Optimal Selection for Hidden layers and Neurons parameters in a Neural Network Configuration: Critical Review

Talha Ahmed Khan[1, 2], Zeeshan shahid[3], Umar Iftikar[4], Muhammad Alam[2, 4], Safdar A. Rizvi[1], Ghulam Muhammad[1], M.S Mazliham[2]

[1]Computer Sciences, Bahria University Karachi Campus, Karachi, Pakistan
[2]Multimedia University-MMU, 63100 Cyberjaya, Selangor, Malaysia
[3]Faculty of Engineering and Computing, Al-Ghurair University Dubai, UAE
[4]Department of Computer and Information Systems Engineering, NED University, Karachi
[5]Riphah Institute of System Engineering (RISE), Faculty of Computing, Riphah International University, Islamabad, Pakistan

***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 perceptrons 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 revealed that the justified network architecture would contain justified number of neurons and layers as more number of neurons and layers increase more computational resources and training time. It was suggested that a neural network architecture should be proposed comprising of minimum 2 to 5 layers. Entropy and Mean square error was considered as a yardstick to measure the neural network architecture performance. Results depicted that t an effective neural network architecture must initially be simulated or checked with minimum number of instances to evaluate the model.**

Keywords—Neural Network, Neurons, Layers, Neural Network Architecture, Perceptron

I. Introduction

Recognition of the optimal number of neurons and layers in proposed neural network architecture can be acknowledged as very complex as it may contain different perceptrons and 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 results therefore its appropriate selection is mandatory. In a research experiment the number of layers and numbers of neurons were investigated in feed forward propagation neural network to determine the actual impact of the variations. The particular neural network was 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 was considered as a yardstick to measure the neural network architecture performance [2].

II. Literature Review

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 was achieved by the multifunctional layered network [5]. The generalization capabilities for three layered recurrent neural networks were investigated. Three layer recurrent neural network comprised 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 were varied for the best optimal results. The performance was 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 were modified for the optimal results [8]. Collaborative behavior and activates were studied to determine the correlation of neurons behavior associated to the neural network architecture [9]. Genetic Algorithm was 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 were compared with other existing algorithms [10]. For the perfect fitting of functions high precision multi layered neural networks were 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 were introduced with the error tolerance capabilities of neurons. Extended back propagation was 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 was 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 demonstrated that using this suggested network will surely enhance the performance of neural network architecture [13]. Two phase method was 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 worked better for the classification problems [15]. Multilayer perceptron based neural network architecture was proposed in the research for the optimal results of the classification [16]. Multi-layer perceptron neural network architecture was built for multi modal classification application and the network was tuned with the suitable number of applications and layers.

III. METHODOLOGY

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

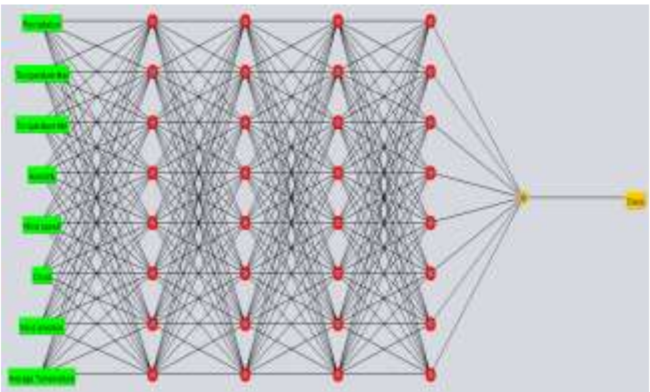

Figure 1. Neural Network Architecture (4 layers, 8 Neurons)

Fig.1 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 validation. 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. The Mean square error was found to be 0.006715 by using the following equation:

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

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 1: Parametric Evaluation Analysis

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

Table no. 1 showed that the mean squared error was found to be 0.07231 with elapsed time of 3.213 seconds. Remembering that number of layers was 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.

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

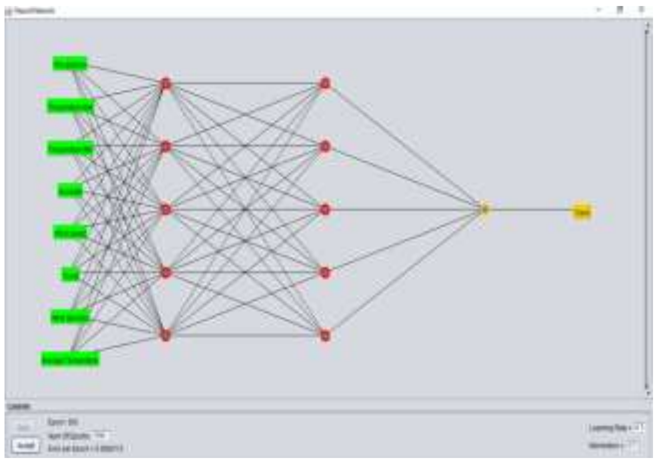

Figure 2. Neural Network Architecture (2 layers, 5 Neurons)

Fig. 2 showed 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

Table 2: Parametric Evaluation Analysis

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

Table no. 2 represented that the mean squared error was found to be 0.006715with elapsed time of 151.33 sec seconds. Keeping in the mind that number of layers was 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.

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

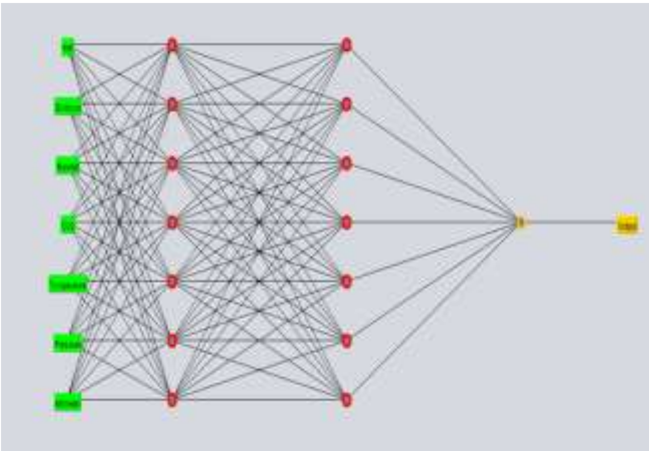

Figure 3. Neural Network Architecture (2 layers, 7 Neurons)

Figure 3 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.

Table 3: Parametric Investigation Analysis

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

Table no. 3 represented that the mean squared error was found to be 0.0333with elapsed time of 151.33 sec seconds. Keeping in the mind that number of layers was 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.

IV. RESULTS

A. Evaluation Metrics

Table 4: Evaluation Metrics

Architecture Model	MSE	Elapsed Time
Model 1	0.07231	253.71 sec
Model 2	0.06523	98.21sec
Model 3	0.03330	151.33 sec

Table 4 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. Result showed that the variation in the neurons and layers impacted the results in terms of the mean square error and elapsed time. This analysis was completed to evaluate the algorithm performance variations according to the change of number of layers and neurons. It was observed that the increasing number of errors unnecessarily will surely increase the 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 was suggested from the results that the before designing a justified and powerful network architecture 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.

Table 5: Critical Review for optimizing Artificial Neural Network Architecture

Research	Technique	Features	Domain
L. Thomas, "Discovery of optimal neurons and hidden layers in feed-forward Neural Network," ***2016*** [1]	Self-organizing ANN 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," *Proceedings 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
9. 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
10. F. Guan, J. Shi, X. Ma, W. Cui and J. Wu, "A Method of False Alarm Recognition Based on k-Nearest Neighbor," **2017** [25]	k-Nearest Neighbor	K-means is simple clustering based	False alarm recognition based on KNN KNN Classification
11. P. Sun, Z. Wu, H. Yang, X. Liu and K. Chen, "Sensors Validation Based on Bayesian Classifiers," **2017** [26]	Bayesian Classifiers	Tree Augmented Naive BAYESIAN CLASSIFIER was applied	Comparative analysis was performed between NBC and TAN
12. .Ankur Kulhari, Avinash Pandey, "Unsupervised Data Classification Using Modified Cuckoo Search Method" **2018** [27]	Cuckoo search with probability distribution	Clustering	Unsupervised data classification
15.Suwannee Phitakwinai, Sansanee Auephanwiriyakul*, "Multilayer Perceptron with Cuckoo Search in Water Level Prediction for Flood Forecasting", **2016** [28]	MLP-CS	Hybrid Algorithm Multilayer perceptron with the combination of cuckoo search	Flood prediction based on water level detection

Table 5 represented the critical review for optimizing the tuning parameters of Artificial Neural Network Architecture.

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. It has been observed from the extensive literature review that regulate the artificial neural network one is number of layers and number of nodes in hidden layers. Experimental analysis revealed that the criteria for the number of selection and neurons can be defined manually by checking the output each time after the design configuration of the neural network. Means testing should be performed for tuning the hyper parameters. But it can be more time taking and inadequate therefore some state of the art Artificial Intelligence based algorithms were suggested by the researchers. In the future the AI based criteria for the selection of neuron and layers will be added in the research.

REFERENCES

[1] L. Thomas, Manoj Kumar M V and Annappa B, "Discovery of optimal neurons and hidden layers in feed-forward Neural Network," *2016 IEEE International Conference on Emerging Technologies and Innovative Business Practices for the Transformation of Societies (EmergiTech)*, 2016, pp. 286-291, doi: 10.1109/EmergiTech.2016.7737353.

[2] I. Shafi, J. Ahmad, S. I. Shah and F. M. Kashif, "Impact of Varying Neurons and Hidden Layers in Neural Network Architecture for a Time Frequency Application," *2006 IEEE International Multitopic Conference*, 2006, pp. 188-193, doi: 10.1109/INMIC.2006.358160.

[3] I. Shafi, J. Ahmad, S. I. Shah and F. M. Kashif, "Impact of Varying Neurons and Hidden Layers in Neural Network Architecture for a Time Frequency Application," *2006 IEEE International Multitopic Conference*, 2006, pp. 188-193, doi: 10.1109/INMIC.2006.358160.

[4] T. Kouhara and Y. Okabe, "Learning algorithm based on moderationism for multi-layer neural networks," *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)*, 1993, pp. 487-490 vol.1, doi: 10.1109/IJCNN.1993.713960.

[5] S. Maruno, T. Kohda, Y. Kojima, S. Sakaue, H. Yamamoto and Y. Shimeki, "Multifunctional layered network with quantizer neurons," *[Proceedings] 1991 IEEE International Joint Conference on Neural Networks*, 1991, pp. 541-546 vol.1, doi: 10.1109/IJCNN.1991.170456.

[6] H. Ninomiya and A. Sasaki, "A study on generalization ability of 3-layer recurrent neural networks," *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*, 2002, pp. 1063-1068 vol.2, doi: 10.1109/IJCNN.2002.1007641.

[7] S. Abe, M. Kayama and H. Takenaga, "Synthesizing neural networks for pattern recognition," *[Proceedings] 1991 IEEE International Joint Conference on Neural Networks*, 1991, pp. 1105-1110 vol.2, doi: 10.1109/IJCNN.1991.170544.

[8] Ming Zhang, Shuxiang Xu and Bo Lu, "Neuron-adaptive higher order neural network group models," *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339)*, 1999, pp. 333-336 vol.1, doi: 10.1109/IJCNN.1999.831513.

[9] N. Ishii, T. Deguchi, M. Kawaguchi and H. Sasaki, "Colaborative Activities of Layered Neural Network," *2011 First ACIS/JNU International Conference on Computers, Networks, Systems and Industrial Engineering*, 2011, pp. 417-422, doi: 10.1109/CNSI.2011.37.

[10] W. Huang, S. Oh and W. Pedrycz, "Hybrid Fuzzy Wavelet Neural Networks Architecture Based on Polynomial Neural Networks and Fuzzy Set/Relation Inference-Based Wavelet Neurons," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3452-3462, Aug. 2018, doi: 10.1109/TNNLS.2017.2729589.

[11] Qianyi Wang, T. Aoyama and U. Nagashima, "High precision expansion of multi-layer neural networks," *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*, 2002, pp. 1493-1498 vol.2, doi: 10.1109/IJCNN.2002.1007738.

[12] T. Horita, T. Murata and I. Takanami, "A Multiple-Weight-and-Neuron-Fault Tolerant Digital Multilayer Neural Network," *2006 21st IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems*, 2006, pp. 554-562, doi: 10.1109/DFT.2006.8.

[13] Shengtun Li, Yiwei Chen and E. L. Leiss, "GMDP: a novel unified neuron model for multilayer feedforward neural networks," *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, 1994, pp. 107-112 vol.1, doi: 10.1109/ICNN.1994.374147.

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

[15] I. Aizenberg, D. V. Paliy, J. M. Zurada and J. T. Astola, "Blur Identification by Multilayer Neural Network Based on Multivalued Neurons," in *IEEE Transactions on Neural Networks*, vol. 19, no. 5, pp. 883-898, May 2008, doi: 10.1109/TNN.2007.914158.

[16] T. A. Khan, M. M. Alam, Z. Shahid and M. M. Su'ud, "Prior Recognition of Flash Floods: Concrete Optimal Neural Network Configuration Analysis for Multi-Resolution Sensing," in *IEEE Access*, vol. 8, pp. 210006-210022, 2020, doi: 10.1109/ACCESS.2020.3038812.

[17] T. A. Khan, M. M. Alam, Z. Shahid and M. M. Su'ud, "Investigation of Flash Floods on Early Basis: A Factual Comprehensive Review," in *IEEE Access*, vol. 8, pp. 19364-19380, 2020, doi: 10.1109/ACCESS.2020.2967496.

[18] T. A. Khan, M. Alam, Z. Shahid, S. F. Ahmed and M. Mazliham, "Artificial Intelligence based Multi-modal sensing for flash flood investigation," *2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, 2018, pp. 1-6, doi: 10.1109/ICETAS.2018.8629147.

[19] T. A. Khan, M. Alam, K. Kadir, Z. Shahid and M. S. Mazliham, "A Novel Approach for the Investigation of Flash Floods using Soil Flux and CO2: An Implementation of MLP with Less False Alarm Rate," *2018 2nd International Conference on Smart Sensors and Application (ICSSA)*, 2018, pp. 130-134, doi: 10.1109/ICSSA.2018.8535606.

[20] I. Karabayir, O. Akbilgic and N. Tas, "A Novel Learning Algorithm to Optimize Deep Neural Networks: Evolved Gradient Direction Optimizer (EVGO)," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 685-694, Feb. 2021, doi: 10.1109/TNNLS.2020.2979121.

[21] M. R. M. Romlay, M. M. Rashid and S. F. Toha, "Development of Particle Swarm Optimization Based Rainfall-Runoff Prediction Model for Pahang River, Pekan," *2016 International Conference on Computer and Communication Engineering (ICCCE)*, 2016, pp. 306-310, doi: 10.1109/ICCCE.2016.72.

[22] S. Zhang, L. Lu, J. Yu and H. Zhou, "Short-term water level prediction using different artificial intelligent models," *2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 2016, pp. 1-6, doi: 10.1109/Agro-Geoinformatics.2016.7577678.

[23] Fister jr, Iztok & Fister, Dušan & Fister, Iztok. (2013). A comprehensive review of Cuckoo search: Variants and hybrids. International Journal of Mathematical Modelling and Numerical Optimisation. 4. 387. 10.1504/IJMMNO.2013.059205.

[24] L. S. Solanki, S. Singh and D. Singh, "An ANN approach for false alarm detection in microwave breast cancer detection," *2016 IEEE Congress on Evolutionary Computation (CEC)*, 2016, pp. 1370-1374, doi: 10.1109/CEC.2016.7743948.

[25] F. Guan, J. Shi, X. Ma, W. Cui and J. Wu, "A Method of False Alarm Recognition Based on k-Nearest Neighbor," *2017 International Conference on Dependable Systems and Their Applications (DSA)*, 2017, pp. 8-12, doi: 10.1109/DSA.2017.11.

[26] P. Sun, Z. Wu, H. Yang and X. Guo, "Sensors Validation Based on TAN Classifiers," *2016 12th International Conference on Computational Intelligence and Security (CIS)*, 2016, pp. 475-478, doi: 10.1109/CIS.2016.0115.

[27] A. Kulhari, A. Pandey, R. Pal and H. Mittal, "Unsupervised data classification using modified cuckoo search method," *2016 Ninth International Conference on Contemporary Computing (IC3)*, 2016, pp. 1-5, doi: 10.1109/IC3.2016.7880262.

[28] S. Phitakwinai, S. Auephanwiriyakul and N. Theera-Umpon, "Multilayer perceptron with Cuckoo search in water level prediction for flood forecasting," *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 519-524, doi: 10.1109/IJCNN.2016.7727243.