Lung Cancer Recognition using Radon Transform and Adaptive Neuro Fuzzy Inference System

M. Obayya
Electronics and Communication
Engineering Department
Faculty of Engineering
Mansoura University
Mansoura, Egypt

Mohamed Ghandour
Electronics and Communication
Engineering Department
Faculty of Engineering
Mansoura University
Mansoura, Egypt

ABSTRACT

In this paper, we represent computer aided diagnosis (CAD) system for recognition of lung cancer by analyzing CT images of chest. CAD system helps to improve the diagnostic performance of radiologists in their image interpretations. The proposed system relies on three stages mainly; firstly, the CT image is enhanced. Secondly, the lung and tumor are segmented from the input CT image by separating them from other organs in the CT scan. This is done using region growing Algorithm for segmenting the lung parenchyma and a set of morphological operations to detect the tumor. Thirdly, the geometrical information and transformed based features such as Radon transform based features obtained from the extracted tumor are used to classify the lung tumor into benign and malignant employing adaptive neuro fuzzy inference system (ANFIS) classifier. Correct Classification rate of 98% is obtained by using geometric features.

Keywords
Radon transform, lung cancer, region growing algorithm, ANFIS classifier.

1. INTRODUCTION
Lung cancer is the deadliest cancer in the world for both men and women. Lung cancer has surpassed breast cancer as the leading cause of cancer deaths in women [1]. Various treatment options are used for lung cancer patients such as surgery, radiotherapy and chemotherapy. If lung nodules can be identified accurately at an early stage, the patient's survival rate can be increased by a significant percentage .According to GLOBOCAN 2012 statistics, it was found that around 14.1 million new cases were diagnosed and around 8.2 million deaths occurred in 2012, due to cancer which is quiet high when compared to statistics of 2008 which was 12.7 million new cases and 7.6 million deaths due to cancer. According to statistics released by the National Cancer Institute, lung cancer cases in Egypt represent 8.2 percent of total cancer cases in men and about 2.4 percent of total cases among women. According to CanSurvice, it was found that 4488 deaths occurred out of 5017 diagnosed cases in 2012. Tumor is a mass produced as a result of uncontrolled growth of the lung tissues. It is mainly classified as benign and malignant. Benign tumors are noncancerous cells, but if they can harm other tissues if not treated in time; while Malignant tumors are cancerous cells. The differentiation of tumors can be performed by the help of statistical and shape features tumors. For example, malignant nodules tend to be more complex and irregular whereas benign tend to be rounder with well-

defined borders. The malign nodules, however, showed relatively higher variance values. Fig.1 shows this difference.
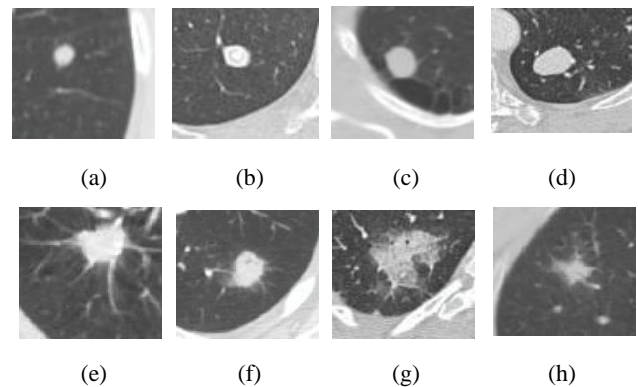

Fig.1 Examples of pulmonary nodules: (a, b, c, d) benign lung nodules, (e, f, g, h) malignant lung nodules.

Computerized tomography (CT) is the most frequently used imaging technique in the diagnosis of lung cancer. Nodules and pathological residues with varied diameter can be comfortably viewed by CT [2]. Nodules on the lung are classified as benign or malignant. During diagnosis, malignant nodules that are solid and atypical can be assessed as benign in some cases. However, in most cases, a solid nodule is usually classified as malignant [3]. It is crucial to diagnose nodules at early stages in order to accelerate the treatment process.

In the literature, there are some studies regarding early diagnosis of lung cancer and identification of nodules. In [4], a CAD system is designed to provide segmentation of nodules on the lobes with neural networks model of Self-Organizing Maps (SOM) and ensures classification between benign and malignant nodules with the help of ANN using GLCM features. Performance values of 90.63% accuracy and 92.30% sensitivity were acquired in the CAD system which utilized a total of 128 CT images obtained from 47 patients. In [5], a proposed method to classify the nodules based on 5 clinical parameters, using ANN and Logistic regression models; the accuracy was found out to be 84.6%, the results were compared by ROC curve analysis. In [6], a computer-aided classification method in CT images of lungs developed. Classification process is done by feed forward and feed forward back propagation neural networks and statistical features. This work resulted in 93.3%, specificity of 100% and

sensitivity of 91.4%. In [7], the effectiveness of geometric feature descriptors is examined , for false positive reduction and for classification of lung nodules in low dose CT (LDCT) scans using active appearance models (AAM); which are then used to detect candidate nodules based on optimum similarity measured by the normalized cross -correlation (NCC). In [8], weighted multi-scale convergence index (WMCI) and fisher linear discriminant (FLD) were combined. The system was evaluated on 154 chest radiographs from JRST database with 100 malignant and 54 benign nodules which results in the correct classification of 78.1% of the nodules in the JRST test set and average of 4.0 false positives per image.

The rest of this paper is organized as follows; Section 2 data collection. Section 3 briefly explains the methodologies adopted. Section 4 presents the experimental results and Section 5 gives the concluding remarks.

2. DATA COLLECTION

Performance of the proposed CAD system in the present study is evaluated using the LIDC database [9]. It is publicly available in the National Biomedical Imaging Archive (NBIA), and its nodules have been fully annotated by multiple radiologists. A total of 100 CT images of 512×512×8 bit from 50 different patients were collected. Based on the annotation Data, 50 of these nodules were benign and 50 were malignant. All images are available in Digital Images and Communication in Medicine (DICOM) format.

3. METHODS

The proposed system is divided into 4 phases as shown in Fig.2. The first part is image enhancement to improve the quality of the input image. The second part is image segmentation to extract the tumor. The third part is feature extraction and selection to extract the main features of the tumor object using both geometric and transformed domain information of the tumor. The fourth part is the ANFIS classifier to distinguish the tumor whether it is benign or malignant.
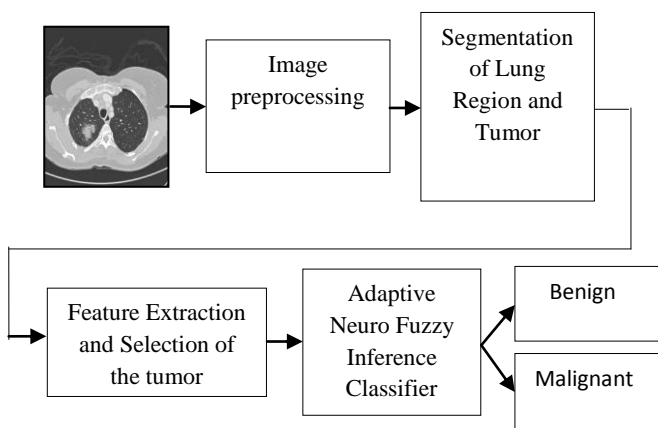

Fig.2. Block diagram of the proposed system

3.1. Image Enhancement

Medical images are corrupted with noise and artifacts due to body movements. Thus wiener filter [10] of mask size 3*3 is used to remove noise as shown in Fig.3.
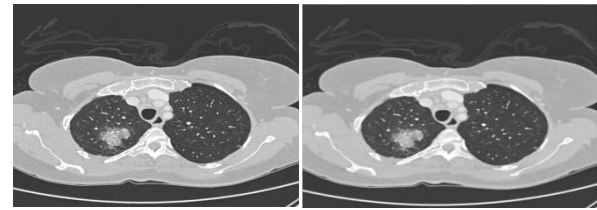

Fig.3 (a) Raw chest CT image. (b) Filtered Image using 3×3 wiener filter

3.2 Segmentation of Lung Region and ROI

Segmentation is the process of partitioning an image into distinct non overlapping regions. In this study, Segmentation is based on region growing Algorithm for segmenting the lung object and a set of morphological operations to detect the tumor. Fig.4. (a-f) shows the steps of segmenting both lung and tumor.

3.2.1 Region Growing Algorithm

Region growing is one of the most popular techniques for segmentation of medical images due to its simplicity and good performance [10]. Region growing algorithm starts with a seed pixel, examines neighboring pixels that surrounds it, and, if it meets certain criteria, it is included in the region. This process is followed until no additional pixels can be added. The region is grown by examining all 4-neighboring pixels to the region. The difference between a pixels intensity value and the regions mean is used as a similarity measure, and the pixel with the smallest difference measured this way is allocated to the iteratively grown region, so the region growing process is completed when the intensity difference between region mean and new pixel is larger than a certain threshold [11]. For the region growing to be effectively finished, the crucial part is the position of the seed pixel, and it must be selected from which the region growing may start. Therefore, selecting a good set of initial seeds is very important. The first goal is segmenting out lung from dark area that doesn't belong to lungs, so that we must select a set of seed points from dark region. The dark regions of lung parenchyma are tagged before applying region growing. Once the seed pixels are selected, region growing start marking all pixels which lie in the 4-connected neighborhood. This process is repeated for each pixel until all black pixels (background pixels) are marked. This will delete dark regions from whole image.

3.2.2 Edge Detection

Edge detection is a technique in which the points where image brightness changes sharply or formally are identified. These points are organized under line segments called edges. Edge detection also aims to classify and place discontinuities in an image [12]. In this work, Canny edge detection algorithm [13] is utilized to find the boundaries of the region growing image.

3.2.3 Morphological operations

After edge detection, morphological binary closing is applied. The morphological close operation is a dilation followed by erosion, using the same structuring element for both operations. Disk structuring element of radius 3 pixels is used to preserve the circular nature of the lung regions then a filling operation is applied on the area inside the lung parenchyma boundaries to create the binary mask which is then multiplied by the input image to obtain the lung parenchyma with its original gray levels. Finally, threshold method [14] followed by opening morphological operation to

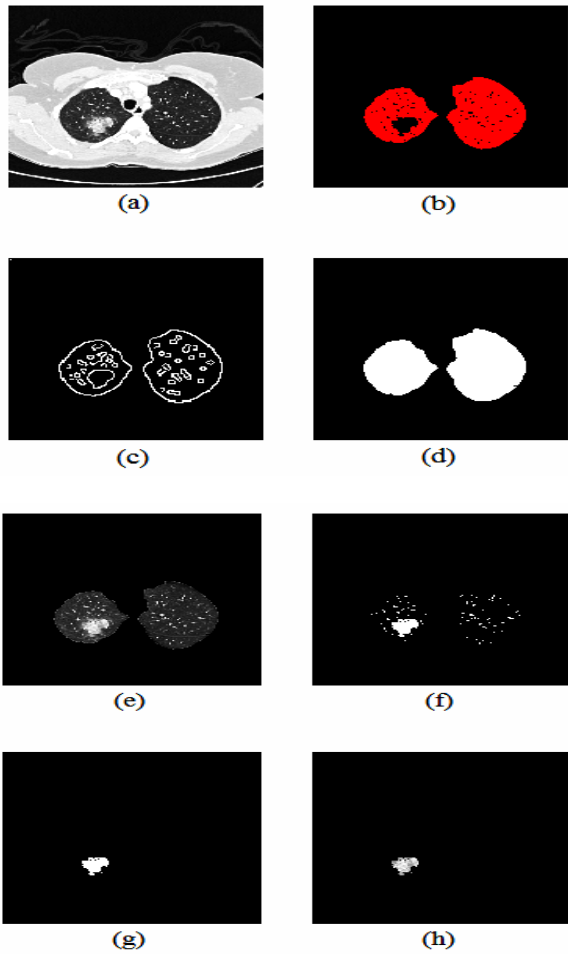rid of the small white areas in the binary image and detect the tumor.

Fig.4: (a) input image (b) region growing image (c) canny edge detection (d) filled image (e) lung parenchyma (f) applying threshold (g) extracted binary tumor (h) the extracted tumor with its original grey levels

3.3 Feature Extraction and Selection

The third step is extracting the most salient features of tumor and reducing the dimensionality of acquired data [15]. Two sets of features were utilized and their performance was evaluated. The techniques used to calculate Features were: (1) geometric features (2) Features derived from radon transform [16].

3.3.1 Geometric Features

Geometric features or structural features describe the shape of the ROI. Malignant nodules tend to be more complex and irregular whereas benign tend to be rounder with well defined borders. That's why geometrical features method is a powerful tool for distinguishing benign from malignant tumors. Features used in this study are:

1. **Area**: Area of the segmented tumor is computed by counting the number of pixels which have the value "1" in a binary image.

2. **Convex Area:** It gives the number of pixels in convex image of the ROI.

3. **Perimeter:** It is the distance between each adjoining pairs of pixel around the border of the ROI.

4. **Equivalent Diameter:** It is the diameter of a circle with the same area as the ROI.

$$E_{diameter} = \sqrt{\frac{4*Area}{\pi}} \qquad (1)$$

5. **Solidity**: It is the proportion of the pixels in the convex hull that are also in the ROI.

$$Solidity = \frac{Area}{Convex\ Area} \qquad (2)$$

6. **Irregularity Index**: is the measure of how closely the shape of an object approaches that of a circle.

$$I = \sqrt{\frac{4\pi*Area}{(Perimeter)^2}} \qquad (3)$$

Table 1 shows an example of the Geometric features values calculated from a benign tumor and a malignant tumor.

Table 1. Geometrical Features

Geometric features	Benign	Malignant
Area	378	476
Convex Area	386	544
Perimeter	21.9382	24.6183
Equivalent Diameter	0.97927	0.875
Solidity	71.9411	103.154
Irregularity Index	0.9178	0.56214

3.3.2 Radon transform based features

The Radon transform computes projections of an image matrix along specified directions. A projection of a two-dimensional function $f(x, y)$ is a set of line integrals. The Radon function computes the line integrals from multiple sources along parallel paths, or beams, in a certain direction. The beams are spaced 1 pixel unit apart. To represent an image, the radon function takes multiple, parallel-beam projections of the image from different angles by rotating the source around the centre of the image [17]. Projections can be computed along any angle θ, by use general equation of the Radon transformation.

$$R_\theta(x') = \iint_{-\infty}^{\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - x')dx\,dy \qquad (4)$$

Where $\delta(\cdot)$ is the delta function with value not equal zero for augment equal 0,

$$x' = x\cos\theta + y\sin\theta \qquad (5)$$

x' is the perpendicular distance of the beam from the origin and θ is the angle of incidence of the beams.

The Radon-pixel image brings a large advantage to achieving global geo-metric affine invariance. This is because Radon pixel image has more geometric information than the original pixel image. It can be seen that one Radon-pixel corresponds to a to a line segment which needs two pixels in the original image to describe. Second, a single Radon-pixel contains the information of a line segment in the original image. This property makes Radon-pixels more robust to image noise and it also fits our motivation of using global features to describe textures [18]. An example showing difference between Radon transformed image of a benign and malignant tumor is shown in fig.5.
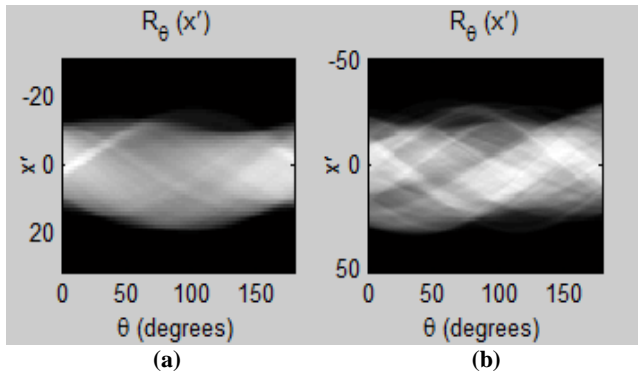

Fig.5 (a) Radon transform of a benign tumor (b) Radon transform of a malignant tumor

In this study, we have computed the Radon transform of the tumor image at 60 different directions ranging from 0-180 degrees with the interval of 3 degrees. To produce feature vector we don't use all values from the Radon transformed image. The reduction the transformed image is possible by the resizing operation. There are several interpolation techniques that have been documented in the past. The widely used techniques are nearest neighbor, bilinear, bicubic, B-splines, lanczos2, discrete wavelet transform, Kriging [19-21]; Bicubic interpolation technique is used in this study to obtain a scaled version of the Radon transformed image of size 8×8. The next step of vector features preparing is concatenate operation in which the rows of pixels in the resized image are concatenated resulting in a single row with 64 elements. The final step is removing the low intensity value pixels in both ends of the concatenated vector so as to minimize the size of the feature vector up to 16 elements. Fig.6 shows the steps of preparing the feature vector.
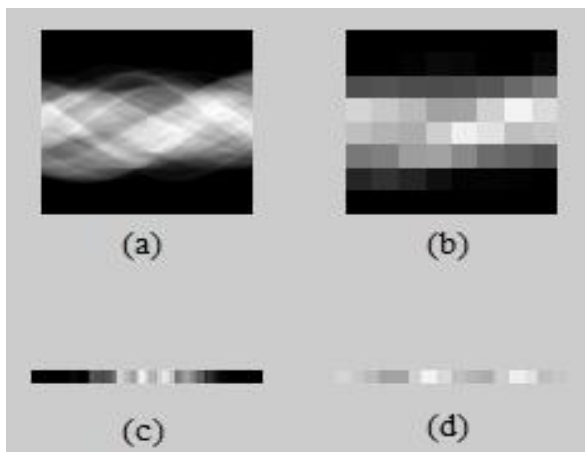

Fig.6 (a) Radon transform image (b) resized image 8×8 (c) concatenated image (d) final feature vector

3.4 Classification Using Artificial Neuro Fuzzy Interference System (ANFIS) classifier

ANFIS is one of hybrid Neuro-Fuzzy Inference Expert Systems and it works in Takagi-Sugeno-type fuzz inference system, which was developed by Jang [22]. ANFIS is a Fuzzy Inference System formulated as a Feed-Forward Neural Network. Hence, the advantages of a fuzzy system can be combined with a learning algorithm. ANFIS architecture consists of five layers of nodes. Out of the five layers, the first and the fourth layer consist of adaptive nodes there are fuzzification and defuzzication, while the second, third and fifth layers consist of fixed nodes there are Rules, Normalization, and Summation neuron. The adaptive nodes are associated with their respective parameters, get duty updated with each subsequent iteration while the fixed nodes are devoid of any parameters. To present the ANFIS architecture, two fuzzy if then rules based on a first order Sugeno model are considered: [23]

$$Rule\ 1: If\ (x\ is\ A1)and\ (y\ is\ B1)then$$
$$(f1 = plx + qly + r1)$$
$$Rule\ 2: If\ (x\ is\ A2)and\ (y\ is\ B2)then$$
$$(f2 = p2x + q2y + r2)$$

Where X and Y are predefined membership functions, Ai and Bi are membership values, fi are the outputs and pi, qi and ri are the consequent parameters that are updated in the forward pass in the learning algorithm. The ANFIS architecture to implement these two rules is shown in Fig.7, in which a circle indicates a fixed node, whereas a square indicates an adaptive node.

ANFIS in MATLAB implements two methods to identify the interference parameters, namely grid partitioning and subtractive clustering. Grid partitioning is a subjective approach, in which the number of clusters, into which each variable is segmented, is initially provided by user. The number of rules is determined by multiplying the number of clusters of each input variable.
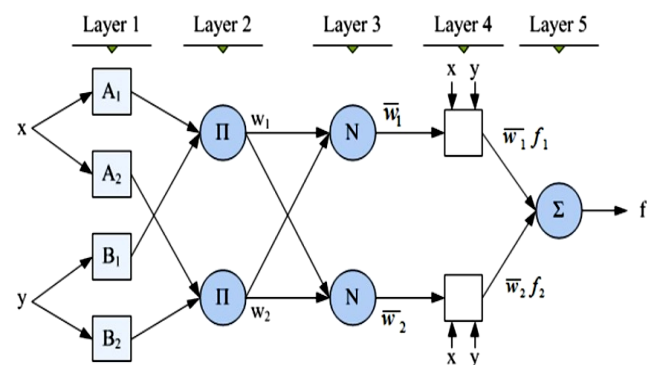

Fig.7 ANFIS Architecture for a two rule Sugeno System

This introduces dimensionality problem, in which as the number of variables increase the number of rules also increase. Subtractive clustering, proposed by Chiu [1994], is much more intuitive and does not suffer from dimensionality issues. In that, the rule extraction method first determines the number of rules and antecedent membership functions and then uses linear least square estimation to determine each

rule's consequent equations. Subtractive clustering method of ANFIS identifies interference parameters of TSK fuzzy interference system structure. It should be noted that, in subtractive clustering ANFIS method, one fuzzy set is identified for each input variable in each rule [24]. The two ANFIS parameter optimization method options available for FIS training are hybrid (the default, mixed least squares and back propagation) Hybrid Learning algorithm was used for all our ANFIS classifiers in training and subtractive clustering method was used to avoid the dimensionality problem of grid clustering method. The number of epochs is determined according to the accepted error measure, fixed by user in the present study.

Table.2 Comparison of ANFIS Structures for the Proposed System

Features	No. of Inputs	No. of fuzzy rules	No. of training epochs	Test error
Geometric	6	2	10	0.18381
Radon transform	16	3	12	0.39711

4. RESULTS

The output plots between the training and test data of the ANFIS classifier is represented in Fig.8 for geometric features, and Radon transform based features with the misclassified tumors enclosed in green circles. In geometric based features Experiment, as shown in Table 3, all the 25 benign tumors were correctly identified as benign (TN), zero tumors were misclassified as malignant (FP). On the other hand, 24 tumors were correctly identified as malignant (TP) and only 1 tumor was misclassified as benign (FN). In Radon transform based features Experiment, as shown in Table.4, 23 tumors were correctly identified as benign (TN), 2 tumors were misclassified as malignant (FP). On the other hand, 20 tumors were correctly identified as malignant (TP) and 5 tumors were misclassified as benign (FN).

(a)

(b)
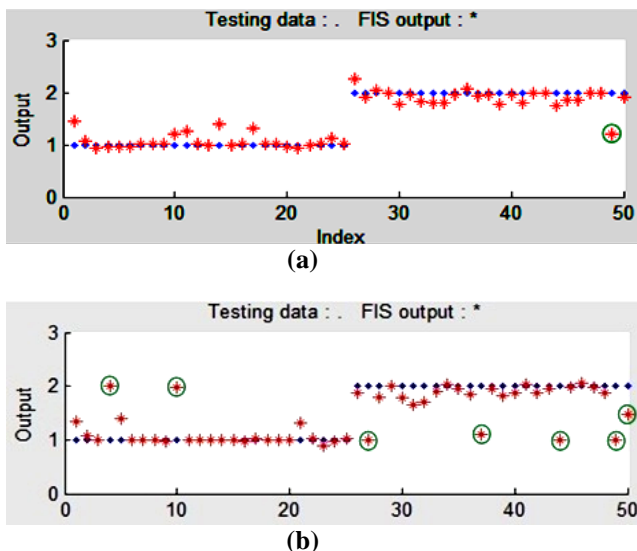
Fig.8 (a) ANFIS output of training data against test data using geometrical features (b) ANFIS output of training data against test data using Radon transform based features.

Table.3 Confusion Matrix Of Obtained Results From Test Set using geometric features

CAD System	Positive	Negative
Positive(Malignant)	24	1
Negative(Benign)	0	25

Table.4 Confusion Matrix of Obtained Results from Test Set using Radon based features

CAD System	Positive	Negative
Positive(Malignant)	20	5
Negative(Benign)	2	23

To evaluate the performance of the proposed system, the following performance measures were used:

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FB} \times 100 \qquad (6)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \times 100 \qquad (7)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \times 100 \qquad (8)$$

Where TP, TN, FN, FP stands for true positive, true negative, false positive and false negative respectively.

Table.5 Performance Measures of the Proposed System

Feature Extraction Technique	Accuracy	Sensitivity	Specificity
Geometric	98%	96%	100%
Radon transform	86%	80%	92%

5. CONCLUSION

In this study, an automatic CAD system that successfully differentiates the lung nodules as benign or malignant on CT images is presented. The system consists of four main parts: image preprocessing, segmentation of Lung region and ROI, feature extraction and selection, and fuzzy network classifier. The segmentation process is done using Region Growing algorithm besides a set of morphological operations including Adaptive thresholding method for tumor detection. A set of features were extracted using different features extraction techniques. These are: Geometric features which describe the morphological shape of the tumor and transformed features based on Radon transform. The Classification experiment is done using ANFIS classifier. A correct classification rate of 98% is achieved using the geometric features.

6. REFERENCES

[1] American Cancer Society, Cancer Facts and Figureures 2014. Atlanta, GA:American Cancer Society, 2014

[2] M. Dolejsi, Detection of Pulmonary Nodules from CT Scans, Czech Technical University, Faculty of Electrical Engineering,Center of Machine Perception, Prag, 2007.

[3] The international early lung cancer action program investigators,Survival of patients with stage I lung cancer detected on CTscreening, N Engl J Med., 355, pp. 1763-1771, 2006.

[4] E.Dandil, M. Cakiroglu, Z.Eksi, M.ozkan, Oz. Kar Kurt, Ar. Canan, "Artificial Neural Network-Based Classification System for Lung Nodules on Computed Tomography Scans", International Conference of Soft Computing and Pattern Recognition, 978-1-4799-5934-1/14/$31.00 ©2014 *IEEE*

[5] Yongjun WU, Na Wang, Hongsheng Zhang, Lijuan Qin, Zhen YAN, Yiming WU," Application of Artificial Neural Networks in the Diagnosis of Lung Cancer by Computed Tomography", Sixth International Conference on Natural Computation,2010.

[6] J. Kuruvilla, K. Gunavathi, "Lung cancer classification using neural networksfor CT images", Computer Methods and Programs in Biomedicine, vol. 113, pp. 202–209, 2014.

[7] Amal Farag, Asem Ali, James Graham, Aly Farag, Salwa Elshazly and Robert Falk, "Evaluation of geometric feature descriptors for detection and classification of Lung Nodules in low dose CT scans of the chest" ,*IEEE* transactions, PP.169-172, 2011.

[8] R. C. Hardie, S. K. Rogers, T. Wilson, and A. Rogers, "Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs " Medical Image Analysis vol. 12, pp. 240-258, 2008.

[9] Lung Image Database Consortium (LIDC): https://imaging.nci.nih.gov/ncia/login.jsf

[10] N. M. Saad, S.A.R. Abu-Bakar, M. Sobri Muda and A.R.A. Mokji, Automated Region Growing for Segmentation of Brain Lesion in Diffusion-weighted MRI, Proceeding of the international MultiConference of Enginneers an Computer Scientists, vol. 1, IMECS 2012, March 14-16, Hong Kong

[11] Nihad Mesanovic, 2Haris Huseinagic, 3Samir Kamenjakovic, Automatic Region Based Segmentation and Analysis of Lung Volumes from CT Images, IJCST Vol. 4, Issue Spl - 2, April - June 2013

[12] Rupika Rana1 , Ashish Verma, Comparison and Enhancement of Digital Image by Using Canny Filter and Sobel Filter , IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 1, Ver. IX (Feb. 2014), PP 06-10.

[13] J. F Canny." A Computational Approach to Edge Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 8, Nov 1986.

[14] Nobuyuki Otsu, "A Threshold Selection Method from Gray-Level Histogram".

[15] Padma, A. and Sukanesh, R. 2013. "SVM based classification of soft tissues in brain CT images using wavelet based dominant gray level run length texture features", middle-east journal of scientific research 13(7): 883-888.

[16] A. L. Spitz, "Determination of script and language content of document images", IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 19, No.3, pp. 235–245, 1997.

[17] Mirosław Miciak, radon transformation and principal component analysis method applied in postal address recognition task, International Journal of Computer Science and Applications, Techno mathematics Research Foundation© Vol. 7 No. 3, pp. 33 - 44, 2010

[18] S.Sahu , S. Kumar, T. Mohapatra, Digital Image Texture Classification and Detection Using Radon Transform, I.J. Image, Graphics and Signal Processing, 2013, 12, 38-48

[19] F. A. Jassim and F. H Altaany., Image Interpolation Using Kriging Technique for Spatial Data, Canadian Journal on Image Processing and Computer Vision Vol. 4 No. 2, 2013.

[20] R. S Asamwar. K. M. Bhurchandi and A. S Gandhi. , Interpolation of Images Using Discrete Wavelet Transform to Simulate Image Resizing as in Human Vision, International Journal of Automation and Computing, 7(1), 2010, 9-16, DOI: 10.1007/s11633-010-0009-7.

[21] R Roy., M Pal. and T Gulati., Zooming Digital Images using Interpolation Techniques, International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 2, Issue 4, 34-45, 2013, ISSN 2319 – 4847.

[22] Jyh Shing and Roger Jang., "ANFIS: Adaptive-Network-Based Fuzzy Inference System ," computer methods and programs in biomedicine , IEEE Transactions on Systems, University of California,1993

[23] T. M. Nazmy, H. El Messiry, B. Al- Bokhity, "Classification of Cardiac Arrhythmia based on Hybrid System", International Journal of Computer Applications (0975 – 8887) Volume 2 – No.4, June 2010.

[24] Asli Celikyilmaz, I. Burhan Türksen,"Theory and Practice of Uncertain. Programming", 2009. ISBN 978-3-540-89483-4. Vol. 240.