A New Approach for Pattern Recognition with Neuro-Genetic System Using Microbial Genetic Algorithm

Tanvir Ahmad Tarique, Muhammad Ahsan Zamee, Md. Imran Khan
Electrical and Electronic Engineering
World University of Bangladesh
Dhaka, Bangladesh
tanvir.tarique5@gmail.com, zamee_mist@yahoo.com, imrankhankuet@gmail.com

***Abstract*—Artificial Intelligence (AI) has been used extensively to solve different types of pattern recognition problems such as disease diagnosis and classification problems. Due to poor performance in recognition of patterns, several methods have been used. Among them genetic algorithm (GA) have shown better performance than other algorithms. Microbial Genetic Algorithm (MGA) or Bacterial Genetic Algorithm is one of the newer branch of GA. MGA follows the evaluation procedure of microbial which gives better results in pattern recognition. In this paper, microbial genetic algorithm with neural network approach for fitness calculation has been developed and it is used for performance analysis of different pattern recognition problems. Proposed algorithm is named as Microbial Neuro-Genetic Algorithm (MNGA). Advantages of MNGA over simple genetic algorithm (SGA) have also been discussed. XOR, Breast Cancer, Diabetes, Heart Diseases, Glass and Card classification problems are taken from UCI machine learning repository dataset as sample problems for performance analysis which shows that this method provides good performance for different types of problems and thus reduces the need for different types of methods for different types of problems.**

Keywords- MGA, SGA, AI, Pattern recognition, UCI machine learning repository dataset, Bacterial Genetic Algorithm

I. INTRODUCTION

This is the era of computers & automation. Now-a-days there are too many diseases and many people are dying for the lack of proper treatment. And also the complete automation for the comfort in human life is very much desired. This problem can be solved by using such machines which can do everything what we want in a very short period of time. For this reason, these machines have to be a robot or something like that in which there is some built-in programs installed and by using them these robots or machines can perform almost everything, certainly for which that the robots or machines are designed.

Some machines, like EEG (Electro Encephalo Gram) for brain scan, ultra-sono gram, X-ray & others for the detection of different kinds of diseases are used now-a-days. And for that reason, different types of machines along with different types of algorithms or methods are required. Also these have no provisions for telling the patients automatically that they are suffering from any diseases or not. A patient have to do this experiments or check-ups and then go to a doctor for assuring that what is happened, is there any disease or not. For the sake of automation, some methods are need to be built up which results in some programs which will be installed in some devices (in future) & tell readily after the check-up that whether a patient is suffering from any diseases or not. For these reasons, different types of patterns with their datasets such as XOR with 2-inputs, 3-inputs, 4-inputs, Cancer, Diabetes, Heart Diseases, Glass and Card problems are analyzed. W. H. Wolberg and O. L. Mangasarian have used Multisurface Method of Pattern Separation for breast cancer diagnosis. [3]. Bamrungsap S, Chen T, Shukoor MI, Chen Z, Sefah K, Chen Y and Tan W have used Pattern recognition of cancer cells using aptamer-conjugated magnetic nanoparticles [4]. Jalil Addeh and Ata Ebrahimzadeh have used Novel Hybrid Intelligent Method for breast cancer diagnosis [5]. Pinaki Ray, Greg Matian, Aparna Srinivasan, David Rodbard and David Price have used Systems and Methods for Pattern Recognition in Diabetes Management [6]. Ian W. Evett and Ernest J. Spiehler have shown the methods for Pattern Recognition in glass type determination in Rule Induction in Forensic Science [7]. It is seen from the previous works that different types of problems are analyzed with different types of algorithms with different types of controllers. But this paper represents that all of the above mentioned pattern recognition problems are analyzed only with MNGA and SGA and their results are compared. And MNGA provides better result over SGA.

This paper is structured as follows. Section II part A describes simple genetic algorithm, part B describes Microbial genetic algorithm, part C shows Microbial Neuro-genetic algorithm and part D shows the pseudocode of Microbial neuro-genetic algorithm. Section III shows simulation with Microbial neuro-genetic algorithm approach, Section IV is for results and comparative analysis, and finally Section V is for conclusion.

II. GENETIC ALGORITHM APPROACH

A. Simple Genetic Algorithm (SGA):

A genetic algorithm is a search technique used in computing, to find true or approximate solutions to optimization and search problems, and is often abbreviated as

GA. Genetic algorithms are categorized as global search heuristics. Genetic algorithms are a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover (also called recombination).

Genetic algorithms are implemented as a computer simulation in which a population of abstract representations (called chromosomes or the genotype or the genome) of candidate solutions (called individuals, creatures, or phenotypes) to an optimization problem evolves towards better solutions [1].

Simple Genetic Algorithm starts with a randomized initial population. For the reason that initial population are randomly assigned, the actual outputs differ much from the target outputs. So error is initially high. As the fitness function is represented by equation 1, there is a negative relationship between error and fitness function, so the initial fitness function is low. The relation between error and fitness function is given below:

$$\text{Fitness function, } f = 1 / (\text{Offset} + \text{Error}) \quad \ldots\ldots\ldots (1)$$

Where, $\text{Error} = (0.5/N) \times \Sigma (\text{Target output} - \text{Actual output})^2$
& N = total number of patterns to be recognized

There are many types of genetic algorithm available. One of them is Microbial Genetic Algorithm (MGA). In this paper MGA is applied with the neural network which leads to the creation of MNGA. Both MGA and MNGA are described here.

B. Microbial Genetic Algorithm (MGA):

There are many varieties of Evolutionary Algorithms, many different ways to implement, for problem solving, the three main requirements of Heredity, Variation and Selection. Here Microbial Genetic Algorithm will be described which little is known but effective method that is so simple to implement that the core of the program can be reduced to a single line of code. It is called Microbial Genetic Algorithm because it is loosely based on the way microbes can exchange genetic material, DNA, 'horizontally' between different living members of the population as an alternative to 'vertically' from one generation to the following one [2].

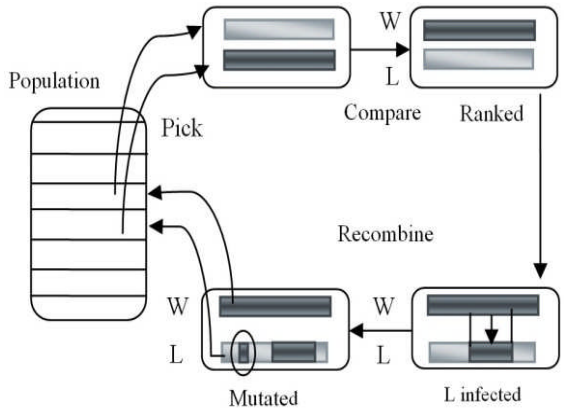

Figure 1. The basic Microbial tournament [2]

This Microbial GA (MGA) obeys the 3 rules such as Heredity, Variation and Selection and yet is so simple that it can be reduced to a single line of code. It is very much effective. [2]

C. Microbial Neuro Genetic Algorithm (MNGA):

MNGA is a modified version of MGA. Where, two random strings are selected based on their fitness. The string with better fitness will replace their certain percentage of string bits with the remaining one. This process is called crossover. The crossover probability is varied from 0.05 to 25 %. After the crossover is completed a certain percentage of bits will be altered. This alteration of bits is called mutation. The mutation probability is varied from 0.05 to 1 %. After completion of mutation a new generation is found whose fitness have to be measured again. If the fitness reaches the desired value or error rate is within the specified limit then the process has to be stopped.

The fitness is evaluated using neural network with a single input, single hidden and single output layer which makes the calculation less complex but gives better result than other algorithms. Sigmoid function is used as transfer function. After achieving the desired fitness, testing is done by finding the Testing Error Rate (TER) for each population strings. Equation of TER is given later.

The major difference in the proposed algorithm with other algorithms is it uses artificial neural network's back-propagation theory for evaluating the fitness of the population which is much easier to implement, but gives better result than the other algorithms.

D. Pseudocode of proposed Microbial Neuro-Genetic Algorithm (MNGA):

Microbial Neuro-Genetic Algorithm (MNGA) has been proposed in this paper. The pseudocode of MNGA is given bellow:

Steps for Training:
Step 1: Initialization of population.
Step 2: Fitness calculation of each individual using neural network.
Step 3: Selection of winner and looser based on their fitness after choosing two individuals randomly.
Step 4: Performing crossover of looser only from winner.
Step 5: Performing mutation of looser only.
Step 6: Checking the ending criteria using individual fitness and average fitness.
Step 7: Stop if it is converged else repeat from Step 2 with new population.
Step 8: Saving the weights values of each population strings.

Steps for Testing:
Step 1: Loading the weights values of each individual.
Step 2: Finding the output values using the weights with neural network.
Step 3: Checking the TER for each individual by comparing the actual output with outputs from neural network.
Step 4: Finding the Average TER.

III. SIMULATIONS WITH MNGA APPROACH

Training datasets for several diseases are collected from UCI machine learning repository [8] and then used for pattern recognition with MNGA approach. For each case average fitness curve is shown against generation number. From these curve convergence rate can be easily determined.

a. XOR Gate with 2 Inputs:

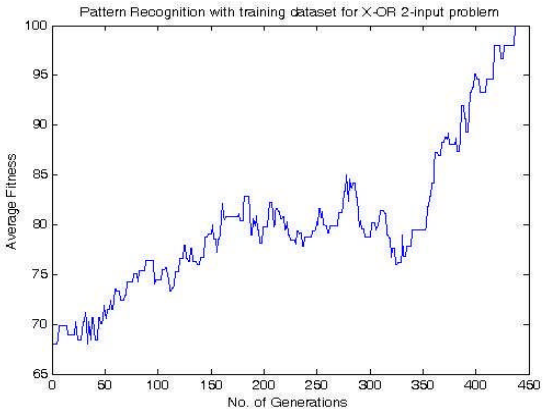

Figure 2. Average fitness vs. Generations curve for Pattern Recognition with training dataset for X-OR 2-input problem with MNGA

b. XOR Gate with 3 Inputs:

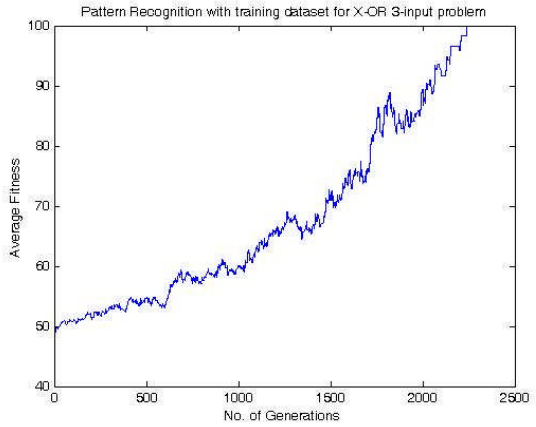

Figure 3. Average fitness vs. Generations curve for Pattern Recognition with training dataset for X-OR 3-input problem with MNGA

c. XOR Gate with 4 Inputs:

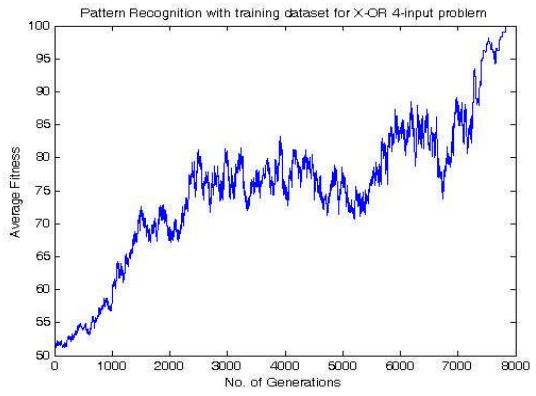

Figure 4. Average fitness vs. Generations curve for Pattern Recognition with training dataset for X-OR 4-input problem with MNGA

d. Cancer Disease problem:

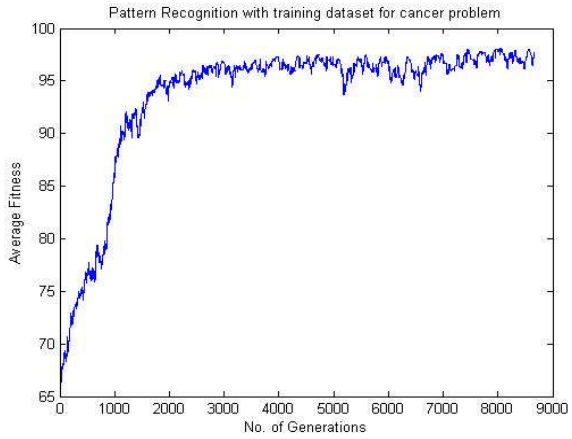

Figure 5. Average fitness vs. Generations curve for Pattern Recognition with training dataset for Cancer disease problem with MNGA

e. Diabetes problem:

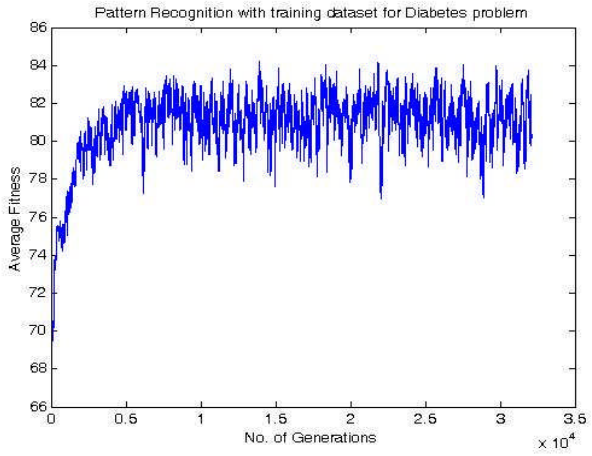

Figure 6. Average fitness vs. Generations curve for Pattern Recognition with training dataset for Diabetes problem with MNGA

f. Heart Disease problem:

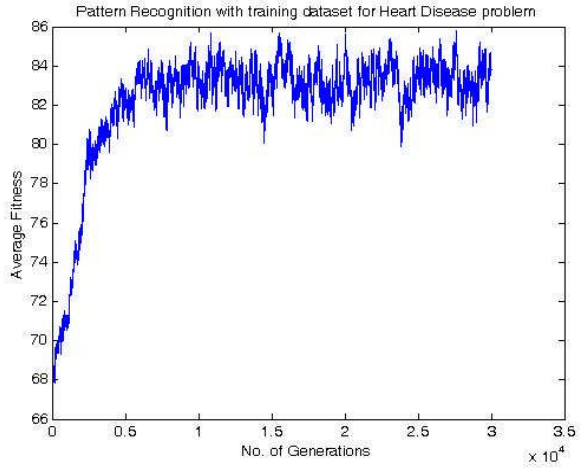

Figure 7. Average fitness vs. Generations curve for Pattern Recognition with training dataset for Heart disease problem with MNGA

g. Glass problem:

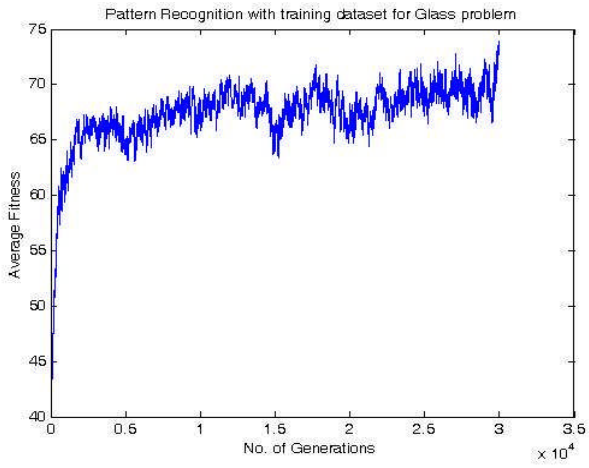

Figure 8. Average fitness vs. Generations curve for Pattern Recognition with training dataset for Glass problem with MNGA

h. Card problem:

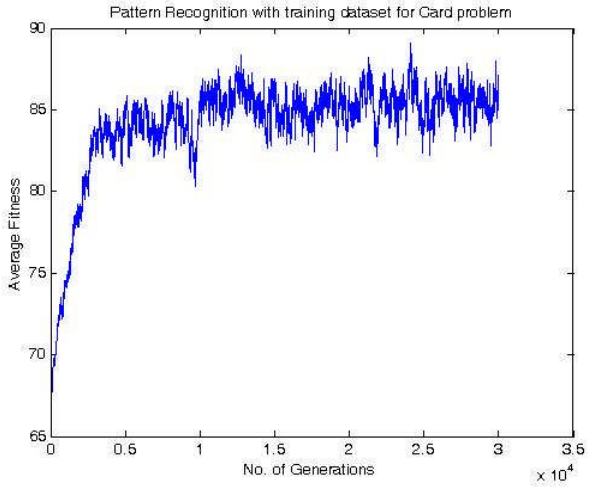

Figure 9. Average fitness vs. Generations curve for Pattern Recognition with training dataset for Card problem with MNGA

IV. RESULT AND COMPARATIVE ANALYSIS

Testing is done by finding the Testing Error Rate (TER) for each population strings. TER is defined as follows:

$$TER = (Tm / Tp) \times 100\ \%$$

Where, Tm = Total number of Misclassification
Tp = Total number of testing pat terns

TER is used to determine how much error is occurred after training the classification problems. Table I shows the no. of inputs and outputs and also the amount of dataset. Table II shows the comparison of performance in between SGA and MNGA.

TABLE I. NO. OF INPUTS/OUTPUTS AND DATASETS FOR EACH CASE

Problems	No. of Inputs in Input Layer	No. of Inputs in Hidden Layer	No. of Outputs in Output Layer	Datasets for Training, Validation & Testing
2 input XOR	2	3	2	4,4
3 input XOR	3	4	2	8,8
4 input XOR	4	5	2	16,16
Cancer	9	3	2	350,175,174
Diabetes	8	3	2	384,192,192
Heart Diseases	35	3	2	460,230,230
Glass	9	7	6	107,54,54
Card	51	3	2	345,173,172

TABLE II. NO. OF ITERATIONS REQUIRED AND TER FOR EACH CASE

Problems	Iterations (MNGA)	Iterations (SGA)	%TER (MNGA)	%TER (SGA)
2 input XOR	450	170	0	0
3 input XOR	2,250	290	0	0
4 input XOR	8,000	650	0	0
Cancer	8,900	1,000	2.42	4.62
Diabetes	30,000	2,500	26.042	30.89
Heart Disease	30,000	2,500	19.565	24.542
Glass	30,000	3,500	30.189	36.258
Card	30,000	2,500	19.767	25.321

V. CONCLUSION

Sometimes it becomes difficult for designers to select the appropriate pattern recognition technique due to diversity of algorithms based on problem pattern. This newly proposed and applied MNGA shows that only this can be used for any kind of pattern recognition problems. It has better flexibility to different types of problem for pattern recognization which will certainly reduce the time for researcher and designers to solve any problem.

REFERENCES

[1] David E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", University of Alabama, 1989.

[2] Inman Harvey, "The Microbial Genetic Algorithm" School of Cognitive and Computing Sciences, University of Sussex, Brighton BN1 9QH, UK, 1996.

[3] W H Wolberg, and O L Mangasarian, "Multisurface method of pattern separation for medical diagnosis applied to breast cytology", Proceedings of the National Academy of Sciences, December 1, 1990 vol. 87 no. 23.

[4] Bamrungsap S, Chen T, Shukoor MI, Chen Z, Sefah K, Chen Y, and Tan W, "Pattern recognition of cancer cells using aptamer-conjugated magnetic nanoparticles", US National Library of Medicine, National Institutes of Health, 2012 May 22.

[5] Jalil Addeh, and Ata Ebrahimzadeh, "Breast Cancer Recognition Using a Novel Hybrid Intelligent Method", Journal of Medical Signals and Sensors, Apr-Jun 2012.

[6] Pinaki Ray, Greg Matian, Aparna Srinivasan, David Rodbard, and David Price have used "Systems and Methods for Pattern Recognition in Diabetes Management", 2008-09-25.

[7] Ian W. Evett and Ernest J. Spiehler., "Rule Induction in Forensic Science", Central Research Establishment, Home Office Forensic Science Service, Aldermaston, Reading, Berkshire RG7 4PN.

[8] Bache, K. & Lichman, M. (2013). *"UCI Machine Learning Repository"*.[Online].Available: http://archive.ics.uci.edu/ml/datasets.html [Accessed: Mar. 27, 2014].