Advanced Science Letters
Vol. 23, 2296–2298, 2017

Naïve Bayes Algorithm for Lung Cancer Diagnosis Using Image Processing Techniques

Kusworo Adi[1,2,*], Catur Edi Widodo[1,2], Aris Puji Widodo[1,3], Rahmat Gernowo[1,2],
Adi Pamungkas[2], and Rizky Ayomi Syifa[2]

[1] *Master of Information Systems, Diponegoro University, Semarang, Indonesia*
[2] *Department of Physics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia*
[3] *Department of Informatics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia*

Lung cancer is a disease has highest mortality rates in the world. Based on data from the International Agency for Research Center (IARC) in 2012, there are 19.4% of people in the world die from lung cancer. Microscopic examination process of biopsy still has some drawbacks. In the process of diagnosis, medical practitioners still use visual observation, so that analysis results are subjective and takes a long time. In this research, a system developed by microscopic analysis of biopsy with digital image processing techniques. The process of identification of cancer cells in the biopsy sample is done through the stages of feature extraction using Gray Level Co-Occurrence Matrix (GLCM) and classification using a Naive Bayes algorithm. The results of image classification biopsy showed the accuracy of 88.57% with combination of parameters contrast and homogeneity. Digital image processing techniques can be implemented in the process of microscopic examination of biopsy.

Keywords: Lung Cancer, Biopsy, Image Processing, GLCM, Naive Bayes.

1. INTRODUCTION

Lung cancer prevalence is among the highest of all cancers, at 18%.[1] In 2010, lung cancer sat on 3rd rank in terms of cancer incidence and was on top of the list of mortality rate due to cancer around the world. Moreover, lung cancer occupied top spot for both incidence and mortality in males, and sat on rank for both facets in females (only next to breast cancer, cervical cancer, and colorectal cancer).[2]

Lung cancer examination is carried out in three stages, i.e., CT scan imaging, sputum examination, and lung tissue sampling (biopsy). That first step is usually an X-ray imaging. Should this step reveal a cancer suspect. Next step, albeit sputum examination is microscopic and meant to find out whether cancer cells are present in the lungs. Last step, biopsy is aimed at showing the presence of cancer cells in the chest. Medical practitioners still rely on subjective visual observation. The medical practitioners must be thorough observation and accurate analysis in detecting lung cancer in patients. Hence, there is a need a system that is capable of detecting lung cancer automatically on microscopic biopsy images. This will improve the accuracy and efficiency of lung cancer detection. Digital image processing is a discipline that studies image processing as to make it better interpreted. This technique has been applied in numerous medical applications such as tuberculosis bacteria detection on microscopic sputum image,[3,4] detection of malaria causing plasmodium falciparum phase.[5–7] Some research for detection of lung cancer objects in CT scan images[8–10] and the microscopic sputum sample analysis for lung cancer.[11–13] These researches show that digital image processing can be applied in the medical field, especially in lung cancer detection.

Based on previous researches, this research designs a system of lung cancer detection based on analysis of microscopic lung biopsy image. This system will help improving lung cancer examination by providing an automatic and objective technique that will certainly be beneficial in aiding medical practitioners in accurately diagnosing and treating lung cancer.

2. EXPERIMENTAL DETAILS

Image processing procedures involve the conversion of RGB image into grayscale, texture feature extraction and classification using Naive Bayes algorithm. Block diagram of the system diagnosis of lung cancer with digital image processing techniques are shown in Figure 1. The algorithm for the detection of cancer and non cancer with GLCM and Naive Bayes is as follows:

1. Load image.
2. RGB image is converted into grayscale using the equation:

$$0.2989 * \mathrm{R} + 0.5870 * \mathrm{G} + 0.1140 * \mathrm{B}$$

[*] Author to whom correspondence should be addressed.

1936-6612/2017/23/2296/003 doi:10.1166/asl.2017.8654

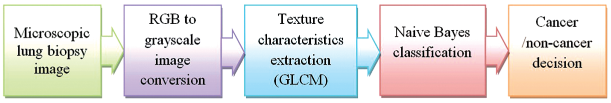

Microscopic lung biopsy image
RGB to grayscale image conversion
Texture characteristics extraction (GLCM)
Naive Bayes classification
Cancer /non-cancer decision

Fig. 1. Block diagram of the microscopic biopsy image analysis system.

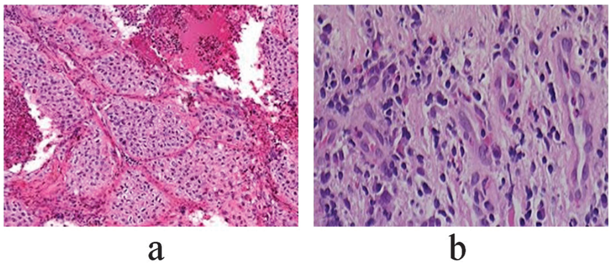


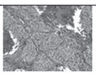
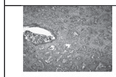
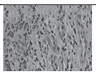
Fig. 2. Samples of microscopic lung biopsy images. (a) Cancer. (b) Non-cancer.

3. Converted grayscale images are then extracted texture using the Gray Level Co-occurrence Matrix method as to obtain texture parameters of contrast, correlation, energy, and homogeneity.
4. Classified into two classes of cancer and non-cancer using the Naive Bayes algorithm.

3. RESULTS AND DISCUSSION

Lung cancer biopsy samples have irregular network structure and agglomerate while normal lung biopsy samples had more regular network structure. The images of biopsy used in this study amounted to 35 images consisting of 16 categories of non cancer and 19 categories of cancer. Biopsy samples for lung cancer and normal lung biopsy is shown in Figure 2.

Table I. Samples features from microscopic lung biopsy images.

No	Category	Grayscale Image	GLCM Parameter			
			Contrast	Correlation	Energy	Homogeneity
1	Cancer		2.1377	0.5070	0.0483	0.6207
2			1.4934	0.5288	0.0682	0.6594
3	Non Cancer		0.3867	0.8502	0.1578	0.8396
4			0.3932	0.8182	0.1679	0.8368

The process of features extraction is carried out with texture analysis using the Gray Level Co-Occurrence Matrix (GLCM) method. This method works on the principle of calculating the probability of nearest neighbor between two pixels on certain distance and angular orientation. This approach builds co-occurrence matrices of image data, which in turn determine characteristics as the matrix function of those images.

Co-occurrence means happening at the same time. This translates to the probability of one level of a pixel value being nearest to a value level of another pixel at certain distance (d) and angular orientation (θ). Distance is stated as pixels, while orientation is in degrees. Orientation is made up of four angular directions, each with a 45° interval. They are; 0°, 45°, 90°, and 135°, whereas the distance between two pixels is given as 1 pixel. Then from that co-occurrence matrix, parameters of contrast, correlation, energy, and homogeneity are extracted as texture features. Samples of extracted microscopic lung biopsy image features are given in Table I.

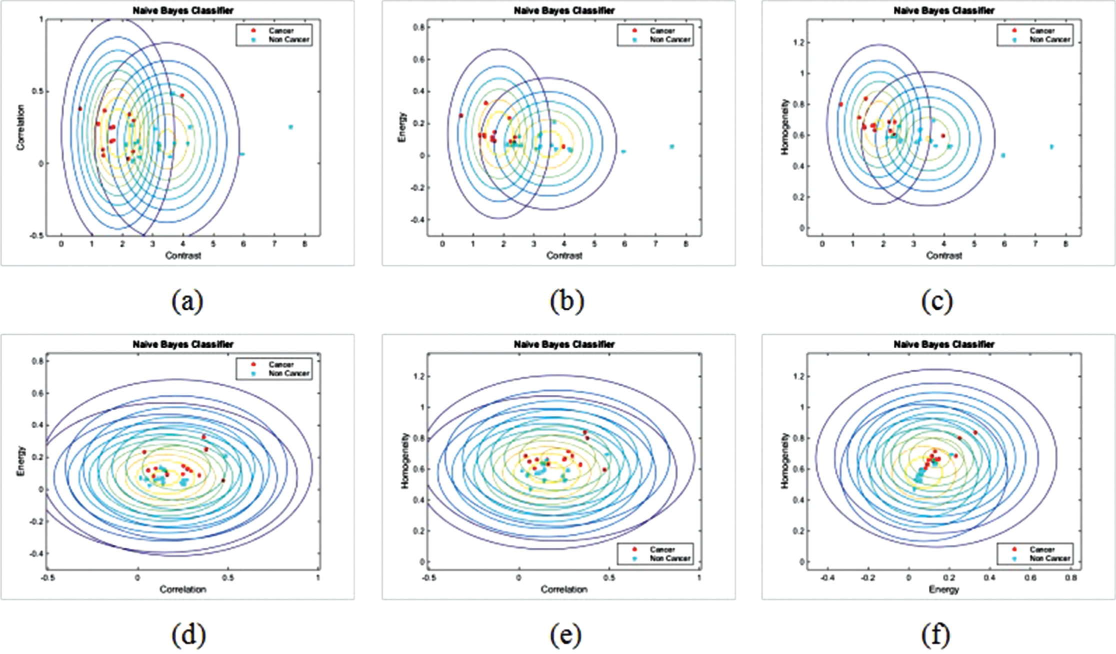


Fig. 3. Distribution graphic of Naive Bayes classification. (a) Contrast and correlation. (b) Contrast and energy. (c) Contrast and homogeneity. (d) Correlation and energy. (e) Correlation and homogeneity. (f) Energy and homogeneity.

Table II. **Results of the biopsy overall image classification.**

No	Parameters	Accuracy (%)
1	Contrast and correlation	85.71
2	Contrast and energy	85.71
3	Contrast and homogeneity	88.57
4	Correlation and energy	71.43
5	Correlation and homogeneity	85.71
6	Energy and homogeneity	74.28

Image classification process biopsies were performed using Naive Bayes algorithm by combining two of the four parameters of feature extraction results. Distribution graph image classification biopsy using Naive Bayes algorithm is shown in Figure 3. On the results of the classification Naive Bayes algorithm with the parameters of contrast and homogeneity there are four images wrongly classified so that the accuracy obtained is:

$$\text{Accuracy} = 31/35 \times 100\% = 88.57\%$$

The results of the biopsy overall image classification are shown in Table II. Table II shows that out of the six combinations of parameters that are used, the highest accuracy achieved by 88.57% by the parameters of contrast and homogeneity. The accuracy value shows that the Naive Bayes algorithm can be implemented in image classifying lung biopsy.

4. CONCLUSION

This research has been done of digital image processing to analyze the image of microscopic lung biopsy using Gray Level Co-Occurrence Matrix and Naive Bayes algorithm to classify the image into the cancer or non-cancer class. Feature extraction process is done by combining two of the four parameters that characterize the texture contrast, correlation, energy, and homogeneity. The results of image classification biopsy showed the highest accuracy of 88.57% on the combination of parameters contrast and homogeneity. This indicates that the digital image processing techniques can be implemented in the process of microscopic examination of biopsy.

Acknowledgments: This research was funded by the Indonesian Directorate General of Higher Education Program in 2016. The images of biopsy were used in this research is downloaded in http://www.eusatlas.ro, http://www.pathologyoutlines.com, and http://www.siumed.edu/, the authors would like to thank to supported image sample of biopsy.

References and Notes

1. World Health Organization, WHO report on the Global Tobacco Epidemic **(2008)**.
2. World Health Organization, Media Centre Cancer **(2010)**.
3. K. Adi, R. Gernowo, A. Sugiharto, A. Pamungkas, A. B. Putranto, and N. Mirnasari, *Proceedings the 7th International Conference on Information and Communication Technology and Systems (ICTS)* **(2013)**, pp. 9–13.
4. K. Adi, R. Gernowo, A. Sugiharto, K. S. Firdausi, A. Pamungkas, and A. B. Putranto, *International Journal of Innovative Research in Science, Engineering and Technology* 2 **(2013)**.
5. K. Adi, S. Pujiyanto, R. Gernowo, A. Pamungkas, and A. B. Putranto, *International Journal of Applied Engineering Research (IJAER)* 9, 13917 **(2014)**.
6. A. Pamungkas, K. Adi, and R. Gernowo, *International Journal of Applied Engineering Research* 10, 4043 **(2015)**.
7. K. Adi, S. Pujiyanto, R. Gernowo, A. Pamungkas, and A. B. Putranto, *International Journal of Applied Engineering Research* 11, 8754 **(2016)**.
8. G. Vijaya, A. Suhasini, and R. Priya, *International Journal of Research in Engineering and Technology (IJRET)* 3 **(2014)**, e-ISSN: 2319-1163, p-ISSN: 2321-7308.
9. V. A. Gajdhane and L. M. Deshpande, *IOSR Journal of Computer Engineering (IOSR-JCE)* 16, 28 **(2014)**.
10. Neha and J. Shekhar, *International Journal of Engineering Development and Research* 3, 1290 **(2015)**.
11. F. Taher, N. Werghi, H. Al-Ahmad, and C. Donner, *Algorithms* 6, 512 **(2013)**.
12. V. Kumar and H. Sharma, *International Journal of Computer Applications* 72, 35 **(2013)**.
13. S. Kaur and S. Kaur, *International Journal of Innovative Research in Computer and Communication Engineering (IJIRCE)* 3, 7446 **(2015)**.
14. M. Roumi, Thesis: Computer Engineering, Mekelweg 4, Delft University of Technology, The Netherlands **(2009)**.
15. C. N. Rao, S. S. Sastry, K. Mallika, H. S. Tiong, and K. B. Mahalakshmi, *International Journal of Innovative Research in Science, Engineering and Technology* 2, 4531 **(2013)**.
16. N. K. Korada, N. S. P. Kumar, and Y. V. N. Deekshitulu, *International Journal of Information Sciences and Techniques (IJIST)* 2, 63 **(2012)**.

Received: 12 October 2016. Accepted: 2 November 2016.