LUNG CANCER DETECTION USING IMAGE PROCESSING

Bharti Sayare, Prof. K.V. Warkar
Department of Computer Engineering BDCE Wardha

Abstract - Lung cancer is the most important cause of cancer death for both men and women. The early detection of cancer can be helpful in curing the disease completely. So the requirement of techniques to detect the occurrence of cancer nodule in early stage is increasing. Computer Aided Diagnosis (CAD) is becoming one of the most popular and effective methods for diagnosing many diseases including cancer. This paper provides a Computer Aided Diagnosis System (CAD) for early detection of lung cancer nodules from the Chest Computer Tomography (CT) images.

There are five phases involved in the proposed CAD system. They are: extraction of lung region from chest computer tomography images, segmentation of lung region, feature extraction from the segmented region, formation of diagnosis rules form the extracted features and classification of occurrence and nonoccurrence of cancer in the lung. The approaches proposed in this study are: Segmentation of Lung Region using Modified Fuzzy Possibilistic C-Means (MFPCM) and Classification of Nodules employing Support Vector Machines (SVM). For segmentation, Modified Fuzzy Possibilistic C-Mean algorithm is used because of its better segmentation features. It modifies the membership function according to its weight and segmentation can be performed with better accuracy. For classification, Support Vector Machine (SVM) is used. SVM is used because of its simplicity and accuracy. The usage of SVM helps in better classification of cancer nodules. The experimentation is performed with the help of real time chest computer tomography images.

Keywords: Cancer, Benign, Tumor

I. INTRODUCTION

The World Cancer Research Fund International emphatically states on its official website that lung cancer is the most common occurrence of all cancer cases around the world. A cursory web search brings to light that cancer of the lungs takes more lives every year than breast, colon, and prostate cancer combined. These sobering revelations were enough to make us consider investigating the current prognostic techniques in detail and possibly propose one of our own which might hopefully be good enough to supersede the existing ones. We seek to build a robust Computer Aided Diagnosis (CAD) system that allows detection of lung cancer through analysis of Chest Computed Tomography (CT) images of the lung region and classify the tumor as Benign or Malignant. Incorrect and/or late diagnoses are the two leading causes for cancer related deaths. Lung cancer in particular, is a major killer the world over. Poor prognosis means that less than 15% of registered cases survive past five years of diagnosis. The poor prognosis is a consequence of inefficient diagnostic methods for early detection and successful treatment. However, it has been observed that patients in whom lung cancer has been detected early suffer a lower risk of lung cancer-related mortality than those at an advanced stage of the disease, suggesting that early detection and treatment of lung cancer can go a long way in the successful treatment of cancer. The objective of undertaking this project is to facilitate doctors to provide the best possible treatment by providing useful insights through analysis and diagnosis of lung cancer treatments with the help of predictive models.

II. LITERATURE SURVEY

[1] Shijun Wang and Ronald M. Summers, " Machine Learning and Radiology " , Published in final edited form as:Med Image Anal. 2012 Jul; 16(5): 933–951.Published online 2012 Feb 23. doi: 10.1016/j.media.2012.02.005.

Description: In this paper, a brief introduction to machine learning and its applications in radiology is presented. It focuses on applications like medical image segmentation and computer aided detection and diagnosis. Machine learning identifies complex patterns automatically and helps radiologists make intelligent decisions on radiology data such as conventional radiographs, CT, MRI, and PET images and radiology reports. In many applications, the performance of machine learning-based automatic detection and diagnosis systems has shown to be comparable to that of a well-trained and experienced radiologist.

Radiologic imaging is of increasing importance in patient care. The rapid expansion is a consequence of the need for more rapid, accurate, cost-effective, and less invasive treatment. Technologic advancements in radiologic imaging equipment have also fueled the utilization of imaging. Such technologic advancements include the ability to acquire higher and higher resolution images, enabling visualization of smaller anatomic structures and abnormalities. The higher resolution comes at the cost of an ever increasing average number of images per patient. Radiologists need to interpret these images and as the number of images increases, radiologists' workload increases as well. The increasing number and complexity of the images threatens to overwhelm radiologists' capacities to interpret them. In many real radiologic practices, automated and intelligent image analysis and understanding are becoming an essential part or procedure, such as image segmentation, registration, and computer-aided diagnosis and detection. In addition, in the area of cancer prognosis and treatment, automated and intelligent algorithms have a large market and are welcomed broadly, in areas such as radiation therapy planning or automatic identification of imaging biomarkers from radiological images of certain diseases, etc. Machine learning algorithms underpin the algorithms and software that make computer-aided diagnosis/prognosis/treatment possible.

In the daily practice of radiology, medical images from different modalities are read and interpreted by radiologists. Usually radiologists must analyze and evaluate these images comprehensively in a short time. But with the advances in modern medical technologies, the amount of imaging data is rapidly increasing. For example, CT examinations are being performed with thinner slices than in the past. The reading and interpretation time of radiologists will mount as the number of CT slices grows.

Machine learning provides an effective way to automate the analysis and diagnosis for medical images. It can potentially reduce the burden on radiologists in the practice of radiology. The applications of machine learning in radiology include medical image segmentation (e.g., brain, spine, lung, liver, kidney, colon); medical image registration (e.g., organ image registration from different modalities or time series); computer-aided detection and diagnosis systems for CT or MRI images (e.g., mammography, CT colonography, and CT lung nodule CAD); brain function or activity analysis and neurological disease diagnosis from fMR images; content based image retrieval systems for CT or MRI images; and text analysis of radiology reports using natural language processing (NLP) and natural language understanding (NLU).

In machine learning research, kernel learning and probabilistic models play key roles in radiology applications. kernel methods provide powerful tools for data analysis and have been found to be successful in a number of real applications. Here kernel means a matrix which encodes similarities between samples (evaluated by a certain kernel function which is a weighting function in the integral equation used to calculate similarities between samples). SVMs try to minimize the empirical classification error and maximize the geometric margin simultaneously on the training set which leads to high generalization ability on the new samples.

Inference: Radiologic imaging is of increasing importance in patient care. The rapid expansion is a consequence of the need for more rapid, accurate, cost-effective, and less invasive treatment. Machine learning provides an effective. Machine learning provides an effective way to automate the analysis and diagnosis for medical images. It can potentially reduce the burden on radiologists in the practice of radiology. Technological advancements in machine learning and radiology will ultimately benefit from each other in the long run.

[2] M.Gomathi, “A Parameter Based Modified Fuzzy Possibilistic C-Means Clustering Algorithm for Lung Image Segmentation”, Global Journal of Computer Science and Technology Vol. 10 Issue 4 Ver. 1.0 June 2010.

Description: In this paper, FCM and different extension of FCM Algorithm is discussed. The unique FCM algorithm yields better results for segmenting noise free images, but it fails to segment images degraded by noise, outliers and other imaging artifact. This paper presents an image segmentation approach using Modified Fuzzy Possibilistic C-Means algorithm (MFPCM). This approach is a generalized adaptation of standard Fuzzy C-Means Clustering (FCM) algorithm and Fuzzy Possibilistic C-Means algorithm. The drawback of the conventional FCM technique is eliminated in modifying the standard technique.

The Modified FCM algorithm is formulated by modifying the distance measurement of the standard FCM algorithm to permit the labeling of a pixel to be influenced by other pixels and to restrain the noise effect during segmentation. Instead of having one term in the objective function, a second term is included, forcing the membership to be as high as possible without a maximum frontier restraint of one. Experiments are carried out on real images to examine the performance of the proposed modified Fuzzy Possibilistic FCM technique in segmenting the medical images. Standard FCM, Modified FCM, Possibilistic C-Means algorithm (PCM), Fuzzy Possibilistic C-Means algorithm (FPCM) are compared with Modified FPCM to explore the accuracy. The three most important parameters used to determine the accuracy of the Modified FPCM are similarity, false positive and the false negative ratio.

Inference: Modified FPCM achieves higher accuracy of 94.25% while Fuzzy Possibilistic C-Means Clustering (FPCM) achieves 92.50% accuracy , Possibilistic C-Means Clustering Algorithm(PCM) achieves 91.00% accuracy, Modified FCM achieves 89.50% accuracy and standard FCM shows least accuracy of 86.03% .

[3] Swati P. Tidke, Vrishali A. Chakkarwar, “Classification of Lung Tumor Using SVM”, International Journal Of Computational Engineering Research (ijceronline.com) Vol. 2 Issue. 5 September 2012

Description: In this paper, SVM classifier is used for classifying the tumor as Benign or Malignant. Once the CAD

(Computer Aided Diagnosis) system for finding the lung tumor using the lung CT images is in place.

Support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification the basic SVM takes a set of input data and for each given input, predicts which of two classes forms the input, making it a non-probabilistic binary linear classifier. From given set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. In the proposed method we are using linear classifier.

Best hyper plane is the one that represents the largest separation or margin between the two classes. So we choose the hyperplane so that the distance from it to the nearest data point on each side is maximized. If such a hyperplane exists, it is known as the maximum margin hyper plane and the linear classifier it defines is known as a maximum classifier. In case of SVM classifier out of 9 features at a time only two features are selected for classification, which produces result as either benign or malignant.

Inference: SVM provides the accuracy of 92.5%.The accuracy of system can be improved if training is performed by using a very large image database.

[4] S.Shaik Parveen, C.Kavitha, "Classification of Lung Cancer Nodules using SVM Kernels", International Journal of Computer Applications (0975 – 8887) Volume 95– No.25, June 2014

Description: In this paper, Threat pixel identification with region growing method is used for segmentation of focal areas in the lung. For feature extraction gray level co- occurrence Matrix (GLCM) is been used. Extracted features are classified using different kernels of Support Vector Machine (SVM). The experimentation is performed with the help of real time computer tomography images.

CAD acts as a second reader and assists radiologist for accurate and efficient detection of cancer cells in the earlier stages. Thus the radiologist use CAD scheme to improve the detection accuracy.

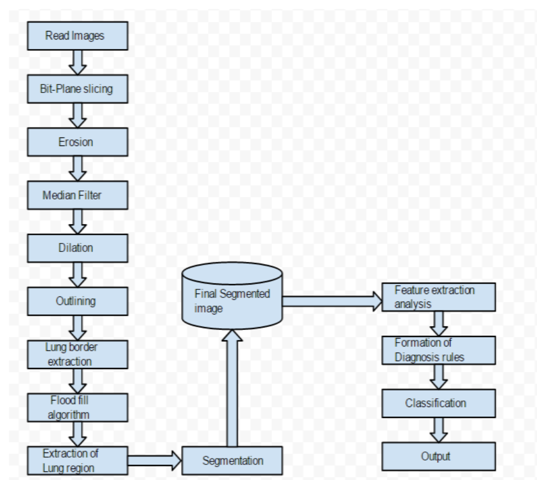
The proposed system consists of five steps for the classification of lung cancer nodules. They are as follows: Data collection,Extraction of lung region from Computer Tomography (CT) images with different preprocessing techniques, Segmentation of lung region using threat pixel identification with region growing method, Feature extraction using Gray level co- occurrence Matrix (GLCM) & Classification to identify normal, benign and malignant cancer of the lung using different Support Vector Machine (SVM) kernels.

After performing the segmentation, features need to be extracted for detecting the cancer in the lung region correctly. This step is concerned with two feature extractions: Gray level co- occurrence Matrix (GLCM), and Texture features. Texture feature are used in identifying normal and abnormal patterns. Texture is an alteration and variation of surface of the image, characterized as the space distribution of gray levels in neighborhood. A gray level co-occurrence matrix (GLCM) contains information about the positions of pixels having similar gray level values. Texture descriptors derived from GLCM are contrast, Energy, Homogeneity and Correlation.

For binary classification SVM is used to determine an Optimal Separating Hyper plane (OSH) which produces a maximum margin between two categories of data. A regularization parameter is employed to find the tradeoff between maximum margin and minimum classification error.

Inference: Linear kernel function, Polynomial kernel function and Radial basis function are the 3 variants of SVM presented in this paper and the conclusion drawn is that RBF kernel works better than the other kernels of SVM.

III. PROPOSED DESIGN



IV. IMPLEMENTATION

Steps:

1. Image Acquisition

First step is to acquire the CT scan image of lung cancer patient. The lung CT images are having lownoise when compared to X-ray and MRI images; hence they are considered for developing the technique. The main advantage of using computed tomography images is that, it gives better clarity and less distortion. For research work, the CT images are acquired from NIH/NCI Lung Image Database Consortium (LIDC) dataset. DICOM (Digital Imaging and Communications in Medicine) has become a standard for medical Imaging. The acquired images are in raw form. In the acquired images lot of noise is observed. To improve the contrast, clarity, separate the background noise, it is required to pre-process the images. Hence, various techniques like smoothing, enhancement are applied to get image in the required format.

2. Image Pre-processing

1) Smoothing

It suppresses the noise or other small fluctuations in the image; equivalent to the suppressions of high frequencies in the frequency domain. Smoothing also blurs all sharp edges that bear important information about the image. To remove the noise from the images, median filtering is used. Median filtering is a non-linear operation often used in image processing to reduce salt and pepper noise. In general, the median filter allows a great deal of high spatial frequency detail to pass while remaining very effective at removing noise on images where less than half of the pixels in a smoothing neighborhood have been affected. B=medfilt2(A,[m,n]) performs median filtering of the matrix A in two dimensions. Each output pixel contains the median value in the m x n neighborhood around the corresponding pixel in the image. Medfilt2 pads the image with 0's on the edges, so the median values for points within one -half the width of the neighborhood ([m,n]/2) of the edges might appear distorted.

2) Enhancement

Enhancement technique is used to improve the interpretability or perception of information in image for human viewers, or to provide better input for other automated image processing techniques. Image enhancement can be classified in two main categories, spatial domain and frequency domain. here Bit plane slicing is used for enhancement purpose.

3) Segmentation

The segmentation is performed for determining the cancer nodules in the lung. This phase will identify the Region of Interest (ROI) which helps in determining the cancer region. Modified Fuzzy Possibilistic C-Mean (MFPCM) is used in the proposed technique for segmentation because of better accuracy of MFPCM.

4) Feature extraction

Feature extraction is the important stage in this work. It uses different methods and algorithms to extract the features from the segmented image. Based on the extracted features normality and abnormality of the lung are decided. The features which we are extracted are area, perimeter and average intensity. Segmented images have only two values 1 and 0. Nodule part will be represented with value 1. Then area of the nodule can be calculated by finding number of pixel with value 1. Perimeter of the nodule means the number of pixels in the boundary region of the nodule. Average intensity is another feature which is used for the purpose of cancer detection. Select two threshold values for mean intensity values, and then calculate the average intensity value for the candidate region. If the average intensity value is between the threshold values then this part is assumed to be cancerous otherwise not. Based on area of the nodules the cancer caused nodules are identified. If the nodule size is greater than 25mm then it is assumed as abnormal image. If the nodule size is less than 25mm then it can be assumed as a normal image.

5) Classification

Support vector machines are supervised learning models with associated learning algorithms that
analyze data and recognize patterns, used for classification. The basic SVM takes a set of input data and for each given input, predicts which of two classes forms the input, making it a non-probabilistic binary linear classifier.

V. CONCLUSION

The difficulty in the early detection of lung cancer nodules is overcome in this paper. This paper provides a computer aided diagnosis system for early detection of lung cancer. The chest computer tomography image is used in this paper. In the first phase of the proposed technique, the lung region is extracted from the chest tomography image. The different basic image processing techniques are used for this purpose. In the second phase, extracted lung is segmented with the help of modified fuzzy possibilistic c-means algorithm. The next phase is extraction of features for diagnosis from the segmented image. Next, the diagnosis rules are generated from the extracted features. Finally with the obtained diagnosis rules, the classification is performed to detect the occurrence of cancer nodules. For the purpose of evaluation, the different real time chest computer tomography image is used. The experimental result reveals the advantage of the proposed CAD system for detecting lung cancer. The experiment shows that the usage of SVM will result in better accuracy of classification.

VI. REFERENCES

[1] Shijun Wang and Ronald M. Summers, " Machine Learning and Radiology " , Published in final edited form as:Med Image Anal. 2012 Jul; 16(5): 933–951.Published online 2012 Feb 23. doi: 10.1016/j.media.2012.02.005.

[2] M.Gomathi, “A Parameter Based Modified Fuzzy Possibilistic C-Means Clustering Algorithm for Lung Image Segmentation”, Global Journal of Computer Science and Technology Vol. 10 Issue 4 Ver. 1.0 June 2010.

[3] Swati P. Tidke, Vrishali A. Chakkarwar, “Classification of Lung Tumor Using SVM”, International Journal Of Computational Engineering Research (ijceronline.com) Vol. 2 Issue. 5 September 2012

[4] S.Shaik Parveen, C.Kavitha, “Classification of Lung Cancer Nodules using SVM Kernels”, International Journal of Computer Applications (0975 – 8887) Volume 95– No.25, June 2014

[5] M.Gomath , Dr.P.Thangaraj, ”A Computer Aided Diagnosis System for Detection of Lung Cancer Nodules Using Extreme Learning Machine”, International Journal of Engineering Science and Technology,ISSN: 0975-5462,Vol. 2(10), 2010.

[6] Sruthi Ignatious, Robin Joseph, “Computer Aided Lung Cancer Detection System”, Global Conference on Communication Technologies (GCCT 2015)

[7] Erkan Emirzade., “A Computer Aided Diagnosis System for Lung Cancer Detection using SVM”, a thesis submitted to the Graduate School of Applied Sciences of Near East University, NICOSIA, 2016