

Detection of Cancerous Nodule in Lung Using KNN Classifier

^{1,2}Sakshi Wasnik, ²Pallavi Parlewar, ³Prashant Nimbalkar

^{1,2}Shri Ramdeobaba College of Engineering and Management, Nagpur, India

³Precision Scan and Research Centre, Nagpur, India

Email: ¹wasnikss_1@rknec.edu, ²parlewarpk@rknec.edu, ³precisionscan@gmail.com

Received: 06th October 2019, Accepted: 20th November 2019, Published: 31st December 2019

Abstract

The paper focuses on detection of cancerous nodule in lung using KNN classifier for the medical images to help the physicians to interpret the computer tomography (CT) scan images. The combination of various image processing tools to effectively accomplish the required objectives. Pre-processing of an image is a done to remove the noise in the image. Threshold value is calculate using Otsu thresholding. Tumor is extracted by thresholding the image and is subtracted by the removal of small object and is extracted from the fore ground. The features like area, perimeter, major axis length, minor axis length, eccentricity and convex area are obtained. K nearest neighbour algorithm is used to find the cancerous tumour. The proposed algorithm uses ---- samples for each iteration and achieved accuracy 96.25%, sensitivity 98.36% and specificity and 89.47%

Keywords

Computer Tomography, Segmentation, Tumour, k Nearest Neighbour.

Introduction

Lung cancer is the most common cancer across the world. To save human lives it is very necessary to detect the lung cancer at an early stage. CT is one of the modest medical imaging methods to diagnose the cancer. The lungs are the essential part of the body that is use to breathe [1]. Lung cancer, also known as lung carcinoma, is a malignant tumour characterized by uncontrolled growth of the cell. It is therefore necessary to treat this to avoid spreading its growth to other parts of the body [2]. Long-period tobacco smoking is the primary factor for 85% of lung cancers [3]. Computer tomography (CT) are the conventional methods to detect the presence of lung cancer. Hence, it is essential to determine a new robust method to diagnose the lung cancer at an earlier stage [4].

The system is separated into four stages. In the first stage enhancing of CT images is done by the usage of enhancement techniques like median filter. Thus the noise in an image gets reduced and also smoothens the edges. To separate out various parts from the image, so that the tumor is clearly visible, segmentation is done. Then various features of an image are computed such as area, perimeter, major axis length, minor axis length, convex area and eccentricity so that it can be used for training of the system. After that KNN classification algorithm is applied to detect the cancerous tumour. To surpass the system, there is a need to experimentally test all the tumour and non-tumour images from the dataset.

Methodology

In the proposed system the different image processing techniques are combined in order to segment the lung tumour and also to classify the image into benign or malignant.

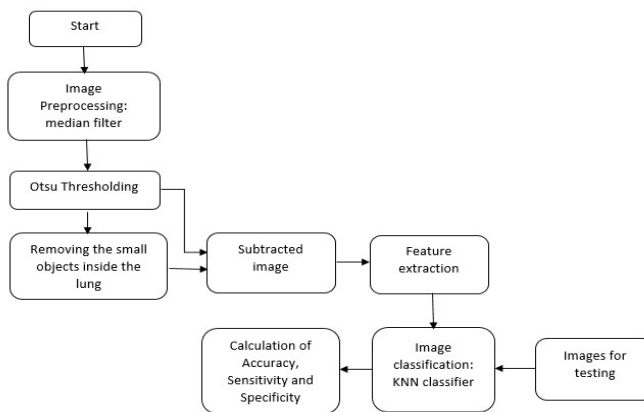


Fig 1: Flowchart of the Proposed System

1. Image Pre-processing: Median Filtering:

To reduce the noise present in an image and to produce clearer image, image preprocessing is used [5]. Median filter is used to lessen the salt-and pepper noise in an image and it normally take care of all the features and image edges [6].

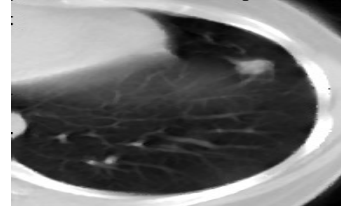
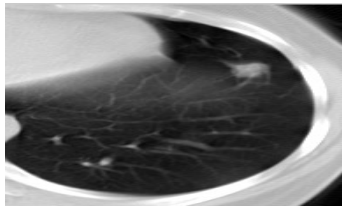


Fig.2 (a): Original Image of CT Scanned

Fig. 2 (b): Median Filtered Image Lung of Patient 1

As its name implies, it replaces the value of a pixel by the median of the intensity levels in the neighborhood of that pixels [6].

It is given by: $F(x, y) = \text{median}(S_{s,t})[g(s, t)]$ (1)

Where, $S_{s,t}$ represents the set of coordinate in a rectangular sub-image window (neighborhood) of size $m*n$, $g(s, t)$ represents original image of size $s*t$, $F(x,y)$ represents the median filtered image. Fig. 2 (b) shows the median filtered image.

2. Otsu's Thresholding

Thresholding image segmentation is an easy but strong approach to segmenting the images with light objects on dark background. The Fig. 3 shows the thresholding process of a lung CT image. Clearly, the tumor portion is properly separated from other components of the image, such as the lungs and the other structure. In our system, we have used Otsu thresholding because of its simplicity and the difference in graylevels intensities which are required for different graylevel images [7]. The Fig.3(c) shows the thresholded image.

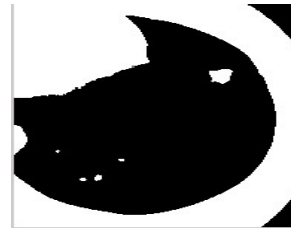
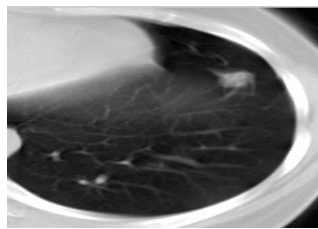
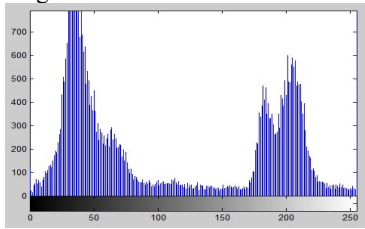


Fig.3 (a): Histogram of Original Image

Fig. 3 (b): Original Image

Fig. 3 (c): Thresholded Image

3. Small Object Removal:

Here the images start an operation of area opening. The basic effect of an opening is to remove some of the foreground (bright) pixels from the edges of region of foreground part. Using this method, the unwanted parts are removed from an image. This method is to remove all linked objects with pixels below the set value into a binary image.

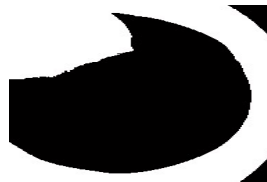
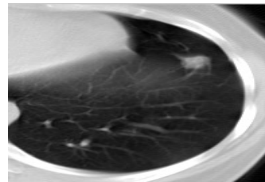


Fig. 4 (a): Original Image

Fig. 4 (b): Small Object Removed Image

It demonstrates the outcomes of using a method or function that removes objects linked to a tumor to make segmentation simpler. However, this method is used in this research to remove tumor in order to maintain only the other parts (i.e. the darker pixel intensity values) in the image. Figure 4 (b) shows small object removed image.

4. Image Subtraction:

In this stage, the previous image i.e. object removed image is subtracted from the thresholded image. Thus we will get a tumour part only, which is the segmented image.

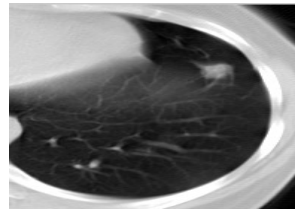


Fig. 5 (a): Original Image

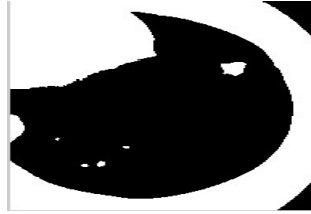


Fig. 5 (b): Thresholded Image



Fig. 5 (c): Small Object Removed Image



Fig 5 (d): Subtracted Image

The pixel values of the output are provided by: $O(m,n) = R(m,n) - S(m,n)$ (2)

Where, $O(m,n)$ represents subtracted image (Figure 5 (d)), $R(m,n)$ represents thresholded image (Figure 5(b)), $S(m,n)$ represents small object removed image (Figure 5 (c))

Or if it is simply desired to subtract a constant value C from a single image then: $O(m,n) = R(m,n) - C$ (3)

Where, C is a constant value.

5. Feature Extraction

Subsequently, the segmented area of nodules is further used for extraction of features. It is the operation to remove distinct characteristics from an image such as area, perimeter, major axis length, minor axis length, convex area and eccentricity etc. that distinguishes the most important characteristics from each other.

a) Area: The region is calculated from the tumor segmented portion containing only ' white ' pixels; that is the pixel with the value ' 1. ' [8]. It is described by a formula below: $Area(A) = n\{w\}$ (4)

Where $n\{w\}$ is the number of white pixels of the tumor portion.

b) Perimeter: Perimeter is calculated by counting the amount of pixels within the region boundary [8].

$$P = |dn d1| + \sum_{i=1}^{n-1} |didi + 1|$$
 (5)

Where $d1$ is the first parametric coordinate of the region of interest, dn is the last parametric coordinate of the region of interest

c) Convex area: The convex area of region is the total number of pixels in an image [9].

d) Major Axis Length: The major axis is the longest diameter of an ellipse. In the Figure 6, the Y-axis represents a major axis of a tumour [9].



Fig. 6: Axis Length for a Nodule

e) Minor Axis Length: The shortest diameter of an ellipse is the minor axis. In the Figure 6, the X-axis represents a minor axis of a tumour [9].

f) Eccentricity: The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1 [9].

Fields	Area (mm ² or cm ²)	Major axis length (mm or cm)	Minor axis length (mm or cm)	Eccentricity (no unit)	Perimeter (mm or cm)	Convex area (mm ² or cm ²)
1.	223	33.7913	9.2832	0.9615	74.4310	236
2.	7	3.8270	2.7591	0.6930	7.2200	8
3.	15	4.7610	4.2227	0.4619	10.8670	16
4.	29	7.6013	5.0513	0.7473	17.4170	31
5.	11	4.4969	3.2775	0.6847	8.7910	11

Table 1: Feature Extraction Parameters of Patient 1

The same procedure is repeated on patient 2 shown in Figure 7
 Patient 2:



Fig. 7 (a): Original CT Image of Patient 2

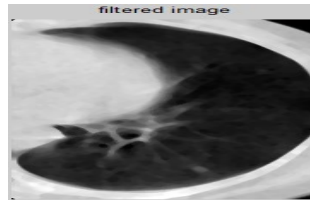


Fig.7 (b): Filtered Image

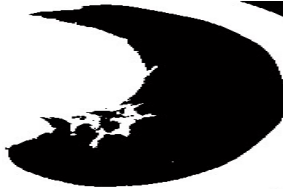


Fig. 7 (c): Thresholded Image



Fig. 7(d): Small Object Removed Image



Fig. 7 (e): Subtracted Image

Fields	Area (mm ² or cm ²)	Major axis length (mm or cm)	Minor axis length (mm or cm)	Eccentricity (no unit)	Perimeter (mm or cm)	Convex area (mm ² or cm ²)
1.	9	4.0537	3.0245	0.6658	7.6830	9
2.	1	1.1547	1.1547	0	0	1
3.	9	4.4308	2.08842	0.7591	8.2370	10
4.	1	1.1547	1.1547	0	0	1

Table 2: Feature Extraction Parameters of Patient 2

Thus from the various features of different patients, it is clear that the features which were extracted are efficient to classify the image into benign or malignant. The result of two patients are shown, and it can be concluded that the patient 1 is having a malignant tumour, while patient 2 is a benign one. Table 3 compares the results of patient 1 and patient 2 and concludes that there is a lot of difference between the malignant and benign nodule.

	Patient 1	Patient 2
AREA	223	9
MAJOR AXIS LENGTH	33.7916	4.0537
MINOR AXIS LENGTH	9.2832	3.0245
ECCENTRICITY	0.9615	0.6658
CONVEX AREA	236	9
PERIMETER	74.310	7.6830

Table 3: Comparison of Features of Patient 1 and Patient 2

6. KNN Classification

K nearest neighbours is a standard algorithm that uses the entire dataset as part of its training stage. Wherever another information comes, the prediction is necessary for a non-observed information, it will look through the complete data training set for k-most like or likelihood cases, and the data with the most similar cases is finally returned as the prediction model [10]. In this system, total 130 images are used, out of them 50 images are used for training while the 80 images are used for testing purpose. As the system is trained using MATLAB, it basically stores all the features of the trained image and accordingly decides whether the new image is belonging to class 0 or 1. With the help of k nearest neighbours, it can look the features of new images and try to figure out that in which class it belongs. Here class 0 shows that it is a benign tumour, while class 1 shows that it is a malignant tumour.

Results and Discussion

The suggested algorithm is tested on various images and through visual observation verified the accuracy of segmentation. The following outcomes for the images were achieved. By evaluating the percentages of Sensitivity (SE), Specificity (SP) and Accuracy (AC), the classification performance (KNN) algorithm can be calculated using a confusion matrix shown in Table 4 and the respective definitions are as follows.

Sensitivity: It is used to evaluate the percentage of those with the disease who test positive for the disease.
 $SE = T_p / (T_p + F_N) * 100$
 $SE = 60 / (60 + 1) * 100 = 98.36\%$ (6)

Specificity: It is used to evaluate the percentage of negatives that are recognized properly.

$$SP = T_N / (T_N + F_P) * 100, \quad (7)$$

$$SP = 17 / (17 + 2) * 100 = 89.47\%$$

Accuracy: It reflects how often the proper output is given by the classifier.

$$AC = (T_P + T_N) / (T_N + T_P + F_N + F_P) * 100, \quad (8)$$

$$AC = (60 + 17) / (17 + 60 + 1 + 2) * 100 = 96.25\%$$

	Predicted No	Predicted Yes	
Actual No	$T_N = 17$	$F_P = 2$	19
Actual Yes	$F_N = 1$	$T_P = 60$	61
	18	62	

Table 4: Confusion Matrix

Whereas the terms T_N , T_P , F_N , F_P are having a specific meaning.

True positives (T_P): Predicted that there is a disease in an individual and they have a disease as well.

True negatives (T_N): Predicted that there is no disease in an individual and that they don't have a disease.

False positives (F_P): Predicted in this case that a person has a disease, but they don't actually have a disease.

False negatives (F_N): Predicted in this situation that the individual does not have a disease, but they do.

Conclusion

This work discusses the need for techniques and tools that help doctors optimize the screening and diagnostic workflow of lung cancer to improve the precision of cancer diagnosis and decrease exposure to costs and radiation. The proposed algorithm is tested under various condition and found to be 96.25% accurate for classifying the nodule using KNN classifier. The sensitivity and specificity of the proposed system is 98.36% and 89.47% respectively.

References

1. Gulshan Sharma, James Goodwin, "Effect of aging on respiratory system Physiology and Immunology", *Clinical Interventions in Aging*, vol. 1, issue 3, pp. 253-260.
2. A. A. Brindha, S. Indirani, and A. Srinivasan, "Lung cancer detection using SVM algorithm and optimization techniques," *Journal of Chemical and Pharmaceutical Sciences*, vol. 9, no. 4, 2016.
3. M. Kurkure and A. Thakare, "Introducing automated system for lung cancer detection using Evolutionary Approach," *International Journal of Engineering and Computer Science*, vol. 5, no. 5, pp. 16736-16739, 2016.
4. B. Rani, A. K. Goel, and R. Kaur, "A modified approach for lung cancer detection using bacterial forging optimization algorithm," *International Journal of Scientific Research Engineering and Technology*, vol. 5, no. 1, 2016.
5. B. Chitradevi, P. Srimathi, "An overview of image processing techniques", *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, issue 11, Nov 2014.
6. Unnikrishnan R., Pantofaru C., Hebert M., "Towards objective evaluation of image segmentation algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 929-944, 2007.
7. Shakuntala Satyawana, Surendra Kumar Agarwal, "A Review Paper on Image Segmentation and Object Recognition Procedures," *International Journal of Science, Engineering and Technology*, vol. 4 Issue 1, 2016.
8. Sakshi Wasnik, Pallavi Parlewar, Prashant Nimbalkar, "Nodule detection in lung using multi-threshold segmentation," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, ISSN: 2278-3075, vol. 8, Issue. 9, July 2019.
9. Khin Mya Mya Tun, Aung Soe Khaing, "Implementation of Lung Cancer Nodule Feature Extraction Using Digital Image Processing," *International Journal of Scientific Engineering and Technology Research*, ISSN 2319-8885 vol. 3, Issue. 9, May 2014.
10. Frank C. Detterbeck, Peter J. Mazonne et al., "Screening for lung cancer diagnosis and management of lung cancer", *American college of Chest Physicians evidence-based clinical practice guidelines*, May 2013.