

## **HYBRID NSGA-II OPTIMIZATION FOR IMPROVING THE THREE-TERM BP NETWORK FOR MULTICLASS CLASSIFICATION PROBLEMS**

<sup>1,2</sup>Ashraf Osman Ibrahim, <sup>2</sup>Siti Mariyam Shamsuddin &  
<sup>3</sup>Sultan Noman Qasem

<sup>1</sup> *Faculty of Computer and Technology, Alzaiem Alazhari University,  
Khartoum North, Sudan*

<sup>2</sup> *UTM Big Data Centre, Universiti Teknologi Malaysia, Malaysia*

<sup>3</sup> *College of Computer and Information Sciences, Al Imam Mohammad Ibn  
Saud Islamic University (IMSIU), Riyadh, Saudi Arabia*

ashrafosman2@gmail.com; mariyam@utm.my;  
sultann.noman@ccis.imamu.edu.sa

### **ABSTRACT**

Recently, hybrid algorithms have received considerable attention from a number of researchers. This paper presents a hybrid of the multiobjective evolutionary algorithm to gain a better accuracy of the final solutions. The aim of using the hybrid algorithm is to improve the multiobjective evolutionary algorithm performance in terms of the enhancement of all the individuals in the population and increase the quality of the Pareto optimal solutions. The multiobjective evolutionary algorithm used in this study is a non-dominated sorting genetic algorithm-II (NSGA-II) together with its hybrid, the backpropagation algorithm (BP), which is used as a local search algorithm to optimize the accuracy and complexity of the three-term backpropagation (TBP) network. The outcome positively demonstrates that the hybrid algorithm is able to improve the classification performance with a smaller number of hidden nodes and is effective in multiclass classification problems. Furthermore, the results indicate that the proposed hybrid method is a potentially useful classifier for enhancing the classification process ability when compared with the multiobjective genetic algorithm based on the TBP network (MOGATBP) and certain other methods found in the literature.

**Keywords:** Artificial Neural Network, hybridization technique, genetic algorithm, NSGA-II, multiobjective optimization.

## INTRODUCTION

Artificial neural networks (ANNs) are the most commonly used classifiers due to their high ability for prediction and adaptability (Fan, Chang, Lin, & Hsieh, 2011). The determining of the optimal structure of the ANN is a very difficult task and depends on the user experience. In most cases, the selection of the parameters is performed manually, and, in this case, the adjustment and tuning of the parameters of the ANN may be effective for the performance of the network. There has been a remarkable increase in the use of evolutionary algorithms (EAs) for solving the optimization problems in recent years (Almeida & Ludermir, 2010). Therefore, there is a need to use other methods rather than manual adjustment and the usual choice of the parameters. The optimization of the parameters of ANNs is considered to be one of the most important problems that need to be solved using this kind of algorithm. Several studies have used the Pareto optimal concept in the classification problems using multiobjective optimization techniques (Abbass, 2002; Fernandez, Martínez, Hervás, & Gutiérrez, 2010; Qasem & Shamsuddin, 2011).

In addition, a multiobjective evolutionary algorithm (MOEA) has been shown to be useful for improving the performance of solving the multiobjective optimization problem (MOOP) (Zhou, Qu, Li, Zhao, Suganthan & Zhang, 2011). Therefore, multiobjective evolutionary algorithms (MOEAs) are used to produce and optimize the parameters of artificial neural networks (ANNs) with the optimization of two or more conflicting objectives (Fernández, Hervás, Martínez-Estudillo, & Gutiérrez, 2011; Goh, Teoh, & Tan, 2008; Qasem, Shamsuddin, Hashim, Darus, & Al-Shammari, 2013). These kinds of algorithm are applied to improve the generalization of the training and unseen data in the network. In this respect, MOEAs are used in the process of simultaneously optimizing two or more conflicting objectives. Recently, the MOEA research area has become one of the ‘hottest’ areas in the field of evolutionary computation (Zhou et al., 2011). Furthermore, MOEAs are convenient for producing and designing the appropriate and accurate ANNs with the optimization of the objectives simultaneously. Hence, due to their ability to improve structural performance, MOEAs have been applied successfully to optimize the network structure, connection weights and train the network.

Hybridization is two or more soft computing techniques that are integrated to leverage the advantages of individual methods (Bonissone, Chen, Goebel, & Khedkar, 1999). In MOEAs, hybridization is the combination of the different characteristics and advantages of different algorithms. Some recent works

have categorized the hybridization according to the different search methods, search and updating methods, and different methods in different search phases (Zhou et al., 2011). Hybridization offers a better speed of convergence to the evolutionary approach and better accuracy of the final solutions, and, in recent times, hybrid algorithms have had extensive success in solving many real world complex problems. Based on the fact that most existing algorithms have their own weakness, the importance of hybrid algorithms is indisputable, since hybridization aims to solve the weaknesses of the existing algorithms by benefitting from the advantages of the other algorithms. In addition, the use of a computational method has been gradually growing. Therefore, continuous research efforts to employ hybridization techniques for the classification problems have been focused upon since the last decade (Idris, Selamat, & Omatu, 2014). However, more intensive works are needed to design and develop the ANNs classifier for the classification problems using hybrid algorithms.

In this paper, an attempt is made to propose hybrid multiobjective evolutionary algorithms (HMOEAs) for the ANN to improve the generalization of the training and unseen data. In the proposed approach, the NSGA-II hybrid with a local search algorithm was applied to optimize the two objectives simultaneously, the number of hidden nodes in the hidden layer and the error rates of the ANN to solve the multiclass classification problem. The local search algorithm was used to enhance all the individuals in the population and increase the quality of the Pareto optimal solution.

## **LITERATURE REVIEW**

Over the past few years, much research has been undertaken on the rapid expansion of developing algorithms to overcome the problems of ANNs. Yao (1999) provided a general framework to use EAs for evolving ANNs as one of the most successful applications in this area of research. He employed genetic algorithm (GA) for optimizing ANNs, as this technique has been successful in the multiobjective optimization to solve ANNs. For example, multiobjective genetic algorithm optimization was used by Pettersson, Chakraborti, and Saxén (2007) for training a feed forward neural network, number of nodes, the architecture, as well as the weights. A Pareto front was effectively constructed by minimizing the training error and the network size using noisy data. Another study, by Fernandez et al. (2010), introduced two methods called MPENSGA2E and MPENSGA2S. In this study, a memetic Pareto evolutionary neural network technique was used to solve the two class

and multiclass classification problems by using a multilayer perceptron neural network hybrid with the NSGA2 algorithm. Similarly, Goh, Teoh & Chen, (2008) used the hybrid multiobjective evolutionary method and artificial neural networks based on a micro-hybrid genetic algorithm for classification of the medical data and other data. A multiobjective GA was presented by Ahmad, Isa, Hussain, and Sulaiman (2012) using Pareto optimal optimization of the ANN for the classification of breast cancer diagnosis problems. In addition, the hybrid model using genetic algorithm (GA) and backpropagation (BP) networks for the diagnosis of diabetes diseases used GA to optimize the network connection weights, which were introduced (Karegowda, Manjunath, & Jayaram, 2011). Also, a general framework using GA for designing neural network ensembles was presented (García-Pedrajas, Hervás-Martínez, & Ortiz-Boyer, 2005). Another method used to generalize multi-layer perceptrons (MLP) improved the performance of the evolutionary model (García-Pedrajas, Ortiz-Boyer, & Hervás-Martínez, 2004). Delgado, Cuéllar, and Pegalajar (2008) proposed a hybrid MOGA method based on the SPEA2 and NSGA2 algorithms. The proposed method was made to optimize the training and the topology of the recurrent neural network (RNN) simultaneously in time-series prediction problems.

Further, some researchers considered hybrid techniques to enhance the performance of ANNs. Therefore, many studies in the literature focused on hybrid algorithms that combine ANNs with other techniques. Such hybrid algorithms are artificial bee colony algorithm (Qiongshuai & Shiqing, 2011), artificial fish swarm optimization (Hasan, Quo, & Shamsuddin, 2012), naive bayes classifier (Adam et al., 2011), hidden markov model (Tang, 2009), multiple linear regression models (Khashei, Zeinal Hamadani, & Bijari, 2012), fuzzy c-means clustering (Egrioglu, Aladag, & Yolcu, 2013), hybrid multilayered perceptron and radial basis function (Isa, Mashor, Othman, & Zamli, 2005). In the same context, hybridization is also an important feature in the area of EAs, which has received relative attention over the past few years. As such, hybrid global and local search algorithms have been one of the new research areas, known as memetic algorithms (MAs). Due to the phenomenal success of this kind of algorithm, MAs have been successfully used in different applications to solve many problems. One of the most successful domains to apply MAs is multiobjective optimization problems (Ibrahim, Hasan, & Noman, 2014; Ishibuchi, Hitotsuyanagi, Tsukamoto, & Nojima, 2009; Zhou et al., 2011). In contrast, several previous studies have presented prospective achievements by using ANNs, MOEAs and local optimizers to speed up the convergence (H.A. Abbass, 2003; Jin, Sendhoff, & Körner, 2006; Wiegand, Igel, & Handmann, 2004). The research by H.A. Abbass (2003) studied the

advantages of hybridizing Pareto differential evolution with the BP algorithm as a local search algorithm for a training method to speed up convergence and a long training time. Furthermore, one of the most famous works in this area, by H. A Abbass (2002) concluded that the memetic Pareto artificial neural network (MPANN) method based on a Pareto optimal solution has better generalization, and positive results were obtained in the context of medical applications. For example, a multiobjective evolutionary learning algorithm was introduced (Cruz-Ramírez, Hervás-Martínez, Fernández, Briceño, & de la Mata, 2012) using an improved version of the NSGA-II algorithm called MPE NSGA-II hybridized with a local search algorithm for training ANNs with generalized radial basis functions. An intelligent classifier approach was proposed by Ibrahim, Shamsuddin, Ahmad, and Qasem (2013), by using the three-term backpropagation (TBP) network based on the Elitist Multiobjective Genetic Algorithm. This is used to reduce or optimize the error rate and network structure of TBP network simultaneously to achieve accurate classification results on the medical disease diagnosis classification. However, the present study proposes optimizing the structure of the TBP network and its parameters simultaneously. Furthermore, it includes a local search to speed the convergence and improve the performance of the TBP network. Subsequently, this present work develops the TPB network by optimizing the structure and error rates simultaneously.

## **BACKGROUND MATERIALS**

### **Three-term Backpropagation Algorithm (TBP)**

The three-term backpropagation algorithm proposed by Zweiri, Whidborne, and Seneviratne (2003) employs the standard architecture and procedure of the standard backpropagation algorithm. However, in addition to the learning rate and momentum parameters, the third parameter, called the proportional factor (PF) is introduced. This has proven to be successful in improving the convergence rate of the algorithm and speeding up the weight adjusting process. Moreover, the PF represents the difference between the output and the target at each iteration and increases the backpropagation learning speed. It contains several layers in which each layer interacts with the upper layer connected to it by the connection link. The connection link specifically connects the nodes within the layers with the nodes in the adjacent layer that builds a highly interconnected network. The bottom-most layer, called the input layer, will accept and process the input and pass the output to the next adjacent layer, called the hidden layer. Figure 1 depicts the general architecture of an ANN.

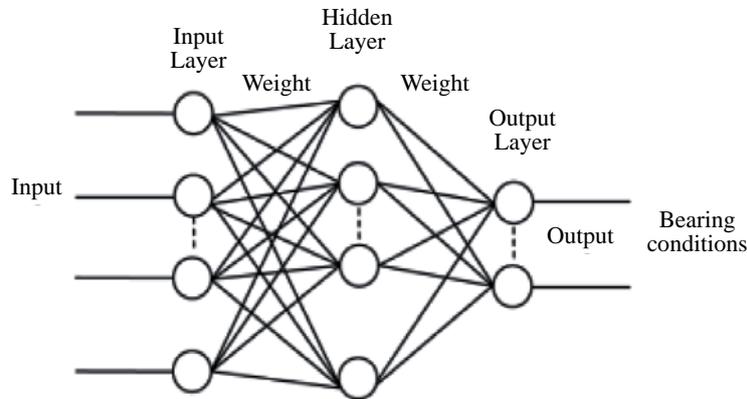


Figure 1. Artificial neural network architecture

### Non-dominated Sorting Genetic Algorithm (NSGA-II)

The non-dominated sorting genetic algorithm-II (NSGA-II) proposed by Deb, Pratap, Agarwal and Meyarivan (2002) was an upgraded and improved version of NSGA (Srinivas & Deb, 1994). Both of these algorithms were based on GA operators. Moreover, NSGA-II is a fast and elitist multiobjective genetic algorithm, which was used to obtain the set of Pareto optimal front solutions. The good performance of global searching of a non-dominated sorting multiobjective optimization genetic algorithm has become the preferred method of optimization algorithm. It proposes a new method and a new arithmetic operator, which is a fast non-dominated sorting approach with a crowded comparison operator. NSGA-II is acknowledged to be one of the most famous Pareto optimal solution algorithms. It requires the simultaneous minimization or maximization of two or more objective functions. To date, much work concerning the optimization and design has been done (Ak, Vitelli, Zio, Droguett & Jacinto, 2012; Cruz-Ramírez, Hervan-Marti, Fernandez, Briceno & de la Mata, 2012; Qasem, Shamsuddin, & Zain, 2011; Ramesh, Kannan, & Baskar, 2011). All of these studies demonstrated that the genetic algorithm and its upgraded derivatives are feasible for optimal design.

The NSGA-II algorithm begins by generating a random population of chromosomes or solutions of size  $N$ . Firstly, both the parent population and offspring population are combined to form a combined population of size  $2N$  instead of finding the non-dominated fronts of the offspring population only. Subsequently, the non-dominated sorting procedure is performed on the entire

population. This procedure allows a global non-domination. It checks between the offspring and parent solutions, and enables NSGA-II to converge faster. Figure 2 illustrates the scenario of the NSGA-II.

The crowding distance is used in the selection of parents for a new individual and the selection of a new population based on a comparison of the congestion around a solution. A greater crowding distance is preferred in order to maintain the diversity of the solutions.

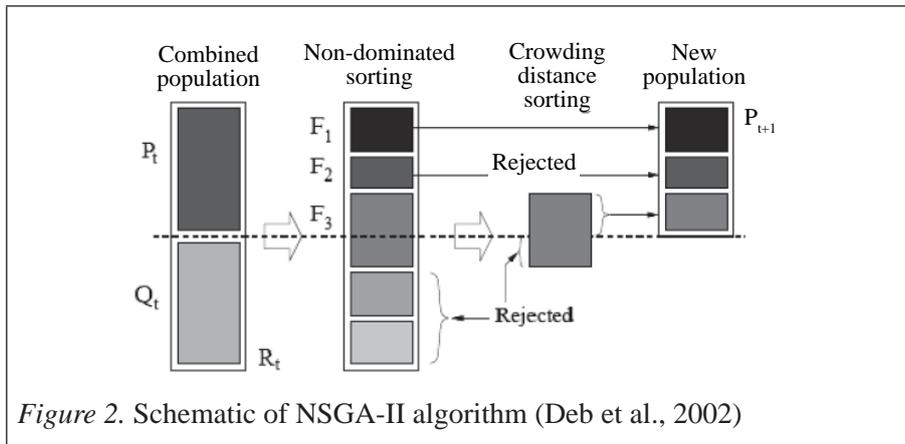


Figure 2. Schematic of NSGA-II algorithm (Deb et al., 2002)

### Local Search Algorithm

Local search algorithms are widely used for several problems in different areas, but have received more attention in computer science and engineering, particularly artificial intelligence applications. It is known that the local methods are able to find the local optimum when searching in a small area of space. Hybrid global and local search algorithms, known as memetic algorithms (MAs), are one of the new research areas. Due to the phenomenal success of this kind of algorithm, MAs have been successfully used in different applications to solve numerous problems (Neri & Cotta, 2012). Furthermore, MAs are able to provide not only the best speed of convergence to the evolutionary approach, but also the best accuracy for the final solution (Lara, Sanchez, Coello Coello, & Schutze, 2010). In this study, a classical BP algorithm was used as the local search method.

## METHODOLOGY

In this section, the details of the proposed hybrid method are given. In fact, the proposed method adapts the non-dominated sorting genetic algorithm (NSGA-

II) (Deb et al., 2002). The hybrid, with the backpropagation (BP) algorithm as a local search algorithm to enhance all individuals in the population, is a good option to improve the performance of the network. The hybrid non-dominated sorting genetic algorithm based TBP network is implemented. The proposed algorithm evolves the network architecture and accuracy simultaneously with each individual being a fully specified TBP network. This algorithm has been proposed to determine the best performance and the corresponding architecture of the TBP network.

However, the proposed method begins with the first step, which is collecting, normalizing and reading the dataset, followed by dividing the dataset into training data and testing data. Then, the minimum and maximum number of hidden nodes, and the maximum number of iterations are set. In addition, the individual length is computed. Furthermore, the parameters of the TBP network are determined by the traditional algorithms. Then the population of the NSGA-II is generated and initialized, and the local search algorithm is run to every individual. Every iteration is evaluated based on the objective functions. Next, the NSGA-II is used to obtain a list of the population. Then, a rank value is given to each individual equal to its dominance level and crowding distance. After the maximum iterations are reached, the proposed method stops and outputs a set of non-dominated TBP networks. To evaluate the TBP network performance of the proposed method, two objective functions were used in this study. The first was the fitness function, which is the performance of the network (Accuracy) based on the mean square error (MSE) of the training set. The other fitness function is the complexity of the network based on the number of hidden nodes in the hidden layer of the TBP network. Figure 3 shows the general optimization framework for the proposed method.

## **EXPERIMENTS**

The performance of the hybrid NSGA-II algorithm is tested on all the datasets with the same parameter for training the TBP network. Here the experimental results of the study on the hybrid NSGA-II algorithm for the TBP network are presented based on two objectives: the complexity of the network in terms of the number of hidden nodes, and the accuracy of the network. The proposed algorithm is evaluated using the 10-fold cross-validation technique. For the experimental design, the five multiclass datasets in Table 1, are the most widely used datasets in pattern classification. They were also considered and obtained from the UCI machine learning repository (Asuncion & Newman, 2007).

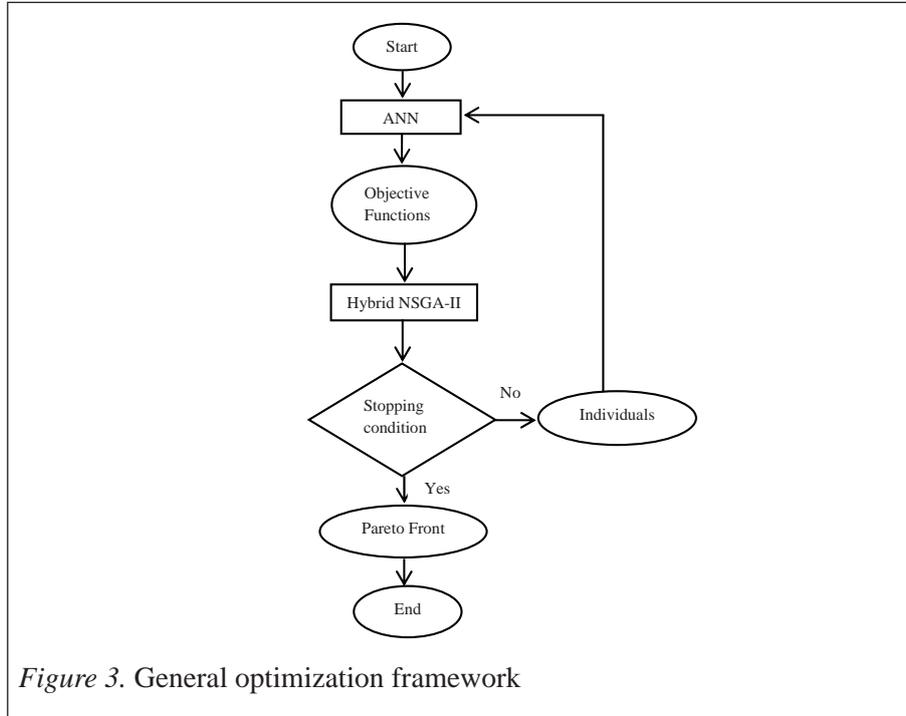


Figure 3. General optimization framework

### Experimental Setup

To be consistent with the literature (Sultan Noman Qasem et al., 2013; Sultan Noman Qasem, Shamsuddin, & Zain, 2012) the initial population size is set to 100. The crossover and mutation rates used are 0.90 and  $1/N$ , respectively, where the “ $N$ ” refers to the dimension of the individual. The maximum number of iterations is 1000. The fitness values are the hidden nodes and network training error or performance of the network. For the local search algorithm, the BP algorithm with the value of 0.01 is used, which is set to the learning rate and the number of iterations is 5.

### Performance Measurements

Regarding the measurements of the proposed method, statistical measures are used. Among these are sensitivity (sens) to identify the correct positive samples; this depends on the number of true positives and false negatives. Specificity (spec) predicts the correct negative samples, which depends on the number of true negatives and false positives. The accuracy is the measure to produce the level of accurate results for the proposed method. Equation (1) shows the calculation of these statistical measures as follows:

$$\begin{aligned}
 \text{Sensitivity} &= \frac{TP}{TP + FN}, \\
 \text{Specificity} &= \frac{TN}{TN + FP}, \\
 \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN}
 \end{aligned}
 \tag{1}$$

Where,  $TP$  is true positive,  $FP$  is false positive,  $TN$  is true negative and  $FN$  is false negative.

Table 1

*The Summary of the Datasets Used in the Experiments*

| Dataset | Number of features | Number of classes | Number of patterns |
|---------|--------------------|-------------------|--------------------|
| Iris    | 4                  | 3                 | 150                |
| Wine    | 13                 | 3                 | 178                |
| BTX     | 3                  | 7                 | 63                 |
| Segment | 19                 | 7                 | 2310               |
| Yeast   | 8                  | 10                | 1484               |

Source: (Asuncion & Newman, 2007).

## RESULTS AND DISCUSSION

In this section, the results of five datasets that used the proposed algorithm are compared with the results of other algorithms. In Table 2, the results of the training and testing error rates of the proposed algorithms are compared with the multiobjective genetic algorithm for the TBP network (MOGATTBP) (Ibrahim, Shamsuddin, & Bahiah, 2013) for all the datasets. The results show the generalization error of both methods. From Table 2, it can be clearly observed that in all the datasets, the proposed method gives more promising results in performance (training and testing error) than the other methods. Moreover, the training and testing error are the average of the errors obtained in a single run of the hybrid NSGA-II for the TBP network and they are reasonable error values that are seen as the average and standard deviation values in the same table.

The statistical results for sensitivity (Sens), specificity (Spec) and accuracy, which appear in Table 3, provide the detailed information about the proposed method compared with the MOGATTBP for the training and testing data. Among all the datasets used in this study, it can be seen from Table 3 that the

proposed method achieved better accuracy than the MOGATTBP. In addition, from the same Table, it can be clearly seen that the results of the proposed method for the iris data obtained an accuracy of 83.11%, 77.94% for wine data, 85.98% for BTX data, 86.93% for segment data and 90.02% for the yeast data.

Table 2

*The results of the Training and Testing Errors*

| Dataset |      | MOGATTBP       |               | Proposed Method |               |
|---------|------|----------------|---------------|-----------------|---------------|
|         |      | Training error | Testing error | Training error  | Testing error |
| Iris    | Mean | 0.1645         | 0.1654        | 0.1158          | 0.1179        |
|         | STD  | 0.0239         | 0.0224        | 0.0139          | 0.0148        |
| Wine    | Mean | 0.1686         | 0.1682        | 0.1311          | 0.1379        |
|         | STD  | 0.0394         | 0.0433        | 0.0249          | 0.0320        |
| BTX     | Mean | 0.1175         | 0.11853       | 0.1122          | 0.1132        |
|         | STD  | 0.0075         | 0.0067        | 0.0171          | 0.0165        |
| Segment | Mean | 0.0832         | 0.0914        | 0.0758          | 0.0871        |
|         | STD  | 0.0150         | 0.0095        | 0.0106          | 0.0050        |
| Yeast   | Mean | 0.0816         | 0.0816        | 0.0808          | 0.0807        |
|         | STD  | 0.0088         | 0.0088        | 0.0040          | 0.0043        |

Table 3

*The Results of Training and Testing Accuracy*

| Dataset | Methods |       | MOGATTBP |          |         |       | Proposed Method |       |         |          |       |       |       |
|---------|---------|-------|----------|----------|---------|-------|-----------------|-------|---------|----------|-------|-------|-------|
|         | Measure | Data  | Training |          | Testing |       | Training        |       | Testing |          |       |       |       |
|         |         | Sens  | Spec     | Accuracy | Sens    | Spec  | Accuracy        | Sens  | Spec    | Accuracy |       |       |       |
| Iris    | Mean    | 34.89 | 99.41    | 78.17    | 34.00   | 99.33 | 77.56           | 51.85 | 98.52   | 82.96    | 52.00 | 98.67 | 83.11 |
|         | STD     | 27.07 | 1.08     | 8.19     | 24.84   | 2.11  | 7.73            | 19.18 | 3.426   | 7.16     | 18.80 | 3.22  | 6.89  |
| Wine    | Mean    | 23.32 | 98.66    | 73.35    | 99.12   | 74.29 | 74.29           | 59.46 | 95.25   | 78.87    | 54.90 | 92.24 | 77.94 |
|         | STD     | 35.71 | 2.39     | 10.42    | 2.02    | 11.95 | 11.95           | 12.77 | 2.763   | 7.26     | 12.86 | 7.54  | 7.446 |
| BTX     | Mean    | 3.93  | 99.28    | 85.79    | 2.58    | 99.64 | 85.77           | 5.44  | 99.58   | 85.99    | 8.21  | 99.05 | 85.98 |
|         | STD     | 12.42 | 1.60     | 0.24     | 5.92    | 0.80  | 0.19            | 10.01 | 0.80    | 0.66     | 17.33 | 1.66  | 0.65  |
| Segment | Mean    | 0.00  | 100      | 85.71    | 0.00    | 100   | 85.71           | 4.42  | 99.84   | 86.95    | 3.65  | 99.65 | 86.93 |
|         | STD     | 0.00  | 0.00     | 0.00     | 0.00    | 0.00  | 0.00            | 3.94  | 0.38    | 1.03     | 3.19  | 0.70  | 1.05  |
| Yeast   | Mean    | 0.00  | 100      | 90.00    | 0.00    | 100   | 90.01           | 0.00  | 100     | 90.00    | 0.00  | 100   | 90.02 |
|         | STD     | 0.00  | 0.00     | 0.00     | 0.00    | 0.00  | 0.00            | 0.00  | 0.00    | 0.00     | 0.00  | 0.00  | 0.00  |

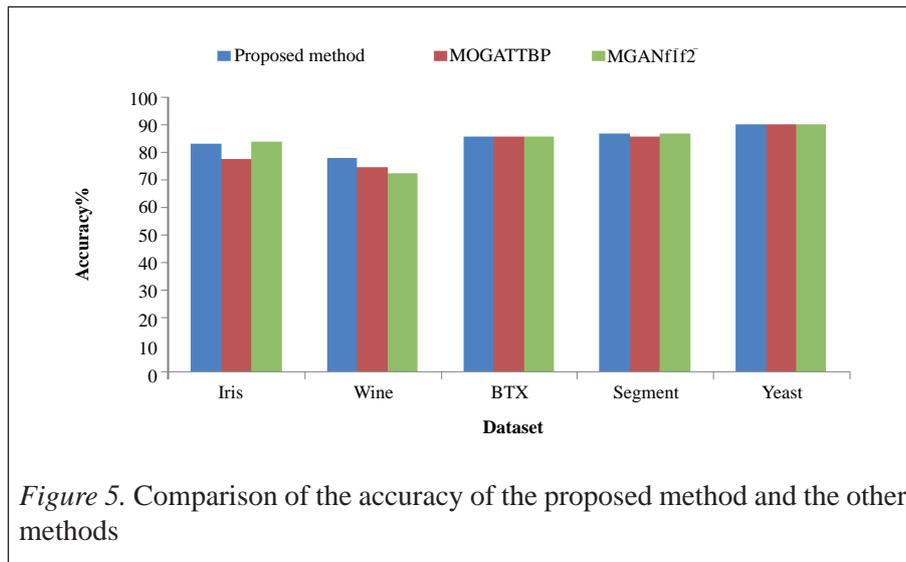
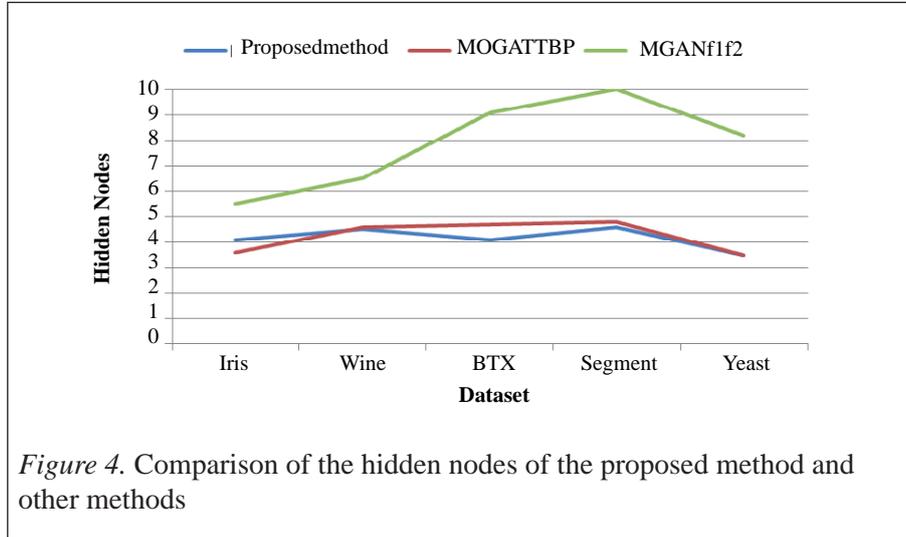
Similarly, for the sensitivity, the proposed method achieved 52.00% for iris, 54.90% for wine, 8.21% for the BTX data, 3.65% for segment and 0.00% for the yeast dataset. The sensitivity of the BTX, segment and yeast datasets is very difficult, due to their unbalanced data. Moreover, besides accuracy and sensitivity, Table 3 shows the specificity for all datasets, from which it can be noted that the specificity rates achieved are as follows: iris data 98.67%, wine data 92.24%, BTX data 99.05%, segment 99.65% and the yeast dataset 100%, which indicates the highest specificity result.

Table 4

*The Comparison of The Accuracy and Hidden Nodes Obtained by the Proposed Method and Other Methods*

| Methods | Proposed Method   |                      | MOGATTBP          |                      | MGANf1f2          |                      |
|---------|-------------------|----------------------|-------------------|----------------------|-------------------|----------------------|
|         | Accuracy<br>$f_1$ | Hidden<br>node $f_2$ | Accuracy<br>$f_1$ | Hidden<br>node $f_2$ | Accuracy<br>$f_1$ | Hidden<br>node $f_2$ |
| Iris    | 83.11             | 4.1                  | 77.56             | 3.6                  | 83.78             | 5.5                  |
| Wine    | 77.94             | 4.5                  | 74.29             | 4.6                  | 72.18             | 6.5                  |
| BTX     | 85.98             | 4.1                  | 85.77             | 4.7                  | 85.71             | 9.1                  |
| Segment | 86.93             | 4.6                  | 85.71             | 4.8                  | 86.90             | 10.0                 |
| Yeast   | 90.02             | 3.5                  | 90.01             | 3.5                  | 90.00             | 8.2                  |

It is clearly seen from Table 3, that the results are improved compared to the MOGATTBP for the training and testing data. In terms of accuracy, when both the algorithms are compared, it is seen that the proposed hybrid algorithm outperforms the MOGATTBP for all the datasets. After comparing the results for accuracy in Table 3, the results are reported in Table 4 in terms of the analysis results of the accuracy ( $f_1$ ), and the number of hidden nodes ( $f_2$ ). The results were compared to MGANf1f2 (Qasem et al., 2013), and MOGATTBP. In terms of the results for accuracy, the proposed method achieved a better result than the other methods for all the datasets except for the iris data, in which MGANf1f2 reported a better result. In respect of the complexity, the proposed method achieved better results than all the methods for all the datasets except for the iris data, in which the MOGATTBP obtained a better result. The same result was obtained for the yeast data for the same average of the number of hidden nodes by the proposed method and the MOGATTBP.



From Figure 4, the comparison of the proposed method and other methods for the complexity of the network results can be seen for all the datasets. Moreover, Figure 5 shows the comparison of the classification accuracy results of the proposed method and other methods. Regarding the complexity of the network, the proposed method is compared with MGANf1f2 and MOGATTBP. It can be observed from Figure 4, that the proposed method, with the blue colour line, indicates a less complex network than all the compared methods except for the iris data, for which the MOGATTBP obtained the best result.

## CONCLUSION

In the present paper, a hybrid NSGA-II to train the TBP network has been proposed to achieve optimization of the two objectives. First, is the accuracy of the network, while the other is the complexity of the network. The hybrid algorithm achieved a high accuracy with low complexity of the network. As a result, it can be said that when using the BP algorithm as a local search algorithm, it will improve the quality of the solutions of the NSGA-II in terms of the individuals in the population. From the experimental studies and graphic illustrations, it can be concluded that the proposed method improved the results of the MOGATTBP in terms of the number of hidden nodes and accuracy rate. Moreover, the experimental results of five multiclass data indicate that the proposed method demonstrated that it has the ability and effectiveness to deal with the classification problems.

## ACKNOWLEDGMENTS

This work is supported by the Ministry of Higher Education (MOHE) under the Fundamental Research Grant Scheme (FRGSJ130000.7828.4F347). The authors would like to thank the Research Management Center (RMC), Universiti Teknologi Malaysia (UTM) for the support in R & D, UTMJ Big Data Centre and the Soft Computing Research Group (SCRG) for the inspiration for making this study a success.

## REFERENCES

- Abbass, H. A. (2002). An evolutionary artificial neural networks approach for breast cancer diagnosis. *Artificial Intelligence in Medicine*, 25(3), 265-281.
- Abbass, H. A. (2003). Speeding up backpropagation using multiobjective evolutionary algorithms. *Neural Computation*, 15(11), 2705-2726.
- Adam, A., Chew, L. C., Shapiai, M. I., Jau, L. W., Ibrahim, Z., & Khalid, M. (2011). A hybrid artificial neural network-naive bayes for solving imbalanced dataset problems in semiconductor manufacturing test process. In *Hybrid Intelligent Systems (HIS), 2011 11th International Conference on* (pp. 133-138). IEEE.
- Ahmad, F., Isa, N. A. M., Hussain, Z., & Sulaiman, S. N. (2012). A genetic algorithm-based multi-objective optimization of an artificial neural network classifier for breast cancer diagnosis. *Neural Computing and Applications*, 23(5), 1427-1435.

- Ak, R., Li, Y., Vitelli, V., Zio, E., Droguett, E. L., & Jacinto, C. M. C. (2012). NSGA-II-Trained neural network approach to the estimation of prediction intervals of scale deposition rate in oil & gas equipment. *Expert Systems with Applications*, 40(4), 1205-1212.
- Almeida, L. M., & Ludermir, T. B. (2010). A multi-objective memetic and hybrid methodology for optimizing the parameters and performance of artificial neural networks. *Neurocomputing*, 73(7), 1438-1450.
- Asuncion, Arthur, & Newman, David. (2007). UCI machine learning repository.
- Bonissone, P. P., Chen, Y. T., Goebel, K., & Khedkar, P. S. (1999). Hybrid soft computing systems: industrial and commercial applications. *Proceedings of the IEEE*, 87(9), 1641-1667.
- Collins, M. (2002). Discriminative training methods for hidden Markov models: theory and experiments with perceptron algorithms. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1–8.
- Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., & Singer, Y. (2006). Online passive-aggressive algorithms. *The Journal of Machine Learning Research (JMLR)*, 7, 551–585.
- Cruz-Rami, M., Hervás-Martí, C., Fernández, J. C., Briceño, J., & de la Mata, M. (2012). Multiobjective evolutionary algorithm for donor-recipient decision system in liver transplants. *European Journal of Operational Research*, 222(2), 317-327.
- Dekel, O., Keshet, J., & Singer, Y. (2004). Large margin hierarchical classification. In *Proceedings of the 21st International Conference on Machine Learning (ICML)*, 27.
- Doddington, G., Mitchell, A., Przybocki, M., Ramshaw, L., Strassel, S., & Weischedel, R. (2004). The automatic content extraction (ACE) program—tasks, data, and evaluation. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, 837–840.
- Finkel, J., Grenager, T., & Manning, C. (2005). Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL)*, 363–370.
- Fleischman, M. (2001). Automated subcategorization of named entities. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL)*, 25–30.
- Fleischman, M., & Hovy, E. (2002). Fine-grained classification of named entities. In *Proceedings of the 19th international Conference on Computational Linguistics (COLING)*, 1–7.

- Florian, R., Ittycheriah, A., Jing, H., & Zhang, T. (2003). Named entity recognition through classifier combination. In *Proceedings of the 7th Conference on Natural Language Learning (CoNLL-2003)*, 168–171.
- Freund, Y., & Schapire, R. (1999). Large margin classification using the perceptron algorithm. *Machine learning*, 37(3), 277–296.
- Grishman, R., & Sundheim, B. (1996). Message understanding conference-6: a brief history. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, 466–471.
- Isozaki, H., & Kazawa, H. (2002). Efficient support vector classifiers for named entity recognition. In *Proceedings of the 19th International Conference on Computational Linguistics*, 1–7.
- Johansson, R. & Nugues, P. (2008). Dependency-based semantic role labeling of propbank. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 69–78.
- Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. A. M. T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *Evolutionary Computation, IEEE Transactions on*, 6(2), 182-197.
- Delgado, M., Cuéllar, M. P., & Pegalajar, M. C. (2008). Multiobjective hybrid optimization and training of recurrent neural networks. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 38(2), 381-403.
- Egrioglu, E., Aladag, C. H., & Yolcu, U. (2013). Fuzzy time series forecasting with a novel hybrid approach combining fuzzy c-means and neural networks. *Expert Systems with Applications*, 40(3), 854-857.
- Fan, C. Y., Chang, P. C., Lin, J. J., & Hsieh, J. C. (2011). A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. *Applied Soft Computing*, 11(1), 632-644.
- Fernandez, C., J. C., Martínez, F. J., Hervás, C., & Gutiérrez, P. A. (2010). Sensitivity versus accuracy in multiclass problems using memetic Pareto evolutionary neural networks. *Neural Networks, IEEE Transactions on*, 21(5), 750-770.
- Fernández, J. C., Hervás, C., Martínez-Estudillo, F. J., & Gutiérrez, P. A. (2011). Memetic Pareto Evolutionary Artificial Neural Networks to determine growth/no-growth in predictive microbiology. *Applied Soft Computing*, 11(1), 534-550.
- García-Pedrajas, N., Ortiz-Boyer, D., & Hervás-Martínez, C. (2004). Cooperative coevolution of generalized multi-layer perceptrons. *Neurocomputing*, 56, 257-283.
- García-Pedrajas, N., Hervás-Martínez, C., & Ortiz-Boyer, D. (2005). Cooperative coevolution of artificial neural network ensembles for pattern classification. *Evolutionary Computation, IEEE Transactions on*, 9(3), 271-302.

- Goh, C. K., Teoh, E. J., & Chen Tan, K. (2008). Hybrid multiobjective evolutionary design for artificial neural networks. *Neural Networks, IEEE Transactions on*, 19(9), 1531-1548.
- Hasan, S., Quo, T. S., & Shamsuddin, S. M. (2012). Artificial fish swarm optimization for multilayer network learning in classification problems. *Journal of Information & Communication Technology*, 11, 37-53.
- Ibrahim, A. O., Hasan, S., & Noman, S. (2014). Memetic Elitist Pareto evolutionary algorithm of three-term backpropagation network for classification problems. *Int. J. Advance Soft Compu. Appl*, 6(3).
- Ibrahim, A. O., Shamsuddin, S. M., Ahmad, N. B., & Qasem, S. N. (2013). Three-Term Backpropagation Network based on elitist multiobjective genetic algorithm for medical diseases diagnosis classification. *Life Science Journal*, 10(4), 1815-1822.
- Ibrahim, A. O., Shamsuddin, S. M., & Bahiah, N. (2013). Multi objective genetic algorithm for training three term backpropagation network. 32-38.
- Idris, I., Selamat, A., & Omatu, S. (2014). Hybrid email spam detection model with negative selection algorithm and differential evolution. *Engineering Applications of Artificial Intelligence*, 28, 97-110.
- Isa, N. A. M., Mashor, M. Y., Othman, N. H., & Zamli, K. Z. (2005). Application of artificial neural networks in the classification of cervical cells based on the Bethesda System. *Journal of Information & Communication Technology*, 4, 77-97.
- Ishibuchi, H., Hitotsuyanagi, Y., Tsukamoto, N., & Nojima, Y. (2009). Use of biased neighborhood structures in multiobjective memetic algorithms. *Soft Computing*, 13(8-9), 795-810.
- Jin, Y., Sendhoff, B., & Körner, E (2006). Simultaneous generation of accurate and interpretable neural network classifiers. In *Multi-Objective Machine Learning* (pp. 291-312): Springer Berlin Heidelberg.
- Karegowda, A. G., Manjunath, A. S., & Jayaram, M. A. (2011). Application of genetic algorithm optimized neural network connection weights for medical diagnosis of pima Indians diabetes. *International Journal on Soft Computing*, 2(2), 15-23.
- Khashei, M., Hamadani, A. Z., & Bijari, M. (2012). A novel hybrid classification model of artificial neural networks and multiple linear regression models. *Expert Systems with Applications*, 39(3), 2606-2620.
- Lara, A., Sanchez, G., Coello Coello, C. A., & Schutze, O. (2010). HCS: A new local search strategy for memetic multiobjective evolutionary algorithms. *Evolutionary Computation, IEEE Transactions on*, 14(1), 112-132.
- Neri, F., & Cotta, C. (2012). Memetic algorithms and memetic computing optimization: A literature review. *Swarm and Evolutionary Computation*, 2, 1-14.

- Pettersson, F., Chakraborti, N., & Saxén, H. (2007). A genetic algorithms based multi-objective neural net applied to noisy blast furnace data. *Applied Soft Computing*, 7(1), 387-397.
- Qasem, S. N., & Shamsuddin, S. M. (2011). Radial basis function network based on time variant multi-objective particle swarm optimization for medical diseases diagnosis. *Applied Soft Computing*, 11(1), 1427-1438.
- Qasem, S. N., Shamsuddin, S. M., Hashim, S. Z. M., Darus, M., & Al-Shammari, E. (2013). Memetic multiobjective particle swarm optimization-based radial basis function network for classification problems. *Information Sciences*, 239, 165-190.
- Qasem, S. N., Shamsuddin, S. M., & Zain, A. M. (2012). Multi-objective hybrid evolutionary algorithms for radial basis function neural network design. *Knowledge-Based Systems*, 27, 475-497.
- Qiongshuai, L., & Shiqing, W. (2011). A hybrid model of neural network and classification in wine. In *Computer Research and Development (ICCRD), 2011 3rd International Conference on* (Vol. 3, pp. 58-61). IEEE.
- Ramesh, S., Kannan, S., & Baskar, S. (2011). Application of modified NSGA-II algorithm to multi-objective reactive power planning. *Applied Soft Computing*, 12(2), 741-753.
- Srinivas, N., & Deb, K. (1994). Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary computation*, 2(3), 221-248.
- Tang, X. (2009). Hybrid Hidden Markov Model and artificial neural network for automatic speech recognition. Paper presented at the *Circuits, In Circuits, Communications and Systems, 2009. PACCS'09. Pacific-Asia Conference on* (pp. 682-685). IEEE .
- Wiegand, S., Igel, C., & Handmann, U. (2004). Evolutionary multi-objective optimisation of neural networks for face detection. *International Journal of Computational Intelligence and Applications*, 4(03), 237-253.
- Yao, X. (1999). Evolving artificial neural networks. *Proceedings of the IEEE*, 87(9), 1423-1447.
- Zhou, A., Qu, B. Y., Li, H., Zhao, S. Z., Suganthan, P. N., & Zhang, Q. (2011). Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm and Evolutionary Computation*, 1(1), 32-49.
- Zweiri, Y. H., Whidborne, J. F., & Seneviratne, L. D. (2003). A three-term backpropagation algorithm. *Neurocomputing*, 50, 305-318.