Bayesian Classifier and FCM Segmentation for Lung Cancer Detection in early Stage

Bhagyarekha U. Dhaware[1], Anjali C. Pise[2]
[1,2,]Electronics and Telecommunication Department
SKN Sinhgad College of Engineering, Korti, Pandharpur, MS, India

Abstract— Image enhancement and classification is a big task, especially while performing in medical field. Enhancing and image classification is used for analysis of texture computed tomography (CT). In this paper images of lungs were taken for find various parameters of the texture. Mainly CT images of lungs can be categorized into normal and abnormal. Classification is based on the features extracted from the taken image. Implementation of the system focuses on texture based features e.g. GLCM (gray level co-occurrence matrix) feature plays a vital role in medical field. Selection is based on the twelve various statistical features and seven shape for extraction by applying sequential forward selection algorithm. After application of sequential forward selection algorithm Bayesian classifier was applied among classified data to get best classification

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

I. INTRODUCTION

Lung cancer is considered to be the major cause of cancer death worldwide, symptoms appear only at advanced stages causing the mortality rate to be the largest in all other types of cancer [1], due to this detection is difficult in it's early stages. Most of the people death is due to the lung cancer than any other types of cancer such as: breast, colon, and prostate cancers. There is prominent proof signifying that the early detection of lung cancer will decrease the mortality rate. The recent estimates provided by world health organization shows that around 7.6 million deaths worldwide per year. Moreover, mortality due to cancer are supposed to continue rising, to become around 17 million worldwide in 2030[1]. The early detection of lung cancer in the primary stage is a challenging problem, because of complicated structure of the cancer cells, where most of the cells are overlapped to each other. It is a computational procedure that classifies the images into groups based on similarities. use Histogram Equalization for preprocessing of the images, feature extraction process and Bayesian classifier confirms the health of a patient in beginning stage to identify it's 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 assess the performance of the variousclassifiers. The performance is based on effective classification by the classifier.

Pre-processing the given test image for reducing noise and for enhancing the contrast by using Contrast Limited Adaptive Histogram Equalization (CLAHE). Then, 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 various directions and distances. In various features extracted select the distinct features that will be applied for classification purpose. For selecting the features SFS (Sequential Forward Selection) is used. Bayesian classifier is used to classify whether the test image comes under normal or abnormal. The proposed technique consists of a set of stages starting from collecting Lung CT images. The main steps are shown fig No.1.

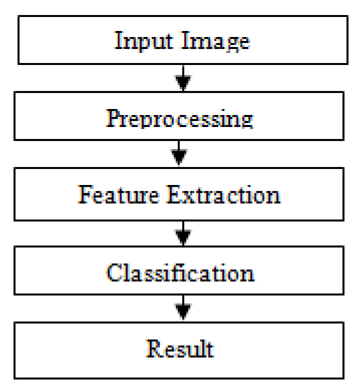


Fig.1. General Block diagram of LCD System

The images we get from CT scan are in DICOM format, therefore matlab commands cannot be applied directly. For making images MATLAB friendly, all images have been converted to BMP format. Work flow is shown in fig.no.2. Test image is taken for analysis as it may possess noise, it is diverted to preprocessed for reduction in noise and contrast enhancement.

Preprocessing is completed by applying "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.

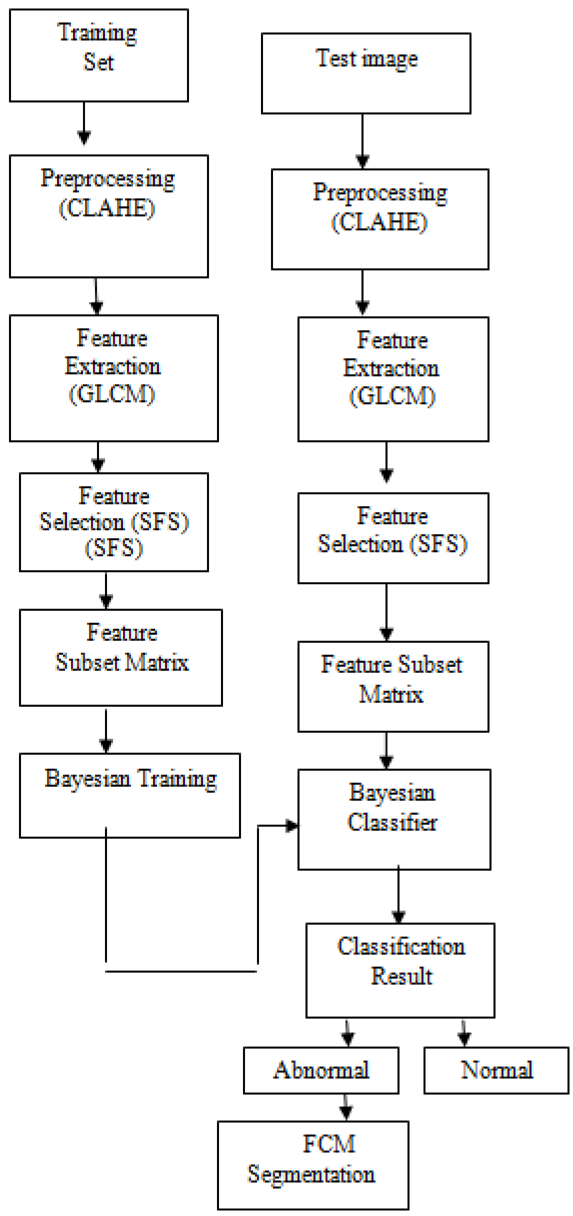


Fig.2. Main Block diagram of LCD System

II. Pre-processing of CT image for to Enhance the Contrast

The CLHAE is modified version of adaptive histogram equalization. The CLHAE algorithm separate images into contextual regions and used the histogram equalization to every image. This helps in making hidden features of the image more visible. The contrast enhancement depends on the cumulative distribution function (CDF) at particular intensity level. Hence contrast enhancement will be limited by controlling slope of the CDF. The slope of CDF is decided on the basis of height of the histogram for that bin. Hence, it results in height limitation of the histogram results in limiting the slope of CDF it results in enhancement in contrast.

III. Feature Extraction

Various techniques can be used to extract features from images. Here will use GLCM (Gray level co-occurrence matrix) for the texture feature extraction form CT scan image. Feature extraction is a dimensionality reduction. Transforming input data into set of features.

A. Extraction of texture feature

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

A Gray Level Co-Occurrence Matrix (GLCM) has information regarding the position of pixels possessing similar gray level values.

Table.1. Four state space features.

GLCM

1	1	5	6	8
2	3	5	7	1
4	5	7	1	2
8	5	1	2	5

	1	2	3	4	5	6	7	8
1	1	2	0	0	1	0	0	0
2	0	0	1	0	1	0	0	0
3	0	0	0	0	1	0	0	0
4	0	0	0	0	1	0	0	0
5	1	0	0	0	0	1	2	0
6	0	0	0	0	0	0	0	1
7	2	0	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0

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

B. Feature selection (SFA):

Feature selection algorithms helps in recognising and classifying systems. If a feature space with large dimension is used, there are chances of decrease in classifier performance in connection to execution time and recognition rate.

For determining the best feature subset in same case, automatic feature selection technique can be used for completion of feature space, by varying the number of selected features from 1 to m. By using sequential forward selection algorithm, best features can be easily extracted.

After selecting the final feature subset will be sent to Bayesian classifier for the classification.

Algorithm: Feature selection
1. Start with the empty set $Y_0 = \{\Phi\}$
2. Select the next best feature

$$X^+ = =\text{argmax}\left[J\left(Y_k + X\right)\right];\ x¢Y_k$$

3. Update $Y_k + 1 = Y_k + X^+; k = k + 1$
4. Go to step 2

C. Feature Subset Matrix

It is a empty matrix for storing features from SFS. Effectiveness of the SFS will be best when optimal subset has a small number of features. Potential evaluation of the large number of states can be possible when search is near to empty set. Four features of state space are shown in table no.1. Disadvantage of SFS is, it may be unable to remove feature which may get obsolete after addition of other features.

IV. IMAGE CLASSIFICATION

Bayesian Classifier is used 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 provisionally independent given the classification hypothesis. Multivariate multinomial distribution for discrete data that fit assumes each individual feature follows a multinomial model within a class.

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.

V. SEGMENTATION

Once classification of CT scan image is done, then project the affected portion for segmentation by applying FCM technique. FCM is technique applied when clustering is needed, and this technique allows one piece of data to belong to more than one cluster. Assigned membership values are between 0 to 1, and membership values for each data from all the clusters will add to 1. The improved FCM algorithm uses a concept of data compression so the dimensionality of the input is highly reduced.

VI. RESULT AND DISCUSSION

The Lung Data base is collected from PRISM Diagnostics centre, Solapur, Maharashtra. All Images are taken from CT scan machine. For Analysing result the first image is a input Test CT Scan image in fig.4 then apply pre processing for reduction in noise and for enhancement in contrast because of the overlapping of cells then the filtered image is shown in fig.5, After that we plot Histogram of before and after equalization of test image in fig.6after that gray level distribution of test image and filtered image, after that 12 texture features are extracted and then shape features are calculated and then by applying Bayesian classifier for classification of Test image is shown in fig.7. If the result of classification is abnormal then we apply segmentation by this we can obtain highlighted piece of tumour from background.

The parameters which define the hypothesis were "learned" by using the training set. After that feature set has been calculated for each pixel, it will be used by a classifier to decide whether each pixel represents a tumour pixel or a normal pixel. The classification stage has two components, a training phase and a testing phase. At the training stage, pixel features and their corresponding manual labels shows the input and the output is a model that uses the features for predicting the corresponding label. In this training phase carried out only once, since the model can then be used to identify new data. The input for the testing phase is a learned model and pixel features without equivalent classes, and the output of the testing phase is the predicted classes for the pixels depending on their features

1. **Take Data base:** Data base which is taken of both cancer effected and non-cancer effected patients and selected some images.

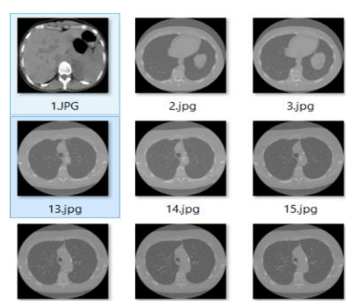


Fig.3. Data Base

2. **Input Image:** Test Image as the first patients image

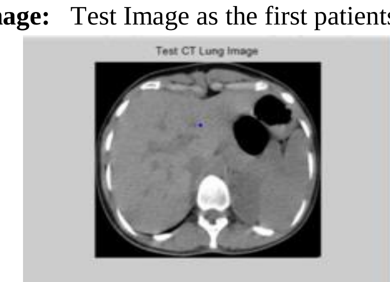


Fig.4. Input Test Image

3. **Preprocess the input image to get filtered output**: In this process noise is removed

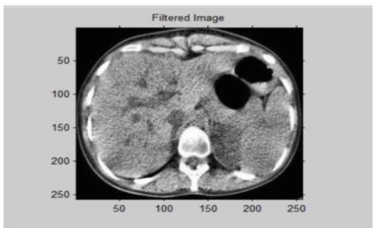


Fig.5. Filtered Output Image

4. **Apply CLAHE before & after filtering**:

The Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied. The contrast limited adaptive histogram equalization algorithm separates the images into contextual regions and applies the histogram equalization to all. This evens out the allocation of applied gray values and thus makes hidden features of the image more visible.

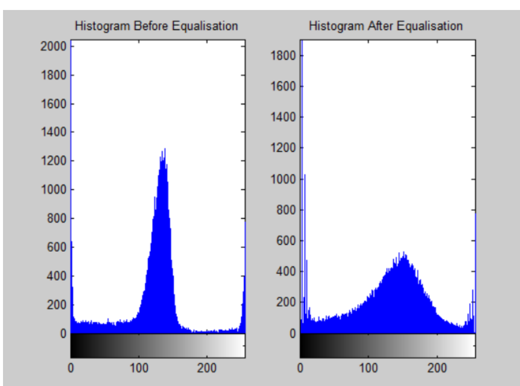


Fig.6. CLAHE before & after filtering

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

- Obtain Gray level distribution for test image
- Obtain Gray level distribution for Filtered image
- Obtain Shape Features it give it gives effective shapeof the tumour or cancerous part of Lung

6. **Classification:** Apply Bayesian Classifier to classify the image is Normalor Abnormal Bayesian Classifier gives accurate result

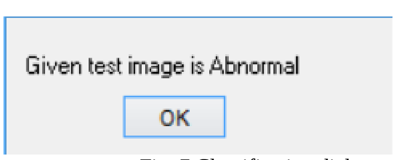


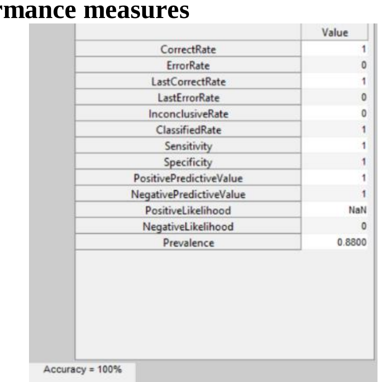
Fig: 7 Classification dialogue box

7. **Segmentation:** After classification if the Image is Abnormal then apply FCM Segmentation to get actual cancerous part of Tumour and eliminates background part

Fig.8. Segmented Image

8. **Detail Analysis:** Into the whole analysis of this results performance measures are calculated in terms of correct rate, error rate, Inconclusive rate, Sensitivity, Specificity, positive predictive value, positive likelihood, negative likelihood, and prevalence as shown in Fig.9 confusion matrix is drawn in Fig.10and classifier performance is analyzed further in table no. 2 tabulation of Individual feature analysis is given

Performance measures



	Value
CorrectRate	1
ErrorRate	0
LastCorrectRate	1
LastErrorRate	0
InconclusiveRate	0
ClassifiedRate	1
Sensitivity	1
Specificity	1
PositivePredictiveValue	1
NegativePredictiveValue	1
PositiveLikelihood	NaN
NegativeLikelihood	0
Prevalence	0.8800

Accuracy = 100%

Fig.9 Performance measures

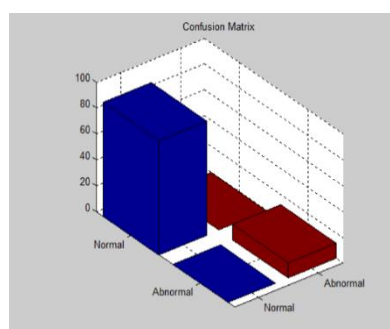


Fig.10 Confusion Matrix

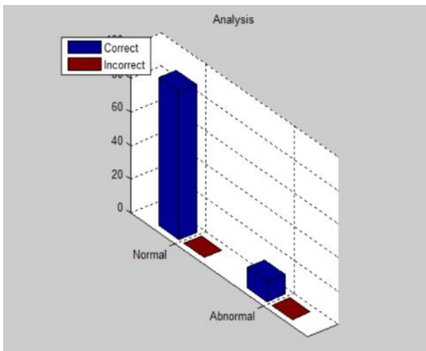


Fig.11 Analysis of Classifier

Table.2.Tabulation of individual feature analysis

Feature no	Feature Name	% Accuracy
1	Contarst	70
2	Correlation	70
3	Autocorrelation	0
4	Energy	30
5	Entropy	30
6	Inverse different Moment	70
7	Cluster Prominence	70
8	Cluster Shade	70
9	Maximum Probabilty	20
10	Dissimalrity	20
11	Homogeneity	20
12	Variance	70
	All 1 to 12	100
	1,2,6,7,8,12	92

VII. CONCLUSION

While implementing this research work, Bayesian algorithm is used for classifying the input of CT lung images for deciding normal or abnormal. Implemented method found useful and it gives more accurate result for the methods better than mentioned in literature. Achieving high accuracy is been made possible by using 12 unlike statistical features contrast, correlation and variance. Inverse different moments and cluster prominence and cluster shade with six different and most efficient features are used to improve accuracy and for extraction of all 12 features.

REFERENCES

[1] Fatma Taher, Rachid Sammouda, "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, Sonit Sukharaj 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, Naoufel Werghi, Hussain Al-Ahmad and Rachidb Sammouda "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 NaturalImages" 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] FatmabTaher and Rachid Sammouda "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 Preserving Histogram 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.

[16] Bhagyarekha Dhaware and Anjali C. Pise. "Lung Cancer Detection System by Using Bayesian Classifier" proceedings of VESCOMM 2016.

[17] Bhagyarekha Dhaware and Anjali C. Pise. "Lung Cancer Detection by Using Bayesian Classifier and FCM Segmentation" proceedings of ICACDOT 2016(*I²IT), Pune*