Automatic Detection of Synchronous Multiple Primary Lung Cancer in Lung CT images using PML Techniques

G. Niranjana* and M. Ponnavaikko**

ABSTRACT

Lung cancer growth is classified into 4 stages based on T, N and M descriptors. Stage I corresponds to the initial stage and stage IV corresponds to the final stage of the lung cancer. Identifying the stages of the lung cancer is vital as treatment as treatment is based on the stage. Stage IV is termed as Metasis stage where the cancer spreads from one organ to another organ in the human body. Synchronous multiple primary nodules may correspond to the metasis stage of the lung cancer. In this paper we are presenting scoring method, a pixel based machine learning (PML) technique for identifying possible metasis stage of the cancer with the property that the nodules extended from one lung region to another or having multiple malignant nodules. The proposed method is tested with 8 patients with a total of 86 nodules. We observed 87.2% of accuracy in the results.

Keywords: lung cancer, metasis, PML, nodules

1. INTRODUCTION

Lung cancer is main cause for around 20 billion deaths worldwide annually. It is most general cause of cancer-related death in men and the second most common in women [1]. The identification of lung cancer is done using Chest Radiographs and Computed Tomography (CT) scans and it is confirmed by biopsy. Lung cancer can start from any portion of the lung. From there it spreads to other organs. Initially it starts to invade the nearby lymph nodes, followed by mediastinum. The right lung and left lung are separated by mediastinum. In the mediastinum, the lung cancer moves towards the other side. Once the mediastinum midline is crossed it is treated as more advance disease. Lung cancer can also spread to other distant organs in the human body which correspond to the final stage called stage IV [3].

Staging is the process of determining where the cancer is located and how much it is spread in the body. Staging of lung cancer is crucial as treatment choices are based on the stage of the cancer. The primary reason is to depict the level of severity of an individual's disease [4]. Staging information which is obtained prior to surgery with the aid of imaging modalities such as X-Rays, CT scans and PET scans are called clinical staging and staging by biopsy is called as pathological staging. We are concentrating on clinical staging in this paper.

The tool used for doctors to identify the different stages of cancer is based on certain standard criteria in the TNM classification system [5]. The size of the tumour (T), the extend of the spread to nearby lymph nodes (N) in the lung, and the presence of metasis (M) where the cancer moves into other organs of the body are the factors to decide TNM staging.

One possible occurrence of metasis stage is having multiple nodules in the lung region. Multiple nodules may occur either simultaneously or at different time. The simultaneous presence of second tumour is the

---

*   Assistant Professor, SRM University, Chennai, Tamilnadu, India.

**   Vice-Chancellor, Bharath University, Chennai, Tamilnadu, India.

characterisation of Synchronous Multiple Primary Lung Cancer (SMPLC). The occurrence of multiple primary nodules may vary from 1 to 16% [20]. Martini and Melamed [22] suggested that the lesions occur in different segments of the same lobe or in different lobes is the most important diagnostic criteria for SMPLC. . So far there is no common method to follow the diagnosis of SMPLC.

There are two main computational systems developed to assist radiologists. They include Computer Aided Detection (CAD) and Computer Aided Diagnosis (CADx) systems. CAD systems are designed to identify lesions in medical image by marking suspected areas of regions in the image. The lesion characteristics are measured using CADx systems i.e determining the malignancy and identifying its stage [2]. The purpose of the CADx system is to enhance the sensitivity, specificity, efficiency and cost of lung cancer screening programs.

Several CADx systems are available for diagnosing lung nodules as benign or malignant. In the proposed work we focused on the design of a CADx system that concentrates on identifying MPLC of the lung which may constitute to the metasis stage. It would be helpful for assisting radiologist as a second opinion to stage the disease since the treatment is treatment is highly based on the stage of the disease.

2. MATERIALS AND METHODS

The proposed work consists of the following stages. Preprocessing of the input image, Extraction of the Lung region, Removal of blood vessels, Feature extraction and detection of the Nodule using pixel based machine learning and Classification.

The input image need to preprocessed as it may contain noise and other unwanted information in it. The unwanted noise in the image is removed using Median filter. Left and Right lungs are extracted separately. Blood vessels are removed with the help of morphological operator. Nodules are segmented by applying multi level thresholding. Statistical count of the number of nodules and its extension to other parts are identified. Using these measures Metasis stage is identified in the image. Fig. 1 shows the stages of the proposed method.

Proposed CADx system consists of several stages as presented below.

2.1. Preprocessing

The input for the proposed work is a DICOM image of a chest CT scan of size 512 x 512 pixels. Hanson et al [7] stated that random noise is the major source of noise in the CT image, but Gaussian noise is also present in CT images. We used Median filtering in our approach. Median filtering is a statistical approach to filter out noise that has corrupted image. "salt and pepper" noise [8] is removed using a nonlinear operation called Median filtering. A median filter works better than the convolution filter where the objective of the median filter is to reduce noise and preserve edges simultaneously.

Figure 1: Stages of the proposed method

2.2. Lung Extraction

Segmentation of lungs can be done with Region Growing method [10]. Two user defined seed points are given as inputs one for the left lung and other for the other lung. User defined seed points may be given interactively as a single click by the user. The corresponding coordinates are taken as input seed points for segmenting the objects. User clicks at any point inside the left and right lung region. That point is selected as a seed point for extracting lungs. Using these seed points as reference points their neighbor values are calculated. The decision to include the neighbors as object is decided by the mean and the standard deviation. Thereby thesholding is combined with Region Growing method for segmenting lungs.

Segmentation of Left and Right lungs has the following steps:

- Click any point inside the left and right lung regions.
- Calculate the coordinates of the clicked points and mark it as initial seed points for left and right lungs.
- Visit every neighbor of the seed point that are 8-connected to it.
- Calculate the mean and the standard deviation of the visited neighbors

Mean is defined as the measure of the average intensity of the selected region

$$\text{Mean} = \sum_{i=0}^{L-1} zi\ p(zi) \tag{1}$$

Standard deviation is the measure of average contrast of the selected region

$$\sigma = \sqrt{\sum_{i=0}^{L-1} (Zi - m)^2\ p(Zi)} \tag{2}$$

- Thresholding is applied for lung extraction using mean and standard deviation as threshold limits.
- Apply mask to separate lung from the original input image.

Fig. 2 shows the output of pre-processing and the segmented lungs.

2.3. Removal of blood vessels

As lung is an organ filled with air, fat tissues and blood vessels, air is removed when preprocessing the image. Blood vessels are removed with the help of existing morphological operators. Since blood vessels have tubular shape and nodules have disk shape we can remove blood vessels by open and closing the morphological operator.

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)

Figure 2 (a) Input image (b). Median Filtered image (c). Segmented lung regions

2.4. Separating Lung Regions

Separate the Lung region from other parts using thresholding technique. Left and the right lungs are separated. The missing boundaries in the lung region are filled using the Rolling ball operator. The most important step in all cancer detection applications is the nodule segmentation. Accurate nodule segmentation is of critical importance since it is the basis for diagnostic and treatment of lung cancer [11] including diagnosing the tumour growth rate in consecutive follow up CT scans [12, 13] and also monitoring tumour response to therapy [14,15]. Computer aided lung cancer screening is used for early detection of tumour malignancy [16, 17].

After the blood vessels are removed the next step is to find the region of interest (ROI). Nodules may be either isolated or non isolated connected with other stuructues in the lung. It may be attached to chest wall, blood vessels and parenchyma. Several techniques are used for segmenting the nodules. Some of the techniques are discussed here. Yongqiang Tan et al[18] used watershed, geometric active contours as well as Markov random field (MRF) for segmenting the nodule with manually selecting the region of interest. Xujiong Ye et al [19] proposed a shaped nodule detection technique which uses shape index map and dot map for detecting lung nodules. We combined Region Growing method and thresholding for segmenting the lung regions[24,25]. Initial seed point is selected, mean and standard deviation of the neighbours is calculated. Then the object may be segmented based on region growing method.

2.5. Scoring method for Classification

Once the left and right regions are segmented, a scoring method [23, 26] is used for combining of output pixel. A score for a given ROI defined as

$$S = fW(x, y) \times O(x, y)$$

Where $fW(x, y) = 1/\sqrt{2\pi\sigma e} - \dfrac{x2 + y2}{2\sigma 2}$ is the focused local window of the selected pixel and O(x, y) is

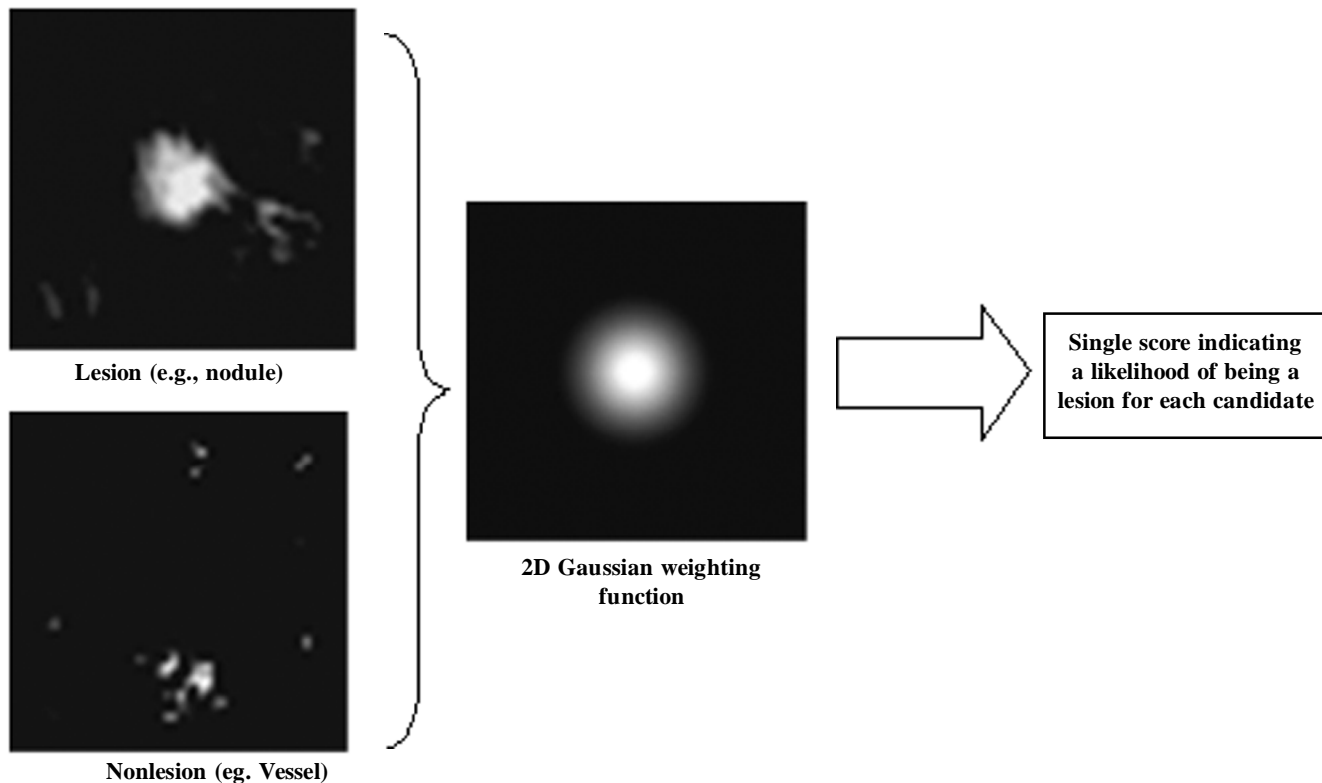the output image used for scoring method. This score represents the weighted sum of the estimates for the

Lesion (e.g., nodule)

Nonlesion (eg. Vessel)

2D Gaussian weighting function

Single score indicating a likelihood of being a lesion for each candidate

Figure 3: Scoring method for combining pixel-based output responses

| (a) | (b) | (c) |
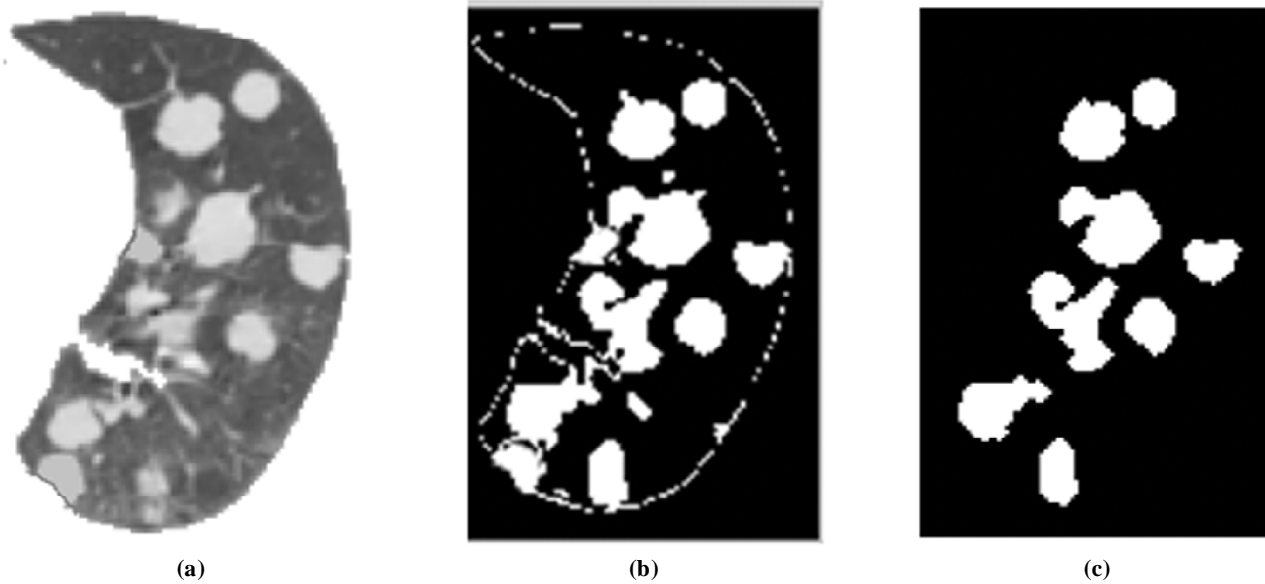
Figure 4: (a) Segmented left lung (b) Outlined lung and nodule region (c) Segmented Nodules
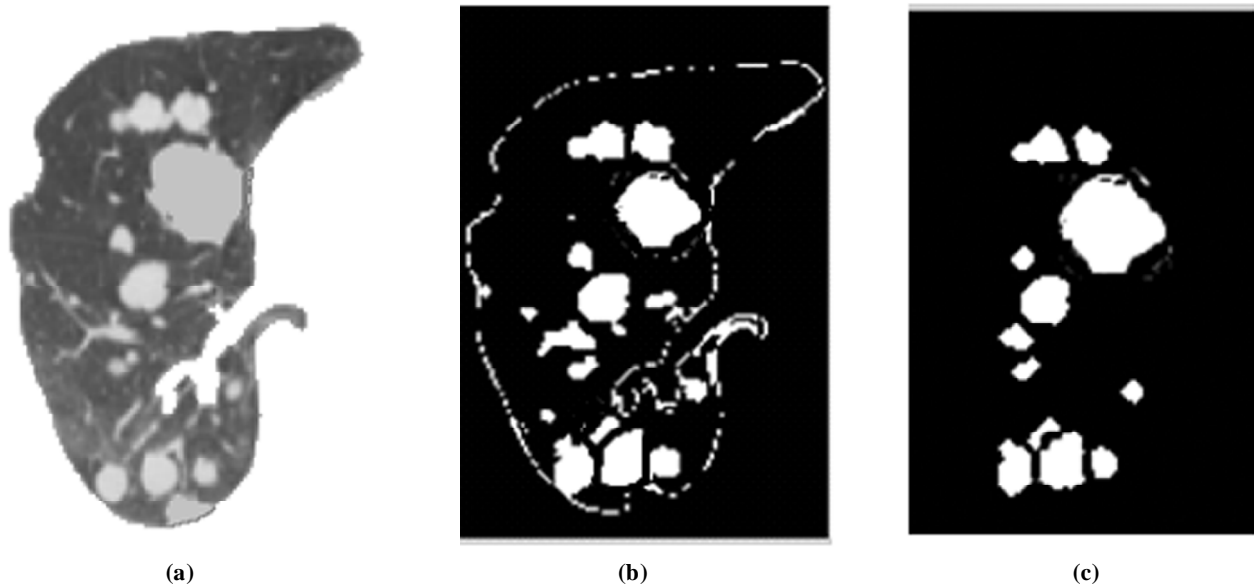

| (a) | (b) | (c) |

Figure 5: (a) Segmented right lung (b) Outlined lung and nodule region (c) Segmented Nodules

likelihood that the ROI (lesion candidate) contains a lesion near the center; that is, a higher score would indicate a lesion and a lower score would indicate a nonlesion. Thresholding is then performed on the scores for distinction between lesions and non-lesions. Fig. 3 shows the Scoring methods using Gaussian function. Fig. 4 represents the output images that are obtained for right lung and Fig. 5 shows the left lung.

2.6. Parameter Estimation

Suspected nodules are identified and segmented in the previous stage. The feature that is required for identifying MPLC is the total number of nodules in each lung and also to check whether both lungs have primary nodules. If there are more than 2 two primary nodules then MPLC is identified. If both lungs have primary nodules then it is reported as metasis stage.

3. EXPERIMENTAL RESULTS

The proposed method is tested with the sample images of 8 patients with a total of 86 nodules. Out of total 86 nodules 74 nodules were identified correctly and 12 nodules are not identified. The nodules that are

Img id	Left Lung		Right Lung		Total No.
	Actual no of nodules	*No of nodules identified*	*Actual no of nodules*	*No of nodules identified*	*of nodules not identified*
1	5	4	7	7	1
2	4	2	3	2	3
3	3	3	3	2	1
4	2	2	4	4	0
5	23	21	7	6	3
6	7	5	5	5	2
7	3	3	1	1	0
8	5	4	4	4	1

attached to the parenchyma wall is mostly not identified. Table 1 shows the observed results. The specificity obtained is 87.2%.

3.   CONCLUSION

A simple hybrid approach for identifying Synchronous Multiple Primary Lung Cancer is discussed in this paper. The proposed approach is tested with 8 patients of total 86 nodules. Among which 75 nodules are identified correctly and 11 nodules are not identified. The nodules attached to the parenchyma wall are not identified. The total specificity obtained is 87.2%. This work may further be extended to remove the false positives in the images.

REFERENCES

[1] American Cancer Society "Cancer Prevention & Early Detection", 2015-2016

[2] Mehul P. Sampat, Mia K. Markey, Alan C. Bovik, "Computer-Aided Detection and Diagnosis in Mammography" , Elsevier Academic Press, 2005

[3] Vaporciyan AA, Nesbitt JC, Lee JS, et al., " Cancer of the Lung", chapter 88 NCBI book shelf, 5th Edition

[4] Bernard J. Park and Valerie W. Rusch, "Lung Cancer Workup and Staging", Chapter 16 pp 241-252.

[5] Saeed Mirsadraee, Dilip Oswal, Yalda Alizadeh, Andrea Caulo, Edwin JR van Beek, " The 7th lung cancer TNM classification and staging system: Review of the changes and implications" World Journal of Radiology, April 2012.

[6] Yongqiang Tan, Lawrence H. Schwartz, and Binsheng Zhao, "Segmentation of lung lesions on CT scans using watershed, active contours, and Markov random field" , Med. Phys. 40 (4), April 2013

[7] HANSON, K. M. 1981. Noise and contrast discrimination in computed tomography. In *Radiology of the Skull and Brain*, vol. 5, T. H. Newtons and D. G. Potts, Eds., Mosby, 3941–3955.

[8] Mohsen Keshani,ZohrehAzimifar n, FarshadTajeripour,RezaBoostani, *"Lung nodule segmentation and recognition using SVM classifier and active contourmodeling:Acompleteintelligentsystem "*,Computers in Biology and Medicine 43 (2013) 287–300

[9] Yuhua Gu, VirendraKumar, LawrenceO.Hal, DmitryB.Goldgof, Ching-YenLi , Rene Korn, Claus Bendtsen, EmmanuelRiosVelazquez, AndreDekker, HugoAerts, PhilippeLambin, XiuliLi, Jie Tian, RobertA.Gatenby, RobertJ.Gillies "Automated delineation of lung tumors from CT images using a single click ensemble segmentation approach" Pattern Recognition 46 (2013) 692–702

[10] Y. Yim, H. Hong, and Y. G. Shin, "Hybrid lung segmentation in chest CT images for computer-aided diagnosis," in *7th International Workshop on Enterprise Networking and Computing inHealthcare Industry, HEALTHCOM2005*, pp. 378–383, kor, June 2005.

[11] I. Sluimer, A. Schilham, M. Prokop, and B. Van Ginneken, "Computer analysis of computed tomography scans of the lung: a survey," *IEEE Transactions on Medical Imaging*, vol. 25, no. 4, pp. 385–405, 2006.

[12] D. F. Yankelevitz, A. P. Reeves, W. J. Kostis, B. Zhao, and C. I. Henschke, "Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation," *Radiology*, vol. 217, no. 1, pp. 251–256, 2000.

[13] J. P. Ko, H. Rusinek, E. L. Jacobs et al., "Small pulmonary nodules: volume measurement at chest CT— phantom study," *Radiology*, vol. 228, no. 3, pp. 864–870, 2003.

[14] C. C. Jaffe, "Measures of response: RECIST, WHO, and new alternatives," *Journal of Clinical Oncology*, vol. 24, no. 20, pp. 3245–3251, 2006.

[15] M. A. Gavrielides, L. M. Kinnard, K. J. Myers, and N. Petrick, "Noncalcified lung nodules: volumetric assessment with thoracic CT," *Radiology*, vol. 251, no. 1, pp. 26–37, 2009.

[16] Y. Kawata, N. Niki, H. Ohmatsu et al., "Classification of pulmonary nodules in thin-section CT images based on shape characterization," in *Proceedings of the International Conference on Image Processing (ICIP '97)*, vol. 3, Part 2 (of 3), pp. 528–530, October 1997.

[17] K. Suzuki, J. Shiraishi, H. Abe, H. MacMahon, and K. Doi, "False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network," *Academic Radiology*, vol. 12, no. 2, pp. 191–201, 2005.

[18] Yongqiang Tan, Lawrence H. Schwartz, and Binsheng Zhao, "Segmentation of lung lesions on CT scans using watershed, active contours, and Markov random field" , Med. Phys. 40 (4), April 2013

[19] Xujiong Ye", Xinyu Lin, Jamshid Dehmeshki, Greg Slabaugh, and Gareth Beddoe, "Shape-Based Computer-Aided Detection of Lung Nodules in Thoracic CT Images", IEEE Transactions on Biomedical Engineering, vol. 56, no. 7, July 2009

[20] XinYing Xue, QingLiang Xue, Na Wang, Lina Zhang, Lina Guo, Xinfu Li, Junping Sun Jianxin Wang,"Early clinical diagnosis of synchronous multiple primary lung cancer", oncology letters, 2011

[21] Angeliki A. Loukeri, Christos F. Kampolis, Anna Ntokou, George Tsoukalas, Konstantinos Syrigos," Metachronous and Synchronous Primary Lung Cancers: Diagnostic Aspects, Surgical Treatment,and Prognosis", Clinical Lung Cancer January 2015

[22] Martini N, Melamed MR. Multiple primary lung cancers. J Thorac Cardiovasc Surg 1975; 70:606-12.

[23] Kenji Suzuki, "Pixel-Based Machine Learning inMedical Imaging" International Journal of Biomedical Imaging, Volume 2012

[24] Niranjana .G, Ponnavaikko M. "Statistical analysis of features and classifiers in identifying nodules and its T staging in lung CT images", ARPN Journal of Engineering and Applied sciences, Volume 11, Issue 5, 1 March 2016, Pages 2928-2938

[25] Niranjana .G, Ponnavaikko. M "An improved hybrid algorithm for accurate determination of parameters of lung nodules with dirichlet boundaries in CT images" Indian Journal of Science and Technology Volume 9, Issue 9, 1 March 2016.

[26] Sowmya. S, Niranjana. G, "Enhanced false positive reduction of lung nodule using pixel based machine learning and Fuzzy Min-max Neural Network " International Journal of Control Theory and Applications" Volume 9, Issue 2, 2016, Pages 1325-1331