

*Chapter 4: Features extraction for
quantitative measurements of
histopathological images of breast cancer*

Abstract:

The proposed method is based on the dysplastic features that work on the computation of features for differentiation of benign and malignant cells. Morphological measures are significantly utilized to analyze these features. The purpose of choosing morphological operators is based on the fact that these operators principally utilize regularities and distribution of the structural features of cells. The analysis of cell morphology is an essential factor that aids in complete evaluation of microscopic cells, and examination of cell behaviours. In addition, it provides the quantitative measure of area, perimeter, intensity, and texture, etc. present in large populations of cells. For the implementation of the proposed method, (UCSB) available image breast data set of 58 images (26 malignant and 32 benign) have been utilized here. It is observed that malignant cells have the considerably greater magnitude for computed features as compared to benign cells. Significant variation in features values are also found in case of malignant cells. Furthermore, an efficient approach of segmenting cells has been adopted here to represent the histopathology images. These images will provide assistance to the pathologist to identify malignant cells. The results reported here have the potential to classify the benign and the malignant cell. Henceforth, the above mentioned part of our work is pivotal for performing this research.

4.1 Introduction:

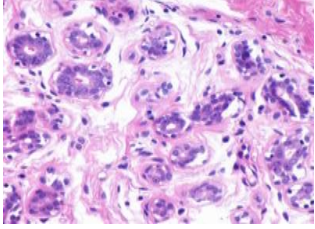
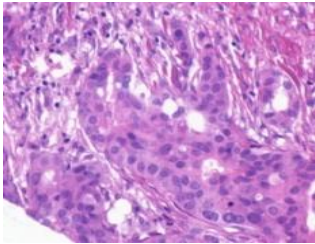
Pathologists used to diagnose tissues based on histopathological images significant features that distinguish among the normal and malignant tissues. Histopathological images are multi-channel in nature and provide rich, minute structural information. Further, it contains meaningful pathological information at various scales (Dermir *et al.*, 2005). Significantly, diagnosis from the histopathological images symbolizes the 'gold standard' in detecting a large number of diseases including cancerous cells (Rubin *et al.*, 2008). For cancer detection, morphological feature extraction is the main tool for analyzing the cellular organization, abnormality, and changes in the physiological state of the cells (Zhang *et al.*, 2015; Gurcan *et al.*, 2009).

Till now, analysis of histopathology images has been done manually via observing dysplastic appearances such as minute structures, distribution and regularities of cells and nuclei, distortion in the shapes and size of cells and change in the density of cluster of cells across the tissues by the pathologist to decide whether it is benign or malignant. This whole process is very time consuming and cumbersome that requires much experience (Veta *et al.*, 2014; Chen *et al.*, 2012; Liu *et al.*, 2007). To overcome this problem, there is the requirement for the development of automation of image analysis. Therefore, computer-aided diagnosis techniques are being introduced for fast, reliable and accurate diagnosis of cancer. Therefore, automation prevails over limitations of manual microscopy based detection method to help pathology practice. Several images analysis software like Cell profiler, Mac Biophotonics Image J and Scion has played a major role in the analysis of cellular images. CAD has been successfully applied in breast cancer, cervical cancer, prostate cancer and lung cancer analysis (Zhang *et al.*, 2015).

4.2 Feature extraction:

Feature extraction is one of the major step to identify the class. Subsequent to the detection of ROI, features extracted from the ROI include cellular level features for better prediction of types of class. Feature extraction and selection are crucial for many image processing and computer vision applications. A feature can be defined according to experience, previous similar studies or visual observation (Rodenacker *et al.*, 2003; Dermir *et al.*, 2005). The objective of feature extraction is to reveal all possible information from input data that are expected to be effective in diagnosis (Antani, 2002; Zitová *et al.*, 2003). Research on useful features for cancer classification and diagnosis has often been approached by the definition of features flagged by clinicians as important features for the diagnosis process (Velkamp, 2001; Chaddad *et al.*, 2010). Pathologists retrieve different features to derive clinically significant information in the stained slide of a tissue sample under the microscope for cancer detection including shape and size of cells, shape, and size of cell nuclei, and distribution of the cells (Veta *et al.*, 2014). Table 4.1 depicts the brief description of these features to difference between normal and cancerous cells.

Table 4.1: Features of benign and malignant cells.

Type	Image of cells	Criteria of abnormality
<p>Benign breast (Non-cancerous)</p>		<ul style="list-style-type: none"> • Similar size and regular shape of cells • Round and single nucleus • Large cytoplasm • Monochromatic nuclei • Fine chromatin structure • Myoepithelial cells presence on outer membrane
<p>Malignant breast (Cancerous)</p>		<ul style="list-style-type: none"> • Enlarged and irregular cells (pleomorphic) • Multiple and irregular shape of nucleus • Diffusion of cytoplasm all over the tissue • Hyperchromatic nuclei • Coarse chromatin structure • Myoepithelial cells are not present • Change in the density of cluster of cells

Feature extraction algorithms include essentially three groups of features: **(1)** morphology features (size and shape), **(2)** intensity features and **(3)** texture features. In the following sub-section, each group of features are explored and the methods of extraction and the associated application within the features group are examined.

4.2.1 Morphological features:

Morphological features provide information about the size and shape of the described region, object or image. Those features are extracted based on the object's inherent information (Esgiar *et al.*, 1998; Dermir *et al.*, 2005). The information about size and shape of the cells are described by morphological features.

This section deal with the features extraction process, in which image features are extracted according to morphological characteristic including area, perimeter, diameter, and circularity (Dermir *et al.*, 2005; Guillaud *et al.*, 2005). Features relating to size include area, perimeter length of a major and minor axis of a cell. Circularity and eccentricity are included in the shape category of a cell (Liu *et al.*, 2007; Yuan *et al.*, 2008). Morphological features have been used to analyze the characteristics of the histology image. These features are used to distinguish between benign and malignant cells by quantifying the properties of size and shape followed by statistical measurements. This method is used to identify the presence of cancer cells (Guillaud *et al.*, 2004; Lassouaoui *et al.*, 2005; Thiran *et al.*, 1996). The structures of cancerous cells are pleomorphic (variation in size and shape) in nature. Histopathology images of abnormal cell tend to vary in size and shape. The extent of cellular polymorphisms generally correlates with the degree of anaplasia. These features are utilized to identify abnormal cells shapes. The cell may be round, tadpole, bizarre, oval or caudate (Nauth *et al.*, 2007). The abnormal cells shape is irregular, with the loss of roundness, indentation, and lobes (Nauth *et al.*, 2007). In comparison, normal cell nuclei have a smooth nuclear envelope, circular shape (round nucleus), small chromocenter and regular shape (Nauth *et al.*, 2007).The normal cells have definite shapes that abet in their respective normal functionalities. Cancerous cells function abnormally and their shapes are often distorted (Nauth *et al.*, 2007).

4.2.2 Intensity features:

Intensity features provide information about the gray level or color of pixels located in the ROI (Belsare *et al.*, 2012). Feature extraction approach uses different color spaces. Intensity based features are used for describing pixel level characteristics of images (Wiltgen *et al.*, 2003; Weyn *et al.*, 1999; Weyn *et al.*, 1998). Pixel-based features are based on the properties of individual pixels, either one-by-one or in a local neighbourhood. Many pixel-based features are based on pixel intensity, edge strength, curvature or texture. These changes will be quantified by the features extracted using two types of information in the image: the intensity values of the pixels and the spatial arrangements of pixels (spatial coordinates) (Wang *et al.*, 2009).

4.2.3 Texture features:

Texture features are based on the inter-pixel variation in various directions and resolutions. Texture features may be measured by obtaining an object of interest's local density variation (Doi, 2007). The texture is defined as something consisting of mutually related elements (Sonka *et al.*, 2000). A texture may be fine, coarse, smooth, or grained, depending upon its tone and structure, where the tone is based on pixel intensity properties, the structure is the spatial relationship between pixels (Haralick, 1979). Further, the spatial arrangement of texture primitives or texture element is termed as textone. It is arranged in more or less periodic manner, where texture primitive is a group of pixels representing the simplest or basic sub pattern.

Feature representation by a texture can be applied to histology images if the texture of the nuclei can be extracted. The texture of abnormal nuclei tends to be coarse and hyperchromatic (Nauth *et al.*, 2007). These features provide information about the variation of intensities presented in the region that aids in tissue identification. In pattern recognition, texture features are useful features in classification tasks (Guillaud *et al.*, 2004). The texture is an important characteristic utilized for classification purposes. The most frequently used textural feature is based on the gray level co occurrence matrix. It measures different combinations of pixel brightness values present in an image, object or local neighbourhood. These features have been shown to be effective in discriminating histopathological images of cancer tissues and cells (Gonzales *et al.*, 2007; Haralick, 1979; Esgiar *et al.*, 1998).

Partitioning and characterisation of histopathological cancer cells are the main focus of the CAD system. When the features are extracted, it is estimated that the features sets will extract appropriate characteristics of the input data. The CAD for grading diseases based on histology images depends upon capturing the variation of cell structure (at the cellular level) and changes (such as the growth of abnormal cell at tissue) that are present at the tissue level (Nauth, 2007; Dermir *et al.*, 2005). For a CAD system, after image segmentation, image features are extracted from the regions of interest to detect and grade potential diseases.

This is implemented by applying CAD systems to identify particular disease signatures with their image features. Image features are of major importance in the segmentation of images. The extraction of features is important and strongly effect classifier design and greatly effects the image classification (Dhawan *et al.*, 2003).

4.3 Materials and Methods:

Cellular image dataset (Breast cancer-UCSB) of benign and malignant cells has been provided by Yale Tissue Microarray Facility placed in www.bioimage.ucsb.edu (Centre for Bio-image Informatics, University of California, Santabarbara (UCSB)).

4.3.1 Dataset preparation:

The dataset is utilized for pre-processing and testing segmentation capability of algorithms along with the classification systems. This dataset is composed of total 58 images (Benign images = 32 images and Malignant images = 26). All benign and malignant cells have been labelled by expert oncologists on the site. Afterward, we have used the designed dataset for testing the performances of classification systems. The breast cancer cells have been cropped including the region of interest. The two main dataset groups are termed as “single cells” and “group cells”.

Both the groups include labelled benign and malignant cells as acquired from dataset images. The dataset of “single cells” consists of 223 benign and 240 malignant cells. Furthermore, the dataset of “group cells” consists of 75 benign and 73 malignant cells respectively. Total 30 features have been used for feature extraction of benign and malignant cells in single cells and group cells dataset. Figure 4.1 depicts the region of interest based cropped images of single cells dataset and group cells dataset respectively.

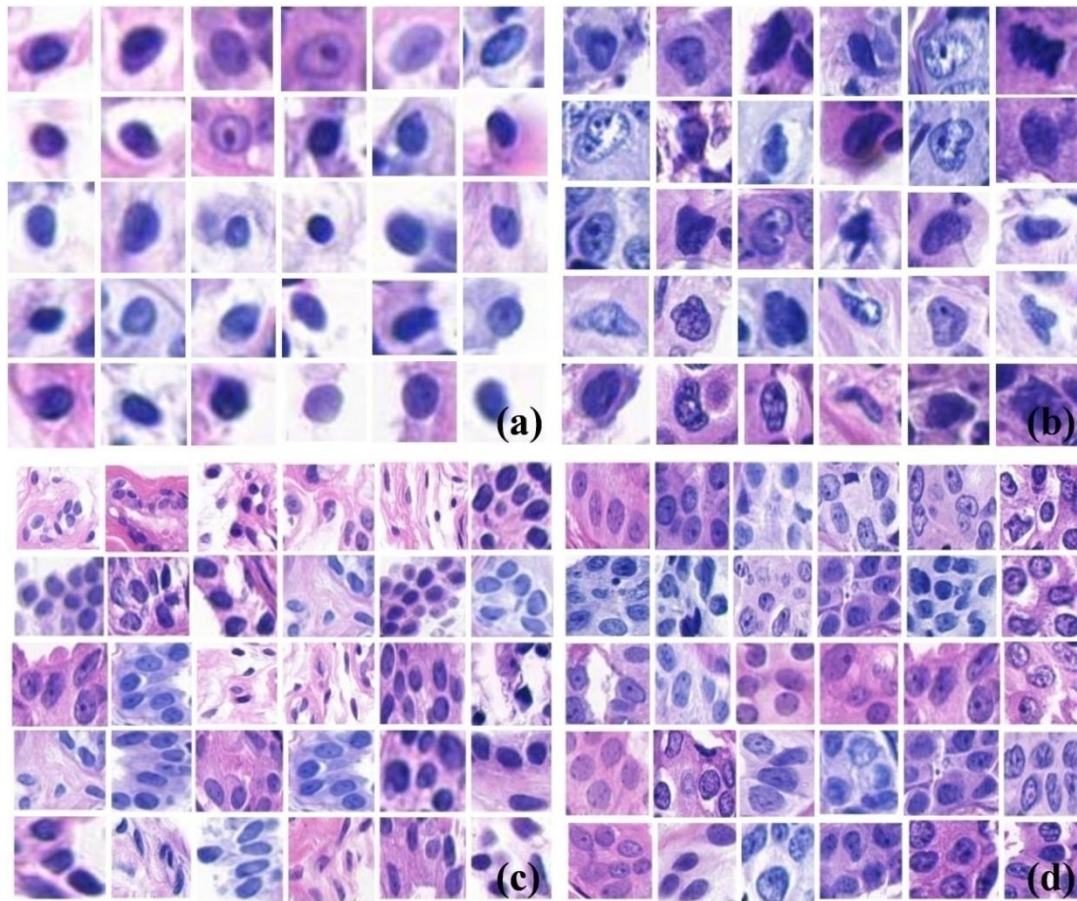


Figure 4.1: Morphological variability of cells.

(a) Single benign cell. (b) Single malignant cell. (c) Group benign cells. (d) Group malignant cells.

The pre-processing module firstly enhances the input image by reducing the acquisition noise and background non-uniformities. Secondly, the background removal of the input image has been done by using a proper segmentation technique. After that, features are extracted from the segmented image. In order to classify the cell, three classifiers k-Nearest Neighbours (k-NN), Artificial Neural Network (ANN), and Support Vector Machine (SVM) have been used to process those morphological indexes separately. By comparing the results, the best classifier for this task has been chosen and the cells test is labelled as benign and malignant in chapter 5 of this thesis. Experiments have been implemented on a 3.40 GHz CPU with 4 GB RAM, 64 bits, Windows 7 operating system, with MATLAB. Figure 4.2 represents the flow chart of the proposed system and basic steps involved in the automated image analysis for the detection of cancer cells.

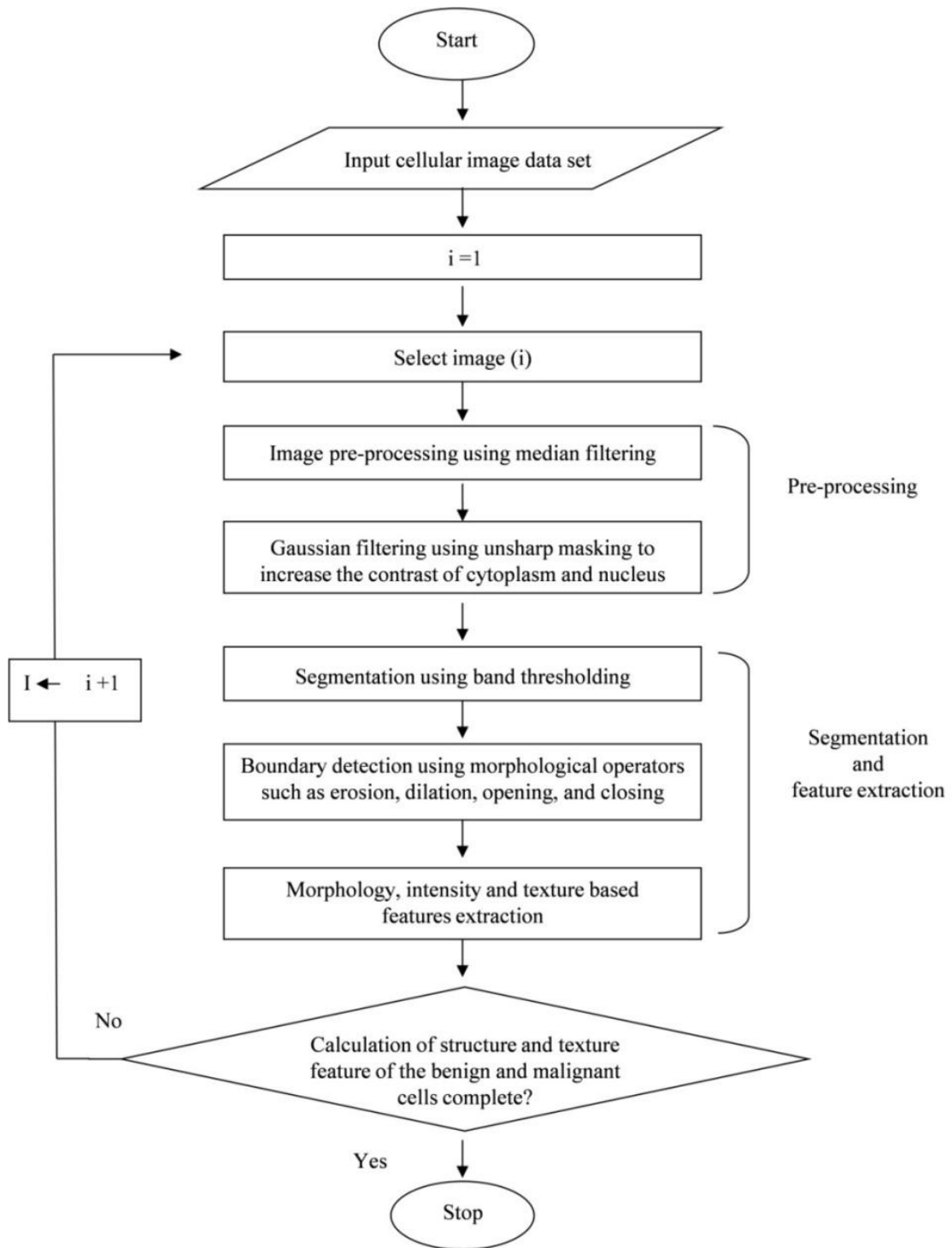


Figure 4.2: Schematic flowchart of the proposed method.

4.3.2 Preprocessing:

The main purpose of the pre-processing stage was to reduce the background noise and to enhance the image to improve the image quality (Cohen *et al.*, 1995; Arif *et al.*, 2007; McAndrew *et al.*, 2014). In this chapter, median filtering was implemented to preprocess the images to eliminate graininess (Gonzalez *et al.*, 2004). The basic fundamental of median filtering is that every output pixel comprises the median value in the 5-by-5 neighborhood around the equivalent pixel in the input image. The image was padded with zeros on the edges, so the median values for the points of 3 pixels of the edges may appear distorted. Afterward, the contrast is enhanced between the cytoplasm, nucleus and extracellular components using unsharp masking. Unsharp masking is a spatial filtering technique which can be used to enhance the edges in an image (Dermir *et al.*, 2005; McAndrew *et al.*, 2014). The unsharp masking operation is carried out by subtracting a scaled unsharp version of the image from the image itself. The result is an image with enhanced edge pixel values. The filter was applied to the image by subtracting the multiplied scaled factor, and gaussian filtered from the input image. A rotationally symmetric gaussian low pass filter with a standard deviation of 50 pixels was used, with a total filter size of 15-by-15 pixels. The scaling factor was 0.35. To enhance the boundaries of the objects in the image Gaussian high pass filter is used. The Gaussian filter gives very high rated results and used very extensively to enhance the finer details of the objects (Gonzalez *et al.*, 2002).

4.3.3 Segmentation:

Segmentation is the process of dividing an image into regions having similar properties such as brightness, color, texture etc. (Gonzalez *et al.*, 2004; Pratt, 2001; Pal *et al.*, 1993). In the process of automated segmentation, the problem encountered includes different tissue incurs very close gray level values for e.g. normal tissue and tumour tissue. Furthermore, as ideal imaging conditions are never realized in practice, noises are present in images e.g. blurring of image, and staining artifact. Owing to it, the image causes a problem in the application of simple gray level dependent segmentation techniques (Ramesh *et al.*, 1912; Pal *et al.*, 1993).

Segmentation is the process where an image is divided into the different regions on some similarity basis. The basic purpose of segmentation is the extraction of important features from the image, from which information can easily be perceived. The decomposition of the image in multiple segments (sets of pixels) is called segmentation (Belsare *et al.*, 2012; Madabhushi, 2009). Accurate image segmentation is essential for the identification of the cancer cells for the morphological structures like size, shape, intensity and texture (Rahmadwati *et al.*, 2010; Dermir *et al.*, 2005).

The accurate feature extraction and classification proportionately depend upon the correct segmentation of the maximized and cropped cellular images. Basically, one operation has been performed in this segment (isolation of the cell). The input for the segmentation is the cropped cell image. The main task is to segment the cell separately. The segmentation is used for background removal of the image. In this chapter, band thresholding has been implemented to group pixels lying in the cellular region for segmentation. Using proper segmentation technique; the object of interest has been segmented out of the image with a black background.

Background removal:

The cropped gray scale cell image is the input of this system. The cell has to be segmented from the rest of the image components so that using the features of these modules, the classification can be performed. For segmentation, first edge detection operation has been performed here. There are several filters available for edge detection (sobel, canny, prewitt etc.). In our work, a canny filter has been used for its better precision in edge detection. Canny filter reconstructs the border of all the cells present in the image. As the edges created by the filter consist of the single pixel only, the borders may contain breakings and may be non continuous. So, it needs to be dilated and joined before further processing. The structural element has been prepared using type square and morphological operation, 'dilation' has been performed that connects the membrane pixels properly. Now, connected membranes have been created for the cells which are completely present in the image. To highlight the connected element present in the image, the 'hole-filling' operation has been done.

Henceforth, using this operation, the internal holes of the connected components in the image have been filled. Morphological operations were done with suitable structure elements. Basic morphological operations such as filling with holes, opening, closing dilation, erosion were performed to plot the boundary of the cells. This procedure provides user to see different outlined cells (Sharma *et al.*, 2009; Xu *et al.*, 2013; Naik *et al.*, 2008; Vahadane *et al.*, 2013).

The next step includes removal of the other non connected components present in the image. Morphological 'opening' has been performed to get the expected output. In the morphological 'opening' operation, the 'erosion' and 'dilation' operations have been performed on an image sequentially using the same structural element. For implementing dilation, an arbitrary sized structuring element has been used. Further, for erosion a disk sized structuring element has been used. The new binary image has been produced comprising biggest connected component with the biggest area.

Isolating the breast cancer cell:

The cropped breast cell has been taken for isolation of cells. The other components present in the image are stroma, tubules, and image background. As stated previously, the image pre-processing includes background removal steps. The breast cell has been isolated using background removal processes and the images are converted to binary format. The opaque area may contain some error pixels. Morphological operator for hole-filling has been used to fill up the errors. Consequently, the background was removed, and the single cell and group cell dataset images were converted to binary format (Bhattacharjee *et al.*, 2015). After this, the binary image of the breast cell dataset is ready for feature extraction related processes.

4.3.4 Feature extraction from segmented image:

Creating a set of features out of a large number of verified dataset is defined by feature extraction. In feature extraction, the acquired data from the image is transformed and labelled to a particular set of features that is applied for further classification (Belsare *et al.*, 2012; Dermir *et al.*, 2005; Thiran *et al.*, 1996).

The multiple algorithms are a prerequisite if the image dataset is desired to be processed fully with its data of every pixel and that seems to be redundant (more data but less information). In order to perform the desired task with reduced representation instead of the full size input, it is expected that if the set of features has been carefully chosen, the features set will provide the relevant information from the input data. Figure 4.3 depicts the binary equivalent images produced by the segmentation technique on the breast cells are used to extract the morphological features defined below.

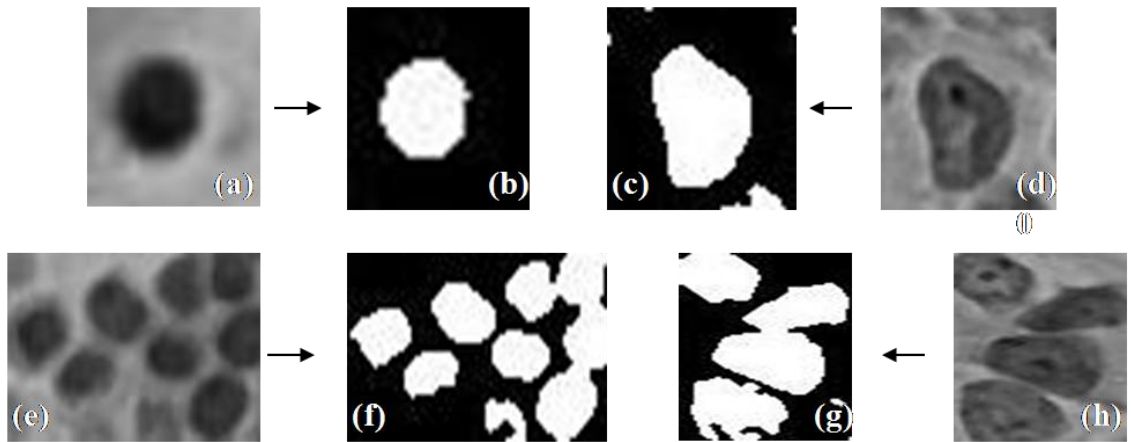


Figure 4.3: Image segmentation method for background removal.

(a) Single benign cell. (b) Single binary opaque of benign cell. (c) Single binary opaque of malignant cell. (d) Single malignant cell. (e) Group benign cells. (f) Group binary opaque of benign cells. (g) Group binary opaque of malignant cells. (h) Group malignant cells.

The regions of interest (ROI) of the segmented cells were considered for feature manipulation (Chaddad *et al.*, 2011; Liu *et al.*, 2007). Un-weighted centroid and the weighted centroid have been marked by blue and red color boundaries respectively. Measurement of standard deviation has been performed. Subsequently, it has been converted to the gray level image having one bounding box marked by yellow color (Zhao *et al.*, 2015; Esgiar *et al.*, 1998). Figure 4.4 depicts the results obtained by implementing the steps discussed above.

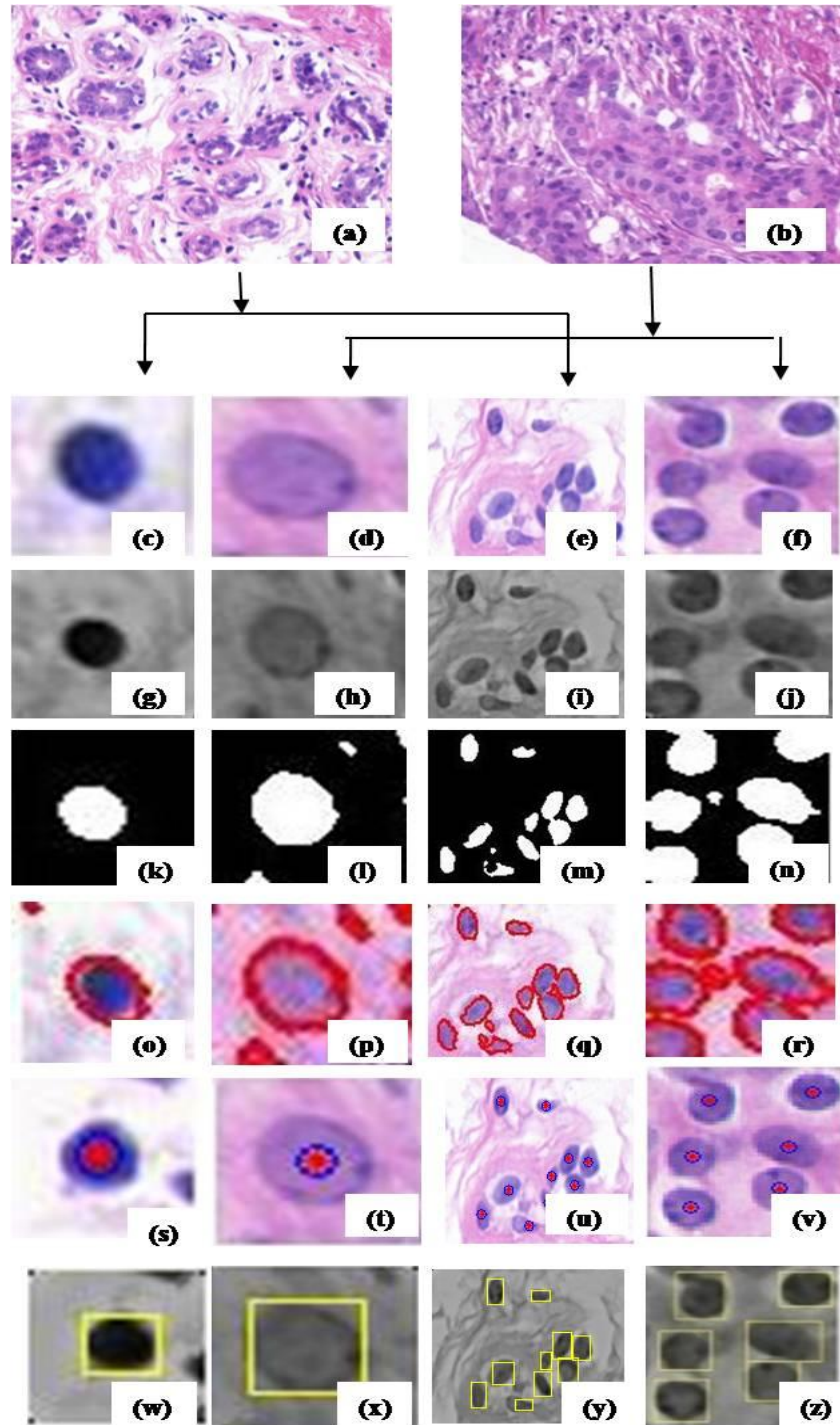


Figure 4.4: Step in CAD method of histopathological breast cancer images.

(a) Benign cell. (b) Malignant cells. Selected ROI of cells in RGB (c) Single benign cell (d) Single malignant cell (e) Group benign cells. (f) Group malignant cells. (g, h, i, j) converted into gray scale image respectively. (k, l, m, n) after band thresholding respectively. (o, p, q, r) distinct cells respectively. (s, t, u, v) weighted (red) and unweighted (blue) marked centroid respectively. (w, x, y, z) cells in the bounding box respectively.

Table 4.2 shows the features of particular ROI extracted to distinguish different types of cells. They are differentiated as benign and malignant based on their morphology, intensity, and texture features in single cells and in the group of cells.

Table 4.2: The distribution of various features extracted from images and their ranges.

Name of features	Number of features (range F1–F30)
Morphology features	8 (F1–F8)
Intensity features	4 (F9–F12)
Texture features	18 (F13-30)

All these features were computed from the segmented cell images. Furthermore, the 30 features have been computed and stored for each cell present in the color and binary images.

Morphology features: (Size and shape features) (F1–F8)

The most significant portion of this work is the computation of features. The morphological appearance of structures like size, shape, intensity and texture features, were important factors for the identification of the cancer cells (Zhao *et al.*, 2015; Dermir *et al.*, 2005). Single cells and group cells dataset have been taken into account for segmentation and feature extraction. For each group of cells, the features were taken as an average of the total number of cells present in the group. The features taken for the group of cells might mislead the classifier while classifying between benign and malignant cells, so single cells were also taken into account and segmented. The quantification of these features helps to differentiate the malignant cells from benign cells. Moreover, the statistics computed on these properties is used to identify cancerous tissue. This work considers the following features as input to the classification process: area, diameter, and roundness parameters of morphological features.

The eight (8) features (F1) to (F8) and equations 4.1 to 4.7 illustrates morphology based features applied in this chapter are as follows: area (F1), convex area (F2), solidity (F3), perimeter (F4), major axis length (F5), minor axis length (F6), circularity (F7) and eccentricity (F8) respectively.

(F1). Area: The area can be represented by cell region containing a total number of non-zero pixels in the region.

$$Area = \sum_{i=1}^n \sum_{j=1}^m B(i, j) \quad (4.1)$$

‘**A**’ is cell area and ‘**B**’ is the segmented image of rows i and j columns.

(F2). Convex area: Scalar that specifies the number of pixels in a convex image.

(F3). Solidity: Scalar was specifying the proportion of the pixels in the convex hull that is also in the region, computed as Area/Convex Area.

$$Solidity = \frac{Area}{Convex Area} \quad (4.2)$$

(F4). Perimeter: Perimeter calculates the distance between each adjoining pair of pixels around the border of the region. It is defined as:

$$Perimeter = Even\ count + \sqrt{2}(odd\ count)\ unit. \quad (4.3)$$

(F5). Major axis length: It specifies the length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region.

$$Major\ axis\ length = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (4.4)$$

x_1, y_1 and x_2, y_2 are end points on the major axis.

(F6). Minor axis length: It specifies the length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region.

$$\text{Minor axis length} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (4.5)$$

x_1, y_1 and x_2, y_2 are end points on the minor axis.

(F7). Circularity: The shape factor is used to measure or assess the shape of the cells. The shape factor is measured by the roundness factor. The roundness factor is used to measure or assess the shape of the cells as presented in the equation.

This dimensionless parameter is calculated by area and perimeter.

$$\text{Circularity} = \left(4 * \pi * \frac{\text{Area}}{\text{perimeter}^2} \right) \quad (4.6)$$

(F8). Eccentricity: The ratio of major axis length and minor axis length is known as eccentricity and defined as:

$$\text{Eccentricity} = \frac{\text{Length of Major Axis}}{\text{Length of Minor Axis}} \quad (4.7)$$

Intensity features: (F9–F12)

The texture features provide information about the variation in the intensity of a surface and quantify properties such as regularity, coarseness, and smoothness. The texture is a connected set of pixels that repeatedly occur in an image (Gonzales *et al.*, 2007). The texture properties have been derived using: **(1)** First order statistics and **(2)** Second order statistics that are computed from spatial gray level co-occurrence matrices (GLCMs).

First order statistical features:

They are also called as amplitude or histogram features as they are computed indirectly in terms of the histogram of image pixels within a neighbourhood. The statistical moments are calculated from the first order gray level histogram, which is

defined as the distribution of the probability of occurrence of a gray level in the image (Gonzalez *et al.*, 2004).

Pixel based features provide information about the intensity (gray level or color) histogram of the pixels located in cells (Bhattacharjee *et al.*, 2014; Belsare *et al.*, 2012). These features were extracted from the gray-level or color histogram of the image. These types of features do not provide any information about the spatial division of the pixels.

Let z be a random variable denoting gray levels and let $p(z_i)$, $i = 0, 1, 2, \dots, L-1$, be the corresponding histogram, where L is the number of distinct gray levels. The n^{th} moment of z about the mean is given by:

$$\mu_n(z) = \sum_{i=0}^{L-1} (z_i - m)^n p(z_i)$$

The different parameters that are computed by using statistical moments are as follows: max intensity (F9), min intensity (F10), mean intensity (F11) and standard deviation (F12) to obtain first order statistical information about the texture of tissues.

The four (4) features (F9) to (F12) describe the max intensity (F9), min intensity (F10), mean intensity (F11) and standard deviation (F12) correspondingly. The equation 4.8 describes the formula of standard deviation applied here for image processing purposes.

(F9). Max intensity: Scalar was specifying the value of the pixel with the greatest intensity in the region.

(F10). Min intensity: Specifying the value of the pixel with the lowest intensity in the region.

(F11). Mean intensity: It specifies the mean of all the intensity values in the region. The mean just tells us the average gray level of each region and is gives us only a rough idea about the intensity, not really texture.

(F12). Standard deviation: The standard deviation is much more informative, it indicates the variability of gray levels i.e., change in the contrast.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (4.8)$$

Note-These functions can provide useful information about the texture of an image but cannot provide information about shape, i.e., the relationships of pixels in an image. Another statistical method that considers the spatial relationship of pixels is the gray level co-occurrence matrix (GLCM), also known as the spatial gray level dependence matrix.

Texture features: 18 (F13-30)

Second - order statistical features based on gray level co-occurrence matrix (GLCM):

The 2nd order statistics are based on pairs of a pixel and takes into account the spatial distribution of the pixels gray level. The co-occurrence matrix is the second order statistic method, which characterizes the spatial interrelationships of the gray tones in an image. Haralick *et al.* (1979) suggested the use grey level co occurrence matrices (GLCMs) for a definition of textural features. The texture analysis techniques based on the GLCM (gray level co occurrence matrix) is applied to histopathological images analysis (Gonzalez *et al.*, 2002; Esgiar *et al.*, 1998).

The values of the co-occurrence matrix element present relative frequencies with which two neighboring pixels separated by distance d and angle Θ appear on the image where one of them has gray level i and the other j figure 4.5 and their joint probability of occurrence is given by $P(i,j)$. Such matrix is symmetric and also a function of the angular relationship between two neighboring pixels. The co-occurrence matrix can be calculated on the whole image, but by calculating it in a small window, which scans throughout the image, the co-occurrence matrix can be associated with each pixel.

According to Haralick *et al.* (1997) non normalized frequencies $P(i,j)$ of the co-occurrence matrix for defined window $M \times N$, distance d and angles quantized to 45° intervals are defined by:

$$P(i, j, d, 0^\circ) = \#\{(k, l), (m, n) \mid k - m = 0, |l - n| = d, I(k, l) = i, I(m, n) = j\}$$

$$P(i, j, d, 45^\circ) = \#\{(k, l), (m, n) \mid (k - m = d, l - n = -d), (k - m = -d, l - n = d), I(k, l) = i, I(m, n) = j\}$$

$$P(i, j, d, 90^\circ) = \#\{(k, l), (m, n) \mid |k - m| = d, l - n = 0, I(k, l) = i, I(m, n) = j\}$$

$$P(i, j, d, 135^\circ) = \#\{(k, l), (m, n) \mid (k - m = d, l - n = d), (k - m = -d, l - n = -d), I(k, l) = i, I(m, n) = j\}$$

where # denotes number of elements in the set and $I(m,n)$ is the pixel value at point (m, n)

Note: The above matrix are symmetric; i.e. $P(i, j, d, a^\circ) = P(j, i, d, a^\circ)$.

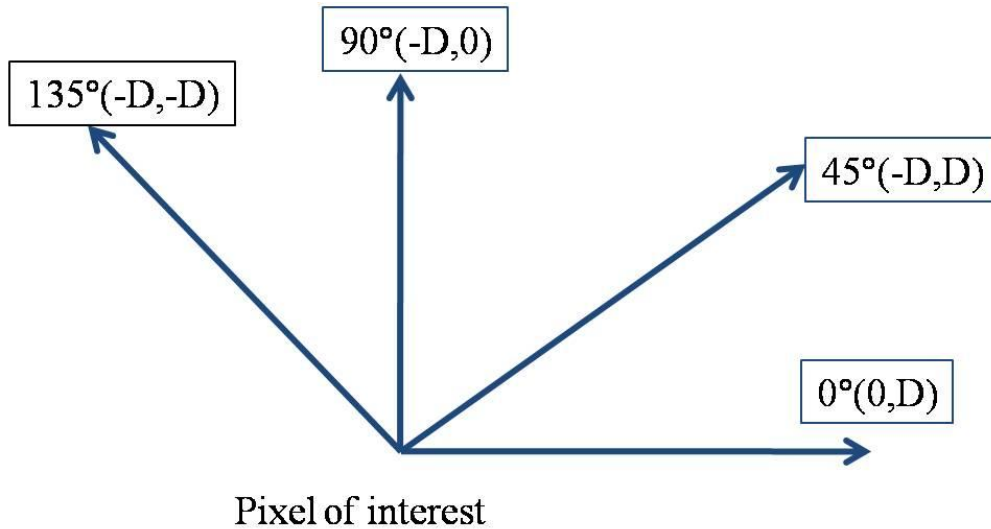


Figure 4.5: Directionality use in calculating GLCM (D is the distance between two pixels)

Depending on the domain application, only some of those features can be used. The features (F13) to (F30) and equations 4.9 to 4.26 illustrate the gray level co-occurrence matrix GLCM quantifies the various texture-based features applied in this chapter as follows: From each of these matrices eighteen (18) texture features are computed and they are as follows:

Autocorrelation (F13), contrast (F14), correlation (F15), cluster prominence (F16), dissimilarity (F17), energy (F18), entropy (F19), homogeneity (F20), maximum probability (F21), sum of squares (F22), sum of average (F23), sum of variance (F24), sum of entropy (F25), difference variance (F26), difference entropy (F27), information measure of correlation 2 (IMC-2) (F28), inverse difference normalized (IDN) (F29) and inverse difference moment normalize (IDMN) (F30) etc. (Haralick *et al.*, 1979).

For an $N_x \times N_y$ image, with each pixel quantised to N_g levels, let L_x be the horizontal spatial domain, L_y the vertical spatial domain, and G the set of quantised gray levels, such that $L_x = \{1,2,\dots,N_x\}$, $L_y = \{1,2,\dots,N_y\}$, and $G = \{1,2,\dots,N_g\}$. N_g = number of gray levels. This measure provides evidence of how sharp the structural variations in the image are.

$$p_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)$$

where $k = 2,3,\dots, 2N_g$

$$p_{x-y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)$$

where $k = 0,1,\dots, N_g-1$

(F13). Autocorrelation: Autocorrelation is a measure of the magnitude of the fineness and coarseness of texture.

$$Autocorrelation = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i, j) i, j \quad (4.9)$$

(F14). Contrast (inertia): Contrast measure provides evidence of how sharp are the structural variations in the image and is defined as the amount of local variation present.

$$Contrast = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i - j|^2 p(i, j) \quad (4.10)$$

(F15). Correlation: Correlation is a measure of the linear dependency of gray levels on those of neighbouring pixels or specified points. It indicates the local gray-level dependency on the texture image; higher values are obtained for similar gray-level regions.

$$Correlation = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P(i, j) ij - \mu_x(i) \mu_y(j)}{\sigma_x(i) \sigma_y(j)} \quad (4.11)$$

(F16). Cluster Prominence: Cluster prominence is also a measure of asymmetry. When the prominence cluster value is high, the image is less symmetric. Furthermore, when prominence cluster value is low, there is a peak in the GLCM matrix around the mean values.

$$Cluster\ prominence = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \left\{ i + j - \mu_x(i) - \mu_y(j) \right\}^4 \times P(i, j) \quad (4.12)$$

(F17). Dissimilarity: Dissimilarity is a measure of local intensity variation defined as the mean absolute difference between the neighbouring pairs. A larger value correlates with a greater disparity in intensity values among neighboring voxels.

$$Dissimilarity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i - j| P(i, j) \quad (4.13)$$

(F18). Energy or Angular second moment (ASM): Energy gives a strong measure of uniformity (homogeneity) of the texture present in/ of an image and can be defined as

$$Energy = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)^2 \quad (4.14)$$

Higher non-uniformity values provide evidence of higher structural variations.

(F19). Entropy: Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image. Entropy is defined as

$$Entropy = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \log_2\{p(i,j)\} \quad (4.15)$$

(F20). Homogeneity: Homogeneity is the measure that increases with less contrast in the window. It is defined as or Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$Homogeneity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1 + |i - j|} \quad (4.16)$$

(F21). Maximum Probability: Maximum Probability is occurrences of the most predominant pair of neighboring intensity values.

$$Maximum Probability = \max (p(i,j)) \quad (4.17)$$

(F22). Sum of squares: Sum of squares is the measure that tells us by how much the gray level are varying from the mean.

$$Sum\ squares = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu_x)^2 p(i,j) \quad (4.18)$$

(F23). Sum of average: Sum Average measures the relationship between occurrences of pairs with lower intensity values and occurrences of pairs with higher intensity values.

$$sum\ average = \sum_{k=2}^{2N_g} p_{x+y}(k)k \quad (4.19)$$

(F24). Sum of variance: Sum variance is a measure of heterogeneity that places higher weights on neighboring intensity level pairs that deviate more from the mean.

Using coefficients p_{x+y} and Sum Entropy (SE) calculate and return the mean Sum Variance.

$$sum\ variance = \sum_{k=2}^{2N_g} (k - SE)^2 p_{x+y}(k) \quad (4.20)$$

(F25). Sum of entropy: Sum Entropy is a sum of neighborhood intensity value differences.

$$sum\ entropy = \sum_{k=2}^{2N_g} p_{x+y}(k) \log_2\{p_{x+y}(k)\} \quad (4.21)$$

(F26). Difference variance: Difference Variance is a measure of heterogeneity that places higher weights on differing intensity level pairs that deviate more from the mean.

Using coefficients p_{x+y} and DifferenceAverage (DA) calculate and return the mean Difference Variance.

$$Difference\ variance = \sum_{k=0}^{N_g-1} (1 - DA)^2 p_{x+y}(k) \quad (4.22)$$

(F27). Difference entropy: Difference Entropy is a measure of the randomness/variability in neighborhood intensity value differences.

$$\text{Difference entropy} = \sum_{k=0}^{N_g-1} p_{x-y}(k) \log_2\{p_{x-y}(k)\} \quad (4.23)$$

(F28). Information measure of correlation2: the additional properties not included in

Using coefficients HXY and HXY2, calculate and return the mean Informal Measure of Correlation 2

$$\text{Informal measure of correlation1 (IMC 1)} = \frac{HXY - HXYI}{\max\{HX, HY\}}$$

$$\text{Informal measure of correlation2 (IMC 2)} = (1 - \exp[-2.0(HXY2 - HXY)])^{1/2}$$

$$HXY = - \sum_i \sum_j p(i,j) \log(p(i,j))$$

$$HXYI = - \sum_i \sum_j p(i,j) \log\{p_x(i)p_y(j)\}$$

$$HXY2 = - \sum_i \sum_j p_x(i)p_y(j) \log\{p_x(i)p_y(j)\} \quad (4.24)$$

Where HXY and HXY are entropies of p_x and p_y

(F29). Inverse difference normalized (IDN): IDN (inverse difference normalized) is another measure of the local homogeneity of an image. Unlike Homogeneity1, IDN normalizes the difference between the neighboring intensity values by dividing by the total number of discrete intensity values.

$$\text{Inverse difference normalized} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1 + \left(\frac{|i-j|}{N_g}\right)} \quad (4.25)$$

(F30). Inverse difference moment (IDM): Inverse difference moment measure the local homogeneity of an image and is defined as

$$\text{Inverse difference moment} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1 + (i - j)^2} \quad (4.26)$$

IDM (inverse difference moment) is a measure of the local homogeneity of an image. IDM weights are the inverse of the Contrast weights (decreasing exponentially from the diagonal $i=j$ in the GLCM).

The data used in present algorithm is of normalized type; the data so generated contain normalized features vectors computed around each pixel. The normalized features vectors contain altogether 30 features computed over the window size of $n \times n$ size pixel matrix. The feature vector includes four features derived from first order statistic and 18 features derived from second order statistic.

4. 4 Results and discussion:

In this work, morphological, intensity and texture based features of histopathological images of breast cancer have been extracted for benign and malignant cells. The single benign cells figure 4.6 (a) and group benign cells figure 4.6 (c) have smaller and almost similar in size, regular in shape, proper distribution of cells, single and round nucleus, monochromatic nuclei, smooth nuclear border while single malignant cells figure 4.6 (b) and group malignant cells figure 4.6 (d) have enlarged size, irregular shape, disorganized arrangement of cells, multiple nucleus, hyperchromatic nuclei, irregular and thickened nuclear membrane.

Further, this observation was confirmed by obtained numerical values related to the area, perimeter, major axis length, minor axis length, circularity, eccentricity and max intensity are shown in table 4.3 (single benign cells), table 4.4 (single malignant cells), table 4.5 (group benign cells) and table 4.6 (group malignant cells).

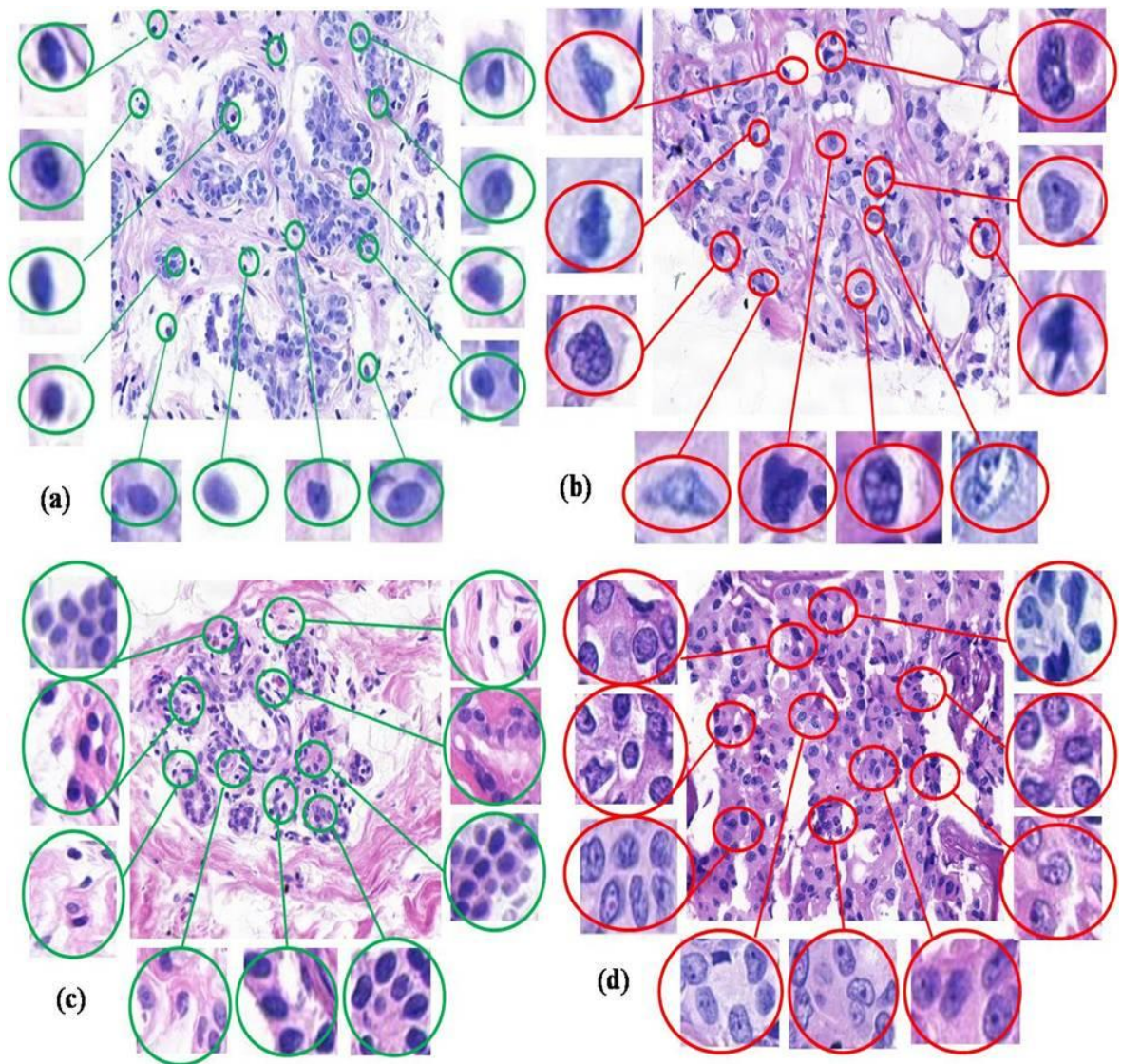


Figure 4.6 Histopathology breast cancer dataset.

(a) Single benign cells dataset. (b) Single malignant cells dataset. (c) Group benign cells dataset.
(d) Group malignant cells dataset.

Table 4.3: Representing area (F1), perimeter (F4), major axis length (F5), minor axis length (F6), circularity (F7), eccentricity (F8) and max intensity (F9) levels of 10 images of single benign cells of breast cancer.

Single Benign Cells	Area	Perimeter	Major axis length	Minor axis length	Circularity	Eccentricity	Max intensity
Image 1	257	56.97	19.30	17.11	0.99	0.46	92
Image 2	150	43.65	14.72	13.34	0.98	0.42	92
Image 3	426	77.79	26.93	20.54	0.88	0.64	92
Image 4	206	51.31	16.59	16.07	0.98	0.24	92
Image 5	415	72.87	24.18	22.05	0.98	0.40	92
Image 6	336	68.97	25.78	16.80	0.88	0.75	92
Image 7	435	75.21	24.85	22.49	0.96	0.42	92
Image 8	334	66.62	24.06	17.80	0.94	0.67	92
Image 9	275	58.97	19.70	17.95	0.99	0.41	92
Image 10	449	76.62	25.61	22.53	0.96	0.47	92

Table 4.4: Representing area (F1), perimeter (F4), major axis length (F5), minor axis length (F6), circularity (F7), eccentricity (F8) and max intensity (F9) levels of 10 images of single malignant cells of breast cancer.

Single Malignant Cells	Area	Perimeter	Major axis length	Minor axis length	Circularity	Eccentricity	Max intensity
Image 1	882	112.42	37.07	30.74	0.87	0.55	92
Image 2	1011	117.01	38.43	33.74	0.92	0.47	103
Image 3	513	96.04	34.29	20.87	0.69	0.79	107
Image 4	1722	184.18	53.21	45.16	0.63	0.52	94
Image 5	710	116.52	39.22	24.33	0.65	0.78	161
Image 6	478	82.97	29.03	21.64	0.87	0.66	92
Image 7	799	163.59	45.51	28.39	0.37	0.78	131
Image 8	737	104.04	35.98	26.77	0.85	0.66	94
Image 9	836	129.25	43.80	25.95	0.62	0.80	103
Image 10	623	97.11	35.56	22.94	0.83	0.76	93

Table 4.5: Representing area (F1), perimeter (F4), major axis length (F5), minor axis length (F6), circularity (F7), eccentricity (F8) and max intensity (F9) levels of 10 images of group benign cells of breast cancer.

Group Benign Cells	Group of Cells	Average Area	Perimeter	Major axis length	Minor axis length	Circularity	Eccentricity	Max intensity
Image 1	11 cells	242	67.49	22.83	14.95	0.66	0.72	94.66
Image 2	10 cells	243	60.05	23.36	13.38	0.84	0.79	91.44
Image 3	9 cells	275	60.98	21.46	16.52	0.92	0.57	91.66
Image 4	6 cells	216	61.21	23.56	12.31	0.72	0.82	91.90
Image 5	7 cells	250	76.63	28.48	13.13	0.53	0.88	94
Image 6	8 cells	236	66.87	26.97	12.38	0.66	0.80	94.33
Image 7	6 cells	242	59.11	22.44	14.22	0.87	0.72	91.66
Image 8	4 cells	228	60.11	24.21	12.01	0.79	0.86	92
Image 9	6 cells	251	68.99	30.25	10.89	0.66	0.92	92
Image 10	5 cells	283	64.80	24.51	14.93	0.84	0.78	92

Table 4.6: Representing area (F1), perimeter (F4), major axis length (F5), minor axis length (F6), circularity (F7), eccentricity (F8) and max intensity (F9) levels 10 of images of group malignant cells of breast cancer.

Group Malignant Cells	Group of Cells	Area	Perimeter	Major axis length	Minor axis length	Circularity	Eccentricity	Max intensity
Image 1	4 cells	541.5	87.99	31.21	21.22	0.87	0.72	99.5
Image 2	5 cells	584	112.59	30.10	26.13	0.57	0.47	104.33
Image 3	5 cells	578.33	90.40	31.77	23.40	0.88	0.66	100.66
Image 4	5 cells	649.66	100.32	35.02	24.27	0.81	0.67	98.33
Image 5	6 cells	765.25	138.39	40.64	27.13	0.50	0.66	98.5
Image 6	6 cells	1220.5	204.37	56.67	32.35	0.36	0.81	104.5
Image 7	4 cells	390.66	78.35	25.14	20.13	0.79	0.59	96.33
Image 8	6 cells	367.57	73.72	25.00	19.63	0.84	0.50	98.85
Image 9	4 cells	386.33	72.29	24.11	20.44	0.92	0.46	95
Image 10	5 cells	408.5	92.42	31.02	19.30	0.60	0.64	113.25

For better identification of every cell present in an image, we have considered both frequency values (no of cells present in specific range of area) and average values (total number of cell area divided by number of cells) in representing the dataset of single cells and group cells that belong to benign and malignant cells respectively.

Seven morphological and one intensity features of benign and malignant cells for single cells datasets and group cells dataset are being plotted as frequency graph shown in figure 4.7 to 4.13 to show the variations of no of cells in terms of variations of morphological features in context to that, figures 4.7 to 4.10 shows four main features related to the size such as area (F1) (figure 4.7), perimeter (F4) (figure 4.8), major axis length (F5) (figure 4.9), minor axis length (F6) (figure 4.10) and figures 4.11 to 4.12 shows two features related to shape such as circularity (F7) (figure 4.11) and eccentricity (F8) (figure 4.12).

Further, figure 4.13 shows one feature related to maximum intensity (F9) is also plotted for single cells and group cells dataset for analysis. Similarly the for group cells datasets also the graph is being plotted for the same morphological and intensity features. This plot will yield significant analysis for single and group cells dataset of benign and malignant cells.

The figure 4.7 (a) reveals that the total area (F1) of single benign cells (blue colored) having/showing (150 to 550 pixels² (unit) and single malignant cells dataset (red colored) (650 to 1750 pixels²) respectively. Henceforth, the figure 4.7 (a) clearly implies that single malignant cells show larger areas in comparison to single benign cell dataset areas. Furthermore, figure 4.7 (b) shows the similar trend of the larger area by the group of malignant cells dataset (red colored) (500 to 900 pixels²) as compared to the group of benign cells dataset (blue colored) (100 to 450 pixels²). Henceforth, both the figure 4.7 (a) and 4.7 (b) depicts that malignant cells dataset have larger areas in comparison to benign cells dataset.

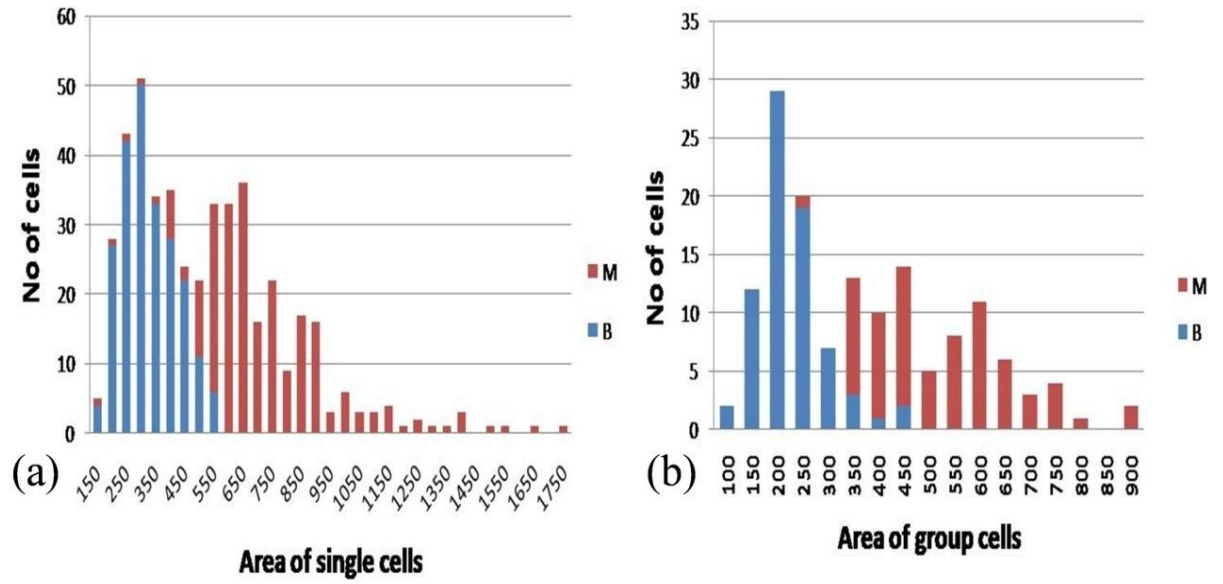


Figure 4.7: Area (F1) feature based variations of values in breast cancer cells.
 (a) Single benign (B) and malignant (M) cell dataset. (b) Group benign (B) and malignant (M) cells dataset.

The figure 4.8 (a) shows that the perimeter (F4) of single benign (blue colored) (110 to 190) pixels and single malignant cells (red colored) are having (40 to 100) pixels respectively. Henceforth, the figure 4.8 (a) clearly implies that single malignant cells occupy larger magnitude of the perimeter in comparison to the amount of single benign cell perimeter. Furthermore, figure 4.8 (b) shows the similar trend of the larger perimeter by group malignant cells (red colored) (120 to 210 pixels) as compared to group benign cells (blue colored) (50 to 110 pixels). Henceforth, both the figure 4.8 (a) and 4.8 (b) depicts that malignant cells have the larger perimeter in comparison to benign cells.

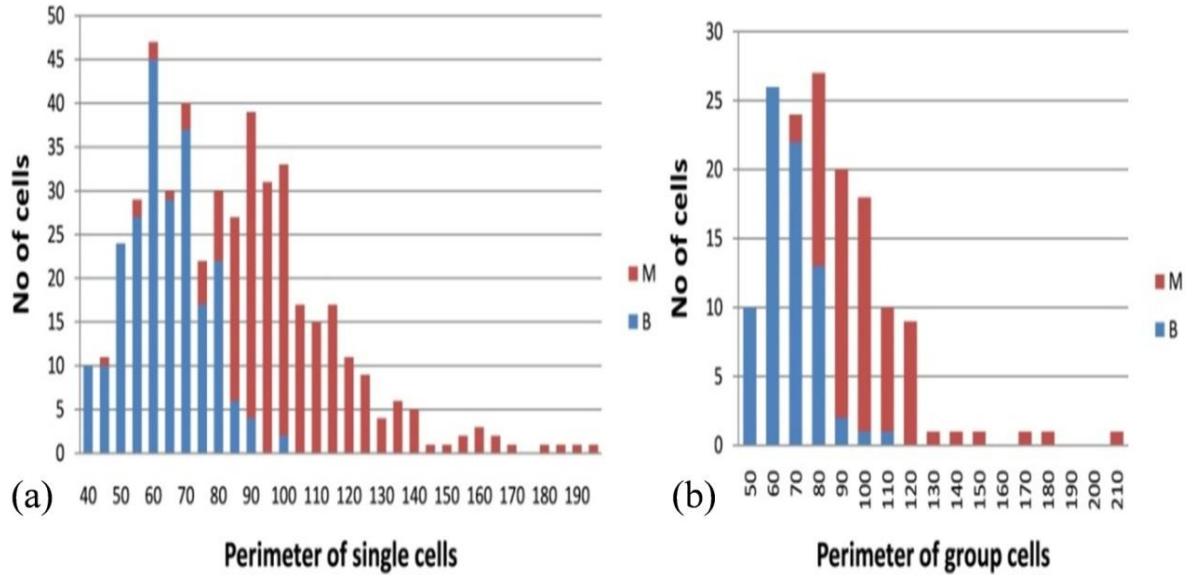


Figure 4.8: Perimeter (F4) feature based variations of values in breast cancer cells.
 (a) Single benign (B) and malignant (M) cell dataset. (b) Group benign (B) and malignant (M) cells dataset.

The figure 4.9 (a) shows that the major axis length (F5) of single benign cells (blue colored) (14 to 34 pixels) and single malignant cells (red colored) (38 to 70 pixels) respectively. Henceforth, the figure 4.9 (a) clearly implies that single malignant cells larger values of major axis length in comparison to the single benign cell major axis length. Furthermore, figure 4.9 (b) shows the similar trend of the larger major axis length by group malignant cells (red colored) (34 to 56 pixels) as compared to group benign cells (blue colored) (16 to 32 pixels). Hence, both the figure 4.9 (a) and 4.9 (b) depicts that malignant cells have the larger major axis length in comparison to benign cells.

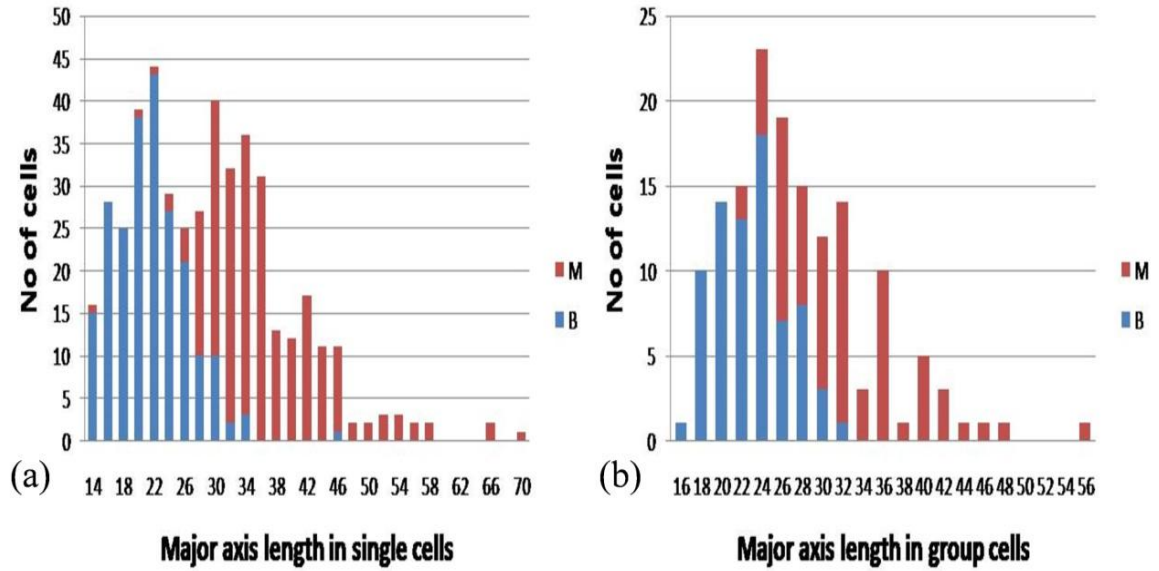


Figure 4.9: Major axis length (F5) feature based variations of values in breast cancer cells.
 (a) Single benign (B) and malignant (M) cell dataset. (b) Group benign (B) and malignant (M) cells dataset.

The figure 4.10 (a) shows that the minor axis length (F6) of single benign cells (blue colored) (12 to 24 pixels) and single malignant cells (red colored) (26 to 46 pixels) respectively. Henceforth, the figure 4.10 (a) clearly implies that single malignant cells larger values of minor axis length in comparison to the single benign cell minor axis length. Furthermore, figure 4.10 (b) shows the similar trend of the larger minor axis length by group malignant cells (red colored) (22 to 34 pixels) as compared to group benign cells (blue colored) (10 to 20 pixels). Hence, both the figure 4.10 (a) and 4.10 (b) depicts that malignant cells have the larger minor axis length in comparison to benign cells.

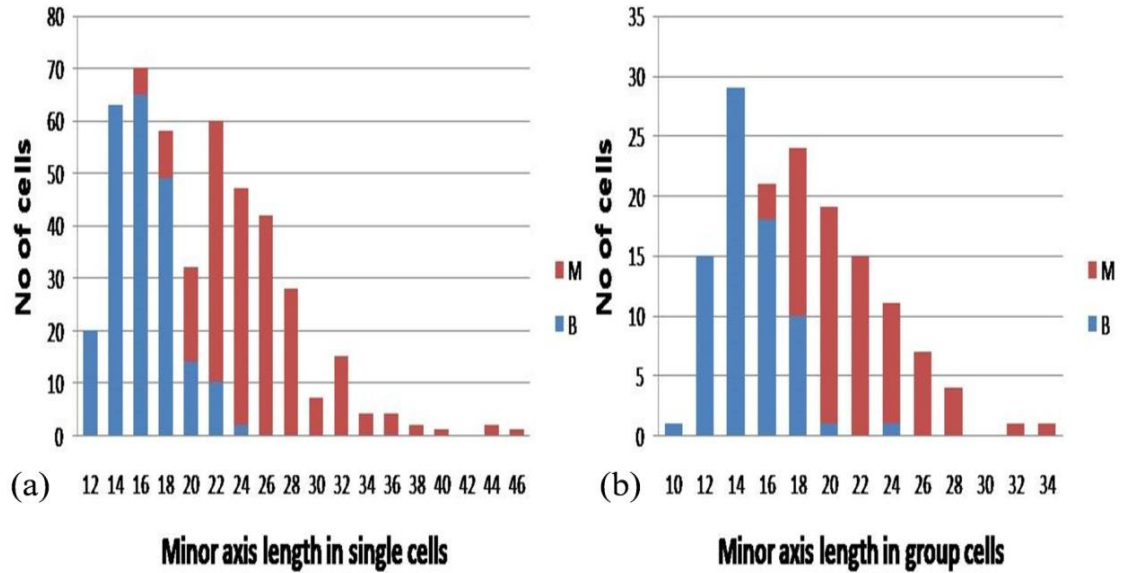


Figure 4.10: Minor axis length (F6) feature based variations of values in breast cancer cells.

(a) Single benign (B) and malignant (M) cell dataset. (b) Group benign (B) and malignant (M) cells dataset.

The figure 4.11(a) shows that the circularity (F7) of single benign (blue colored) 1(one) and single malignant cells (red colored) 0 (zero) respectively. The figure 4.11 (a) clearly implies that single benign cells dataset values (1 to 0.9) tends to move towards 1, that implies near to circularity shape features. On the contrary, the malignant cells datasets (red colored) values (0.85 to 0.5) tends to move towards 0, in reverse direction as compared to benign cells that indicate diminishing of circular properties and occurrence of elongated structures in malignant cells. The figure 4.11(b) shows fewer circularity features of malignant group cells (red colored) ranging from 0.55 to 0.3 in contrast to group benign cells (blue colored) that ranges from 1 to 0.6 respectively. Henceforth, both the figure 4.11(a) and 4.11(b) depicts that malignant cells are less circular and more elongated in comparison to benign cells respectively.

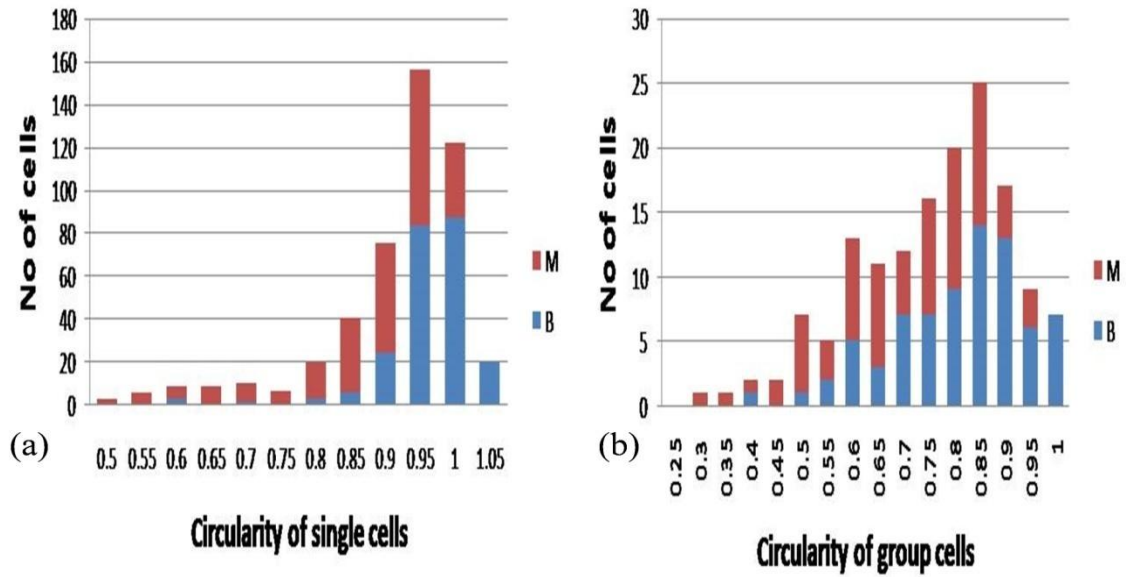


Figure 4.11: Circularity (F7) feature based variations of values in breast cancer cells.
 (a) Single benign (B) and malignant (M) cell dataset. (b) Group benign (B) and malignant (M) cells dataset.

The figure 4.12 (a) shows that the eccentricity (F8) of single benign cells (blue colored) 0 and single malignant cells (red colored) 1 respectively. The figure 4.12 (a) clearly implies that single benign cells dataset values (0.7 to 0.05) tends to move towards 0, that implies nearly to circular shape features. On the contrary, the malignant cells datasets (red colored) values (0.7 to 0.99) tend to move towards 1, in reverse direction as compared to benign cells that indicate circularity of the cells is diminishing. The figure 4.12 (b) shows more eccentricity (elongated) features of malignant group cells (red colored) ranging from (0.6 to 1) in contrast to group benign cells (blue colored) that ranges from (0.55 to 0.3) respectively. Henceforth, both the figures 4.12 (a) and 4.12 (b) depicts that malignant cells are more eccentric and elongated in comparison to benign cells respectively.

Furthermore, our result analysis resembles with findings of other research articles (Nauth *et al.*, 2007) that observes less circularity, more eccentricity and elongated aspects of malignant cells in comparison to benign cells.

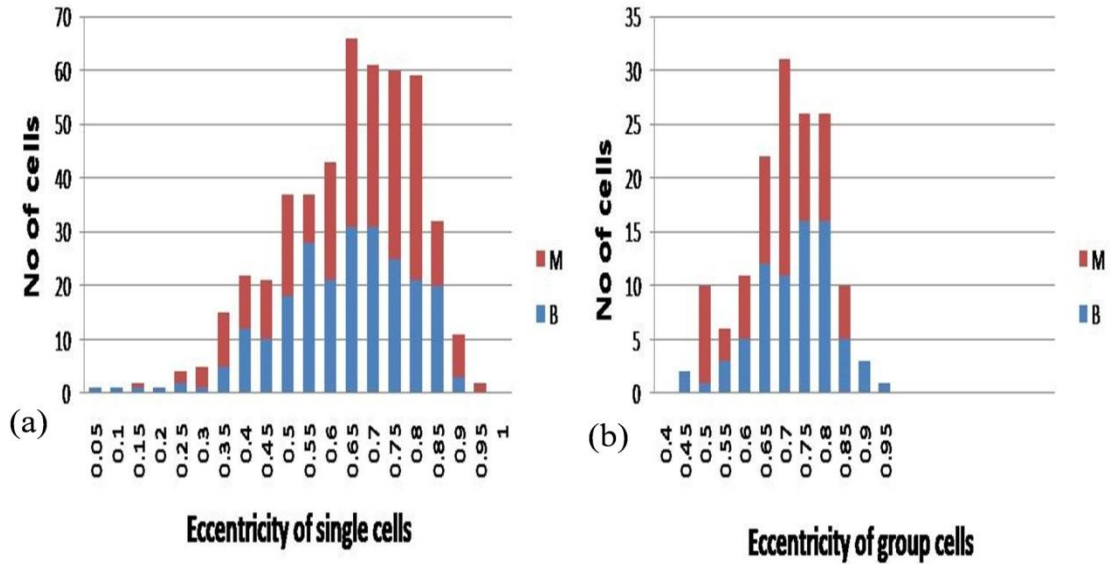


Figure 4.12: Eccentricity (F8) feature based variations of values in breast cancer cells.
 (a) Single benign (B) and malignant (M) cell dataset. (b) Group benign (B) and malignant (M) cells dataset respectively.

The figure 4.13 (a) reveals that the max intensity (F9) of single benign cells (blue colored) typically possess comparative pixel value of 92. On the contrary, figure 4.13 (b) represents single malignant cells (red colored) that exhibits diverse pixel values such as: 97,100,103,104,112,115,130,175 respectively. Furthermore, in figure 4.13 (b) the comparative pixel values of benign group cells typically possess 92 (blue colored). In contrast, the figure 4.13 (b) shows different pixel values (94, 95, 96, 98, 100, 105, 110, 115, 120, 125 and 130) (red colored) for malignant group cells respectively. Henceforth, both the figure 4.13 (a) and 4.13 (b) depicts that malignant cells exhibit diverse pixel values for max intensity in comparison to benign cells.

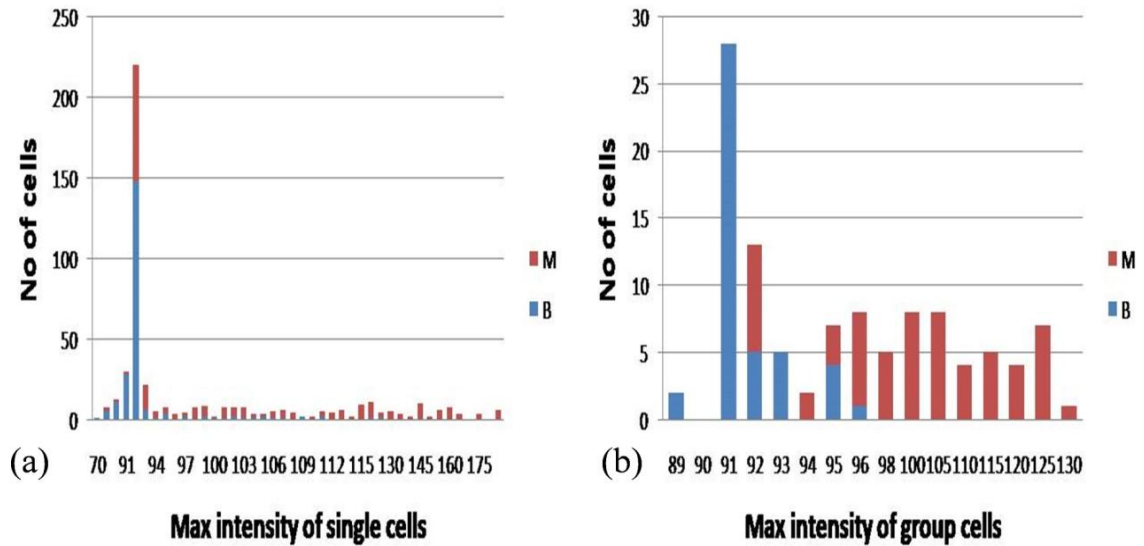


Figure 4.13: Max intensity (F9) feature based variations of values in breast cancer cells
 (a) Single benign (B) and malignant (M) cell dataset. (b) Group benign (B) and malignant (M) cells dataset.

Total 30 features included 8 morphology, 4 intensity and 18 texture features of benign and malignant cells for single cells datasets and group cells dataset are being plotted as average values graph shown in figure 4.14 to 4.17 to show the variations of average values of features in terms of variations of different features. These features are area (F1), convex area (F2), solidity (F3), perimeter (F4), major axis length (F5), minor axis length (F6), circularity (F7), eccentricity (F8), intensity features include max intensity (F9), mean intensity (F11) and texture include autocorrelation (F13), contrast (F14), cluster prominence (F16), sum of squares (F22), sum of average (F23), sum of variance (F24), sum of entropy (F25), and information measure of correlation² (F28).

Table 4.7 show various features of morphology, intensity and texture variation in terms of average values. This processing yielded significant results for benign and malignant cells based single and group cells dataset. The results presented here are expressed as Mean \pm S.D. Statistical analysis has been performed using Graph Pad Prism software (version 5.1). To perform unpaired, two-tailed students t-tests, p-value $<$ 0.05 was used for significance.

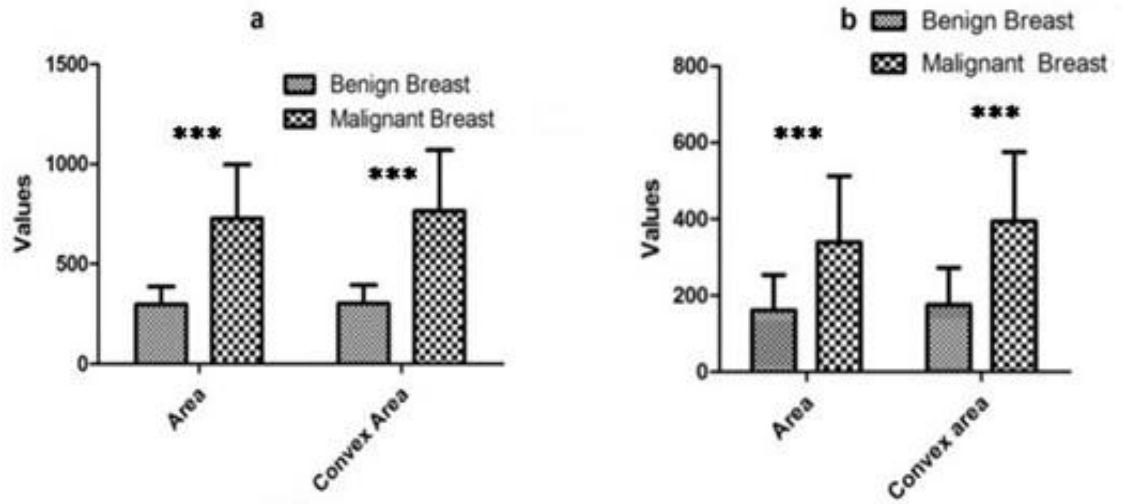


Figure 4.14: Area (F1) and convex area (F2) feature based variations of values in breast cancer cells.

(a) Single benign and malignant cell dataset. (b) Group benign and malignant cells dataset.

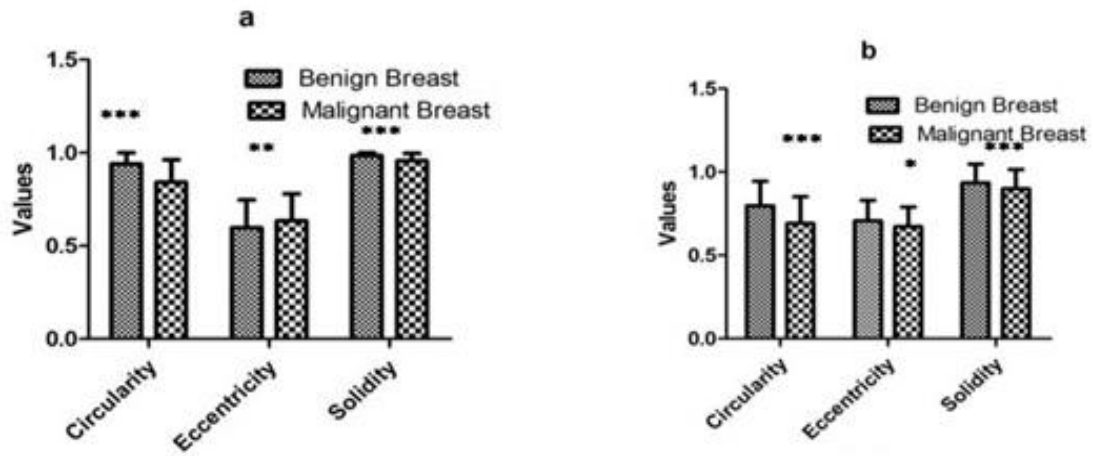


Figure 4.15: Circularity (F7), eccentricity (F8) and solidity (F3), feature based variations of values in breast cancer cells.

(a) Single benign and malignant cell dataset. (b) Group benign and malignant cells dataset.

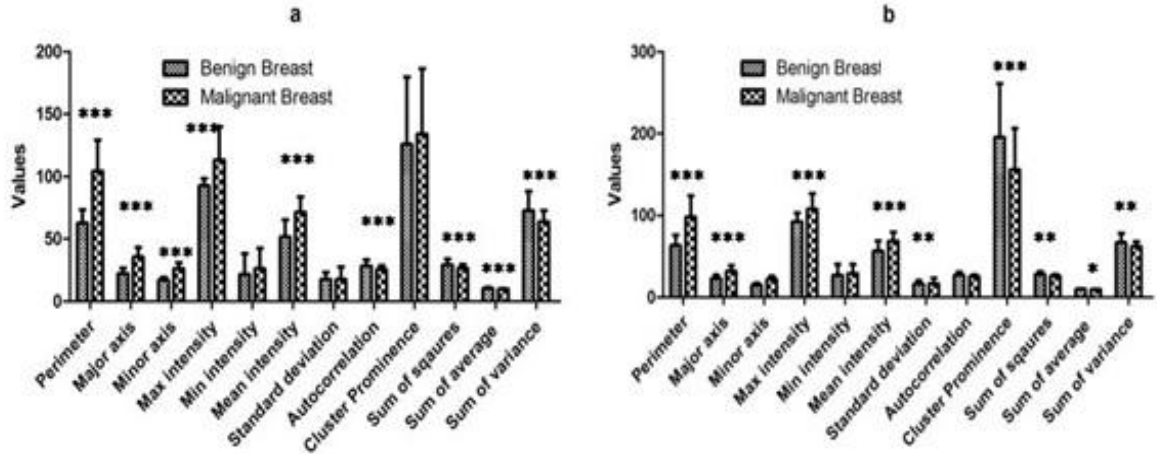


Figure 4.16: Perimeter (F4), major axis length (F5), minor axis length (F6), max intensity (F9), min intensity (F10), mean Intensity (F11), standard deviation (F12), autocorrelation (F13), cluster prominence (F16), sum of squares (F22), sum of average (F23), sum of variance (F24) feature based variations of values in breast cancer cells.

(a) Single benign and malignant cell dataset. (b) Group benign and malignant cells dataset.

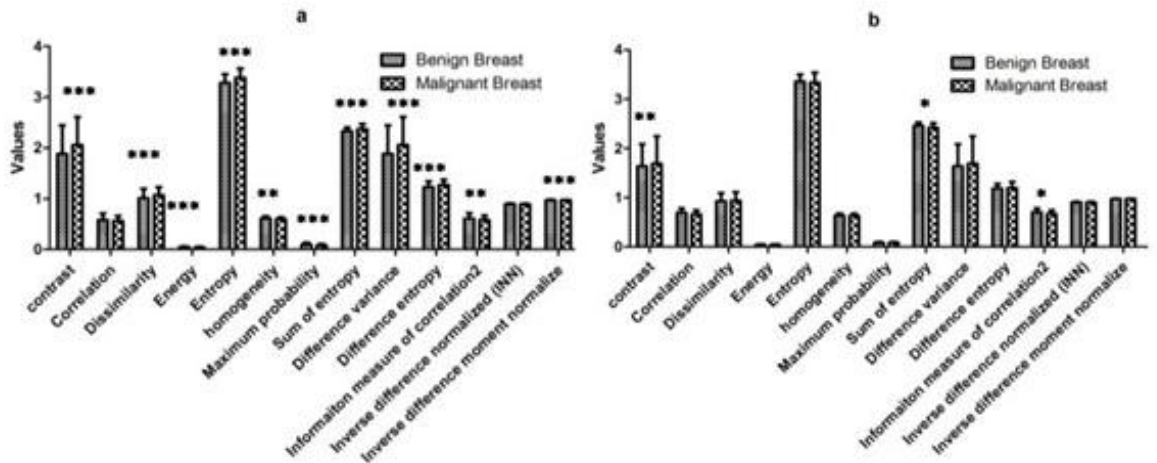


Figure 4.17: Contrast (F14), correlation (F15), dissimilarity (F17), energy (F18), entropy (F19), homogeneity (F20), maximum probability (F21), sum of entropy (F25), difference variance (F26), difference entropy (F27), information measure of correlation2 (IMC-2) (F28), inverse difference normalized (IDN) (F29) and inverse difference moment normalize (IDMN) (F30) feature based variations of values in breast cancer cells.

(a) Single benign and malignant cell dataset. (b) Group benign and malignant cells dataset.

Table 4.7: The comparative parameter of single cells and group cells dataset of benign and malignant cells of breast cancer image

S. No.	Name of features	Benign breast single cells	Malignant breast single cells	<i>p</i> -value single cells	Benign breast group cells	Malignant breast group cells	<i>p</i> -value group cells
F1	Area	298.4 ± 88.31	727.5± 270.28	<0.0001	253.41±69.72	512.32±167.19	<0.0001
F 2	Perimeter	62.71 ± 10.76	104.4 ± 24.79	<0.0001	63.48 ± 12.61	98.09 ± 26.06	< 0.0001
F 3	Convex area	303.26±91.64	765.00±305.85	<0.0001	273.06±78.40	575.29±213.06	<0.0001
F 4	Circularity	0.94±0.06	0.84±0.12	<0.0001	0.79±0.14	0.69±0.16	<0.0001
F 5	Eccentricity	0.60± 0.15	0.63 ± 0.15	<0.0096	0.71± 0.12	0.67±0.12	<0.0300
F 6	Major axis length	22.32 ± 4.49	35.61 ± 7.57	<0.0001	22.90±4.35	31.92 ±7.04	<0.0001
F 7	Minor axis length	17.04 ± 2.40	26.24 ± 4.77	<0.0001	14.76 ±2.67	21.77 ±4.10	<0.0001
F 8	Solidity	0.98 ± 0.01	0.96 ± 0.04	<0.0001	0.94 ±0.04	0.90 ±0.11	<0.0001
F 9	Max intensity	92.78± 5.70	113.49±26.5	<0.0001	92.34± 10.73	107.75±19.04	<0.0001
F 10	Min intensity	21.64± 16.64	26.33± 16.26	<0.1377	26.95±13.38	29.21±11.11	<0.2752
F 11	Mean intensity	51.65± 13.76	71.38± 12.30	<0.0001	56.39±12.54	68.63±10.98	<0.0001
F 12	Standard deviation	17.81± 5.47	17.55± 10.21	<0.8636	16.17±4.30	16.70±6.80	<0.5592
F 13	Autocorrelation	28.46±4.93	25.66±3.06	<0.0001	27.12± 3.83	25.41±1.96	<0.0052
F 14	Contrast	1.88±0.56	2.06±0.55	<0.0008	1.64±0.45	1.69±0.56	<0.5951
F 15	Correlation	0.58±0.13	0.56±0.10	<0.1288	0.69±0.09	0.66±0.08	<0.0652
F 16	Cluster prominence	125.89±53.86	134.16±52.30	<0.0986	195.62±65.53	156.03±50.30	<0.0008
F 17	Dissimilarity	1.01±0.18	1.05±0.16	<0.0001	0.93±0.16	0.94±0.18	<0.7439
F 18	Energy	0.05±0.01	0.04±0.01	<0.0001	0.04±0.01	0.04±0.01	<0.3360
F 19	Entropy	3.28±0.18	3.38±0.18	<0.0001	3.36±0.14	3.33±0.21	<0.5268
F 20	Homogeneity	0.61±0.05	0.59±0.04	<0.0013	0.63±0.04	0.63±0.04	<0.8211
F 21	Maximum probability	0.09±0.04	0.08±0.03	<0.0001	0.08±0.02	0.08±0.02	<0.8565
F 22	Sum of squares	29.24±4.72	26.54±3.03	<0.0001	27.78±3.654	26.11±1.89	<0.0042
F 23	Sum of average	10.38±0.85	9.84±0.58	<0.0001	10.02±0.69	9.74±0.39	<0.0140
F 24	Sum of variance	72.79±15.49	63.68±9.29	<0.0001	66.88±11.27	62.10±5.94	<0.0080
F 25	Sum of entropy	2.33±0.07	2.37±0.10	<0.0001	2.45±0.07	2.42±0.09	<0.0134
F 26	Difference variance	1.88±0.56	2.06±0.55	<0.0008	1.64±0.45	1.68±0.56	<0.5951
F 27	Difference entropy	1.23±0.12	1.27±0.10	<0.0001	1.18±0.09	1.19±0.12	<0.6421
F 28	Information measure of correlation 2 (IMC-2)	0.61±0.11	0.58±0.08	<0.0037	0.70±0.08	0.67±0.08	<0.0284
F 29	Inverse difference normalized (IDN)	0.89±0.02	0.89±0.01	<0.0011	0.90±0.01	0.90±0.01	<0.7824
F 30	Inverse difference moment normalizes (IDMN)	0.97±0.01	0.97±0.01	<0.0009	0.97±0.01	0.97±0.01	<0.6466

The frequency graph figures 4.7 to 4.10, average value graph figure 4.14 and 4.16 and average values present in (table 4.7) of all morphology based features related to size show that malignant single cells and group cells dataset have the greater magnitude of area (F1), convex area (F2), perimeter (F4), major axis length (F5) and minor axis length (F6) in comparison to benign single cells and group cells dataset.

The structure of malignant cells is pleomorphic with various cell sizes (Nauth, 2007; Koss *et al.*, 2005). The main reason for these outcomes is that benign cells grow and divide when they receive signals from the surrounding cells and does not exhibit contact inhibition phenomenon while malignant cells have uncontrolled cell division and grow faster (Rubin *et al.*, 2008; Chen *et al.*, 2012). Benign cells undergo aging and senescence process, as well as repair their physiological and chromosomal abnormalities (e.g. apoptosis; while malignant cells show neither repair nor induce apoptosis, also benign cells become specialized or mature so that they are able to carry out their function in the body. While malignant cells often reproduce very quickly and do not exhibit mature phenotypes (Singh, 2012; Belsare *et al.*, 2012; Madabhushi, 2009).

Further, shape based two other features as shown in the frequency graph for circularity (F7) figures 4.11 and eccentricity (F8) figure 4.12 and average value graph figure 4.15 of the cell have also been taken into consideration. Shape factor circularity is taken between 0 and 1, the circularity value is 1 then the object is a perfect circle. In our case, the average value (table 4.7) for benign cells, it is found to be nearly 1 i.e. 0.94 and in malignant cells, it is found to be nearly 0 i.e., 0.84. This shows that benign cells have a circular structure as compared to malignant cells. The eccentricity of an ellipse gives a measure of just how squashed it is. If the eccentricity is 0, it is not squashed at all and so remain a circle. If it is 1 then completely squashed and look like a line. As per consideration of eccentricity, in our case, the average value (table 4.7) for benign cells, it is found to be nearly 0 i.e. 0.5 and in malignant cells, it is found to be nearly 1 i.e. 0.7.

This shows that malignant cells have an elongated structure as compared to benign cells. Shape based features related to circularity and eccentricity show that malignant single cells and group cells dataset have the elongated shape in comparison to benign single cells and group cells dataset. The main cause of the results obtained as malignant cells image result shows generally elongated, distorted or blebs shape of cells (Nauth *et al.*, 2007). This becomes physiologically nonfunctional such types of shape is useful for malignant cells to exhibit random migration i.e. metastasis. In normal tissues, the cells stay together and adhere to each other through specific microstructures that assist in governing the cellular function (Dermir *et al.*, 2005; Guillaud *et al.*, 2005).

Our report is in agreement with the data reported by several authors (Kasmin *et al.*, 2012; Belsare *et al.*, 2012; Mulrane *et al.*, 2008) extracted the features of microscopic biopsy images including (area, perimeter, convex area, solidity, major axis length, eccentricity, the ratio of cell and nucleus area, circularity, and mean intensity) of cytoplasm. Basavanhally *et al.* (2010) quantified the morphological features that classify their structure in a histopathological slide image which leads to discrimination of a cell into a particular class for the purpose of diagnosis. Sinha *et al.* (2003) extracted some features of histopathological images that contain the area of cells, area ratio, eccentricity, compactness, and average values of color components, energy, correlation, and entropy. Guillaud *et al.* (2005) computed morphological features for quantitative analysis of CIN cervical histological images. Rahmadwati *et al.* (2012) use the morphological method for quantifying the characteristics of normal and abnormal nuclei in cervical histology images.

In concern to intensity based features for maximum intensity (F9) show that in frequency graph figures 4.13, average value graph figure 4.16 and table 4.7 pixel values are greater magnitude and diverse in malignant cells as compared to benign cells where maximum intensity pixel typically possesses comparative pixel value of 92 (values found to be almost identical) to normal cells. Mean intensity values are higher in malignant cells as compared to benign cells.

The possible reason for this observation shows the presence of high amount of deoxyribose nucleic acids (DNA) or increase amount of nucleoprotein synthesis in the malignant cells. This results in the larger nucleolus and dark-staining nuclei which referred as hyperchromatism (Thiran *et al.* 1996). Zhao *et al.* (2007) also worked on the pixel of the benign and malignant nucleus. Demir *et al.* (2005) reported that the intensity based approach is employed to calculate the intensity value of pixels to define the features in a histopathological slide image (Bhattacharjee *et al.*, 2014).

The figure 4.16, 4.17 and Table 4.7 represents analogous values for both the single cells and the group cells dataset that implies certain features such as standard deviation, minimum intensity, possess insignificant relationship and possess irrelevant features from the viewpoint of differentiation.

Further, the computed texture based features as shown in the average value graph (figure 4.16, figure 4.17 and table 4.7) can also depict the difference between benign and malignant cells. The significant features in single cell datasets include dissimilarity (F17), energy (F18), entropy (F19), homogeneity (F20), maximum probability (F21), difference variance (F26), difference entropy (F27), inverse difference normalized (IDN) (F29) and inverse difference moment normalized (IDMN) (F30). However, the all those important features are irrelevant in group cells dataset. Texture based feature are also helpful in distinguishing benign and malignant cells (Chaddad *et al.*, 2011;Haralick *et al.*, 1973; Doyle *et al.*, 2008).

Hamilton *et al.* (2007) worked on texture analysis to develop criteria for the automatic identification of colorectal dysplasia from a background through focal areas of histologically normal tissue. Mouelhi *et al.* (2013) classified the cancerous cells from histopathological images by using Haralick's textures features, color component and the histogram of oriented gradients (HOG). This is based on statistical moments (CCSM) of feature selection and extraction approaches. Morphology (size, shape), intensity and texture features have been effective in discriminating benign and malignant cells of histopathological images of breast cancer cells. Features extraction is needed to quantify the cellular changes in the tissue. Therefore, morphological features, texture features and intensity based features may be used for classification.

4.5 Conclusion:

In this chapter, the histopathological cellular image of suspected breast cancer has been analyzed using structure, intensity, and texture based morphological features. The developed algorithm for automated analysis and evaluation of histopathological images will assist the pathologists in the diagnosis of cancer and reduce the occurrence of human error. Such automated cancer diagnosis facilitates mathematical judgment to the pathologist. The future work would include more features in the algorithm for efficient differentiation between benign and malignant cancer cells so that suitable classifiers may be designed.