A Review on an Innovative Approach for Lung Cancer Detection and Classification from Medical Images

[1]Swati Sendre, [2]Shailendra Kumar Singh
[1]MTech Scholar, [2]Asst. Professor
[1]Eletectronics & Telecommunication, [2]Eletectronics & Telecommunication
[1]GD-Rungta College of Engineering & Technology, Bhilai, India
[2]GD-Rungta College of Engineering & Technology, Bhilai, India

Abstract: In previous years, image processing had covered a large area over biomedical applications in diagnosis of several of diseases in medical images. Lung Cancer detection and classification is one of the most widely used applications by many researchers. In this review paper we treat a very important research subject that affects directly the Lungs of human body Lung cancer detection and classification techniques verify the presence the cancer in Lung and check the probability of cancer. This study of this paper is aims to highlight the significance of data analytics and machine learning in prognosis in health sciences, particularly in detecting life threatening and terminal diseases like cancer. Here, we consider lung cancer for our study. For this purpose, preexisting lung cancer patient's data are collected to get the desired results. This paper presents the better Computer Aided Diagnosing (CAD) system for automatic detection of lung cancer. The initial process is lung region detection by applying basic image processing techniques such as Bit-Plane Slicing, Erosion, Median Filter, Dilation, Outlining, Lung Border Extraction and Flood-Fill algorithms to the CT scan images. After the lung region is detected, the segmentation is carried out with the help of Mean Shift clustering algorithm. With these, the features are extracted and the diagnosis rules are generated. These rules are then used for learning with the help of Random Forest.

***Keywords* – Lung cancer detection, CT scan images, Mean Shift clustering algorithm, Computer Aided Diagnosing, Lung Border Extraction.**

I. INTRODUCTION

The term cancer is a technical term refers to uncontrolled cell growth in tissues leading to malfunctioning of body organs at extreme state of influence cause major suffering and even death. Cancer leads to excessive multiplication of abnormal cells without control and are able to affect other tissues. Cancer cells infect neighboring part of human body through connective tissue. Most of the cancers are named for the organ where cancer starts – rapid growth of abnormal cancer cells in the colon is called colon cancer; that of skin is called basal carcinoma.

Lung cancer is a disease of abnormal cells multiplying and growing into a tumor. Cancer cells can be carried away from the lungs in blood, or lymph fluid that surrounds lung tissue. It is the second most common disease in men and women. When symptoms are seen in person, lung cancer tests are done to determine the type of lung cancer and if it has spread. Lymph flows through lymphatic vessels, which drain into lymph nodes located in the lungs and in the center of the chest. Lung cancer often spreads toward the center of the chest because the natural flow of lymph out of the lungs is toward the center of the chest. Metastasis occurs when a cancer cell leaves the site where it began and moves into a lymph node or to another part of the body through the blood stream [1]. Cancer that starts in the lung is called primary lung cancer. There are several different types of lung cancer, and these are divided into two main groups: Small cell lung cancer and non-small cell lung cancer which has three subtypes: Carcinoma, adenocarcinoma and Squamous cell carcinomas.

Figure 1 shows a general description of lung cancer detection system that contains four basic stages. The first stage starts with taking a collection of CT images (normal and abnormal) from the available Database from IMBA Home (VIA-ELCAP Public Access). The second stage applies several techniques of image enhancement, to get best level of quality and clearness. The third stage applies image segmentation algorithms which play an effective rule in image processing stages, and the fourth stage obtains the general features from enhanced segmented image which gives indicators of normality or abnormality of images.

Objective

We presented a novel method to classify a given CT scan images as normal or abnormal. The aim of this research was to detect features for accurate images comparison as pixels percentage and mask-labelling.

II. LITERATURE REVIEW

Segmentation is a pre-processing stage where images are divided into some separate regions and each is a set of pixels. Numerous image segmentation approaches were projected in the literature. On the other hand, a single technique may not be well-organized for a precise image class.

C. T. Henschke, et al. [2000] proposed an automatic computer aided diagnosing system for identifying lung cancer by analyzing lung CT images. Wiener filter, erosion, slicing techniques was used by authors to extract lung image region from CT image. Bit image slicing was used to convert CT images to binary image region growing segmentation. Algorithm is used to segment extracted lung regions. After segmentation of lung region rule based model was used to classify the cancer nodules. It was observed that the proposed method achieved the overall accuracy of 80%.

R. P. Petersen et. al. [2006] has proposed a method to obtain accurate results using enhancement and segmentation techniques. The image processing procedures: pre-processing; segmentation and feature extraction are used. Fast Fourier transform techniques were used for image enhancement from original image. In the segmentation stage the Watershed and thresholding segmentation were used and comparison had been made.

Gajdhane, M. V. A et. Al. [2014] developed a computerized system to detect lung nodules with the help of CT scan images. Identification of lung cancer was done in two stages, segmentation, enhancement and feature extraction. Threshold segmentation technique was applied to remove background and extract nodule from an image. Abnormal region is extracted by using feature vector analysis. Authors have tried to detect the smallest nodules for early diagnosis of lung cancer.

Schilham, A.M.R [2006] Authors have extracted the nodules using preprocessing, feature extraction, neural classifier. Noise reduction was mainly concentrated in preprocessing by using median filters and average filter, Gabor filter was used for image enhancement features like perimeter, area, eccentricity were extracted to localize ROI (Region of Interest) Support Vector Machine (SVM) was used for classification purpose.

C. A. Van Iersel, et al. [2006] Authors used Hopfield Neural Network (HNN) and a Fuzzy C-Mean (FCM) clustering algorithm, for early detection of lung cancer. To improve the accuracy and to decease the expertise required preprocessing (threshold-ing) technique was used to separate between nucleus and cytoplasm , because most of procedures are depending on nucleus feature. HNN has shown a better classification result than FCM, HNN succeeded in extracting the nuclei and cytoplasm regions.

W. J. Kostis, et al [2004] Authors have tried to group the images based on the similarity of their appearance. Preprocessing (histogram) of the images, feature extraction process and neural network classifier is used to check the state of patient to decide condition of patient. Classifier was tested initially for its correct operation neural network algorithm was implemented using open source, compared with available classification algorithms.

III. PROBLEM IDENTIFICATION

Lung cancer is the second most common cancer in both men and women (not counting skin cancer), and is by far the leading cause of cancer death among both men and women. Each year, more people die of lung cancer than of colon, breast, and prostate cancers combined. Most lung cancers could be prevented, because they are related to smoking (or secondhand smoke), or less often to exposure to radon or other environmental factors. But some lung cancers occur in people without any known risk factors for the disease. It is not yet clear if these cancers can be prevented. Most lung cancers have already spread widely and are at an advanced stage when they are first found. These cancers are very hard to cure. But in recent years, doctors have found a test that can be used to screen for lung cancer in people at high risk of the disease. This test can help find some of these cancers early, which can lower the risk of dying from this disease.

IV. PROPOSED ALGORITHM

As in the figure 1 we have some of the steps to detect Lung Cancer

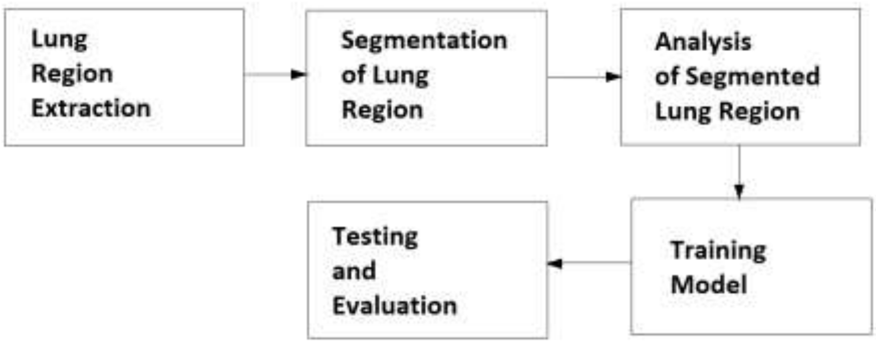

Figure 1- Lung Cancer Detection System

CT images acquisition

We load the CT image data which is store in our system. We collect the number of sample data from the radiologist and test this data in our system. We collect the high resolution structural and functional CT image data in normal and abnormal patient with different types of brain tumors.

Lung region Extraction

The goal of the lung extraction step is to separate the voxels corresponding to lung tissue from those belonging to the surrounding anatomical structures. Since the gray level of the background is close to the gray level of the lung tissues, it is advantageous to remove the background (all pixels outside the chest cavity) from each slice before doing extraction of the lung regions, in order to increase the accuracy of segmentation. Background is simply removed as follows: starting from the edges of the CT slices, all pixels that have similar gray levels, within a certain range, are removed. This process is applied, on each slice, from the left to right comers (i.e., horizontally or row-wise) as well as vertically (column-wise), in order to cover the whole slice. This simple approach has been shown to be quite adequate for removal of the background pixels outside the chest cavity region. After removing the background from each slice, we assume that each CT slice (image) contains only two types of pixel: 1) Lungs - pixels within very dense regions in the CT scan. These pixels have an average gray level gb.; and 2) other tissues organs – low density pixels within the lungs and the surrounding regions (e.g., ribs, heart, liver, and other parts in the chest cavity).

Detection of Abnormal Tissue in Lung Cancer

We assume that the segmented lung volume is formed of two types of voxels: normal, which describes the healthy lung tissue, and abnormal, which describes all abnormalities in the lungs, the bronchi and bronchioles. The first step to detect the abnormalities is to remove the normal tissues. The gray level histogram shows the distribution of the intensity of lung, which is usually a homogenous region. Lung abnormalities show in the histogram of the CT slices as bright (or homogeneous) areas with distinct gray levels from the surrounding lung tissues. Therefore, it is logical to assume that the abnormalities will correspond to the peaks (usually one) in the gray level histogram. Therefore, as a first step in isolating the abnormities, we threshold all pixels having a gray level below the gray level value of the peak of the histogram. These pixels will correspond to normal lung tissues.

The abnormalities, in general, appear as elliptical-shaped regions in the lungs. These elliptical shapes can be isolated by the analysis of the distribution of the gradient maxima in the neighborhood of each pixel. Ellipses and rings are patterns symmetrical relatively to their centroids. In the case of the polar co-ordinates, the expression of the intensity and spatial distribution of the most significant edges has characteristics properties of symmetry and uniformity. In the case of CT images, the gradient should show local maxima at the point on the border of the bronchi, bronchioles, and abnormal tissues.

Training Model using Random Forest algorithm

The random forest starts with a standard machine learning technique called 'Decision Tree' which, in ensemble terms, corresponds to the weak learners. In a decision tree, an input is entered at the top and as it traverses down the tree the data gets bucketed into smaller and smaller sets. The random forest takes this notion to the next level by combining trees with the notion of ensemble. Thus, the ensemble terms, the trees are weak learners and random forest is a strong learner.

Random forest runtimes are quite fast, and they are able to deal with unbalanced and missing data. Random Forest weaknesses are that when used for regression they cannot predict beyond the range in the training data, and that they may over-fit data sets that are particularly noisy.

Testing and Evaluation

(i) Cross-validation

We test our classifier using a technique called "cross-validation": train the classifier on all projects except for one. Here the projects mean the partition of the data set. Of course, we also know the ground truth for this held-out project, so we can see how well the classifier does on it, without cheating by training the classifier on the hold-out.

(ii) Mean Precision

The effectiveness of the classifier is the distance between the two means, which does not vary as threshold changes. One way of measuring the effectiveness of the classifier is the "precision". Precision is the number of truly correct items ("hits") divided by the number of items that the classifier says are correct (hits + false alarms). "Mean precision" takes into account the issues with choosing a threshold, noted above, by performing this calculation at a range of thresholds and taking the mean.

V. CONCLUSION

In this study we goes to develop an automated Lung cancer detection and find the percentage of cancer tissue. In this paper a partial survey of various methods of computer aided diagnosis were described here. This consists of CAD system in lungs, brain, liver, bone etc. A comparative study was made with various techniques and after simulation of these techniques that clarifies different methods which can locate the tumor perfectly and an exact result can produce. The future expansion is possible with some algorithm for different areas and that will provide more exact result than the existing. By complaining and selecting different methods and algorithms, it's possible to develop a new CAD system for different diseases. This leads to a good design of CAD.

VI. EXPECTED OUTCOME

This research was conducted to be detect Lung cancer using medical imaging techniques. The main technique will be used segmentation, which will be using a method based on threshold segmentation, watershed segmentation and morphological operators. We have introduced a novel approach for automatic lung cancer segmentation and visualization of the chest cavity from CT scans. The resolution of the CT scans is increased, the accuracy of the reconstructed 3-D volume will increase; hence, the diagnosis of abnormal tissues our current efforts are focused on improving the segmentation, and validating the 3-D models with respect to human experts. Long-term focus is to develop an expert system for chest cancer screening based on CT and X-ray scans.

VII. ACKNOWLEDGMENT

Expression of giving thanks is just a part of those feeling which are too large for words, but shall remain as memories of wonderful people with whom I have got the pleasure of working during the completion of this work. I am grateful to GD-Rungta College of engineering and technology, which helped me to complete my work by giving encouraging environment. I would like to express my deep and sincere gratitude to my supervisor, Assistant Professor, Mr. Shailendra Kumar Singh. His wide knowledge and his logical way of thinking have been of great value for me. His understanding, encouraging and personal guidance have provided a good basis for the present work.

REFERENCES

[1] Ch. Hima Bindu , and K. Satya Prasad, “An Efficient Medical Image Segmentation Using Conventional OTSU Method,” Int. Journal of Advanced Science and Technology, January 2012 Vol. 38.

[2] B. Wu and Y. Yang, [2012], "Local-and global-statistics-based active contour model for image segmentation," Mathematical Problems in Engineering, vol. 2012.

[3] M. Erdt, S. Steger, G. Sakas, [2012], “Regimentation: A New View of Image Segmentation and Registration,” Journal of Radiation Oncology Informatics ISSN: 1663-618X, J Radiat On col Inform.

[4] A. Nyma, M. Kang, Y. KeunKwon, C. HongKim, and J. MyonKim, [2012] “A Hybrid Technique for Medical Image Segmentation,” Hindawi Publishing Corporation Journal of Biomedicine and Biotechnology, Article ID 830252, 7 pages doi:10.1155/2012/830252.

[5] N. T. N. Anh, J. Cai, J. Zhang, and J. Zheng, [2012],"Constrained active contours for boundary refinement in interactive image segmentation," IEEE Int. Symposium on Circuits and Systems (ISCAS), pp. 870-873.

[6] D. Kaushik, S, Utkarsha , P. Singha, and V. Singh, [2013]“Medical Image Segmentation using Genetic Algorithm,” Int. Journal of Computer Applications (0975 – 8887) , November,Volume 81.

[7] A. Norouzi1, M. Rahim, A. Altameem, T. Saba, E. Abdolvahab, A. Rehman and M. Uddin,[2014], “Medical Image Segmentation Methods, Algorithms, and Applications,” IETE Technical review ,vol 31.

[8] BingquanHuo, Guoxin Li, and Fengling Yin ,[2015] “Medical and Natural Image Segmentation Algorithm using M-F based Optimization Model and Modified Fuzzy Clustering: A Novel Approach,” Int. Journal of Signal Processing, Image Processing and Pattern Recognition, Vol.8.

[9] Gawade Prathamesh Pratap, R.P. Chauhan [2016],"Detection of Lung cancer cells using image processing techniques", IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), ISBN: 978-1-4673-8587-9.

[10] Ping Liu, Jianqi Sun, Yijing Guan[2016], "Detection of Lung Cancer with Phase-Contrast X-ray Imaging Using Synchrotron Radiation", Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE, ISBN: 1-4244-0032-5.

[11] S Avinash, K. Manjunath, S. Senthilkumar[2017],"Analysis and comparison of image enhancement techniques for the prediction of lung cancer", 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), ISBN: 978-1-5090-3704-9.

[12] Pradeep Singh,Namrata Singh,[2017],"Intelligent approaches for prognosticating post-operative life expectancy in the lung cancer patients", IEEE International Conference on Inventive Computing and Informatics (ICICI), ISBN: 978-1-5386-4031-9.

[13]Anjali Kulkarni, Anagha Panditrao [2016],"Classification of lung cancer stages on CT scan images using image processing",IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), ISBN: 978-1-4799-3914-5.

[14]Lilik Anifah,Haryanto,Rina Harimurti ,[2017],"Cancer lungs detection on CT scan image using artificial neural network backpropagation based gray level coocurrence matrices feature", IEEE International Conference on Advanced Computer Science and Information Systems (ICACSIS), ISBN: 978-1-5386-3172-0.

[15]Snehal Dabade, Shubhangi Chaudhari, Sneha Jadhav [2017],"A review paper on computer aided system for lung cancer detection ", IEEE International Conference on Big Data, IoT and Data Science, ISBN: 978-1-5090-6593-6.