Content Based Retrieval of Small Cell Lung Cancer Images from Clinical Database

Vijayalaxmi Rathod*1, Rani Shetty*2

¹Student Computer Science and Engineering Department SDMCET-Dharwad, Karnataka, India

²Assistant Professor Computer science and engineering Department SDMCET-Dharwad, Karnataka, India

ABSTRACT

Lung Cancer is a standout amongst the most deadly ailment influencing humanity on the planet. Small cell lung tumor is starting phase of Lung carcinoma disease, initially it found in the just a single side of lung and later it will spreads to other parts of human body. The proposed Content based image retrieval system for Small Cell Lung Cancer images, system retrieves the similar images from database that match the input query image condition and also decides that the input CT lung image is influenced by malignancy or not. Efficiency of algorithms are also calculated, for content based Small Cell Lung Cancer image retrieval system the Precision and Recall values are also studied.

Keywords: Image Enhancement, Gabor Filter, Segmentation, Feature extraction, Feature vector

I. INTRODUCTION

Cancer is gathering of infections, it is described as strange and uncontrolled cell development in body. Cancer is the out of control boom of unusual cells in the human body. Most cancers develops whilst the frame is normal manage mechanism stops operating. These unusual cells may body a mass of tissue, known as cancer tumors. Dangerous cancerous tumors are risky, which implies they could spread into, or attack to, adjacent tissues in the human body. Lung disease is additionally called as Lung carcinoma, a dangerous type of cancer tumor initially develops in the phones of lung at first and It can likewise spread to different parts of the human body.

A lung malignancy is classifies as non– small cell lung cancer and Small cell lung carcinoma in light of the kind of mobile in which the sickness commenced. Small cell lung cancer (SCLC) speaks that around 15 to 20 percents of lung cancer growth and remaining 80 to 85 percents of lung malignancies are Non-Small Cell Lung Cancer (NSCLC). Oat cell lung (SCLC) malignancy is the beginning level of lung cancer.

Anybody can get lung growth, yet 90 percent of lung tumor cases are the after effect of smoking. Small cell lung disease is quite often connects with substantial smoking. From the studies, we can conclude that the early detection of lung carcinoma can decrease the death rate to great extent.

Therefore, there is need for automatic detection of Small cell lung cancer. This has lead to the associated diagnostic details so that it can be used for a medical decision support system framework. Presenting an efficient technique for retrieving similar case's data (CT images) from the database, that matches fulfil the condition specified in the query image.

II. METHODS AND MATERIAL

In this paper, we are trying for generating algorithm that may use for successful detection of Small cell lung cancer and developing content based system for same. The contemporary work is organised as follows

A. *Image Enhancement*

In this paper we are working only on CT images of Small Cell Lung Cancer, This format is useful compared to others medical image format because CT images enable radio therapist and doctors to distinguish inward structures of the body system and see their shape, size, thickness. The objective of image pre-preparing is change of the image information that stifles undesirable twists or improves some image features essential for additionally processing, and it lessens additionally processing time.

In the next phase of Image Pre-processing is image upgrades. Image enhancement is the way toward changing computerized images with the goal that the results make, image more appropriate for visualising or further image processing method..

In this paper we are using Gabor filter as image enhancement method it's far a linear filter, this kind filters more preferably used in the area of image texture analysis, identification purpose of object edges, extraction of features from input.

These filtering techniques are unique instructions for filter bandpass, it means that this filtering technique considers only particular frequency band and rejects the frequencies of other than defined band. When image undergoes by this kind filtering technique, this method offers the very best reaction at the boundaries of object when texture of image changes. g(x, y; λ, θ, ψ, σ, γ)=

$$\exp\left(-\frac{X'^2 + \gamma^2 Y'^2}{2\sigma^2}\right) \exp(i(2\pi \frac{x'}{\lambda} + \psi)) \qquad (1)$$

where $X' = xCos\theta + ySin\theta$

The various parameters affecting the Gabor filtering method are variance value σ of the Gaussian function and γ specifies the ellipticity property of the Gabor function

When a Gabor filter is applied to an image, it produces the highest response at edges and at points where texture changes in an image. A Gabor filtering method is more responds for edges and texture changes in an image. When we say that a filter responds to a particular feature, it means that the filter has a distinguishing value at the spatial location of that feature in the image.

B. *Image Segmentation*

The next phase of this image processing work is Image segmentation.

Segmentation of image is the method for partitioning an image into more than one segment in manner that it is useful for further processing. The main intension of image segmentation is to easier the image representation in way that is more meaningful then original image and easier to research as compared to original image. The image segmentation is generally used for discovering the items and boundaries like (lines, curves, and so on.) in images. Precisely describing, the image segmentation is the procedure of assignment of a pixel labelling to each pixel in an image, where each pixel in equal vicinity stocks the similar characteristic. The achievement of image evaluation depends on reliability of segmentation of enter image

In this paper, we are describing the Marker Controller Watershed algorithm for image segmentation, which is very flexible and robust method for segmentation of images in medical imaging field.

Marker-controlled watershed segmentation

Watersheds method is one of the oldest concepts in the topographic surface area. This is kind of transformation where a single line that specifies the drop of water that will falls into a unique vicinity.

Mathematically in morphology process consideration of gray-level images are topographic relives. When representing image in topographic manner, the value of intensity considered as an elevation.

In image processing, segmentation using Marker controlled method is an exhausting method. This method is used for separation of object, which is in touch with location as background in image. This segmentation method proved as one of the best and sturdy and flexible for image segmentation, for images, which are having closed contours, here ridges are described as edges of an object. The resulting image of this process is binary image contacting single point of marker or connected points of marker as a larger region.

Each single markers has one-to-one relationship initially for specifying the region of watershed, as a result count of single marker point is same as final count of watershed regions. The segmentation results as ridges are organized in preferred manner based on the bounds of watershed areas, the results is segregation of objects from its surrounding pixels. The selection of marker can also be done manually. However high throughput experiments regularly generates markers automatically to save time and resources of person.

Conventional ways for predetermination is wide variety and approximate place of the areas furnished by using the watersheds approach is composed inside the amendment of the homotopy of the characteristic to which the set of rules is applied. The change method is finished by a mathematical morphology operation and reconstruction, with the aid of which the characteristic is modified so that the minima may be imposed by using an outside characteristic (the marker feature).

Segmentation steps are as follows:
- ✓ Reading of enhanced input image
- ✓ Computation of segmentation function.
- ✓ Foreground markers computation.
- ✓ Background markers computation.
- ✓ Modification of segmentation function
- ✓ For modification of segmentation function computing the watershed transform.

c. Feature Extraction

Feature extraction is essential part in the image processing technique in this phase feature image is extracted from previously segmented image. This extracted image is required for further processing. Characteristic extraction from image is also a kind of dimensionality reducing method, where it generates compact vector of features, where this vector contains the exciting components of an image. Here in this work morphological operations are finished to extract the specified function from segmented image. The extracted image capabilities include the useful image details. Feature extraction strategies are applied to get features on the way to be beneficial in recognition of images.

The more common morphological operations are erosion and dilation. Dilation operation defined as, in image adding the pixels to edges of item or object and similarly erosion operation is removing of pixels from the edges of object. Addition or subtraction of objects from an image is depending on size and shape of the structuring element of morphological operator.

1. Features of Erosion

This operation is performed by removing the items from the binary image, where its size is actually less than that of structuring element

In binary image, the object is removed completely from the binary image, where object is having size very lesser than the element. And this process

eliminates the pixels which are in the perimeter of object, and which is actually larger.

It is written as

$$A \, \ominus \, B = \prod_{b \in B} A_{-b} \qquad (2)$$

2. Feature of Dilation

This operation is performed by filing the broken regions and holes in a binary image, where it actually joins the region which might be separated by the areas of object, where these objects are having smaller size than structuring element.

In binary image, this process joins the area of objects in an image by adding pixels, which is separated by the region which is having values less than that of its structuring elements. Dilation technique is executed by sliding the structuring element of B on image of A.

It can be described as :
When structuring element B slides on image A, if the starting point of element coincides with pixel having value 1 in image that is white pixel then no alteration is performed, just move to neighbouring pixel. Otherwise if it coincides with pixels having value 0, then all the pixels covered by this structuring element is converted to black pixels.

Written as

$$A \, \oplus B = \prod_{b \, \in \, B} A_b \qquad (3)$$

D. *FEATURE CALCULATIONS*

As soon as the function is extracted from the segmented image some statistical function values are calculated to recognize and classify image. The function extracted image is binary image that incorporates only the extracted a part of cancer nodule inside the lung image, for this image some physical and texture statistical feature values are calculated and in addition they may be saved in database for content material based retrieval purpose.

These are some features which are calculated

1. **Area:** It's far the scalar price it results the count of nodule object pixels inside the (Region of interest) extracted from the binary image [4].

$$\text{Area} = A_{(i,j)} \, (\text{X ROI}[A_{(i)}] \, , \, \text{Y ROI}[A_{(j)}] \,) \qquad (4)$$

Where, the vector X contains the x coordinate points of ROI and vector Y contain the y coordinate of ROI.

2. **Perimeter:** It's far a linear data that offers real count of nodule object pixel in the binary image. This function may produce a 1d array or vector of boundary of object, this vector contains the value varies between zero and 255 [4].

$$\text{Perimeter} = P_{(i,j)} \, (\text{X edge}[P_{(i)}], \quad \text{Y edge}[P_{(j)}]) \qquad (5)$$

Where, vectors are represented by the edges, edge X represents one vector and vector Y represents another edge

3. **Eccentricity:** This matrix costs are likewise called as roundness of improper complex (I) equal, The max eccentricity is equals to 1, if the object is rounder then its eccentricity value is closer to 1 [5].

$$\text{Eccentricity} = \frac{\text{Length of Major Axis}}{\text{Length of Minor Axis}} \qquad (6)$$

And some texture capabilities are calculated using gray-degree Co-incidence Matrix. This approach is a manner of extracting the texture statistical capabilities from an image.

In GLCM, wide variety of rows and columns of matrix are same to the gray scale degree, in graph of image G. Matrix defines the element co-ordinate (Δx, Δy) is the relative frequency with two pixels values, where these pixels are segregated by the distance among point Δx and point Δy, in the image by its adjacent pixel points

4. Entropy: It suggests the how much quantity of image records is wanted for the image suppression. It calculates the quantity information in an image [5].

$$Entropy = \sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} -Pij * logPij \qquad (7)$$

5. Correlation: It can be defined for calculating the direct dependency of binary levels of next pixel values. This Correlation method allows for monitoring and image registration methodology. It is used for determine the displacement, pressure [5].

$$Correlation = \frac{\sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1}(i,j)\,p(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y}$$

(8)

6. Contrast: It described by calculating the intensity evaluation for connecting pixel and their next pixels, in the entire image. At regular picture comparison value is 0. After i and j are identical i.e. I-j=0. No evaluation is there. When i and j are differ by using 1, small comparison is there. After i and j range with the aid of 2, the evaluation is expensing and weight [6]

$$Contrast = \sum_{i,j=0}^{N-1} P\,i,j\,(i-j)^2 \qquad (9)$$

7. Homogeneity: It passes the cost that calculates the tightness of distribution of the factors within the GLCM to the GLCM diagonal. For diagonal GLCM its price is 1 and its variety is [0,1].Opposite of comparison weight is homogeneity weight values, with weight decreases exponentially free from the diagonal [6].

$$Homogeneity = \sum_{i,j=0}^{N-1} P(i,j)/R \qquad (10)$$

8. Energy: It considers the image texture for its calculation. It offers the addition of rectangular elements in GLCM matrix. It is fully extraordinary from entropy. Whilst the

window is talented orderly, electricity cost is high .The square root of ASM(Angular 2d second)

texture function is used for energy. Its range varies between[0 1] [6].

$$Energy = \sum_{i,j=0} P(i,j)^2 \qquad (11)$$

III. SYSTEM DESIGN

In this phase, we are discussing the process of defining the architecture of system and its sub-modules that fulfils the requirement of system.

A. *Small Cell Lung Cancer detection System*

It describes the way of information flows through a process or system. It includes data inputs and outputs, data stores, and the various sub-processes the data moves through in system framework.
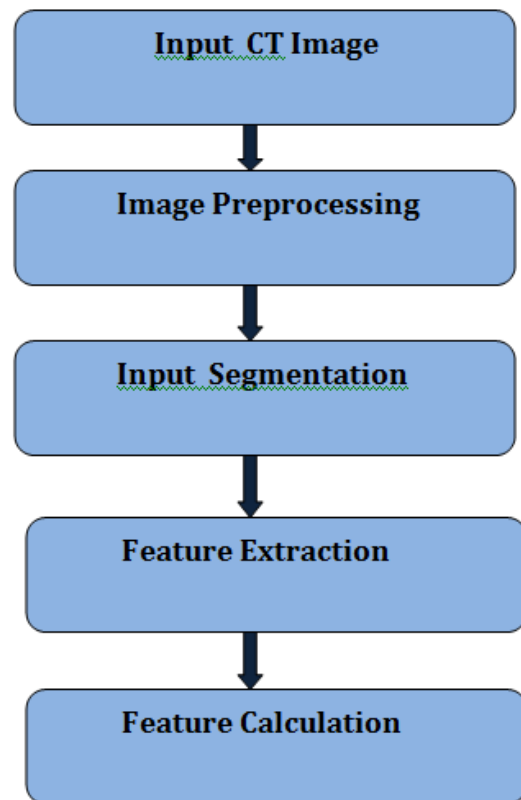

Figure 1. Proposed system for Small Cell Lung cancer Detection.

A. *Content Based Retrieval System Small Cell Lung cancer*

This CBIR system for Small Cell Lung Cancer, it is for retrieving image from the database that is similar to query image. In this system initially image is

enhanced to improve its visually quality and this enhanced image is segmented, the segmented image is in the binary form after this region of interest is extracted that is only the cancerous tumor in the binary image and then physical and texture statistical features are calculated for input image.

The statistical feature information from the images is stored in the database for all images for training purpose in the .mat file format, and When query image is selected for testing first image is enhanced, segmented and feature is extracted for query image and statistical feature information is calculated for this query image and stored in feature vector for the query image and then similarity matching is done with database stored images. Image Id's are image number in the database that is used for indexing during the query, the feature of query image is generated and images are found by a matching engine with similar features. Euclidean distance and its variations are commonly used to define the similarity of features of two images.
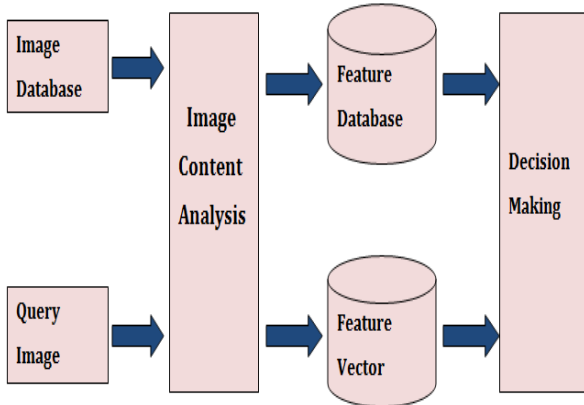

Figure 2. Content Based Retrieval System for Small Cell Lung Cancer images.

IV. RESULTS AND DISCUSSION

In this section of paper we described the result we got from implementing proposed system.

The database contains total of 50 images both normal lung image and Small Cell Lung Cancer Defected CT image for content based image retrieval

When system accuracy is calculated manually it gives result about 84.74%. Precision and Recall efficiency is also calculated for Content based Retrieving of Small Cell Lung Cancer images from Database, The proposed system is having higher precision efficiency value for 44 images, the precision value is more than 87% and Recall values also calculated.
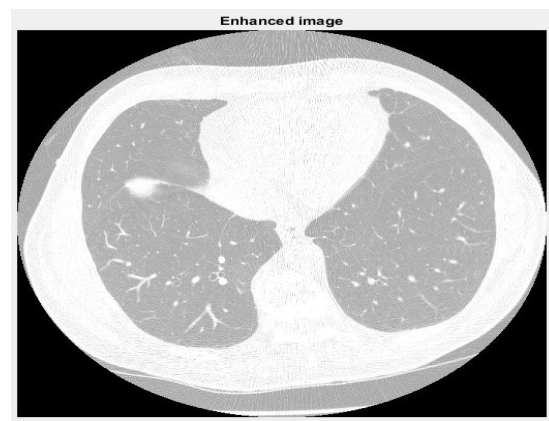

Figure 3. Input image

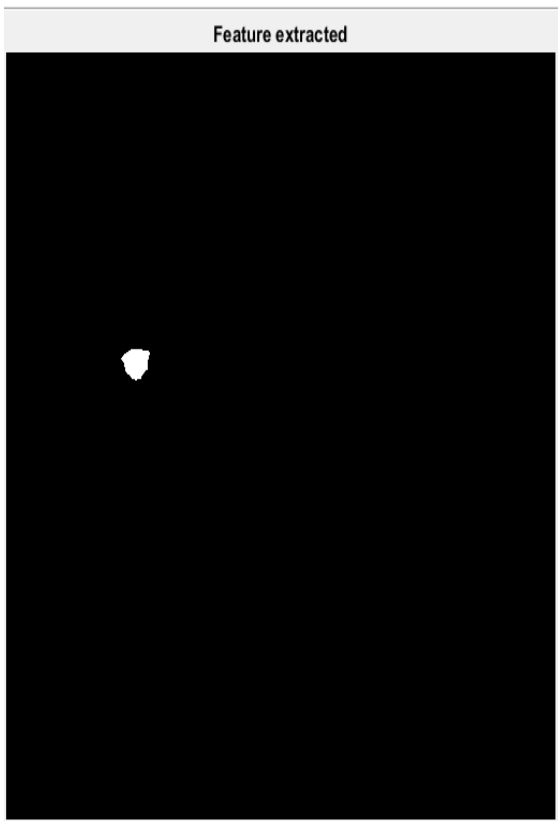**Figure 4.** Gabor Filter enhanced image

Figure 5. Segmented Image

Figure 6. Feature Extracted

Figure 8. GUI for Content based Image Retrieval system for Small Cell Lung Cancer images


```
Statistical information of extracted features are

statistical_properties =

            Area: 390
    Eccentricity: 0.54943
       Perimeter: 73.026


GLCM_properties =

        Contrast: 0.0002064
     Correlation: 0.93067
          Energy: 0.99682
     Homogeneity: 0.9999
```

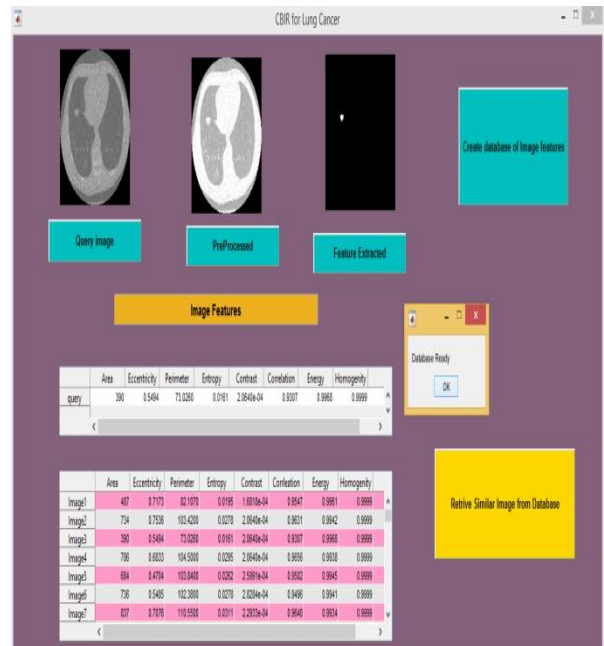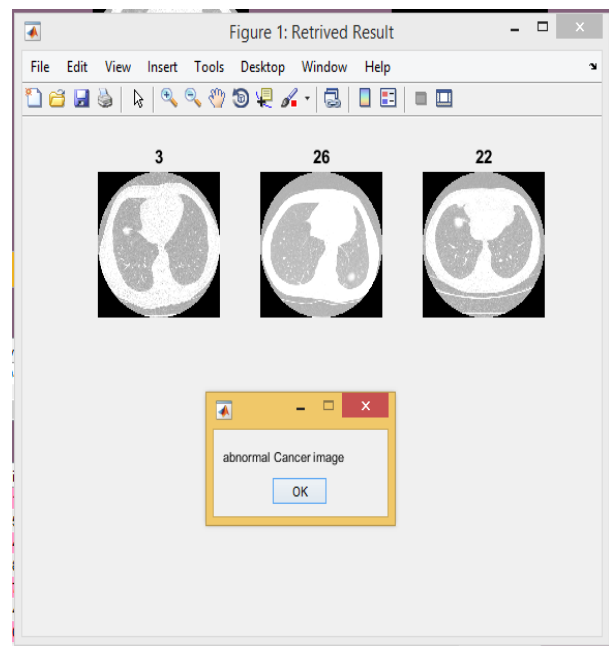**Figure 7.** Statistical information from feature extracted image

Figure 9. retrieving of similar image from database

V. CONCLUSION

The proposed Content based Small Cell Lung Cancer image retrieval system has main modules which are Image acquisition, Image enhancement and Image Segmentation then Feature extraction and Calculating the distance between query image and

feature database images and retrieving of similar images from the database. Our overall performance is achieved is 84.75% when accuracy is calculated manually, Precision and Recall values are also calculated and for 44 images system shows the 100% result for retrieving . Overall, our system achieved satisfactory result

VI. REFERENCES

[1]. Bhagyashri G. Patil, "Cancer Cells Detection Using Digital Image Processing Methods," International Journal of Latest Trends in Engineering and Technology (IJLTET), Vol. 3 Issue 4 March 2014.

[2]. Georgy George, Nisha J.S, A novel approach for the detection of small cell lung cancer on entropy and PSNR value

[3]. Mokhled S. AL-TARAWNEH, "Lung Cancer Detection Using Image Processing Techniques".

[4]. Gawade Prathamesh Pratap and R.P. Chauhan ," Detection of Lung Cancer Cells using Image Processing Techniques".

[5]. Aniket Gaikwad, Azharuddin Inamdar, Vikas Behera , ,"Lung cancer detection using digital Image processing".

[6]. Khin Mya Mya Tun, Aung Soe Khaing, "Feature   Extraction and Classification of Lung Cancer Nodule using Image Processing Techniques"

[7]. Er. Kanchan Sharma, Er. Priyanka , Er. Aditi Kalsh, Er.Kulbeer Saini "GLCM and its Features"

[8]. P. Mohanaiah, P. Sathyanarayana, L. GuruKumar, "Image Texture Feature Extraction using GLCM Approach".

[9]. "www.lungcancer.org.

[10]. www.healthline.com.