

**The Journal of
Mathematics and Computer Science**

Available online at

<http://www.TJMCS.com>

The Journal of Mathematics and Computer Science Vol. 4 No.2 (2012) 237 - 245

An Intelligent Diagnostic System for Detection of Hepatitis using Multi-Layer Perceptron and Colonial Competitive Algorithm

**Khosro Rezaee^{*1}, Mohammad Rasegh Ghezelbash², Nasim Chagha Ghasemi³
and Javad Haddania⁴**

¹Department of Biomedical Engineering, Hakim Sabzevari University, Sabzevar, Iran,
Rezaeekhosro@ymail.com

²Department of Biomedical Engineering, Hakim Sabzevari University, Sabzevar, Iran,
M.rasegh@gmail.com

³Department of Electronic Engineering, Islamic Azad University,
South Tehran Branch, Tehran, Iran,
Nasim.ghasemi87@gmail.com

⁴Department of Biomedical Engineering, Hakim Sabzevari University, Sabzevar, Iran,
Haddadnia@sttu.ac.ir

Received: January 2012, Revised: April 2012

Online Publication: June 2012

ABSTRACT

This article proposes an intelligent diagnostic system for the diagnosis of Hepatitis based on a new algorithm including MLP and ICA which is more accurate and faster than the similar algorithms in terms of performance. At first, colonial competitive algorithm seeks to find the best solution in neural network training, then the MLP will be designed which can intelligently diagnose Hepatitis. Providing a certain solution and the ability to analyze complex, large-scale problems are among the advantages of this algorithm over similar optimization algorithm in diagnosis of Hepatitis. For neural network training and sample data testing 100 and 55 sample data were used respectively. Taken from UCI database, the data were applied to the system, revealing the effectiveness of the proposed algorithm in diagnosis of Hepatitis with less than 5% error.

Keywords: Hepatitis diagnosis, ICA, MLP, GA.

* Corresponding Author

INTRODUCTION

Hepatitis is the inflammation of the liver which is commonly known as jaundice [1]. Most people suffering from Hepatitis type B and C do not have any symptoms [2]. When a person becomes infected with Hepatitis his/her liver becomes inflamed. It destroys the normal tissue and in the end, only the fibrous worn-out tissues remain [3]. At the present, there are about 2 million people infected with Hepatitis B virus, and 200,000 to 300,000 people infected with Hepatitis C [4]. Various intelligent methods have been proposed for the diagnosis of Hepatitis. The first intelligent method was proposed by Eason et al. [5]. It used Linear Detection Analysis (LDA) and ANFTS to detect Hepatitis by drawing on the properties and categories of data with 94.16% accuracy. In 2009, Rouhani et al. [6] by using Support Vector Machine (SVM) and neural network, proposed a method for the diagnosis of Hepatitis. In addition to the diagnosis of the disease, this system detected its type and stage of development with 97% accuracy. Sathyadevi [7] created a new algorithm based on CART, C4.5 and ID3 algorithms for diagnosis of Hepatitis by comparing their results with each other. Dogantekin et al. [8] proposed a counseling system for quick diagnosis of Hepatitis method called DIAGFH. In this method, the disease information were organized in a tree structure divided into two branches, namely positive or negative Hepatitis, which aided the diagnosis of the disease. Jilani et al. [9] proposed a system based on neural network and PCA for detection of Hepatitis which comprised of two stages: first, qualities were extracted using PCA techniques, and then categorization was made using neural network. The accuracy of this system was 99.13% and 100% for training and testing of the data respectively. Uttreshwar et al. [10] proposed an automatic system for Hepatitis B diagnosis using Fuzzy deduction and generalized regression neural network. It comprised of disease detection using an intelligent system along with the prediction of its development stage by neural network.

MATERIALS AND METHODS

Hepatitis data were taken from reference database [11] that included 155 samples with 19 inputs (features) and one output. The output indicates whether the person infected with Hepatitis is dead or alive. There are 6 discrete and 13 binary features. Of the total 155 subjects, 32 died of Hepatitis. The existing features used in selection, training and testing of data have been presented in Table 1 and Table 2.

Table1. The process of data selection, training and testing

S. No	Variable	Description	S. No	Variable	Description
1	Sex	Male, female	11	Spiders	No, Yes
2	Age	10, 20, 30, 40, 50, 60, 70, 80	12	Ascites	No, Yes
3	Steroid	No, Yes	13	Varices	No, Yes
4	Antiviral	No, Yes	14	Bilirubin	0.39, 0.80, 1.20, 2.0, 3.0, 4.0
5	Fatigue	No, Yes	15	Alk Phosphat e	33, 80, 120, 160, 200, 250

6	Malaise	No, Yes	16	SGOT	13, 100, 200, 300, 400, 500
7	Anorexia	No, Yes	17	Albumin	2.1, 3.0, 3.8, 4.5, 5.0, 6.0
8	Liver Big	No, Yes	18	Protime	10, 20, ... , 90
9	Liver Firm	No, Yes	19	Histology	No, Yes
10	Pleer Palpable	No, Yes	20	Class	Die, Live

Table 2: class statistics, number of data, training data, testing data

Class	Number of Data	Training Data	Testing Data
Live	123	80	43
Die	32	20	12

COLONICAL COPETITVE ALGORITHM

Colonial competitive algorithm is a method used in the field of evolutionary computation which helps find optimal answer to various optimization problems. This algorithm, drawing on the mathematical modeling of social-political evolution, offers a solution for optimization problems in mathematics [13]. To make an algorithm, first an array of optimal variables is formed. The array is known as the “Country” Figure 1 presents the “country” array which has been formed based on categorized factors.

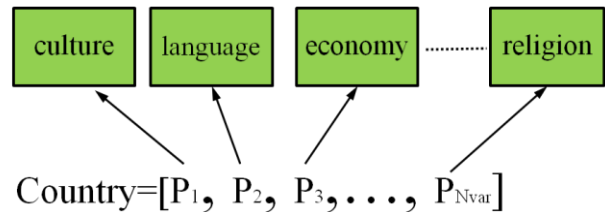


Fig 1: “Country” array formed based on such factors as culture, language, economic and etc.

In solving an optimization problem using the proposed algorithm, we look for the best country (a country with the best social-political features). In doing so, we, in fact, find the best parameters of the problem which generate the lowest value of cost function. The cost of a country is calculated using equation (1):

$$cost_i = f(country_i) = f(P_1, P_2, P_3, \dots, P_{N_{var}}) \tag{1}$$

In an attempt to distribute N original colonies (the weak optimization) among imperialists (the strong optimization) proportional to their power, some colonies are assigned to them. To do this, with the estimated cost of all imperialists at hand, their normalized cost is defined according to equation (2):

$$C_n = \max_i \{c_i\} \tag{2}$$

Where C_n is the n imperialist cost, $\max_i \{C_i\}$ is the maximum cost of all imperialists and C_n is the normalized cost of the target imperialist. The imperialist with the higher cost would be basically the weaker imperialist, having the lower normalized cost. With the normalized cost available, the relatively normalized power of each imperialist can be calculated according to equation (3). Then, based on the result of this equation, the colonies are distributed among the imperialists.

$$P_n = \left| C_n \times \left(\sum_{i=1}^{N_{imp}} C_i \right)^{-1} \right| \quad (3)$$

From another perspective, the normalized power of an imperialist is the ratio of colonies run by that imperialist. Therefore, the number of colonies of an imperialist is calculated according to equation (4).

$$N.C_n = \text{round} \{ P_n . (N_{col}) \} \quad (4)$$

In which $N.C_n$, is the original the number of colonies of an empire and N_{col} is the total number of colonies in the population of the original country. If we display the distance between the colonizer and the colonies with d , equation (5) shows how the colony moves toward the colonizer:

$$x \sim U(0, \beta \times d) \quad (5)$$

Where $2 \geq \beta > 1$ and if $\beta > 1$, the colony would move toward the colonizer from different directions. In moving toward the colonizer, the colony may find a better situation than the colonizer. In this algorithm each empire needs to increase the number of its colonies; otherwise, it will lose its power in imperialistic competition, conquered by the stronger one. In this competition model, the total cost of emperor is calculated according to equation (6).

$$NTC_n = \max_i \{ TC_i \} - TC_n \quad (6)$$

In which $T \times C_n$ is the total cost of the n emperor and $N \times T \times C_n$ is the normalized cost of the emperor. The empire with the lowest cost will be the most powerful. With the normalized total cost at hand, the possibility of possessing the competitive colony by each empire is equal to equation (7):

$$P_{P_n} = \left| (N \times T \times C_n) \times \left(\sum_{i=1}^{N_{imp}} N \times T \times C_i \right)^{-1} \right| \quad (6)$$

As this process continues, the weak emperors are eliminated until the last one and the algorithm becomes convergent.

NEURAL NETWORK

Artificial Neural Networks due to their ability to model complex and nonlinear systems are very common in different branches of engineering. Layers serve as the link of neurons in neural networks, the weight of which is considered as the degree of neurons impact on each other with a constant added as the bias. In 1943, McClouth and Pitts proposed the simplified mathematical models of biological neurons [13]. Figure 3 displays how the weight and bias influence the output of single input neuron with activation.

A. MULTI-LAYER PERCEPTRON NEURAL NETWORK (MLP)

Multilayer Perceptron Neural Networks are the most common neural networks which have the most application among different types of neural structures. Leaving its impact, the output of each layer becomes the stimulus input function of the layer. The input of the desired network contains nineteen numbers. The best result is yielded when 9 layers are used. Consequently, 9 layers have been selected as the middle layer with only one output which indicates whether a person is infected with hepatitis or not. The layer structure of the proposed neural network is shown in Figure 2.

B. THE LEARNING OF THE NEURAL NETWORK BY COLONICAL COMPETITIVE ALGORITHM

Learning is a process in which the free parameters of a neural network adapt themselves with the network environment through applying a stimulation process. In fact, in neural networks, learning is the adjustment of weight values, bias and the slope of activation function. To measure the accuracy of the network, the measuring error between the actual output and network output is calculated in progressive neural network. If the network errors decrease, therefore, the accuracy of the network increases. This is a result of back-propagation method. In this algorithm, the propagation channel of network output error is from output layer to the hidden layer, adjusting the synaptic weights and other network parameters. Given the similarity of neural network learning and optimization problems, the colonial competitive algorithm has been used for teaching MLP neural network.

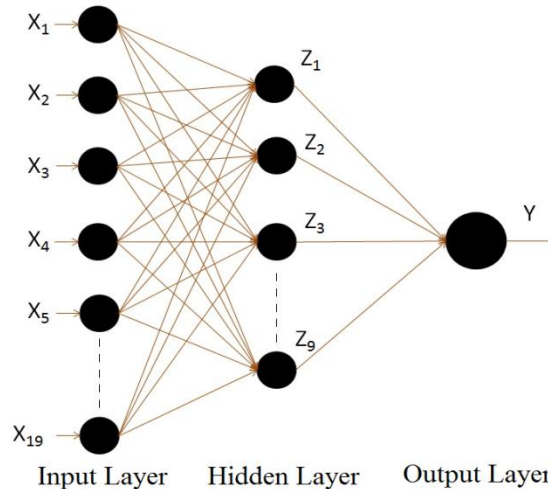


Fig 2: The proposed neural network model: input layer, hidden layer and output layer.

To learn such a neural network, 19 neurons in the first layer, 9 neurons in the middle layer and 1 neuron in the output layer has been used. The original weight and network bias in colonial competitive algorithm is the original population of the country. The network error is a function of the cost in each country $cost_i$. Table 3 displays the related parameters and rates of each parameter. Given the intense complexity of parameters in neural network learning, the presence of evolutionary algorithms requires high accuracy.

Table (3): Parameter and rates of each parameter in colonial competitive algorithm

Parameters	<i>Countries</i>	<i>Imperials</i>	<i>Decades</i>	<i>Revolution</i>
Rates	200	10	350	0.3

To demonstrate the accuracy of colonial competitive algorithm, a comparison of it has been made with genetic algorithm to minimize 12-dimensions Ackley rastrigin function. Despite high calculations, genetic algorithms had a slow pace in finding the optimal solution, the only difference being the rate of convergence. The important point, however, is the certainty in finding solution which genetic algorithm, having trapped in local optimization, has failed to find the desired optimization unlike colonial competitive algorithm which has been effective in finding the optimization solution.

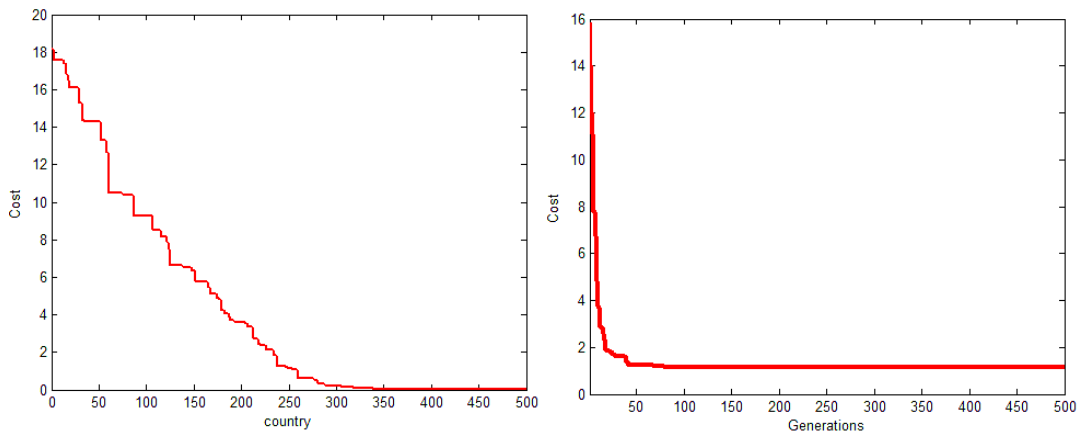


Fig 3: The right diagram shows the convergence of genetic algorithm in finding the optimization solution in Ackley rastrigin or 12-dimensions function and the left diagram shows the convergence of colonial competitive algorithm in finding solution in the same Ackley rastrigin.

As Figure 3 shows, the colonial competitive algorithm has almost reached its minimum value, i.e. zero (optimal point) in 340th generation while genetic algorithms have converged in 150th generations, reaching the minimum value of 1.15. Table 4 displays the algorithm and its entire implementation process.

Table (3): The stages of the proposed algorithm

The algorithm of the proposed system

1. Start

- The original data related to Hepatitis are loaded.
- The data are normalized according to cost function and MLP neural network input
- The data related to network training and testing are separated.

2. The structure of MLP neural network

- The network training using 100 types of data out of 155 total sample data.
- The evaluation of network training which is required in the learning of that colonial algorithm

2.1 Function cost

2.1.1 Selecting the unique features of the country array

2.1.2 Substituting the selected features of country array as original weights of neural networks and selecting the best colonial as the best weight vector.

2.1.3 Putting weights in the target places within network.

3. Output

3.1 Calculating the error induced by applying the final cost data and its comparison with the actual output.

3.2 Analyzing the neural network output after training and testing of the desired data.

3.3 Announcing the positive or negative Hepatitis.

End

THE PRACTICAL RESULTS

To prove the effectiveness of the proposed algorithm, 155 sample data taken from database were applied to the target network, 100 samples for the purpose of network training and 55 for network of testing. In Table V the results of the proposed algorithm along with similar methods used in the diagnosis of Hepatitis [6][8][9][14] have been compared . Figure 4 (Left side) displays the system training using 100 samples from a total 155 samples. Figure 4 (Right side) also shows the test data using the remaining 55 samples. The results of the comparison indicate the high accuracy of the proposed algorithms in the diagnosis of Hepatitis. The 0.0163, 0.0454, 98.38% and 95.46% values presents the mean square of error for the data used in training and testing system respectively as well as the ultimate accuracy of the program in data training and testing.

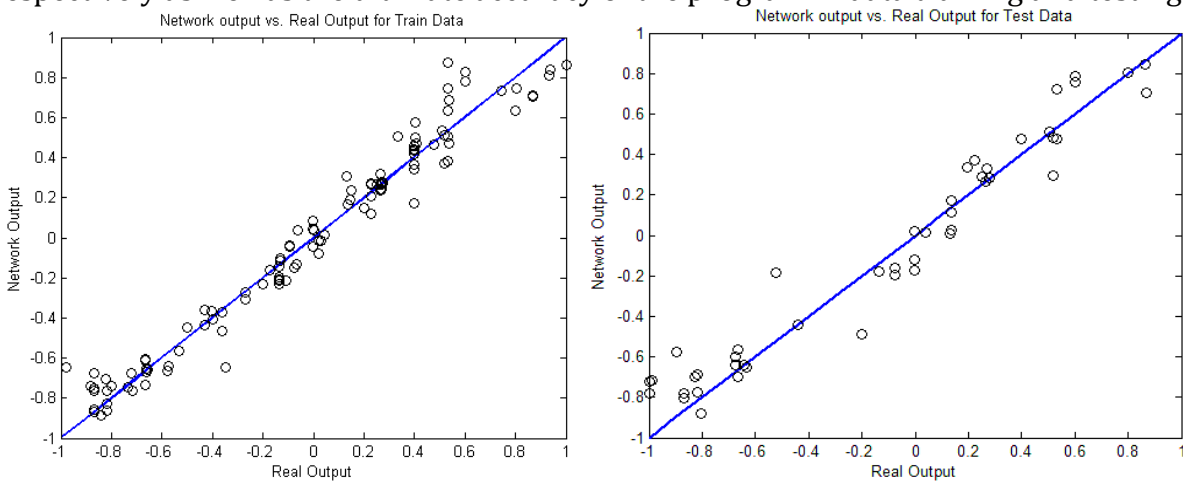


Fig 4: Training of the proposed system using 100 samples (Left side) and Testing of the proposed system using 55 samples (Right side)

Table (3): Review of the previous works and their comparison with the proposed algorithm

<i>Method</i>	<i>Accuracy</i>
MLP+BP	82.1%
PCA-AIRS	94.12%
GA-MLP	93.55%
FS-Fuzzy-AIRS	94.12%
CART	82.7%
IncNet	86.00%
SVM-NN	97.00%
LDA-ANFIS	94.16%
Proposed Algorithm	95.46%

CONCLUSION

This article attempted to propose an intelligent system with high accuracy and speed in diagnosis of Hepatitis by drawing on MLP and colonial competitive algorithm. The results reveal the high effectiveness of the system and its capacity to properly response to multi-dimensional complex problems. Certainty in obtaining the correct solution and large-scale analysis of complex problems are among the major advantages of this algorithm in neural network training which distinguishes it from similar optimization algorithms. The comparison of colonial competitive algorithm with genetic algorithm displays a plethora of calculations in the latter which is indicative of its slow pace in achieving the optimization solution. The important point, however, is the certainty with which it finds the solution because genetic algorithm, being trapped in local optimizations, is unable to find the desired optimization while the colonial competitive algorithm acts efficiently in finding optimal solution. In general, MLP neural network and colonial competitive algorithm achieved 95.48% accuracy in diagnosis of hepatitis.

REFERENCE

- [1] www.medicinenet.com/viral_hepatitis/article.html.
- [2] Shariati, S.; Haghghi, M.M. Comparison of anfis Neural Network with several other ANNs and Support Vector Machine for diagnosing hepatitis and thyroid diseases. *Computer Information Systems and Industrial Management Applications (CISIM)*, (2010), vol(10) pp 1101-1109.
- [3] Peng Guan, De-Sheng Huang, Bao-Sen Zhou. Forecasting model for the incidence of hepatitis A based on artificial neural net work. *China World Journal of Gastroenterol*, 10 (24), pp. 3579-3582. (2004).
- [4] <http://www.pezeshk.us/?p=2867>.

- [5] G. Eason, B. Noble, and I. N. Sneddon. On certain integrals of Lipschitz-Hankel type involving products of Bessel functions. *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551. (1955)
- [6] Rouhani, M.; Haghghi, M.M. The Diagnosis of Hepatitis Diseases by Support Vector Machines and Artificial Neural Networks. *Computer Science and Information Technology - Spring Conference. IACSITSC*. 10.1109/IACSIT-SC.2009.25. (2009)
- [7] Sathyadevi, G. Application of CART algorithm in hepatitis disease diagnosis. *Recent Trends in Information Technology (ICRTIT), International Conference on Digital Object Identifier*: 10.1109/ICRTIT.2011.5972349. (2011).
- [8] E. Dogantekin and A. Dogantekin, Automatic hepatitis diagnosis system based on Linear Discriminant Analysis and Adaptive Network based on Fuzzy Inference System. *Expert Systems with Applications* vol(36), pp 11282–11286. (2009).
- [9] Tahseen A Jilani, Huda Yasin and Madiha Mohammad Yasin. PCA-ANN for Classification of Hepatitis-C Patients. *International Journal of Computer Applications* 14(7):1–6. (2011)
- [10] Uttreshwar, G.S.; Ghatol, A.A Hepatitis B Diagnosis Using Logical Inference And Generalized Regression Neural Networks. *Advance Computing Conference, IEEE International Digital Object Identifier*: 10.1109/IADCC.2009.4809255. (2009).
- [11] Blake, C. L., & Merz, C. J. UCI repository of machine learning databases. Available from: <<http://www.ics.uci.edu/~mllearn/MLRepository.html>>. (1996).
- [12] Atashpaz Gargari E. developing of Social Optimization Algorithm and Surveying its efficiency. *The thesis of graduate top in school of electrical and computer engineering in Tehran University*. (2009).
- [13] Shea, P.M., and Liu, F. Operational Experience With a Neural Network in the Detection of Explosives in Checked Airline Luggage, *IJCNN San Diego, IEEE Press*, Vol. 2, pp. 175-178. (1990).
- [14] K.Polat and S. Gunes, Prediction of hepatitis disease based on principal component analysis and artificial immune recognition system. *Applied Mathematics and Computation* 189– 1282 (2007).