International Journal of Scientific Research & Engineering Trends
Volume 5, Issue 2, Mar-Apr-2019, ISSN (Online): 2395-566X

Lung Cancer Detection Using Digital Image Processing and Artificial Neural Networks

M. Tech Scholar Shubham Bhardwaj , Asst. Prof. Shalini Bhadola, Asst. Prof. Kirti Bhatia
Department of CSE
Sat Kabir Institute of Technology & Management
Bahadurgarh, Haryana, India

Abstract- **Cancer detection is generally carried out manually by trained professionals and these techniques are majorly helpful in the advanced stage detection, also it involves a very tedious procedure and highly dependent on the given individual. This introduces the high possibility of human error in the detection process which necessitate an automated process. Hence, this paper aims at early detection of cancer through an automated process to minimize human error and making the process more accurate and hassle-free. In the proposed work, image processing algorithms and artificial neural network have been employed to design an automated process for early stage detection of lung cancer. In today's world ,image processing methodology is very rampantly used in several medical fields for image improvement which helps in early detection and analysis of the treatment stages ,time factor also plays a very pivotal role in discovering the abnormality in the target images like-lung cancer ,breast cancer etc. this research focuses upon image quality and accuracy. image quality assessment as well as improvement are dependent upon enhancement stage where low pre-processing techniques are used based upon gabor filter within Gaussian rules; thereafter the segmentation principles are applied over the enhanced region of the image and the input for feature extraction is obtained, further depending upon the general features, a normality comparison is made .in the following research the crucial detected features for accurate image comparison are pixel percentage and masking labeling. In this research we have done classification based upon artificial neural networks which is more satisfactory than other current classification methods.**

Keywords- **Lung Cancer Detection, Enhancement, Feature Extraction, segmentation, neural network, image processing, artificial network, histogram.**

I. INTRODUCTION

1. Overview

Lung cancer is a disease, also referred as lung carcinoma which is a malignant lung tumor and is characterized by the highly uncontrolled cell growth arising in the lung tissues. Tumors composed of cancer cells are called malignant tumors and the tumor composed of mainly non-cancerous cells are referred as benign tumors. If left unchecked, this cancer can grow and spread to other parts of the body. There are broadly 2 types of lung cancer namely, small-cell lung carcinoma (SCLC) [1] and non-small-cell lung carcinoma (NSCLC) [1]. NSCLC contributes to more than 80 % of all lung cancer cases, while SCLC accounts for about 15 % of the cases [1].

The grade of tumor depends on the appearance of tumor microscopically and cell's growth rate. In grade 1: The cells look nearly like the normal lung cells. Theses cancerous cells are generally slow at spreading and multiplying. In grade 2, the cancerous cells become relatively more abnormal and differentiable as compared to grade 1, and spread faster as well. More advanced stages include grades 3 and 4, which tend to present themselves as extremely abnormal and fast spreading cancerous cells [2]. The mortality rate of cancer has never been higher. The lung or respiratory system cancer has been ranked 1 by the world cancer research fund international (WCRF) in a list of the most diagnosed cancers worldwide [3]. About 1,825,000 new cases of lung cancer were diagnosed in 2012, being 13 per cent of all cancers worldwide, indicating its immense threat worldwide [3].

2. Corresponding Author

1,242,000 men were diagnosed worldwide in 2012, which is about 16.7 per cent of all the cancers in the male population being the most dangerous cancer to men, while 583,000 diagnosed cases in 2012 were women, which is 8.8 per cent of all the cancers in the female population making it the third most dangerous cancer to women according to WCRF, hence lung cancer poses a bigger threat to men than women as indicated by the statistics [3]. Tackling this feral disease has always been a tedious procedure but with new technological advancements an improvement in the qualities of remission and detection

© 2019 IJSRET

has been accomplished, but it's still a long and expensive procedure nonetheless. It is also important to note that that in developing countries, a large proportion of patients afflicted with cancer are poor. Imaging processes for cancer detection are quite expensive [4], which in turn makes it difficult for such patients to pay hefty consultation charges to get diagnosed. Thus, our design also aims to make the cancer detection process affordable for such a demographic scenario where one does not have a proper access to expensive health care.

We have used the Lung CT-Diagnosis database from The Cancer Imaging Archive [5],[6],[7], a publicly available archive which consists of a large number of medical images for various cancers. Our database consisted of diagnostic contrast enhanced CT scans stored in Dicom format, taken from studies of 61 patients. The database has a total of 4682 such images.

We have proposed a design that reads JPEG converted Dicom Format images of lungs (AS DESCRIBED IN THE DATABASE) and scans these images for any abnormality through image processing techniques. Once the system has completed the scanning process, it calculates certain features of the abnormality and feeds them into a system which is trained to detect if the abnormality is cancerous. The training system is a Back Propagation Artificial Neural Network.

The image processing steps include conversion into grayscale, Histogram Equalization, Thresholding and Feature extraction. The Back Propagation Network (BPN) is trained using 70 images, and contains 6 input neurons and 2 output neurons, along with 12 hidden layer neurons. The output indicates whether the tumor is malignant or benign. Our design was found to be 78% accurate.

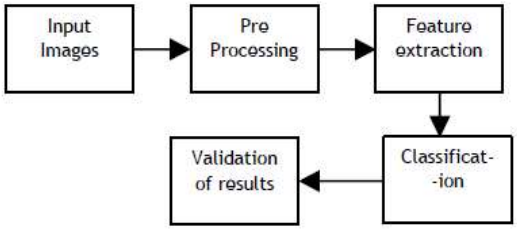

Fig.1 A Flowchart Showing the proposed method of Cancer Detection incorporated in our design.

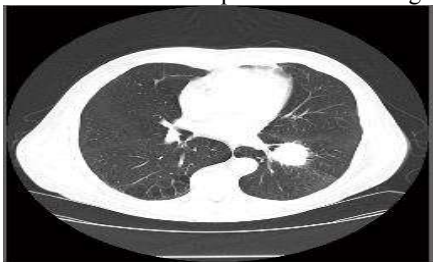

Fig. 2 A grayscale converted image.

II. LITERATURE REVIEW

Avinash.S, Dr. K Manjunath, Dr.J.Senthilkumar explained lung cancer detection method using Gabor filter and watershed segmentation techniques. To overcome the drawbacks of FFT method proposed method is explained. This new technique with Gabor filter and watershed segmentation can be used for quick detection of lung cancer [2]

Rachid Sammouda proposed a model to segment extracted lung regions from chest computer Tomography images. In this paper three diagnostic rules are verified as well-defined filters of candidate cancerous regions from the status of candidate [4]

Summrina Kanwal wajid, Kaizhu Huang, Amir Hussain, Wadii Bouliia explained feature extraction technique using Local energy-shape Histogram(LESH). For research experiments the JSRT digital image database of radio chest radiograph is selected. The enhancement of radiograph images was using a contrast limited adaptive Histogram equalization [CLAHE] approach. Simulation results evaluated using classification accuracy performance measure [7]

Bhagyarekha U. Dhaware, Anjali C. Pise explained lung cancer detection method using Bayesian classifier and FCM Segmentation. In this paper feature selection is based on the statistical features by applying sequential forward algorithm

III. PROPOSED METHOD

In this research, to obtain more accurate results we divided our work into the Following three stages:

- Image Enhancement stage: at this stage we improve the image and eradicate any kind of noise, Corruption or interference from it. The following three methods are used for this purpose: Gabor filter (has the best results), Auto enhancement algorithm, and FFT Fast Fourier Transform (shows the worst results for image segmentation).
- Image Segmentation stage: at this stage we divide and segment the enhanced images, the used Algorithms on the ROI of the image (just two lungs), are: Thresholding Approach and Marker-Controlled Watershed Segmentation approach (this approach give improved results than thresholding approach).
- Features Extraction stage: at this stage the general features of the enhanced segmented image are extracted using Binarization and Masking Approach.

The proposed technique starts with preprocessing, histogram equalization, followed by thresholding and feature extraction. We used Matlab2016a to implement this design.

1. Pre-Processing

Extraction of images is performed accurately when the image is in binary form, i.e. a black and white image.

© 2019 IJSRET

Images in their default form are in Red Green Blue (RGB) form. This means that such images are logically represented as a combination of 3 matrices, each having a dimension of 1 × n, where n is the number of pixels. Such data is complex to process and therefore we convert this into a grayscale image. Grayscale images are logically a matrix with each pixel represented as a discrete level out of 0 to 255 [8]. The minimum level represents the darkest possible colour while the highest level represents the brightest colour.

The dimension of the matrix matches the dimension of the image represented in terms of pixels. Generally medical images are grayscale by default but we still check if there are any other components and if so, we convert it into grayscale to remove any such components. This also helps in removing hardware level errors that might have occurred during the scanning process.

2. Histogram Equalization

The Histogram of an image is a plot between the tones that can be represented in the image and the number of pixels that share the respective tones. A histogram which is high at lower levels indicates a dark image while the opposite indicates an overexposed image. The process of smoothening out the histogram of an image is called "Histogram Equalization".

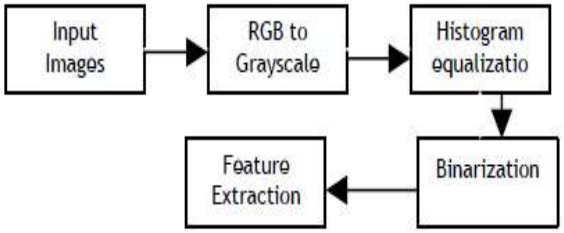

Fig.3 A Flowchart Showing the Pre-Processing techniques used in our design.

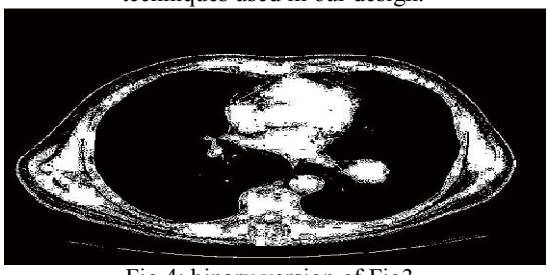

Fig.4: binary version of Fig3.

Histogram Equalization leads to a clearer and crisper image as compared to the original image. This includes sharper borders and highlighted objects in the image. It generates an image which has increased dynamic range and increased contrast [9]. So, we perform Histogram Equalization on the grayscale image to obtain its improved version with better contrast.

3. Thresholding

Thresholding is a process which is used for segmentation of an image by removing the data of pixels which are above or below a constant discrete level in a grayscale image. This constant level is called the Threshold [9]. It retains the required information in the image, while neglects the information that are irrelevant .We use thresholding in our system for segmentation and ROI (Region of Interest) selection. This also converts the image into a black and white image or binary image. The pixels of a binary image are represented in terms of two discrete levels, generally a 1 representing white or presence of data ,and a 0 i.e. absence of data. This helps in converting the image into a form that is ideal for feature extraction.

4. Feature Extraction

The next step involves the calculation of certain properties of the objects left in the binary image. The range of the values of these properties are used to the train the neural network and ultimately for decision making. This process is called Feature Extraction. We used the region props function in Matlab for this process. It is capable of calculating 21 types of features of objects identified in the image. However, we observed that feeding more number of features into the neural network decreased its efficiency.

So we decided to extract 6 features of the objects retained after thresholding out of the 21.These are Area, Perimeter, Major Axis Length, Minor Axis Length, Eccentricity and Solidity. For Each image, we took the mean value of these features of objects in the binary image. We removed insignificant features by running a loop that calculated these features for objects that were having area greater than 2 percent of the area of the largest object.

The mean values of all these features were computed and fed into a 6×70 matrix for all 70 images and this matrix was fed into the neural network.

V. PRRPOSED FLOW CHART

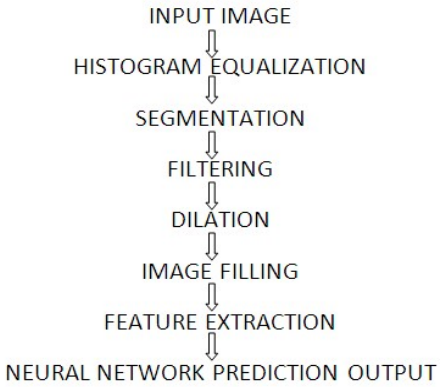

Fig.5 Proposed Flow Chart.

© 2019 IJSRET

At first, from the lung cancer patient the CT scan image is acquired. The main benefit of using computed tomography images is less distortion. CT images have low noise in contrast to X-ray and MRI images; henceforth they used for developing the technique. The CT images are acquired from NIH/NCI Lung Image Database Consortium. Dataset.

1. Histogram Equalization- It subdue the noise or further small instabilities in the image; equivalent to the suppressions of high frequencies in the frequency domain. All sharp edges that bear essential data about the image are also blurred during smoothing. Many methods, irrespective of implementation, share the same basic idea noise reduction through image blurring. Blurring can be done locally, as in the Gaussian smoothing model or in anisotropic filtering by calculating the dissimilarities of an image.

One of the most common problem in image processing is White noise. Even a high resolution Photo is bound to have some noise in it. For a high resolution picture a simple box blur may be sufficient. Mainly the idea of neighborhood filter is to calculate pixel weights, depending on their colors similarity. We describe two methods: the 'Median Filter' and 'Weiner filters'.

2. Segmentation- Image segmentation is most essential procedure for image analysis subsequent tasks. It divides the image into multiple segments of constituent regions or objects. Segmentation of medical images in 2D, slice by slice has many useful applications for the medical professional such as: visualization and volume estimation of objects of interest, detection of abnormalities (e.g. tumors, polyps, etc.), tissue quantification and classification, and more. [6].

The goal of segmentation is to simplify and/or change the representation of the image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics [7]. This results in a set of segments that collectively cover the entire image, or a set of contours extracted from the image (edge detection).

All pixels in a given region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s). Segmentation algorithms are based on one of two basic properties of intensity values: discontinuity and similarity. The first category is to partition the image based on abrupt changes in intensity, such as edges in an image. The second category is based on partitioning the image into regions that are similar according to a predefined criterion. 'Histogram thresholding' approach falls under this category.

3. Filtering An Image

Image filtering is useful for many applications, including smoothing, sharpening, removing noise, and edge detection. A filter is defined by a kernel, which is a small array applied to each pixel and its neighbors within an image. In most applications, the center of the kernel is aligned with the current pixel, and is a square with an odd number (3, 5, 7, etc.) of elements in each dimension. The process used to apply filters to an image is known as convolution, and may be applied in either the spatial or frequency domain. See Overview of Transforming Between Image Domains for more information on image domains.

Within the spatial domain, the first part of the convolution process multiplies the elements of the kernel by the matching pixel values when the kernel is centered over a pixel. The elements of the resulting array (which is the same size as the kernel) are averaged, and the original pixel value is replaced with this result. The CONVOL function performs this convolution process for an entire image.

Within the frequency domain, convolution can be performed by multiplying the FFT (Fast Fourier Transform) of the image by the FFT of the kernel, and then transforming back into the spatial domain. The kernel is padded with zero values to enlarge it to the same size as the image before the forward FFT is applied. These types of filters are usually specified within the frequency domain and do not need to be transformed. IDL's DIST and HANNING functions are examples of filters already transformed into the frequency domain. See Windowing to Remove Noise for more information on these types of filters.

The following examples in this section will focus on some of the basic filters applied within the spatial domain using the CONVOL function:

- Low Pass Filtering
- High Pass Filtering
- Directional Filtering
- Laplacian Filtering

Since filters are the building blocks of many image processing methods, these examples merely show how to apply filters, as opposed to showing how a specific filter may be used to enhance a specific image or extract a specific shape. This basic introduction provides the information necessary to accomplish more advanced image-specific processing.

3. Dilation

Is one of the two basic operators in the area of mathematical morphology, the other being erosion. It is typically applied to binary images, but there are versions

© 2019 IJSRET

that work on grayscale images. The basic effect of the operator on a binary image is to gradually enlarge the boundaries of regions of foreground pixels (i.e. white pixels, typically). Thus areas of foreground pixels grow in size while holes within those regions become smaller.

4. Image Filling

The imfill function performs a flood-fill operation on binary and grayscale images. For binary images, imfill changes connected background pixels (0s) to foreground pixels (1s), stopping when it reaches object boundaries. For grayscale images, imfill brings the intensity values of dark areas that are surrounded by lighter areas up to the same intensity level as surrounding pixels. (In effect, imfill removes regional minima that are not connected to the image border. See Finding Areas of High- or Low-Intensity for more information.) This operation can be useful in removing irrelevant artifacts from images.

5. Feature Extraction

This stage is an important stage that uses algorithms and techniques to detect and isolate various desired portions or shapes of a given image. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant, then the input data will be transformed into a reduced representation set of features. The basic characters of feature are area, perimeter and eccentricity. These are measured in scalar.

6. Neural Networks Output

Artificial neural networks are reckoning systems made up of numerous simple and highly interconnected processing elements, which process information by their dynamic state response to external inputs. In this paper, feed forward neural network with back propagation algorithm was used. The back propagation looks for the least of the error function in the weight space using the method of gradient descent. The weights are altered such that, the error function has the minimum value.¬ The algorithm has 252 input nodes, 20 hidden nodes and a couple of output nodes.

IV. RESULT

1. GUI Design- GUI Design of lung cancer detection

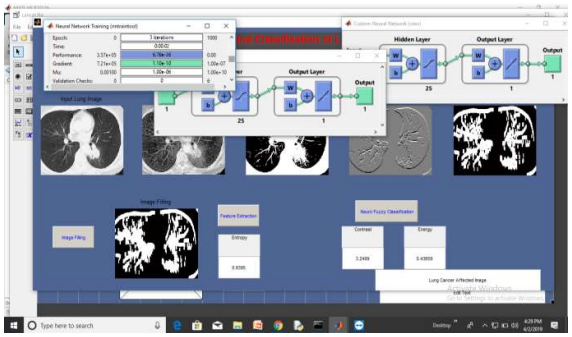

Fig. 6 Result of all training from input, sample and target.

V. CONCLUSION

The presented work is the detection of lung cancer nodules by applying implementation on image pre processing and segmentation. By implementing these steps the nodules are detected and then some features are extracted. Then the obtain features are used for the classification of the disease stages. Through the obtained nodules feature more information about the condition of lung cancer at the early stages. After that we have applied prediction model by applying that we predict from the obtained dataset from feature extraction to know how many people suffering from cancer or not. This technique helps the radiologists and the doctors by providing more information and taking correct decision for lung cancer patient in short time with accuracy. Therefore, this method is less costly, less time consuming and easy to implement.

REFERENCES

1. Ilya Levner, Hong Zhangm ,"Classification driven Watershed segmentation ", IEEE TRANSACTIONS ON IMAGE PROCESSING VOL. 16, NO. 5, MAY 2016
2. Mr.Vijay A.Gajdhane, Deshpande, "Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques"IOSR Journal of Computer Engineering (IOSR-JCE),e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 5, Ver. III , Sep – Oct. 2014
3. Am J Respir Crit Care Med Vol. 182, P1-P2, 2015. Online Version Reviewed September 2015, ATS Patient Education Series © 2010 American Thoracic Society
4 http://www.owlnet.rice.edu/~elec539/Projects99/ BAC /proj2/wiener.html (lastaccess:4/30/2017‖4:52am)
5. Mokhled S. AL-TARAWNEH, "Lung Cancer Detection Using Image Processing Techniques", Leonardo Electronic Journal of Practices and Technologies,(ISSN 1583-1078), Issue 20, p. 147-158, January-June 2014.
6. Venkateshwarlu K., Image Enhancement using Fuzzy Inference System, in Computer Science & Engineering, Master thesis, 2015.
7. Shapiro L.G., Stockman G.C., Computer Vision: Theory and Applications, Prentice Hall,2011.
8. Huang Q., Gao W., Cai W., Thresholding technique with adaptive window selection for uneven lighting image, Pattern Recognition Letters, Elsevier, 2014, 26, p. 801-808.
9. Suzuki K., "False-positive Reduction in Computer-aided Diagnostic Scheme for Detecting Nodules in Chest Radiographs",Academic Radiology, Volume 13, Number 10, pp.10-15, February 2015.
10. Levner I., Zhang H., Classification-Driven Watershed Segmentation, IEEE Transactions on Image Processing, 2007, 16(5), 1437-45.

© 2019 IJSRET

11. Non-Small Cell Lung Cancer, Available at: http://www.katemacintyrefoundation.org/ pdf/non-small-cell.pdf, Adapted from National Cancer Institute (NCI) and Patients Living with Cancer (PLWC), 2007, (accessed July 2011).
12. Tarawneh M., Nimri O., Arqoub K., Zaghal M., Cancer Incidence in Jordan 2008, Available at: http://www.moh.gov.jo/MOH/Files/Publication/Jordan%20Cancer%20Registry_2008%20Report_1.pdf, 2008,
13. Lung Cancer Database, Available at: https://eddie.via.cornell.edu/cgibin/ datac/signon.cgi,
14. Gonzalez R.C., Woods R.E., Digital Image Processing, Upper Saddle River, NJ Prentice Hall, 2008.
15. Cristobal G., Navarro. R., Space and frequency varient image enhancment based in Gabor representation, Pattern Recognition Letters, Elsevier, 1994, 15, p. 273-277.
16. Krishan A., Evaluation of Gabor filter parameters for image enhancement and segmentation, in Electronic Instrumentation and Control Engineering, Master. Punjab: Thapar University, 2009, p. 126.
17. Nunes É.D.O., Pérez M.G., Medical Image Segmentation by Multilevel Thresholding Based on Histogram Difference, presented at 17th International Conference on Systems, Signals and Image Processing, 2010.
18. Venkateshwarlu K., Image Enhancement using Fuzzy Inference System, in Computer Science & Engineering, Master thesis, 2010.
19. Anita chaudhary, Sonit Sukhraj Singh,"Lung Cancer Detection on CT Images Using Image Processing ", International transaction on Computing Sciences, VOL 4, 2012
20. B.V. Ginneken, B. M. Romeny and M. A. Viergever, "Computer-aided diagnosis in chest radiography: a survey", IEEE, transactions on medical imaging, vol. 20, NO.12,DEC-2001
21. S.Sivkumar,"Lung Nodule Detection Using Fuzzy Clustering and Support Vector Machine ", International Journal of Engineering and Technology Volume 5, No. 1, ISSN: 0975-4024, March 2013.
22. The DICOM Standards Committee. DICOM homepage. http://medical.nema.org/, September 2004.

© 2019 IJSRET