

Classification of Lung Tumour on Ct Images Using Glcm Based Svm Classifier

A.Alaimahal¹, P.C.keerthana², M.Sangeetha³, S.S.Shwetha⁴

¹ Assistant Professor, ^{2,3,4}UG students

Department of Electronics and Communication Engineering
Velammal College of Engineering and Technology, Madurai.

Abstract: Lung Cancer is the leading cause of death worldwide. The mortality rate of lung cancer is the highest among all other types of cancers, contributing about 1.3 million deaths/year globally. Lung cancer can be detected using Computer-Aided Detection (CAD) system which can provide an effective solution by assisting doctors for early detection of diseases. Image processing is one of most recent researches these days and now it is very much integrated with the medical and biotechnology field. The primary phase consists of 2 stages - preprocessing and segmentation. In the preprocessing stage noise removal is done by median filter followed by contrast enhancement using contrast limited adaptive histogram equalization (CLAHE). In the second stage, segmentation using adaptive thresholding is carried out to extract the lung tissues. The segmented lung tissues are obtained as the end result of the primary phase of the lung nodule detection system. In this paper we have presented an application of gray level co-occurrence matrix (GLCM) to extract second order statistical texture features for motion estimation of images. Classification is done by using SVM classifier in order to identify whether the given image is normal or malignant.

Keywords: CLAHE, Median filter, Adaptive Thresholding, GLCM, SVM Classifier.

I. INTRODUCTION:

Lung cancer has been considered as the major cause of cancer deaths over the past few years. The presence of lung nodules which are spherical or oval spots on lungs having a size of 1-30 mm is considered as an indicator of lung cancer's early stage. So it is very important to distinguish cancerous and non-cancerous lung nodules as early as possible. About 40 percent of lung nodules can be cancerous. Timely diagnosis of potentially harmful cancerous lung nodules can help to decrease the death rates of lung cancer patients. The diagnostic challenge lies in the proper classification of lung nodule. Nowadays computed tomography (CT) is used as a standard diagnostic tool and for patient evaluation. In this paper we apply different image processing techniques such as preprocessing which is used for denoising purpose, Histogram Equalization for image enhancement, segmentation is done by adaptive thresholding, is an effective and better method to segment and extract the lung tissues from the CT images. It is suitable for lung tissue extraction which converts the input grayscale images into binary images, feature extraction is done by GLCM method in which important features to detect the lung tumour are extracted. Finally classification is done by SVM classifier to identify whether the given input image is normal or malignant.

II. Methodology:

The proposed algorithm to detect the lung tumor has the following phases.

1. Preprocessing phase
2. Segmentation phase
3. Feature extraction phase
4. Classification phase

The flow diagram of the proposed work is depicted in Fig.1. The primary phase focuses on pre-processing and segmentation stages. The pre-processing stage involves 2 steps: noise removal and contrast enhancement. The pre-processing is followed by segmentation stage to extract the lung tissues. The segmented output is given as input to the feature extraction stage in which features such as energy, correlation, homogeneity, contrast, size, shape, density are extracted. The feature extraction is followed by classification stage which is used to identify whether the given image is malignant or not.

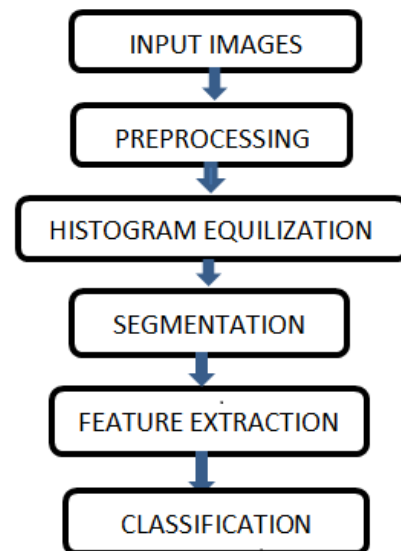


Fig.1 PROPOSED FLOW DIAGRAM

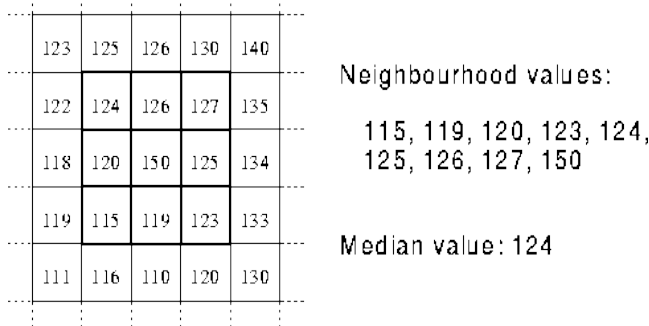
1. PREPROCESSING:

Image pre-processing is the technique for the enhancement of the data images. Pre-processing of images are used to remove the low frequency background noise, normalization

of the intensity of the individual particles in images, removing the reflections and masking the portions of images. In the proposed system, pre-processing stage consists of 2 stages namely noise removal stage and contrast enhancement stage. Noise removal is done using median filter. Contrast enhancement can be done after the filtering for the improvement of the contrast of images so that the visibility of structures can be improved.

1(a).Noise removal using Median filter:

The median filter is a non-linear filtering tool used for removal of noise. Its implementation (hardware) is straightforward and does not require many resources. This filter is used traditionally to remove impulse noise as it is the most popular used non-linear filter. The median filter considers each pixel in the image in turn and looks at its nearby neighbors to decide whether or not it can be a representative of its surroundings. Despite replacing the pixel value with the *mean* of neighboring pixel values, it replaces it with the *median* of those values. The median is calculated by sorting of all the pixel values from the surrounding neighborhood into numerical order and then replacing the pixel being considered with the middle pixel value. Figure 2 illustrates an example calculation.



1(b).Contrast Enhancement Using CLAHE:

Image enhancement is a process of improvement of quality of an image by improving its features. In the proposed system the Contrast Limited Adaptive Histogram Equalization (CLAHE) is used. This evens out the distribution of used grey values and thus makes hidden features of the image more visible. To express the image, the full grey spectrum is used. It optimizes the contrast enhancement on local image data in a divide and conquer mode and hence reduces the global noise. In this method, image is divided into sub-images or blocks, and histogram equalization is performed to each one. The CLAHE presented clip limit to overcome the problem due to noise. The CLAHE restricts the intensification by clipping the histogram before computing the CDF at a predefined value. The value at which the histogram is clipped (clip limit) is dependent on both the histogram normalization and the size of the neighborhood region. During the redistribution few of the pixels are added over the clip limit and it results in an effective clip limit that is larger than the predefined limit. It is to be noted that the accurate clip limit value

depends on the original image. Clip limit is used in limiting the maximum slope of all histograms. The algorithm of the function can be summarized as:

- 1) Calculate a grid size based on the maximum dimension of the image.
- 2) Choose the grid size as the default window size, if a window size is not specified.
- 3) Recognize the grid points in the given image and it has to be done from top left corner. Each one is separated by grid size pixels.
- 4) For each chosen point the histogram for the area surrounding the grid point is computed, having area equal to window size and centered at the chosen point.
- 5) For predefined clip limit, the histogram computed can be trimmed above to that level and then use the new histogram to calculate the cumulative probability function (CPF).
- 6) For each pixel in the input image the steps 6-8 is performed after calculating the mappings for each grid point. The mapping of new pixel in case of uniform distribution is given by equation (5).

$$g = (g_{max} - g_{min}) * P(X_k) + g_{min} \quad (5)$$
 Where, g_{max} is the maximum gray level value, g_{min} is the minimum gray value, g is the computed pixel value and $P(X_k)$ is the cumulative probability distribution.
- 7) Find the four nearest neighboring pixels for each pixel that has been chosen.
- 8) Use the cumulative probability function find the mapping at 4 grid points by making use of the intensity value of the pixel as an index.
- 9) These values are interpolated to obtain the mapping for the present pixel. Plot this intensity to the range [min: max] and put it in the output image.

2. SEGMENTATION:

The preprocessing stage is followed by the segmentation stage where segmentation using adaptive thresholding is used in this work. Thresholding is the simplest segmentation method where the pixels are partitioned depending on their intensity value. Adaptive thresholding typically takes a grayscale or colour image as input and, in the simplest implementation, outputs a binary image representing the segmentation. For each pixel in the image, a threshold has to be calculated. If the pixel value is below the threshold it is set to the background value, otherwise it assumes the foreground value. There are two main approaches to finding the threshold: (i) the *Chow and Kaneko* approach and (ii) *local* thresholding. The assumption behind both methods is that smaller image regions are more likely to have approximately uniform illumination, thus being more suitable for thresholding. Chow and Kaneko divide an image into an array of overlapping subimages and then find the optimum threshold for each subimage by investigating its histogram.

The threshold for each single pixel is found by interpolating the results of the subimages. The drawback of

Neighbour pixel value → ref pixel value.	0	1	2	3
0	0,0	0,1	0,2	0,3
1	1,0	1,1	1,2	1,3
2	2,0	2,1	2,2	2,3
3	3,0	3,1	3,2	3,3

this method is that it is computational expensive and, therefore, is not appropriate for real-time applications. An alternative approach to finding the local threshold is to statistically examine the intensity values of the local neighborhood of each pixel. The statistic which is most appropriate depends largely on the input image. Simple and fast functions include the *mean* of the *local* intensity distribution

$$T = \text{mean}$$

the *median* value,

$$T = \text{median}$$

or the mean of the minimum and maximum values,

$$T = (\text{max} + \text{min}) / 2$$

The size of the neighborhood has to be large enough to cover sufficient foreground and background pixels, otherwise a poor threshold is chosen. On the other hand, choosing regions which are too large can violate the assumption of approximately uniform illumination. This method is less computationally intensive than the Chow and Kaneko approach and produces good results for some applications.

3. FEATURE EXTRACTION:

Feature extraction is a method used for capturing visual content of images for indexing and retrieval. Primitive or low level image features can be either general features, such as extraction of color, texture, size and shape or domain specific features. This paper presents an application of gray level co-occurrence matrix (GLCM) to extract second order statistical texture features for motion estimation of images.

3(a). Extraction of GLCM:

According to the number of intensity points or pixels in each combination, statistics are classified into first-order, second-order and higher-order statistics. The Gray Level Cooccurrence Matrix (GLCM) method is a way of extracting second order statistical texture feature. A GLCM is a matrix where the number of rows and columns is equal to the number of gray levels, G , in the image. The matrix element $P(i, j | \Delta x, \Delta y)$, can be considered as the relative frequency with which two pixels, separated by a pixel distance $(\Delta x, \Delta y)$, occur within a given neighborhood, one with intensity 'i' and the another with intensity 'j'. A large number of intensity levels G implies storing a lot of temporary data, i.e. a $G \times G$ matrix for each combination of $(\Delta x, \Delta y)$ or (d, θ) . Due to its large dimensionality, the GLCM's are very complex to the size of the texture samples on which they are estimated. Thus, the number of

gray levels is often reduced. GLCM matrix formulation is explained with the example illustrated in fig 3 for four different gray levels. Here one pixel offset is used (a reference pixel and its immediate neighbour). If the window is large enough, using a larger offset is also possible. The top left cell will be filled with the number of times the combination 0,0 occurs.

Table 3. GLCM Calculation.

4. CLASSIFICATION:

SVM Classifier:

Support vector machine (SVM) is supervising learning models, associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. In the given set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, building it a non-probabilistic binary linear classifier.

An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap which is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. Data classification is a common task in machine learning. Suppose some given data points each belong to one of two classes, and the goal is to decide which class a new data point will be in. In the case of support vector machines (SVM), a data point is observed as a p -dimensional vector (a list of p numbers), and we want to know whether we can separate such points with a $(p - 1)$ -dimensional hyper plane. This is known as a linear classifier. There are many hyper planes that might classify the data. One reasonable choice as the best hyper plane is the one which represents the largest separation or margin, between the two classes. So we select the hyper plane so that the distance from it to the nearest data point on each side is maximized.

If such a hyper plane exists, it is called the maximum-margin hyper plane and the linear classifier it defines is known as a maximum margin classifier or equivalently, the preceptor of optimal stability.

Support Vector Machines

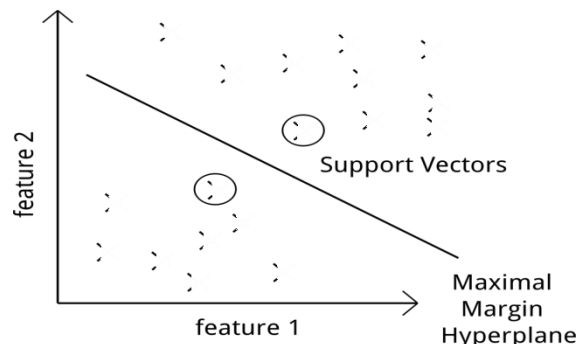


Fig 4. SVM Classifier

III. RESULTS AND DISCUSSION:

The obtained results of all the phases of the proposed method are shown and discussed. All phases of proposed method are implementing using MATLAB software.

1. Preprocessing:

The input image is pre- processed for noise removal and contrast enhancement using median filter and contrast limited adaptive histogram equalization (CLAHE). The input image may contain additive noise from various sources. This may create difficulties in subsequent processing stages while retrieving features and during classification. To remove the noise components the input image is subjected to median filtering.



Fig 5. Input CT image

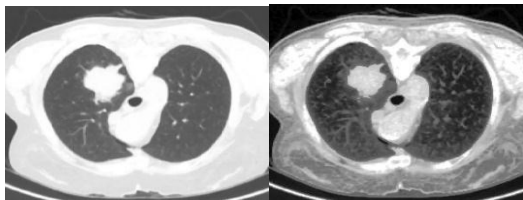


Fig6.Filtered and 7. CLAHE enhanced image.

2. Segmentation:

The pre-processed image is segmented in the next stage in order to extract the lung parenchyma. The segmentation is done by using thresholding method. The thresholding approach uses a threshold value to convert a gray-scale image into a binary image. The threshold is used to isolate lung tissues from the rest of the original image. In this work thresholding approach used to extract the lung is the adaptive thresholding.



Fig 8.Segmented image.

3. Feature Extraction:

The segmented image can be distinguished as either cancerous or not using their texture properties. Gray Level Co-occurrence Matrix (GLCM) is one of the most popular ways to describe the texture of an image. Energy, correlation, homo geneity, contrast such features are extracted.

4. Classification:

An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. One reasonable choice as the best hyper plane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyper plane so that the distance from it to the nearest data point on each side is maximized. If such a hyper plane exists, it is known the maximum-margin hyper plane and the linear classifier it defines is known as a maximum margin classifier or equivalently, the preceptor of optimal stability.

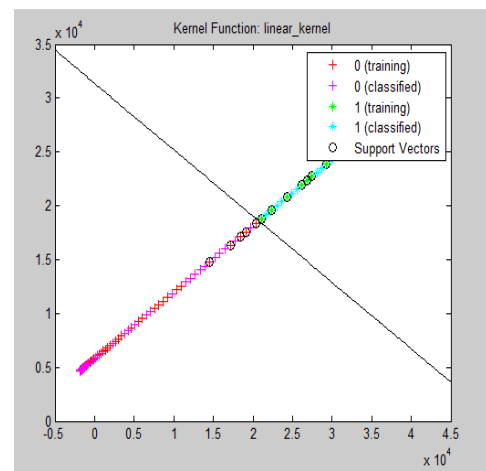


Fig 9.SVM Classifier graph.

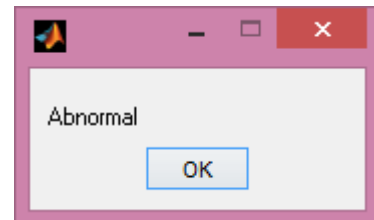


Fig 10.Classified output image.

IV. CONCLUSION:

An efficient algorithm for classification of lung tumor on lung CT images using SVM classifier based GLCM features is proposed. The final classification that is whether the given CT image is normal or abnormal is obtained by comparing GLCM test features and training features by using SVM classifier. The proposed method outperforms other existing methods and also outperforms non-expert human disease identification. The system intends to help the ophthalmologists in the classification of lung tumor screening process to detect symptoms faster and without doubt.

REFERENCE:

- [1] Sasidhar B, Ramesh Babu D R, RaviShankar M, BhaskarRao N “ Automated Segmentation of Lung Regions using sMorphological Operators in CT scan” International Journal of Scientific & Engineering Research, Volume 4, Issue 9, September 2013 1114 ISSN 2229-5518.
- [2] Disha Sharma and Gagandeep Jindal “Computer Aided Diagnosis System for Detection of Lung Cancer in CT Scan Images” International Journal of Computer and Electrical Engineering, Vol. 3, No. 5, October 2011.
- [3] Mr. Vijay A. Gajdhane , Prof. Deshpande L.M. “Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques” IOSR Journal of Computer Engineering (IOSR-JCE) .
- [4] Prof. B. C. Preethi, Gia Elizabeth Abraham “Lung Tissue Extraction Using OTSU Thresholding in Lung Nodule Detection from CT Images” International Journal of Current Trends in Engineering & Technology.
- [5] P. Mohanaiah, P. Sathyanarayana, L. GuruKumar“Image texture feature extraction using GLCM approach”International Journal of Scientific and Research Publications, Volume 3, Issue 5, May 2013.
- [6] Ruchika, AshimaKalra“Detection of lung cancer in CT images using Mean Shift Algorithm”International Journal of Advanced Research in Computer Science and Software Engineering, volume 5, Issue 5, May 2015.
- [7] Jyothsna ,Dr.G.R.Udupi”Adaptive K-means Clustering for Medical image segmentation”International Journal of Technical Research and Applicationse-ISSN: 2320-8163.