SSURJET

Sir Syed University Research Journal of Engineering & Technology
2023, Vol. 13, No. 2,
https://doi.org/10.33317/ssurj.438

Appropriate Selection for Numbers of Neurons and Layers in a Neural Network Architecture: A Brief Analysis

Asif Aziz[1], Talha Khan[1], Umar Iftikhar[2], Irfan Tanoli[3] and, Asif Khalid Qureshi[3]

[1]*Department of Computer Sciences, School of Engineering Applied Sciences, Bahria University, Karachi, Pakistan*
[2]*Department of Computer and Information Systems Engineering, NED University, Karachi, Pakistan*
[3]*Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Karachi, Pakistan*

Correspondence Author: Talha Khan (talha_khann@hotmail.com)

Received February 03, 2022; Revised June 16, 2023; Accepted June 16, 2023

Abstract

Identification of optimal number of neurons and layers in a proposed neural architecture is very complex for the better results. The determination of the hidden layer number is also very difficult task for the proposed network. The recognition of the effective neural network model in terms of accuracy and precision in results as well as in terms of computational resources is very crucial in the community of the computer scientists. An effective proposed neural network architecture must comprise the appropriate numbers of perceptron and number of layers. Another research gap was also reported by the researchers' community that the perceptron stuck during the training phase in finding minima or maxima for stochastic gradient to solve any engineering application. Therefore, to resolve the problem of selection of neurons and layers an analysis was performed to evaluate the performance of the neural network architecture with different neurons and layers on the same data set. The results have revealed that the justified network architecture would contain justified number of neurons and layers as more is the number of neurons and layers, the more will be needed computational resources and training time. It has been suggested that a neural network architecture should be proposed consisting of minimum 2 to 5 layers. Entropy and Mean square error was considered as a yardstick to measure the neural network architecture performance. Obtained results showed that an effective neural network architecture must initially be simulated or checked with minimum number of instances to evaluate the model under consideration.

Index Terms: *Layers, Neural Network, Neural Network Architecture, Neurons, Perceptron.*

I. INTRODUCTION

Recognition of the optimal number of neurons and number of layers in the proposed neural network architecture can be acknowledged as very complex process as it may contain different number of perceptron and similarly varying number of layers which may have great impact on the results of neural network architecture model [1]. It has been observed that varying number of neurons and layers have great impact on neural network architecture, and hence on the performance results therefore its appropriate selection is mandatory. In a research experiment the number of layers and numbers of neurons have been investigated in feed forward propagation neural network to determine the actual impact of the variations. The particular neural network has been trained so many times with different frequency domains and it was revealed that the variations have influenced the neural network architecture greatly. Entropy and Mean square error have been considered as a yardstick to measure the neural network architecture performance [2]. An algorithm was designed for the automatically adjustment and appropriate selection of neurons and layers in multi-layer neural network architecture. The developed algorithm results were compared with the pursuit learning network. The pursuit learning network can be acknowledged as the most famous modular structure. Results depicted that algorithm performed better for solving the regression problems [3]. Moderation concept of neurons were adopted to update the neurons on input and output of neural network architecture [4]. The developed multifunctional layered network was found to be speedier compared to the traditional backward propagation based neural network. Ninety-seven percent accuracy has been achieved by the multifunctional layered network [5]. The generalization capabilities for three layered recurrent neural networks have been investigated. Three-layer recurrent neural network is made of the feed forward propagation and backward propagation modes. It was observed that the neurons and its weights values are always not equal according to the layers it can be changed [6]. It was highlighted in a research paper that pattern identification can be synthesized by varying the hyper planes. In back propagation based neural network architecture the weight values are varied for the best optimal results. The performance has been drastically enhanced due to the tuning of the weights in back propagation neural network architecture [7]. In a suggested algorithm the parameters of neural network architecture have been modified for the optimal results [8].

Collaborative behavior and activation functions have been studied to determine the correlation of neurons behavior associated to the neural network architecture [9]. Genetic Algorithm has been applied to optimize the parameters of developed Hybrid Fuzzy Wavelet Neural Network integrated with Fuzzy set inference based wavelet neurons and polynomial neural networks. Results have been compared with other existing algorithms [10]. For the perfect fitting of functions high precision multi layered neural networks have been studied. Actually the neural network was expanded by using sigmoid function, sine-function series and variable dumping factors. The suggested expansion enhanced the accuracy multi layered neural network [11]. Utilization of multiple weights values concepts have been introduced with the error tolerance capabilities of neurons. Extended back propagation has been adopted to improve the suggested deep learning method. The efficiency was found to be competitive compared to the recent trends for the working of neural network architecture [12]. A unique neuron model has been developed for the neural network architecture with feed forward propagation. Usually it has been reported that higher order networks are considered as competent in terms of computation of difficult applications. Generalized Multi-Dendrite Product (GMDP). Results have demonstrated that using this suggested network will surely enhance the performance of neural network architecture [13]. Two phase method has been also introduced by the researchers for the optimal selection of number of neurons in hidden layer of neural network architecture. Initially the number of neurons are calculated in first phase by using back propagation. In second phase the neurons are calculated using generalizing capacity [14]. Researchers have also proposed the multi layered neural network architecture that is based on the multiple values of neurons. The back propagation is derivative free. The suggested algorithm has worked better for the classification problems [15]. Multilayer perceptron based neural network architecture has been proposed in the research for the optimal results of the classification [16]. Multi-layer perceptron neural network architecture has been built for multi modal classification application and the network was tuned with the suitable number of applications and layers. It has also been observed that generated output results also depend on some other important factors such as activation functions, epochs number, error in epochs, learning rate, momentum, layers number, neurons number and application attributes, instances number. These all parameters have significant impact on any designed or proposed neural network architecture. Weights estimation for the neurons can also be considered as an important parameter for the better optimal results of any proposed neural network architecture. Activation functions are generally used to estimate the output of the proposed neural network architecture. It can also be categorized into two categories such as linear and non-linear activation functions. In linear functions the output is very complex to understand as it is usually not confined between specified ranges. Linear functions ranges from $-\infty$ to $+\infty$. While on the other hand the non-linear activation functions are frequently used in neural network architectures as it is easy to understand and generalize the models. In differential or derivative functions, the output y varies with the variations in x. The non-linear functions are generally classified according to their curves and output shapes. For the prediction of the probability between 0 and 1 the most frequently non-linear activation function is known as sigmoid activation function. The sigmoid function produces S shape curve as it ranges between 0 and 1. Another non-linear function named as hyperbolic tangent function is acknowledged for the classification application by using neural network architectures. The tangent function output caries form -1 to 1. It usually adopted to classify between different classes. Tangent activation function produces also slightly S shaped curve compared to the sigmoidal curve. Rectified Linear unit function is also applied to resolve the convolutional neural network architecture problems. It ranges from zero to infinity. Rectified linear unit function has a de-merit that it converts the negative values into zero immediately that is complex to map on the graphs as well. Similarly, the use of Gaussian activation functions is also reported in designing the neural network architecture [16]. It has also been studied in past research that neural network architectures that are designed for different purposes such as classification and regression models have significant and direct impact of number of layers and number of neurons. Therefore, for the better optimal results the appropriate selection for the preference of number of layers and neurons are strongly needed. Exhaustive comparative analysis was performed for the complete detailed analysis to determine the RMSE, absolute error, accuracy precision, reliability and computational resources. The output results revealed that the optimal selection of number of neurons and layers are mandatory for the designing of vigorous neural network [17-25]. Many algorithms have been implemented for the selection of neurons and layers in the past researchers using the back propagation neural network.

Table I demonstrated the highlighted research papers which have been completed to analyze the concrete concepts and methodology related to the selection of Artificial Neural Network architecture characteristics. It can be observed by the exhaustive literature review that the optimized architecture for back propagation neural network or feed forward propagation is necessary otherwise the proposed architecture may produce garbage values. In Back propagation neural network, the random and assumptions for the number of layers and neurons make the framework worst as back propagation takes output as feedback and try to remove the error. Therefore, optimization algorithm is needed for the smooth results of neural network.

II. PROBLEM STATEMENT

It was revealed from exhaustive literature review that appropriate selection of number of neurons and layers in proposed multi-layer perceptron is very challenging and complex. If the number of neurons and layers are increased the processing time and computational complexity also increases and requires higher computational resources. Results are varied by varying the number of the neurons and layers. The back propagation neural network and feed forward propagation neural network need appropriate selection of number of neurons and layers. It was also found during the detailed research survey that stochastic gradient stuck in the training phase while finding the minima and maxima. Training stop in neural network can be considered as crucial problem as network doesn't know that when training should be stopped.

The training should be stopped where over fitting is minimized and data set is validated.

Table I: Artificial Neural Network Parameters Optimization Techniques Survey

S. No.	Research	Technique	Features	Domain
1.	L. Thomas, "Discovery of Optimal Neurons and Hidden Layers in Feed-Forward Neural Network" , **2016** [1]	Self- organizing Artificial Neural Network Architecture	A hypothesis was developed to select the optimal number of Layers and Neurons	Self- organizing optimal Neural Network Architecture
2.	I. Shafi, "Impact of Varying Neurons and Hidden Layers in Neural Network Architecture for a Time Frequency Application" , **2006** [2]	Time Frequency Distribution	Entropy and Mean Square Error were used as the Yardstick to measure the NN performance	Improved Architecture
3.	H. Ninomiya, "A Study on Generalization ability of 3-Layer Recurrent Neural Networks" , Proceedin gs of the **2002** [6]	3 layer recurrent Neural Network	Step Functions	Results compared with the feed forward propagation and performed better
4.	K. Shin-ike, "A two Phase Method for determining the number of Neurons in the Hidden Layer of a 3-Layer Neural Network", **2010** [14]	Back propagation method and generalization capacity	Comparison with trial and minimized error	Results were superior to the traditional method
5.	I. Karabayir, O. Akbilgic and N. Tas, "A Novel Learning Algorithm to Optimize Deep Neural Networks: Evolved Gradient Direction Optimizer (EVGO)", **2021** [20]	Gradient based Algorithm for Optimizing Parameters	Evolved Gradient Detection Optimizer	EVGO Performed outclass compared to the existing methods
6.	"Development of Particle Swarm Optimization Based Rainfall-Runoff Prediction Model for Pahang River, Pekan", **2016** [21]	Optimizing Algorithm	Multiple Perceptron (MLP) is type of ANN	AI, Particles were trained and learnt from its own knowledge and neighbor particle knowledge
7.	Lizhen Lu, Shuyu Zhang, "Short-Term Water Level Prediction using Different Artificial Intelligent Models" , **2016** [22]	Intelligent Algorithms for getting optimal results	ANN, SVM, ANFIS	Artificial Intelligent Model
8.	Iztok Fister, Dušan Fister, "A Comprehensive Review of Cuckoo Search: Variants and Hybrids", **2013** [23]	Cuckoo search Algorithm of Optimization and Classification Network	Cuckoo search with variants	Comparative Analysis
10.	L. S. Solanki, S. Singh and D. Singh, "An ANN Approach for False Alarm Detection in Microwave Breast Cancer Detection", **2016** [24]	ANN	Antenna for Biological sensing was designed	Positive False Alarm Detection Negative False Alarm detection

III. PROBLEM STATEMENT

It was revealed from exhaustive literature review that appropriate selection of number of neurons and layers in proposed multi-layer perceptron is very challenging and complex. If the number of neurons and layers are increased the processing time and computational complexity also increases and requires higher computational resources. Results are varied by varying the number of the neurons and layers. The back propagation neural network and feed forward propagation neural network need appropriate selection of number of neurons and layers. It was also found during the detailed research survey that stochastic gradient stuck in the training phase while finding the minima and maxima. Training stop in neural network can be considered as crucial problem as network doesn't know that when training should be stopped. The training should be stopped where over fitting is minimized and data set is validated.

IV. METHODOLOGY

A. Neural Network Architecture Analysis with 8 Neurons and Four Layers (Feed Forward Propagation)-Architecture Model- I

Figure I demonstrated the feed forward propagation based neural network architecture comprising of four layers and eight neurons in each layer. The neural network architecture was designed for the classification purpose. Data set comprised of total nine attributes named as precipitation, maximum temperature, minimum temperature, humidity, wind speed, cloud, wind direction and average temperature. Testing mode was selected 10 cross validations. Sigmoid activation function was used in this feed forward propagation based neural network architecture. Elapsed time was found to be 154.13 second to build the architecture and to train the 731 instances of the data set.

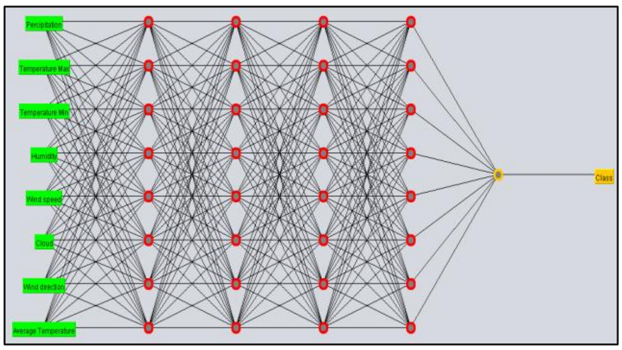

Figure I: The Neural Network Architecture (4 layers, 8 Neurons)

The Mean square error was found to be 0.006715 by using the following equation:

$$\frac{1}{v}\sum_{i=1}^{v} Actual - Calculate \quad (1)$$

Another important parameter epoch was also varied from 200 to 500 but it impacted results not very much. Maximum training speed was achieved 300 instances per second.
Table II shows that the mean squared error has been found to be 0.07231 with elapsed time of 3.213 seconds. Remembering that number of layers has been selected to be four and eight neurons now the neural network architecture would be investigated with minimized number of layers and neurons for the analysis.

Table II: Parametric Evaluation Analysis

S. No.	Parameters	Values
1.	Epochs	200-500
2.	MSE	0.07231
3.	E	0.07715
4.	Training Speed	13000/sec
5.	Elapsed Time	253.71 sec
6.	Learning Rate	0.3
7.	Cross Fold Validation	5-10
8.	Momentum	0.2

B. Neural Network Architecture Analysis with 5 Neurons and Two Layers (Feed Forward Propagation-Architecture Model-II

Figure II shows the neural network architecture comprising of two layers and five neurons. Moreover, the epochs were set to 500. Learning rate and momentum was set to be 0.3 and 0.2 respectively. The sigmoid function was used for the classification multilayer perceptron neural network architecture.

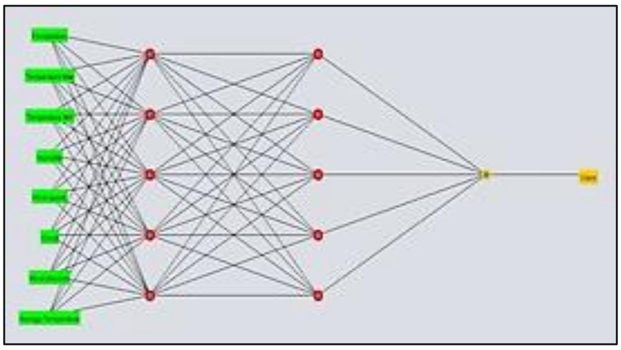

Figure II: The Neural Network Architecture (2 layers, 5 Neurons)

Table III has represented that the mean squared error has been found to be 0.006715with elapsed time of 151.33 sec seconds. Keeping in the mind that number of layers was selected is to be two and five neurons per layer now the neural network architecture would be investigated with minimized number of layers and neurons for the analysis.

Table III: Parametric Evaluation Analysis

S. No.	Parameters	Values
1.	Epochs	200-500
2.	MSE	0.006523
3.	E	0.006715
4.	Training Speed	17000/sec
5.	Elapsed Time	151.33 sec
6.	Learning Rate	0.3
7.	Cross Fold Validation	5-10
8.	Momentum	0.2

C. Neural Network Architecture Analysis with 7 Neurons and Two Layers (Feed Forward Propagation) Architecture Model-III

Figure III shows a multi-layer neural network architecture that was designed on 'WEKA', with 7 neurons in each layer. The neural network architecture consisted of two layers and seven neurons per layer.

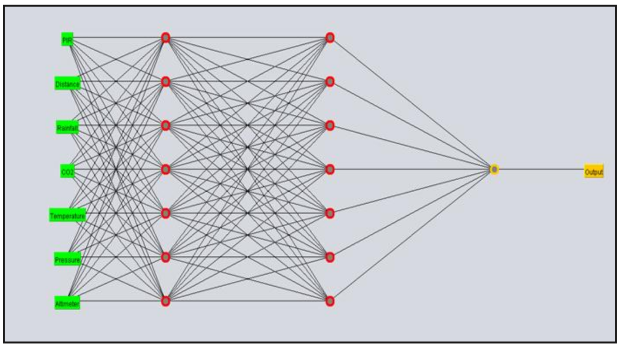

Figure III: The Neural Network Architecture (2 layers, 7 Neurons)

Table IV has represented that the mean squared error is found to be 0.0333with elapsed time of 151.33 sec seconds. Keeping in the mind that number of layers has been selected to be two and five neurons per layer now the neural network architecture would be investigated with minimized number of layers and neurons for the analysis. Some highlighted research papers have been highlighted to have concrete and deep analysis for the neural network architecture parameters.

Table IV: Parametric Investigation Analysis

S. No.	Parameters	Values
1.	Epochs	200-500
2.	MSE	0.0333
3.	E	0.033
4.	Training Speed	23000/sec
5.	Elapsed Time	98.21sec
6.	Learning Rate	0.3
7.	Cross Fold Validation	5-10
8.	Momentum	0.2

Table V has explained that the mean square error and elapsed time can be considered as the major yardstick to measure the mean squared error and elapsed time for the neural network architecture.

Table V: Results and Discussion

S. No.	Architecture Model	MSE	Elapsed Time
1.	Model 1	0.07231	253.71 sec
2.	Model 2	0.06523	98.21sec
3.	Model 3	0.03330	151.33 sec

Results have shown that the variation in the neurons and layers impacted the results in terms of the mean square error and elapsed time. This analysis has been completed to evaluate the algorithm performance variations according to the change of number of layers and neurons. It has been observed that the increasing number of errors unnecessarily has resulted in the increase of absolute error per epoch which will be reflected in the main mean squared error as well that can be acknowledged as the main yardstick to evaluate the performance of neural network architecture. Initially many models were tested with different number of layers and neurons but only these three models were selected for the analysis it revealed some significant changes in the results. It can be noted from the obtained results that the before designing a justified and powerful network architecture, a suitable number of layers and neurons must be selected after getting the simulation results tested before training the whole data set so that the appropriate number of neurons and layers should be selected. Moreover, other tuning parameters epochs, learning rate and momentum must be taken account for the better results.

V. CONCLUSION AND FUTURE ENHANCEMENT

This analysis and investigation depicted results on the basis of these results a self-organizing and self-adjusting neurons algorithm would be developed. A machine learning based approach would be followed in the extension of the research in which optimal selection of neuron and layers would be done by that algorithm. Model 2 was found to be very appropriate design as it has minimum MSE value of 0.06523 with 98.21 seconds. It has been concluded that the appropriate selection is mandatory for the better optimal results of neural network architecture. It was also observed that random selection of weight, learning rate, layers, and neurons through trials and error improvement can be acknowledged as highly efficient and erroneous. Therefore, back propagation strategy is strongly suggested for the better results. Moreover, addition of bias can also be a good option to avoid over fitting and under fitness.

Acknowledgment

The authors would like to thank the management of Bahria University, Karachi, Pakistan, for their support and their assistance throughout this study.

Authors Contributions

The contribution of the authors was as follows: Talha Khan’s contribution to this study was the concept, technical implementation, and correspondence. The methodology to conduct this research work was proposed by Asif Aziz. Other authors (Umar Iftikhar, Irfan Tanoli and, Asif Khalid Qureshi) performed Data collection, supervision, data compilation, project administration, and paper writing.

Conflict of Interest

The authors declare no conflict of interest and confirm that this work is original and not plagiarized from any other source.

Data Availability Statement

The testing data is available in this paper.

Funding

This research received no external funding.

References

[1] Thomas, L., Kumar, M., & Annappa, B. (2016, August). Discovery of optimal neurons and hidden layers in feed-forward Neural Network. In *2016 IEEE International conference on emerging technologies and innovative business practices for the transformation of societies (EmergiTech)* (pp. 286-291). IEEE.

[2] Khare, S. K., & Bajaj, V. (2020). Time–frequency representation and convolutional neural network-based emotion recognition. *IEEE transactions on neural networks and learning systems, 32*(7), 2901-2909.

[3] Shafi, I., Ahmad, J., Shah, S. I., & Kashif, F. M. (2006, December). Impact of varying neurons and hidden layers in neural network architecture for a time frequency application. In *2006 IEEE International Multitopic Conference* (pp. 188-193). IEEE.

[4] Kouhara, T., & Okabe, Y. (1993, October). Learning algorithm based on moderationism for multi-layer neural networks. In *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)* (Vol. 1, pp. 487-490). IEEE.

[5] Maruno, S., Kohda, T., Kojima, Y., Sakaue, S., Yamamoto, H., & Shimeki, Y. (1991, November). Multifunctional layered network with quantizer neurons. In *[Proceedings] 1991 IEEE International Joint Conference on Neural Networks* (pp. 541-546). IEEE.

[6] Ninomiya, H., & Sasaki, A. (2002, May). A study on generalization ability of 3-layer recurrent neural networks. In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)* (Vol. 2, pp. 1063-1068). IEEE.

[7] Abe, S., Kayama, M., & Takenaga, H. (1991, November). Synthesizing neural networks for pattern recognition. In *[Proceedings] 1991 IEEE International Joint Conference on Neural Networks* (pp. 1105-1110). IEEE.

[8] Zhang, M., Xu, S., & Lu, B. (1999, July). Neuron-adaptive higher order neural network group models. In *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339)* (Vol. 1, pp. 333-336). IEEE.

[9] Ishii, N., Deguchi, T., Kawaguchi, M., & Sasaki, H. (2011, May). Colaborative Activities of Layered Neural Network. In *2011 First ACIS/JNU International Conference on Computers, Networks, Systems and Industrial Engineering* (pp. 417-422). IEEE.

[10] Huang, W., Oh, S. K., & Pedrycz, W. (2017). Hybrid fuzzy wavelet neural networks architecture based on polynomial neural networks and fuzzy set/relation inference-based wavelet neurons. *IEEE transactions on neural networks and learning systems, 29*(8), 3452-3462.

[11] Wang, Q., Aoyama, T., & Nagashima, U. (2002, May). High precision expansion of multi-layer neural networks. In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)* (Vol. 2, pp. 1493-1498). IEEE.

[12] Horita, T., Murata, T., & Takanami, I. (2006, October). A multiple-weight-and-neuron-fault tolerant digital multilayer neural network. In *2006 21st IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems* (pp. 554-562). IEEE.

[13] Li, S., Chen, Y., & Leiss, E. L. (1994, June). GMDP: a novel unified neuron model for multilayer feedforward neural networks. In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)* (Vol. 1, pp. 107-112). IEEE.

[14] Shin-ike, K. (2010, August). A two phase method for determining the number of neurons in the hidden layer of a 3-layer neural network. In *Proceedings of SICE Annual Conference 2010* (pp. 238-242). IEEE.

[15] Aizenberg, I., Paliy, D. V., Zurada, J. M., & Astola, J. T. (2008). Blur identification by multilayer neural network based on multivalued neurons. *IEEE Transactions on Neural Networks, 19*(5), 883-898.

[16] Khan, T. A., Alam, M. M., Shahid, Z., & Su’ud, M. M. (2020). Prior recognition of flash floods: concrete optimal neural network configuration analysis for multi-resolution sensing. *IEEE Access, 8*, 210006-210022.

[17] Khan, T. A., Alam, M. M., Shahid, Z., & Su’Ud, M. M. (2020). Investigation of flash floods on early basis: A factual comprehensive review. *IEEE Access, 8*, 19364-19380.

[18] Khan, T. A., Alam, M., Shahid, Z., Ahmed, S. F., & Mazliham, M. S. (2018, November). Artificial Intelligence based Multi-modal

sensing for flash flood investigation. In *2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS)* (pp. 1-6). IEEE.

[19] Khan, T. A., Alam, M., Kadir, K., Shahid, Z., & Mazliham, M. S. (2018, July). A Novel approach for the investigation of flash floods using soil flux and CO 2: An implementation of MLP with less false alarm rate. In *2018 2nd International Conference on Smart Sensors and Application (ICSSA)* (pp. 130-134). IEEE.

[20] Karabayir, I., Akbilgic, O., & Tas, N. (2020). A novel learning algorithm to optimize deep neural networks: evolved gradient direction optimizer (EVGO). *IEEE transactions on neural networks and learning systems, 32*(2), 685-694.

[21] Romlay, M. R. M., Rashid, M. M., & Toha, S. F. (2016, July). Development of Particle Swarm Optimization Based Rainfall-Runoff Prediction Model for Pahang River, Pekan. In *2016 International Conference on Computer and Communication Engineering (ICCCE)* (pp. 306-310). IEEE.

[22] Zhang, S., Lu, L., Yu, J., & Zhou, H. (2016, July). Short-term water level prediction using different artificial intelligent models. In *2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)* (pp. 1-6). IEEE.

[23] Fister Jr, I., Fister, D., & Fister, I. (2013). A comprehensive review of cuckoo search: variants and hybrids. *International Journal of Mathematical Modelling and Numerical Optimisation, 4*(4), 387-409.

[24] Solanki, L. S., Singh, S., & Singh, D. (2016, July). An ANN approach for false alarm detection in microwave breast cancer detection. In *2016 IEEE Congress on Evolutionary Computation (CEC)* (pp. 1370-1374). IEEE.

[25] Khan, T. A., Alam, M., Kadir, K., Shahid, Z., & Mazliham, M. S. (2019). A comparison review based on classifiers and regression models for the investigation of flash floods. *Editorial Preface From the Desk of Managing Editor…, 10*(9).