Genetic Algorithm Based Feature Selection In a Recognition Scheme Using Adaptive Neuro Fuzzy Techniques

M. Bhattacharya, A. Das

Mahua Bhattacharya
Indian Institute of Information Technology & Management
Morena Link Road, Gwalior-474003, India
E-mail: mb@iiitm.ac.in

Arpita Das
Institute of Radio Physics & Electronics
University of Calcutta
92, A.P.C. Road, Kolkata-700009
E-mail: arpita.rpe@caluniv.ac.in

Abstract:
The problem of feature selection consists of finding a significant feature subset of input training as well as test patterns that enable to describe all information required to classify a particular pattern. In present paper we focus in this particular problem which plays a key role in machine learning problems. In fact, before building a model for feature selection, our goal is to identify and to reject the features that degrade the classification performance of a classifier. This is especially true when the available input feature space is very large, and need exists to develop an efficient searching algorithm to combine these features spaces to a few significant one which are capable to represent that particular class. Presently, authors have described two approaches for combining the large feature spaces to efficient numbers using *Genetic Algorithm* and *Fuzzy Clustering* techniques. Finally the classification of patterns has been achieved using adaptive neuro-fuzzy techniques. The aim of entire work is to implement the recognition scheme for classification of tumor lesions appearing in human brain as space occupying lesions identified by CT and MR images. A part of the work has been presented in this paper. The proposed model indicates a promising direction for adaptation in a changing environment.
Keywords: Adaptive neuro- fuzzy, Genetic algorithm, Feature selection, pattern recognition.

1 Introduction

The boundary detection based on Fourier Descriptors introduces a large number of feature vectors in a pattern recognition scheme. To classify different boundaries, any standard classifier needs large number of inputs and to train the classifier large number of training cycles and huge memory are also required. A complicated structure of the classifier invites the problem of over learning, and which may cause for misclassification [2]. Therefore need exists for significant feature selection for efficient pattern recognition scheme. Among many existing methods for solving feature selection problem (FSP), pruning methods for neural network [7],[8], classification trees [9] fuzzy clustering [10] may be referred. GA is an efficient search algorithm based on the mechanics of natural selection and natural genetics [1]. It combines survival of the fittest among string structures with a structured yet randomized information exchange to form a search algorithm with some of the innovative flair of human search. Since genetic algorithm is invented to simulate evolutionary processes observed in nature the goal of survival or optimization in a changing environment could be achieved [3]. However, GA [1],[4],[5],[6],[11] differs

from other searching algorithm in that sense, it does not deal with the neighborhood of a single current solution. GA use a collection (or population) of parameters, from which using selective crossover and mutation strategies, better solutions may come out. In present paper, the network architecture used for final classification is ANFIS adaptive neuro fuzzy inference system. ANFIS [13],[14] architecture for Sugeno fuzzy model is an innovative soft computing expert system that removes the limitations of conventional neural networks [12],[13],[15]. The proposed method of feature selection has been compared with Fuzzy clustering theory where GA based feature selection shows the improvement over fuzzy clustering due to natural selection mechanisms. Proposed FSP methodology combined with ANFIS classifier is an intelligent, expert system that gives the user accurate detection even in presence of additive noise. The objective of entire work is to identify the different space occupying lesions appearing in human brain as tumor / cancer lesions in different grades of benignancy / malignancy using boundary as feature. Presently a part of the work has been presented considering few pattern boundaries in order to develop an accurate classification technique using GA based feature selection.

2   Proposed Methodology

In the proposed method the significant boundary of ROI is

extracted and GA has been applied to reduce the feature vector size. These reduced and significant features are then fed to ANFIS Sugeno fuzzy network for classification. A comparative study has been conducted for efficient feature selection using both GA and FCM and finally to classify patterns using ANFIS. This study effectively gives the superior results for GA based feature selection. The method is summarized in Figure-1.
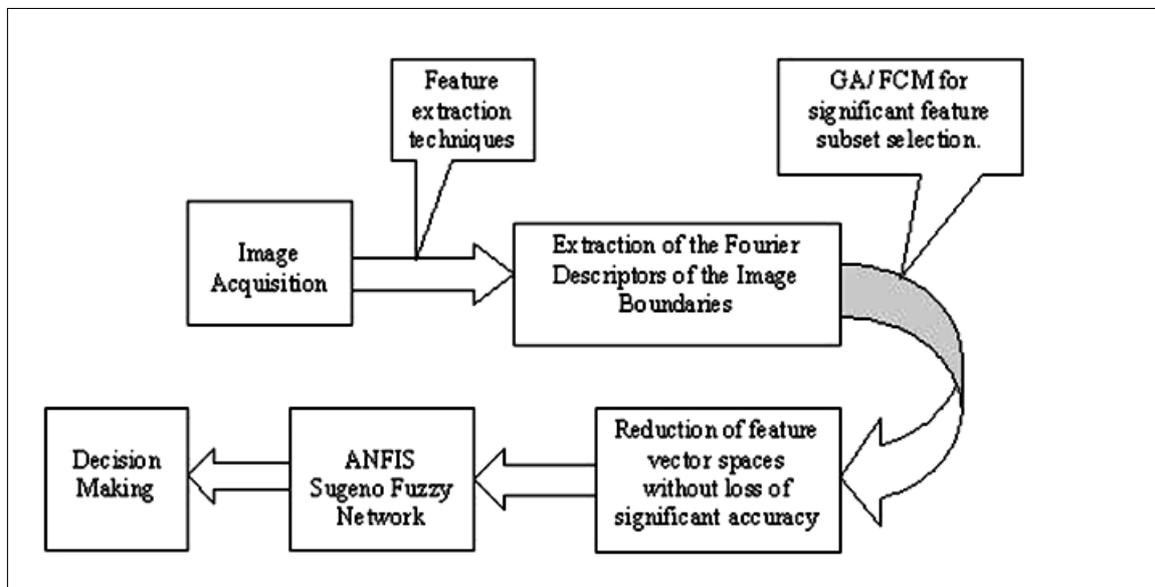

Figure 1: Proposed technique

2.1   Boundary Extraction using Fourier Descriptors

Feature selection is the choice of descriptors in a particular application. The boundary of pattern to be analyzed has been detected by implementing *Canny edge detector* and *Fourier Descriptors* of the edges then used as shape information. A figure with k-points digital boundary in the x-y plane as, $x(k) = x_k$,

$y(k) = y_k$ can be represented as

$$s(k) = [x(k)y(k)] \text{ for } k = 0, 1, 2, ..., k-1. \tag{1}$$

Each co-ordinate pair can be treated as a complex number so that,

$$s(k) = x(k) + j * y(k) \text{ for } k = 0, 1, 2, ..., k-1. \tag{2}$$

The Discrete Fourier Transform (DFT) of s(k) is given below

$$a(u) = \frac{1}{k} * \sum_{k=0}^{K-1} s(k) * e^{-\frac{j2(\pi)uk}{K}} \text{ for} k = 0, 1, 2, ..., k-1. \tag{3}$$

The complex coefficient $a(u)$ is called the Fourier Descriptor of the edge points. Let us suppose that instead of all Fourier coefficients, only the first 'P' coefficients are used. This is equivalent to set $a(u) = 0$ for $u > (P-1)$. The overall global shape of the images has been identified (It can be shown that if $P \approx u/3$, approximate boundary detection would be possible). Thus a few Fourier descriptors can be used to capture the gross essence of a boundary. This property is valuable, because these coefficients carry shape information and can be used as the basis for differentiating between distinct boundary shapes.

2.2   Genetic Algorithm for Feature Selection

GA manipulates chromosomes, which are the encoded string set of parameters of a target system to be optimized. Presently different boundaries extracted from CT and MR images of section of human brain having space occupying lesions are recognized on the basis of Fourier Descriptors and which play the role of payoff values (objective function) associated with individual strings. In GA, a new set of offspring has been created in every generation on the basis of the fittest of old generation. GA efficiently exploits the historical information to speculate on new search points with expected improved performance. It is the best learned from the careful study of biological example that, where robust performance is desired, nature does it better which is the secret of adaptation and survival. GA uses three operators: selection (or reproduction), crossover and mutation to achieve the goal of evolution [1],[3].

2.3   Fuzzy C-Means Clustering Algorithm for Feature Selection

In the proposed method, fuzzy c-means clustering algorithm used for reduction of input feature vector sizes without loss of accuracy level of detection.

Algorithm 1. *Let $X = \{x_1, x_2, ..., x_n\}$ be a set of given data. A fuzzy c-partition of X is a family of fuzzy subsets of x, denotes by $P = \{A_1, A_2, ..., A_c\}$, which satisfies*

$$\sum_{i=1}^{c} A_i(x_k) = 1 \text{ for all } k \in N_n \tag{4}$$

and $i = 1$

$$0 < \sum_{k=1}^{n} A_i(x_k) < n \text{ for all } i \in N_c \tag{5}$$

where c is a positive integer Given a set of data $X = \{x_1, x_2, ..., x_n\}$, where x_k, in general is a vector, for all $k \in N_n$, the problem fuzzy clustering is to find a fuzzy pseudo partition and the associated cluster centers by which the structure of the data is represented as best as possible. To solve the problem of fuzzy

*clustering, we need to formulate a performance index. Usually, the performance index is based upon
cluster centers, $v_1, v_2, ..., v_c$ associated with the partition are calculated by the formula.*

$$v_i = \frac{\sum_{k=1}^{n} [A_i(x_k)]^m x_k}{\sum_{k=1}^{n} [A_i(x_k)]^m} \qquad (6)$$

*for all $i \in N_c$, where $m > 1$ is a real number that governs the influence of membership grades. Observe
that the vector v_i calculated by above equation is viewed as the cluster center of the fuzzy class A I, is
actually weighted average of data in A_i. The performance index of a fuzzy pseudo partition P, $J_m(P)$, is
defined in terms of the cluster centers by the formula*

$$J_m(P) = \sum_{k=1}^{n} \sum_{i=1}^{c} [A_i(x_k)]^m \|x_k - v_i\|^2 \qquad (7)$$

*where $\|x_k - v_i\|^2$ represents the distance between x_k and v_i . Clearly, the smaller the value of $J_m(P)$, the
better the fuzzy pseudo partition P. Thus, the goal of fuzzy c-means clustering method is to find a fuzzy
pseudo partition P that minimizes the performance index $J_m(P)$.*

2.4   Classification of Features using ANFIS model

A generalized ANFIS model based on Sugeno fuzzy architecture is utilized for classification of
significant features. The numbers of input nodes are equal to the reduced input feature space sizes. The
number of membership functions in each of the input node is continually adjusted to achieve the optimum
classification results. To adapt the model with ever-changing environments, hybrid-learning rule is used.
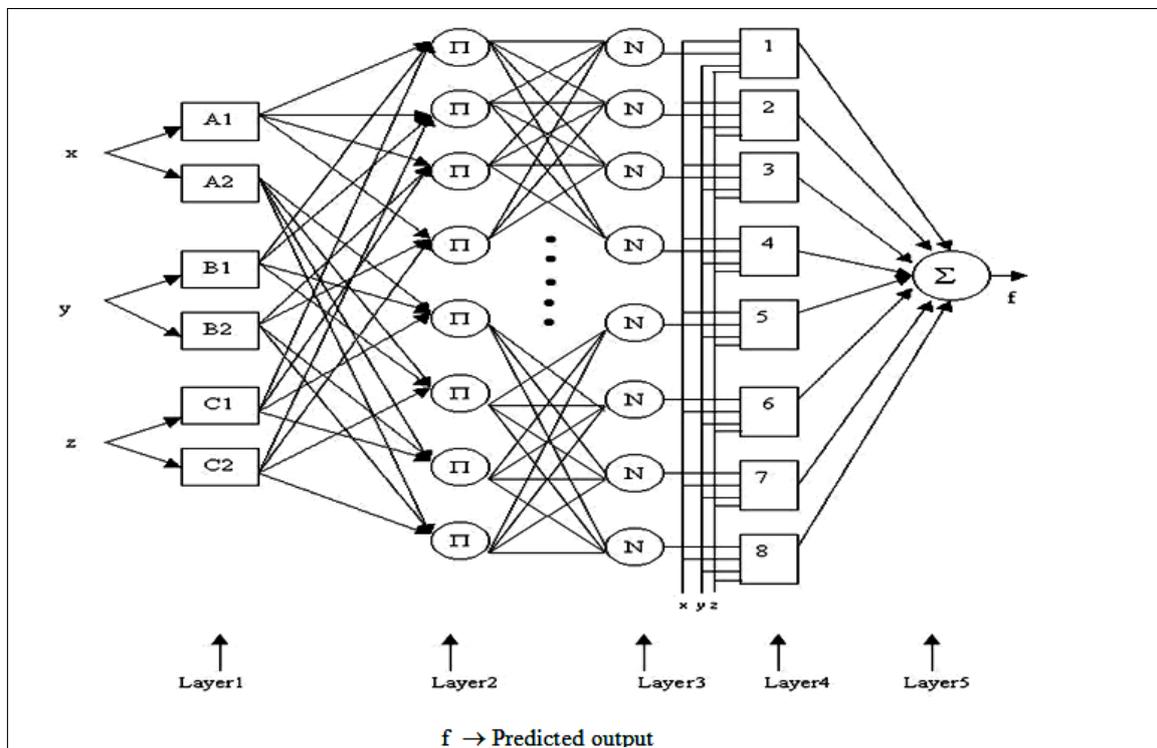

Figure 2: The ANFIS Model for Final classification.

Figure-2 illustrates the reasoning mechanism of the Sugeno fuzzy ANFIS architecture for bound-
ary detection and texture analysis of masses respectively, where nodes of the same layer have similar
functions as described below.

Layer 2. Every node I in this layer is an adaptive node with a node function $O_{1i} = \mu_{A_i}(x)$ for $i = 1, 2$. and $O_{1i} = \mu_{B_i}(y)$ where x (or y) is the input to node i and A_i (or B_i) is a linguistic label (such as large or small) associated with this node. In other words O_{1i}, i is the membership grade of fuzzy set $A(A_1, A_2)$ or $B(B_1, B_2)$. Here the membership function for A can be any appropriate parameterized membership function, such as generalized bell function:

$$\mu_A(x) = \frac{1}{1 + |\frac{(x-c_i)}{a_i}|^{2b}} \tag{8}$$

where a_i, b_i, c_i is the parameter set. As the values of these parameters change, the bell-shaped function varies accordingly. Parameters of this layer are referred to as premise parameters.

Layer 3. Every node in this layer is a fixed node labeled $\prod$, whose output is the product of all the incoming signals:

$$O_{2,i} = w_i = \mu_{A_i}(x)\mu_{B_i}(x) \text{ for } i = 1, 2. \tag{9}$$

In general, any T-norm operator that performs fuzzy AND can be used as the node function in this layer.

Layer 4. Every node in this layer is a fixed node labeled N. The ith node calculates the ratio of the rule's firing strength to the sum of all rule's firing strengths:

$$O_{3,i} = \overline{w_i} = \frac{w_i}{(w_1 + w_2)} \text{ for } i = 1, 2. \tag{10}$$

For convenience, outputs of this layer are called normalized firing strengths.

Layer 5. Every node i in this layer is an adaptive node with a node function

$$O_{4,i} = \overline{w_i}f_i = \overline{w_i}(p_i x + q_i y + r_i) \tag{11}$$

where w_i is a normalized firing strength from layer 3 and p_i, q_i, r_i is the parameter set of this node. Parameters of this layer are referred to as consequent parameters.

Layer 6. The single node in this layer is fixed node labeled $\sum$, which computes the overall output as the summation of all incoming signals:

$$\text{Overall output } = O_{5,i} = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{12}$$

Thus ANFIS architecture is functionally equivalent to a Sugeno fuzzy model.

Hybrid leaning rule combines steepest decent method and least-squares estimator for fast identification of parameters in ANFIS model. For hybrid learning to be applied in a batch mode, each epoch is composed of a forward pass and a backward pass. In the forward pass, after an input vector is presented, node outputs go forward until layer 4 and consequent parameters are identified by the least squares method. In the backward pass, the error signals propagate backward and the premise parameters are updated by gradient decent. The hybrid approach converges much faster since it reduces the search space dimensions of the original pure back propagation.

2.5 Decision Making Logic

The ANFIS model is trained with targets for each of the output classes, which are well separated, and then membership functions are generated for detecting the possible range of output values. Each membership function corresponds to each of the output class; the overlapping regions between two or more classes give the possibility of existing of the particular pattern in all of the overlapped classes. But highest membership grade determines that the particular image pattern belongs to the corresponding

class. Thus to construct a boundary region for a particular class, we design a decision rule using fuzzy if-then conditions that states:

$2.5 <=$ output value $<= 7.5$, test image belongs to class-A;

$7.5 <=$ output value $<= 12.5$, test image belongs to class-B;

$12.5 <=$ output value $<= 17.5$, test image belongs to class-C.

The decision making membership function through the range of all possible output values is given in Figure-3. Each membership function is a generalized bell shaped curve, which corresponds to each output classes; the overlapped region between two or more classes gives the possibility of existing of the particular pattern in all of the overlapped classes.
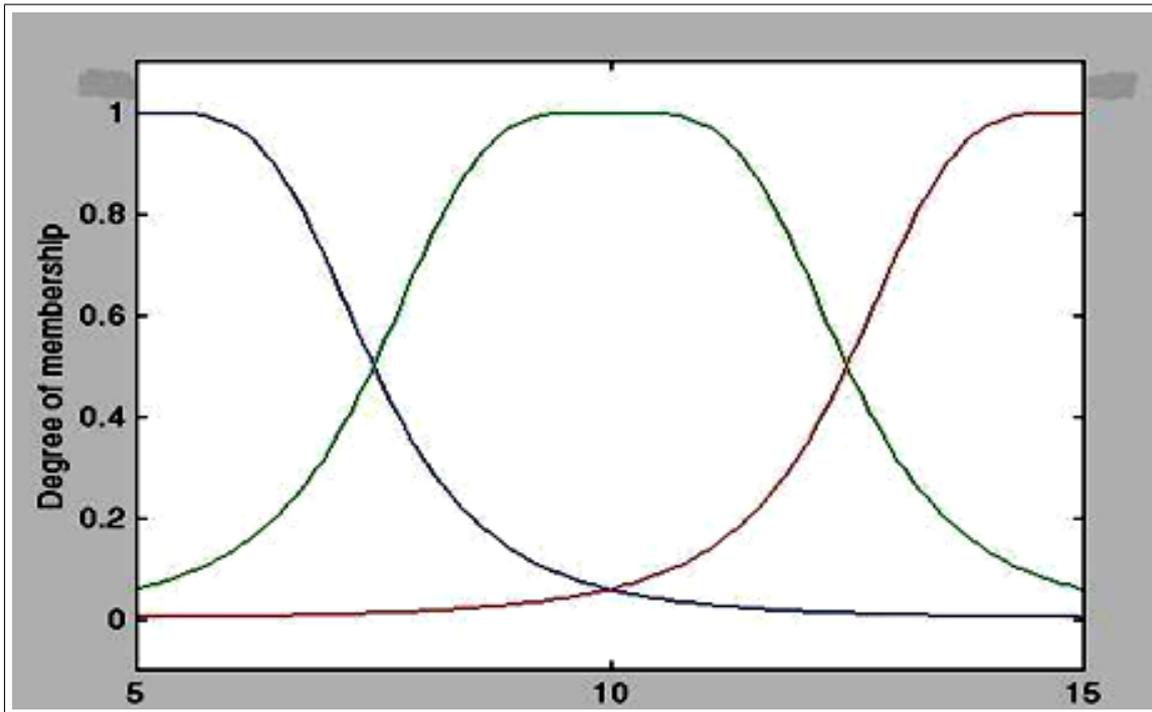

Figure 3: Output decision making membership function

3   Experimental Results

The experiment has been conducted with three distinct boundary shapes extracted from CT and MR images for section of human brain having space occupying lesions shown in Figure 4 belonging to class-A, class-B & class-C. Two membership functions are chosen for each input terminal of the network, to obtain the best possible classification result. The superiority of GA is investigated over the conventional FCM clustering technique to classify the noisy images.

3.1   Choice of String Length in GA Based Feature Subset Selection Problem

In genetic algorithm a particular string of length l contains 2^l search points. As a result, a population of size n contains some where between 2^l to $n*2^l$ search points, depending upon the population diversity. Now among these large numbers of search points, only a few are processed in a useful manner. The reproduction, crossover and mutation operators determine the exponential growth or decay of important search points from generation to generation. It has been observed that GA with samples containing less number of bit strings which are shifted towards the enumerative search. With string length 20,
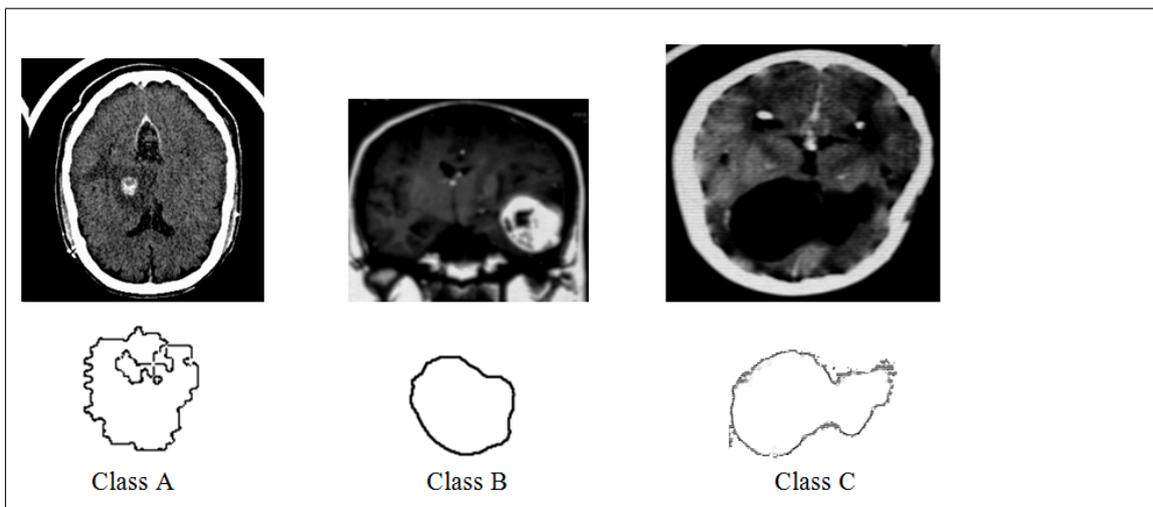
Figure 4: Boundary Features of tumors in human brain

there are a least $2^{20} = 1.04 * 10^6$ search points in the search space. Thus GA converges rapidly with the samples containing large string length. But too much increase of string length is not profitable for computational enumeration. Figure 5 shows an optimum sting length which is 20 and is acceptable for efficient feature subset selection. The variations of average and maximum values of objective function and the corresponding population size for each generation with different string lengths are shown below in Figure $5, 6, 7$ respectively.
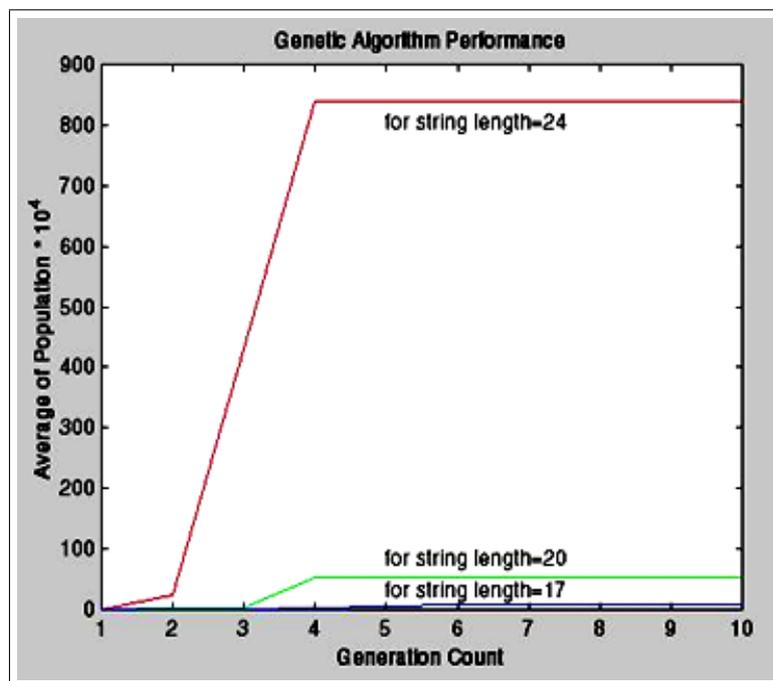

Figure 5: Variations of average value with different string length

It is also viewed from above results that GA with samples containing less number of bit string, shifted towards the enumerative search or random walk. But with string length 24 or more, there are at least $2^{24} = 1.68 * 10^7$ search points in the search space and random walk or enumeration would not be profitable. Thus GA converges rapidly with the samples containing large number of bit string.
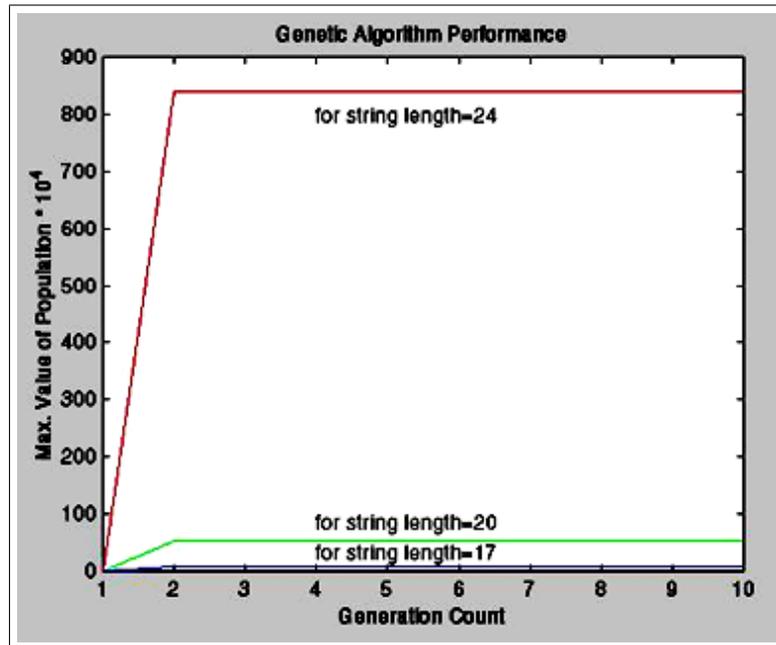
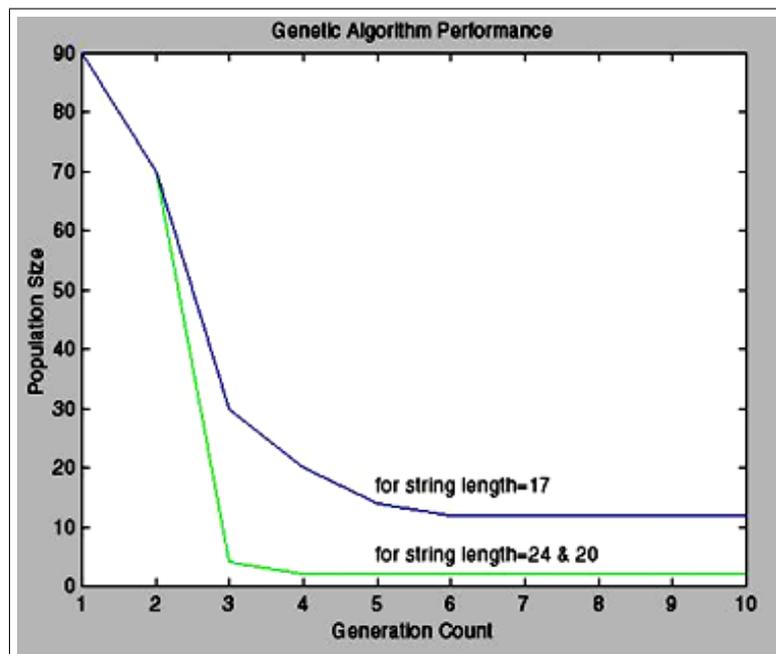Figure 6: Variations of maximum value with different string length

Figure 7: Variations of Population size with different String Length.

Table 1: Recognition of Distinct Noise Free Test Images Using GA and ANFIS Model

| Tested Image | Classification Rate | | Decision |
	Training Value	Tested Value	
Boundary-1	5.0000	4.9496	Belongs to class-A.
Boundary-2	10.000	9.8162	Belongs to class-B.
Boundary-3	15.000	14.8400	Belongs to class-C.

Table 2: GA based

| Noise | Classification Rates | | | Decision |
	Training Value	Tested Value	Classification	
0.002	5.0000	5.5309	class-A	Correct
0.004	5.0000	4.5801	class-A	Correct
0.006	5.0000	5.9943	class-A	Correct
0.008	5.0000	6.6275	class-A	Correct
0.010	5.0000	5.0767	class-A	Correct
0.015	5.0000	3.0135	class-A	Correct
0.020	5.0000	8.6202	class-B	Misclassification

3.2 Results of Experiment to Recognize the Distinct Noise Free Test Images using GA Based Feature Subset Selection & ANFIS Model

ANFIS Sugeno fuzzy model is implemented to recognize the image boundary-2. The network with three significant input features and two optimum membership functions on each would result in $2^3 = 8$ fuzzy if-then rules and thus the input space is partitioned with 8 grids.

3.3 Comparative study of GA & FCM based Feature Subset Selection (FSS) Mode in presence of Noise

In the proposed model, the inputs of ANFIS network are GA based feature subset. This reduced feature subset helps to form a simple ANFIS classifier. Table-2 and Table-3 compare the classification rates of GA and FCM Based FSS model respectively for Image Boundary-1 in presence of Gaussian noise.

Table 3: FCM based

| Noise | Classification Rates | | | Decision |
	Training Value	Tested Value	Classification	
0.002	20.0000	39.6124	class-A	Correct
0.004	20.0000	83.5869	class-C	Incorrect
0.006	20.0000	33.3989	class-A	Correct
0.008	20.0000	96.9512	class-C	Incorrect
0.010	20.0000	34.6393	class-A	Correct
0.015	20.0000	92.7572	class-C	Incorrect
0.020	20.0000	91.0327	class-C	Incorrect

4   Discussions

Authors have presented a pattern recognition scheme by efficiently selecting the significant features and finally using adaptive neuro-fuzzy techniques for design of classifier. For efficient feature selection, two approaches like *Genetic Algorithm* and *Fuzzy Clustering* techniques have been implemented. Finally the classification of patterns has been achieved using adaptive neuro-fuzzy techniques. The aim of entire work is to implement the recognition scheme for classification of tumor lesions appearing in human brain as space occupying lesions identified by CT and MR images. The comparative study of GA and FCM based feature subset selection (FSS) reveals that there is a large possibility of misclassification if FCM is used for significant FSS in presence of noise. GA based FSS is resistant from noise up to a certain level and classification rate is improved for GA based FSS model. This is because, FCM has partitioned the large number shape descriptors such that the degree of association is strong for the descriptors within the same cluster and weak for the descriptors in different clusters. Genetic Algorithm (GA) searched the significant shape descriptors by applying the beauty of natural argument. Using three operators like reproduction, crossover and mutation, GA is capable to select significant feature subset.

Bibliography

[1] I. D. E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning. Reading, MA: Addison-Wesley, 1989.

[2] B. K. Fukunaga and R. R. Hayes, "Effects of sample size in classifier design," IEEE Trans. Pattern Anal. Mach. Intell., vol. 11, pp. 873-885, Aug. 1989.

[3] D'haeseleer, P. "Context preserving crossover in genetic programming"' Proc. of the 1994 IEEE World Congress on Computational Intelligence, vol. 1, pages 256-261, Orlando, FL, USA. IEEE Press, 1994.

[4] [4]. Burke, E., Gustafson, S., and Kendall, G. Diversity in genetic programming: An analysis of measures and correlation with fitness. IEEE Transactions on Evolutionary Computation, 8(1): pp. 47-62, 2004.

[5] J. Yang and V. Hanovar, "Feature subset selection using genetic algorithm", Journal of IEEE Intelligent Systems, vol. 13, pp. 44-49, 1998.

[6] S. S. Sanz, G.C .Valls, F. P. Cruz, J. S. Sanchis, C. B. Calzn, "Enhancing Genetic Feature Selection Through Restricted Search and Walsh Analysis", IEEE Trans. on Systems, Man, and Cybernetics, Vol. 34, No. 4, November 2004.

[7] P. Leray and P. Gallinari, "Feature selection with neural networks," Behaviormetrika, vol. 26, Jan. 1999.

[8] B. Hassibi and D. G. Stork, "Second order derivatives for network pruning: optimal brain surgeon," in Advances in Neural information Processing Systems, S. J. Hanson, J. D. Cowan, and C. L. Giles, Eds. San Mateo, CA: Morgan Kaufmann, 1993, vol. 5, pp. 164-171.

[9] L. Breiman, J. Friedman, R. Olshen, and C. Stone, Classification and Regression Trees, 3rd ed. London, U.K.: Chapman & Hall, 1984.

[10] T. E. Campos, I. Bloch, and R. M. Cesar Jr., "Feature selection based on fuzzy distances between clusters: first results on simulated data," Lecture Notes in Computer Science, vol. 20, no.13, pp. 186, 2001.

[11] E. Nabil; A. Badr; I. Farag; "An Immuno-Genetic Hybrid Algorithm", International Journal of Computers, Communications & Control, vol. IV, no. 4, ISSN 1841 - 9836; E-ISSN 1841-9844, 2009.

[12] Adlassnig, K. P., "Fuzzy neural network learning model for image recognition." Integrated Computer-Aided Engineering, pp. 43-55, 1982.

[13] Kim, J.S. and H. S. Cho, "A fuzzy logic and neural network approach to boundary detection for noisy images." Fuzzy Sets and Systems, pp. 141-159, 1994.

[14] Jang, J.-S.R., C.-T. Sun, E. Mizutani, "Neuro-Fuzzy and Soft Computing, A Computational Approach to Learning and Machine Intelligent" Pearson Education.

[15] C. Muńoz, F. Vargas, J. Bustos, M. Curilem, S. Salvo ; H. Miranda; "Fuzzy Logic in Genetic Regulatory Network Models", International Journal of Computers, Communications & Control, vol. IV, no. 4, ISSN 1841 - 9836; E-ISSN 1841 - 9844, 2009.

Mahua Bhattacharya, an Associate Professor of Indian Institute of Information Technology & Management, Gwalior, India is working in the area of medical image analysis more than a decade in various fields of bio - medical applications like multimodal medical image fusion and registration, mammographic image analysis, classification of tumor / cancer lesion in CNS, computational techniques for study of neuro- degeneracy in brain, study of bone degeneracy and erosion. She had her B.Tech and M.Tech degree and from the institute of Radio Physics and Electronics, University of Calcutta. She worked as a research scientist at Indian Statistical Institute, Calcutta from 1995 till 2000 Calcutta and got her Ph.D degree in the area of Multimodal Medical Image Processing and Analysis Used Knowledge Based Approach in 2001 She was recipient of Frank George award for the paper - *Cybernetic Approach To Medical Technology : Application To Cancer Screening And Other Diagnostics'* WOSC - The World Organization Of Systems & Cybernetics, UK. She has published more than 70 papers in international journals and conference proceedings and as book chapters.

Arpita Das is an Assistant Professor of Institute of Radio Physics & Electronics, University of Calcutta, India. She received her B.Tech. and M.Tech. degree in Radio Physics and Electronics, University of Calcutta, in 2004 and 2006, respectively. Presently she is pursuing her Ph.D. on 'Some Studies on Medical Image Processing Methods And Their Implementation'. She was a senior research fellow under CSIR. Her research interests include image processing, pattern recognition, soft computing approaches for biomedical applications