

# IDENTIFYING LUNG CANCER IN ITS EARLY STAGE USING NEURAL NETWORK AND GA ALGORITHM

Nancy,Paramjit Kaur

**Abstract**— Medical data mining is one of the major issues in this modern world. Medical problems are often in each and every human being. Cancer is one of the most dangerous diseases a human can ever had. Lung cancer is one of them. Lung cancer is a disease that occurs due to the uncontrolled cell growth in tissues of the lung. It is very difficult to detect it in its early stages as its symptoms appear only in the advanced stages. Aim of this paper is to automate the classification process for the early detection of Lung Cancer. To justify this research, it includes classification algorithm i.e. Neural Network and for optimization GA (Genetic Algorithm) is used. Evaluation would be done on the basis of correctly classified sample data. For testing and training diacom images has been used.

**Index Terms**— ANN, GA, NSLC, SMLC, GLCM, BFO, BPNN.

## I. INTRODUCTION

Lung cancer is a disease that occurs due to the 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. Treatment depends on the type of cancer, the stage, and the patient's performance and will power. Though there are various possible treatments include surgery, chemotherapy, and radiotherapy. Survival totally depends upon stage, overall health, and other factors. Symptoms that may suggest lung cancer includes:

- dyspnea (shortness of breath with activity),
- hemoptysis (coughing up blood),
- chronic coughing or change in regular coughing pattern,
- wheezing,
- chest pain or pain in the abdomen,
- cachexia (weight loss, fatigue, and loss of appetite),
- dysphonia (hoarse voice),
- clubbing of the fingernails(uncommon),
- dysphasia(difficulty swallowing),
- Pain in shoulder ,chest , arm,
- Bronchitis or pneumonia,
- Decline in Health and unexplained weight loss.

It is important to find out what kind of lung cancer a person has. The different types of carcinomas, involving different

regions of the lung, may cause different symptoms and are treated differently. Various techniques used for diagnose lung cancer are Chest Radiography (x-ray), computed Tomography (CT), Magnetic Resonance Imaging (MRI scan) and Sputum Cytology. However, these techniques are expensive and detecting the lung cancer in its advanced stages, where the chance of survival is very low. The early detection of cancer can be helpful in curing the disease completely. So the requirement of techniques to detect the occurrence of cancer nodule in early stage is increasing. Aim of this study is to automate the classification process for early detection of Lung Cancer. To justify research, it includes classification algorithm i.e. Neural Network and for optimization Genetic Algorithm is used. Evaluation would be done on the basis of correctly classified sample data. For training and testing diacom images has been used.

## II. METHODOLOGY

This system is fully implemented in MATLAB and tested using diacom images.

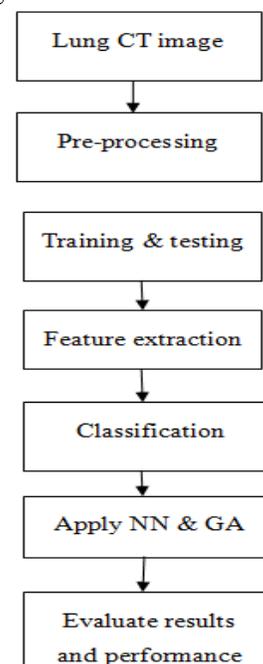


Fig. Methodology

*A. Images Collected*

909 scan images have been collected from the private hospital. Lung cancer images are stored in .dat file. Large database has been stored to run the propose work.

*B. Preprocessing*

Pre-processing is the initial step for detecting the lung cancer. Image denoising algorithms is used in image processing. White noise is one of the most common problems in image processing. Even a high resolution Photo is bound to have some noise in it. The input image is a normal RGB image. The RGB image is converted into grey scale image because the RGB format is not supported in Matlab.

*C. Training and testing of samples*

Training and testing of lung cancer samples is done by using own database. The neural net has to be trained to perform own pattern classification. Both feed forward and feed forward back propagation neural networks are used for classification. When a network has been structured for a particular application, that network is ready to be trained. To start this process the initial weights has to be chosen randomly. Then, the training, or learning, begins.

*D. Feature Extraction*

In order to decrease the complexity of processing for an exacting task, it is very important to recognize and extract required features. It is not necessary that all the attributes of an image are useful for knowledge extraction Geometrical features like Autocorrelation, contrast, cluster prominence, cluster shade, dissimilarity, energy, entropy, homogeneity, maximum probability, sum variance, sum entropy, difference variance, difference entropy and information measure have been estimated from the separated lung nodules.

*E. Classification*

In recent years, many advanced classification approaches, such as neural networks, fuzzy-sets, expert system and SVM have been widely applied for image classification. In this paper we are using Neural Network for classification.

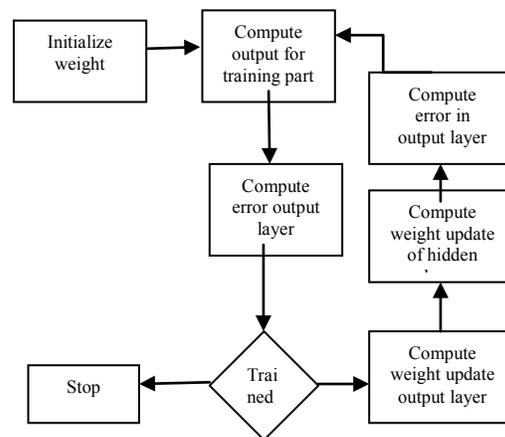
*F. Neural Networks (NN):*

An artificial neural network is a mathematical model based on biological neural networks. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation.

Working of Neural Network:

- Load images

- Create neural network
- Train neural network
- Test targets
  - cancer samples classified as cancerous
  - cancer samples classified as normal
  - normal samples classified as normal
  - normal samples classified as cancerous
- Classification matrix in percentage.



**Fig. FLOWCHART OF NEURAL NETWORK**

**Advantages of neural network:**

The various advantages of using neural network are given below:

- Neural Networks are flexible in changing environment. Rule based systems or programmed systems are limited to the situation for which they have been designed.
- Through learning, system is developed not by programming; neural nets teach themselves the pattern.
- Neural network learns to recognize the patterns which exist in the data set. Pattern Recognition is a powerful technique for harnessing the information in data and generalizing about it.
- Neural networks can easily build informative models because they can handle very complex situations

Back propagation is a method of training artificial neural network. It requires a desired output for each value in order for calculation of loss function gradient. Trained method is offered by back propagation method.

**BACKPROPAGATION NEURAL NETWORK**

Back-propagation learning algorithm

Basically the error back-propagation process consists of two passes through the different layers of the network a forward pass and a backward pass. The algorithm is as follows:

- (1) Step 0. Initialize weights. (Set to small random values)

- (2) Step 1. While stopping condition is false do steps 2–9
- (3) Step 2. For each training pair do steps 3–8 Feed-forward
- (4) Steps 3: Each input unit ( $X_i, i = 1 \dots n$ ) receives input signal  $X_i$  and sends signal to all units in the layer above (the unseeable units).
- (5) Steps 4: Each hidden unit ( $Z_j, j = 1, \dots, p$ ) sums its weighted input Signals

**GENETIC ALGORITHM (GA)**

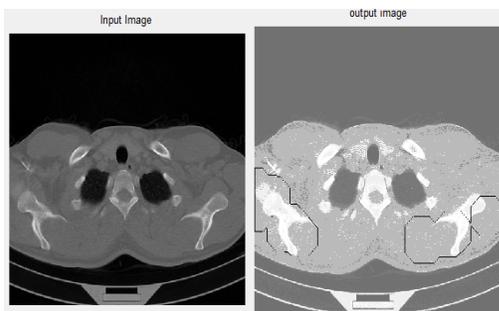
Genetic algorithms (GA's) are search algorithms that work via the process of natural selection. They begin with a sample set of potential solutions which then evolves toward a set of more optimal solutions.

GA is machine learning model, and following steps are taken in order to solve the problem :

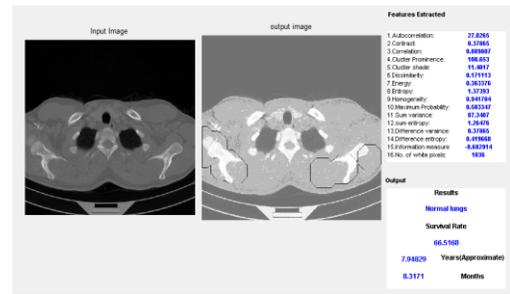
- Initialize random population with time
- Evaluate fitness function
- Test for termination case
- Initialize time counter
- Select sub population
- Select parents
- Evaluate new fitness function
- Agitate mated population
- Select survivors from fitness function
- End GA

**III. RESULTS**

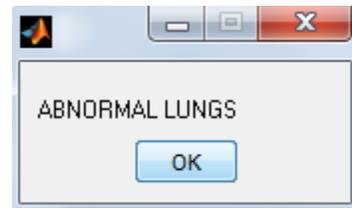
Early detection of cancer can be helpful in curing the disease completely. So the requirement of techniques to detect the occurrence of cancer nodule in early stage is increasing. Neural Networks play a vital role in automating the classification process for early detection of Lung Cancer. Prognosis of early diagnosis of Lung cancer with BPNN models has the best performance in large data sets and for more accuracy, after that result has been passed through GA algorithm for optimization.



**Fig.** After the feature extraction, classification is done.



**Fig.** On the basis of features extracted, classification has been done, that either lung belongs to normal or abnormal section.



**Fig:** Classification using Neural Network has been done, and from the features extracted, we came to know that lung finds out abnormal.

**IV. FUTURE SCOPE**

Future Scope lies in the technique where we can use BFO, Bacterial foraging optimization algorithm (BFOA) has been widely accepted as a global optimization algorithm of current interest for distributed optimization and control. BFOA is inspired by the social foraging behavior of *Escherichia coli*. BFOA has already drawn the attention of researchers because of its efficiency in solving real-world optimization problems arising in several application domains.

**V. REFERENCES**

- [1] Sandeep Kumar Saini, Gaurav, Amita Choudhary, "Detection of Lung Carcinoma using Fuzzy and ACO Techniques," IJERT, Vol.3 Issue 8, pp.903-906, 2014.
- [2] Ada Rajneet Kaur, "Feature Extraction and Principal Component Analysis for Lung Cancer Detection in CT scan Images," IJARSSCE, Vol.3, pp.187-190, 2013.
- [3] A. D. Back, T. Chen, "Universal Approximation Of Multiple Non Linear Operators By Neural Networks, Neural Computation," 2002, pp. 2561 – 2566.
- [4] Ankit Agrawal and Alok Choudhary, "Identifying HotSpots in Lung Cancer Data Using Association Rule Mining," IEEE, pp. 995-1002, 2011.
- [5] C. Bishop, "Curvature Driven Smoothing: A learning Algorithm for Feed Forward Networks," IEEE, vol. 4, 1993, pp. 461-489.
- [6] Derong Liu; Zhongyu Pang; Lloyd S.R, "A Neural Network Method for Detection of Obstructive Sleep Apnea and Narcolepsy" - Based on Pupil Size and EEG. 2008V-19 I-2 18. D. Itchhaporia.
- [7] D. Wang, "Pattern recognition : neural networks in perspective," IEEE, vol. 8, 1993, pp.5-60.
- [8] Frenster, J.H, "Neural Networks for Pattern Recognition in Medical Diagnosis", Annual International Conference in the IEEE Engineering in Medicine and Biology Society, vol. 12, No. 3, issued 1990, pp. 1423-1424.

- [9] F. Takeda, S. Omatu, "High Speed Paper Currency Recognition by Neural Networks," IEEE, vol. 6, 1995, pp. 25 – 30.
- [10] H.A. Rowley, S. Baluja, T. Kanade, "Neural Network-Based Face Detection," PAMI, 1998, pp. 23 – 28.
- [11] Hernandez, C.A. et al. "How to Choose the Training Data for Neural Network Medical Diagnosis Systems", ISA, pp. 283-290 (1993).
- [12] I. Sluimer, A. Schilham, M. Prokop, and B. Ginneken, "Computer Analysis of Computed Tomography Scans of Lungs: A Survey," IEEE Transactions on Medical Imaging, vol. 25, no. 4, 2006.
- [13] Kiyani, T., and Yildirim, T, "Breast Cancer Diagnosis Using Statistical Neural Networks." International XII. Turkish Symposium on Artificial Intelligence and Neural Networks. University Besiktas, Istanbul, Turkey.2003.
- [14] K.S. Ravichandran, P. Thirunavukarasu, R.allaswamy, R. Babu, "Estimation Of Return On Investment In Share Market Through Ann," JATIT, 2007, pp. 44 – 54.

**Nancy** is Pursuing Mtech in Computer Science from Indo Global College of Engineering, Abhipur, Distt.Mohali. Her current field of working is "Identifying Lung Cancer in its early stage using Neural Network and GA algorithm."