

Triplet Contents based Medical Image Retrieval System for Lung Nodules CT Images Retrieval and Recognition Application

Ranjit Biswas, Sudipta Roy, Abhijit Biswas

Abstract: Content Based Medical Image Retrieval (CBMIR) has found its relevance in medical diagnosis by processing massive medical databases based on visual and semantic features and user preferences. In this paper we address two issues such as retrieval and recognition. We present a novel method called Triplet-CBMIR for lung nodules CT images retrieval and recognition application. A Triplet CBMIR is a combination of three properties: Visual Features (Shape and Texture), Semantic Features and Relevance Feedback. Dataset training is done using: Preprocessing, Feature Extraction, Selection, Nodules Sign Detection and Clustering. In preprocessing we perform image scaling, denoising and normalization. In feature extraction, two methods are presented such as Hybrid Discrete Wavelet Transform (DWT) and Discrete Cosine Transform (DCT), Bounding-Box based Convolutional Network (CNN) for visual and semantic features extraction. Then optimum set of feature vectors are selected using Mutual Information based Neighborhood Entropy (MINE). Based on selected features, lung nodule sign is detected using K-nearest Neighbor (KNN) algorithm in which Hassanat Distance used and similar images are grouped using Multi-Self organizing Map (SOM). For similarity measurement, d_1 distance metric is used. Benchmark dataset such as LISS and LIDC are used for the study. Performance matrices such as Average Precision Rate (APR), Average Retrieval Rate (ARR), Average Recognition Rate (ArR), Running Time found in the simulation results are compared with some other already present state-of-the-art works. The proposed method shows a significant improvement as compared to other existing methods.

Keywords : Content based medical image retrieval, KNN with Hassanat distance, Multi-SOM, bounding Box based Convolutional Neural Network.

I. INTRODUCTION

Medical images such as computed tomography (CT), magnetic resonance imaging (MRI), X-rays and Positron Emission Tomography (PET) have been increased recently with the dawn of latest technologies in image acquisition and enhancement [1]. Today many hospitals and medical

research centers generate enormous medical images with different modalities, which cause large database formation [2]. To handle such huge amount of medical images, Content based Medical Image Retrieval (CBMIR) is introduced. CBMIR is currently emerging and used in many real-time medical image applications [3]. Over the past few decades CBMIR has started research for radiologists and doctors. Several medical image databases are emerging in CBMIR which includes MRI, CT, Ultrasound, PET, and many more. The aim of CBMIR is to retrieve similar images for the given query image. For that several CBMIR methods have been proposed and there are different issues which were raised [4], [5], [6]. Feature extraction and reduction is one of the important and challenging issues. However, medical images are retrieved using low level and high level feature contents. Low level features represent visual contents and high level features represent semantic contents. A good image must compose of these two features. Previous solutions are based on visual contents such as color, shape, texture, and other information [7], [8]. However design a fast and accurate CBMIR is a challenging task of today's researchers. Medical image consists of texture, and shape oriented features and these features are primary features. Among several features, texture feature is essential for CBMIR and today it is widely supported in several medical applications (disease diagnosis) [9], [10]. For CBMIR, similarity between features is focused. Individual and local feature information helps to improve average precision rate (APR) and average retrieval rate (ARR). However, only texture feature information are not sufficient to retrieve accurate results [11]. Many new feature descriptors, such as, LWP (Local Wavelet Pattern) and CS-LBP, for a number of CBMIR systems have been proposed. Furthermore, dictionary learning, and deep convolutional neural network [15] is proposed for CBMIR. Several distance functions viz., Euclidean distance, Manhattan distance, Cosine similarity, Canberra distance, Minkowski distance, and Chi-square distance are used for similarity computation between the query image and the large collection of image database [12], [13], [14], and [15].

According to distance functions, images with high similarity values are retrieved as an output. The primary objectives of this paper are follows: (1). to extract visually similar features for enhancing the average retrieval rate (ARR) in CBMIR, and (2). to use fast similarity function to reduce the retrieval time and improve the CBMIR system performance.

Revised Manuscript Received on October 30, 2019.

* Correspondence Author

Ranjit Biswas*, Department of Information Technology, Ramkrishna Mahavidyalaya, Kailashahar, India.

Sudipta Roy, Department of Computer Science & Engineering, Assam University, Silchar, India.

Abhijit Biswas, Department of Computer Science & Engineering, Assam University, Silchar, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

The rest of the paper is organized as follows: preliminary literature related to CBMIR is presented in Section II. A detailed comprehensive review of recent state-of-the-art in CBMIR is given in Section III. In this section we also presented the limitations of each study. In section IV, we present our solution and give a brief description of the proposed Triplet CBMIR.

In section VI, we describe the experiment results and analysis. Finally, In section VII, work is concluded.

II. CBMIR FOR LUNG CT IMAGING

In this section we present the preliminaries of our CBMIR system for Lung CT images, which help radiologists to retrieve similar lung nodules for sign detection. Lung cancer is one of the terminal diseases that occur in man and women. Computer Aided Diagnosis (CAD) plays a significant role in the detection of lung cancer [16], [17]. Hence the CAD system analyzes lung images in different steps. Firstly a pre-processing step for image resizing, noise reduction and contrast enhancement is implemented. Image segmentation step is used to show the difference between ROI and original image. In general, lung CT images are based on rich set of features (contents). Therefore, several features such as geometrical, textural, statistical, and semantic features are extracted. Finally evaluation or detection is completed to diagnose the images using extracted features. For any kind of operation such as recognition or retrieval, these features must be accurately extracted since every image in a medical database consists of special characteristics. Hence CBMIR is the requirement for early diagnosis of nodules sign in the large database [18], [19]. A general CBMIR system for Lung CT imaging is depicted in fig.1

III. LITERATURE REVIEW

In computer aided diagnosis, lung nodules CT images play a significant role. In this section we deliberate the state-of-the-art in CBMIR for the recent available works. Previous works of CBMIR were focused on effective feature (local and global) feature representation such as color, shape and texture, and indexing images based on the contents of feature vectors. Disadvantages of earlier works are follows: (1). Sematic gap between low level and high level image features. Thus, it is very tough to choose the appropriate features, and (2). User preferences changes, so accurate images must retrieve for a given query. In [20] comprehensive review has been made on lung CT image signs detection for publicly available LISS database. This database consists of 271 CT images (677 abnormal CISL regions). Different abnormal CISL regions are ground glass opacity, calcification, cavity and vaculous, pleural indentation, air bronchogram, obstructive pneumonia, bronchial mucus plugs, speculation, etc. Authors in [21] have proposed a weighted sum method (multi-classifier fusion) for lung nodules recognition. The confusion matrix is constructed, which shows the historical decision making reliability.

Two components are combined to find the weight of confidence value for classifier. Classifier fusion process is ineffective since the range of classification is not uniform and hence normalization of confidence value is important.

A deep learning model is proposed for classification of lung CT imaging sign. This deep learning model is implemented for small data, which determine signs of lung nodules. Visual features extraction and sign detection is executed over deep learning model. Here CNN is used as a

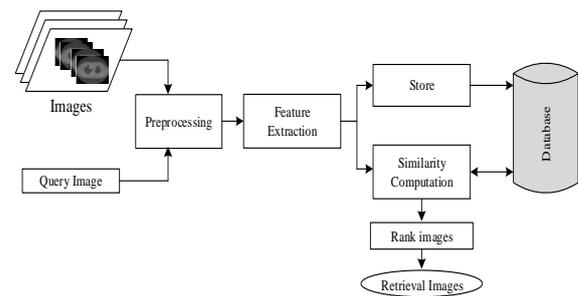


Fig.1. A General CBMIR for Lung CT Imaging

learning model for sign detection and for testing real samples are used. Accurate features extraction is required for lung nodules sign detection [22]. A new hybrid method is proposed for hybrid features selection, which are called genetic optimization and fisher criterion [23]. In the work [21] addressed the problem of lung nodules CISL recognition, which is the significant part of CAD and CBMIR. A new hybrid feature selection method effectively addresses these issues and tested over five classifiers. Among these they have concluded that support vector machine (SVM) is the best classifier for sign detection. In general CISL of LISS database consists of noises which must be removed for sign detection and image retrieval. Furthermore other preprocessing steps such as contrast enhancement and wavelet decomposition are required. In [24] a topic and location model is proposed for CBMIR. Information of topic is created using Guided Latent Dirichlet Allocation (GuideLDA) model whereas the location model is used to incorporate the spatial information of visual words. In order to find the order of images for retrieval, a new precision metric is used, which is called as $wPrecision$ (Weighted Precision). Weighted precision is used to find the top-k results for the given query image. One major problem in weighted precision is it produces infrequent feature vectors, which is less weight precision value. Infrequent classes of the weight precision may hide and also it leads to poor performance, but infrequent classes are also important to retrieve the relevant images. K is static and fixed i.e. the number of topics are fixed in guided LDA, and hence it affects the quality of CBMIR system in terms of ARR and APR. In [25] authors have proposed unsupervised CBMIR scheme using visual words spatial matching. Here spatial similarity of visual words is matched based on the Skip Similarity Index (SSI). This paper majorly has two contributions including location features and SSI. The SSI can effectively measure the sequential location and similarity between each visual word by skipping the unmatched location features. Hamid et al.

[26] have discussed a new image dense-sampling descriptor called Encoded Local Projections (ELP), which solve classification and search problems. For CBMIR, local binary pattern (LBP) is used and also deep features extracted using pre-trained networks. The ELP image descriptor produces semantically better matches and also LBP and HOG of deep features, which are used for accurate medical image retrieval. Deep features are extracted using VGG network with 16 hidden layers, which require large set of images for training and it deliver the results of $O(1)$, but training require exponential time. Unsupervised learning algorithm called SOM is proposed [27] for CBMIR. In training stage, local texture image features are extracted for large image database. Then SOM is applied on texture features to retrieve the set of medical images. It only considers texture features for retrieval and it also extracts texture features in 3×3 neighborhood information. Pixel-wise computation for retrieval using SOM is time consuming and also texture features extraction only does not provide the accurate results. Researchers in [28] have proposed a simple CBMIR scheme, which primarily considered texture features for retrieval. Initially two filtering schemes are employed such as Gabor and Schmid. Then filtered images are partitioned into patches which are non-overlapping. It provides extensive local texture information for medical images. Finally, bag-of-words model is presented to acquire image represented as features. The rate of ARR and APR is very low due to simple method for CBMIR. Deep convolutional neural network is proposed in [29] for image classification. The proposed scheme is tested on intermodal database, which consists of nearly twenty four classes for five medical image modalities.

Singh et al. [30] have proposed an automated and effective content based mammogram retrieval on the basis of two approaches including SOM and wavelet features based CS-LBP. It is very fast, effective and automatic. To improve the system performance three preprocessing steps are included: automatic labeling scratches suppression, automatic removal of pectoral muscle, and image quality improvement. The segmentation of selective threshold in seeded region growing is proposed. Wavelet features are extracted in the segmented region using CS-LBP. Finally, SOM is used for cluster formation. There are two limitations in SOM, which are classes/groups, do not necessarily match informational categories of interest and SOM is very limited control of clusters (number of classes and identities). Hepatobiliary images retrieval is focused in which Hepatobiliary [31] diseases related images are retrieved for the human part of biliary and liver systems. Three feature extraction methods are used: (1). Hu-moments, (2). Scale Invariant Feature Transform, and (3). GLCM. Finally these feature vectors are fused into single vector and compared with image database using Euclidean Distance. Experiments are conducted and the performance is evaluated for three metrics: precision, recall and error rate. Among various distance measures, Euclidean Distance is not effective distance metric since the distance between two feature vectors may be large even though the two feature vectors are similar. In order to fast retrieval of images, domain index is employed in [32]. It is an index constructed for different categories. For content based queries searching in domain

index, KNN is proposed. Experiment is tested for mammogram public database. KNN is a very simple technique that does not support of similarity computation for very large database. Error rate during similarity computation is large and it takes large computation time. K-SVD algorithm is proposed, which uses clustering method for dictionary learning for similarity analysis on group of large medical image collections [33]. A user query image is matched using orthogonal matching unit (OMP) algorithm. Supervised information is used for image retrieval. OMP and K-SVD are time consuming process. In [34] authors have proposed LWP (local wavelet pattern) which characterizes the query image for retrieval of similar images. The LWP addresses the issues of LBP where information determines from centre pixel information and it is matched with the LWP decomposed values. The LWP feature descriptor outperforms in CT image retrieval and it is better than the LBP method. Even though, LWP is better than the LBP, much better feature descriptor is required for medical image retrieval. Relevance feedback is a new concept which is focused on CBMIR. In [41] Lenient Relevance Feedback mechanism is proposed. This mechanism better captures the intention of users. The Meta-Knowledge is mined for multiple users and through that experience, it is able to increase the precision of subsequent image recovery results. There are two set of techniques proposed such as redundancy detection and confidence quantization. For weight interface identification, binary search and best first search are proposed. These two searching strategies increase the speed of retrieval and increase the interface computation, but it failed to produce accurate results

IV. PROBLEM STATEMENT

In this paper we consider massive database consists of thousands of images. Each image may contain rich set of features or low level features. These features are extracted and reduced, which produce the accurate results during retrieval. We addressed the following research questions:

RQ1: What kind of large database can be supported for CBMIR, particularly to lung nodules sign diagnosis applications?

RQ2: How to reduce the search/retrieval time in very large collection of dataset?

RQ3: How to measure the similarity between query image and set of images in database?

In [35] a new CBMIR is employed by fusion of context-sensitive similarity. Visual and semantic features are extracted and the similarity between the query image and image database are fused according to pair-wise similarities. Then weighted graph is constructed in which node refers to image and edges represent the pair-wise similarity in the given dataset. Our problems stated in this paper is related to medical image retrieval efficiency, which is very less since searching top-k results require large amount of time. Clustering/ classification of similar images are required to enhance the efficiency of retrieval system. Currently, Medical sign detection is one of the important tasks in lung nodules (benign/malignant) classification [36].

The major contributions of this paper are four-fold: (1). Deep learning algorithm is proposed for semantic features extraction, which efficiently represents the medical sign information, (2). High-dimensional image features are translated into compact binary codes using supervised hashing and principal component analysis (PCA), (3)

Similar images of lung nodules are retrieved using Adaptive Weighted Similarity Technique, and (4). Recognizing nodules by retrieved results. Two sets of database (LIDC-IDRI, and LISS) are experimented. Our problem stated in this paper is related to image scaling (resize), which is implemented using linear interpolation method. It is a very simple method generally apt for small scale images. It does not smoothes the edges of the image. Query images are currently noise (salt and pepper), which degrade the retrieval performance [37]. Semantic features extracted using CNN are not efficient and effective. KNN is the nearest neighbor algorithm proposed for nodule sign detection by computation of Euclidean Distance (ED). This metric is not suitable to find the similar images for nodule sign detection. An integration of Tamura texture features and wavelet transform is applied for medical image retrieval using Hassanat distance. This Hassanat distance metric defines the overall feature set similar values. The main objective of this retrieval system is to retrieve top-k similar (accurate) images. Experiments conducted for two medical image databases: lung CT database and MRI database. Our problem statement in this paper is related to image preprocessing. This step is required since similarity measure Hassanat distance not reduce noise and abnormal points of medical images. Importance of wavelet transform is not defined. In [38] two-step CBIR method is proposed for lung nodules image retrieval. There are two similarity metrics which were used such as visual similarity and semantic relevance. Firstly, each given ROI with top-k similar ROI references are determined using semantic relevance measure. Then, weight value is assigned for each retrieved ROI. The proposed TSCBIR scheme is validated for LIDC-IDRI lung CT images. Three different texture features (LBP, Gabor, and Haralick) group is designed to represent ROI. In this work, Mahalanobis distance is used to preserve the semantic relevance, and European distance is used to describe the visual similarity, which does not produce the accurate results. Our problem stated in this paper is related to feature extraction. Feature extraction time using LBP, Gabor, and Haralick are high. In [39], authors of this paper are presented a new similarity measure to retrieve top-k results for the query image. Experiment is conducted for lung mass dataset, which assembled with 746 mass ROIs. A semisupervised distance metric is considered to map the input feature space to the image repository. Feature extraction, similarity measurement and ROI extraction are implemented in this study. More effective features such as semantic and visual features are required for medical image retrieval and thus the proposed scheme lacks robustness. In [40] inter observer variability is reduced when retrieving similar images. This system is very fast, simple and reliable for image retrieval. However image retrieval for lung nodules is based on the segmented part of the nodule. For this purpose, semi-automated nodule segmentation method is proposed. SVM classifier is used for nodule classification. Our problem stated in this paper is

related to insufficient feature set, database size degradation, poor segmented result and classification. For effective CBMIR, the optimum sets of features are required.

Our research aim in this paper is to develop a fast and accurate CBMIR, which find the top-k medical images that are visually and semantically similar to the query image (target image). Assume that the images of similar semantic meaning or category form a cluster in feature vector length. In given set of samples, user's preference is determined for a given query to produce the better results for retrieval

V. PROPOSED TRIPLET CBMIR SYSTEM

In this paper we developed the CBMIR and Recognition system based on the Triplet contents of the given query image viz., Visual features (Shape and Texture), Semantic features, Relevance feedback Steps were followed to develop an accurate and fast CBMIR described in Algorithm 1. Our proposed work is implemented using LISS and LIDC-IDRI Databases. We present the system architecture in fig.2 There are two core phases: (1). Offline Phase, and (2). Online Phase.

In offline phase, images in database are trained and then implemented preprocessing steps, where we perform image scaling, denoising, and image normalization. After preprocessing, feature extraction process is initiated in which multiple features are extracted. Then we implement feature selection, where redundant features are removed and only optimum set of features are considered. Then we detect nodule signs. However, it is varied for one another. In Total, 9 different nodules sign were detected from the given databases that are: (1). Grand Grass Opacity (GGO), (2). Lobulation, (3).Calification, (4).Cavity &vacuolous, (5).Spiculation, and (6). Pleural dragging, (7). Bronchial Mucus Plugs (BMP), (8). Air Bronchogram, (9).Obstructive Pneumonia (OP). To improve retrieval efficiency in terms of time of result deliver, we cluster the similar images into group.

In online phase, user will give the query image. Before processing into the system, relevance feedback is employed for a user.

Relevance feedback is a technique that considers the user preference during medical image retrieval process. When the retrieval system receives the query image, it provides some relevant images from the large collection of database and user can checks the relevancy for the resultant images and then select the most relevant image for accurate retrieval. Relevance feedback and clustered image database environment have provided high performance results. For the query image our proposed system implements the process of preprocessing, feature extraction, feature selection and then similarity is computed using five distance metrics: Manhattan Distance, Euclidean Distance, Canberra Distance, Chi-square distance, and d_1 Distance Indexing is performed to retrieve most similar top-k results. We find the most similar image from the retrieved results for nodule signs detection. When retrieval is over, then nodules sign detection is initiated. For recognize the lung nodules sign, we presented KNN based Hassanat distance metric.

In the following sections we illustrate proposed system architecture (Illustrated in fig 2). Algorithm for the same is given below.

A. Image Preprocessing

Image preprocessing is a significant part of this paper, which is done to improve the image quality and minimize the redundant data. There are three preprocessing steps

incorporated.

- **Scaling**

Image scaling is the resizing of image, where we gain aliasing and blurring effects. If the image is scaled down to the optimum size, then we can increase

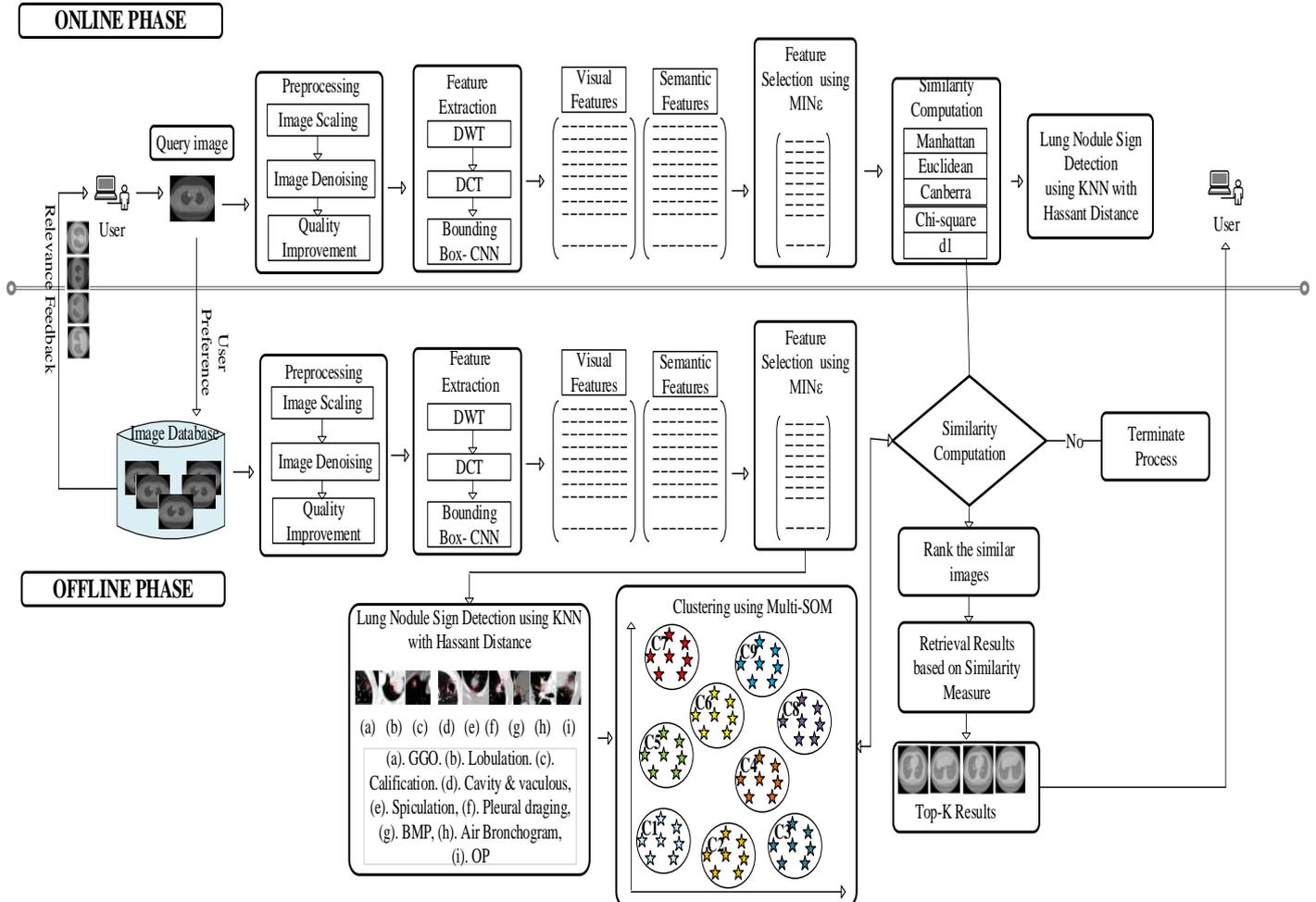


Fig.2. System Architecture

the optimum size, then we can increase the speed of entire process. In this paper, cubic interpolation method is used for image scaling

- **Denoising**

Image denoising is performed, which reduces noise (salt and pepper) using adaptive median filtering. Noise in image is reduced in left to right directions. Adaptive median filtering works on the image pixels median values. It is computed pixel by pixel and obtain median values for the whole image. It is computed in adaptive way i.e. median value for pixels in the entire image is different from one another. This kind of filtering technique is effective for retrieval and are suited when images are corrupted by noise.

- **Contrast Enhancement**

Then we implement Histogram Equalization to reduce the redundant pixels on each image. This process not only increase contrast level but also it increases the speed of computation. Histogram equalization is used for contrast enhancement, which also reduces computational burden in CBMIR.

B. Feature Extraction

In this step, the pre-processed image is fed into feature extraction. For detecting nodules sign, visual and semantic characteristics are considered. For apparent features, multiple geometric parameters are considered such as shape, elongation, size, speculation, density, suspected lesion and other features. Hybrid DWT and DCT are employed for feature extraction. Initially DWT separate image data into different frequency components and then each component is studied to its scale. We utilized Haar in wavelet since it provides the high recognition rate. In addition, DWT compute four coefficients such as cA (All components), cH (Horizontal Component), cV (Vertical Component), and cD (Diagonal Component). Among these, cA coefficients are used to get a reduced set of visual features. Still, DWT is in Spatial Domain.

To deal with this issue, we consider DCT, which transforms into Frequency Domain. In addition, DCT lay a ground for better feature selection. These are combined to extract the wavelet features. In parallel mode, semantic features are extracted using Bounding-box based CNN. Table.1. shows the list of features. The retrieval efficiency for any feature extractor is depending upon the feature extraction required to key features using Spectrum achieved. The unique combination of hybrid DWT and DCT, and bounding box based CNN extracts all salient features.

Algorithm 1: Triplet - CBMIR

// Offline Phase

- Step 1)**Begin
 - Step 2)**Define the image sets (LISS Database)
 - Step 3)**For each image do
 - Step 4)**Image preprocessing
 - Step 4.1) Image scaling by Cubic Interpolation Method
 - Step 4.2)Denoising by Adaptive Median Filter
 - Step 4.3) Quality improve by Histogram Equalization
 - Step 5)**Implement feature extraction (Visual & Semantic)
 - Step 5.1) Image decomposition by 3 levels of DWT
 - Step 5.2) Then apply DCT
 - Step 5.3) Semantic features extraction using CNN based Bounding Box
 - Step 6)**Features selected by *MINε*
 - Step 7)**Lung nodule sign detection by KNN (Hassanat metric)
 - Step 8)**Similar images clustering by Multi-SOM
 - Step 9)**End For // **Terminate offline phase**
- // Online Phase
- Step 10)** Given query image
 - Step 11)** For a query image, return relevance feedbacks
 - Step 12)** Implement steps (4) – (6)
 - Step 13)** Compute similarity index between query image and every database images using similarity measures.
 - Step 14)** Retrieve top-k images
 - Step 15)** End for// **Terminate online phase**

Table-I: Image Features and their Properties

Type of Features	Feature Name
Texture features	Correlation, Contrast, Homogeneity, Sum of square variance, Spectral, Spatial, and Entropy
Shape features	Area, Irregularity, Roundness, Perimeter, Circularity
Intensity features	Intensity, Mean, Standard Variance Kurtosis, Skewness, Median
Geometric features	Eccentricity, Compactness, Roughness, Local Area Integral Invariant, and Radial Distance Signatures.

C. Feature Selection.

The result from feature extraction consists of huge feature vectors. In order to reduce and to gain the most essential features needed for recognition and retrieval, feature selection is performed using Mutual Information based Neighborhood Entropy (MINε). It is the combination of two methods such as information entropy and mutual information. A single feature selection method include mutual information, Shannon entropy, mRMR, and rough introduces more errors in evaluating and analysing the prominent part of the image. It is often called as the Significance of Decision (SoD). The presence of one feature

in the query image helps in another aspect so that we compute the neighborhood entropy for feature selection. It is computed by the eqn. (1)

$$S(w) = \varphi_B(w) \times SoD(w) \tag{1}$$

Where *SoD(w)* is the any feature value and $\varphi_B(w)$ is the Banzhaf value. Before starting feature selection, we present the details of the information system $U = \{NES, C, D\}$, where NES is a Non-Empty finite Set of objects, C is a Non-Empty finite set of conditional attributes, and D is the attribute for making the decision. Since it is very difficult to find the sign of lung nodules and semantic features are calculated to predict the sign of lung nodules. In this paper we proposed bounding box based CNN for semantic features extraction in region of interest level. The set of semantic features and the characteristics are described in following.

- Sphericity: the shape of nodule related to its roundness
- Malignancy: the assessment of likelihood of malignancy of the nodule
- Margin: the description of how well defined the nodule margin
- Texture: nodule composition or internal texture related to solid and ground glass of components
- Subtlety: it is very difficult in sign detection and it represents the contrast between the lung nodules and the surrounding.
- Spiculation: It is a degree of spicules exhibition
- Radiographic Solidity (Texture): It is a consistency of nodule, and internal texture
- Calcification: Its presence and pattern of calcification.
- Internal structure: Expected Internal Composition of the Nodule

The above set of nodule features are rated from 1 to 5 or 6. It gives the image pixel level characteristics to radiologists. Bounding box is used to detect and locate the nodule signs precisely. The structure of the network is similar to Region Proposal Network. Firstly we use the layers from 1-5, whose intention is to extract the image semantic features. This model is pre-trained on the lung nodules dataset which produces the higher performance for features extraction.

D. Nodule Sign Detection

After the optimum sets of features are extracted, a modified version of KNN is used. KNN is used as a classifier to classify an unlabeled test image based on the k-nearest neighbors that are very close to the test image. It is based on the distance between the test image and training images stored in the database. Similar samples are found by a specific distance measure, which is called as Euclidean Distance.



Algorithm 2: Hassanat Distance Metric-based KNN

// Lung Nodules Sign Detection

Step 1) Begin

Step 2) Define the feature vector sets for training set t_s

Step 3) For each image T_i feature set F_s do

Step 4) Sign Detection Initiated.....

Step 4.1) Initialize the variable K

// K = Pre-defined small integer value

Step 4.2) Calculate the distance between T_i and t_s

Step 4.3) Select K in t_s close to T_i

Step 4.4). Allocate the most frequent class close to the distance to T_i

Step 5). Compute the class label for all images in t_s

Step 6). End for

Distance computation between the test image and training images using Euclidean Distance is not effective and it leads to the miss-classification and noisy data are not trained properly. Hence selecting the most adopted distance function is needed in the KNN classifier. Therefore in this paper, we proposed Hassanat distance metric, which outperforms than several distance metrics in [42]. Algorithm for the modified version of KNN classifier is following.

Algorithm description is follows:

The training set and the class label of each test image from t_j is stored in $|DB|$. Hence no missing data is allowed and no non-numerical data is allowed. Each test image is correctly assigned by the class label using Hassanat distance metric. Then K - nearest neighbors are chosen from t_j . The Hassanat distance function D is computed by eqn (2-3)

$$HD(X, Y) = \sum_{i=1}^N D(X_i, Y_i) \quad (2)$$

$$D(X, Y) = \begin{cases} 1 - \frac{1+Min(X_i, Y_i)}{Max(X_i, Y_i)}, & Min(X_i, Y_i) \geq 0 \\ 1 - \frac{1+Min(X_i, Y_i)+|Min(X_i, Y_i)|}{Max(X_i, Y_i)+|Min(X_i, Y_i)|}, & Min(X_i, Y_i) < 0 \end{cases} \quad (3)$$

$D(X, Y)$ is bounded by 0 and 1.

E. Clustering

Clustering is used to categorize the CT image database for speeding up the CBMIR and enhance the retrieval accuracy and efficiency. Clustering process is implemented using multiple SOM. Our proposed work solves the two major limitations [8] of conventional SOM. For all images that are stored in the databases, feature vectors are pre-computed and similar feature vectors of the images are grouped, which are stored in a database.

F. Similarity Computation

A good retrieval system for content based medical images is not only based on feature descriptor or classifier, but it is also based on the best similarity metric. For a similarity computation between a query image and images stored in a database, we used several distance functions include Manhattan Distance, Euclidean Distance, Canberra Distance, and d_1 Distance that are represent by the following eqns (4-7).

- **Manhattan Distance**

$$d(Q, |DB|) = \sum_i |f_i(Q) - f_j(|DB|)| \quad (4)$$

It is also known as L_1 City Block Distance.

- **Euclidean Distance**

$$d(Q, |DB|) = \sum_i |f_i(Q) - f_j(|DB|)|^{1/2} \quad (5)$$

It is also known as L_2 City Block Distance

- **Canberra Distance**

$$d_s(Q_M, |DB|_M) = \sum_{i=1}^{F_{LE}} \frac{|f_{|DB|,i} - f_{Q_M,i}|}{|f_{|DB|,i} + f_{Q_M,i}|} \quad (6)$$

- **d_1 Distance**

$$d(Q, |DB|) = \sum_{i=1}^{F_{LE}} \frac{|f_{|DB|,i} - f_{Q,i}|}{|f_{|DB|,i} + f_{Q,i}|} \quad (7)$$

Where Q is query image, F_{LE} is the length of feature vectors, $|DB|$ is the database, $f_{i,i}$ is the feature i in the database, $f_{Q,i}$ is the feature i in the database of query image Q .

In this stage, relevance feedback is incorporated. Initially query image for round 1 is implemented, which perform on the basis of visual and semantic features extracted by the radiologists. Running of this triplet-CBMIR system is based on top-k images stored in the database and similarity computation with the initial query feature vectors and k is defined by user and it is typically 100 (maximum). After the query submission, user gives feedback for setting the relevant and non-relevant images. Accordingly, the user feedback for initial query image feature vector is properly analyzed with their position and semantic marks. Query reformulation is recommended from users when query image is not matched with the database and the new updated feature vector incorporated both radiologists and images. Finally the top “K” images are retrieved using initial query vector and relevant image features vector.

VI. EXPERIMENTAL RESULT AND DISCUSSION

A. Experimental Setup

For implementation we use MATLAB programming. In this paper Matlab code is used for execution. This programming language works quicker and in a effective way where it is operating with image or video. The system configuration for experimentation and full installation of the proposed system can be seen in table.2.

Table-II: Configuration Setup

Parameter	Value
Operating System	Windows 7 (64-bit OS)
Processor	Intel (R) Pentium (R) CPU G2030@2.20GHz
Disk Space	500 GB
RAM	2.00GB
Graphics Card	No specific requirement

For retrieval of lung CT images, some of the training and testing datasets are required. Currently, some datasets are publicly available for evaluation of several CBMIR methods. These datasets are described in this section.

B. LISS Dataset: A Public Database of Common Imaging Signs of Lung Diseases [20].

We abbreviate this data set as LISS CISL. It is a publicly available database consists of 511 2D CISLs for 252 subjects, 166 3D ones for 19 subjects. In totally it consists of nine categories, which are acquired from the Cancer Institute and Hospital at Chinese Academy of Medical Sciences.

It captured using GE LightSpeed VCT 64 and Toshiba Aquilion 64 CT Scanners and images are saved in the format of DICOM 3.0. In LISS database, 2D instances cover all nine categories (Calcification, Cavity and Vacuolous, Spiculation, Lobulation, Pleural Dragging, Bronchial Mucus Plugs, Obstructive Pneumonia and Air Bronchogram) of lung nodules sign, whereas the 3D instances cover the single category i.e. Grand Grass Opacity (GGO). In this database all CT scans are labeled and the corresponding regions are annotated the signs. Size of each image is 512×512 . The main intention of this database is to provide massive large database support for CAD research. It consists of 9 classes of lung CT images of lung diseases with slice is 0.418mm – 1mm and the mean value is 0.664mm. . Table 3 illustrates the number of CT scans and annotated ROIs given in this database. Fig 3 shows the lung nodules sign for all images in a dataset.

C. Retrieval and Recognition Performance parameters

In order to evaluate the performance of the CBMIR system, we considered four

Table-3.:LISS Database Description [20]

Lung Sign	CT Imaging Scans	Annotated ROIs
GGO	25 – 2D 19 - 3D	45 – 2D 166 – 3D
Lobulation	21	41
Calcification	20	47
Cavity and Vacuolous	75	147
Spiculation	18	29
Pleural Dragging	26	80
Air Bronchogram	22	23
Bronchial Mucus Plugs	19	81
Obstructive Pneumnia	16	18
Sum	271	677

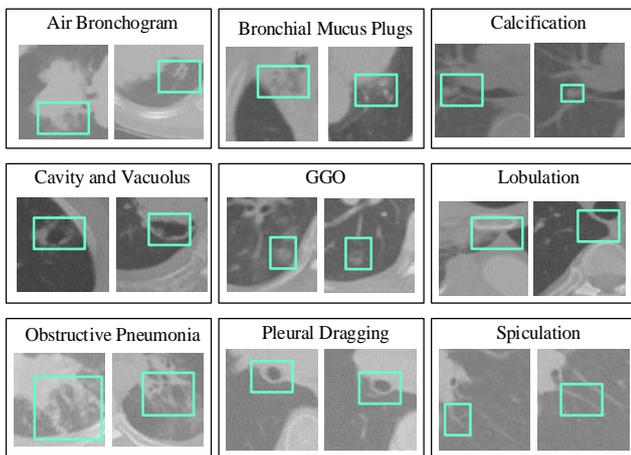


Fig.3 Lung Nodules Sign Detection (LISS Database)

parameters such as Average Precision Rate (APR), Average Retrieval Rate (ARR), Average Recognition Rate (ArR) and Running Time. Definition for each metric is given in this section.

• **APR**

Precision defined as the most relevant retrieved and it indicates the fraction of retrieved items that are most similar to the user query and thus we can estimate the ability of the method for relevant items retrieval. APR is defined as the average precision values where a relevant item has been retrieved, which depends on the User-Defined Ranking Precision values. It is computed by:

$$APR = \sum Precision \tag{8}$$

Recall is also known as sensitivity, which refers to the fraction of relevant items to be retrieved and it shows the capability to retrieve the relevant items only by the normalized plot unit. ARR is defined as the average recall values. It is computed by:

$$AR = \sum Recall \tag{9}$$

• **ARR**

It is defined as the ratio of user retrieval sessions that lead to retrieval success. This metric represents the retrieval method performance. In other words it is defined as the rate of items that are retrieved to be relevant. It is computed by

$$ARR = \frac{1}{|DB|} \sum_{i=1}^{|DB|} r(K_i, n) \tag{10}$$

Where $|DB|$ is the total number of images in the database, K_i is the image which has the rank i

• **ArR**

It is defined as the success rate of recognition. It differs from the feature dimension and also it is evaluated based on feature extraction and selection methods. It is computed by

$$ArR = \frac{1}{|DB|} \sum_{i=1}^{|DB|} R(K_i, n) \tag{11}$$

• **Running Time**

It is defined as the amount of time required for retrieving all relevant images from the database. It is negative metric so it must be small to show the system has given the better performance. In this paper we define this metric by time required to retrieve for the similar images for individual user query. It is computed by,

$$RT = User_Query_{Given\ Time} - Retrieval\ Completion\ Time \tag{12}$$

• **Precision-Recall**

It is the most common measure for information retrieval. Precision-Recall is the integration of two measures, represented in a single plot. Visually, it is represented into three variations such as Non-Interpolated, Inter-Polated, and the Ten/Even-Point Interpolated Plots. Precision and recall metrics formulation is given in eqn. (14), and (15).

$$Precision = \frac{\#.relevant\ images\ retrieved}{sum\ of\ images\ retrieved} \tag{13}$$

$$Recall = \frac{\#.relevant\ images\ retrieved}{sum\ of\ relevant\ images\ in\ the\ database} \tag{14}$$

When we retrieve all images, recall is 1 while precision is high by retrieving only few images.

D. Comparative Results

In this section we detailed the description of the comparison process. We compare the proposed method with several state-of-the-art works include Ma et al. [35], Zhao et al. [36], Xiaoming et al. [37], Wei et al. [38], [39], and Dhara et al. [40].

1) Comparison on APR

Fig 4 represents the comparison of APR between the proposed Triplet-CBMIR system and previous works include WT+TF, DL+PCA, Semi-supervised, LBP+Haralick and Self-Learning Tool.

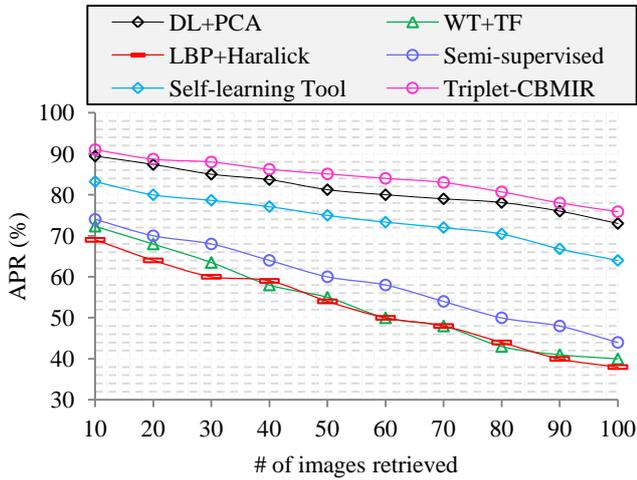


Fig.4 Comparison on APR

APR is a single number (in percentage) that gives the overall performance of the CBMIR system. The higher the value shows the better performance. This metric is increased by the effective process of feature representation, and similarity measurement. From the fig 4 we can see that the APR of our proposed Triplet-CBMIR method gradually increases with the increased number of images. Accordingly our proposed method average APR score is 89%, which is relatively higher than other methods. With the use of triplet

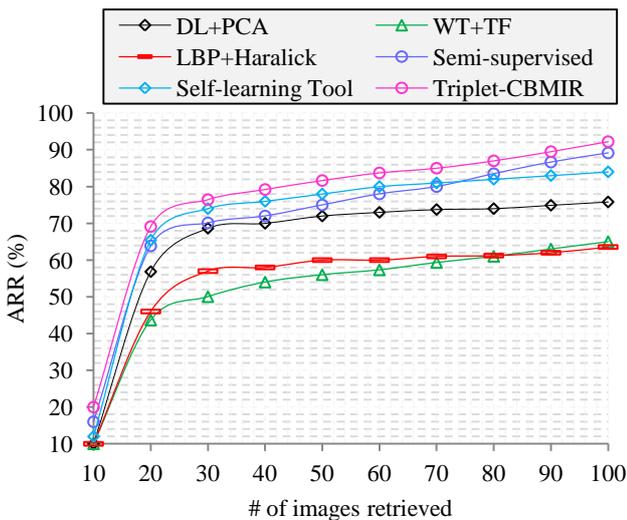


Fig.5 Comparison on ARR

methods such as visual, semantic and relevance feedback, our proposed triplet method has attained better performance. Among the previous methods, self-learning tool has given the foulest performance which does not scale well for CBMIR.

2) Comparison on ARR

In this section we draw the methods of proposed triplet method and previous methods, respectively.

The result of ARR is shown in fig.5. From the fig 5 we can observe that the ARR curve of the proposed method has given the best results than other methods. One of the big reasons of why triplet-CBMIR method can attain the best performance is the proper clustering. Database images are correctly trained and grouped with set of images so it does not lead to any misclassification issues. As a result, our ARR score is the very highest one for both datasets, 85.3%, which

is obviously better than the DL+PCA, WT+TF, LBP+Haralick, Semi-supervised and Self-Learning tool. With the poor development of CBMIR system related to feature extraction, selection and similarity measurement, previous methods are suffered to attain the highest ARR.

3) Comparison on ArR

ArR indicates the performance of lung nodules sign detection. After the images retrieval, the operation of

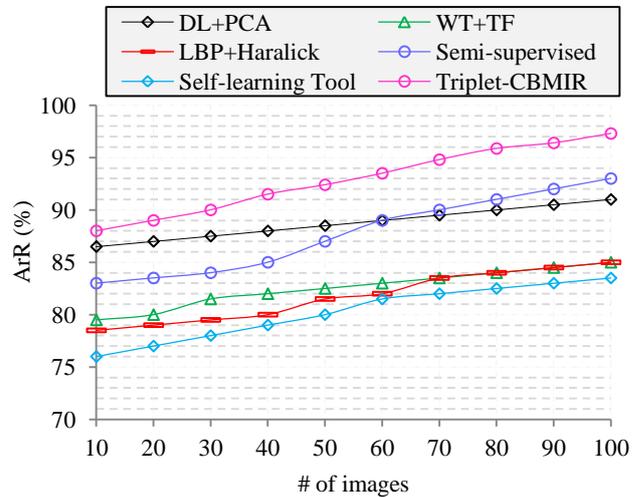


Fig.6. Comparison on ArR

recognition is invoked. For recognition several objectives are needed to be considered. Recognition method must be computationally simple and fast, with very shorter training time. Fig 7 shows the ArR results on the proposed and previous methods.

From the fig 6 we can observe that our proposed triplet-CBMIR method demonstrates the best results than others. In DL with PCA method, KNN algorithm is used. Firstly DL framework is applied for images training and PCA is applied for feature selection. Secondly similar images are retrieved and then recognition is executed using KNN. In KNN, Euclidean distance metric is used which has given incorrect results and also leads to misclassification. To overwhelm the problems, in this paper, we proposed KNN with Hassanat distance. It is used to classify the retrieval images into nine classes. Data analysis in fig 6 indicated that the ArR performance comparison of the proposed method and previous methods. However the ArR performance of the proposed method has reached larger performance, which is very stable for both small scale and large scale dataset.

4) Comparison on Running Time

To demonstrate and verify the robustness of our proposed method on both lung images retrieval and recognition systematically, the performance of the proposed Triplet scheme is compared with the previous methods presented for retrieval. DL with PCA is utilized for image retrieval and recognition. Fig. 7 shows the running time for this scheme is less due to deep learning framework, but indexing images using either clustering or classification is essential, which improves CBMIR feasibility, scalability and reliability.



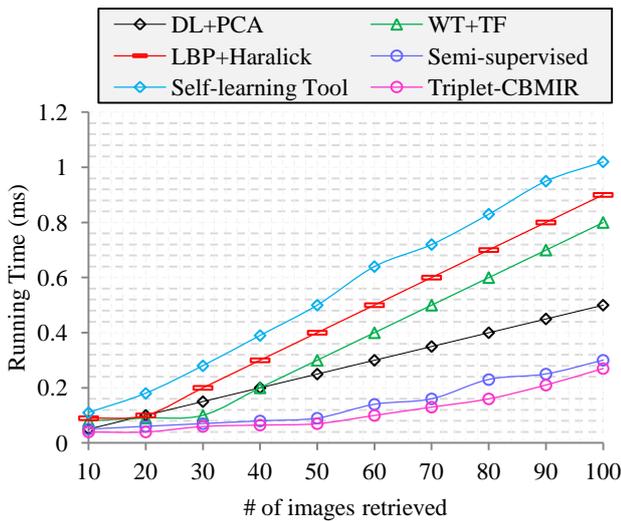


Fig.7. Comparison on Running Time

E. Result Analysis

In this section we explain how the proposed Triplet CBMIR system has achieved good performance than previous CBMIR methods. Firstly we demonstrate the results gain in the implementation and then present the results summary. The preprocessing results (image scaling, denoising and contrast enhancement) are depicted in fig.8. After preprocessing, feature extraction is implemented where we extract visual and

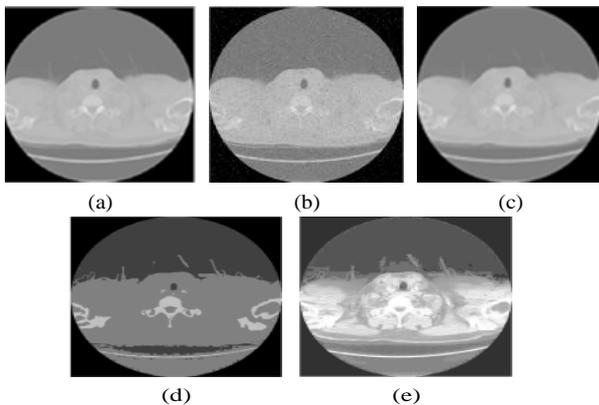


Fig.8 (a). Original Image, (b). Noised Image, (c).Image Scaling, (d).Denoised Image, (e). Contrast Enhancement Image

semantic features using Hybrid DWT and DCT, and Bounding Box-based CNN. In DWT, we decompose the image into third level If feature extraction is implemented, then feature selection and similarity computation is initiated for most similar images retrieval. Fig 9 shows the three level decomposition of lung CT image using DWT.

Fig 10 shows the retrieval results for a query image. For the user given query image, top-k (K=7) images are retrieved. In this stage, we used four different similarity metrics, and retrieved with different set. Since similarity for a query image exist in all images therefore well representation of feature

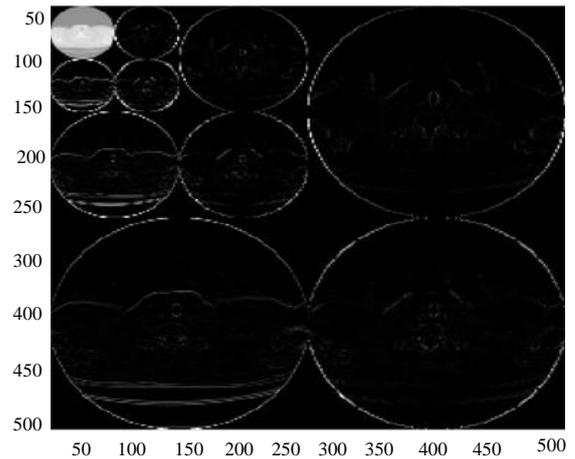


Fig.9 Third Level DWT

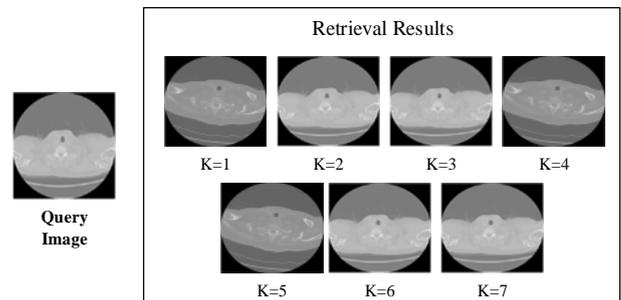


Fig.10 Retrieval Results (K=7) for a query image extraction and distance metric is needed. We compared with those distance metrics for attaining the best performance. Finally we compare our proposed method with the other previous methods in the view of all performance metrics. Table V shows the comparison results of proposed method with some state-of-the-art methods.

Table-V: Average Comparison Results with others existing work.

Methods	APR (%)	ARR (%)	ArR (%)	Running Time (s)
DL+PCA	81.28	64.90	88.75	0.275
WT+TF	53.88	51.93	82.55	0.377
LBP+Haralick	52.60	53.87	81.75	0.459
Semi-supervised	59.00	71.42	87.75	0.143
Self-Learning Tool	74.06	71.54	80.25	0.56
Proposed	84.00	76.38	92.61	0.114

VII. CONCLUSION

In this paper we proposed a novel Triplet-CBMIR for retrieval and recognition of LISS and LIDC-IDS datasets for lung nodules CT images. Our proposed method addresses several issues in CBMIR. Three preprocessing steps (image scaling, de-noising and contrast enhancement) are included to improve the quality of image. Image resizing for 48x48 pixels is implemented using Cubic Interpolation method, this interpolation aids to keep smooth edges and also increase contrast level.

Then Hybrid wavelet decomposition and bounding box based CNN is applied for visual and semantic features extraction, respectively. Visual features are extracted using wavelet decomposition Hybrid DWT and DCT. Then optimum set of features are selected for accurate retrieval and recognition by MINE. Lung nodule sign detection is implemented using KNN-Hassanat Distance metric, which outperforms than KNN. Then similar images (same sign) are grouped into clusters by Multi-SOM. Clustering using Multi-SOM also require $O(1)$ for retrieval, but training we do not require much time. User has given a query image for top-k results retrieval. In order to compute the similarity between the query images and images stored in the database d_distance metric is proposed which compute the distance between feature vectors and it outperforms than three other distance measures such as Canberra Distance, Manhattan, and Euclidean distance. Finally experiments conducted for the proposed method and previous methods; from the results we can be seen that the proposed method has produces higher performance in terms of all performance metrics.

REFERENCES

1. M. Kumar, Kh. M. Singh, "Content Based Medical Image Retrieval System (CBMIRS) to Diagnose Hepatobiliary Images", *Smart and Innovative Trends in Next Generation Computing*, PP. 663-676
2. P. Das, A. Neelima, "An Overview of Approaches for Content-based Medical Image Retrieval", *International Journal of Multimedia Information Retrieval*, Vol.6, Issue.4, PP. 271-280.
3. Q. Sun, Y. Yang, J. Sun, Z. Yang, J. Zhang, "Using Deep Learning for Content-based Medical Image Retrieval", *Medical Imaging 2017: Imaging Informatics for Healthcare, Research, and Applications*, Proc. of SPIE, Vol. 10138, PP. 10113812 (1-11)
4. A. Kumar, S. Dyer, J. Kim, C. Li, Philip H.W. Leong, Michael Fulham, Dagan Feng, "Adapting content-based image retrieval techniques for the semantic annotation of medical images", *Computerized Medical Imaging and Graphics*, Vol. 49, PP. 37-45, 2016
5. S. Mishra, M. Panda, "Medical Image Retrieval using Self-Organizing Map on Texture Features", *Future Computing and Informatics*, Vol. 3, Issue.2, PP. 259-370, 2018
6. J. Y. Priya, B. Nithya, "Fuzzy based Grey Wolf Optimization for Effective Medical Image Retrieval System", *International Conference on Communication and Signal Processing (ICCSPP)*, 2018
7. S. K. Sundararajan, B. Sankaragomathi, D. S. Priya, "Deep Belief CNN Feature Representation based Content based Image Retrieval for Medical Images", *Journal of Medical Images*, Vol. 43, Issue.174, PP. 1-9, 2019
8. L. Shuang, C. Deyun, C. Zhifeng & P. Ming, Multi-feature fusion method for medical image retrieval using wavelet and bag-of-features , *Computer Assisted Surgery*, pp. 1-10, 2019
9. N. Malviya, N. Choudhary, Kalpana Jain, "Content Based Medical Image Retrieval and Clustering based Segmentation to Diagnose Lung Cancer", *Advances in Computational Sciences and Technology*, Vol. 10, Issue.6, PP. 1577-1594, 2017
10. C. Qiu, Y. Cai, X. Gao, Y. Cui, "Medical Image Retrieval Based on the Deep Convolution Network and Hash Coding", 10th *International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2017
11. G. Wei, H Ma, Wei Qian, Xinzhuo Zhao, "A content-based image retrieval scheme for identifying lung nodule malignancy levels", 29th *Chinese Control And Decision Conference*, 2017
12. D. Perez and Y. Shen, Joan Dayanghirang and Jiang Li, Shengli Wang and ZezhongZheng, "Deep Learning for Pulmonary Nodule CT Image Retrieval — An Online Assistance System for Novice Radiologists", *IEEE International Conference on Data Mining Workshops (ICDMW)*, 2017
13. R. Marcos. Nesso-Jr., Mirela T. Cazzolato, Lucas C. Scabora, Paulo H. Oliveira,GabrielSpadon, Jessica A. de Souza, Willian D. Oliveira, Daniel Y. T. Chino,Jose F. Rodrigues-Jr., Agma J. M. Traina, Caetano Traina-Jr. RAFIKI: Retrieval-Based Application for Imaging and Knowledge Investigation. *IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, 2018
14. M. Vinicius N. Bedol & D. Pereira dos Santos1&Marcelo Ponciano-Silva2&Paulo Mazzoncini de Azevedo-Marques3&André Ponce de León Ferreira de Carvalho1&Caetano TrainaJr, "Endowing a Content-Based Medical Image Retrieval System with Perceptual Similarity Using Ensemble Strategy. *Journal of Digital Imaging*", Vol. 29, Issue. 1, PP. 22-37, 2015
15. M. G. Ayadi, R. Bouslimimi, J. Akaichi, "A medical image retrieval scheme with relevance feedback through a medical social network. *Social Network Analysis and Mining*", Vol. 6, Issue.1, 2016.
16. S. Mehre, A. Dhara, A. K. Garg, M., Kalra, N., Khandelwal, N., & Mukhopadhyay, S, Content-Based Image Retrieval System for Pulmonary Nodules Using Optimal Feature Sets and Class Membership-Based Retrieval, *Journal of Digital Imaging*, 2018
17. Ling Pan, Yan Qiang, Jie Yuan, Lidong Wu, "Rapid Retrieval of Lung Nodule CT Images Based on Hashing and Pruning Methods", *Biomed Research International*, PP. 1-10, 2016
18. D. j Ferreira de Lucena, José Raniery Ferreira Junior, AydanoPamponetMachadoand Marcelo Costa Oliveira, "Automatic weighing attribute to retrieve similar lung cancer nodules". *BMC Medical Informatics and Decision Making*, Vol. 16, Issue.S2, PP. 1-15, 2016
19. X. Wang, K. Mao, L. Wang, P. Yang , D. Lu, P. He, An Appraisal of Lung Nodules Automatic Classification Algorithms for CT Images", *Sensors*, Vol. 19, Issue. 1, PP. 194 (1-35)
20. 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 Transactions on Biomedical Engineering*, Vol.62, Issue.2, PP. 648-656, 2015
21. L. Ma, X. Liu, L. Song, C. Zhou, X. Zhao, Y. Zhao, "A New Classifier Fusion Method based on Historical and On-Line Classification Reliability for Recognizing Common CT Imaging Signs of Lung Diseases", *Computerized Medical Imaging and Graphics*. Vol. 40, PP. 39-48, 2015
22. G. He, "Lung CT Imaging Sign Classification through Deep Learning on Small Data", 2019
23. X. Liu, Ling Ma, Li Song, Yanfeng Zhao, Xinming Zhao, Chunwu Zhou, "Recognizing Common CT Imaging Signs of Lung Diseases through a New Feature Selection Method based on Fisher Criterion and Genetic Optimization", *IEEE Journal of Biomedical and Health Informatics*, Vol. 19, Issue.2, PP. 635-647, 2015
24. P. Shamna, V. K. Govindan, K.A. Abdul Nazeer, "Content based Medical Image Retrieval using Topic and Location Model", *Journal of Biomedical Informatics*, Vol. 91, Issue.103112, PP. 1-42, 2019
25. P. Shamna, V. K. Govindan, K. A. AbdulNazeer, "Content-based Medical Image Retrieval by Spatial Matching of Visual Words", *Journal of King Saud University- Computer and Information Sciences*, Open Access, PP. 1-14, 2018
26. R. Hamid. Tizhoosh, Babaie, "Representing Medical Images with Encoded Local Projections", *IEEE Transactions on Biomedical Engineering*, Vol. 65, No.10, PP. 2267-2277, 2018
27. S. Mishra, M. Panda, "Medical Image Retrieval using Self-Organizing on Texture Features", *Future Computing and Informatics Journal*, Vol.3, Issue.2, PP. 359-370, 2018
28. R. Lan, S. Zhong, Z. Liu, Z. Shi, X. Luo, "A Simple Texture Feature for Retrieval of Medical Images", *Multimedia Tools and Applications*, Vol. 77, PP. 10853-10866, 2018
29. A. Qayyum, S. M. Anwar, M. Awais, M. Majid, "Medical Image Retrieval using Deep Convolutional Neural Network", *Neurocomputing*, Vol. 266, PP. 8-20, 2017
30. V. P. Singh, R. Srivastava, "Automated and Effective Content-based Mammogram Retrieval using Wavelet based CS-LBP feature and Self-Organizing Map", *Biocybernetics and Biomedical Engineering*, Vol. 38, Issue.1, PP. 90-105, 2018
31. M. Kumar, Kh. M. Singh, "Content based Medical Image Retrieval System (CBMIRS) to Diagnose Hepatobiliary Images", *CCIS*, Vol. 828, PP. 663-676, 2018
32. H. P. Oliveira, C. Lucas. Scabora, T. C. Mirela, D. Willian . O. S. Rafael. Paixao, A. J. M. Traina, Caetano Traina-Jr, "Employing Domain Indexed to Efficiently Query Medical Data from Multiple Repositories", *IEEE Journal of Biomedical and Health Informatics*, Vol. 99, PP. 1-10, 2018
33. M. Srinivas, R. Ramunaidu, C.S. Sastry, C. Krishna Mohan, "Content based Medical Image Retrieval using Dictionary Learning", *Neurocomputing*, Vol. 168, PP. 880-895, 2015
34. S. R. Dubey, Rajat Kumar Singh, "Local Wavelet Pattern: A New Feature Descriptor for Image Retrieval in Medical CT Databases", *IEEE Transactions on Image Processing*, Vol. 24, Issue.12, PP. 5892-5903, 2015

Triplet Contents based Medical Image Retrieval System for Lung Nodules CT Images Retrieval and Recognition Application

35. Ling Ma, Xiabi Liu, Yan Gao, Yanfeng Zhao, Xinming Zhao, Chunwu Zhou, "A New Method of Content based Medical Image Retrieval and Its Applications to CT Imaging Sign Retrieval", *Journal of Biomedical Informatics*, Vol. 66, PP. 148-158, 2017
36. J. Zhao, L. Pan, P. F. Zhao, X. Tang, "Medical Sign Recognition of Lung Nodules based on Image Retrieval with Semantic Features and Supervised Hashing", *Journal of Computer Science and Technology*, Vol. 32, Issue.3, PP. 457-469, 2017
37. Sun Xiaoming, Zhang Ning, Wu Haibin, Yu Xiaoyang, Wu Xue, Yu shang, "Medical Image Retrieval Approach by Texture Features Fusion based on Hausdorff Distance", *Mathematical Problems in Engineering*, PP. 1-12, 2018
38. Guohui Wei, Hui Cao, He Ma, Shouliang Qi, Wei Qian, Zhiqing Ma, "Content-based Image Retrieval for Lung Nodule Classification using Texture Features and Learned Distance Metric", *Journal of Medical Systems*, Vol. 42, Issue.1, PP.1-7, 2018
39. Guohui Wei, He Ma, Wei Qian, Min Qiu, "Similarity Measurement of Lung Masses for Medical Image Retrieval using Kernel based Semisupervised Distance Metric", *Medical Physics The International Journal of Medical Physics Research and Practice*, Vol. 43, Issue.12, PP. 6259-6269, 2016
40. A. K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, 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, Issue.1, PP. 63-77, 2017
41. R.I. Heaven Rose, A.C. Subajini, "An Efficient Medical Content based Medical Image Retrieval with Lenient Relevance Feedback", *International Journal of Engineering and Technology*, Vol. 7, Issue. 3.6, PP. 175-178, 2018
42. V. B. S. Prasath, H. A. A. Alfeilat, O. Lasassmeh, A. B. A. Hassanat, "Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier A Review", arXiv:1708.04321v1 [cs.LG] 14 Aug 2017

AUTHORS PROFILE



Ranjit Biswas, received his B.Tech. degree in Information Technology in 2007 and M.Tech. degree in Information Technology in 2009. Currently, he is pursuing Ph.D degree at Assam University, Silchar. Currently, he is working as an Assistant Professor in the Department of Information Technology at Ramkrishna Mahavidyalaya, Kailashahar, Unakoti Tripura, India. His areas of interest are digital image processing and content based image retrieval.



Sudipta Roy, is currently serving as Professor in the department of CSE, TSSOT, Assam University. He is presently holding the post of Dean, TSSOT. He has received his MCA from BITS, Mesra and ME degree from Jadavpur University, Kolkata. He received his PhD from Assam University, Silchar, in the year 2010. He has guided many PhD theses and also has completed a number of research projects and Industrial projects. He also have a huge number of national and international publication. His research interest includes, image processing, soft computing, AI and neural networks to name a few.



Abhijit Biswas, is currently working as an Assistant Professor in the department of CSE, TSSOT, Assam University since October 2009. He is a alumni of NIT Bhopal and have received his M.Tech and PhD degrees from NERIST in the year 2009 and 2018 respectively. His research interest includes, network routing, Network – On – Chip topologies and Data Science.