Microscopic Image Processing Of Automated Detection And Classification For Human Cancer Cell

Laith Muayyad Abdul-Hameed Al-Hayali, Dr. M. Morsy, M. M. Abdul-Razak

Abstract: Automated Detection for Human Cancer Cell is one of the most effective applications of image processing and has obtained great attention in latest years, therefore. In this study, we propose an automated detection system for human cancer cells based on breast cancer cells. This study was conducted on a set of Fine Needle Aspiration (FNA) biopsy microscopic images that have been obtained from the "Pathology Center - Faculty of Medicine - Mansoura University Hospital - Egypt" is made up of 72 microscope image samples of benign, 72 microscope image samples of malignant. The purpose of this study is to detect and classify the benign and malignant cells in the breast biopsy. The images are exposed to a series of pre-processing steps, which include resizing image such as 1024*1024, 512*512, enhance images by remove noise through (Median Filter) and contrast enhancement through (Unsharp Masking – Adjust Intensity). The system depends on breast cancer cells detection using clustering-based segmentation (K-means clustering, Fuzzy C-means clustering) and region-based segmentation (Watershed). Shape, Texture and Color features are extracted for Detection. The results show high Detection Rate for breast cancer cells images either (Benign or Malignant). Finally classification stage by using (Support Vector Machine, K-Nearest Neighbors and Back-Propagation Neural Networks). The final classification with the best accuracy in SVM is (97.22%), in K-NN and BPNNs is (98.61%).

Keywords : Digital Image Processing, Breast Cancer Cells, Contrast Enhancement, K-Means Clustering, Fuzzy C-Means, Watershed, Features Extraction, Classification, SVM, K-NN, Back Propagation Neural Networks.

I. INTRODUCTION

Cancer starts when cells begin to grow out of control. If the cells can grow and invade the surrounding tissues or spread to distant areas of the body the tumor is malignant. For Breast cancer which is our concern occurs almost entirely in women, but men could be vulnerable, too. Breast cancers can start from different parts of the breast. In breast cancer primary tumor, a desmoplastic reaction usually arises, creating a microenvironment which is suitable for a cross talk between stromal fibroblasts and malignant cells. A few studies have explored the epithelial–mesenchyme interactions through co-culture systems and their results suggest that fibroblasts originated from normal tissue tend to inhabit, contrary to fibroblasts obtained from tumors, which tend to motivate, epithelial cell proliferation. In addition, it seems likely that direct contact between these cell types or soluble factors secreted by them may differentially interfere with epithelial cell proliferation [1-4].

II. DIGITAL IMAGE PROCESSING (METHODS)

The procedure of Breast Cancer Cells (BCCs) detection in microscopic images consists of pre-processing (resize, remove noise and contrast image), segmentation using (k-means clustering, fuzzy c-means and watershed), feature extraction (shape – texture – color), detection will display the Benign and Malignant cells.

- *Laith Muayyad Abdul-Hameed Al-Hayali, AL Mansoura University Faculty of Engineering Electronics & Comm. Dept., Mansoura - Egypt.*
- *Dr. M. Morsy, AL Mansoura University Faculty of Engineering Electronics & Comm. Dept., Mansoura - Egypt.*
- *Prof. M. M. Abdul-Razak, AL Mansoura University Faculty of Engineering Head of Electronics & Comm. Dept. ,Mansoura - Egypt.*

Finally, we apply three classification algorithms which are (SVM – K-NN – BPNNs) to get best performances. The proposed system is shown in Fig. 1.

A. Pre-processing

Pre-processing methods can be divided into the three groups according to the goal of the processing:

A.1. Image resizing

Resizing images in database by using several sizes such as 512 * 512 and 1024 * 1024.

A.2. Remove Noise

Digital images are exposed to a various types of noise. Noise is the consequence of blunders in the image acquiring process that outcome in pixel values that don't mirror the intensities of the genuine scene. There are several ways that noise can be introduced into an image, by using:

A.2.1 Median Filter:

The best-known order-statistics filter is the median filter, which replaces the value of a pixel by the median of the gray levels in the neighborhood of that pixel. The original value of the pixel is included in the computation of the median. Median filters are quite popular because, it preserves edges while removing noise, with quite less blurring than linear smoothing filters of similar size [5][6].

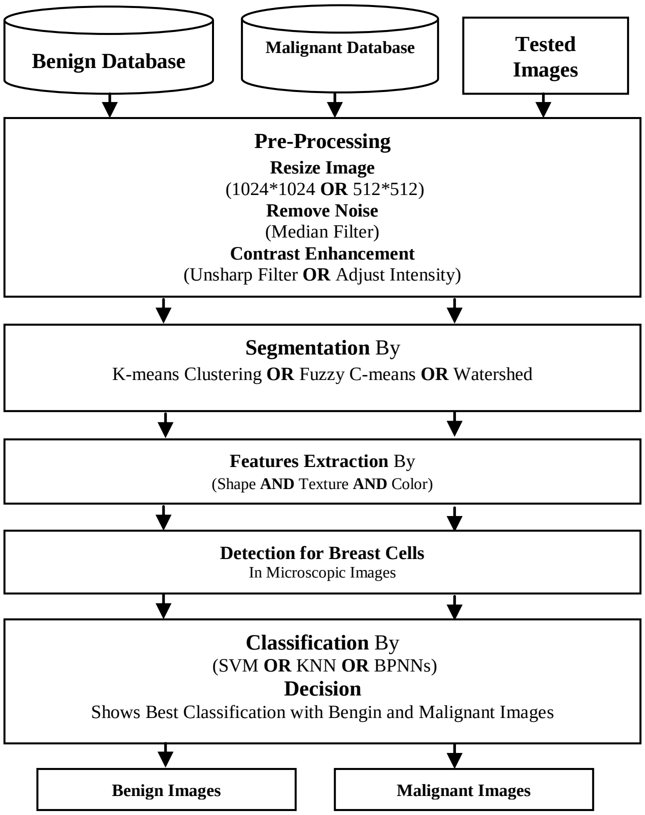


Fig.1. System Overview

A.3. Contrast Enhancement
That is a technique to improve the quality of a color image by using:

A.3.1. Unsharp Masking:
Unsharp masking enhancement process is one of the most widespread algorithms used for image enhancement [7], unsharp masking is a flexible and powerful way to increase sharpness

A.3.2. Adjust Intensity:
Intensity adjustment is a technique for mapping an image's intensity values to a new range [8].

B. Segmentation
Segmentation is to perform the detection of Breast Cancer Cells (BCCs) from the microscopic images using color based clustering. Initial segmentation can achieve by K-means clustering, Fuzzy C-means and Watershed.

B.1. K-means clustering
K-means clustering followed by nearest neighbor classification in L*a*b* space. K-means is a semi supervised clustering technique, which is a use to create K clusters from n observations. It attempts to achieve partition such that objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible [9]. Object's each pixel is classified into four clusters based on the corresponding a* and b* values in L*a*b* color space. In order to overcome the undesirable overlapping of regions, a second stage segmentation is performed using nearest neighbor classification. In the second stage we select a sample region randomly from each of the four clusters obtained using K–means. The mean color of each sample regions are calculated in a*b* space and those values act as color indicators. Now each pixel in the L*a*b* space is classified into any of the four classes by calculating the Euclidean distance between that pixel and each color indicator. Each of the whole image pixels will be labeled to a particular color depending on the minimum distance from each indicator [10].

B.2. Fuzzy C-means clustering
Fuzzy c-means (FCM) is a data clustering method in which a dataset is gathered into n clusters with each data point in the dataset having a place in each cluster with a specific degree. For example, a certain data point that lies near the center of a cluster will gain a high rank of belonging or membership to that cluster and data point that lies far away from the center of cluster will reduce its rank of belonging or membership to that cluster.

B.3. Watershed
The main goal of watershed segmentation algorithm is to find the "watershed lines" in an image in order to separate the distinct regions. A watershed definition for the continuous case can be based on distance functions. Relying on the distance function used, one may arrive at different definitions. but other choices have been proposed as well [11]. Segmentation using the watershed transforms works well if you can identify and "mark" the foreground objects and background locations.

C. Morphorgical Filter
Morphological Filtering: An image is partitioned into several regions depending on the features to be extracted. Image enhancement ensures that perceptibility and visibility of these regions gets improved [12] Many of the above enhancing techniques were employed on the breast tissue specimen that is obtained from biopsy. Some of the methods were[13]:

Sobel: it is a discrete differentiation operator, processing an approximation of the gradient of the image intensity function. At every point in the image, the aftereffect of the Sobel operator is either the corresponding gradient vector or the norm of this vector. The Sobel operator depends on convolving the image with a small, separable, and integer valued filter in horizontal and vertical direction and is therefore relatively inexpensive in terms of processing.

Dilation: Dilation is the way toward adding pixels to the edges of objects in an image, while erosion is the way toward eliminating them. The parameter called structuring element decide the quantity of pixels to be added or eliminated from the objects. In the morphological dilation operation, the condition of any given pixel in the yield image is dictated by applying a standard to the comparing pixel and its neighbors in the input image. The guideline used to handle the pixels characterizes the operation as dilation [14].

D. Feature Extraction:
Feature extraction in image processing involves reducing the amount of resources required to describe a large set of data. In the present paper, broadly three types of features are extracted (Shape, Texture and Color Features)

1) Shape Feature: According to pathologist which describes how the specimen looks, including the shape, color, size, and other features. Region and boundary based shape features are extracted and analyzed for the shape of cancer cells. All the features are extracted from the binary equivalent image of the cancer cells with non-zero pixels representing the cancer cells region. The quantitative evaluation of each breast cancer cells images are do using the extracted features under two classes i.e. region based and boundary based. The features are as follows:

- Area: The area was determined by counting the total number of non-zero pixels within the image region.
- Perimeter: It was measured by calculating distance between sequent boundary pixels.
- Compactness: Compactness or roundedness is the measure of a cells as defined in (1).
-

$$Compactness = \frac{Perimeter^2}{Area} \qquad (1)$$

- Solidity: The actual area divided by the convex hull area is known as solidity and is also an essential feature for cancer cell classification. This measure is defined in (2).

$$Solidity = \frac{Area}{ConvexArea} \qquad (2)$$

- Eccentricity: This parameter is used to measure how much a shape of an uncontrolled growing cells deviates from being normal. It's an important feature since normal cells are in a uniform way than that in cancer cells. To measure this a relation is defined in (3).

$$Eccentricity = \frac{\sqrt{a^2 - b^2}}{a} \qquad (3)$$

where "a" is the major axis and "b" is the minor axis of the equivalent ellipse representing the cells region.

- Elongation: is measured in terms of a ratio called elongation. This is defined as the ratio between maximum distance (R_{max}) and minimum distance (R_{min}) from the center of gravity to the cancer cells and is given by (4).

$$Elongation = \frac{R_{max}}{R_{min}} \qquad (4)$$

where R_{max} and R_{min} are maximum and minimum radii respectively.

- _ Formfactor: This is a dimensionless parameter which changes with surface irregularities and is defined as (5).

$$Formfactor = \frac{4 * pi * Area}{Perimeter^2} \qquad (5)$$

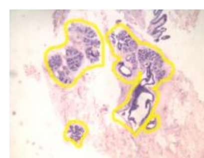
2) Texture Feature: Cell texture measurements were performed on gray scale version of the breast cancer images. These features were computed from the co-occurence matrices for each breast cancer image. This includes:

- Homogeneity: It is a measure of degree of variation.
- Energy: Is used to measure uniformity.
- Correlation: This represents correlation between pixel values and its neighborhood.
- Entropy: Usually used to measure the randomness.
- Contrast: It is a measure of the intensity contrast between a pixel and its neighbor over the whole image.

3) Color Feature: Since color is an important feature that human comperhend while visualizing it is considered for extraction from cancer cell regions. Hence, for each breast cancer cells image the mean color values in RGB is found suitable for feature extraction [15].

E. **Detection**:

The purpose for detection the stage, that comparing the results of the segmentation process resulting from the three algorithms images of our existing work that has been segmented by an expert and to detect the affected areas and non-infected areas. The following figure 2 shows the result of the segmentation process and detection of the breast cancer cells with size 1024*1024 and median with unsharp masking



Detection by Expert

Detection by K-Means Detection by C-Means Detection by Watershed

Fig.2 Shows result for detection by using K-Means, C-Means and Watershed

F. Classification:

Classification is the task of assigning to the unknown test vector, a label from one of the known classes. Image classification analyze the numerical properties of various image features and organize data into categories. Classification algorithms typically employ two stages of processing: training and testing. In the underlying training stage, characteristic properties of an ordinary image features are isolated and, based on these, a unique description of each classification category, i.e. training class, is made. In the subsequent testing stage, these feature-space partitions are used to classify image features.

F.1. Support Vector Machine (SVM)
In our framework, we used the Support Vector Machine (SVM) classifier that represents the features as points in the space. The SVM classifier aims to find the optimal hyperplane that separates the data features into separate classes[16].

F.2. K-Nearest Neighbors (K-NN)
Its a supervised learning method. Medical diagnosis using the K-NN is very interesting. The K-NN is a method based on the nearest training examples in the feature space for classifying the objects, K-NN is among the simplest of all Machine Learning (ML) algorithms: a majority vote of its neighbors is responsible for classyfying an object, with the object being assigned to the class most common amongst its k-nearest neighbors (k is a positive small integer). If k = 1, then the object is simply assigned to the class of its nearest neighbor. For largest value of k , the algorithm assigns the most common value among the k nearest training examples. Ties can be broken at random. The k – nearest neighbor algorithm used neighborhood classification as the prediction value of the new query instance. The neighbors have taken from a set of objects for which the correct classification is known. Basically, an object is classified by the “distance” from its neighbors, with the object being assigned to the class most common among its k distance-nearest neighbor. Distance is a key word in this algorithm. Each object in the space is represented by position vectors is a multi-dimensional features space and the Euclidean distance is used to calculate distance between two vector positions[17].

F.3. Back-Propagation Neural Network (BPNNs)
One of the most popular method of training artificial neural networks used in conjunction with an optimization method such as gradient descent. The method computes the gradient of loss function with respect to all weights in the network. To update the weights, a gradient which fed to the optimization method is used, in aim to minimize the loss function. Back propagation (BP) in order to compute the loss function gradient, it requires a knowledge about the desired output for each input value. A supervised learning method for this is usually considered, although it is as well used in some unsupervised networks such as auto encoders. It is a generalization of the delta rule to multi-layered feedforward networks, made possible by employing the chain rule to iteratively calculate gradients to each layer. Backpropagation requires that the activation function used by the artificial neurons (or "nodes") be differentiable.

III. EXPERIMENTAL RESULT
The proposed technique has been applied on 142 Breast Cancer Cells images obtained from “Pathology Center - Faculty of Medicine - Mansoura University Hospital - Egypt” is made up of 72 microscope image samples of Benign , 72 microscope image of Malignant.

A. Experiment
The experiment work of proposed system consist several steps, all image are preprocessed by MATLAB 2014 platform to resize and remove noise and then contrast enhancement. Then the computer aided detection system segments all dataset breast cancer cells images (benign and malignant) using clustering-based segmentation by (K-means Clustering, C-means Clustering) and region-based segmentation by (Watershed) , and we extract all features (Shape – Texture – Color). Finally, we used two algorithms for classification, that (SVM, K-NN and BPNNs).

B. Result Analysis for Detection
The experimental result has been developed by taking the entire dataset breast cancer cells image. The entire dataset images are gone through the Preprocessing – Segmentation – Features Extraction for detection the breast cancer cells. Now we will review the results to detect the microscopic images of the breast cancer cells through three algorithms (K-Means, C-Means, and Watershed) used to gain access to the best result in the Aided –Computer Detection, as in the following Figure [3]. **Figure 3** Shows Detection Rate by using segmentation algorithms (K-Means, C-Means and Watershed) for dataset Benign and Malignant image that (72 images for every one) with the size 512*512 and the applied the remove noise filter and contrast enhancement techniques.

Segmentation Algorithms	K-means		C-Means		Watershed	
Size Images	512*512		512*512		512*512	
Remove Noise & Contrast Enhancement	Median & Unsharp	Median & Intesity	Median & Unsharp	Median & Intesity	Median & Unsharp	Median & Intesity
Detection Rate Benign	90.2%	87.5%	87.5%	77.7%	83.3%	76.8%
Detection Rate Maligant	88.8%	87.5%	86.1%	76.8%	81.9%	77.7%

Fig .3 Shows Detection Rate for Benign and Malignant images with size is 512*512

Figure 4 Shows Detection Rate by using segmentation algorithms (K-Means, C-Means and Watershed) for dataset Benign and Malignant image that (72 images for every one) with the size 1024*1024 and the applied the remove noise filter and contrast enhancement techniques.

Segmentation Algorithms	K-means		C-Means		Watershed	
Size Images	1024*1024		1024*1024		1024*1024	
Remove Noise & Contrast Enhancement	Median & Unsharp	Median & Intesity	Median & Unsharp	Median & Intesity	Median & Unsharp	Median & Intesity
Detection Rate Benign	98.60%	95.80%	97.20%	97.20%	97.20%	93.05%
Detection Rate Maligant	97.20%	95.80%	97.20%	95.80%	97.20%	91.60%

Fig .4 Shows Detection Rate for Benign and Malignant images with size is 1024*1024

Through **figure 3** and **4**, we find that: First, was the best detection of the images segmented process, that the results of a 1024 * 1024 sizes better than the image size 512 * 512 as shown previous tables where the ratio was better detection the filter Median in the noise removal with a filter (Unsharp and Intensity) to contrast enhancement. In addition to, we get high detection as shown in figure (4).

C. Result Analysis for classifiction

Here we used three algorithms they (SVM – K-NN –BPNNs) to make the classification process after ignoring the noise filter Wiener to remove its impact on segmentation stage. He was relying on the proposed system on the filter Median to remove noise and contrast enhancement, which (Unsharp - Intensity). It will be the process of classification algorithms for the three segmentation algorithms, that (K-means – C-Means - Watershed) and the addition of sizes images of the database (512 * 512 and 1024 * 1024). We will review the results of microscopic breast cancer cell images classification, whether benign or malignant. Through the available database is consisting of 144 images, which (72 benign and 72 malignant Images). It was developed half a database for training and the other half of the test in order to make a classification. Images for training, that (36 benign images - 36 malignant images) as well as test images for classification, that (36 benign images - 36 malignant images). Table 1 and 2. Shows the result with Median Filter for remove noise from images and contrast enhancement with Unsharp Masking and Adjust Intensity. We use sizes for microscopic images (512*512 – 1024*1024), then we apply three different algorithms for segmentation to detect breast cancer cells, that (K-Means – C-Means – Watershed). Moreover, features are extracted (Shape, texture and Color features). In addition, classification with SVM, K-NN and BPNNs algorithms. Finally we show the accuracy for classification as shown in next tables

	Segmentation Algorithm	K-means		C-Means		Watershed	
	Size Image	512*512		512*512		512*512	
	Remove Noise & Contrast Enhancement	Median & Unsharp	Median & Intensity	Median & Unsharp	Median & Intensity	Median & Unsharp	Median & Intensity
SVM	Accuracy	87.50%	87.67%	83.33%	81.94%	86.11%	83.56%
KNN	Accuracy	90.27%	88.88%	88.88%	86.11%	88.88%	88.88%
NN-BP	Accuracy	93.05%	91.66%	90.27%	88.88%	91.66%	88.88%

Table 1 : Accuracy For size 512*512

	Segmentation Algorithm	K-means		C-Means		Watershed	
	Size Image	1024*1024		1024*1024		1024*1024	
	Remove Noise & Contrast Enhancement	Median & Unsharp	Median & Intensity	Median & Unsharp	Median & Intensity	Median & Unsharp	Median & Intensity
SVM	Accuracy	97.22%	95.83%	94.44%	90.27%	93.05%	91.66%
KNN	Accuracy	98.61%	97.22%	95.83%	94.44%	95.83%	94.44%
NN-BP	Accuracy	98.61%	95.83%	97.22%	95.83%	95.83%	94.44%

Table 2: Accuracy For size 1024*1024

Finally, by comparison the previous tables (1,2) , we get the best accuracy rate when we follow these steps : Resize images (1024*1024) – Used Median Filter for remove noise from microscopic images – Used Unsharp Masking for contrast enhancement for microscopic images and used K-means for segmentation to detect breast cancer cells. When we used SVM as classifier, we get the high accuracy rate with value (97.22%) and when we used KNN and BPNNs as classifier, we get the high accuracy with (98.61%)

IV. CONCLUSION

This paper has present Computer Aided Detection system to detect breast cancer cells in microscopic images either (Benign – Malignant) using database 72 Benign images and 72 Malignant images. The experiments has been performing in computer type Dell, Processor Intel(R) core (TM) i5-4200 CPU @ 1.60GHz, Memory (RAM) 6.00 GB, System type 64-bit Operating System. This paper discussion multi technique in preprocessing and segmentation stage to detect the breast cancer cells for microscopic images. The system achieved better Detection Rate. We applied three classification algorithms, that (SVM-KNN –BPNNs) to get the best accuracy rate. We achieved high accuracy is value (97.22%) with SVM and (98.61%) with K-NN and BPNNs classifier.

References

[1] Shekhar MPV, Wrdell J, Santner SJ, Pauley RJ, Tait L. "Breast stroma plays a dominant regulatory role in breast epithelial growth and differentiation: implications for tumor development and progression.", Cancer Res, Vol: 61, No: 4, PP:1320-1326,2001.

[2] Sadlonova A, Novak Z, Johnson MR, Bowe DB, Gault SR, Page GP, et al. "Breast fibroblasts modulate epithelial cell proliferation in three-dimensional in vitro co-culture", Breast Cancer Res ,Vol : 7, No:1, PP: 46-59 , 2005.

[3] Kunz-Schughart LA, Heyder P, Schroeder J, Knuechel R. "A heterologous 3-D coculture model of breast tumor cells and fibroblasts to study tumor-associated fibroblast differentiation.", Exp Cell Res, Vol : 266, No:1, PP: 74-86, 2001.

[4] Hu M, Yao J, Carroll DK, Weremowicz S, Chen H, Carrasco D, et al. "Regulation of in situ to invasive breast carcinoma transition.", Cancer Cell, Vol:13, No:5, PP:394-406,2008.

[5] R. Yang, L. Lin, M. Gabbouj, J. Astola, and Y. Neuvo, "Optimal Weighted Median Filters Under Structural Constraints," IEEE Trans. Signal Processing, Vol. 43, PP. 591-604, Mar 1995.

[6] Pei-Eng Ng and Kai-Kuang Ma, "A Switching Median Filter with BDND for Extremely Corrupted Images", IEEE Trans Image Processing, Vol. 15, No. 6, PP. 1506-1516, June 2006.

[7] M. Fischer, J. L. Paredes, G. R. Arce "Image sharpening using permutation weighted median". In Proc. European Signal Processing Conference (EUSIPCO '00) Tampere, Finland, September 2000

[8] C. M. Tsai and Z. M. Yeh,"Contrast enhancement by automatic and parameter-free piecewise linear trasformation for color images", IEEE Trans. On Consumer Electro., Vol: 54, No: 2, PP: 213-219, 2008.

[9] K. S. Ravichandran and B. Ananthi. "Color Skin Segmentation using K–Means Cluster." International Journal of Computational and Applied Mathematics, Vol: 4, No: 2, PP:153 – 157, 2009.

[10] S. Mohapatra and D. Patra, "Automated Cell Nucleus Segmentation and Acute Leukemia Detection in Blood Microscopic Images ," International Conference on Systems in Medicine and Biology. Vol: 5, No: 1, PP: 16-18 December 2010, IIT Kharagpur, India

[11] Najman, L., and Schmitt, M." Watershed of a continuous function". Signal Processing, Vol: 38 ,PP: 99–112,1994

[12] Preteux, F. "On a distance function approach for gray-level mathematical morphology". In Mathematical Morphology in Image Processing, Vol:10, PP: 323–349, 1993.

[13] Diaz, R., S.Guichard, et. al., "Segmentation of Bone Marrow Cell Images for Morphological Classification of Acute Leukemia", Paper presented at the Twenty-Third International Florida Artificial Intelligence Research Society Conference, Florida,USA.,2010.

[14] H. J. A. M. Heijmans: "Morphological Image Operators", Academic Press, New York.1994

[15] Anil Z Chitade., "Color Based image Segmentation using K-Means Clustering", IJEST: International Journal of Engineering Science and Technology, Vol.2, 2010.

[16] Dhoha Abid," Segmentation of Tumor Regions in Microscopic Images of Breast Cancer Tissue", thesis for master degree, PP: 23,2016.

[17] M. Sarkar and T.Y. Leong, "Application of K-Nearest Neighbors Algorithm on Breast Cancer Diagnosis Problem" AMIA, 1067-5027, PP: 759-736,2000.