



International Journal OF Engineering Sciences & Management Research

AN EFFICIENT LUNG CANCER DETECTION BASED ON ARTIFICIAL NEURAL NETWORK

Shashi Kiran.S

* Assistant Professor, JNN College of Engineering, Shimoga, Karnataka, India

Keywords: Artificial Neural Network (ANN), Classification, Lung Cancer, MLP, Wavelet Transform.

ABSTRACT

Cancer is the most important cause of death for both men and women. The early detection of cancer can be helpful in curing the disease completely. In this paper an ANN base approach is proposed to identify lung cancer from the raw chest x-ray images. It mainly consists of three stages; they are pre-processing, feature extraction and classification. The pre-processing involves resizing so that the further processing is easier. Feature Extraction involves extracting the features. The classification stage involves Artificial Neural network. Raw X-ray chest images are valuable in lung cancer diagnosis. On this basis, a lung cancer identification system is developed to analyze those features to judge whether lung cancer is present or not.

INTRODUCTION

Cancer is the uncontrolled growth of abnormal cells in the body. Cancerous cells are also called malignant cells. Symptoms of cancer depend on the type and location of the cancer. Cancer grows out of normal cells in the body. Normal cells multiply when the body needs them, and die when the body doesn't need them. Cancer appears to occur when the growth of cells in the body is out of control and cells divide too quickly. It can also occur when cells forget how to die.

There are different kinds of cancer. Cancer can develop in almost any organ or tissue, such as the lung, colon, breast, skin, bones, or nerve tissue. Lung cancer is the leading cause of cancer deaths in both men and women worldwide. Lung cancer is responsible for 29% of cancer deaths, more than those from breast cancer, colon cancer, and prostate cancer combined.

Lung cancer is a disease that occurs because of uncontrolled cell growth in tissues of the lung. This growth may lead to metastasis, which is the invasion of adjacent tissue and infiltration beyond the lungs.

A physical exam helps to determine if lymph nodes (glands) are enlarged, which may be a sign of lymph node involvement. The general practitioner will usually order a chest x-ray before referring the patient to a specialist.

Survival from lung cancer is directly related to its growth of its detection. The earlier the detection is, the higher the chances of successful treatment are [1].

Treatment depends on the histological type of cancer, the stage (degree of spread), and the patient's performance status. Possible treatments include surgery, chemotherapy, and radiotherapy. Most lung nodules do not have any symptoms and are found "accidentally" when a chest x-ray or computerized tomography (CT) scan is done for some other reason. There are many techniques to diagnose lung cancer. Most of these techniques are detecting the lung cancer in its advanced stages, where the patients' chance of survival is very low. Therefore, there is a great need for a new technology to diagnose the lung cancer in its early stages.

Medical diagnosis is one of major problem in medical application. Several research groups are working world wide on the development of neural networks in medical diagnosis. Neural networks are used to increase the accuracy and objectivity of medical diagnosis [2].

The flow of the topics is as follows. In section 2, we provide a brief review of the related work, in section 3, system design is presented. In section 4, pre-processing is discussed. In section 5, feature extraction technique is dealt. The method of training of Neural Network and the classification work carried out is explained in section 6. The results obtained during experimentation and conclusions drawn are elaborated in section 7.

RELATED WORK

The literature shows the various methods for the detection of lung cancer. K.A.G. Udeshani, R.G.N. Meegama and T.G.I. Fernando have used a novel approach to detect lung cancer from raw chest X-ray images [3]. At the first stage, pipeline of image processing routines have been done to remove noise and the lung from other anatomical structures in the chest X-ray is segmented and extracted regions that exhibit shape characteristics of lung nodules. Subsequently, first and second order statistical texture features are considered as the inputs to train a neural network to verify whether a region extracted in the first stage is a nodule or not. Zhi-Hua Zhou, Yuan Jiang, Yu-Bin Yang and Shi-Fu Chen have used an ensemble of Artificial Neural Networks to detect cancer cells in specimen images of needle biopsies [4]. A template matching algorithm and genetic algorithms have been used to detect lung nodules. These are quite fundamental approaches to detect nodules where the similarity between an unknown object

International Journal OF Engineering Sciences & Management Research

and a template are compared. Abhinav Vishwa, Alka Vishwa and Archana Sharma implemented an Artificial Neural Network to recognize the presence of cancer from the symptoms present in a patient [5]. Vinod kumar & Dr. Kanwal Garg uses a technique to remove noise using various filters and segment the lung to detect abnormal regions in the X-ray image and extracted regions that demonstrate area, perimeter and shape characteristics of lung nodules. These shape features are considered as the inputs to train a neural network and to verify whether a region is a malignant nodule or not [6]. Yongjun WU, Na Wang, Hongsheng ZHANG, Lijuan Qin, Zhen YAN, Yiming WUC developed computer-aided diagnostic scheme of the CT for the diagnosis of lung cancer based on Artificial Neural Networks (ANN) to assist radiologists in distinguishing malignant from benign pulmonary nodules. CT images of pulmonary nodules were analyzed. 21 CT radiological features of each case were carefully selected and quantified by three experienced radiologists. The 21 features and 5 clinical parameters were used as ANN input data. The result of ANN was compared with those of logistic regression by ROC curve analysis [7]. Almas Pathan and Bairu.K.saptalkar presented a method where image segmentation is done using Artificial neural network for detecting the lung cancer in its early stages [8]. Hence, in recent literature it is observed that principles of neural networks have been very widely used for the detection of lung cancer in medical images.

SYSTEM DESIGN

X-ray chest films are valuable in lung cancer diagnosis. General chest X-ray of Lung cancer is as shown in Fig. 1. The system first preprocesses the raw chest x-ray image and then extracts set of features to represent information of the image for Artificial Neural Network for recognition as shown in the fig. 2.



Figure 1: Chest X – Ray of Lung cancer

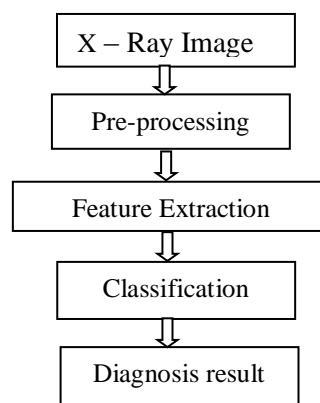


Figure 2: Stages of the proposed algorithm

PREPROCESSING

Pre-processing involves normalizing of the x-ray image so that the further processing is easier. The raw x-ray image is often of different sizes. This needs to be pre-processed in order to achieve the correct classification. The Preprocessing in the proposed lung cancer detection system involves the following two steps.

International Journal OF Engineering Sciences & Management Research

In the first step, resizing of the image is done. The chest x-ray images obtained will not be of same size and they need to be converted to a standard size. This is required so that the set of features obtained for all the x-ray images will be same. Here the resizing is done to get [50 X 50] matrix size image.

In the second step, conversion of the image to gray scale image is done. Along with the resizing of the image, conversion is done to get a gray scale image from RGB image. This is because for the RGB image, each pixel will be represented in three planes and the set of features obtained are more. This step is done to have reduced set of features.

FEATURE EXTRACTION

When the input data to an algorithm is too large to be processed and it is suspected to be redundant, then the input data will be transformed into a reduced representation set of features. Transforming the input data into the set of features is called features extraction.

Wavelet transforms decomposes a signal into a set of basis functions unlike Fourier transforms, whose basis functions are sinusoids. Wavelet transforms are based on small wave called wavelets of varying frequency and limited duration [9].

Wavelet can be applied on portions of an unknown signal to extract information from the unknown signal using a “shift, multiply and sum” technique called convolution [10]. Wavelet transform represents windowing technique with variable-sized regions. Wavelet analysis allows the use of long time intervals where we want more precise low-frequency information, and shorter regions where we want high-frequency information. Hence wavelet transform has been used for feature extraction.

If the features extracted are carefully chosen, it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. Wavelet transform is used to extract features of images. The number of features obtained for each of the x-ray image are 500. These features are then fed to the Classifiers for further processing.

The flowchart given in the fig. 3 describes the steps involved in feature extraction.

CLASSIFICATION

Artificial neural network is used as classifier in this system. Multi-layer feed forward neural networks are used for classification with only one hidden layer.

Number of feature vector coefficients decides the number of nodes in the input layer. Since the wavelet feature has 500 coefficients, input layer of 500 nodes are used. The number of nodes in the hidden layer is 300. One node is used in the output layer. In the output layer, the neuron is trained to produce output (logic 1) only when a cancerous image input pattern is presented and produce output (logic 0) for healthy image pattern. Therefore the number of nodes in the output layer is one.

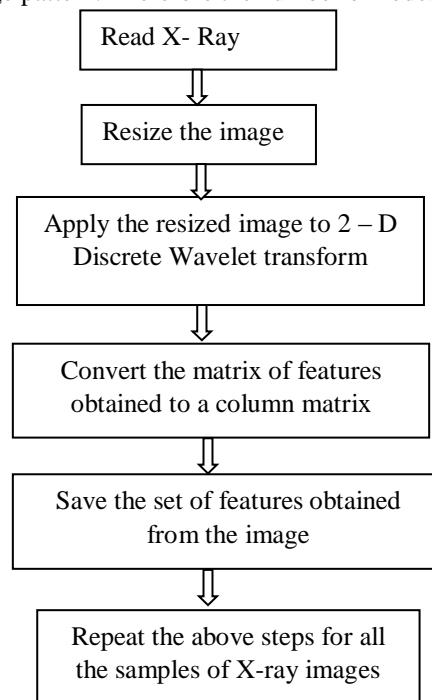


Figure 3: Flow chart for feature extraction

Training of Neural Network

A general three-layered Multi Layered Perceptron (MLP) back-propagation network is as shown in Fig. 4 with connection weights [11]. There are ‘I’ number of nodes in the input layer, ‘J’ number of nodes in the hidden layer and ‘K’ number of nodes in the output layer. The inputs to the layer are z_i , for $i = 1, 2, \dots, I$. The hidden and output values are denoted as y_j and O_k respectively where $j = 1, 2, \dots, J$ and $k = 1, 2, \dots, K$. The weight V_{ji} connects the output of i th neuron with the input to the j th neuron. The weight W_{kj} connects the output of j th neuron with the input to the k th neuron.

The network needs to be trained in a supervised mode. During the training process, a set of pattern examples is used, each example consisting of a pair with the training pattern input and corresponding target output. The patterns are presented to the network sequentially, in an iterative manner. The appropriate weight corrections are being performed during the process to adapt the network to the desired behavior. The iterative procedure

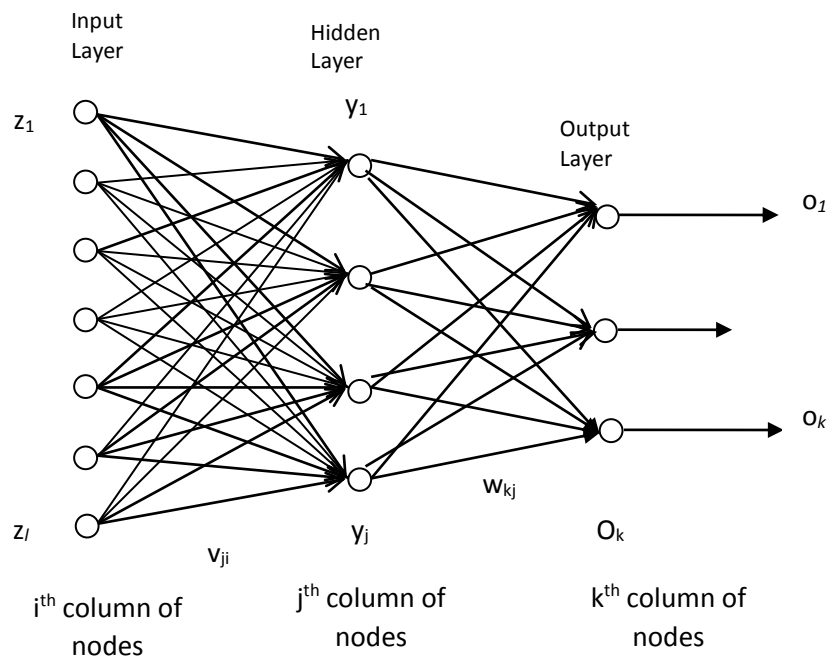


Figure 4: Three layered feed forward multi-layer neural classifier

continues until the connection weight values allow the network to perform the required mapping.

The minimization of the error function is carried out using a gradient-descent technique. The necessary corrections to the weights of the network for each iteration n are obtained by calculating the partial derivative of the error function in relation to each weight, which gives a direction of steepest descent. A gradient vector representing the steepest increasing direction in the weight space is thus obtained. Since minimization is required, the weight update value uses the negative of the corresponding gradient vector component for that weight. The delta rule determines the amount of weight update based on this gradient direction along with a step size.

The error function is defined as the mean square sum of differences between the output values of the network and the desired target values. The error function is expressed as

$$E_p = \frac{1}{2} \sum_{k=1}^k (d_{pk} - o_{pk})^2 \quad \text{----- (1)}$$

for a specific pattern ‘p’. The subscript ‘p’ refers to a specific pattern that is

at the input and produces the output error, ‘d’ is desired response vector and ‘o’ is the output vector.

Activation function used for the output of a neuron is the unipolar sigmoid function given by

$$f(net) = \frac{1}{1 + \exp(-\lambda net)} \quad \text{----- (2)}$$

where λ is the slope of the sigmoid function and ‘net’ is the net input of the neuron.

Each neural classifier has to be first trained to recognize the cancerous and healthy image and produce the desired output when it is presented with the wavelet features of a particular image. Error Back Propagation training method is employed for this purpose.

International Journal OF Engineering Sciences & Management Research

Weights are updated using the error (difference) between the actual output of the network and the desired (target) output, during training. Weights are adjusted repeatedly and error is calculated either until error reduces to 0.01 or 5000 epochs of training is reached. Training graph of neural network is as shown in Figure 5. The final weights at the end of the training represent the memory of neural network. They are saved and used for lung cancer detection. For each of lung cancerous and healthy image, 30 samples are used for training. The steps involved in training of neural network is shown in the flow chart in Figure 6.

Testing of Neural Network

In the testing phase, the features of the testing image are fed to the network and feed forward computations are carried out to obtain the output of the network. The output of the network lies in the range 0 to 1 due to sigmoid function. These are rounded to nearest binary value ['0' or '1'].

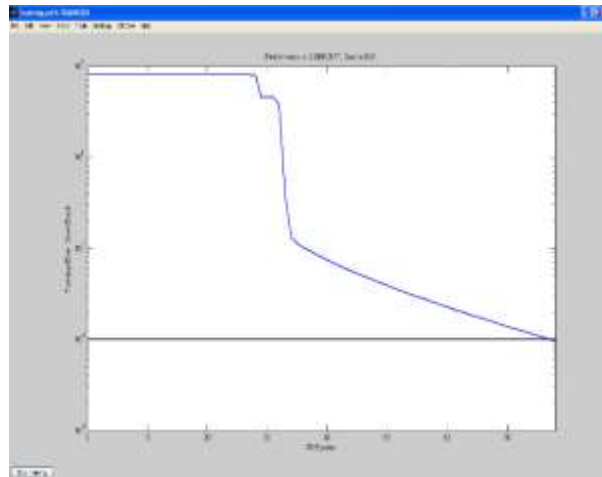


Figure 5: Training graph of Neural Network

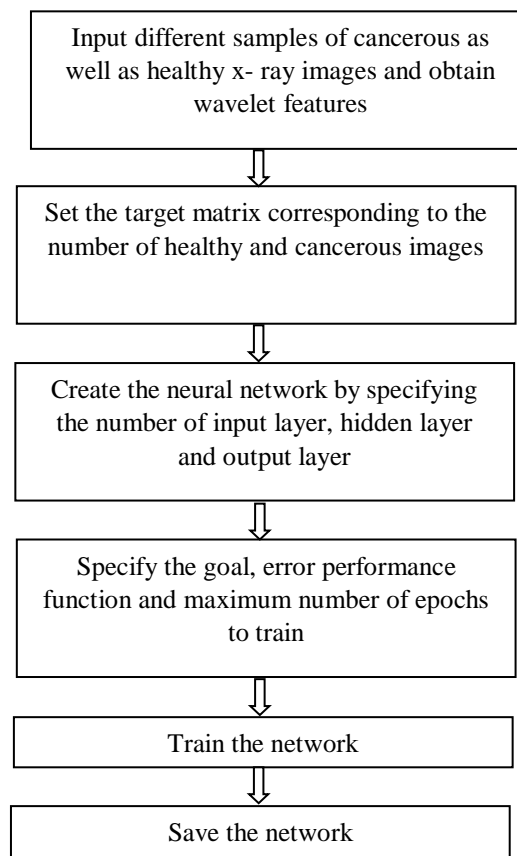


Figure 6: Flowchart for training process



International Journal OF Engineering Sciences & Management Research

This binary output is compared with target value of the learnt patterns. If it matches with the target value, the pattern is said to be recognized and is displayed. Once the Artificial Neural Network is trained to recognize the cancerous and healthy images, it is ready to use for recognizing the x-ray image as cancerous or healthy. During recognition phase, Artificial Neural Network has the capacity to generalize and identify the cancerous images with little variations when compared to the images used for training.

RESULTS AND CONCLUSION

The proposed system describes a novel procedure which uses the Wavelet features and Artificial Neural Network for the detection of lung cancer in chest x-ray images. Separate training network is considered to train the cancerous and healthy images. The performance of the system was evaluated for 70 different images and observed to possess a recognition accuracy of 92%.

The codes have been written for the images to be recognized. Experimental results show that cancerous as well as healthy images are recognized correctly. The percentage of error can be reduced by extracting more features from input images or by taking more number of samples for training.

REFERENCES

- [1] Jun WEI , Yoshihiro HAGIHARA , Akinobu SHIMIZU, Hidefumi KOBATAKE, *Optimal image feature set for detecting lung nodules on chest X-ray images*, CARS 2002, Computer Assisted Radiology and Surgery, 2002, pp 706-711
- [2] N. Ganesan, K. Venkatesh, M. A. Rama & Malathi Palani, *Application of Neural Networks in Diagnosing Cancer Disease Using Demographic Data*, International Journal of Computer Applications (0975 - 8887), Volume 1 – No. 26, pp. 76 – 85.
- [3] K.A.G. Udeshani, R.G.N. Meegama & T.G.I. Fernando, *Statistical Feature – based Neural Network Approach for the detection of Lung cancer in Chest X – Ray images*, International Journal of Image Processing (IJIP), volume (5); Issue (4), 2011
- [4] Zhi-Hua Zhou, Yuan Jiang, Yu-Bin Yang, Shi-Fu Chen, *Lung Cancer Cell Identification Based on Artificial Neural Network Ensembles*, Artificial Intelligence in Medicine, 2002, vol.24, no.1, pp.25-36, Elsevier.
- [5] Abhinav Vishwa, Alka Vishwa, Archana Sharma, *Pre – diagnosis of Lung Cancer using Feed Forward Neural Network and Back Propagation Algorithm*, International Journal on Computer Science and Engineering (IJCSSE), ISSN : 0975-3397 Vol. 3 No. 9 September 2011.
- [6] Vinod Kumar & Dr. Kanwal Garg, *Neural Network Based Approach for Detection of Abnormal Regions of Lung Cancer in X-Ray Image*, International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 1 Issue 5, July - 2012
- [7] Yongjun WU, Na Wang, Hongsheng ZHANG, Lijuan Qin, Zhen YAN, Yiming WUC, *Application of Artificial Neural Networks in the Diagnosis of Lung Cancer by Computed Tomography*, Sixth International Conference on Natural Computation (ICNC 2010), 2010
- [8] Almas Pathan, Bairu.K.saptalkar, *Detection and Classification of Lung Cancer Using Artificial Neural Network*, International Journal on Advanced Computer Engineering and Communication Technology Vol-1 Issue:1 :ISSN 2278 – 5140
- [9] Liyakathunisa, C.N. Ravi Kumar, “A Novel and Efficient Wavelet based Denoising for Super Resolution Reconstruction of Low Resolution Mammogram Images” in Proc Medical Imaging: Perspectives on Perception and Diagnostics”, 2-Day International Symposium Organized in Conjunction with the Seventh Indian Conference on Computer Vision, Graphics And Image Processing (ICVGIP) 2010, 9-10 December 2010, IIT Delhi.
- [10] Raghuvver.M.Rao and Ajit.S.Bopardikar, *Wavelet Transforms, Introduction to Theory and Applications*, Pearson Education, 2000
- [11] Jacek.M.Zurada, *Introduction to Artificial Neural Systems*, Jaico publishing house, 3rd Edition, 1999.