CAD system for lung cancer with remote authentication via biometrics

[1]Ms. C.Rekha, [2]B.Deepa

Asst.Prof.Gr.II/ Department of IT
Sri Sai Ram Institute of Technology,ch-44

***Abstract* — The computer aided diagnosis system for lung cancer is an effective method of assisting radiologist in the early detection of cancer. The objective of this project is to develop a system for diagnosis of lung cancer using image retrieval techniques for retrieval of similar computed tomography images of lung. The proposed framework for lung cancer diagnosis consists of uthentication, noise reduction, segmentation, ROI selection, features extraction and image classification. Initiallynoise reduction is employed using component median filter in the preprocessingstep. Then Main features are extracted from theROI based on Haralick and discretewavelet transformation methods. In this paper, neural fuzzy model is used to extract the diagnosis rules for detecting the pulmonary nodules. Further, retrieval of similar images based on k- nearest neighbour, statistical similarity matching are implemented for diagnosis purposes.Support vector machine technique is used to improve the nodule detection rate. The detection rate of the proposed method is 85 %, and the false positive is approximately 0.3 per image . This result demonstrates that our method improves the detection rate and reduces false positive compared toother approaches. This also implies potential of his system in clinical practice.**

Index Terms: Noise reduction, Lung segmentation ,ROI selection, Rule based neural fuzzy model , K-nearest neighbour (KNN), statistical similarity matching, support vector machine(SVM).

I. INTRODUCTION

Computer Aided Diagnosis Systems allow the detection of diseases through, analysis of various modality images. Diagnosis parameters acquisition is achieved with image processing methods. The system is capable of recognizing and locating diseases. It improves the quality of diagnosis, increases the therapy success by early and avoids unnecessary biopsies.

Ilias et al. introduced classification for the recognition of idiopathic pulmonary fibrosis in microscopic images using Radial Basis Function neural networks classifier. The resulting scores derived by visual assessment of the images by expert pathologists were compared with the RBF and SVM classification outcome. The neural network had a better performance than SVM classifier. The advantages of this Radial Basis Function neural network classification for faster training procedure and better approximation capability.

Binsheng et al. presented a computerized method for automated identification of small lung nodules on CT images. The method consists of separation of the lungs from the other anatomic structures, detection of nodule candidates in the extracted lungs, reduction of false-positives among the detected nodule candidates. A three-dimensional lung mask can be extracted by analyzing density histogram of volumetric chest images followed by a morphological operation.

2. AUTHENTICATION AND METHODOLOGY

The prototype of system for pulmonary nodules detection consists of two stages, the preprocess procedure and lung field images diagnosis procedure. Lung nodule detection is a feature searching and analyzing problem. Therefore, we segment the lung field area from the original images. Then, the region of interest (ROI)selected from lung field. To utilize these ROIs of lung fields, extract the features representing lung nodules. Finally, the neural fuzzy model is employed to obtain the identification rules and to discriminate true or false nodules from ROIs . Further, retrieval of similar images based on k- nearest neighbor, statistical similarity matching are implemented for diagnosis purposes. Support vector machine technique is used to improve the nodule detection rate. Figure 1 shows architecture diagram of diagnosis system of lung. The details of each procedure is illustrated in the following sections.

Chaotic Encryption:

Generate encrypt biometric signals to allow for natural authentication, involve a Chaotic Pseudo-Random Bit Generator (C-PRBG) to create the keys that trigger the whole encryption to increase security, and the encrypted biometric signal is hidden , which can reliably be detected in modern applications that involve teleconferencing. biometric signal is encrypted by incorporating a chaotic pseudo-random bit generator and a chaos-driven cipher.

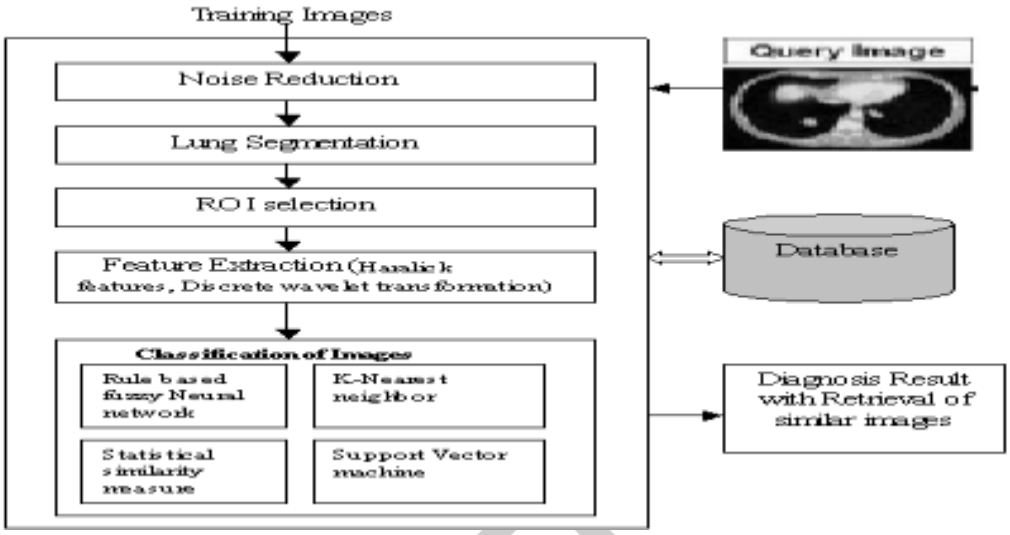


Figure 1. Architecture diagram for Diagnosis system of Lung cancer with authentication

Extraction Process:
The -object (or a distorted version of it) has reached its destination , the encrypted biometric signal is initially extracted by following a reverse (to the embedding method) process. Towards this direction let us assume that the recipient of the -object has also received the size of the encrypted 2-D biometric signal the scaling constants and possesses the original host video object (or he/she has the algorithm to segment it from the initial head-and body image).

QSWT Estimation & Hiding:
Host video object is decomposed into two levels using the shape-adaptive discrete wavelet transform (SA-DWT). By applying the SA-DWT once to an area of arbitrary shape, four parts of low, middle, and high frequencies (LL1, HL1, LH1, and HH1) are produced. Initially the extracted host object is decomposed into two levels by the separable 2-D wavelet transform, providing three pairs of sub bands (HL2, HL1), (LH2, LH1) and (HH2, HH1). Afterwards, the pair of sub bands with the highest energy content is detected and a QSWTs approach is incorporated in order to select the coefficients where the encrypted biometric signal should be casted. Finally, the signal is redundantly embedded to both sub bands of the selected pair, using a non- linear energy adaptable insertion procedure.

Noise Reduction

Noise removal based on component median filter. Initially filter the image with a 3 x 3 component median filter (CMF), after zero padding at the image borders. Pad Array method is to pads the array with 0's (zeros). Pad size is a vector of positive integers that specifies both the amount of padding to add and the dimension along which to add it. The value of an element in the vector specifies the amount of padding to add. Median is to calculate the median value, which is the 50th percentile of a sample. The median is a robust estimate of the centre of a sample of data. For matrices, median (X) is a row vector containing the median value of each column. Since median is implemented using sort.

Lung segmentation

The main task is to segment the lung field on CT images. An unsuitable segmentation process of lung field may cause damage to the ROI (region of interest) that we shall place intention. Therefore, we must preserve the features of the lung field as completely as possible during the segmentation process, and avoid artifacts. Several techniques have been proposed for automatic segmentation in medical images, such

as thresholding, edge detection, region growing, and morphological operations.
Voxels having a density lower than the threshold values were recognized as lung candidates and assigned the value of 1 and appear white, whereas other voxels are assigned the value of 0 and appear black. Due to their low densities, both the lung parenchyma and background will be classified as the ‘‘lung’’ on the resultant binary images. As the lung parenchyma is usually completely isolated from the background by the chest wall, it was readily determined by labeling 3D connected components. To obtain a complete, lung mask, morphological operation as applied. To compute a global threshold (level) that can be used to convert an intensity image to a binary image with im2bw. Level is a normalized intensity value that lies in the range [0, 1]. The gray thresh function uses Otsu's method, which chooses the threshold to minimize the intra class variance of the black and white pixels.

Median filtering is a nonlinear operation often used in image processing to reduce noise. Median filtering is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges. Dilates the grayscale, binary, or packed binary image to dilated image. The structuring element object, or array of structuring element objects is returned by the strel function. Continuously, we apply morphology closing (dilation and erosion) on the image to fill the indentation caused by the pulmonary vessels. Besides, to fill some large indentations within the lung region, region growing is applied. Finally, the lung field is extracted from slice image.

Region of Interest (ROI) selection

The goal of this module is to detect ROI, which may represent lung cancers at earlier and potentially more curable stages. Lung nodule was detected based on region selection and connected components methods.
Step 1: Region selection.
It returns the region of interest selected by the polygon described by vectors . The output image is the same size as input image with 0's outside the region of interest and 1's inside. Pseudo code for nodule detection is shown below.
{
Load input image;
Select a region of interest within an image; Returns a binary image that use as a mask for masked filtering;
Returns an m-by-n matrix of zeros;
Set output image as 0's outside nodule and 1's
nodule;
}
Output: ROI images.

Features Extraction

Nodule features namely orientation, texture, mean, variance were extracted .Texture feature extracted using discrete wavelet transform. It performs a single-level two- dimensional wavelet decomposition with respect to either a particular wavelet filters (Lo_D and Hi_D) .

2.4.1 Haralick Texture Features

There has been considerable work on texture features since 1960 and many theoretical models have been proposed. The most well known texture feature extraction algorithm is the Grey Level Co occurrence Matrix method developed by Haralick . A list of haralick features is in table 1. Assuming we are working with images that have 256 grey levels, this method involves first generating a grey level co-occurrence matrix with 256(i) columns and 256(j) rows. An entry in the matrix is the frequency of occurrence of pixels with grey level i and j level separated by a displacement d in a particular orientation. Most work uses a displacement of 1. There are four principal orientations namely 0^{o}; 45 ^{o}; 90 ^{o} and 135 ^{o} so four matrices are generated. Thirteen second order statistical features namely energy, entropy, correlation, contrast, homogeneity, variance, sum mean, Inertia, shade, tendency, Inverse variance, max probability are then calculated from each matrix. The features for the four principal orientations are then averaged giving a total of 13 Haralick features.

2.4.1.1 Grey Level Co-occurrence Matrix (GLCM)

GLCM is a tabulation of how often different combinations of pixel brightness (grey levels) occur in an image. Basically, GLCM considers the relation between two neighboring pixels at a time called the reference and the neighbor pixel. The grey value relationships in a target are transformed into a co- occurrence matrix space.
A co-occurrence matrix, also referred to as a co-occurrence distribution, is defined over an image to be the distribution of co-occurring values at a given offset. Mathematically, a co- occurrence matrix **C** is defined over an **n x m** image **I**, parameterized by an offset **(Δx,Δy)**.The 'value' of the image originally referred to the grayscale value of the specified pixel. The value could be anything, from a binary on/off value to 32-bit color and beyond. Note that 32-bit color will yield a $2^{32}x2^{32}$ co-occurrence matrix.It is also possible to define the matrix across two diffrent images.

Neural Fuzzy Model

The final procedure of the proposed system is to confirm the suspicious region and determine if it is a true nodule utilizing features obtained from previous stages. Rule-based criteria and neural network are popular methods. The rule based criteria, used with fuzzy inference in general, must predefine the fuzzy membership functions. It is hard to define the suitable rules for all of the cases need. Due to these drawbacks and uncertainty of selected features, established neural fuzzy model to classify the lung nodules as shown in Fig. 2**.** The learning mechanism and high computational power of neural networks are brought into fuzzy inference system. This model can revise the membership function of

Rule k = min(fp1, fq2, fr3)

where fp1, fq2, and fr3 is one of the membership value of features connected to the k th rule.

2.3.4. Defuzzification layer

In the defuzzification step, a weight parameter W_{ij} is assigned for each rule. The weight W_{ij} are random values initially .Then take the sum of product of each rule's output and their
associate weights to compute the defuzzification output through sigmoid function $f(0_X)$. The equation is computed as follows:
0x = sum of(Rule k WK) f(0x) = 1/[1+exp(-0x)]
each feature And the corresponding weights of defuzzification by itself. This model is organized layer by layer including input layer, fuzzification layer, rule inference layer. and defuzzification layer. The function of each layer is illustrated in the following sections.

Input layer

To applied features, mean, variance ,haralick 12 features , discrete wavelet transformation features, orientation to the input layer. The input nodes I_1, I_2, I_3, I_4, I_5, I_6 are values of mean, variance , discrete wavelet transformation features as mean, variance, orientation, haralick features calculated from image respectively, and are normalized to speed up the training process.

2.3.2 Fuzzification layer

The high membership function is formed by one positive sigmoid function, and median membership function is formed by the summation of two sigmoid functions, one positive and one negative. The symbol Fj is a sigmoid function defined as

Fj=1/[1+exp (-Ii X Wij)]
Where Ii is normalised feature value of either one of the feature value of the input nodes and Wij is the weight parameter on the connection between input Ii and sigmoid function j, j=1,2,3,…30. W_{ij} is assigned randomly at the beginning.

2.3.3. Rule inference layer

After obtaining eighteen membership functions from the previous fuzzification layer, at most six inference rules can be generated by one of the membership values from each feature. For each rule, the minimum membership value is the output of the inference rule
Support vector machines **(**SVM**)**

SVMs are a set of related supervised learning methods used for classification and regression. Viewing input data as two sets of vectors in an *n*-dimensional space, an SVM will construct a separating hyperplane in that space, one which maximizes the margin between the two data sets. To calculate the margin, two parallel hyperplanes are constructed, one on each side of the separating hyperplane, which are "pushed up against" the two data sets.
Formalization

To given the training data, a set of points of the form, where the c_i is either 1 or −1, indicating the class to which the point belongs. Each is a *p*-dimensional real vector. To give the maximum-margin hyperplane which divides the points having $c_i = 1$ from those having $c_i = -1$. Any hyperplane can be written as the set of points satisfying Maximum-margin hyperplane and margins for a SVM trained with samples from two classes. Samples on the margin are called the support vectors.The vector is a normal vector: it is perpendicular to the hyperplane. The parameter determines the offset of the hyperplane from the origin along the normal vector .

Statistical Similarity measurement of features

Features are used to capture the information in the image in a large number of points. Similarity between images can then be checked in several different way. In the first place , the texture as haralick features, discrete wavelet transformation features,and local features are used to identify the points of the data as identical to the points of the query. A measure of similarity between the feature values measured resulting from ROI images consists of a Mahalanobsis distance between the feature vector composed of the features.
The features of the ROI images are collected in a vector F, the distance between q and d is given by

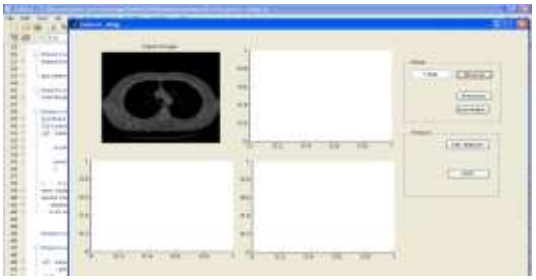
$$D_{q,d} = [(Fq-Fd)^T \; E^{-1} \; (Fq-Fd)]^{1/2}$$

Where E is a diagonal weights matrix set by the user. The strategy is used when comparing texture features derived from different ROI images. This model has been used in retrieval of similar images of the query image.

3. RESULTS AND DISCUSSION

Authentication &Noise Reduction

Filter the image with a 3x 3 component median filter, after zero padding at the image borders. The input image is displayed in figure 3.1.



4. CONCLUSIONS AND FUTURE WORK

The diagnosis process is automated in this paper.This is achieved through by series techniques as noise reduction, lung segmentation, and ROI selection and features extraction of the lung CT-scan images. The system must extract features from the lung CT - scan images using efficient methods for the diagnosis of lung cancer with effective retrieval of images.

Based on the features extracted, the images can be classified using different classifier as K-Nearest Neighbour classifier, Statistical distance measure, Support vector machine and Neural network classifier. Neural fuzzy model was designed to extract suitable diagnosis rules, and classify the true nodules from the ROIs. Meanwhile, the false positive per image was reduced as low as 0.3. The retrieval performance, which clearly showed the advantage of similar images based on K-Nearest Neighbour classifier and Statistical distance measurement.

Furthermore, identification method can be implemented in on- line diagnosis system efficiently and the results of these classifiers are correlated to improve the retrieval efficiency.

REFERENCES

[1] Xiaofan Zhang, WeiLiu, Murat Dundar Sunil Badve, andShaoting Zhang(2015) “Towards Large-Scale Histopathological Image Analysis: Hashing-Based Image Retrieval ”IEEE Transaction on Medical Imaging.V0l.34,No.2, pp 496-506

[2] Klimis ntalianis, and nicolas tsapatsoulis(2015) “Remote authentication via biometrics: a robust ideo-object steganographic mechanism over
Wireless networks” IEEE transactions on emerging topics in computing , January 2015

[3] A. Smeulder, M.Worring, S. Santini, A.Gupta, And R. Jain, “Content Based Image Retrieval At The End Of The Early Years ,” IEEE Transactions Pattern Analysis and Machine Intelligence , Vol. 22, No. 12, pp. 1349–1380.

[4] Ilias Maglogiannis , Haralambos Sarimveis, And C.T.Kiranoudis (2008), "Radial Basis Function Neural Networks Classification For The Recognition Of Idiopathic Pulmonary Fibrosis In Microscopic Images" , IEEE Transactions On Information Technology In Biomedicine, Vol 12. pp. 42-54.

[5] Jia Tong, Zhao Da-Zhe,Wei Ying, And Wang Xu (2007), "Computer-Aided Lung Nodule Detection Based On CT Images" , IEEE International Conference On Complex Medical Engineering, pp. 816-819.

[6] Md. Mahmudur Rahman, Prabir Bhattacharya, And Bipin
C. Desai (2007),"A Framework For Medical Image Retrieval Using Machine Learning And Statistical Similarity Matching Techniques With Relevance Feedback", IEEE Transactions On Information Technology In Biomedicine, Vol. 11, No. 1, pp. 58-69.

[7] M. M. Rahman, P. Bhattacharya, And B. C. Desai (2005), “Probabilistic Similarity Measure In Image Databases With SVM-Based Categorization & Relevance Feedback”, In International Conference on Image Analysis and Recognition, Springer Lecture Notes In Computer Science, Vol. 3656, pp.601–608.

[8] O. Chapelle, P. Haffner, And V. Vapnik (2003), “SVMs For Histogram-Based Image Classification”, IEEE Transaction On Neural Network, Vol. 10, No. 5, pp. 1055–1064.

[9] T. Lehmann, M. Guld, C. Thies, B.Fischer, K.Spitzer, D.Keysers, H.Ney, M. Kohnen, H. Schubert, And B. B.Wein (2004), "Content-Based Image Retrieval In Medical Applications" , Methods Information on Medical, Vol. 43, No. 4, pp. 354–361.
[10] Qiang Li, Shusuke Sone, And Kunio Doi (2003), "Selective Enhancement Filters For Nodules, Vessels, And Airway Walls In Two And Three Dimensional CT Scans" , American Association Of Physicists In Medicine, pp. 2040-2051.

Content-Based Image Retrieval", IEEE Transactions On Circuits System Of Video Technology, Vol. 8, No. 5, pp. 644–655.
[11] Yu-Li You And M.Kaveh (2000),"Fourth-Order Partial Differential Equations For Noise Removal", IEEE Transactions On Image Processing, Vol. 9, No. 10, pp. 1723-1730.