EasyChair Preprint

№ 3076

# Prediction of Lungs Cancer Using Machine Learning

Prasanta Das, Biplab Das and Himadri Sekhar Dutta



April 3, 2020

# Prediction of Lungs Cancer using Machine Learning

Prasanta Das[1], Biplab Kanti Das[2] and Himadri Sekhar Dutta[3]
[1,2]Calcutta Institute of Technology, Uluberia, India; [3]Kalyani Government Engineering College, Kalyani.
pd.researcher13@gmail.com, biplab118@gmail.com, himadri.dutta@gmail.com

**Abstract**

There is lot of progress made in the field of treatment of lung cancer in the last years (adjuvant chemotherapy, radio therapy, individualized therapy). Nonetheless, lung cancer is still remained the threat of society and cause of death of thousands of people in all over the world. This paper is all about detection of lungs cancer. Here Computer Tomography (CT) images are used to detect lungs cancer. There are several algorithms are used to detect Lungs cancer accurately. Here unsharp masking filter is use to filtering the image. Adaptive Canny edge detection algorithm is used to detect the edges and cancer affected areas. Neural network is used to classify the features and predict the probability of lung cancer. K-Nearest Neighbors is used to segment the cancer from lungs. Finally achieved the classification accuracy near about 99.5% by using Bayesian Regularization Neural Network (BRNN) and performance measure by Mean square error (MSE) noted as 0.0166.

***Keywords***: Lungs Cancer, K-Nearest Neighbors, Bayesian Regularization Neural Network

# 1 Introduction

Another name of lung cancer is carcinoma. It's actually a malignant tumor which is characterized by the uncontrolled growth of cell tissue. Lung cancer happened mainly because of tobacco smoking for a long period. It's a fact that 85% of lung cancer happened because of tobacco smoking. But it's a fact that 10%-15% of lung cancer affected people are never smoke in their life. Actually, it's happened because of passive smoking, air pollution, asbestos, radon gas [1]. Doctors can able to examine the interior body parts by the help of several testing process. Lung cancer is the most dangerous cause for man and women both. It's not an easy task to detect lung cancer in earlier stages. But there are some major facts that indicate that if we can detect it on earlier stage then it can help to reduce the mortality rate. Approximate 80% of lung cancer affected people are diagnosis correctly at the middle or advance stage. There is an estimation that 1.2 million of people every diagnosis with this disease which is 12.3% of total number of cancers diagnosed and approximate 1.1 million peoples are dying because of this disease which is 17.8% of total number of cancer death [3]. Image processing is an effective way to detect lung cancer easily because there are several useful tools which help in detection. In this proposed work multiple filtering and segmentation are used to detect lung cancer on earlier stage and analyze the data more accurately.

# 2 Literature survey

Prediction and detection of cancer in its early stages is main target of all researchers. For this they always try to develop a system which helps to predict and detect cancer in its primary stage. Different algorithms are used to improve the accuracy of the Early Prediction and Detection system by preprocessing, segmentation feature extraction and classification techniques. K. Senthil Kumar, Venkatalakshmi and K. Karthikeyan proposed in their work that cancer is a malignant tumor. Several algorithms like K-means clustering, Particle Swarm Optimization (PSO), Inertia-Weighted, Particle Swarm Optimization (IWPSO) and Guaranteed Convergence Particle Swarm Optimization

(GCPSO) also used for the experiment to compare the accuracy between them on different images. But here Adaptive median filter and GCPSO gives the highest accuracy which is 95.8079% [1]. A. Asuntha, A. Brindha, S.Indirani, Andy Srinivasan proposed in their work that there is multiple choice for choosing images which may be a CT, MRI or Ultrasound image. Gabor filter, Layer separation technique, Super pixel segmentation technique is use to separate the object or the region of interest from the other part of image. PSO, Genetic Algorithm and SVM are used to get a better result. Here accuracy is about 89.5% with reduction in false positive [2]. Babita Rani, Ashok Kumar Goel, Ravneet Kaur proposed in their work that modified approach of some algorithm which can achieve the better result. In this paper hybridization of three techniques are utilized that is Bacterial Forging Optimization (BFO), Principle Component Analysis (PCA) and Back Propagation Neural Network (BPNN). Here accuracy is 98.1881% [3]. Manasee Kurkure, Anuradha Thakare used CT images for the experiment. They used canny edge detection algorithm where canny filter used in feature extraction and output value output value gives to naïve bayes classification which is optimized by Genetic Candidate Group Search (GCGS) algorithm. Genetic Algorithm (GA) used to identify genes that help to classify patient lung cancer status with a notable predictive performance [4]. Neha Panpaliya, Neha Tadas, Surabhi Bobade, Rewti Aglawe, Akshay Gudadhe proposed a system where they use Histogram Equalization and neural network classifier to check the state of patient whether it is normal or abnormal. This will generate very accurate result for detection and prediction of lung cancer. So, early detection and prediction of lung cancer should play a important role in the diagnosis process and also improve the survival rate of patient [5]. Fatma Taher, Naoufel Werghi, Hussain Al-Ahmad, Rachid Sammouda had proposed the early detection of lung cancer is a challenging problem, due to the structure of the cancer cells, where most of the cells are overlapped with each other [6]. This paper presents KNN algorithm for segmentation and BRNN for classification.

# 3 Proposed Model for Lungs Cancer Detection

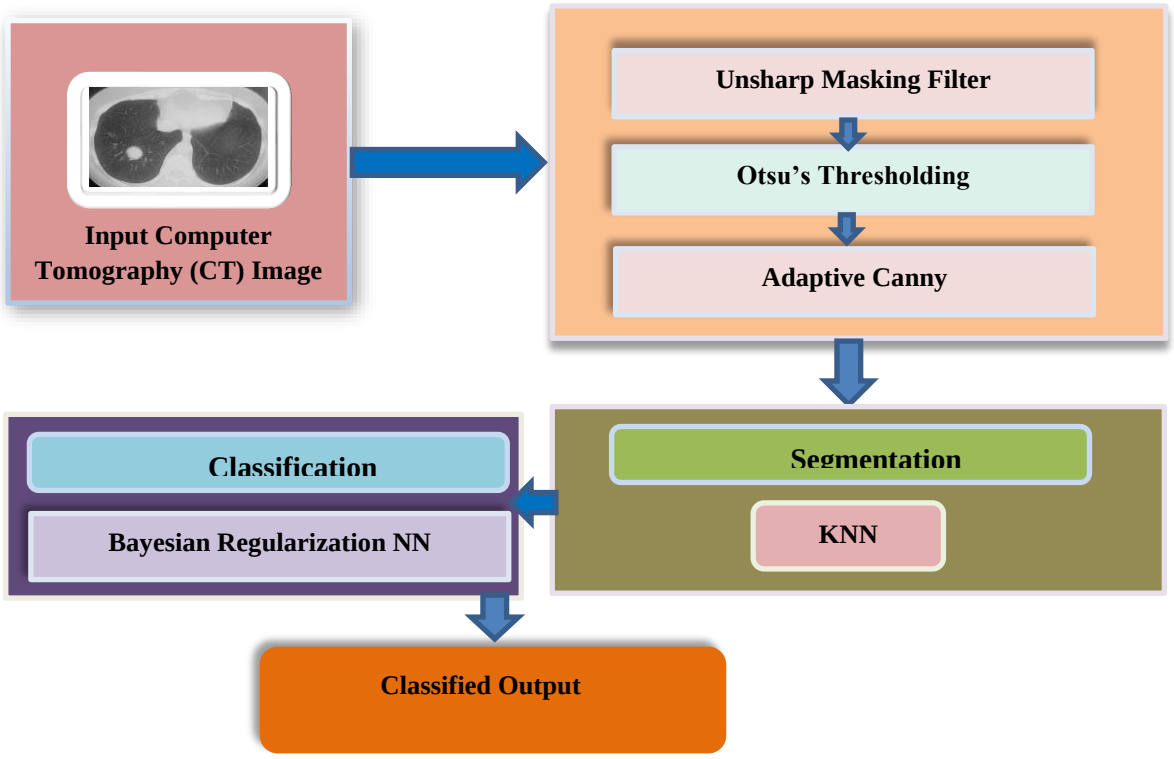



Fig 1: propounded model for lungs cancer detection.

## 3.1 Unsharp Masking Filter

To make an image sharp "unsharp mask" is used, though the name means opposite. Sharpening used to emphasize detail and texture, also it applied at the time of post processing of digital image. It is a technique for image sharpen which is more advantageable for the segmentation process. Unsharp masking used several purposes of an image pre-processing where we used for the blur image convert to the original sharpen or smooth image.

We compute the equation for unsharp masking,

$$P(\alpha, \beta) = T(\alpha, \beta) - T_1(\alpha, \beta) \quad (1)$$

where $T_1(\alpha, \beta)$ is a smoothed version of $T(\alpha, \beta)$ and $P(\alpha, \beta)$ is an input edge map of image.

## 3.2 OTSU'S Thresholding

Many more researchers are used different thresholding algorithm for the calculate the image background and foreground color intensity level. We were used Otsu's thresholding because of some advantage make it flexible. Basically, Otsu's thresholding method compute the three parameters which are weight, mean, variance of background and foreground of each level. It calculates the threshold with in class variance and end of the calculation we add two variances after that result multiplied their corresponding weight. Otsu's method compute with in class variance to the following equation,

$$\varphi_\delta^2 = \delta_\omega \varphi_\omega^2 + \delta_\vartheta . \varphi_\vartheta^2 \quad (2)$$

Where $\delta_\omega$is the weight and $\varphi_\omega^2$ is the variance of the background pixel. For the foreground pixel $\delta_\vartheta$ is represent the weight and $\varphi_\vartheta^2$ is represent the variance.

## 3.3 Adaptive Canny

In this section we were propounded the adaptive canny algorithm for detection of cancer region or find the actual effected area. Generally, edge map is little bit of necessary for the proper results, some techniques are failed to find the proper edge. In this paper we used adaptive canny which successfully perform of our lungs cancer detection model. Adaptive canny compute through Otsu's global thresholding techniques.

## 3.4 Propounded K-Nearest Neighbor for Segmentation

This section we proposed the KNN algorithm which successfully perform by the segmentation process and achieve the better segmentation result. In pattern recognition KNN needs to predicted value of K-based and training data for computation. It has some powerful application, such as pattern recognition, machine learning, text categorization, data mining, object recognition etc. In this segmentation process the KNN used to predict the model as well as used for predict the segmented result.

## 3.5 Bayesian Regularization

Here the neural network has become independent, using BR neural network we get the classifications results. Using BRNN we compute the classification process where BRNN is the most outstanding learning ability. NN provide some specific parameter which necessary to classify dataset. It contains layers and neurons, which get some inputs and generate corresponding output. this out passed through the layer into next layer of neurons and finally BRNN compute the specific fitting model which has less loss over the results. BRNN is predictive learning ability which is helps to get the better accuracy almost than some others method. It also handles gradient based optimization problem and measurement the performance using MSE which value noted as 0.0166. finally, we achieve the classification accuracy which near about 99.5%.

# 4 Result & Discussion

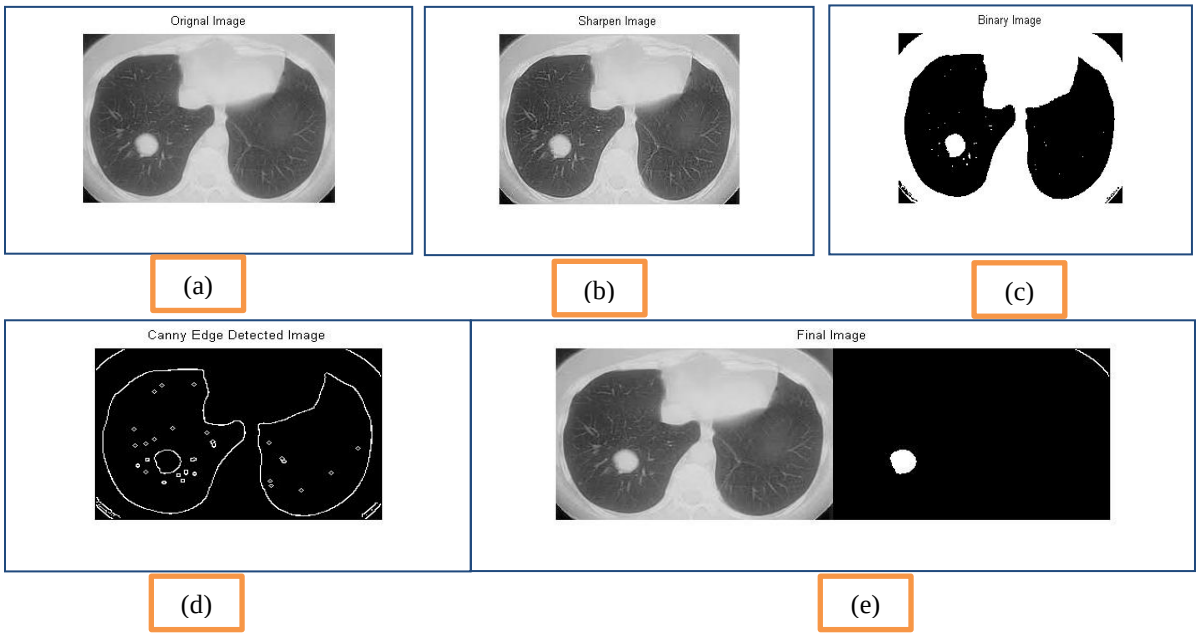
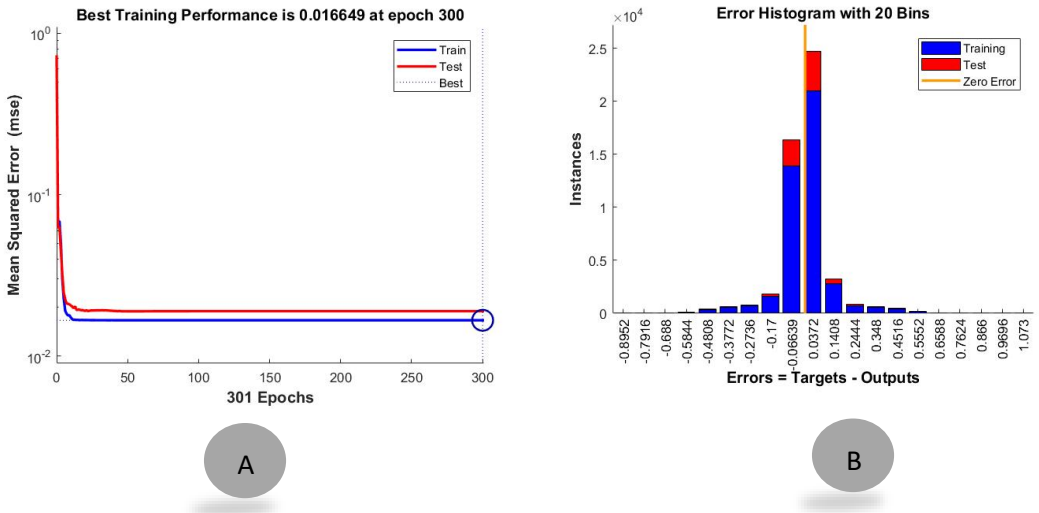



Fig 2: Using proposed algorithms detect lungs cancer (result).

Here all the figures show the different step of processing. JPG format 512 X 512-pixel image of lung be taken as for the experiment. Here lung image is sharpening for the easier edge detection. Otsu global thresholding is used for set the threshold value of canny edge detection where proposed edge detection algorithm can easily detect the edges from the binary image. But there are lots of small pixels and lung edges also detected which is not required to show the result. So, these edges are removed and the affected area is filled using whole filling algorithm.

## 4.1 Evaluate the Performance

Evaluate based performance show the result of graphical form which include the training, testing and validation-based performance .BR algorithm is including probabilistic distribution, sum of squares of network weight and gradient based optimization. fig 3A shown as the training and testing state, fig 3B shown as the error histogram. Fig 3C and 3D shown as the gradient based optimization of the given data and the training, testing and target set visualization.



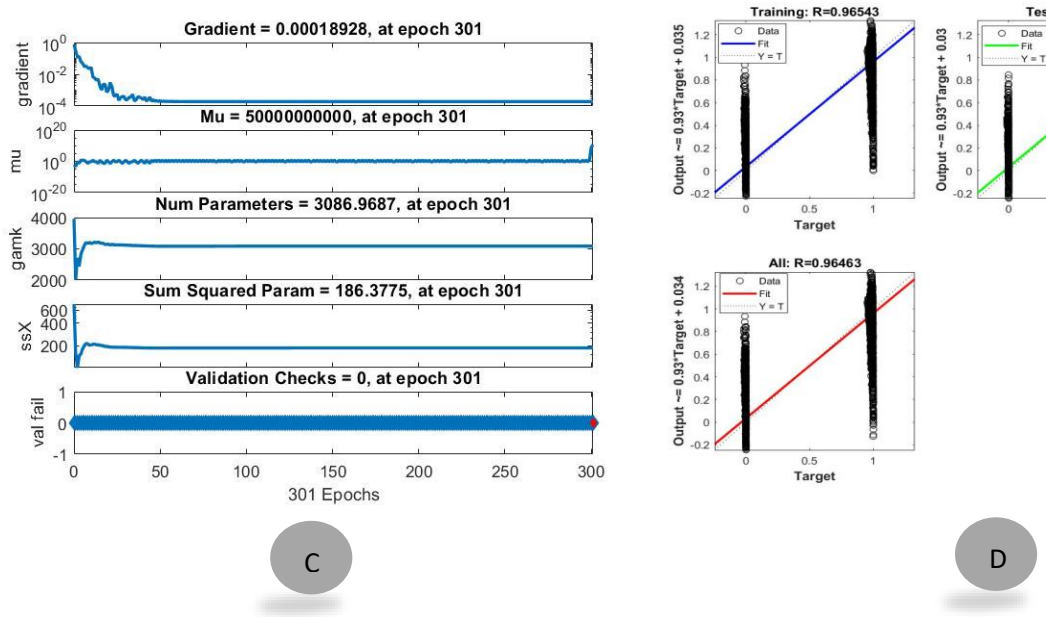



Fig 3(A, B, C, D): Evaluate best training, testing, validation and MSE.

# 5 Conclusion.

This paper is basically focused upon segmentation and classification of lung cancer. The proposed unsharp masking and adaptive canny algorithm with Otsu's thresholding method applied for filtering which can able to detect the interest region. This is successfully implemented, but this system needs a lot of improvement which can be achieved by using KNN for segmentation and Bayesian Regularization Neural network for classification.

# Reference


1. K. Senthil Kumar, K. Venkata Lakshmi, and K. Karthikeyan (2019), Lung Cancer Detection Using Image Segmentation by means of Various Evolutionary Algorithms: Computational and Mathematical Methods in Medicine Volume 2019, Article ID 4909846.

2. A. Asuntha, A. Brindha, S. Indirani1, Andy Srinivasan, Lung cancer detection using SVM algorithm and optimization techniques: ISSN: 0974-2115

3. Babita Rani, Ashok Kumar Goel, Ravneet Kaur (2016), A Modified Approach for Lung Cancer Detection Using Bacterial Forging Optimization Algorithm: ISSN 2278 – 0882

4. Manasee Kurkure, Anuradha Thakare (2016), Introducing automated system for Lung Cancer Detection using Evolutionary Approach ISSN: 2319-7242

5. Neha Panpaliya1, Neha Tadas, Surabhi Bobade, Rewti Aglawe, Akshay Gudadhe (2015), A Survey on Early Detection and Prediction of Lung Cancer: ISSN 2320–088X

6. Fatma Taher, Naoufel Werghi, Hussain Al-Ahmad, Rachid Sammouda (2012), Lung cancer detection by using artificial neural network and fuzzy clustering methods: American Journal of Biomedical Engineering 2012, 2(3): 136-142

7. Yupeng Wang & Jiangyun Li (2017), An Improved Canny Algorithm with Adaptive Threshold Selection.

8. Andrea Polesel, Giovanni Ramponi, and V. John Mathews (2000) Image Enhancement via Adaptive Unsharp Masking.

9. Kayri, Murat. (2016). Predictive Abilities of Bayesian Regularization and Levenberg–Marquardt Algorithms in Artificial Neural Networks: A Comparative Empirical Study on Social Data. Mathematical and Computational Applications. 21. 1-11. 10.3390/mca21020020.