SEGMENTATION OF LUNG NODULE IN CT DATA USING K-MEAN CLUSTERING

[1]ANIKETBOMBALE, [2]C.G.PATIL

Sinhgad Academy Of Engineering

Abstract- Lung cancer is a rapidly growing deadly disease among human beings. Lung cancer occurs because of smoking, radon gas, air pollution, asbestos,and genetics. The disease can be cured if detected in early stage. The proposed algorithm in this work performsfiltering of CT images followed by morphological operations to extract actual lung region in an image.K-means segmentation is performed to detect probable cancerous areas. Features of this probable area are extracted to perform classification using support vector machines as normal, moderate or severe. The results of the method are encouraging to have an automated method for detection.

Index Terms- Computed Tomography, K- mean clustering, SVM

I. INTRODUCTION

Cancer is a group of diseases begins in cells that are the basic building blocks of a body. There are different types of cancers but all starts with the cells growing out of control. The type and stage of cancer can be determined by determining where the abnormal growth occurs in the body and the type of cells thatstart to grow abnormally. The type of cancers often found in adults isendometrial cancer, skin cancer, lung cancer,ovarian cancer andbreast cancer. Of these cancers, the most common type of cancer is the Lung cancer.This type of cancer is common due to smoking.The earlier it is detected; the better is the chance of curing. According to the estimated statistics released by American Cancer Society in the United States for May 2016 about 1 out of 4 cancer deaths are from lung cancer. Among both men and women lung cancer by far is the leading cause of cancer deaths.There are two major types of lung cancer: small cell lung cancer and non-small cell lung cancer. Out of this non-small cell, lung cancer is very often found.For lung cancer detection one of the most important and fundamental steps is screening. Screeningis the process used for identification of nodule. A nodule is a white color spot present on lungs that is visible on an X-ray or Computed Tomography (CT) scansImages. Lung nodules will appear in X-ray or CT scan image if and only if its diameter is about 1 cm.A nodule may be of two types: Either a benign or a mass. A nodule that is 3 cm or less in diameter is called a Pulmonary or benign nodule. These types of nodule are non-cancerous. Another type of nodule whose size is larger than 3 cm is in diameter is called as a lung mass. This type of nodule is more likely to be cancerous and needs to be detected as early as possible.

II. RELATED WORK

Detection of anodule is one of the fundamental problems but is a very important step in medical image processing, In ParinazEskandarian, JamshidBagherzadeh [1], a nodule detection system based on support vector machine is implemented.The main objective of this research was to reduce the errorrate. Errors may bethe image of cancer nodule detectionSystem which may recognizea cancerous image as a normal image (false negative) or a healthy image as a cancerous image(falsepositives).

In [2] Imran FareedNizami, SaadUlHasan and Ibrahim Tariq Javed presented a new technique for Segmentation of lung region from CT scan.They employed Wavelet Packet Frame (WPF)technique to acquire spatialfrequencyrepresentations and applied k-Means clustering for better segmentation of lungtissues.This proved the technique to be robust and is able to efficiently segment lung regions from multiple images from various scans. In [3] Hao Han, Lihong Liproposed a new Computer Aided Design for easy, fast and accurate detection of pulmonary nodules in CT scans. Based on their previous work on self-adaptive VQ for image segmentation [4], they developed a hierarchical high-level VQ scheme to detect INC's. With high-level VQ,it becomes possible to change the thresholding scheme that is used to extract lungs with lower processing time and higher accuracy. The SVM classifier results show that gradient features Hessian, geometric, and intensity performed best against any other group. The forward feature selection strategy indicates that the SVM classifiers performed best in "Gradient + Intensity" feature space insteadof any other feature combinations. According to [5] Awai et al. researched that the detection rate oflung cancer is 2.5 to 10 times greater using CT than by using previous analogradiography.However,due to the number of patients increasing day by day it is the workload of radiologists who need to analyze the tests in a short time is also increasing. Due to this,the radiologists may misinterpret causing errors in detection. Therefore, CAD systems that can detect nodules efficiently and effectively within a short duration of time is needed. The two main CAD systems used byradiologists to assist them, they are:

CADe– These systems are used only to detect a tumor.CADx– Theses are used to check the characteristics of a tumor. Earlier these systems were separate but now a day these systems are combined to a single system and are called as a Computer - Aided Detection and Diagnosis (CAD).

III. THEORY OF TECHNIQUES

A. COMPUTED TOMOGRAPHY(CT) SCAN

Computed Tomography (CT) scan is a technique in which a 3 cross-sectional view of an image is derived by capturingseveral two-dimensional x-ray images of an object.A CT scan continuously transmits a narrow width beam through the body that gives more detailed information than the standard X-ray. The advantage of CT scan over conventional X-ray is it is able to distinguish tissues inside the organ. [2]

B. Low-dose computerized tomography (ldct):

An LDCT scan looks for abnormalities in the lungs that could be cancer or may turn into it. A conventional CT scan is more sensitive than LDCT but still LDCT is preferred over CT because of its ability to continuously acquire data, less scanning time and lower radiation exposure dose. In LDCT a 3-dimensional image of the lungs is taken and nodules with size as small as that of a grain of rice can be detected. As LDCT scans are very much sensitive they always show nodules that may not be cancer. These nodules need to be followed over time to check if they are growing. The larger the nodule more is its possibility of being cancer. Thus, a nodule needs to be under observation.

C. Margin Of Nodule

The margin of nodule means the area or boundaryor edge where the nodule is in contact with the normal lungtissue. The margins of many cancerous nodules are uneven and look spiky and are generally termed as speculated. Most of the nodules which are non-cancerous havea very smooth or round margin.

D. K-Means Clustering

The K-means algorithm is the superior algorithm to solve the clustering problem. The k-Means algorithm runs multiple times to make a group of a cluster. The algorithm works in following steps:

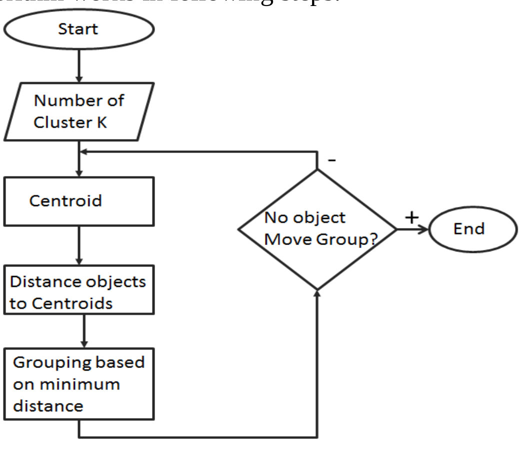


Fig (1). Flow Chart of K-mean clustering algorithm

Step 1: Initially begin by assuming the value of K (no. of clusters).

Step 2: Then divide the data into k partitions known as clusters. The training samples may be assigned systematically or randomly by using the following procedure:

1. Single - element clusters may be taken as the first k training samples.
2. Then the remaining (N-K) samples may be assigned with the values nearest to the centroid. After assigning each sample, the centroid is re-computed.

Step 3: Now take each sample one at a time and re-compute its distancefrom the centroid of the nearby clusters. If the sample does not belong to the cluster with the nearest centroid then shift this sample to that cluster with the nearest centroid and update the clusters which are gaining and losing the sample.

Step 4: Repeat this step 3 until all the samples have been analyzed.

III. METHODOLOGY

Here, the proposed method is based on detection of a nodulefrom CT images using K-mean clustering algorithm. The two most important tasks for detecting lung nodules are segmentation and clustering. This section details the proposed computer-aided detection of pulmonary nodules described. A high-level diagram of the proposed system in Figure 2 is shown.

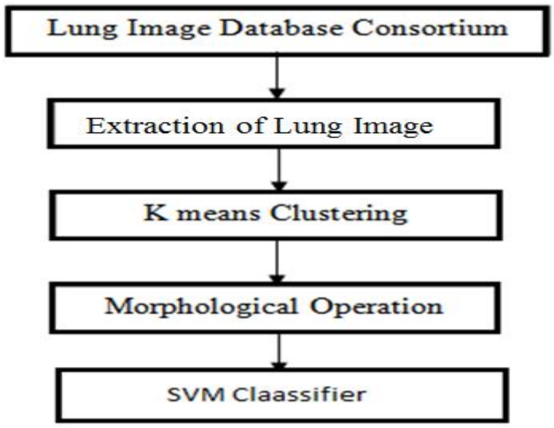


Fig.2: Flowchart of proposedalgorithm

The proposed system for detection of pulmonary nodules consists of following stages:

IV. LUNG IMAGE DATABASE CONSORTIUM (LIDC)/IMAGE DATABASE RESOURCE INITIATIVE (IDRI)

The Lung Image Database Consortium (LIDC)or Image Database Resource Initiative (IDRI) is the name given to a DataBase Management System (DBMS) by National Cancer Institute (NCI) that is available publically for the medical imaging research community for image processing or CAD based algorithms.

PRE-PROCESSING: The pre-processing isdone before the main data is processed. The main objective of pre-processing is to improve the quality of the image that may be corrupted due to noise during data acquisition. Some of the important techniques used for data pre-processing are Median Filtering [07],Histogram Equalization [07], Fast Fourier Transform [08], WaveletTransform [08], and Noise Correction [09].

SEGMENTATION: Segmentation is a method used to separate outpart of an image. Here, the segmentation is carried out to extract only the part of lungs from the CT scan image. Some of the important techniques used for image segmentation are Thresholding,Compression based method, Clustering, Histogram-based method, Edge detection, Region growing and Graphical Partitioning method.

NODULE DETECTION: Nodule detection is the most important step in the detection of lung cancer. nodule detection includes its identification and its location in the lung field. The success of this process greatly depends on the previous stages. 4 types of nodules given byKostis, W.J., Reeves, A.P., Yankelevitz, D.F., et al.in [10] are,

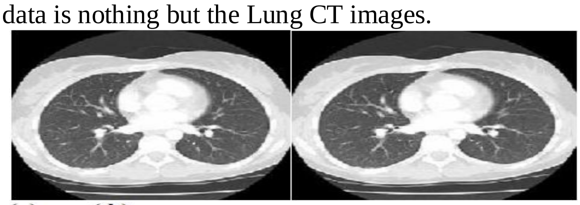
i. Well-circumscribed:In this case, the nodules are not connected to vasculature but are at the core of the lung tissue.
ii. Juxta - vascular:In this case, the nodules are at the center of the lung field and are connected to thesurrounding lung vessels.
iii. Pleural Tail:These types of nodule areconnected by a thin structure and are located near the pleural surface.
iv. Juxta-pleural:Here a thin structure is connected by the substantial portion of the nodule.

NODULE CLASSIFICATION:

After the nodule detection, the next step isthe classification of the nodule as benign or malignant. The nodules are classified based on their morphological features like Shape and size, appearance and growth rate. Support vector machine is as a classifier. Most of the pulmonary nodules are benign but may represent an early stage of lung cancer. If a malignant noduleis detected at an early stage the survival rate of the diseased may increase.The proposed CAD system is designed.

Support Vector Machine (SVM):

SVM Consist of learning model and classification model. The Learning model understands the data with a limited number of inputs. The classification model classifies the input data based on learning model. The data is nothing but the Lung CT images.



(a) (b)

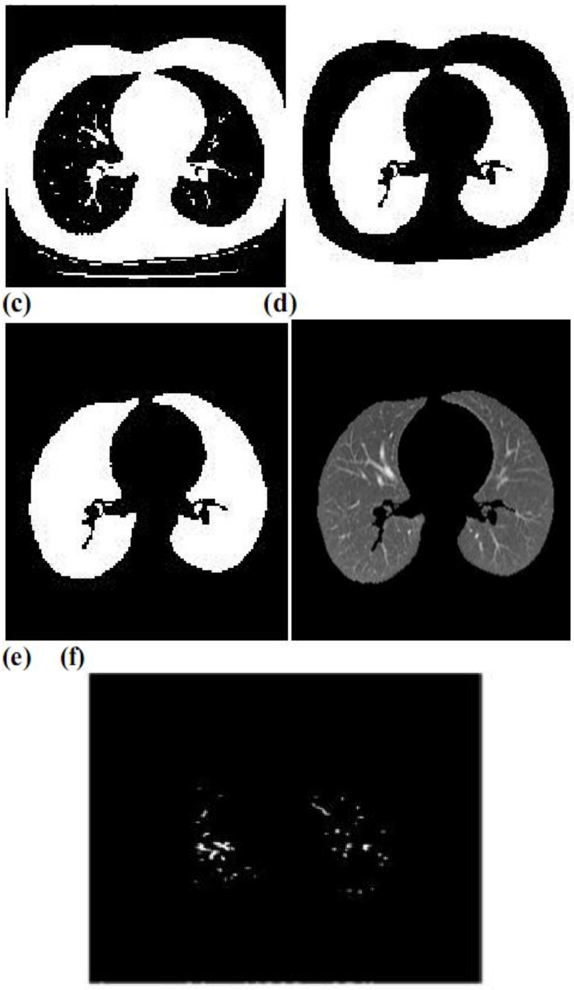


Fig.3: Extracted Lung Image

Fig(a)Input CT Image (b)Enhanced Image after Pre-Processing (c) Binary Image (d) Hole Filled Image (e) Approximate Lung Image (f)Clusterized lung image (g)Extracted nodules

V. EXPERIMENTAL RESULT

The proposed CAD system is evaluated on datasets taken from hospitals in the metro region and publically available datasets.For this study, we collected 525 samples of chest CT scans. These samples are arranged randomly forming the dataset. The dataset containing lung CT images is classified into stages of cancer such assevere, moderate, normal using SVM classifier. In this particular aspect, this study is unique.

$$\text{Detection Rate} = \frac{\text{Correctly detected samples}}{\text{Total no. of samples}}$$

Table 1. Evaluation of Proposed System

Total no. of Samples (525)	Detected (494)	Not Detected (31)	Detection Rate 94.09
Severe	222	13	94.46

(235)			
Moderate (150)	140	10	93.33
Normal (140)	132	8	94.28

The comparison with ground truth shows that the results are very good and well comparable with the manually segmented images. The results of proposed systems are shown in Fig.3. Most of the samples are segmented correctly.

CONCLUSION

The CT images are collected from the hospital. The different images are tested. The images are classified according to normal, moderate and severe. The automated segmentation and detection of lung cancer by using SVM classifier make it easy for Doctor to diagnosis a patient.The Proposed CAD system yields the good detection accuracy and computational speed in comparison with existing systems. The outcome of proposed CAD system yields an overall accuracy of around 94.09%, which demonstrates the quality of system for medical utility.

REFERENCES

[01] ParinazEskandarian,JamshidBagherzadeh "Computer-Aided Detection of Pulmonary Nodulesbased on SVM in Thoracic CT Images", IKT2015 7th International Conference on Information and Knowledge Technology.978-1-4673-7485-9/15/$31.00 ©2015 IEEE

[02] Imran FareedNizami*, SaadUlHasan†, Ibrahim Tariq Javed, "A Wavelet Frames + K-means based AutomaticMethod for Lung Area Segmentation in MultipleSlices of CT Scan "ISBN: 978-1-4799-5754-5/14/$26.00 ©2014 IEEE 245

[03] Hao Han, Lihong Li,Fast and Adaptive Detection of Pulmonary Nodules in Thoracic CT Images Using a Hierarchical Vector Quantization Scheme", IEEE Journal of Biomedical and Health Informatics 2168-2194 (c) 2013 IEEE.

[04] D. Chen, Z. Liang, M. R. Wax, L. Li, B. Li, and A. E. Kaufman, "A novel approach to extract colon lumen from CT images for virtual colonoscopy," IEEE Trans. Med. Imag., vol. 19, no. 12, pp. 1220-1226, 2000.

[05] Kazuo A, Kohei M, Akio O, Masanori K, Haruo H, Shinichi H, Yasumasa "Pulmonary nodules at chest CT: effect of Computer-aided diagnosis on radiologists' detectionperformance. Radiology 2004, 230:347–352.

[06] Gomathi M, Thangaraj P: Computer aidedmedical diagnosis system for detection of lung cancer nodules: a survey. Int J ComputIntell Res 2009, 5:453–467.

[07] Ashwin S, Kumar SA, Ramesh J, Gunavathi K: Efficient and reliable lung nodule detection using a neuralnetwork based computer aided diagnosis system. In Emerging Trends in Electrical Engineering and EnergyManagement (ICETEEEM), 2012 International Conference On. Chennai; 2012:135–142.

[08] Ye X, Lin X, Dehmeshki J, Slabaugh G, Beddoe G: Shape based computer-aided detection of lung nodules inthoracic CT images. Biomed Eng IEEE Trans 2009, 56(7):1810–1820.

[09] Arimura H, Magome T, Yamashita Y, Yamamoto D: Computer-aided diagnosis systems for brain diseases inmagnetic resonance images. Algorithms 2009, 2(3):925–952.

[10] Kostis, W.J., Reeves, A.P., Yankelevitz, D.F., et al., "Three-dimensional segmentation and growth-rateestimation of small pulmonary nodules in helical CT images", IEEE Trans., Medical Imaging 22, pp.1259–1274 ,2003

[11] El-Baz A, Beache GM, Gimel'farb G, Suzuki K, Okada K, Elnakib A, Soliman A, Abdollahi B: "computer-aided diagnosis systems for lung cancer: Challenges and methodologies". Int J Biomed Imaging 2013, 2013:1–46.

★★★