Diagnosis of Lung Disorder Using Immune Genetic Algorithm and Fuzzy logic to Handle Incertitude

T.Illakiya, O.Pandithurai, V. Swetha

Department of Computer Science and Engineering, Rajalakshmi Institute of Technology
*E-mail: pandics@gmail.com,illakiya.t@ritchennai.edu.in

Abstract

In this paper, we present an immune based fuzzy-logic approach for computer-aided diagnosis scheme in medical imaging. The scheme applies to lung CT images and to detect and classify lung nodules. Classification of lung tissue is a significant and challenging task in any computer aided diagnosis system. This paper presents a technique for classification of lung tissue from computed tomography of the lung using the Gaussian interval type-2 fuzzy logic system. The type-2 Gaussian membership functions (T2MFs) and their footprint of uncertainty (FOU) are tuned by immune, genetic algorithm, which is the combination of immune genetic algorithm (GA) and local exploration operator. An immune, genetic algorithm estimates the parameters of the type-2 fuzzy membership function (T2MF). By using immune, genetic algorithm, converging speed is increased. The proposed local exploration operator helps in finding the best Gaussian distribution curve of a particular feature which improves the efficiency and accuracy of the diagnosis system.

Keywords: Lung disorder, Immune Genetic algorithm, Classification, Type 2 fuzzy logic

1. Introduction

Research in computer-aided diagnosis CAD is a rapidly growing dynamic field with new computer techniques, new imaging modalities and new interpretation tasks. CAD is defined as a diagnosis made by a radiologist who uses the output from a computerized analysis of medical images as a second opinion in detecting lesions, assessing the severity of disease and making diagnostic decisions. Diagnosis of lung disorder involves 5 main steps, 1.Image preprocessing, 2.Image segmentation, 3.ROI Segmentation, 4.Feature extraction and 5.Classification. This system basically employs the fuzzy logic for classification. The type - 2 Gaussian fuzzy logic is initially generated using various features obtained from image data sets. Subsequently, the Gaussian membership functions are optimized using an immune genetic-algorithm learning process. After optimization, the system is used for discrimination of diseases.

A Fuzzy logic system (FLS) is a rule-based expert system with the ability to model a human subjective decision-making model through its linguistic rules. When the information is so fuzzy that even defining the membership function values in an interval [0, 1] is difficult, a type-2 membership function (T2MF) is beneficial. A type-2 fuzzy logic system (T2FLS) deals with fuzzy inference that is based on type-2 fuzzy sets (T2FSs). T2MFs and their FOUs are tuned according to the uncertainties in the training dataset by a combination of genetic algorithm (GA) and local search. Genetic Algorithms (GA) are direct and stochastic methods

for global search and optimization, which mimics the evolution of the living beings. Genetic Algorithms work with a set of individuals, representing possible solutions to the given task. The selection principle is practiced by using a criterion, giving an evaluation for the individual with respect to the desired solution. The best individuals create the next generation. Every chromosome represents a solution of the given problem. From other hand, the genes could be integers, floating point or string variables, as well as any combination of the above. A set of different chromosomes (individuals) forms a generation. By means of genetic operators, like selection, recombination and mutation an offspring population is created. In the nature, the selection of individuals is performed by survival of the fittest. On EA the selection of the best individuals is based on an evaluation of the fitness function. Many selection procedures have been proposed to accomplish this idea, such as roulette-wheel selection, universal selection, ranking selection and tournament selection. This generational process is repeated until a termination condition has been reached.

Certain optimization problems cannot be solved by means of genetic algorithms. This is because of poorly known fitness functions which generate bad chromosome blocks in spite of the fact that only good chromosome blocks crossover. There is no absolute assurance that a genetic algorithm will find a global optimum. Unfortunately, due to the inherent nature of GA, it may often take a long time to locate the local optimum in a region of convergence and may sometimes not optimize with sufficient precision. One method to solve this problem is to hybridize GA with exploration operators which are capable of fine-tuning and improving the

solutions generated by the GA more precise and efficient. This also reduces the premature convergence of the genetic algorithm. Immune concepts and methods in GA is theoretically to utilize the locally characteristic information for seeking the ways and means of finding the optimal solution when dealing with difficult problems. It utilizes the local information to intervene in the globally parallel process and restrain or avoid repetitive and useless work during the course, so as to overcome some of

the blinded action of the crossover and mutation. During the operation, IGA refrains the degenerative phenomena arising from the evolutionary process, and makes the fitness of the population to increase steadily [17]. Because this course is very similar to that of an immune phenomenon in nature, the algorithm based on the above idea is used for the purpose of simplicity and directness-perception. This approach is applied to classify lung nodules in a lung CAD system.

The rest of the article is organized as follows. In Section II related works are discussed. The System architecture is discussed in Section III. Experimental results are analyzed in Section IV. Conclusion and future work are presented in Section V.

2. Related works

One of the main challenges in designing a type-2 fuzzy logic system (FLS) is how to estimate the parameters of the type-2 fuzzy membership function (T2MF) and the footprint of uncertainty (FOU) from imperfect and noisy data sets. This paper presented an automatic approach to learn and tune Gaussian interval type-2 membership functions (IT2MFs) with application to multidimensional pattern classification problems. T2MFs and FOUs are tuned according to the uncertainties in the training dataset by a combination of genetic algorithm (GA) and cross validation techniques. In GA-based approach, the structure of the chromosome has fewer genes than other GA methods, and chromosome initialization is more particular. This paper addresses the application of the interval type-2 fuzzy logic system (IT2FLS) for the problem of nodule classification in a lung computer-aided detection system [1]. The performance of Fuzzy Rule-Based Classification Systems is increased by means of the theory of Interval-Valued Fuzzy Sets and a post-processing genetic tuning step. In order to create the Interval-Valued Fuzzy Set, a new function called weak ignorance for modeling the uncertainty associated with the definition of the membership functions is defined. Fuzzy partitions are adapt to the problem in an optimal way through a cooperative evolutionary tuning which can handle both the degree of ignorance and the lateral position (based on the

2-tuples fuzzy linguistic representation) of the linguistic labels [2].

A feature subset selection method based on genetic algorithms is used to improve the performance of false positive reduction in lung nodule computer-aided detection (CAD). It is combined with a classifier based on support vector machines. These approaches determine automatically the optimal size of the feature set, and choose the most relevant features from a feature pool [3]. Spiro metric Measurement Post-processing deals with determining and predicting the respiratory system pathologies through the instrumentation, which mainly carries out measurements on the volume and the air flow expired from the lungs. This is a rapid technique for helping hysiopathologists to extract information from a incomplete expiration curve as spirometric post-processing. The two techniques are based on work of breath (WOB) and controlled genetic algorithm (CGA), respectively. A comparison is performed between the two techniques; the WOB is calculated by assuming classes of fixed resistance R according to the age, to the sex, to the previous pathologies, etc., while the CGA technique provides a strict monitoring of GA steps in order to reduce uncertainty of final results [4].

Hybrid diversification operator-based evolutionary approach towards tomographic image reconstruction algorithm introduces a new and efficient hybrid diversification operator (HDO) in the evolution cycle to improve the tomographic image reconstruction and diversity in the population by using simulated annealing (SA), and the modified form of decreasing law of mutation probability. The Algorithm was designed to address the observation that the convergence of a genetic algorithm slows down as it evolves. The HDO is shown to yield a higher image quality as compared with the filtered back-projection (FBP), the multi scale wavelet transform, the SA, and the hybrid continuous genetic algorithm (HCGA) techniques [5]. Application of Evolutionary Fuzzy Cognitive Maps for Prediction of Pulmonary Infections is a new evolution-based fuzzy cognitive map (FCM) methodology was described to cope with the forecasting of the patient status in the case of pulmonary infections, to improve the

efficiency of the prediction. It got succeeded with a new data fuzzification procedure for observables and optimization of gain of transformation function using the evolutionary learning for the construction of FCM model, results emerged had less prediction errors for the examined data records than those produced by the conventional genetic-based algorithmic approaches [6].

3. System architecture

The system architecture of this work is presented in the Fig. 1. The major components of our work are image preprocessing, image segmentation and ROI segmentation, feature extraction and classification.
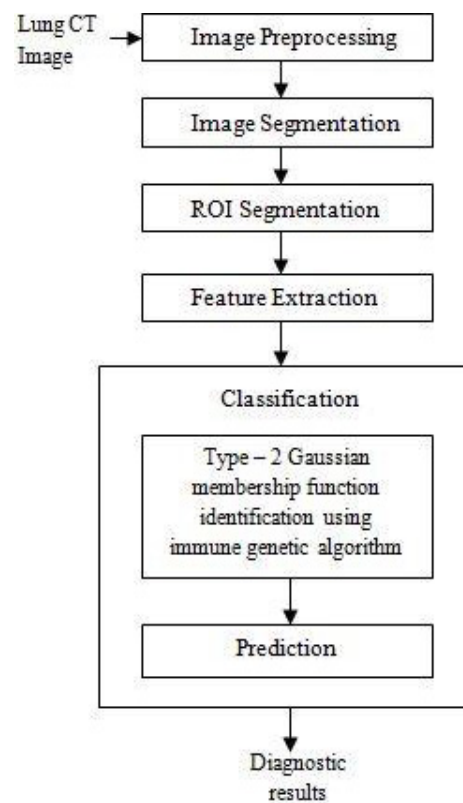

Fig. 1 System architecture for CAD system

3.1. Image Preprocessing

The input to this subsystem is a JPEG image of a chest CT scan of size 512×512 pixels. The aim of pre-processing is to improve CT image data so that it suppresses undesired distortions and enhances image features that are relevant for further processing such as enhance the visual appearance of images and improve the manipulation of data sets. 2-D adaptive noise-removal filtering is used to remove the Gaussian

noise from the CT image. A piecewise Wiener filter is used for eliminating Gaussian white noise present in the CT lung image as defined by Equation 1.

$$b(n_1, n_2) = \mu + (\sigma^2 - v^2/\sigma^2)(a(n_1, n_2) - \mu) \quad (1)$$

Where v2 is the noise variance, and μ and σ2 are the estimates of the local mean and variance around each pixel. The denoised CT chest image B obtained as the output of the Weiner filter is deconvolved with a point spread function (PSF) using the Wiener filter algorithm to get the deblurred image.

3.2. Image Segmentation

Image segmentation is the process of partitioning a diagnosed lung image into multiple segments. The aim of segmentation is to simplify and change the representation of an image into more meaningful and easier to analyze. The segmentation technique used in this paper involves the conventional optimal threshold technique and operations based on convex edge and centroid properties of the lung region [17]. This segmentation technique detects the pathology bearing regions (PBRs) even if they are in the periphery of the lung.

The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image.

The thresholding operator identifies the nodules adjacent to the pleura (juxtapleura nodules) as belonging to the chest and therefore suppresses them. So it is necessary to rebuild the pulmonary lobe parts that have been erroneously eliminated. The segmented lung image is shown in the fig. 2.

Input: Lung image with the right and left lobes of the lung separated.

Process:

Step 1. The image is converted into a single connected component by applying a morphological close operation using a structuring element of appropriate size; the size is determined by finding the separation between the closest pair of points in the two largest connected components.

Step 2. The convex image of this single connected component is determined.

Step 3. The center coordinates of the single connected component and the centroid of the convex image are determined.

Step 4. The edge of the convex image is obtained using the Canny [1986] operator.

Step 5. The edge is aligned to the coordinate system of the thresholded image.

Step 6. The edge at the neighborhood of the lung junction is eliminated.

Step 7. A logical OR operation is performed between the edge image and the thresholded image with the right and left lobes of the lung separated.

Step 8. An image is generated with the holes filled. These steps would restore the edge of the lungs even in the presence of peripherally placed PBRs.
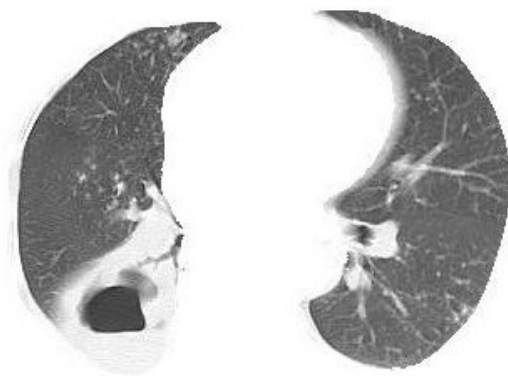
Output: The segmented lung image.

Fig. 2 Snapshot – Segmented Lung CT image

3.3. ROI Segmentation

ROI's are defined as regions containing user defined objects of interest, and an algorithm is developed for the detection of such regions. The main goal of this step is to determine the (ROI) region of interests in medical CT images. The ROIs for our system are suspected defect tissues. The regions other than these tissues are removed from morphological operations. Pixel based segmentation is used to extract the PBRs (Pathology bearing regions) from the entire lung region. Then the PBRs are labeled as shown in the fig. 3.

Fig. 3 Snapshot – ROI Segmentation.

3.4. Feature Extraction

Feature extraction is a process of transforming the input data into the set of features. As the input data to classification algorithm is too large to be processed and it is suspected to be notoriously redundant, then the data will be transformed into a reduced set of features. A feature is a significant piece of information extracted from an image which provides a more detailed understanding of the image. GLCM features such as Area, Convex area, Equiv Diameter, Eccentricity, Solidity, Energy, Contrast, Correlation, Homogeneity, Mean, Standard deviation, Smoothness, Third Moment, Uniformity Entropy is taken into account [18] as shown in the fig 4.

	Contrast	Correlation	Cluster Shade	Homogeneity
Roi_1	0.003599	0.709120379	2.138470962	0.003765522
Roi_2	0.002969	0.801415891	3.422464399	0.00359928
Roi_3	0.005624	0.70856662	3.449402369	0.005396342
Roi_4	0.012921	0.904321838	32.59304951	0.02204818
Roi_5	0.010568	0.846916614	14.18738075	0.014612749
Roi_6	0.005426	0.814401339	5.872877483	0.007038947
Roi_7	0.004001	0.668113207	2.274498353	0.003228287

Fig. 4 Snapshot – Feature Extraction.

3.5. Classification using Fuzzy logic and Immune Genetic Algorithm

Genetic Algorithm is used here to tune and optimize the Gaussian type 2 fuzzy membership functions. In order to construct the FootPrint of uncertainity (FOU) of an interval type-2 fuzzy set (IT2FS), the Gaussian distribution properties have been taken into account [9]. The lower and upper bound parameters of a Gaussian IT2MF with mean m and standard deviations are considered as given in the equation (2) and (3).

$$m' = m + k_m s, \quad m'' = m - k_m s, \quad k_m \in [0, 1] \quad (2)$$

$$s' = s \times k_v, \quad s'' = s / k_v, \quad k_v \in [0.3, 1] \quad (3)$$

where k_m and k_v are the parameters to tune the FOU, m' and m'' are the lower and upper bounds of the mean, and s' and s'' are the lower and upper bounds of the standard deviation of the IT2MF. In a Gaussian distribution with mean m and standard deviations, 99.8% of the probability masses are in the interval [m – 3s, m + 3s], and for each of the inputs we consider a minimum of three linguistic terms with a Gaussian distribution; therefore, k is considered in the interval [0, 1]. These parameter model the FOU of a Gaussian IT2MF with uncertain mean and standard deviation.
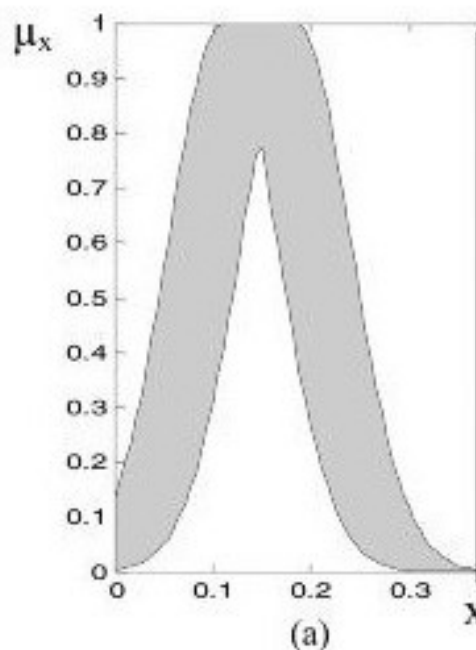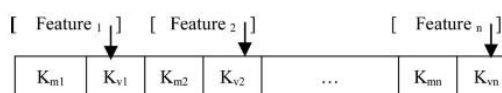

Fig. 5 Gaussian distribution of a feature.

Fig. 6 Structure of a chromosome in GA initialization

3.6. The Immune Genetic Algorithm

The idea of immunity is mainly realized through two steps based on reasonably selecting vaccines, i.e., a vaccination and an immune selection, of which the former is used for increasing

the fitness and the latter is for preventing the deterioration. Now they are explained below,

1. The Vaccination step: Given an individual, a vaccination means changing the genes on some bits in accordance with prior knowledge so as to gain higher fitness with greater probability. Suppose a population is C= $(x_1, x_2,…, x_{n0})$, and then the vaccination on C means the operation carried out on $n_\alpha=\alpha an$ individuals which are selected from C in proportion as α.

2. The Immune Selection Step: This operation is accomplished by the following two steps. The first one is the immune test, i.e., testing the antibodies. If the fitness is smaller than that of the parent, which means serious degeneration must have happened in the process of crossover or mutation, then instead of the individual the parent will participate in the next competition; the second one is the annealing selection [13], i.e., selecting an individual in the present offspring to join in the new parents with the probability shown in equation 4.

$$P(x_i) = \frac{e^{\frac{f(x_i)}{Tk}}}{\sum_{i=1}^{n_0} e^{f(x_i)/T_k}}$$

(4)

Where f (xi) is the fitness of the individual xi {Tk} and is the temperature-controlled series approaching 0. Immune genetic algorithm used in this paper helps in optimizing the parameters of the Gaussian type 2 fuzzy membership functions. Fig. 5 shows the Gaussian distribution function of a feature. The proposed genetic algorithm with exploration operator is described as follows.

Input: Initial Population with k_m and k_v range.

Output: Best solution for k_m and k_v value.

Step 1: The GA algorithm initializes the chromosome randomly, with $k_m \in [0, 1]$ and $k_v \in [0.3, 1]$.

Step 2: The fitness function is used to evaluate the fitness of each chromosome.

Step 3: Roulette wheel selection is performed to select the individual which passes the fitness evaluation process.

Step 4: The crossover and mutation operators are used in recombination. Crossover rate was fixed as 0.25, and the mutation rate was 0.1.

Step 5: Perform vaccination and obtain immune selection and obtain the next parent.

Step 6: Step (2), (3) and (4) are repeated until the terminating condition is met.

Step 7: The best solution of k_m and k_v are obtained using the local exploration operator.
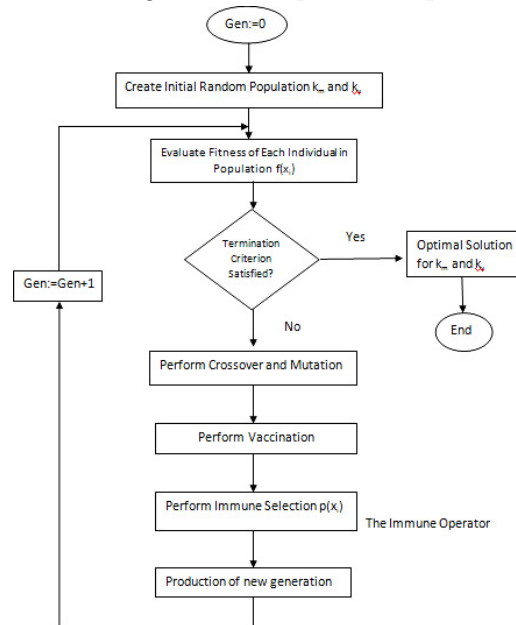

Fig. 7 Flowchart of Immune genetic algorithm

In GA applications, the parameters of the search space are encoded in the form of strings, so called chromosomes, representing a solution of problems and being encoded by a binary number. The GA initializes the chromosome k_m randomly interval [0, 1] and the k_v interval [0.3, 1]. The structure of a chromosome is depicted in fig 6.

The k_m and k_v values which satisfies the equations (2) and (3) are considered to be fittest and selected using the Roulette wheel selection. The crossover operation refers to the exchange of substrings of two individuals to generate two new individuals. The third operator, mutation, enhances the ability of the GA to find near-optimal solutions. Mutation is the occasional alternation of the binary value at a particular string position. Crossover rate was fixed as 0.25, and the mutation rate was 0.1.

Genetic algorithm finds the k_m and k_v values which satisfies the fitness function. Local exploration operator uses the spline interpolation to check the points which are intersected in the Gaussian curve of a feature. This approach identifies the best upper membership function

and lower membership function which makes optimal distribution in type-2 fuzzy logic system. Premature convergence in genetic algorithm is overthrown by using this approach as shown in the fig 7.

3.7. Prediction

The database contains 81 nodules that are identified in 40 patient scan sets, with slice thicknesses ranging from 0.5 to 2.0 mm and X-ray tube currents ranging from 80 to 300 mA. The prediction is done on each nodule based on a deterministic number which defines whether an object is a nodule (1) or a non-nodule (0). Upper and lower membership functions are identified using the obtained k_m and k_v values. For each of the features in the pattern recognition system, the training data set is clustered and fitted to several normal Gaussian functions equivalent to each of the linguistic terms, such as small, medium, and large. Standard fuzzy clustering methods can be applied. Rules are generated based on the expert's knowledge and classification is done using type – 2 fuzzy classifier. 14 rules are used based on their features.

4. Result analysis

The immune, genetic algorithm used in this paper is more efficient and better than genetic algorithm since it converges to an optimal result within a reasonable amount of iterations regardless of the initial population size. The optimal k_m and k_v values are determined by local exploration operator as shown in the Table 1. The operator helps in finding the best Gaussian distribution curve of a particular feature which improves accuracy of the type – 2 Gaussian fuzzy classifier.

The classification was implemented in MATLAB using the fuzzy logic toolbox. In learning and tuning the type 2 fuzzy logic, the tenfold cross validation was applied. The optimization algorithm considers (Receiver operating characteristic) ROC accuracy as an indicator of the classification performance. The ROC average accuracy using type - 2 fuzzy logic and genetic algorithm is 83% and type - 2 fuzzy logic and immune genetic algorithm is 89%.

	k_m and k_v values selected by proposed immune genetic algorithm		k_m and k_v values selected by genetic algorithm	
Feature 1	0.76	0.89	0.56	0.81
Feature 2	0.57	0.71	0.65	0.46
Feature 3	0.94	0.68	0.86	0.92

Table 1. Generated k_m and k_v values by GA and Improved GA

5. Conclusion and future work

An approach to diagnose lung disorder using an immune genetic algorithm and local exploration operator has been proposed in this paper. This method generates the FOU of a Gaussian interval type-2 membership function based on training data set. The FOU is learned and tuned using the Immune Genetic Algorithm, with the chromosome structure and initialization technique, in combination with a local exploration operator. This method is applied to classify lung nodules in a lung CAD system. In order to have an unbiased view of the system performance with the minimum dependence on the dataset, the training and testing datasets were selected randomly using the tenfold cross-validation technique. Our future work is to learn rules for uncertain environments where defining the system rules can be challenging and to generate optimized solutions which scale well with complexity.

6. References

1. Rahil Hosseini, Salah D. Qanadli, Sarah Barman, Mahdi Mazinani, Tim Ellis, and Jamshid Dehmeshki, "An Automatic Approach for Learning and Tuning Gaussian Interval Type-2 Fuzzy Membership Functions Applied to Lung CAD Classification System", IEEE Transactions on Fuzzy Systems, Vol. 20, No. 2, 2012.

2. J. Sanz, A. Fernández, H. Bustince ,F.Herrera "A genetic tuning to improve the performance of Fuzzy Rule-Based Classification Systems with Interval-Valued Fuzzy Sets: Degree of ignorance and lateral position", International Journal of Approximate Reasoning (2011) 751-766.

3. Lilla B¨or¨oczky, Luyin Zhao, and K. P. Lee. "Feature Subset Selection for Improving the Performance of False Positive Reduction in Lung Nodule CAD", IEEE Transactions on Biomedical Engineering, Vol. 58, 2011.

4. Aimé Lay-Ekuakille, Giuseppe Vendramin, and Amerigo Trotta, "Spirometric Measurement Postprocessing: Expiration Data Recovery",IEEE Sensors Journal, Vol. 10, 2010.

5. Shahzad Ahmad Qureshi, Sikander M. Mirza, Nasir M. Rajpoot, and Muhammad Arif, "Hybrid Diversification Operator-Based Evolutionary Approach Towards Tomographic Image Reconstruction", IEEE Transactions on Image Processing, Vol. 20, 2011.

6. Elpiniki I. Papageorgiou and Wojciech Froelich "Application of Evolutionary Fuzzy Cognitive Maps for Prediction of Pulmonary Infections", IEEE Transactions on Information Technology In Biomedicine, Vol. 16, No. 1, 2012.

7. Antonelli,M., Lazzerini.B, Marcelloni.F. "Segmentation and reconstruction of the lung volume in CT images", In Proceedings of the ACM Symposium on Applied Computing. 255–259, 2005.

8. Howarth.P and Ruger. S. M. "Evaluation of texture features for content-based image retrieval. In Proceedings of the 3rd International Conference on Conference on Image and Video Retrieval. 326–334, 2004

9. N. N. Karnik and J. M. Mendel, "Introduction to type-2 fuzzy logic systems," in Proc. IEEE Int. Conf. Fuzzy Syst., Anchorage, AK, 1998, pp. 915–920.

10. D. K. Iakovidis, S. Tsevas, M. A. Savelonas, and G. Papamichalis, "Image Analysis Framework for Infection Monitoring", IEEE Transactions on Biomedical Engineering, Vol. 59, No. 4, 2012.

11. Lilla B¨or¨oczky, Luyin Zhao, and K. P. Lee, "Feature Subset Selection for Improving the Performance of False Positive Reduction in Lung Nodule CAD", IEEE Transactions on Information Technology in Biome Dicine, Vol. 10, No. 3, 2006.

12. Licheng Jiao and Lei Wang, "A Novel Genetic Algorithm Based on Immunity", IEEE Transactions on Systems, Man, And Cybernetics—Part A: Systems and Humans, Vol. 30, No. 5, 2000.

13. J. S. Zhang, Z. B. Xu, and Y. Liang, "The whole annealing genetic al-

14. gorithms and their sufficient and necessary conditions of convergence,"

15. Science in China, vol. 27, no. 2, pp. 154–164, 1997

16. Licheng and Lei Wang, "Quantum-Inspired Immune Clonal Algorithm for Global Optimization", IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, Vol. 38, No. 5, 2008.

17. M.N. Gurcan, B. Sahinder, N. Petrick et al., "Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system", Med. Phys. 29 (11) (2002), pp.2552-2558.

18. Otsu, N. "A threshold selection method from gray-level histograms", IEEE Trans. Syst. Man Cybern. 9, 62–66, 1979.

19. Licheng Jiao and Lei Wang, "A Novel Genetic Algorithm Based on Immunity", IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems And Humans, Vol. 30, No. 5, September 2000.

20. D. Shiloah Elizabeth, H. Khanna Nehemiah, C. Sunil Retmin Raj, and A. Kannan "A Novel Segmentation Approach for Improving Diagnostic Accuracy of CAD Systems for Detecting Lung Cancer from Chest Computed Tomography Images , ACM Journal of Data and Information Quality, Vol. 3, No. 2, Article 4, 2012.