Empirical Analysis of Supervised and Unsupervised Filter based Feature Selection Methods for Breast Cancer Classification from Digital Mammograms

Subodh Srivastava
School of Biomedical Engg.
Indian Institute of Technology(BHU)
Varanasi-221005, UP, India

Neeraj Sharma
School of Biomedical Engg.
Indian Institute of Technology(BHU)
Varanasi-221005, UP, India

S.K. Singh
Department of CSE
Indian Institute of Technology(BHU)
Varanasi-221005, UP, India

ABSTRACT
In the design and development of an automated CAD tool for breast cancer detection and diagnosis, the various steps include enhancement, segmentation, feature extraction, feature selection and classification. The feature selection plays an important role in the design of the said CAD tool as it aims towards the redundant feature elimination and relevant feature selection. The selected feature set also decides the efficacy of the chosen classifier for classification of mammograms. In literature, various filter based feature selection methods exists under unsupervised and supervised categories based on different basis criterion. The filter based feature selection methods ranks the extracted feature sets based on some criteria in descending order of their importance. The various methods produce different feature subsets which are associated with different performance measures. In this paper, an evaluation and comparative study of various unsupervised and supervised feature selection methods are presented for breast cancer classification from digital mammograms though various classifiers. The study aims towards finding out the better feature selection method and associated classifier which gives better performance.

General Terms
Digital Mammography, CAD Tool, Breast Cancer detection, Feature Selection, Pattern Classification

Keywords
Supervised feature selection methods, unsupervised feature selection methods, comparative study, classifier selection, CAD tool, breast cancer detection, MIAS database.

1. INTRODUCTION
Breast cancer is the most common form of cancer and it is the second leading cause of cancer death after lung cancer. Women in the India have about a 1 in 11 lifetime risk of developing invasive breast cancer. An early detection of breast cancer increases the survival rate and increases the treatment options. For early breast cancer detection, mammographic X-ray images are playing an important role due to its cost effectiveness and capability to detect the breast to cancer. Annual mammogram screenings, combined with other imaging based examination such as ultrasounds and MRIs, significantly increase the detection of cancer in women who had high risk of breast cancer. It has been reported [1] that out of more than 2,600 women who were at increased risk of breast cancer and they either had dense breast tissue and at least one other risk factor, such as a family history of the disease, 53 percent of the cancers were detected through mammograms[1-2]. Radiologists visually search mammograms for specific abnormalities. Screening mammography, radiographic imaging of the breast is currently the most effective tool for early detection of breast cancer. The various types of abnormalities which are indicator of breast cancer include micro-calcifications, tumours, and architectural distortions etc. Calcifications are deposits of calcium in breast tissue. Micro calcifications are usually associated with extra cell activity in the breast tissue and are grouped in clusters that can be a sign of developing malignant tumor. Scattered micro calcifications are usually a part of benign breast tissue. In mammograms calcifications are seen as bright dots of different sizes. The exact position of micro calcifications cannot be predicted, as well as their number [1]. Breast tumors and masses usually appear in the form of dense regions in mammograms. Breast density is a measure of the extent of radio dense fibro glandular tissue in the breast which has the potential to be used as a predictor of breast cancer risk, it is a measure of how well tissue can be seen on mammogram [1]. Some tissue, such as the milk gland, is dense and appears white on a mammogram. This density makes it hard for medical professionals to observe tumors, which also appear white. Fatty tissue is less dense and appears clear on the mammogram, allowing better tumor detection [3]. Architectural distortion is defined as distortion of normal architecture with no definite mass visible, including speculation radiating from a point and focal retraction or distortion at the edge of the parenchyma. Architectural distortion of breast tissue can indicate malignant changes especially when integrated with visible lesion such as mass, asymmetry or calcifications. A mass is defined as a space occupying lesion seen in at least two different projections. If a potential mass is seen in only a single projection it should be called asymmetry or asymmetry density [2]. Masses have different density such as fat containing masses , low density , iso-dense, and high density;, different margins such as circumscribed , micro lobular , obscured, indistinct; and speculated and different shapes such as round , oval , lobular , irregular round and oval [4-5]. Masses must be classified as benign or malignant. A typical benign mass has a round, smooth and well circumscribed boundary. A malignant tumor usually has a speculated, rough, and blurry boundary. There are subtle signs that can also lead to breast cancer diagnosis, such as architectural distortion and bilateral asymmetry [4-6]. Based on information discussed as above as well as other related pathological signs the radiologists arrive at conclusions to detect cancer from mammograms. In large scale mammogram screening program where numerous mammograms have to be examined by the various radiologists observers there may be chances of error in breast cancer detection, results may not be consistent due to human observer variability, and may be time consuming. Also in the

examination of single mammogram there may be chances of error in breast cancer detection due to lack of expert professionals. Hence, an automated computer aided detection and diagnosis (CAD) [5-6] tool may be used for primary examination of mammograms for breast cancer detection and classification. If the CAD tool detects the cancer from mammogram the second opinion of radiologists may be sought for second opinion thus reducing the cancer detection and diagnosis time. In the design and development of a CAD tool for breast cancer detection the various primary steps involved include the restoration, enhancement, and segmentation of mammograms. The segmentation step aim towards the extraction of abnormalities such as micro-calcifications, tumours, and architectural distortions. The next steps in the design of CAD tool include the features extraction from segmented image, feature selection method to select minimum redundant and maximal relevant features, and finally classification of mammograms in to normal and abnormal classes based on selected features. All these above mentioned steps play an important role in the design and development of a CAD tool.

In this paper, the feature selection and classification steps are addressed. This is due to the facts that in literature there are numerous feature selection and classification methods which can be used in the design of the said CAD tool and all these methods are associated with different performance measures due to their inherent characteristics and design features. Further, an empirical evaluation and comparative study of various supervised and unsupervised filter based features selection methods in terms of various classifiers are presented for mammogram dataset derived from MIAS database.

2. FEATURE SELECTION: BACKGROUND

In the design and development of a CAD tool for breast cancer detection from mammograms extracted feature sets from segmented mammogram are used for classification task. The extracted feature set may also contain redundant and non-relevant features hence an appropriate feature selection method is applied. To select the relevant features from the large set of candidate features are called feature selection. If we extract N number of features for each mammogram in database then a feature vector is formed for each mammogram and all feature vectors for whole mammogram present in database constitute a feature matrix. After constituting a feature matrix, the next task is to select the most relevant features which best describes the mammographic image content from the overall feature space. This can be done by computing some sort of feature scores for each images and sorting or ranking them according to their relevance. The main objective of features selection are that it reduces the dimensionality of feature space, speed up and reduce the cost of learning algorithms, improve the predictive accuracy of classification algorithm, and also improve the visualization and the comprehensibility of the induced concepts. The feature selection algorithms may be based on three major criterions such as based on some evaluation measure; based on search organization; and based on the generation of successors [7, 9] as presented in Table 1. The features selection algorithms [9-10] are broadly divided into three categories: filter based, wrapper based, and hybrid.

Table 1. Feature selection methods characterizations based on different criterion and their types

Characterization criterion	Types
Evaluation measure	Distance based
	Divergence based
	Information theoretic based
	Dependence measure based
	Accuracy based
Search organization	Exponential
	Sequential
	Random
Generation of successors	Forward selection
	Backward selection
	Compound selection
	Random selection
	Weights based selection

Filter based features selection method use general characteristics of the data independently from the classifier for the evaluation process. In wrapper based methods, the evaluation process is classifier-dependent and uses the learning algorithm as a subroutine. The general argument in favour of this scheme is to equal the bias of both the feature selection algorithm and the learning algorithm that is used later to assess the goodness of the solution. The main disadvantage associated with wrapper based scheme is the extra computational cost that comes from calling classifier algorithm to evaluate each subset of considered features. For optimal feature selection in wrapper based method, the classifier error rate is minimized and a feature subset associated with minimum misclassification error is selected. The wrapper based feature selection method losses its generality, but gain accuracy towards the classification task and is computationally extensive. 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. Several methods may be used to evaluate the discrimination power of a feature which includes distance, divergence, information, dependence, and accuracy based criterions. In information theoretic based approaches the mutual information is used for measuring the feature and data relations. Inter cluster and inner-cluster affinity characterizes the relationship between features and classes; thus, they are for different attributes of feature–data relations.

All these three categories (filter, wrapper, and hybrid) [9] can be divided into supervised and unsupervised basis; further supervised and unsupervised can be categories as multivariate and univariate. In supervised learning, the data is assigned to be known before computation and are used in order to 'learn' the parameters that are really significant for those clusters. Here each object in the data set comes with a pre assigned class label. The main task is to train a classifier to do the labelling but often the labelling process cannot be described in an algorithmic form, hence the machines is equipped with learning skills and present the labelled data to it. The classification knowledge learned by the machine in this process might be obscure, but the recognition accuracy of the classifier will be the judge of its adequacy. In unsupervised learning, the datasets are assigned to segments, without the clusters being known. In literature, numerous clustering algorithms have been and are being developed for unsupervised learning. Some examples include k-means clustering and fuzzy c-means clustering approaches [8]. In unsupervised learning approach, different algorithms exists with different structures for the same set of data and the disadvantages and advantages of these approaches are that there is no ground truth against which the results of an algorithm can be compared. The only indication of how good the result is probably the subjective estimate of the user. Further the feature selection process may be defined for univariate and multivariate data as the case may be for both types of learning approaches viz. supervised and unsupervised. In univariate data analysis, it is assumed that the response variable is influenced only by one other factor whereas in multivariate data analysis it is assumed that the response variable is influenced by multiple factors and even combinations of factors. In mathematics, univariate refers to an expression, equation, function or polynomial of only one variable. Objects of any of these types but involving more than one variable may be called multivariate.

2.1 Filter based feature selection methods

In this paper, an empirical evaluation of filter based feature selection methods [9-10] are presented in terms of various classifier performance measures. As discussed above, the filter based feature selection methods may be broadly categorized into two categories namely unsupervised and supervised defined for univariate or multivariate data. A classification of some of the popular filter based feature selection methods under these categories are presented in Table 2. Further classification of filter based feature selection methods can be done according to the basis criterion used for evaluating the discrimination power of features. The various types of evaluation functions are presented in Table 3. The categorization of some of the popular filter based feature selection methods which uses different basis criterion [8-9, 11--17] or evaluation function are presented in Table 4.

Table 2. Classification of some filter based feature selection methods on the basis of supervised and unsupervised learning approaches

Filter based feature Selection methods	Supervised		Unsupervised	
	Multivariate	**Univariate**	**Multivariate**	**Univariate**
Relief F [11]	**Yes**	No	No	No
mRmR [12]	**Yes**	No	No	No
FCBF [13]	**Yes**	No	No	No
GFLIP [14]	**Yes**	No	No	No
Fisher score [8]	No	**Yes**	No	No
SVM- RFE [15]	**Yes**	No	No	No
t-test [8-9]	No	No	No	**Yes**
Bhattacharya distance [8-9]	No	No	No	**Yes**
Wilcoxon paired test [8-9]	No	No	No	**Yes**
ROC based [8-9]	No	No	No	**Yes**
Entropy based [8-9]	No	No	No	**Yes**
PCA [8]	No	No	**Yes**	No
Laplacian Score [16]	No	No	No	**Yes**

Table 3. Evaluation functions used by filter based feature selection methods [9]

Basis Criterions/ Evaluation function used	Examples
Distance based measures	Euclidean distance
Information theory based measures	Entropy, information gain, mutual information
Data dependency measures	Correlation coefficient
Consistency based measures	Minimum features bias

Table 4. Brief description of some of the popular filter based feature selection methods

Filter based feature selection methods	Major basis criterion

Supervised feature selection methods	Fisher Score [8]	Distance based, univariate filter method evaluating each feature individually
	ReliefF [11]	A multivariate filter method taking into account dependencies between features.
	mRmR [12]	Information theory based and uses mutual information (MI)
	FCBF [13]	Based on information gain, FCBF is Fast Correlation-Based Filter
	SVM-RFE [15]	SVM-RFE method ranks features based on their corresponding coefficients in the SVM classifier
	GFLIP (greedy feature flip) [14]	Margin based greedy feature selection algorithm
Unsupervised rank based feature selection methods	ttest scores	Statistical, rank based feature selection methods [8-9]
	Bhattacharyya distance	
	Wilcoxon signed-rank test	
	Entropy Rank feature	
	Receiver operating characteristic (ROC) based	
	Principal component analysis (PCA) [8]	PCA finds a ***linear*** projection of high dimensional data into a lower dimensional subspace such as: the variance retained is maximized; the least square reconstruction error is minimized.
	Laplacian Score based [16]	Laplacian Score (LS) is an unsupervised feature selection algorithm based on Laplacian Eigen maps and Locality Preserving Projection. The basic idea of LS is to evaluate the features according to their locality preserving power.

3. METHODS AND MODELS

This paper presents the empirical analysis of supervised and unsupervised filter based feature selection methods for breast cancer classification from digital mammograms. For experimentation purposes the 322 mammogram images available in MIAS database were used. For the evaluation of the feature selection and classification methods following steps were used in the design of the CAD tool.

3.1 Enhancement of mammograms: The first step in the design of the CAD tool is enhancement of mammograms for highlighting the abnormalities such as micro-calcifications, tumours etc. In this step, a contrast limited histogram equalization (CLAHE) [18] method combined with unsharp masking and crispening [19] were used.

3.2 Segmentation of mammograms: For the segmentation of abnormalities present in the mammograms a modified fuzzy c-means based thresholding [20] method was used.

3.3 Feature Extraction [21]: After segmentation process, the 88 hybrid features were extracted for each of the segmented mammogram. The various features extracted belong to the various categories which are histogram based features, shape based features, texture features, wavelet based features, and Gabor features. The brief descriptions of extracted hybrid features (total 88 numbers) from each segmented mammograms are given as below:

Histogram based features [19,21] (F1-F16]: Mean, Standard Deviation, Gray Level Local Variance, Variance, Kurtosis, Skewness, Entropy, Histogram Range, Mean Absolute Deviation (MAD), Second Order Moment (Var), Mean of Z-score, Normalized gray level variance, Mean energy of gradient, Threshold Gradient, Squared gradient, Spatial Frequency.

Texture Features [21-22]: (F17-F37) Angular Second Moment (ASM), Contrast, Correlation, Variance, Standard deviation, Dissimilarity, IDM, Energy, Entropy, Cluster shade, Cluster Prominence, Sum Average, Sum Entropy, Sum Variance, Difference Variance, Difference Entropy, Information measure and others. These features have been derived from the gray level co-occurrence matrix (GLCM) probabilities.

Geometric or shape Features [19, 21]: (F38-F44) Area, Perimeter, Orientation, Equivalent diameter, Euler number, Eccentricity, Image Curvature. Geometric features describe the geometric properties of the region of interest (ROI). It is represented as a collection of pixels in an image.

Wavelet based features [21, 23-24]: (F45-F52) Mean Entropy, Energy, Contrast, Homogeneity, Sum of wavelet coefficients, Variance of wavelet coefficients, Wavelet ratio.

Gabor Features [21, 25]: (F53-F88) Mean square energy Orientation i.e. Mean Amplitude (at 3rd level of decomposition having 6 orientations i.e. 36 features)

3.4 Feature Selection:

The next step in the design of the CAD tool is the feature selection and evaluation of various filter based supervised and unsupervised feature selection methods are the main focus of this paper. A brief description of various feature selection methods are presented in section 2. The various supervised filter based feature selection methods considered in this paper are as follows: ReliefF [11], mRmR [12], FCBF [13], GFLIP [14], Fisher score [8], and support vector machine –recursive feature elimination (SVM-RFE) [15]. Further, the various unsupervised filter based feature selection methods in consideration includes various rank based feature selection methods [8-9] such as t-test scores, Bhattacharyya distance, Wilcoxon signed-rank test, entropy based, receiver operating characteristic (ROC) based; principal component analysis (PCA) based [8], and Laplacian score [16] based methods.

3.5 Classification

[8]: The various supervised classifiers used for the evaluation of various feature selection methods for the design and analysis of the CAD tool include Naïve Bayes , k-nearest neighbour (k-NN), linear discriminant analysis (LDA), artificial neural networks (ANN), and support vector machines [26] for its various kernel functions such as linear, RBF, quadratic, polynomial, and multilayer perceptron (MLP).

4. RESULTS AND ANALYSIS

4.1 Dataset description

For evaluating the performance measures of various filter based feature selection methods, the mammograms available in MIAS database were used. The MIAS database [27] consists of 322 mammographic images which contain 207 images for normal cases and 115 images for abnormal cases .The MIAS database images includes radiologist's "truth"-markings on the locations of any abnormalities that may be present. Normal cases mean only normal breast tissues. Abnormal cases include both benign and malignant images, which includes radiologist's "truth" markings on the locations of any abnormalities that may be present. The various pathological features present in abnormal cases are as follows [27]:

Character of background tissue: Fatty, Fatty-glandular, and Dense-glandular.

Class of abnormality present: Calcification, Well-defined/circumscribed masses, speculated masses, miscellaneous other ill-defined masses, architectural distortion, asymmetry.

4.2 Performance measures

The performance measures used for the evaluation of the feature selection methods in terms of chosen classifier include accuracy, sensitivity, specificity, balanced classification ratio (BCR), F-measure, and area under curve (ROC). These measures are defined in terms of confusion matrix elements TP(true positive), TN (true negative), FP (false positive), and FN (false negative) as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (1)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

$$Specificity = \frac{TN}{FP+TN} \quad (3)$$

Balanced classification rate,

$$BCR = \frac{1}{2}\left[Sensitivity + Specificity\right] \quad (4)$$

F-measure is harmonic mean between precision and recall and defined as

$$F\_measure = 2 * \left[\frac{(\mathrm{Pr}ecision \times \mathrm{Re}call)}{(\mathrm{Pr}ecision + \mathrm{Re}call)}\right] \quad (5)$$

where $\mathrm{Pr}ecision = \frac{TP}{TP+FP}$ and $\mathrm{Re}call = Sensitivity$.

Area under curve (AUC) is defined as the area under the receiver operating characteristic (ROC) curve.

4.3 Performance analysis and discussions

For testing purposes, top 50 most relevant features out of total 88 hybrid features were selected for each feature selection method. For hybrid features, the feature matrix size reduced to 322x50 after feature selection. The images in MIAS database were categorized into two groups namely normal and abnormal images and a class or group of size 322x1 was used. In next step, 10-fold cross-validations were applied to split the 322x50 size feature matrix and 322x1 class size into two sets viz. train and test data sets and groups respectively. The size of train feature matrix was 289x50 and that of the test feature matrix was 33x50. Finally, selected feature subsets by each of the feature selection methods were used to evaluate the performances of the various supervised classifiers to find the best combination of feature selection method and classification model for the said CAD tool. For training, testing, and measuring the classifier's performance measures, 10-fold cross-validations were used and the results are reported for the average of 100 runs for all the cases presented in this paper.

Tables 5, present the comparative analysis of the various *supervised feature selection methods* in consideration. Fig 1 shows comparison of supervised filter based feature selection methods in terms of classifier's maximum accuracy , sensitivity, and specificity values , and Fig 2 shows comparison of supervised filter based feature selection methods in terms of classifier's maximum AUC, BCR, and F-measures values. From Table 5 and Figs. 1-2 following observations are made:

- ReliefF feature selection method is performing better for SVM-MLP classifier with accuracy 84.37%;
- mRmR feature selection method is performing better for SVM-MLP classifier with accuracy 87.50%;

- FCBF is performing better for SVM-Quadratic with 78.12%, however SVM-RBF is associated with 87.50 % of accuracy but sensitivity value is 1 and specificity value is 0 giving boundary conditions i.e. SVM-RBF is not capable of classifying both positive and negative samples simultaneously for the dataset in consideration.
- GFLIP feature selection method is performing better for k-NN classifier with accuracy 81.25%;
- Fisher score based method is performing better for SVM-MLP classifier with 81.81% of accuracy,
- SVM-RFE is performing better for k-NN classifier with 84.37% of accuracy.
- The better performance of most of the supervised filter based classifier are associated with either SVM-MLP classifier or k-nearest neighbour classifier (k-NN) for k=5.

Therefore, from above observations it can be concluded that mRmR based feature selection method is performing better with 87.50% accuracy for SVM-MLP classifier followed by SVM-RFE and ReliefF methods with 84.37% accuracy of each for SVM-MLP and k-NN classifiers respectively. On further observation of other values of performance measures, in addition to accuracy of classifier, following conclusions are made:-

- Performance measures for mRmR for SVM-MLP classifier are:

Accuracy: 87.50%, Sensitivity: 85%, Specificity: 90%, AUC: 0.76, BCR: 0.88, and F-measure: 0.90.

- Performance measures for SVM-RFE for k-NN classifier are:

Accuracy: 84.37%, Sensitivity: 83.33%, Specificity: 87.50%, AUC: 0.6111, BCR: 0.8541, and F-measure: 0.8888

- Performance measures for ReliefF for SVM-MLP classifier are:

Accuracy: 84.37%, Sensitivity: 84.37%, Specificity: 60%, AUC: 0.71, BCR: 0.77, and F-measure: 0.89.

In view of above observations the order of supervised filter based feature selection methods in terms of their performance are as follows:

mRmR>SVM-RFE>ReliefF>Fisher score>GFLIP>FCBF.

Therefore, from above observations it can be concluded that mRmR feature selection is performing better in comparison to all other methods closely followed by SVM-RFE and ReliefF methods.

Further, Table 6 presents the performance comparison of various unsupervised filter based feature selection methods in consideration. Fig 3 shows comparison of unsupervised filter based feature selection methods in terms of classifier's maximum accuracy , sensitivity, and specificity values , and Fig 4 shows comparison of unsupervised filter based feature selection methods in terms of classifier's maximum AUC, BCR, and F-measures values.

From Table 6 and Figs 3-4 following observations are made:

- t-test based feature selection method is performing better for ANN classifier with accuracy 87.50% followed by SVM-Polynomial classifier with accuracy 81.81%;
- Bhattacharya rank feature based method is performing better for k-NN classifier with accuracy 81.81%;
- Wilcoxon rank feature based method is performing better for k-NN classifier with 84.37% accuracy;
- ROC rank feature is performing better for SVM-MLP classifier with accuracy 84.375%;
- Entropy rank feature based method is performing better for k-NN classifier with accuracy 75%;
- PCA based feature selection method is performing better for k-NN classifier with accuracy 81.81%;
- Laplacian Score based method is performing better for k-NN classifier with 81.25% accuracy.
- Most of the unsupervised feature selection methods are performing better for k-NN classifier. Some of them are performing better for ANN and SVM-MLP classifier.

Hence, from Table 6, it can be concluded that t-test method is associated with better performance of 87.50% for ANN classifier followed by Wilcoxon rank feature based method for k-NN classifier with 84.37% accuracy.

5. CONCLUSIONS

In this paper, an empirical analysis of various supervised and unsupervised filter based feature selection methods for various classifiers was presented for breast cancer classification from mammograms. For experimentation purposes 322 mammograms available in MIAS database were used. The total of 88 hybrid features were extracted from each mammogram in database after applying enhancement and modified fuzzy c-means based segmentation approach. The efficacy of various filter selection method were evaluated for top 50 features out of 88 hybrid features extracted from 322 mammograms in terms of accuracy, sensitivity, specificity, BCR, F-measure and AUC. From the obtained results, it can be concluded that mRmR based feature selection method is performing better with 87.50% accuracy for SVM-MLP classifier in comparison to other methods followed by SVM-RFE and ReliefF methods under supervised category. Under unsupervised category, t-test method is performing better with 87.50% accuracy for ANN classifier followed by Wilcoxon rank feature based method for k-NN classifier with 84.37% accuracy. In addition to evaluation of feature selection methods, it is also observed that in most of the cases k-NN and SVM-MLP classifiers are performing better. The overall performance of mRmR feature selection method is better in comparison to all filter based feature selection methods in consideration under both supervised and unsupervised categories.

6. REFERENCES

[1] Cancer facts and figures 2011. http://www.cancer.org/Cancer/BreastCancer /DetailedGuide/ breast-cancer-key-statistics.

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

[3] Subashini T.S. , Ramalingam V., Palanivel S. 2010. Automated assessment of breast tissue density in digital mammograms. Computer Vision and Image Understanding. 114, 33–43.

[4] Rangayyan, R.M., Ayres, F.J., Desautels, J.E.L 2007. A Review of Computer-Aided Diagnosis of Breast Cancer: Toward the Detection of Subtle Signs. Journal of the Franklin Institute. 312–348.

[5] Bozek J., Mustra M., Delac K., and Grgic M. 2009. A Survey of Image Processing Algorithms in Digital Mammography. Rec. Advances in Mult. Sig. Process. And Commn. (Grgic M. Eds.) SCI 231. Springer-Verlag Berlin Heidelberg. 631-656.

[6] Tang J., Rangayyan R., Xu J., Naqa I. El, Yang Y. 2009. Computer-aided detection and diagnosis of breast cancer with mammography. IEEE Transactions Recent

advances, Information Technology in Biomedicine, 13(2), 236–251.

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

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

[9] Guyon I. , and Elisseeff A. 2003. An introduction to variable and feature selection. Journal of Machine Learning Research, 3:1157–1182.

[10] Kohavi R. and John G.H. 1997. Wrapper for feature subset selection. Artificial Intelligence. 97(1-2), 273–324..

[11] Marko R. S., Igor K. 2003. Theoretical and empirical analysis of Relief and ReliefF. Machine Learning Journal. 53, 23-69.

[12] Peng H., Long F., and Ding C. 2005. Feature selection based on mutual information: criteria of max-dependency, max- relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence. 27(8), 1226-1238.

[13] Yu L. , and Liu H. 2003. Feature selection for high-dimensional data: A fast correlation-based filter solution. In proceedings of the ICML'03 conference. 856-863.

[14] Gilad-Bachrach R., Navot A., and Tishby N. 2004. Margin based feature selection - theory and algorithms. In Proceedings of 21st ICML conference. 337–344.

[15] Furey T.S., Cristianini N., Duffy N., Bednarski D.W., Schummer M., and Haussler D. 2000. Support Vector Machine Classification and Validation of Cancer Tissue Samples Using Microarray Expression Data. Bioinformatics. 16(10), 906-914.

[16] He, X., Cai, D., Niyogi, P. 2005. Laplacian score for feature selection. In: Advances in Neural Information Processing System, vol. 17. MIT press, Cambridge.

[17] Yang F. and Mao K.Z. 2011. Robust Feature Selection for Microarray Data Based on Multicriterion Fusion. IEEE/ACM Transact. On Comput. Biology and Bioinformatics. 8(4), 1080-1092.

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

[19] Jain, A.K. 2006. Fundamentals of Digital Image Processing. PHI, India.

[20] Cai, W., Chen, S. and Zhang, D. 2007. Fast and robust fuzzy c-means clustering algorithm incorporating local information for image segmentation. Pattern Recognition. 40(3), 825-838.

[21] 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. 13(3), 270–300.

[22] Haralick, R.M., Shanmugam, K. and Dinstein, I. 1973. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics. SMC-3 (6), 610-621.

[23] Gonzalez, R.C. 2003. Digital Image Processing. Prentice Hall, India.

[24] Huang, K. and Aviyente, S. 2008. Wavelet Feature Selection for Image Classification. IEEE Transactions on Image Processing, 17(9),1709-1720.

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

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

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

Table 5. Comparison of supervised filter based feature selection methods in terms of various classifier performance measures for the given mammogram data set

Feature Selection Methods	**Classifier for detection**		**Classifier Performance Measures**					
			Accuracy	Sensitivity	Specificity	AUC	BCR	F- Measure
Relief	Naïve Bayes		71.87	0.8636	0.400	0.6400	0.6318	0.8085
	KNN		81.25	1	0.400	0.5215	0.7000	0.8800
	DA		71.87	0.9090	0.300	0.2976	0.6045	0.8163
	ANN		81.25	0.6666	0.8275	0.5277	0.7471	0.4000
	SVM	Linear	81.25	0.86	0.70	0.40	0.78	0.86
		RBF	87.50	1	0	0.43	0.50	0.93
		Quadratic	65.62	0.66	0.62	0.65	0.64	0.74
		Polynomial	68.75	0.70	0.60	0.60	0.65	0.79
		MLP	**84.37**	**0.95**	**0.60**	**0.71**	**0.77**	**0.89**
mRmR	Naïve Bayes		76.92	0.9615	0	0.7016	0.4807	0.8695
	KNN		69.69	0.76190	0.5833	0.5873	0.6726	0.7619

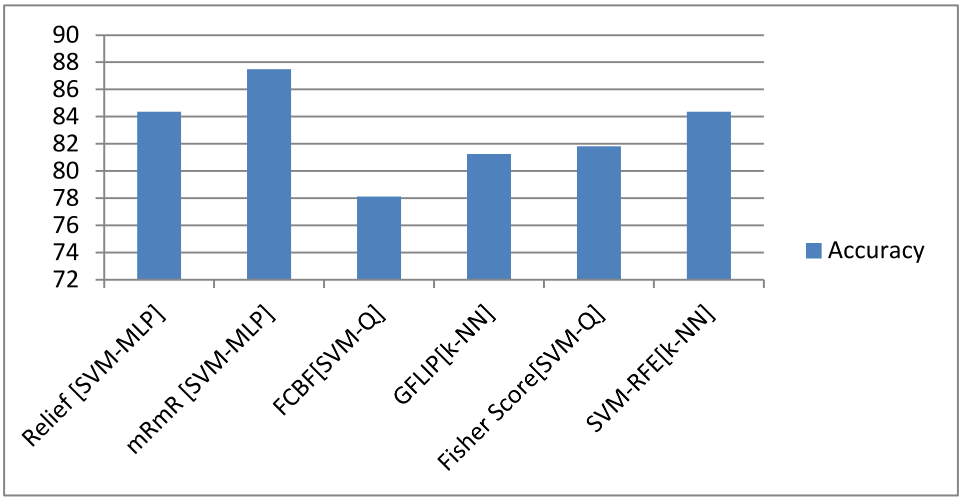
	DA		68.75	0.7619	0.5454	0.5108	0.6536	0.7619
	ANN		81.25	1	0.8064	0.6156	0.9032	0.2500
	SVM	Linear	78.12	0.92	0.28	0.58	0.60	0.86
		RBF	78.12	0.76	0.81	0.71	0.79	0.82
		Quadratic	66.66	0.69	0.69	0.56	0.78	0.81
		Polynomial	65.62	0.71	0.71	0.43	0.62	0.73
		MLP	**87.50**	**0.85**	**0.90**	**0.76**	**0.88**	**0.90**
FCBF	Naïve Bayes		61.53	0.700	0.4800	0.6437	0.5900	0.6913
	KNN		75	0.8500	0.5833	0.4742	0.7166	0.8095
	DA		65.62	0.70833	0.500	0.4458	0.60416	0.7555
	ANN		75.00	1	0.7419	0.4502	0.87096	0.2000
	SVM	Linear	68.75	0.77	0.50	0.35	0.63	0.77
		RBF	84.37	1	0	0.33	0.50	0.91
		Quadratic	**78.12**	**0.82**	**0.66**	**0.52**	**0.74**	**0.84**
		Polynomial	75	0.69	0.88	0.40	0.79	0.80
		MLP	75	0.86	0.44	0.41	0.65	0.83
GFLIP	Naïve Bayes		64.06	0.7500	0.4583	0.6705	0.6041	0.7228
	KNN		**81.25**	**0.8400**	**0.7143**	**0.3922**	**0.7771**	**0.8750**
	DA		68.75	0.8421	0.4615	0.5039	0.6518	0.7619
	ANN		78.12	NaN	0.7812	0.4087	NaN	NaN
	SVM	Linear	68.75	0.77	0.50	0.64	0.63	0.77
		RBF	84.3	0.93	0	0.38	0.46	0.91
		Quadratic	71.87	0.70	0.75	0.51	0.72	0.79
		Polynomial	68.75	0.70	0.62	0.56	0.66	0.77
		MLP	75.00	0.7917	0.6250	0.2684	0.7083	0.8261
Fisher Score	Naïve Bayes		64.61	0.7209	0.5000	0.733	0.6104	0.7294
	KNN		78.78	0.9130	0.500	0.6666	0.7065	0.8571
	DA		78.12	0.8888	0.200	0.6227	0.5444	0.8727
	ANN		78.12	0.7500	0.7857	0.5314	0.7678	0.4615
	SVM	Linear	68.75	0.71	0.63	0.40	0.67	0.75
		RBF	81.81	1	0	0.36	0.50	0.900
		Quadratic	**81.81**	**0.86**	**0.70**	**0.58**	**0.78**	**0.86**
		Polynomial	72.72	0.70	0.76	0.46	0.73	0.75
		MLP	75	0.800	0.57	0.85	0.68	0.83
SVM-RFE [SVM based Recursive Feature Elimination]	Naïve Bayes		68.25	0.7619	0.5000	0.7221	0.6309	0.7529
	KNN		**84.37**	**0.8333**	**0.8750**	**0.6111**	**0.8541**	**0.8888**
	DA		81.25	1	0	0.3506	0.500	0.8965
	ANN		84.37	0.6666	0.8620	0.3982	0.7643	0.4444
	SVM	Linear	78.12	0.80	0.75	0.71	0.77	0.82
		RBF	87.50	1	0	0.51	0.50	0.93
		Quadratic	71.87	0.72	0.700	0.43	0.71	0.78
		Polynomial	81.25	0.93	0.70	.47	0.81	0.81
		MLP	78.12	0.76	0.83	0.51	0.80	0.85

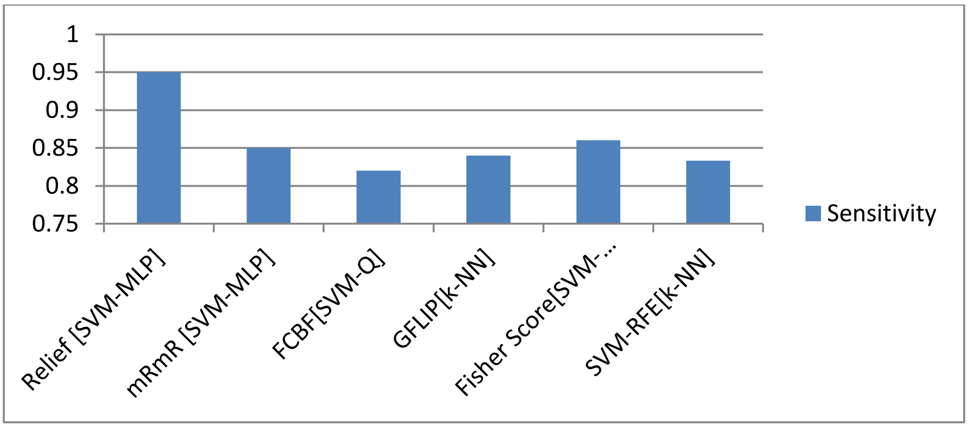
Table 6. Comparison of filter based unsupervised feature selection methods in terms of various classifier performance measures for the given mammogram data set

Feature Selection Methods	Classifier for detection	Classifier Performance Measures					
		Accuracy	Sensitivity	Specificity	AUC	BCR	F-Measure
t-test	Naïve Bayes	68.75	0.7560	0.5652	0.6405	0.6606	0.7560
	KNN	81.25	0.80000	0.8571	0.5333	0.8285	0.8695

	DA		75	0.95652	0.2222	0.5974	0.5893	0.8461
	ANN		**87.500**	**0.7500**	**0.9166**	**0.8750**	**0.8333**	**0.7500**
	SVM	Linear	71.875	0.7619	0.6363	0.5333	0.6991	0.78048
		RBF	90.90	1	0	0.6356	0.5000	0.9523
		Quadratic	69.69	0.8181	0.4545	0.2077	0.6363	0.7826
		Polynomial	81.81	0.8076	0.8571	0.5304	0.8324	0.8750
		MLP	72.72	0.8636	0.4000	0.5318	0.6075	0.8235
Bhattacharya distance	Naïve Bayes		67.69	0.7906	0.4545	0.6716	0.6226	0.7640
	KNN		**81.81**	**0.88000**	**0.62500**	**0.4106**	**0.7525**	**0.8800**
	DA		69.69	0.88461	0	0.6086	0.4423	0.8214
	ANN		84.37	NaN	0.84375	0.6153	NaN	NaN
	SVM	Linear	75.75	0.8000	0.6923	0.6272	0.7461	0.8000
		RBF	81.250	1	0	0.4541	0.5000	0.8965
		Quadratic	78.12	0.8571	0.6363	0.5367	0.7467	0.8372
		Polynomial	75	0.8333	0.5000	0.4534	0.6666	0.8333
		MLP	78.12	0.6818	1	0.6681	0.8409	0.8108
Wilcoxon test	Naïve Bayes		73.43	0.9347	0.2222	0.7067	0.5785	0.8349
	KNN		**84.37**	**0.9200**	**0.5714**	**0.314**	**0.7457**	**0.9019**
	DA		75	0.9130	0.3333	0.4615	0.6231	0.8400
	ANN		81.25	0.5000	0.8333	0.3885	0.6666	0.2500
	SVM	Linear	71.87	0.8181	0.5000	0.3816	0.6590	0.8000
		RBF	87.50	1	0	0.6450	0.5000	0.9333
		Quadratic	75	0.8461	0.3333	0.7125	0.5897	0.8461
		Polynomial	71.87	0.7857	0.2500	0.5748	0.5178	0.8301
		MLP	75	0.7058	0.8000	0.4000	0.7529	0.7500
ROC based	Naïve Bayes		66.15	0.7619	0.4782	0.6312	0.6200	0.7441
	KNN		81.25	0.9166	0.5000	0.7552	0.7083	0.8800
	DA		71.87	0.7619	0.6363	0.5020	0.6991	0.7804
	ANN		81.25	0.75000	0.8214	0.4636	0.7857	0.5000
	SVM	Linear	71.875	0.86363	0.4000	0.5500	0.6318	0.8085
		RBF	84.375	1	0	0.3882	0.5000	0.9152
		Quadratic	68.750	0.77777	0.5714	0.5833	0.6746	0.7368
		Polynomial	81.818	0.91666	0.5555	0.4448	0.7361	0.8800
		MLP	**84.375**	**0.90476**	**0.7272**	**0.4166**	**0.8160**	**0.8837**
Entropy based	Naïve Bayes		74.60	0.8604	0.4545	0.6740	0.6575	0.8043
	KNN		**75**	**0.8181**	**0.60000**	**0.5182**	**0.7090**	**0.8181**
	DA		71.87	0.9583	0	0.3645	0.4791	0.8363
	ANN		81.25	NaN	0.81250	0.4687	NaN	NaN
	SVM	Linear	75	0.8400	0.42857	0.3000	0.6342	0.8400
		RBF	84.84	1	0	0.3270	0.5000	0.9180
		Quadratic	78.12	0.8695	0.55555	0.4114	0.7125	0.8510
		Polynomial	78.12	0.7500	0.83333	0.4666	0.7916	0.8108
		MLP	75	0.7500	0.75000	0.6969	0.7500	0.7500
PCA	Naïve Bayes		67.18	0.7619	0.50000	0.6146	0.6309	0.7529
	KNN		**81.81**	**0.9545**	**0.54545**	**0.5845**	**0.7500**	**0.87500**
	DA		65.62	0.7142	0.54545	0.6666	0.6298	0.7317
	ANN		78.78	NaN	0.7878	0.4939	NaN	NaN

	SVM	Linear	75	0.761	0.7272	0.4227	0.7445	0.8000
		RBF	84.37	1	0	0.4882	0.5000	0.9152
		Quadratic	66.66	0.7000	0.6153	0.5436	0.6576	0.7179
		Polynomial	87.87	0.9090	0.8181	0.4125	0.8636	0.9090
		MLP	72.72	0.7500	0.6666	0.4272	0.7083	0.8000
Laplacian Score	Naïve Bayes		68.75	0.8222	0.3684	0.6068	0.5953	0.7872
	KNN		**81.25**	**0.8333**	**0.7500**	**0.5079**	**0.7916**	**0.8695**
	DA		65.62	0.6470	0.6666	0.5060	0.6568	0.6666
	ANN		81.25	0.5555	0.9130	0.3028	0.7342	0.62500
	SVM	Linear	71.87	0.8000	0.5833	0.4632	0.6916	0.78048
		RBF	87.50	1	0	0.2636	0.5000	0.93333
		Quadratic	65.62	0.7000	0.5833	0.5024	0.6416	0.71794
		Polynomial	71.87	0.7826	0.5555	0.6032	0.6690	0.80000
		MLP	78.12	0.7727	0.8000	0.4727	0.7863	0.82926





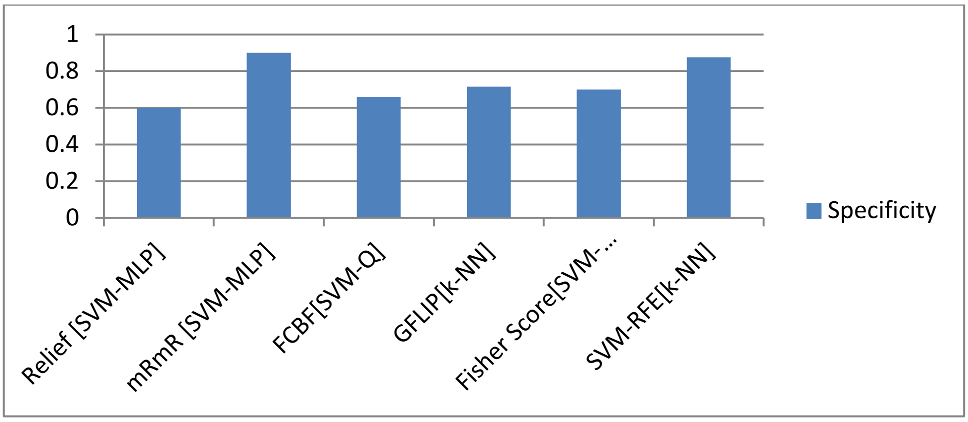
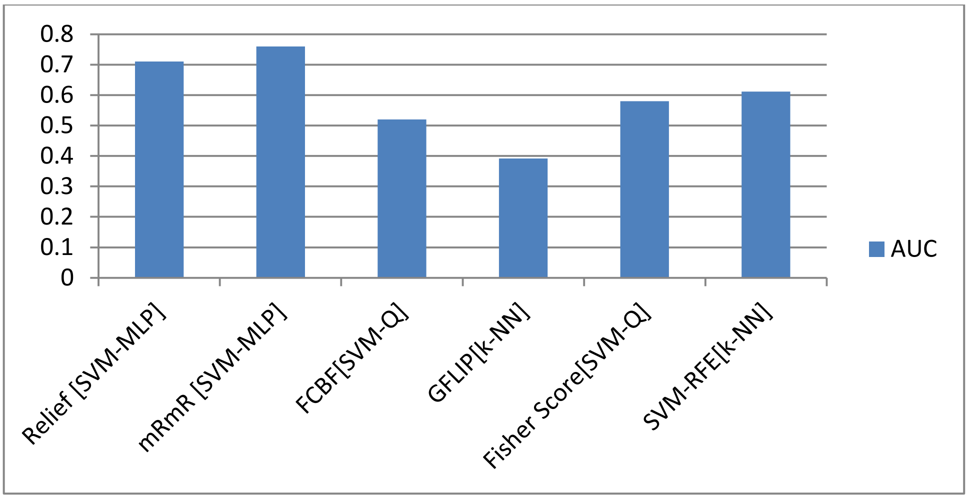
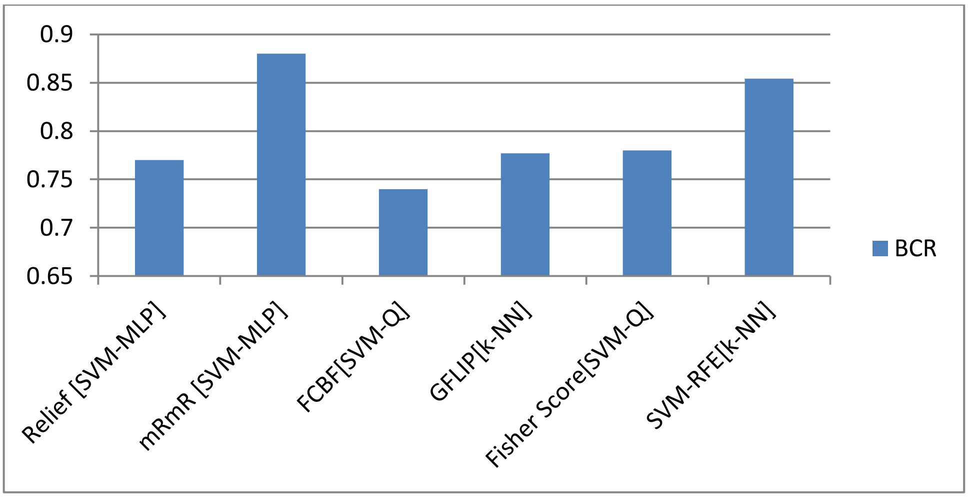


Fig 1: Comparison of supervised filter based feature selection methods in terms of classifier's maximum accuracy , sensitivity, and specificity values





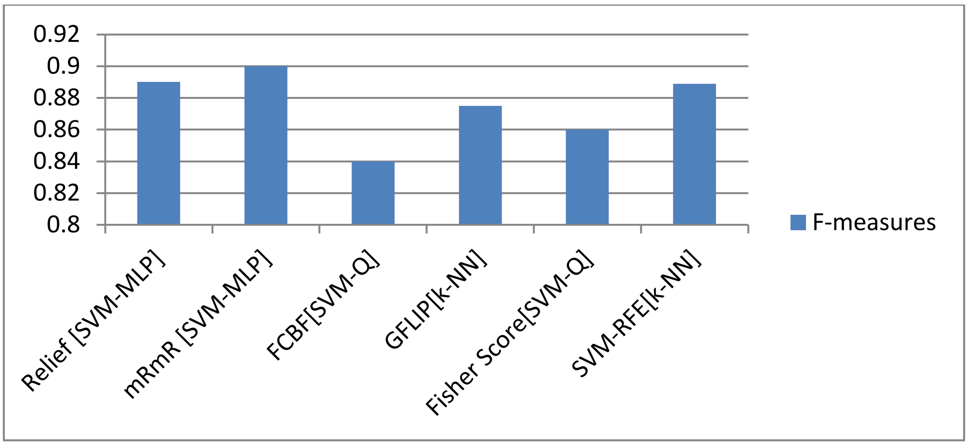
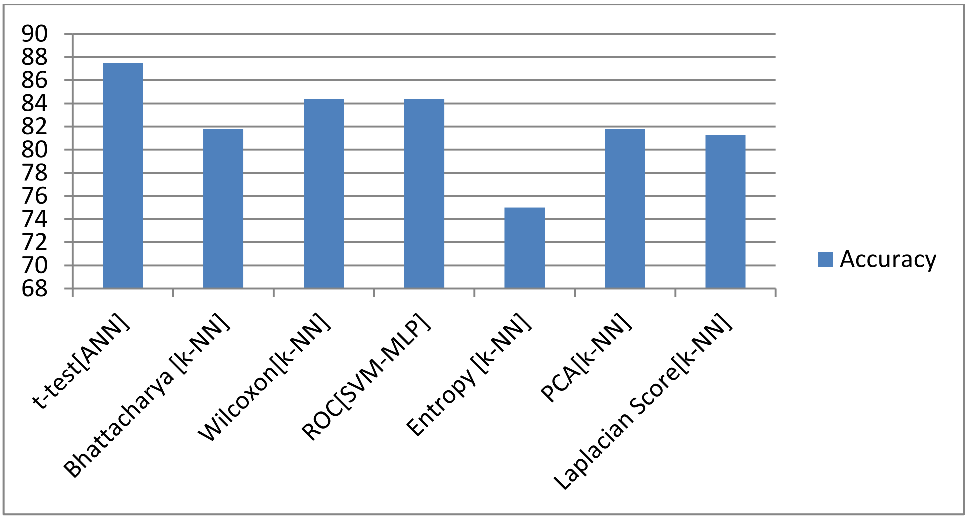
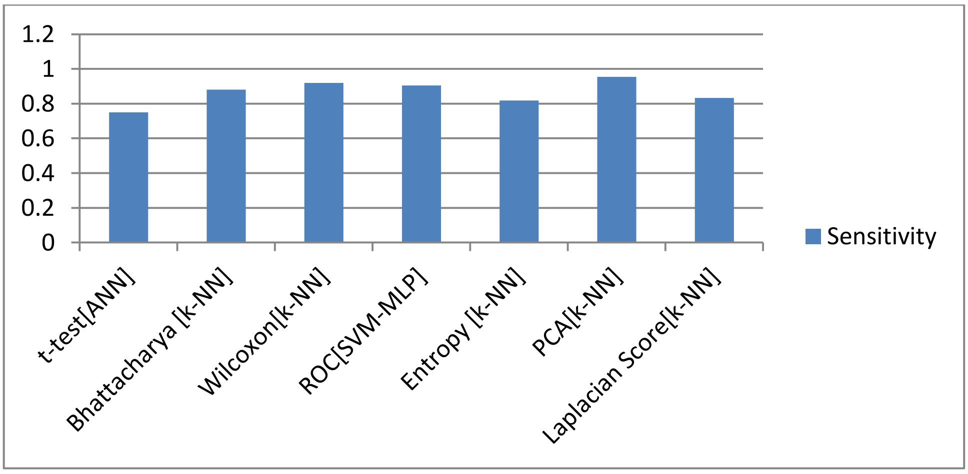


Fig 2: Comparison of supervised filter based feature selection methods in terms of classifier's maximum AUC, BCR, and F-measures values





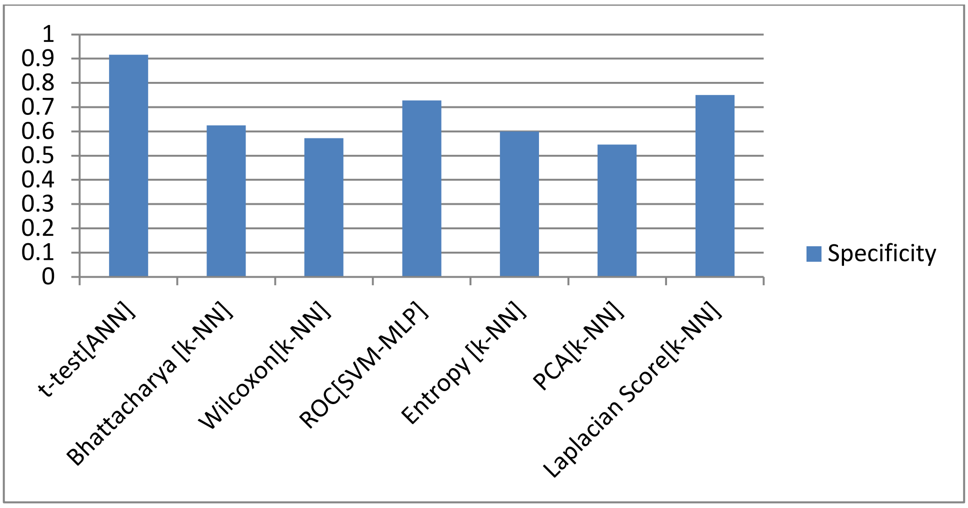
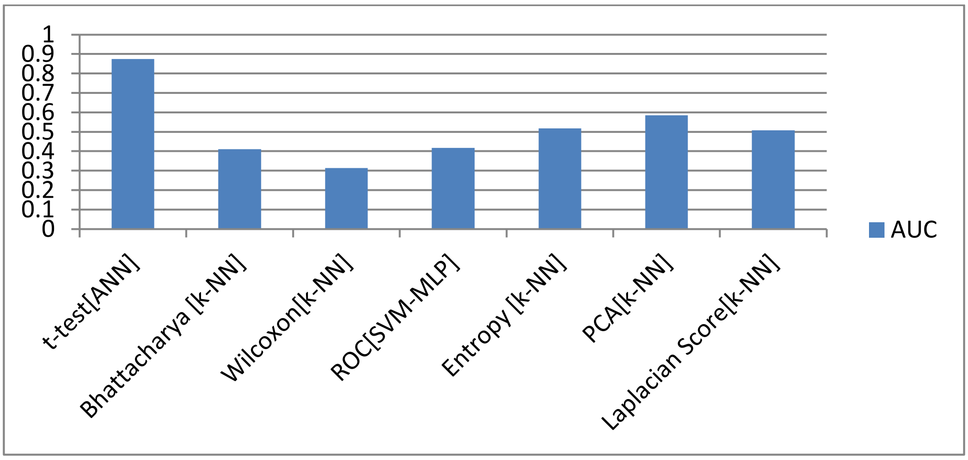
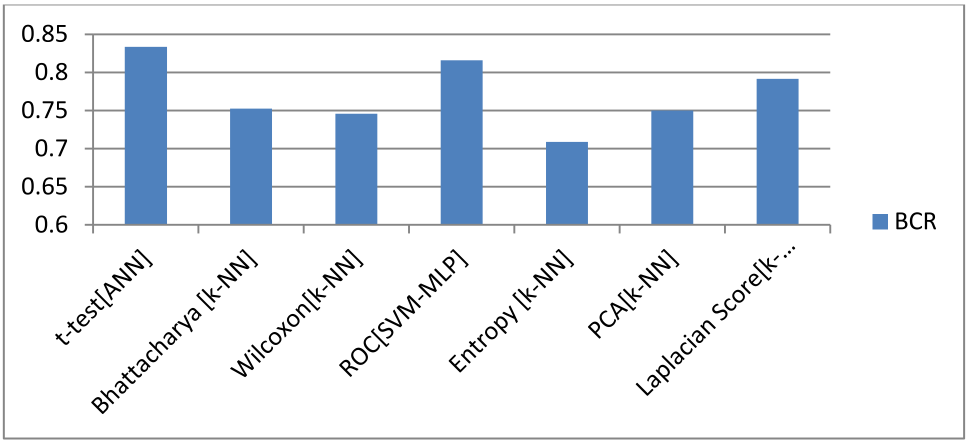


Fig 3: Comparison of unsupervised filter based feature selection methods in terms of classifier's maximum accuracy , sensitivity, and specificity values





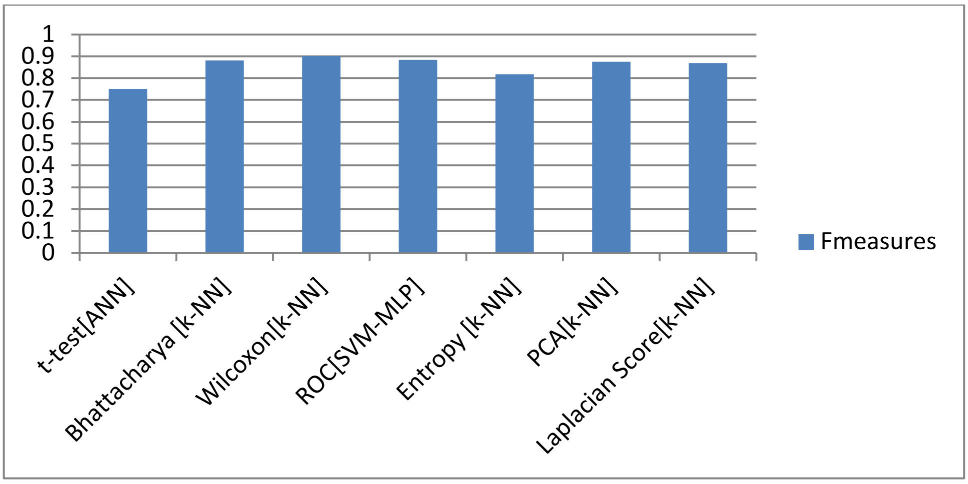


Fig 4: Comparison of unsupervised filter based feature selection methods in terms of classifier's maximum AUC, BCR, and F-measures values