# *Chapter 6: Comparative study of different classifiers for classification of histopathological images of breast cancer*

**Abstract:**

This chapter presents the computer aided diagnosis techniques for classification of histopathological images of breast cancer. The comparative study of different classifiers for histopathological image of breast cancer classification and detection using morphological features based on all cells present in the image are presented. The manual assessment of disease is time consuming and varies with the perception and the level of expertise of the pathologists. The judgment is based on the tissue structures, distribution of cells in tissue and the irregularities of cell shape and size. To overcome the limitation of manual diagnosis, a computer aided diagnosis based on the morphological features has been implemented for accurate and reliable detection of cancer. A dataset of 70 histopathological images of benign and malignant cells has been selected. The contrast limited adaptive histogram equalization (CLAHE) approach was used to highlight the details of the cellular structures. Trainable weka segmentation algorithm was used for the segmentation of background cells. It performs better in comparison to robust automatic threshold selection (RATS), Simple interactive object extraction (SIOX), mixture modeling and threshold segmentation methods. In other segmentation techniques cells are overlapping but in TWS, there is no overlapping has been visualized. The shape and morphology based features are proposed to extract the feature from the segmented images. The Multilayer perceptron (MLP), Logistic modal tree (LMT), naive Bayes, Sequential minimal optimization (SMO), Random Forest, Rotation forest, J-Rip, and PART classifiers were used for classification. The performance of Rotation forest was found better among others classifiers which have the accuracy of 85.7 % and with maximum BCR value 0.806. The proposed work aims at developing the technique that uses reliable quantitative measures for providing objective and reproducible information complementary to that of a pathologist.

**Keywords:** Image Processing, Classification, Fiji, Morphological, Feature, Cancer, Weka.

## 6.1 Introduction:

In this chapter we have done the comparative study of different classifiers for cancer detection and classification based on all cells in the image using morphological features. For the identification of benign and malignant cells the following CAD steps are follow which include pre-processing, segmentation, feature extraction and classification. Enhancement method is performed to improve the quality of the image and visible cells better. Different segmentation methods are used to select the best one to perform better segmentation of cells from the background of the image. Morphology and intensity based features of benign and malignant cells are used. Eight types of classifiers have been chosen to test their classification accuracy with the extracted features. The eight classifiers are as follows: (1) Multi-layer perceptron (MLP) (2) Logistic modal tree (LMT) (3) Rotation Forest, (4) Random forest, (5) Naïve Bayes, (6) Sequential minimal optimization (SMO) (7) J-Rip and (8) PART are utilized here to classify the features into two sets, such as benign and malignant and described as follows.

**Multi-layer perceptron (MLP):**

MLP is a classifier based on feed forward artificial neural network modal that uses back propagation to classify instances. It has much triumphant application in data classification. It consists of different layers having various nodes, which represents directed graph and every layer is fully connected with the further layer. The supervised learning process consists of input data y and target P, requires the objective function (Z, P) in order to evaluate the divergence of the predicted output values, Z=MLP(Y; K) from the observed data values P and employ that evaluation for the convergence towards an optimal set of weights k. Many MLP training algorithms used $\partial M/\partial K$ radiant information whether directly or indirectly (Silva *et al.*, 2008).

**Logistic modal tree (LMT):**

The LMT is a classification replica, which has an affiliated supervised learning algorithm that amalgamates logistic regression (LR) and decision tree learning. It is made of standard decision tree having logistic regression functions at the leave nodes, which is based on the concept of a modal tree.

The leave nodes contain two child nodes. One of the child nodes represents left branch and other represent right branch by threshold values. Feature value which is smaller than a threshold is sorted to left and greater than a threshold is sorted to right branch (Mahesh *et al.,* 2009).

**Random Forest:**

Random forest proposed by Breiman is one type of ensemble learning process for classification and regression. A random forest is a multiway classifier composed of some trees, and each tree grows using randomization. The leaf nodes of each tree are labeled by approximation of the posterior distribution over the classes of image. This test has been done to split the space of data to be classified. Every interior node that contains a test that best splits it. Classification of an image takes place by sending it down to every tree and after that aggregating the reached leaf distributions. Randomness can be inserted at two points during training and testing. This concept is used so that training process can be done by using different data subset. Randomness can be injected in selecting the node tests (Breiman *et al.,* 2001). Large scale sample sets are trained that is based on decomposition and iteration. These methods decrease accuracy.

**Rotation forest:**

Rotation forest is assembled with independent decision trees. Each tree is trained with complete information system with a rotated feature space. It uses hyperplanes parallel to the feature axes and a small rotation of the axes guide to diverse trees. Rodriguez *et al.,* 2006) done the comparative study and proved that rotation forest performs better than random forest, bagging, and AdaBoost. It is devised that rotation forest produces more accurate classifiers than AdaBoost which are also more diverse than bagging.

**Sequential minimal optimization (SMO):**

SMO was introduced by John Pitt in 1998 at Microsoft research to solve problems. It is used to solve the quadratic programming (QP) problem that appears during the training of support vector machines. SMO disintegrates the (QP) problem into sub problems, using Osuna's theorem which selects to resolve the smallest

feasible optimization problem at every step. The smallest feasible optimization problem for the standard SVM-QP problem involves two large range multipliers that must obey a linear equality constraint. It selects two lagrange multipliers jointly to optimize at every step and tries to find the optimal values for these multipliers. After that updates, the SVM reflects the new optimal values which solve lagrange multipliers analytically (Platt *et al.,* 1998).

**Naïve Bayes:**

Naïve Bayes classifiers are based on a probabilistic approach for classification hinged on Bufe's theorem with strong independence assumptions between the features. These classifiers are highly scalable. Naïve Bayes nearest neighbor classifier (NBNN) is a non-parametric approach for image classification introduced by Bioman (Timofte *et al.*, 2012).

**J-Rip:**

J-Rip is used to learn propositional rules by frequently developing rules and trimming them. Precursors are appended greedily until a termination condition is satisfied during the growth phase. After that antecedent is pruned in the upcoming phase on a pruning metric on one occasion, the rule set is generated. Optimization is required for the rules, which are evaluated by some criteria and deleted by their performance against those criteria on randomized data (Cohen *et al.,* 1995).

**PART:**

PART produces rules through frequently creating decision trees from data. The algorithm acquires a separate and conquers strategy in that. It abolishes instances covered by the ongoing rule set during processing. Essentially a rule is generated by constructing a pruned tree for the present set of instances; the leaf with the maximum coverage is converted into a rule (Witten *et al.,* 1998*).*

After classification, ranking of the features has been performed using relief- F algorithms. This has been done to observe which features play important role in classifying the benign and malignant.

## 6.2 Ranking of the features:

Recent studies have shown that histopathological features data is useful for differentiating between benign and malignant tissues (Demir *et al.*, 2005). Cancer classification using features data have the major challenge because of the following characteristics: (1) The number of features greatly exceeds the number of instances (tissue samples) and (2) Most features are not related to the given cancer classification problem. The selecting a small set of revealing features improve classification accuracy. Feature ranking used in features selection method. In this features ranking methods, each feature is evaluated separately and assigned a score reflecting its correlation with the class according to certain criteria. Features are then ranked by their scores and the top ranked ones are selected. Feature subset selection is a method for reducing the attribute of a feature set. It is identifying a subset of features by removing irrelevant or redundant features. A good feature set contains a highly relevant feature which helps to improve the efficiency of the classification algorithms and accurate classification of the cells. Relief-F is a feature selection algorithm for random selection of instances for feature weight calculation. The Relief-F algorithm adopts the random selection of instances for weight estimation. In the machine learning field, one of the most successful individual feature filtering algorithms is the Relief-F algorithm. This algorithm has been successfully used in many large subset feature selection tasks (Kononenko, 1994).

## 6.3 Relief-F algorithms:

To perform the Relief experiment we used Weka a powerful open-source Java based machine learning tool that can be run on any computer that has a Java run time environment installed. The Relief-F algorithm was first described by Kira and Rendell, (1992) as a simple, fast, and effective approach to attribute weighing. The output of the Relief algorithm is a weight between −1 and 1 for each attribute, with more positive weights indicating more predictive attributes. The weight of an attribute is updated iteratively as follows. A sample choose from the data, and the nearest neighbouring sample that belongs to the same class (nearest hit) and the nearest neighbouring sample that belongs to the opposite class (nearest miss) are identified.

A change in attribute value accompanied by a change in class leads up to the weighting of the attribute based on the intuition that the attribute change could be responsible for the class change. On the other hand, a change in attribute value accompanied by no change in class leads to down-weighting of the attribute based on the observation that the attribute change had no effect on the class. This procedure of updating the weight of the attribute is performed for a random set of samples in the data or for every sample in the data. The weight updates are then averaged so that the final weight is in the range (−1, 1). The attribute weight estimated by Relief-F has a probabilistic interpretation. It is proportional to the difference between two conditional probabilities, namely, the probability of the attribute's value being differently conditioned on the given nearest miss and nearest hit respectively (Robnik-Sikonja and Kononenko, (2003).

## 6.4 Materials and Methods:

Breast cancer cellular datasets of benign and malignant cells used in present work has been obtained from www.bioimage.ucsb.edu. Images were captured from 5 microarrays (ytma 10,12,23,49 and 55) with 6, 6, 6, 34 and 6) images captured per array, respectively from the Yale Tissue Microarray Facility placed in (Centre for Bio-image Informatics, University of California, Santabarbara (UCSB).

### 6.4.1 Data set preparation:

The study consists of a dataset using 70 breast cancer histopathology images (35 benign and 35 malignant). Enhancement, segmentation and feature extraction based image analysis has been performed using software Fiji (www.fiji.net).. Morphology and intensity based 16 features have been acquired for classification of benign and malignant cells. Further, feature extraction based output result of 35 benign and 35 malignant images using Fiji software are provided in appendix-3(a) for 35 benign breast images and appendix-3(b) for 35 malignant breast images from (F1-F16) of this present thesis. After features extraction, a dataset of order 70 x16 in .arff (attribute relation file format) format is prepared. Within this 70 instances and 16 attributes are available. Table 6.1 represents .arff format based on sixteen (16) features and description of features are provided in appendix- 4 of this present thesis.

The sixteen (16) features mainly includes (F1) to (F16) are as follows: count (F1), total area (F2), average size (F3), area fraction (F4), perimeter (F5), major axis length (F6), minor axis length (F7), angle (F8), circularity (F9), solidity (F10), feret (F11), feret X (F12), feret Y (F13), feret angle (F14), minimum feret (F15) and integrated density (F16).

**Table 6.1: Represent the attribute relation file format (.arff) file with 16 attributes.**

| | |
|---|---|
| @RELATION | Breast Cancer |
| @attribute Count | Numeric |
| @attribute Total Area | Numeric |
| @attribute Average Size | Numeric |
| @attribute Area Fraction | Numeric |
| @attribute Perimeter | Numeric |
| @attribute Major | Numeric |
| @attribute Minor | Numeric |
| @attribute Angle | Numeric |
| @attribute Circularity | Numeric |
| @attribute Solidity | Numeric |
| @attribute Feret | Numeric |
| @attribute Feret X | Numeric |
| @attribute Feret Y | Numeric |
| @attribute Feret Angle | Numeric |
| @attribute Min feret | Numeric |
| @attribute Int Den | Numeric |
| @attribute Class {Yes, No} | |
| @Data | |

Classification purpose experiments have been carried out using Weka data mining tool. 10 fold cross-validation approaches are used for training and testing of samples. Selected features are introduced into eight classifiers. Figure 6.1 depicts the flowchart for the present work in which describes basic steps involved in the cells morphology based image analysis.

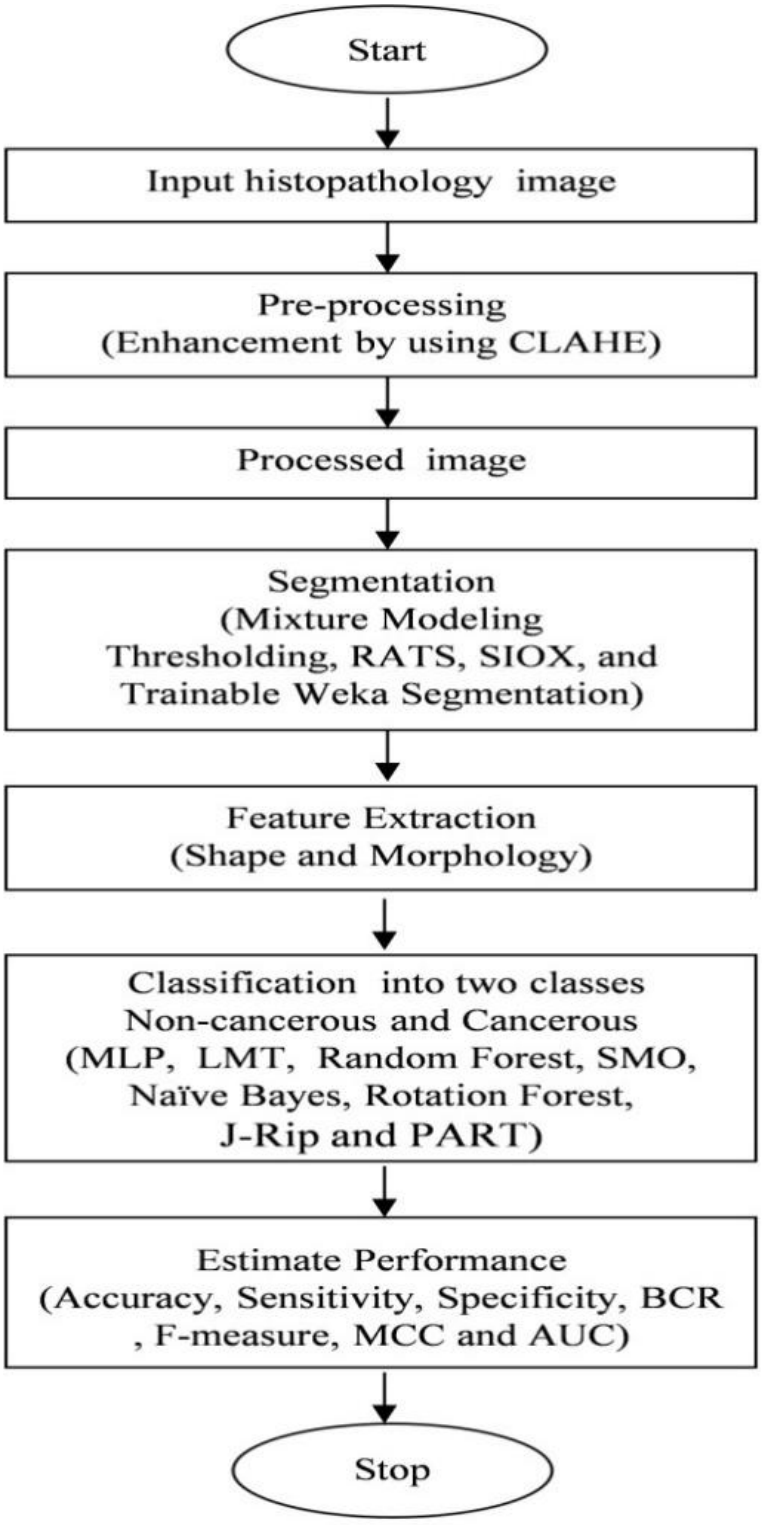


**Figure 6.1: Schematic flowchart of the proposed method.**

### 6.4.2 Preprocessing:

In histopathology images, the blurriness, artifacts, weak boundary detection and overlapping problem occurred due to uneven staining of the slide as a result of human error. To eradicate these types of irregularities or uneven staining, the CLAHE method is proposed. CLAHE algorithm improves the image contrast by improving the local contrast present in an image and also by enhancing the weak boundary edges in each pixel of an image through limited amplification (Zuiderveld *et al.*, 1994). Digital image processing techniques interpret the result in a much better way than conventional methods. So it is well suited for features enhancement of histopathology images. Figure 6.2 represents CLAHE method has been used for pre-processing of images.

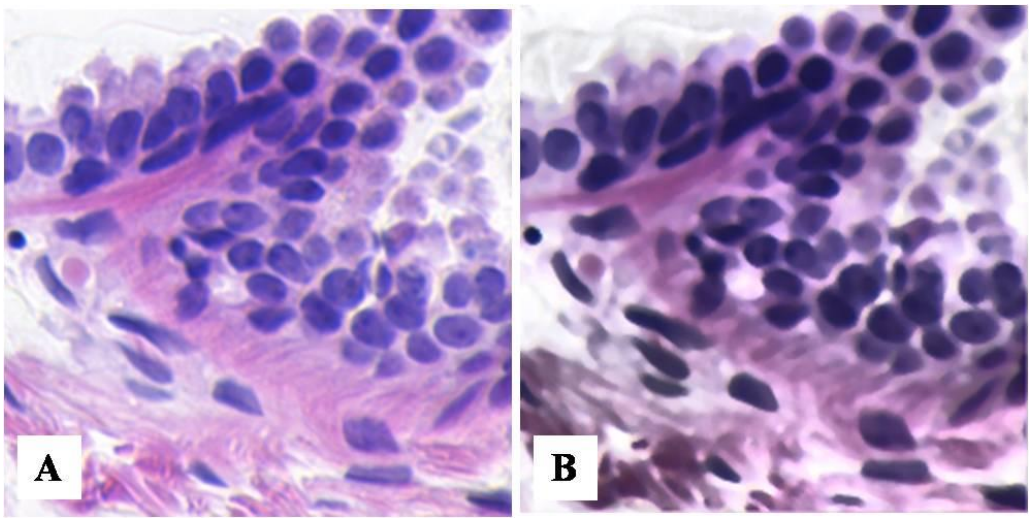


**Figure 6.2: Enhancement method.**
(A) Original benign breast image. (B) Enhanced image using CLAHE method.

### 6.4.3 Segmentation:

In digital pathology, segmentation of histopathology sections is a ubiquitous requirement due to the large variability of histopathology tissue. Further machine learning techniques play a vital role in delivering superior performance over standard image processing methods. During image analysis, the segmentation process is an essential domain. It is used to locate objects and boundaries in an image (Sharma *et al.*, 2009).

The proposed method, pre-processing steps involve, removing noise and enhancing the contrast for segmentation purpose. The basic purpose of segmentation is to extract the important features from the image and perceive the information. Selection of appropriate segmentation methods depends on the type of features that has to be maintained for detection. Segmentation methods like Mixture Modeling Thresholding (MMT), Simple Interactive Object Extraction (SIOX), Robust Automatic Threshold Selection (RATS), and Trainable Weka Segmentation (TWS) has proposed from Fiji open access free software for image analysis (Schindelin *et al.,* 2012).

Mixture modeling algorithm uses gaussian model to separate the histogram of an image into two Gaussian classes based on average, standard deviation and thresholding (Huang *et al.,* 2008). SIOX is a method used for extracting foreground information from a colored (RGB) image (Friedland *et al.,* 2005). RATS measure the threshold map of an image based on pixels value and the corresponding gradients value (Wilkinson. *et al.,* 1998). TWS is a pixel-based segmentation method which combines machine learning algorithms with a selected set of image features (Arganda *et al.,* 2014).

The performance of various segmentations is quantified regarding the global consistency error (GCE), a variation of information (VI) and probabilistic rand index (PRI) of the segmented image with the ground truth image. The brief description of Global Consistency Error (GCE), Variation of Information (VI) and Probabilistic Rand Index (PRI) performance measures are already described in **chapter-3** of this present thesis.

GCE and VI should be low, whereas PRI should be high for a better-segmented cell in the image. The MMT, SIOX, and RATS method have high GCE and VI whereas low PRI in comparison to TWS, which shows an edge of proposed TWS method over conventional methods. Figure 6.3 (F) for benign cells and figure 6.4 (F) for malignant cells depicts TWS gives a better result because TWS uses random forest machine learning algorithm for image segmentation. There is no overlapping in the cells and shows cells separated well from each other. This is providing the most accurate shape of the cells as compared to other methods.

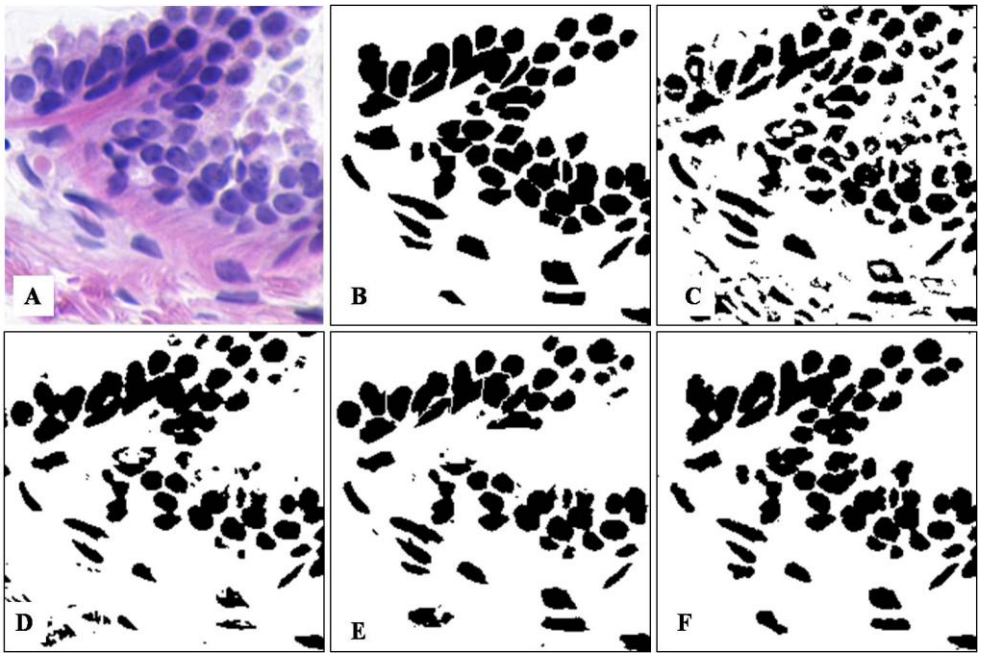


**Figure 6.3: Segmentation of benign cells from histopathology images by using different methods.**

A: Original image. B: Ground truth image. C: Mixture Modeling Thresholding (MMT). D: Simple Interactive Object Extraction (SIOX). E: Robust Automatic Threshold Selection (RATS). F: Trainable Weka Segmentation (TWS).

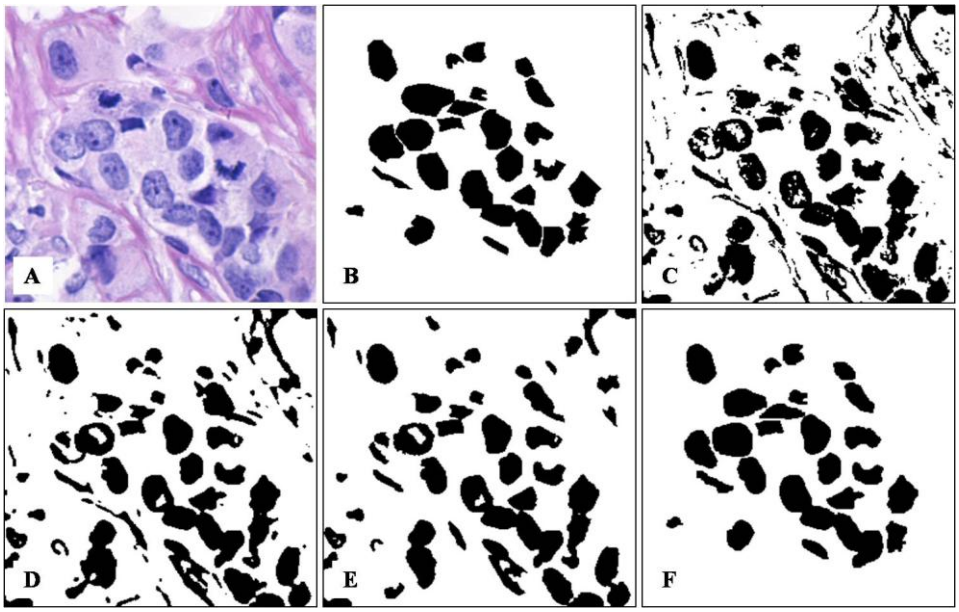


**Figure 6.4: Segmentation of malignant cells from histopathology images by using different methods**

A: Original image. B: Ground truth image. C: Mixture Modeling Thresholding (MMT). D: Simple Interactive Object Extraction (SIOX). E: Robust Automatic Threshold Selection (RATS). F: Trainable Weka Segmentation (TWS).

The ROI of the segmented histopathology image is compared to ground truth images for the quantitative assessment of different segmentation methods by GCE, VI, and PRI, for 25 sample images as acquired from histopathology dataset. Table 6.2 and figure 6.5 for PRI, figure 6.6 for GCE, figure 6.7 for VI represents TWS is associated with the lower value of GCE, VI and higher value of PRI in comparison to other better performing methods regarding all parameters. Henceforth, TWS is chosen as the segmentation method in the proposed work for cancer detection from histopathology images.

**Table 6.2: Quantitative comparison of segmentation methods on the basis of average values of 25 images**

| Segmentation Methods | PRI (Higher better) | GCE (Lower better) | VI (Lower better) |
|---|---|---|---|
| **M MT** | 0.95038 | 0.028408 | 0.303852 |
| **SIOX** | 0.9734 | 0.01608 | 0.209856 |
| **RATS** | 0.975016 | 0.015312 | 0.201652 |
| **TWS** | 0.976124 | 0.013844 | 0.19144 |

PRI- Probabilistic Rand Index, GCE- Global Consistency Error, VI- Variation of Information

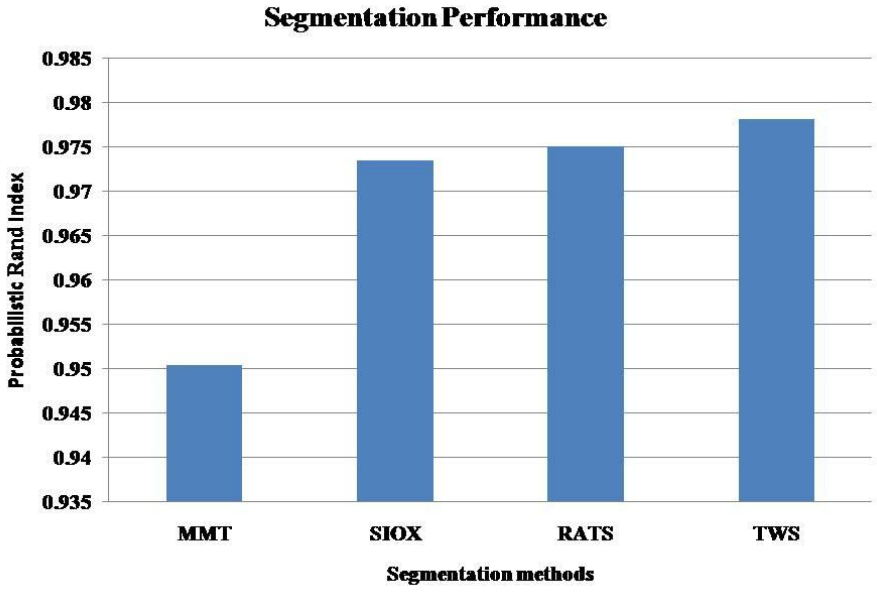


**Figure 6.5: Comparison of segmentation methods on the basis of average values of PRI for 25 sample images.**

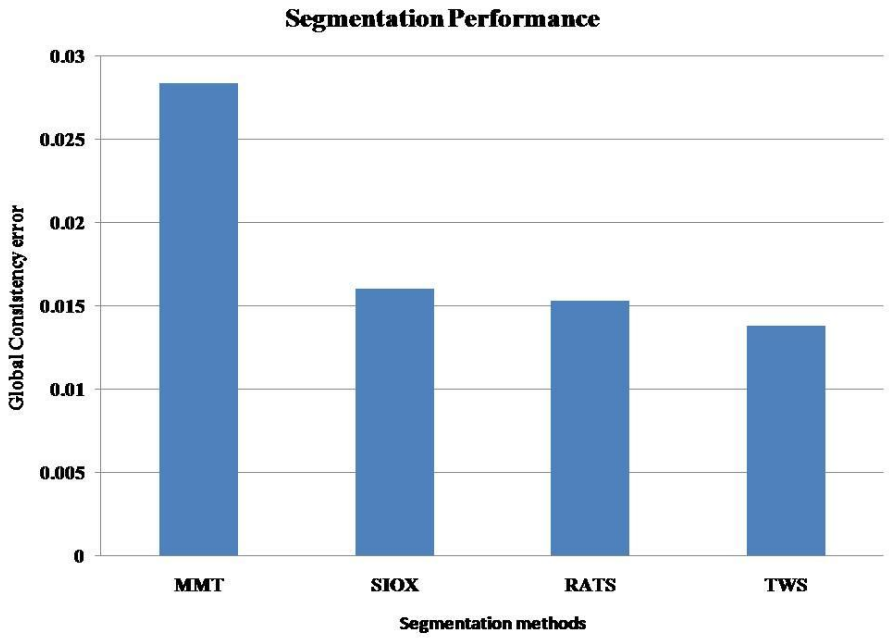


**Figure 6.6: Comparison of segmentation methods on the basis of average values of GCE for 25 sample images.**

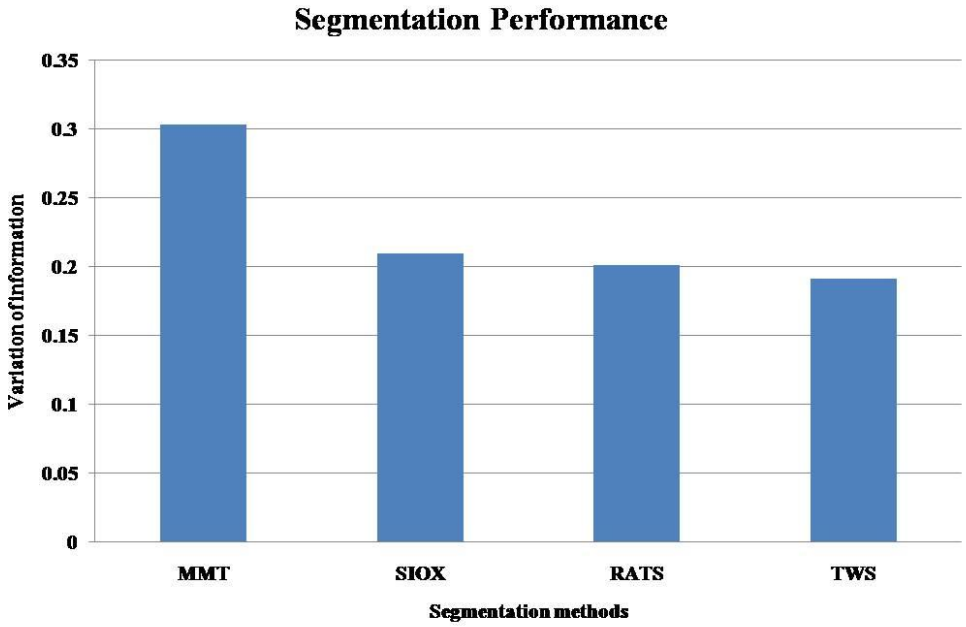


**Figure 6.7: Comparison of segmentation methods on the basis of average values of VI for 25 sample images.**

### 6.4.4 Feature Extraction:

Image morphology is a very powerful tool for analyzing the shapes of the objects and to extract the image features, which are necessary for object recognition (Zhao *et al.*, 2015). The most significant portion of this work is the computation of features. Morphological and shape based features have been extracted after segmentation of image for further classification purpose. These features provide information regarding the size and shape of cells (Anuranjeeta *et al.*, 2016). Figure 6.3 (F) and 6.4 (F) depicts TWS method is considered for features extraction from the segmented cells of the images.

Total 16 features (F1) to (F16) have been used in this chapter. The quantification of these features helps to differentiate the malignant cells from benign cells. The features used in this thesis as described from (F1) to (F16) are as follows: count (F1), total area (F2), average size (F3), area fraction (F4), perimeter (F5), major axis length (F6), minor axis length (F7), angle (F8), circularity (F9), solidity (F10), feret (F11), feret X (F12), feret Y (F13), feret angle (F14), minimum feret (F15) and integrated density (F16). The features perimeter (F5), major axis length (F6), minor axis length (F7), circularity (F9) and solidity (F10) are already discussed in the chapter - 4, page no-104-105 from equations 4.2 to 4.6. The descriptions of rest of the morphological features are as follow:

**(F1). Count:** No of cells present in segmented ROI of images.

**(F2). Total Area:** The sum of the area of individual cells in a particular segmented image.

**(F3). Average Size:** The total area of the cells presents in the segmented image divided by no of cells present in that image.

**(F4). Area fraction:** For thresholded images is the percentage of pixels in the image or selection that have been highlighted in red using Image. For nonthresholded images is the percentage of non-zero pixels.

**(F8). Angle:** (0-180 degrees) is the angle between the primary axis and a line parallel to the x-axis of the image.

**(F11). Feret:** It is defined as the longest distance between any two points along the selection boundary.

**(F12). Feret X:** Feret X is the starting coordinates of the Feret's diameter along the x-axis.

**(F13). Feret Y:** Feret Y is the starting coordinates of the Feret's diameter along the y-axis

**(F14). Feret Angle:** 0-180 degrees is the angle between the Feret's diameter and a line parallel to the x-axis of the image.

**(F15). Min Feret:** It is the minimum caliper diameter.

**(F16). Integrated density:** Integrated density is known as the sum of the values of the pixels in the selected part of the image.

### 6.4.5 Classification:

Classifications of benign and malignant cells are performed based on the extracted features. Factors such as staining, artifact, noise, and blurriness cause variation in the image and results in misclassification. Hence, a good classifier should be able to overcome these flaws (Spanhol *et al.,* 2015).

Moreover, the choice of classifier must be made by fast computation and it must be proficient enough to meet good classification. Supervised and unsupervised machine learning approaches have been used on the dataset of benign and malignant histopathology images for classification.

**Data Spreading in Weka experimental editor:**

Classification part has been carried out using the Weka (Waikato Environment for Knowledge Analysis) data mining tool. In this 70 instances and 16 attributes are available. Based on 16 features classification of cells has been done to classify benign and malignant.

These features in a dataset of order 70x16 .arff (attribute relation file format) format are loaded into the Weka toolbox for classification purpose. Figure 6.8 depicts the difference in different feature based on benign and malignant cells in weka data mining tool. After loading the features into weka in the .arff file format. The blue and red colored zone indicates benign and malignant cells features respectively.

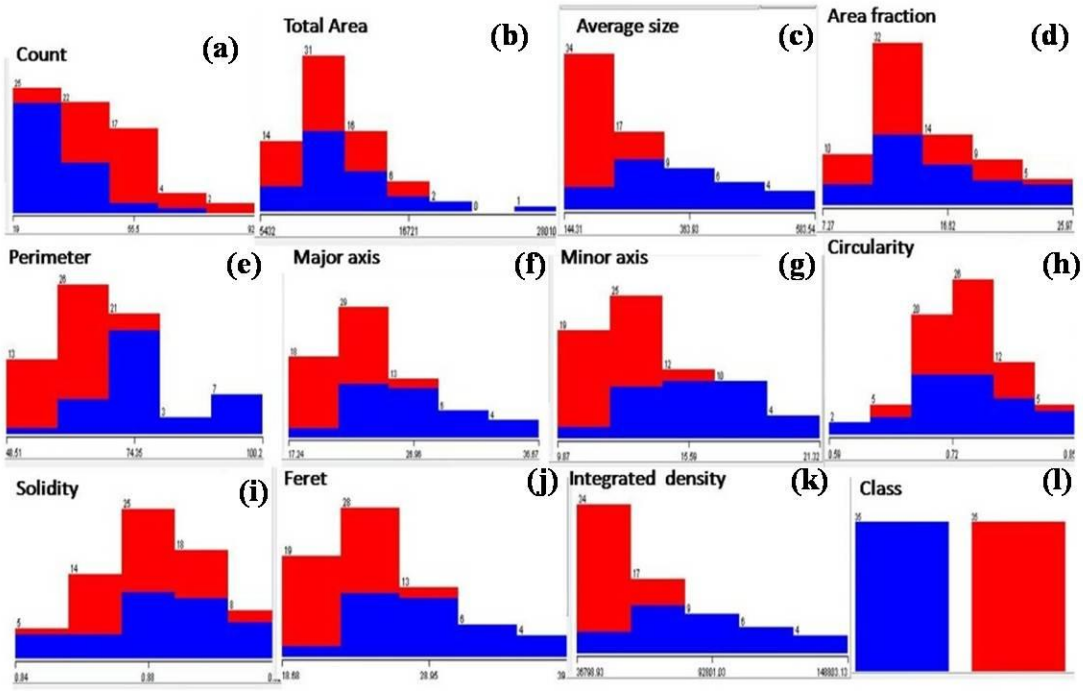


**Figure: 6.8: Features based difference of the benign (blue color) and malignant (red color) cells are observe while loaded in the weka data mining software.**
(a) (F1) count, (b) (F2) total area, (c) (F3) average size, (d) (F4) area fraction, (e) (F5) perimeter, (f) (F6) major axis length, (g) (F7) minor axis length, (h) (F9) circularity, (i) (F10) solidity, (j) (F11) feret, (k) (F16) integrated density, (l) class (two class benign and malignant).

For classification, selected features get feed into eight types of classifiers are as follow: **(1)** Multi-layer perceptron (MLP) **(2)** Logistic modal tree (LMT) **(3)** Rotation Forest, **(4)** Random forest, **(5)** Naïve Bayes, **(6)** Sequential minimal optimization (SMO) **(7)** J-Rip and **(8)** PART are utilized here to classify the features into two sets, such as benign and malignant.

**Measures of performance evaluation of classifier:**

Performance evaluation of each classifier is considered using confusion matrix (2 × 2) of size. The value of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) is calculated. Further, the definition of TP, TN, FP and FN and performance parameters of classifier like accuracy, sensitivity, specificity, balanced classification rate (BCR), F-measure (F-m), Matthews' correlation coefficient (MCC) and area under the curve (AUC) are defined to assess the success of the diagnostic system has been discussed in the **chapter-5** of this thesis.

## 6.5 Results and discussion:

The proposed methodologies are implemented with image analysis software Fiji (www.fiji.net) for enhancement, segmentation and feature extraction on the dataset of digitized at 40x, the magnification on PC with 3.4 GHz Intel Core i7 processor, 2GB RAM, and Windows 8.1 platform. For experimentation purposes, a total of seventy (70) histopathological images have been used. The dataset includes benign and malignant images. The given methodology for diagnosis of cancer from histopathology images consists of image enhancement, segmentation, feature extraction, and classification. The CLAHE method has been used for enhancement of histopathology images as it has shown better results. It highlights the region of interests in the images as tested through experimentation. The original image has been processed through following two pre-processing steps **(1)** contrast enhancement, **(2)** bilateral filtering to remove the artifact, blurriness that has been introduced during the staining process and to produce a better contrast image of good quality as shown in figure 6.2. The segmentation has been done by following methods MMT, SIOX, RATS and TWS, afterward, their results have been compared. TWS performs better in comparison to other methods as shown in figure 6.3 (F) for benign cells and figures 6.4 (F) for malignant cells. In other segmentation techniques cells are overlapping but in TWS, no overlapping has been visualized.

**Table 6.3: Comparative performances of various classifiers.**

| **Classifier** | **Accuracy** | **Sensitivity** | **Specificity** | **BCR** | **F- m** | **MCC** | **AUC** |
|---|---|---|---|---|---|---|---|
| MLP | 0.800 | 0.829 | 0.771 | 0.701 | 0.794 | 0.601 | 0.892 |
| LMT | 0.829 | 0.914 | 0.743 | 0.710 | 0.813 | 0.667 | 0.920 |
| Random Forest | 0.800 | 0.829 | 0.771 | 0.701 | 0.794 | 0.601 | 0.886 |
| **Rotation Forest** | **0.857** | **0.829** | **0.886** | **0.806** | **0.861** | **0.715** | **0.884** |
| Naïve Bayes | 0.829 | 0.857 | 0.800 | 0.740 | 0.824 | 0.658 | 0.855 |
| SMO | 0.857 | 0.914 | 0.800 | 0.764 | 0.848 | 0.719 | 0.857 |
| J Rip | 0.829 | 0.857 | 0.800 | 0.740 | 0.824 | 0.658 | 0.821 |
| PART | 0.771 | 0.771 | 0.771 | 0.676 | 0.771 | 0.543 | 0.749 |

BCR- Balanced Classification Rate, F-m- F-measure, MCC- Matthews's Correlation Coefficient and AUC-Area under the Curve

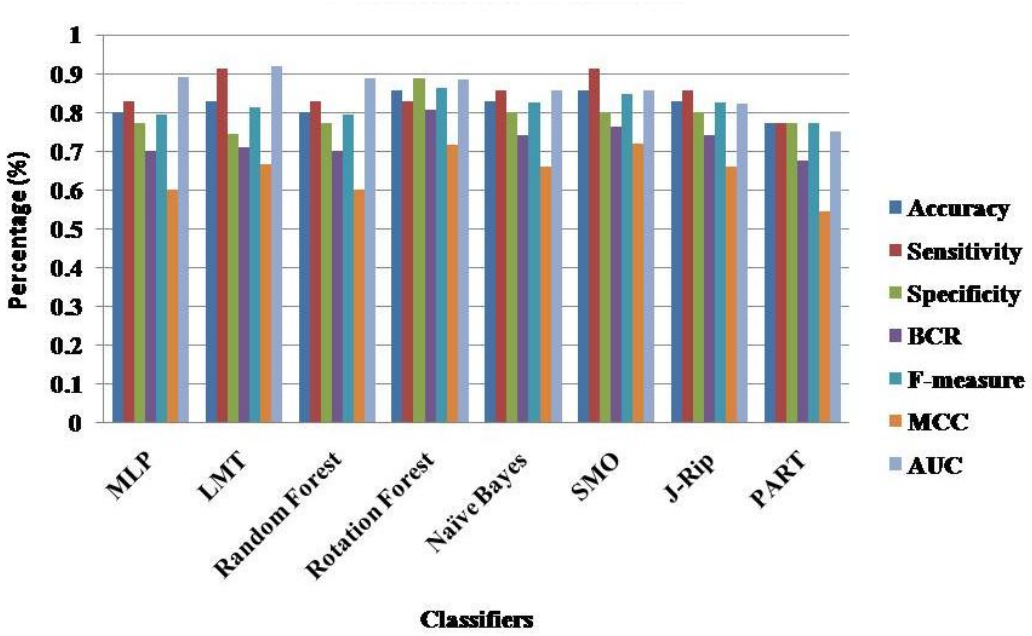


**Figure 6.9: Graph for comparative performances of various classifiers.**

In feature extraction phase, morphology based features as shown in F1 to F16 have been extracted from the segmented images. Finally, a 2D matrix of order (70×16) feature has formed using all the feature sets, where 70 breast histopathological images in the dataset and 16 total numbers of features has been extracted and further, these features has been used for classification. The experiment has been performed using 10-fold cross validation approach. Table 6.3 and figure 6.9 represent the proposed framework for different histopathology images containing benign and malignant features of cells has been tested using eight popular classifiers like **(1)** MLP, **(2)** LMT, **(3)** Random forest, **(4)** Rotation forest, **(5)** SMO, **(6)** Naïve Bayes, **(7)** J-Rip and **(8)** PART.

Among all these classification methods rotation forest differentiated better between benign and malignant cells with the accuracy of 85.7% and with maximum Balanced Classification Rate (BCR) value 0.806. The superiority of rotation forest measure lies in the application of rotation matrix, created by linear transformed subsets.

Ranks of all the features have obtained in the features vector by applying Relief-F algorithms (Kira *et al.*, 1992; Wang and Makedon *et al.*, 2004) in weka 3.8. Relief-F is to draw instances at random, compute their nearest neighbours, and change a feature weighting vector to give more weight to features that differentiate the instance from neighbours of different classes. In particular, it tries to get a better estimate of the following probability to allocate as the weight for each feature f. The pseudo code for Relief is shown below.

$w_f = P$ (different value of f | different class) – $P$ (different value of f | same class)

This approach has shown good performance in various domains (Robnik-Sikonja and Kononenko, 2003).Table 6.4 and figure 6.10 depicts the ranks of the features of cells have been investigated.

**Table 6.4: Ranking of morphological features.**

| **Feature Rank** | **Attributes Name** | **Maximal Relevance Factor** |
|---|---|---|
| 1 | Minor axis length | 0.12016 |
| 2 | Average Size | 0.11966 |
| 3 | Integrated density | 0.11966 |
| 4 | Min feret | 0.11798 |
| 5 | Perimeter | 0.11634 |
| 6 | Major axis length | 0.08956 |
| 7 | Feret | 0.08578 |
| 8 | Count | 0.08147 |
| 9 | Solidity | 0.0109 |
| 10 | Total Area | 0.0066 |
| 11 | Feret X | 0.00578 |
| 12 | Area fraction | 0.00562 |
| 13 | Circularity | 0.00419 |
| 14 | Feret Angle | 0.00164 |
| 15 | Angle | 0.00155 |
| 16 | Feret Y | 0.00132 |

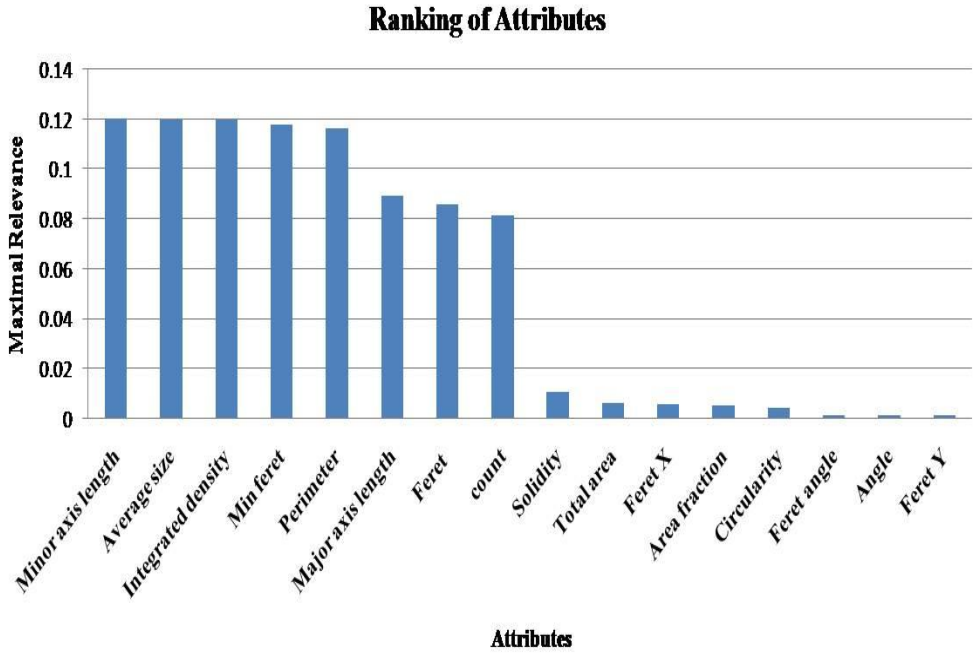


**Figure 6.10: Graph for ranking of maximal relevance factor.**

Maximal relevance factor has been derived for obtaining important features by Relief-F. Ranking of the Maximal relevance factor gives appropriate results than taking a large number of features.

## 6.6 Conclusion:

In this chapter, an effective computer aided technique is proposed and utilized for pre processing, segmentation and classification purposes. The comparative study of different classifiers for histopathological image of breast cancer classification and detection using morphological features based on all cells present in the image have been describe. The cells are classified by morphological features. This research work deals issues related to staining and with colour consistency problems. The developed technique for automated analysis and evaluation of histopathology images will assist the pathologists and reduces the human error. Such automated cancer diagnosis facilitates improved judgment by the pathologist.