

A Hybrid ANN-GWO Algorithm for Prediction of Heart Disease

Hamza Turabieh

Information Technology Department, CIT Collage, Taif University, Taif, KSA
Email: turabieh@gmail.com

Received 23 November 2015; accepted 5 March 2016; published 8 March 2016

Copyright © 2016 by author and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The paper investigates the powerful of hybridizing two computational intelligence methods viz., Gray Wolf Optimization (GWO) and Artificial Neural Networks (ANN) for prediction of heart disease. Gray wolf optimization is a global search method while gradient-based back propagation method is a local search one. The proposed algorithm implies the ability of ANN to find a relationship between the input and the output variables while the stochastic search ability of GWO is used for finding the initial optimal weights and biases of the ANN to reduce the probability of ANN getting stuck at local minima and slowly converging to global optimum. For evaluation purpose, the performance of hybrid model (ANN-GWO) was compared with standard back-propagation neural network (BPNN) using Root Mean Square Error (RMSE). The results demonstrate that the proposed model increases the convergence speed and the accuracy of prediction.

Keywords

Artificial Neural Network, Gray Wolf Optimizer, Back-Propagation, Heart Disease

1. Introduction

Recently, heart problems are the major cause of deaths; human life depends on the performance of heart, which is pumping the blood to the whole parts through the body. Any heart problem will influence the other parts such as brain, kidney, lung, ... etc. Many reasons will increase the risk of heart problems (diseases) such as: high or low blood pressure, smoking, high level cholesterol, obesity and lack of physical exercise. The World Health Organization (WHO) has estimated that 17.5 million deaths occur worldwide in 2012. About 7.4 million were dead based on heart disease; 6.7 million were due to stroke. As a result, heart disease is the number one cause of deaths. WHO estimated that by 2030, almost 23.6 million people will die due to heart disease [1].

Discovering of heart disease by doctors based on symptoms, physical examinations and sings of patient body,

which is hard task in the medical field. This is a multi-layered problem, which may lead to wrong presumptions and unpredictable effects. As a result, healthcare industry today develops huge amounts of complex data about patients, hospitals resources, disease diagnosis, electronic patient records, and medical devices... etc. The huge amount of data (records) is considered as key resources to be processed and analyzed for knowledge extraction that enable doctors to take the correct presumptions, which enhance the probability of survive.

In the medical field, researchers preferred to use back-propagation neural network (BPNN) to model unstructured problems due to its ability to map complex non-linear relationships between input and output variables. The back-propagation algorithm is a local search algorithm that uses gradient descent to iteratively modify the weights and biases. Minimizing the fitness function measured based on the root mean square error between the actual and ANN predicted output. Easy trapped in local minimum and slow convergence is the drawback of back-propagation algorithm. Back-propagation algorithm used to generate different set of weights and biases every re-run during training phase. As a result, each run will produce different prediction results and convergence speed.

To overcome the drawback of back-propagation algorithm, gray wolf optimizer (GWO) has been used to find the optimal starting weights and biases for back-propagation algorithm. GWO is one of the latest bio-inspired optimization algorithms, which mimic the hunting activity of gray wolves in the wildlife. GWO search for optimal solutions in different directions in order to minimize the chance of trapped in local minimum and increase the convergence speed.

The rest of the paper is organized as follows: Section 2 presents the background and literature on heart disease and Section 3 describes the methodology, in which neural networks molding of heart disease and its optimization of gray wolf optimizer through training process. Section 4 presents a discussion of the experimental results and Section 5 concludes the presented work in this paper.

2. Background

A huge number of research works have been published in the prediction area for medical field. Dilip *et al.* [2] applied back-propagation neural network to predict neonatal disease diagnosis. The authors applied standard back-propagation on different data-sets of neonatal disease. The accuracy of the proposed model is 75% with higher stability. Yuehjen *et al.* [3] proposed a several hybrid models to predict heart disease such as logistic regression (LR), multivariate adaptive regression splines (MARS), artificial neural network (ANN) and rough set (RS). The performance of the proposed models is better than artificial neural network. Humar and Novruz [4] proposed a hybrid neural network that includes artificial neural network (ANN) and fuzzy neural network (FNN) to predict diabetes and heart diseases. The proposed model is able to support good accuracy values 84.24% and 86.8% for Pima Indians diabetes data-set and Cleveland heart disease data-set, respectively.

Milan and Sunila [5] investigated several data mining algorithms to solve cardiovascular disease such as: Artificial Neural Network (ANN), Support vector Machine (SVM), Decision Tree and RIPPER classifier. The accuracy of the ANN, SVM, Decision Tree and RIPPER are 80.06%, 84.12%, 79.05% and 81.08% respectively. SVM is able to predict the cardiovascular disease with higher accuracy. Mazurowski *et al.* [6] investigated the performance of class imbalance in training data using neural network for medical diagnosis. Two algorithms have been used standard back-propagation (BP) and particle swarm optimization (PSO). The results show that BP outperforms the PSO for small and large data-sets.

Resul *et al.* [7] introduced a methodology which uses SAS base software 9.1.3 for predicting of the heart disease. The proposed system is developed based on neural networks. The obtained accuracy is 89.01% which is considered good enough but not optimal. Archana and Sandeep [8] introduced a novel hybrid prediction algorithm with missing value imputation (HPM-MI). The authors proposed a hybrid model based on K-means clustering with Multilayer Perceptron. The performance of the proposed algorithm is investigated on three benchmark medical data sets namely Pima Indians Diabetes, Wisconsin Breast Cancer, and Hepatitis from the UCI Repository of Machine Learning. The results are very strong and the proposed model works fine when numbers of missing value are large in the data-set. Bajaj *et al.* [9] applied three data mining classification algorithm for heart disease such as decision tree, split validation and apply model. The proposed system is able to reduce medical mistakes and provide high accuracy prediction. Beheshti *et al.* [10] proposed a new meta-heuristics approach named Centripetal Accelerated Particle Swarm Optimization (CAPSO). The proposed approach is used to enhance the performance of ANN learning and accuracy. The performance has been evaluated over nine

standard medical data-sets and the results are promising. Interested readers can find more related work on [11]-[13].

3. Methodology

This section gives a detailed account of the basic ANN and GWO, followed by a discussion on the ANN-GWO.

3.1. Artificial Neural Network

Artificial Neural Network (ANN) is adopted from the biological complex system, the human brain, which consists of a huge number of highly connected elements called neurons. ANN tries to find the relationships between input-output data pairs. In general, the collected data were randomized and split into three groups; Training (70% of the data-set), Validation and Testing (30% of the data-set) data-sets. The training data-sets used to learn the ANN based on finding the relationship between input and out pairs by adjusting the weights and biases using back-propagation algorithm. However; through learning process, there is a probability of the neural network to over-fit or over-learn the input-output data-set. This problem will generate a weak mapping between input-output especially for unseen data-set. Once over-fitting occurs, the validation data-set is used through learning process to guide and stop training if the validation error begins to rise. The prediction evaluation of the developed ANN model is done after finishing training phase though testing data-set [14] [15]. Artificial neural network have been successfully applied to solve hard and complex problems in the field of industry and research. A huge number of publications have proved the strength of ANN in the medical field [16] [17].

The basic model of ANN consists of three layers, where each layer has different number of neurons. The three layers are input, hidden and output layers. All these layers are connected to each other in such a way so that each neuron in one layer is connected to all neurons in the following layer. The basic diagram for a network with a single neuron is illustrated in **Figure 1**.

The neurons of the input layer take the input data-set from the real environment. The input vector (x_i) is transmitted using the connection that multiplies its strength by a weight (w) to produce the product (w_{ij}). Each neuron has a biases (θ_i). The output of the neuron is generated by $f(\cdot)$ and $a(\cdot)$, which is a summation function and activation function. An activation function consists of two algebraic formulas linear and non-linear ones. These two functions enable the neural network to find the relationships between input and output. The outputs either send to other interconnected neurons or directly to the environment. The difference between the output and neural network output is considered as error.

3.2. Gray Wolf Optimizer

Gray Wolf Optimizer (GWO) is a new bio-inspired evolutionary algorithm proposed by Seyedali *et al.* [18]. GWO simulates the hunting process of gray wolves in the wildlife. Wolves used to live in a pack and there are two gray wolves (male and female) managing the other wolves in the pack. A very strong social dominate hierarchy is founded inside every pack. The size of each pack usually from 5 to 12 on average. According to [18] the social hierarchy of the pack is organized as following:

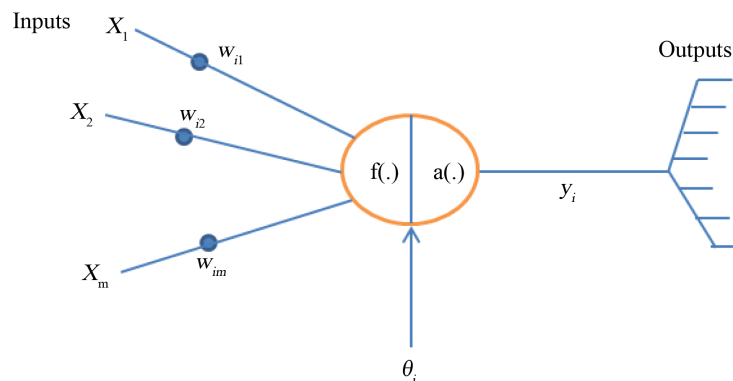


Figure 1. Mathematical principal of a neuron.

- 1) The alphas wolves (α): The leading wolves in the pack and responsible for making decisions. The alphas orders are dictated to the pack.
- 2) The betas wolves (β): The second level wolves after alphas in the pack. The main job of betas wolves to help and support alphas decisions.
- 3) The deltas wolves (δ): The third level in the pack is delta wolves. They used to follow alpha and beta wolves. The delta wolves have 5 categories as following:
 - a) Scouts: wolves are used to control and monitor the boundaries of the territory and alert the pack in case of danger.
 - b) Sentinels: protect and guarantee the safety of the pack.
 - c) Elders: strong wolves used to be alpha or beta wolves in the future.
 - d) Hunters: wolves used to help alpha and beta though hunting prey and providing food the pack.
 - e) Caretakers: wolves are responsible for caring the ill, wounded and weak wolves.
- 4) The omegas wolves (ω): are the lowest levels in the pack and they have to follow alpha, beta and delta wolves. They are the last wolves that are allowed to eat.

GWO algorithm consider alpha (α) wolves are the fittest solution inside the pack, while the second and third best solutions are named Beta (β) and delta (δ) respectively. The result of solutions inside the pack (population) are considered omega (ω). The process of hunting a prey is guided by α , β and ω .

The first step of hunting a prey is circling it by α , β and ω . The mathematical model of circling process as shown in equations 1, 2, 3 and 4.

$$\mathbf{X}(t+1) = \mathbf{X}_p(t) + \mathbf{A} \cdot \mathbf{D} \quad (1)$$

where \mathbf{X} is the gray wolf position. t is the number of iteration. \mathbf{X}_p is prey position and \mathbf{D} is evaluated from Equation (2).

$$\mathbf{D} = |\mathbf{C} \cdot \mathbf{X}_p(t+1) - \mathbf{X}(t)| \quad (2)$$

the \mathbf{A} and \mathbf{C} are coefficient vectors are evaluated based on Equations (3) and (4) respectively.

$$\mathbf{A} = 2a \cdot \mathbf{r}_1 - a \quad (3)$$

$$\mathbf{C} = 2\mathbf{r}_2 \quad (4)$$

where a is a linearly decreased from 2 to 0 through the number of iterations, which is used to control the tradeoff between *exploration* and *exploitation*. Equation (5) used to update the value of variable a , where $NumIter$ is the total number of iterations. Two random vectors between [0,1] namely r_1 and r_2 to simulate hunting a prey (find the optimal solution). The solutions alpha, beta and delta are considered to have a good knowledge about the potential location of prey. These three solutions helps others wolves (omega) to update their positions according to the position of alpha, beta and delta. Equation 6 presents the formula of updating wolves' positions.

$$a = 2 - t(2/NumIter) \quad (5)$$

$$\mathbf{X}(t+1) = (\mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3)/3 \quad (6)$$

The values of \mathbf{X}_1 , \mathbf{X}_2 and \mathbf{X}_3 is evaluated as in Equations (7) (8) and (9) respectively.

$$\mathbf{X}_1 = |\mathbf{X}_\alpha - \mathbf{A}_1 \cdot \mathbf{D}_\alpha| \quad (7)$$

$$\mathbf{X}_2 = |\mathbf{X}_\beta - \mathbf{A}_2 \cdot \mathbf{D}_\beta| \quad (8)$$

$$\mathbf{X}_3 = |\mathbf{X}_\delta - \mathbf{A}_3 \cdot \mathbf{D}_\delta| \quad (9)$$

The \mathbf{X}_1 , \mathbf{X}_2 and \mathbf{X}_3 are the best 3 solutions in the population at iteration t . The values of \mathbf{A}_1 , \mathbf{A}_2 and \mathbf{A}_3 are evaluated in Equation (3). The values of \mathbf{D}_α , \mathbf{D}_β and \mathbf{D}_δ are evaluated as shown in Equations (10) (11) and (12) respectively.

$$\mathbf{D}_\alpha = |\mathbf{C}_1 \cdot \mathbf{X}_\alpha - \mathbf{X}| \quad (10)$$

$$\mathbf{D}_\beta = |\mathbf{C}_2 \cdot \mathbf{X}_\beta - \mathbf{X}| \quad (11)$$

$$D_\delta = |C_3 \cdot X_\delta - X| \tag{12}$$

The values of C_1 , C_2 and C_3 is evaluated in Equation (4). The pseudo code for GWO is shown in **Figure 2**.

3.2. Gray Wolf Optimizer of Neural Networks (ANN-GWO)

To find an accurate ANN model and reduce the drawback of back-propagation algorithm, GWO is hybridized with ANN. The proposed idea consists of two major steps. In the first one, ANN is trained using GWO. GWO is used to find the optimal initial weight and biases. The second step involves training the neural network using back-propagation algorithm. The weights and biases evolved from GWO. This idea will enhance the performance of back-propagation to search for global optima model. The weights and biases are evaluated as a vector of variables for the proposed model. The fitness of each vector is evaluated based on the Root Mean Square Error (RMSE), which finds the error between the actual input and predicted output. Equation (13) presents the RMSE, where T_i is the target output and P_i is predicted value from ANN. A lower value of *RMSE* indicates a better model. **Figure 3** illustrates the pseudo code of the hybrid ANN-GWO algorithm.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (T_i - P_i)^2} \tag{13}$$

4. Experimental Results

In the experiments, the proposed ANN-GWO is implemented using MATLAB and simulation are performed on an Intel Pentium 4, 2.33 GHz computer. We execute 11 independent runs on the data-set.

4.1. Problem Description

In this work, the data-set is adopted from UCI Machine Learning Repository, Cleveland database [19] which is a medical data-set related to heart disease. Cleveland database used to classify person into four classes (Normal (0), First Stroke (1), Second Stroke (2) and End of life (3)) regarding to the heart disease. The data-set has 166

```

Input:  n number of wolves in the pack.
          NumIter total Number of iteration for optimization.
Output:  $x_a$  Optimal gray wolf position.
            $f(x_a)$  Best fitness value.
Begin
    Generate a population of n gray wolves positions randomly.
    Determine  $\alpha$ ,  $\beta$  and  $\delta$  based on the fitness value.
    While (termination condition dose not met)
    {
        for each  $wolf_i \in pack$  do
            Update  $wolf_i$  based on equation 6.
        end for
        Update  $\alpha$ ,  $A$  and  $C$ .
        Evaluate the positions of individual wolves.
        Update  $\alpha$ ,  $\beta$  and  $\delta$ .
    }
End
    
```

Figure 2. Gray wolf optimizer algorithm.

```

Training ANN using GWO
    set number of pack (population).
    set NumIter total Number of iteration for optimization.
    generate ANN model based on back-propagation algorithm.
    run GWO to find the best values of weights and biases (vector) based on equation 13.
    return the initial optimal weights and biases ( $\alpha$ ).
Training ANN using back-propagation:
    Use GWO results as initial weights and biases.
    return the trained ANN model.
    
```

Figure 3. The hybrid ANN-GWO algorithm.

records with 13 attributes (inputs) for each record and 4 classes (outputs). **Table 1** and **Table 2** illustrate the benchmark data-set and classes respectively.

4.2. Tuning Parameters for ANN and GWO

To find an accurate ANN model, a good parameters setting should be used for training and testing the ANN model. GWO is used to search for the initial optimal weights and biases. The parameters setting used for the proposed ANN-GWO after some preliminary experiments are shown in **Table 3** and **Table 4**. GWO is able to find an optimal initial weights and biases in 51 generations, and the time used to reach optimal value of RMSE 2.812 mm was evaluated in 29.532 s. **Figure 4** presents the convergence process in smooth curve with a fast decrease at the start and gradually slow down.

Table 1. Benchmark data-sets attributes.

Attributes	Description	Range
Age	Age in years	Continuous
Sex	& (1 = male; 0 = female)	0, 1
Cp	Value 1: typical angina Value 2: atypical anginal Value 3: non-anginal pain Value 4: asymptotic	1, 2, 3, 4
trestbps	Resting blood pressure (in mm Hg)	Continuous
chol	Serum cholesterol in mg/dl	Continuous
fbs	Fasting blood sugar. 120 mg/dl (1 = true; 0 = false)	0, 1
restecg	electrocardiography results Value 0: normal Value 1: having ST-T wave abnormality (T wave inversions and/or ST Elevation or depression of > 0.05 mV) Value 2: showing probable or definite left	0, 1, 2
Thalach	Maximum heart rate achieved	Continuous
Exang	Exercise induced angina (1 = yes; 0 = no)	0, 1
Oldpeak	ST depression induced by exercise relative to rest	Continuous
Slope	The slope of the peak exercise ST segment Value 1: up sloping Value 2: flat Value 3: down sloping	0, 1, 2
Ca	Number of major vessels (0 - 3) Colored by fluoroscopy	Continuous
Thal	Normal, fixed defect, reversible defect	3, 6, 7

Table 2. Benchmark datasets classes.

Class	Description
Class 0	Normal Person
Class 1	First Stroke
Class 2	Second Stroke
Class 3	End of Life

Table 3. Parameters setting for GWO.

Parameter	Value
Iteration	100
Number of pack (population size)	10

Table 4. Parameters setting for ANN.

Parameter	Value
Iteration	5000
Number of Neurons in Input layer	13
Number of Neurons in Hidden layer	8, 5, 2
Number of Neurons in output layer	1

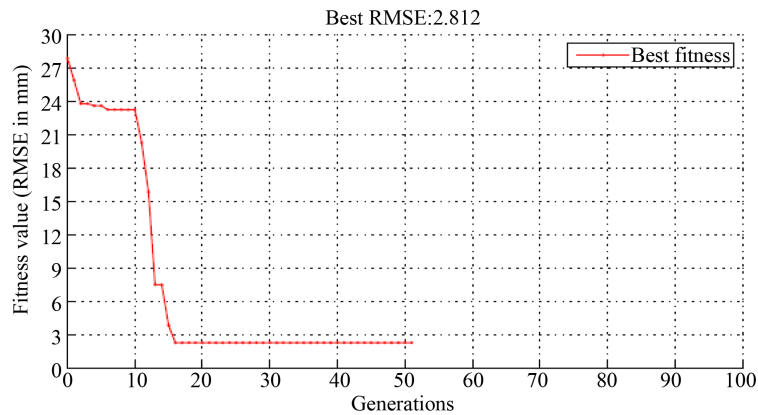


Figure 4. Evolving optimal weights and biases using gray wolf optimizer.

4.3. Results

The performance of GWO in finding the initial optimal weights and biases is founded by minimizing the RMSE (Equation (13)). The experiments for finding the initial weights and biases were implemented eleven times to ensure that RMSR reach the optimal value. In the heart disease problem, the RMSE converges to a value 2.812 mm using GWO, which is very close to 0.

In order to evaluate the performance of ANN-GWO, a comparison is done between standard ANN and ANN-GWO. **Table 5** presents the results obtained for both models. The neural network architecture used for modeling ANN-GWO and standard ANN (13-8-5-2-1) is trained using back-propagation algorithm. The proposed model (ANN-GWO) is able to achieve a high quality performance goal 0.0029 in 40 epochs taking 2.435 s as shown in **Figure 5**. Whilst, the standard neural network based on back-propagation algorithm took 3000 epochs and 73.43 s to reach the 0.0019 goal as shown in **Figure 6**.

4.4. Discussions

The performance of ANN-GWO in tracking the original and predicted output for training and testing data-set is illustrated in **Figure 7** and **Figure 8** respectively. Analyzing the results shows that the performance of standard ANN is easy to trap in local minima and slow convergence. Whilst, ANN-GWO is faster convergence and reach optimal values better than the standard ANN. **Figure 7** and **Figure 8** show the values of the original and predicted results of the ANN-GWO and standard ANN respectively. It is clear that the performance of ANN-GWO is much better than standard ANN.

Figure 9 and **Figure 10** show the box plots that illustrate the distribution results of the ANN-GWO and standard ANN respectively. It is clear that the ANN-GWO reduces gab between the best, average, and worst prediction

Table 5. Results of RMSE (mm).

Run	ANN-GWO			Standard ANN		
	Training	Validation	Testing	Training	Validation	Testing
Run 1	2.213	3.196	3.688	3.325	3.228	3.689
Run 2	2.813	3.123	3.899	4.429	4.832	4.699
Run 3	1.992	3.221	3.614	3.639	3.701	3.894
Run 4	2.119	3.001	3.505	3.759	3.779	3.942
Run 5	2.871	3.001	3.445	3.682	3.701	3.783
Run 6	2.613	3.306	3.692	3.432	3.604	3.731
Run 7	2.412	3.145	3.244	3.592	3.481	3.691
Run 8	2.001	2.991	3.502	3.841	3.992	4.829
Run 9	2.413	3.001	4.392	3.958	4.002	4.314
Run 10	1.913	2.991	3.402	3.491	3.509	3.393
Run 11	2.319	3.219	3.589	3.495	3.349	3.321

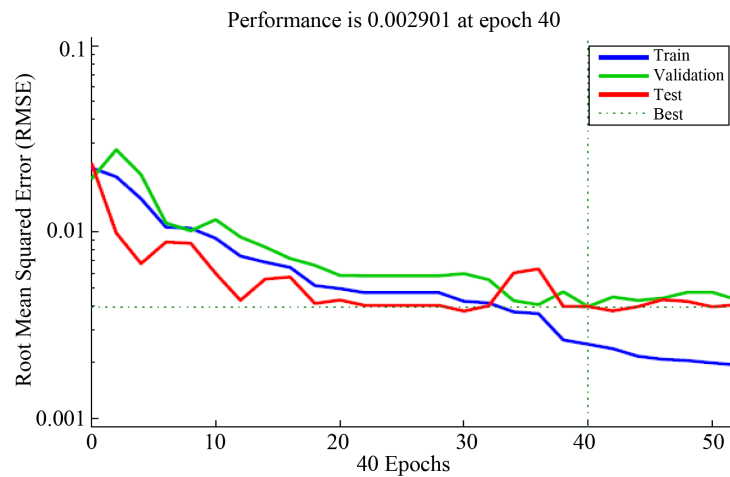


Figure 5. Training of ANN-GWO model.

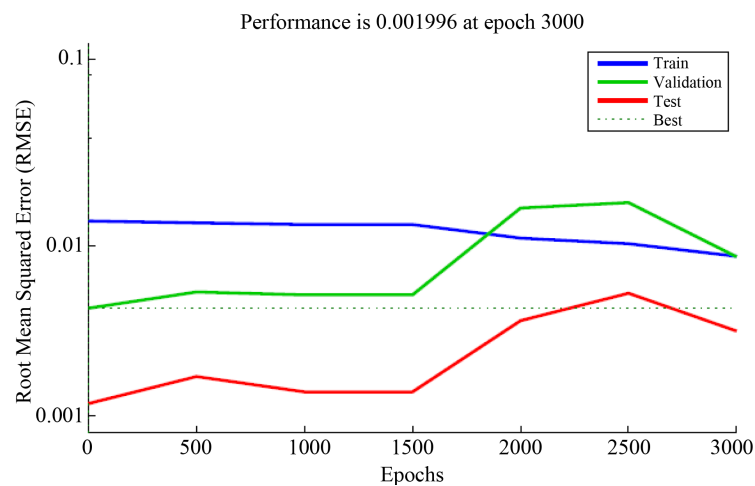


Figure 6. Training of standard ANN model.

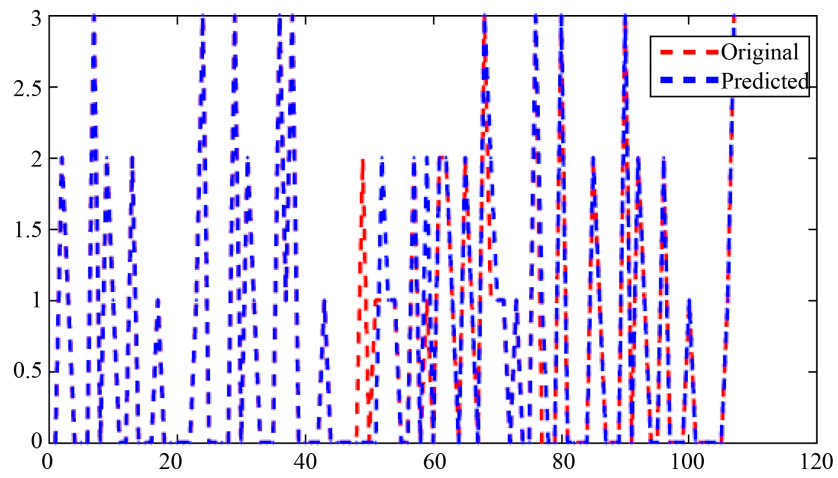


Figure 7. Original and predicted output for training data-set.

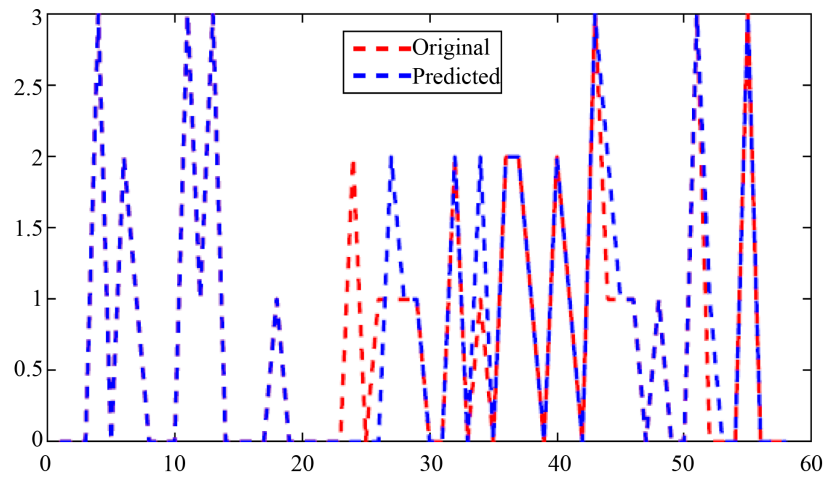


Figure 8. Original and predicted output for testing data-set.

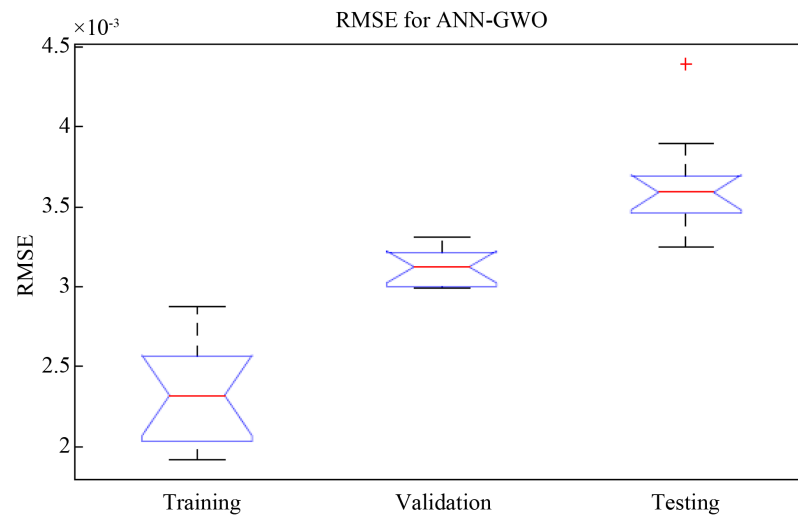


Figure 9. Box plot of RMSE for ANN-GWO.

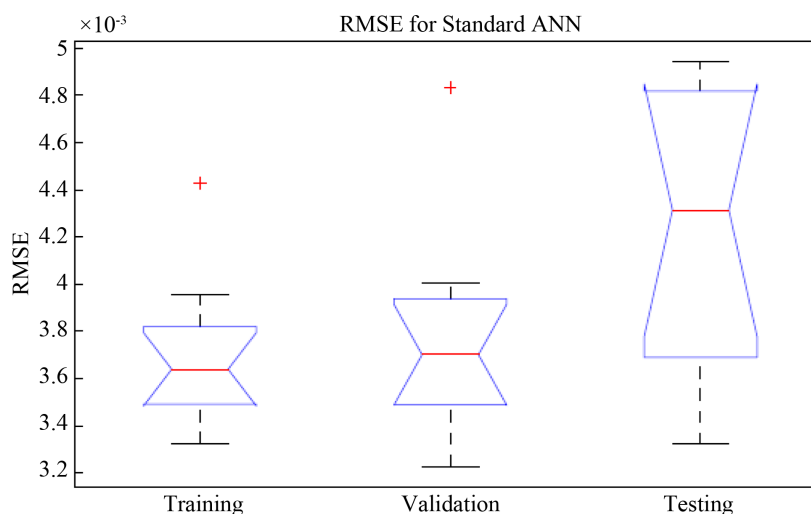


Figure 10. Box plot of RMSE for standard ANN.

qualities, which shows that the proposed method is robust and much better than the standard ANN. We believe that this is because the gray wolf optimizer helps the back-propagation to overcome the drawback of standard ANN.

5. Conclusion

This work has proposed a new hybrid algorithm between artificial neural network and gray wolf optimizer to enhance the performance of back-propagation algorithm and overcome the drawback of getting stuck at local minima. Based on the results of this work, GWO helps ANN to find optimal initial weights and biases, which speed up the convergence speed and reduce the RMSE error. The proposed hybrid model uses GWO as a global search algorithm, while back-propagation as a local search one. This kind of hybridization makes a balance between exploration and exploitation. Moreover, the ANN-GWO model in comparison with standard back-propagation ANN, took almost more than half the time to find the optimal model. However, we believe that a future study is needed to investigate the optimal design of the neural network architecture such as number of hidden layers, number of neurons, transfer functions and learning functions.

References

- [1] World Health Organization (WHO), Cardiovascular Diseases. http://www.who.int/cardiovascular_diseases/en/
- [2] Dilip, R.C., Mridula, C. and Samanta, R.K. (2011) An Artificial Neural Network Model for Neonatal Disease Diagnosis. *International Journal of Artificial Intelligence and Expert Systems*, **2**, 96-106.
- [3] Shao, Y.E., Hou, C.-D. and Chiu, C.-C. (2014) Hybrid Intelligent Modeling Schemes for Heart Disease Classification. *Applied Soft Computing*, **14**, 47-52. <http://dx.doi.org/10.1016/j.asoc.2013.09.020>
- [4] Kahramanli, H. and Allahverdi, N. (2008) Design of a Hybrid System for the Diabetes and Heart Diseases. *Expert Systems with Applications*, **35**, 82-89. <http://dx.doi.org/10.1016/j.eswa.2007.06.004>
- [5] Kumari, M. and Godara, S. (2011) Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction 1. *International Journal of Computer Science and Technology*, **2**, 304-308.
- [6] Mazurowski, M.A., Habas, P.A., Zurada, J.M., Lo, J.Y., Baker, J.A. and Tourassi, G.D. (2008) Training Neural Network Classifiers for Medical Decision Making: The Effects of Imbalanced Datasets on Classification Performance. *Neural Networks*, **21**, 427-436. <http://dx.doi.org/10.1016/j.neunet.2007.12.031>
- [7] Das, R., Turkoglu, I. and Sengur, A. (2009) Effective Diagnosis of Heart Disease through Neural Networks Ensembles. *Expert Systems with Applications*, **36**, 7675-7680. <http://dx.doi.org/10.1016/j.eswa.2008.09.013>
- [8] Purwar, A. and Singh, S.K. (2015) Hybrid Prediction Model with Missing Value Imputation for Medical Data. *Expert Systems with Applications*, **42**, 5621-5631. <http://dx.doi.org/10.1016/j.eswa.2015.02.050>
- [9] Bajaj, P., Choudhary, K. and Chauhan, R. (2015) Prediction of Occurrence of Heart Disease and Its Dependability on RCT Using Data Mining Techniques. *Advances in Intelligent Systems and Computing, Springer India*, **340**, 851-858.

- [10] Beheshti, Z., Shamsuddin, S.M., Beheshti, E. and Yuhaniz, S.S. (2014) Enhancement of Artificial Neural Network Learning Using Centripetal Accelerated Particle Swarm Optimization for Medical Diseases Diagnosis. *Soft Computing*, **18**, 2253-2270. <http://dx.doi.org/10.1007/s00500-013-1198-0>
- [11] Seera, M. and Lim, C.P. (2014) A Hybrid Intelligent System for Medical Data Classification. *Expert Systems with Applications*, **41**, 2239-2249. <http://dx.doi.org/10.1016/j.eswa.2013.09.022>
- [12] Woźniak, M., Graña, M. and Corchado, E. (2014) A Survey of Multiple Classifier Systems as Hybrid Systems. *Information Fusion*, **16**, 3-17. <http://dx.doi.org/10.1016/j.inffus.2013.04.006>
- [13] Nishara Banu, M.A. and Gomathy, B. (2014) Disease Forecasting System Using Data Mining Methods. *International Conference on Intelligent Computing Applications (ICICA)*, Coimbatore, 6-7 March 2014, 130-133.
- [14] Hagan, M.T., Demuth, H.B. and Beale, M. (1996) *Neural Network Design*. PWS Publishing Co., Boston.
- [15] Zhao, Y., Gao, J. and Yang, X.Z. (2005) A Survey of Neural Network Ensembles. *International Conference on Neural Networks and Brain*, **1**, 438-442.
- [16] James, A.R. (2014) A Fuzzy Nearest Neighbor Neural Network Statistical Model for Predicting Demand for Natural Gas and Energy Cost Savings in Public Buildings. *Expert Systems with Applications*, **41**, 1813-1829.
- [17] Wang, W.-C., Chau, K.-W., Qiu, L. and Chen, Y.-B. (2015) Improving Forecasting Accuracy of Medium and Long-Term Runoff Using Artificial Neural Network Based on EEMD Decomposition. *Environmental Research*, **139**, 46-54. <http://dx.doi.org/10.1016/j.envres.2015.02.002>
- [18] Mirjalili, S., Mirjalili, S.M. and Lewis, A. (2014) Grey Wolf Optimizer. *Advances in Engineering Software*, **69**, 46-61. <http://dx.doi.org/10.1016/j.advengsoft.2013.12.007>
- [19] UCI, Machine Learning Repository, Cleveland Database. <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>