INVESTIGATING THE PERFORMANCE OF OPTIMIZATION TECHNIQUES WITH SVM FOR MEDICAL IMAGE CLASSIFICATION

Dr. N.T. Renukadevi[1] , Ms. S.Nandhinidevi[2]

[1, 2]Assistant Professors, Department of CT-UG, Kongu Engineering College, Perundurai (India)

ABSTRACT

Computerized tomography uses X-rays, and a computer to create detailed body image. CT scan images are called tomograms having more detail than standard X-rays. Content-based image retrieval (CBIR) of medical images, according to its domain specific image features, is a valuable tool for physicians. A method for automatic classification of computed tomography (CT) images of different types is presented in this paper. The proposed method has three major steps: 1. Feature are extracted from the CT images using Coif let wavelets; 2. The features extracted are classified using Support Vector Machine; 3. The parameters of the SVM are optimized using Particle Swarm Optimization (PSO), and modified PSO with a genetic algorithm

Keywords- Content-Based Image Retrieval, Computed Tomography, Coif Let Wavelets, Particle Swarm Optimization, Genetic Algorithm, Support Vector Machine

I INTRODUCTION

Computerized tomography uses X-rays, and a computer to create detailed body image. CT scan images are called tomograms having more detail than standard X-rays. A CT scan produces images of structures within the body like internal organs, blood vessels, bones and tumours [1]. The various types of CT scans that help investigate particular body areas include: Head scans, which can check for suspected brain tumours and arterial bleeding/swelling; head scans also investigate the brain after a stroke (when blood supply to a part of the brain is cut). Abdominal scans can detect tumours and diagnose conditions causing internal organs like liver, kidneys, pancreas, intestines or lungs, to become enlarged or inflamed. Vascular scans assess conditions affecting blood flow to various parts of the body. Bone scans assess bone injuries and disease, specially the spine.

Support Vector Machines (SVMs) achieve good empirical on different learning problems when compared to the other machine learning methods [2]. Though, the accomplishments of SVMs are governed by the adequate choice of parameters of the kernel and the regularization parameters. The parameter selection is treated as an optimization problem wherein a search technique is used to the optimal parameters to maximize the SVM performance [3][4][5]. Even though search techniques represent a systematic approach for parameter selection in SVM, it can also be expensive if the number of parameters to be evaluated during the search process is large [6].

Classification includes a broad range of decision-theoretic approaches that is used for identification of images. All classification algorithms are mainly based on the assumption that the image in question portrays one or more features and each of these features belongs to one of several distinct and exclusive classes. The class specifies a prior by an automatically clustered into sets of prototype classes, where the analyst merely specifies the number of desired categories.

Image classification analyzes the numerical properties of various image features and organizes the data into categories. Classification algorithms mainly employs into two phases of processing such as training and testing. In initial training phase, the characteristic properties of typical image features are isolated based on the training class, is created. In the subsequent testing phase, these feature-space partitions are used for classification of image features. The description of training classes is an extremely important component of the classification process.

- Supervised classification: here the statistical processes or distribution-free processes can be used for extraction of class descriptors.
- Unsupervised classification: it relies on clustering algorithms for automatically segment the training data into prototype classes.

A support vector machine classification can be constructed as a hyper plane or set of hyper planes in a high or infinite-dimensional space, which are used for classification. A separation is achieved by the hyper plane that consists of the largest distance to the nearest training data points of any class so-called as functional margin, since in general the larger the margin the lower the generalization error of the classifier. SVM is a binary classifier and it is immune to noise.

A different method to SVM parameter selection is based on Meta-Learning wherein the SVM parameter selection is treated as supervised learning tasks [7]. In this, the characteristics of the training examples and the performance achieved for a set of parameters for the problem are stored. Meta-learners on the basis of the set of training examples received as input, predicts the best values for the parameters for a new problem. The Meta-Learning is also used as it is less expensive when compared to the search approach.

The SVM parameters are often selected by calculating different combinations of parameters and utilizing the combination which achieves the best performance for the particular dataset. To automatize the search process, various search and optimization techniques are used [8][9]. The search space is made up of a set of possible combination of parameters and a fitness function corresponding to performance measure achieved by the SVM is considered. Different search techniques available in the literature are based on Evolutionary Algorithms [10], gradient based techniques [11] Tabu search [12] and so on.

II OVERVIEW OF GA

Population Initialization

The parameters in problem space are not calculated directly, but instead the feasibility solutions of the problem to be solved that are expressed as Chromosomes or individuals in genetic space by encoding. The population of kernel function parameter can be initialized.

Fitness Function

Fitness functions are used to distinguish the quality of individuals in population and it is the only guide for natural selection, which is usually derived from objective function. If the fitness function value is larger then there will be better in quality of individual [14][15]

Selection

By selection relay, tremendous individuals is found from old population, and new population can be build for reproduction of the next generation individual. If the fitness function value of individual is larger than the probability of being selected is higher. For instance, an selection method to fitness Proportion, is performed and the probability of being selected to the individual i is

$$p_i = \frac{F_i}{\sum_{j=1}^{n} F_j} \quad (1)$$

In above formula, Fi is the fitness function values of the individual i, n is population size.

There are different types of selection process of which Roulette wheel selection is considered.

Roulette–Wheel selection : In this, the strings with higher fitness values has higher probability of being selected in the new mating pool. Fig shows that, the top surface area of Roulette wheel is divided as N parts according to their fitness values where higher fitness value string consumes larger area. Each part of the wheel represents a string like f1, f2, f3,....fn. A fixed pointer is used here for identifying the selected string or winner of the rotation and the wheel is rotated either in clockwise or anti-clockwise. After N rotation, N number of string can be selected to form a new population of size N. The probability of a string to be selected after a rotation is proportional to its area or its fitness value.

Crossover

It is used to find two individuals randomly from the present population, whose chromosome information's are exchanged and combined for each other to pass outstanding characteristics of father string down to son string, in order to reproduce the new admirable individuals. The original crossing method is adopted here because all individuals are encoded with real.

$$a_{kj} = a_{ij}(1-d) + a_{li}d$$
$$a_{lj} = a_{lj}(1-d) + a_{ki}d \quad (2)$$

In above formula, d is a random number on interval (0, 1) [18].

Different types of crossovers are

- Single-point crossover: Here crossover operation is done at a single point i.e. a single point which is selected between 1 and (L-1) at random where L is the length of the string.
- Two-point crossover: The only difference is that, it has two crossover sites rather than one as in single point cross over.
- Multi point crossover: In multiple crossover points are selected along the string length randomly with no duplicates and are sorted into ascending order. The bits between successive crossover points are exchanged alternatively between two parents in order to produce new offspring.
- Uniform crossover: It is more general version of multi point crossover. At each bit position there will be a tossing of coin which decides whether crossover will occur at that bit position or not. The mask consists of same length as the chromosome structure is created at random and the parity of the bits in the mask indicates which parent will supply the offspring with which bits. The “1” in the random mask means bits swapping and the “0” means bit replacing

Mutation

It helps to reproduce a better individual, which are randomly found from the present population and mutated slightly. Mutation aims at maintaining the diversity of population. The operation of mutation to the jth gene of the ith individual as follows

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{\max}) * f(g), & r \geq 0.5 \\ a_{ij} + (a_{\min} - a_{ij}) * f(g), & r < 0.5 \end{cases} \quad (3)$$

In above formula, max a and min a are upper and lower bounds of gene aij. r is a random number on interval (0 , 1), g is the present number of iterations, maxG is the maximum evolutionary generation

In the current work, PSO is implemented for parameter selection of the RBF kernel of SVM. “Swarm intelligence” is usually used for optimization, to maximize or minimize the cost function by searching for a set of variable is termed optimization. Swarm optimizations are based on the collective behavior of the bees or ants, or social behavior of bird flocking and fish schooling. The Particle Swarm Optimization algorithm is a population-based stochastic search algorithm and is efficient in solving complex non-linear optimization problem. The PSO is popular as it is easily implemented, computationally inexpensive. To prevent premature convergence in PSO, the PSO is modified using genetic algorithm (GA).

III RELATED WORKS

Kharrat et al (2011) proposed a new approach for automatic classification of Magnetic Resonance (MR) human brain images as normal and abnormal. Wavelets Transform (WT) is used as input module to Genetic Algorithm (GA) for feature selection and Support Vector Machine (SVM) for classifying the MR images. The GA requires very less computation when compared with Sequential Floating Backward Selection (SFBS) and Sequential Floating Forward Selection (SFFS) methods. A reduction rate of 88.63% is realized and classification rate of 100% was obtained using the support vector machine.

Zhang et al (2010) presented adaptive chaotic particle swarm optimization (ACPSO) for optimizing parameters. The methodology is used to classify MR brain image as normal or abnormal. Wavelet transforms are used to extract features and principle component analysis (PCA) is applied to reduce the dimensions of features. Feed forward neural network is used to classify the features. To enhance the generalization, K-fold stratified cross validation was applied. The proposed method was evaluated using 160 images (20 normal, 140 abnormal), and classification accuracy of 98.75% was achieved.

Gal et al (2012) presented a multi-disciplinary approach to address the classification problem. The proposed methodology combined image features, meta-data, textual and referential information for classification of medical images. Image CLEF 2011 medical modality classification data set was used to evaluate the system's accuracy. Multiple kernels were used for classification; significantly better classification accuracy was achieved as the kernels were selected for different features. Best classification accuracy of 88.47% obtained and outperforms the other methods available in the literature.

Uma maheswari et al (2012) proposed PSO SVM for classification of DICOM images. The proposed method was used to recognise and classify brain images as normal and abnormal. Optimal recognition and detection of disease in DICOM images is crucial for the diagnosis process. The proposed method focused on recognition and classification based on combined approach of digital image processing incorporating PSO, GA and SVM. The combined approach by using PSO-SVM achieves high approximation capability and faster convergence.

IV MATERIALS AND METHODS

Feature extraction using Coif let wavelet

Coif lets are discrete wavelets designed to have scaling functions with vanishing moments [23]. The wavelet is near symmetric with N/3 vanishing moments and N/3-1 scaling functions. The function Ψ has 2N moments equal to 0 and the function φ has 2N-1 moments equal to 0. The support length of the two functions is 6N-1. The coifN Ψ and φ are considerably more symmetric than the dbNs. The coifN are compared to *db3N* or *sym3N* when considering the support length. When number of vanishing moments of Ψ is considered, coifN is compared to *db2N* or *sym2N*.[19]

If s is a sufficiently regular continuous time signal, for large j the coefficient [23] $\langle s, \phi_{j,k} \rangle$ is approximated by $2^{-j/2} s\left(2^{-j} k\right)$.

If s is a polynomial of degree d, d ≤ N - 1, then the approximation becomes equality.

Support Vector Machine (SVM)

Given a set of features that can be represented in space, SVM maps features non-linearly into n dimensional feature space when provided with features set that can be represented in space. When a kernel is introduced with high computation the algorithm uses inputs as scalar products with classification being solved by translating the issue into a convex quadratic optimization problem with a clear solution being obtained by convexity [7]

In SVM, an attribute is a predictor variable and a feature a transformed attribute. A set of features describing an example is a vector. Features define the hyperplane. SVM aims to locate an optimal hyper plane separating vector clusters with a class of attributes on one side of the plane with the on the other side. The margin is the distance between hyper plane and support vectors. SVM analysis orients the margin that space between it and support vectors is maximized. Figure 4.1 shows a simplified SVM process overview.

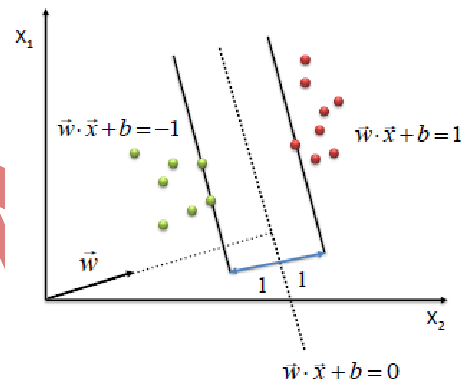

Fig 4.1: Support vector machine

Given a training set of $(x_i, y_i), i1,2,....l$ where $x_i \in R^n$ and $y \in \{1,-1\}^l$, SVM has to solve the optimization problem of:

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C\sum_{i=1}^{l} \xi_i \quad (4)$$

$$\text{Subject to } y_i\left(w^T\phi(x_i)+b\right) \geq 1-\xi_i \text{ and } \xi_i \geq 0. \quad (5)$$

The function ϕ maps the vectors x_i in higher dimensional space. C>0 is penalty parameter of the error term.

This optimization model is solved through the use of the Lagrangian method, equal to the method for solving optimization problems in a separable case. One maximizes the dual variables Lagrangian:

$$\underset{\alpha}{Max}\ \ L_D(\alpha)=\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{m}\alpha_i\alpha_j y_i y_j\left\langle x_i.x_j\right\rangle \quad (6)$$

Subject to: $0\le\alpha_i\le C\ \ i=1,..,m$ and $\sum_{i=1}^{m}\alpha_i y_i=0$

To find the optimal hyper plane, a dual LagrangianLD(α) should be maximized as regards non-negative α_i under the constrains $\sum_{i=1}^{m}\alpha_i y_i=0$ and $0\le\alpha_i\le C$. The penalty parameter C, now the upper bound on α_i, is user determined.

A kernel function is defined as $K(x_i,x_j)=\phi(x_i)^T\phi(x_j)$. The Radial Basis function is given as follows:

$$K(x_i,x_j)=\exp\left(-\gamma\left\|x_i-x_j\right\|^2\right)\ ,\gamma>0 \quad (7)$$

A proper parameter setting improves SVM classification accuracy. There are two parameters to be determined in the SVM model with the RBF kernel: C and gamma (γ). The value of C and γ influence the learning performance of the SVM. Intuitively the γ parameter defines the distance a single training example can reach, with low values meaning 'far' and high values meaning 'close'. The C parameter trades off training examples misclassification against decision surface simplicity. A low C ensures a smooth decision surface while a high C attempts to classify training examples correctly. Experiments are undertaken to evaluate SVM performance through variations of the γ and C parameters [20][21].

4.1. GA-SVM Algorithms

There are two methods used to combine Genetic algorithm and SVM are as follows:

- Dealing with dataset: The initial training dataset are optimized by including the important samples in order to find a sample subset that preserves or improves the discrimination ability of SVM. Training on the subset is equal to the initial sample sets and the training time is greatly reduced. Main idea is to find out the important examples that affect the classification results significantly. If the feature vectors are removed, then the separating boundary can be changed.

- Defining parameters: The value of parameters is important in case of Support Vector Machines for better performance of algorithm. A methodology is presented to train SVM where the regularization parameter (C) was determined automatically by the use of an efficient Genetic Algorithm in order to solve multiple category classification problems [13].

GA-SVM algorithm is proposed using the character of GA in which searching the optimal combination is made to combine SVM classification algorithm to which it gets the tool of optimal columns combination. Fig. shows the process details of the GA-SVM algorithm. After users input for the dataset and column features, GA-SVM algorithm first acquire the feature of column dataset in which values are initialized with chromosome, and encodes the chromosome to create different combination columns of sub-dataset. The fitness function of chromosome is evaluated to create fitness value of each chromosome based on condition of good or bad. Before the condition of satisfaction closed down, it continually passed through the evolution procedure such as selection, crossover and mutation in order to find the highest categories of predicted accuracy of the combination columns for the selecting key combination columns. The optimal combination searching of GA-SVM algorithm effectively reduces the dataset columns. The following figure 4.2 shows the process of GA-SVM algorithm. Also it is used to find the important columns that affects classification predicted accuracy and survive as the main column of dataset in classification predicted analysis to reduce the complexity of dataset.

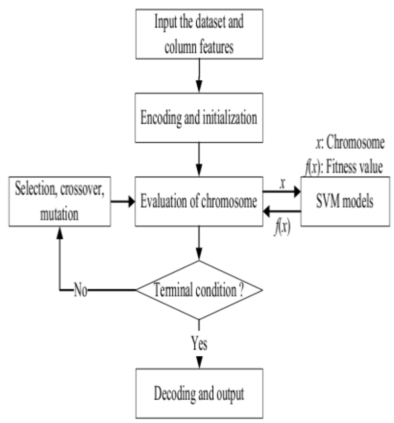

Fig 4.2 Process Flow Diagram of GA-SVM algorithm

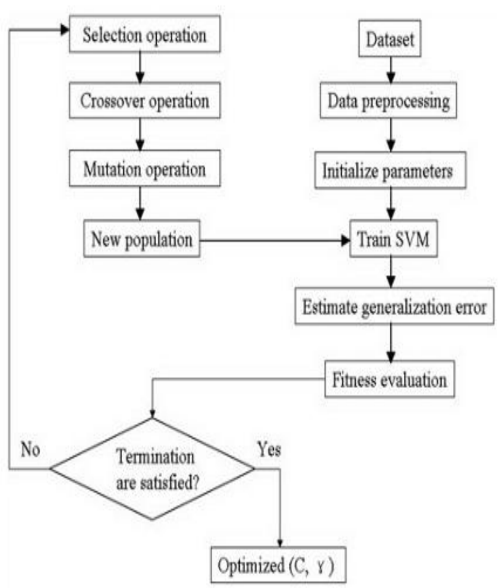

Fig 4.3 Flow Chart of GA-Based Parameter Optimization

- Steps for New Generation Formation: A new generation can be formed by (a) selection of some individuals from current population based on their fitness values, (b) by exchanging part of the genes for the chromosomes in current generation with a certain probability using a crossover operator, (c) by modifying one or more genes in some chromosomes with a certain probability using a mutation operator, (d) selection of new off springs to substitute old individuals according to their fitness values and (e) rejecting others to keep the population size constant. The flow chart of the GA-based parameter optimization algorithm of the SVM is given away in following figure 4.3.

4.2 PSO-SVM Algorithms

To optimize the parameters C and γ, PSO is adapted to execute the search for optimal combination (C, γ). The objective function for evaluating the quality of combination of parameters is based on Root Mean Squared Error (RMSE) achieved by the SVM in a 10-fold cross validation experiment. Thus, the PSO finds the combination of parameters with the lowest RMSE[16].

Each particle *i* represents a combination of parameters which also indicates the position of the particle in the search space. The velocity of the particle is indicative of the direction of the search of the particle. The PSO algorithm keeps updating the position and velocity of the particle in each iteration which leads to best regions in the search space. The velocity and the position of the particle are updated as follows:

$$\begin{aligned} v_i^d &= wv_i^d + c_1 r_1\left(p_i^d - x_i^d\right) + c_2 r_2\left(p_g^d - x_i^d\right) \\ x_i^d &= x_i^d + v_i^d \end{aligned} \qquad (8)$$

where w is the Inertia weight; d represents the number of dimensions; i is the size of the population; the two "best" values - pbest and gbest - of a particle where 'pbest' $\left(p_i^d\right)$ is the best solution achieved by the particle till then and 'gbest' $\left(p_g^d\right)$ is the best value obtained till then by any particle in the population; c_1, c_2 are positive constants; r1 and r2 are random values with value between (0, 1).

The parameters w, c_1, c_2, r_1, r_2 in the PSO affect the performance of the algorithm significantly. The inertia weight controls the exploration and exploitation; generally 0 < w < 1 for the particles to converge. Higher value of w (near 1) favours global search and lower values less than 0.5 favours local search. The random numbers r1 and r2 are with value between (0, 1). The coefficients c_1 and c_2 are usually equal (i.e., c_1=c_2) and has a value in the range of (0, 4). The value of c_1 and c_2 are significant as convergence is dependent on these values.

Convergence is feasible when $1 > w > \frac{1}{2}(\phi_1 + \phi_2) - 1$ where $\phi_1 = c_1 r_1,\ \phi_2 = c_2 r_2$ Also for stochastic ϕ_1 and ϕ_2, convergences results when $\phi_{ratio} = \frac{\phi_{crit}}{c_1 + c_2}$ is close to 1 and $\phi_{crit} = \sup \phi\left[0.5\phi - 1 < w,\ \phi \in \left(0, c_1 + c_2\right)\right]$

To avoid premature convergence and to combine the coordinates to achieve high convergence speed, the classical PSO is modified using GA. The GA incorporated, coordinates the relationship of the PSO parameters to maximize the performance. The GA generates a population by encoding the PSO parameters. The fitness value is calculated based on ϕ_{ratio} with the objective to maximize it. Genetic operators, selection, crossover and mutation, are applied to generate the next generation. On termination of the GA algorithm, the PSO parameters obtained are updated into PSO algorithm.

V RESULTS AND DISCUSSION

Experiments were conducted using 150 CT scans images of brain, chest and colon. Features were extracted using Coif let wavelet. Experiments were conducted to evaluate the classification accuracy for SVM-RBF, with PSO and with modified PSO. All the experiments were conducted for 10-fold cross validation. The classification accuracy and the root mean square error (RMSE) achieved is tabulated in Table 5.1. Shows the classification accuracy

Table 5.1 Classification Accuracy

Methods	Classification accuracy	
	PCA	Kernel PCA
SVM - GA	87.88	89.42
SVM – PSO`	87.93	91.15
SVM – GA-PSO	91.52	92.67

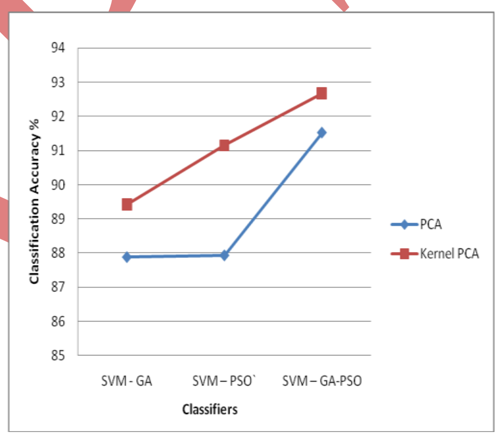

Figure 5.1 Classification Accuracy

From figure 5.1 and table 5.1 it is observed that the Classification Accuracy of SVM C and Gamma GA PSO with local search increases by 4.54 % than SVM C and Gamma PSO and increases by 4.6% than SVM C and Gamma GA with Kernel PCA.

VI CONCLUSIONS

In this paper, to improve the performance of the SVM-RBF for classifying the CT images, the SVM parameters C and Gamma (γ) are optimized. Particle Swarm Optimization (PSO) is implemented to select the values of two SVM parameters for classification problems. To avoid premature convergence and to combine the coordinates to achieve high convergence speed, the classical PSO is modified using GA. The experiments were conducted for 10-fold cross validation. The classification accuracy and the root mean square error (RMSE) achieved for the proposed modified PSO is significantly better.

REFERENCES

1. Kalender, W. A. Computed tomography. (Wiley-VCH 2011).
2. C. Soares, P. Brazdil, and P. Kuba, "A meta-learning approach to select the kernel width in support vector regression," Machine Learning, vol. 54, no. 3, pp. 195–209, 2004.
3. Lorena and A. de Carvalho, "Evolutionary tuning of SVM parameter values in multiclass problems," Neuro computing, vol. 71, pp. 16–18, 2008.
4 S. Lessmann, R. Stahlbock, and S. Crone, "Genetic algorithms for support vector machine model selection," International Joint Conference on Neural Networks, pp. 3063–3069, 2006.
5 X. Guo, J. Yang, C. Wu, C. Wang, and Y. Liang, "A novel LS-SVMS hyper-parameter selection based on particle swarm optimization," Neurocomputing, vol. 71, pp. 3211–3215, 2008.
6 S. Ali and K. A. Smith-Miles, "A meta-learning approach to automatic kernel selection for support vector machines," Neuro computing, vol. 70, no. 1-3, pp. 173–186, 2006.
7 F. Friedrichs and C. Igel, "Evolutionary tuning of multiple svm parameters," Neurocomputing, vol. 64, pp. 107–117,2005.
8 B. de Souza, A. de Carvalho, and R. Ishii, "Multiclass SVM model selection using particle swarm optimization," in Sixth International Conference on Hybrid Intelligent Systems, pp. 441–446, 2006.
9 S. Lessmann, R. Stahlbock, and S. Crone, "Genetic algorithms for support vector machine model selection," in International Joint Conference on Neural Networks, pp. 3063–3069, 2006.
10 O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee , "Choosing multiple parameters for support vector machines," Machine Learning, vol. 46, no. 1, pp. 131–159, 2002.
11 Xiaoyong LIU "A Hybrid Algorithm for Text Classification Problem", PRZEGL ĄD ELEKTROTECHNICZNY (Electrical Review), ISSN 0033-2097, 2012.
12 Fengchun Tian "Classification of Electronic Nose Data on Wound Infection Detection Using Support Vector Machine Combined GA", Journal of Computational Information Systems 8: 8, 2012.
13. Fengmei Gao, "Application of SVM Optimization Based on GA in Electronic Sphygmomanometer Data Fusion", IJCSI International Journal of Computer Science Issues, Vol. 10(1) ,2013.
14. D.E. Goldberg, K. Deb, " A comparison of selection schemes used in genetic algorithms," In G. Rawlins, (Ed.), Proceedings of Foundation of Genetic Algorithms-, San Mateo: Morgan Kaufmann,pp.69-93, 1991.

15. D.K. Pratihar, "Soft Computing ", IIT Kharagpur, department of Mechanical Engg, ISBN 978-81-7319-866-3
16. Satyobroto Talukder, "Mathematical Modelling and Applications of Particle Swarm Optimization, Master's Thesis submitted to the School of Engineering at Blekinge Institute of Technology, 2011.
17. Kharrat, A., Gasmi, K., Ben Messaoud, M., Benamrane, N., & Abid, M. "Medical Image Classification Using an Optimal Feature Extraction Algorithm and a Supervised Classifier Technique". International Journal of Software Science and Computational Intelligence (IJSSCI), *3*(2), 19-33, 2011.
18. Zhang, Y., Wang, S., & Wu, L. "A novel method for magnetic resonance brain image classification based on adaptive chaotic PSO". Progress in Electromagnetics Research, 109, 325-343, 2010.
19. G. Beylkin, R. Coifman, and V. Rokhlin ,Fast wavelet transforms and numerical algorithms, Comm. Pure Appl. Math., 44, pp. 141-183, 1991.
20. B. E. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers". In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, pages 144-152. ACM Press, 1992.
21. P.Brazdil C.Soares and A. Woznica , "Exploiting sampling and meta-learning for parameter setting support vector machines". Proceedings of the IBERAMIA, pages 217–225, 2001.
22. Umamaheswari, J., & Radhamani, G., "An Amalgam Approach for DICOM Image Classification and Recognition. World Academy of Science, Engineering and Technology", pg. 62, 2012.
23. Ingrid Daubechies, "Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics", 1992.