

Advancements in Multi-Layer Perceptron Training to Improve Classification Accuracy

K. Hemalatha
Research Scholar
Dept. of Computer Science
Sri Padmavati Mahila Visvavidyalayam
Tirupati, Andhra Pradesh
e-mail:hemalathakulala@gmail.com

K. Usha Rani
Professor
Dept. of Computer Science
Sri Padmavati Mahila Visvavidyalayam
Tirupati, Andhra Pradesh
e-mail: usharanikuruba@yahoo.co.in

Abstract— Neural Networks are the popular classification tools used in Medical diagnosis for early disease detection. The performance of Neural Networks is highly depended on the training process. In the training process, the individual weights between each of the neuron are adjusted for better classification results. Many Gradient-based and Meta-heuristic training algorithms are proposed and used by the researchers to improve the training performance of Neural Network. However, there are some limitations in both Gradient-based and Meta-heuristic algorithms when there are used individually. To overcome these limitations and to improve the Multi-Layer Perceptron Network performance Hybrid algorithms are useful. In this study, a review on advancements in Multi-Layer Perceptron Network training process for the improvement of classification performance is presented.

Keywords- Neural Networks, Meta-heuristic, Multi-Layer Perceptron, Training, Classification

I. INTRODUCTION

Neural Network (NN) is the model inspired by the biological neurons, the basic elements of the human brain. A biological neuron consists of Soma, Dendrites and Axon. Soma is the cell body where the cell nucleus is located. Dendrites are the tree like network made of nerve fiber connected to the cell body. An axon is a long connection extended to carry the signal from the neuron. The end of the axon splits and terminated into a small bulb like organ called Synapses. Through synapses neuron passes signals to another nearby neuron [1]. The structure of biological neuron [2] is presented in figure 1.

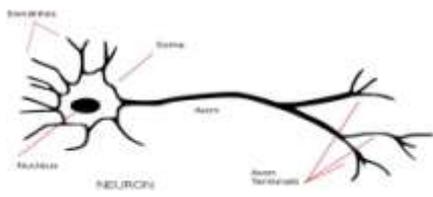


Figure 1. Structure of Biological Neuron

An artificial neuron processes and transmits the information as a biological neuron does. It can receive more than one input and perform necessary computations and generates the outputs. Basic structure of NN is shown in figure 2.

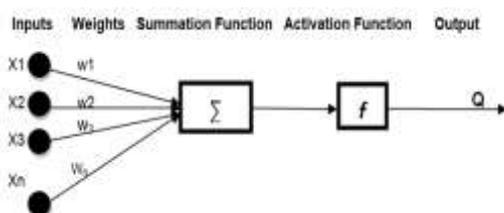


Figure 2. Basic Structure of Neural Network

Each input in the NN is multiplied by a weight value and the result becomes input for each neuron. Neuron uses transfer function to produce output for each layer. Along with input layer and output layer, hidden layers may present to perform necessary intermediate computations [3].

Neural Network forms many different simple and complex structures such as Feed Forward Neural Network, Recurrent Neural Network etc., to solve non-linear problems. Among them Feed Forward Neural Network is the well-known type of NN, in which inputs from neurons in the previous layer will be passed to the next layer after processing. Multi-Layer Perceptron (MLP) is a widely used Feed Forward NN. Many training methods are available to train MLP network and each method has both its own advantages and disadvantages [4].

II. MULTI-LAYER PERCEPTRON NERURAL NETWORK

Multi-Layer Perceptron is one of the most commonly used Neural Network architecture due to its lower complexity and ability to produce satisfactory result for non-linear relationships. This network is trained using supervised learning. A MLP usually consists of three or more layers: an input layer, one or more hidden layers and an output layer [5]. The general structure of MLP is presented in figure 3.

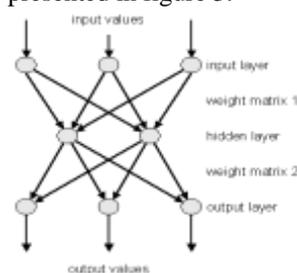


Figure 3. Structure of MLP Neural Network

The most significant elements of MLP are the connection weights and biases. The output of each node is calculated in two steps [6]. In first step, the weighted summation of the input is calculated using equation:

$$S_j = \sum_{i=1}^n w_{ij} I_i + \beta_j \quad (1)$$

where I_i is the input variable, w_{ij} is the connection weight between I_i and hidden neuron j , β_j is bias.

In second step, an activation function is used to generate the output of neurons based on the calculated weighted summation value. Different types of activation functions such as Logistic, Hyperbolic, Exponential and Sigmoid can be used in MLP. Sigmoid function is the most applied activation function in literature. The equation for sigmoid function is:

$$f_i(x) = \frac{1}{1+e^{-s_j}} \quad (2)$$

Once the output of each hidden neuron is calculated, the final output of MLP is calculated using equation:

$$y_k = \sum_{i=1}^m w_{kj} f_i + \beta_k \quad (3)$$

III. TRAINING OF NEURAL NETWORK

NNs are successfully implemented in Computer Aided Diagnosis (CAD) systems because of their capability to learn in an efficient manner. Learning algorithms to train are significantly influence the classification performance of NN. Training is the process in which the best set of weights to minimize the error is evaluated using different learning algorithms. Generally NN iteratively make changes in its weights and biases. NN becomes more knowledgeable after each iteration of training process. NN Training methods are classified into three types: Supervised, Unsupervised and Reinforced methods.

Supervised Method: In Supervised method a teaching element will guide the NN during training process.

Unsupervised Method: In Unsupervised method training process will be done in the absence of a teacher. No one will guide the training process in this method.

Reinforced Method: Reinforced method is similar to supervised method. In Supervised method the correct target and output values for each input pattern are known. But in Reinforced method less information is available. Not the exact information, only critic information will be available. Learning based on this critic information is called Reinforced method [7].

A. Training of MLP :

MLP is trained using Supervised Method. The classification performance of NN depends on selection of learning algorithms. In general learning methods are mainly classified into Gradient-based and Meta-heuristic search methods.

1) *Gradient-based algorithms:* Gradient-based algorithm involves computing the gradient of some form of the performance measure for the network in weight space either exactly or approximately. By using the calculated gradient the

weight changes are determined with the suitable approach. In this approach, the performance measure is the error between actual and desired output. Gradient-based algorithms are widely used in Back Propagation algorithms [8].

Training of MLP using Back Propagation algorithm includes two phases: Propagation and Updation of weights. In first phase the information presented in input vector is propagated forward through the layers of the network. The error values between actual output and desired output is evaluated and these errors are propagated backwards. Gradient value is calculated using these error values. In second phase the weights are updated by the optimization of the calculated gradient. Optimization is the process to achieve the best outcome of a given operation while satisfying certain conditions. The training process will be continued until maximum number of epochs or the minimum acceptable error reached. After completion of the training process, testing on another set of inputs will be done to determine how well Back Propagation generalized on the untrained inputs. The performance is evaluated generally using the measures such as Mean Squared Error (MSE) or Root Mean Squared Error (RMSE). MSE and RMSE values can be calculated with the Equations:

$$MSE = \frac{\sum t_i - o_i^2}{n} \quad (4)$$

and

$$RMSE = \sqrt{MSE} \quad (5)$$

Where t_i is the required output, o_i is the actual output and n is the total number of samples in the dataset.

Different types of Back Propagation algorithms can be applied to train MLP network. In this section the theoretical background of widely used training algorithms [9] such as Levenberg-Marquardt, Quasi-Newton, Resilient Back Propagation, Conjugate gradient and Gradient descent with variable learning rate are presented.

Levenberg-Marquardt (LM): Levenberg-Marquardt algorithm is the evolution of Newton's algorithm [10] given by:

$$w_{k+1} = w_k - H_k^{-1} g_k \quad (6)$$

Where w is the weight matrix, H is Hessian matrix and g is the gradient. The main objective of the LM algorithm is updating weights. For weights updation hessian matrix should be calculated. Hessian matrix is calculated by the second order derivative of total error function which is complicated. Therefore, instead of calculating Hessian matrix, Jacobian matrix is introduced and the method is known as Gauss-Newton method. The advantage of this method is its convergence about minimum and providing more accurate results. LM is the fastest training algorithm but it requires more memory compared with other training methods [11].

Quasi-Newton (QN): The computation of second derivatives is complicated in Newton's method. To overcome this complexity, the Quasi-Newton method is developed. In QN method no need to calculate second derivatives. The principle of this algorithm is based on the updation of an approximate Hessian matrix at each iteration [12]. Faster computation time and no need of iteratively solve linear system of equations are the main advantages of QN algorithm.

Resilient Back Propagation (RBP): Sigmoid activation function is the generally used activation function for hidden layer of feed forward multilayer neural networks. In sigmoid function the input values takes large value as the slope come closer to zero. When steepest descent is used to train MLP with sigmoid activation function a problem occurs. The magnitude of the gradient takes very small value. Therefore, small changes in weights occur even though the weights and biases are far from its optimal value. Hence, it is hard to train the network. To overcome this problem, Resilient Back Propagation algorithm was generated. RBP removed the above mentioned unwanted effects by considering the direction used instead of gradient [13].

Gradient descent with variable learning rate: In the training process, the learning rate is fixed with standard steepest descent. The high performance of learning in the training also depends on the learning rate. Too high or too low learning rate may take long time for the convergence. In practice, it is very hard to decide the best value for leaning rate factor. If larger learning rate gives stable results then the learning rate parameter can be increased [14].

Generally selection of training algorithm depends on factors such as structure of the network, the number of hidden layers, weights and biases, Error rate of learning etc. Slow convergence, high dependency on initial parameters and tendency to be trapped in local minima are the limitations of the Gradient-Based algorithms [15]. By considering the advantages and limitations of each Back Propagation Training algorithm the suitable one may be selected for a particular problem.

2) *Meta-heuristic algorithms:* Meta-heuristic algorithms are the procedures that guide heuristic techniques such as construction based and local search based approaches in search space domain. Meta-heuristic algorithms were proposed by the researchers as alternatives to the Gradient-based algorithms for MLP training. Efficient implementation, fine-tuning and faster convergence are some of the advantages of Gradient-based algorithms. However, these methods are also facing local minima issue. To overcome the local minima problem Meta-heuristic algorithms were developed. The main aim of these algorithms is to go for global optimum. In general Meta-heuristic methods improve the learning process of NN than traditional Gradient-based algorithms. For certain types of NNs these methods are not applicable. But most of the Meta-heuristic optimization methods can be applied on MLP [16].

Evolutionary and Swarm - based algorithms are the two main families of Meta-heuristic algorithms. Evolutionary algorithms incorporate randomness as the main mechanism to generate, evolve and update population based solutions until an adequate solution is found or maximum number of iterations is reached. Swarm-based algorithms are inspired by the nature such as movements of birds, insects and other creatures in nature. These methods update the generated random solutions using some mathematical models until an optimal solution is found. Many Evolutionary algorithms and Swarm-based algorithms were used to train MLP. In this section, some of the Meta-heuristic algorithm methodologies are presented.

Genetic Algorithms (GA): Genetic algorithm is an evolutionary algorithm inspired by the Darwin's theory of evolution and natural selection. GA solves the optimization problems with the help of population of chromosomes. It randomly generates an initial population of chromosome which serves as parents. Once the operation started the current population of chromosomes is replaced by the new population of chromosomes until the stopping condition is satisfied [17]. In each iteration the population evolves best solution. The evolution occurs by performing operations such as Selection, Crossover, Mutation and Replacement. In Selection operation two parents are selected based on fitness value. Genes are updated randomly in Mutation operation. Crossover operation combines the distinct features from parents. Replacement operator eliminates the inferior chromosomes for better solution [18].

Particle Swarm Optimization (PSO): Particle Swarm Optimization is motivated by the behavior of birds flocking proposed to explore optimal solution. It can be implemented and applied successfully to solve various optimization problems. It consists of particles population in search space. Each potential solution associated with random velocity is called as particle in the swarm. These particles are located randomly when the swarm moves to find the optimal fitness value. Based on the fitness value the previous best solution for the particle and overall best solution for the swarm are calculated after every iteration. The particles in the problem space will fly by following the best solutions so far [19]. Achieving high-quality solutions in less time, flexibility in balancing global and local exploration and easy implementation are the main advantages of PSO. PSO also maintain the track of previous best solutions to avoid the loss of previously learned knowledge [20].

Ant Colony Optimization (ACO): Ant Colony Optimization (ACO) is originated from the food foraging behavior of ants. It can be used to solve multi-objective problems with high computation complexities. To determine the shortest path a moving ant lay the pheromone which acts as a sign for other ants to follow. The higher probability to follow it is calculated. The emergence of collective behavior forms the best path for ants to follow and make pheromone more stable for transferring the food back to the nest [21]. The main idea of ACO inspired by the behavior ants is a parallel search over several constructive computational processes. This process depends on the local problem and dynamic memory structure containing the information about the quality of previously obtained result [22].

Firefly Algorithm (FA): Firefly Algorithm is the recent swarm intelligence based Meta-heuristic algorithm that has successfully solved many optimization problems. FA is based on the bioluminescent communication behavior of fireflies. It mainly depends on the light intensity changes occur in fireflies. Different flashing intensities show various indications such as attracting mating partners, attracting hunt, etc [23]. FA is modeled by the level of interaction between fireflies. Each fly has its own attractiveness to attract other fireflies. Each firefly is attracted to the one which has higher glow. FA capability in dealing complex problems with no or less knowledge of the

search space made it powerful swarm algorithm. Flexible structure, clear and simple modeling are the important advantages of FA. It also includes the significant features adopted from the existed evolutionary approaches such as a random walk strategy, a simulated annealing convergence strategy and movement rules to increase FA effectiveness so that premature convergence will be avoided [24].

IV. APPROACHES TO IMPROVE MLP PERFORMANCE

Efficiency of NN is highly influenced by its learning process. The goal of learning process in MLP network is to find the best set of weights that can minimize the classification or prediction error. In this section few studies on Gradient - based, Meta-heuristic and Hybrid algorithms to improve the MLP performance are presented.

A. Gradient-based algorithms:

Mohammed Sarhan et.al [25] mentioned that existing studies do not mention the relationship between training rate and momentum with gross weight that reduces the training accuracy. So, they suggested the creation of dynamic training rate with boundary and momentum terms with an inverse relationship between them to escape gross weight training to maintain high training accuracy.

Hamed Azami et.al [4] used an NN ensemble, which is a learning concept in which a number of NN outputs are combined to improve the accuracy of classification. Training algorithms such as BP with adaptive learning rate, BP with adaptive learning rate and momentum, Polak–Ribikre Conjugate Gradient Algorithm (CGA), Fletcher-Reeves CGA, Powell–Beale CGA, scaled CGA, Resilient BP (RBP) and One Step Secant and Quasi-Newton methods are used to develop NN Ensemble. The classification accuracy is improved using NN ensemble approach.

N.V. Saiteja Reddy et.al [26] analyzed the performance of Back Propagation algorithm with and without bias thresholds. The algorithm with bias threshold performed more efficiently than the BP algorithm without threshold. They concluded that practically it is very difficult to get a good NN topology just by using the number of inputs and outputs. Further they stated that the requirement of number of hidden layer neurons depends on the type of the dataset samples and unnecessary increase in number of hidden layer may causes over-fitting problem.

B. Meta-heuristic algorithms:

Emina et.al [27] used GA to improve the classification accuracy on Breast Cancer Dataset. They showed that GA is able to increase the training efficiency of NN.

Sankhadeep Chatterjee et.al [28] used PSO based NN for the Dengue Fever classification to improve the classification performance of MLP.

Christian Blum et.al [29] trained Feed-Forward Neural Network (FFNN) with Ant Colony Optimization algorithm. In that study ACO algorithm outperformed to improve the classification performance of FFNN.

Hossam Faris et.al [15] trained Feed-Forward Neural Network with Lighting Search Algorithm (LSA) to overcome the local minima problem occurs in Neural Network. LSA-based training classified 16 standard datasets efficiently. It showed the fast convergence speed.

Ibrahim Aljarah et.al [6] proposed new training algorithm based on Whale Optimization Algorithm (WOA) to train MLP Neural Network. WOA utilizes a population of search agents to determine the global optimum for optimization problems. In this study WOA is applied to train MLP with single hidden layer. The performance of proposed algorithms is tested with 20 standard datasets. WOA-based trainer performed better on majority of datasets than existing algorithms. Along with higher accuracy WOA-based trainer showed fast convergence speed.

C. Hybrid Training algorithms:

According to the No-Free-Lunch (NFL) theorem [30] no optimization technique can solve all optimization problems. So, new optimization techniques should be developed for training MLP. This motivation leads to develop new hybrid techniques in the literature for MLP training.

Artificial Bee Colony (ABC) algorithm was proposed based on the Honey Bee Swarms intelligent behavior in searching and to obtain solutions for problems related to optimization and classification. Habib Shah et.al [31] improved the training efficiency of MLP with a new hybrid technique Global Artificial Bee Colony-Levenberg-Marquardt (GABC-LM) algorithm. GABC-LM performed better and produced prominent results.

Fatma Mazen et.al [32] introduced a new hybrid approach called Genetic Algorithm based Firefly algorithm for training MLP neural network. This approach optimized the weights between layers and biases of neurons to minimize the fitness function. This simulation increased the performance of MLP than existed methods.

Waheed Ali et.al [33] developed a hybrid technique which is a combination of ABC and PSO algorithms to train FFNN. To get rid of imperfections occurred in traditional training algorithms this hybrid approach is used. The hybridization of ABC and PSO performed well than ABC and PSO.

V. CONCLUSION

In this paper the advancements in Gradient-based and Meta-heuristic training algorithms to increase the MLP performance are presented. Local minima is the major issue occurred when MLP is trained with Traditional Back Propagation algorithms such as Levenberg-Marquardt, Quasi-Newton, Resilient Back Propagation, Conjugate gradient, etc. A review on Meta-heuristic algorithms such as Genetic Algorithm, Particle Swarm Algorithm, and Artificial Bee Colony which are used to train MLP network with improved classification accuracy to overcome the local minima issue is presented. Further, the role of Hybridization of Gradient-based and Meta-heuristic training algorithms to improve MLP performance is studied in this paper. From this study it can be concluded that Hybrid training algorithms may improve the MLP classification performance significantly than the individual training methods.

REFERENCES

- [1] Sivanandam S N, S.N. Deepa, Principles of Soft Computing, Wiley India Edition, New Delhi, 2007
- [2] <http://durofy.com/machine-learning-introduction-to-the-artificial-neural-network/neuron-2>

- [3] Ishani Arora, Anju Saha, Comparison of Back Propagation Training Algorithms for Software Defect Prediction, 2016
- [4] Hamed Azami, Javier Escudero, A Comparative Study of Breast Cancer Diagnosis based on Neural Network Ensemble via Improved Training Algorithms, 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Milan, Italy, 26/08/15 - 29/08/15, pp. 2836-2839, 2015
- [5] Azar, Ahmad Taher, and Shaimaa Ahmed El-Said, Probabilistic neural network for breast cancer classification. *Neural Computing and Applications*, Springer, VOL:23, Issue:6, PP: 1737-1751, 2013.
- [6] Ibrahim Aljarah, Hossam Faris, Seyedali Mirjalili, Optimizing connection weights in neural networks using the whale optimization algorithm, *Soft Comput* DOI 10.1007/s00500-016-2442-1, 2016
- [7] Vaibhav Kant Singh, Mathematical Analysis for Training ANNs Using Basic Learning Algorithms, *Research Journal of Computer and Information Technology Sciences*, Vol. 4(7), 6-13, 2016
- [8] Rumelhart DE, Hinton GE, Williams RJ, *Neurocomputing: Foundations of research*. MIT Press, Cambridge, MA, USA, chap Learning Representations by Back-propagating Errors, pp 696–699, 1988
- [9] Yavuz Sari, Performance Evaluation of the Various Training Algorithms and Network Topologies in a Neural-network-based Inverse Kinematics Solution for Robots, *International Journal of Advanced Robot Systems*, 2014
- [10] Mohamad N, F. Zaini, A. Johari, I. Yassin, and A. Zabidi, Comparison between Levenberg-Marquardt and Scaled Conjugate Gradient Training Algorithms for Breast Cancer Diagnosis Using MLP, in *Signal Processing and Its Applications (CSPA)*, 2010 6th International Colloquium on, May, pp. 1–7, 2010
- [11] Hagan MT, Menhaj MB, Training feed forward networks with the Marquardt algorithm. *IEEE Trans Neural Networks*, 5(6), 989–93, 1994
- [12] Rob Haelterman, Alfred Bogaers, Joris Degroote, Silviu Cracana, Coupling of Partitioned Physics Codes with Quasi-Newton Methods, 2017
- [13] Riedmiller M, Braun H., A direct adaptive method for faster back propagation learning: The RPROP algorithm. In: *Proceedings of the IEEE international conference on neural networks*. p. 896, 1993
- [14] Rasit Koker, Necat Altinkok, Adem Demir, Neural network based prediction of mechanical properties of particulate reinforced metal matrix composites using various training algorithms, *Materials and Design* 28, 616–627, 2007
- [15] Hossam Faris, Seyedali Mirjalili, Nailah Al-Madi, Optimizing the Learning Process of Feedforward Neural Networks Using Lightning Search Algorithm, *International Journal on Artificial Intelligence Tools*, Vol. 25, No. 6, 2016
- [16] Gandomi M, Soltanpour M, Zolfaghari MR, Gandomi AH, Prediction of peak ground acceleration of Iran's tectonic regions using a hybrid soft computing technique, *Geosci Front*. <http://www.sciencedirect.com/science/article/pii/S1674987114001376>, 2014
- [17] Cheng-Hong Yang, Yu-Da Lin, Li-Yeh Chuang, Hsueh-Wei Chang, Evaluation of Breast Cancer Susceptibility Using Improved Genetic Algorithms to Generate Genotype SNP Barcodes, *IEEE-ACM Transactions On Computational Biology And Bioinformatics*, 2017
- [18] Fahimeh Mostofi, Fahreddin Sadikoglu, Discovering SNP interactions associated with breast cancer using evolutionary algorithms, *Procedia Computer Science* 102, 562 – 56, 2016
- [19] Chellamuthu Gunavathi, Kandasamy Premalatha, A Comparative Analysis of Swarm Intelligence Techniques for Feature Selection in Cancer Classification, Hindawi Publishing Corporation, *The Scientific World Journal*, 2014
- [20] Rui Xu, Georgios C. Anagnostopoulos, and Donald C. Wunsch II, Multiclass Cancer Classification Using Semisupervised Ellipsoid ARTMAP and Particle Swarm Optimization with Gene Expression Data, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, VOL. 4(1), 2007
- [21] Anand Nayyar, Rajeshwar Singh, Simulation and Performance Comparison of Ant Colony Optimization (ACO) Routing Protocol with AODV, DSDV, DSR Routing Protocols of Wireless Sensor Networks using NS-2 Simulator, *American Journal of Intelligent Systems*, 2017
- [22] Sapna Katiyar, Ibraheem, Abdul Quaiyum Ansari, Ant Colony Optimization: A Tutorial Review, Conference Paper, 2015
- [23] Kemal Keskin, Abdurrahman Karamancioglu, Application of Firefly Algorithm to Train Operation, 2016 IEEE 8th International Conference on Intelligent Systems, 2016
- [24] Ilyes Khennak, Habiba Drias, A Firefly Algorithm-based Approach for Pseudo-Relevance Feedback: Application to Medical Database, *Journal of Medical Systems*, 2016
- [25] Mohammed Sarhan Al_Duais, F. S. Mohamad, A Review on Enhancements to Speed up Training of the Batch Back Propagation Algorithm, *Indian Journal of Science and Technology*, Vol 9(46), 2016
- [26] Saiteja Reddy N.V., T. Srikanth, Class Label Prediction using Back Propagation Algorithm: A comparative study with and without Thresholds (Bias), *International Journal of Computer Sciences and Engineering*, Vol 3 (7), 2015
- [27] Emina Alickovic, Abdulhamit Subasi, Breast cancer diagnosis using GA feature selection and Rotation Forest, *Neural Comput & Applic*, DOI 10.1007/s00521-015-2103-9, 2015
- [28] Sankhadeep Chatterjee, Sirshendu Hore, Nilanjan Dey, Amira S. Ashour, Dengue Fever Classification Using Gene Expression Data: A PSO Based Artificial Neural Network Approach, *Advances in Intelligent Systems and Computing*, 2017
- [29] Christian Blum, Krzysztof Socha, Training Feed-Forward Neural Networks with Ant Colony Optimization: An Application To Pattern Classification, IRIDIA Technical Report Series Technical Report No. TR/IRIDIA/2005-038, 2005
- [30] Wolpert DH, Macready WG, No free lunch theorems for optimization. *IEEE Trans Evol Comput* 1(1):67–82, 1997
- [31] Habib Shah, Rozaida Ghazali, and Nazri Mohd Nawi and Mustafa Mat Deris, Global Artificial Bee Colony-Levenberg-Marquardt Algorithm (GABC-LM) Algorithm For Classification, DOI: 10.4018 / jaec. 2013070106, 2013
- [32] Fatma Mazen, Rania Ahmed AbulSeoud, Amr M. Gody, Genetic Algorithm and Firefly Algorithm in a Hybrid Approach for Breast Cancer Diagnosis, *International Journal of Computer Trends and Technology (IJCTT) – Volume 32 Number 2*, 2016
- [33] Waheed Ali H, M. Ghanem, Aman Jantan, Using Hybrid Artificial Bee Colony Algorithm and Particle Swarm Optimization for Training Feed-Forward Neural Networks, *Journal of Theoretical and Applied Information Technology*, Vol.67, No.3, 2014