

Design and Development of Expert System for Lung image Analysis

N.Sharanya¹, P.S.N.Sumalya², R.Nageshwara rao³, V.Dinesh⁴, B.Sridhar⁵
^{1,2,3,4} Lendi institute of engineering & technology, jonnada, Vizianagaram.

⁵ Professor, ECE Department, Lendi institute of engineering & technology, jonnada, Vizianagaram.

Abstract: The known fact of wide impression of digitization on every aspect makes us refer to the artificial intelligence in lowering the difficulty of analyzing a medical image. Considering that the neural network applications in computer aided diagnosis represents the main stream of computational intelligence in medical imaging, it focuses on recent neural network developments. Keeping in mind the extremity of the requirement of how much perfectly it analyzes an image that well it can be justified, providing a clear view on the developments made by Mathematical morphology, back propagation neural networks in computer-aided detection, diagnosis and simulation describing number of applications. Representative techniques and algorithms are explained in detail illustrating how neural networks with fixed structure and training procedure could be applied to resolve a medical imaging problem followed by getting analyzed, processed and characterized by it. In the concluding section, an emphasis of comparisons among many neural network applications is included to provide the extent of neural computation in medical imaging.

Key words: Morphological operators, Medical image, Lung cancer, edge detection Back propagation neural networks

I. Introduction

Cancer is a class of diseases characterized by out-of-control cell growth. There are over 100 different types of cancer, and each is classified by the type of cell that is initially affected. Lung cancer is most common type of cancer. Cigarette smoking is the principal risk factor for development of lung cancer. Passive exposure to tobacco smoke also can cause lung cancer. There are two types of lung cancer, which grow and spread differently, are the small cell lung cancers (SCLC) and non-small cell lung cancers (NSCLC). The stage of lung cancer refers to the extent to which the cancer has spread in the body.

From the observations made on the paper "A Computer Based Feature Extraction of Lung Nodule in Chest X-Ray Image" by Nitin S. Lingayat, [1] we felt it is not too good to completely rely on filters whose ideality is not practically possible and for this reason we make use of gradient operators also in the process of gathering information related to cancer.

This paper is organized as follows: Detection and diagnosis of lung diseases are presented in detail in section 1. The definitions of operators in mathematical morphology are given in section 2. Back propagation of neural networks is in section 3. The new algorithm is proposed in detail in section 4. Results and discussion of the new model are presented in section 5. Concluding remarks are given in section 6.

II. Detection and diagnosis of lung diseases:

Making use of improved computer-aided diagnosis (CAD) scheme developed by Xu et al. the automated detection of lung nodules in digital chest images is retrieved to extract the image which is to be given as the input [2]. In the CAD scheme, nodule candidates were selected initially by multiple gray-level thresholds of the difference image (subtraction of a signal enhanced image and a signal suppressed image) and then classified into six groups. Between 50% and 70% of false positives were eliminated by adaptive rule-based tests and an ANN.

Automatic pathological diagnosis procedure proposed by Zhou et al. [3] named neural ensemble-based detection (NED) that utilizes an ANN ensemble to identify lung cancer cells in the specimen images of needle biopsies obtained from the bodies of the subjects to be diagnosed. An ANN ensemble is a learning paradigm where several ANNs are jointly used to solve a problem. The ensemble is built on two-level ensemble architecture and the predictions of those individual networks are combined by plurality voting. Keserci et al. [4] developed a computer-aided diagnosis scheme for automated detection of lung nodules in digital chest radiographs based on a combination of morphological features and the wavelet snake. In their scheme, an ANN was used to efficiently reduce false positives by using the combined features. The scheme was applied to a publicly available database of digital chest images for pulmonary nodules. Qian et al. [5] trained a computer-aided cytologic diagnosis (CACD) system to recognize expression of the cancer biomarkers histone H2AX in lung cancer cells and then tested the accuracy of this system to distinguish re-sected lung cancer from pre-

neoplastic and normal tissues. The major characteristics of CACD algorithms are to adapt detection parameters according to cellular image contents. Coppini et al. [6] described a neural network-based system for the computer-aided detection of lung nodules in chest radiograms. The approach is based on multi-scale processing and feed-forward neural networks that allow an efficient use of a priori knowledge about the shape of nodules and the background structure.

The methodology that we have implemented for the detection of cancer part is as follows in Fig 1.

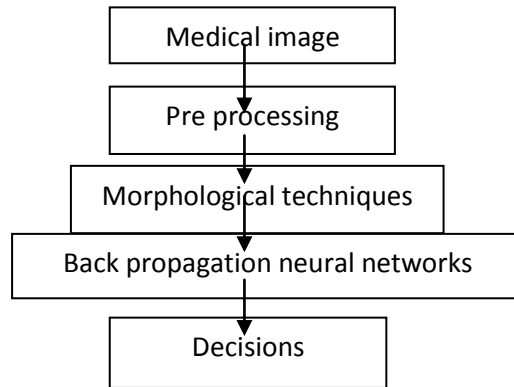


Figure: 1

III. Mathematical morphology:

The field of mathematical morphology contributes a wide range of operators to image processing, all based around a few simple mathematical concepts from **set theory**. The operators are particularly useful for the analysis of **binary images** and common usages include edge detection, noise removal, image enhancement and image segmentation. **Structure element:**

Morphological techniques typically probe an image with a small shape or template known as a **structuring element**. The structuring element is positioned at all possible locations in the image and it is compared with the corresponding neighborhood of pixels.

Erosion:

The erosion "shrinks" an image according to the shape of the structuring element, removing objects that are smaller than the shape. To erode an image, use the 'imerode' function. The imerode function accepts two primary arguments:

- The input image to be processed (gray scale, binary, or packed binary image)
- A structuring element object, returned by the strel function, or a binary matrix defining the neighbourhood of a structuring element

imerode also accepts three optional arguments: SHAPE, PACKOPT, and M.

$$E(A, B) = A \ominus (-B) = \bigcap_{\beta \in B} (A - \beta)$$

Dilation:

It makes objects grow in size.

$$D(A, B) = A \oplus B = \bigcup_{\beta \in B} (A + \beta)$$

Opening:

An opening is defined as erosion followed by a dilation using the same structuring element for both operations. It is used to remove small objects from an image while preserving the shape and size of larger objects in the image. For example, you can use the 'imopen' function to remove all the circuit lines from the original circuit image.

$$O(A, B) = A \circ B = D(E(A, B), B)$$

Closing: It defined simply as dilation followed by erosion using the same structuring element for both operations.

$$C(A, B) = A \bullet B = E(D(A, -B), -B)$$

Opening by reconstruction:

Suppose we want to identify image components containing a tall vertical segment. We can do this by opening with a vertical structuring element.

Erode first:
Then dilate:

Or you can do the opening in a single step by calling imopen:

The dilation step in the opening operation restored the vertical strokes, but the other strokes of the characters are missing. To get the entire characters containing vertical strokes morphological reconstruction can be used. For binary images, reconstruction starts from a set of starting pixels (or "seed" pixels) and then grows in flood-fill fashion to include complete connected components.

To get ready to use reconstruction, first define a "marker" image. This is the image containing the starting or seed locations. For our text example, the marker images will be the output of the erosion.

Next, define mask image. The flood-filling will be constrained to spread only to foreground pixels in the mask image. We can use the original text image as our reconstruction mask.

Finally, call "imreconstruct" to perform the operation. Performing morphological reconstruction, using the eroded image as the marker and the original image as the mask, is called "opening by reconstruction."

Gradient operators:

First-order derivatives of a digital image are based on various approximations of the 2-D gradient. The gradient of an image $f(x, y)$ at location (x, y) is defined as the vector.

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}.$$

It is well known from vector analysis that the gradient vector points in the direction of maximum rate of change of f at coordinates (x, y) . An important quantity in edge detection is the magnitude of this vector, denoted by $|\nabla f|$, where

$$|\nabla f| = \text{mag}(\nabla f) = [G_x^2 + G_y^2]^{1/2}.$$

This quantity gives the maximum rate of increase of $f(x, y)$ per unit distance in the direction of $|\nabla f|$. It is a common (although not strictly correct) practice to refer to $|\nabla f|$ also as the gradient. The direction of the gradient vector also is an important quantity. Let $\alpha(x, y)$ represent the direction angle of the vector $|\nabla f|$ at (x, y) . Then, from vector analysis,

$$\alpha(x, y) = \tan^{-1}\left(\frac{G_y}{G_x}\right)$$

where the angle is measured with respect to the x-axis. The direction of an edge at (x, y) is perpendicular to the direction of the gradient vector at that point. Computation of the gradient of an image is based on obtaining the partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$ at every pixel, location. The gradient operator is discussed as below.

- The Sobel method finds edges using the Sobel approximation to the derivative. It returns edges at those points where the gradient of I is maximum.
- The Prewitt method finds edges using the Prewitt approximation to the derivative. It returns edges at those points where the gradient of I is maximum.
- The "Roberts method finds edges using the Roberts approximation to the derivative. It returns edges at those points where the gradient of I is maximum.
- The "Laplacian of Gaussian method" finds edges by looking for zero crossings after filtering I with a Laplacian of Gaussian filter.
- The zero-cross method finds edges by looking for zero crossings after filtering I with a filter you specify.
- The "Canny method" finds edges by looking for local maxima of the gradient of I . The gradient is calculated using the derivative of a Gaussian filter. The method uses two thresholds, to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges. This method is therefore less likely than the others to be fooled by noise, and more likely to detect true weak edges.

IV. Back propagation neural networks:

Stimulation is applied to the inputs of the first layer in figure 2, and signals propagate through the middle (hidden) layer(s) to the output layer.

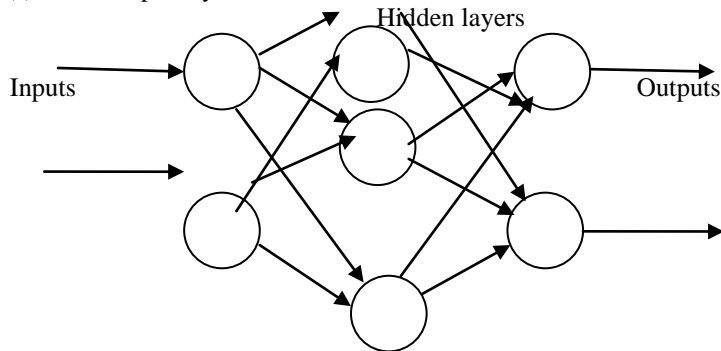


Fig: 2 a generalized network

Each link between neurons has a unique weighting value, Inputs from one or more previous neurons are individually weighted, then summed. The result is non-linearly scaled between 0 and +1, and the output value is passed on to the neurons in the next layer in figure 3

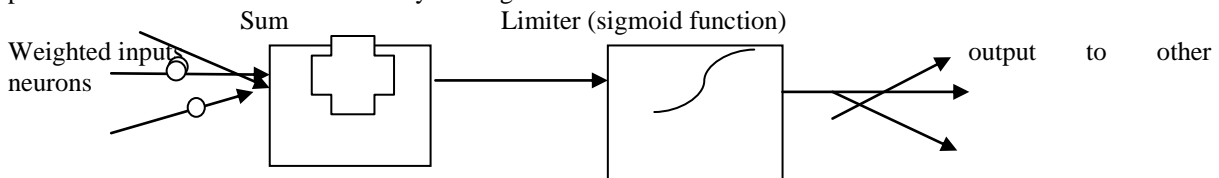


Fig: 3 the structure of neuron

V. Proposed method:

The method of detection procedure as given below. The procedure is programmed in MATLAB 7.8 and results are discussed in the next section.

Algorithm 1:

1. Read the image.
2. Apply the edge detection to identify the edges of the image using sobel operator.
3. Choose structure element "line" that can exhibit manipulation in both horizontal and vertical directions individually.
4. Perform image erosion.
5. Perform image dilation.
6. Apply image opening to the dilated image with 'disk' structure element.
7. Apply again erosion to the eroded image.
8. Perform opening by reconstruction for the results obtained by steps 6 & 7.
9. Perform image closing to the resulting image after carrying out opening in step 6.
10. Dilate the obtained image at step 8 with structure element 'disk'.
11. Opening by reconstruction of the complimented images obtained in steps 8 & 10.
12. Obtain regional maximus from step 11.
13. Superimpose the acquired image, marking on the original image.
14. Then transform the image into pseudo color image format.
15. Halt.

Neural network:

- Open neural network GUI interface in MATLAB.
- Choose pattern recognition tool.
- Select the input data & target data from the lung cancer database.
- Validation & test data sets to 15% of original data.
- Training the network to increase the iteration & mean square erosion percentage error percentage error to be reduced nearly to zero value.
- Generate the ROC curve & validate the algorithm.

VI. Results & discussion:

Taking the available techniques into considerations and trying to envelope them with analyzing much deeply the cancer prone areas making it more reliable, using the tools in MATLAB like spatial, frequency & averaging filters, morphological methods such as erosion, dilation, edge detection & thresholding. Using neural networks it is very clear how we can design an expert system that is equipped with above features can analyze a medical image effectively. Here are the inputs, x-ray of lungs of a normal, healthy person & those of patients who are supposed to have cancer. Using the proposed algorithm, the results are obtained, clearly explaining healthy and cancer effected lungs respectively in addition to that the areas are specified vividly.

In the Figure 3 below, the former one is an x-ray of a healthy person whose processing results are obtained as the latter one.

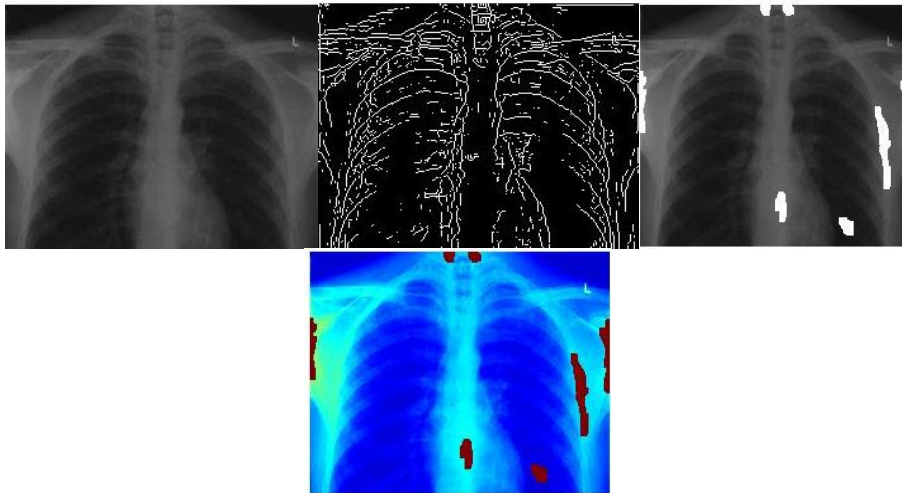


Figure 3 Lung x-ray without cancer and processed outputs

Here Fig 4 & 5 the three examples of x-rays of lung cancer patients on the left to the obtained results pointing out exactly where the cancer is.

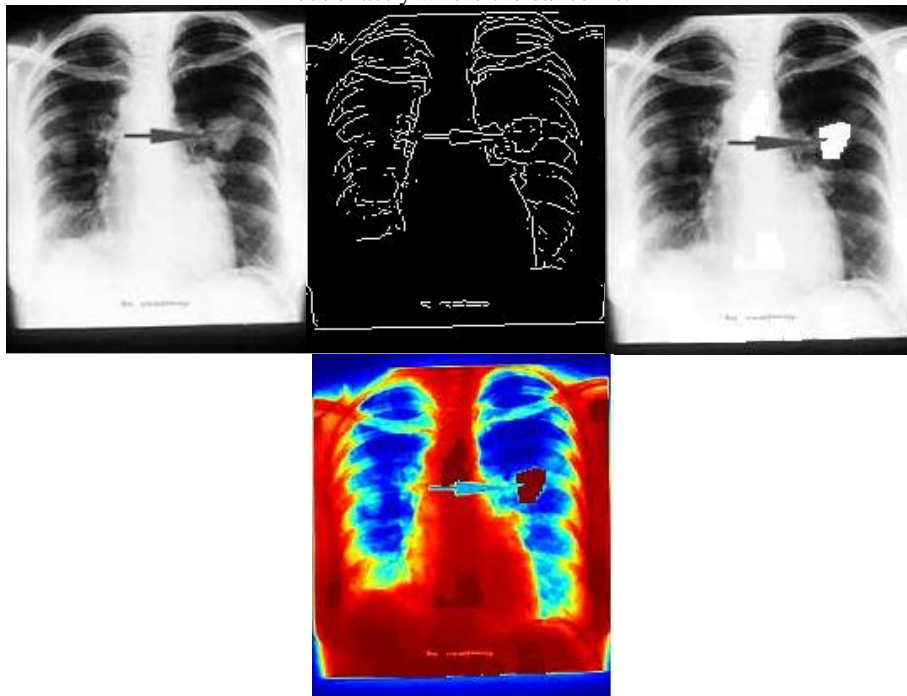


Figure 4 Lung x-ray with cancer marked by radiologist and processed outputs

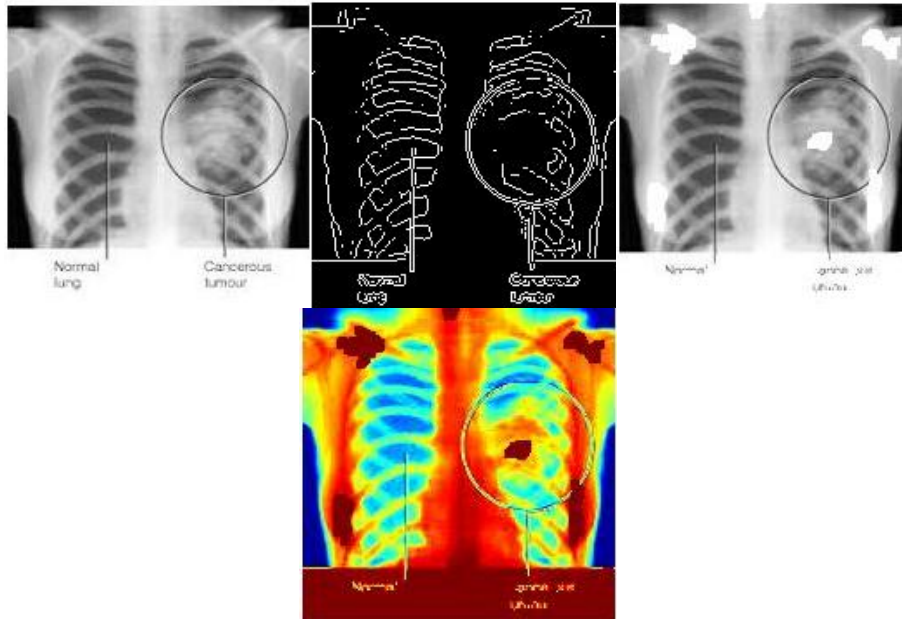


Figure 5 Lung x-ray with cancer marked by radiologist and processed outputs

Here the lung cancer x-ray data base has taken which consists both cancer without cancers. The results were stored in the output results are stored in the another database. The performance of the method is calculated by receiver operating characteristics(ROC) shown in the figure 6. This is the expected roc plot using npr tool that can justify the performance and efficiency in detecting the cancer cells. The true positive and true negative denotes the existence of cancer where it can identify and where it fails to, respectively. The false positive is where cancer does not exist and is sensed to be a cancer whereas false negative determines no cancer response at an unaffected area.

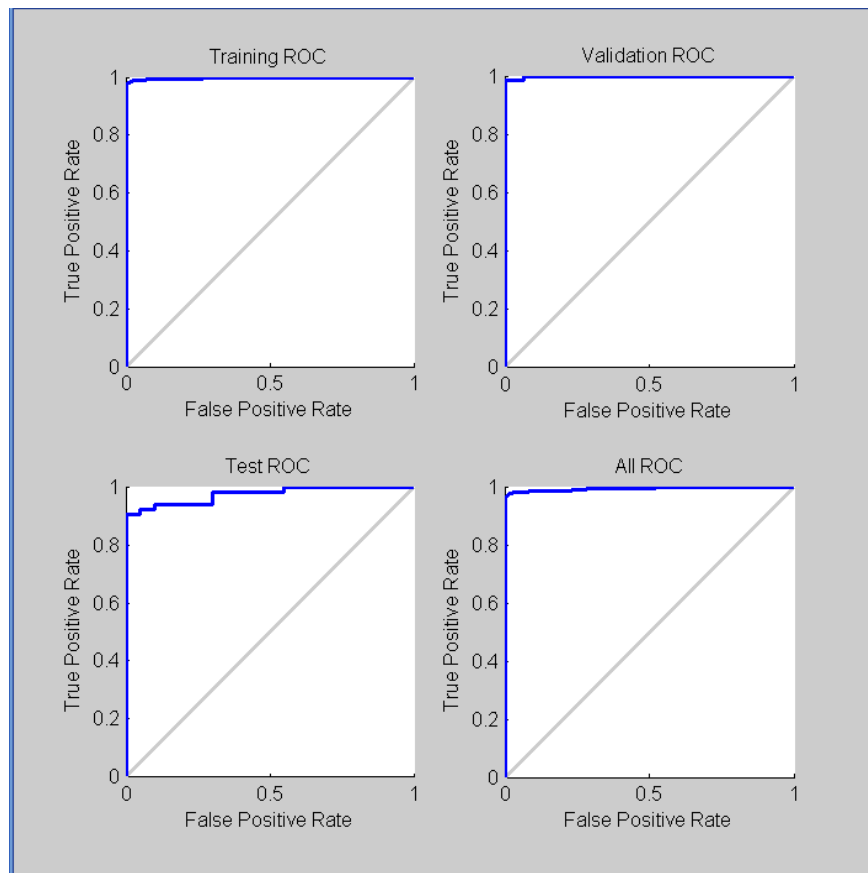


Figure 6:ROC curve

VII. Conclusion:

In this paper presented a morphological and neural network based lung image analysis and detection of the cancer. The performance of the algorithm is determined by Roc curve. In this method we observed the morphological operators are very powerful tools to mark the cancer and identify the area by using various structure elements. Here, Disc and line structure elements. ROC shows the method increases true positives and reduces true negatives. It is also observed that extra marked areas are shown others parts of the lung image. In future work It is to be minimized and the simulated method will be implemented on Hardware such as FPGA system.

Acknowledgements:

The authors shows their gratitudes to the college principal, management, faculty who were constantly encourage and made the work success, also thankful to the Govt Hospital, Vizayanagaram, who were provide the database for our work.

References:

- [1] Nitin S. Lingayat and Manoj R. Tarambale, "A Computer Based Feature Extraction of Lung Nodule in Chest X-Ray Image," *International Journal of Bioscience, Biochemistry and Bioinformatics* vol. 3, no. 6, pp. 624-629, 2013.
- [2] Xu XW, Doi K, Kobayashi T, MacMahon H, Giger ML. Development of an improved CAD scheme for automated detection of lung nodules in digital chest images. *Medical Physics* 1997;24(9):1395-403.
- [3] Zhou ZH, Jiang Y, Yang YB, Chen SF. Lung cancer cell identification based on artificial neural network ensembles. *Artificial Intelligence in Medicine* 2002;24(1):25-36.
- [4] Keserci B, Yoshida H. Computerized detection of pulmonary nodules in chest radiographs based on morphological features and wavelet snake model. *Medical Image Analysis* 2002;6(4):431-47.
- [5] Qian W, Zhukov T, Song DS, Tockman MS. Computerized analysis of cellular features and biomarkers for cytologic diagnosis of early lung cancer. *Analytical and Quantitative Cytology and Histology* 2007;29(2):103-11.
- [6] Coppini G, Diciotti S, Falchini M, Villari N, Valli G. Neural networks for computer-aided diagnosis: detection of lung nodules in chest radiograms. *IEEE Transactions on Information Technology in Biomedicine* 2003;7(4): 344-57.