Multidisciplinary International Journal http://www.mijournal.in

(MIJ) 2018, Vol. No. 4, Jan-Dec **e-ISSN: 2454-924X; p-ISSN: 2454-8103**

EMPLOYBILITY OF ARTIFICIAL NEURAL NETWORK EARLY DETECTION OF LUNG CANCER

Sachin Gautam

ABSTRACT

Cancer detection is generally carried out manually by trained professionals and these techniques are majorly helpful in 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 to make the process more accurate and hassle-free. In the proposed work, image processing algorithms and artificial neural network have been employed to design anautomated process for early-stage detection of lung cancer.

In the recent era, image processing technique could be very rampantly used in numerous clinical fields for image development which allows in early detection and analysis of the remedy ranges, and time aspect also plays a totally pivotal role in coming across the abnormality within the target pictures like-lung most cancers, breast most cancers and so forth. This research focusses upon photograph excellent and accuracy. Image pleasant assessment in addition to improvement are structured upon enhancement degree where low preprocessing strategies are used based upon Gabor clear out inside Gaussian rules; thereafter the segmentation ideas are applied over the improved area of the photograph and the center for function extraction is obtained, similarly depending upon the general capabilities, a normality contrast is made .within the following research the essential detected features for correct photo contrast are pixel percent and overlaying labeling. In this research, we've got carried out classification based upon synthetic neural networks that are greater satisfactory than other contemporary classification strategies.

Keywords: *Neural network, artificial neural network, Lung Cancer Detection, Enhancement method, integration, image processing, Histogram.*

INTRODUCTION

Overview

Lung cancer is a disease, also referred to 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 to as benign tumors. If left unchecked, this cancer can grow and spread to other parts of the body. There are broadly two 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 the tumor depends on the appearance of tumor microscopically and the cell's growth rate. In grade 1: The cells look nearly like the normal lung cells. These 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 one by the world cancer research fund international (WCRF) in a list of the most diagnosed cancers worldwide [3]. About 1,825,000new cases of lung cancer were diagnosed in 2012, is 13 percent of all cancers worldwide, indicating its immense threat worldwide [3].

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 583,000diagnosed cases in 2012 were women, which is 8.8 percent 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 has been accomplished, but it's still a long and expensive procedure nonetheless.

It is also important to note 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 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 forvarious 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 THEDATABASE) 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 abnormalityis cancerous. The training system is a Back Propagation Artificial Neural Network.
The image processing steps include conversion into gray scale, Histogram Equalization, Thresholding, and Feature extraction. The Back Propagation Network (BPN) is trained using 70 images, and contains six input neurons and two 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.

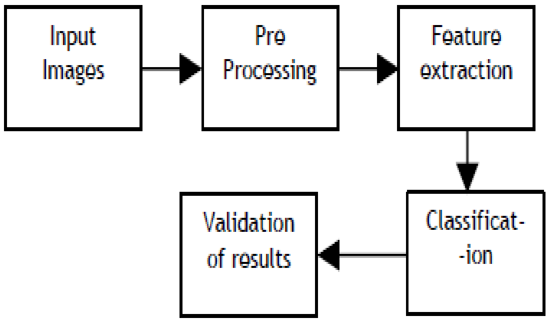

Fig.1: A Flowchart Showing the proposed method of Cancer Detection incorporated in our design.

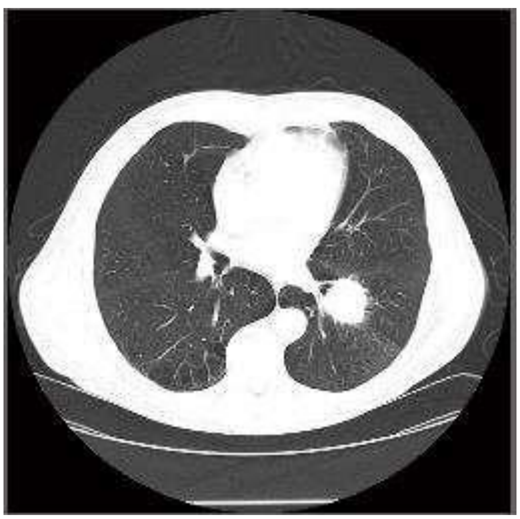

Fig 2: A grayscale converted image.

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 the 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]

SummrinaKanwalWajid, Kaizhu Huang, Amir Hussain, WadiiBouliia 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. Delaware, 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 a sequential forward algorithm.

PROPOSED METHOD

In this research, to extract greater correct consequences we divided our paintings into the following three stages: 1. picture Enhancement level: at this stage, we improve the photograph and eradicate any type of Noise, Corruption or interference from it. The subsequent three methods are used for this purpose: Gabor filters out (have the pleasant consequences), automobile enhancement algorithm, and FFT fast Fourier transform (suggests the worst results for picture segmentation). 2. photograph Segmentation stage: at this degree, we divide and section the improved pics, the used Algorithms on the ROI of the photo (just two lungs), are Thresholding method and Marker-managed Watershed Segmentation method (this approach supply stepped forward results than thresholding method). Three Capabilities Extraction level: at this level, the overall capabilities of the improvedsegmented photo have extracted the usage of Binarization and protected approach.
The proposed approach begins with preprocessing, histogram equalization, accompanied with the aid of thresholding and characteristic extraction. We used Matlab2016a to enforce this layout.

A. Preprocessing

Extraction of images is performed accurately when the image is in binary form, i.e., a black and white image. 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 $1 \times n$, where n is the number of pixels. Such data is complex to process, and therefore we convert this into a grayscale image. Gray scale 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 color, while the highest level represents the brightest color. The dimension of the matrix matches the dimensionof 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 helpsin removing hardware-level errors that might have occurred during the scanning process.

B. 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 Over exposed image. The process of smoothening out the Histogram of an image is called "Histogram Equalization."

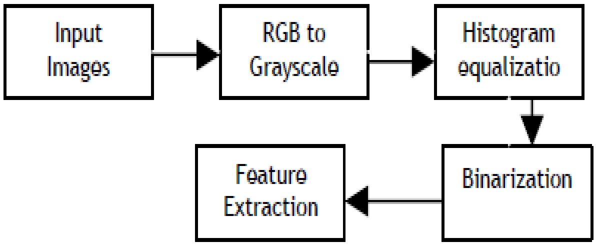

Fig.3: A Flowchart Showing the Preprocessing techniques used in our design.

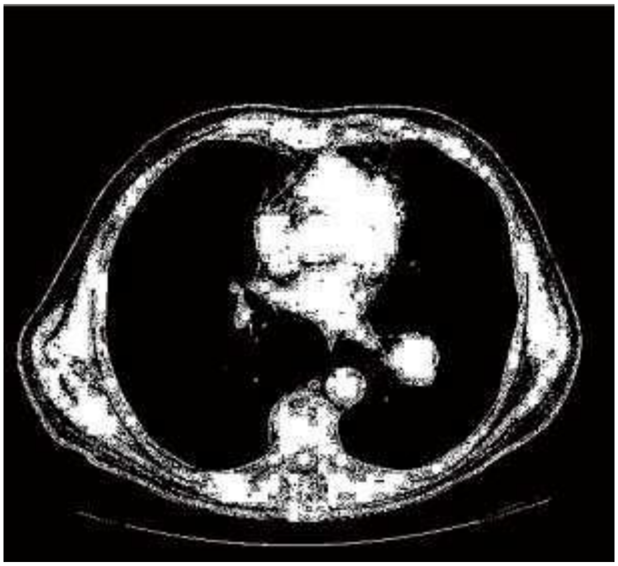

Fig.4: the binary version of Fig3.

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.

C. 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 is 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 is represented in terms of two discrete levels, generally, a one 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.

D. 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 is used to 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 number of features into the neural network decreased its efficiency. So we decided to extract six features of the objects retained after thresholding out of the 21. These are Area, Perimeter, Major Axis Length, Minor Axis Length, Eccentricity, and Solidity.
Each image, we took the mean value of these features of objects in the binary image. We removed in significant features by running a loop that calculated these features for objects that were having an 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.

PROPOSED FLOW CHART

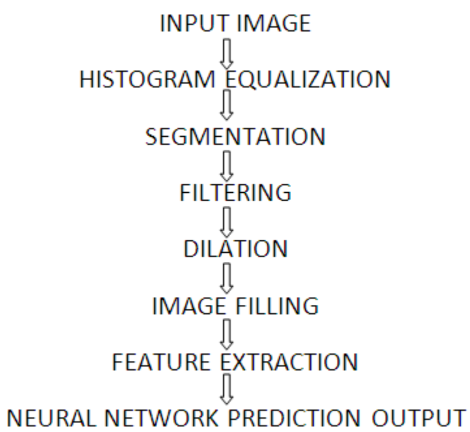

Fig.5: Proposed Flow Chart

At first, from the lung cancer patient, the CT scan image is collected. The main advantage of using computed tomography pictures is much less distortion. CT pictures have low Noise in comparison to X-ray and MRI snapshots; henceforth they used for growing the approach. The CT images are received from the NIH/NCI Lung Picture Database Consortium dataset.

Histogram Equalization: It subdues the Noise or in addition, small instabilities inside the photo; equivalent to the suppressions of high frequencies in the frequency area. All sharp edges that undergo critical records approximately the photo are also blurred during smoothing. Many techniques, regardless of implementation, proportion the identical fundamental idea noise reduction through picture blurring. Blurring may be finished locally, as inside the Gaussian smoothing version or in anisotropic filtering by way of calculating the dissimilarities of an image. One of the maximum commonplace hassles in photo processing is White Noise. Even a high decision image is sure to have some noise in it. For a high-resolution image, an easy box blur can be sufficient. Especially the idea of neighborhood clears out is to calculate pixel weights, relying on their colorings similarity. We describe two methods: the 'Median clear out' and 'Weiner filters.'

Segmentation: image segmentation is an essential technique for photo evaluation subsequent responsibilities. It divides the picture into multiple segments of constituent areas or gadgets. Segmentation of clinical photos in second, slice by means of the slice has many useful programs for the scientific expert together with: visualization and quantity estimation of items of interest, detection of abnormalities (e.g., tumors, polyps, and so on.), tissue quantification and category, and extra. [6].

The aim of segmentation is to simplify and/or change the representation of the photograph into something this is greater significant and less complicated to analyze. Image segmentation is generally used to find items and obstacles (traces, curves, and so on.) in snapshots. Greater exactly, photo segmentation is the procedure of assigning a label to every pixel in a picture such that pixels with the equal label proportion certain visual traits [7]. This effects in a hard and fast of segments that collectively cover the whole photograph or a fixed of contours extracted from the photo (edge detection). All pixels in a given vicinity are similar with recognize to a few functions or computed belongings, such as coloration, intensity, or texture. Adjoining areas are considerably unique with admire to the identical function(s). Segmentation algorithms are based on considered one of the fundamental houses of depth values: discontinuity and similarity. The first category is to partition the photograph primarily based on abrupt adjustments in-depth, which includes edges in a photo. The second one category is primarily based on partitioning the photo into regions which can be comparable consistent with a predefined criterion. 'Histogram thresholding' approach falls below this category.

Filtering an Image

Image filtering is beneficial for plenty of packages, such as smoothing, polishing, getting rid of the Noise, and edge detection. A filter out is described by means of a kernel, that's a small array applied to every pixel and its friends inside an image. In most packages, the middle of the kernel is aligned with the current pixel and is rectangular with an abnormal variety (three, five, 7, and many others.) of elements in every measurement. The manner used to use filters to a photograph is called convolution and may be implemented in either the spatial or frequency domain. See an overview of remodeling among image domain names for greater data on image domains.

Within the spatial area, the first part of the convolution procedure multiplies the elements of the kernel via the matching pixel values when the kernel is centered over a pixel. The elements of the ensuing array (that's the identical length as the kernel) are averaged, and the authentic pixel price is changed with this end result. The CONSOL function plays this convolution technique for a whole picture.

Within the frequency area, convolution can be accomplished with the aid of multiplying the FFT (rapid Fourier remodel) of the image by using the FFT of the kernel and then transforming again into the spatial domain. The kernel is padded with 0 values to enlarge it to the same size as the picture before the head FFT is implemented. Those kinds of filters are commonly distinctive inside the frequency area and do now not need to be transformed. IDL's DIST and HANNING features are examples of filters already transformed into the frequency domain. See Windowing to put off Noise for greater facts on those forms 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

On the grounds that filters are the constructing blocks of many image processing techniques, those examples simply show the way to follow filters, as opposed to displaying how a particular filter can be used to decorate a specific photo or extract a selected shape. This basic introduction gives the statistics important to perform more advanced photograph-unique processing.

DILATION

It's far one of the primary operators in the location of mathematical morphology, the other being erosion. It's far usually applied to binary pictures. However, there are versions those paintings on grayscale photos. The primary effect of the operator on a binary image is to steadily extend the boundaries of areas of foreground pixels (i.e., white pixels, usually). Consequently, regions of foreground pixels develop in length whilst holes within the one's regions emerge as smaller.

IMAGE FILLING

The infill function performs a *flood-fill* operation on binary and grayscale images. For binary images, infill changes connected background pixels (0s) to foreground pixels (1s), stopping when it reaches object boundaries. For grayscale images, infill brings the intensity values of dark areas that are surrounded by lighter areas up to the same intensity level as surrounding pixels. (In effect, infill 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.

FEATURE EXTRACTION

This level is a crucial stage that uses algorithms and techniques to stumble on and isolate various desired quantities or shapes of a given photo. Whilst the input information to an algorithm is simply too big to be processed, and it is suspected to be notoriously redundant, then the enter facts might be transformed into a reduced illustration set of functions. The primary characters of function are vicinity, perimeter, and eccentricity. Those are measured in a scalar.

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, a 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.

RESULT

4.1 GUI Design: GUI Design of lung cancer detection

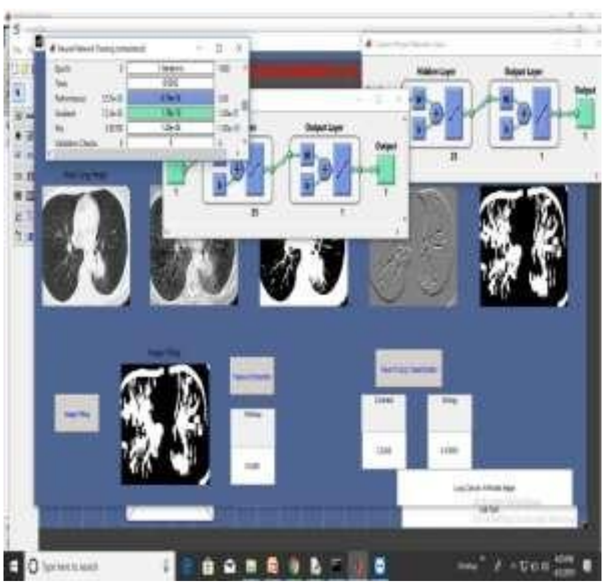

Fig 6: Result of all training from the input, sample, and target

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 the 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 method enables the radiologists and the doctors through providing extra data and taking correct selection for lung cancer patient in quick time with accuracy. Therefore, this technique is less high priced, much less time-ingesting, and easy to put into effect.