RESEARCH ARTICLE OPEN ACCESS

Lung cancer detection using Mean Shift Clustering Algorithm and Bayesian Decision Theory

V.Sowmya[1], Dr.Judith Justin[2]
1,2 (Department of Biomedical Instrumentation Engineering,
Faculty of Engineering, Avinashilingam University, Coimbatore)

Abstract:

Early detection of lung cancer is a demanding problem due to both the structure of the cancer cells and the staining method employed in the formulation of the sputum cells. The manual analysis of the sputum sample is time consuming and inaccurate due to overlapping of the cells. Computer Aided Diagnosis (CAD) system helps to detect cells accurately, thereby improving the chances of survival of the patient. In the proposed study, Wiener filter is used to denoise the image and the quality metrics such as Peak Signal to Noise Ratio (PSNR) and Root Mean Squared Error (RMSE) are used to assess the performance. The sputum cells are segmented using Mean Shift Clustering Algorithm and to extract the homogeneous tissues in the image. Bayesian Decision Theory classifies images as normal or abnormal.

Keywords **— Lung cancer, Sputum Cytometry, Color Sputum Image, Mean Shift Clustering, Bayesian Decision Theory**

I. INTRODUCTION

Lung cancer is considered to be the main cause of death worldwide, and it is difficult to detect it in its early stage as the symptoms appear only in the advanced stages. Sputum is a non-invasive accessible body fluid in the lungs, which contains exfoliated bronchial epithelial cells and Molecular study of sputum can detect tumour cells. Non-Small Cell Lung Cancer (NSCLC) is the most common type of lung cancer, mainly consisting of Adeno Carcinoma (AC) and Squamous Cell Carcinomas (SCC) [1]. Different diagnostics techniques such as Chest Radiograph (X-ray), Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and Sputum Cytology were used to detect the lung cancer in its advanced stage [2].

Good quality sputum is obtained from the Lower Respiratory Tract (LRT). Bad quality sputum is obtained from the Upper Respiratory Tract (URT) which consists of only saliva and epithelial cells. Sputum with good quality is used for diagnosis of Lung cancer. Fig 1 shows a sputum slide image which consists of pus cells (neutrophils), Squamous Epithelial Cells (SEC), gram-positive and gram-negative bacteria.

The sputum image is pre-processed with denoising filters such as Wiener, Gaussian and Median filter. A comparison of performance of the filters is done using PSNR and RMSE. The region of interest is highlighted and segmented using Mean Shift Clustering Algorithm. The segmented image is classified as either normal or abnormal using Bayesian classifier.

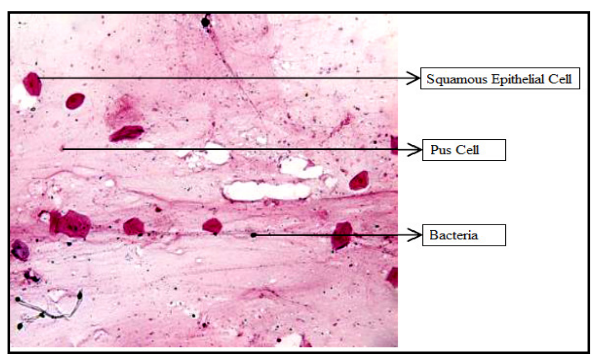


Fig.1. Sputum image under x10 magnification

II. MATERIALS AND METHODS

A CAD system is illustrated in Fig.2. The input image is a digitized Sputum image, which is obtained from the database[16]. In order to remove the noise in these images Wiener filter is used since its performance is better than the Gaussian and Median filter in terms of PSNR and RMSE. The region of interest is segmented with clean differentiation between sputum cells, body fluids, debris and background. Mean Shift Clustering is used in the proposed method for segmentation. The clusters with sputum cells are considered for classification. Bayesian Classifier helps to identify the abnormalities.

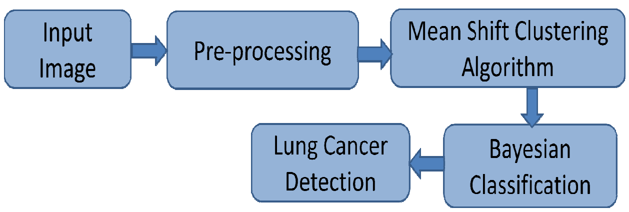


Fig. 2. Steps for Lung Cancer Detection

A. *Preprocessing*

The sputum image is pre-processed with denoising filters like Wiener, Gaussian and Median. Performance of these filters are compared using PSNR and RMSE values and their results are shown in Table.1 and Table.2. Wiener filter gives best performance. Higher values of PSNR and lower values of RMSE indicates best noise removal. A total of thirty normal and pathological images are used for testing.

1. Peak Signal to Noise Ratio (PSNR):

Peak signal to noise ratio (PSNR) [13] is used to measure the difference between the original and the filtered images of size MxN and is estimated using equation (1). It is expressed in decibel (dB).

$$\mathrm{PSNR} = 10 \log \frac{g_{max}^2}{\mathrm{MSE}} \quad (1)$$

where g2max is the maximum intensity in the gray scale image and MSE is the Mean Squared Error and is given in equation (2).

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(x(i,j)-\hat{x}(i,j))^2 \quad (2)$$

TABLE. 1. PSNR VALUES OF NORMAL AND ABNORMAL IMAGE

Filters	Normal	Abnormal
Wiener	**38.365**	**43.131**
Median	23.752	29.076
Gaussian	16.432	23.011

2. Root Mean Squared Error (RMSE):

Root Mean Squared Error (RMSE)[14] is the square root of the squared error averaged over MxN window and is calculated using the equation (3).

$$\mathrm{RMSE} = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(x(i,j)-\hat{x}(i,j))^2} \quad (3)$$

TABLE.2. RMSE VALUES OF NORMAL AND ABNORMAL IMAGE

Filters	Normal	Abnormal
Wiener	**0.4045**	**0.233**
Median	1.7106	0.845
Gaussian	12.4611	13.187

To asses the visual quality, the denoised images are shown in Fig.3.

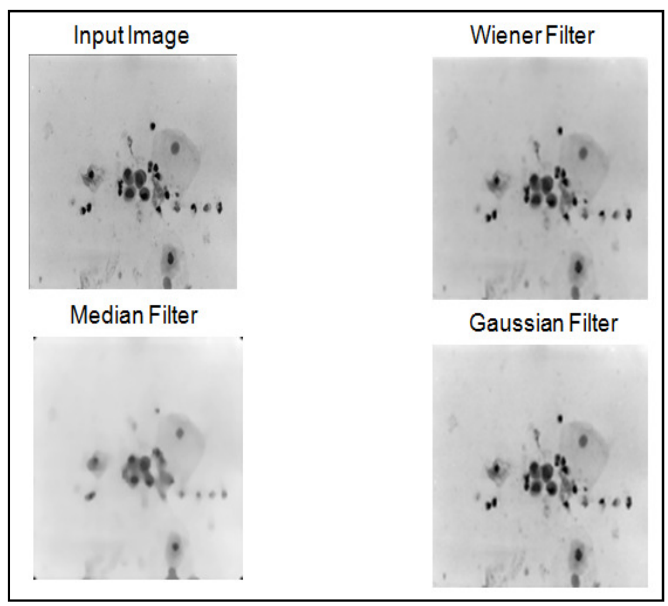


Fig. 3. Comparison of Filters

3. Classification Accuracy:

Classification accuracy is calculated as the sum of correct classifications divided by the total number of classifications. The classification accuracy Ai is evaluated as in equation (4).

$$Ai = (t/N)*100 \qquad (4)$$

where,
i depends on the number of samples correctly classified,
t is the number of samples correctly classified, and
N is the total number of samples

TABLE. 3. CLASSIFICATION ACCURACY

Image type	Total number of images	Classification accuracy of proposed method
Normal	12	98%
Abnormal (Cancerous)	18	90%

B. Segmentation

Mean shift clustering considers the feature space as an empirical Probability Density Function (PDF). Each pixel of the image is considered as a vector x in the feature space and this vector can be built from different features like spatial position, color components and normal vectors. Mean shift cluster is a set of vectors as sampled from the underlying PDF. The denoised image is separated into four regions.

After repeated iterations, the clusters were clearly visible at the 200th iterations. Fig.4 shows the result obtained after clustering. Fig.5 shows the segmentation output.

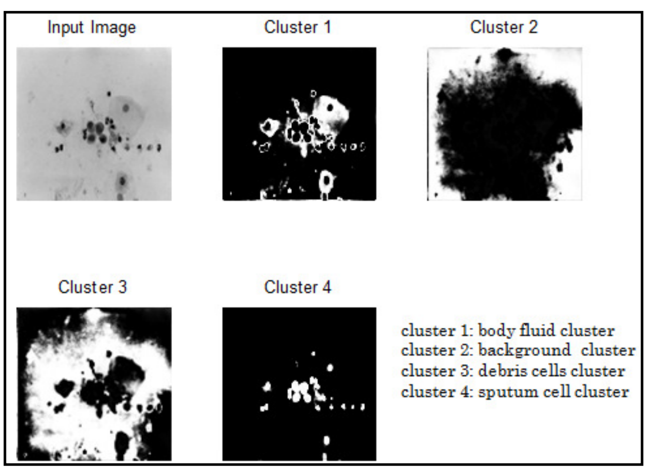


Fig. 4. Clusters after 200th iteration

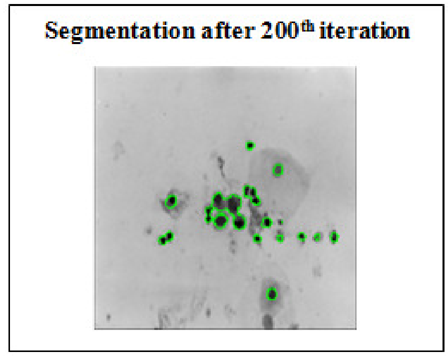


Fig. 5. Segmentation after 200th iteration

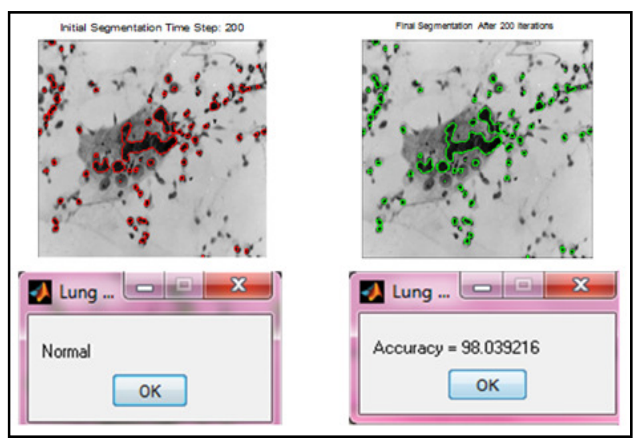
C. Classification

Bayesian decision theory is a basic statistical approach to the problem of classification. It is considered as an ideal case in which the probability structure underlying the categories is known perfectly. The Baye's theorem can be given as shown in equation (5).

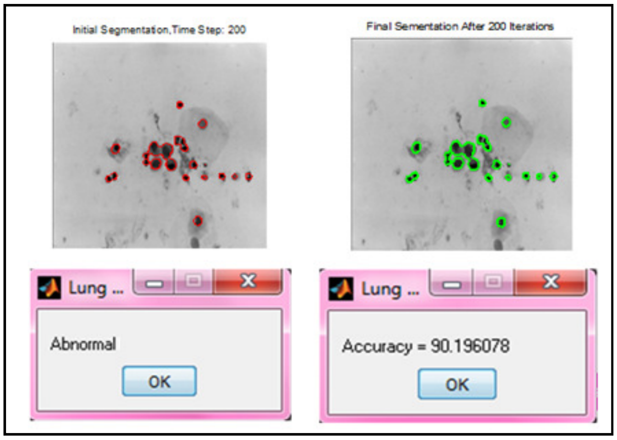
$$P(c_j|d) = \frac{P(d|c_j)P(c_j)}{P(d)} \qquad (5)$$

where,
$P(c_j|d)$ = probability of instanced being in class c_j
$P(d|c_j)$ = probability of generating instanced has given class c_j,
$P(c_j)$ = probability of occurrence of class c_j,
$P(d)$ = probability of instanced is occurring

This classifier is used in the proposed method, since it is robust to noise. The classification shows an accuracy of normal image is 98% and is shown in Fig.6 (a) and for abnormal image the accuracy is 90% and is shown in Fig.6 (b).



(a)



(b)
Fig. 6. Classification Output (a) Normal Image
(b) Abnormal Image

III. CONCLUSION

The Wiener filter is used to remove the noise from the input image for the effective denoising. Based on the robust and accurate spatial information provided by the mean shift convergence frequency and locations (creating map images), the over-segmentation occurring in the non-parametric segmentation method can be reduced. This enhanced joint spatial-range mean shift segmentation method is less sensitive to bandwidth selection according to the property of the mean shift convergence which also holds in the temporal sense.

The lung cancer detection can be made effective using the mean shift clustering algorithm and Bayesian decision theory. The mean shift clustering algorithm is used to classify the clusters based on the sputum cells, background, debris cells and body fluid to obtain accurate segmentation of sputum cells with higher accuracy. The Bayesian classifier categorizes the images as normal or abnormal. In the proposed system the accuracy of the proposed system is 90%.

REFERENCES

1. *Dignam JJ, Huang L, Ries L, Reichman M, Mariotto A, Feuer E (2009). Estimating cancer statistic and other-cause mortality in clinical trial and population based cancer registry cohorts, Cancer 10, August.*
2. *W. Wang and S. Wu (2006). A Study on Lung Cancer Detection by Image Processing, Proceeding of the IEEE conference on Communications, Circuits and Systems, pp. 371-374.*
3. *A. Sheila and T. Ried (2010). Interphase Cytogenetics of Sputum Cells for the Early Detection of Lung Carcinogenesis, Journal of Cancer Prevention Research,Vol. 3, Issue. 4, pp. 416-419, March.*
4. *D. Kim, C. Chung and K. Barnard (2005). Relevance Feedback using Adaptive Clustering for Image Similarity Retrieval, Journal of Systems and Software, Vol. 78, pp. 9-23, October.*
5. *S. Saleh, N. Kalyankar, and S. Khamitkar (2010). Image Segmentation by using Edge Detection, International Journal on Computer Science and Engineering (IJCSE), Vol. 2, Issue. 3, pp. 804-807.*
6. *L. Lucchese and S. K. Mitra (2001). Color Image Segmentation: A State of the Art Survey, Proceeding of the Indian National Science Academy (INSA), New Delhi, India, Vol. 67, Issue.2, pp. 207-221.*
7. *F. Taher and R. Sammouda (2007). Identification of Lung Cancer based on shape and Color, Proceeding of the 4th International Conference on Innovation in Information Technology, pp. 481-485, Dubai, UAE, November.*
8. *M. G. Forero, F. Sroubek and G. Cristobal (2004). Identification of Tuberculosis Based on Shape and Color, Journal of Real time imaging, Vol. 10, pp. 251-262.*
9. *Y. HIROO (2003). Usefulness of Papanicolaou stains by dehydration of air dried smears , Journal of the Japanese Society of Clinical Cytology, Vol. 34, pp.107-110, Japan.*
10. *R. Sammouda, N. Niki, H. Nishitani, S. Nakamura and S. Mori (1998). Segmentation of Sputum Color Image for Lung Cancer Diagnosis based on Neural Network, IEICE Transactions on Information and Systems. Vol.8, pp. 862-870, August.*
11. *F.Taher and R. Sammouda (2010). Lung Cancer Detection based on the Analysis of Sputum Color Images, Proceeding of the International Conference on Image Processing, Computer Vision, & Pattern Recognition (IPCV'10: WORLDCOMP 2010), Las Vegas, USA, 12-15, pp. 168-173, July.*

12. *H. Sun, S. Wang and Q. Jiang (2004). Fuzzy C-Mean based Model Selection Algorithms for Determining the Number of Clusters, Pattern Recognition,Vol. 37,pp. 2027-2037.*

13. *D. Sakrison (1997). On the role of observer and a distortion measure in image transmission, IEEE Transactions on Communications, Vol.25, no.11, pp.1251-1267.*

14. *R.E. Gonzalez and R.C. Woods (2008)(3rd ed), Digital Image Processing, Pearsons Education, Inc Prentice Hall, India.*

15. *Naman Chopra and Mr.Anshul Anand (2014). Despeckling of Images Using Wiener Filter in Dual Wavelet Transform Domain, Vol.5, pp:4069-4071.*

16. *http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3276027/*