of feature is equal to 15 or level of wavelet decomposition is 2. It tends to decrease as the wavelet decomposition increases. This can be explained due to smaller number of features. A drastic reduction of features, however, can lead to a decrease in the testing performance.

5.3. Experiment 3: performance evaluation with optimization in different runs

In this experiment, for evaluating the performance of the proposed algorithm, five different runs have been performed at second level of decomposition. The Genetic algorithm finds the best combination of the free parameters of SVM classifier to gain the fitness function maximum. The optimal values of SVM classifier parameters, i.e. γ parameter and optimal values of C parameters estimated by proposed algorithm are shown in Table 5. The good agreement has been observed between the estimation and preselected param-

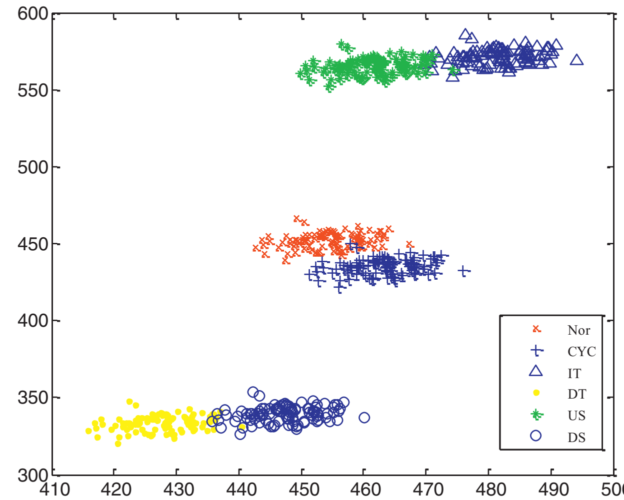

Fig. 7. Clustering of sample data set which the level 5 approximated coefficients are computed for each of the CCPs.

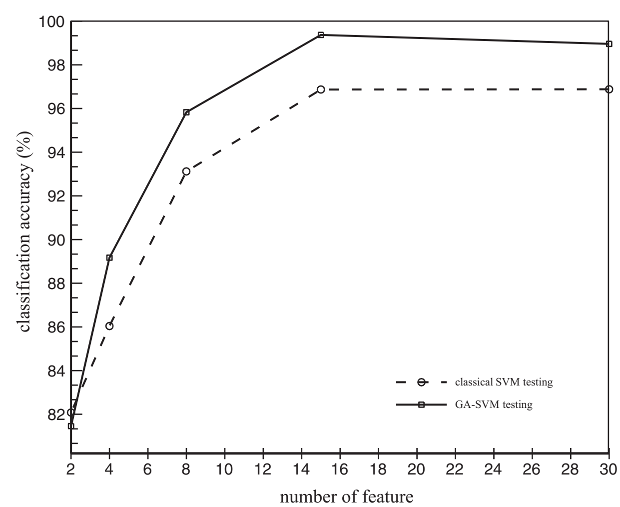

Fig. 8. Classification performance comparison for different number of feature.

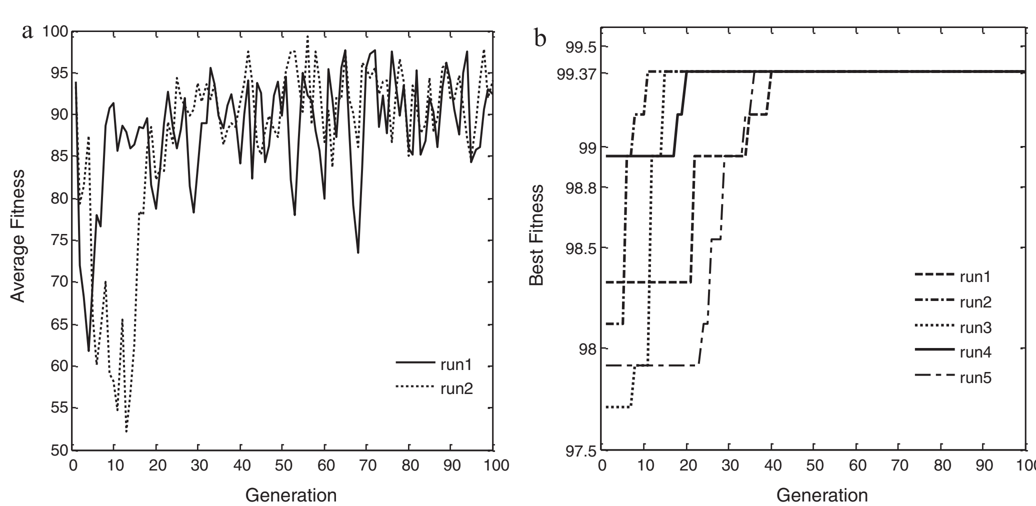

Fig. 9. Average (a) and the best fitness (b) evolution of fitness function for different runs.

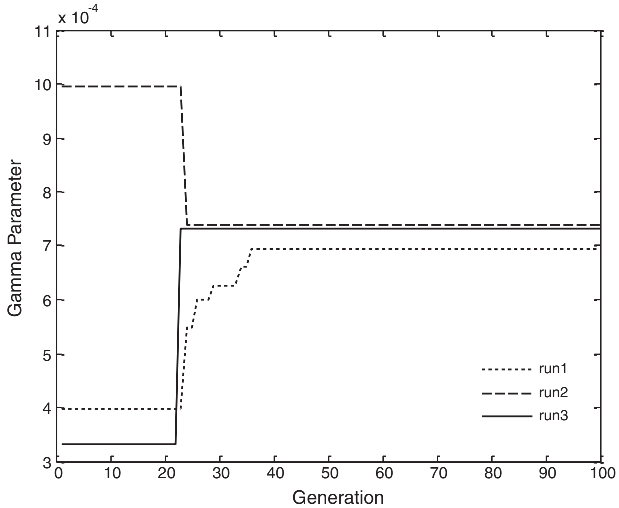

Fig. 10. Estimated parameter (γ) for GRBF kernel function.

Table 6
Comparing the performance of the classifier with various wavelet and various level.

Levels wavelet	Level 1	Level 2	Level 3	Level 4	Level 5
Haar	98.96	99.37	95.83	89.17	81.46
Db2	98.96	98.33	97.08	92.50	80.21
Db3	98.75	98.33	96.46	93.12	88.75
Db4	98.75	98.96	98.33	97.29	95.21
Db5	98.54	97.50	97.92	96.67	89.17
Db6	98.75	98.12	98.12	93.33	89.79
Db7	98.75	98.12	98.54	96.67	89.38
Db8	98.96	98.33	99.17	97.08	90.83
Db9	98.96	98.54	98.75	96.25	91.46
Db10	98.75	98.75	98.12	96.87	91.67
Coif2	98.96	98.96	97.29	91.87	87.08
Coif3	98.96	98.33	97.92	94.17	90.42
Coif4	98.96	98.96	97.08	95.21	90.42
Coif5	98.96	98.54	97.29	94.37	91.67
Sym2	98.96	98.33	97.08	92.50	80.21
Sym3	98.75	98.33	96.46	93.12	88.75
Sym4	98.96	98.75	96.46	90.63	84.17
Sym5	99.17	98.96	98.54	95.63	85.21
Sym6	98.96	98.75	98.75	93.54	84.79
Sym7	98.96	98.54	97.92	95.62	86.25
Sym8	98.96	98.33	96.46	92.71	86.04
Bior1.3	98.96	98.75	94.58	91.04	86.25
Bior1.5	99.17	98.96	98.33	95.63	87.71
Bior2.2	98.75	97.50	96.88	90.62	80.00
Bior2.4	98.75	97.92	97.71	91.87	87.08
Bior2.8	98.96	97.50	96.67	94.58	92.29
Bior3.1	97.92	92.50	76.25	70.63	55.83
Bior3.3	98.33	97.50	91.67	91.67	81.46
Bior3.5	98.54	95.83	96.67	96.04	90.00
Bior3.7	98.54	97.08	92.29	93.33	88.96
Bior3.9	98.33	93.75	90.83	94.58	92.92
Bior4.4	98.75	99.17	98.33	91.87	86.46
Bior5.5	98.75	98.75	97.29	90.00	85.42
Bior6.8	98.96	98.12	96.67	93.96	89.17

Table 8
Comparison the performance of proposed classifier (HIM) with other classifiers.

Classifier	Parameters	Recognition accuracy (%)
PNN	Spread = 13	96.04
RBF	Spread = 35	95.83
MLP (BP)	Hidden neurons = 35	92.7
MLP (RP)	Hidden neurons = 46	93.75
SVM	C = 500 σ = 0.0001	97.29
HIM (proposed method)	C = 328 σ = 0.000749	99.37

eter. The proposed algorithm successfully finds the global optimum just with 100 generations.

Fig. 9 shows a typical increase of both the fitness (classification accuracy) of the best individual and the average fitness of the population obtained from HIM for different runs. As indicated in Fig. 9b, its fitness curves gradually improved from iteration 0 to 100, and exhibited no significant improvements after iteration 40 for the five different runs. The optimal stopping iteration to get the highest validation accuracy for the five different runs was around iteration 10–40.

Value of gamma parameter of the SVM classifier in three different runs of the program with 100 iterations is presented in Fig. 10 In each different runs of program, GA first generates random value gamma (as seen in the figure) then it searches for better values of them that produce better fitness. Usually after 30 iterations, the algorithm converges to the best value of gamma parameter of the SVM classifier.

5.4. Experiment 4: performance evaluation HIM with different mother wavelets

We have claimed that Haar wavelet has better performance than others. In order to indicate this term, we have evaluated the performance of the recognizer (HIM) with different mother wavelets. The maximum value of the fitness function (classification accuracy) based on HIM with the other mother wavelet at the various levels of wavelet decompositionis shown in Table 6. The HIM applied to achieve the free parameters of SVM classifier (C, γ). From Table 6, it is found that the Haar wavelet with the second level of decomposition shows better recognition accuracy than the other wavelet with various levels. Table 7 shows the confusion matrix for the best performances of recognizer with the Haar wavelet at second levels of the decomposition wavelet.

5.5. Experiment 5: comparison with different classifier

The performance of the proposed classifier has been compared with other classifiers for investigating the capability of the proposed classifier, as indicated in Table 8. In this respect, probabilistic neural networks (PNN) [26], radial basis function neural network (RBFNN) [27] and Multilayered perceptron (MLP) neural network with different training algorithm such as: Back propagation (BP) learning algorithm [28] and with Resilient propagation (RP) learning algorithm [29] are considered. They comprise parameters

Table 7
Confusion matrix of Haar wavelet and level = 2.

	Normal	Cyclic	Up trend	Down trend	Up shift	Down shift
Normal	100%	0	0	0	0	0
Cyclic	0	100%	0	0	0	0
Up trend	0	98.75%	0	1.25%	0	0
Down trend	0	0	100%	0	0	0
Up shift	0	0	0	0	100%	0
Down shift	0	0	0	2.5%	0	97.5%

Table 9
A summary of different classification algorithms together with their reported results used measures of the accuracy.

Ref. no.	Number of CCP types	Features	Classifier	Total recognition accuracy (%)
[5]	6	Un-processed data	MLP (RSFM)	97.46
[6]	6	Un-processed data	MLP	94.30
[7]	6	Un-processed data	PNN	95.58
[8]	6	Un-processed data	MLP	93.73
[30]	4	Un-processed data	MLP (SPA)	96.38
[9]	6	Un-processed data	LVQ	97.7
This work	6	MRWA	GA-SVM	99.37

which should be readjusted in any new classification. Furthermore, those parameters regulate the classifiers to be best fitted in for classification task. In most cases, there is no classical method for obtaining the values of them and therefore, they are experimentally specified through try and error. It can be seen from Table 8 that the proposed method has better recognition accuracy than other classifiers.

5.6. *Comparison of the proposed method (HIM) with other methods in the literature*

Several researchers have addressed the arrhythmia detection and classification problem using the CCP signals directly or by analyzing the pattern rate variability signal in the past. Direct comparison with other works is difficult in control chart pattern recognition problem. This is mainly because of the fact that there is no single unified data set available. Different setup of patterns (in case of number of training and testing samples and the number of patterns) will lead to different performance. Besides, there are many different kinds of benchmarking systems used for system's quality. This causes difficulties for direct numerical comparison. A summary of different methods together with their reported results used measures of accuracy is summarized in Table 9.

As for neural network-based CCPs recognizers, Le et al. [5] introduced a new ANN model and their numerical simulations showed that this model a recognition accuracy about 97.46% for recognition of six types of CCP. Pham and Oztemel [6] reported a generalization rate of 94.30%. Cheng and Ma [7] have gained recognition accuracy (RA) about 95.58%. However, the performance for lower patterns is reported to be less than 90%. The proposed method in [8] reached a RA about 93.73% of classification accuracy for recognition of six types of CCPs. In [30], Guh and Tannock proposed a sequential pattern analysis method and reported a classification rate about 96.38%. In [9] the authors used LVQ neural network and achieved a RA about 97.70. Comparing to these papers, an effective algorithm is proposed in the current work which provides a better accuracy over a wider range of different types of CCPs (six different classes).

6. Conclusion and discussion

Control chart patterns (CCPs) are important statistical process control tools for determining whether a process is run in its intended mode or in the presence of unnatural patterns. In this paper, we proposed a method for CCP classification based on multi resolution wavelet and HIM. This paper focuses on the improvement of the classical SVM model by means of the integration of GA and SVM. This study presents the methods for improving SVM performance in two aspects: feature extraction and parameter optimization. GA is used to select appropriate parameters of SVM classifier. This paper applies, the proposed HIM model to the classification problem using a data set. We evaluated the proposed model using the data set and compared it with other models. The results showed that the proposed model was effective in finding the parameters of SVM, and that it improved classification accuracy. The results also demonstrate that the choice of the feature has an influence on the appropriate kernel parameters. The simulation results, using the novel HIM model, show that the highest recognition accuracy (RA) (99.37%) is achieved at the second level of wavelet decomposition with the Haar wavelet.

References

[1] D.C. Montgomery, Introduction to Statistical Quality Control, 5th ed., John Wiley, Hoboken, NJ, USA, 2005.
[2] J.A. Swift, J.H. Mize, Out-of-control pattern recognition and analysis for quality control charts using lisp-based systems, Computers and Industrial Engineering 28 (1995) 81–91.
[3] C.H. Wang, W. Kuo, Identification of control chart patterns using wavelet filtering and robust fuzzy clustering, Journal of Intelligent Manufacturing 18 (2007) 343–350.
[4] C.H. Wang, R.S. Guo, M.H. Chiang, J.Y. Wong, Decision tree based control chart pattern recognition, International Journal of Production Research (2008) 124–134.
[5] Q. Le, X. Goal, L. Teng, M. Zhu, A new ANN model and its application in pattern recognition of control charts, in: Proc. IEEE. WCICA, 2008, pp. 1807–1811.
[6] D.T. Pham, E. Oztemel, Control chart pattern recognition using neural networks, Journal of Systems Engineering 2 (1992) 256–262.
[7] Z. Cheng, Y. Ma, A research about pattern recognition of control chart using probability neural network, in: Proc. ISECS, 2008, pp. 140–145.
[8] S. Sagiroujlu, E. Besdoc, M. Erler, Contro chart pattern recognition using artificial neural networks, Turkish Journal of Electrical Engineering 8 (2000) 137–147.
[9] D.T. Pham, E. Oztemel, Control chart pattern recognition using linear vector quantization networks, International Journal of Production Research 32 (721–729) (1994) 256–262.
[10] R.S. Guh, Robustness of the neural network based control chart pattern recognition system to non-normality, International Journal of Quality & Reliability management 19 (2002) 97–112.
[11] S.K. Gauri, S. Chakraborty, A study on the various features for effective control chart pattern recognition, International Journal of Advanced Manufacturing Technology 34 (2007) 385–398.
[12] Y. Gou, K.J. Dooley, Identification of change structure in statistical process control, International Journal of Production Research 30 (1992) 1655–1669.
[13] M.A. Wani, S. Rashid, Parallel algorithm for control chart pattern recognition, in: Proc. IEEE the Fourth International Conference on Machine Learning and Applications (ICMLA'05), 2005.
[14] R.S. Guh, Y.R. Shiue, On-line identification of control chart patterns using self-organized approaches, International Journal of Production Research 43 (2005) 1225–1254.
[15] M. Pacella, Q. Semeraro, A. Anglani, Adaptive resonance theory-based neural algorithms for manufacturing process quality control, International Journal of Production Research 42 (2004) 4581–4607.
[16] A.M. Al-Ghanim, L.C. Ludeman, Automated unnatural pattern recognition on control charts using correlation analysis techniques, Computers and Industrial Engineering 32 (1997) 679–690.
[17] J.H. Yang, M.S. Yang, A control chart pattern recognition scheme using a statistical correlation coefficient method, Computers and Industrial Engineering 48 (2005) 205–221.
[18] E.A. Mohamed, andA.Y. Abdelaziz, A.S. Mostafa, A neural network-based scheme for fault diagnosis of power transformers, Electric Power Systems Research 75 (1) (2005) 29–39.
[19] V. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.
[20] S.A. Mallat, A Wavelet Tour of Signal Processing, Academic Press, New York, 1998.
[21] C. Cortes, V. Vapnic, Support Vector Network, Machine Learning 20 (1995) 1–25.
[22] C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2 (1998) 121–167.
[23] Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs, 3rd ed., Springer, NY, USA, 1999.
[24] J.F. Doyle, Determining the size and location of a crack in a frame structure, in: Proceeding of the 7th Inverse Problems in Engineering Seminar, OSU, USA, 1995.

[25] http://archive.ics.uci.edu/ml/databases/synthetic_control/synthetic_control.data.html.
[26] D.F. Specht, Probabilistic neural networks, Neural Networks 3 (1990) 109–118.
[27] T. Poggio, F. Girosi, Networks for approximation and learning, Proceedings of the IEEE 78 (1990) 1481–1497.
[28] S. Haykin, Neural Networks: A Comprehensive Foundation, MacMillan, New York, 1999.
[29] M. Riedmiller, H. Braun, A direct adaptive method for faster back propagation learning: the RPROP algorithm, in: Proceedings of the IEEE Int. Conf. On Neural Networks, San Francisco, CA, March 28, 1993.
[30] R.S. Guh, J.D.T. Tannock, A neural network approach to characterize pattern parameters in process control charts, Journal of Intelligent Manufacturing 10 (1999) 449–462.