An approach for Lung Cancer Pattern Recognition in Computer Aided Detection Systems

Rahil Hosseini, Mahdi Mazinani
Islamic Azad University, Shahryar-shahr-e-Qods branch, Tehran, Iran
Rahilhosseini@gmail.com

Paper Reference Number: 0102-676

Name of the Presenter: Rahil Hosseini

Abstract

A computerized medical image analysis technology suffers from imperfection, imprecision and vagueness of the input data in medical images and its propagation in all individual components of the technology including image processing and pattern recognition. In addition to the above, a Computerized Medical Image Analysis System such as computer aided detection (CAD) technology deals with another source of uncertainty which comes from inter and intra uncertainties in medical diagnosis. These sources of uncertainties in disease pattern recognition in medical images incur to incorrect diagnosis which involves the human life. Therefore, modeling uncertainties in disease pattern recognition in computer aided detection systems is a very vital task to be taken into account in developing a CAD system. In this paper, the sources of the uncertainty in the design of a lung CAD system using CT images of thoracic will be addressed. A fuzzy model is proposed to tackle the problem of uncertainty in the pattern recognition component of the lung CAD system. The result is promising to improve the performance of nodule pattern diagnosis in the lung and improve medical diagnosis using CAD systems.

Key words: Computer Aided Detection, Fuzzy Approach, Pattern Recognition, Uncertainty Modeling.

1. Introduction

One of the common characteristics of information derived from an image device is its imperfection. The acquired image can be inconsistent, incomplete and uncertain. Consequently, a major hidden barrier to better performance of a digital image analysis technology results from imprecision and vagueness in every level of the processes (figure 1) and input data; such as uncertainty in gray level, texture, contours, edge detection, converting a 3-D image to a 2-D image, relationship between two segments of an image. In addition, in medical diagnosis as a 'human-centred' problem we deal with other sources of uncertainties: 1) uncertainties in both 'medical science' and incomplete knowledge of human; 2) uncertainties related to subjective information in diagnosis and medical decision making. Therefore, computerized medical image analysis (CMIA) systems such as computer aided detection (CAD) suffer inherently from uncertainty and vagueness both from digital image analysis technique and medical diagnosis.

In a CMIA system, a classifier is the component which is significantly analogous to human judgment in the process of analysis and making final decision. Pattern recognition was defined by Gonzalez [1] as a high-level processing which involves "making sense" of an ensemble of recognized patterns, analysis of the image, and, performing the cognitive function normally associated with human vision. In addition, ROC analysis [2] which is an approach for measuring the performance of the CAD systems is a reflection of its pattern recognition results.

One of the challenges in the pattern recognition component of a CAD system is to reduce the vagueness and imprecision during the classification. It means the classifier must have the ability of managing imprecision in features extracted from different segments of an image, mathematical modeling of them and uncertainty in making decision about the class of the patterns. Moreover, there are cases in which a recognized pattern does not certainly belong to one class and it may belong to several classes with a different degree of belief. Ignoring these issues in classification of a CAD, incur to poor performance and inaccurate results.

In the other hand, in clinical context, there are always variations in experts' final decision about one tissue in the body which is natural in clinical diagnosis because of experts' previous knowledge, their observations in laboratory examinations and their subjective information, and patient history. In addition, their final decision usually has a confidence level [13]. Furthermore, the process of a disease is dynamic; it begins from an early stage and can continue to an end stage. The main sources of uncertainty in such systems can be summarized as follow:

1. Imprecision in input data and features of patterns and noisy measurements (Input data),
2. Uncertainty in final decision of one expert about one pattern (intra-uncertainty [18], is called intra-operator observer variability in clinical context),
3. Uncertainty between experts for making decision about one pattern (inter-uncertainty [18], is called inter-operator observer variability in clinical context),
4. Different meanings of linguistic terms from various experts' point of view (words perception).
5. Dynamic aspect of disease progress for diagnosis of its stage and (Non-stationary features)
6. Uncertainty of mathematical models for measuring the complex features of images (such as degree of circularity, darkness, etc.)
7. Uncertainty in all process of image enhancement and segmentation, edge detection, converting a 3-D image to a 2-D image, gray level, texture.

The history of pattern recognition technologies are described in sectionn2. In section 3, the applied methodology and theory and concept of it is described in details. The result of applying the T2FLS for lung cancer pattern recognition is discussed in section 4 and the paper is included in section5.

2. Literature review

In this section, a brief overview of related existing approaches for dealing with vagueness and imprecise problems in pattern recognition in medical application will be presented. There are various approaches for designing a pattern recognition system. Menahem and Kandel [19] defined the main taxonomy as: Distance function and clustering, Statistical, neural networks, syntactic, and fuzzy approaches. In addition, there are combinations of mentioned approaches for different application such as fuzzy-neural networks [6-8, 14], fuzzy clustering [4, 22- 23]. Between them, fuzzy models are the most robust, mature and pervasive applied methods for tackling the uncertainty problems. The theory of fuzzy set was first introduced in 1995 by Zadeh [38]. Bezdek and Pal [11] collected seminal published papers on the theory and application of pattern recognition based on fuzzy sets in their book. Moreover, it is a robust and practical classifier with high interpretability and usability [39-40]. This is the reason that the focus of this study is on rule-based classification. In rule-based classifier, rules show expert's knowledge and are expressed by using linguistic terms. Fuzzy logic systems (FLSs) define the general architecture for designing a classifier with linguistic terms [39].

Type-2 fuzzy sets (T2FSs) are proposed as a potentially powerful method to overcome abovementioned problems in the classification of the CAD system.T2FSs (an extension of fuzzy set) are known to fuzzy-fuzzy sets in which membership function is not a real number in [0, 1] and is a fuzzy set [3]. T2FS was first introduced by Zadeh in 1975 and then Karnik and Mendel extended it in 1998 [10]. Mendel introduced uncertain rule-based fuzzy logic systems [3]. He addressed the problems of noisy data, intra-uncertainty, inter-uncertainty, word perception, non-stationary features and mathematical ill-defined systems as situations where using T2FSs are beneficial [5, 15]. Moreover, it was issued in various applications that the type-2 fuzzy logic systems (T2FLSs) have the potential to provide better performance than type-1 fuzzy sets [16-17]. The application of T2FSs in medical diagnosis has been taking place at the same time of extension of the theory and it is an open research area now.

The main advances of T2FSs in pattern recognition [28-30] in CAD systems can be summarized as follow: Zeng and Liu present the state-of-the-art of type-2 pattern recognition and the success of that for solving the problem of randomness and fuzziness. In addition, the robustness of T2FSs for handling the uncertainty in feature and hypothesis space is demonstrated in comparison to statistical pattern recognition such as Bayesian methods [12]. They proposed type-2 fuzzy random Markov method for handling uncertainty in speech recognition and Chinese character recognition [28]. Mitchell proposed a similarity measure for measuring compatibility between two T2FSs and applied it to the problem of automatic detection of welded structures in radiographic images [29]. Choi and Rhee proposed three methods for automatically defining internal type-2 membership function from pattern data which can be used as an important component in pattern recognition. Their methods are based on heuristic, histogram, and interval type-2 fuzzy C-mean. In order to evaluate the performance of the classification, the methods are applied to back-propagation neural network and the histogram based method demonstrates the best performance results [36].

The aim of this research is to address the uncertainty sources and handle the uncertainty issues of pattern recognition of the CAD system through applying a T2FLS.

3. Methodology

In this section, the proposed methodology is demonstrated though type-2 fuzzy logic architecture. The important issues which must be taken into account for designing a T2FLS for classification in a CAD system and the systems requirements are described. The problem description is:

Problem statement: Designing an interval T2FLS for pattern recognition in a CAD system.

Inputs: Assume we are given p classes, n segmented pattern, m features and $X_i = (x_{i1}, x_{i2,}, \dots, x_{im})$ is feature vector for pattern i; in which, x_{ij}, is the jth extracted feature for the ith pattern. Output: Define the class of the pattern and the membership of each pattern to different classes. There are three important components in the architecture of the T2FLS; 1) fuzzifier, 2) inference engine and rules, and 3) output producer includes type reducer and defuzzifier (figure 3). Main uncertainty issues which are required to be incorporated in this architecture can be expressed briefly as follow:

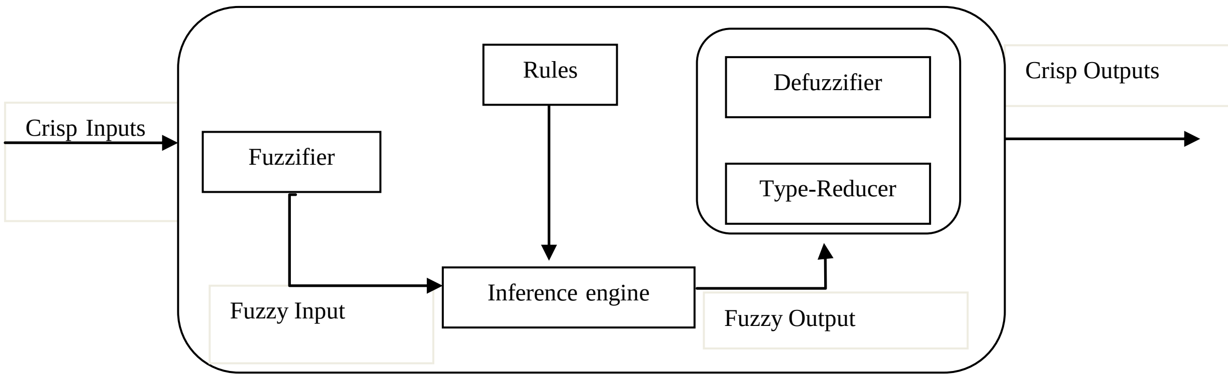


Figure 3: Type-2 Fuzzy Logic System (T2FLS) Architecture

Noisy measurements of inputs and vagueness in some features value of a segmented patterns
Inherited uncertainty from all process of the image enhancement and segmentation, i.e. edge detection, converting a 3-D image to a 2-D image, gray level, and texture.
Ill-defined mathematical models for measuring the complex features of images (such as degree of circularity, darkness, etc.) which complicates the fuzzification process.
Inter and intra-operator observer variability in the description of systems rules which affects inference engine
Variation in words perception complicates defining fuzzy sets and membership functions of fuzzy input sets and rules antecedents and consequents

The requirements of a typical T2FLS can be defined in the following main categories:
Medical images: including a dataset of images from patients with annotated abnormalities in order to train and test the system
Features: are extracted from segmented patterns in an image and their value is measured
Fuzzification: defines the type of features; i.e. crisp, T1FS and T2FS, and the method of defining membership function of fuzzy sets
Inference requirements: includes extracted rules from several experts for distinguish between abnormal and healthy patterns and the method of inference
Output: defines the method of type-reduction and aggregation
Software requirements: a CAD system for measuring features values, visual C++ for feature analysis and training, Matlab fuzzy logic toolboxes, Matlab curve fitting toolboxes, Interval Type-2 Fuzzy Logic Toolbox on Matlab developed in [37].

3.1. An overview of theory and Concepts of type-2 fuzzy set

In this section, a brief overview of a T2FS, according [3, 24, 32], is demonstrated. In type-1 fuzzy sets, membership function is a crisp number whereas in T2FSs, membership function is a subset of fuzzy set and is itself fuzzy. The membership function of a type-2 fuzzy set $\widetilde{A}$ (figure 4), is characterized by function $\mu_{\widetilde{A}}(x,u)$, where $0 \le \mu_{\widetilde{A}}(x,u) \le 1$; x is domain and $x \in X$ (X is the universe of discourse)

$$\mu_{\widetilde{A}}(x,u) : X \to [0,1]^{[0,1]} \qquad (1)$$

$u \in J_x \subseteq [0,1]$. Fuzzy set $\widetilde{A}$ can be defined as:

$$\widetilde{A} = \left\{\left((x,u), \mu_{\widetilde{A}}(x,u)\right) \middle| \forall x \in X, \forall u \in J_x \subseteq [0,1]\right\} \qquad (2)$$

The domain of a secondary membership function is called primary membership function of x. In equation (1), $\mu_{\widetilde{A}}(x')$ is secondary membership function and $J_{x'}$ is primary membership function at $x = x'$ (figure 4).

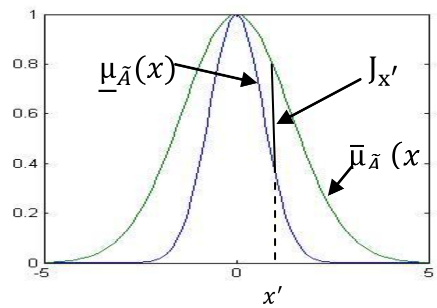


Figure 4: Primary Membership Function of Type-2 Fuzzy Set at $x = x'$

Uncertainty in the primary memberships of a type-2 fuzzy set , $\widetilde{A}$ includes a bounded and blurred region which is called the Footprint of Uncertainty (FOU). It can be written as the union of all primary memberships:

$$\mathrm{FOU}\,(\widetilde{A}) = \bigcup_{x \in X} J_x \qquad (3)$$

An upper and a lower membership function bounds FOU of a type-2 fuzzy sets (figure 5). The upper membership function is a type-1 membership function:

$$\bar{\mu}_{\widetilde{A}}(x) \equiv \overline{\mathrm{FOU}(\widetilde{A})} \qquad \forall x \in X \qquad (5)$$

, and the lower membership function is also a type-1 membership function:

$$\underline{\mu}_{\widetilde{A}}(x) \equiv \underline{\mathrm{FOU}(\widetilde{A})} \qquad \forall x \in X \qquad (4)$$

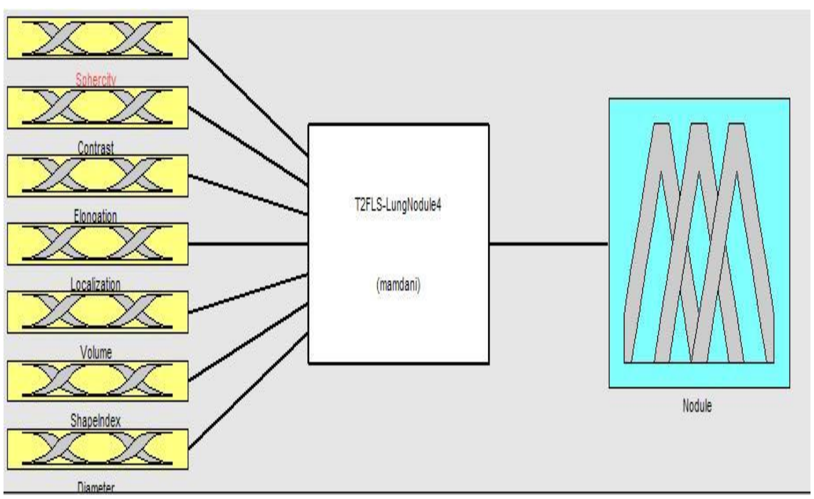
Type-2 membership function, which is the secondary membership function, can be any type such as Gaussian, trapezoidal, triangular or interval. When the secondary membership functions are interval sets, we have interval type-2 membership function. It means the uncertainty is uniform along its primary membership function (domain) and is equal to unity:

$$f_x(u) = 1 \qquad \forall u \in J_x \subseteq [0,1] \qquad (5)$$

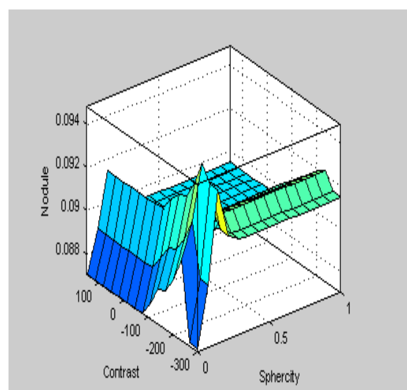
Interval type-2 fuzzy systems are the practical type-2 fuzzy sets and are frequently used because of the lowest computational complexity and high simplicity.

4. Results and Analysis

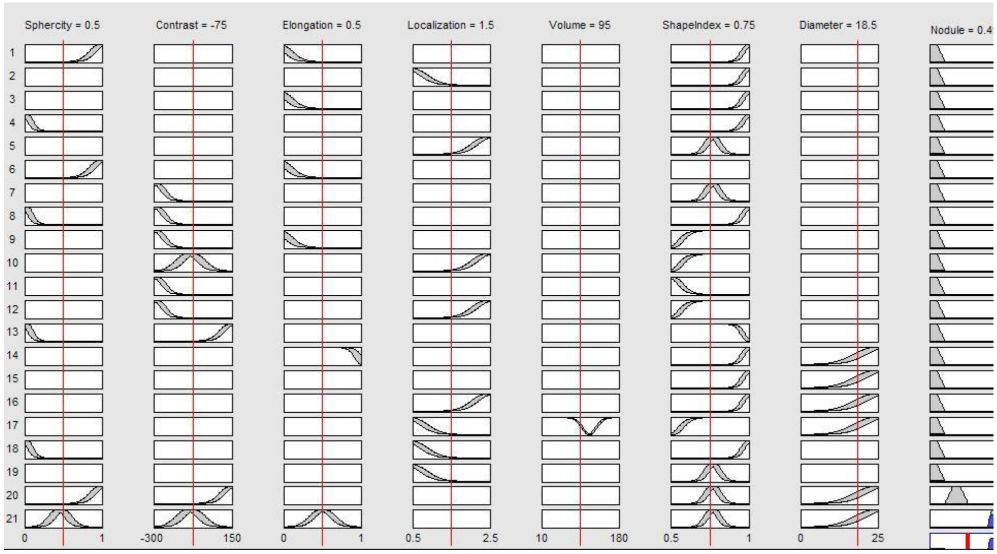
In this study for measuring and improving performance of CAD systems, first the concepts and theory of medical image processing techniques was learned by attending in several digital image processing courses and study proposed references. Then current methods have been surveyed and main challenges have been defined. One of the crucial deficiencies in designing an accurate CAD technology is uncertainty issues in pattern recognition component of a CAD system and inherited imperfection and vagueness in computerized image-based practices of medicines which lead to low performance of the system and high rate of false positive detection. For this reason, existing methods for tackling uncertainty issues in classification have been surveyed. Between them, T2FLSs seem to be a suitable framework for overwhelming classification problem in the CAD system. This approach has high potential for handling all aforementioned sources of uncertainty and is promising to achieve better performance. For this reason first the T1FLS will be designed and implemented for an instance application such as lung nodule CAD, then a IT2FLS will be designed for that application and uncertainty will be incorporated in design of the system. Finally, performance of two fuzzy logic systems is measured in comparison to each other and further analysis will be done. Figure 8 demonstrates some designed properties of implemented T1FLS for lung nodule CAD application.



(a)



(b)



(c)

Figure 8: T1FLS classification components for the lung nodule CAD application; (a) Classification inference engine (b) 3D representation of contrast and sphericity features, (c) Systems Rules and Rules aggregation.

The General architecture of a T2FLS [3] is demonstrated in figure 3. For the reason of design an IT2FLS for classification, a computer aided detection (CAD) for detecting lung nodule within a CT scan data is applied. Dataset consists of 40 CT scan images from Lausanne hospital. In this application, rules for deciding about a nodule or non-nodule pattern are annotated by radiologists.

A non-singleton type-2 FLS [3] is applied as a system which accounts for noise in input data in addition to uncertainty in antecedent and consequents of the rules. An example of such a rule in the lung nodule CAD systems is: Rule1: “If the intensity (HU) of a pattern is high and sphericity is strong then it is very likely to be a nodule”.

Refer to lung CAD application, one of the complex features to measure is the degree of sphericity. The radiologists’ linguistic terms for sphericity are strong, moderate, and weak which can be considered as fuzzy sets. Considering the patterns in training set which are annotated as nodule, there is even vagueness and difficulty in defining the membership function of this fuzzy set as a precise number in [0,1]. Furthermore, the feature is mathematical ill-defined; i.e. in study reported in [33] two features with different mathematical models are defined as indicator of the sphericity for nodule and non-nodule patterns in training data. Moreover, the fact that each method produces an error must be taken into account. For handling sphericity uncertainty, the feature value is considered as a T2FS.The membership function for sphericity of one pattern is demonstrated in (figure 9). By training the system and interpolating different mathematical methods for nodule and non-nodule patterns in training data sets, lower and upper bounds of uncertainty can be defined (figure 9).

Performance of a CAD system is measurement through ROC (Receiver Operation Characteristic) analysis. ROC curve demonstrates the tradeoffs between sensitivity (i.e. the probability that an actually abnormal image is classified as “positive”) and specificity (i.e., the probability that an actually normal image is classified correctly as “negative”) [2]. The performance of the proposed model was measured based on the ROC accuracy which is 91%

and is a very reasonable and competitive with other methods and model the uncertainties as well.

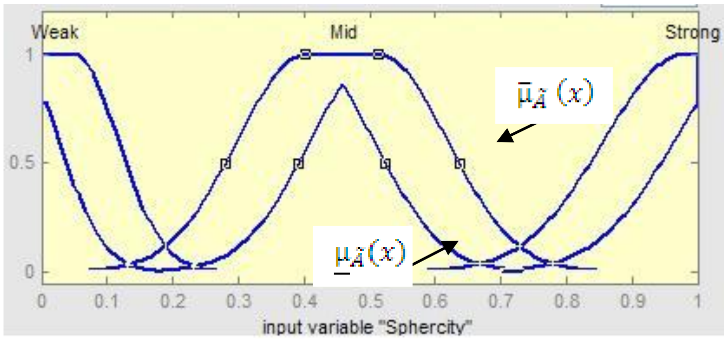


Figure 9: Sphericity Membership Function - FOU and the Lower and Upper bounds In addition, a paper entitled "modeling uncertainty in classification design of CAD" has been submitted to international conference.

5. Conclusions

In this paper, the main uncertainty sources in CAD systems are described. Two new approaches for tackling the uncertainty issues in the CAD system are presented and have built on the classification of the system.

The proposed methods are applied for the lung nodule CAD application. The results after incorporating uncertainty issues in the pattern recognition components of the CAD system is 91% which is satisfactory and reveals the superiority of T2FS for tackling the problem of inter and intra operator observer variability. Our future work is to apply our proposed model for diagnosis of the pattern of various disease and analysis the results.

References

[1] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, Digital image processing using MATLAB, Prentice Hall, (2004).
[2] J. Beutel, M. Sonka, H. L. Kundel, J. M. Fitzpatrick, and R. L. V. Metter, Handbook of Medical Imaging: Medical image processing and analysis, SPIE P, (2000).
[3] J. M. Mendel, Uncertain rule-based fuzzy logic systems: introduction and new directions, Prentice-Hall (2001).
[4] E. E. Kerre, M. Nachtegael, Fuzzy Techniques in Image Processing: Techniques and Applications, Studies in fuzziness and soft computing Vol. 52, Springer, (2000).
[5] J. M. Mendel, Advances in type-2 fuzzy sets and systems, Information Sciences 177, 84–110, (2007).
[6] C. M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, (2005).
[7] S Abe, Pattern classification; Neuro-fuzzy methods and their comparison, UK, Springer Verlag, (2001).
[8] M. Friedman, A. Kandel, Introduction to Pattern Recognition: Statistical, Structural, Neural and Fuzzy Logic Approaches, Vol. 102, (1999).
[9] Z. Chi, H. Yan, and T. Pham, Fuzzy algorithms: with applications to image processing and pattern recognition, Word Scientific (1996).
[10] N. N. Karnik, J. M. Mendel, "Introduction to Type-2 Fuzzy Logic Systems", in Proc. of IEEE Fuzzy Conference, 915--920, AK, (1998).
[11] J. C. Bezdek, S. K. Pal, Fuzzy models for pattern recognition: methods that search for structures in data, IEEE press, (1992).
[12] J. Zeng, and Z.Q. Liu, Type-2 Fuzzy Sets for Pattern Recognition: The State-of-the-Art, Journal of Uncertain Systems, Vol.1, No.10, 163--177, (2007).
[13] J. Dehmeshki, M.C. Valdivieso, M. Siddique, J. Karani, and J. Costello, "Identification of Pulmonary Nodules on Thoracic CT Scans Using Image Processing Techniques", RSNA, USA, (*2002*) .
[14] S. Pal, S. Mitra, Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing, John Wiley, USA, (1999).

[15] J. Mendel, and R. I.B. John, Type-2 Fuzzy Sets Made Simple, IEEE Trans. on Fuzzy Systems, Vol. 3, No. 2, (2002).
[16] J. M. Mendel, R. I. John, and F. Liu, Interval Type-2 Fuzzy Logic Systems Made Simple, IEEE Trans. on Fuzzy Systems, Vol. 14, No. 6, (2006).
[17] S. Coupland, R. John, Type-2 Fuzzy Logic and the Modelling of Uncertainty, Fuzzy Sets and Their Extension: Representation, Aggregation and Models, Springer (2008).
[18] R.I. John, P.R. Innocent, M.R. Barnes, Neuro-fuzzy Clustering of Radiographic Tibia Image Data Using Type 2 Fuzzy Sets, Information Sciences 125, 65—82 (2000)
[19] T. Ozen, and J.M. Garibaldi, Investigating Adaption in Type-2 Fuzzy Logic Systems Applied to Umbilical Acid-Based Assessment, In Proc. of European Symp. on Intelligent Technologies, 289--94, Finland, (2003).
[20] M. Jonathan, Garibaldi, and T. Ozen, Uncertain Fuzzy Reasoning: A Case Study in Modelling Expert Decision Making, IEEE Trans. on Fuzzy Systems, Vol. 15, No. 1, 16--30, (2007).
[21] P.R. Innocent, R.I. John, Computer aided fuzzy medical diagnosis, Information Sciences 162, 81–34, (2004).
[22] R. I. John and P. R. Innocent, Modelling Uncertainty in Clinical Diagnosis Using Fuzzy Logic, IEEE Trans. on Systems, Man, and Cybernetics, Vol. 35, No. 6, (2005).
[23] P.R Innocent, R.I. John, I. Belton, and D. Finlay, Type-2 fuzzy representations of lung scans to predict pulmonary emboli, In Proc. of NAFIPS, 1902--1907 vol.4, Canada, (2001).
[24] O. Castillo, P. Melin, Type-2 Fuzzy Logic: Theory and Applications, Studies in Fuzziness and Soft Computing, Springer, (2008).
[25] H. R. Tizhoosh, Image thresholding using type II fuzzy sets, Pattern Recognition, 38, 2363-- 2372, (2005).
[26] H. R. Tizhoosh, Type II Fuzzy Image Segmentation, Fuzzy, Springer, Berlin, Vol. 220, 607-619, (2008).
[27] P. Ensafi and H.R. Tizhoosh, Type-2 Fuzzy Image Enhancement, ICIAR 2005, LNCS 3656, Vol. 3656, 159--166, Springer, Berlin, (2005).
[28] Jia Zeng Zhi-Qiang Liu, Type-2 Fuzzy Sets for Handling Uncertainty in Pattern Recognition, IEEE Int. Conf. On Fuzzy Syst, Vancouver, BC, 1247-1252, (2006).
[29] H.B. Mitchell , Pattern recognition using type-II fuzzy sets, Information Sciences,vol. 170, 409–418, (2005)
[30] Denisse Hidalgo, Oscar Castillo and Patricia Melin, Interval type-2 fuzzy inference systems as integration methods in modular neural networks for multimodal biometry and its optimisation with genetic algorithms, Int. Journal of Biometrics, vol. 1, No. 1, (2008)
[31] L. A. Zadeh, From Computing with Numbers to Computing with Words- Form Manipulation of Perceptions. IEEE Trans. on Circuit and Systems, 45:35--119, (1999).
[32] Q. Liang, and J. M. Mendel, Interval type-2 fuzzy logic systems: theory and design, IEEE Trans. on Fuzzy Systems, Vol. 8, No. 5, (2000).
[33] X. Ye, X. Lin, J. Dehmeshki, G. Slabaugh, G. Beddoe, Shape Based Computer Aided Detection of Lung Nodules in Thoracic CT Image, IEEE Trans. on Biomedical Engineering, vol. 56, pp. 1810 - 1820, (2009).
[34] J. M. Mendel and F. Liu, “Super-exponential convergence of the Karnik-Mendel algorithms for computing the centroid of an interval type-2 fuzzy set,” IEEE Trans. on Fuzzy Systems, (2006).
[35] H. Wu, and J. M. Mendel, Uncertainty Bounds and Their Use in the Design of Interval Type-2 Fuzzy Logic Systems, IEEE Trans. on Fuzzy Systems, Vol. 10, No. 5, (2002).
[36] Byung-In Choi, Frank Chung-Hoon Rhee, Interval type-2 fuzzy membership function generation methods for pattern recognition, Information Sciences, vol. 179, 2102–2122, (2009)
[37] R. Castro, O. Castillo, P. Melin, L. G. Martínez, S. Escobar and I. Camacho, Building Fuzzy Inference Systems with the Interval Type-2 Fuzzy Logic Toolbox, Analysis and Design of Intelligent Systems using Soft Computing Techniques Springer Berlin, Vol. 41, pp. 53-62,(2007).
[38] L. A. Zadeh, Fuzzy sets, Information and Controls, 8:10108--10510, (1965).
[39] H. Ishibuchi, and Y. Nojima, Pattern Classification with Linguistic Rules, Fuzzy Sets and their Extensions; Representation, Aggregation and Models, Vol. 220, 1077--1095, Springer, Berlin (2008).
[40] K. Arakawa, Fuzzy Rule-Based Image Processing with Optimization, Fuzzy Techniques in Image Processing, Studies in Fuzziness and Soft Computing, Springer (2000).
[43] N. N. Karnik and J. M. Mendel, “Centroid of a type-2 fuzzy set,” Information Sciences, vol. 132, 195--220, (2001).