A review of artificial neural network learning rule based on multiple variant of conjugate gradient approaches

*AG Farizawani, M Puteh, Y Marina, A Rivaie
Faculty of Computer and Mathematical Science, Universiti Teknologi MARA, Terengganu, Malaysia
Faculty of Computer and Mathematical Science, Universiti Teknologi MARA, Shah Alam, Malaysia

*farizawani@uitm.edu.my

Abstract. The evolution of Artificial Neural Network (ANN) begins in 1940s when McCulloch and Pitts published research articles in 1943 discussing about the idea of neural network in general. Basically, the concept of ANN has been inspired by biological human brain model. Then, this concept is transformed into a mathematical formulation and lastly become a machine learning used to solve many problems in this world. Mathematic formulations, design concept, algorithm and computer program can be constructed from ANN. Artificial neural network had undergone many changes on its algorithm and its execution. Otherwise, areas of applications are numerous involving different techniques and approaches of algorithms. ANN algorithm use optimization techniques as a way to find the best outcome based on the problem to be solved. Conjugate Gradient (CG) is one of the popular optimization practices used in ANN to improve learning algorithm nowadays. Therefore, this paper is intended to find out the function and role of modified conjugate gradient method in neural networks and potential related approaches along the age of its advancement. This paper also projected to give an overview approach of ANN with CG method especially on modified CG and overalls performances of those selected models.

Keywords: Artificial Neural Network, Learning Algorithm, Modified Conjugate Gradient, Optimization

1. Introduction
ANN is a part of machine learning that has become most significant in research and developments today. The concept of machine learning is an ability of the computer to understand the structure of data by using mathematical or statistical model. However, artificial neural network is not just a common process, it needs more process and deals with complicated patterns in large amount of data. This require a deep learning of methodology and systematic algorithm structure. ANN applies several techniques in its execution process such as; supervised learning, unsupervised learning or reinforcement learning. Until now, research on neural network learning algorithm is still being explored and accepted by communities.

This paper is intended to discover the evolvement of ANN learning algorithm using conjugate gradient method and how far the research and development has been made since it had been
discovered. The objectives of this paper are to analyze the related works of ANN learning algorithm based on CG method, its advances until today. This paper also focuses on the implementation of conjugate gradient models in the learning algorithm, the way how those models had been manipulated and how it had improved the overall performance of the outcomes. The paper as well reviews an area of applications and type of CGs used in ANN and the growth of learning algorithm using CGs algorithm.

2. Brief history of artificial neural network

ANN is inspired by biological human brain that consist of up to 60 trillion of interconnected set of neurons to perform network pattern of decision making. Based on this root idea, artificial neural network process begin with very simple interconnected neurons that act as a single processor. The concept of perceptron had been introduced based on Mc Culloch and Pits’ neuron model [56]. The foundation of an ANN is made up from a single layer of input, process and output elements. As a result, from a very basic concept of information processing cycle, ANN then performs as a complex mathematical formulation in order to produce an optimum result for any datasets or problem segments. To complete a network cycle, neuron should be tested in two ways; feed-forward and backward algorithm. Backward algorithm or back propagation is the area most discussed and studied by many researchers before and nowadays.

2.1. The ANN main process

For the beginning, the number of layers usually depends on the complexity of a problem to be solved. There are four main steps in neural network; (i) Initialisation, (ii) Activation, (iii) Weight training and (iv) Iteration used in classification task. However, the steps or activation functions change based on the problem to be solved. The basic information processing elements of neural network begins with neuron connected by links. Each link had its own weight and every neuron consists of more than one weight as well as the adjusted weight. From input neuron, links will move towards the other nodes. This is known as feed-forward (FF) neural network. Each link has numeric and associated weight that connect it through the output. All connected links are called as perceptron. After the input links have been individually weighted, they will be summed together to form as an activation function. In this phase, CG formulations had been applied to find optimum weight connected to the next layer; layer1, (l1) to 2, 3 or up to layerm (l_m). The number of layers in certain network depend on the complexity of a problem to be solved. The following diagram shows an elementary of NN process with feedforward and backpropagate algorithms.

![Artificial Neural Network Process with MCGs Formulations](image-url)

**Figure 1.** Artificial neural network process with MCGs formulations
Meanwhile, as shown in the following diagram, ANN process had been detailed by inserting a CG formulation for each input and proceed with the same formula into the next layer of the network. This level will excluded the error corrections because target of the process is to find optimum output value based on given input value. Application of CG formulations play as main role in order to get better result compare by using any other related NN formulations.

![Feed Forward NN](image)

**Figure 2.** Artificial neural network process in feedforward approach

### 2.2. The error corrections

The error corrections Back Propagation training algorithm is the error corrections where backward process of multilayer network check the fault from the output layer and work backward to the hidden layer. The similar four main steps are also been taken to this level. Normally, CG is use for solving nonlinear functions where back-propagation formula is modify to obtain a new faster learning algorithm. The goal of back-propagation is to update each of the weights in the network so that they could cause the actual output to be closer than the target output, thereby minimizing error for each output neuron and also the network as a whole. The following figure portrayed the process of implementation of MCGs in and backward or backpropagation method. This focus area (BP) in neural network is still evolving and always being discussed until now. Variety of approaches, methods, algorithms, formulations etc can be apply in neural network to find the best solution that will produce a compatible, stable and accurate result among all.

![Back Propagate NN](image)

**Figure 3.** Artificial neural network process in backpropagate approach
3. Review of methodology and materials
This section presents the methodology for review and analysis on the publication papers. A total number of 55 publication papers were collected and analyzed to obtain information based on the number of publication. The reviews are sorted based on modification techniques used, the output result, area of application, and publication year. The published papers involved journals, proceeding conference and technical reports. The keyword ‘artificial neural network using conjugate gradient’ is used to find related papers according to the topic discussed. However, only 32 references are selected which related to CG formulations. Another respective of papers were chosen to show the application areas of ANN. Hence, focus of this paper is to concentrate on the evolvem
tent of CG formulation in ANN most potential aspect to be discovered in future.

3.1. ANN Advancement
Research of ANN are widely improved and evolved since it was discovered in 1950s. Variety of techniques found in ANN to improve performance of learning and not restrict to use CG formulation only. However, many researchers use other approaches of bio-inspired algorithm such as genetic algorithm, particle swarm optimization, bee colony optimization or others.

The following table shows problem and modification technique used by researchers in ANN. Result for each research also is represented to show that those modifications are able to improve the performances of the experimental testing data. The table also shows variety of CG that have been manipulated in ANN problems. It involve with feed-forward and/or with back-propagation algorithms. The performance of implementing CG shows better improvement with accuracy precisely, robustness, reducing errors and faster convergence. Every result paper are summarized and the average are calculated as shown in Table 1.

Table 1. Modification technique and performance of ANN result

| Author | Problem | Technique | Result |
|--------|---------|-----------|--------|
| [22]   | To access CG methods in large-scale typical minimization problems | CGFR, PR, Powell, Shanno-Phua | Result are performed in graph for numerical results on the tested problem. CGR & Powell perform better on 1st test problem. 2nd prob fit better on Shanno-Phuaand Powell. |
| [52]   | A study of NN applications in business - The characteristics Feasibility in buss apps & examples ANN drawback | Basic RFN, FFNN | Reviewed of NN common characteristics and feasibilities in business fields |
| [11]   | To describe the similarities and differences of ANN and AIS to both nervous and immune system | ART, ART1 | ANN show better robustness: – very flexible and damage tolerant, memory: -pattern synaptic strength, learning: - global & local learning rule |
| [20]   | To review the ANN applications in forecasting | ANNs approach | ANNs is: Satisfy the performance in forecasting |
Many factors can affect performance of ANNs

Increase convergence rate and proposed uniform distributed weights performed better than algo in normal distribution. It able to apply on network with different activation function.

Tested on pyrolysis mass spectra for microorganism identification show ANNs more robust and rapid with higher accuracy and cost effectiveness.

RBF showed poorer performance.

MLP & NRBF equally well but NRBF has faster training

The mean of load forecasting using NN was more accurate than traditional procedure. Weather ensemble prediction in NN have strong potential to expand in future.

SBLLM – high speed, min error, good performance, homogeneous behavior

The average result for all 4 classification problems (total time convergence)

Trainbfgs – 4.465s
BFGS – 4.755s
BFGS/AG - 3.180s

Proposed algorithm is generic and easy to implement in all commonly used gradient based optimization processes. It also robust and potentially to enhanced the computational efficiency of the training process.

M-FLANN performed better interpolation (0.03595) and extrapolation (0.2116) ability compare to other NN variants.

ANN has shown better performances; high accuracy, noise tolerance, ease
| Reference | Description |
|-----------|-------------|
| [2]       | To improve the training efficiency of BP-NN algo by adaptively modifying the initial search direction. CG based NN-algo, FRCG, BP-NN. New technique based on CG updates improve about 10-15% by adaptively modify the search direction. It also robust and potentially enhance the computational efficiency. |
| [41]      | To optimize the total structural cost, under constraints related to minimum target for different limit states or performance requirements. NN for mean & standard deviation for dynamic analysis. NN able to represent the relationship between structural responses and intervening random variables also between achieved reliabilities and design parameters. |
| [28]      | To propose modified algs for FFNN by using gradient of performance function (energy & error). MMFNN; traincgAJ, traincgll & traincgak. TraincgAJ performed better result on 3 tested problem with stable result. |
| [8]       | To solve imbalance dataset problems by focusing on optimizing decision boundary of a step function at ANN and applied PSO to train real predicted output. Std ANN, PSO, Modified ANN. Modified ANN with G-Mean trained performed better result compare to conventional ANN. Modified NN: Average : 64.965 |
| [5]       | To study the minimum temperature for Chandigarh city by using past of 10 years data and trained its network to analyzed the result by applying NN models. MLP, BPA. Result shown in graph of accuracy for weather prediction plotted for 10 years. No solid figure/ value given for this study. |
| [8]       | To solve imbalance dataset problems (majority, minority). BPN, PSO, G-mean Train. Measurement of G-mean Train & Test of proposed ANN show an increment compare to conventional ANN. |
| [47]      | To improve ABCA based on BPNN for fast and improved convergence rate of hybrid NN learning method. ABCPNN, GABPNN. ACBPNN more stable compare to GABPNN with average result 96.145% for all four data set. |
| [44]      | To review the performance of BPA and several variations to improve the performance assessed by G-mean. BPA, BP-AG, BP-AGMAL. Result for different problems analysis; Accuracy – BPAGMAL -93.1% Training time- BP-AL – 0.00203 |
| [23]      | To propose a CGNNA in order to achieve a high order accuracy in approximating the second-curvature information of FNN, CGNNA. Average performance result for every CG Models are: PR – 85.7% PR* - 87% MPR* - 87.78% |
To find better accuracy in forecasting by training ANN using several CG models. Numerous of climatic parameter input have been used to test the process.

Performance result in % (average) of error prediction:

- BPCGPR – 3.825%
- BPCGFR – 2.275%
- BPCGPB – 4.01%
- STD BP – 5.47%

To analyze and investigate aspect of QN based on BFGS formula for Hessian approximation with a line search convergent assumption of linearly independent iterates.

Theoretical and numerical aspect of quasi-newton methods are tested based on BFGS update. Modified quasi-Newton are very competitive.

To demonstrate better performance of hybrid technique using LM algorithm

- ABC-BP, ABC-LM, MABCNN, B PNN, APSO_LM

Proposed APSO_LM is much better compare to other algorithms (BPNN, ABCNN, ABC-BP, ABC-LM) in terms of MSE, SD and accuracy.

The convergence speed of trainlm & trainscg higher than other training functions and less number of iterations

To compare the performances of three types of training algorithms for making analysis in medical image of brain hematoma

- GD, CG, QN, trainrp, trainscg, trainlm

The convergence speed of trainlm & trainscg higher than other training functions and less number of iterations

To promote a modified PRPCG algorithm for solving nonsmooth unconstrained convex minimization problem

- CGPRP

PRP is claimed as most effective CG methods for nonsmooth convex optimization problems.

To compare the minimization error in training the prediction models

- GD, Scaled CG, MNN

Scaled CG show faster convergence compared with GD and CG. Able to to train the network to an accuracy level of $10^{-2}$ and $10^{-5}$

To find the best BPA to solve problem of analyzing early detection of DHF

- GD, BQN, CGD, RB and LM

LM showed the best outcome compared to other algorithms for solving the earlier problem detection

4. ANN and MCG, research area

ANN has contribute a lot of growth in many areas of research. In medical and prediction area, ANN is very suitable to be applied. The range of applications is getting wider and among others in medical, mathematical formulations, business, engineering as well as other sectors.

Based on Figure 4, the number of publications in medical and prediction & forecasting have dominated in ANN research area. Factor that contribute to this domination because the nature of NN itself is suit to be implemented in both fields. For the prediction area no matter what in any kind of applications; weather forecast, agriculture, or even in stock market analysis will able to produce better
result by implementing neural network algorithm for their problem solver. Other area such as mathemetic and engineering, neural network can also benefit in analysis and designing a new framework or structure based on requirements of an organization. Meanwhile, other fields such as business, image processing, neurocomputing, meteorology and variety data testing also play as part of neural network research contributors. In image processing area, neural network become one of a main popular field that widely been implemented into this research study. However, image processing focus more on different angle of network architectures factors and approaches such as pattern recognition and data classification. The Self Organizing Mapping (SOM) network in image processing plays as a part of neural network subclasses heritance. By comparing image processing to medical and prediction or forecasting fields, optimization technique with MCGs are more regularly being employed in neural network and became the target of improvement in future studies.

![Figure 4. Application of neural network in multiple areas](image)

5. Result and discussion

5.1. MCGs in ANN contributions

Numerous studies have demonstrated modified conjugate gradient algorithm in neural network. Starting from classical CG and basic Gradient Descend (GD), the area of optimization are broadly used in order to find the most accurate result, minimal cost function and least error square (MSE) in neural network problems. Table 2 shows multiple type of MCGs used in NN problems based on identical research and similar formulations.

| MCGs  | Description                                           | Research Publication |
|-------|-------------------------------------------------------|----------------------|
| GD    | Gradient Descent – basic and initial CG formulation  | [10] [12] [14] [55]  |
| CGFR  | Fletcher-Reeves improve the GD                        | [2] [4] [22] [39]   |
| CGBFGS| Broyden-Fletcher-Goldfarb-Shanno extended formulation of FR | [35] [49]           |
| CGPB  | CG formula with Broyden-Powell                       | [4]                  |
**CGPR**  CG Polak-Ribiére improve the previous formula [22] [23]

**CGPRP**  CG Polak-Ribiére-Polyak update the previous PR method [19]

**CGDP**  CG Descent – Powell alike as Quasi-Newton method [55]

**trainCG**  trainedCG with modification of formulation based on combination of standard CG algorithm [10] [28]

**CGNNA**  CG with combination of NN functions [23]

### 5.1. A Progression of Research Publications

The development of research and study of neural network had grown since 1980s until now. Not restrict to one specific CGs technique only, numerous number of publications evolve synchronize with the growth of modified CG formulation. This happen because the environment of CG itself able to support the optimization function to reach the optimum result of tested data set. Therefore, potential modification algorithm and improvement factors can be done to NN in order to obtain a better outcome among all. To support this claim, Table 3 illustrate the number of publication papers in 10 years gaps. The result shows for the past of 20 years ago, research of feedforward NN was actively made to find the best output role (output gaining for each problem specification). Nevertheless, starting from new century of 2000s, the error corrections function had took place as the priority target of formulation of neural network with CGs optimization in the training algorithm.

**Table 3.** Publication Year of ANN with CGs and Multiple Techniques

| Publication Year | 1980s | 1990s | 2000s | 2010s until recent |
|------------------|-------|-------|-------|-------------------|
| Reference No     | [3] [6] [22] [28] [34] | [1] [11] [15] [20] [22] [31] [32] [45] [46] [52] | [2] [7] [14] [17] [18] [21] [24] [26] [29] [33] [36] [37] [39] [41] [43] [51] [54] | [4] [5] [8] [9] [10] [12] [13] [16] [19] [23] [25] [27] [30] [35] [38] [40] [42] [44] [47] [48] [49] [50] [53] [55] |
| NN Type          | Feedforward NN | Backpropagation NN |
| Problem emphasis | Output gaining | Error corrections |
| Total            | 5              | 10             | 17             | 23             |

### 5.1. Other Modification Techniques in ANN

In spite of using concrete MCG formulation in a research, some researchers have discover a new technique with better solution of improving learning algorithm of neural network. Combination of multiple techniques have led to a new discovery and better performance compare to previous discoveries. Thus, Table 4 show a summarization of the new modification practices in neural network that not only use the conjugate gradient technique, but a hybrid learning algorithm from variety types of other optimization approaches.
### Table 4. Summary of Discovery Modification Technique in ANN

| Author/s | Modification Technique                        | Problem specifications                                                                 | Analysis                                                                 |
|---------|-----------------------------------------------|----------------------------------------------------------------------------------------|--------------------------------------------------------------------------|
| [39]    | **Adaptive gain**                            | To improve the training efficiency of gradient descent method                          | Robust, easy to compute and easy to implement for most nonlinear CG algorithm. |
|         | - modify on the parameter setting.            | To modify the gradient based search direction                                           |                                                                          |
|         | modifies the gradient based search direction  |                                                                                        |                                                                          |
|         | by changing the gain value for each node     |                                                                                        |                                                                          |
| [57]    | **Adaptive regulation**                      | To minimize generalization error, the sum of the bias, the variance and inherent noise | Improved algorithm for adaptation of regularization parameters.          |
|         | - iteratively adapts regulation parameters by |                                                                                        |                                                                          |
|         | minimizing validation error using conjugate   |                                                                                        |                                                                          |
|         | gradient                                     |                                                                                        |                                                                          |
| [58]    | **Cuckoo Search (CS) & Cuckoo Optimization** | To minimize the cost function of MSE between actual and target input                   | ANN-COA produced slightly better result for predicting students’ performance. COA and CS are able to search optimal or near optimal solution in terms of RMSE, MAPE and R. |
|         | Algorithm (COA)                              |                                                                                        |                                                                          |
|         | - minimize a cost function defined as MSE     |                                                                                        |                                                                          |
|         | between actual and target output by adjusting |                                                                                        |                                                                          |
|         | weight and biases.                           |                                                                                        |                                                                          |
| [34]    | **Three-term CGA**                           | To overcome problems of NN search directions if the objective function is scaled       | Promote the sufficient descent and global convergent.                    |
|         | - apply a vector based training algorithm     |                                                                                        |                                                                          |
|         | established the convex function.             |                                                                                        |                                                                          |

### 6. Conclusion

This paper demonstrates the implementation of MCG in ANN and its advancement in many business areas throughout years of publication. Based on the review, it shows numbers of modified conjugate gradient that have been applied in ANN since 1980s with variety of even single modified or combination of several CG algorithms. Majority of ANN research focused on improving training efficiency of learning algorithm. This main goal will be achieved by applying proper formulations and algorithms to increase the performance of overall testing data. Otherwise, the review also performs CG as a main factor of correcting errors in back-propagation algorithm. However, intensive research with experimental testing need to be done in order to prove the performance, effectiveness and efficiency of selected technique. The study can also be extended to explore multiple variant of NN agents and performance factor that can improve or probably could be the best solution in that field. Therefore, this review can be extended in the future to analyse a different characteristics of neural network learning algorithm. Other modification factors such as the network parameters influencer, numbers of neurons and hidden layers, weight adjustment or learning rate can be explored deeper in order to find the most optimal and better result among all. Consequently, to determine the new formulation, techniques and...
models will be the next goal of the following future research by testing it with suitable data set in variety of neural network problem areas.

References

[1] Goh ATC 1995. Back-propagation Neural Networks for Modelling Complex Systems, *Artificial Intelligence in Engineering*, 9(3), pp. 143-151

[2] Abbas YA, Najmaddin AS, Gulnar WS 2009. A Modified Conjugate Gradient Formula for Back Propagation Neural Network Algorithm, *Journal of Computer Sciences*, 5(11) pp. 849-856

[3] Abraham A 2004. Meta-Learning Evolutionary Artificial Neural Networks, *Neurocomputing* 56.

[4] Adiwijaya K 2013. An Improvement of Backpropagation Performance by Using Conjugate Gradient on Forecasting of Air Temperature and Humidity in Indonesia *Far East Journal of Mathematical Sciences*, pp. 57-67

[5] Amanpreet K and Harpreet S 2011. Artificial Neural Networks in Forecasting Minimum Temperature *International Journal of Electronics & Communication Technology*, 2(3) pp. 101-105

[6] Anthony M 1995. Interpolation and Learning in Artificial Neural Networks, International Conference on Neural Networks

[7] Goel AK, Saxena SC, and Bhanot S, 2008. Modified Functional Link Artificial Neural Network, *International Journal of Computer, electrical, Automation, Control and Information Engineering*, pp. 530-538

[8] Adam A, Shapiai I, Ibrahim Z, Khalid M, Chew LC, Jau LW, and Watada J 2010. A Modified Artificial Neural Network Learning Algorithm for Imbalanced Data Set Problem, *International Conference on Computational Intelligence, Communicatin Systems and Networks* pp. 44-48

[9] Adam A, Shapiai I, Ibrahim Z, Khalid M, Chew LC, Jau LW, and Watada J 2012. A Two-Step Supervised Learning Artificial Neural Network for Imbalanced Dataset Problems *Journal of Innovative Computing, Information and Control* pp. 3163-3172

[10] Sharma B and Venugoapan K 2014. Comparison of Neural Network Training Functions for Hematoma Classification in Brain CT Images *Journal of Computer Enggineering*, 16(1), pp.31-35

[11] Dasgupta D 1997. Artificial Neural Networks and Arificial Immune Systems: Similarities and Differences, *IEEE International Conference on Systems, Man, and Cybernatics*. pp. 873-878

[12] Kumar D, Gupta S, and Sehgal P, 2014. Improved Training of Predictiv ANN with Gradient Techniques, *International Multi Conference of Engineers and Computer Scientist*, pp. 978-988

[13] Agliari E, Barra A, Galuzzi A, and Isopi M 2014. Multitasking Attractor Networks with Neuronal Threshold Noise, *Neural Networks*, pp. 19-29

[14] Castillo E, Berdinas BG, Romero OF, Betanzos AA, 2006. A Very Fast learning Method for Neural Networks Based on Sensitivity Analysis. *Journal of Machine Learning Research*, pp. 1159-1182

[15] Moller MF 1990. A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, *Neural Networks* 6(4), pp. 525-533.

[16] Navarro FF, Martinez CH, Guttierrez PA, , and Penna-Barragan JM, and Lopez-Gradanos FL 2012. Parameter Estimation of Q-Gaussian Radial Basis Function Neural Networks with Hybrid Algorithm for Binary Classification, *Neurocomputing*, 75(1), pp.123-134

[17] Geem ZW 2006. Parameter Estimatin for the Nonlinear Muskingum Model using BFGS Technique, *Journal of Irrigation and Drainage Engineering*, pp. 474-478
[18] Di Caro GA, Ducatelle F, and Gambardella LM 2005. AntHocNet: An Adaptive Nature-Inspired Algorithm for Routing in Mobile Hoc Networks, European Transaction on Telecommunications, pp. 443-455

[19] Gonglin Y, Wei Z, and Li G 2014. A Modified Polark-Ribiere-Polyak Conjugate Gradient Algorithm for Nonsmooth Convex Program, Journal of Computational and Applied Mathematics, 255, pp.86-96

[20] Zhang G, Patuwo BE, and Hu MY 1998. Forecasting with Artificial Neural Networks: The State of Art International Journal of Forecasting, 14(1), pp. 35-62

[21] Basheer IA and Hajmeer M 2000. Artificial Neural Networks: Fundamental Computing Design and Application, Journal of Microbiological Methods, 43(1), pp. 3-31.

[22] Navon IM and Legler DM 1987. Conjugate-Gradient Methods for Large-Scale Minimization in Meteorology. American Meteorology Society, 115 pp.1479-1502

[23] Livieris IE, Pintelas P, 2012. An Advanced Conjugate Gradient Training Algorithm Based on a Modified Secant Equation, ISRN Artificial Intelligence, pp.1-9.

[24] Taylor JW and Buizza R 2002. Neural Network Load Forecasting with Weather Ensemble Predictions, IEEE Trans. on Power Systems, 17(3), pp. 626-632.

[25] Wang J, Jiang H and Hu C, and Ma T 2014. Convergence Behavior of Delayed Discrete Cellular Neural Network without Periodic Coefficients, Neural Network, pp. 61-68

[26] Yum JYF and Chow TWS 2000. A Weight Initialization Method for Improving training Speed in Feedforward Neural Network, Neurocomputing, 30(1-4), pp. 219-323

[27] Ko CN 2013. Reinforcement Radial Basis Function Neural Networks with an Adaptive Annealing Learning Algorithm, Applied Mathematics and Computing, 221, pp.503-513

[28] Luma NM and Tawfiq AM (n.d.) Improving Conjugate Method for Training Feed Forward Final Year Project Report, Dept of Mathematics, University of Baghdad retrieved from http://qu.edu.iq/cmjou/wp-content/uploads/2015/03/47617.pdf

[29] Rafiq MY, Bugmann G, and Easterbrook DJ 2001. Neural Network Design for Engineering Applications. Computer & Structures, pp.1541-1552

[30] Rivaie M, Mamat M, June LW, and Mohd I 2012. A New Class of Nonlinear Conjugate Gradient Coefficients with Global Convergence Properties, Applied Mathematics and Computation, 218(22), pp 11323-11332

[31] Moller M 1993. A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Neural Network, 6(4), pp. 525-533

[32] Adya M and Collopy F 1998. How Effective are Neural Networks at Forecasting and Predicting? A Review and Evaluation, Journal of Forecasting, 17, pp. 481-495

[33] Alsmadi MKS, Omar K and Noah SA 2009. Back Propagation Algorithm: The Best Algorithm Among the Multi-Layer perceptron Algorithm, International Journal of Computer Science and Network Security, 9(4), pp.378-383.

[34] Abbas HT 1998. Improved Three-Term Conjugate Gradient Algorithm for Training Neural Network, Journal of Kufa for Mathematics and Computer, 2(3), pp. 17-21

[35] Nazri NM, Rehman MZ, and Ghazali MI 2011. Noise-Induced Hearing Loss Prediction in Malaysian Industrial Workers using Gradient Descent with Adaptive Momentum Algorithm International Review on Computers and Software, 6(5), pp.740-748

[36] Nawi NM, Ransing MR, and Ransing RS 2008. An Improved Conjugate Gradient Based on Learning Algorithm for Back Propagation Neural Networks, International Journal of Information and Mathematical Sciences, 2(6), pp. 2062-2071

[37] Hung NQ, Babel MS, Weesakul S, Tripathi NK. 2008. An Artificial Neural Network Model for Rainfall Forecasting in Bangkok, Thailand Hydrology and Earth System Sciences Discussions European Geosciences Union 5(1), pp. 183-218

[38] Nazri MN, Abdullah K, Rehman MZ, Maslina AA, Herawan TT, and Jemal HJ 2014. An Accelerated Particle Swarm Optimization Based Levenberg Marquardt Back Propagation Algorithm, Lecture Notes in Computer Science pp. 245-253
[39] Nazri MN, Ransing MR, and Ransing RS 2006. An Improved Learning Algorithm Based on the Broyden-Fletcher-Goldfarb-Shanno (BFGS) Method for Back Propagation Neural Networks, *Sixth International Conference on Intelligent Systems Design and Applications*, pp. 152-157

[40] Bigdeli N, Farid Y, Afshar K 2012. A Novel Image Encryption/Decryption Scheme Based on Chaotic Neural Networks *Engineering Applications of Artificial Intelligence*, 25(4), pp. 753-765.

[41] Moller O, Foschi RO, Quiroz LM and Rubinstein M 2009. Structural Optimization for Performance-based Design in Earthquake Engineering: Applications of Neural Networks, *Structural Safety*, 31(6), pp. 490-499.

[42] Parvinder SS and Shalini C 2011. A Comparative Analysis of Conjugate Gradient Algorithms & PSO Based on Neural Network Approaches for Reusability Evaluation of Procedure based Software Systems, *Journal Science*, pp.123-135

[43] Kannan SR 2004. Application of Particle Swarm Optimization Technique and its Variants to Generation Expansion Planning Problem, *Electric Power System Research*, pp.203-210

[44] Saduf MA 2013. Comparative study of Back Propagation learning Algorithms for Neural Networks, *International Journal of Advanced Research in Computer Science and Software Engineering*, pp.1151-1156

[45] Patil S, Henry JW, Rubinefire M, and Stein PD 1993. Neural Network in the Clinical Diagnosis of Acute Pulmonary Embolism Chest, 104(6), 1685-1689

[46] Horikawa SI, Furushashi T, Uchikawa Y 1992. On Fuzzy Modeling Using Fuzzy Neural Networks with the Back-Propagation Algorithm, *IEEE Transaction on Neural Networks*, 3(5), pp. 801-806

[47] Sudarshan N, Partha PS, Das A 2012. Training a Feed Forward Neural Network with Artificial Bee Colony Based Back-Propagation Method, *International Journal of Computer Science and Information Technology* pp. 33-45

[48] Svetlana I, Jelena D, Jelena P, Zorica D 2012. Artificial Neural Network in Evaluation and Optimization of Modified Release Solid Dosage Forms, *Pharmaceutics*, 4(4) pp.531-550.

[49] Adewale TA, Oruh BI 2013. A Theoretical and Experimental Study of the Broyden-Fletcher-Golfarb-Shano (BFGS) Update *African Journal of Mathematics and Computer Science Research*, 6(8) pp.166-176, 2013.

[50] Santhanam T and Subhajini AC 2011. An Efficient Weather Forecasting System using Radial Basis Function Neural Network, *Journal of Computer Science*, pp.962-966

[51] Hager WW, and Zhang H 2005. A Survey of Nonlinear Conjugate Gradient Methods. Pasific Journal of Optimization 2, pp. 35-58.

[52] Li EY 1994. Artificial Neural Networks and Their Business Applications *Information and Management*, 27(5), pp.303-313

[53] Ma Y, Xiao Y, Wei G, and Sun J 2014. Fetal ECG Extraction using Adaptive Functional Link Artificial Neural Network, *Signal and Information Processing Association Annual Summit and Conference* pp. 1-4

[54] Gaur P 2009. Neural Network in Data Mining *Journal of Theoretical and Applied Information Technology*, 5(1), pp. 37-42

[55] Priyo SS, Wibawana HA, Maulana F, Bhatiar N 2017. Performance Comparison Neural Network Models for Dengue Fever Disease Detection, *International Conference on Informatics and Computational Sciences* pp.183-187.

[56] Negnevitsky M 2009. Artificial Intelligence: A Guide to Intelligent Systems. Addison Wesley: Longman Publishing

[57] Goutte C and Larsen J 2000 Adaptive Metric Kernel Regression. *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, 26(1-2), pp. 155-167

[58] Jeng FC, Quang HD, and Hsieh HN 2015. Training Artificial Neural Networks by a Hybrid PSO-CS Algorithm. *Algorithm* 8(2), pp. 292-308