IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 6, Ver. III (Nov.-Dec. 2016), PP 60-65
www.iosrjournals.org

Optimizing the Parameters of SVM with RBF Kernel for Classifying Breast Tumors in Ultrasound Images

P. K. Saranya[1], Dr. E. S. Samundeeswari[2], Dr. R. Manavalan[2+]
Research Scholar, Department of Computer Science, Vellalar College for Women, Tamil Nadu, India[1]
Associate Professor, Department of Computer Science, Vellalar College for Women, Tamil Nadu, India[2]
Assistant Professor, Dept. of Computer Science Arignar Anna Govt. Arts College, Villupuram, Tamil Nadu, India[2+]

Abstract: *Breast cancer is one of the leading causes of death among women worldwide accounting more than 8 million deaths. The most effective way to reduce breast cancer deaths is to detect it earlier. Ultrasound imaging is found to be a real time, non-ionizing tool used mostly for early detection of breast cancer. It is an operator dependent tool, as it needs a physician to analyze the internal echo in making diagnostics decision. Computer Aided System acts as a second reader to physician to make correct diagnosis. In this work, 84 Breast Ultrasound (BUS) images are considered, and 28 morphological and 43 textural features are extracted from each image. Principal Feature Analysis (PFA) dimensionality reduction is applied to extract uncorrelated features and to identify optimal features from uncorrelated feature set, GA, PSO, ACO and ABC optimization techniques are considered. For improving the classification performance, Pattern Search optimization is incorporated in SVM with RBF classifier to optimize Maximal Margin and Kernel Scale parameters. From the experimental results, it is found out that the Optimized Support Vector Machine (SVM) with RBF kernel classifier yields 99.69% accurate classification than the conventional SVM with RBF.*

Keywords: *Breast Ultrasound, MAVM filter, Regularized K-Means Clustering, Feature Selection, Artificial Neural Network, Support Vector Machine, Pattern Search*

I. Introduction

Early diagnosis of Cancer requires an accurate and reliable procedure that allows the physician to distinguish benign tumor from malignant one without going for painful surgical biopsy. Among medical imaging modalities, Breast Ultrasound (BUS) is identified as a non-invasive, less expensive and painless tool which is reliable to screen women with dense breast and adults. It is an operator dependent tool that requires a well trained physician to examine the image. But physicians are often faced with the dilemma of trying to do what is best for the patient, sometimes decide to use a computational system. Designing a Computer Aided System for BUS images will serve as a second reader or beneficial opinion to physician in image interpretation. This research work is carried out to identify a suitable classifier for the last phase of Computer Aided System which will determine the type of tumor appropriately. For the past decades, researchers' analyzed various machine learning algorithms, but the critical task is to identify a most suitable classifier for a particular problem. This work considers Support Vector Machine with Radial Basis (RBF) kernel function and optimized its parameters to improve the classification accuracy.

II. Materials And Methods

Totally 84 Breast Ultrasound images, including 49 benign and 35 malignant categories are collected from ultrasoundcases.info database with HON code (Health on the Net Foundation), a standard code that gives the trustworthy of Health Information [1]. The etiquette information is approved and delivered globally by the Gelderse Vallie Hospital, Ede, Netherlands with the patients consent. The collected images are used for developing a better CAD system for Breast Ultrasound Images. Fig. 1 depicts the stages of CAD System and the classifiers considered in classification phase.

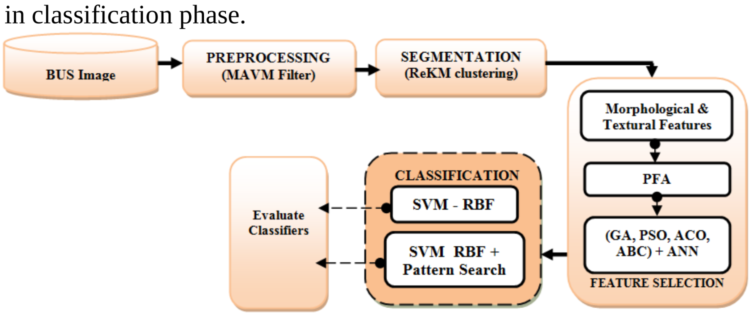


Fig. 1. Classification phase in proposed CAD System

2.1 Preprocessing and Segmentation

Preprocessing BUS image is the first and foremost step to suppress the noises and to make image more suitable for further diagnosis. In this work, Modified AVM (MAVM) Filter, a spatial filter with 5x5 kernel size proposed by Saranya et al. (2016) is used for preprocessing. The filtering process of MAVM is carried out by summing up the medians of row median and row variance, and the medians of column median and column variance of the kernel and is averaged. Then the central pixel value of the kernel is averaged with the resultant valve and replaced [2]. Then the preprocessed image is segmented using Regularized K-Means (ReKM) Clustering algorithm proposed by Samundeeswari et al. (2016) for extracting the ROI. The ReKM is a variant of K-Means clustering concept where the cluster centroids are initialized using Ant Colony Optimization technique and the Euclidean distance metrics is modified with an additive regularization parameter λ [3]. After Segmentation, the morphological operations such as Dilation, Erosion, Open and Close are applied to extract the ROI precisely [3].

2.2. Feature Extraction and Selection

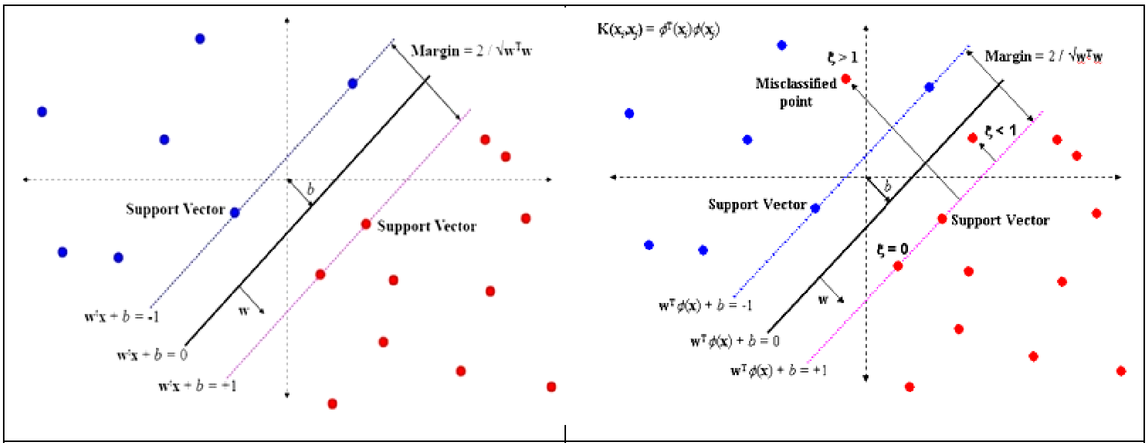
Feature extraction is the essential stage for breast cancer detection and classification. An optimum feature set containing discriminate features will help to classify the dataset accurately with minimum computation time. In this work, 28 morphological and 43 textural features are extracted [4] where the morphological features represent the shape and contour properties of lesion and textural features represent the repeating pattern of local variations in gray level intensity. The 28 shape features include Perimeter, Area, ENC, ENS, LS_Ratio, TCA_Ratio, TEP_Ratio, TEP_Difference, TCP_Ratio, TCP_Difference, AP_Ratio, Form_Factor, Roundness, Aspect_Ratio, Solidity, Convexity, Extent , Volume, Surface Area, Compactness, NRL mean, Sphericity, NRL entropy, NRL ratio, Roughness, Eccentricity, Smoothness and Speculation [4]. The 43 textural features include six Histogram, twenty two Gray Level Co-occurrence Matrix (GLCM), eleven Grey-Level Run-Length Matrix (GLRLM), Fractal Dimension and three Tamura features [4]. The resultant 71 features from each image form a large dimensional dataset. Hence to reduce the dimensionality and to extract uncorrelated features, Principal Feature Analysis (PFA) technique is used. This method computes the coefficient of Principal Components (PCs) and arranges the column vector of these PCs in descending order of Eigen values. The PCs with Eigen value greater than one are selected to form a new matrix that contains highly uncorrelated features. This matrix is subjected to clustering process using K-Means clustering method with the initialization of *k* number of clusters i.e., the number of features required. The features nearest to its cluster centroids are chosen as the principal features. These principal features form an uncorrelated feature set that is optimal in terms of high spread in lower dimension and insensitive to noise [5]. With the aim of finding the discriminating features to form the minimal number of optimal feature subset, this work considers Evolutionary algorithms like Genetic algorithm (GA) and Swarm algorithms such as Particle swarm optimization (PSO) and Artificial Bee Colony (ABC) techniques. These techniques are wrapped with Artificial Neural Network (ANN) as fitness evaluator. The Evolutionary algorithm, GA, a fitness value is assigned to each candidate representing its ability to 'compete'. Three biological processes like Selection, Crossover and Mutation are used to find the best candidate among individuals in a specified search space. Swarm algorithms, inspired by the social behavior of bird flocking and food foraging, keeps track of its solution space in search for food and the solution achieved by it so far is called as personal best, ***pbest*** and the best value that is achieved by it so far and among other populations is called as global best, ***gbest***. In this work, ANN is used as a fitness evaluator for each population. After reaching maximum iteration in each optimization technique, the globally best one is chosen to form the optimal feature subset that yields better classification accuracy.

2.3. Classification

Classification process discovers the relationship between the attributes and undergoes a two step process, consisting of a learning step where a classification model is constructed using training dataset and a classification step where the model is used to predict class labels for new or test dataset. A good and well trained classifier will examines the features or attributes of each instance and classify the instances appropriately to result in correct class label.

2.3.1. Support Vector Machine

Support Vector Machine (SVM), proposed by Vapnik and his co-workers in 1998 is a machine learning method based on the concept of decision planes that define the boundaries and separate the set of objects which are having different class memberships [6] [7]. SVM classifier handles both linear and nonlinear data by transforming them from original feature space to very high dimensional space. For linear data, SVM separates data objects of different class memberships with a line and constructs a hyper-plane in high dimensional space. The best hyper-plane will be represented with large separation or margin between the class labels, i.e. the similar data objects on either side is far apart from the margin. Also larger margin will lower the generalization error of the classifier and such kind of large hyper-plane is called as Optimal Separating Hyper-plane (OSH) or Maximal Margin Hyper-plane (MMH) linear classifier. The separating maximal margin hyper-plane is written as $\boldsymbol{y = w \cdot \varphi(x) + b}$, where $\boldsymbol{w}$ is the weight, $\boldsymbol{x}$ is the feature vector, and $\boldsymbol{b}$ is the bias to separate the samples into two class problems i.e. if $y > 0$, then the sample belongs to class 1 else to class 0. Fig. 2 shows the SVM hyper-plane classifier for linear and nonlinear data.



2-D Linear data –separable by dot product	**2-D Nonlinear data –separable by kernel function**

Fig. 2. SVM classifier separates linear and nonlinear data

For nonlinear data, a straight line hyper-plane can't give optimal solution and hence the nonlinear SVM classifier maximizes the margins and transfers non-separable data to separable data into a dimension of original feature space using a set of mathematical function called kernel function (kernel trick) K. Hence to transform the nonlinear data, the dot product $(x_i \cdot x_j)$ of linear hyper-plane classifier is replaced with the kernel function $K(x_i, x_j)$.

For any two feature vector x_i and x_j of nonlinear data, the hyper-plane of linear data $y = w \cdot \varphi(x)$ is reconstructed as in Eq.(1)

$$y = \sum_i \alpha_i y_i K(x_i, x_j) \text{ and } \quad (1)$$

$$K(x_i, x_j) = \langle \varphi(x_i) . \varphi(x_j) \rangle$$

where x_i , x_j are feature vector, i and j=1,2,..,N, $\phi(x_i)$, $\phi(x_j)$ are the nonlinear mapping functions, $\mathbf{y}$ is the hyper-plane and $\boldsymbol{\alpha}$ belongs between 0 and 1[8]. For any two vectors x_i and x_j,

Radial Basis Function (RBF kernel) is calculated as in Eq.(2)

$$K(x_i, x_j) = \exp^{\left(-\gamma \|x_i - x_j\|^2\right)} \quad (2)$$

where $\gamma = \frac{1}{2\sigma^2}$ and $\|x_i - x_j\|^2$ is the squared Euclidean distance. SVM with Radial Basis Function gives the same type of hyper-plane like neural network and so called as a radial basis function network.

2.3.2. Optimizing Support Vector Machine Parameters using Pattern Search

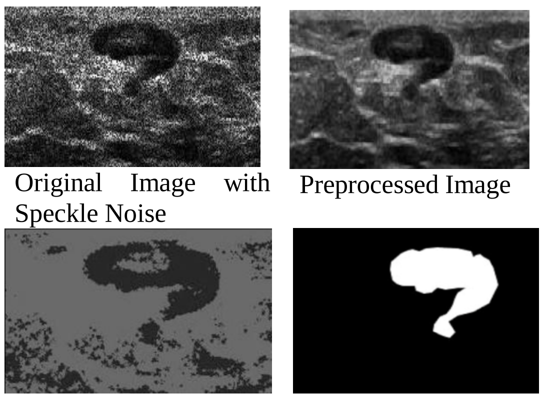
Pattern Search is proposed by Hooke and Jeeves in 1961. It describes sequential examination of trial solutions and compared each trial solution with the best obtained so far to that time to determine the next trial or move. It involves two types of moves such as Exploratory search and Pattern move. Exploratory search is a local search where a current point X_0 looks for an improving direction to move on. Pattern move explores its larger search in improving direction [9].

Initialize
Let n be the number of points and Starting point be X_0, acceleration factor a, perturbation vector P_0, perturbation tolerance vector **T** = $(t_1,t_2,t_3,\dots t_n)$. Let $P \leftarrow P_0$, $f_{best} \leftarrow f(X_0)$
Do
Start
Start Exploratory search around X_0 to get next best improved point X_1 based on the quadratic objective function and yield $f(X_1)$.
If $f(X_0) < f(X_1)$ then $f_{best} \leftarrow f(X_1)$ and reset the perturbation vector
and Go to Pattern Move
Else reset all perturbations to ½ of its current size i.e. $P \leftarrow P/2$
If any P is smaller than its corresponding tolerance in **T** then exit with X^0 as solution **Else** go to start
Pattern Move
Obtain next improved point $X_2=X_0+a[X_1-X_0]$ where **a** is mostly 2.
Start its Exploratory move and calculate fitness $f(X_2)$
If $f(X_2) < f(X_1)$ then update $X_0 \leftarrow X_1$, Go to start
Else $f_{best} \leftarrow f(X_2)$ then update $X_0 \leftarrow X_1$ and $X_1 \leftarrow X_2$ and go to Pattern move
Until (when there is no further improved points)

Hence the f_{best} point will be considered as the optimal point and is used for the box constraints i.e. maximal margin value and kernel scale value while classifying nonlinear data using SVM with RBF kernel classifier.

III. Experiments And Discussion

The 84 Breast Ultrasound images, including 49 benign and 35 malignant categories are collected from ultrasoundcases.info database. The experimental analysis is done in MATLAB 7.0 -2013 B version (Mathworks Inc, USA) environment on a computer system with Intel Core i3 processor, 4 GB RAM and Windows XP operating system. The images are preprocessed using MAVM filter with 5x5 kernel and segmented using ReKM algorithm with regularized parameter as 0.008 [2] [3]. Fig. 3 shows the preprocessed and segmented sample BUS image. From the ROI of all the BUS Images, 28 Morphological and 43 Textural features are extracted and indexed [4]. The extracted numerical values form a 84 x 71 large dimensional dataset and is reduced to 84 x 35 size dataset using PFA and indexed.



Segmented Image

Fig. 3. BUS image- Preprocessed by MAVM filter and segmented by ReKM clustering with λ-0.008

To yield good classification accuracy with minimal number of optimum features, the number of features required is fixed as 15. The three Optimization Techniques such as GA, PSO and ABC are applied to select the most significant and relevant features. Also the common features found in more than one optimization techniques are selected and considered as another subset. Table 1 shows that the common features found in any two optimization techniques.

Table 1. Common Features extracted from subset 15 of GA, PSO and ABC

Methods	Feature subsets	Features Obtained	Common Features
GA	15	1, 2, 5, 8, 9, 10, 12, 13, 14, 16, 22, 23, 27, 32, 33	2, 3, 6, 7, 12, 14, 16, 18, 22, 27, 33, 34
PSO		1, 2, 3, 6, 7, 9, 14, 18, 20, 21, 22, 24, 28, 33, 34	
ABC		2, 3, 6, 7, 12, 16, 17, 18, 22, 25, 26, 27, 31, 33, 34	

It is observed from Table 1, 12 features are found to be common in more than one optimization technique. The 12 features are ENS (elliptic-normalized skeleton), LS_Ratio (long axis to short axis ratio), Roundness, Convexity, Volume, NRL (Normalized Radial Length) mean, NRL (Normalized Radial Length) ratio, Speculation, Homogeneity, Inverse difference normalized (INN), Fractal Dimension and Tamura Coarseness.

To evaluate the classifier qualitatively, Fmeasure (Accuracy) is used. The measure is calculated using Confusion Matrix, which contains True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values.

The accuracy measure is calculated as **Accuracy= TP + TN / (TP + FP + TN + FN)**.

Table 2. Accuracy of GA, PSO and ABC for 15 features

Classifiers	PFA + OPTIMIZATION TECHNIQUES			
	15 Feature subset			12 Features
	GA	PSO	ABC	
ANN	78.57	94.05	65.48	95.42
SVM+ RBF	96.81	99.56	96.62	97.12

From Table 2, it is observed that the ANN yields a classification accuracy of 94% for the 15 features obtained by PSO where as ANN yields 20% to 30% less accuracy for the feature subsets formed by GA and ABC than PSO.

It is also observed from Table 2, the ANN and SVM classifier for classifying 12 features yields a good classification accuracy of 95.42% and 97.12% respectively. Also it is shown that the SVM classifies all the 15 feature subsets and 12 features with an average classification accuracy of 98%. From these analyses, it is found out that SVM with RBF kernel classifier yields a classification accuracy of 97.12% which is less than the classification accuracy for 15 features of PSO with 99.56%. Hence to improve SVM with RBF kernel classification accuracy for a minimal number of features, the parameters of SVM with RBF kernel are optimized using Pattern Search Optimization technique and found 99.69% classification accuracy for 12 features. The performance improvement in classification accuracy after optimizing the parameters of SVM with RBF kernel is shown in Fig. 4.

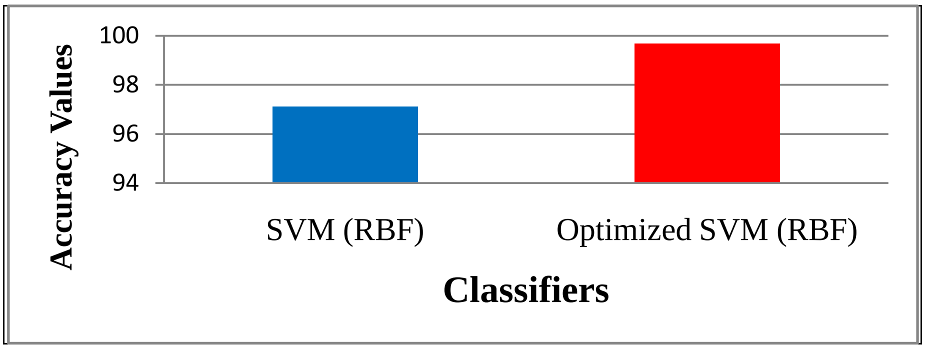


Fig. 4. Accuracy comparison of conventional and optimized SVM + RBF kernel

All these analyses shows that the SVM with RBF kernel is more suitable for classifying the feature set formed from breast ultrasound images. Also the dataset formed with 15 or 12 optimal features are good to differentiate the type of lesion appropriately. The graph also shows that the optimized SVM with RBF kernel will correctly identify and classify the True Positive samples and False Negative samples. Hence the optimized SVM with RBF classifier suits well for the classification phase in the Computer Aided system to distinguish the BUS image as Benign and Malignant accurately.

IV. Conclusion

This paper assesses and identifies the improvement in classification performance after incorporating Pattern Search optimization method to tune the parameter of SVM with RBF kernel classifier. A 84x71 dimensional feature set is subjected to the PFA dimensionality reduction technique to extract uncorrelated features. GA, PSO and ABC, as feature selectors are used to select the significant features to form a set with minimal number of optimal features. The 12 significant features found common in more than one optimization technique feature subset are also selected and classified using the conventional and optimized SVM with RBF kernel function for distinguishing benign from malignant lesion in BUS image. From the experimental analysis, the Optimized SVM with RBF kernel classifier is recognized as a better classifier for classification phase in the CAD system with 99.69% classification accuracy.

References

[1]. Ultrasoundcases. http://ultrasoundcases.info/category.aspx?cat=67. Accessed May 2014.

[2]. Saranya P K, Samundeeswari E S, Manavalan R, Noise Reduction In Breast Ultrasound Images Using Modified AVM Filter, *International Journal of Informative & Futuristic Research (IJIFR)*, 15 December; 3(4), ISSN: 2347-1697, 2015.

[3]. Samundeeswari E S, Saranya P K, Manavalan R. Segmentation of Breast Ultrasound Image Using Regularized K-Means (ReKM) Clustering. *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, IEEE *Xplore* Digital Library, Sep 2016. DOI: 10.1109/WiSPNET.2016.7566362

[4]. Saranya P K, Samundeeswari E S, A Study on Morphological and Textural Features for Classifying Breast Lesion in Ultrasound Images, *International Journal of Innovative Research in Science, Engineering and Technolo*gy, 2016; 5(3).

[5]. Yijuan Lu, Ira Cohen, Xiang Sean Zhou, Qi Tian, Feature Selection Using Principal Feature Analysis. *ACM Multimedia*, Augsburg, Germany, 2007 September; 23-29.

[6]. Vapnik V, *Statistical Learning Theory*, New York: John Wiley & Sons, 1998.

[7]. Christianini N, Shawe-Taylor J, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, United Kingdom: Cambridge University Press, 2000.

[8]. Justice Kwame Appati, Gideon Kwadzo Gogovi, Gabriel Obed Fosu, On the Selection of Appropriate Kernel Function for SVM in Face Recognition, *International Journal of Advanced Research in Computer Science and Software Engineering*, 2014 March; 4(3).

[9]. John W. Chinneck, *Pattern Search for Unconstrained NLP, Practical Optimization: a Gentle introduction*, http;//www.scs.carleton.ca/faculty/chinneck/po.htm, 2014