CHROMOSOMES CLASSIFICATION BASED ON NEURAL NETWORKS, FUZZY RULE BASED, AND TEMPLATE MATCHING CLASSIFIERS

Ahmed M. Badawi[*], *Kahled G. Hasan*[**], *Emam-Elhak A. Aly*[**], *and Rimon A. Messiha*[***]
** PHD., Associate Professor of Biomedical & Systems Engineering, Faculty of Engineering, Cairo University, Egypt, ambadawi@esbme.com*
*** MSC., Biomedical Engineering, Faculty of Engineering, Cairo University, IBM Software Design Engineer*
*** BSC., Biomedical Engineering, Faculty of Engineering, Cairo University

Abstract *- A new features extraction algorithms for G-banded chromosomes classification system based on neural networks, fuzzy rule based and template matching classifiers are proposed. Chromosomes image is acquired and processed, geometrical features and gray-scale features are extracted for 872 chromosomes. Neural networks, fuzzy rule based, template matching classifiers results were compared. Classification rates were found to be over 99% for training and over 96% for testing sets.*

I. INTRODUCTION

Great efforts to develop automatic chromosome classification techniques have been made during the last 25 years. However, all have had limited success and have yielded lower classification results compared to those of a trained cytotechnician [1-8]. In this paper we propose an automatic Karytyping system for G-banded images based on neural networks, fuzzy rule based and template matching classifiers.

II. CHROMOSOMES IMAGE SEPARATION

The image is first segmented (Black and white) using OTSU algorithm [9]. Figure 1a shows a grayscale image and Figure 1b shows its segmented image. Contours were then detected, traced, regions were then thinned and processed to get its full medial axis as shown in Figure 2. Finally using a new chromosomes separation algorithm [10], image was separated as shown in Figure 3a. Manually classified image by an expert is shown in Figure 3b.

III. FEATURES EXTRACTION

Figure 4 shows a flowchart for different processes used to extract features and matching process. Final features were extracted through many processes and will be described in the following sections.

A. Chromosome Real Banding Profile Extraction

The profile is extracted along the chromosome medial axis and is represented as a gray-level function of the medial axis points or as an average of the points perpendicular to the medial axis. Maxima and minima points and real band borders points were extracted. Figure 5 shows a chromosome profile, its maxima-minima points, and its real banding thickness. Real banding associated with number of maxima-minima points are estimated by taking the graylevel derivatives along band medial axis. Real bands thicknesses were calculated by getting the point of maximum grayscale transition between every maxima and minima point. Figure 6 shows one of the satellite chromosomes with its automatically

Figure 1
a) Original G-banded image b) Segmented image using OTSU algorithm

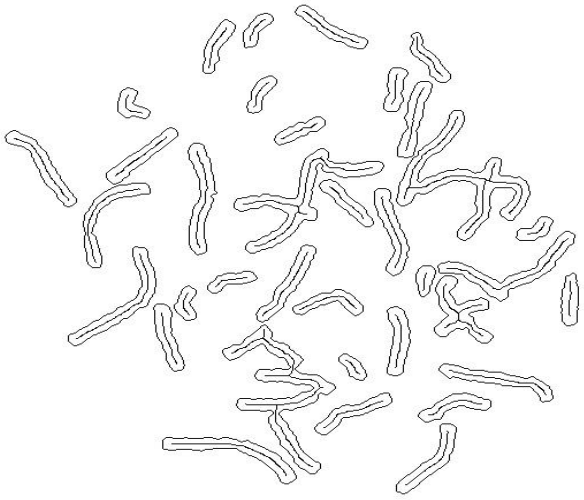

Figure 2
Contours and its medial axis skeleton

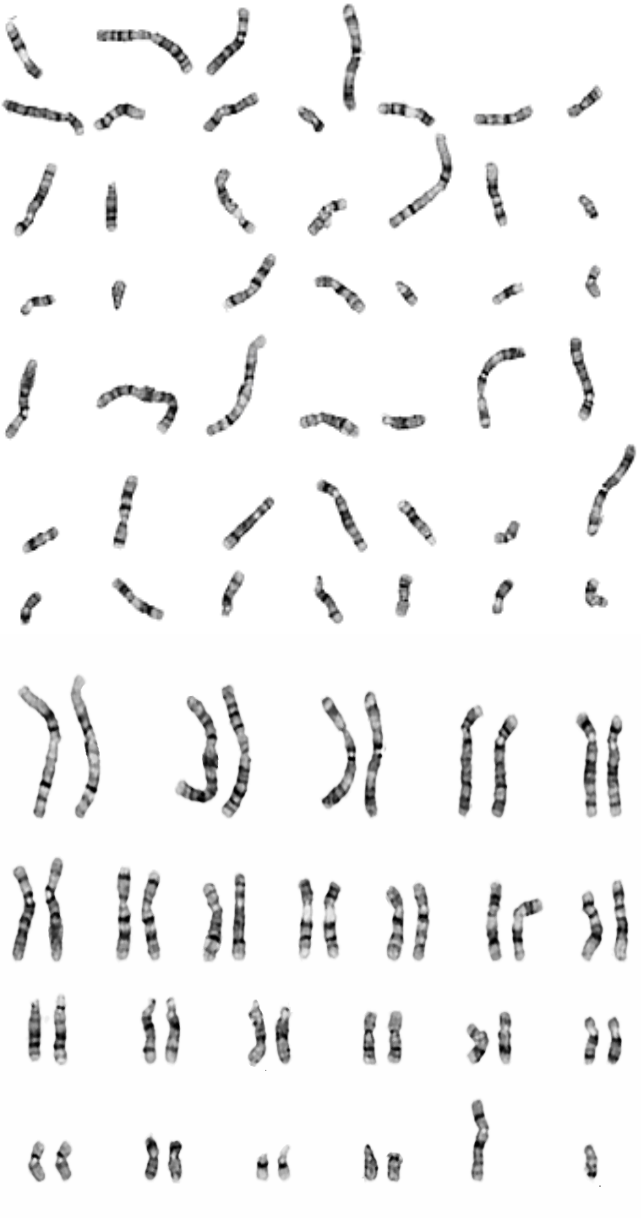

Figure 3
a) Separated image b) Manual classified and aligned image

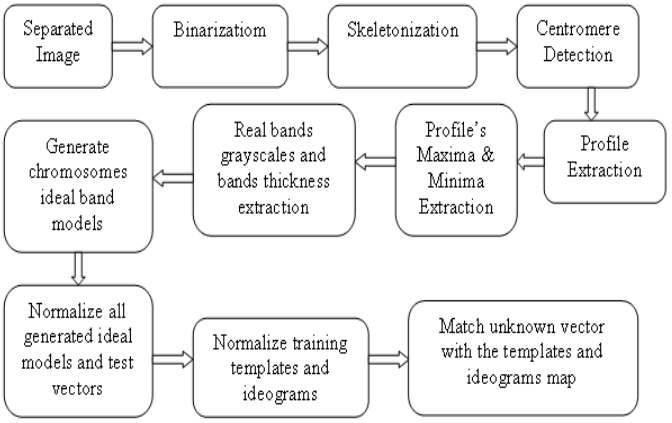

Figure 4
Flowchart for features extraction and matching process

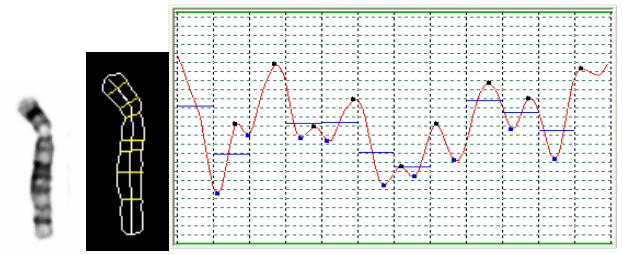

Figure 5
Profile for the shown chromosome

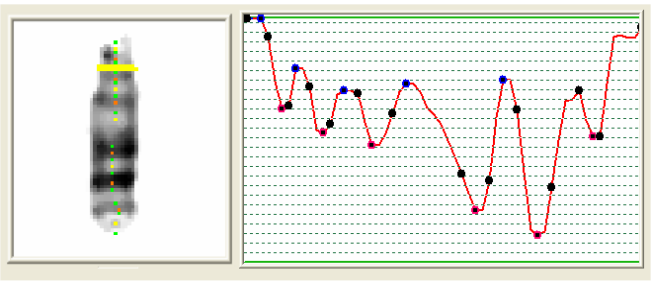

Figure 6
a) Maxima-minima, real thickness points b) Chromosome profile with the points

extracted centromere position [10] together with the maxima and minima points as well the points of maximum grayscale transition along the profile (Band borders points).

B. Proposed Chromosome Features Categories

Geometric and grayscale features were extracted [10]. Geometric features were normalized using the maximum length chromosome in the image. The following categories of features were then extracted:

1) G-Banding grayscale profile features:

These are 12 fixed thickness grayscale average features.

2) Global chromosome's features:

a) Medial axis length (Normalized)
b) Contour length (Normalized)
c) Area (Normalized)
d) Mass (Normalized)

3) Chromosome's centromeric features:

a) Medial axis length ratio of P/Q (Normalized)
b) Contour length P/Q ratio (Normalized)

c) Area P/Q ratio (Normalized)
d) Mass P/Q ratio (Normalized)

4) Number of real bands:

Number of maxima & minima in profile.

5) Real gray level banding:

These are the average graylevel of the real band (sampled to 42).

6) Distances between centers of bands:

These are the distances between the bands centers (sampled to 42).

7) Real bands thickness:

These are the real banding thickness (sampled to 42).

All these proposed features are of unequal-sizes and depends on specific chromosome type. Unequal real banding parameters were stretched (sampled) to a size of 42 which is the maximum number of hills and valleys found in the profile. Figure 7 shows point (Along medial axis points) based real banding model and its automatically extracted centromere position. Figure 8 shows average (Average of grayscale along a perpendicular line of medial axis points) banding profile points. Figure 9 shows cytogenetic ideogram map indicating real number of bands, normalized thickness of bands, and exact centromere position.

IV. Proposed Matching Methods

A dataset of 20 images of 46 chromosomes each are divided into two sets, one to get training templates while the other to test the proposed matching system. All the features were extracted for all the training and testing chromosomes. Three different classification methods were used and were compared for the optimal chromosome classification technique to be used later.

A. Neural Networks

Training different multi-layer perceptron Neural Network architecture is shown in Figure 10 (60 Hidden neurons, 30000 Epochs,0.001 goal MSE, 0.02 Learning factor, and 0.7 Momentum term). The algorithm used to train this class of networks is the *Multi-Layer BackPropagation* training algorithm [11].

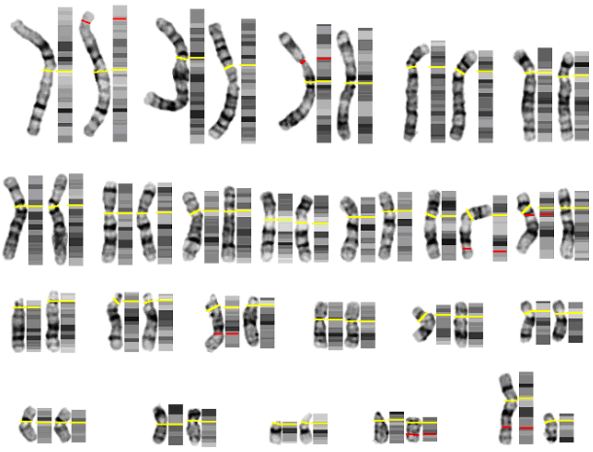

Figure 7
Point based generated banding model

Figure 8
Average based generated banding model

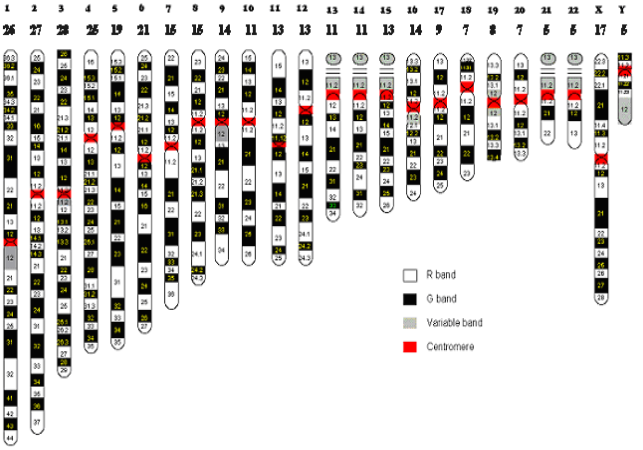

Figure 9
Ideograms map

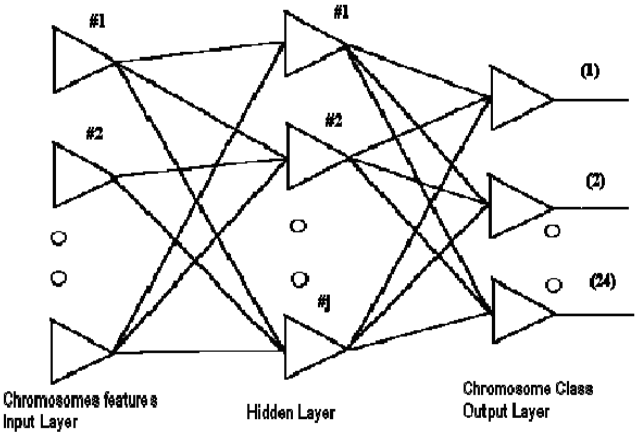

Figure 10
Neural network architecture for multilayer network

B. Fuzzy Rule Based

Features vectors were first fuzzified using its defined membership functions of 11 fuzzy sets each as shown in Figure 11. The training set of 593 vectors for 24 classes of chromosomes were used to derive the fuzzy rules. 279 vectors are used to test the system. Output fuzzy sets are shown in Figure 12. Fuzzy rules were first generated,

validated, and then were used to evaluate the system via an inference mechanism using *SUP MIN* compositional rule of inference [12-19].

C. Template Matching

The 24 template chromosomes types are from chromosome 1-22, X, and Y chromosomes. For example among the 40 chromosomes of type 1, 28 templates were averaged to get the average template of chromosome 1 while 12 are used to test the template matching technique. Features were normalized and stretched to get an overall average template of fixed length for every type [10]. After extracting the mentioned categories of fixed length and variable length features, we stretched banding to the maximum number of maxima and minima found in the profile (42 bands) so that all variable parameters are sampled to a size of 42. Matching is done with both the training template and the cytogenetic ideograms map. In this stage we have a set of reference patterns (templates) and we have to decide which one of these reference patterns an unknown one (test pattern) matches best. A defined simple measure for fixed length features is the Euclidean distance cost measure [11] between template and test vectors. Euclidean cost measures for all features after stretching to fixed size were calculated. These cost measures were weighted average to get an overall cost measure of similarity using a neural network [10], where we trained a net taking the inputs as the costs and the output is either one if correctly classified and zero if not. Then in classification, first we get the costs of different features of this chromosome with either the templates or the whole references, then entering these costs to the net, the one having highest output will be classified to that chromosome average template or reference. A test vector is assigned to that class having the minimum overall cost (Sum). Twelve cost functions (contour-length, medial-axis length, area, mass, contour length-centromeric-index, medial-axis-length-centromeric index, mass-centromeric-index, banding-numbers, 12-fixed-length-banding-profile, real bands grayscale, and real-band-thicknesses) were weighted using neural networks for maximum recognition [10].

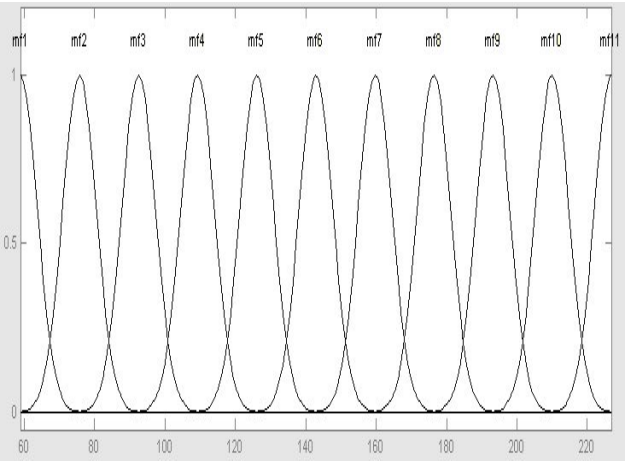

Figure 11
Example of a fuzzy membership functions of grayscale band

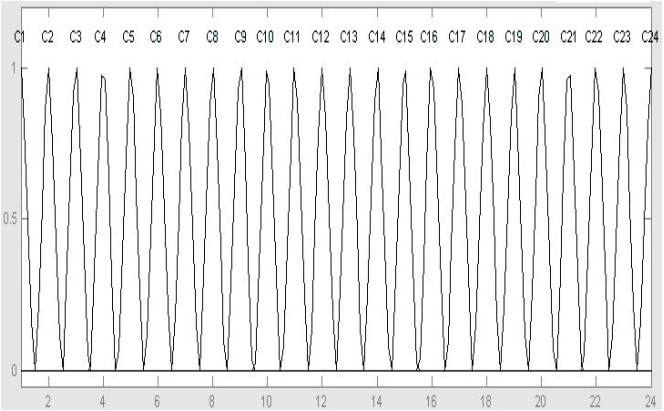

Figure 12
Fuzzy membership functions for the output

V. RESULTS

A Neural Networks classification, Fuzzy Rule Based classification, and Template Matching classification results were compared and are shown in table 1. Figure 13 shows a Klinefelter chromosomes metaphase image. Figure 14 shows its manual Karytyping image. Figure 15 shows output of a Neural Network classification for the image shown in Figure 13, where a misclassified chromosome (5) that should be (4), and a misclassified chromosome (12) that should be chromosome (9). Table 2 shows a comparison of testing times using different classification methods using Pentium 4, 2.4 Mhz, and 512 MRAM computer.

Table I
Classification Results Comparison

Classification Method	Training %	Testing %
Fuzzy Rule Based Classification	100	93.54
Neural Networks Classification	98.81	94.76
Matching with 24 average templates	96.9	95.96
Matching with all reference vectors	97.65	96.89

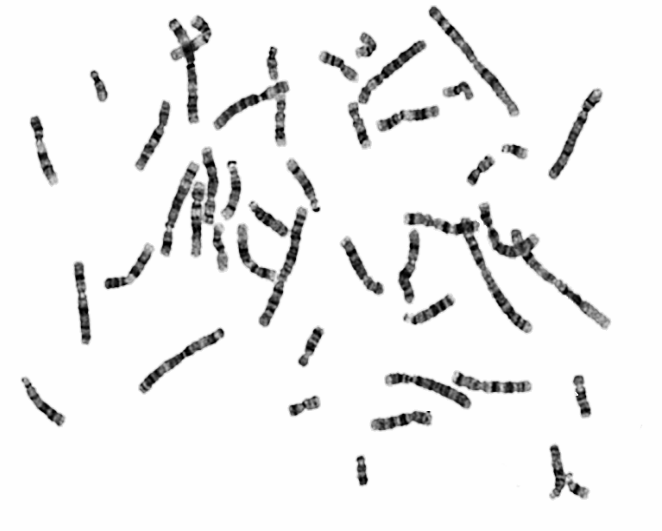

Figure 13
Klinefelter image

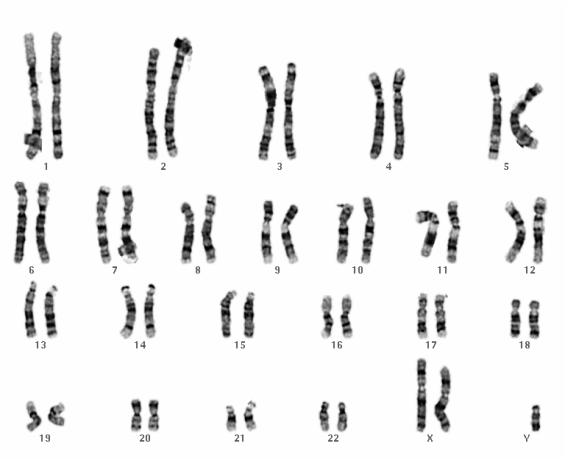

Figure 14
Manual Karyotyping

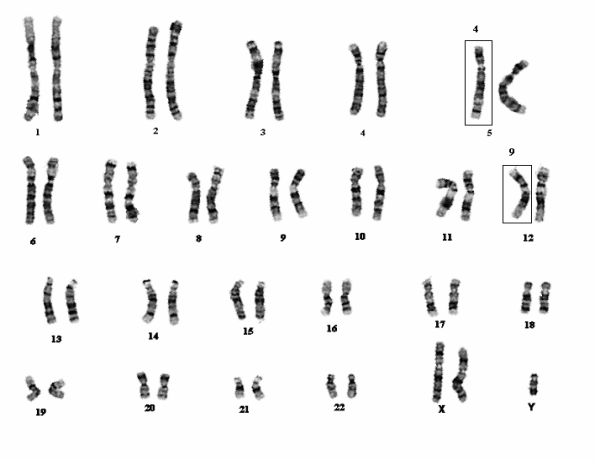

Figure 15
Neural Networks (95.56%) classification result.

Table II
Overall Testing Time Comparison

Testing Time	Seconds
Fuzzy rule based	~0.5 sec
Neural networks	~0.1 sec
Matching with 24 templates	~0.1 sec
Matching with all references	~2.5 sec

VI. CONCLUSIONS

A new features extraction and classification algorithms for real G-banded chromosomes images is proposed. Real bands grayleves features and real band thicknesses features as well as real number of bands (hills and valleys) were automatically extracted for every chromosome together with the shape and fixed banding features. Classification results were found to be more than 93% and were compared for different classification methods. The matching comparison results of the three classification methods using the whole features categories suggests the usage of template matching method with all the references since it has the highest classification rate among other methods (96.89%). Real banding graylevels and thickness will be suitable to analyze chromosome deletions and translocations via bands registration to diagnose structural abnormalities.

VII. ACKNOWLEDGMENTS

This work was partially supported via a Cairo University grant for chromosomes separation and classification (2002). We are grateful to Professor of Clinical Pathology, Dr. Mervat El-Ansari at Cairo Medical School for her valuable suggestions and data preparation of this work.

VIII. REFERENCES

[1] G. Agam, I Dinstein, “Geometric separation of partially overlapping non-rigid objects applied to automatic chromosome classification,” IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-19, pp. 1212-1222, 1997.

[2] B. Lerner, “Toward a completely automatic neural network based human chromosome analysis, ”IEEE Trans. Syst. Man Cyber. Special issue on Artificial Neural Network, vol. 28, pt. B, pp. 544-552, 1998.

[3] J. Liang, “Fully automatic chromosome segmentation,” Cytometry, vol. 17, pp. 196-208, 1994.

[4] P. Mousavi et al., “Feature Analysis and Centromere Segmentation of Human Chromosome Image Using an Iterative Fuzzy Algorithm,” IEEE Transactions on Biomedical Engineering, vol. 49, No. 4, April 2002.

[5] J. Graham, “A neural network classifier for chromosome analysis,” Biology and Biochemistry , Handbook of Neural Computation release 97/1 , 1997 IOP Publishing Ltd and Oxford University Press.

[6] B. Lerner, and Neil D. Lawrence, “A Comparison of State-of-the-Art Classification Techniques with Application to Cytogenetics,” Neural Computing & Applications (2001) 10:39-47, 2001 Springer-Verlag London Limited.

[7] C. Konig and I. Ebert, “Computer-aided quantitative analysis of banded karyotypes, exemplified in G-banded Hyscinthoides italica s.l. (Hyacinthacase),” Caryologa vol. 50, n. 2:105-116, 1997.

[8] Lucas J. van Vliet, Ian T. Young, Ton K. ten Kate, Brain H. Mayall, Frans C.A. Groen, Roelof Roos, “Athena: A Macintosh-Based Interactive Karyotyping System,”Automation of Cytogenetics, C.Lundsteen & J. piper (eds), Springer-Velag, 1989, 47-66.

[9] M. Sonka et al., “Image Processing, Analysis, and Machine Vision,” Chapter 5, Second edition”, PWS Publishing, 1999.

[10] Ahmed M. Badawi, "Cairo university chromosomes image separation and classification system," Cairo University grant technical report 2003.

[11] Sergios Theodoridis et al, ”Pattern Recognition,” Chap. 8, Template Matching, Academic Press, 1999.

[12] L. A. Zadeh,”Outline of a new approach to the analysis of complex systems and decision process,” IEEE Trans. Sys., Man., Cybern., 3, pp., 1973.

[13] Li-Xin Wang and J. M. Mendel,”Generating fuzzy rules by learning from examples,” IEEE Trans. Sys., Man., Cybern., 22, 6,pp. 1414-1427, 1992.

[14] K. Abraham,”Fuzzy techniques in pattern recognition,” John Wiley & Sons Inc., 1982.

[15] C. C. Lee,”Fuzzy logic in control systems,” (Parts I,II) IEEE Trans. Syst. Man, Cybern, 20, 2, pp. 404-435, 1990.

[16] J. K. George and T. A. Folger,”Fuzzy sets: Uncertainity, and information,” Prentice-Hall,1992.

[17] T. Toshiro et al.,”Fuzzy systems theory,” Academic Press, 1992.

[18] Y. Yee et al. ”An algorithm for multiple attribute, multiple alternative decision problem based on fuzzy sets with application to medical diagnosis,” IEEE Trans. Sys., Man., Cybern., 20, Oct., 1980.

[19] Ahmed M. Badawi, Ahmed S. Derbala and Abou-Bakr M. Youssef, "Fuzzy logic algorithm for quantitative tissue characterization of diffuse liver diseases from ultrasound images," International Journal of Medical Informatics 55-2 pp. 135-147, 1999.