International Journal of Enhanced Research in Management & Computer Applications, ISSN: 2319-7471
Vol. 2 Issue 5, May-2013, pp: (63-68), Available online at: www.erpublications.com

Comprehensive Study and analysis of image processing algorithms for early detection of lung cancer

Suman Dahiya[1], Manoj Dahiya[2]

[1]Assistant Professor, ECE Dept., GJUST, Hisar, Haryana
[2]Lecturer, Dept. of ECE, Govt. Polytechnic, Nathusari Chopta, Sirsa, Haryana

ABSTRACT

Many research papers have been published on Lung Cancer in medical as well as in Computer Science related journals. There are many researches been carried out for early diagnosis of Lung cancer. This paper focuses on early stage lung cancer detection. Lung cancer is prominent cancer as it states large number of deaths of more than a million every year. It creates need of detecting the lung nodule at early stage in Computer Tomography medical images. So to detect the occurrence of cancer nodule at early stage, the requirement of methods and techniques is increasing. There are different methods and techniques existing but none of them provide a better accuracy of detection. One of the techniques is content based image retrieval Computer Aided Diagnosis System (CAD) for early detection of lung nodules from the Chest Computer Tomography (CT) images. This optimization algorithm allows physicians to identify the nodules present in the CT lung images in the early stage hence the lung cancer. The performance measures like the classification rate and the false positive rates are also studied.

Key words: Lung Cancer Survey Data Mining Medical Image Processing.

1. INTRODUCTION

The term Lung cancer, also known as lung carcinoma, means a lung tumor malignancy characterized by uncontrolled cell growth in tissues of the lung. This growth can spread beyond the lung by the process called metastasis into nearby tissues and even other parts of the body, if it is not treated in the early stage. Primary lung cancers are the cancers which start in the lung. The two main types are Small Cell Lung Cancer (SCLC) and Non-Small-Cell Lung Cancer (NSCLC). The most common symptoms are cough (including coughing up blood), weight loss, shortness of breath, fever and chest pains. Long term smoking is the primary cause for the cancer of its type which takes a share of 85- 90% and only 10 – 15 % for nonsmokers case. These cases are often caused by a combination of genetic factors as well. In the year 2012, because of lung cancer 90% of total victims 1.8 million victims) dead around the globe. This makes it as the most common cause of cancer-related death in men and second most common in women after breast cancer. Overall, 17.4% of people in the United States diagnosed with lung cancer survive five years after the diagnosis, while in the developing countries this survival rate is observed as average to poor. About 8% of lung cancer is observed because of family background that is genetic reason. Signs and symptoms which may suggest lung cancer include:

- Respiratory symptoms: coughing, coughing up blood, wheezing, or shortness of breath.
- Systemic symptoms: weight loss, weakness, fever, or clubbing of the fingernails.

- Symptoms due to the cancer mass pressing on adjacent structures: chest pain, bone pain, superior vena cava obstruction, or difficulty swallowing.

After the discovery of X-Rays, medical imaging had begun with radiography. Radiography is an imaging system that makes utilization of the X-Rays, which were utilized as a part of demonstrative methods previously the impacts that are exceptionally unsafe for individuals because of the radiation of ionization, was found. Infiltration of X-beams inside the body and the ingestion of their radiation are differential and rely upon the tissue's thickness.

The X-Ray picture is created on the fluorescent screen or the photographic film by the diverse densities of the tissues in the body. The two types of radiography are fluoroscopy and projectional radiography. Inspite of the progression of 3D tomography advancements, these 2D innovations are still being used because of less cost, high-determination and low radiation applications. Positron Emission Tomography (PET): Some of the atomic medication contemplates, permit imaging of the entire body in light of certain cell receptors or capacities. PET outputs or PET/CT filters and octreotide checks are a portion of the illustrations.

Despite the fact that there are some clinical applications where comparative demonstrative data is given by SPECT and PET, as a rule the application have a tendency to be distinctive in view of the outcome from attributes of the radio nuclides that are utilized. Fudiciary markers can relate the utilitarian data from SPECT or positron emanation tomography to anatomical data gave by attractive reverberation imaging (MRI) One of the primary investigative strides towards the recognition of the tumor nearness in the jump is performing chest radiography, when a man accompanies remarkable indications. CT imaging is another alternative to give more data about the sort and degree of infection and it is being utilized as a part of all most all multispecialty healing centers and finding focuses.

The CT check picture was first taken in the year 1960 however later on from 1970 or 1980 onwards it turned out to be exceptionally unique for precise clinical choices conceivable. In 1979 two researchers, Godfrey Hounsfield and Allan Cormack, shared the Nobel Prize for their work building up the CT scanner. CT is additionally called Computerized Axial Tomography or CAT. A CT scanner utilizes x-beams similarly as a regular x-beam however as opposed to taking one picture a CT scanner takes numerous pictures, or cuts. An introduced PC program gathers every one of the pictures and incorporates them to make a 3D picture of the inside structures being analyzed. Headways in this field have enhanced the precision and affectability and helpfulness of CT scanners since they were produced.

The analysis arrangement of discovery of lung tumor utilizes processed tomography with uncommon three dimensional CAD frameworks and a volumetric dataset with up to 3,000 single pictures. This kind of frameworks are called Computer Aided Diagnosis (CAD) frameworks. Round sores running from 1 mm are discernible. Round injuries shift from 5– 10 mm are not entirely obvious if the tumor sizes are high as far as centimeters. The standard utilization of CAD Chest Systems may distinguish little changes without introductory doubt. Philips was the primary merchant to show a CAD for early identification of round lung sores on x-beam pictures. A few highlights which incorporate size of the tumor zone of the tumor will recognize generous conditions from conceivable lung malignancy. On the off chance that the extent of the tumor is little, the hazard for dangerous disease is likewise little. Benevolent makes tend have an all around characterized fringe, while lobulated injuries or those with an unpredictable edge stretching out into the neighboring tissue have a tendency to be threatening.

REVIEW OF LITERATURE

There are two image databases which are freely available for the researchers on medical image processing and they are Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) both give the Lung tumor CT images with different number of slices. At present, for the early detection of malignancy or cancer of lung and treatment in the early stage, the digital image processing approaches are most widely used and applied in medical fields in which the time factor is very important to discover the disease in the patient as soon as possible, especially in tumors with variable sizes. For detecting the tumor or classification of the tumor, the image processing involves various steps which include

image pre-processing, enhancement, segmentation, feature extraction and finally the classification. Some of the methodologies found during the review of literature and they are briefed below.

Ada and Rajneet Kaur [1] proposed a method to evaluate the survival rate of a patient if has the symptoms mentioned above. For that they proposed a method which uses Histogram Equalization for preprocessing of the images and binarization and GLCM(Grey Level Co-occurrence Method) algorithms for feature extraction process and neural network classifier for the classification to check the state of a patient in its early stage whether it is normal or abnormal.

Avinash et al.[2] proposed a method for detection of cancerous tumor in lung CT image using Gabor filter for Image enhancement instead of DWT and Histogram equalization, for segmentation stage they used watershed segmentation algorithm and for feature extraction, masking approach was used instead of Binarization. With this approach they were able to reduce the detection time of cancer.

Sanghamitra et al.[3] came with new method of detection of tumor with four stages of processing. First CT images are collected with different features, preprocessing is done with median filter, wiener filter or Gabor filter. Next, for segmentation, EK-means clustering method was used and the features extraction was done using gray-level co-occurrence matrix (GLCM) method and Back Propagation Network (BPN) is applied for classification of disease stages. They classified the given CT image as normal lung or cancerous lung.

Raajan.P et al. [4] developed a method derived from the Wiener filter techniques. Appling the wiener filters in an image adaptively, tailoring itself to the local image variance. It smoothen the image at low variance. Similarly, it also smoothen the image more when the variance high. This filter provides better results compared to the median filter.

Premalatha et al. [5] proposed a system consists of pre-processing, segmentation, feature extraction and final classification. The primary stage is preprocessing which was implemented with convolution filter and at the second stage the image enhancement was done by thresholding. For the segmentation the watershed segmentation algorithm used to detect the vey small nodules with better accuracy. The proposed technique gives very promising results comparing with other used techniques. The CT scan images which are used for processing were collected from the reputed hospitals, Hyderabad. This image dataset contains lung CT scan images with tumor and without tumor.

METHODOLOGIES TO DETECT LUNG CANCER

Diagnosis of lung cancer includes the following stages:

1. Images captured
2. Preprocessing of images
3. Image segmentation
4. Feature extraction
5. Principle component analysis
6. Neural network classifier
7. Diagnosis result
8. Prediction process

1. Images captured or collected:

Primarily, cancer and non-cancer patient's data or CT-Scan images will be collected from different diagnostic centers. The digitized images are stored in the DIACOM format with a resolution of 8 bits per plane [1]

.

2. Preprocessing of images:

The image Pre-processing stage in this system begins with image enhancement which aims to improve the interpretability or sensitivity of information included in them to provide better input for other programmed image processing techniques.

Image enhancement techniques can be divided into two wide types: Spatial domain methods and frequency domain methods. On the other hand, when image enhancement techniques are used as preprocessing tools for other image processing techniques, the quantifiable measures can determine which techniques are most suitable. In the image enhancement stage we will be using the Histogram Equalization. The pre-processing of image aims for selective elimination of the redundancy in scanned images without affecting the original image, this play a vital role in the diagnosis of lung cancer. Therefore, Histogram- Equalization becomes the crucial step in preprocessing. Hence, each image is preprocessed to enhance its superiority.

3. Image Segmentation:

Image segmentation is a crucial process for most image analysis consequent tasks. Especially, most of the existing techniques for image description and recognition are highly depend on the segmentation results. Segmentation splits the image into its constituent regions or objects. Segmentation of medical images in 2D has many beneficial applications for the medical professional such as: visualization and volume estimation of objects of concern, detection of oddities, tissue quantification and organization and many more. The main objective of segmentation is to simplify and change the representation of the image into something that is more significant and easier to examine. Image segmentation is usually used to trace objects and borders such as lines, curves, etc. in images. More accurately, image segmentation is the process of allocating a label to every pixel in an image such that pixels with the same label share certain pictorial features.

The outcome of image segmentation is a set of segments that collectively cover the entire image, or a set of edges extracted from the image i.e. edge detection. In a given region all pixels are similar relating to some distinctive or computed property, such as texture, intensity or color. With respect to the same characteristics adjacent regions are significantly different. One of two basic properties of intensity values Segmentation algorithms are based on: discontinuity and similarity. In the first group we partition the image based on abrupt changes in intensity, such as edges in an image. The next group is based on segregating the image into regions that are alike according to a predefined criterion. Histogram thresholding methodology comes under this group.

4. Feature Extraction:

Image features Extraction stage is a crucial stage that uses algorithms and methods to detect and separate various preferred portions or shapes of an inputted image. The following two methods are used to predict the probability of lung cancer presence: Binarization and GLCM, both methods are based on facts that strongly related to lung anatomy and information of lung CT imaging.

5. PCA (Principle Component Analysis)

PCA is a technique to normalize the data in image. Real-world data sets generally display associations among their variables. These associations are frequently linear, or at least practically so, making them agreeable to common examination techniques. One such technique is principal component analysis ("PCA"), which rotates the original data to new coordinates, making the data as "even" as possible. The features mined are delivered to the PCA data mining for better sorting [1]

.

The following steps takes place in PCA:-

i. Calculating the mean and standard deviation of the features in the image.
ii. Subtracting the sample mean from each observation, and then dividing by the sample standard deviation. This scales and centers the data.
iii. Then we calculate the coefficients of the principal components and their relevant changes are done by finding the Eigen function of the sample covariance matrix.

iv. This matrix holds the coefficients for the principal constituents. The diagonal elements store the modification of the relevant principal constituents. We can mine the diagonal.
v. The maximum variance in data results in maximum information content which is required for better classification [1]
.

6. Neural Network Classifier

Supervised feed-forward back-propagation neural network ensemble used as a classifier tool. Neural network contrasts in different means from traditional classifiers like Bayesian and k – nearest neighbor classifiers. Linearity of data is one of the major variances. Other existing classifiers like Bayesian and k – nearest neighbor entails linear data to work properly. But neural network works as well for nonlinear data because it is simulated on the reflection of biological neurons and network of neurons. Training the neural network with wide range of input data will increase the detection accuracy, in other words the system will get biased with a small set of data or large set of similar data. Hence neural network classifier needs a large set of data for training and also it is time consuming to train to reach the stable state. But once it is trained it works as fast and quick as biological neural network by transmitting signals as fast as electrical signals.

7. Diagnosis Result

After completion of all the processes in the last stage i.e. in the diagnosis stage or in diagnosis result the proposed system show whether the image is in normal or in abnormal state.

8. Prediction process

There is no remedy for cancer after completely affected. Death is inevitable. So the ability to predict Lung cancer plays an important role in the diagnosis process. In this paper we have proposed an effective Lung cancer prediction system based on data mining. This lung cancer risk prediction system should prove helpful in detection of a person's predisposition for lung cancer. We will be considering various risk factors which includes-age, gender, hereditary, previous health examination, use of ant hypersensitive drugs, smoking, food habit, physical activity, obesity, tobacco, genetic Risk, environment, mental trauma, uptake of red meat, balance diet, hypertension, heart disease, excessive alcohol, radiation therapy and chronic lung diseases. Various algorithms will be used such as decision tree algorithm for the prediction process. Figure below illustrates the working of decision tree algorithm. After its processing, system will predict the chances of patient to get affected by lung cancer in the years to come.

CONCLUSION

Thus, we can conclude that using the combination of neural network classifier along with Binarization techniques will increase the accuracy of lung cancer detection process. This system will also decrease the cost and time required for cancer detection. Also if the patient is not detected with the lung cancer the system will proceed further for the prediction process. As this system will be available online, patients from remote areas can also avail its benefits. So this system is beneficial for huge number of people all over the world. Also tests required for cancer detection is required.

REFERENCES

[1]. T. Sowmiya, M. Gopi, M. New Begin L. Thomas Robinson "Optimization of Lung Cancer using Modern data mining techniques." International Journal of Engineering Research ISSN:2319- 6890)(online),2347-5013(print)VolumeNo.3,Issue No.5, pp : 309-3149(2014)
[2]. Ada[1], Rajneet Kaur[2] "Early Detection and Prediction of Lung Cancer Survival using Neural Network Classifier", (IJAIEM)Volume 2, Issue 6, June 2013
[3]. Dasu Vaman Ravi Prasad,"Lung cancer detection using image processing techniques", International journal of latest trends in engineering and technology.(2013)
[4]. Ada, Rajneet Kaur "Early Detection and Prediction of Lung Cancer Survival using Neural Network Classifier"IJAIAM. Volume 2, Issue 6, June 2013.

[5]. Avinash.S, Dr.K.Manjunth, Dr.S.Senthilkumar,"An Improved Image Processing Analysis for the Detection of Lung Cancer using Gabor Filters and Watershed Segmentation Technique",IEEE,2016.
[6]. P.B. Sangamithraa, S. Govindaraju.," Lung Tumour Detection and Classification using EK-Mean Clustering", IEEE –WiSPNET conference, 2016.
[7]. Md. Badrul Alam Miah, Mohammad Abu Yousuf," Detection of Lung Cancer from CT Image Using Image Processing and Neural Network", Electrical Engineering and Information & Communication Technology (ICEEICT) 2015.
[8]. Taruna Aggarwal, Asna Furqan, Kunal Kalra," Feature Extraction and LDA based Classification of Lung Nodules in Chest CT scan Images",IEEE,2015.
[9]. Dasari Hemalatha, Raja Rao.Ch, S.J.Sugumar," Detection of Lung Cancer Using Marker Controlled Watershed Transform", International Journal & Magazine Engineering, technology, management Research ,Volume3,Isuue no.10,2016.
[10]. Raajan.P, Muthuselvi.S, Agnes Saleema. A," An Adaptive Image Enhancement using Wiener Filtering with Compression and Segmentation", International Journal of Computer Applications,2015.
[11]. L. Berlin, "Liability of performing CT screening for coronary artery disease and lung cancer," Amer. J. Roentgenol., vol. 179, pp. 837–42, 2002.
[12]. S. J. Swensen et al., "Screening for lung cancer with low-dose spiral computed tomography," Amer. J. Respir. Crit. Care Med., vol. 165, pp. 508–13, 2002.
[13]. J. C. Nesbitt et al., "Survival in early-stage lung cancer," Ann. Thorac. Surg., vol. 60, pp. 466–472, 1995. [32] C. I. Henschke et al., "CT screening for lung cancer: Suspiciousness of nodules according to size on baseline scans," Radiology, vol. 231, pp. 164–8, 2004.
[14]. L. L. Humphrey, S. Teutsch, and M. Johnson, "Lung cancer screening with sputum cytologic examination, chest radiography, and computed tomography: An update for the U.S. Preventive Services.