Pattern recognition Using Genetic Algorithm

Majida Ali Abed , Ahmad Nasser Ismail and Zubadi Matiz Hazi

Abstract-- **The recognition processes is among the many intelligent activities of the human brain system . This paper is concerned with the Pattern recognition (isolated Arabic characters) using genetic algorithm to satisfy a successful recognition operation. The unknown character is read from a file and many operations will perform on it to manipulate it and extract its features, to compare these features with saved template's features The ratio of successful was over 95%. The proposed system has been implemented and tested on Delphi 6 environment.**

Index Terms— **Artificial Intelligence, Pattern Recognition, Genetic Algorithm, Delphi 6 environment.**

I. INTRODUCTION

Recognition is regarded as a basis attribute of human beings, as well as other living organisms. A pattern of an object. We are performing acts of recognition every instants of our life. We recognize the objects around us, and move and act in relation to them. We can recognize the voice of known person; read handwriting and analyze fingerprints and distinguish between a happy face and an angry face. A human being is a very sophisticated information system, partly because he/she possesses a superior pattern recognition capability. Many definitions of pattern recognition have been proposed, pattern recognition can be defined as a process, which leads to a decision. The quality of this decision can only be measured by statistic relating to the number of "good" and "bad" classifications. Also pattern recognition can be defined as an area of science concerned with discriminating objects on the basis of information available about them. Each distinct of information about objects is called a feature. The problem of pattern recognition may regarded as one of discriminating the input data, not between individual patterns but between populations, via the search for features or invariant attributes among members of a population[1]. Recognition methods can be roughly classified into three major groups: statistical, structural and syntactical ,and neural network methods. Sometimes different methods are combined for example simple methods are used for pre-classification and final decision is made with more sophisticated methods the recognition procedure is basically very simple: after preprocessing some features are extracted from the unknown character, which is then classified to the class whose members have the most similar features. The character recognition problem has been approached in many ways and various recognition methods have been suggested. Some of the methods have been especially developed for the character recognition but most of them are borrowed from other fields of pattern recognition,

Department of Computer Science. University of Tikrit , Iraq

signal processing, and image analysis. The recognition of machine-printed characters and handwritten a machine-printed characters and handwritten characters are the first practical application in the field of pattern recognition [2].character recognition system can contribute tremendously to the advancement of the automation process and can improve the interaction between man machine in many application , including office automation ,check verification and large variety of banking ,business and data entry applications.

The character recognition is often called "optical characters that are magnetically [3].

II. GENETIC ALGORITHMS

Genetic algorithms are a stochastic search algorithm, which uses probability to guide the search. It was first suggested by John Halland in the seventies. Over the last twenty years, it has been used to solve a wide range of search, optimization, and machine learning. Genetic algorithms are a class of parallel adaptive search algorithms based on the mechanics of natural selection and natural genetic system.

It can find the near global optimal solution in a large solution space quickly. It has been used extensively in many application areas, such as image processing, pattern recognition, feature selection, and machine learning[4].It is a powerful search technique that mimics natural selection and genetic operators. Its power comes from its ability to combine good pieces from different solutions and assemble them into a single super solution[5].genetic algorithms are initial population of solution called individuals is (randomly) generated, the solutions are evaluated. The algorithm creates new generations of population by genetic operations, such as reproduction, crossover and mutation. The next generation consists of the possible survivors (i.e. the best individuals of the previous generation) and of the new individuals obtained from the previous population by the genetic operations.

The best source of information about Gas is Holland's adaptation in natural and artificial systems, Holland uses terms borrowed from mendelian genetics to describe the process:each position in the string is called a gene. The possible values of each gene are called alleles.A particular string is called a genotype. The population of strings also called the gene pool. The organism or behavior pattern specified by a genotype is called a phenotype. If the organism represented is a function with one or more inputs, these inputs are called detectors[6].

III. THE SIMPLE GAS

The algorithm (pseudo code) of the simple GAs in Figure (1) illustrates the main steps that should be performed to

produce the required solution [7], [8]

```
Algorithm simple GAs:
Initialization [population];
Evaluation [population];
Generation:=0;
-do
Seected-parents:= selection[population];
Created-offspring:=recombination
   [selected-parents];
Mutation [created-offspring];
Population:=created-offspring;
Evaluation [population];
Generation:=geaeration+1;
UNTIL stop-criterion;
```

Fig. 1) the simple Gas

A. Population

A population consists of n individuals where N is chosen by the designer of the GAs. Every individual has a chromosome which consists of L genes. The chromosome is often referred to as the genotype of an individual. The following notation will be used to describe the chromosome of individual number i

Where $1 <= i <= N$:

Chromosome I = gene1 gene 2…….gene j…..gene L

When a binary encoding has been used every gene $1 <= J <= L \in \{0,1\}$.

As an example assume L=6, then a chromosome can look like:
Chromosome I =110101

B. Initialization

Initialize the genes of all individuals randomly with 0's and 1's(assuming a binary encoding for simplicity). These individuals are the starting points in the search space for the simple GAs

C. Evalaution

Calculate the fitness of each individual by decoding each chromosome and applying the fitness function to each decode individuals. The decoding creates a phenotype based on a genotype.

D. Selection

Select a specific individuals from the population to be the parents that will used to create new individuals there is many methods are used to choose those parents the most popular is the roulette wheel selection (RWS) which select the individuals with higher fitness with a higher probability("selection of the fitter individuals"). In this research we choose the parents that have specific fitness[9].

E. Recombination

Individuals from the set selected-parents are mated at random and each pair created offspring using 1-point crossover or 2-point crossover.

F. Mutation

Mutation is a random change of one or more genes. Every chromosome is simply scanned gene by gene and with a mutation rate Pm a gene is changed/swapped, i.e. 0 → 1 and 1 → 0 the probability for a mutation is usually kept small, i.e. $Pm = 1/L$ such that we can expect one mutated gene per chromosome[10].

G. Stop Criterion

A simple and easy to implement stopping criterion is to stop the simple GAs if no improvement of the best solution has been made for a (large) predefined number of generations, where one generation is one turn through the do-until loop in algorithm in Figure(1).

IV. PROPOSED MODEL

In this section we will describe the techniques we have adopted for the character recognition as shown in figure(2) the proposed model consist of two segments:

1) Process the character before recognize : and this is done through several procedures we can explain it as follows: a procedure takes the unknown letter, return it without empty rows and columns in an array b, then another procedure isolate the unknown character in array b from points if found, return the result in array c, the next procedure takes array c and extract the courner feature of the unknown character and save the result in four 3*3 arrays , the last procedure in this group translate the features of the character which stored in four 3*3 arrays into one vector called unn. Finally this vector is used in the search operations and calculate fitness.
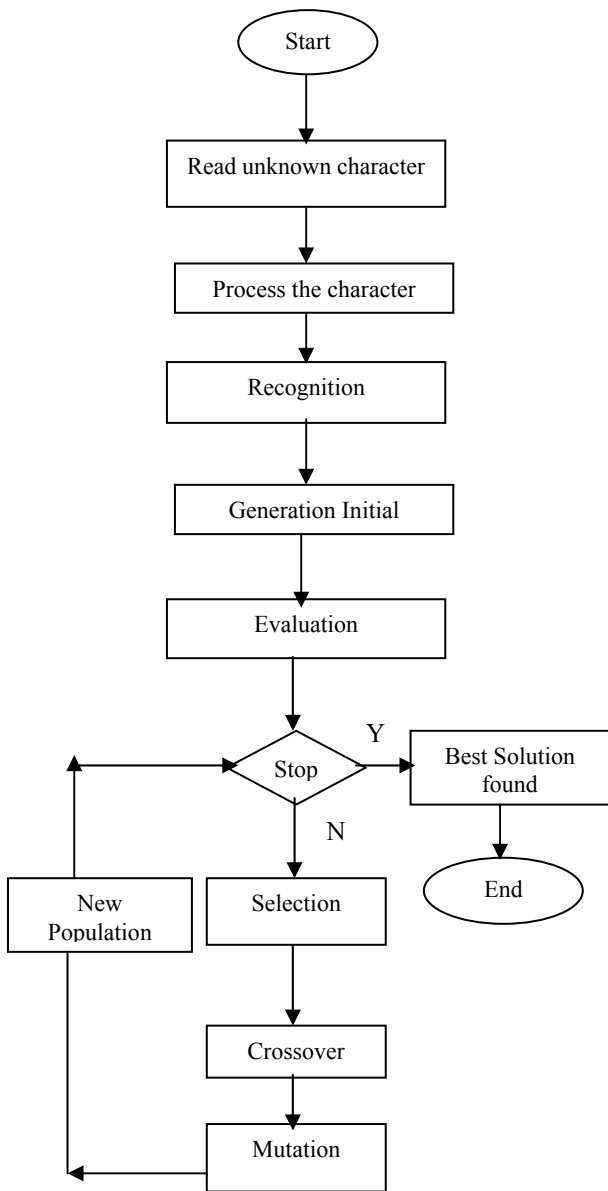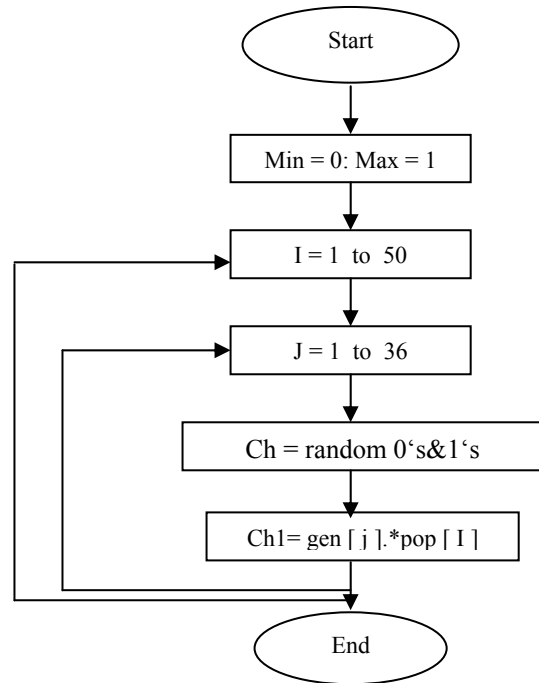
Fig. 1) Circle Life of the Genetic Algorithm

2) Recognition: it represent the genetic algorithm operations such as reproduction, crossover, mutation and replacement. This can be discussed as follows:
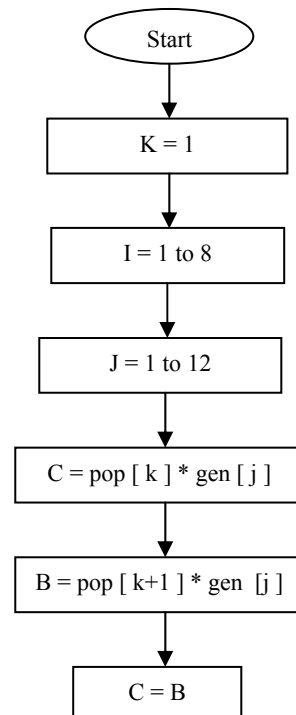
Initialize: create an initial population made of 50 individuals each one is length 36 gene randomly. And the algorithm goes as follow:
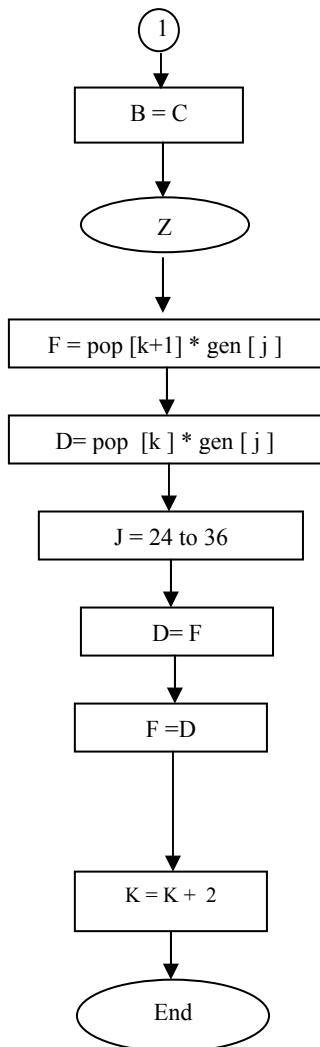

Evaluate: calculate the fitness of each individuals in the population according to the differences between them and the vector unn.
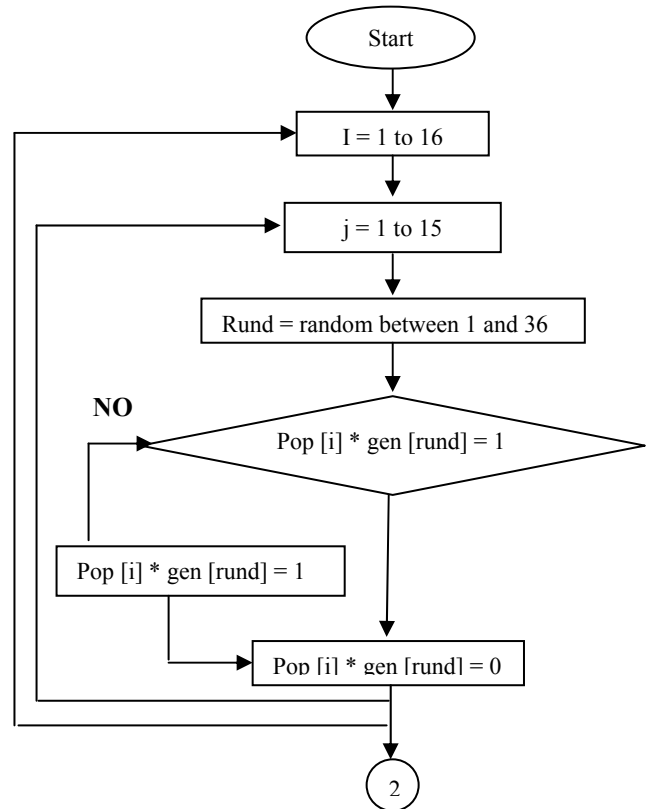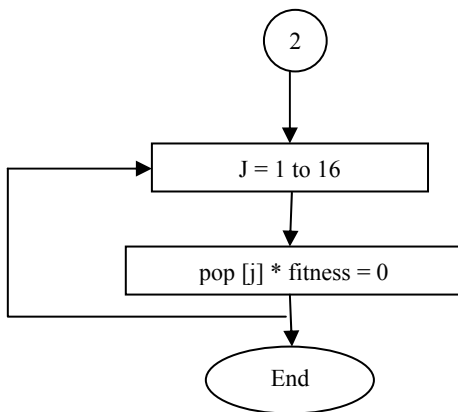
Choice parents : select 16 individuals that have fitness between 15-20 randomly from the population to be the parents.

Cross2x: perform 2-piont crossover (2x) on the parents to create new individuals, and the algorithm goes on:
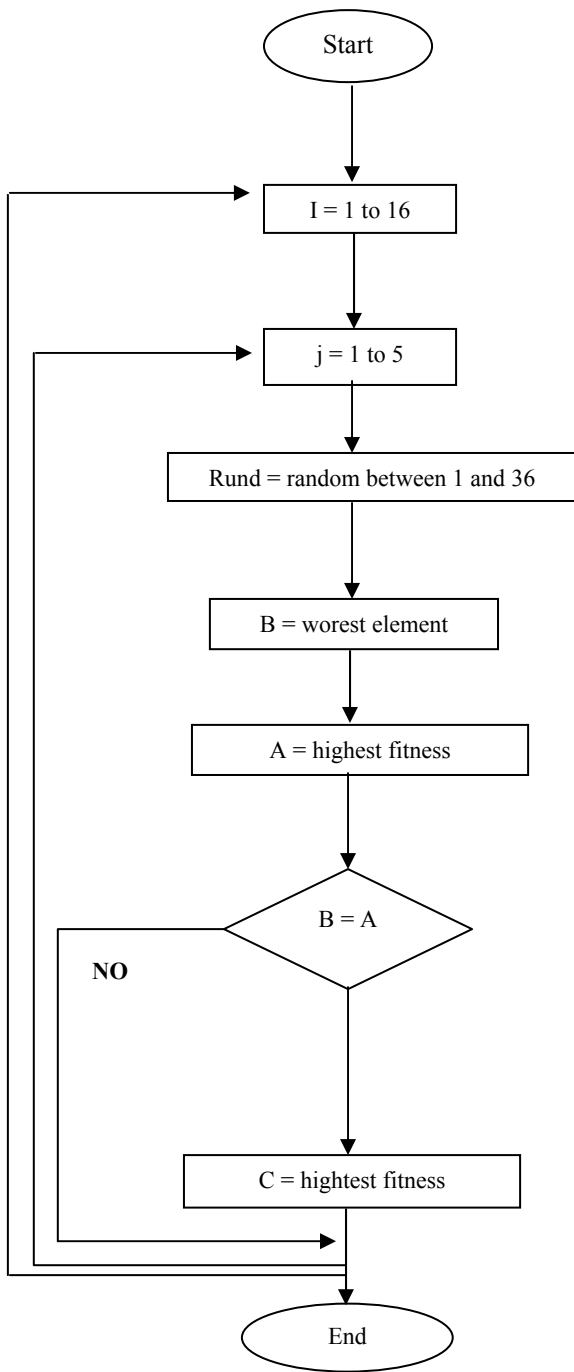
Mutation: it takes the new individuals created by cross2x procedure and randomly changes 15 genes from it, the algorithm goes on:

Replacement: it take five element randomly from the population, choose worst one and then replace it with a new individuals ,the algorithm goes on:

Start

↓

I = 1 to 16

↓

j = 1 to 5

↓

Rund = random between 1 and 36

↓

B = worest element

↓

A = highest fitness
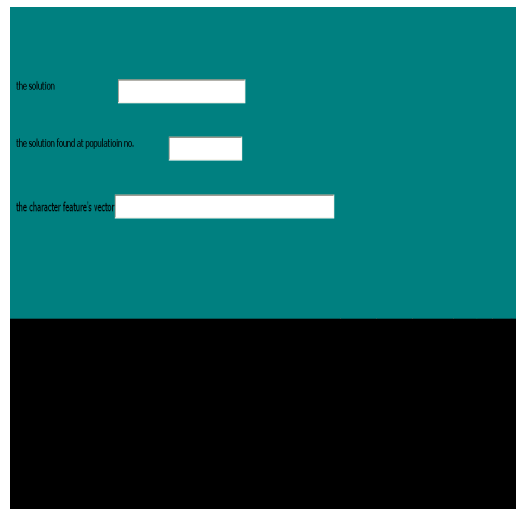
↓

B = A

NO

↓

C = hightest fitness

↓

End

Figure (3)

To load the unknown character we click on the load the letter bottom, a window will appear as shown in Figure(4), then we choose the file where the character stored as shown in fig
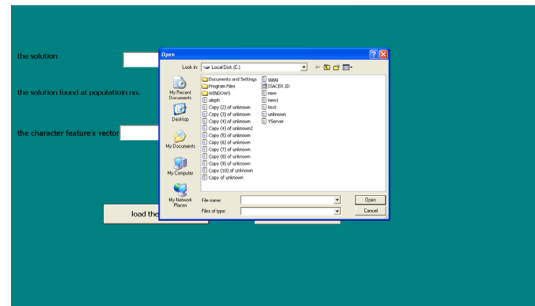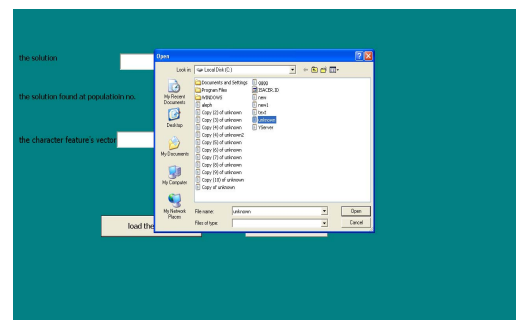

Figure (4)

Figure (5)

Checkstop : stop the system if best solution found which it's fitness should be less than 3 or the generation no. is more than 500.

V. RESULTS

The proposed system when we run it will be shown as in Figure(3) .

When we click on the open bottom the procedures will perform , the character is stored in 50 rows and 50 column. The features will extracted from it, the genetic algorithm operations will perform the result after recognition will be as follows in figure(5):
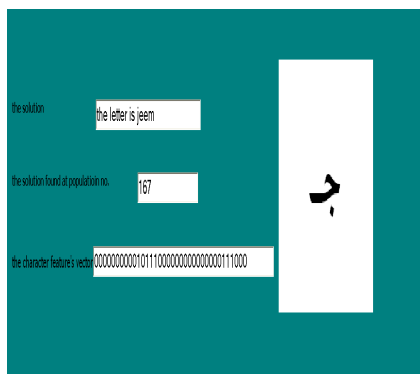
Figure (6)

Majida Ali Abed received his B.S. in mathematic science from Basrah university, 1978 , the M.S. in computer science from Basrah University in 1986 address of her thesis is "perceptron that can recognize and learn " . She is an assistant professor in the Department of Computer Science at the Tikrit University. She is the author of more than 30 publications in technical journals and conferences. She served on the several universities and computer centers in Iraq Her research interests are in the areas of Artificial intelligence (expert systems, pattern recognition, processing of natural languages),image processing.

The proposed system recognize the character jeem, show its picture, the solution found after no. of generation shown in the Figure(6) as the solution found at population no.) textbox, and the feature of the character represent in the (the character feature's vector) textbox, the speed of the proposed system is very high, when we enter five different sizes of letter for example jeem the system recognize the letters all as jeen and the ratio of recognition is over 95%.

VI. CONLUSIONS

Pattern recognition is one of the most important stages for any pattern recognition system. This paper proposed a method for recognizing Pattern (isolated Arabic characters). patterns recognition method, which is based on genetic algorithm, has presented. The method has been tested with a lot of patterns, high recognition rate has recorded and its recognition rate is over 95%. The proposed method is applied for (isolated form of Arabic characters), but it can be used for recognizing other three forms (beginning, middle and end form) with a little change in algorithm architecture.

REFERENCES

[1] Sturat, J .Russell, and Peter Norving, "Artificial Intelligence a modern approach",2nd Edition,Preentice Hall.2003.
[2] Koray Korkut , Bilat Alatas , "Mining Classification Rule by using Genetic Algorithms with non-random initial population and uniform operator ", Turk Jelec Engine,2004, Vol.12,pp:43- 52.
[3] Shyu, M. and Leou, J., "A Genetic Algorithm Approach To Color Image Enhancement", Pattern Recognition, Vol.33, No.7, PP.871-880, 1998.
[4] Jasradj U. Dange, "Introduction to Genetic Algorithms" 2001.
[5] Phili Kohn, "Combing Genetic Algorithm and Neural Networks" M.Sc. Thesis,University of Tennessee,1994.
[6] Rasheed, Sh. A., "Genetic Algorithms Application in Pattern Recognition", Master's thesis, National Computer Center Higher Education Institute ,2000.
[7] Roger L. Wainwright, "Introduction to Genetic Algorithms Theory and Applications",Addison-Wesley publishing.1993.
[8] Schamidt, M. and Stidsen, T., "Hybrid Systems: Genetic Algorithms, Neural Networks and Fuzzy Logic", Aarhus University.
[9] Franti, P., Kivjarvi, J., Kaukoranta, T. and Nevalainen, O., "Genetic Algorithm for Large- Scale Clustering Problems", The Computer Journal, 1997 Vol.40, No.9.
[10] Mauldin, M. L., "Maintaining Diversity in Genetic Search", The National Conference on Artificial Intelligence (AAAI- 84), August 1984.