Expert Systems with Applications 37 (2010) 3177–3183

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine

İsmail Babaoglu *, Oğuz Findik, Erkan Ülker

Department of Computer Engineering, Selcuk University, Konya, Turkey

ARTICLE INFO

Keywords:
Binary particle swarm optimization
Genetic algorithm
Support vector machine
Exercise stress testing
Coronary artery disease

ABSTRACT

The aim of this study is to search the efficiency of binary particle swarm optimization (BPSO) and genetic algorithm (GA) techniques as feature selection models on determination of coronary artery disease (CAD) existence based upon exercise stress testing (EST) data. Also, increasing the classification performance of the classifier is another aim. The dataset having 23 features was obtained from patients who had performed EST and coronary angiography. Support vector machine (SVM) with *k*-fold cross-validation method is used as the classifier system of CAD existence in both BPSO and GA feature selection techniques. Classification results of feature selection technique using BPSO and GA are compared with each other and also with the results of the whole features using simple SVM model. The results show that feature selection technique using BPSO is more successful than feature selection technique using GA on determining CAD. Also with the new dataset composed by feature selection technique using BPSO, this study reached more accurate values of success on CAD existence research with more little complexity of classifier system and more little classification time compared with whole features used SVM.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Exercise stress test (EST) is most commonly used method on diagnosis of angina. It is a noninvasive, inexpensive, easy operable, safe and reproducible method. The main cause of angina is coronary artery disease (CAD) and occurs due to atherosclerosis of the cardiac arteries (Sprigings, 2002). Therefore, EST is one of the first choice noninvasive diagnostic tools in the diagnosis of suspected CAD. Nonetheless, the relatively low sensitivity and specificity of EST for diagnosing CAD has led to limit its clinical usage (San Roman, Vilacosta, Castillo, et al., 1998; Thom et al., 2006).

Artificial intelligence techniques are commonly used in medical diagnosis with an amazing increment day by day, and also it could be seen in the literature.

Fuzzy discrete hidden Markov model is used to classify transcranial Doppler signals to predict the patients whether they are brain diseased or not (Uğuz, Öztürk, Saraçoğlu, & Arslan, 2008). Sepehri et al. proposed a method for automated screening of congenital heart diseases in children by means of heart sound analysis techniques. The method relies on categorizing the pathological murmurs by examining the heart sound energy over specific frequency bands based on the heart sections initiating those (Sepehri et al., 2008). Tantimongcolwata et al. proposed a method for the interpretation of ischemic heart disease pattern of magnetocardiography recordings using backpropagation neural network and direct kernel self-organizing map machine learning approaches (Tantimongcolwata, Naennab, Isarankura-Na-Ayudhyaa, Embrechtsc, & Prachayasittikula, 2008). Least squares support vector machine and backpropagation artificial neural network methods are employed to classify the extracted features obtained from Doppler signals of the heart valve (Comak et al., 2007). Zhidong proposed noninvasive diagnosis method of coronary artery disease based on the instantaneous frequency estimation of diastolic murmurs and support vector machine (SVM) classifier (Zhidong, 2005). Kurt et al. compare performances of machine learning approaches which are logistic regression, classification and regression tree, multi-layer perceptron, radial basis function and self-organizing feature maps in order to predict the presence of CAD by using demographic and medical data (Kurt, Ture, & Kurum, 2008).

Since the use of optimization and feature selection techniques, these literature studies could be more effective and less complex. Binary particle swarm optimization (BPSO) is used as a feature selection method by implementing to the data obtained by mutual information and rough set to increase the effectiveness of the SVM classifier and the classification accuracy (Zhou, Zhou, Liu, & Zhu, 2006). Chuang et al. used improved BPSO in cancer-type classification based on the gene expression profiles (Chuang, Chang, Tu, &

* Corresponding author. Tel.: +90 332 223 20 63; fax: +90 332 241 06 35.
E-mail address: ibabaoglu@selcuk.edu.tr (İ. Babaoglu).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.09.064

Yang, 2008). Yang et al. proposed a feature selection and classification method for hyperspectral images by combining the global optimization ability of particle swarm optimization algorithm and the superior classification performance of a SVM (Yang, Zhang, Deng, & Du, 2007).

In this paper, SVM is used as the classifier. The efficiency of BPSO and genetic algorithm (GA) techniques as feature selection models on determination of CAD existence based upon EST data is investigated. Also, increasing the classification performance of the classifier is aimed.

2. Materials and methods

2.1. Data acquisition

Four hundred eighty patients who underwent EST and coronary angiography (CAG) were included in this study. A total of 23 features are obtained from EST data. Basal demographic characteristics, rest and peak exercise heart rate, blood pressure and exercise time were recorded. The EST results were evaluated by two experienced cardiologists. ST segment depression and elevation occurred 60 ms after the J point were recorded at each derivation in peak exercise. EST was accepted as positive in case $\geqslant 1$ mm ST depression or ST elevation in $\geqslant 2$ contiguous leads seems. Within the first month following the EST, CAG was performed to all patients, and the angiographic images were evaluated by two skilled cardiologists. Presence of $\geqslant 50\%$ narrowing in left main coronary artery or $\geqslant 70\%$ narrowing in other major epicardial coronary arteries indicated severe CAD. Patients with bundle branch blocks (right or, left bundle branch block), pre-excitation syndromes, atrial fibrillation, left ventricular hypertrophy and taking the digoxin were excluded from the study.

2.2. Support vector machine

Support vector machine has been invented by Vapnik (1995) and proposed for classification and regression tasks. SVM has been constructed on a strong statistical learning theory including Vapnik–Chervonenkis dimension and structural risk minimization. Since SVM includes many reliable properties for learning and presents good experimental results, it has been used in many application fields (Kulkarni, Jayaraman, & Kulkarni, 2004; Takeuchi & Collier, 2003; Chen & Wang, 2007).

2.2.1. Linear SVM classifier

Employed training data obey a form; $(x_1, y_1), \ldots, (x_n, y_n)$, $x \in R^N$ and $y \in \{+1, -1\}$. Each data is formed with N dimensional vector and belonging only one of two classes (+1 or −1). Hyperplanes separate two classes from each other to provide following forms for all training data. Thus,

$$\begin{aligned} &(w \cdot x_i) + w_0 \geq +1 - \xi_i, \quad \text{if } y_i = +1 \\ &(w \cdot x_i) + w_0 \leq -1 - \xi_i, \quad \text{if } y_i = -1 \\ &\text{or} \\ &y_i[(w \cdot x_i) + w_0] \geq 1, \qquad i = 1, \ldots, n, \end{aligned} \tag{1}$$

where $\xi_i \geqslant 0$ are slack variables and used for providing a tolerance to some data with small error. If all data satisfy (1) correctly, ξ_i variables will not be used. Optimal hyperplane among all hyperplanes is found by minimizing following formula

$$C \sum_{i=1}^{n} \xi_i + 1/2 ||w||^2, \tag{2}$$

where C is a regularization parameter and providing a trade-off between complexity and classification performance. In other words,

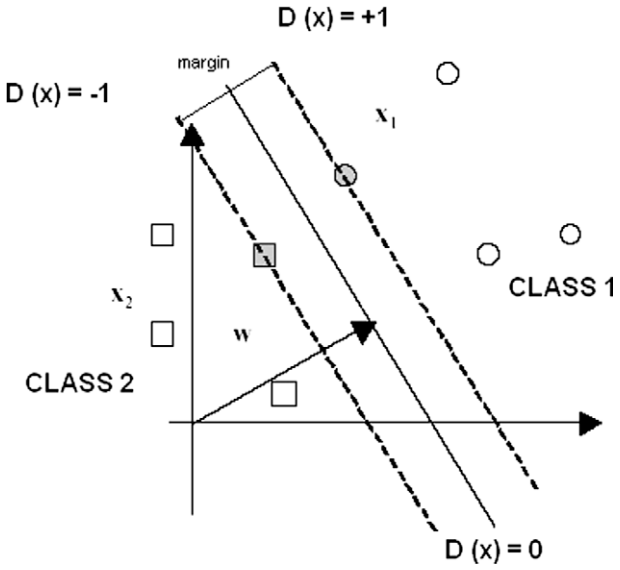

Fig. 1. SVM structure.

the optimal separating hyperplane maximizes the margin illustrated in Fig. 1. Problem is transformed into following dual form of quadratic optimization problem

$$\begin{aligned} &\text{Maximize} \quad w(\alpha) = \sum_{i=1}^{n} \alpha_i - 1/2 \sum_{i,k=1}^{n} \alpha_i y_i \alpha_k y_k (x_i, x_k), \\ &\text{Subject to} \ \sum_{i=1}^{n} \alpha_i y_i = 0, \quad \alpha_i \geq \forall i. \end{aligned} \tag{3}$$

According to α_i Lagrange multipliers computed in (3), following decision function is built

$$f(x) = sign\left(\sum_{i=1}^{sv} \alpha_i y_i (x, x_i) + b\right). \tag{4}$$

2.2.2. Nonlinear SVM classifier

In a nonlinear input space (including all training data), SVM fails to build optimal separating hyperplane. In this case, nonlinear input space is transformed into higher dimensional linear feature space via several kernel functions. A kernel function can be defined as following formula

$$K(x, x') = (\Phi(x) \cdot \Phi(x')) = \Phi(x)\Phi(x'). \tag{5}$$

Kernel functions must satisfy the Mercer's condition. Thus, aim of quadratic optimization problem and decision function of SVM in Section 2.2.1 are transformed into following formula

$$\begin{aligned} &\text{Maximize} \quad w(\alpha) = \sum_{i=1}^{n} \alpha_i - 1/2 \sum_{i,k=1}^{n} \alpha_i y_i \alpha_k y_k K(x_i, x_k), \\ &f(x) = sign\left(\sum_{i=1}^{sv} \alpha_i y_i K(x, x_i) + b\right). \end{aligned} \tag{6}$$

Commonly used kernel functions are as follows:

- Dot product kernels: $K(x, x') = x \cdot x'$.
- Polynomial kernels: $K(x, x') = (x \cdot x' + 1)^d$; where d is the degree of kernel and positive integer number.
- RBF kernels: $K(x, x') = \exp(-||x - x'||^2/\sigma^2)$; where σ is a positive real number.

2.3. Binary particle swarm optimization

Particle swarm optimization is an optimization technique based on swarm intelligence such as fish schooling and bird flocking. The aim of this technique is to find optimum solution in the solution set.

The PSO is firstly developed by Kennedy in 1995 (Kennedy & Eberhart, 1995). This technique is developed to optimize the problems that could be solved using real numbers. Later, it is successfully implemented in many research areas.

In PSO technique, particles are composed of cells called position. The swarm composed from these particles separates in the solution space randomly. Every particle in the swarm is a part of the solution set. Best values of each particle (local best value – $p_{best_{i,j}}$, global best value – $g_{best_{i,j}}$) in the swarm and the swarm itself are accumulated to be used in the next step and also to obtain optimum values. The velocity and the position of the particle are calculated as follows

$$v_{i,j}(t+1) = w v_{i,j}(t) + c_1 R_1 (p_{best_{i,j}} - x_{i,j}(t)) + c_2 R_2 (g_{best_{i,j}} - x_{i,j}(t)), \quad (7)$$

$$x_{i,j}(t+1) = x_{i,j}(t) + v_{i,j}(t+1), \quad (8)$$

where i is the index of particle, j is the index of position in particle, t shows the iteration number, $v_{i,j}(t)$ is the velocity of the ith particle in swarm on j th index of position in particle $v_{\min} \leqslant v_{i,j}(t) \leqslant v_{\max}$ and $x_{i,j}(t)$ is the position. R_1 and R_2 are the random numbers uniformly distributed between 0 and 1. c_1 and c_2 are the acceleration numbers and default values 2 and w is the inertia weight and is usually used less than 1. PSO is briefly illustrated in Fig. 2.

Binary particle swarm optimization is introduced in 1997 firstly by Kennedy and Eberhart. Like GA, BPSO could be effectively utilized in binary optimization problems (Kennedy & Eberhart, 1997). In the BPSO technique, the probability of the particle being as 0 or 1 is specified by the velocity value using sigmoid function (Kennedy & Eberhart, 1997). This determination of the position is performed using the following formula

$$X_{i,j}(t+1) = \left\{ \begin{array}{ll} 0 & \text{if } rand() >= S(v_{i,j}(t+1)) \\ 1 & \text{if } rand() < S(v_{i,j}(t+1)) \end{array} \right\} \quad (9)$$

where $rand()$ is the random numbers uniformly distributed between 0 and 1. $S(\cdot)$ is the sigmoid function and it is given as follows

$$S(v_{i,j}(t+1)) = \frac{1}{1 + e^{-v_{i,j}(t+1)}}. \quad (10)$$

2.4. Genetic algorithm

Based on long-term observation, Darwin asserted his theory of natural evolution. According to this theory, living beings compete with each other to survive. At the end of this competition, the successful beings transfer their genes to the beings in the next generation.

Inspiring by Darwin's evolution theory, Genetic algorithm was first introduced by Holland as a powerful computational model in 1975 (Holland, 1975). It is commonly used for optimization problems which could take discontiguous or continuous values. Prime aim of GA is to find optimum solution within the potential solution set. Each solution set is called as population. Populations are composed of vectors, namely, chromosome or individual. Each item in the vector is called as gene. The structure of the GA is given in Fig. 3 and described as follows.

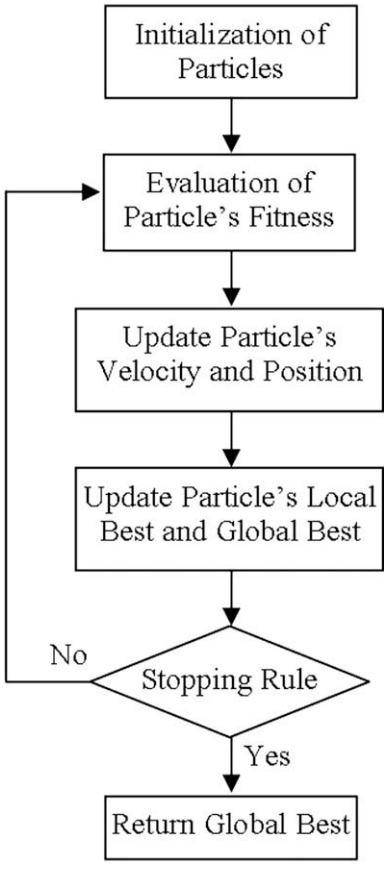

Fig. 2. The PSO algorithm.

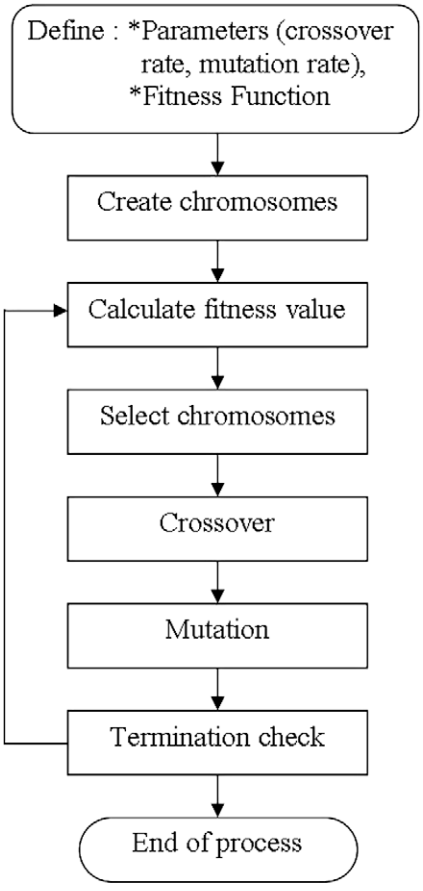

Fig. 3. The structure of the GA.

X_1	X_2	X_3	X_4	.....	X_n
V_1	V_2	V_3	V_4	.....	V_n

Fig. 4. A structure of the particle.

2.4.1. Initialization

At the beginning of the process, each chromosome is randomly composed within the values 0 and 1. These chromosomes are individual solutions in the potential solution set.

2.4.2. Evaluation

In the optimization process with GA, the proper degree of each chromosome in the potential solution set is calculated using fitness function. One of the most important factors that effect the achievement of GA is the selection of the fitness function. Thus, fitness function must be adopted effectively focused to the solution.

2.4.3. Selection

This is the process of the selection of the chromosomes to transfer their genes to the next generation considering fitness values. Selection process is implemented using roulette wheel selection technique. The probability of the selection of each chromosome using roulette wheel selection technique is calculated as follows

$$P(c_i) = \frac{\bar{f}_i}{\sum_{j=1}^{N} f_i i}, \tag{11}$$

where c_i is the chromosome in question, f_i is the fitness value of the chromosome, N is the number of the chromosomes in the population.

2.4.4. Crossover

The chromosome pairs are crossed over to generate the chromosomes in the next generation using a predetermined crossover rate. Crossover is the process of replacement of one or more segments which are selected randomly.

2.4.5. Mutation

Mutation process is utilized to enhance the variation of the population. Value of the each gene in the chromosomes is changed considering mutation rate.

2.4.6. Termination

The selection, crossover and mutation processes are repeated to the end of the iteration. The algorithm is also terminated in the situation of obtaining desired fitness value.

3. Proposed feature selection methods

3.1. Feature selection technique using BPSO (BPSO-FST) architecture

In the BPSO–FST, each particle is composed of 23 binary cells, which refer to the whole features in the dataset. The value of these cells shows whether the feature that the cell refers to would be selected. A cell value of 1 shows the feature that the cell refers to is selected, a cell value of 0 shows the feature that the cell refers to is not selected into the dataset. The structure of the particle is given below, where n is the number of the features, x is the value of the cell and $x \in \{0, 1\}$, and v denotes the velocity of the cell.

BPSO is iterated 200 times to find the optimum solution set for each value of the parameters c and γ that effect SVM's performance.

Population includes 10 particles, and the values of these particles are taken as 0 at the beginning of the process. In the optimization process, training and test sets are composed considering the features defined by the particles, and SVM is trained and tested using mentioned datasets. As a result, classification accuracy rate, training error rate and the sum of the times elapsed for training and test processes are obtained for each particle. The success rate of each particle is calculated using the following fitness function formula

$$f(i) = A(i) - E(i), \tag{12}$$

where $f(i)$ is the success rate, $A(i)$ is the classification accuracy rate and $E(i)$ is the training error rate of ith particle. Velocity of the particles is calculated using (7) and $v_{\min}$ and $v_{\max}$ are used as -6 and 6, respectively. The value of the cells within each particle is updated using (8). For each iteration, $p_{best_{ij}}$ and $g_{best_{ij}}$ is updated if necessary. At the end of the optimization process, $g_{best_{ij}}$ is found as the optimum solution.

3.2. Feature selection technique using GA (GA–FST) architecture

In the GA–FST, each chromosome is composed of 23 genes, which refers to the whole features in the dataset similarly particles used in BPSO–FST. The value of these genes shows whether the feature that the gene refers to would be selected. A gene value of 1 shows the feature that the gene refers to is selected, a gene value of 0 shows the feature that the gene refers to is not selected into the dataset.

The same SVM kernel parameters ranges are used in the GA–FST to compare with BPSO–FST. Also, GA–FST is implemented for 200 iterations.

The size of the population, the mutation rate and the crossover rate is used as 10, 0.05 and 0.25, respectively, in this study. At the beginning of the process, the values of each chromosome are used as 0. In the optimization process, training and test sets are composed considering the features defined by the chromosomes, and

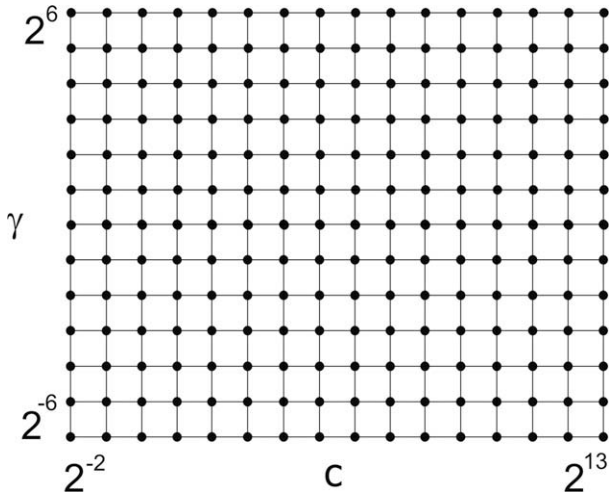

Fig. 5. SVM parameter search grid.

SVM is trained and tested using mentioned datasets. As a result, classification accuracy rate, training error rate and the sum of the times elapsed for training and test processes are obtained for each chromosome. The success rate of each chromosome is calculated using (12). Chromosomes are selected using roulette wheel composed according to the success rates mentioned. Chromosomes are crossed over and mutated according to the given crossover and mutation rates. At the end of the process, the chromosome that has the optimum fitness value is found as the optimum solution.

Table 1
Test results.

Method	c	γ	ERR (%)	ACC (%)	TIME (s)	NoF
BPSO–FST	4	8	0.26	81.46	0.52	11
GA–FST	512	4	0.00	79.17	0.53	12
SVM	8	1	0.31	76.67	0.67	23

Method, Method used in classification; c, RBF kernels parameter; γ, RBF kernels parameter; ERR, Training error; ACC, Classification accuracy; TIME, Sum of the time elapsed in training and test processes of SVM classification; NoF, Number of features selected by the feature selection method; BPSO–FST, SVM classification utilizing BPSO–FST; GA–FST, SVM classification utilizing GA–FST; SVM, Simple SVM classification.

4. Results and discussion

Fig. 4 Datasets are implemented by normalizing into the range [−1, 1]. Radial basis function (RBF) kernel, which is commonly used for SVM, is employed for the default kernel. The values of the RBF kernel parameters c and γ are found using grid search algorithm (Cormen, Leiserson, Rivest, & Stein, 2001) and both of the values of these parameters are used as 2^n. The search grid used in this study is diagrammatized in Fig. 5. In this technique, n is used in the range [−2, 13] and [−6, 6] for c and γ, respectively.

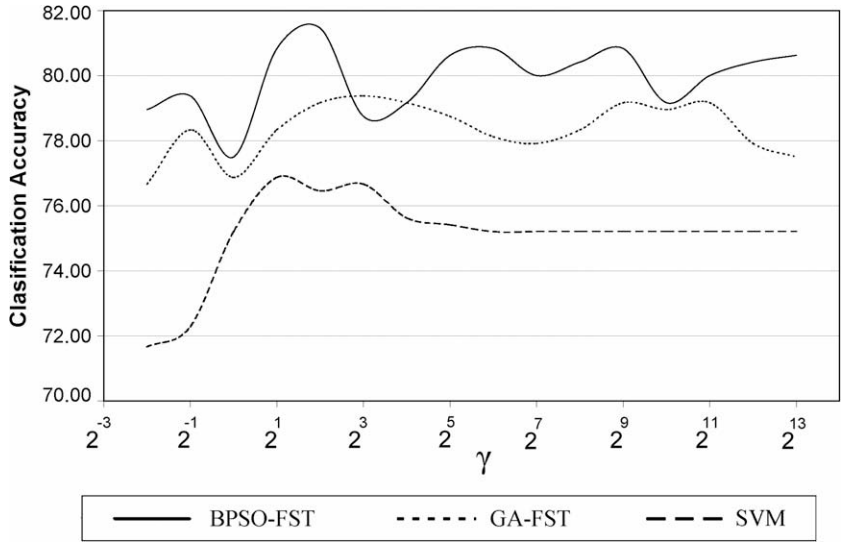

Fig. 6. Classification accuracy graph depend on γ. The values of the fixed c parameters are 4, 512 and 8 for BPSO–FST, GA–FST and simple SVM classification model, respectively. BPSO–FST, SVM classification utilizing BPSO–FST; GA–FST, SVM classification utilizing GA–FST; SVM, Simple SVM classification.

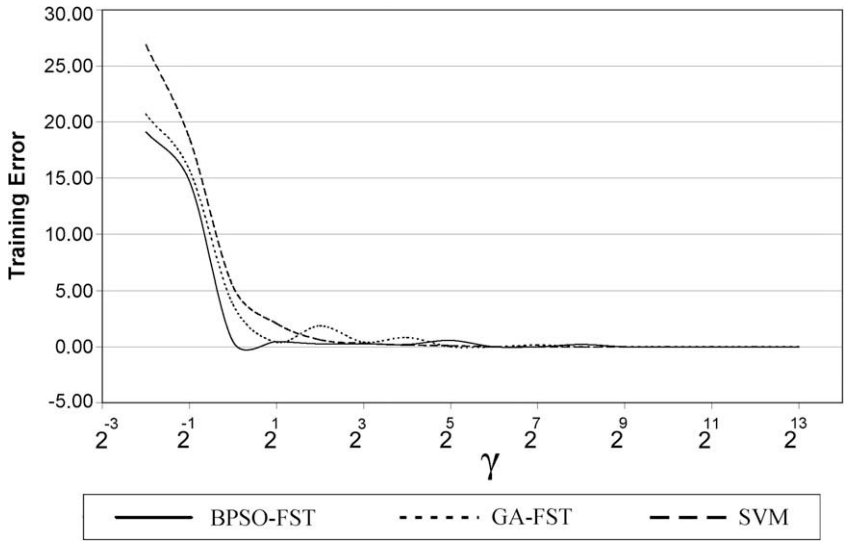

Fig. 7. Training error graph depend on γ. The values of the fixed c parameters are 4, 512 and 8 for BPSO–FST, GA–FST and simple SVM classification model, respectively. BPSO–FST, SVM classification utilizing BPSO–FST; GA–FST, SVM classification utilizing GA–FST; SVM, Simple SVM classification.

The results of the BPSO-FST, the results of the GA–FST and the results of the simple SVM classification are compared using the optimum c and γ parameters obtained for each technique.

Classification accuracy, training error and the sum of the training and test time is considered finding the best classification architecture. k-Fold cross-validation method is employed using $k = 5$ to improve the reliability of the results (Shao, 1993; An, Liu, & Venkatesh, 2007).

The implementations are employed using LIBSVM package (Chang & Lin, 2001) in the Matlab 7.0 application platform in a computer with Pentium IV with a 3.2 GHz CPU and 1 GByte memory.

The optimum parameters c and γ are found as 4 and 8, respectively, in the BPSO–FST. The optimum particle has 11 features selected, and the sum of the training and test times is 0.52 s.

The optimum parameters c and γ are found as 512 and 4, respectively, in the GA–FST. The optimum particle has 12 features selected, and the sum of the training and test times is 0.53 s.

The optimum parameters c and γ are found as 8 and 1, respectively, in the simple SVM classification technique. Whole 23 features are used in this technique, and the sum of the training and test times is 0.67 s.

The values of the optimum parameters and the result of the SVM classification process for all techniques are given in Table 1.

In each technique, the classification accuracy and training error graphs depend on the γ parameter of SVM for the results based on fixed c parameters in the optimum solutions are given in Figs. 6 and 7. The classification accuracy and training error graphs depend on the c parameter of SVM for the results based on fixed γ parameters in the optimum solutions are given in Figs. 8 and 9.

As seen in Figs. 6–9, BPSO–FST has highest classification accuracy rate considered to GA–FST and simple SVM classification technique for each c and γ parameters within the search grid. Sum of the training and test times depends on number of the features used in SVM classification process. Hence, BPSO–FST has the minimum

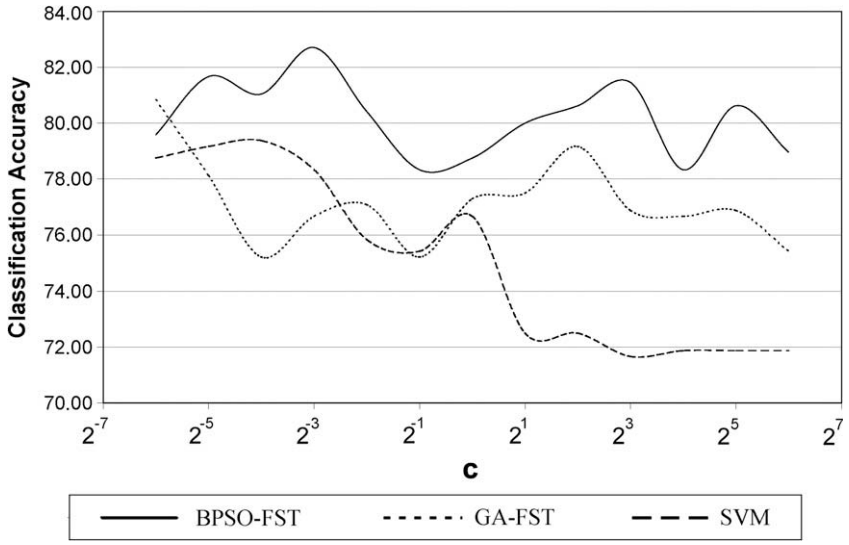

Fig. 8. Classification accuracy graph depend on c. The values of the fixed γ parameters are 8, 4 and 1 for BPSO–FST, GA–FST and simple SVM classification model, respectively. BPSO–FST, SVM classification utilizing BPSO–FST; GA–FST, SVM classification utilizing GA–FST; SVM, Simple SVM classification.

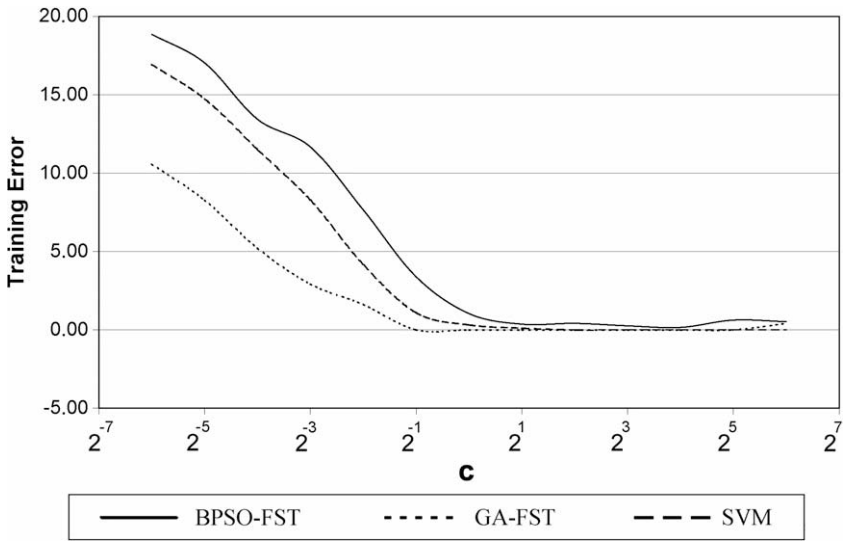

Fig. 9. Training error graph depend on c. The values of the fixed γ parameters are 8, 4 and 1 for BPSO–FST, GA–FST and simple SVM classification model, respectively. BPSO–FST, SVM classification utilizing BPSO–FST; GA–FST, SVM classification utilizing GA–FST; SVM, Simple SVM classification.

feature size, so that the sum of the training and test times is also minimum compared to the other classification techniques in addition to providing the best classification accuracy.

5. Conclusion

Instead of using the whole features in the dataset, SVM classification process is implemented employing the reduced dataset on the determination of CAD using EST data. The dataset's dimension is reduced by utilizing BPSO–FST or GA–FST. The classification process implemented by feature selection techniques achieves more successful classification accuracy. Besides, they decrease the complexity of the system by reducing the dimensions of the dataset. Classification processes implemented by using mentioned techniques are compared to each other. The classification process implemented by utilizing BPSO–FST has the best classification accuracy and minimal process time compared to others.

References

An, S., Liu, W., & Venkatesh, S. (2007). Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression. *Pattern Recognition, 40*, 2154–2162.

Chang, C. C., & Lin, C. J. (2001). LIBSVM: A library for support vector machines, available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

Chen, K. Y., & Wang, C. H. (2007). A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. *Expert Systems with Applications, 32*(1), 254–264.

Chuang, L. Y., Chang, H. W., Tu, C. J., & Yang, C. H. (2008). Improved binary PSO for feature selection using gene expression data. *Computational Biology and Chemistry, 32*, 29–38.

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

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2001). *Introduction to algorithms* (2nd ed.). MIT Press.

Holland, J. (1975). *Adaptation in natural and artificial systems*. The Michigan University Press.

Kennedy, J., & Eberhart, R. C. (1995). Particle swarm optimization. In *Proceedings of IEEE international conference on neural networks* (vol. 4; pp. 1942–1948).

Kennedy, J., & Eberhart R. C. (1997). A discrete binary version of the particle swarm algorithm. In *Proceedings of 1997 conference systems man and cybernetics* (pp. 4104–4108).

Kulkarni, A., Jayaraman, V. K., & Kulkarni, B. D. (2004). Support vector classification with parameter tuning assisted by agent-based technique. *Computers and Chemical Engineering, 28*, 311–318.

Kurt, I., Ture, M., & Kurum, A. T. (2008). Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. *Expert Systems with Applications, 34*, 366–374.

San Roman, J. A., Vilacosta, I., Castillo, J. A., Rollan, M. J., Hernandez, M., Peral, V., et al. (1998). Selection of the optimal stress test for the diagnosis of coronary artery disease. *Heart, 80*(4), 370–376.

Sepehri, A. A., Hancq, J., Dutoit, T., Gharehbaghi, A., Kocharian, A., & Kiani, A. (2008). Computerized screening of children congenital heart diseases. *Computer methods and programs in biomedicine, 92*(2), 186–192.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of American Statistical Association, 88*, 486–494.

Sprigings, D. (2002). Exercise stress testing. *Medicine, 30*(3), 31–33.

Takeuchi, K., & Collier, N. (2003). Bio-medical entity extraction using support vector machines. *Artificial Intelligence in Medicine, 33*(2), 125–137.

Tantimongcolwata, T., Naennab, T., Isarankura-Na-Ayudhyaa, C., Embrechtsc, M. J., & Prachayasittikula, V. (2008). Identification of ischemic heart disease via machine learning analysis on magnetocardiograms. *Computers in Biology and Medicine, 38*, 817–825.

Thom, T., Haase, N., Rosamond, W., Howard, V. J., Rumsfeld, J., Manolio, T., et al. (2006). Heart disease and stroke statistics – 2006 update: A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation, 113*, e85–e151.

Uğuz, H., Öztürk, A., Saraçoğlu, R., & Arslan, A. (2008). A biomedical system based on fuzzy discrete hidden Markov model for the diagnosis of the brain diseases. *Expert Systems with Applications, 35*, 1104–1114.

Vapnik, V. (1995). *The nature of statistical learning theory*. New York: Springer.

Yang, H. C., Zhang, S. B., Deng, K. Z., & Du, P. J. (2007). Research into a feature selection method for hyperspectral imagery using PSO and SVM. *Journal of China University of Mining and Technology, 17*(4), 473–478.

Zhidong, Z. (2005). Noninvasive diagnosis of coronary artery disease based on instantaneous frequency of diastolic murmurs and SVM. In *Proceedings of the 2005 IEEE engineering in medicine and biology 27th annual conference*, Shanghai, China (pp. 5651–5654).

Zhou, W., Zhou, C., Liu, G., & Zhu, H. (2006). *Feature selection for microarray data analysis using mutual information and rough set theory*. Boston: Springer.