CHAPTER 4: MICROSCOPIC BIOPSY IMAGES SEGMENTATION USING HYBRID COLOR K-MEANS APPROACH

The color image segmentation is a fundamental requirement for microscopic biopsy image analysis and disease detection. In this chapter, a hybrid combination of color k-means and marker control watershed based segmentation approach is proposed to be applied for the segmentation of cell and nuclei of microscopic biopsy images. The proposed approach is tested on breast cancer microscopic data set with ROI segmented ground truth images. Finally, the results obtained from proposed framework are compared with the results of popular segmentation algorithms; Fuzzy c-means, color k-means, texture based segmentation as well as adaptive thresholding approaches. The experimental analysis shows that the proposed approach is providing better results in terms of accuracy, sensitivity, specificity, FPR (false positive rate), Global consistency error (GCE), Probability random index (PRI), and variance of information (VOI) as compared to other segmentation approaches.

4.1 Introduction

Segmentation of microscopic biopsy images is typical task due to its complex nature. The cellular components of the microscopic biopsy images are segmented on the basis of intensity, color, shape, and texture features. The microscopic biopsy has several segmentation issues and challenges like high quality segmentation with low computational cost. There are many approaches which are reported in the literature for the segmentation of microscopic biopsy images (Pal, N. R., & Pal, S. K. (1993)) but most of the segmentation algorithms are time consuming, and even some times not accurate up to the mark. Some of the popular segmentation approaches are adaptive thresholding (Pal, N. R., & Pal, S. K.,1993) watershed based approach (Beucher S, Lantu´ejoul C (1979).), color k-means (Ray, S., & Turi, R. H. (1999)), fuzzy c-means (Lim, Y. W., & Lee, S. U. (1990).) and texture based segmentation (Lorigo, L. M,et al (1998)), and every approach has its own advantages and drawbacks. Thresholding approaches are roughly categorized into local, global, and adaptive thresholding approaches. The adaptive thresholding is the most popular approach for creating the binary images. The computational complexity of thresholding approaches are low but its losses the colour information of microscopic biopsy images. Therefore, it does not use alone for the segmentation of microscopic biopsy images. The watershed segmentation algorithms is an example of region based segmentation approach. It is popular segmentation from the field of mathematical morphology and proposed by (Boucher *et al*.,1979) the watershed can be in intuitively thought as a landscape or topographic relief which is flooded by water, and watersheds are the dividing lines of the domains of attraction of rain falling over the region. The intensity value is represented by the height of each point. The gradient of the original image is taken as input of the watershed transform. However, catchment basin boundaries are located at high gradient points (Pal, N. R., & Pal, S. K. 1993). The watershed transform has good properties that make it useful for many different image segmentation applications: it is simple, easy to implement and intuitive. It can also be parallelized (Beucher S, Lantu´ejoul C , 1979) and always produces a complete division of the image. The major drawback of the watershed segmentation includes over segmentation and sensitive to the false edges. Some works reported in literature for the segmentation of microscopic biopsy images are described as follows; a novel synergistic boundary and region- based active contour model were presented by authors, to demonstrate for level set formulation with automated initialization based on watershed. Digitized histopathology images of breast, and prostate biopsy specimens are demonstrated as an application of these synergistic active contour models using multiple level sets, to segment nuclear and glandular structures (A. Mouelhia *et al,* 2013).The qualitative and quantitative evaluation were tested on 100 prostate and 14 breast cancer histology images for the task of detecting and segmenting nuclei, and lymphocytes reveals that the model easily outperforms two state of the art segmentation schemes ,and on average is able to resolve up to 91% of overlapping/occluded structures in the images. The watershed based algorithm, and concave vertex graph approaches were proposed for touching nuclei to perform accurate quantification of the different stains. Two datasets of breast cancer cell images containing different level of malignancy were tested on these segmentation algorithms. The segmentation accuracy in term of cancer nuclei number is over than 97%, reaching an improvement of 3–4% over earlier methods for the complete image database.

4.2 Related works

A blood vessels segmentation using watershed based approach is presented by (Rodríguez *et al.,* 2005) The number of blood vessels present in an image is easily identified by this approach. The results obtained by the approach are compared with manual segmentation of blood vessels with 10% false positive and 0 % false negative. Automated cell nuclei segmentation for breast tissues is presented by (Georgeet *et al*., 2013) The circular Hough transform is proposed for the cell, and nuclei segmentation form breast microscopic biopsy. The noisy circles and blood cells are extracted by Otsu's thresholding the fuzzy c-means approach. The corresponding sensitivity, and specificity obtained is 95.49 %, and 83.16 % respectively. Xu, J. *et al., (2010*) presented a weighted normalized cut active contour model for segmentation of 60 microscopic biopsy of prostate cancer(Xu, J. *et al* . ,2010). Total accuracy obtained by weighted normalized cut active contour model is 84%. Sertel, O. *et al.,* (2009) proposed a component wise thresholding approach for segmentation of neroblastic (NB) biopsy images. The sensitivity, and false positive rate obtained by this approach are 81.14%, and 12.2% respectively. Demir, C. *et al.,* (2005), presented object graph approach for the segmentation of colon gland microscopic images and average accuracy obtained by this approach is 85 % . Kong, H. *et al.,* (2011) presented supervised color-texture segmentation of cell and nuclei of follicular lymphoma and total error rate reprted in this approach is 5.25% . Dundar, M. M. *et al.,* (2011) proposed a Gaussian mixture model based segmentation of breast microscopic biopsy. The average accuracy reported in this chapter is 87.9%. Basavanhally, A. *et al.,* (2011) presented color gradient based active contour model (CGAC) and hieratical normalized cut approach for the segmentation of cell and nuclei of H&E strain microscopic biopsy of breast tissue. The average reported accuracy in this paper is 86% . Tosun, A. *et al.,* (2011), proposed a graph run length matrix based approach for the segmentation of microscopic biopsy images of breast sample. In this paper 150 images were taken for testing and experimentation purpose. The average accuracy, specificity, and sensitivity obtained by this approach are 94.7%, 95.2% and 92 % respectively. Table 4.1 shows the comparison of various segmentation approaches with proposed approach.

The proposed frame work of this chapter is organized as follows: the section first presents the introduction and brief literature, for the segmentation of microscopic biopsy images. The section second represents the materials and methods of proposed approach, the section four represents the results and discussions, and finally section five provides the conclusion and future works.

4.3 Materials and Methods

The proposed framework is implemented on MATLAB 2013b on data set of 58, Hematoxylin and Eosin (H&E) strain microscopic biopsy images of the breast tissues (http://www.bioimage.ucsb.edu/images/stories/BioImage/research/Benchmark/BREAST_CANCER/BreastCancerCell_dataset.tar.gz), on the PC with 3.4 GHz Intel Core i7 processor, 2GB RAM, and windows7 platform. The microscopic biopsy data set contains 31 benign and 27 malignant images of 896×768. The ROI (region of interest) selected ground truth of all 58 images are also available for this data set. For testing and experiment purpose, the various segmentation algorithms are compared with proposed frame work for cell and nuclei segmentation from microscopic biopsy. Finally, the ROI segmented image of microscopic biopsy is compared to ground truth images. The quantitative evaluation of various segmentation approaches for all 58 sample images are performed. The performance of the various segmentation approaches such as color K-means (Ray, S., & Turi, R. H. ,1999) fuzzy c-means (A. Mouelhia *et al.,,* 2013) texture based segmentation (Sumi, M. A. *et al.,* (2010).) were evaluated in terms of various popular parameters of segmentation measures. These parameters include accuracy, sensitivity, specificity, false positive rate (FPR), probability random index (RI), global consistency error (GCE), and variance of information (VOI).

In proposed approach at first, original image is used to get clustered image using color k-means approach. The color k-means clustering provides primary segmentation. The color k-means clustering is simple to implement, and it has low computational cost as compared to Fuzzy c-means (FCM). After getting clustered image provides as output of k-means, Sobel filter is applied on clustered image to get smooth and differencing edges of the cell, and nuclei of microscopic biopsy images. In place of Sobel filter, the canny edge detection operations can be used, however canny edge operator has high complexity; on other hand Sobel operator provides both differencing and smoothing effect of the image. Finally, the resultant segmented image is obtained by using marker control watershed approach. The flow diagram of proposed framework is illustrated in Figure 3.1.

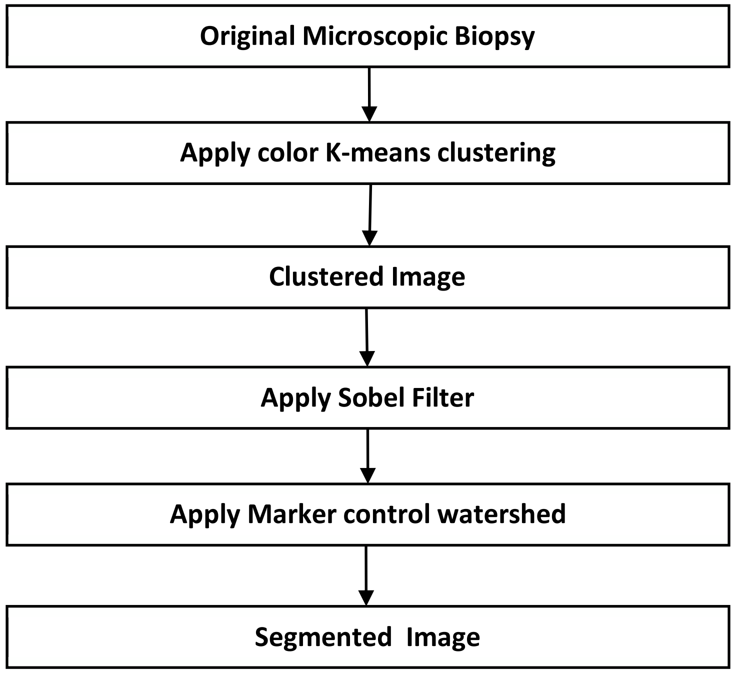

Figure 4.1: Frame work for hybrid color k-means image segmentation

4.3.1 **K-means Segmentation:** The k-means is unsupervised clustering based algorithm to provide a basic segmentation of images. In microscopic biopsy images of breast tissues, there are many regions with the similar intensity, and color. Every region has its own local minima, that increases the over segmentation by directly applying the watershed segmentation approach. Thus, due to simple and low complexity properties of k-means approach is suitable for the segmentation of microscopic biopsy images.

In clustering based approaches, there is need of prior knowledge to the number of clusters, before processing to classify the pixels. A similarity measure is defined among the pixels are grouped together to form a new clusters. The Euclidean distance metric is used to measure the difference between various colors. K-means clustering treats each object as having a location in space (Kumar, R., Srivastava, R., & Srivastava, S. 2015). The color k-means segmentation find the partitioned in such a way that objects of the same clusters are as closer as possible to each other, on other hand the object of the other clusters as for as possible. The main drawback of k-means clustering is that to specify the number of clusters to be partitioned, and a distance metric to quantify how close two objects are to each other. Since, the color information exists in the 'a*b*' space, and number of color components are pre specified on microscopic biopsy image for pixels with 'a*' and 'b*' values.

The microscopic biopsy of beast tissues consist of regions representing the cell and nuclei (blue) , connective tissues (pink), and background (white). The value of k is chosen as 3 for k-means segmentation because in H &E strain microscopic biopsy image, the three colors pink, blue, and white are associated with various components. For example, pink is associated with connective tissues, blue is associated with cell and nuclei, and white is associated as other background components. In the first pass of the data the initial cluster is chosen. The data set is partitioned into k clusters, and assigned to the clusters that have the similar number of pixels. For each pixel, the Euclidian distance is calculated from the pixels to the mean of the cluster. If the pixel is not near to its cluster, it will have to be shifted in to nearest cluster. If pixel is already nearest to its own cluster, it will not shift. The whole process is repeated until cluster mean, do not shift more than a given threshold value or number of iteration is reached. Figure 2 shows, the segmentation results after the color k-means clustering is applied on it.

4.3.2 **Watershed segmentation:** If foreground and background of the image is easily identifiable, watershed based segmentations are well suited in such type of applications (Ng, H. P *et al.,* 2006)).The gradient magnitude of primary segmentation is obtained by applying Sobel filter. The marker control watershed segmentation includes following basic steps:

- A segmentation function trying to segment image, whose dark regions are cell and nuclei to be segmented.
- The foreground marker are computed the connected tissues of the pixels with in the image.
- Calculating the background marker and the pixels that are not parts of any object are selected.
- The segmentation is modified in such a way that only one minimum should be presents at the foreground and background marker location.
- Finally, from the modified segmentation function the watershed transform are performed to obtain the segmented image.

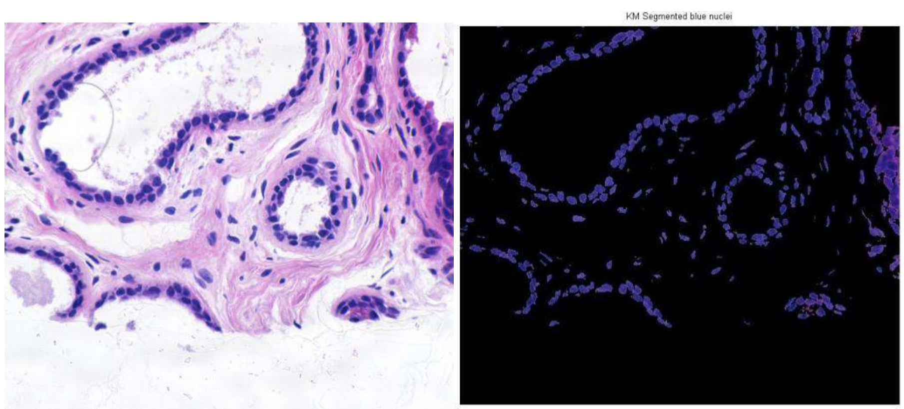

Figure 4.2: Original (left) and segmented microscopic biopsy image with color K-means clustering approach (right)

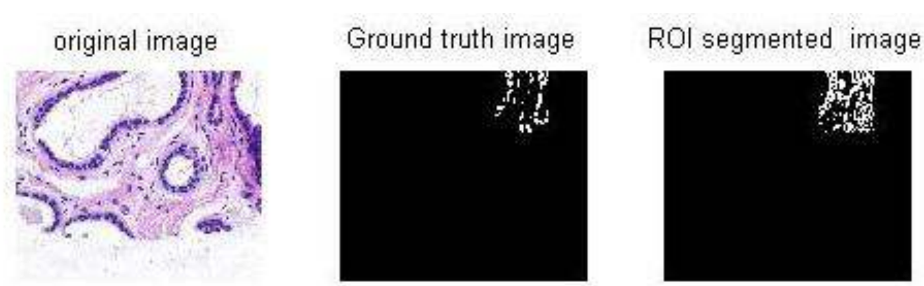

Figure 4.3: Original, Ground truth, and ROI Segmented by Texture based segmentation

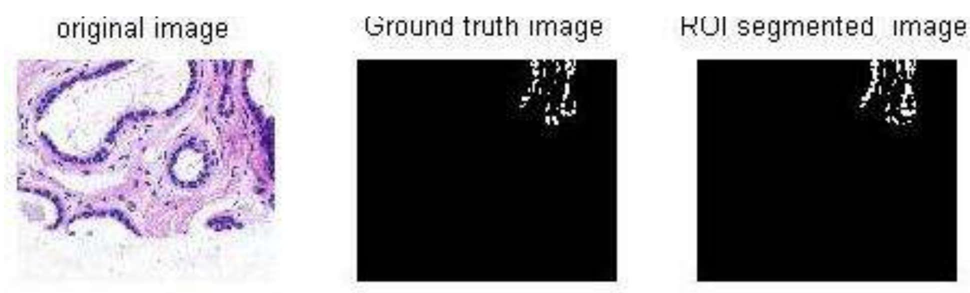

Figure 4.4: Original, Ground truth, and ROI Segmented by FCM segmentation

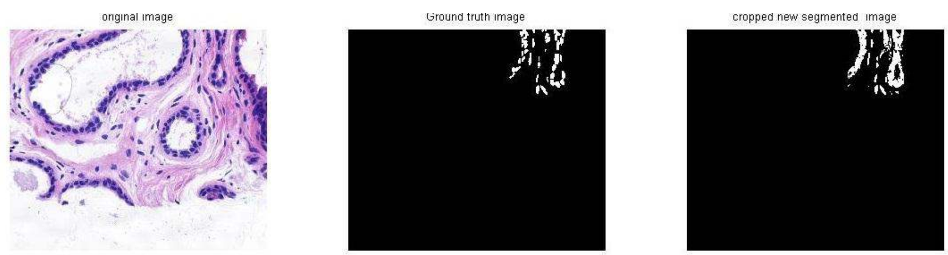

Figure 4.5: Original, Ground truth, and ROI Segmented by color k-means segmentation

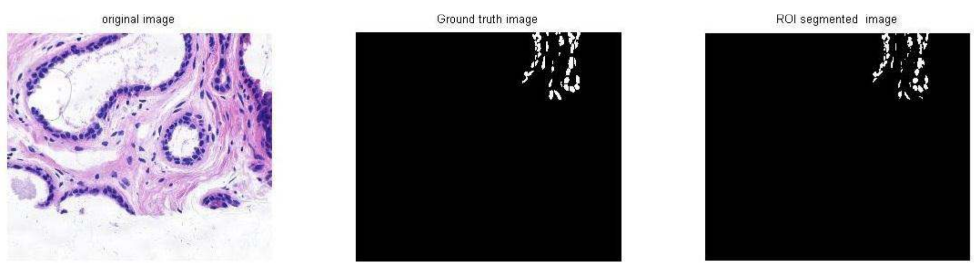

Figure 4.6: Original, Ground truth, and ROI Segmented by proposed segmentation

4.4 Results and discussions

Table 4.1: Comparison of various segmentation approaches with proposed approach

Authors(Year)	**Methods used for Segmentation**	**Data set used**	**Performance (%)**
Xu, J. et al., (2010)	Geodesic active contour	Microscopic biopsy images	Accuracy=84
Sertel, O *et al.*, (2009)	Nonlinear color quantization	Follicular Lymphoma	Accuracy=81 FPR=12.2
Demir, C *et al.*,	Object graph	Colon Glands	Accuracy=90

(2005)	approach		
Kong, H *et al.,* (2011)	Color texture cell segmentation	Follicular Lymphoma	Accuracy=94.75
Dundar, M. M *et al.*, (2011)	Gaussian mixture model based segmentation	Microscopic biopsy image of breast	Accuracy=87.9
Basavanhally, A, *et al.,* (2011)	Color gradient based active contour model	Microscopic biopsy image of breast	Accuracy=86
Tosun, A, *et al.,* (2011)	Graph run length matrix for breast tissue segmentation	Microscopic biopsy image of breast tissue of size 1920×2560	Accuracy=94 Specificity=92 Sensitivity =95
Yan Xu, *et al,.* (2014)	Hybrid approach	Follicular Lymphoma	Accuracy=96
Y.M .George *et. al., (2*013)	FCM based hybrid approach	92 images of size 640 ×480 breast cytology images	Specificity=83.16 Sensitivity =95.49
Proposed approach	Color K-means based hybrid segmentation	58 Microscopic biopsy image of breast tissues of 896×768 with its ground truth image	Accuracy=99.03 Specificity=84.30 Sensitivity=99.20

Table 4.1 shows the comparison of various segmentation algorithms with proposed approach. Figures 4.2, 4.3, 4.4, 4.5, and 4.6 provide the visual observations of results of the ROI segmentation with proposed approach and others approaches.

Table 4.2 shows quantitative evaluation of various segmentation methods on the basis of average values of various performance metrics for a set of 58 microscopic images of breast tissues, using the popular segmentation approaches of microscopic biopsy images. Table 4.2 and Figure 4.7 show the comparison of various segmentation algorithms on the basis of average accuracy, sensitivity, specificity, FPR, FNR, PRI, GCE, and VOI for 58 sample images taken from breast image data set.

From Table 4.2 and Figure 4.7 following observations are made for sample of microscopic biopsy of breast tissue:

- For the segmentation of microscopic biopsy images of breast tissues in the case of Texture based segmentation approach, the average value of accuracy, sensitivity, specificity, FPR, FNR, PRI ,GCE, and VOI are 0.966093, 0.245813 , 0.991565, 0.008435, 0.754187, 0.934675, 0.028149, and 0.278278 respectively.
- For the segmentation of microscopic biopsy images of breast tissues in the case of Fuzzy c means based segmentation approach, the average value of accuracy, sensitivity, specificity, FPR, FNR, PRI, GCE and VOI are 0.983846, 0.529358, 0.998609, 0.001391, 0.470642, 0.968331, 0.016659, and 0.156123 respectively.

- For the segmentation of microscopic biopsy images of breast tissues in the case of color k-means based segmentation approach, the average value of accuracy, sensitivity, specificity, FPR, FNR, PRI ,GCE and VOI are 0.990097, 0.826871, 0.991793, 0.008207, 0.173129, 0.980435, 0.009699, and 0.10325 respectively.
- For the segmentation of microscopic biopsy images of breast tissues in the proposed segmentation approach the average value of accuracy, sensitivity, specificity, FPR, FNR, PRI, GCE and VOI are 0.990368, 0.843017, 0.99216, 0.00784, 0.156983, 0.980964, 0.010295, and 0.102975 respectively.

From the Table 4.2 and Figure 4.7, it observed that the proposed framework for segmentation of microscopic biopsy images of breast cancer tissues is performing better in comparison to the color k-means, fuzzy c-means, , and texture based methods in terms of accuracy, specificity, and PRI measures. Only, the proposed segmentation algorithm is associated with larger value of accuracy, sensitivity, specificity, and random index (RI), and smaller value of FPR, FNR, GCE, and VOI in comparison to other methods. Hence, it is performing better in terms of all parameters.

Table 4.2: Quantitative evaluation of segmentation methods on the basis of average values of various performance metrics for a set of 58 microscopic images of breast tissues

Methods	**Accuracy**	**Sensitivity**	**Specificity**	**FPR**	**FNR**	**PRI**	**GCE**	**VOI**
Texture	0.9660	0.2458	0.9915	0.0084	0.7541	0.9346	0.0281	0.2782
FCM	0.9838	0.5293	0.9986	0.0013	0.4706	0.9683	0.0166	0.1561
Color km	0.9900	0.8268	0.9917	0.0082	0.1731	0.9804	0.0096	0.1032
Proposed	0.9903	0.8430	0.9921	0.0078	0.1569	0.9809	0.01029	0.1029

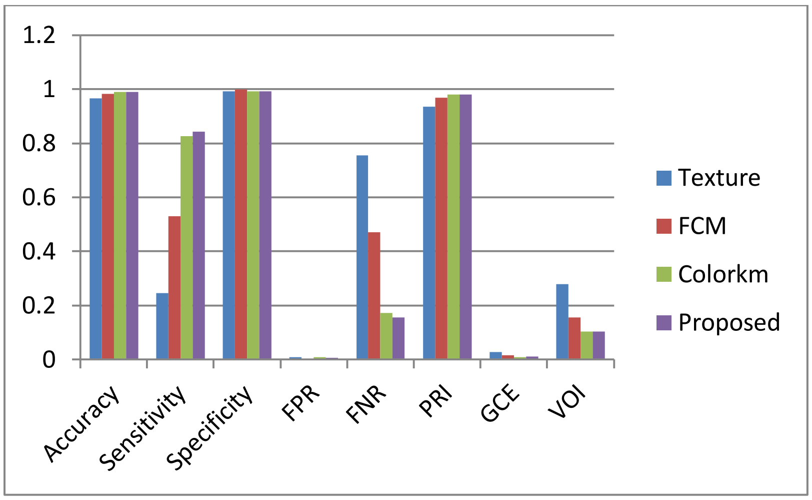

Figure 4.7: Comparisons of various segmentation methods on the basis of average Accuracy, Sensitivity, Specificity, FPR, FNR, PRI, GCE, and VOI for 58 sample images from breast tissues

4.5 Conclusion

A hybrid color k-means and watershed based segmentation approach has been proposed for segmentation of cell and nuclei from breast tissue of microscopic biopsy image. The proposed approach was based on tissues level microscopic observations of cell and nuclei for breast biopsy images of benign and malignant tissue. For testing and experimentation purpose, 31 benign and 27 malignant images of 896×768 were taken from breast tissue dataset. Finally, the

ROI segmented image of microscopic biopsy was compared to ground truth images. The quantitative and qualitative evaluation of various segmentation approaches for all 58 sample images were performed. From Table 4.2 and Figure 4.7, it was observed that the proposed framework is associated with larger value of accuracy, sensitivity, specificity, random index (RI), and smaller value of FPR, FNR, GCE, and VOI in comparison to other methods. The performance measured for microscopic biopsy of breast tissues dataset in terms of the accuracy, sensitivity, specificity, FPR, FNR, PRI, GCE, and VOI were 0.990368, 0.843017, 0.99216, 0.00784, 0.156983, 0.980964, 0.010295 and 0.102975 respectively. Hence, we are in position to conclude that, the proposed approach is performing better in terms of all parameters and it is suitable for the segmentation of microscopic biopsy images.