ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization

Vol. 5, Special Issue 4, March 2016

National Conference on Signal Processing, Instrumentation and Communication Engineering (SPICE' 16)

Organized by

Dept. of ECE, Mar Baselios Institute of Technology & Science (MBITS), Kothamangalam, Kerala-686693, India

Computer Automated System for Effective Lung Nodule Classification from CT scan Image Using Image Processing Techniques

Angel Varghese[1], Preena K.P[2]

PG Student [IP], Dept. of CSE, Model Engineering College, Ernakulum, Kerala, India [1]

Assistant Professor, Dept. of CSE, Model Engineering College, Ernakulum, Kerala, India [2]

ABSTRACT: Lung cancer is the leading cause of cancer death worldwide. As occurs in almost all types of cancer, its cure depends in a critical way on it being detected in the initial stages, when the tumor is still small and localized. This paper proposes a new Computer Aided Diagnosis (CAD) system for detecting nodules from the slice of DICOM Lung CT images for the identification of cancerous nodules.CAD system helps to improve the diagnostic performance of radiologists in their image interpretations. The basic steps involved in the CAD systems are 1)Pre-processing of the input CT image 2)Segment the area of interest (lung) and then analyzes the separately obtained area for nodule detection in order to diagnosis the disease 3)Feature extraction 4)Classification using a classifier. The proposed method gives much better results as compared to previous methods.

KEYWORDS: Feature Extraction, Classification, SGLDM

I. INTRODUCTION

Lung cancer becomes the prime factor in cancer deaths due to increasing rate of smoking and air pollution in different countries. Lung cancer stands highest in mortality rate among all other types of cancer. Lung cancer appears in the form of nodules or lesions in the lungs that are either solitary or attached to the lung wall (juxtapleural nodules), major vessels, or other structures. Survival chance is directly proportional to the growth at its detection time. Earlier the detection, higher the chances for survival. Survey report shows that 31% of cancer deaths for males and 26% for females are caused by the lung cancer. There has been a great effort to improve the diagnosis, However, most of these techniques are expensive and time consuming. Majority of the techniques detects the lung cancer in its advanced stages, where the patients chance of survival is very low. This situation leads to a great need for a new technology to diagnose the lung cancer in its early stages. Image processing techniques acts as a good quality tool for improving the manual analysis and treatment of the lung cancer. In current clinical practice, hundreds of thin-sectional CT images are generated and are evaluated by radiologist by looking at each image in axial mode. It can makes difficulty for a radiologist to interpret it and acts as a very time consuming which cause missing a cancer. This results errors in detection or misinterpretation. Therefore, computational systems are needed to assist radiologists for the image interpretation, nodule detection and determination of their characteristics are needed.

Computers acts as a second reader that analyses digitized forms of the images, following rules similar to those that would normally guide a radiologist. Prediction of lung cancer at earliest should play an inevitable role in the diagnosis process also for an effective preventive strategy. The different existing techniques used to diagnose lung cancer are Chest Radiograph (x-ray), Computed Tomography (CT), Magnetic Resonance Imaging (MRI scan) and Sputum Cytology. But most of these techniques are detecting the lung cancer in its advanced stages, where the patient's chance of survival is very low. As a result, there is a great need for a new technology to diagnose the lung cancer in its early stages

A .Types of nodules

A lung nodule is an abnormal spot or shadow that appears on an imaging test, typically on an X-ray or on CT scan. It measures approximately 1.5 cm or less in diameter .Any lung spot that is larger than this, considered to be a lung mass and is more likely to be malignant in nature. About 150,000 lung nodules are spotted in X-rays or CT scans in people in the United States every year. There are two types of lung nodule

1) Benign

They are non cancerous, will not spread to other areas of the body. If the nodule is benign, then the size of the tumor is less than 3mm. This is starting level of cancer tumor. However presence of benign nodules, especially if they are large can cause health problems. Under this category is easily curable.

2) Malignant

They are cancerous, spread to different areas of the body. If the nodule is malignant, then the tumor size is greater than 3mm. This is an uncontrollable level of cancer tumor. Under this category is not curable

B. Structure of a CAD system

The computer-aided diagnosis (CAD) system is used for early detection of lung cancer by analysing chest 3D computed tomography (CT) images. There are four main phases involved in the proposed CAD system. They are image pre-processing, segmentation of lung region, feature extraction from the segmented region, classification of lung cancer as benign or malignant.

1) Pre-processing: Pre-processing is an important step that is used mainly to reduce the noise artifacts in the image, boundary identification. It improves the image quality by eliminating the regions which reduce the accuracy. This stage removes defects caused by the image acquisition process, for example, noise and lack of contrast etc. The main techniques for pre-processing are: Median Filtering, Enhancement Filter, Contrast Limited Adaptive Histogram Equalization, Auto enhancement, Wiener filter, Fast Fourier Transform, Wavelet Transform, Antigeometric Diffusion, Erosion Filter, Smoothing filters and Noise Correction [13].

2) Segmentation: This stage has the function to separate the study region from other organs and tissues in radiographic images in order to reduce the computational cost of the next stages. Segmentation of an image entails the division or separation of the image into regions of similar attribute .Several techniques like boundary extraction, region growing, gray-level histogram thresholding, deformable models (snakes - active contour maps), level set methods, can be used. The success of a particular technique can be measured in terms of accuracy, processing time, and automation level. Segmentation methods can be divided into eight categories (a) thresholding approaches, (b) region growing approaches, (c) classifiers, (d) clustering approaches, (e) Markov random field (MRF) models, (f) artificial neural networks, (g) deformable models, and (h) atlas-guided approaches.

3) Feature Extraction: Feature extraction is the process of defining a set of features, or image characteristics, which will most efficiently or meaningfully represent the information that is important for analysis and classification. This stage is an important stage that uses algorithms and techniques to detect and isolate various desired portions or shapes of a given mage. When the input data to an algorithm is too large to be processed and it is suspected to be redundant, then there occurs the necessity for the input data to be transformed into a reduced representation set of features. The basic characters of feature are area, perimeter and eccentricity [10].

4)Classification: The final procedure of the CAD system is to confirm the suspicious region and determine if it is a true nodule utilizing features obtained from previous stages. Different classifiers used for classification are rule based methods, minimum distance classifier, cascade classifier, Bayesian classifier, Multilayer perception, Radial Basis Function network (RBF), Support Vector Machine (SVM), Artificial Neural Networks, Fuzzy logic etc[15].

II. SURVEY OF THE CAD SYSTEM

In 1999, Samuel G. Armato et al. [2] were ones of the leading researchers on the lung nodule detection. They applied gray level thresholding on each slice of a CT scan set to segment the thorax from the background, and then the lungs from the thorax. They then applied rolling ball algorithm to the lung segmentation contours to avoid the loss of juxta- pleural nodules. Afterward, they applied multiple gray level thresholds to the volumetric lung regions to identify nodule candidates. For each nodule candidate geometric and gray level features are computed. They computed nine features for each nodule candidate; six geometric (volume sphericity, radius of the equivalent sphere, maximum compactness, max- imum circularity, and max eccentricity) and three gray-level features. Then, the values of these features were analysed by a Linear Discriminant Analysis (LDA) classifier, which has substantially reduced the number of false positives. Wail A. H. Mousa et al [11] used support vector machines (SVM) for classification of lung nodules. They used features which were extracted from CT images to train the SVM classifier. SVM classifier maps the input into high dimensional feature space through nonlinear mapping. The optimal hyper plane was constructed to differentiate the nodules. By using the kernel mappings such as radial basis function a nonlinear hyper plane was constructed. In 2001, Lee et al. [7] have developed a technique using genetic algorithm and template matching for detecting pulmonary nodules. False positives were eliminated through rules based on the characteristics of the nodules found. The system had a sensitivity of 72% with 25.3% false positives per case. In the validation of the system 98 nodules that possessed dimensions smaller than 10mm were used. A.Amutha, Wahidabanu [8], presented level set Active contour model for the detection of lung tumor. This method was based on kernel function having the minimum mean square error value. Then second order features were calculated which were based on the histogram of the noise free image [5]. The classification between the normal and abnormal lung image was made on these features. The drawback of this system, it was only able to work on 2-D images.

III. PROPOSED METHOD

A. Pre-processing stage

Pre-processing is the basic step in the diagnosis of lung nodules. The lung CT images usually contain impulsive noise. Impulsive noise consists of random noise and salt and pepper noise. Salt and pepper noise can be effectively removed by median filter. The goal of median filtering is to filter out noise that has corrupted image. It is based on a statistical approach. Typical filters are designed for a desired frequency response. Median filtering is a nonlinear operation often used in image processing to reduce salt and pepper noise. Median is calculated by sorting all the pixel values from the neighbourhood and replaces the pixel being considered with the middle pixel value. Here used the most advanced form of median filter ie. Untrimmed decision based median filter. Untrimmed decision based median filter removes the pixel values 0s and 255s in the image (pixel values responsible for salt and pepper noise), calculate the median value of remaining pixels thereby replace the noisy pixels with the median value. Fig 1 shows the pre-processing stage. It gives much better results as compared to previous methods

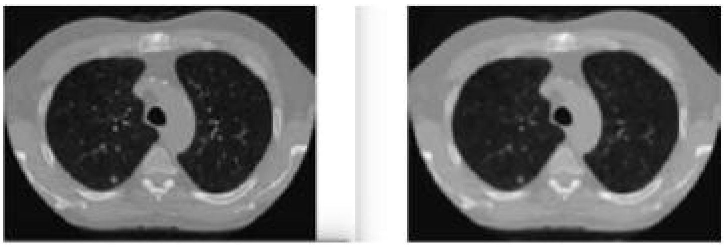

Fig. 1. a)Original image b)Applying Untrimmed decision based median filter

B. Segmentation

This stage has the function to separate the study region from other organs and tissues in CT images in order to reduce the computational cost of the next stages. Segmentation of an image entails the division or separation of the image into regions of similar attribute. The basic aim of doing the segmentation procedure is to extract more information from the given image so that maximum information can be extracted. Depending on the characteristics of the image a particular segmentation technique is used. Thresholding is the best segmentation method that is used in the medical field, especially in the CT images. In order to overcome the disadvantages of ordinary otsu's thresholding hybrid otsu is used .Thereby ,individual lung nodules are extracted by geodesic opening and closing. The important steps followed in this phase.

1) Obtain the histogram values (h) of the image I
2) Set the initial threshold value T in = Σ(h □ total shades)/Σh
3) Segment using T in .This will produce two groups of pixels: C1 and C2.
4) Repeat step - 3 to obtain the new threshold values for each class T c1 and T c2
5) Compute new threshold T=(T C1 + T C2 /2)
6) Repeat the steps 3 - 6 until the difference in T in successive iterations is not tends to zero
7) Now apply the Conventional Otsu method for the obtained threshold value for further segmentation process.
8) Thereafter to extract the individual nodules, apply geodesic opening and closing
9) Geodesic opening=Geodesic erosion followed by geodesic dilation
10) Geodesic closing=Geodesic dilation followed by geodesic erosion
11) Geodesic dilation->dilate marker image F with B, intersect the result with mask image G

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 4, March 2016*

National Conference on Signal Processing, Instrumentation and Communication Engineering (SPICE' 16)

Organized by

Dept. of ECE, Mar Baselios Institute of Technology & Science (MBITS), Kothamangalam, Kerala-686693, India

12) Geodesic erosion->erode marker image F with B, union the result with mask image G

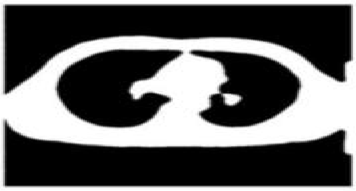

Fig. 2. a)Output of otsu's thresholding

Fig. 3. a)Output of segmented lung nodule

C. Feature extraction

Feature extraction is basically it separates the visual information from the image and stores them in the form of feature vectors in a feature database. These feature value (or a set of values) called feature vectors of image finds the image information from the feature extraction. Here, texture features are considered. For extracting features, Spatial Grey Level Dependency Method(SGLDM) is used. In SGLDM, the features are extracted by creating Grey level co occurrence matrix(GLCM),thereby 13 features like energy, entropy, variance,..are extracted.

D. Classification

Support Vector Machine is a powerful classification method based on the statistical learning theory. In general, classification algorithms aim at finding patterns in empirical data (training data or input data) with regard to label classes. The resulting classification model is used to make a prediction for new unlabeled data. In a sense, supervised learning concludes in finding a function f which fits the training data in the best way possible. It's a non-linear classifier which is a newer trend in machine learning algorithm and is popularly used in many pattern recognition problems, including texture classification. Here the kernel used is RBF kernel.

IV. EXPERIMENTAL RESULTS

The proposed method is tested using different lung CT images obtained from the image dataset ELCAP. The figure 5 shows the experimental result obtained for a lung CT image. It shows the results obtained at the different stages of the proposed CAD method. Thereby the nodule can be correctly classified as benign or malignant. The hybrid segmentation method used here gives better segmentation results which in turn gives better feature extraction and classification results through SGLDM and SVM classifier. Fig 4.shows the mean of 13 features obtained by SGLDM feature extraction method.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 4, March 2016*

National Conference on Signal Processing, Instrumentation and Communication Engineering (SPICE' 16)

Organized by

Dept. of ECE, Mar Baselios Institute of Technology & Science (MBITS), Kothamangalam, Kerala-686693, India

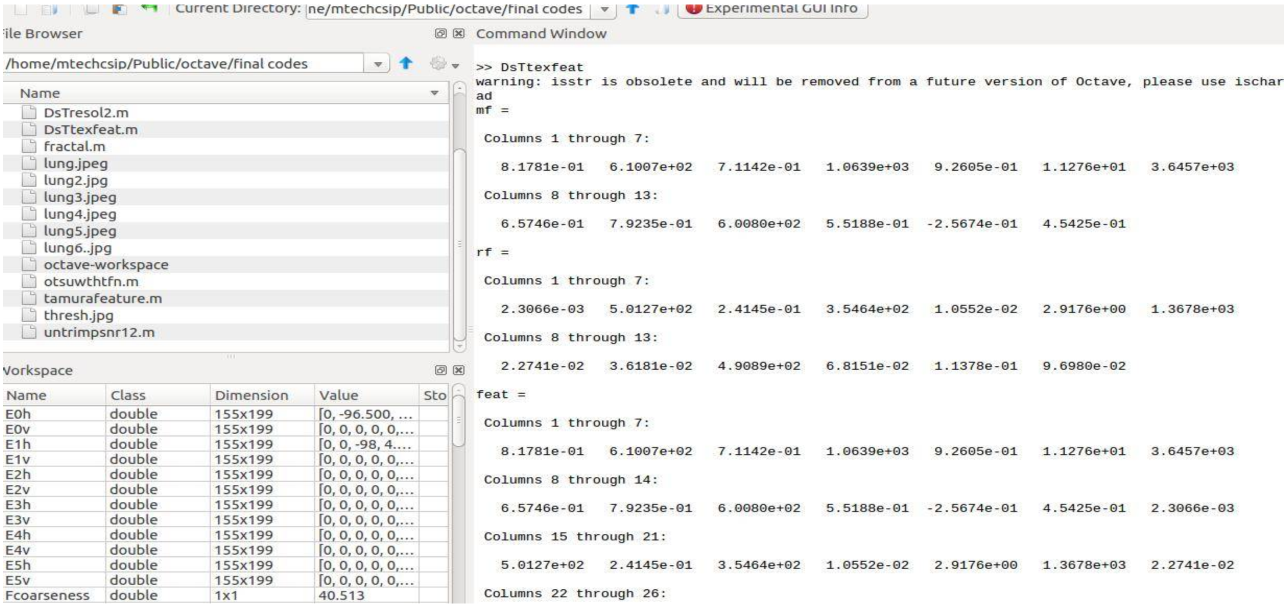

Fig.4:.GLCM features

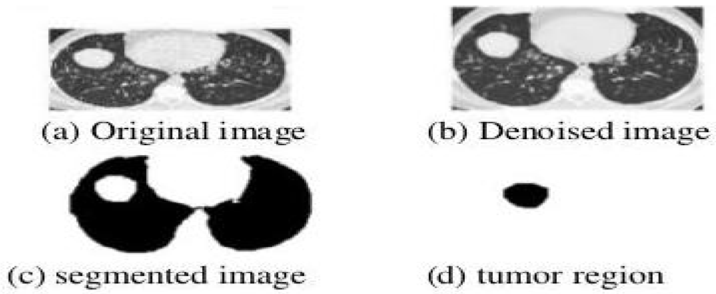

Fig 5: Results of the different stages obtained in the CAD system

V. CONCLUSION

In this paper, a new CAD system that accurately identifies and classifies the lung nodules in its early stage is proposed. In this method median filter is used in pre-processing to enhance the image, otsus thresholding for lung nodule segmentation. In this method SGLDM feature extraction are used. Feature vectors that are fed to the SVM classifier accurately classify the nodule. The proposed method gives much better result as compared to the previous methods.

REFERENCES

[1] Jhilam Mukherjee,Amlan Chakrabarti and Soharab Hossain Skaikh,Automatic Detection and Classification of Solitary Pulmonary Nodules from Lung CT Images ,978-1-4799-4272-5/14 2014 IEEE DOI 10.1109/EAIT.2014.64

[2] Samuel G. Armato et al,Computer Aided Diagnosis System for the identification and classification of lessions in lungs,International Journalof Computer Trends and Technology- May to June Issue 2011 P. K. Visscher, "How Self-Organization Evolves," Nature, vol. 421, pp. 799–800

[3] S.Shaik Parveen,Dr.C.Kavitha,A Review on Computer Aided Detection and diagnosis of lung cancer nodules ,International Journal of Computers and Technology Volume 3 No. 3, Nov-Dec, 2012 Feb.2003

[4] Xiabi Liu, Ling Ma, Li Song, Yanfeng Zhao, Xinming Zhao, and Chunwu Zhou Recognizing Common CT Imaging Signs of lung Diseases through a new feature selection method based on Fisher criterion ang Genetic Optimization' IEEE Journal of Biomedical and

Health Informatics,vol.19 NO.2 March 2015

[5] Ankit Agarwal,Sanchit Misra,Ramanathan Narayanan,Lalith Poleddiu,Alok Choudhary, A Lung Cancer Outcome Calculator Using Ensemble Data Mining on SEER data,ACM International conference on BIOKDD,pp.1-9,2011

[6] Anam Tariq ,M.Usman Akram and M.Younus Javed,Lung Nodule Detection in CT Images using Neuro Fuzzzy Classifier,Fourth International Workshop on Computational Intelligence in Medical Imaging(CIMI),pp:49-53,2013

[7] Younbun Lee,Takshi Hara,Hiroshi Fujita,Shigeki Itoh, and Takeo Ishi- gaki,Automated Detection of Pulmonary Nodules in Helical CT Images Baes on an Improoved Template-Matching Technique,IEEE Transaction on Medical Imaging ,Vol.20 page 595-604,july 2001.

[8] A.Amutha,Dr .R.S.D Wahidabanu,Lung Tumor Detection and Diagnosis in CT scan Images,International conference on Communication and signal Processing,April3-5,IEEE,2013‘

[9] S. Esakkirajan S. Jayaraman and T.Veerakumar. Digital Image Processing.McGraw Hill Education, 2009

[10] Noriyasu Homma, Kazuhisa Saito, Tadashi Ishibashi, Madan M. Gupta, Zeng-GuangHou, and Ashu M. G. Solo ,Shape Features Extraction from Pulmonary Nodules in X-ray CT Images,Neural Networks IJCN. (IEEE World Congress on Computational Intelligence).IEEE International Joint Conference , pp 3396-3400,June 1-8,Hong Kong,2008.

[11] Wail A.H Mousa and Mohammad A. U Khan, Lung Nodule Classification Utilizing Support Vector Machines, Image Process ing.Proceedings.InternationalConference, vol.3, pp.153 -156, 2002.

[12] Suzuki K, III SGA, Li F, Sone S, Doi K: Massive training artificial neural network (mtann) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography. Med Phys 2003, 30(7):16021617.

[13] Macedo Firmino , Antnio H Morais, Roberto M Mendoa, Marcel R Dantas, Helio R Hekis and Ricardo Valentim ”Computer-aided detection system for lung cancer in computed tomography scans: Review and future prospects”,Firmino et al. BioMedical Engineering OnLine 2014

[14] Mr. Vijay A.Gajdhane,Prof.Deshpande L.M,”Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Tech- niques”,IOSR Journal of Computer Engineering Sep-Oct.2014

[15] S.Sivakumar , Dr.C.Chandrasekar,”Lung Nodule Detection Using Fuzzy Clustering and Support Vector Machines”,International Journal of Engineering and Technology (IJET) Feb-March 2013

[16] Disha, Gagandeep,Identifying Lung Cancer Using Image Processing Techniques, International conference on computational techniques and artificial intelligence ICCTAI 2011.