Lung Cancer Detection &Classification Using MVMN Bayesian Classifier & Modified FCM Algorithm

Prof.Tejashri Patil[1], Ms. Kalyani Zagade[2], Mr. Nitesh Waghmare[3]

[1,2,3]E&TC, PGMCOE, SPPU, (India)

ABSTRACT

Image Processing is a technique to enhance raw images received from cameras/sensors placed on satellites, space probes and aircrafts or pictures taken in normal day-to-day life for various applications. Various techniques have been developed in Image processing during the last four to five decades. Most of the techniques are developed for enhancing images obtained from unmanned space crafts, space probes and military reconnaissance flights. Image Processing systems are becoming popular due to easy availability of powerful personnel computers, large size memory devices, graphics software etc. Medical image segmentation & classification play an important role in medical research field. The patient CT lung images are classified into normal and abnormal category. Then, the abnormal images are subjected to segmentation to view the tumor portion. Classification depends on the features extracted from the images. We mainly are concentrating on feature extraction stage to yield better classification performance. Texture based features such as GLCM (Gray Level Co-occurrence Matrix) features play an important role in medical image analysis. Totally 12 different statistical features were extracted. To select the discriminative features among them we use sequential forward selection algorithm. Afterwards we prefer multinomial multivariate Bayesian for the classification stage. Classifier performance will be analyses further. The effectiveness of the modified weighted FCM algorithm in terms of computational rate is improved by modifying the cluster center and membership value updating criterion.

Keywords: feature extraction, fuzzy c mean algorithm, histogram equalization, image segmentation, neural network classifier

I. INTRODUCTION

Lung cancer is considered to be the main cause of cancer death worldwide, and as we know that it is difficult to detect lung cancer in its early stages because symptoms appear only at advanced stages causing the mortality rate to be the highest among all other types of cancers. 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. 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.

In recent years, doctors use wide range of diagnostic procedures to detect the cancer. Chest X-ray, Computer tomography (CT) are the standard for pulmonary imaging. There are many techniques to diagnosis lung cancer, such as Chest Radiograph (x-ray), Computed Tomography (CT), Magnetic Resonance Imaging (MRI scan) and Sputum Cytology. However, most of these techniques are in current clinical practice, hundreds of thin-sectional CT images are generated for each patient and are evaluated by radiologist by looking at each image in axial mode. It is very difficult to interpret and very time consuming which cause missing a cancer. However, the extreme variation in the grey level and the relative contrast among the images make the segmentation results less accurate, for this reason we attempt to use automatic diagnostic system for detecting lung cancer in its early stages based on the analysis of the sputum colour images. In this work we used Histogram Equalization for pre-processing of the images and multivariate multinomial Bayesian classifier to check the state of a patient in its early stage whether it is normal or abnormal. Experimental analysis is made with database to evaluate the performance of the different classifiers. The performance is based on the classification of the classifier.

In order to formulate a rule, we have developed a modified FCM technique for unsupervised segmentation of the sputum colour image to divide the images into several meaningful sub regions. Image segmentation has been used as the first step in image classification and clustering. In this work, we used Contrast Limited Adaptive Histogram Equalization for pre-processing of the images and multivariate multinomial Bayesian classifier to check the CT scan image in its early stage whether it is normal or abnormal. There are a few fundamental problems that are always relevant in the field and form the basis for the development of various applications. They are problems like structure recovery from images, image enhancement, object recognition, classification and tracking, rendering of CT scan images.

II. SYSTEM ARCHITECTURE

System architecture is the conceptual model that defines the structure, behavior, and more views of a system. System architecture is as given in below figure 3.1.

An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures of the system.

Flow of the work is given into the right side of the 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

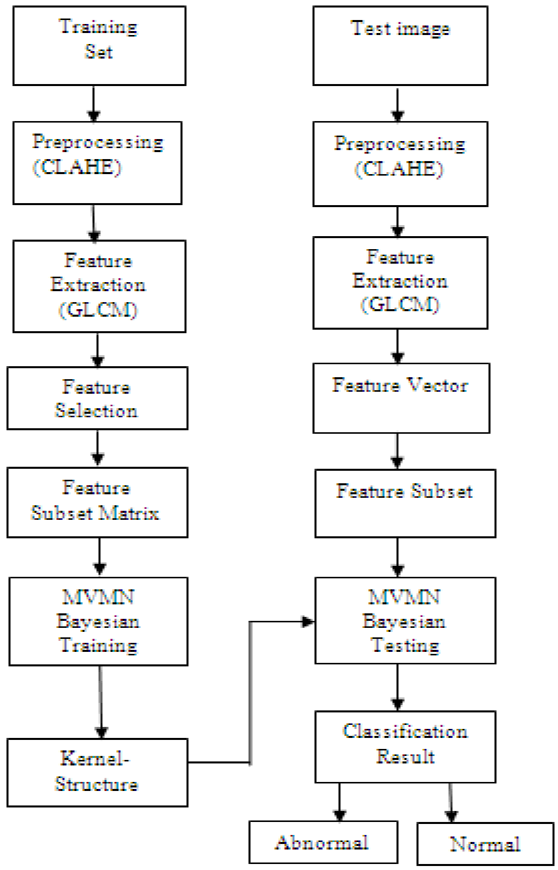

Fig 1. Block Diagram

III. WORKING

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 the height 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.Clipping the histogram itself is not quite straight forward because the excess after clipping has to be redistributed among the other bins, which might

IJARSE
ISSN: 2319-8354

increase the level of the clipped histogram. Hence the clipping should be performed at a level lower than the specified clip level so that after redistribution the maximum histogram level is equal to the clip level

A gray level co-occurrence matrix (GLCM) contains information about the positions of pixels having similar gray level values. GLCM is also called as Gray Level Dependency Matrix. It is defined as “A two-dimensional histogram of gray levels for a pair of pixels, which are separated by a fixed spatial relationship

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.SFS (Sequential Forward Selection) 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 region examined by SFS is narrower since most of the features have already been selected. The search space is drawn like an ellipse to emphasize the fact that there are fewer states towards the full or empty sets.
The main disadvantage of SFS is that it is unable to remove features that become obsolete after the addition of other features.

IV.RESULT

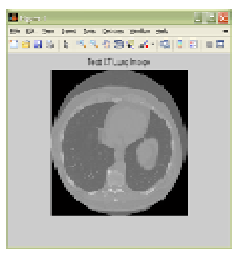

Fig.2Test CT Lung Image

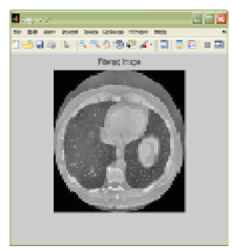

Fig.3 Filtered Image

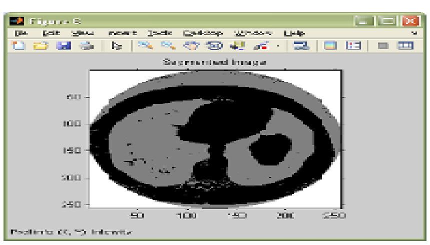

Fig.4 Segmented Image

Fig.5 Classification dialogue box

V. CONCLUSION

The report aims at using kernelised Bayesian to classify the input into normal and abnormal condition. We intend to prove that this kennel technique will help to get more accurate result. Thus we have achieved 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 we can get 90% accurate classification with the help only these six features but to get zero error and more accurate result we have to extract all 12 features.

From the previous research work it is found that the HNN segmentation results are more accurate and reliable than FCM clustering in all cases. The HNN succeeded in detecting and segmenting the nuclei and cytoplasm regions. However FCM failed in detecting the nuclei, instead it detected only part of it. In addition to that, the FCM is not sensitive to intensity variations as the segmentation error at convergence is larger with FCM compared to that with HNN. Further we have to use modified FCM algorithm.

REFERENCES

[1] Robert M. Haralick, k. Shanmugam, Dinstein, "Texture Feature for Image Classification",IEEE Transactions on Systems, Vol. no.3,Issue 6, 1973, Pages: 610-621.

[2] Felipe J.C., Traina A.J.M., Traina, C., "Retrieval by content of medical images using texture for tissue identification", Computer-Based Medical Systems, Proceedings 16th IEEE Symposium USA, 26-27 June 2003, Page(s): 175 – 180.

[3] Felipe J.C., Traina A.J.M., Traina, C., "Retrieval by content of medical images using texture for tissue identification", Computer-Based Medical Systems, Proceedings 16th IEEE Symposium USA, 26-27 June 2003, Page(s): 175 – 180.

[4] S.heppard, 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] Fatma Taher, Naoufel Werghi, Hussain Al-Ahmad and Rachid Sammouda "Lung Cancer Detection by Using Artificial Neural Network and Fuzzy Clustering Methods", American Journal of Biomedical Engineering, pp 136-142, 2012.