Content Based CT Image Sign Retrieval using Fast Discrete Curvelet Transform and Deep Learning

R. Biswas[1], S. Roy[2]

[1] Department of Information Technology, Ramkrishna Mahavidyalaya, Tripura, India, ranjitit1984@gmail.com

[2] Department of Computer Science and Engineering, Assam University, India, sudipta.it@gmail.com

ABSTRACT

In the world, one of the most critical diseases is lung cancer that leads to death of almost all affected human beings due to uncontrolled growth in the cell. These abnormal cells grow rapidly and divide to form tumor in the lungs. For lung cancer detection, the CAD system divided in four parts in CT images, such as noise removing process, segmentation of lung, detection of lung nodule and classification. The Visual information of similar nodules helps radiologists to detect the disease. This paper contains the Content-Based Image Retrieval (CBIR) system which is used for nodule retrieval in lung CT images. The CBIR system of pulmonary nodules retrieval system consists of pre-processing, feature extraction, feature selection, retrieval and classification. Our proposed method used manual cropping for segmentation phase to extract Region of Interest (ROI). The Gray-Level Co-Occurrence Matrix (GLCM) is used to extract features and the extraction of feature is done by Fast Discrete Curvelet Transform (FDCT). To select the perfect features, enhanced moth flame optimization algorithm is used and other best features are filtered by Deep Neural Networks (DNN's). We used the Euclidean distance to retrieve similar lung nodules from ROI database. The proposed method has been tested on the LISS database. Finally, we have achieved the content based image retrieval from ROI image database.

Key words: Content-based image retrieval, Curvelet Transform, Lung nodule.

1. INTRODUCTION

Now-a-days the medical image is most essential, and the medical imaging database has more content, quantity and dimension. The hospitals and clinics are generating multiple numbers of lung images day by day and these CT images are saved in Picture Archival and Communication System (PACS) [1]. This lung image contains computed tomography (CT) X-ray, ultrasound, etc. There are two approaches in image retrieval process such us Text-Based Image Retrieval (TBIR) and Content-Based Image Retrieval (CBIR). The important challenge is to efficient access, search and retrieval in this large lung nodule database [2]. To rectify this issue of lung image, the Content Based Image Retrieval (CBIR) method is introduced. To achieve the retrieval of lung nodule in separated lungs area, it is segmented in further to extract the exact ROI. The ROI mainly defined boundaries of a tumor and measured its size [3]. The CBIR mainly depends on the accurate extraction of visual features and it utilized visual contents of an image. The visual

features divide into two parts *viz*., Global descriptors and Local descriptors. The global descriptor defines the whole image visual features and local descriptor defines the object or region visual features. The texture based features works at important in visual features. The Proposed algorithm has efficient determining of similarities and it is done through pre-processing, segmentation of lung region, feature extraction from this segmented region, feature selection, retrieval and classification.

This paper proposed the Fast Discrete Curvelet Transform (FDCT) for efficient extraction of the feature. In image analysis, the gray level co-occurrence matrix (GLCM) features are used. Our proposed FDCT method reduces the loss of information in CT images. Figure 1 illustrates the basic block diagram of content based image retrieval using curvelet transform.
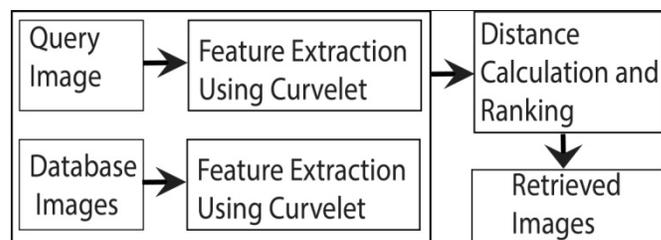

Figure 1: Image Retrieval using Curvelet Transform

The rest of the paper is structured in the following manner: Section 1 defines the Content-Based Image Retrieval (CBIR) in pulmonary nodule. Section 2 contains the various surveys about content base retrieval system of CT images in image processing, Section 3 describes the proposed method, ROI decomposing and feature extraction, classification and retrieval algorithms for retrieval of pulmonary nodule from CT images, and Section 4 present the experimental results and comparative analysis. Finally, in section 5, conclude the work with future directions.

2. RELATED WORKS

This section provides a discussion of variety of surveys about content base retrieval system of CT images in image processing. Lam et al. [4] used the BRISC framework with various measures, and it's determining the relationship of two nodules. This paper contains lung image database consortium (LIDC) database, the similar nodules are collected in this database. It consist three types of feature

extraction methods, such as Gabor filters, Haralick co-occurrence, and Markov random fields. It acts better for recover the identical nodules, after that the next process Haralick co-occurrence techniques will done to increase the accuracy. In this paper, defines two groups of CBIR medical systems: 1. retrieving the overall basic structures, 2. Retrieving the abnormalities or pathologies of anatomical structure. It mainly focused pulmonary nodules more than entire lung images. Agarwal et al. [5] proposed the content based retrieval for determining lung nodules from the (CT) Computer Tomography images. In CAD system, this paper explained various phases, such us lung region feature extraction, segmentation, feature selection, and classification. In this paper contains the Support Vector Machine for learning purpose.

Deep learning based algorithms has been proposed for content-based retrieval [6], [7], where different feature extraction methods were implemented. The content based retrieval mainly depends on similarity measurements and feature representation. Here, in deep learning the deep belief network (DBN) method is defined for feature extraction and classification, because this method generates large number of data. This DBN extracts large amount of data for learning and provides best classification. This paper contains the disadvantage, that didn't use some more good algorithms for the reason of computation time. The introduced method contains the simulation for testing and evaluation results proved the performance in high positive deviation. In analysis, they explained how to improve the proposed method and how to apply for CBIR method. The dataset contains 1000 images, so the accuracy rate would be 98.6%. The main focus is to reduce the time complexity requirement. Qayyum et al. [7] proposed the Convolutional neural network system for classification of Interstitial Lung Diseases (ILDs). The evaluation results proved the performance of classification as 85.5% in lung patterns characterizing and achieved the average precision 0.69. This paper contains also Boltzmann machine based method for convolutional classification. Later on Bhende et al. [8] proposed novel image based retrieval system for retrieving CT images. Here introduced CT (CBCT) latest method for content-based image retrieval. The training contains two important phases such us querying, database construction, and clearly explained these two techniques. The training of database construction consist three main steps 1. Analysis of 3D symmetry for segmentation of lesions, 2. Embedding of 3D symmetry characteristics, 3. Pyramid matching for feature extraction. The evaluation results proved the performance with some measures such us mean average, precision, Dice coefficient and normalized value. The quantitative results proved the proposed method had high efficient and effective also explains CBIR in clinical CBCT applications. In this CBIR Normalized Discounted Cumulated Gain (NDCG) was obtained. Chung et al.[9] proposed the deep learning representation for Content-Based Image Retrieval. This paper contains the two steps; first step is to classify the query according to the label of image. The

second step is retrieving of most identical images. It introduced new algorithm named FSSEM (feature subset selection using expectation-maximization clustering) for improve the retrieval precision. Shamna et al. proposed in [10] the topic and location model for retrieving the image from database. . The text-based retrieval was directly proposed to (DICOM), Digital Imaging and Communications in Medicine attributes. This paper clearly explained the Topic Modeling for retrieving hidden data from large database. CBMIR mainly consist visual features; it defines the determination of similarity of images and it classifies low-level visual features. The similarities divided into inter-class similarity and intra-class dissimilarity. The Topic Models developed from Bag of Words (BoW) concept, and it represents traditional document of words. The Bag of Visual Words (BoVW) means histogram visual words of image representation. The important method in topic modeling is Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA), and Probabilistic Latent Semantic Analysis (PLSA). This paper clearly explained about these methods and applications. Mehre et al. proposed in [11] the optimal set of feature and membership-based class Retrieval for retrieving. This paper contains the CBIR system for lung nodules with learning and optimal features to improving retrieval performance. The limited Redundancy with high Relevance method was used for feature set selection. The simple distance based retrieval (SDR) method was used to improve the precision performance. The conventional classifier-based retrieval (CCBR) method was used for improving classification performance. This method gives the hundred percentage of precision rate for right classification. This paper proposed class membership-based retrieval (CMR) technique for texture based retrieval. The texture features contains margin and shape features. The proposed method examines with two date set such us public database LIDC/IDRI. The CMR based retrieval system provides the best performance compared than CCBR and SDR in optimal features. In zero precision, the standard deviation of precision and average precision also calculated. The fusion of medical knowledge in Content-based medical image retrieval (CBMIR) for lung nodule classification [12] . It contains the two databases such us LIDC and NCI. The lung nodules classification is Malignant, Benign, and Metastasis. The experimental results proved the measures of average precision, accuracy and specificity is 92.8%, 88.6% and 89%. And proved the parameters, it improved the performance

Pang et al. [13] used deep preference learning method for medical image retrieving. Here the deep learning method was proposed to extract the compact features and maximum level of features. This deep feature extraction process applied on hidden layer of high resolution image, and represents multiple layer of abstraction and improved the retrieving and indexing of medical images. This paper contains multi layered DNN's (Deep Neural Networks) named convolutional neural networks (CNN), Stacked Denoising Auto Encoders (SDAE).They introduced learning method for

finding the similarity images. For this training, the preference learning technology was used. It explains, efficient biomedical image indexing algorithm for detailing images. This proposed method had high efficiency and high indexing ability for retrieving biomedical images. In [14], [15] proposed the new pattern that is local ternary of quantized extra pattern for CT image retrieving & indexing. It defined as (DLTerQEP). This pattern was divided into standard local ternary patterns (LTPs) and local binary patterns (LBPs). And the proposed method used a ternary pattern, its starts from horizontal–vertical–diagonal–antidiagonal (HVDA7). The evaluation examines in two types of medical database such as LIDC-IDRI-CT, VIA/I-ELCAP-CT images and OASIS-MRI brain Magnetic Resonance Imaging (MRI) image and results proved with some measures named Average Retrieval Precision (ARP) and Average Retrieval Rate (ARR). The Local directional of ternary quantized extrema pattern is retrieving spatial relation with pair of neighbors. It captured more dimensional information compared than LBP. In proposed approach the GLCM feature was used for mapping pixels. Wei et al. proposed in [16] the Deep Convolutional Neural Network for determining the similarity in medical images. It rectifies the acceptable gap between minimum levels of visual information in devices. The proposed framework was used for training the classification of medical images. The dataset contains classes with twenty four and modalities with five, and then collected from various areas of medical imaging in target areas. These dataset are used for learning the medical image. The neural network was trained these datasets successfully. After the features are learned, the classification is used to retrieve the medical image.. The proposed CNN's framework is compared with some traditional framework in classification accuracy, average precision. Terasa et al. [17] used Wavelet Transform methods for retrieving content based image. This paper explained proximity measures; it calculates similarities between the medical images and improves the similarity search for CBIR. There are two techniques used to improve the search similarity between images such us Discrete Wavelet Transform and Convolutional Neural Network. This paper focused two phases such us feature representation and similarity measures. In extraction the minimum level of characteristics are shape, color and texture. For feature representation the conventional neural network was used. The proposed retrieval algorithm reduces the computation time and increase user's interaction.

Varish et al [18] used gray level co-occurrence matrix descriptors for discrete cosine transform, and it is based the residual image for image retrieving. In this paper, the feature extraction was done by block level Discrete Cosine Transformation (DCT) method, so easily extract the visual content of images. Alternatively other visual features are extracted. This paper explains the combination of DC coefficients and GLCM features in more efficient. This GLCM features are efficient and effective and it constructed using Discrete Courier Transform (DCT) of image plane.

Testing is done by using RGB color space and experimental results are proved the effectiveness with standard two two image databases. The comparative results are proved the efficient performance of proposed system with historical CBIR schemes. This paper mainly focused DCT information and GLCM parameters of proposed image. The DC and GLCM features are normalized in effective manner. Paper have advantage of minimum searching time and maximum image retrieval results. In another work Huneiti et al. [19] proposed the Discrete Wavelet Transform and Self Organizing Map (SOM) for efficient image retrieving. Retrieving mainly depends comparing of important features without comparing whole image features. Proposed CBIR method extracts the features like color and some textures using Discrete Wavelet Transform. The Euclidean distance was used to compares the identity measures and texture features. Using Self Organizing Map (SOM), the other relevant features are retrieved. The extracted images are converted using DWT and mean value for each block of pixels.

Thereafter, to achieve better retrieval results, Zhang et al. [20] proposed the DWT-DCT frequency domain based image retrieval. In this paper, initially in frequency domain the biomedical image was encrypted and after that feature was extracted. In encrypted medical image database, the encrypted image and eigen vector are uploaded. To represent the two encryption medical image, the Normalized correlation Coefficient (NC) was used. The evaluation results are proved the introduced algorithm had high robustness across geometric attacks and conventional attacks. Paper explained the Wavelet transform and four parts of transformation named, low resolution, vertical (LH), horizontal (HL) and diagonal (HH). They explained Peak Signal to Noise Ratio (PSNR) feature vector and it compares with NC feature vector. It explicated Gaussian Noise, Median Filter, Rotation Attack, Scaling Attack, Translation Attack and Cropping Attack. Shinde et al.[21] proposed the fast discrete curvelet transform for biomedical of image retrieval. In this paper, the curvelet transform was applied in image the feature vector was determined. In this work contains three types of database such us series of Open access, Emphysema-CT & NEMACT. The experimental results proved the proved the accuracy of proposed method with other traditional approaches. The proposed method had the advantage of less computational time. Sarala et al. [22] proposed an efficient region based image retrieval method in a hierarchical manner using color and texture features for image retrieval. Kumar and Ramanaiah proposed a ROI extraction method based on lossless compression for compression of different type of medical images [23].

3. PROPOSED METHOD

This section describes the proposed CBIR system for extracting pulmonary nodules from CT images, which consists pre-processing, ROI segmentation, feature extraction, feature selection, retrieval and classification. The detailed flow diagram of the proposed method is shown in

Figure 2 and different steps are described in the following sections.

3.1 Pre-processing

It provides better visual information and reduces the noise in lung CT images. It's most difficult to segment the lung region, without using pre-processing. The focus of this method is to delete the unwanted region for further processing. It upgrades some picture highlights significant for further processing. Using adaptive median filter the noise reduction is done and histogram equalization is used to improving the lung image quality. We have mentioned the enhanced lung image attained by applying histogram equalization in figure 3.

3.2 Segmentation

The segmentation of region of interest (ROI) is one of the important steps to reduce the search space for locating lung nodules. The image is manually segmented using coordinate information provided by the radiologist for simplifying segmentation task. The main aim of the lung segmentation process is to separate out regions corresponding to its CT scan slices.

A. ROI Extraction

The nine categories of lung images are obtained from LISS database [24] such as grand grass opacity (GGO), lobulation, calification, cavity & vacuolous, spiculation, pleural dragging, bronchial mucus plugs, air bronchogram, obstructive pneumonia (OP).

In ROI extraction, the lung image needs to crop ROI manually based on given co-ordinates information. This process is used to fill the excluded sign region part of the lung nodule CT image and deletes the undesirable part with low intensity region from the image. The nine categories of database involves with smaller rectangular bonding boxes. After cropping the single sign region in rectangular bounding box indicating the lesion region, we have to resize them in 48×48 pixels using cubic Interpolation method.

B. Co-efficient Extraction

After segmentation process the feature extraction done, using Fast Discrete Curvelet Transform (FDCT). It's used to extract the co-efficient from images. This frequency co-efficient are used to extract feature vector and these vectors are reduced for further processing. Generally curvelet represents the multi-scale object representation. In 48x48 ROI image, based fast discrete curvelet transform decomposition we have to select the correspond features. From decomposition the scaled and oriented curvelet components are achieved. For our 48x48 ROI signs image we have decomposed using 4 levels curvelet transform and curvelet coefficients are computed.

Fast Discrete Curvelet feature extraction

The curvelet features are considered for improving object determination and it's based on wrapping of Fourier samples. The curvelet transform is one of the multi-scale pyramids

with various directions and positions. The curvelet transform function is defines by equation (1).

$$c(j,l,k) = [f,\emptyset\_(j,l,k)] \tag{1}$$

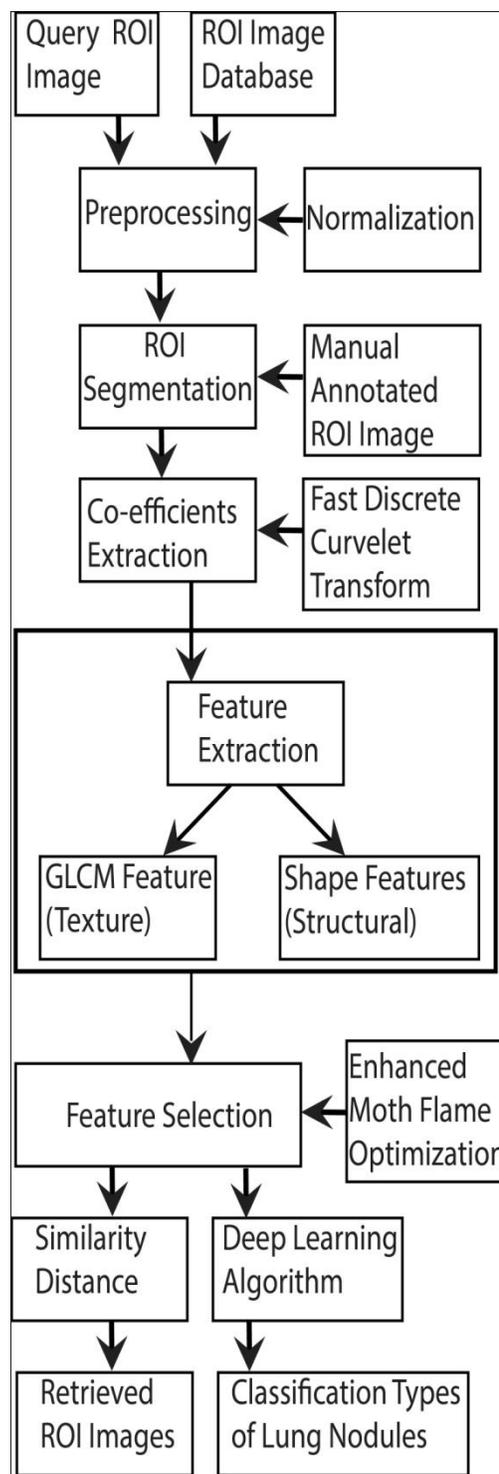Where, j,l,k defined parameters of scale, direction and position.

Figure 2: Flow diagram of Lung CT Image Retrieval.

In fast discrete curvelet transform, input is given by feature extracted image and output will get as co-efficient extraction.

Generally the curvelet is used to represent the edges in objects of lung image. First, we initialize the data structure with shift condition. The image is decompose into set of bands, so we find the pyramidal scale decomposition using M_{pyr1} and M_{pyr2}. the pyramid is used to produce various level of sub-bands. Every sub band is smoothly windowed into "squares". Then we find the smooth periodic extension of high frequencies using mentioned equation. After that we compute the window scale and wrapping window to get the efficient curvelet co-efficient. The wrapping mainly consists periodizing and re-indexing the windowed frequency data. Then we compute the normalization for x co-efficient and y coefficient using this wrapping functions. Then we find the low-pass filter with mentioned condition. After scale decomposition we find angular decomposition for horizontal band, vertical band, left and right bands. Finally we find the curvelet level using $curvelet_{wav}$=fftshift(ifft2(L_{wed}))* $\sqrt{sixe(roi_{ex})}$ function with minimum transform.

Algorithm 1: Fast Discrete curvelet Transform

Input: Feature Extracted image (roi_{ex})
Output: coefficient Extraction F_{cur}
Procedure:
Let N_s be the Number of scales
To initialize the data structure,
Finest, F_e=3;
Let shift condition,
  Img_{sh}=fftshift(ifftshift(roi_{ex}))/$\sqrt{size(roi_{ex})}$
[N_e N_v]=size(Img_{sh})
Where,
N_e be the Number of edges
N_v be the Number of vertex
To find the pyramidal scale decomposition
  $M_{pyr1}=\frac{N_e}{3}$; $M_{pyr2}=\frac{N_v}{3}$;
To compute smooth periodic extension of high frequencies,
If F_e ==1
  SN_e=2*floor(2*M_{pyr1})+1
  SN_v=2*floor(2*M_{pyr2})+1
Img_{sh}=Img_{sh}(equ_{ind1}, equ_{ind2});
To compute windows length,
w_{len1}=floor(2* M_{pyr1})-floor(M_{pyr1})-1-mod(($N_{e,3}$)==0);
w_{len2}=floor(2* M_{pyr2})-floor(M_{pyr2})-1-mod(($N_{v,3}$)==0);
co_x=0:(1/w_{len1}): 1;
co_y=0:(1/w_{len2}): 1;
To compute wrapping window,
 w_{r1}=zeros(size(co_x))
w_{l1}=zeros(size(co_x))
$w_{r1}(co_x \le 0)$=1
$w_{r1}(co_x > 0)$&(co_x < 1)=exp(1-exp(1-1/co_x)(co_x > 0)&(co_x < 1));
$w_{l1}(co_x \ge 0)$=1
$w_{l1}(co_x > 0)$&(co_x < 1)=exp(1-exp(1-1/co_x)(co_x > 0)&(co_x < 1));
Let find the low pass filter,
L_{pf1}=[w_{r1}, $ones(1,2 * floor(M_{pyr1} + 1))$]

To compute angular decomposition,
n_{qua}=2+2*(nor_r);
for q_{re}=1: n_{qua}
  M_{hor} =M_{pyr2}*(mod(q_{re},2)==1)+ M_{pyr1}*(mod(q_{re},2))==0;
M_{ver} =M_{pyr1}*(mod(q_{re},2)==1)+ M_{pyr2}*(mod(q_{re},2))==0;
If mod(n_{qua},2)
W_t=[W_{t_left} W_{t_rite}]
End
End
To compute left corner wedge,
l=l+1
Let regular wedge,
To compute right corner wedge,
  $slow_{re}$=round(2*floor(4*M_{ver})2*n_{qua} + 1)-
  lcw_{re}(end)/floor(4*M_{ver}))
$midcw_{re}$=floor(4*M_{ver})-floor(M_{ver})+ $slow_{re}$
Let regular wedge,
L_{wed}=floor(4*M_{ver})-floor(M_{ver})
Wr_{data}=$\sqrt{2 * img(roi_{ex})}$
To compute wavelet level,
  $curvelet_{wav}$=fftshift(ifft2(L_{wed}))* $\sqrt{sixe(roi_{ex})}$
End
End

3.3 Features Extraction

The feature extraction acts as important role in lung image retrieval. It transforms the high-dimensional nodule images into a lower-dimensional space with maintaining the essential CBIR. The features are extracted for given query image. The texture related features and shape related features are extracted; the texture related features are represents by Gray level Co-occurrence matrix. And the shape related features are represents by structural attributes.

i)   GLCM (Gray level Co-occurrence matrix)

GLCM method defines with geometric method of texture feature. The texture feature of lung image defines the spatial relationship between the pixels. That means it characterizes the texture related features of lung image by calculating the spatial relation of pixels. It can be used to reverse, the overall average for degree of correlation between pairs of pixels in homogeneity and uniformity. GLCM features are estimated at different angle such as 0^0, 45^0, 90^0 and 135^0. The gray level pairs of frequency pixels divided by distance d and GLCM matrix will calculate by displacement "d". The texture related features are contrast, energy, correlation and homogeneity etc using equations (3)-(9). In proposed method the GLCM is used to find the rough, smooth, silky, or bumpy of the pixel intensities of lung region, here intensity values are represented as gray level. In GLCM Count all the pairs of pixels, the first pixel has the value "i" and second pixel has "j" value. These values are entered in matrix by rows and columns. The elements "i" and "j" normalized by dividing the pair of pixels. The GLCM normalization can found by equation (2).

$$N[i,j] = \frac{p[i,j]}{\Sigma_i \Sigma_j p[i,j]} \qquad (2)$$

Some of the GLCM features

Autocorrelation $= \sum_{x=1}^{R} \sum_{y=1}^{C} (x^* y^* I_{clu}(x,y))$ (3)

Entropy $= \sum_{x=1}^{R} \sum_{y=1}^{C} (I_{clu}(x,y) * \log I_{clu}(x,y) + \in))$ (4)

Sum of entropy $= \sum_{t=1}^{(2*R)-1} -(S(t) * \log(S(t) + \in))$ (5)

$$S(x+y-1) = \lim_{x \to 1 R} \lim_{y \to 1 C} I_{clu}(x,y)$$

Sum of variance $=$
$\sum_{t=1}^{(2*R)-1} ((t+1) - Sum\ of\ entropy))^2$ (6)

Sum of average $= \sum_{t=1}^{(2*R)-1} ((t+1) * S(t))$ (7)

Energy $= \sum_{x=1}^{R} \sum_{y=1}^{C} (I_{clu}(x,y))^2$ (8)

Homogeneity $= \sum_{x=1}^{R} \sum_{y=1}^{C} \frac{I_{clu}(x,y)}{(1+(x-y)^2)}$ (9)

The GLCM the number of gray level images is in same size (2D). GLCM features are extracted using single distance d=1and four directions 0^0, 45^0, 90^0 and 135^0. In our method, we have used 22 GLCM features.

ii) *Shape features*

Shape features acts as powerful feature, that may recognized from the image outline, that means extract the shape. It allows measuring the similarity lung region shapes. It mainly focused to describe the image content. Some of the shape features are edge length and edge smoothness. In lung image shape features are used to retrieving the similar shapes from the database. The shape descriptor acts an effective way to find the defecting region of shape and noisy affected shapes. It achieved to perform lung image retrieval for maximal types of shapes. Some of the shape features are calculated using equations (9), (10) and (11).

Area $=$ bwarea (I_{clu})   (9)

Perimeter $=$ bwperim (I_{clu})   (10)

Smoothness: $S = 1 + \frac{1}{1 + \frac{\sum (I_{clu} - \mu)^2}{N}}$ (10)

3.4 Feature Selection, *classification and retrieval*

From extracted features the irrelevant features are deleted and obtain the limited features. And it is done by Enhanced Moth Flame Optimization algorithm. Then detect the various types of lung nodules by deep leaning algorithm, where ROI segmented results are used. It is the machine learning algorithm for classification and regression. It's used to determine the pulmonary nodule with respect to other independent lung region.

Algorithm 2: Enhanced Moth Flame Optimization
Input: Feature Extracted F_{ext}
Output: optimized Sequences f_{fn}
Procedure:
Let N_{fea} be the Number of Features

Let F_{ext} be the Features extracted
To initialize the population of moths,
M_{ite} =100; //maximum iteration
SA_n= size(Tr_{fea}, 1) // number of search agents
l_{bou}=min(min(Tr_{fea})) ////lower bound
U_{bou}=max(max(Tr_{fea})) ///upper bound
di_{bou}=size(Tr_{fea}, 2) // dimension matrix
  N_{bou}=size(U_{bou}, 2) //no of boundaries
If N_{bou} == 1
    M_{pos}=rand(SA_n, di_{bou})*(U_{bou} * l_{bou})+ l_{bou})
End

If N_{bou}>1
 For i=1: N_{bou}
  U_{bou}_i=U_{bou}(i)
l_{bou}_i=l_{bou}(i)
    M_{pos}(:,i)=rand(SA_n, di_{bou})*(U_{bou} * l_{bou})+ l_{bou}
End
End
While iter<M_{ite}+1
    f_{pos}=round(SA_n,- iter)*(iter-1)/ M_{ite}
To compute fitness of moths
 For ii=1:size(M_{pos}(:,i))
    M_{pos}(:,i)=rand(SA_n, di_{bou})*(U_{bou} * l_{bou})+ l_{bou}
End
If iter==1
 $fitness_{sort}$=sort(M_{pos})
End
To compute flames,
 $best_{sort}$=sorted population
End
To compute the sort moths,
 For ii=size(F_{ext}, 1)
 If F_{ext}(ii)~=0
    ∂_n(ii,:)= ∂_n(ii,:)
ii=ii+1
 end
end
    for jj=1:size(∂_n, 1)

 fitensss value,F_n= 20*exp(2*$\sqrt{(M_{pos}^2)}$/ di_{bou} -
exp(sum()cos(2π*M_{pos}))/ di_{bou} +20exp(1)
End
End
 mF_n=min(F_n) ///minimum fitness value
While(mF_n > γ_v)
    $\rho\vartheta_n$= ϑ_n + n // update the best fitness value
If ($\rho\vartheta_n$ < mF_n)
 mF_n =$\rho\vartheta_n$
End
End
End
λ=1
For nn=1:length($\rho\vartheta_n$)
 $\rho\vartheta_n$=∂_n(: ,nn)
 λ = λ+1
end
end

To update the best fitness value,

$\pi\rho\vartheta_n$=rand($£_n$, α_b, β_b)

best solution, $\rho s_n = \rho\vartheta_n.* \lambda$

To update the bound,

$t_b = \rho s_n$ // temporary bound,

$\tau = t_b < \alpha_b$

$t_b(I) = \alpha_b(I)$

$J = t_b > \beta_b$

$t_b(J) = \beta_b(J)$

$\beta_{sol} = t_b$

▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪

In Enhanced Moth Flame Optimization algorithm, the input is extracted features and the output will get by optimized sequences. It's one of the efficient optimization algorithms and it's also population based algorithm. It directly proposed with moths and flames solution. First we find the population of moths in number of iterations. Then we calculate number of search, agents, lower bound, upper bound, dimension matrix and the number of boundaries. In each iteration we compute the pair of moths and flames fitness values. After updating the list of moths, the moths are sorted based on their fitness values. Finally we update the minimum fitness value and update the best fitness value in maximum number of iterations.

Using Convolutional Neural Networks the optimized sequences are classified. Firstly, we initialize the training set size and testing set size (80 % training and 20% testing) to create a list of features set. Then we initialize the CNN parameters, and back propagation is used to get the accurate calculation of derivatives. It contains different layers like Convolutional layer, Pooling Layer, Fully connected Layer, Activation Layer. After calculating these layers the accuracy will find in different cases.

In classification according to the neural networks algorithm, we determined the most similar nodule image using the calculation of distance between the query image and retrieval results. It mainly considers the influence of the most similar lung image.

Algorithm 3: Deep learning algorithm

Input: optimized Sequences f_{fn}

Output: Classified output

Procedure:

C be the Size of the Total sequences to be classified

N_l=3 // no. of layer

N_r=1 // no of runs

Let N_s //no of samples

Let N_f //no of features

Initialize $Training\ Set\ Size$ ($t_{\phi c}$) and

$testing\ Set\ size$ ($t_{\psi c}$)

// 80 % training and 20% testing

Extract features from the sequences and create a list of features set

Let f_{fn} be the feature set extracted

Let initialize CNN parameters,

 Let batch_size =1

No of the epochs=1

Let Lb_s be the labels corresponding to the selected features

Let N_c be the number of classes to be identified

Load f_{fn} //load optimized data

For i=1 to N_l

Split F_{fs} (feature set) into Γ (feature subset)

For j=1 to Γ

To Find back_propagation_cnn,

For i=1:size($t_{\psi c}$)

$N_l . f_{fn} = t_{\psi c}(i,:)$

End

For i=2: N_l

Layer=i

 If $N_l(i) = t_{\psi c}(i,:)$

Val=$N_l(i) * t_{\psi c}(i,:)$

 For j=1:length(Val)

 Z=0;

For k=1: length(Val)

 Kk=kk+1;

Val=$N_l(i-1) * t_{\psi c}(k)$

Val1=$N_l(i) * t_{\psi c}(k(:,:,1))$

End

End

End

 Let f_{fn} is to be feature in $F_{fs(i,j)}$

 Trained T_{CLF} estimates f_{fn}

Let R_{sort}=sort(T_{out}) //rank level

Let accuracy = mean(T_{CLF} ,1)

$T_{CLF} = \sum_i T_{out} / R_{sort}^2$

$cnt_i = \Sigma(f_{fn})$ in belonging to $samples\ N_s$

End

Compute Total count as $Cnt_T = \sum_{i=1}^{N} cnt_i$

Compute probabilistic Components for each class as

For i=1 to N_c

$P_{comp}(i) = cnt_i / Cnt_T$

End

End

Clustering/Indexing

For improving the retrieval efficiency in terms of time, we require to Cluster the similar images using clustering technique. This technique reduces the error without multiple calculations. It mainly used for classification.

At the same time it improves the retrieval performance by compared the similarity between classified texture image and query image. The query image is made by user. The clustering approach makes the iterations in several times till the clusters are combined.

$$Distance(D_k^2) = \sum_{k=1}^{k} \|X - M_k\| \qquad (11)$$

Similarity Measure:

Similarity measure compare the given query feature vector, with all database feature vectors based on Euclidean distance using equation (12).

$$E_d(Q,T) = \left(\sum_{i=1}^{S} (Q_{i-} T_i)^2 \right)^{1/2} \qquad (12)$$

The similarity function measures the similarity between the object template $[E_1, E_2 \ldots \ldots E_n]$ and undermined objects $[E_{i1}, E_{i2} \ldots \ldots E_{in}]$

$M = \min (\Delta E_1 \Delta E_2 \ldots \ldots \Delta E_m)$

M defines the maximum similarity feature vector of object template.

Performance Evaluation:

The performance of the proposed method for CT image lung nodules retrieval is measured using precision and recall.

$$\text{Precision} = \frac{Number\ of\ relevant\ image\ retrieved}{Total\ number\ of\ images\ retrieved} \quad (13)$$

$$\text{Recall} = \frac{Number\ of\ relevant\ image\ retrieved}{Total\ number\ of\ relevant\ images\ in\ the\ dataset}$$
(14)

4. RESULTS AND ANALYSIS

The proposed method is tested on Lung CT Imaging Signs (LISS) database. This section contains the performance results of CT imaging signs retrieval and classification in terms of various performance measures. Figure 3 show the normalization of lung CT image and Figure 4 represent the ROI extraction process.
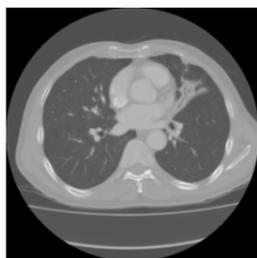

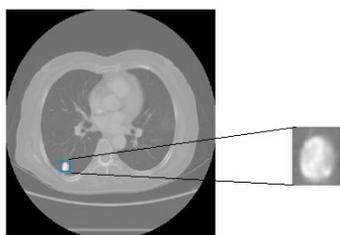**Figure 3:** Lung image after normalization

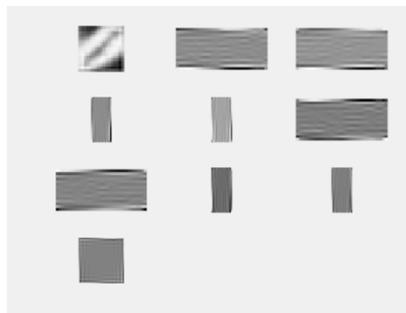**Figure 4:** ROI Segmentation of lung nodule sign

Figure 5: Lung Nodule and it's Curvelet Coefficients

Figure 5 shows the curvelet coefficients of ROI image. Table 1 shows the performance measures of the proposed method. The measures are True positive, False positive, True negative, False negative, Specificity, Sensitivity, Recall, Jaccard Coeff, Dice Coeff, Kappa Coeff and accuracy rate.
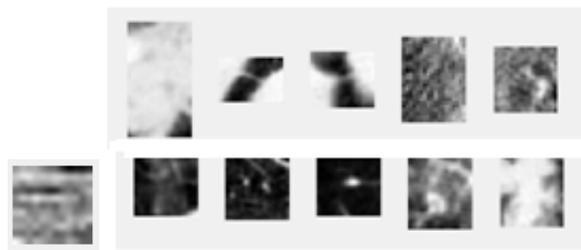

Figure 6: Retrieval results of ROI CT images

Table 1: Different Measures of Proposed Method

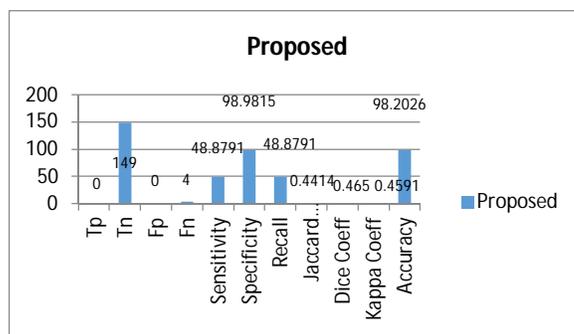| Measures	Proposed
True positive	0
False positive	0
True negative	149
False negative	4
Specificity	98.9815
Sensitivity	48.8791
Recall	48.8791
Jaccard Coeff	0.4414
Dice Coeff	0.465
Kappa Coeff	0.4591
Accuracy	98.2026

Figure 6: Classification Measures for Proposed FDCT

Figure 6 shows the True positive, True negative, False negative, False positive, Specificity, Sensitivity, Recall, Jaccard Coeff, Dice Coeff, Kappa Coeff and accuracy rate measures of Fast Discrete Curvelet Transform. Figure 7 plots the retrieval performance in terms of precision measures for every categories CT imaging signs retrival of lung nodules. In Figure 8 shows the graphical plot of recall of the proposed method. Figure 9 shows the graph between precision and recall. In Figure 10, plots the different method of precision rate and compared with the proposed FDCT Method.
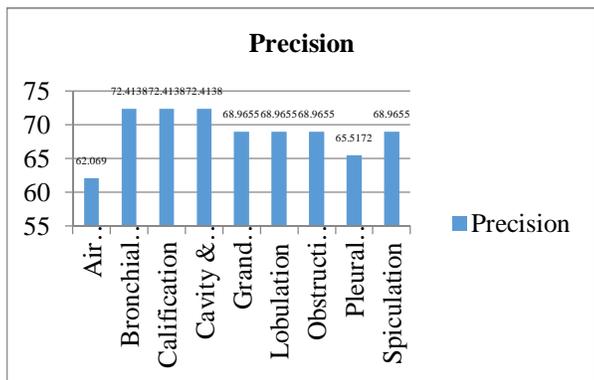
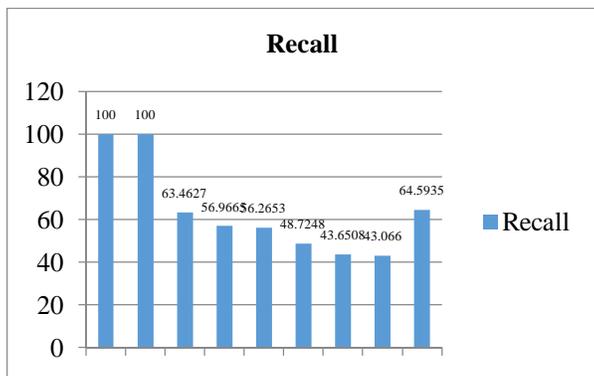**Figure 7:** Precision measures for proposed 9 nodule signs

Figure 8: Recall measures of proposed method for each category nodule signs.
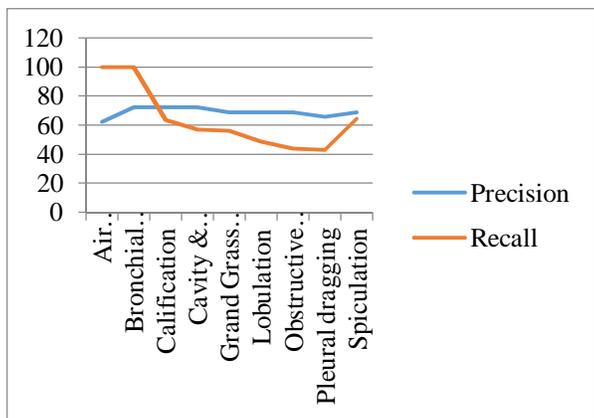

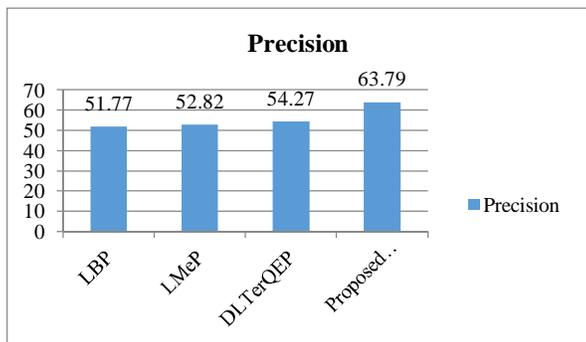**Figure 9**: Precision and Recall measures of proposed method

Figure 10: Comparison of Precision rate with various methods

5. CONCLUSION

We have proposed an efficient ROI extraction method of CT images for retrieving lungs nodules. This paper achieved content based CT imaging signs ROI retrieval. Fast discrete curvelet transform is used, for feature extraction of lung nodules. Deep learning method is employed for classification. The GLCM and shape features are used for feature extraction and enhanced moth flame optimization algorithm is used for selecting optimum features. The evaluation results show the improvement in terms of precision and recall. Our proposed method acts as an efficient algorithm for retrieving the nodules as compared to LBP, LMeP and DLTerQEP.

The various deep learning methods may be used for further improvement of retrieval results.

REFERENCES

[1] M. Loyman and H. Greenspan, **"Lung nodule retrieval using semantic similarity estimates,"** Proc. SPIE 10950, Medical Imaging*: Computer-Aided Diagnosis*, 109503P, March 2019.
https://doi.org/10.1117/12.2512115

[2] A. Alzu'bi, A. Amira, and N. Ramzan, **"Content-based image retrieval with compact deep convolutional features,"** *Neurocomputing,* vol. 249, pp. 95-105, 2017.
https://doi.org/10.1016/j.neucom.2017.03.072

[3] A. K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, and N. Khandelwal, **"Content-based image retrieval system for pulmonary nodules: assisting radiologists in self-learning and diagnosis of lung cancer,"** *Journal of digital imaging,* vol. 30, pp. 63-77, 2017.
https://doi.org/10.1007/s10278-016-9904-y

[4] M. Lam, T. Disney, M. Pham, D. Raicu, J. Furst, and R. Susomboon, **"Content-based image retrieval for pulmonary computed tomography nodule images,"** in *Medical Imaging 2007: PACS and Imaging Informatics*, 2007, p. 65160N.
https://doi.org/10.1117/12.710297

[5] R. Agarwal, A. Shankhadhar, and R. K. Sagar, **"Detection of lung cancer using content based medical image retrieval,"** in *2015 Fifth International Conference on Advanced Computing & Communication Technologies*, 2015, pp. 48-52.
https://doi.org/10.1109/ACCT.2015.33

[6] R. R. Saritha, V. Paul, and P. G. Kumar, **"Content based image retrieval using deep learning process,"** *Cluster Computing,* pp. 1-14, 2018.
https://doi.org/10.1007/s10586-018-1731-0

[7] A. Qayyum, S. M. Anwar, M. Awais, and M. Majid, **"Medical image retrieval using deep convolutional neural network,"** *Neurocomputing,* vol. 266, pp. 8-20, 2017.
https://doi.org/10.1016/j.neucom.2017.05.025

[8] P. Bhende and A. Cheeran, **"Content based image retrieval in medical imaging,"** *International Journal of*

Computational Engineering Research, vol. 3, pp. 10-15, 2013.

[9] Y.-A. Chung and W.-H. Weng, **"Learning deep representations of medical images using siamese cnns with application to content-based image retrieval,"** *arXiv preprint arXiv:1711.08490,* 2017.

[10] P. Shamna, V. Govindan, and K. A. Nazeer, **"Content based medical image retrieval using topic and location model,"** *Journal of biomedical informatics,* vol. 91, p. 103112, 2019.
https://doi.org/10.1016/j.jbi.2019.103112

[11] S. A. Mehre, A. K. Dhara, M. Garg, N. Kalra, N. Khandelwal, and S. Mukhopadhyay, **"Content-Based Image Retrieval System for Pulmonary Nodules Using Optimal Feature Sets and Class Membership-Based Retrieval,"** *Journal of digital imaging,* vol. 32, pp. 362-385, 2019.
https://doi.org/10.1007/s10278-018-0136-1

[12] P. Aggarwal, R. Vig, and H. Sardana, **"Lung cancer detection using fusion of medical knowledge and content based image retrieval for lidc dataset,"** *Journal of Medical Imaging and Health Informatics,* vol. 6, pp. 297-311, 2016.
https://doi.org/10.1166/jmihi.2016.1703

[13] S. Pang, M. A. Orgun, and Z. Yu, **"A novel biomedical image indexing and retrieval system via deep preference learning,"** *Computer methods and programs in biomedicine,* vol. 158, pp. 53-69, 2018.
https://doi.org/10.1016/j.cmpb.2018.02.003

[14] A. B. Spanier, N. Caplan, J. Sosna, B. Acar, and L. Joskowicz, **"A fully automatic end-to-end method for content-based image retrieval of CT scans with similar liver lesion annotations,"** *International journal of computer assisted radiology and surgery,* vol. 13, pp. 165-174, 2018.
https://doi.org/10.1007/s11548-017-1687-1

[15] G. Deep, L. Kaur, and S. Gupta, **"Local mesh ternary patterns: a new descriptor for MRI and CT biomedical image indexing and retrieval,"** *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization,* vol. 6, pp. 155-169, 2018.
https://doi.org/10.1080/21681163.2016.1193447

[16] G. Wei, H. Ma, W. Qian, H. Jiang, and X. Zhao, **"Content-based retrieval for lung nodule diagnosis using learned distance metric,"** in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 3910-3913.
https://doi.org/10.1109/EMBC.2017.8037711

[17] A. Terasa, M. Waghela, A. Jagtap, and S. Chauhan, **"Content Based Image Retrieval Using Wavelet Transform,"** *Research Journal of Engineering Technology and Management,* vol. 2, 2019.

[18] N. Varish and A. K. Pal, **"A novel image retrieval scheme using gray level co-occurrence matrix descriptors of discrete cosine transform based residual image,"** *Applied Intelligence,* vol. 48, pp. 2930-2953, 2018.
https://doi.org/10.1007/s10489-017-1125-7

[19] A. Huneiti and M. Daoud, **"Content-based image retrieval using SOM and DWT,"** *Journal of Software Engineering and Applications,* vol. 8, p. 51, 2015.
https://doi.org/10.4236/jsea.2015.82007

[20] C. Zhang, J. Li, S. Wang, and Z. Wang, **"An encrypted medical image retrieval algorithm based on DWT-DCT frequency domain,"** in *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, 2017, pp. 135-141.
https://doi.org/10.1109/SERA.2017.7965719

[21] A. A. Shinde, A. D. Rahulkar, and C. Y. Patil, **"Fast discrete curvelet transform-based anisotropic feature extraction for biomedical image indexing and retrieval,"** *International Journal of Multimedia Information Retrieval,* vol. 6, pp. 281-288, 2017.
https://doi.org/10.1007/s13735-017-0132-0

[22] D. Sarala, T. Kanikdaley, S. Jogi, R. K. Chaurasiya, "Content-based image retrieval using hierarchical color and texture similarity calculation," *International Journal of Advanced Trends in Computer Science and Engineering,* vol.7(2), pp. 11-16,2018
https://doi.org/10.30534/ijatcse/2018/02722018

[23] B. P. Kumar, K. V. Ramanaiah, **"Effective ROI extraction methods for hybrid medical image compression"** *International Journal of Advanced Trends in Computer Science and Engineering,* vol.8(2), pp. 277-284, 2019
https://doi.org/10.30534/ijatcse/2019/29822019

[24] G. Han, X Liu, F. Han et al. **"The LISS-A public database of common imaging signs of lung diseases for computer aided detection and diagnosis research and medical education"**. *IEEE Trans. Biomedical Engineering,* 62(2): 648-656, 2015.
https://doi.org/10.1109/TBME.2014.2363131