Volume 3, Issue 3, March 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Feature Extraction and Principal Component Analysis for Lung Cancer Detection in CT scan Images

Ada
Student of masters of technology Computer Science
Department of Computer Science and Engineering
Sri Guru Granth Sahib World University
Fatehgarh Sahib, Punjab, India.

Rajneet Kaur
Assistant Professor
Department of Computer Science and Engineering
Sri Guru Granth Sahib World University
Fatehgarh Sahib, Punjab, India.

Abstract- A hybrid technique based on feature extraction and Principal Component Analysis (PCA) is presented for lung detection in CT scan images. Lung cancer, if detected successfully at early stages, enables many treatment options, reduced risk of invasive surgery and increased survival rate. .In this paper features are extracted using principal component analysis and Histogram Equalization is used for preprocessing of the images. The system produces promising results for lung cancer detection.

Keywords- Lung Cancer Detection, CT-scan Images, Principle Component Analysis, Histogram Equalization.

I. INTRODUCTION

Lung cancer is the growth of a tumor, referred to as a nodule that arises from cells lining the airways of the respiratory system. These cells are often in bright contrast in chest X-rays and take the shape of a round object. However, these nodules that can be seen in a chest X-ray may not necessarily be a lung cancer; it can be due to some other disease such as pneumonia, tuberculosis or calcified granuloma. As such, the detection of lung cancer has been a tedious task in medical image analysis over the past few decades. If lung nodules can be identified accurately at an early stage, the survival rate of the patients can be increased by a significant percentage. In the health industry, chest X-rays are considered to be the most widely used technique for the detection of lung cancer. However, because it is difficult to identify lung nodules using raw chest X-ray images, analysis of such medical images has become a tedious and complicated task. This paper presents a novel technique that can be used to detect lung cancer in early stages [1]. The rank order of cancers for both males and females among Jordanians in 2008 indicated that there were 356 cases of lung cancer accounting for (7.7 %) of all newly diagnosed cancer cases in 2008. Lung cancer affected 297 (13.1 %) males and 59 (2.5%) females with a male to female ratio of 5:1 which Lung cancer ranked second among males and 10th among females [2]. Figure 1 shows a general description of lung cancer detection system that contains four basic stages. The first stage starts with taking a collection of CT images (normal and abnormal) from the available Database The second stage applies several techniques of image enhancement, to get best level of quality and clearness. The third stage obtains the general features from enhanced image which gives indicators of normality or abnormality of images.

Our system has been fully implemented (in matlab) and tested with real CT scan images. The objective is to support efficient image data processing and feature extraction. Obviously, to deal with real image data, the image processing tool must possess important characteristics such as being noise-tolerant, efficient, practical, and convenient to use. The aim of this research was to detect features for accurate images.

II. METHODS

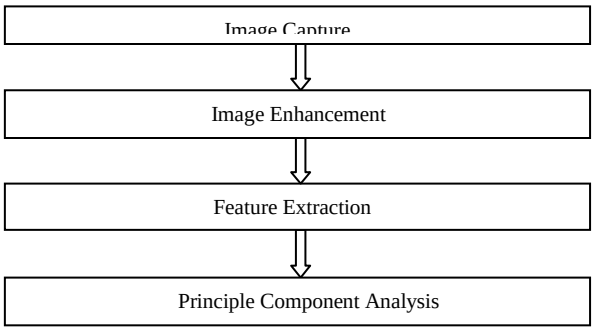

Fig 1:Stages of Lung Cancer Detection

In this research, to obtain more accurate results we divided our work into the following three stages:

1. Image Enhancement stage: to make the image better and enhance it from noising, corruption or interference. For this we used the Histogram Equalization.
2. Features Extraction stage: to obtain the general features of the enhanced segmented image using Binarization and Masking Approach.

III. RESULTS AND DISCUSSIONS

A. Image Enhancement

The image Pre-processing stage starts with image enhancement; the aim of image enhancement is to improve the interpretability or perception of information included in them image for human viewers, or to provide better input for other automated image processing techniques.

Image enhancement techniques can be divided into two broad categories: Spatial domain methods and frequency domain methods. Unfortunately, there is no general theory for determining what "good" image enhancement is when it comes to human perception. If it looks good, it is good. However, when image enhancement techniques are used as pre-processing tools for other image processing techniques, the quantitative measures can determine which techniques are most appropriate.

Lung Cancer Detection Using Image Processing Techniques processing tools for other image processing techniques, the quantitative measures can determine which techniques are most appropriate [5]. In the image enhancement stage we used the Histogram Equalization.

B. Feature Extraction

Image features Extraction stage is an important stage that uses algorithms and techniques to detect and isolate various desired portions or shapes (features) of a given image. To predict the probability of lung cancer presence, binarization approach is used.

i. Binarization Approach

Binarization approach has been applied for detection of cancer. In this we extract the number of white pixels and check them against some threshold to check the normal and abnormal lungs.

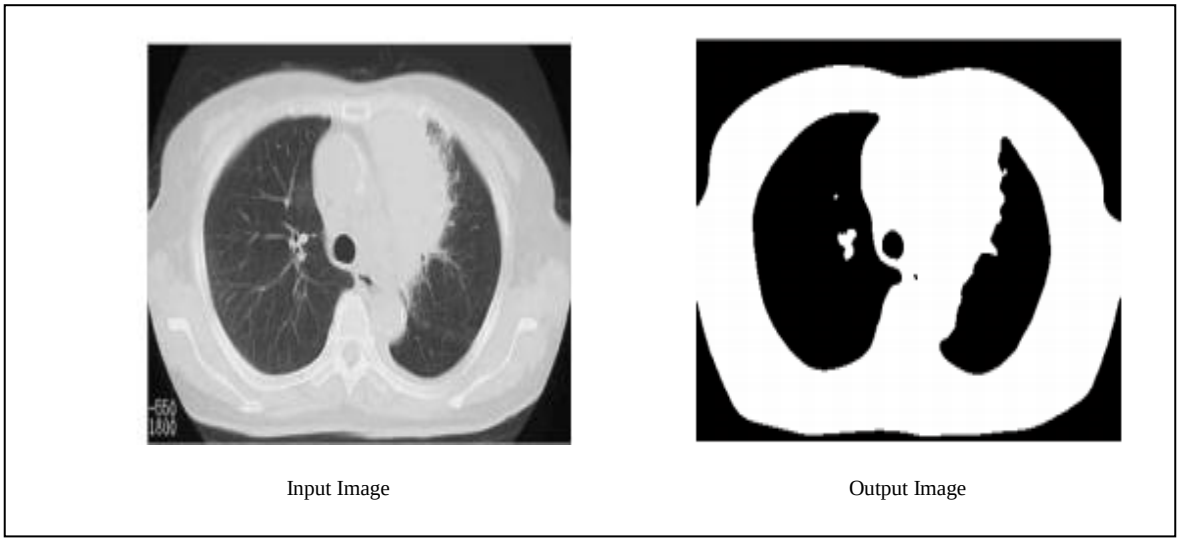

Fig 2: Shows the Histogram Equalization on CT scan image

If the number of the white pixels of a new image is less that the threshold, then it indicates that the image is normal, otherwise, if the number of the white pixels is greater than the threshold, it indicates that the image in abnormal. The threshold value that is used in this research is 255. Figure 3 shows the binarization check method flowchart.

ii. GLCM (Grey Level Co-occurrence Method) [4]

The GLCM is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image. Firstly we create gray-level co-occurrence matrix from image in MATLAB. Then we normalize the GLCM using the following formula

$$P_{i,j} = \frac{V_{i,j}}{\sum_{i,j=0}^{N-1} V_{i,j}}$$

where i is the row number and j is the column number

From this we calculate texture measures from the GLCM.

The following features are extracted using this method-

- Contrast
- Energy
- Entropy
- Homogeneity
- Maximum Probability

Combining Binarization and GLCM approaches together will lead us to take a decision whether the case is normal or abnormal.

C. PCA (Principle Component Analysis) [3]

PCA is to standardize the data in image. Real-world data sets usually exhibit relationships among their variables. These relationships are often linear, or at least approximately so, making them amenable to common analysis techniques. One such technique is principal component analysis ("PCA"), which rotates the original data to new coordinates, making the data as "flat" as possible.

The features extracted are passed through the PCA data mining for better classification. The following steps takes place in PCA:-

i. Calculate the mean and standard deviation of the features in the image.
ii. Subtract the sample mean from each observation, then dividing by the sample standard deviation. This centers and scales the data.
iii. Calculating the coefficients of the principal components and their respective variances is done by finding the Eigen functions of the sample covariance matrix.
iv. The matrix contains the coefficients for the principal components. The diagonal elements store the variance of the respective principal components. We can extract the diagonal.
v. The maximum variance in data results in maximum information content which is required for better classification.

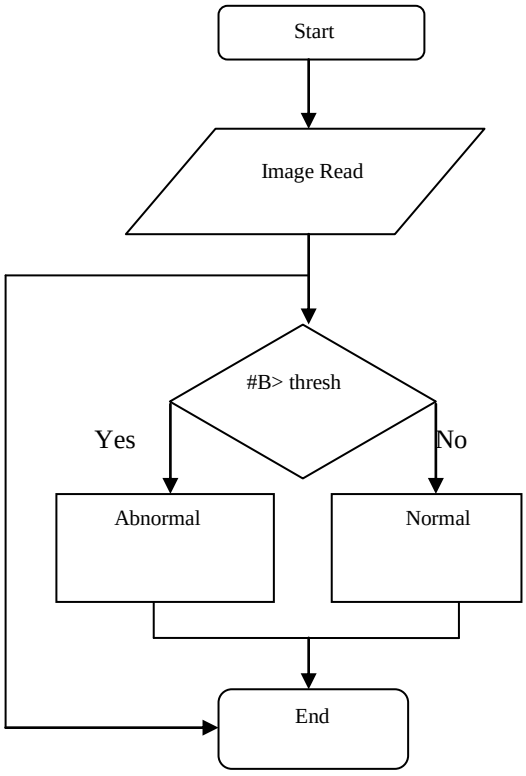

Fig1 :Binarization check method

IV. CONCLUSION AND FUTURE WORK

In this research, we have successfully developed a solution for the detection of lung cancer using image processing algorithms. Image quality and accuracy is the core factors of this research, image quality assessment as well as enhancement stage where were adopted on low pre-processing techniques based on Histogram Equalization. The proposed technique gives very promising results comparing with other used techniques. Relying on general features, a normality comparison is made. In future work, we can also plan to do similar analysis for other cancers.

REFERENCES:

[1] K.A.G. Udeshani, R.G.N. Meegama, T.G.I. Fernando, "Statistical Feature-based Neural Network Approach for the Detection of Lung Cancer in Chest X-Ray Images," *International Journal of Image Processing (IJIP)*, Volume (5), Issue (4) , 2011.

[2] Mokhled S. AL-TARAWNEH, "Lung Cancer Detection Using Image Processing Techniques," *Leonardo Electronic Journal of Practices and Technologies*, Issue 20, January-June 2012.

[3] Taranpreet Singh Ruprah, "Face Recognition Based on PCA Algorithm," *Special Issue of International Journal of Computer Science & Informatics (IJCSI)*, 2231–5292, Vol.- II, Issue-1, 2.

[4] Dr. H.B.Kekre, Sudeep D. Thepade, Tanuja K. Sarode and Vashali Suryawanshi, "Image Retrieval using Texture Features extracted from GLCM, LBG and KPE," *International Journal of Computer Theory and Engineering*, Vol. 2, No. 5, October, 2010, 1793-8201.

[5] Vertika Agarwal, "Analysis of Histogram Equalization in Image Preprocessing," *BIOINFO Human-Computer Interaction*, Volume 1, Issue 1, 2011, pp-04-07.

[6] G. N. Srinivasan, and Shobha G, "Statistical Texture Analysis," *PROCEEDINGS OF WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY* , VOLUME 36 , DECEMBER 2008.

[7] Ö M Soysal and J Chen, " A new spectral feature for shape comparison," *Proceedings of the Int'l Conf. in Image Processing, Computer Vision, & Pattern Recognition, IPCV'08 in WORLDCOMP'08,* pp.23-27, 2008

[8] M.N. Gurcan, B. Sahiner, N. Petrick, H.P. Chan, E.A. Kazerooni, P.N. Cascade and L. Hadjiiski, " Lung nodule detection on thoracic computed tomography images," *Preliminary evaluation of a computer-aided diagnosis system, Medical Physics*, vol. 29, N 11, pp. 2552, 2002.

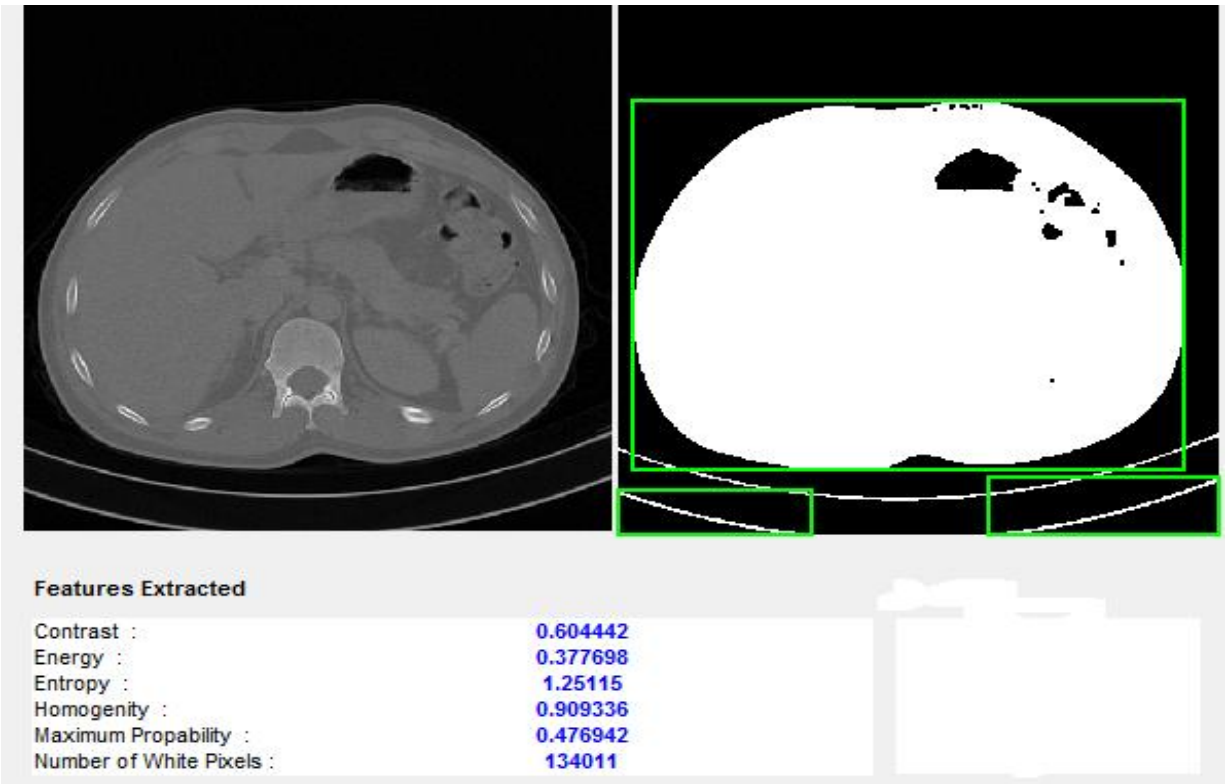

Fig2 Shows Features Extracted of an abnormal lung

