Lung Cancer Identification Using CT Images

Sunanda Biradar[1], Kavya Agalatakatti[2]

[1]Assistant Professor, Dept of CSE, B.L.D.E.A's CET, Vijayapur
sunanda_biradar@rediffmail.com

[2]Dept of CSE, B.L.D.E.A's CET, Vijayapur

kavyabec09@gmail.com

Abstract*: Lung cancer is one of the most difficult and dangerous cancer to cure and the number of deaths due to cancer is increasing day by day. The death rate of patients will be reduced if lung cancer is detected in the early stages. This paper aims at developing a Computer Aided Diagnosis (CAD) system for detection of lung cancer by analyzing the Computed Tomography (CT) images of lungs. The images are collected from the LIDC dataset and enhanced to increase the contrast of images by Median filter. After enhancement morphological segmentation is used to segment lungs region from the CT images. Then segmented images are used to identify and classify the cancerous and noncancerous nodules using the Support Vector Machine (SVM) classifier. The SVM polynomial has given the 96.6% accuracy and SVM quadratic function has given the 92% accuracy.*

Keywords: CT images of Lung, Morphological Segmentation, Feature Extraction, and SVM Classifier.

1. Introduction

Medical imaging is a technique and process of creating visual representations of interior of a body for clinical analysis and medical intervention. It helps to disclose internal structures hidden by skin and bones, and it is also used to diagnose and treat disease. Many types of medical images are their like Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound images etc. in our proposed method CT images are used. Computed Tomography (CT) images are also called as "CAT scanning" (Computed Axial Tomography). It provides different forms of imaging known as cross-sectional imaging. A CT imaging system produces the cross sectional images or slices of anatomy, like the slices in a loaf of bread. The cross sectional images are used for a variety of diagnostic and therapeutic process. A CT scan takes data from several X-ray images of structures inside the body and then converts them into pictures. Lung cancer is the most cause of cancer death in the worldwide. Uncontrolled cell growth causes cancer. Lung cancer occurs for out-of-control cell growth in one or both lungs. Cigarette smoking is the one of reason for lung cancer. The growth of cell can form a tumors. Tumors are either non-cancerous or cancerous. Identification of cancerous image is a great challenge.

C. Bhuvaneshwari et al. [1] has detected and classified the diseases of pleural effusion and normal lung cancer. Here, they have used feature extraction through moment invariants for classification of lung diseases. Then, they have used genetic algorithm for feature selection technique. Then naive bayes and decision tree classifier are used to classify the result. Here, the decision tree classifier has given the more accuracy then the navie bayes classifier. Kuruvilla et al. [2] proposed early detection of lung cancer using

mathematical morphological operations. This paper introduces a Computer Aided Detection method for CT images of lungs. The preprocessed method is used for denoising, and then lung region is segmented using morphological operations from the CT images. Nitesh Sharma Sheenam et al. [3] proposed the OTSU'S thresholding technique to segment the lung images. And then they are calculated hope shape parameters. Based on this parameter the lung cancer is classified using the back propagation network. Muthukumaran K et al. [4] proposed the basic mathematical morphological theory and operations for the edge detection and cancer cell identification. The results show the proposed algorithm is more efficient for denoising, edge detection and identification of cancer cells than the usually used edge detection algorithm and morphological edge detection algorithm. Ginneken, et al. [5] introduces ANODE09, which is available in website http://anode09.isi.uu.nl. A dataset contains 55 scans from a lung cancer screening program and web-based framework for objective evaluation of nodule detection algorithm. It deals with the automatic detection of pulmonary nodules in computed tomography scans. Results show a substantial performance difference between algorithms and demonstrate that combining output of algorithm leads to marked performance improvement. Kostis et al. [6] proposed small pulmonary nodules are common radiographic. Based on the nodule growth they have distinguishing the malignant and non-malignant pulmonary nodules. Penedo et al. [7] proposed the CAD system, based on two-level artificial neural network architecture. Curvature peaks computed for all pixels in each region in used as input for the second ANN. Using these values they have segmented the lung region and classified the cancerous and noncancerous lung nodule in the CT image.

Al-Kadi et al. [8] proposed the contrast-enhanced computed tomography (CT) images to differentiate between aggressive and non aggressive malignant lung tumor. They have considered a fractal process as branching of blood vessels. The classification accuracy is 83.3%. Mesanovic et al. [9] proposed automatic region based segmentation for CT images. This algorithm starts with selecting the initial thresholding for segmenting the lung images. After giving the initial threshold value it set to all pixels under the selected value to the background. It looks for first threshold value if it is found the region will grow. Likhitkar et al. [10] proposed an automated Computer Aided Diagnosis (CAD) system is proposed for the detection of lung nodule from the computed tomography (CT) images. They have considered the growth rate, density, boundary of the nodule as feature values. Then the lung nodule is classified as cancerous and noncancerous with the help of CAD system. Nathaney et al. [11] proposed the development of programmed system for detecting lung cancer and to build more efficient outcome within an interactive time frame with least false negative rate. A new approach is used to segment the images and to classify focal areas in lung region. Instead of simple thresholding, adaptive threshold method is applied images and ROI processing is used to segment the cautious region. Akkasaligar et al. [12] proposed the Classification of medical ultrasound images of kidney using k-nearest neighbour's classifier (k-NN). Here, they have used median filter to enhance the images. Then texture features are extracted. After that cystic and normal kidney's are identified using k-NN classifiers.

2. Methodology

The methodology of the proposed work consists of six stages, which are shown in the block diagram of the proposed work.

```
┌─────────────────────────┐
│   Image Acquisition     │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│     Enhancement         │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│     Segmentation        │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│   Feature extraction    │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│ Cancer cell identification │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│       Analysis          │
└─────────────────────────┘
```
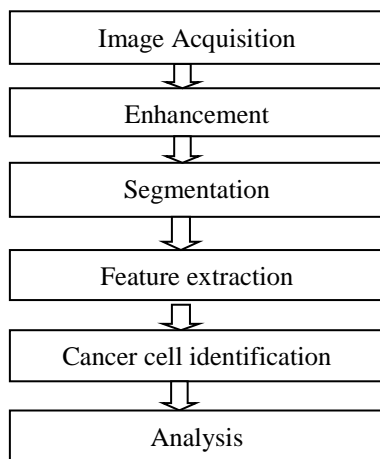
Figure 1: Block diagram of proposed work

The images are collected from the public dataset Lung Image Database Consortium (LIDC) which is available in the website http://imaging.nci.nih.gov/ncia/. The dataset contains two types of images such as cancerous and noncancerous images. In the proposed method median filtering technique is used to remove the noise. It replaces each pixel with the median of neighbouring pixels. The preprocessed image is converted into binary image. All the pixels in the input image with a intensity greater than a threshold level is

replaced with value '1'and all pixel values with a intensity less than threshold level is replaced with value '0'. The threshold level is calculated by Otsu method. Segmentation means it subdivides the region into small regions. In our proposed method we have used morphological opening operation for segmentation. It extracts the image components for representation of region such as shape, boundaries using structuring element. The structuring element is used to probe an input image. Structuring element is a matrix consisting of 0's and 1's that can have any arbitrary shape or size. The centre of the structuring element pixel is called origin of the image and it identifies the pixel of interest. The morphological opening of P by Q is given by P Q. It is defined as erosion of P by Q followed by dilation of the result by Q.

$$P \circ Q = (P \ominus Q) \oplus Q$$

The morphological opening operation will completely remove the region of an object that cannot contain the structuring element. For extracting the image regions we have used three structuring elements.

1. Se1 is periodicline, it contains the $2 \times P + 1$ elements.
2. Se2 is line of length 2 and degree 0.
3. Se3 is line of length 2 and degree 90.

The steps involved in the process of segmentation are represented by Algorithm 2.1

Algorithm 2.1: Segmentation by morphological operation
  Input: Medical CT image of lung.
  Output: Segmented region of lung.
  Start
  Step 1. Read input image.
  Step 2. Convert to grayscale image.
  Step 3. Apply FIR filter for enhancement
  Step 4. Resize the image
  Step 5. Convert it to binary image
  Step 6. Apply morphological operation
  Step 7. Invert the image obtained in previous step
  Step 8. Apply clear border operation
  Step 9. Segmented output
  Stop

After applying inversion and clear border operation, we get the segmented image. The segmented image is shown in the figure. 2(c).
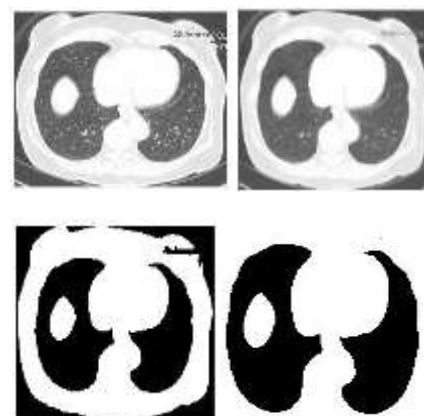
Figure 2: (a) Original image (b) filtered image (c) After morphological operation (d) Segmented image

Then segmented image is superimposed on to the original image to extract the features of the image. Feature extraction is very important and essential step to extract properties of region of interest (ROI). In our proposed method statistical features are calculated. The mean, standard deviation, skewness, kurtosis, fifth central moment and sixth central moment are considered as feature values. The fifth and sixth central moments are considered as high order moments. This will provide the information about the distribution. Then we have created the feature vector for these values. The feature vector is given as input for the SVM classifier.

2.1 Mean

m is the mean of the pixels present in the image.

$$m = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} c(i,j) \tag{1}$$

2.2 Standard deviation

The standard deviation, s is the mean square deviation of the calculated pixel value c (i, j) from its mean value m.

$$s = \sqrt{\frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} (c(i,j) - m)^2} \tag{2}$$

2.3 Skewness

Asymmetry in statistical distribution is called skewness.

$$sk = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \left[\frac{(c(i,j)-m)^3}{s} \right] \tag{3}$$

2.4 Kurtosis

Kurtosis determines the peakness of distribution related to the normal distribution.

$$ks = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \left[\frac{(c(i,j)-m)^4}{s} \right] \tag{4}$$

2.5 Central moments

Central moments provide clear information about the distribution's shape. The higher central moments are fifth and sixth central moments.

2.5.1 Fifth central moment

$$g = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \left[\frac{(c(i,j)-m)^5}{s} \right] \tag{5}$$

2.5.2 Sixth central moment

$$h = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \left[\frac{(c(i,j)-m)^6}{s} \right] \tag{6}$$

3. Classification

We have used the segmented image for the classification further. Support Vector Machines (SVM) are supervised learning models with associated learning algorithms that analyze data and recognize patterns. It is used for classification and regression analysis. A support vector machine builds a hyper plane or set of hyper planes in a high- or infinite-dimensional space, which is used for classification. If hyper plane is achieved the good separation and that has the largest distance to the nearest training data points of any class. If classifier having the larger margin then generalization error will be low. The SVM classifier differentiates the non-cancerous from cancerous lung image. A classification task usually contains training and test sets which consist of data objects. Each object in the training set contains one target value also called as class label and lot of attributes or features. The goal of a classifier is to produce a model able to define target values of data instances in the testing set, for which only the attributes are known. For classification the 60% of the images are given to the training set and 40% of the images are given to the testing set.

4. Experimental results

The statistical parameters are calculated for segmented lungs and given as input to the Support vector machine. The training set for the SVM consists of 60% of the total images and the testing set is 40% of the total images. Images for training and testing refer to the segmented single slices containing 2 lungs. In the testing set the features are calculated for each image then it is given as input to the classifier. The SVM will classify the lung image as cancerous and noncancerous based on the feature values. The True positive rate and True Negative rate and accuracy are calculated for both SVM polynomial and quadratic kernel functions.

4.1 True Positive rate

The true positive rate is also called as sensitivity. It determines the proportion of actual positive values which are correctly identified.

True Positive Rate (TPR) $= \frac{TP}{TP+FN}$

4.2 True negative rate

The negative rate is also called as specificity. It determines the proportion of negative values which are correctly identified.

True Negative Rate $= \frac{TN}{TN+FP}$

4.3 Accuracy

It determines the statistical measure of how well a classifier has correctly classified the images. The accuracy is the percentage of true results of both true positive and true negative values.

Accuracy $= \frac{TP+TN}{TP+TN+FP+FN}$

The classified results are shown in the bellow table. The table shows the results of both SVM polynomial and quadratic functions in that the SVM polynomial gives the more accuracy than the SVM quadratic function.

Table 1: Classification results

Parameters	SVM Polynomial	SVM quadratic
True Positive Rate	96.9%	90.9%
True Negative Rate	95%	93%
Accuracy	96%	92%

Conclusion

A computer aided segmentation and classification method is proposed. Morphological operations are used for segmentation and classification is done by Support vector machine. The statistical parameters are used as features for classification. The proposed method is classified using two kernel functions. The polynomial kernel function has given the more accuracy than the quadratic kernel function. The work can be used to identify the stages of the lung cancer and the work can also be extended for identification of cancer in other organs of the human body. We can also find out the type of the cancer present in the image as a future extension of this work.

References

[1] C.Bhuvaneswari P.Aruna, D.Loganathan, "Classification of Lung Diseases by Image Processing Techniques Using Computed Tomography Images", pp.2277-7970, March 2014.

[2] Jinsa Kuruvilla, Dr.K.Gunavathi, "Detection of Lung Cancer using Morphological Operations", pp. 2229-5518, 2013.

[3] Nitesh Sharma, Sheenam, "Size Estimation of Lung Cancer Using Image Segmentation and Back Propagation", pp. 2348-1439, 2014.

[4] Sivappriya T, Muthukumaran K, "Cancer Cell Detection Using Mathematical Morphology", pp. 2320-9798, 2014.

[5] B. van Ginneken, et al, "Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study, Medical Image Analysis", pp.707-722, 2010.

[6] W.J.Kostis, A.P.Reeves, D.F.Yankelevitz, C.l. Henschke, "Three dimensional segmentation and growth rate estimation of small pulmonary nodules in helical CT images", IEEE Transactions on Medical imaging 22 pp.1259-1274, 2003.

[7] R.C.Gonzalez, R.E.Wood, "Digital image processing", 3rd ed., Pearson Prentice Hall, 2008.

[8] M.G.Penedo, M.J.Carreira, A. Mosquera, "Computer-aided diagnosis a neural-network-based approach to lung nodule detection, IEEE Transcations on Medical imaging", pp. 872-880, 1998.

[9] O.S. Al-Kadi, D. Watson, "Texture analysis of aggressive and non aggressive lung tumor CE CT images, IEEE Transactions Biomedical Engineering 55", pp. 1822–1830, 2008.

[10] Mr. Vaibhav K. Likhitkar, Prof. Ujwalla Gawande, Mr. Kamal O. Hajari, "Automated Detection of Cancerous Lung Nodule from the Computed Tomography Images", pp.5-10, 2014.

[11] Ms. Gangotri Nathaney, Prof. Kanak Kalyani, "Lung Cancer Detection System on Thoracic CT Images based on ROI Processing", 2015.

[12] Prof. Prema T. Akkasaligar, Prof. Sunanda Biradar, "Classification of Medical Ultrasound Images of Kidney", 2014.

Author Profile

Sunanda Biradar received B.E. and M.Tech degrees in Computer Science and Engineering. Pursuing P.hd in Medical images and working as Assistant Professor in Dept, of Computer Science and Engineering, BLDEA CET, Vijaypur.

Kavya Agalatakatti received B.E degree in Information Science and Engineering from Basaveshwara Engineering College, Bagalkot. Pursuing M.Tech in Computer Science and Engineering in BLDEA CET, Vijaypur.