Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 10, Number 6 (2017) pp. 1577-1594

http://www.ripublication.com

Content Based Medical Image Retrieval and Clustering Based Segmentation to Diagnose Lung Cancer

Neha Malviya[1], Dr. Naveen Choudhary[2], Kalpana Jain[3]

[1]Department of Computer Science and Engineering, MPUAT College of Technology and Engineering, Udaipur, India.

[2]Professor, Department of Computer Science and Engineering, MPUAT College of Technology and Engineering, Udaipur, India.

[3]Assistant Professor, Department of Computer Science and Engineering, College of Technology and Engineering, Udaipur, India.

Abstract

Now a day lung cancer is most serious health problem in the world which causes multiple deaths every year. There are various techniques available for diagnosis of the lung cancer such as CT image, MRI image, X-Ray Image etc. but the CT scan image provides greater details about multiple organs of lungs. Hence today the medical images are generated more and more in their daily activities which are millions in size. Retrieving medical images from the large collection is a challenging task, therefore it emerges content based medical image retrieval system (CBMIR) system. The retrieval system proposed clustering based segmentation for diagnoses of the lung cancer. Basically it has three phases. First is segmentation for segment out the lung image into particular regions, second phase describes the texture feature extraction of lung regions and third is clustering which is used to classify and arranged into images in particular cluster which is further improved the speed and accuracy of system by retrieving images. It is analysis and measures the performance in terms of precision and recall with respect to time.

Keywords: CBMIR, segmentation, ROI, GLCM matrix, k-means clustering, hierarchical clustering.

1. INTRODUCTION

Lung cancer is most serious health problem in all over the world**.** It is second most common cancer [22] in both men and women. Lung cancer would estimate report for about 13% of all cancer diagnosis and 28% for all cancer deaths. The survival rate of lung cancer is 4 To 5 years and there having only 15% is surviving if lung cancer is diagnosed at its early stage. This rate is increases to 49% while it is still localized and identified and medical images support in clinical decisions. There is rapid development of medical science and technology which generates more and more digital medical imaging techniques. The digital medical image data has been needed of efficient retrieval system. Due to this problem arises [9] when retrieving images from the large collection set of images. The retrieval of such image in specific application data affects the efficiency, retrieving speed and scalability of system. In medical field retrieving image from large dataset is a tedious task. The lung cancer is diagnosed by Lung CT image [12] which is used for monitoring therapy and diagnosis diseases and also it has better clarity with less distortion. A lung CT images are usually followed the content based approach. However it is carefully extracted and classified features [10] of medical image with efficient techniques for easily retrieval since the each medical image in dataset have some special characteristics. First of all each medical image has decomposed into modality and anatomic region because medical images [25] are differ from one another. They are selected on the basis of the image modality. Therefore content based medical image retrieval process has well known diagnosis process for estimate accurate result. Generally the medical image has highly inconsistency and different minor structures, so there it is need to be requirement of the feature extraction and classification [24] of image for efficient retrieval. The main aim of this dissertation is retrieved the images from the large volumes of medical image with high accuracy by carrying out the feature extraction and classification process and these retrieved images play an important role in surgical planning, medical training and diagnosis of lung cancer using lung CT images.

The purpose of this research is improved the performance of CBMIR system including domain specific knowledge. In this experimentation the feature extraction process based on the segmentation of lung CT images and texture feature based retrieval process made by gray level co-occurrence matrix (GLCM). According to its structural functions it is organized in various sections. In the first section it gives overview about CBMIR system, second section discussed drawbacks of present techniques. Section third explains methodology and clustering based segmentation framework. A comparative analysis of proposed techniques present in section 4. It followed by discussion and proceeds by conclusion and result analysis.

1.1 Background

Regarding the rapid development of multimedia technology and information technology which produces huge amount of multimedia data thus the dataset capacity becomes larger and deciding for how to efficiently locate desired image from the large collection of the image dataset. To resolve this problem (CBIR) content based image retrieval [9] is considered for retrieving images from the large dataset according to the content of images. CBIR introduced in 1990 it is based on automated features extraction methods. In recent years there is quick development in the medical imaging [21] technology which improves in the medical services to support for clinical decision making in many image modalities such as CT –scan, MRI, PET, X-rays are considered. In the CBIR system process it basically depends on three features such as color, texture and shape. These features extracted [10] the property of query image and also extracted the features of various images available in the dataset. This method represents only the property of image but does not describe how to handle the inter-relationships, objects or regions. The medical imaging modality is depended on spatial relationships [3] between images. These relationships are presented by content based medical image retrieval (CBMIR) system to locate the spatial relationships between such objects which are extracted from images. The image retrieval system in medical domain provide efficient classification as well as some of the present woks related to medical field in which ASSERT [13] system for HRCT lung image dataset, Image Map [15] system handle multiple organs of medical image, IRMA system [14] describe information model using CAS image dataset and SPRIS [16] system is associated to retrieval of digitized spine x-rays.

1.2 Content Based Medical Image Retrieval

The CBMIR system mainly focuses on spatial relationships between regions or objects of image. It is an interactive system to identify and extracting regions from all segmented images. This methodology is focused on the regions and relationships [3] between the objects. It is also recognized as a new model for medical images where images are first decomposed into regions or objects using segmentation techniques. The main opinions are as follow

1. Presents an effective retrieval method for examining retrieval process of the digital medical images.
2. Define a method for efficient storage and retrieval of medical images based on image attributes. The frame work of CBMIR system present in Figure 1.

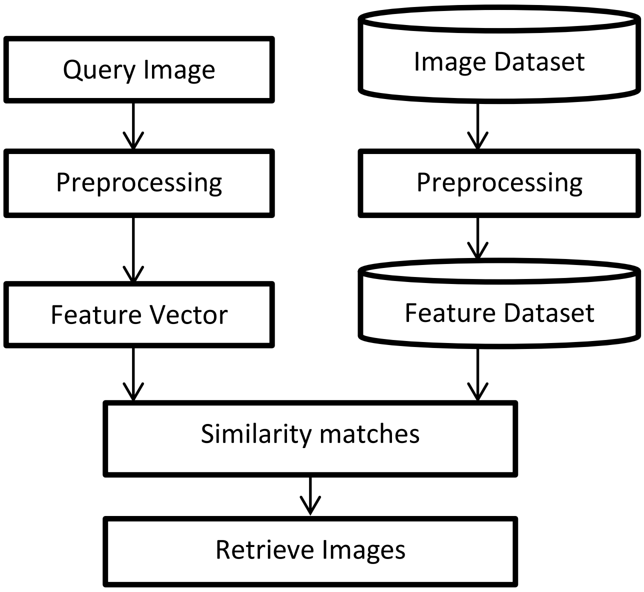


Figure 1: CBMIR System

1.3 Lung cancer

The lungs are main organ of the respiratory system in which the normal cells of lungs are generated, grown and divided into new cells. When this process goes wrong, it generates abnormal cells; thereby it comes into tumor [22] and spread other area of human body. Lung's tumors can be benign and malignant. The main complicated part of system is diagnosis of the lung cancer where multiple techniques and imaging technology available for monitoring and diagnosis of the lung cancer including PET image, MR Image, Mammogram image, x-ray image and CT scan image etc. but CT scan image gives better clarity and low distortion result this is made the easy calculation and the evaluation of texture feature of the medical image.

1.4 Computed Temography (CT) image

There are multiple techniques available for diagnosis of the lung cancer [12] including MRI, PET, X-ray, but the CT scan image is considered as one of the best method which gives detailed description about lung nodules. It has low noise and less distortion and gives better clarity than other images, thereby easily calculate and estimate texture feature of image.

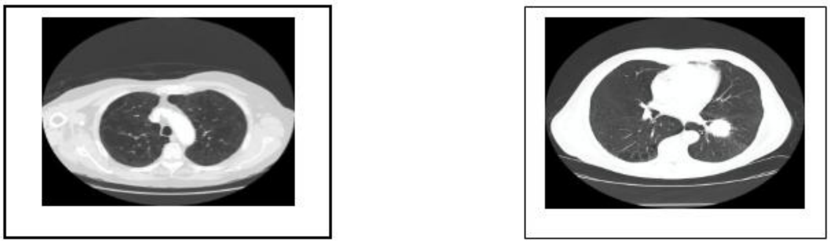


Figure 2: Lung Ct Normal and Abnormal image

1.5 Dataset

Lung image dataset is acquired from NIH/NCI (national informative health/ national cancer imaging) organization (LIDC) lung image dataset consortium that provides lung computed temography image which is available on the web cancer [20] imaging archive site. It is publically available and easily accessible.

2. LITERATURE REVIEW

The image retrieval is the process of the current research [9] area where image mining proposed searching and retrieving techniques basically depends on texture feature of the image which presents the image very huge. The present technique [1] color image retrieval based on RGB color component of image taken from each image where each image lies between 0 to 255 values in two dimensional. The top most similar images are clustered according to their texture feature and texture based classification which depends on the entropy based method where images are grouped under high, average and low texture values because entropy checks the variation and measures the randomness of image. The result of retrieved images depends on the pre-clustered images based on the fuzzy c-means clustering and to check the similarity between the images where entropy based method is used. The present technique is based on standard color images and clustered by FCM clustering method but this method is not suitable in noisy environment because it requires less number of iteration with improper cluster. It has also uncertainty about clustering of medical images and it gives improper result in noisy environment. In gray level based image retrieval method including the medical image retrieval where each image represents gray level distribution of pixels available in image. A gray scale image values typically lay between 0 to 255 ranges thus 0 for black pixel value, 1 for white pixel value and remaining shows the shades of gray. Medical image retrieval is robust process which is based on translation and scaling of the objects and depends on gray level distribution of pixels. Therefore it needs to be requirement of efficient retrieval system which is capable of retrieving medical images. The CBMIR employs [3] as new modal where each image is first decomposed image into regions. It is also focused on the spatial relationships between pixels of the image. Before the classification of medical image it needs to be requirement of segmentation [24] of image for extract particular region of interest (ROI). Hence the extraction of region and detection of lung's tumor from the image is a tedious and error prone task. For segmentation it considers ROI as exterior body part. The segmentation process [17] separate out the suspected nodule area from image and it is easily to distinguish the lung from the whole body structure of image. Sometime it is included the excluded body part of the lung as part of ROI or ROI with black islands. So the segmentation process must be needed for extract lung regions from image. To segment out the lung region form CT-scan body, it is necessity of iterative optimal thresholding method where it is illustrate and analysis of the lung images. The extractions of regions are recognized by the texture feature [6] classification which measures various aspect of the image in the form of color, texture and shape. In medical imaging field all visual information are in form of health information stored in the scanned images. The analysis of texture feature [11] method of medical images is based on GLCM matrix. This process is based on the statistical method and spatial distribution of the gray level. This method is efficient for texture feature classification and computes the various statistical measurements to increase the resemblance between images. Where the target images are retrieved and process will be based on accurate classification of images. The images retrieval process of target images will be fast only when it is clustered in right manner and clustering based segmentation improve the accuracy of system. Clustering is offered the superior organization of multidimensional [6] data for well-organized retrieval process. This process depends on the prior knowledge of cluster and wants to require finite cluster set. For extract the significant phrases [8] of image it is needed efficient cluster method however with hierarchical clustering generate hierarchy of similar feature and all similar feature confined into one cluster. But it does not represent as heuristic method so it needs to be requirement of efficient methods which gives appropriate description about cluster. To get accurate favored images from cluster [7] for analysis the of medical images which would help to identify correct evaluation of image and analyzed the performance of system based on partitioned based [6] method where it is presenting fixed clustering of images related only one cluster rather than belonging to two or more clusters. CBMIR system application is based on lung cancer diagnosis process. The accurate extraction [10] of images from large dataset is obtained from (LIDC) [27] lung image dataset consortium of lung CT image that is provided by NIH/NCI [26] organization. Similarity comparison between images based on texture features of images. There are different metric functions available for determining the similarity degree of image. A distance function such as Euclidean distance [3] method is used to calculating distance between each pair of signature of image. The similarity indexing method is used to efficiently locate the signatures of image where are similar to query point and the corresponding resemblance images returned to the user.

3. PROPOSED METHODOLOGY

The objective of the CBIR system in the medical domain is allowed to radiologist to retrieve similar images based on resemblance feature that can be focused on diagnosis process of input image. The proposed CBMIR framework is shown in the figure 3, in which lung CT scan image given as input in the system. CBMIR such a method cannot directly apply on medical images first it is needed to be requirement of segmentation process to segment out image into particular regions. In this study the extraction of texture feature basically depends on accurate evaluation of lung region thereby we extract texture features of regions of lung image. In this proposed system it has three sections first is segmentation process to extract lung images out of the CT scan body it requires the iterative optimal thresholding process. The main aim of segmentation process is to separate out the lung region that characterized by different anatomical structures. Since the gray level variations in each image for the specific case is quite large. So, second stage depends on texture feature extraction process. The extracted lung region is used for texture feature calculation, the (GLCM) gray level co-occurrence matrix contain information about the distribution of pixels and which is similar to gray level values. GLCM matrix is derived the various texture feature descriptor which is stored in feature vector dataset. The same process is applied for the query image which is submitted by user and stored all information about query in the query vector. The third stage is clustering for classification is used to improve the efficiency of system, For reduce the searching time of images in dataset, it clustered similar type of feature of images and check the similarity between the query image and dataset images and similar images are reported. The proposed system is considered a domain specific based search engine for lung CT scan image to diagnosis of lung cancer and it is shown in Figure 3.

3.1 Algorithm

There are two dataset implemented in this research.

1. All are real time 500 lungs CT scan images are saved in dataset $D_{d.}$
2. The features of images saved in dataset are named $D_{f.}$
3. The proposed CBMIR system is divided in four phases

a. Phase1: CBMIR algorithm for load dataset images.

Step1: first load real time lung image dataset $D_{d.}$

Step 2: At preprocessing stage filter out and segment the particular region (ROI).

*/(et. Figure 3)

b. Phase 2: CBMIR algorithm for creating the feature dataset.

Step1: for k=1: n */read the k number of images of dataset D_d.

Step2: for i= 1: f_k */is the number of feature of k^{th} images.

Step3: read i; */read features I from dataset

Calculate F_{i1}, F_{i2}, $F_{i3.}$

Step 4: save k, I in $D_{f.}$

end;

end;

c. Phase 3: CBMIR algorithm for classifying images.

Step1: Input I feature of dataset $D_{f.}$ */ classified D_f dataset into C clusters.

Step2: Establish hierarchy of i features.

Generate cluster Ci ={C_1, C_2, C_3, … C_n}

*/ grouping the similar features of image using hierarchical clustering.

Step3: Run the K clusters each have N_k patterns related to cluster centers m.

*/ assign fixed cluster using k-means clustering.
For all j ∈ X_j to cluster Kj.
Step 4: Partitioning and update mean ($\{X_j : X_i \in K_j\}$)
Partitioning K_1, K_2, … K_m.
end;

d. Phase 4: CBMIR algorithm for query matching.

Step1: Input Q segmented into region (ROI); */ enter query image.
Find F_q ;*/ calculate feature vector of query image. like Phase 2;
Step2: for Y=1 : n */ extract number of features from dataset D_f.
Step 3: Compare Q to cluster K_m; */ search most similar images from cluster.
Step 4: Compare Q; */ search similar images to closest cluster;
*/ compare min distance between query image feature and cluster image feature.
Step 5: report similar images.
end;

3.2 Preprocessing

Image preprocessing stage can considerably increase the reliability of visual inspection of medical image.

a. Input CT Image: In this CT lung image is taken as input image. First image is resized from 512 by 512 to 256 by 256 and gray scale image is extracted. In this research, [26] the (LIDC) lung image dataset is considered for diagnosis process. It contains 1000 patient's lung images and each size has 512*512. The real time 500 samples of the lung images are selected for testing the diagnosis process of the lung cancer based on application of CBMIR.

b. Median filter: Filtering is a technique for modifying and enhancing an image [12] and suppresses unwilling distortions or enhances some image feature. It is also used for improving image and removes small variations of image available in form salt and pepper noise. Median filter is a suitable method for smoothing out image and remove noise, distortions or small image artifacts available in the image.

c. Segmentation: Medical image segmentation method is referred to as the image partitioning method where image is classified into distinct regions [17] thereby grouping together with neighborhood pixels based on some predefined similarity criteria.

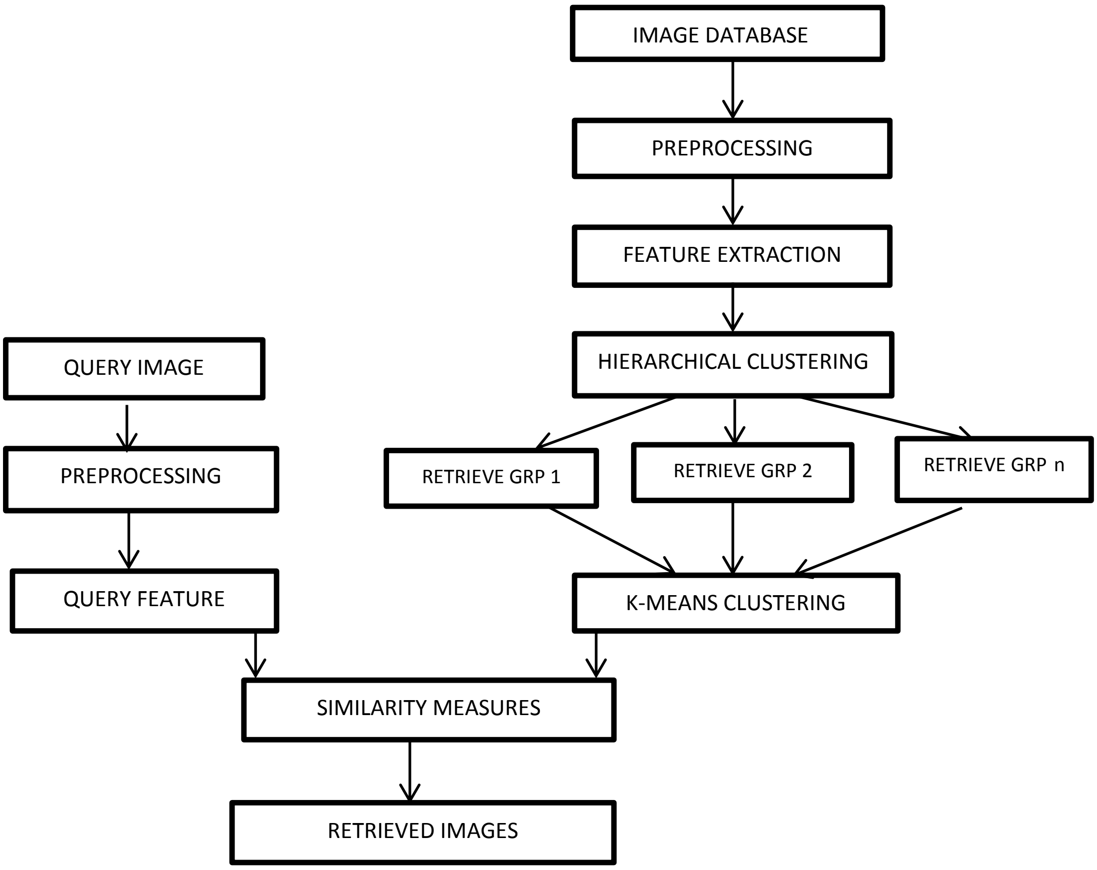


Figure 3: Block Diagram of proposed CBMIR System

3.3 Segmentation

In medical imaging field, segmentation is important for feature extraction, image measurements and also useful to classifying image pixels into anatomical regions such as tumors, tissues, bones and blood vessels etc. the segmentation method is processed to image as an important tool for image processing. The main aim of the lung segmentation process is to separate out regions corresponding to its CT scan slices from the surrounding area of the lung anatomy. To extract lung out of the whole CT image body proposed a technique that utilize the iterative optimal thresholding process. The region of interest (ROI) [24] of image is segmented out from the whole lung body using optimal thresholding with image morphological operation. This method basically divides into three sections such as thresholding, morphological operation and region filling approach which is shown in Figure 4.

a. Thresholding: It is converts the gray scale image into binary image and calculates the threshold value of image that divide the image into two parts first is region of interest and second is background, for ROI extraction it uses high contrast CT volume image since [25] it is easier to extract and distinguish the lungs out the remaining body. The initial result of thresholding method is imperfect because it is considered the exterior part of the lung body.

b. **Morphological operation:** It is used to remove exterior part of lungs inside and outside of lung body in the form of small lung regions [24] and edge detection method which allow enhances the borders of the lungs

c. **Region filling approach:** This process is used to fill the excluded region part [24] of the lung image and removes the undesirable part and low intensity region from the image.

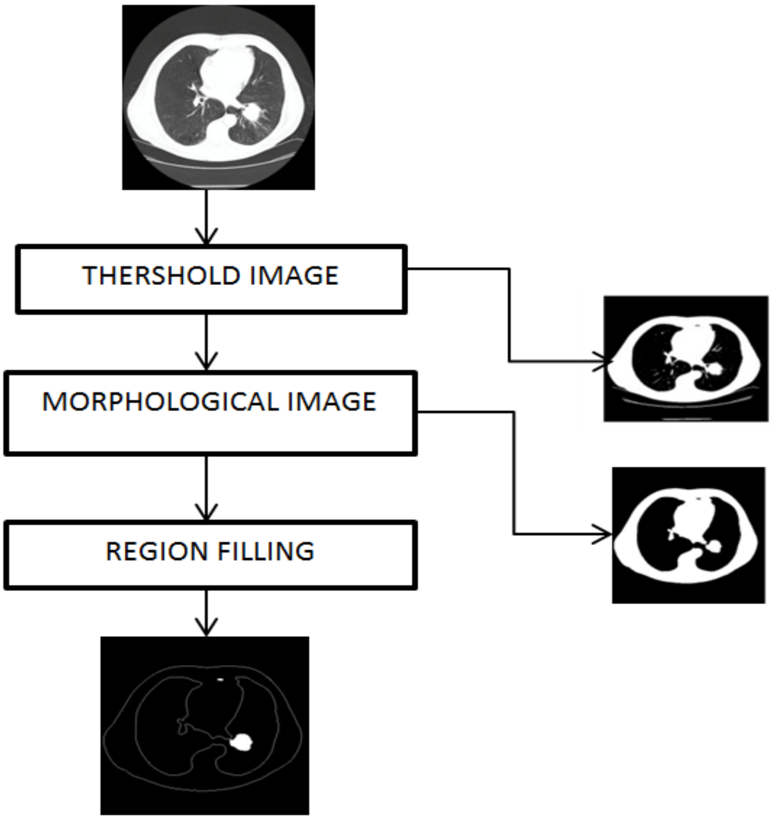


Figure 4: Segmentation of lung CT Image

3.4 Feature Extraction

The visual feature of medical image [19] is analysis by texture feature of image. It comes under the statistical approach for statistical measures of pixel value**.** It contain the information about the structural arrangement of surface such as cloud, bricks etc. and also represent the relationship between the surface and its surrounding environment. The statistical methods are used for analysis the gray level spatial distribution of image. It is performed and computing by the (GLCM) gray level co-occurrence matrix for texture feature extraction thereby it calculates the pair of pixel with specific value and also specified the relationship when creating the GLCM. The number of statistical method [20] based on GLCM computation. This function describes how a specific gray level pixel i value related to j gray level pixel value. GLCM features are extracted using one distance and four directions {0, 90, 180 and 270} thus calculate the second order method when it is define by GLCM matrix. This method calculates [6] the conditional joint probability and all pair of gray level of image based on two methods including inter pixel distance(δ) and alignment (θ).

$$P(x) = \{Pij \mid (\delta, \theta)\}$$

Pij = the co-occurrence probability between gray level i and j.

Where Cij is count the number of times of pixels occurrence and its neighborhood pixels where F(y, z) = i and F(y+1, z+1) =j.

$$\textbf{Entropy} = \sum_{i,j} p(i,j) \log p(i,j)$$

$$\textbf{Contrast} = \sum_{i,j} (i-j)^2\, p(i,j)$$

$$\textbf{Mean} = \sum_{i,j} p(i,j)$$

3.5 Clustering

Image classification and clustering is used to categorization image dataset for speeding up image retrieval process for large databases and improving the efficiency and accuracy of system. Image clustering process basically depends on similarity measures [6] of image and performing image automatic annotation. Image clustering [18] process basically depends on image categorization and similarity measures method and they are together formed the efficient retrieval process. They play an important role in the domain specific application that makes clustering as intelligent support system. On the other hand order it improve the performance of the retrieval system by comparing the similarity between classified texture image and query image which is made by user. For reducing the searching time and scalable duration for image retrieval from large dataset here we have used the combined approach of k-means and hierarchical clustering. The both approach are frequently used in this literature for pattern recognition. In the hierarchical clustering where similar images are grouped into cluster and k-means clustering is iterative refinement algorithm that iterates the process until the cluster is converged. Both methods come under the pattern recognition literature.

a. Hierarchical clustering algorithm: It is provide the better organization of image dataset. It is build a binary tree of the data that merges the similar groups of point and present the hierarchy of similar type of groups which are generated by hierarchical structural tree in Figure 5(a). It is merges the similar instance and maintain the set of clusters. The result of hierarchical cluster [8] analysis is created and explained by a dendrogram or tree approach. It is an effective method to classify the relevant phrases and extract the abstract feature of image.

Algorithm: Input X= $\{x_1, x_2 \ldots\ldots x_n\} \in R^m$ and distance is measured by
D = $R^m * R^m \rightarrow R$

Step 1: Establish cluster $C_i = \{x_n\}$ let $C_i = \{C_1, C_2 \ldots Cn\}$.

Step 2: While $|C| \neq 1$ do
For all pair of all clusters $< C_i, C_j \neq 1 \in C * C >$ calculate $<C_i, C_j>$

Step 3: For best $<C_i, C_j> = \forall \{C_K \neq i, C_l \neq k \in C * C\}$
Where $[D(C_i, C_j) \leq D(C_k, C_l)]$

Step 4: Let $C^{new} = (C / \{C_i, C_j\})$
Update $C = C^{new} \cup C_{ij}$.

b. K-means clustering algorithm: it is partitioning based method for cluster analysis which is extensively used in data mining applications. The k-means clustering is iteratively heuristic partitioning based approach where it iterates the process several times until the clusters are converged. It is simple and efficient approach for grouping the texture feature [6] of images of dataset into k clusters $\{C_1, C_2, C_3 \ldots\ldots C_k\}$ in Figure 5(b), where each have N_k patterns they are related to cluster centers M_k with minimum cost function D^2_k such as.

Centroid (M_k) = $\frac{1}{N_k} \sum_{k=1}^{k} X$

Distance (D^2_k) = $\sum_{k=1}^{k} ||X - M_k||$

Algorithm: Input X $=\{x_1, x_2 \dots x_n\} \in R^m$ number of clusters

Step 1: Initialize K cluster centroids $u_1^{(0)}, u_2^{(0)} \dots u_k^{(0)} \in R^m$, set t=0;

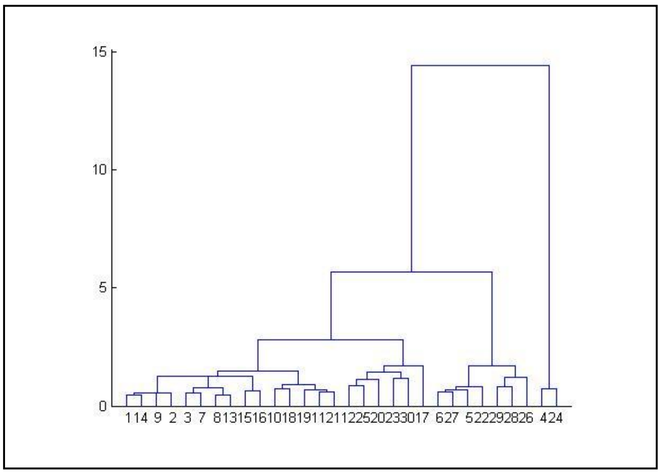
Step 2: For all i $\in x_i$ to cluster g_i

Where $g_i^2 = \sum_{j=1}^{k} ||X_i - U_j||$ and update $C_j = \{g_i = j\}$

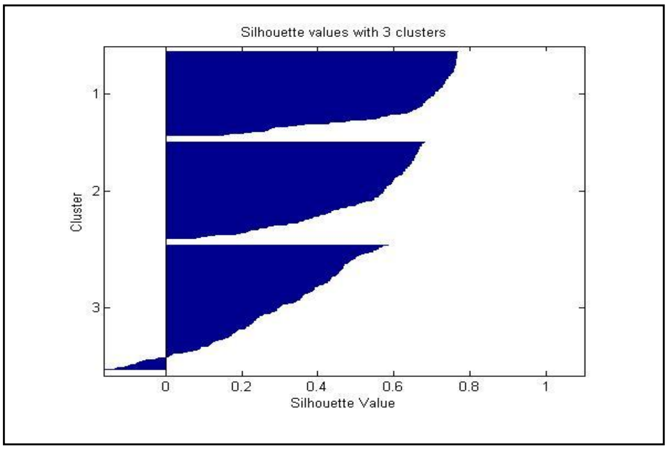
Step 3: Update centroid $U_j^{(t+1)} = \text{mean}(\{X_i: X_i \in C_j\})$

Step 4: Repeat step 2, until no change requires.

Step 5: Partitioning $C_1, C_2 \dots C_k$.



(a)



(b)

Figure 5: Clustering (a) Hierarchical Cluster Tree (b) K-means Cluster

3.6 Image similarity measures and retrieval

Searching and browsing of images from large dataset is a challenging task and this task depends on texture feature of image. It is search based engine basically used for calculate similarity between query image and dataset image [3] and it is ranked the images by sorting their similarity. According to this it is basically depends on two methods.

a. Find the confined cluster: It is first check the distance between the cluster image [6] and query image. Distance is defined by compare the similarity between the mean of cluster image and mean of query image and here the mean of the clustered image is referred to as the mean of all images which are available in the cluster. Such a process is repeated at every cluster and estimate the closest cluster which has minimum distance to query image. The distance is calculated by Euclidean method such as.

$$\mathbf{D}(\mathbf{x_i}) = \sqrt{(\boldsymbol{w} - \boldsymbol{x_i})^2}$$

Where $\mathbf{D}(\mathbf{x_i})$ = distance between cluster image and query image, $\mathbf{w}$ = mean of query image $\mathbf{x_i}$ = mean of cluster image.

b. Find the similar images: In this method the feature of query image is compared with feature of cluster images and it computes by Euclidean method which is calculated the minimum distance. It is retrieved the top most 15 images and located according its mean value where signature of images are close to query image. The similar images are reported.

4. RESULT ANALYSIS

There are many distance method have been created for measure the performance of the system since the estimation of the retrieval performance [3] of the medical images is a critical problem. The retrieval efficiency is measures by most common method such as precision and recall which has calculated the precision and recall values with five sample images shown in Table1. It has achieved the high precision value at 67% and usually presented by the precision and recall graph in Figure 6.

a. **Precision** = $\dfrac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ images\ retrieved}$

b. **Recall** =

$$\frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ relevant\ images\ in\ the\ dataset}$$

Table 1: Precision and recall value in (%)

Query Image	Precision (%)	Recall (%)
Image 19	**60**	**5.7**
Image 94	**53.33**	**6.2**
Image 138	**60**	**4.5**
I Image 278	**66.66**	**6.5**
Image 400	**60**	**7.3**

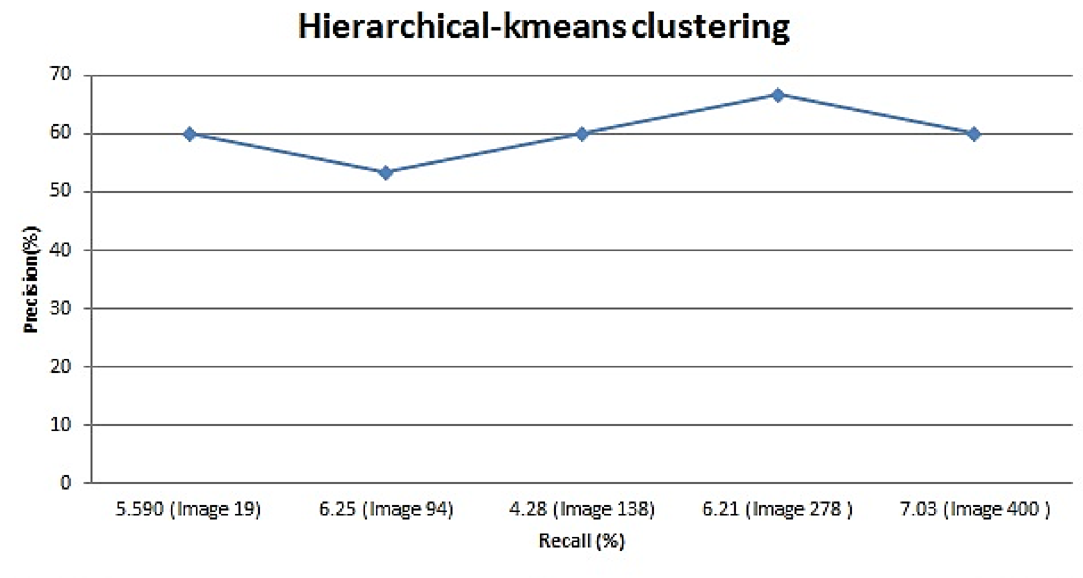


Figure 6: Graphical representation of precision and recall

4. CONCLUSION

The main aim of this study is present a simple and efficient approach for searching and retrieving the lung CT images from the large medical image dataset**.** The combined approach of hierarchical and k-means clustering provide the precise and proficient image retrieval system and computes the efficient result than other clustering technique. The content based medical image retrieval system with clustering can evaluate the fast image retrieval system. Matlab image processing tool box with workspace is used with 500 real time lung CT scan images for testing and implementing proposed the CBMIR system.

5. FUTURE ENHANCEMENT

The performance of the content based medical image retrieval system will be enhanced the system further optimized by combining the various techniques for give better performance and better result in minimum time with efficient accuracy. This system can be used in future to classify another type of medical images in order to predict right disease for proper diagnosis.

6. ACKNOWLEDGEMENT

The authors would like to thanks Dr. Christopher Nimsky from the university of Marburg and Siemens healthcare. Special thanks to his online support for providing LIDC-IRDI dataset of lung CT scan image. This dataset is freely available and easily accessible at http:// www.cancerimagingarchive.net for testing CBMIR system.

7. REFERENCES

[1] Kannan, A., Mohan, V., & Anbazhagan, N. 2010, December. Image Clustering and Retrieval Using Image Mining Techniques. In *IEEE International Conference on Computation Intelligence and Computing Research,*2.

[2] Chen, Y., Wang, J.Z., & Krovetz, R. 2005. CLUE: Cluster Based Image Retrieval By Unsupervised Learning. *IEEE Transaction on Image Processing,* 14(8):1187-1199.

[3] Pilevar, A.H. 2011.CBMIR: Content Based Image Retrieval Algorithm for Medical Image Databases, *Journal of Medical Signals and Sensor,* 1(1):12.

[4] Hassaballah, M.,Abdelmegeid ,A.A. & Alshazly, H.A.2016. Image Features, Detection and Descriptors and Image Matching. *Image Feature Detectors and Descriptors ,Springer International publishing Switzerland,* pp.11-45.

[5] Bhagat, A., Atique M. 2014, March. Web Based Image Retrieval System Using Connectedness Image Segmentation and Geometric Moments. In *IEEE Computational Science and Computational Intelligence (CSCI), International\Conference on* **1**:208-214.

[6] Ramamurthy, B., Chandran, K.R. 2012.Content Based Medical Image Retrieval with Texture Content Using Grey Level Co-occurrence Matrix and K- Means Clustering Algorithm. *Journal of Computer Science,* 8(7):1070.

[7] Dharamrajan, A.,Velmurugan, T. 2015. Lung Cancer Data Analysis by K-means and Farthest First Clustering Algorithms. *Indian Journal of science and Technology*, 8(15).

[8] Zang, C.L., Huang, S.,Xue, G.R.& Yu,Y. 2006,January. Image Description Mining and Hierarchical on Data Recording using HR. tree. In *Springer Berlin Heidelberg ,Asia- Pacific Web Conference ,pp. 379-390.*

[9] Khadosker A.A., ladhake S.A. 2014. Image Mining: An Overview of Current Research. In *IEEE, Communication Systems and Network Technology (CSNT), 2014, Fourth International Conferences, pp.433-438.*

[10] Kishore, M.R. 2015.An Effective and Efficient Feature Selection for Lung Cancer Detection. *International Journal Computer Science and Information Technology* (*IJCSIT)* **7**.

[11] Lam, M., Disney, T., Pham, M., Raicu, O., Furst, J. & Susomboon, R. 2007, March. Content Based Image Retrieval for Pulmonary Computed Temography Nodule Image. In International Society for Optics and Photonics Medical Imaging ,pp. 65160- 65160N.

[12] Aggarwal, P., Sardana, H.K., & Vig, R., 2010, March. An Efficient Visualization and Segmentation of Lung CT -Scan Images for Early Diagnosis of Cancer. In *National Conference on Computational Instrumentation(NCCI)*.

[13] Shyu, C.R., Brodely, C.E., Kak, A. C., Kosaka, A.,Aisan, A.M. & Broderick, L.S.1999. ASSERT: A Physician-in the loop Content-Based Retrieval System for HRCT Image Databases*. Computer Vision Image Understanding*, **75**(1): 111-132.

[14] Theis, C., Guld, M.O., Fischer B., & Lehmann, T.M. 2004, September. Content Based Queries on the Cas Image Database with in the IRMA Framework. In *Springer Berlin Heidelberg, Workshop of the Cross-Language Evaluation Forum for European Languages, pp.*781-792*.*

[15] Petrakis, E.G.M., Faloutsos, C., & Lin K.I., 2002. Image Map: An Image Indexing Method Based on Spatial Similarity*. IEEE Transaction on Knowledge and Data Engineering,* 14(5): 979-987.

[16] Hsu, W., Antani, S., Long, L.R., Neve, L. & Thoma, G.R. 2009. SPRIS: A Web Based Image Retrieval System for Large Bio Medical Databases. *International Journal of Medical Informatics,* 78: S13-S24.

[17] Sharma, D., & Jindal. 2011. Computer Aided Diagnosis System for Detection of Lung Cancer in CT- Scan Images. *International Journal of Computer Science and Electrical Engineering,* 3(5):714.

[18] Haiwai, P., Li, J., & Wei, Z.,2006. Medical image clustering for intelligent decision Support. In *Engineering in medicine and biology Society ,2005. IEEE-EMBS. 27th Annual international Conference of the IEEE,3308-3311.*

[19] Surpreethi, K.P. Medical Image Retrieval using Visual and Semantic features.

[20] Yadav, N. G.. "Detection of lung nodule using content based medical image retrieval". International Journal of electrical, electronics and data communication, ISSN (P), 2320-2084, 2013.

[21] Lehmann, T. M., Gold, M. O., Thies, C., Fischer, B., Spitzer, K., Keysers, D.,. & Wein, B. B.. "Content-based image retrieval in medical applications." Methods of information in medicine, 43(4), pp. 354-361.

[22] Travis, W.D., Brambila, E., Nicholson, A. G., Yatabe, Y., Austin, J.H., Beasley, M.B., Geisinger , K., 2015. The 2015 World health Organization Classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *Journal of Thoracic Oncology,* 10(9), pp.1243-1260.

[23] Abubakar, F.M., 2013.Study of image segmentation using thresholding technique on a noisy image. *International journal of science and research (IJSR),* 2(1), 49-51.

[24] Jafar, I., Ying, H., Shields, A.F.,& Muzik, O.2006.Cmputerized detection of lung Tumors in PET/CT images. In Engineering in Medicine and Biology Society,2006. EMBS'06.28 *Annual international conference of the IEEE,* pp. 2320- 2323.

[25] Shiying Hu, Eric A.Huffman, and Joseph M. Reinhardt, "Automatic Lung Segementation for Accurate Quantitiation of Volumetric X-Ray CT images", IEEE Transactions on Medical Imaging, vol. 20, No. 6, June 2001.

[26] C. Zeng et. al, Development of a Data Integration and Visualization Software for LIDC, J. Softw., vol. 8, no. 9, pp. 2297-2304, 2013.

[27] LIDC-IDRI, Cancer imaging archive, [online] https://wiki.cancerimaging archive.net/display/Public/LIDC-IDRI.