AN IMPROVED ARTIFICIAL BEE COLONY ALGORITHM FOR TRAINING MULTILAYER PERCEPTRON IN TIME SERIES PREDICTION

HABIB SHAH

A thesis submitted in

fulfilment of the requirements for the award of the

Doctor of Philosophy in Information Technology

Faculty of Computer Science and Information Technology
University Tun Hussein Onn Malaysia

MARCH 2014

ABSTRACT

Learning an Artificial Neural Network (ANN) is an optimization task since it is desirable to find optimal weight sets of an ANN in the training process. Different equations are used to guide the network for providing an accurate result with less training and testing error. Most of the training algorithms focus on weight values, activation functions, and network structures for providing optimal outputs. Backpropagation (BP) learning algorithm is the well-known learning technique that trained ANN. However, some difficulties arise where the BP cannot get achievements without trapping in local minima and converge very slow in the solution space. Therefore, to overcome the trapping difficulties, slow convergence and difficulties in finding optimal weight values, three improved Artificial Bee Colony (ABC) algorithms built on the social insect behavior are proposed in this research for training ANN, namely the widely used Multilayer Perceptron (MLP). Here, three improved learning approaches inspired by artificial honey bee's behavior are used to train MLP. They are: Global Guided Artificial Bee Colony (GGABC), Improved Gbest Guided Artificial Bee Colony (IGGABC) and Artificial Smart Bee Colony (ASBC) algorithm. These improved algorithms were used to increase the exploration, exploitation and keep them balance for getting optimal results for a given task. Furthermore, here these algorithms used to train the MLP on two tasks; the seismic event's prediction and Boolean function classification. The simulation results of the MLP trained with improved algorithms were compared with that when trained with the standard BP, ABC, Global ABC and Particle Swarm Optimization algorithm. From the experimental analysis, the proposed improved algorithms get better the classification efficacy for time series prediction and Boolean function classification. Moreover, these improved algorithm's success to get high accuracy and optimize the best network's weight values for training the MLP.

ABSTRAK

Mempelajari Rangkaian Neural Buatan (ANN) merupakan satu tugas optimimasi bagi mencari set pemberat optima bagi ANN dalam proses latihan. Pelbagai algoritma digunakan bagi melatih rangkaian ini dalam menghasilkan keputusan yang jitu dengan pengurangan ralat latihan dan ujian. Kebanyakan algoritma pembelajaran mengfokus kepada nilai pemberat, fungsi pengaktifan, dan struktur rangkaian bagi menghasilkan keputusan yang optima. Algoritma Rambatan Balik (BP) merupakan satu teknik yang popular bagi melatih ANN. Walau bagaimanapun, beberapa masalah timbul apabila BP mudah terperangkap di dalam minima lokal dan menumpu secara sangat lambat dalam ruangan solusi. Oleh itu, bagi mengatasi masalah terperangkap dalam minima lokal, penumpuan yang lambat, serta kesukaran mencari nilai pemberat optima, maka tiga algoritma baru dibina bagi menambahbaik Artificial Bee Colony (ABC) yang berpandukan kepada perlakuan sosial serangga untuk melatih ANN; yang dinamakan Perseptron Multiaras (MLP). Di sini, tiga kaedah pelatihan terbaru ilham perangai lebah madu buatan dihasilkan bagi melatih MLP. Ini terdiri daripada algoritma *Global Guided* Artificial Bee Colony (GGABC), Improved Gbest Guided Artificial Bee Colony (IGGABC) dan Artificial Smart Bee Colony (ASBC). Algoritma yang ditambaik ini digunakan untuk menaikkan mutu pencarian, eksploitasi dan memberi keseimbangan bagi mendapatkan keputusan optima bagi kes tertentu. Untuk itu, algoritma ini digunakan untuk melatih MLP di dalam dua kes khusus, iaitu jangkaan peristiwa seismik dan klasifikasi fungsi Boolean. Keputusan dari simulasi MLP yang dilatih dengan algoritma yang ditambahbaik ini telah dibandingkan dengan kes yang dilatih melalui algoritma piawai iaitu BP, ABC, Global ABC, dan *Particle Swarm Optimization*. Dari analisa eksperimen, algoritma yang dicadangkan ini telah menunjukkan bukti penambahbaikan dari segi keberkesanan efikasi untuk fungsi Boolean dan jangkaan siri masa. Malah algoritma ini juga berjaya menghasilkan ketepatan yang tinggi dan nilai pemberat rangkaian yang optima bagi melatih MLP.

TABLE OF CONTENTS

LIST OF AWARDS

1. **Gold Medal in Research Festival 2012:** Global Artificial Bee Colony Algorithm for Training Neural Networks.

2. **Bronze Medal in Research and Innovation Competition 2011:** Hybrid Ant Bee Colony Algorithm for Training Neural Networks.

3. **First Position In Research Poster Competition At 3rd International Recognition Seminar 2013:** An Efficient Meta Heuristic Algorithms for Classification and Prediction Tasks.

4. **Third Position In Three Minute Thesis (3 Mt) Presentation At 3rd International Recognition Seminar 2013:** An Efficient Meta Heuristic Learning Algorithm for Seismic Prediction.

5. **Gold Medal in Research and Innovation Festival 2013:** 3G Meta Heuristic Learning Algorithm

6. **Best Applied Science Research Award in Research and Innovation Festival 2013:** 3G Meta Heuristic Learning Algorithm.

LIST OF JOURNAL PUBLICATIONS

1. **Habib Shah**, **Rozaida Ghazali and Nazri Mohd Nawi**"Using Artificial Bee Colony Algorithm for MLP Training on Earthquake Time Series Data Prediction" Journal of Computing (JOC 2011). Vol. 3, Issue 6, pp : 135-142.

2. **Habib Shah**, **Rozaida Ghazali and Nazri Mohd Nawi,** "Boolean Function classification using Hybrid Ant Bee Colony Algorithm". Journal on Computer Science & Computational Mathematics, JCSCM,Vol.2 Issue.11, 2012.

3. **Habib Shah**, **Rozaida Ghazali and Nazri Mohd Nawi, "**G-HABC Algorithm for Training Artificial Neural Networks": *International Journal of Applied Metaheuristic Computing.*Vol.3, Issue3, pp.1-19. 2012

4. **Habib Shah**, **Rozaida Ghazali and Nazri Mohd Nawi "**Global Artificial Bee Colony-Levenberq-Marquardt Algorithm (GABC-LM) Algorithm for Classification*": International Journal of Applied Evolutionary Computation, IGI (IJAEC)* 4(3), pp. 58-74, 2013.

5. **Habib Shah**, **Rozaida Ghazali and Nazri Mohd Nawi, "**Hybrid Global Artificial Bee Colony Algorithm for Prediction Task", *Journal of Applied Sciences Research,* 9(5), pp. 3328-3337, 2013.

6. **Habib Shah, Rozaida Ghazali and Yana Mazwin Mohmad Hassim,** 2G Meta Heuristic Learning Algorithm for Earthquake Time Series Data Prediction. Journal of Applied Sciences Research 9(5): pp. 3328-3337, 2013.

LIST OF CONFERENCE PROCEEDINGS

1. **Habib Shah**, **Rozaida Ghazali and Nazri Mohd Nawi,** “Hybrid Bee Ant Colony Algorithm (HBAC) for training Multilayer perceptron”, 2nd World IT conference Turkey (WCIT 2011).
2. **Habib Shah and Rozaida Ghazali**,"Prediction of Earthquake Magnitude by an Improved ABC-MLP," Developments in E-systems Engineering (DeSE), pp. 312-317, 6-8 Dec. 2011.
3. **Habib Shah**, **Rozaida Ghazali and Nazri Mohd Nawi,** “Artificial Bee Colony Algorithm for Predicting Tsunami Intensity” Malaysian Technical Universities International Conference on Engineering & Technology, proceeding (MUiCET 2011).
4. **Habib Shah**, **Rozaida Ghazali and Nazri Mohd Nawi,** “Artificial Bee Colony Algorithm for Training the Multilayer Perceptron to Predict the Tsunami from Undersea Earthquake Seismic Signals” Post Graduate Seminar 2010 UTHM Proceeding.
5. **Habib Shah**, **Rozaida Ghazali and Nazri Mohd Nawi,** “Hybrid Ant Bee Colony Algorithm for classification and Prediction Mission”. Proceeding in Conference on Computer Science & Computational Mathematics Melaka , Malaysia (CCSCM 2012).
6. **Habib Shah**, **Rozaida Ghazali and Nazri Mohd Nawi**, (2012)"Hybrid Ant Bee Colony Algorithm for Volcano Temperature Prediction Emerging Trends and Applications in Information Communication Technologies." vol. 281, B. S. Chowdhry, *et al*., Eds., ed: Springer Berlin Heidelberg, 2012, pp. 453-465.

7. **Habib Shah**, **Rozaida Ghazali, Nazri Mohd Nawi**, **and Mustafa Mat Deris**, "Global Hybrid Ant Bee Colony Algorithm for Training Artificial Neural Networks Computational Science and Its Applications – ICCSA 2012. In B. Murgante, O. Gervasi, S. Misra, N. Nedjah, A. Rocha, D. Taniar & B. Apduhan (Eds.), (Vol. 7333, pp. 87-100): Springer Berlin Heidelberg.
8. **Habib Shah**, **Rozaida Ghazali and Nazri Mohd Nawi, "**Global **A**rtificial Bee Colony Algorithm for Boolean Function Classification", In A. Selamat, N. Nguyen & H. Haron (Eds.), *Intelligent Information and Database Systems* (Vol. 7802, pp. 12-20): Springer Berlin Heidelberg.
9. **Habib Shah, Rozaida Ghazali and Nazri Mohd Nawi,** Hybrid Global **A**rtificial Bee Colony Algorithm for Training Artificial Neural Networks, The 15th International Multi Topic Conference (INMIC 2012) Islamabad, Pakistan, Accepted.
10. **Habib Shah, Rozaida Ghazali and Nazri Mohd Nawi,** Hybrid Guided **A**rtificial Bee Colony Algorithm for Earthquake Time Series Data Prediction, International Multi-Topic Conference (IMTIC' 13), Mehran University, Jamshoro, Sindh, Pakistan, (To be published in spirnger).
11. **Habib Shah, Rozaida Ghazali and Nazri Mohd Nawi,** Honey Bees Inspired Learning Algorithm: Nature Intelligence Can Predict Natural Disaster, Accepted in SCDM-2014 (To be published in spirnger).

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

$a_{net,j}$	-	Net input activation function for the j^{th} unit
c_j	-	Weighting factor,
D	-	Dimension
FS	-	Food Sources
$maxIter$	-	Maximum iteration number.
L	-	lower bound
$rand$	-	Random number generator built-in function of Matlab
n	-	Number of nodes in the output layer
n	-	total number of input patterns
t_k	-	Desired output of the k^{th} output unit
O_b	-	Onlookers Bees
O_k	-	Network output of the k^{th} output unit
O_j	-	Output of the j^{th} unit
O_i	-	Output of the i^{th} unit
s_i^k	-	Current position of agent i at iteration k,
T_i	-	Predicted data
U	-	Upper Bound
v_i^k	-	Velocity of agent *i* at iteration *k*,
w	-	Weighting function,
$wMax$	-	Initial weight,
$wMin$	-	Final weight, and
w_{ij}	-	Weight of the link from unit i to unit j
w_{jk}	-	Weight of the link from unit j to unit k
x_i	-	Number of input patterns
y_i	-	predicted patterns
θ_j	-	Bias for the j^{th} unit
η_{ij}	-	Visibility of *j* when standing at *i*

τ_{ij}	-	Pheromone level of edge (i, j)
α_k	-	Momentum coefficient
ACO	-	Ant Colony Optimization
ABC	-	Artificial Bee Colony
ASBC	-	Artificial Smart Bee Colony
AIS	-	Artificial Immune Systems
ANN	-	Artificial Neural Networks
BP	-	Back-Propagation
BDU	-	Binary Decision Units
BCO	-	Bee Colony Optimization
BO	-	Bee Optimization
CSA	-	Cuckoo Search Algorithm
CS	-	Colony Size
DABC	-	Discrete Artificial Bee Colony
EC	-	Evolutionary Computation
EP	-	Evolutionary Programming
ES	-	Evolutionary Strategies
EA	-	Evolutionary Algorithms
FLNN	-	Functional Link Neural Networks
GA	-	Genetic Algorithms
GABCS	-	Global Artificial Bee Colony
GABC	-	Guided Artificial Bee Colony
GGABC	-	Global Guided Artificial Bee Colony
GGABC	-	Gbest Guided Artificial Bee Colony
HONN	-	Higher Order Neural Network
HABC	-	Hybrid Artificial Bee Colony
IGGABC	-	Improved Gbest Guided Artificial Bee Colony
IABC	-	Improved Artificial Bee Colony
MLP	-	Multilayer Perceptron
MCN	-	Maximum Cycle Number
MSE	-	Mean Square Error
NMSE	-	Normalized Mean Square Error
NNs	-	Neural Networks
PSO	-	Particle Swarm Optimization
PNN	-	Probabilistic Neural networks
RNN	-	Recurrent Neural Network
S.D-MSE	-	Standard Deviation of Mean Squarer Error
SLP	-	Single Layer Perceptron
SI	-	Swarm Intelligence

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Research Background

Though the rapid developments in various fields, life became easy and world is becoming virtually global village. Systems are happening progressively and more complex, whether it is engineering, transportation, communication resources, diseases, foods, power, water distribution and land division. These complexities may be due to the increase in number of components, increased interdependencies among various components, and population. The complexities further increases when these communications are subjected to natural or man-made disasters. According to the Centre for Research on the Epidemiology of Disasters (CRED), 336 natural disasters and 234 technological disasters were reported worldwide in 2011. In the history, these disasters created many problems to human communities.

In the past decade, the fracture of earth, flow of rocks, movements of tectonic plates, heat waves temperature and the high range of sea waves has been focused by geologists, geophysicists and engineers. These sources may be the most important rule in earthquake, weather temperature, water level height and tsunami occurrence called seismic signals or natural hazards. Seismic events, especially earthquakes and tsunami are the most costly natural hazards faced by the nation in which they occur without prior warning and may cause severe injuries. The intensity of occurrence of such event creates disasters and can change human lives, animals as well as other social swarms.

In recent years much has been learned from natural disasters and risk to infrastructure systems. It is estimated that, the total direct economic loss from natural disasters during the decade of 1987-1997 was 700 billion USD, an average loss of 70 billion USD per year. The number of deaths caused by natural disasters (31,105) is the fourth lowest of the decade, much lower than the peaks of 2004 (242,010 deaths), 2008 (235,272) and 2010 (297,730). The deadliest natural disaster was the earthquake and subsequent tsunami in Japan in March, which killed 19,846 people. The number of deaths is much lower than those caused by the Indian Ocean tsunami in December 2004 (226,408 deaths) and the earthquake of January 2010 in Haiti (222,570 deaths). In 2011, natural disaster costs (US$ 365.6 billion) were the highest of the decade, accounting for almost 1.5 times the direct losses reported in 2005 (US$ 248 billion, 2011 prices) (Ifrcrcs, 2012).

The disasters caused by seismic events are beyond computation, location, time and intensity of a future seismic events occurrence are accurately predicted, and some appropriate precautionary measures can be carried out in advance. The countless lives in earthquake risk areas can be saved, and the human and economic losses caused by these events can be reduced. Each seismic event is related to a different source process, and its time and spatial distribution could be used as elements of an early-warning system of seismic occurrence (Gutierrez *et al*., 2009; Suratgar *et al.*, 2008). The predictions of these seismic events are really crucial to our continuation. The data of these occurring events are depending on geographical areas in real time series form. The behavior of seismic time-series data is quite different among the other data, so the prediction of this nature of data is quite challenging for a scientist. In this regard, to predict these events the scientist used to study the grounds of these trials. These grounds can be physical and non-physical parameters such as; water level changes of wells, temperature changes, radon emission changes, climate changes, weather, earthquake and the changes in earth magnetic fields (Donovan, 2012; Lin & Lin, 2010; William, 2011).

Researchers have focused on seismic prediction using various applications such as geomagnetic field, space technology, satellite information, mathematical approach, electromagnetic fields, weather conditions, unusual clouds, radon or hydrogen gas content of soil or ground water, water level in wells, animal behavior and other methods (Adeli & Panakkat, 2009; Botev & Glavcheva, 2003; Romano *et al.*, 2009; Serebryakova *et al*., 1992). Seismologists have investigated the relationship of future earthquakes with different phenomena such as seismicity patterns (Brown *et al*., 1989) crustal movements (Mogi, 1984), gas emissions from the earth, large-scale changes in soil temperature (Tsunogai & Wakita, 1995), and changes in ion concentration in the ionosphere were used to apply the infrared waveband of the meteorological satellite in the research of earthquake precursors.

Although several works claimed to provide seismic prediction, according to a specific location area, specific span of time, specific magnitude range and specific probability of occurrence. That is, seismic event prediction should state when, where, how big, and how probable the predicted event is and why the prediction is made. Unfortunately, no general useful method developed to predict (seismic signals accurately) has been found yet (Clarence, 1982). And it may never be possible to predict the exact time when a destructive seismic activity will occur, because when enough strain has been built up, a fault may become inherently unstable, and any small background seismic occurrence may or may not continue rupturing and turn into a large seismic events (Reyes *et al*., 2013). Although, the great efforts are made and the multiple techniques are developed by different researchers but no successful system has been found yet, (due to having nature behavior of earthquakes). Also, it may never be possible to determine the correct time, magnitude and location of the next damaging earthquake (Tiampo & Shcherbakov, 2012). The researchers reached to the point that classical techniques are not suitable for prediction of seismic events, and computational techniques like Artificial Neural Network (ANN) attracted them to predict the seismic signals through neural intelligence approach (Alarifi *et al*., 2012; Esteban *et al*., 2010).

ANN incorporates powerful data modeling and predictor tools that is able to capture and represent complex input/output relationships. ANN is known to have the ability to represent both linear and nonlinear relationships and to learn these relationships directly from the Boolean to time series data. The application of ANN to simulation and/or forecasting problems can now be found in various disciplines. The application of ANN in time-series data prediction has shown improved performance in comparison to statistical methods because of their nonlinear and training capability and universal approximator's ability for complex mapping (Ho *et al*., 2002; Yümlü *et al*., 2005). ANN approach has been applied to the comprehensive seismic signal such as earthquake forecasting which obtained satisfying results.

There are many factors, which arose from seismic events; it is very difficult to establish the physical model for their prediction. ANN technology has a unique advantage in constructing and predicting unknown object theoretical model, so it has been widely applied to the prediction. Therefore, ANN can behave as a model for the seismic process prediction tasks (Mart *et al*., 2011). Predictions of seismic event should specify time, intensity, location and probability. However, a statement that does not specify a time or magnitude or a statement that an earthquake will not occur in a particular place or time would be beneficial. ANN technique used by (Lakkos *et al*., 1994), which was simulated using the XERION software package and the Delta-Bar-Delta as the guidance algorithm to predict the magnitude of an impending earthquake and the geographical location. However, the authors do not clarify the magnitude range of the data used for training and testing.

ANN techniques have been focused by many investigators to explore their potential as a tool for simulation of the performance of systems that are managed by nonlinear large multivariate data and generally unknown interconnections within a noisy and poorly-controllable physical environment. The benefit of this framework is that the ANN provides an advance black-box technique, and the user does not need to know much about the nature of the process being simulated. The most widely used ANN models are the feed forward NNs, also called Multilayer Perceptron (MLP) (Rumelhart, 1986), which perform nonlinear regression and classification. The MLP is one of the earlier network models used for different problems such as classification, forecasting, seismic prediction, image processing and clustering (Adeli & Panakkat, 2009; Bock, 1999; Ghazali *et al*., 2011). The training algorithms have the important role in ANN output. Many training techniques defined for MLP such as BP and Gradient Descent (GD) algorithm (Jordan & Rumelhart, 1992).

Normally, ANN based BP training algorithm has a good success in solving many complex problems like classification and time series prediction. However, this method has some shortcomings, such as; the dependence of error surface shape, initial values of connection and parameters. Furthermore, MLP using BP for seismic time series prediction task failed to provided less training and testing error (Alhadi *et al*., 2011; Yue *et al*., 2004).

In order to overcome the drawbacks of standard BP, many evolutionary, population based techniques have been used such as: Genetic Algorithms (Curilem *et al*., 2009; Sexton & Gupta, 2000) Particle Swarm Optimization (PSO) (Alhadi *et al*., 2011; Gudise & Venayagamoorthy, 2003; Merkle & Middendorf, 2005), ACO (Blum & Socha, 2005), and ABC algorithm. ABC is more successful, and robust on multimodal functions included in the set with respect to Differential Evolution (DE) (Christian, 2005). The method of finding a way to search for the approximate optimal weight values as initial weights of training algorithm is needed, which can avoid the BP's trouble of slow convergence speed, trapping local minima and oscillation effects. Furthermore, for getting better accuracy in Boolean function classification and seismic signals prediction, the improved ABC algorithms were used. In this research work, MLP used to train by an improved learning algorithms to predict future values of possibly seismic time series based on past histories and at the same time for Boolean function classification task.

1.2 Problem Statement

Multilayer Perceptron (MLP) is a universal approximator which has been used in various scientific and engineering tasks (Hornik *et al*., 1989). The performance of MLP depends on its training algorithm, weight values, network topology and activation functions. The most common supervised learning algorithm called BP is used to train the weights in order to provide the network with good mapping capability. BP has high achievement ratio in solving many computational problems such as system controller, classification, prediction, function approximation, mathematical modelling, feature selection and other optimization problems (Drndarevic, 2006; Khan *et al*., 2008; Qi & Tufts, 1997; Weiyang, 1999).

Despite the general success of the BP algorithm, there are some drawbacks and restrictions that still exist (Ghaffari *et al*., 2006). These are the existence of temporary local minima resulting from the diffusion performance of the activation function, convergence speed is comparatively slow for network with two or more hidden layers, and some of the adjustments of BP algorithm require complex and costly calculations at each hidden layer and iteration, which offset their faster rates of convergence.

The new system of transformer fault diagnosis based on Dissolved Gases Analysis (DGA) with a BP-ALM algorithm developed by Sun *et al*., (2007) for quick learning and global convergence than previous methods, and a superior performance in fault diagnosis compared to convectional BP-based neural networks. For improving the efficiency of the error minimization process, or in other words the training efficiency (Bishop, 1995) is used. The gain parameter used by (Nawi *et al*., 2006; Sperduti & Starita, 1993) which controls the steepness of the activation function. It has been shown that a larger gain value has an equivalent effect of increasing the learning rate. In computing dynamically the new optimal learning rate method was proposed (Roy, 1994). Although this method could improve the performance of standard BP, the algorithm is computationally complex and might took longer to train MLP than standard BP. Due to these shortcoming in the standard BP, many improvements have been done by researchers to advance the performance of standard BP.

From the last decade, researchers developed an interest in Swarm Intelligence (SI) techniques. These algorithms include Genetic Algorithm, Evolutionary Algorithm (EA), Co-evolutionary Algorithm, Ant Colony Optimization and some other social hybrid algorithms used for training ANN (Blum & Socha, 2005; Carvalho & Ludermir, 2006; Ricardo, 2011). Swarm Intelligence based algorithms have high achievements in various research areas such as clustering (Bharne *et al*., 2011), prediction task (Ping *et al*., 2008), classification (Ozturk & Karaboga, 2011), numerical function optimization (Peng *et al*., 2011) and other mathematical and statistical problems.

Furthermore, for the prediction of seismic signals, researchers used different ANNs models such as Probabilistic Neural Networks (PNN) (Adeli & Panakkat, 2009), Recurrent Neural Networks, Radial-Basis Function (RBF) (Connor *et al*., 1994; Romano *et al*., 2009), however no efficient technique has been establish yet for getting high accuracy and less prediction error (Reyes *et al*., 2013). Due to the random activities of seismic occurrence, it may never be possible to ascertain the exact time, magnitude and location of the next damaging earthquake (Panakkat & Adeli, 2008). These techniques failed to predict the fix location, size and time of seismic occurrence. The swarm based algorithms recently have been famous for prediction of seismic signal prediction such as earthquake magnitude for South California (Shah *et al*., 2011), volcanoes and so on (Martínez *et al*., 2011). These techniques used to train ANNs with optimal weight values using the intelligence behaviours of social insects like particles, ant and honey bees.

From the last decade, honey bee population-based technique becomes famous for training ANN called Artificial Bee Colony (ABC) algorithm. The ABC algorithm is population-based technique that can be used to find approximate solutions to difficult optimization problems (Karaboga, 2005). It is inspired by the aforementioned described foraging intelligent behaviors of bee colonies. However, there is still deficiency in standard ABC algorithm regarding its solution search equation, which is good at exploration but poor at exploitation procedure (Zhu & Kwong, 2010).

To increase and balance the exploration and exploitation procedures of the standard ABC algorithm, and for improving the efficiency for Boolean function classification and time series prediction task, three improved algorithms called Global Guided Artificial Bee Colony (GGABC), Improved Gbest Guided Artificial Bee Colony (IGGABC) algorithms and Artificial Smart Bee Colony (ASBC) algorithm was developed based on honey bees intelligent behavioral approach. These proposed approaches used to update's weight values and bias for training MLP, to minimize the training error for classification of Boolean functions (XOR, 3 Bit parity and 4 Bit Encoder / Decoder) and time-series prediction (earthquake, heat waves temperature and water level height). In this research work, these three improved approaches are going to be used to overcome the limitations of standard BP and ABC by global guided bees, improved gbest guided bees and smart bees for getting high efficiency for Boolean function classification and time series prediction tasks.

1.3 Aims of Research

This research aims to develop improved learning techniques to train the MLP, for searching optimal weight values based on the artificial bee's intelligence behavioural algorithms. These learning techniques increase the effectiveness through average amount of exploration and exploitation based on neighbour information. Furthermore, this research work seeks to find suitable network architecture, which maintains good performance for Boolean function classification and time series prediction, with less training and testing errors. The proposed learning techniques used for time series data prediction such as earthquake magnitude, water level height and heat waves temperature and for Boolean function classification such as XOR, 3 bit parity and 4 bits Encoder / Decoder tasks. Furthermore, the proposed improved learning algorithms used to reduce the training, testing errors and get outstanding performance from standard ABC, PSO, GABC and BP in classification and prediction tasks.

1.4 Objectives of the Research

In order to achieve the research aim, a few specific objectives are set as follows:

(i) To implement and simulate the MLP trained with BP and swarm intelligence algorithms, namely the ABC and PSO for the prediction and classification tasks.

(ii) To propose and construct a hybrid GGABC algorithm for increasing exploration in MLP training.

(iii) To propose and construct an improved IGGABC algorithm for increasing exploitation in MLP training.

(iv) To propose a new ASBC algorithm for training the MLP with enough exploitation and exploration process.

(v) To compare and evaluate the performance of the MLP trained with the proposed GGABC, IGGABC and ASBC algorithms against the benchmarked algorithms BP, ABC, GABC and PSO, in terms of their classification error, prediction error, classification accuracy, prediction accuracy and convergence rate.

1.5 Significance of Research Contribution

This research provides the following contributions to knowledge in the fields of swarm intelligence based learning algorithms for ANN. In SI based learning algorithms the performance depends, on exploration and exploitation procedures.

(i) The proposed Global Guided Artificial Bee Colony (GGABC) algorithm used to increase the exploration procedure through global best and guided of neighbor bees, to find the optimal weight values for MLP, which will provide the high performance for Boolean classification and time series prediction.

(ii) Also, the exploitation process increased by proposed Improved Gbest Guided Artificial Bee Colony (IGGABC) algorithm through improved gbest guided neighbor information to find best weight values for outstanding performance of Boolean function classification and time series prediction.

(iii) The new algorithm Artificial Smart Bee Colony (ASBC) used train MLP through smart neighbor bees, to find the optimal weighs values, which will provide the good presentation for Boolean classification and time series prediction.

(iv) The GGABC, IGGABC and ASBC algorithms were used to provide best performance of Boolean functions classification and time series prediction.

1.6 Scope

The potential application of ANN in various applications with various types of learning algorithms is virtually limitless. In order to place boundaries around the vast topic of classification and prediction using these network models and algorithm, this research work covers and is limited to the training and testing of MLP with GGABC, IGGABC and ASBC algorithm, and their performance is benchmarked with the standard BP, ABC, GABC and PSO. All the algorithms used for the classification of XOR, 3 Bit Parity and 4 Bit encoder/decoder operators and the prediction of heat waves temperature, earthquake magnitude and water level height. The ability of the algorithms for training the MLP on both tasks were evaluated using five performance metrics, namely the Squared Error (MSE), Normalized Mean Squared Error (NMSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) accuracy and Signal to Noise Ratio (SNR).

1.7 Thesis Organization

This thesis is organized and divided into six chapters as follows. The motivation, objectives, and contributions are highlighted in the Chapter One. Chapter Two provides an overview of an ANN, history, types of learning algorithms of MLP and swarm intelligence based algorithms such as PSO, ABC, and ACO learning algorithms. Methodology used to carry out the study systematically is discussed in Chapter Three. The proposed improved swarm based learning algorithms GGABC, IGGABC and ASBC are detailed in Chapter Four. The simulation results of Boolean function classification and time-series data prediction and analysis of data are included in Chapter Five. Finally, conclusion and suggestions for future works are explained in Chapter Six. List of references and appendices section are included at the end the thesis.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

An Artificial Neural Network (ANN) can be defined as the information processing paradigm, which is inspired by the human biological nervous system. It is composed of a large number of highly interconnected processing elements known as neurons to solve computational problems. An important criterion of ANN is the ability of learning from the environment. Synaptic or weight connections that exist between the neurons in the nervous system are adjusted in order to learn. ANN consists of a number of artificial neurons which receive a number of inputs. A function called activation or cost function is applied to these inputs resultant in an activation level of a neuron. Knowledge about the learning task is given in the form of examples called training examples. ANN is defined by architecture, neuron model and the learning algorithm. Architecture refers to a set of neurons and links connecting neurons with weights. Neuron model refers to the information-processing unit of the ANN. Besides that, learning algorithms are used to train the ANN by modifying the weights in order to model a particular learning task correctly on the training examples.

2.2 From Biological to Artificial Neuron

The human brain consists of a large number of neural cells that process information. Each cell works like a simple processor and only the massive interaction between all cells and their parallel processing makes the brain's abilities feasible (Byrne, 1991; Giles & Maxwell, 1987). ANN show the parallel processing ability of human brain system, such as Central Processing Unit which is referred to as the computer brain. Simply because their performance is inspired by the way in which human brain process information (Byrne, 1991; Fogel *et al*., 1966). Figure 2.1 shows the basic structure of the biological neurons.

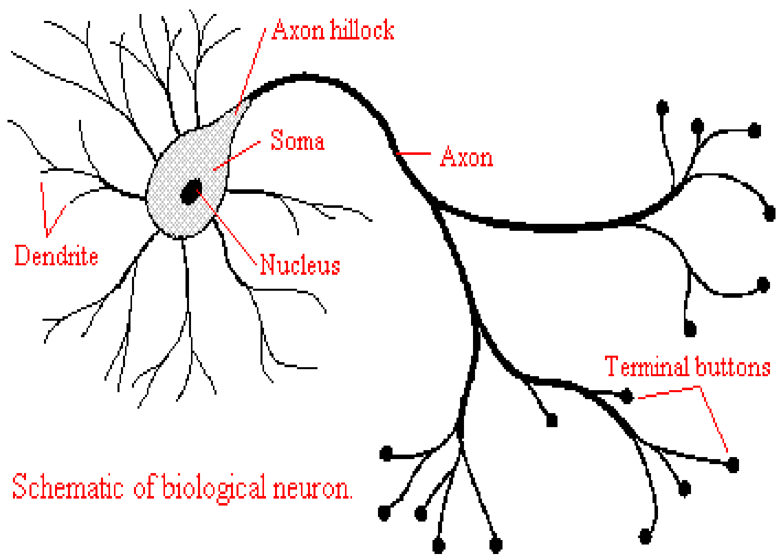

Figure 2.1: Basic structure of biological neuron (Fraser, 1998)

ANN is also referred to as “neural nets," “artificial neural systems," “parallel distributing processing element” and “connectionist system". ANN information is simulated by using inspiration of human brain’s skill and nature of processing information. The inspiration behaviors have the abilities to take a decision for the best solution. The human brain process information that is called neurons using the following function such as, Dendrites, Synapses and Axon (Byrne, 1991; Holland, 1975). The dendrites' function used for getting input from the environment or from other neurons and send for processing to synapses step. Information is transported between neurons in the form of electrical stimulations along the dendrites. Synapses search the favorable solution with random hidden information. The axon will give a response either good or otherwise. Synapses are the elementary structural and functional units that are subsequent to the interconnection between neurons.

The body of the cell contains the nucleus of the cell and transmits the biochemical alterations necessary to synthesize enzyme. It is typically several microns in diameter (a micron is a millionth of a meter). The signal of most real neurons is chemical, and it consists of spikes, short pulses of electric activity. In ANN, these spikes are replaced by continuing variable x_j which may be think of as temporal average pulse. The majority of neurons encodes their outputs as a series of brief voltage pulses. A biological neuron may have as many as 10,000 different inputs and may send its output to many other neurons up to 200,000. The ANN also works and developed using human brain processing techniques using the following model (McCulloch & Pitts, 1943). From the Figure 2.2, the inputs which are shown at start node called input node.

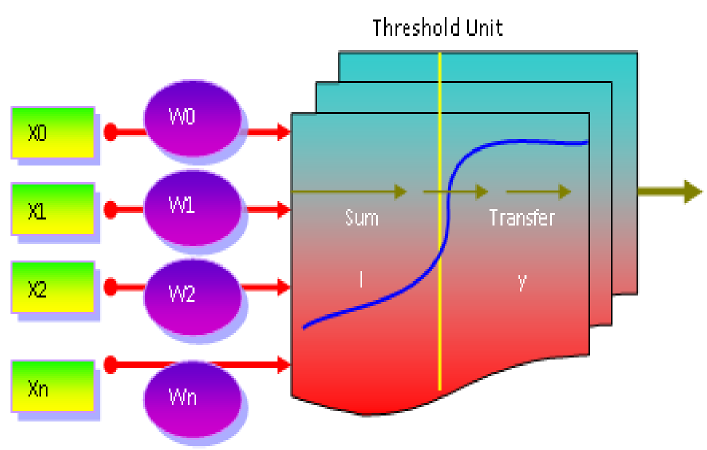

Figure 2.2: Simple mathematic model for neuron (McCulloch, 1943)

It has been established in 1943, that the neuron model proposed by McCulloch and Pitts is implemented as a threshold unit. Weighted inputs to the unit are summed to produce an activation signal, and if the activation signal exceeds some threshold value the unit produces some output response. However if the activation signal does not exceed the threshold, no output is produced by the unit. For instance if there are *n* inputs to the threshold unit with weights $w_1, w_2 .... w_n$ and signals $x_1, x_2 ...... x_n$. The activation α of the unit.

$$\alpha = \sum_{i=1}^{n} w_i x_i \tag{2.1}$$

The output *O* of the threshold unit is given by

$$O = \begin{cases} 1 & \text{if } \alpha \geq \theta \\ 0 & \text{if } \alpha < \theta \end{cases} \tag{2.2}$$

where θ is the threshold and often equal to zero.

2.3 The Working of Artificial Neural Networks

Certainly, ANN can often provide a suitable solution for problems that are generally characterized by nonlinear, high dimensionality, noisy, complex, imprecise, imperfect and/or error-prone sensor data, poorly understood physical and statistical models, and lack of clearly stated mathematical solution or algorithm (Zaknich, 2003). Mostly ANN approaches are capable of solving scientific, electrical engineering, earth knowledge, mathematical and of course statistical tasks (Karaki & Chedid, 1994).

The determination of the network architecture is one of the most important steps in developing a model for a given problem. Although ANN construction has been extensively investigated by various researchers, there is no known procedure or algorithm for this process for the general case. Two approaches have been proposed, namely constructive and destructive methods. In both constructive and destructive methods, the numbers of hidden nodes are considered.

The fundamental composition block of every ANN is the artificial neuron. That is, a simple mathematical functions or model combination such as multiplication, summation and activation. At the doorway of an artificial neuron, the input signals are weighted, implying that every input value is multiplied by individual weight values. The center distribution of an artificial neuron is addition function that sums all weighted values, input signals and bias values. The final step or output layer of an artificial neuron is the sum of earlier weight values; input signals and bias are passing through transfer function also called activation function for getting target values. There are different activation or transfer function uses for getting the best target which depends on the output and behaviors of the data. Figure 2.3 shows the universal ANN structure (McCulloch & Pitts, 1943).

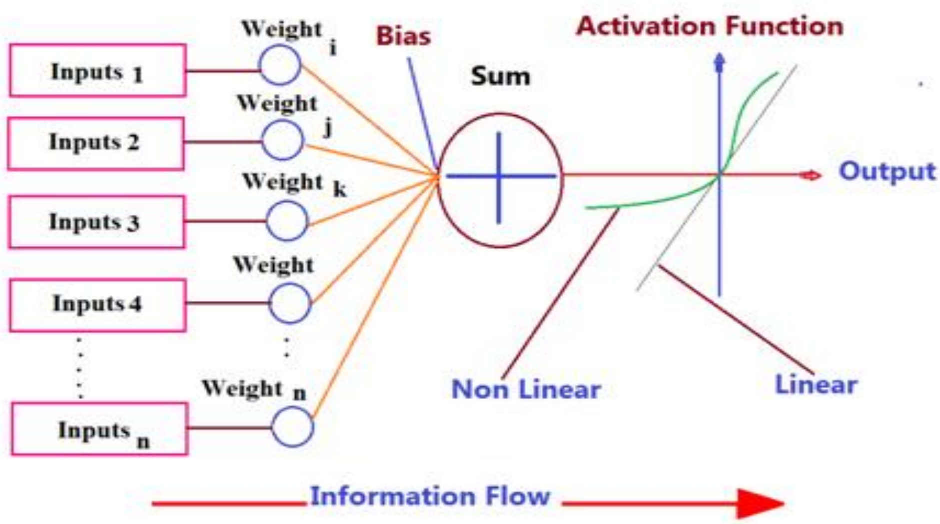

Figure 2.3: Artificial neuron working model (Stamatios, 1996)

Each input has an associated weight *w*, which can be modified in order to model synaptic learning. The unit computes some function *f* of the weighted sum of its inputs by equation (2.3) as:

$$y_i = f_i\left(\sum_{i=1}^{n} w_i x_i + bias\right) \tag{2.3}$$

Figure 2.3 shows the working principle of an artificial neuron with the three mathematical functions such as summation, multiplication and activation function. The output can be linear and nonlinear depending upon the behavior of data and activation function. While from the above working principle's model and simple set of rules of the artificial neurons appear to have nothing special, the full perspective and calculation ability of these models proved to be capable of solving different difficulties, experimentation and analysis when it starts to communicate them into ANN.

2.4 Advantages of Neural Networks

Depending on the nature of the application and the strength of the internal data patterns a network can generally be expected to train quite well. This applies to problems where the relationships may be quite dynamic or non-linear. ANN provides an analytical alternative to conventional techniques which are often limited by strict assumptions of normality, linearity, variable independence etc. Because an ANN can capture many kinds of relationships it allows the user to quickly and relatively easily model, which otherwise may have been very difficult or impossible to be explained. ANN shows the benefits in some important difficulties (Haykin, 1999). These are:

Neurobiological Analogy: The design of ANN is motivated by analogy with the brain, which shows the fault-tolerant parallel processing which is not only physically possible but fast and powerful (Christodoulou *et al*., 2012).

Nonlinearity, which is a very important property of ANN especially if the system which ANN tries to model, is naturally nonlinear. ANN performs the best performance for nonlinear problems (Olyaee *et al*., 2012).

Input Output Mapping: ANN gets input and train with weight values using different algorithms and output as the target. The ANN training methods will try to get the target in supervised or unsupervised learning if the ANN can not find the target; the weights value changes and train repeatedly until it finds another best target value. ANN is a powerful data-modeling approach that is able to capture and represent complex input/output relationships (Lawrence, 1994).

Adaptively: ANN has a built-in capability to adapt their synaptic weights to changes in the environment where they operate. This property leads ANN to human brain processing or thinking properties.

Fault Tolerance: Fault tolerance refers to the capability of a system to function adequately despite the failure of components. ANN has the ability to recover from a tragic failure without disrupting its operations (Jain *et al*., 2000). ANN implemented in hardware form has the potential to be inherently a fault tolerant in the sense that the performance is degraded gracefully under adverse operating condition. ANN show fault tolerance since the information spreads in the connections during the network. Even if few relations are cracked or a few neurons are not working, the information is still conserved due to the spreading nature of the encoded information. Fault tolerance in ANN computing takes several appearances. The first is an inherent tolerance in the exact computation of the network. The second is tolerance to the inexact mapping of the network for the implementation, including the possible malfunction of portions of the implementation, such as failure of individual units (Panduranga *et al*., 2007).

2.5 Learning Algorithms

Artificial Neural Networks is based on human brain processing techniques and gets decision using neuron, and it is connection values to find better results. The most significant aspect of ANN is its ability to learn from its environment, and to improve its performance through learning. ANN learns about its environment or a dynamic system through an iterative process of adjustments applied to its weights and biases. One of the most important characteristics in ANN is its knowledge ability, which makes it generally suitable for the computational purpose whose organization is known or unknown. The decision is based on the synapse's learning strategy. The network becomes more "knowledgeable" about its environment after iteration of the learning process. Like human beings and animals that learn more things. ANN learning is an inferred process which can not be perceived directly, but can be assumed to have happened by observing changes in performance (Zurada, 1992).

Learning in the context of ANN is defined as a process by which the free parameters of ANN are adapted through a process of presenting signals from the environment in which the network is embedded. The type of learning is determined by the way the parameter changes take place (Simon, 1994). The notion of learning in ANN, is the process of guiding the network to provide a particular output or response for a specific given input. Learning is necessary when information about the input-output relationship is unknown or incomplete a-priori. The learning objective is to minimize the cost function which is the difference between desired output and neural network output. The networks were trained for finding optimal weights, which reduce the error until the convergence.

There are two different types of learning, namely supervised and unsupervised learning, which identifies or creates pattern-class information about the learning outcome. In this case, no desired or target classes are known beforehand, and thus no output information is known a-priori. Meanwhile, a supervised learning deals with the desired set of responses, outputs or classes for given inputs which are known and provided during the learning process. In this case, the neural has to learn the function, mapping or transformation that will produce the desired output for new inputs (Zurada, 1992). Supervised and unsupervised learning can thus be distinguished as follows:

The main principle of supervised learning is to "guide or practice" a network to print the good behavior of a system. In this case, there is always a need to have a "training" data set. The network topology and the algorithm that the network is trained with are highly interrelated. In general, a topology of the network is chosen first, and then a proper training approach is used to tune the weights (Brandt & Feng, 1996; Gallant, 1993), which integrates an external teacher, so that each output unit is told what it's required answer to input signals ought to be.

On the other hand, the unsupervised learning uses no external teacher and is based upon only local information. It is also referred to as self-organization, in the sense that it self-organizes data presented to the network and detects their emergent collective properties. This group of network training attempts to cluster around input data without the need for the traditional "learn by example" technique that is commonly used for ANN. Note that, clustering applications tend to be the most popular type of applications that these networks are normally used.

2.6 Backpropagation (BP) Learning Algorithm

Backpropagation (BP) is a renowned supervised form of learning algorithm for obtaining the optimal weight's values in ANN applications, developed by (Rumelhart *et al*., 1986). BP algorithm is widely used to solve many engineering modelling problems (Najaf *et al*., 2013; Zweiri *et al*., 2002). The basic BP algorithm is based on minimizing the error of the network using the derivatives of the error function. The BP used to adjust the network's weight and threshold so as to minimize the error for the different task such as classification, clustering and prediction on the training set. The major advantage of the BP algorithm over the Least Mean Squared Error (LMSE) and perceptron algorithms is in expressing how an error at an upper (or outer) layer of a multilayer network can be propagated backwards to nodes at lower (or inner) layers of the network.

BP algorithm presented in three stages for training MLP. Firstly, the feed-forward phase the input signals are propagated through the input and hidden layers of processing elements, generating an output pattern in response to the input pattern presented. Secondly, in Back-forward phase as shown in Figure 2.4 , each output node compares its activation with the desired output based on these differences, the error is propagated back to all previous nodes Delta Rule. Thirdly, weights of all links are computed simultaneously based on the errors that were propagated back. The three layers MLP with BP learning algorithm is shown in the Figure 2.4.

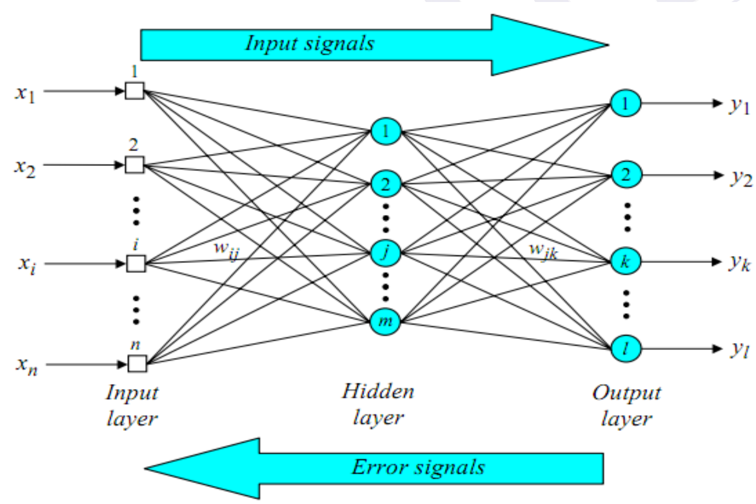

Figure 2.4: Artificial Neural Network with BP learning algorithm

Each node, or artificial neuron (Threshold Logic Unit), is composed of two sections. The first section generates a sum of the products of the weights multipliers and input signals. The second section takes the result of the first section and puts it through its activation function, with scales input to a value between 0 and 1. Signal *e* is the output of the first section, and $y = f(e)$ is the output of the second section. Signal *Y* is also the output signal of an artificial neuron. There are several types of activation function, the most common activation function of a neuron *f (x)* is a sigmoid function (Wang *et al*., 2004) as shown below:

$$f(net_j) = \frac{1}{1+e^{net_j}} \tag{2.4}$$

where:

$net_j = \sum w_{ij}\, a_i,$

a_i is the input activation from unit *i*, and

w_{ij} is the weight connecting unit *i* to unit *j*.

In the next algorithm step, the output signal of the network *y* is compared with the desired output value (the target). The difference is called error signal of output layer neuron, which is calculated as:

$$E = \frac{1}{2}\sum_k (t_k - y_k)^2 \tag{2.5}$$

E = error vector, t_{nk} is the actual output and y_{nk} is the network value. In order to derive the BP learning rule, chain rule use to rewrite the error gradient for each pattern as the product of partial derivatives. Thus, the error gradient becomes:

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2} \cdots\cdots\cdots\cdots \frac{\partial E}{\partial wn} \tag{2.6}$$

The partial derivative reflects the change in error as a function of the net input; the second partial derivative reflects the effect of a weight change on a change in the net input. By using the chain rule with respect to weight and biases, in the BP algorithm, is determined as follows:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial S_i}\frac{\partial S_i}{\partial net_i}\frac{\partial net_i}{\partial w_{ij}} \tag{2.7}$$

where:

w_{ij} is the weight from neuron *j* to *i*,

S_i represent the output of neuron and

net_i is the weighted sum of the inputs of neuron *i*.

The weight will update with the gradient rules, learning rate with derivative to minimize the error function as:

$$W_{ij}(t+1) = W_{ij}(t) - \varepsilon \frac{\partial E}{\partial W_{ij}}(t) \tag{2.8}$$

where:

$W_{ij\,(t+1)}$ shows the new weight value,

W_{ij} represents the old weight values and ε represents the learning rate, which can control the learning and has important effect on convergence time.

The learning rate is a constant used in error BP learning that affects the speed of learning. The smaller the learning rate, the more steps it takes to get to the stopping criterion. A too large or too small η will cause negative inferences to converge . If it is too small, the learning process can be very slow (Knight, 1990). The different combinations of the learning rate and momentum are introduced to try to find the right combination that will allow the solution to escape local minima but not skip over the global solution. To make the learning process more stable, the momentum term is used to the weight changes as:

$$\Delta W_{ij}(t) = -\varepsilon \frac{\partial E}{\partial W_{ij}}(t) + \mu \Delta w(t-1) \tag{2.9}$$

where, $\Delta W_{ij}(t) = W_{ij}(t) - W_{ij}(t-1)$ and the momentum term represented by μ, where the momentum factor 0< μ <1, usually sets around 0.9 (Wasserman, 1989). Using high learning rate, momentum term can avoid the oscillation.

Although the back propagation algorithm is a powerful technique applied to classification, combinatorial problems and for training MLP. The problem's complexity increases (due to increased dimensionality and/or greater complexity of the data), performance of back propagation falls off rapidly because gradient search techniques tend to get trapped in local minima the performance of back propagation falls off rapidly because of the fact that complex space have nearly global minima which are sparse among the local minima. Gradient search techniques tend to get trapped at local minima (Montana & Davis, 1989). When the closely global minima are well hidden among the local minima, back propagation can end up bouncing between local minima, especially for those non-linearly separable pattern classification problems or complex function approximation problem. A second shortcoming is that the convergence of the algorithm is very sensitive to the initial value. So, it often converges to an inferior solution and gets trapped in a long training time.

There are several approaches developed for recovering and updating BP algorithm for different problems. An improved BP algorithm with stochastic attenuation momentum factor proposed by (Jia & Dali, 1993) and compared with the standard BP algorithm, the algorithm claims to effectively cancel the negative effect on the momentum of a network. However, the calculation of this approach is complex since it uses the correlation matrix in defining the momentum. For evolving convergence speed, an adaptive learning rate and momentum coefficient is proposed (Chien & Bin, 2002). In this proposed technique for fast convergence BP used with Adaptive Learning rate and Momentum factor (BPALM). Thus, from avoiding the local minima trapping problem proposed an improved BP where each training pattern has its own activation functions of neurons in the hidden layer and the activation functions are adjusted by the adaptation of gain parameters during the learning procedure (Wang *et al*., 2004). However, this approach did not produce good results on large problems and practical applications.

Due to all these problems in BP, for the last decade swarm Intelligence, an artificial intelligence discipline, is concerned with the design of intelligent multi-agent systems, such as ants, termites, fish, birds, bees, and wasps, by taking inspiration from the collective behaviors of social insects and other animal societies used for different combinatorial tasks (Kennedy *et al*., 1995; Dorigo, 1999; Karaboga *et al*., 2007). They are characterized by a decentralized way of working that mimics the behavior of the swarm. The researchers have replaced the BP algorithm with SI based learning algorithm in order to avoid local minima and slow convergence problems.

2.7 From Artificial Neural Networks to Swarm Intelligence

The ANN working model has different tasks such as, multiplication, summation and activation. In multiplication stage the weights are multiplied by an input signal. The input signals are predefined values while weight values are initialized with different techniques. Initially, weights were chosen randomly but with the passage of time researchers started taking an interest to get optimal weight in weight equation.

There are different techniques uses for finding optimal weight values such as BP, and modified version of BP (Nazri *et al*., 2010), Gradient Descent (Baldi, 1995; Yu & Chen, 1997), Differential Evaluation (DE) (Slowik & Bialko, 2008), Genetic algorithm (GA) (Qiang *et al*., 2005), Improved BP (Nazri *et al*., 2010) and other mathematical approaches. While these algorithms have been shown to be an effective method for training ANN, it typically has a slow convergence rate, and is known to suffer from local minima (Nazri *et al*., 2010). To overcome these limitations, researchers took interested in Swarm Intelligence approaches such as PSO (Kennedy & Eberhart, 1995), Ant Colony Optimization , ABC, Improved ABC algorithm, Hybrid Artificial Bee Colony (Karaboga *et al.,* 2007) and some others hybrid such as BP and Levenberq-Marquardt with ABC algorithm (Bitam *et al*., 2010; Blum & Socha, 2005; Ozturk & Karaboga, 2011). These learning techniques show that ANN processed to swarm based behaviors in the basic part of ANN. These algorithms are easy to implement and found to be robust compared to the standard BP.

2.8 Swarm Intelligence

Swarm Intelligence (SI) is a recent technique, which deals with natural social insects and artificial systems that composed of many individuals' agents based on the study of collective behaviour in decentralized and self-organized systems like the movement and behaviour of natural swarm workers (Bonabeau *et al*., 1999). The ant colonies, bees and bird flocking that can be effectively applied to computational intelligent systems are the basic agent of SI.

2.9 Fundamentals of Swarm Intelligence

SI has two fundamental notions; self-organization and division of labour or agents. The two basic notations, having necessary and sufficient properties to obtain swarm agent's behaviour such as distributed problem solving systems that is self-organized and adapt to the given environment (Karaboga *et al*., 2005). Also flexibility and robustness is influenced by self-organization with SI techniques for different task (Abraham *et al*., 2006). These two fundamentals of SI are discussed below.

1) Self-organization can be defined as a set of dynamical mechanism, which results in structures at the global level through a system by means of interactions among its low-level elements (Bonabeau *et al*., 1999). Through self-organization, the behavior within the group emerges from collective interactions of all individuals. These mechanisms demonstrate basic guidelines for the interactions between the components within the system. The rules ensure that the interactions are executed based on purely local information without any relation to the global method. The social insects seem to have two key priorities in their life time, finding food and defending against enemies. It seems to be a simple life as compared to human beings

2) Inside a swarm, there are different tasks performed simultaneously by specialized individuals. This kind of phenomenon is called division of workers for given tasks. Simultaneous task performance by cooperating specialized individual is believed to be more efficient than the serial task performance by unspecialized individuals. Division of agents also enables the swarm to respond to changed conditions within the search space. The above two fundamental concepts for the collective performance of a swarm presented are necessary and sufficient properties to obtain environment behavior of SI agent to get optimal solution for given problems.

2.10 Types of Swarm Intelligence Algorithms

There are various types of Swarm Intelligence (SI), these include Particle Swarm Optimization (PSO) inspired by the social behaviour of bird flocking or fish schooling (Kennedy & Eberhart, 1995), Ant Colony Optimization which is inspired by the foraging behaviour of ant colonies (Dorigo, 1992), Bee Swarm Optimization (BSO) (Davidović *et al*., 2012; Teodorovic *et al*., 2006), Artificial Bee Colony algorithm which is inspired by the foraging behaviour bee colonies and Cuckoo Search (CS) algorithm which is inspired by the behaviour of cuckoo bird (Yang and Deb 2009). Researchers have widely used SI through hybridization with many other techniques. The PSO algorithm improved by researchers with different new and hybrid strategies (Jun & Xiaohong, 2009; Mohammadi & Jazaeri, 2007). ACO hybrid with BP, LM, PSO, GA and other optimization algorithms for different tasks (Biswal *et al*., 2011; Jung & Lee, 2003; Xiao *et al*., 2009).

Researchers have extends standard ABC algorithm to the Modified Artificial Bee Colony (MABC) (Zhang *et al*., 2011), an Improved Artificial Bee Colony (IABC) (Shah and Ghazali., 2011), PSO-ABC (Tarun *et al*., 2011), the Global Hybrid Ant Bee Colony (GHABC) algorithm, the Hybrid Artificial Bee Colony (HABC), the Hybrid Artificial Bee Colony (Shah *et al*., 2010, 2011, 2012), the Discrete Artificial Bee Colony (DABC), a Combinatorial Artificial Bee Colony (CABC), the parallel Artificial Bee Colony (PABC) (Narasimhan, 2009), the Novel Artificial Bee Colony (NABC), Application Artificial Bee Colony (AABC) and many other types of recent improvements for different mathematics, statistical and engineering problems. Undoubtedly, all types of SI are extremely renowned and focused upon by the researchers for further improvement and increasing their applicability to mathematical, statistical and optimization problems. The ACO, PSO and ABC have the highest ratio of interest as compared to other swarms based approaches (Abraham *et al*., 2006; Kennedy & Eberhart, 1995). There are many other types of SI approaches, however, the distinguished PSO, ACO, and ABC algorithms are detailed in the following sections.

2.10.1 Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) is a population based stochastic optimization method inspired by social behaviour of bird flocking or fish schooling developed (Kennedy & Eberhart, 1995). PSO is a robust technique based on the movement and intelligence of swarms. The system is initialized with a population of random solutions and searches for optima by updating generations. Each particle is treated as a point in an N-dimensional space which adjusts its "flying" according to its own flying experience as well as the flying experience of other particles. However, unlike GA, PSO has no evolution operators such as crossover and mutation (Gudise & Venayagamoorthy, 2003).

REFERENCES

Abraham, A. Guo, H. & Liu, H. (2006). Swarm Intelligence: Foundations, Perspectives and Applications. In N. Nedjah & L. Mourelle (Eds.), *Swarm Intelligent Systems* Springer Berlin Heidelberg, Vol. 26, pp. 3-25.

Adeli, H. & Panakkat, A. (2009). A probabilistic neural network for earthquake magnitude prediction. *Neural Networks*, *22*(7), 1018-1024.

Ahmad, S. Mohammad. (2011). Weather Temperature Forecasting Using Artificial Neural Network. *Journal of Engineering and Development, 15*(2).

Ajith, A. Jatoth, R. K & Rajasekhar, A. (2012). Hybrid Differential Artificial Bee Colony Algorithm. *Journal of Computational and Theoretical Nanoscience*, *9*(2), 249-257.

Alarifi, A., S, N,, Alarifi, N & Al-Humidan, S. (2012). Earthquakes magnitude predication using artificial neural network in northern Red Sea area. *Journal of King Saud University- Science, 24*(4), 301-313.

Alhadi, I., A. A. Hashim, S. Z. M. & Shamsuddin, S. M. H. (2011). *Bacterial Foraging Optimization Algorithm for neural network learning enhancement.* Paper presented at the Hybrid Intelligent Systems (HIS), 2011.

Alpsan, D., Michael W, T., Ozcan, O., Ah Chung, T. & Dhanjoo, G. (1995). Efficacy of modified backpropagation and optimisation methods on a real world medical problem. *Neural Networks,Elsevier, 8*, 945-962.

Altan, D. Omer. & Golcu, M. (2009). Daily means ambient temperature prediction using artificial neural network method: A case study of Turkey. *Renewable Energy, 34*(4), 1158-1161.

Asadi, S., Hadavandi, E., Mehmanpazir, F., & Nakhostin, M. M. (2012). Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction. *Knowledge-Based Systems, 35*(0), 245-258.

Ashena, R. & Moghadasi, J. (2011). Bottom hole pressure estimation using evolved neural networks by real coded ant colony optimization and genetic algorithm. *Journal of Petroleum Science and Engineering, 77*(3–4), 375-385.

Attoh, O. & Nii, O. (1999). Analysis of learning rate and momentum term in backpropagation neural network algorithm trained to predict pavement performance. *Advances in Engineering Software, 30*(4), 291-302.

Awang, M. Khalid, Rahman, M. N, Abdul. & Ismail, M, Ridwan. (2012). Data Mining for Churn Prediction: Multiple Regressions Approach. *Computer Applications for Database, Education, and Ubiquitous Computing* (Vol. 352, pp. 318-324): Springer Berlin Heidelberg.

Baldi, P. (1995). Gradient descent learning algorithm overview: a general dynamical systems perspective. *Neural Networks, IEEE Transactions on, 6*(1), 182-195.

Beheshti, S. Hashemi, M. Sejdic, E. & Chau, T. (2011). Mean Square Error Estimation in Thresholding. *Signal Processing Letters, IEEE, 18*(2), 103-106.

Bergman, D. E. A. (2000). The Determination of Earthquake Parameters. Retrieved 25 november, 2012, from http://www.seismo.com/msop/msop79/par/par.html

Bharne, P. K., Gulhane, V. S., & Yewale, S. K. (2011). *Data clustering algorithms based on Swarm Intelligence.* Paper presented at the Electronics Computer Technology (ICECT), 2011 3rd International Conference on.

Bhattacharya, B. & Solomatine, D. P. (2005). Neural networks and M5 model trees in modelling water level-discharge relationship. *Neurocomput., 63*, 381-396.

Bird, P. & Liu, Z. (2007). Seismic hazard inferred from tectonics. *California Research Letters, 78*(1), 37–48.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*: Oxford University Press, Inc.

Biswal, B., Dash, P. K., & Mishra, S. (2011). A hybrid ant colony optimization technique for power signal pattern classification. *Expert Systems with Applications, 38*(5), 6368-6375.

Bitam, S., Batouche, M., & Talbi, E. g. (2010). *A survey on bee colony algorithms.* Paper presented at the Parallel & Distributed Processing, Workshops and Phd Forum (IPDPSW), 2010 IEEE International Symposium on.

Blum, C., & Socha, K. (2005). *Training feed-forward neural networks with ant colony optimization: an application to pattern classification.* Paper presented at the Hybrid Intelligent Systems, 2005. HIS '05. Fifth International Conference on.

Bock, H.-H. (1999). Clustering and Neural Network Approaches. In W. Gaul & H. Locarek-Junge (Eds.), *Classification in the Information Age* (pp. 42-57): Springer Berlin Heidelberg.

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). Swarm Intelligence: From Natural to Artificial Systems, Oxford University Press, NY,.

Botev, E. A., & Glavcheva, R. P. (2003). *Modern earthquake monitoring in central Balkan region.* Paper presented at the Recent Advances in Space Technologies, 2003. RAST '03. International Conference on. Proceedings of.

Brameier, M. (2009). Data-Mining and Knowledge Discovery, Neural Networks in. In R. A. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science* (pp. 1812-1826): Springer New York.

Brandt, R. D., & Feng, L. (1996). *Supervised learning in neural networks without feedback network.* Paper presented at the Intelligent Control, 1996., Proceedings of the 1996 IEEE International Symposium on.

Brown, D., Li, Q., Nyland, E., & Weichert, D. H. (1989). Premonitory seismicity patterns near Vancouver Island, Canada. *Tectonophysics, 167*(2–4), 299-312.

Byrne, J. H. (1991). Introduction to Neurons and Neuronal Networks. Retrieved 25 May, 2013

Cannon, A. J. (2007). Nonlinear analog predictor analysis: A coupled neural network/analog model for climate downscaling. *Neural Networks, 20*(4), 444-453.

Carvalho, & Ludermir, T., B. (2006). *Particle Swarm Optimization of Feed-Forward Neural Networks with Weight Decay*. Paper presented at the Hybrid Intelligent Systems, 2006. HIS '06. Sixth International Conference on.

Chakraborty, S., & De, D. (2012). *Object oriented classification and pattern recognition of Indian Classical Ragas*. Paper presented at the Recent Advances in Information Technology (RAIT), 2012 1st International Conference on.

Chan, Tze, H., Chun, T., Lye,, & Chee, W., Hooy,. (2010). Forecasting Malaysian Exchange Rate: Do Articial Neural Networks Work?

Chandra, P, & Singh, Y. (2004). An activation function adapting training algorithm for sigmoidal feedforward networks. *Neurocomputing, 61*(0), 429-437.

Chauhan, N. C., Kartikeyan, M. V., & Mittal, A. (2012). Soft Computing Methods *Soft Computing Methods for Microwave and Millimeter-Wave Design Problems* (Vol. 392, pp. 9-23): Springer Berlin Heidelberg.

Chien, C., Yu, & Bin, D., Liu. (2002). *A backpropagation algorithm with adaptive learning rate and momentum coefficient*. Paper presented at the Neural Networks, 2002. IJCNN '02. Proceedings of the 2002 International Joint Conference on.

Christodoulou, M. S., Zunino, F., Zuco, V., Borrelli, S., Comi, D., Fontana, G. (2012). Camptothecin-7-yl-methanthiole: Semisynthesis and Biological Evaluation. *ChemMedChem, 7*(12), 2134-2143.

Clarence, R., Allen. (1982). Responsibilities in earthquake prediction, Bulletin of the Seismological Society of America. 2069–2074.

Commerce, U. S. D. Commerce. (2012). National Oceanic and Atmospheric Administration (NOAA). Retrieved 2012, 2012, from http://www.ndbc.noaa.gov/download_data.php?filename=ptat2l2011.txt.gz&dir=data/historical/wlevel/

Connor, J. T., Martin, R. D., & Atlas, L. E. (1994). Recurrent neural networks and robust time series prediction. *Neural Networks, IEEE Transactions on, 5*(2), 240-254.

Coulibaly, P., Anctil, F., Aravena, R., & Bobee, B. (2001). Artificial neural network modeling of water table depth fluctuation. *Water Resources Researches, 37*(4), 885-896.

Craven, M., P,. (1997). *A Faster Learning Neural Network Classifier Using Selective Backpropagation.* Paper presented at the Fourth IEEE International Conference on Electronics, Circuits and Systems,, Cairo, Egypt.

Curilem, G., Vergara, J., Fuentealba, G., Acuña, G., & Chacón, M. (2009). Classification of seismic signals at Villarrica volcano (Chile) using neural networks and genetic algorithms. *Journal of Volcanology and Geothermal Research, 180*(1), 1-8.

Daizhan, C. (2009). Input-State Approach to Boolean Networks. *Neural Networks, IEEE Transactions on, 20*(3), 512-521.

Dasgupta, B., & Schnitger, G. (1993). *The Power of Approximating: a Comparison of Activation Functions* Paper presented at the Advances in Neural Information Processing Systems, San Mateo, CA. .

David, J., Christini,, Abhijit, K., Srika. R., Eric R. Stutman, Frederick M. Bennett, W Nancy Oriol, F. A., & Lutchen., K. R. (1995). Influence of Autoregressive Model Parameter Uncertainty on Spectral Estimates of Heart Rate Dynamics *Annals of Biomedical Engineering, 23*.

Davidović., T. (2012). Bee colony optimization for scheduling independent tasks to identical processors. *Journal of Heuristics, 18*(4), 549-569.

Donovan, A. (2012). Earthquakes and Volcanoes: Risk from Geophysical Hazards. In S. Roeser, R. Hillerbrand, P. Sandin & M. Peterson (Eds.), *Handbook of Risk Theory* (pp. 341-371): Springer Netherlands.

Dorigo, M., T. Stützle, (2005). Ant Colony Optimization (2004) MIT Press,Cambridge, MA 300 pp. *Artificial Intelligence, 165*(2), 261-264.

Dorigo, M. (1992). *Optimization, Learning and Natural Algorithms.*

Dreiseitl, S., & Ohno, M., Lucila. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics, 35*(5–6), 352-359.

Dreyfus, G. (2005). Neural Networks: An Overview *Neural Networks* (1-83): Springer Berlin Heidelberg.

Drndarevic, D. (2006). *Influence of Inputs in Modelling by Backpropagation Neural Networks.* Paper presented at the Neural Network Applications in Electrical Engineering, 2006. NEUREL 2006. 8th Seminar on.

Dubois, D. (1998). Boolean Soft Computing by Non-linear Neural Networks With Hyperincursive Stack Memory. In O. Kaynak, L. Zadeh, B. Türkşen & I. Rudas (Eds.), *Computational Intelligence: Soft Computing and Fuzzy-Neuro Integration with Applications* (162(0), pp. 333-351): Springer Berlin Heidelberg.

Duch, W., & Jankowski, N. (1999). Survey of neural transfer functions. *Neural Computing Application, 2*, 163–212.

Elizondo, D. (2006). The linear separability problem: some testing methods. *Neural Networks, IEEE Transactions on, 17*(2), 330-344.

Esteban, Álvarez, F, Troncoso, A, Justo, J, L, Rubio-Escudero, C, (2010). Pattern recognition to forecast seismic time series. *Expert Systems with Applications, 37*(12), 8333-8342.

Fahlman, S. (1988). *An empirical study of learning speed in backpropagation networks,.* Paper presented at the Technical Report CMU-CS-88-162,.

Fogel, L. J., Owens, A. J., & Walsh, M. J. (1966). *Artificial Intelligence through Simulated Evolution*: John Wiley & Sons

Franco, L. (2006). Generalization ability of Boolean functions implemented in feedforward neural networks. *Neurocomputing, 70*(1–3), 351-361.

Freitas, A., & Lavington, S. (2000). Data Mining Tools *Mining Very Large Databases with Parallel Processing* (Vol. 9, pp. 51-57): Springer US.

Gallant, S. I. (1993). *Neural Network learning and expert system*: MIT press.

Gang, C., Yong,, Yang, F., jie,, & Sun, J., gui,. (2006). A new Particle swam optimization Algorithm. *journal of Jilin University, 24*(2), 181-183.

Gao, W., & Liu, S. (2011). Improved artificial bee colony algorithm for global optimization. *Information Processing Letters, 111*(17), 871-882.

Gao, W., Liu, S., & Huang, L. (2012). A global best artificial bee colony algorithm for global optimization. *Journal of Computational and Applied Mathematics, 236*(11), 2741-2753.

Ghaffari, Abdollahi, H. Khoshayand, A. Rafiee-Tehrani, M. (2006). Performance comparison of neural network training algorithms in modeling of bimodal drug delivery. *International Journal of Pharmaceutics, 327*(1–2), 126-138.

Ghazali, R., Hussain, A., Al-Jumeily, D., & Merabti, M. (2007). Dynamic Ridge Polynomial Neural Networks in Exchange Rates Time Series Forecasting. In B. Beliczynski, A. Dzielinski, M. Iwanowski & B. Ribeiro (Eds.), *Adaptive and Natural Computing Algorithms* (Vol. 4432, pp. 123-132): Springer Berlin Heidelberg.

Ghazali, R., Hussain, A. J., & Liatsis, P. (2011). Dynamic Ridge Polynomial Neural Network: Forecasting the univariate non-stationary and stationary trading signals. *Expert Systems with Applications, 38*(4), 3765-3776.

Gilde, C. (1996). Time Series Analysis And Prediction Using Recurrent Gated Experts.

Giles, C. L., & Maxwell, T. (1987). Learning, invariance, and generalization in high-order neural networks. *Appl. Opt., 26*(23), 4972-4978.

Gopal, S., & Fischer, M. M. (1996). Learning in Single Hidden-Layer Feedforward Network Models: Backpropagation in a Spatial Interaction Modeling Context. *Geographical Analysis, 28*(1), 38-55.

Gu, W., Yin, M., & Wang, C. (2012). Self Adaptive Artificial Bee Colony for Global Numerical Optimization. *IERI Procedia, 1*(0), 59-65.

Gudise, V. G., & Venayagamoorthy, G. K. (2003). *Comparison of particle swarm optimization and backpropagation as training algorithms for neural networks.* Paper presented at the Swarm Intelligence Symposium, 2003. SIS '03. Proceedings of the 2003 IEEE.

Gutierrez, I. Jesús., C. Guillermo., R. Javier., B. Carmen., T. Virginia., & Alvarez, I. (2009). *Volcano-seismic signal detection and classification processing using hidden Markov models. Application to San Cristóbal volcano, Nicaragua.* Paper presented at the Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009.

Habib, S., Ghazali, R., & Nazri, M, Nawi. (2013). Hybrid Global Artificial Bee Colony Algorithm for Classification and Prediction Tasks. *Journal of Applied Sciences Research, 5*(9).

Hadavandi, E. Farhad, M. N, M. & Masoud. (2012). Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction. *Knowledge Based Systems*, *35*(0), 245-258.

Harun, S., Kuok, K., & Siti, M., Shamsuddin,. (2010). Particle swarm optimization feedforward neural network for hourly rainfall- runoff modeling in bedup basin, Malaysia. *International Journal of Civil & Environmental Engineering, 9*(10).

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation* (second edition ed.). Prentice-Hall, Upper Saddle River, NJ.

Helmstetter, A., Kagan, Y. Y., & Jackson., D. D. (2007). High-resolution time-independent grid-based forecast for M= 5 Earthquakes in California. *Seismological Research Letters 78*(1), 78–86.

Hirose, Y., Yamashita, K., & Hijiya, S. (1989). *Backpropagation algorithm which varies the number of hidden units.* Paper presented at the Neural Networks, 1989. IJCNN., International Joint Conference on.

Ho, S. L., Xie, M., & Goh, T. N. (2002). A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. *Computers & Industrial Engineering*, *42*(2–4), 371-375.

Holland, J. H. (1975). Adaptation in Natural and Artificial Systems, . *University of Michigan Press, Ann Arbor, MI,* .

Hongru, L., Wang, S., & Ji, M. (2012). An Improved Chaotic Ant Colony Algorithm. In J. Wang, G. Yen & M. Polycarpou (Eds.), *Advances in Neural Networks – ISNN 2012* (Vol. 7367, pp. 633-640): Springer Berlin Heidelberg.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Netw., 2*(5), 359-366.

Hruschka, E. R., & Ebecken, N. F. (2000). *Using a clustering genetic algorithm for rule extraction from artificial neural networks.* Paper presented at the Combinations of Evolutionary Computation and Neural Networks, 2000 IEEE Symposium on.

Huang, & Chongfu. (2004). *Fuzzy Logic and Earthquake Research*. Burlington: Academic Press.

Husaini, N., Aida,, Ghazali, R., Mohd, N., Nazri,, & Ismail, L., Hakim,. (2011). Jordan Pi-Sigma Neural Network for Temperature Prediction. In T.-h. Kim, H. Adeli, R. Robles & M. Balitanas (Eds.), *Ubiquitous Computing and Multimedia Applications* (151(0), pp. 547-558): Springer Berlin Heidelberg.

Ifrcrcs.(2012). *World Disaster Report*. 17, Chemin des Crêts, P.O. Box 372 CH-1211 Geneva 19, Switzerland: International Federation of Red Cross and Red Crescent Societies.

Irani, R., & Nasimi, R. (2011). Application of artificial bee colony-based neural network in bottom hole pressure prediction in underbalanced drilling. *Journal of Petroleum Science and Engineering, 78*(1), 6-12.

Jain, A. K., Duin, R. P. W., & Jianchang, M. (2000). Statistical pattern recognition: a review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22*(1), 4-37.

Jia, Y., & Dali, Y. (1993). *Analysis of the misadjustment of BP network and an improved algorithm.* Paper presented at the Circuits and Systems, 1993., ISCAS '93, 1993 IEEE International Symposium on.

Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science, 16*(3), 307-354.

Jun, L., & Xiaohong, Q. (2009). *A Novel Hybrid PSO-BP Algorithm for Neural Network Training.* Paper presented at the Computational Sciences and Optimization, 2009. CSO 2009. International Joint Conference on.

Joutsijoki, H., Meissner, K., Gabbouj, M., Kiranyaz, S., Raitoharju, J., Arje, J, & Juhola, M. (2014). Evaluating the performance of artificial neural networks for the classification of freshwater benthic macroinvertebrates. *Ecological Informatics, 20*(0), 1-12

Jung, L., Zne,, & Lee, W., Li,. (2003). A Hybrid Search Algorithm of Ant Colony Optimization and Genetic Algorithm Applied to Weapon-Target Assignment Problems. In J. Liu, Y.-m. Cheung & H. Yin (Eds.), *Intelligent Data*

Engineering and Automated Learning (2690 (0), pp. 278-285): Springer Berlin Heidelberg.

Kagan, Y., Y, D. D, Jackson & Y. Rong. (2007). A testable five-year forecast of moderate and large Earthquakes in Southern California based on smoothed seismicity. *Seismological Research Letters 78*(1), 94-98.

Kagan, Y., Y, & Jackson, D., D. (1994). Long-term probabilistic forecasting of earthquakes,. *Journal of Geophysical Research*, 685–700.

Karaboga, D. (2005). *An Idea Based On Honey Bee Swarn For Numerical Optimzation*. Türkiye: Engineering Faculty Computer Engineering Department,Erciyes University.

Karaboga, D., Akay, B., & Ozturk, C. (2007). Artificial Bee Colony (ABC) Optimization Algorithm for Training Feed-Forward Neural Networks. In V. Torra, Y. Narukawa & Y. Yoshida (Eds.), *Modeling Decisions for Artificial Intelligence* (Vol. 4617, pp. 318-329): Springer Berlin Heidelberg.

Karaboga, D., & Ozturk, C. (2011). A novel clustering approach: Artificial Bee Colony (ABC) algorithm. *Applied Soft Computing, 11*(1), 652-657.

Karaboga, & Akay, B. (2009). A comparative study of Artificial Bee Colony algorithm. *Applied Mathematics and Computation, 214*(1), 108-132.

Karaboga, D. (2005). *An Idea Based On Honey Bee Swarn For Numerical Optimzation*. Türkiye: Engineering Faculty Computer Engineering Department,Erciyes University.

Karaboga, D., & Akay, B. (2009). A comparative study of Artificial Bee Colony algorithm. *Applied Mathematics and Computation, 214*(1), 108-132.

Karaki, S., & Chedid, R. (1994). *Artificial neural networks: a new approach for treating electrical engineering problems.* Paper presented at the Frontiers in Education Conference, 1994. Twenty-fourth Annual Conference. Proceedings.

Karhunen, J., & Joutsensalo, J. (1995). Generalizations of principal component analysis, optimization problems, and neural networks. *Neural Networks, 8*(4), 549-562.

Ke, J., Liu, X., & Wang, G. (2008). Theoretical and Empirical Analysis of the Learning Rate and Momentum Factor in Neural Network Modeling for Stock Prediction.

In L. Kang, Z. Cai, X. Yan & Y. Liu (Eds.), *Advances in Computation and Intelligence* (Vol. 5370, pp. 697-706): Springer Berlin Heidelberg.

Kennedy, J., & Eberhart, R. (1995). *Particle Swarm Optimization.* Paper presented at the Proceeding of IEEE International Conference on Neural Network 4, Australia,.

Khan, A. U., Bandopadhyaya, T. K., & Sharma, S. (2008). *Comparisons of Stock Rates Prediction Accuracy Using Different Technical Indicators with Backpropagation Neural Network and Genetic Algorithm Based Backpropagation Neural Network.* Paper presented at the Emerging Trends in Engineering and Technology, 2008. ICETET '08. First International Conference on.

Khashei, M., & Bijari, M. (2010). An artificial neural network model for timeseries forecasting. *Expert Systems with Applications, 37*(1), 479-489.

Knight, K. (1990). Connectionist ideas and algorithms. *Commun. ACM, 33*(11), 58-74.

Kollias, S. (1990). A study of neural network applications to signal processing. In L. Almeida & C. Wellekens (Eds.), *Neural Networks* (412(0), pp. 233-242): Springer Berlin Heidelberg.

Kunwar, S., Vaisla,, & Ashutosh, K., Bhatt. (2010). An Analysis of the Performance of Artificial Neural Network Technique for Stock Market Forecasting. *International Journal on Computer Science and Engineering, 2*(6).

Labib, R. (1999). *New single neuron structure for solving nonlinear problems.* Paper presented at the Neural Networks, 1999. IJCNN '99. International Joint Conference on.

Lakkos., S. (1994). *A neural network scheme for Earthquake prediction based on the seismic electric signals.* Paper presented at the Neural Networks for Signal Processing [IV. Proceedings of the 1994 IEEE Workshop.

Lan, Y., & Neagu, D. (2012). A New Approach and Its Applications for Time Series Analysis and Prediction Based on Moving Average of n th -Order Difference. In D. Holmes & L. Jain (Eds.), *Data Mining: Foundations and Intelligent Paradigms* (Vol. 24, pp. 157-182): Springer Berlin Heidelberg.

Lawrence, J. (1994). *Introduction to neural networks: design, theory, and applications*: California Scientific Software Press.

Lee, & Jong, B., Park,. (2006). *Application of Particle Swarm Optimization to Economic Dispatch Problem: Advantages and Disadvantages.* Paper presented at the Power Systems Conference and Exposition, 2006. PSCE '06. 2006 IEEE PES.

Lin, A., & Lin, S., Juan,. (2010). Tree-Ring Abnormality Caused by Large Earthquake: An Example From the 1931 M 8.0 Fuyun Earthquake. In M. Stoffel, M. Bollschweiler, D. R. Butler & B. H. Luckman (Eds.), *Tree Rings and Natural Hazards* (Vol. 41, pp. 417-420): Springer Netherlands.

Yue, L., Wang, Y., Li, Y., Zhang, B., & Wu, G. (2004). Earthquake Prediction by RBF Neural Network Ensemble. In F.-L. Yin, J. Wang & C. Guo (Eds.), *Advances in Neural Networks - ISNN 2004* (Vol. 3174, pp. 962-969): Springer Berlin Heidelberg.

Livieris, I. E., & Pintelas, P. (2013). A new conjugate gradient algorithm for training neural networks based on a modified secant equation. *Applied Mathematics and Computation, 221*(0), 491-502.

Lu, Y., Yang, J., Wang, Q., & Huang, Z. (2012). The upper bound of the minimal number of hidden neurons for the parity problem in binary neural networks. *Science China Information Sciences, 55*(7), 1579-1587. doi: 10.1007/s11432-011-4405-6

Luo, X. W. (2011). The Dynamic Ant Colony Optimization Based on Permutation and its Application. *Advanced Materials Research, 179-180*.

Ma, L., (2010). Earthquake Prediction Based on Levenberg-Marquardt Algorithm Constrained Back-Propagation Neural Network Using DEMETER Data. In Y. Bi & M.-A. Williams (Eds.), *Knowledge Science, Engineering and Management* (Vol. 6291, pp. 591-596): Springer Berlin Heidelberg.

Magonlas, D., G,, M. N. Vrahatis, & Androulakis., a. G. S. (1997). Effective backpropagation training with variable stepsize. *Neural Networks, 10*(1), 69-82.

Marques, J., P, , de, Sa. (2001). Neural Networks *Pattern Recognition* (pp. 147-242): Springer Berlin Heidelberg.

Mart, T., A. Riquelme, J. C. (2011). *Computational intelligence techniques for predicting earthquakes*. Paper presented at the Proceedings of the 6th international conference on Hybrid artificial intelligent systems - Volume Part II.

Martínez, A., F. (2011). Computational Intelligence Techniques for Predicting Earthquakes. In E. Corchado, M. Kurzyński & M. Woźniak (Eds.), *Hybrid Artificial Intelligent Systems* (Vol. 6679, pp. 287-294): Springer Berlin Heidelberg.

Masters. (1993). Practical Neural Network Recipies in C++ [Paperback].

Mazumder, P., Patel, J. H., & Fuchs, W. K. (1988). Methodologies for testing embedded content addressable memories. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on, 7*(1), 11-20.

McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology, 5*(4), 115-133.

Medina, J., Mérida-Casermeiro, E., & Ojeda-Aciego, M. (2003). A neural approach to extended logic programs. In J. Mira & J. Álvarez (Eds.), *Computational Methods in Neural Modeling* Springer Berlin Heidelberg.Vol. 2686, pp. 654-661.

Mentzer, J., & Gomes, R. (1994). Further extensions of adaptive extended exponential smoothing and comparison with the M-Competition. *Journal of the Academy of Marketing Science, 22*(4), 372-382.

Merkle, D., & Middendorf, M. (2005). Swarm Intelligence Search Methodologies. In E. K. Burke & G. Kendall (Eds.), (pp. 401-435): Springer US.

Middendorf, M., Reischle, F., & Schmeck, H. (2002). Multi Colony Ant Algorithms. *Journal of Heuristics, 8*(3), 305-320.

Minku, F. L., & Ludermir, T. B. (2008). Clustering and co-evolution to construct neural network ensembles: An experimental study. *Neural Networks, 21*(9), 1363-1379.

Minsky, M., & Papert, S. A. (1969). *Perceptrons: An Introduction to Computational Geometry*, . MA: Cambridge: MIT Press.

Mogi, K. (1984). Temporal variation of crustal deformation during the days preceding a thrust-type great earthquake — The 1944 Tonankai earthquake of magnitude 8.1, Japan. *Pure and Applied Geophysics, 122*(6), 765-780.

Mohammadi, A., & Jazaeri, M. (2007). *A hybrid particle swarm optimization-genetic algorithm for optimal location of svc devices in power system planning.* Paper presented at the 2nd International Universities Power Engineering Conference,.

Mohsen, H., & Zahra, M. (2008). Application of Artificial Neural Networks for Temperature Forecasting *International Journal of Engineering and Applied Sciences 4*(4).

Montana, D. J., & Davis, L. (1989). *Training feedforward neural networks using genetic algorithms*. Paper presented at the Proceedings of the 11th international joint conference on Artificial intelligence - Volume 1.

Moustra, M., Avraamides, M., & Christodoulou, C. (2011). Artificial neural networks for earthquake prediction using time series magnitude data or Seismic Electric Signals. *Expert Systems with Applications, 38*(12), 15032-15039.

Murtagh, F. (1991). Multilayer perceptrons for classification and regression. *Neurocomputing, 2*(5–6), 183-197.

Najaf, Z. M., Barani, G., Abbas, & Hessami., K, M, Reza. (2013). GMDH based back propagation algorithm to predict abutment scour in cohesive soils. *Ocean Engineering, 59*(0), 100-106.

Narasimhan, H. (2009). *Parallel artificial bee colony (PABC) algorithm.* Paper presented at the Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on.

Nawi, N, Mohd., Ransing, M. R., & Ransing, R. S. (2006). *An Improved Learning Algorithm Based on The Broyden-Fletcher-Goldfarb-Shanno (BFGS) Method For Back Propagation Neural Networks.* Paper presented at the Intelligent Systems Design and Applications, 2006. ISDA '06. Sixth International Conference on.

Nazri, M., Nawi, Ransing, R, S., M, Najib,R, Gazali, & H, Norhamreeza, A, (2010). An Improved Back Propagation Neural Network Algorithm on Classification Problems. In Y. Zhang, A. Cuzzocrea, J. Ma, K.-i. Chung, T. Arslan & X. Song (Eds.), *Database Theory and Application, Bio-Science and Bio-Technology* (Vol. 118, pp. 177-188): Springer Berlin Heidelberg.

Negarestani, A., Setayeshi, S., Ghannadi-Maragheh, M., & Akashe, B. (2002). Layered neural networks based analysis of radon concentration and environmental parameters in earthquake prediction. *Journal of Environmental Radioactivity, 62*(3), 225-233.

NOAA. (2012). National Ocean and Atmosphere Adminstartion *National Climate Data Center*, 2012, from http://www.srh.noaa.gov/oun/?n=climate-okc-heatwave

Noman, S., Shamsuddin, S. M., & Hassanien., A. E. (2009). Hybrid Learning Enhancement of RBN Network with Particle Swarm Optimization In A. E. Hassanien (Ed.), *Foundations of Comput. Intel.* (Vol. 1, pp. 381–397): Springer-Verlag Berlin Heidelberg 2009

Noriega, L. (2005). Multilayer Perceptron Tutorial . School of Computing. Staffordshire University.

Nourani, E., Rahmani, A. M., & Navin, A. H. (2012). *Forecasting stock prices using a hybrid Artificial Bee Colony based neural network.* Paper presented at the Innovation Management and Technology Research (ICIMTR), 2012 International Conference on.

Olyaee, S., Hamedi, S., & Dashtban, Z. (2012). Efficient performance of neural networks for nonlinearity error modeling of three-longitudinal-mode interferometer in nano-metrology system. *Precision Engineering, 36*(3), 379-387.

Ozkan, C., Kisi, O., & Akay, B. (2011). Neural networks with artificial bee colony algorithm for modeling daily reference evapotranspiration. *Irrigation Science, 29*(6), 431-441.

Ozturk, C., & Karaboga, D. (2011). *Hybrid Artificial Bee Colony algorithm for neural network training.* Paper presented at the Evolutionary Computation (CEC), 2011 IEEE Congress on.

Pacharne, M., & Nayak, V. (2011). Feature Selection Using Various Hybrid Algorithms for Speech Recognition. In V. Das & N. Thankachan (Eds.), *Computational Intelligence and Information Technology* (250(0), pp. 652-656): Springer Berlin Heidelberg.

Panakkat, A., & Adeli, H. (2008). Recent Efforts in Earthquake Prediction (1990–2007). *Natural Hazards Review, 9*(2), 70-80.

Panduranga, P. P., Rao, D. H., & Deshpande, A. G. (2007). *Fault Tolerance Analysis of Neural Networks for Pattern Recognition.* Paper presented at the Conference on

Computational Intelligence and Multimedia Applications, 2007. International Conference on.

Peng, G., Wenming, C., & Jian, L. (2011). *Global artificial bee colony search algorithm for numerical function optimization.* Paper presented at the Natural Computation (ICNC), 2011 Seventh International Conference on.

Perez, S., T. (2011). Artificial Neural Network in FPGA for Temperature Prediction. In C. Travieso-González & J. Alonso-Hernández (Eds.), *Advances in Nonlinear Speech Processing* (7015(0), pp. 104-110): Springer Berlin Heidelberg.

Ping, W. (2008). Mechanical Property Prediction of Strip Model Based on PSO-BP Neural Network. *Journal of Iron and Steel Research, International, 15*(3), 87-91.

Plagianakos, V., P, D, G., Sotiropoulos,, & M, N., Vrahatis,. (1998). *An Improved Backpropagation Method with Adaptive Learning Rate* (No. TECHNICAL REPORT No.TR98-02): University of Patras.

Prechelt, L. (1994). *Proben1&mdash,A set of neural network benchmark problems and benchmarking rules*. Fakulta&uml,t fu&uml,r Informatik, Univ. Karlsruhe.

Puspadevi, K. (2008). *Study Of Cost Functions In Three Term Backpropagation For Classification Problems.* Universiti Teknologi Malaysia Johor.

Qi, L., & Tufts, D. W. (1997). Principal feature classification. *Neural Networks, IEEE Transactions on, 8*(1), 155-160.

Qiang, G., Keyu, Q., Yaguo, L., & Zhengjia, H. (2005). *An Improved Genetic Algorithm and Its Application in Artificial Neural Network Training.* Paper presented at the Information, Communications and Signal Processing, 2005 Fifth International Conference on.

Qiuwen, Z., & Cheng, W. (2008). *Using Genetic Algorithm to Optimize Artificial Neural Network: A Case Study on Earthquake Prediction.* Paper presented at the Genetic and Evolutionary Computing, 2008. WGEC '08. Second International Conference on.

Raus, M., & Ameling, W. (1994). A Parallel Algorithm for a Dynamic Eta/ Alpha Estimation in Backpropagation Learning. In M. Marinaro & P. Morasso (Eds.), *ICANN '94* (pp. 639-642): Springer London.

Reyes, J., Morales, E., A,, & Martínez, Á., F. (2013). Neural networks to predict earthquakes in Chile. *Applied Soft Computing*, *13*(2), 1314-1328.

Ricardo, d., A, Araújo. (2011). A class of hybrid morphological perceptrons with application in time series forecasting.*Knowledge-Based Systems*, *24*(4), 513-529.

Romano, M., Liong, S., Yui,, Vu, M., Tue,, Zemskyy, P., Doan, C., Dung,, Dao, M., Ha,, & Tkalich, P. (2009). Artificial neural network for tsunami forecasting. *Journal of Asian Earth Sciences*, *36*(1), 29-37.

Rosenblatt., F. (1958). The Perceptron: A probabilistic model for information storage and organization in the brain. Cornell Aeronautical Laboratory. *65*, 386–408.

Roy, S. (1994). *Factors influencing the choice of a learning rate for a backpropagation neural network.* Paper presented at the Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on.

Rumelhart, D., E,. (1986). Parallel Distributed Processing: Exploration in the Microstruc ture of Cognition,. *MIT Press.*

Rumelhart, D., E,, Hinton, G., E,, & Williams, R., J,. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088).

Sapankevych, N. I., & Sankar, R. (2009). Time Series Prediction Using Support Vector Machines: A Survey. *Computational Intelligence Magazine, IEEE, 4*(2), 24-38.

SCEDC. (2010). Southern California Earthquake Data Center (SCEDC). Retrieved 12 January, 2012, from http://www.data.scec.org/ftp/catalogs/SCSN/2010.catalog

Schultz, K. J. (1997). Content-addressable memory core cells A survey. *Integration, the VLSI Journal*, *23*(2), 171-188.

Serebryakova, O. N. (1992). Electromagnetic ELF radiation from earthquake regions as observed by low‐altitude satellites. *Geophys. Res. Lett., 19*(2), 91-94.

Sexton, R. S., & Gupta, J. N. D. (2000). Comparative evaluation of genetic algorithm and backpropagation for training neural networks. *Information Sciences, 129*(1–4), 45-59.

Shah, H., Rozaida, G., & Nazri, M., Nawi. (2011). Using Artificial Bee Colony Algorithm for MLP Training on Earthquake Time Series Data Prediction. *Journal Of Computing,3*(3),7.

Shah, H., Ghazali, R., & Nawi, N. M. (2012). Hybrid Ant Bee Colony Algorithm for Volcano Temperature Prediction Emerging Trends and Applications in Information Communication Technologies. In B. S. Chowdhry, F. K. Shaikh, D. M. A. Hussain & M. A. Uqaili (Eds.), (Vol. 281, pp. 453-465): Springer Berlin Heidelberg.

Shah, H., R, Gazali, N. M, Nawi. (2012). G-HABC Algorithm for Training Artificial Neural Networks, *International Journal of Applied Metaheuristic Computing, 3*(3), 20.

Shah, H. R, Gazali, N. M, Nawi. (2012). Global Artificial Bee Colony-Levenberq-Marquardt Algorithm (GABC-LM) Algorithm for Classification. *nternational Journal of Applied Evolutionary Computation*.

Shah, H., Ghazali, R., & N. M, Nawi. (2013). Global Artificial Bee Colony Algorithm for Boolean Function Classification. In A. Selamat, N. Nguyen & H. Haron (Eds.), *Intelligent Information and Database Systems* (Vol. 7802, pp. 12-20): Springer Berlin Heidelberg.

Shah, R, Gazali, N. M, Nawi. (2012). Global Hybrid Ant Bee Colony Algorithm for Training Artificial Neural Networks Computational Science and Its Applications – ICCSA 2012. In B. Murgante, O. Gervasi, S. Misra, N. Nedjah, A. Rocha, D. Taniar & B. Apduhan (Eds.), (Vol. 7333, pp. 87-100): Springer Berlin / Heidelberg.

Shamsuddin, S. M., Sulaiman, M. N., & Darus, M. (2001). An improved error signal for the backpropagation model for classification problems. *International Journal of Computer Mathematics, 76*(3), 297-305.

Sharma, A., & Manoria, M. (2006). *A weather forecasting system using concept of soft computing: A new approach.* Paper presented at the International Conference on Advanced Computing and Communications.

Shen, Z., Z, D. D. Jackson, & Y.Y. Kagan. (2007). Implications of geodetic strain rate for future earthquakes, with a five-year forecast of M5 Earthquakes in Southern California. *Seismological Research Letters*, 116–120.

Shi, D., Yang, Ya, l., Y, I,, & Zhi, y., Shan,. (2011). Gbest-guided Artificial Chemical Reaction Algorithm for global numerical optimization. *Procedia Engineering, 24*(0), 197-201.

Shih, H., Tai, Ching-l, L., & Yan-Haw, C. (2009). *Design and Implementation of the Extended Exponentially Weighted Moving Average Control Charts.* Paper presented at the Management and Service Science, 2009. MASS '09. International Conference on.

Shiri., J. (2011). Prediction of Short-Term Operational Water Levels Using an Adaptive Neuro-Fuzzy Inference System. *Journal of Waterway, Port, Coastal, and Ocean Engineering, 137*(6), 344-354.

Shoemaker, P. A., Carlin, M. J., & Shimabukuro, R. L. (1991). Back propagation learning with trinary quantization of weight updates. *Neural Networks, 4*(2), 231-241.

Simon, H. (1994). *Neural networks: a comprehensive foundation*: Macmillan.

Siti, M. S, S. M. Darus and A. S. Shafie. (2008). Improved Backpropagation Error Function for Classification problems. *International Journal of Mathematics and Apllications, 1*(1).

Slowik, A., & Bialko, M. (2008). *Training of artificial neural networks using differential evolution algorithm.* Paper presented at the Human System Interactions, 2008 Conference on.

Smilgyte, K., & Nenortaite, J. (2011). Artificial Neural Networks Application in Software Testing Selection Method. In E. Corchado, M. Kurzyński & M. Woźniak (Eds.), *Hybrid Artificial Intelligent Systems* (Vol. 6678, pp. 247-254): Springer Berlin Heidelberg.

Sonego, P. (2007). A protein classification benchmark collection for machine learning. *Nucleic Acids Res, 35,* D232 - 236.

Sperduti, A., & Starita, A. (1993). Speed up learning and network optimization with extended back propagation. *Neural Networks, 6*(3), 365-383.

Srinivasan, D., & Seow, T. (2005). Particle Swarm Inspired Evolutionary Algorithm (PS-EA) for Multi-Criteria Optimization Problems. In A. Abraham, L. Jain & R.

Goldberg (Eds.), *Evolutionary Multiobjective Optimization* (pp. 147-165): Springer London.

Stork, D. G., & Allen, J. D. (1992). How to solve the N-bit parity problem with two hidden units. *Neural Networks, 5*(6), 923-926.

Stromatias, E. (2011). *Developing a supervised training algorithm for limited precision feed5forward spiking neural networks* Liverpool University.

Sudarshan, R., N, & Moharir, P. S. (1995). Estimation of the order of an auto-regressive model. *Sadhana*, *20*(5), 749-758.

Sun, Y. j. (2007). Improved BP Neural Network for Transformer Fault Diagnosis. *Journal of China University of Mining and Technology, 17*(1), 138-142.

Suratgar, S., F., Salemi, A, H,, & Negarestani, A. (2008). *Magnitude of Earthquake Prediction Using Neural Network.* Paper presented at the Natural Computation, 2008. ICNC '08. Fourth International Conference on.

Suresh, R., Sivagnanam, N., & Chandra, K. S. (1999). Prediction of weather parameters on a very short time scale by an Auto Regressive process for aviation flight planning. *Proceedings of the Indian Academy of Sciences - Earth and Planetary Sciences, 108*(4), 277-286.

Tarun, K., Sharma,, Pant, M., & Bhardwaj, T. (2011). *PSO ingrained Artificial Bee Colony algorithm for solving continuous optimization problems.* Paper presented at the Computer Applications and Industrial Electronics (ICCAIE), 2011 IEEE International Conference on.

Teodorovic, D., (2006). *Bee Colony Optimization: Principles and Applications.* Paper presented at the Neural Network Applications in Electrical Engineering, 2006. NEUREL 2006. 8th Seminar on.

Thompson, J., Plewniak, F., & Poch, O. (1999). BAliBASE: a benchmark alignment database for the evaluation of multiple alignment programs. *Bioinformatics, 15*, 87 - 88.

Thusberg, J., Olatubosun, A., & Vihinen, M. (2011). Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat, 32*, 358 - 368.

Tiampo, K. F., & Shcherbakov, R. (2012). Seismicity-based earthquake forecasting techniques: Ten years of progress. *Tectonophysics, 522–523*(0), 89-121.

Timothy, M. (1993). *Practical neural network recipes in C++*: Boston : Academic Press, 1993.

Trafalis, T., & Kasap, S. (2001). Neural networks for combinatorial optimizationNeural Networks for Combinatorial Optimization. In C. Floudas & P. Pardalos (Eds.), *Encyclopedia of Optimization* (pp. 1679-1687): Springer US.

Trelea, I. C. (2003). The particle swarm optimization algorithm: convergence analysis and parameter selection. *Information Processing Letters, 85*(6), 317-325.

Tseng, L. Y., & Chen, W, C. (2010). Solving Large N-Bit Parity Problems with the Evolutionary ANN Ensemble. In L. Zhang, B.-L. Lu & J. Kwok (Eds.), *Advances in Neural Networks - ISNN 2010* (Vol. 6063, pp. 389-395): Springer Berlin Heidelberg.

Tsunogai, U., & Wakita, H. (1995). Precursory Chemical Changes in Ground Water: Kobe Earthquake, Japan. *Science, 269*(5220), 61-63.

Tuba, M., Bacanin, N., & Stanarevic, N. (2011). *Guided artificial bee colony algorithm*. Paper presented at the Proceedings of the 5th European conference on European computing conference.

Ufnalski, B., & Grzesiak, L. M. (2012). Particle swarm optimization of artificial-neural-network-based on-line trained speed controller for battery electric vehicle, *Bulletin of the Polish Academy of Sciences: Technical Sciences* (Vol. 60, pp. 661).

Vihinen, M. (2012). How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis. *BMC Genomics, 13*(Suppl 4), S2.

Von, L., A., (1988). *Factors influencing learning by backpropagation.* Paper presented at the Neural Networks, 1988., IEEE International Conference on.

Wang, X., G., (2004). An improved backpropagation algorithm to avoid the local minima problem. [doi: 10.1016/j.neucom.2003.08.006]. *Neurocomputing, 56*(0), 455-460.

Wasserman, P. D. (1989). *Neural computing: Theory and practice*: Van Nostrand Reinhold Co.

Wasserman, P. D., & Schwartz, T. (1988). Neural networks. II. What are they and why is everybody so interested in them now? *IEEE Expert, 3*(1), 10-15.

Weiyang, Z. (1999). Verification of the nonparametric characteristics of backpropagation neural networks for image classification. *Geoscience and Remote Sensing, IEEE Transactions on, 37*(2), 771-779.

William, L. (2011). Complexity in Earthquakes, Tsunamis, and Volcanoes, and Forecast, Introduction to. In R. A. Meyers (Ed.), *Extreme Environmental Events* (pp. 68-78): Springer New York.

Xhemali, D., Hinde, C. J., & Stone, R. G. (2009). Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages. *IJCSI International Journal of Computer Science Issues, 4*(1).

Xiao, Y., Xuemei, S., & Zheng, Y. (2009). *Improved Ant Colony Optimization with Particle Swarm Optimization Operator Solving Continuous Optimization Problems.* Paper presented at the Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on.

Xiaodong, L. (2001). *Comparison of neural networks and an optical thin-film multilayer model for connectionist learning.* Paper presented at the Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on.

Xiaoxia, Z. (2010). *A modified particle swarm optimization with differential evolution mutation.* Paper presented at the Natural Computation (ICNC), 2010 Sixth International Conference on.

Xin, Y., & Yong, L. (1997). A new evolutionary system for evolving artificial neural networks. *Neural Networks, IEEE Transactions on, 8*(3), 694-713.

Yamada, T., & Yabuta, T., (1994). *Remarks on neural network controller using different sigmoid functions.* Paper presented at the Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference.

Yang, X., Hiroyuki, K., & Wei, Z. (2009). Back Propagation Wavelet Neural Network Based Prediction of Drill Wear from Thrust Force and Cutting Torque Signals. *Computer and Information Science 2*(3).

Ying, L., Jingsheng, W., & Lixin, W. (2012). *Collaborative optimization based on particle swarm optimization and chaos searching.* Paper presented at the Control Conference (CCC), 2012 31st Chinese.

Yu, X.H., & Chen, G.A. (1997). Efficient Backpropagation Learning Using Optimal Learning Rate and Momentum. [doi: 10.1016/S0893-6080(96)00102-5]. *Neural Networks, 10*(3), 517-527.

Yue, L., (2004). Earthquake Prediction by RBF Neural Network Ensemble. In F.-L. Yin, J. Wang & C. Guo (Eds.), *Advances in Neural Networks - ISNN 2004* (Vol. 3174, pp. 962-969): Springer Berlin Heidelberg.

Yümlü, S., Gürgen, F. S., & Okay, N. (2005). A comparison of global, recurrent and smoothed-piecewise neural models for Istanbul stock exchange (ISE) prediction. *Pattern Recognition Letters, 26*(13), 2093-2103.

Zaknich, A. (2003). *Neural networks for intelligent signal processing*. World Scientific Pub.: River Edge, NJ.

Zhang, G., Eddy Patuwo, B., & Y. Hu, M. (1998). Forecasting with artificial neural networks:: The state of the art. [doi: 10.1016/S0169-2070(97)00044-7]. *International Journal of Forecasting, 14*(1), 35-62.

Zhang, Y. Hu, Michael, Eddy Patuwo, B, C. Indro, Daniel., (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. [doi: 10.1016/S0377-2217(98)00051-4]. *European Journal of Operational Research, 116*(1), 16-32.

Zhu, G., & Kwong, S. (2010). Gbest-guided artificial bee colony algorithm for numerical function optimization. *Applied Mathematics and Computation, 217*(7), 3166-3173.

Zurada, J. M. (1992). *Introduction to artificial neural systems*: Pws Pub Co.

Zweiri, Y. H., Whidborne, J. F. Althoefer, K. Seneviratne, L. D. (2002). *A new three-term backpropagation algorithm with convergence analysis.* Paper presented at the Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on.