International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 3 Issue 1, January-2014, pp: (171-176), Impact Factor: 1.252, Available online at: www.erpublications.com

Fusion of SFS-SVM feature selection methods using robust rank aggregation for optimal feature subset selection for mammogram classification

Subodh Srivastava[1], Neeraj Sharma[2], S. K. Singh[3]
[1,2]School of Biomedical Engineering, [3]Department of Computer Science & Engineering
Indian Institute of Technology (BHU), Varanasi, UP, India

Abstract: Feature selection and classification plays an important role in the design and development of a computer aided detection and diagnostics (CAD) tool for breast cancer detection from mammograms. In literature, the various feature selection methods exists such as filter based, wrapper based, and hybrid methods whose aim is to select the most relevant and minimum redundant features from the extracted feature set. The various feature selection methods are based on various basis criteria on which they are designed;and they produce different feature sets even for the same dataset with varying performance measures for a chosen classifier. Therefore, the basic question arises that which one is better and if all of them performs better with slight variations in their performance measures then what should be the robust method to select the optimal feature subset associated with constant and better performance for a chosen dataset? Hence, in this paper, the above mentioned issues are addressed by proposing a wrapper based feature selection method based on sequential forward selection (SFS) and support vector machine (SVM) for optimal feature subset selection. The proposed method is derived from the fusion of SFS-SVM feature selection methods for various kernel spaces of SVM i.e. the feature sets obtained from SFS-SVM-Linear, SFS-SVM-RBF, SFS-SVM-Quadratic, SFS-SVM-Polynomial, and SFS-SVM-MLP wrapper based feature selection methods are fused using robust rank aggregation method. Finally, an optimal feature subset is obtained from the fused feature set by evaluating the minimum misclassification error of a chosen classifier.

Keywords: Optimal feature set selection, wrapper based feature selection, feature fusion, SFS-SVM, robust rank aggregation, misclassification error, performance evaluation.

1 Introduction

Breast cancer is one of the main cancers in women all around the world which is responsible for high mortality rate [1]. An early examination for the possibility of the breast cancer in women at an early stage may reduce the risk for developing the malignant cancer and reduce the mortality rate by the suitable treatment options. For early detection of breast cancer mammography is one of the popular and cost effective test. Digital mammography is a low dose X-ray examination of breasts by obtaining the mammograms which in turn is examined by the radiologists for the breast cancer detection. In large screening mammography programme at a time hundreds or thousands of the patients are examined. The manual examination of each of the mammogram by different experts may lead to diverse results due to their expertise or human errors. Hence, to reduce the human observer variability in mammogram examination automated computer aided detection and diagnosis (CAD) tool [2-6] may be used to classify the available database of mammograms in to normal and abnormal groups. The abnormal groups may further be classified into benign and malignant groups. The second opinion of radiologists may be obtained for final detection and diagnosis of cancer from the abnormal mammograms for recommendation of further examinations, pathological tests and treatment options. The various steps involved in the design and development of a CAD tool include mammogram enhancement, segmentation, feature extraction, feature selection and classification as given in Figure 1.

For demonstration of the proposed method, here in this paper as an application, the feature selection and classification for a CAD tool [2-6] as illustrated in Figure 1 is considered. The initial database of mammograms consists of 322 mammograms (available from MIAS database [7]). Since the mammograms are low contrast X-ray images, hence at first a contrast limited histogram equalization (CLAHE) [8] method is applied on the mammograms in MIAS database to obtain a better contrast mammograms for further processing. In next step, a three class modified fuzzy c-means segmentation [9] is applied on all mammograms to segment the abnormalities present in them. The various abnormalities that may be present in mammogram include tumours and micro-calcifications. After segmentation of all mammograms the feature extraction step is applied for dimensionality reduction. The various features extracted from each mammogram include the features from various categories such as histogram based features [10], geometric or shape features [10], texture features [11], wavelet based features [12], and Gabor features [13]. In feature extraction step total of 88 features, as mentioned in paper [5], were extracted for 322 mammograms resulting in a feature matrix of size 322x88. Now this feature matrix may consist of minimum relevant and redundant features. Therefore to select the maximum relevant and non-redundant feature set, feature selection step is applied for better classification of available images.

In literature [2-6, 14-15], the various feature selection methods are available which include filter based, wrapper based, and hybrid methods. Filter based features selection method use general characteristics of the data independently from the classifier for the evaluation process. The filter based feature selection methods may be unsupervised or supervised in nature. In literature, various types of supervised feature selection methods exists which have demonstrated their efficacy for different datasets belonging to diverse applications. The various types of supervised filter selection methods can be categorized according to the basis criterion used to design these filters [14]. In wrapper based methods [15], the evaluation process is classifier-dependent and uses the learning algorithm as a subroutine. The wrapper based methods equal the bias of both the feature selection algorithm and the learning algorithm that can be used to assess the effectiveness of the solution but the main disadvantage is the extra computational cost that comes from calling the induction/ classifier algorithm to evaluate each subset of considered features. In wrapper based approach the classifier error rate is minimized. The wrapper based feature selection loss its generality, but gain accuracy towards the classification task and is computationally expensive. The hybrid models use both filtering and wrapping methods for improving the performance of the selection process. Evaluating the discrimination power of the individual feature is a key operation in feature selection processes.Due to the advantages mentioned as above, in this paper, a wrapper based feature selection method based on fusion of SFS-SVM feature selection methods for various kernel spaces of SVM classifier using robust rank aggregation [18] is proposed.

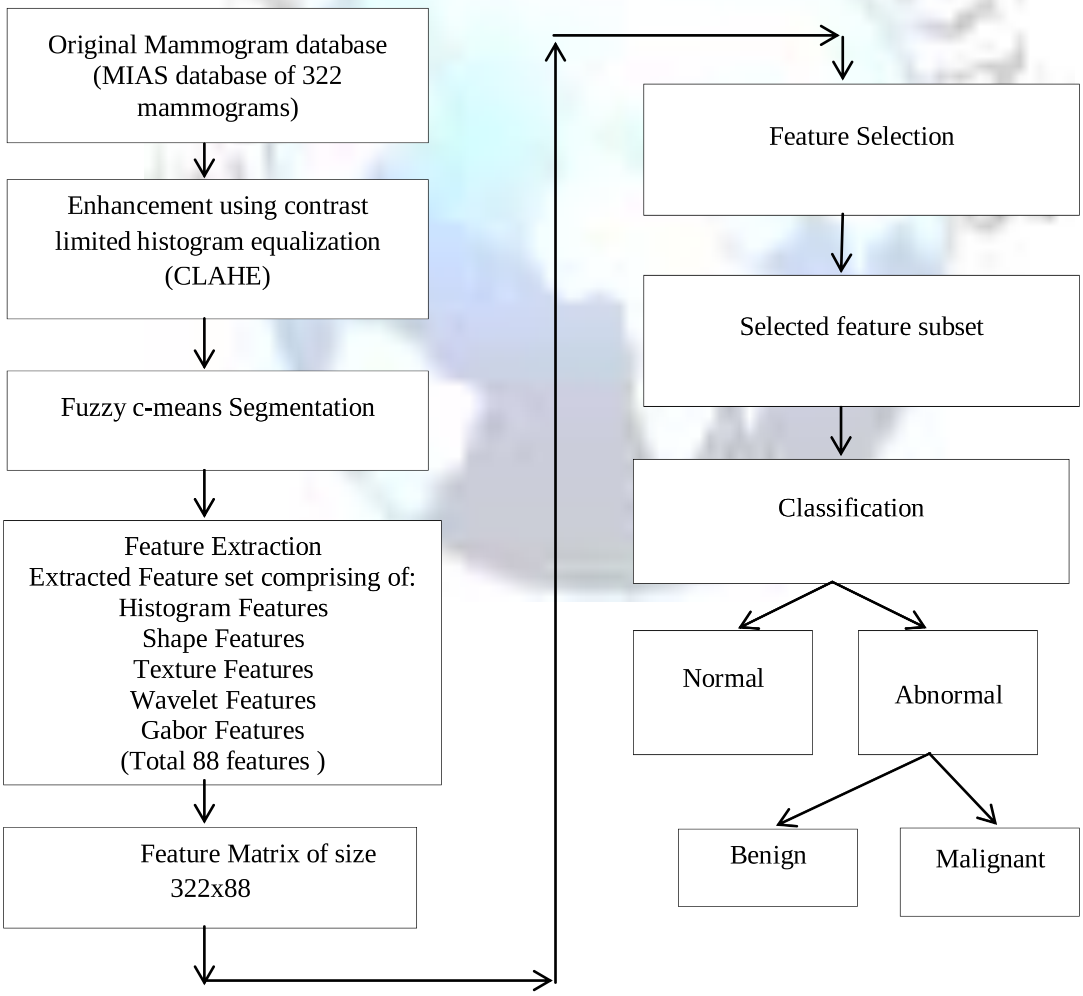

Figure 1: Design steps for a CAD tool for breast cancer detection from mammograms

2 Methods and Models

In this paper, a forward sequential feature selection (SFS) method [15] which uses SVM classifier for various kernel spaces [16-17] are used to propose a new method for feature subset selection. SFS begins with zero attributes and evaluates all features subsets with exactly one feature at a time. It selects the one with the best performance and adds to these subsets the feature that yields the best performance for subsets of next larger size. If evaluation function is a heuristics measure, the feature selection algorithm acts as a filter, extracting features to be used by the main algorithm; and if it is the actual accuracy, it acts as a wrapper around that algorithm.The wrapper based forward sequential search selection (SFS) method employs a classifier to select a best feature subset. These methods use cross-validated misclassification rate of a classifier as a criterion function to select a best feature subset. Since the performance of a SVM classifier also varies for the various kernel spaces it uses, hence the new filter is proposed by the fusion of the feature subsets produced by the various SFS-SVM filters for different kernel spaces using robust rank aggregation. These feature selection methods include SFS-SVM-Linear, SFS-SVM-RBF, SFS-SVM-Quadratic, SFS-SVM-Polynomial, and SFS-SVM-MLP. Each method in this category stops the selection of a feature when a local minimum of the criterion function is reached. However, for testing, purposes total of 10 features in each feature subset were selected. The various feature subsets produced by these methods are listed in Table 1. From Table 1, it is observed that the SFS-SVM classifier produces different feature subset for the same dataset which may be associated with different performance measures for different classifiers. Therefore, the basic question arises that if SFS-SVM feature selection method is used, then which SVM kernel function will be most appropriate for a chosen dataset. Hence to address this issue, a robust method is proposed for optimal feature selection for classification. The proposed model is illustrated in Figure 2. In Figure 2, the feature subsets produced by the above mentioned five feature selection methods are FS1, FS2, FS3, FS4, and FS5 respectively. FS is the feature set obtained by rank aggregation. In the proposed method, a robust rank aggregation method [18] is used in place of conventional majority voting method.The working of robust rank aggregation method as proposed in paper [18] is given as follows:

Let for any normalized rank vector R, let $R_1, R_2 \ldots\ldots, R_n$ be a reordering of ranks R such that $R_1 \le R_2 \ldots\ldots \le R_n$, then the task is to find that how probable it is to obtain $\widehat{R}_k \le R_k$ when the rank order $\widehat{R}$ is generated by the null model i.e. all ranks $\widehat{R}_i$ are sampled from uniform distribution. If $\beta_{k,n}(R)$ denote the probability that $\widehat{R}_k \le R_k$, then under the null model the probability that the order statistic $\widehat{R}_k \le x$ can be represented by the binomial probability as follows since at least k normalized rankings must be in range $[1, x]$.

$$\beta_{k,n}(x) = \sum_{l=k}^{n} \binom{n}{l} x^l (1-x)^{n-l} \quad (1)$$

The $\beta_{k,n}(R)$ can also be expressed through a beta distribution because $\widehat{R}_k$ is order statistic of n independent random variables uniformly distributed over the range [0,1]. As the number of the informative ranks of the features is not known, the final score for the rank feature vector R is defined as the minimum of P-values as follows:

$$\rho(R) = \min_{k=1..n} \beta_{k,n}(R) \ . \ (2)$$

Lastly, all rank features are ordered according to their ρ scores.
After obtaining the fused feature set by robust rank aggregation, the next task is to obtain the top few optimal features. This is being achieved by evaluating the misclassification error of a chosen classifier for various feature subsets whose size varies from 1 to Nf, where Nf is the number of features selected from fused feature set. A feature subset associated with minimum misclassification error is selected.In this paper, the suitability of various classifiers is also evaluated for their efficacy for optimal feature selection.

Table 1: Selected feature sets by different wrapper based classifiers for given mammogram data set

Wrapper based supervised feature selection method		Feature Set Selected in order of their importance (First feature-highest rank/importance)
SFS-SVM	LINEAR	[5,8,14,19,37,39,40,42,55,58]
	RBF	[8,14,37,40,52,55,57,58,67,70]
	Quadratic	[1,2,3,12,22,23,31,40,43,48]
	Polynomial	[11,39,40,42,57,60,68,76,78,86,34]
	MLP	[4,17,20,38,39,43,46,49,53,55]

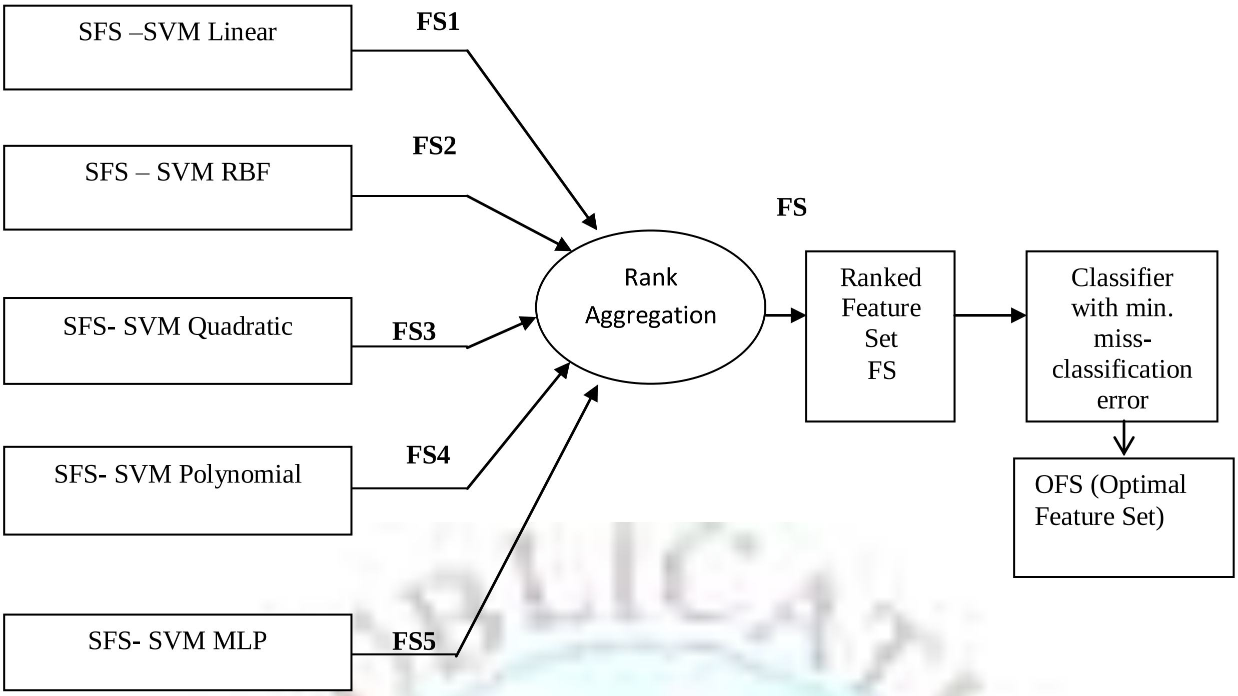

Figure 2: Optimal feature subset selection by fusion of SFS-SVM ensemble feature subsets using robust rank aggregation and by evaluation of minimum miss-classification error of chosen classifier

3 Results and analysis

The proposed method is evaluated on the mammogram dataset/extracted feature set by the procedure as illustrated in Figure 1 and explained in section 1. The fused feature set (FS) and optimal feature subset (OFS) selected by the proposed method are given in Table 2. Table 3 lists the minimum hold out test misclassification errors (MCE) for the various classifiers and selected number of feature subset from the fused feature set (FS).The various classifiers [18-19] chosen for evaluation of the method and selection of OFS are Naïve Bayes, k-nearest neighbour for number of neighbours k=5, linear discriminant analysis (LDA), artificial neural network (ANN), and support vector machines for various kernel spaces such as linear, radial basis function (RBF), polynomial, quadratic and multilayer perceptron (MLP). Figure 3 visually represents the misclassification error associated to all features for all classifiers in consideration. From Table 3 and Figure 3, it is observed that SVM-polynomial classifier is associated with min MCE of 0.0114 for top 12 features from the fused feature set (FS). The next classifiers which are associated with min MCEs after SVM-polynomial are SVM-RBF and k-nearest neighbour (KNN) for top 11 and 15 features respectively. Hence, SVM-polynomial is performing better with optimal feature set (OFS) size of 12 for breast cancer classification from mammograms. From computational complexity point of view kNN is faster in comparison to other methods.

Table 2: Feature subsets fusion using robust rank aggregation and optimal feature subset

Feature selection method	Selected feature subset
Feature Fusion of SFS-SVM-all-kernel-spaces using robust rank aggregation (FS)	[8, 14, 40, 37, 1, 4, 5, 55, 58, 39, 6, 9, 17, 49, 20, 11, 19, 38, 70, 12, 52, 28, 43, 31, 46, 57, 62, 34, 42, 53, 67, 81, 88]
Final feature subset obtained from feature list obtained by robust rank aggregation associated with minimum misclassification error (OFS)	[8,14,40,37,1,4, 5,55,58,39,6,9]: SVM-Polynomial [8, 14, 40, 37, 1, 4, 5, 55, 58, 39, 6]: SVM-RBF [8, 14, 40, 37, 1, 4, 5, 55, 58, 39, 6, 9, 17, 49, 20]: KNN

Table 3: Minimum Classification Errors and final features list obtained from various classification models from the aggregated ranked features

Type of misclassification Error(MCE)	Classification model used to evaluate error		Min. miss-classification error	Indices of aggregated rank feature list related to minimum error
Holdout Test MCE	Naïve Bayes		0.0167	1-2
	KNN		**0.0124**	15-17
	DA		0.0128	22-24
	ANN		0.0130	27
	SVM	Linear	0.0144	6-7
		RBF	**0.0124**	11
		Quadratic	0.0140	5
		Polynomial	**0.0114**	12
		MLP	0.0133	17

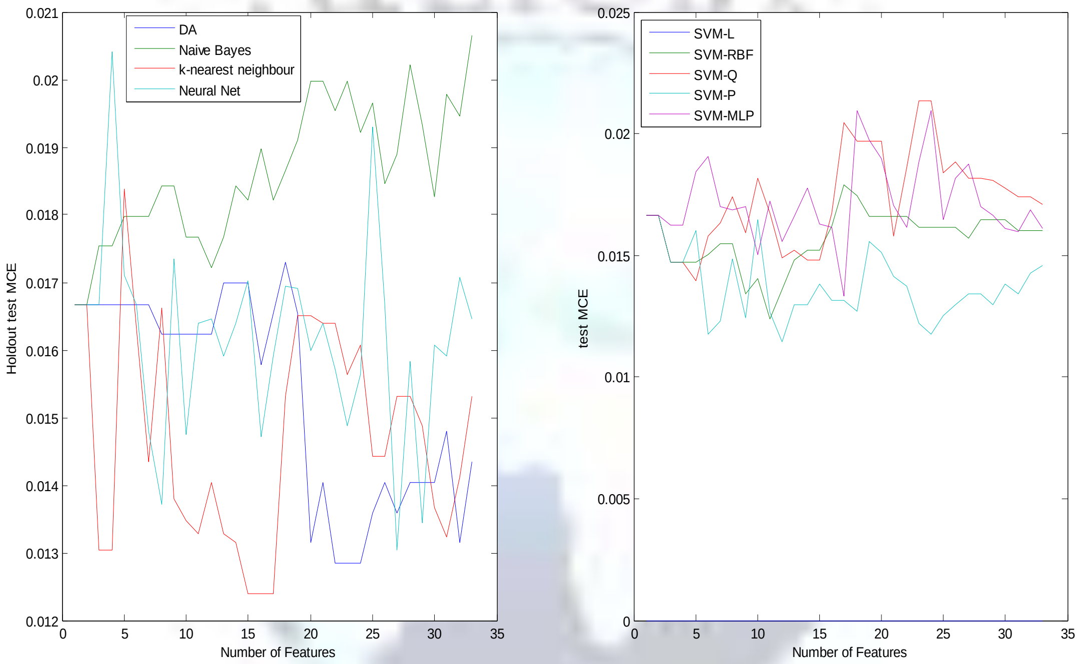

Figure 3: Hold test miss-classification error for selected feature subset by various classifier

4 Conclusions

In this paper, a wrapper based feature selection method based on robust rank aggregation is proposed. The proposed method is derived from the fusion of SFS-SVM feature selection methods for the various kernel spaces of SVM classifier. The five wrapper based feature selection methods fused are SFS-SVM-Linear, SFS-SVM-RBF, SFS-SVM-Polynomial, SFS-SVM-Quadratic, SFS-SVM-MLP. For fusion of feature subsets produced by various individual methods are performed using a robust rank aggregation in place of very common method such as majority voting. The robust rank aggregation method removes the outliers and noises that may be present in features. For optimal feature subset selection, the holdout test misclassification errors for various classifiers were evaluated for top Nf features, where Nf=1 to number of features in FS. The classifier with minimum MCE for top Nf features gives the optimal feature set (OFS). From the obtained results, it is observed that the SVM-Polynomial is performing better in comparison to other classifiers for MCE and OFS evaluation with OFS size being the top 12 features from the fused feature set (FS). The other classifiers which are performing better after SVM-Polynomial are SVM-RBF and KNN (k-nearest neighbour). The processing time for the KNN classifier is less in comparison to all other classifiers; hence it may also be used.

References

[1]. American College of Radiology (ACR) (2003): ACR Breast Imaging Reporting and Data System, Breast Imaging Atlas, 4th Edn., Reston, VA, USA.

[2]. Cheng, H.D., Shi, X.J., Min, R., Hu, L.M., Cai, X.P. and Du, H.N. (2006) 'Approaches for automated detection and classification of masses in mammograms', Pattern Recognition, Vol. 39, pp. 646 – 668.

[3]. J. Tang, R. Rangayyan, J. Xu, I. El Naqa, Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography, IEEE Transactions Recent advances, Information Technology in Biomedicine, Vo. 13, No. 2, 236–251, 2009.

[4]. Ganesan, K., Acharya, U. R. , Chua, C. K., Min, L. C. , Abraham, K. T. and Ng, K.-H. (2013) 'Computer-Aided Breast Cancer Detection Using Mammograms: A Review', IEEE Reviews in Biomedical Engineering, Vol. 6, pp. 77-98.

[5]. Srivastava, S., Sharma, N., Singh, S.K. and Srivastava, R. (2013) 'Design, analysis and classifier evaluation for a CAD tool for breast cancer detection from digital mammograms', Int. J. Biomedical Engineering and Technology, Vol. 13, No. 3, pp. 270–300.

[6]. SubodhSrivastava, Neeraj Sharma, S.K. Singh, "Image Analysis and Understanding Techniques for Breast Cancer Detection from Digital Mammograms", in Research developments in image processing and computer vision, R. Srivastava, S.K. Singh, and K.K. Shukla (Eds.), Chapter 8, pp. 123-148, 2013, IGI Global, USA.

[7]. Mammographic Image Analysis Society (MIAS) database. (2013). http://www.mammoimage.org/databases/

[8]. Zuiderveld, K. (1994) 'Contrast Limited Adaptive Histogram Equalization', Graphic Gems IV. San Diego: Academic Press Professional, pp. 474–485.

[9]. Cai, W., Chen, S. and Zhang, D. (2007) 'Fast and robust fuzzy c-means clustering algorithm incorporating local information for image segmentation', Pattern Recognition, Vol. 40, No. 3, pp. 825-838.

[10]. Jain, A.K. (2006) Fundamentals of Digital Image Processing, PHI, India.

[11]. Haralick, R.M., Shanmugam, K. andDinstein, I. (1973) 'Textural features for image classification', IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-3, No.6, pp. 610-621.

[12]. Huang, K. and Aviyente, S. (2008) 'Wavelet Feature Selection for Image Classification', IEEE Transactions on Image Processing, Vol. 17, No. 9, pp. 1709-1720.

[13]. Grigorescu, S.E., Petkov, N. and Kruizinga, P. (2002) 'Comparison of Texture Features Based on Gabor Filters', IEEE Transactions on Image Processing, Vol. 11, N0. 10, pp. 1160-1167.

[14]. Feng Yang and K.Z. Mao, "Robust Feature Selection for Microarray Data Based on Multicriterion Fusion",IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, VOL. 8, NO. 4, JULY/AUGUST 2011.

[15]. M. Das, H. Liu, "Feature Selection for Classification", Intelligent Data Analysis, vol. 1, pp. 131–156, 1997.

[16]. T.S. Furey, N. Cristianini, N. Duffy, D.W. Bednarski, M. Schummer, and D. Haussler, "Support Vector Machine Classification and Validation of Cancer Tissue Samples Using Microarray Expression Data," Bioinformatics, vol. 16, no. 10, pp. 906-914, Oct. 2000.

[17]. V. Vapnik, Statistical Learning Theory. John Wiley & Sons, 1998.

[18]. RaivoKolde, Sven Laur, Priit Adler and JaakVilo, "Robust rank aggregation for gene list integration and meta-analysis ", BIOINFORMATICS, Vol. 28 no. 4 2012, pages 573–580.

[19]. Duda, R. O., Hart, P.E. and Stork , D.G. (2000) Pattern Classification (2nd Edition), Wiley, India.