Cancer Cell Recognition by Fuzzy Logic

Cheol-Hun Na, *Member, KIMICS*

Abstract— This paper proposes the new method based on fuzzy logic which recognizes between normal and abnormal. The object image was the Thyroid Gland cell image that was diagnosed as normal and abnormal(two types of abnormal : follicular neoplastic cell, and papillary neoplastic cell), respectively. The nuclei were successfully diagnosed as normal and abnormal. The multiple feature parameters (pre-obtained 16 feature parameters of image data) were used to extract the features of each nucleus. As a consequence of using fuzzy logic algorithm, proposed in this paper, average recognition rate of 98.25% was obtained.

Index Terms— *recognition, fuzzy logic, multiple feature parameters.*

I. INTRODUCTION

THIS paper deals with the cancer cell discrimination that calls attentions to the pathologists. The object cell images were Thyroid Gland cells image that diagnosed as normal cell, follicular neoplastic cell, and papillary neoplastic cell, respectively.

The Clinical Cytology which detects the cancer cells by analyze the microscopic images was introduced by Papanicolaou[1]. The Clinical Cytology is the inspection method of detecting the cancer cells by analyzing the microphotographs of cells in medical image processing. Cells are taken from the internal organs of human body and check for the existence of cancer cells. It is a necessary inspection method of detection of the various types of cancers for early diagnosis and treatments. However, discriminations were achieved by human visual system. The digital process of medical image began early 1960' dealing with the microscopic images, X-ray images, and Computer Tomographic(CT) images. Digital image processing methods has been applied to Clinical Cytology[2].

Clinical Cytology is the inspection method of detecting cancer cells by analyzing the microphotographs of cells in medical image processing. Cells are taken from the internal organs of the human body and check for the existence of cancer cells. It is a necessary inspection

method for detecting various types of cancers for early diagnosis and treatments. Digital image processing methods have been applied to Clinical Cytology, but the Clinical Cytology has had many problems for the engineers.

Medical features are difficult to understand for the engineer, and there are various features in every type of cell. The discrimination experiment uses the multiple parameters instead of simple parameters to increase the discrimination rate. The focus of this paper is to find a combination of dominant feature parameters for the recognition of cancer cells, using fuzzy logic and genetic algorithm.

II. FUZZY CLASSIFIER

When the similarity between standard pattern **P** and input pattern **I** is high, the possibility of inclusion in the specific class is high[3]. This simple fact is the basis of classification algorithms. Fuzzy membership function $\mu(x)$ is defined by Eq. (1).

$$\mu(x) = \begin{cases} \dfrac{x - p_M}{P_M - p_L} + 1, & where & p_L \leq x \leq p_M \\ \dfrac{x - p_M}{P_M - p_R} + 1, & where & P_M \leq x \leq p_R \\ 0 & , & otherwise \end{cases}$$
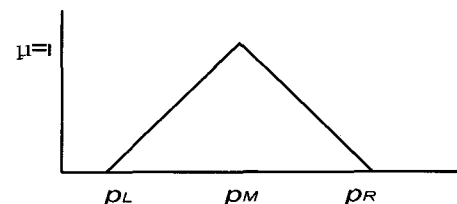
(1)

Fig. 1. Fuzzy membership function.

p_L, p_M and p_R are parameters of the fuzzy membership function. The standard pattern $P(i, j)$ in Eq. (2) is defined by mean values of features in each class. The number of pattern in each class is not the same, but the feature vectors have the same dimension. $I_k(i, j)$ is *i-th* class and *j-th* feature and *k-th* input pattern.

Cheol-Hun Na is with the Department of Information & Communication Engr., Mokpo Nat'l Univ. Chonnam, 534-729, Korea (Email: chna@mokpo.ac.kr)

$$\mathbf{P}(i,j) = \sum_{k=1}^{N_I} \mathbf{I}_k(i,j) / N_I \qquad \begin{cases} i = 1,2,\cdots,N_C \\ j = 1,2,\cdots,N_F \end{cases}$$

N_C : Number   of feature

N_F : Number   of class

N_I : Number   of input   patterns   in each   class

(2)

We apply Eq. (3), which stands for similarity between standard patterns and input patterns, to calculate the fuzzy pattern matching[4] in each class. FPM_{ij} is the value of the fuzzy pattern matching in i-th class and j-th feature.

$$FPM = \{FPM_{ij} \mid 1 \le i \le N_c, 1 \le j \le N_F\}$$
$$= \min_{k=1}^{N_I}(1 - \mid \mu_P(x) - \mu_{I_k}(x) \mid)$$

(3)

$\mu_P(x)$ is a membership function of a standard pattern, and $\mu_{I_k}(x)$ is a membership function of the i-th input pattern. When the similarity between the input patterns between standard pattern is higher, the possibility of inclusion in standard pattern is also higher. To calculate the class with a standard pattern, we first define the degree of matching M_i in the Eq. (4). The s_R is a selection bit and means inclusion or exclusion of features in each rule, N is the number of features in each rule.

$$M_i = \sum_{j=1}^{N_F} \frac{FPM_{ij}}{N}$$

(4)

The class of input pattern from the membership value of input pattern is obtained Eq. (5)

*If* $\text{Max}_{i=1}^{N_c} M_i$ is M_c, Then Class(I) is c

(5)

III. FUZZY RULE GENERATION USING THE GENETIC ALGORITHMS.

Genetic algorithms are a kind of stochastic optimization method, modeling native natural evolution phenomena[4]. Usual Genetic Algorithms consist of three kinds of genetic operations; selection, crossover and mutation. They work as stochastic operations on a population to make a change of generation. Crossover combines substructures of parents to produce new individuals. It is the most characteristic operation as an optimization method because other methods as simulated annealing did not use such global multiple search points.
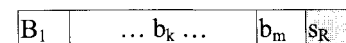
III-1 String Representation

To solve any optimization problem using a genetic algorithm, it is necessary to code scheme and to encode the parameters of the problem into a string. Genetic Algorithms are applied to obtain the shape and number of fuzzy membership function and an inclusion or exclusion of feature for classification. String representation and an example of real fuzzy membership function in Genetic Algorithms are shown in fig. 2. The numeral 1 in the membership function part represents the boundary value in the membership function, and 0 and 1 in the rule selection part represents exclusion and inclusion.

Three successive 1s, except 0 in b_k, are the lower, middle, and upper parameters in membership function. The optimal strings can be produced using Genetic Algorithms[5]. An optimal string is an optimal fuzzy membership function and comprises the maximum recognition rate. Furthermore, the rule selection bit makes for better performance by eliminating the negative effect due to non-effective features. To select effective feature vectors guarantee better than recognition rate. We classify feature vectors into dominant, recessive, and extra feature vectors. The proposed rule selection bit excludes recessive feature vectors and includes dominant feature vectors to increase the recognition rate.
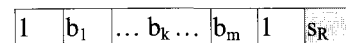
(a) String representation

(b) Reduced string representation
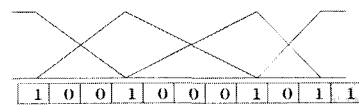
(c) Fuzzy membership function example

Fig. 2. String representation and its example.

III-2 Fitness Function

A fitness function must be relevant to the problem being optimized. The fitness value of each string is computed from the fitness function. A good string is one that scores a high fitness value. The fitness value is defined in the recognition rate (*RR*). We classified the higher 30% of strings to a superior set. All operations, for example reproduction, crossover, and mutation, are conducted as superior strings and the other strings.

III-3 Genetic Operator

A. Reproduction

The reproduction operation is performed as follows:

- Normalize the fitness value of each string such that the sum of the fitness values of all the strings in the current population is equal to 1.

- Partition a unit-length scale into population size N slots, where each slot size is in proportion to the normalized fitness value of a string in the current population

- Generate N random numbers from 0 to 1 and see where the number falls on the scale. The string corresponding to the division where a random number falls is selected to be a member for the new population.

For better performance, superior string must survive in the next generation. The String with the best performance has a higher selection probability. Crossover operation was conducted in the next step. In the reproduction procedure, strings for the crossover operation are selected in superior strings and other strings. The size of population is 20 and the number of superior strings is 6 in every experiment.

B. Crossover Operation

The crossover operator is the most important operator in Genetic Algorithms. It is the mating operator that allows production of new strings through a combination of parts of strings. The probability for crossover operation is 0.5.

- Pairs of members of the newly reproduced strings are randomly selected for mating.

- For each pair of selected strings, parts of the strings are swapped from a pair of new strings. The position of a string to which the swapping takes place is randomly selected.

C. Mutation Operation

Mutation is needed because, even though reproduction and crossover effectively search and recombine extant notions, occasionally they may become overzealous and lose some potentially useful genetic materials. The mutation operator protects against such irrecoverable loss. The probability for mutation operation is 0.05.

IV. EXPERIMENTAL RESULTS

To show the effectiveness of the proposed method, we apply our algorithms 150 IRIS data and 57 Thyroid Gland cancer cells data and use the 16 feature parameters[6].

IV-I   IRIS Data[7]

IRIS data, which is standard data for pattern classification problems consist of 4 features and 3 classes. To show the effectiveness of the proposed method, we show the classification results in Fig. 3. The size of the string is changed by 1 to 7, except with the rule selection bit, and the maximum recognition result was obtained in 10 time experiments.

The results show that the proposed method has a 98.67% recognition rate(148 correct classification in 150 patterns)when the size of string is 4 and the number of rule is 1. Fig. 4 shows the results of the multiple classification rules used by Ishibuchi[8].

Ishibuchi has only one misclassification pattern in 150 patterns using 5 classification rules and has a 100% classification result using 7 rules. Young could classify 146 patterns using only one rule. This result is the same even when we didn't use the rule selection bit.

IV-II Recognition of Thyroid Gland Cancer Cells

Thyroid cancer cells have two types of papillary and follicular cancer cells[6]. The selection rule classifying cancer cells and normal cells with the use of a expert is subjective and too complicated to explain. The input pattern has 16 features, 16 normal cells, 16 papillary cancer cells, and 25 follicular cancer cells. Fig. 5 show recognition rates when the size of the feature vector was changed 4 ~ 1 and rule selection bit was used, or not. The number in parenthesis is the number of misclassified patterns.
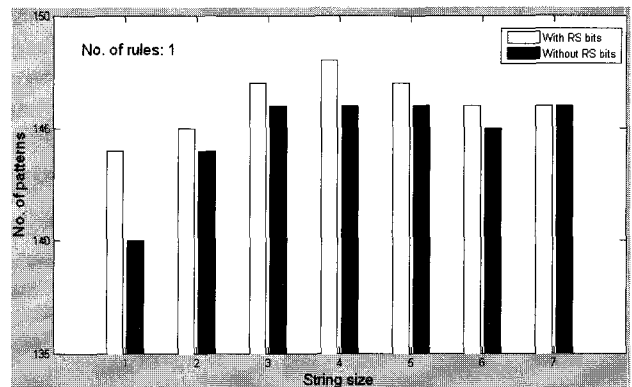

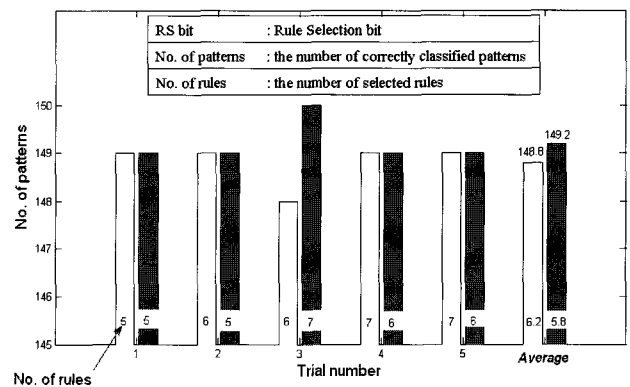Fig. 3. Simulation results with different string size.

Fig. 4. Simulation results with different crossover operations[8].
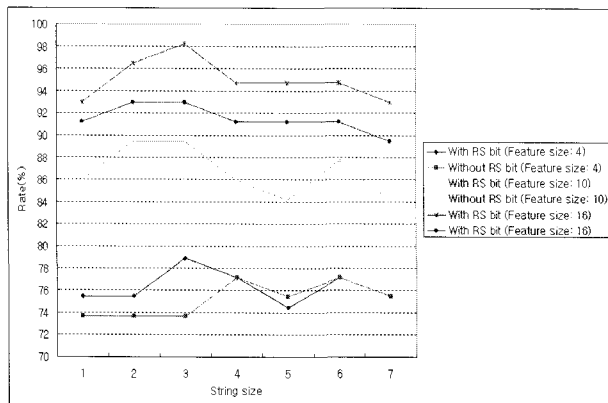
Fig. 5. String size and recognition rates with/without RS bit, Feature size.

The maximum recognition results were obtained when the number of selected features is 10 or 16 and the size of string is 3, using the rule selection bit. The number of misclassified patterns is only one. When we didn't use the rule selection bit, we get maximum recognition rates in the number of features and is 12 or 16, and string size is 2 or 3. The number of misclassification pattern is 4.

The maximum recognition rate was 96.49% and 94.74% when the string size was 2 and 5, respectively. The maximum recognition rate was 88.75% in reference[6] .

V. PERFORMANCE OF REPETITIVE ANALYSIS AND DOMINANT FEATURES

The size of the string is the important part to determine the shape of fuzzy membership functions and has a direct relationship to the performance of recognition rates.

When we guarantee proposed method search of all space, a bigger size of string means better recognition rates.

But it is impossible, or not effective, to guarantee these conditions. We use a repetitive analysis method increases, the search space in small volumes to larger ones. Fig. 6 shows some examples when the string size is 1, 3, and 7. When the string size is 1, the optimal membership function is a part of those in string size of 3. Therefore recognition rate in string size 3 must be the same or at least greater than those in string size 1. The increment of string is shown in fig. 6.

After saturation in the recognition rates occurs, the size of string will be doubled as demonstrated in fig. 6. The recognition rate using the proposed method is shown in fig. 7. Initial string size L_{String} is 1, 2, 4, 6, 8, ... When the string size is doubled, recognition rates will not decrease any more. The recognition rate using the proposed method is shown in fig. 6.

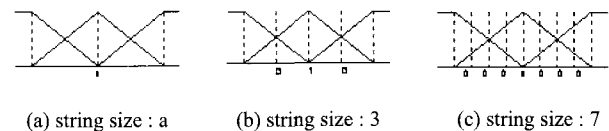(a) string size : a   (b) string size : 3   (c) string size : 7

Fig. 6. Fuzzy membership functions for each string size.

When rule selection bit is applied to classification, additional 1.06% recognition rate in IRIS data and 6.44% recognition rate in gland cancer cell were obtained. The rule selection bit is more effective when the input feature is greater. Three 98.25% maximum recognition rates were obtained by using the dominant input features and fuzzy membership functions. We could get same dominant features, although fuzzy membership functions show change. This result shows that the selection of dominant features is very important in classification problems.

VI. CONCLUSION.

In this paper, a new method of recognition for medical image analysis was studied by using pattern recognition techniques. The focus of this paper is the automatic recognition of cells into normal and abnormal cells.

The object cell images used in this paper were microscopic images of Thyroid Gland cells and a new technique for the recognition of cells image which uses Fuzzy Logic and Genetic Algorithms was proposed. The fuzzy classifier has a simple structure, which contains a classification part based on fuzzy logic theory and a rule generation part using Genetic Algorithms. The rule generation part determines optimal fuzzy membership functions and inclusion or exclusion of each feature in fuzzy classification rules.

We analyzed the recognition rate of a specific object, then added the finer features repetitively, if necessary, to the object with large misclassification rates. Further, we introduce a repetitive analyses method for the minimum size of string and population, and for the improvement of recognition rates. This classifier is applied to two examples of the recognition of iris data and the recognition of thyroid gland cancer cells. The fuzzy classifier proposed in this paper has recognition rates of 98.67% for iris data and 98.25% for thyroid gland cancer cells.
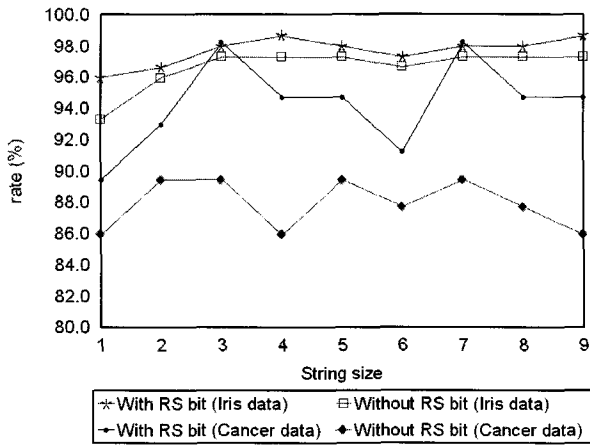
Fig. 7. Recognition rates with/without RS bit using repetitive analyses method for Iris data and Thyroid Gland Cancer cells data(Feature size: 10).

REFERENCES

[1] Sing-Tze Bow, Pattern Recognition and Image Preprocessing, Marcel Dekker, New York, 1992.

[2] A. Khotanzad and R. Kashyap, " Feature Selection for Texture Recognition Based on Image Systhesis", IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC-17, pp. 1087-1095, Nov., 1987

[3] W. Galvraith, et, al., " Studies on Papanicolaou Staining," Analytical and Quantitative Cytology, Vol. 1, No. 3, pp. 160-169, 1979.

[4] J. C. Bezdek and S. K. Pal, Fuzzy Models for Pattern Recognition, IEEE Press, 1992.

[5] R. J. Marks II, Fuzzy Logic Technology and Applications, Technical Activities Boards, 1994.

[6] M. Yamamura, H. Satoh, and S. Kobayashi, "An Analysis of Crossover's Effect in Genetic Algorithms", Proceedings of The 1st IEEE Conference on Evolutionary Computation, pp.613~618, 1994

[7] Cheol-Hun Na et, al., "Cancer Cell Recognition by Fuzzy Logic in Medical Images", GSPx 2005, Santa Clara, CA, U.S.A., 2005.

[8] Ishibuchi, K. Nozaki, N. Yamamoto and H. Tanaka, "Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms", Fuzzy Sets and Systems, vol. 65, pp. 237~253, 1994.
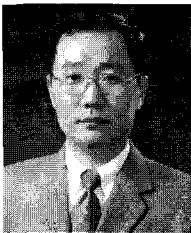
Cheol-Hun Na   Received B.S., M.S. and Ph.D. degree from Elecrtonic Engineering, Chonnam Nat'l Univ. in 1985, 1987 and 1994, respectively. From 1995, he worked as a professor of Mokpo Nat'l University. His research interest is in the area of Digital Images   Processing,   Communication Engineering, and Medical Imaging.