EasyChair Preprint

№ 3076

Prediction of Lungs Cancer Using Machine Learning

Prasanta Das, Biplab Das and Himadri Sekhar Dutta

March 31, 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 possibility or probability of lung cancer. K-Nearest Neighbors is used to segment the cancer from lungs. Finally achieved the classification accuracy near about 99.5%.

KEYWORDS: Lungs Cancer, Segmentation, K-Nearest Neighbors

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 (2019) 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 (2016) 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 (2016) 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 (2015) 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 acurate 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 (2012) 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. This paper presents two segmentation methods, Hopfield Neural Network (HNN) and a Fuzzy C-Mean (FCM) clustering algorithm, for segmenting sputum color images to detect the lung cancer in its early stages. Moreover, it succeeded in determining the best range of thresholding values. To classify the image of N pixels among M classes, the HNN and FCM methods are designed. The HNN succeeded in extracting the nuclei and cytoplasm regions [6].

3 Proposed Model

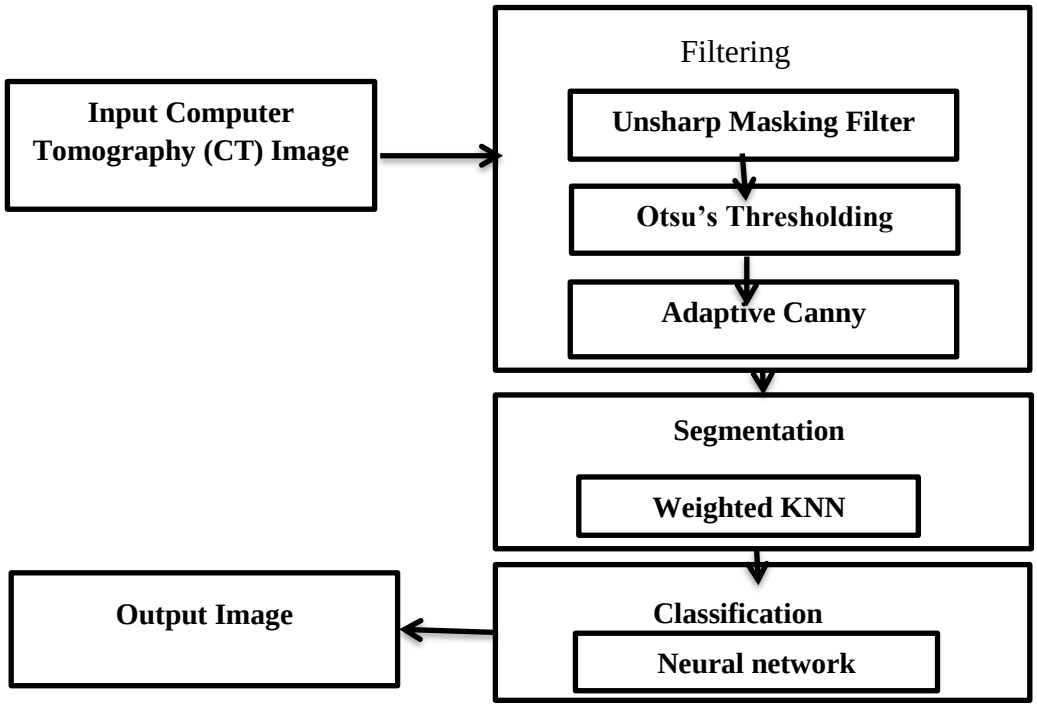


Fig 1: propounded model for lungs cancer detection.

3.1 Unsharp Masking Filter

To make an image sharp "unsharp mask" is used, though the name means opposite. Sharpening used to emphasize texture and detail, and it applied at the time of post processing of digital image. Without creating any additional details, unsharp mask can greatly enhance the appearance of detail by increasing small-scale acutance.
From an input image, Unsharp masking produces an edge image g (x, y)

f (x, y) via g (x, y) = f (x, y) – fsmooth (x, y)
where fsmooth (x, y) is a smoothed version of f (x, y).

3.2 OTSU'S Thresholding

Otsu's thresholding method iterate through all the possible threshold values and calculate a measure of spread for the pixel levels each side of the threshold, i.e. the pixels that either falls in foreground or background.

Algorithm:

1. Compute histogram and probabilities of each intensity level
2. Set up initial $\omega_i(0)$ and $\mu_i(0)$
3. Step through all possible thresholds *t=1*, maximum intensity
    1. Update ω_i and μ_i
    2. Compute $\sigma_b^2(t)$
4. Desired threshold corresponds to the maximum $\sigma_b^2(t)$

3.3Adaptive Canny

The traditional canny algorithm uses fixed spatial scale coefficient of Gauss filter and empirical values of the high and low thresholds, and it has defective in self-adaptability because it is unable to correct parameters according to the actual image. In adaptive canny threshold is calculated using Otsu's global thresholding method.

```
[~,threshOut = edge(InputImage,'Canny');
threshold = threshOut*1.0;
BW = edge(InputImage,'Canny',threshold);
```

Where 'InputImage' is the input image 'BW' is the image after edge detection.

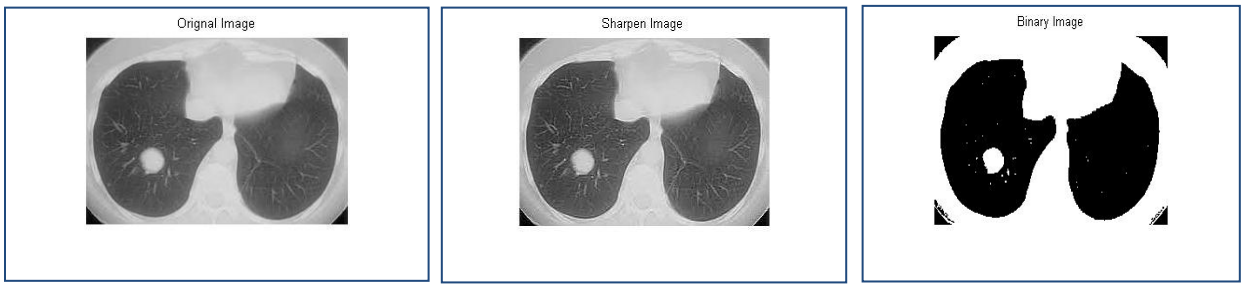
4 Propounded Method for Segmentation

This section we proposed the KNN algorithm which successfully done the segmentation process and achieve the accuracy 99.5%. 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.

5 Proposed Classification process

Here the neural network has become independent, using neural network and KNN algorithm we get the classifications results. Using neural network (NN) we compute the classification process where NN 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 NN compute the specific fitting model which has less loss over the results.

6 Result & Discussion



(a) (b) (c)

(d) (e)

Fig 2: Using proposed method detect lungs cancer

Here all figures show the various stages of processing. 512 X 512-dimensional image of lung taken for the experiment which is in JPG format. 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 canny 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.

7 Conclusion.

This thesis 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 Neural network for classification.

Reference

1. Lung Cancer Detection Using Image Segmentation by means of Various Evolutionary Algorithms (2019): K. Senthil Kumar, K. Venkatalakshmi, and K. Karthikeyan. : Computational and Mathematical Methods in Medicine Volume 2019, Article ID 4909846.
2. Lung cancer detection using SVM algorithm and optimization techniques: A. Asuntha, A.Brindha, S.Indirani1, Andy Srinivasan : ISSN: 0974-2115
3. A Modified Approach for Lung Cancer Detection Using Bacterial Forging Optimization Algorithm (2016) : Babita Rani, Ashok Kumar Goel, Ravneet Kaur : ISSN 2278 – 0882
4. Introducing automated system for Lung Cancer Detection using Evolutionary Approach (2016): Manasee Kurkure, Anuradha Thakare : ISSN: 2319-7242
5. A SURVEY ON EARLY DETECTION AND PREDICTION OF LUNG CANCER (2015) : Neha Panpaliya1, Neha Tadas, Surabhi Bobade, Rewti Aglawe, Akshay Gudadhe : ISSN 2320–088X
6. Lung cancer detection by using artificial neural network and fuzzy clustering methods (2012) : Fatma Taher, Naoufel Werghi, Hussain Al-Ahmad, Rachid Sammouda : American Journal of Biomedical Engineering 2012, 2(3): 136-142
7. An Improved Canny Algorithm with Adaptive Threshold Selection (2017): Yupeng Wang & Jiangyun Li*.
8. Image Enhancement via Adaptive Unsharp Masking (2000): Andrea Polesel, Giovanni Ramponi, and V. John Mathews.
9. ArtificialNeuralNetwork-HopfieldNetworks(Tutorialspoint): https://www.tutorialspoint.com/artificial_neural_network/artificial_neural_network_hopfield.htm

Prasanta Das received his BCA degree from Prabhat Kumar College, Contai under Vidyasagar University, Medinipur. Now he is final year student of MCA (Maulana Abul Kalam Azad University of Technology, Kolkata, West Bengal, India)

Biplab Kanti Das received his M.Tech (IT) degree from Jadavpur University, Kolkata, West Bengal, India, and pursuing Ph.D. in Computer Science & Engineering from Maulana Abul Kalam Azad University of Technology, Kolkata, West Bengal, India. He is presently working as Assistant Professor at MCA Department of Calcutta Institute of Technology, Howrah, West Bengal, India. His research areas include Medical Image Processing, Courseware Engineering. He has published 03 research papers in various international journals and conferences and reviewed papers for reputed Journals. He has also published 4 books for schools and colleges.

Himadri Sekhar Dutta received his B.Tech degree in Electronics and Communication Engineering from Kalyani Government Engineering College, Kalyani, India, M.Tech. degree in Optics and Opto-Electronics from University of Calcutta,Kolkata, India and Ph.D. in Technology from Institute of Radio Physics and Electronics, Kolkata, India respectively. He is presently working as Assistant Professor at ECE Department of Kalyani Government Engineering College, Kalyani. He was the Chairperson of IEEE Young Professional, Kolkata Section for two consecutive years (in 2016 and 2017) and actively participated in different activities conducted by IEEE. His research areas include Medical Image Processing, Embedded Systems and Opto-electronic Devices. He has published more than 70 research papers in various international journals and conferences and reviewed papers for reputed Journals and international conferences.