Volume 8, No. 3, March – April 2017
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info
ISSN No. 0976-5697

Early Detection of Lung Cancer using Artificial Neural Network

Niveditha.B.M
Dept of Computer Science
Christ University
Bengaluru, India

Nachamai.M
Dept of Computer Science
Christ University
Bengaluru, India

Abstract: The aim of the paper is to detect the lung cancer in the early stage. As early detection of the lung cancer is a testing issue, as a consequence of the structure of the lung tumor cells, where the greater part of the cells will be covered or not developed. Lung cancer is the most extreme sort of danger among each one of the disease with less survival rate. It is difficult to break down the development of tumor cells at its underlying stages (Stage I and Stage II). This paper demonstrates the early detection of lung malignancy by utilizing artificial neural network based algorithms such as Levenberg-Marquardt Algorithm, Bayesian Regularization Algorithm and Conjugate gradient Algorithm.

Keywords: Artificial Neural Network, Back propagation, Mean Squared Error, Regression

I. INTRODUCTION

In this advanced period of modernized and totally robotized incline, the impact of medical imaging in the field of healthcare diagnostics is playing a vital role and developing massively. The prediction of the diseases is becoming easier through different medical image techniques such as X-ray radiography, medical photography, Magnetic Resonance Imaging (MRI), medical ultrasonography, endoscopy, Positron Emission Tomography (PET), Computed Tomography (CT) etc. The diseases can be detected effortlessly by utilizing these imaging techniques. The lung cancer is considered as the second significant reason behind the malignancy demise around the world. Especially, the early detection of the lung malignancy is extremely troublesome as the symptoms appear just at the propelled stages of the lung cancer. Thus, in this manner it prompts to an expansion in the death rate among every single other kind of cancer. There is a need of a tool to analyze the lung cancer by distinguishing it in the underlying stages.

II. LITERATURE REVIEW

In [1] Bhagyashri G. Patil, Prof. Sanjeev N. Jain proposed CT Image Segmentation by comparing two different technique of segmentation, for example, thresholding and watershed which are utilized to detect the cancer cell. This paper concludes that the watershed method of segmentation is better which gives an accuracy of 85.27% over the quality than thresholding which gives an accuracy of 81.24% . In [2] Sukhjinder Kaur has introduced two segmentation methods such as Artificial Neural Network (ANN) and a Fuzzy C-Mean (FCM) clustering algorithm, for sputum shading images to essentially identify the lung malignancy in its underlying stages. The manual examination of the sputum image samples is inaccurate, tedious and requires intensive trained individual in order to avoid the diagnostic errors. This paper concludes by comparing two segmentation method along with its advantages so that further work should be possible by using Artificial Neural Network as a part of inclination to Fuzzy C-Mean (FCM) clustering algorithm and also conveys that Fuzzy C-Mean (FCM) clustering algorithm is bad at low intensity variations. In [3] Nikita Pandey, Sayani Nandy presented an approach for distinguishing the carcinogenic cells from lung CT scan images and another strategy to identify the cancerous cells effectively from the CT scan images by diminishing the detection error made by the specialist's exposed eye for therapeutic survey based on Sobel edge recognition and label matrix. In [4] Masaood A. Hussain, Tabassum M. Ansari, Prarthana S. Gawas, Nabanita Nath Chowdhury proposed a well plan of Computer Aided Diagnosis (CAD) framework to detect malignant cell in CT scan images. The proposed CAD framework involves four stages. They are pre-processing, evaluating region of interest (ROI), feature extraction and final classification using Artificial Neural network (ANN). The pre-processing stage comprises of different image enhancement techniques such as noise removal, contrast stretching, thresholding, with the specific end goal to enhance the visibility of tumor cells in CT scan images. The ROI is segmented from the upgraded image and feature extraction is done. The extracted features are given has inputs to the artificial neural network in the classification stage. Once training is completed, testing will be done in which the same steps will be applied to the testing image. Final result will be generated that whether the image is cancerous or not. In [5] Mokhled S. AL-TARAWNEH proposed a way to identify features for accurate image comparison as pixels rate and mask-labelling. In [6] Shraddha G. Kulkarni, Sahebrao B. Bagal presents lung tumor detection framework by using image processing and soft computing techniques which considers the methods such as Gaussian filter and anisotropic diffusion filter for pre-processing of CT images. Necessary features like energy, standard deviation, entropy and mean absolute deviation are calculated for the extracted lung nodule. Based on these parameters, comparative study of classifiers is done using SVM and KNN classifier. In [7] Prashant Naresh, Dr. Rajashree Shettar proposed a method where CT scan images of lung cancer patients are taken as input and during preprocessing Gaussian white noise is removed. The preprocessed image is then fed to feature extraction in which textual and structural features of nodule are extracted. These features are fed to the classifier. The classifier is trained and tested, the stable classifier is used to predict the severity of the cancer patient i.e., stage-I or stage-II. In [8] G. Dheepak, S.

Premkumar, R. Ramachandran detected nuclei and cytoplasm using Fuzzy C-Mean clustering by providing thresholding technique as pre processing step and also the further technique applied is Hopfield Neural Network segmentation in which the nuclei and cytoplasm are detected from the overlapped classes of cancer cells. The accuracy of these methods will be obtained by calculating the standard deviation, entropy and correlation as performance parameters. In [9] Damodar Dipika A. and Prof. Krunal Panchal proposed a comparative analysis using different algorithms such as Bayes classification and active contour modeling, Random walks algorithm and ROI, Click &Grow algorithm, Self-Organizing Maps (SOM) Method and Artificial Neural Network (ANN) , ROI Neuro Fuzzy classifier. In [10] Miss. Sayali Kanitkar, Prof. Nilima D. Thombare, Prof. Sunita S. Lokhande proposed a work on discovering tumor and initial indications of the illness showing up in patient's lungs. The accessible lung cancer images are passed through the framework stages such as pre-processing, feature extraction and classification. Feature extraction is a fundamental stage which represents the ultimate outcome to decide the ordinariness or anomaly of an image. The extracted features are necessary for the classification procedure. The special features are considered to be extracted, for example, normal intensity, area, perimeter and eccentricity etc.

III. IMPLEMENTATION

Back propagation is a typical technique for training an artificial neural network system. The objective of back propagation is to improve the weights so that the neural network can figure out how to effectively outline contributions to yields. The back propagation technique comprises of repeatedly applying the chain rule through all of the possible paths in the network. However, there are an exponential number of directed paths from the input to the output. Back propagation's real power arises in the form of a dynamic programming algorithm, where the reuse of the intermediate results to calculate the gradient. The transmitting of an intermediate error backwards through a network is leading to the name *back propagation*. Back propagation is closely related to forward propagation, but instead of propagating the inputs forward through the network, the error is propagated backwards.

For instance, an artificial neural network for lung cancer detection is characterized by an arrangement of input neurons which will be activated by the pixels of an input image. After the neurotransmitters (synapses) are being weighted and transformed by a sigmoid function the activations of these neurons are then passed on to different neurons. This procedure continues repeating until, an output neuron is activated. This determines the input given is carcinogenic of Stage I or Stage II.

The network is trained by utilizing the back propagation rule based algorithms such as Levenberg Marquardt, Bayesian Regulation, Scaled Conjugate Gradient**.** The dataset is isolated in a ratio for training phase and testing phase of artificial neural network. Finally, the effectiveness of the neural network is calculated by the Mean squared error and the regression.

Algorithms

a. Levenberg-Marquardt Algorithm

trainlm is a neural network training function which redesigns the weight and bias values as indicated by LM optimization. trainlm is considered as the quickest back propagation calculation and is profoundly considered as a primary-decision supervised algorithm. This algorithm requires more space than different algorithms. trainlm supports the training with validation vectors and testing vectors. Validation vectors are utilized to quit training early if and only if the neural network's performance on the validation vectors neglects to enhance or remains the same for maximum epochs. Test vectors are utilized for further verification that the neural network is generalizing well, however it do not have any effects on training. Back propagation is used to ascertain the Jacobian(jinput) of performance(perf) as for the weight and bias of input. Every input is balanced according to Levenberg-Marquardt,

jj = j(nput) * j(input)

je = j(input) * E

d(input) = -(jj + I * mu) \ je

where, E is all errors and I is the identity matrix. Training quits when any of these conditions are satisfied:

- If the maximum no. of epochs (repetitions) are reached.
- If the maximum measure of time is exceeded.
- If the performance is minimized to the target.
- If the performance gradient falls underneath minimum gradient.
- If validation performance has expanded more than maximum fail times since the last time it diminished.

b. Bayesian Regularization Algorithm

The Bayesian regularization happens inside the Levenberg-Marquardt algorithm. Bayesian regularization reduces a linear combination of squared errors and weights. trainbr is a artificial neural network training function which can train any neural network the length of its weight, net input, and transfer functions have derivative functions. This function additionally adjusts the linear combination, towards end of training. Thus, the subsequent neural network has great generalization qualities. Back propagation is used to figure the Jacobian(jInput) of performance(perf) as for the weight and bias of the input. Every input is balanced by Levenberg-Marquardt,

jj = j(input) * j(input)

je = j(input) * E

d(input) = -(jj + I *mu) \ je

where, E is all errors and I is the identity matrix. Training quits when any of these conditions are fulfilled:

- If the greatest no. of epochs are reached.
- If the greatest measure of time is surpassed.
- If the performance is minimized to the targets.
- If the performance gradient falls underneath minimum gradient.
- mu surpasses mu_max.

c. Conjugate gradient Algorithm

traincgp is an artificial neural network training function which trains any neural network the length of its weight, net input, and transfer functions have derivative functions. Back propagation algorithm is for the most part used to compute derivatives of performance(perf) concerning the weight and bias inputs. Each input is balanced accordingly:

input = input + a*d(input);

where, d(input) is the search direction and the parameter 'a' is chosen to minimize the performance along the search direction. searchFcn is a line search function which is used in matlab to find the minimum point. The first search direction is taken as the negative of the gradient of performance. In the succeeding iterations, the search direction is figured from the new gradient and the past search direction.

d(input) = -g(input) + d(input_old) * Z;

where, g(input) is the gradient. The parameter Z can be processed in several ways. For the Polak-Ribiére variation of conjugate gradient, it is registered by,

Z = ((g(input)-g(input_old))' * g(input)) /norm_sqr;

where norm_sqr is the norm square of the past gradient, and g(input_old) is the gradient on the past iteration. Training quits when any of these conditions are satisfied:

- If the maximum no. of epoch is reached.
- If the maximum measure of time is exceeded.
- If the performance is minimized to the target.
- If the performance gradient falls beneath minimum gradient.
- When the validation performance has expanded more than greatest fail circumstances since the last time it diminish.

Training Phase: During the training phase the neural network is trained by utilizing respective algorithms with a certain ratio of input and target samples. The network is additionally balanced according to its error.

Validation Phase: Validation phase is utilized to measure the neural network generalization, and to end training when generalization quits improving.

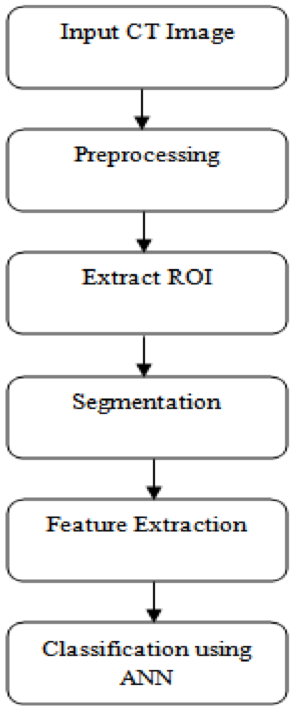
Testing Phase: Testing phase does not have any impact on training phase. Therefore, it provides a free measure of neural network performance amidst and after training.

Parameters

Mean Squared Error (MSE) is the normal of squared distinction among outputs and targets. Lower estimations of MSE are better. O indicates error free.

Regression (R) is a measure of the correlation amongst outputs and targets. The R estimation of 1 indicates a cozy relationship, 0 demonstrates an arbitrary relationship.

IV. BLOCK DIAGRAM



V. RESULTS

Case 1:

Training - 80%
Validation - 10%
Testing - 10%

Table 6.1 **Case 1**

Algorithm	Phase	Samples	MSE	R
Levenberg Marquardt Algorithm	**Training**	**107**	**0.0640546**	**0.863111**
	Validation	**13**	**0.128843**	**0.71395**
	Testing	**13**	**0.0620369**	**0.868457**
Bayesian Regulation Algorithm	**Training**	**107**	**0.0752967**	**0.835963**
	Validation	**13**	**0.00000**	**0.00000**
	Testing	**13**	**0.0281547**	**0.969762**
Scaled Conjugate Gradient Algorithm	**Training**	**107**	**0.107140**	**0.757032**
	Validation	**13**	**0.086478**	**0.822267**
	Testing	**13**	**0.25074**	**0.709850**

When the values of MSE and R are compared from the table 6.1 for the 80% of the training dataset, Bayesian Regulation algorithm gives best classification as MSE=0.0281547 and R=0.969762.

Case 2:

Training - 70%
Validation - 15%
Testing - 15%

Table 6.2 **Case 2**

Algorithm	Phase	Samples	MSE	R
Levenberg Marquardt Algorithm	**Training**	**93**	**0.0649471**	**0.860355**
	Validation	**20**	**0.0146192**	**0.990030**
	Testing	**20**	**0.154433**	**0.639753**
Bayesian Regulation Algorithm	**Training**	**93**	**0.0682161**	**0.852733**
	Validation	**20**	**0.00000**	**0.00000**
	Testing	**20**	**0.0803610**	**0.823783**
Scaled Conjugate Gradient Algorithm	**Training**	**93**	**0.0654125**	**0.859281**
	Validation	**20**	**0.0478676**	**0.901899**
	Testing	**20**	**0.116011**	**0.739337**

When the values of MSE and R are compared from the table 6.2 for the 70% of the training dataset, Bayesian Regulation algorithm gives best classification as MSE=0.0803610 and R=0.823783.

Case 3:

Training - 60%
Validation - 20%
Testing - 20%

Table 6.3 **Case 3**

Algorithm	Phase	Samples	MSE	R
Levenberg Marquardt Algorithm	**Training**	**79**	**0.0541653**	**0.885237**
	Validation	**27**	**0.100366**	**0.785396**
	Testing	**27**	**0.0890963**	**0.802542**
Bayesian Regulation Algorithm	**Training**	**79**	**0.0636845**	**0.863294**
	Validation	**27**	**0.00000**	**0.00000**
	Testing	**27**	**0.12393**	**0.774372**
Scaled Conjugate Gradient Algorithm	**Training**	**79**	**0.078188**	**0.825106**
	Validation	**27**	**0.0822649**	**0.821927**
	Testing	**27**	**0.0881435**	**0.810090**

When the values of MSE and R are compared from the table 6.3 for the 60% of the training dataset, SCG(Scaled Conjugate Gradient) algorithm gives best classification as MSE=0.0881435 and R=0.810090.

VI. CONCLUSION

The detection of lung cancer in the early stages is difficult as the tumour growth is less.By effective preprocessing, segmentation, extraction of ROI from the cancer images ,it can be easily detected using the artificial neural network by training and stabilizing the network.Thus the same network is used to test the new sample datasets. Using the Back propagation principle based algorithm-" **Bayesian Regulation**" implementated on the lung cancer datasets gives us the best results with almost 0 error and R value nearer to 1.

VII. ACKNOWLEDGEMENT

We would like to thank our Head of Department, Prof. Joy Paulose, Dept. of Computer Science, Christ University for the motivation and encouragement given to take up research and update ourselves in the field of research.

VIII. REFERENCES

[1] Bhagyashri G. Patil, Prof. Sanjeev N. Jain, "*Cancer Cells Detection Using Digital Image Processing Methods*", International Journal of Latest Trends in Engineering and Technology (IJLTET), Vol. 3 Issue 4 March 2014.

[2] Sukhjinder Kaur, "*Comparative Study Review on Lung Cancer Detection Using Neural Network and Clustering Algorithm*", International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 4, Issue 2, February 2015.

[3] Nikita Pandey, Sayani Nandy, "*A Novel Approach of Cancerous cells Detection from Lungs CT Scan Images*", International Journal of Advanced Research in Computer Science and Software Engineering (*IJARCSSE*), Vol. 2, Issue 8, 8 August 2012.

[4] Masaood A. Hussain[1], Tabassum M. Ansari[2], Prarthana S. Gawas[3], Nabanita Nath Chowdhury[4], "*Lung Cancer Detection Using Artificial Neural Network & Fuzzy Clustering*", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 3, March 2015.

[5] Mokhled S. AL-TARAWNEH, "Lung *Cancer Detection Using Image Processing Techniques*", Leonardo Electronic Journal of Practices and Technologies Issue 20, January-June 2012, p. 147-158.

[6] Shraddha G. Kulkarni[1], Sahebrao B. Bagal[2], "*LUNG CANCER TUMOR DETECTION USING IMAGE PROCESSING AND SOFT COMPUTING TECHNIQUES*", International Journal of Science Technology and Management, Vol.No.5, Issue No. 05, May 2016.

[7] Prashant Naresh[*], Dr. Rajashree Shettar[**], "*Early Detection of Lung Cancer Using Neural Network Techniques*", Prashant Naresh Int. Journal of Engineering Research and Applications, ISSN: 2248-9622, Vol. 4, Issue 8(Version 4), August 2014, pp.78-83.

[8] G. Dheepak, S. Premkumar, R. Ramachandran, "*LUNG CANCER DETECTION BY USING ARTIFICIAL NEURAL NETWORK AND FUZZY CLUSTERING METHOD*", International Journal of Power Control and Computation (IJPCSC), Vol. 7, No.1 – 2015, pp. 24-28.

[9] [1]Damodar Dipika A. and [2]Prof. Krunal Panchal, "*A Survey of Lung Tumor Detection on CT Images*", JETIR, Volume 2, Issue 12, December 2015.

[10] Miss. Sayali Kanitkar, Prof. Nilima D. Thombare, Prof. Sunita S. Lokhande, "Lung Cancer Detection and Classification: A Review", International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 12, December – 2013.