

Preliminary Detection and Analysis of Lung Cancer on CT images using MATLAB: A Cost-effective Alternative

¹Md. Daud Hossain Khan, ²Mansur Ahmed and ³Christian Bach

Department of Biomedical Engineering, University of Bridgeport, USA

¹mdkhan@my.bridgeport.edu; ²maahmed@my.bridgeport.edu; ³cbach@bridgeport.edu

ABSTRACT

Cancer is the second leading cause of death worldwide. Lung cancer possesses the highest mortality, with non-small cell lung cancer (NSCLC) being its most prevalent subtype of lung cancer. Despite gradual reduction in incidence, approximately 585720 new cancer patients were diagnosed in 2014, with majority from low-and-middle income countries (LMICs). Limited availability of diagnostic equipment, poorly trained medical staff, late revelation of symptoms and classification of the exact lung cancer subtype and overall poor patient access to medical providers result in late or terminal stage diagnosis and delay of treatment. Therefore, the need for an economic, simple, fast computed image-processing system to aid decisions regarding staging and resection, especially for LMICs is clearly imminent. In this study, we developed a preliminary program using MATLAB that accurately detects cancer cells in CT images of lungs of affected patients, measures area of region of interest (ROI) or tumor mass and helps determine nodal spread. A preset value for nodal spread was used, which can be altered accordingly.

Keywords: Lung cancer, CT, MATLAB, Region-of-interest (ROI), area

1 Introduction

According to global reports, cancer is one of the most dominant causes of death worldwide, second only to cardiovascular disease.[1] Of the entire global cancer burden, majority of the incidences occur in low-and-middle-income countries.[2] Though there has been a gradual reduction in cancer incidence in the United States of America over the past decade, studies reveal that approximately 585720 new cancer patients were diagnosed in 2014.[3] Lung cancer will be responsible for 158040 total deaths in the United States alone by the end of 2015.[4] Lung cancer continues to be both the most dominant cause of incidence and death due to cancer, with non-small cell lung cancer (NSCLC) being the most prevalent subtype.[1, 2, 5] For quick, proper screening and diagnosis, the availability of radiotherapy or teletherapy centers equipped with efficient diagnostic or imaging equipment is absolutely necessary. Diagnostic procedures include regular imaging techniques such as X-rays, computed tomography (CT) and positron emission tomography (PET) scan.[6] Recent advancements in technology, both hardware and software, have allowed accurate detection, including precise measurement of tumorous masses inside the thoracic cavity. [7-9] However, there are drawbacks. Availability of advanced, diagnostic equipment for accurate detection of lung cancer, such as PET, Magnetic resonance imaging (MRI),

DOI: 10.14738/jbemi.26.1628

Publication Date: 16th November 2015

URL: <http://dx.doi.org/10.14738/jbemi.26.1628>

fluorodeoxyglucose PET (FDG-PET), helical CT or 3D-CT are limited, due to the overbearing expenses associated with both purchase and maintenance of such equipment, especially in case of limited-income regions.[8, 10-15] Modern computer assisted diagnosis (CAD) systems, though significantly improve the efficiency of these machines, add to the expenses.[9, 15] Along with the complexity involved in staging and late revelation of the symptoms, patients are often diagnosed at late or advanced stages of the disease.[16, 17] This is more so in the case of low-and-middle-income countries (LMICs). Recent meta-analysis reveals that only 4 of the listed 139 LMICs, as according to the World Bank, have the requisite number of teletherapy centers and 39.5% do not possess a single radiotherapy center.[18] The median patient access to radiotherapy centers in the remaining 80 LMICs is only a deplorable 36.7% and projections indicate that by 2020, the patient demand will exceed the facilities at these LMICs by several folds.[18] African nations have the least number of active radiotherapy centers (60% in Egypt and South Africa alone) while almost negligible data is available regarding Eastern European nations.[19] The situation is made even worse by the fact that most of these centers in LMICs do not have properly trained personnel; a study in Ethiopia showed that 81% of all the healthcare providers had basic medical training and 44% lacked access to any form of diagnostic imaging equipment.[20] Therefore the need for a simple, fast and cost-effective computed image processing system enabling detection of cancer cells from PET/ CT or MRI scans is clearly imminent. MATLAB, a commonly used image-processing and graphics-interface program in academia and industry,[21, 22] can pose as an alternate and affordable means of diagnosis. In this independent, graduate study, we aim to develop a fast, reliable and cost-effective program using MATLAB, which would enable detection of cancerous or tumorous lesions in CT images of the lungs of affected patients. In addition, the program would be able to determine the area of any specific labelled region of interest (ROI) or tumor mass, approximate nodal class, thus, facilitating quick and calculated treatment decisions and diminishing the requirement of more advanced diagnostic equipment.

2 Method

The images used for testing the MATLAB program were obtained from “The Cancer Imaging Archive” (TCIA) portal, a joint project of the National Cancer Institute (NCI) and Washington University serving as a public, open-access image database of advanced medical images of several forms of cancer.[23] Since NSCLC is the most prevalent sub-type of lung cancer,[5] 50 thoracic CT images of NSCLC patients were randomly selected, irrespective of the stages of the disease and set in a directory. The program reads each of these images automatically and are subjected to 4 basic types of modifications; filtering, conversion, edge detection and labelling.[22, 24] A detailed schematic is illustrated below in Figure 1.

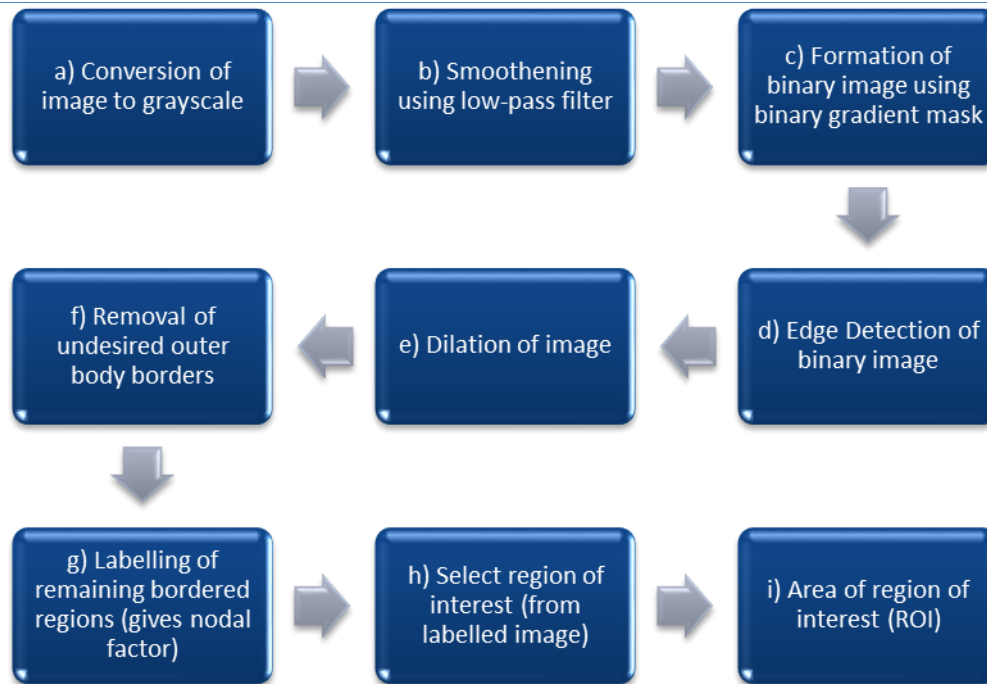


Figure 1: Schematic of program

Conventional image pre-processing techniques of conversion to grayscale, followed by smoothing of image using a customized low-pass filter which is defined by the following equation –

$$LPF(x,y) = \begin{cases} 1 & \text{if } \sqrt{X^2 + Y^2} < CF \\ 0 & \text{if } \sqrt{X^2 + Y^2} > CF \end{cases} \quad [25]$$

Where where (x, y) , is the size of the image and cf is the cut-off value for the low pass filter. After this, thresholding and segmentation techniques were performed respectively.^[24] The “smoothened” image was converted to a “black-and-white” image in order to improve contrast.^[24] A sobel operator using a fudge factor of 0.8 was used for edge detection of the latter image and then tuned to obtain a binary mask that contains the new segmented image.^[26] Sobel operator is shown as follows –

$$\nabla f(x,y) = [G_x \ G_y]^T = \left[\frac{\delta f}{\delta x} \ \frac{\delta f}{\delta y} \right] \quad [27]$$

$$mag(\nabla f) = |\nabla f_{(2)}| = [G_x \ G_y]^{1/2} \quad [27]$$

$$\theta(x,y) = \arctan\left(\frac{G_x}{G_y}\right) \quad [27]$$

The resultant edges in the image were then dilated, removed of undesired borders as much as possible and subsequently labelled.^[26, 28, 29] Selection of any labelled ROI will yield area of that particular region. The “N” value or lymph node involvement, as according to the TNM staging of lung cancer,^[17] was determined as a function of the number of labelled centroid objects. For example, if the number of labelled objects was less than 50, “N” is considered to be “0”, and if less than 150, the value of “N” is “1”. Distant metastasis and metastatic lymph node involvement cannot be determined since the images

used are confined to the thoracic cavity. Therefore the nodal value is limited to “0-2”. Being a completely independent study, the co-authors were responsible for all the finances.

3 Results and Discussion

The program was successfully able to mark and label cancerous lesions or tumor masses in the CT images of the NSCLC patients (Fig. 2), indicate lymph node involvement and determine the cross-sectional area of any given or labelled lesion or tumor mass (ROI). The values of area obtained are all scalar values. These are the summation of the total number of pixels present in the particular ROI. Using the dimensions of the actual CT slices, the exact area of any tumor mass can be easily calculated, thus, facilitating selection of “T” value and type of resection. Figure 2(a) illustrates the step-wise image process modification conducted on a particular CT image of a NSCLC patient. Figure 2(b) illustrates the selection of ROI and cross-sectional area determination.

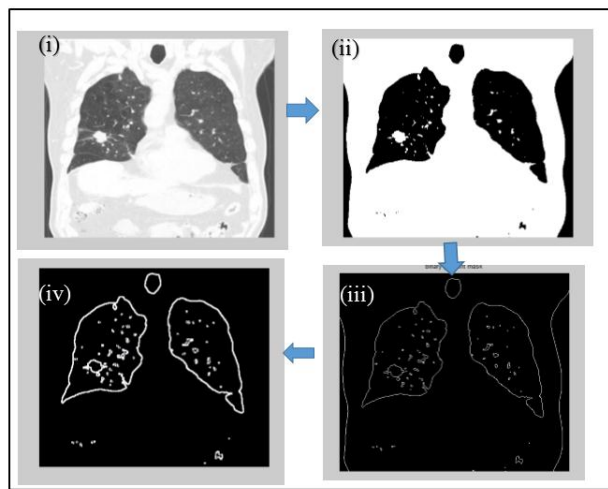


Figure 2(a): (i) Formation of filtered grayscale image; (ii) Conversion from grayscale to black-and-white binary image; (iii) Detection of borders or edges present in binary image; (iv) Dilation and labelling of the detected edges

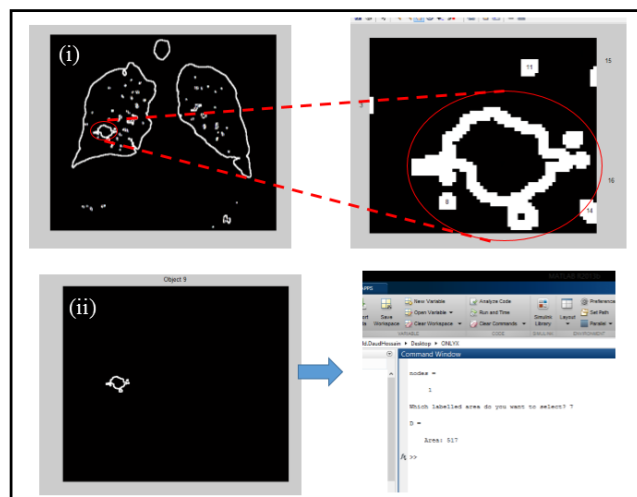


Figure 2(b): (i) Magnification of labelled ROI '7'; (ii) Selection of ROI labelled '7' and window showing cross-sectional area of ROI '7' and nodal-spread value

As mentioned earlier, an assumed criteria for lymph node involvement was used, using the principle that the degree of lymph node spread or involvement is directly proportional to the number of objects detected and labelled in the image. Since the values were assumed, they need rectification from licensed practitioners for validation. The values of cross-sectional areas of the respective ROIs obtained are scalar values. These are the summation of the total number of pixels present in the particular ROI. Using the dimensions of the actual CT slices, the exact cross-sectional area, as well as the volume, of any tumor mass detected in an image can be easily calculated. Summation of the volume of a particular tumor mass from consecutive slices will yield the actual volume of the original tumor mass, thus, facilitating an accurate and cost-effective determination of “T” since this program can be easily implemented in regular low-cost computers, making its clinical availability far superior than the current standards. Another advantage is that the program has the capability to process on all the images in succession, which the user has set in the library. One particular drawback is that this method will not be able to determine the subtype of lung cancer since histological examinations are required for that purpose.

4 Conclusion

Despite a few initial setbacks, the program is highly efficient in detecting tumorous lesions in CT images of lung cancer (NSCLC) patients and determining the cross-sectional areas of any particular lesion or tumor mass. It is adept in determining nodal spread within the thoracic cavity and, with a slight improvement, will be able to determine the volume of any tumor mass accurately. Only a facility with access to CT and a moderate configuration desktop computer is required to run the analysis. The usage is relatively straightforward and does not require highly-trained personnel, thus making it an ideal, cost-effective diagnostic alternative for LMICs. Further modifications are currently being made to incorporate histological data into program and automatically select images or CT slices of interest, though this will inevitably increase operating costs.

REFERENCES

- [1] Jemal, A., et al., *Global cancer statistics*. CA: a cancer journal for clinicians, 2011. 61(2): p. 69-90.
- [2] Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012*. International Journal of Cancer, 2015. 136(5): p. E359-E386.
- [3] Siegel, R., et al., *Cancer statistics, 2014*. CA: a cancer journal for clinicians, 2014. 64(1): p. 9-29.
- [4] Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2015*. CA: a cancer journal for clinicians, 2015. 65(1): p. 5-29.
- [5] Ettinger, D.S., et al., *Non-small cell lung cancer*. Journal of the National Comprehensive Cancer Network, 2012. 10(10): p. 1236-1271.
- [6] Smith, R.A., et al., *American Cancer Society guidelines for the early detection of cancer*. CA: a cancer journal for clinicians, 2002. 52(1): p. 8-22.

- [7] Cerfolio, R.J., et al., *The accuracy of integrated PET-CT compared with dedicated pet alone for the staging of patients with nonsmall cell lung cancer*. The Annals of thoracic surgery, 2004. 78(3): p. 1017-1023.
- [8] Mahadevia, P.J., et al., *Lung cancer screening with helical computed tomography in older adult smokers: a decision and cost-effectiveness analysis*. Jama, 2003. 289(3): p. 313-322.
- [9] Sprindzuk, M.V., et al., *Lung cancer differential diagnosis based on the computer assisted radiology: The state of the art*. Polish Journal of Radiology, 2010. 75(1): p. 67-80.
- [10] Humm, J.L., A. Rosenfeld, and A. Del Guerra, *From PET detectors to PET scanners*. European journal of nuclear medicine and molecular imaging, 2003. 30(11): p. 1574-1597.
- [11] Wisnivesky, J.P., et al., *The cost-effectiveness of low-dose CT screening for lung cancer: preliminary results of baseline screening*. CHEST Journal, 2003. 124(2): p. 614-621.
- [12] Yen, T.-C., et al., *Defining the priority of using 18F-FDG PET for recurrent cervical cancer*. Journal of Nuclear Medicine, 2004. 45(10): p. 1632-1639.
- [13] Anderson, B.O., et al., *Breast Cancer in Limited-Resource Countries: An Overview of the Breast Health Global Initiative 2005 Guidelines*. The breast journal, 2006. 12(s1): p. S3-S15.
- [14] Shyyan, R., et al., *Breast Cancer in Limited-Resource Countries: Diagnosis and Pathology*. The breast journal, 2006. 12(s1): p. S27-S37.
- [15] Keppler, J.S. and P.S. Conti, *A cost analysis of positron emission tomography*. American Journal of Roentgenology, 2001. 177(1): p. 31-40.
- [16] Corner, J., et al., *Is late diagnosis of lung cancer inevitable? Interview study of patients' recollections of symptoms before diagnosis*. Thorax, 2005. 60(4): p. 314-319.
- [17] Rami-Porta, R., J.J. Crowley, and P. Goldstraw, *Review the revised TNM staging system for lung cancer*. Ann Thorac Cardiovasc Surg, 2009. 15(1): p. 5.
- [18] Datta, N.R., M. Samiei, and S. Bodis, *Radiation therapy infrastructure and human resources in low-and middle-income countries: present status and projections for 2020*. International Journal of Radiation Oncology* Biology* Physics, 2014. 89(3): p. 448-457.
- [19] Grover, S., et al., *A systematic review of radiotherapy capacity in low-and middle-income countries*. Frontiers in oncology, 2014. 4.
- [20] Levine, A.C., et al., *Understanding Barriers to Emergency Care in Low-Income Countries: View from the Front Line*. Prehospital and Disaster Medicine, 2007. 22(05): p. 467-470.

- [21] Goering, R., *Matlab edges closer to electronic design automation world*. Electronic Engineering Times, 2004(1341): p. 4-5.
- [22] Guide, M.U.s., *The mathworks*. Inc., Natick, MA, 1998. 5: p. 333.
- [23] Clark, K., et al., *The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository*. Journal of digital imaging, 2013. 26(6): p. 1045-1057.
- [24] Gonzalez, R.C., R.E. Woods, and S.L. Eddins, *Digital image processing using MATLAB2004*: Pearson Education India.
- [25] Mustafa, W.A., H. Yazid, and S. Bin Yaacob. *Illumination correction of retinal images using superimpose low pass and Gaussian filtering*. in *Biomedical Engineering (ICoBE), 2015 2nd International Conference on*. 2015.
- [26] Maini, R. and H. Aggarwal, *Study and comparison of various image edge detection techniques*. International journal of image processing (IJIP), 2009. 3(1): p. 1-11.
- [27] Wenshuo, G., et al. *An improved Sobel edge detection*. in *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*. 2010.
- [28] Anoraganingrum, D. *Cell segmentation with median filter and mathematical morphology operation*. in *Image Analysis and Processing, 1999. Proceedings. International Conference on*. 1999. IEEE.
- [29] Shrivakshan, G. and C. Chandrasekar, *A comparison of various edge detection techniques used in image processing*. IJCSI International Journal of Computer Science Issues, 2012. 9(5): p. 272-276.