Paper ID: VESCOMM29

LUNG CANCER DETECTION SYSTEM BY USING BAYESIAN CLASSIFIER

BhagyarekhaU.Dhaware
Electronicsand Telecommunication Department
SKN Sinhgad College of Engg.,KortiPandharpur,India
Email:bhagyarekha.dhaware28@gmail.com

Anjali C. Pise
Electronicsand Telecommunication Department
SKN SinhgadCollege of Engg.,Korti,Pandharpur,India
Email: prof_anjali@rediffmail.com

Abstract :**Medical image enhancement & classification play an important role in medical research area. To analyse the texture Computed Tomography (CT) images of lungs are taken to find the values of various parameters of texture. Mainly CT lung images are classified into normal and abnormal category. Classification of images depends on the features extracted from the images. Proposed system focusing on texture based features such as GLCM (Gray Level Co-occurrence Matrix) feature plays an important role in medical image analysis. Totally 12 different statistical features & 7 shape features will be extracted. To select the required features among them, use sequential forward selection algorithm. Afterwards prefer Bayesian classifier for the classification stage which gives perfect classification..**

Keywords—LCD,CLAHE,GLCM,CDF,SFA, Bayesian Classifier,Texture Feature Extraction,Lung Cancer Detection System,CT images,IP, MATLAB,

1.INTRODUCTION

Lung cancer is considered to be the main cause of cancer death worldwide, and it is difficult to detect in its early stages because symptoms appear only at advanced stages causing the mortality rate to be the highest among all other types of cancer [1]. More people die because of lung cancer than any other types of cancer such as: breast, colon, and prostate cancers. There is significant evidence indicating that the early detection of lung cancer will decrease the mortality rate. The most recent estimates according to the latest statistics provided by world health organization indicates that around 7.6 million deaths worldwide each year because of this type of cancer. Furthermore, mortality from cancer are expected to continue rising, to become around 17 million worldwide in 2030[1].The early detection of lung cancer in its primary stage is a challenging problem, due to the complicated structure of the cancer cells, where most of the cells are overlapped to each other. It is a computational procedure that sort images into groups according to their similarities. use Histogram Equalization for preprocessing of the images and feature extraction process and multivariate multinomial Bayesian classifier to check the state of a patient in its early stage whether it is normal or abnormal. The manual analysis of the sputum samples is a very time consuming, inaccurate and requires well trained person to avoid diagnostic errors. The quantitative procedure is very helpful for earlier detection of lung cancer. Experimental analysis will be made with dataset to evaluate the performance of the different classifiers. The performance is based on the correct and incorrect classification of the classifier.

Preprocess the given test image for reducing noise and to enhance the contrast by using Contrast Limited Adaptive Histogram Equalization (CLAHE). Afterwards, texture features will be extracted from the test image using GLCM. In feature extraction stage, statistical measurements will be calculated from the gray level co-occurrence matrix for different directions and distances. Among the various features extracted select the distinct features that will be utilized for classification purpose. For the selection of features SFS (Sequential Forward Selection) is used. Bayesian classifier is used to classify whether the test image comes under normal or abnormal. The proposed methodology consists of a set of stages startingfrom collecting Lung CT images. The main steps are shown below.

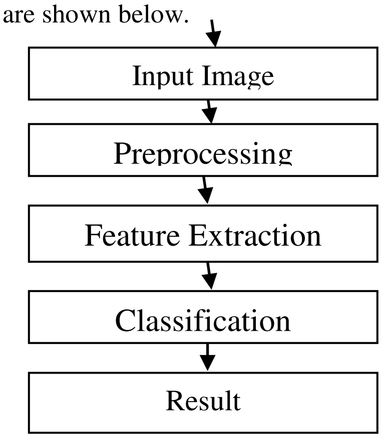


Fig. 1: General Block diagram of LCD System

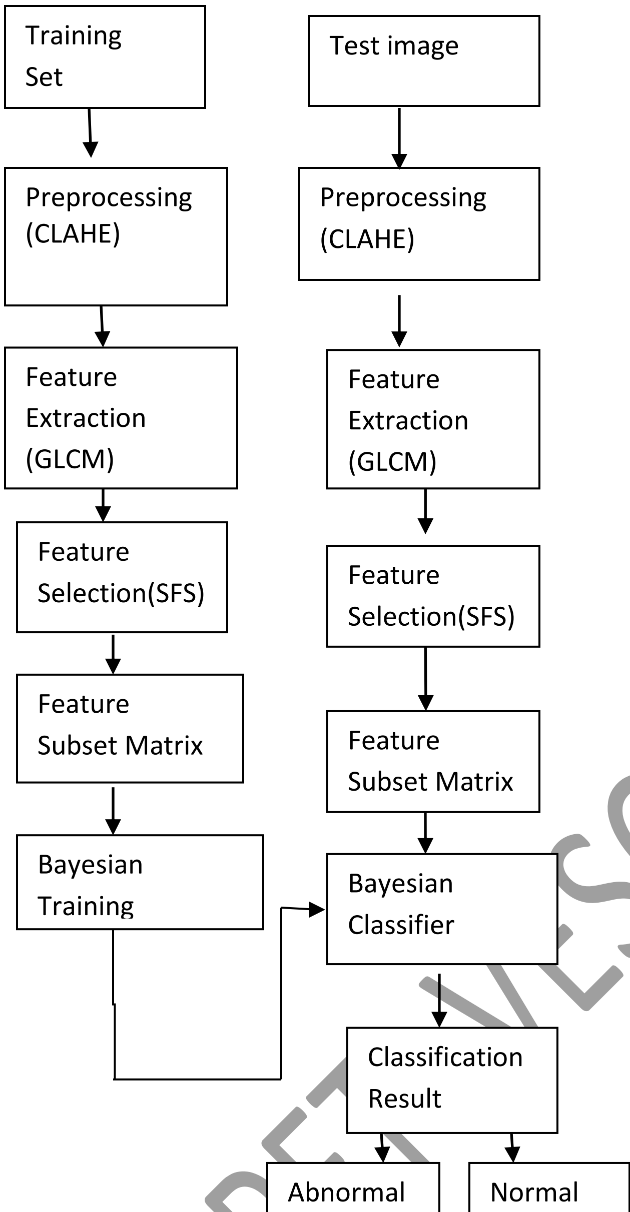


Fig. 2: Main Block diagram of LCD System

The images obtained from CT scan machine were in DICOM format. The commands of MATLAB could not be used on them. Therefore, to work on the images, they have been converted to BMP format, so that they can easily be used in the MATLAB environment.

Flow of the work is shown in above system architecture, one test image is taken, as test image has noise we have to preprocess the given test image for reducing noise and to enhance the contrast. Preprocessing has been done by using The Contrast Limited Adaptive Histogram Equalization (CLAHE) then texture features (GLCM) will be extracted from the test image, By using sequential forward selection feature subset is obtained and then processed towards the Multivariate multinomial Bayesian classifier for classification whether it is normal one or abnormal. Same data sequence flow is used for the training set of images into the left side of the architecture.

Pre-processing of CT image for to Enhance the Contrast

The Contrast Limited Adaptive Histogram Equalization (CLAHE) is an improved version of adaptive histogram equalization. The contrast limited adaptive histogram equalization algorithm partitions the images into contextual regions and applies the histogram equalization to each one. This evens out the distribution of used gray values and thus makes hidden features of the image more visible. The amount of contrast enhancement for some intensity is directly proportional to the slope of the Cumulative Distribution Function (CDF) at that intensity level. Hence contrast enhancement can be limited by limiting the slope of the CDF. The slope of CDF at a bin location is determined by theheight of the histogram for that bin. Therefore the height limitation of the histogram results in limiting the slope of the CDF and hence the amount of contrast enhancement

Feature Extraction

Various techniques can be use to extract features from images. Here will use GLCM (Gray level co-occurrence matrix) for the texture feature extraction form CT scan image.

Extraction of texture feature

Amongthe four approaches (Structural, Statistical, model based and Transform) here will use Statistical approach to texture analysis is the statistical approach does not presume in term of primitive but it draws on the general set of statistical tool. It is the most widely used and more generally applied method because of its high accuracy and less computation time.

A gray level co-occurrence matrix (GLCM) contains information about the positions of pixels having similar gray level values.

From the co-occurrence matrix obtained, we have to extract the 12 different statistical features

1. Contrast 2. Sum of Squares, Variance
3. Correlation 4. Energy 5.Maximum Probability

6. Dissimilarity 7. Autocorrelation 8.Inverse different Moment 9.Entropy 10. Homogeneity 11. Cluster Prominence 12. Cluster Shade

Feature selection(SFA): Feature selection algorithms are important to recognition and classification systems because, if a feature space with a large dimension is used, the performance of the classifier will decrease with respect to execution time and to recognition rate. The execution time increases with the number of features because of the measurement cost. The recognition rate can decrease because of redundant features and of the fact that small number of features can alleviate the course of dimensionality when the training samples set is limited, leading to overtraining. On the other hand, a reduction in the number of features may lead to a loss in the discrimination power and thereby lower the accuracy of the recognition system.
In order to determine the bestfeature subset for some criterion, some automatic feature selection algorithm can be applied to the complete feature space, varying the number of selected features from 1 to m. by Sequential Forward Selection algorithm be select best features.

Feature Subset Matrix: It is a empty Matrix to store Features from SFS. SFS performs best when the optimal subset has a small number of features. ¬When the search is near the empty Set, a large number of states can be potentially evaluated. Towards the full set, the state space for 4 features is shown. The main disadvantage of SFS is that it is unable to remove features that become obsolete after the addition of otherfeatures.

. **Bayesian Classifier:**
Bayesian to classify the input CT lung image into normal and abnormal conditions.This technique will help to get more accurate result. Any kind of abnormalities can be classified, based on a probabilistic model specification Features that describe data instances are conditionally independent given the classification hypothesis. Multivariate multinomial distribution for discrete data that fit assumes each individual feature follows a multinomial model within a class. The parameters for a feature include the probabilities of all possible values that the corresponding feature can take.
Bays Rule is stated as follows,
P (h/d) = P (d/h) P (h) / P (d)
Understanding Baye's rule
d= data
h= hypothesis (model)
-rearranging
P (h/d) P (d) = P (d/h) P (h)
P (d, h) = P (d, h)
The same joint probability on both sides.

RESULT

1.Take Data base:

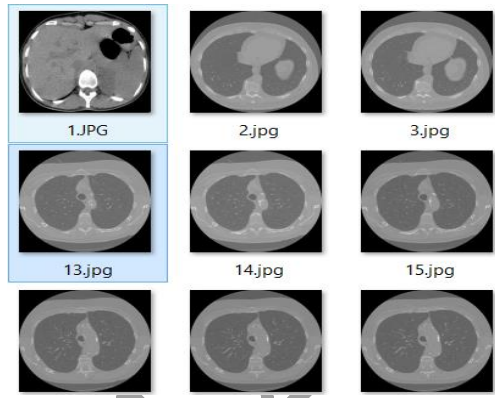


Fig.3

2. Take input i/p Test Image

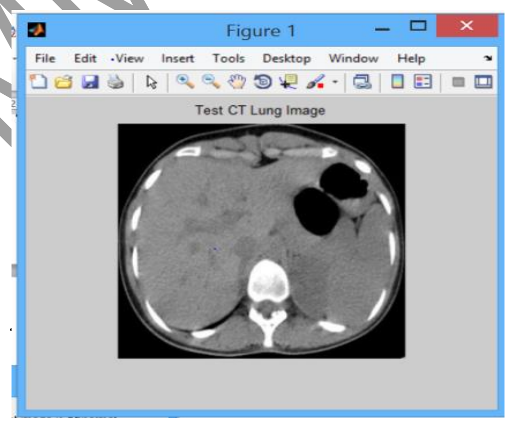


Fig:4

3. Process the i/p image to get filtered o/p

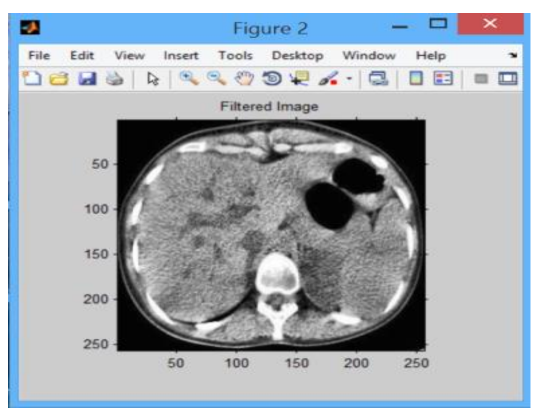


Fig:5

4. Apply CLAHE before & after filtering

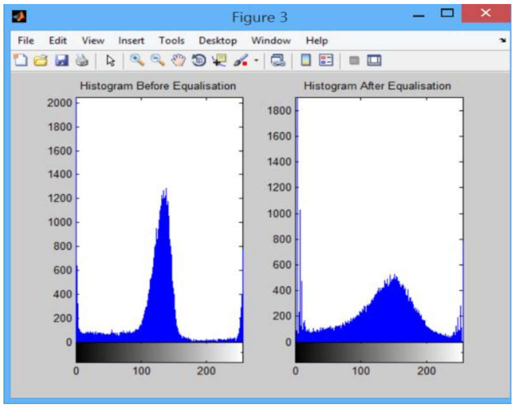
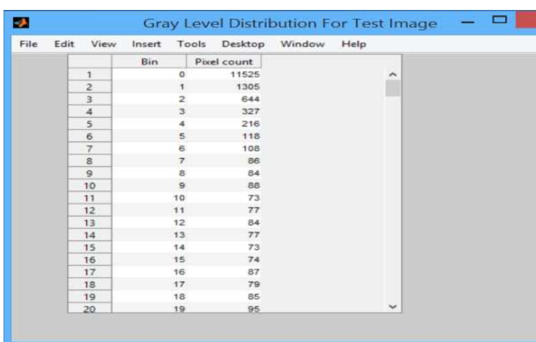


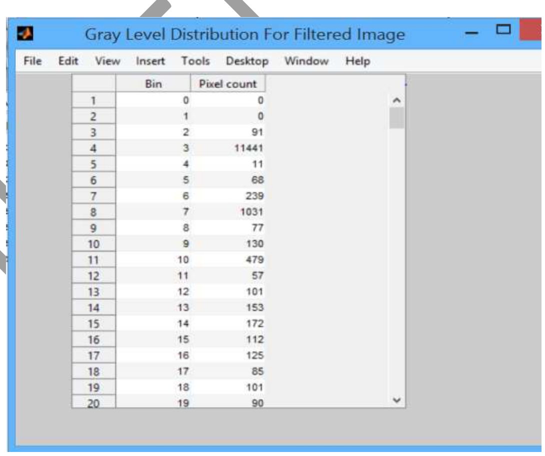
Fig:6

5. Apply GLCM on Test image & filtered image to extract 12 texture features & 7 Shape features



	Bin	Pixel count
1	0	11525
2	1	1305
3	2	644
4	3	327
5	4	216
6	5	118
7	6	108
8	7	86
9	8	84
10	9	88
11	10	73
12	11	77
13	12	84
14	13	77
15	14	73
16	15	74
17	16	87
18	17	79
19	18	85
20	19	95

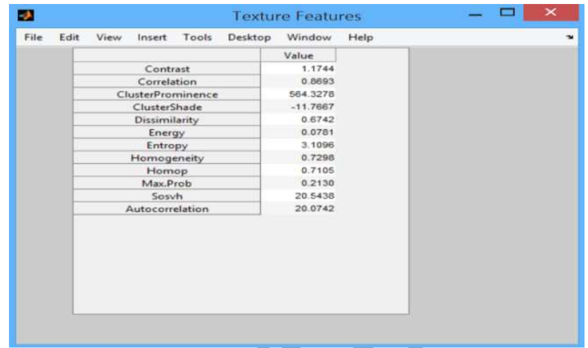
Fig:7



	Bin	Pixel count
1	0	0
2	1	0
3	2	91
4	3	11441
5	4	11
6	5	68
7	6	239
8	7	1031
9	8	77
10	9	130
11	10	479
12	11	57
13	12	101
14	13	153
15	14	172
16	15	112
17	16	125
18	17	85
19	18	101
20	19	90

Fig:8

.



	Value
Contrast	1.1744
Correlation	0.8693
ClusterProminence	564.3278
ClusterShade	-11.7667
Dissimilarity	0.6742
Energy	0.0781
Entropy	3.1096
Homogeneity	0.7298
Homop	0.7105
Max.Prob	0.2130
Sosvh	20.5438
Autocorrelation	20.0742

Fig:9

6. Apply Bayesian Classifier to classify the image is Normal or Abnnormal

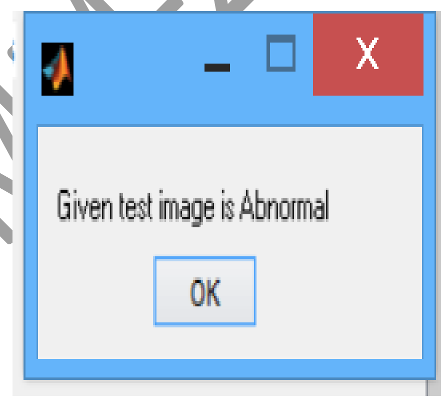


Fig:10

CONCLUSION

The work in this research involves using Bayesian to classify the input which is CT lung image into normal and abnormal conditions. This kennel technique will help to get more accurate result. Thus to achieve high accuracy among the 12 different statistical features contrast, Correlation, Variance, Inverse different Moment, Cluster Prominence and Cluster Shade these six are the most efficient features to get accurate result we have to extract all 12 features

REFERENCES

[1]FatmaTaher, RachidSammouda, "Lung Cancer Detection by using Artifitial Neural Network and Fuzzy Clustering methods",*IEEE GCC Conference and Exibition,*2011, Pages: 295-298.

[2] Chung-Ming W.,Yung-Chang Chen, Kai-Sheng Hsieh, "Texture features for classification of ultrasonic liver images" *Medical Imaging, IEEE*

Transactions on Vol No. 11, Issue 2, 1992, Page(s): 141 - 152

[3] Anita Chaudary, SonitSukharaj Singh, C., "Lung Cancer Detection on CT Images by using Image Processing", *International Conference on Computing Sciences2012*, Page(s): 142 – 146.

[4]Sheppard, M.A.; Liwen Shih; "P6C-2 Image Texture Clustering for Prostate Ultrasound Diagnosis", Ultrasonics Symposium, 2007.*IEEE* 28-31 June 2007, Page(s): 2473 - 2476

[5]FatmaTaher,NaoufelWerghi,Hussain Al-Ahmad and RachidSammouda "Lung Cancer Detection by Using Artificial Neural Network and Fuzzy Clustering Methods", *American Journal of Biomedical Engineering*, pp 136-142, 2012.

[6]ShyamLal and Mahesh Chandra, "Efficient Algorithm for Contrast Enhancement of Natural Images" *The International Arab Journal of Information Technology*, Vol. 11, No. 1, January 2014

[7] M.Gomathi ,Dr. P. Thangaraj, "A Computer Aided Diagnosis System for Detection of Lung Cancer Nodules Using Extreme Learning Machine" , *International Journal of Engineering Science And Technology* Vol.2(10),2010.

[8] FatmaTaher ,RachidSammouda, "Identification of Lung Cancer Based On Shape and Color" Innovations in Information Technology, 2007. IIT '07.*4th International Conference on Medical Engineering, IEEE* 18-20 Nov. 2008.Pages 481 – 485.

[9] JIA Tong, ZHAO Da-Zhe, WEI Ying, ZHU Xin-Hua, WANG Xu, "Computer-Aided Lung Nodule Detection Based On CT Images", *IEEE/ICME International Conference on Complex Medical Engineering*, 2007..

[10] S.A.Patil, Dr. V.R. Udupi ,Dr. C.D. Kane , Dr.A.I.Wasif, Dr.J.V.Desai,A.N.Jadhav, " Geometrical and Texture Feature Estimation of Lung Cancer And TB Image Using Chest X-ray Database", *Biomedical and Pharmaceutical Engineering, 2009. ICBPE '09.IEEE* 2-4 Dec. 2009. Page(s) 1-7.

[11]Tang J., Peli E., and Acton S., "ImageEnhancement using a Contrast Measure in theCompressed Domain," *IEEE Signal ProcessingLetter*, vol. 10, no. 10, pp. 289-292, 2003.

[12]Wadud M., Kabir M., Dewan M., and Chae O.,"A Dynamic Histogram Equalization for ImageContrast Enhancement," *IEEE Transactions onConsumer Electronics*, vol. 53, no. 2, pp. 593-600, 2007.

[13]Wang C. and Ye Z., "Brightness **7.0** PreservingHistogram Equalization with Maximum Entropy:AVariational Perspective," *IEEE Transactionson Consumer Electronics*, vol. 51, no. 4, pp.1326-1334, 2005.

[14]Yeganeh H., Ziaei A., and Rezaie A., "A Novel Approach for Contrast Enhancement Based on Histogram Equalization," *in Proceedings of IEEE International Conference on Computer and Communication Engineering*, Kuala Lumpur, pp. 256-260, 2008.

[15]Yun S., Kim J., and Kim S., "Image Enhancement using a Fusion Framework of Histogram Equalization and Laplacian Pyramid," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2763-2771, 2010.