Assessing Search and Rescue Optimization based DNN Model for Streamflow Data Prediction

Syed H. Hasan
Iowa State University, Ames, IA

Syeda Huyam Hasan
University of Texas at Arlington, Arlington, TX

Syed Hamid Hasan
Department of Information Systems
Faculty of Computing and Information Technology
King Abdulaziz University, Jeddah, Saudi Arabia

Salman Khalid
King Saud University Medical City Riyadh

ABSTRACT
For many activities related to water resource management, such as flood and drought control, reservoir service, water supply planning and hydroelectric power generation, accurate streamflow prediction is important. While both short- and long-term forecasts are important, reservoir activities are usually planned on the basis of monthly periods; monthly streamflow forecasts therefore play a major role in the management of water resources. Therefore, there is need to propose an efficient approach for prediction of streamflow to improve the system efficiency. Hence, in this paper we have developed an adaptive model based on Search and rescue optimization based DNN for prediction of monthly streamflow. The analysis shows that the adaptive model outperforms existing models such as ANN, SVM and OANN. This AI based learning model shows that this model can able to handle huge number of data for prediction of monthly inflow.

Keywords
Data prediction, Deep Neural Network, Streamflow, Optimization, and monthly inflow

1. INTRODUCTION
Streamflow is an essential part of the water cycle of the Earth and has a critical role in a wide range of applications. The prediction of streamflow has several advantages and it would help to provide reliable, useful and important information in large areas such as water resource management. It has been shown over the years that streamflow can be predicted and predicted using Artificial Intelligence (AI)-based models at different timescales [1]. Streamflow forecasting is one of the most important issues in hydrology and is a key measure in the development and planning of water resources. The forecasting of the river flow alerts the impending stages during the floods and helps to regulate the outflow of reservoirs during low river flows for the management of water resources. Accurate streamflow prediction is essential for the proper management. The proper streamflow forecasting may also help in providing information for city planning, hydroelectric projects, efficient management plans preparation, proactive mitigation programs and real-time operation of water resources projects that decreases the climatic events impact on the environment. Thus, the river flow or streamflow forecasting is very important [2].

The process of streamflow is difficult and not easily predictable. This streamflow process is affected by the huge number of parameters like temperature, evapo-transpiration, land use, precipitation and is described by the non-linear link among the streamflow and the watershed. Streamflow prediction can be categorised as physical models and data-driven models. Physically based models are data-intensive and require a wide range of parameters based on intensity and distribution; rainfall levels, land use; watershed physiographic characteristics; and human activities. Data-driven models mathematically present linear or non-linear relationships between streamflow and its parameters [3]. The streamflow prediction is predicted by means of daily stream basis, monthly basis, annual streamflow, seasonal streamflow prediction, etc. [4].

Monthly streamflow prediction with high and stable performance is of great importance and application value in originating the rational allocation and best water resources management and enhancing the depth and breadth of hydrological forecasting integrated services. Meteorological forecasts coupled with hydrological models, and data-driven methods are two main approaches for monthly streamflow forecasting. Monthly meteorological forecasts such as precipitation and evaporation to drive the hydrological models to achieve the monthly streamflow forecast. Data-driven models based on various machine learning algorithms directly build the relationship between predictors and predictand [5-8]. Hence in this paper we are implementing DNN and S-ROA for improving the efficiency by tuning the parameters of DNN. The paper is organized in the manner such as the section 2 reviews the existing models used in streamflow prediction. The section 3 shows the proposed S-ROA based DNN for streamflow data prediction. The section 4 shows the results and discussion followed by conclusion in the section 5.

2. RELATED WORKS
In this section the table 1 reviews the existing methods used for streamflow prediction, study area, performance measures, and drawbacks.

Table 1: Review of Literature

References	**Method**	**Study area**	**Performance measures**	**Drawbacks/ Future scope**
Khatibi R et al. (2017) [9]	MLP with the Levenberg–Marquardt and MLP integrated with the Fire-Fly Algorithm	Bear River, U.S.A	Taylor diagram and Correlation coefficient	In the future, modelling the formation of science and minimising haphazard modelling practises will be developed.
Jaeger KL et al. (2019) [11]	PROSPER model	Pacific Northwest, US	Accuracy, Error rate	Predictor models shall be included in the future for better capturing of local processes and characterize the streamflow permanence
Keteklahijani VK et al. (2019) [12]	Two Global climate model (GCMs), four downscaling methods (DSMs), and four representative concentration pathways (RCPs)	Karaj dam reservoir, Iran	RMSE	Assembly of various hydrological models to boost accuracy in the future
Chu H et al. (2020) [13]	Fuzzy C-means (FCM) and Deep Belief Networks (DBN) called LASSO-FCM-DBN model	Tennessee River, USA	RMSE correlation coefficient and MAE	An efficient parameter optimization in the future
Tikhamarine Y et al. (2019) [15]	Hybrid wavelet support vector regression based on grey wolf optimizer	Hydrometric station, Algeria	Nash–Sutcliffe efficiency (NSE), correlation coefficient, RMSE, and MAE	• Modified GWO for tuning SVR parameters in future

3. PROPOSED S-ROA BASED DNN FOR STREAMFLOW PREDICTION

The main objective of this paper is to develop an efficient adaptive model for monthly streamflow prediction using search and Rescue optimization Algorithm (S-ROA) based Deep Neural Network (DNN). The proposed model is compared with the existing techniques such as Support vector Machine (SVM) , Artificial Neural Network (ANN) and Optimal Artificial Neural Network (OANN).The data have been collected more than 100 years (1871 to 2000)[17] Aswan High Dam, Egypt from which we have evaluated the performance metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Nash-Sutcliffe coefficient (NSE) and Correlation Coefficient (CC). Here 60% of the data is used for training and remaining 40% of the data is used for testing the proposed model.

3.1 Prediction using Optimal Deep Neural Network

At first, initializing the inputs based on input layer weight α_j and the hidden layer weights β_{ij}. Where, Input I_i represent the database. The input layer consists of neurons. The dataset is used for training network which is denoted as $i_1, i_2, \ldots\ldots\ldots i_n$ and inputs are represented as $W_1, W_2, \ldots\ldots\ldots W_n$. The basis function is calculated using the below equation

$$I_b = \sum_{b=1}^{y} W_i \times \omega_{ab} \quad (1)$$

Where, ω denotes the weight, W represents the input value and 'I' represent the bias function. This layer consists of number of neurons as $h_1, h_2, \ldots\ldots\ldots h_n$ the hidden layers are connected to the output layer by using the neurons. The connection between input parameter d , and the hidden layer, h_1, is represented as

$$h_1 = A(w_1 d + bias_1) \quad (2)$$

Where, w_1 represent the the weight and $bias_1$ represent the bias. The connection among the 'm^{th}' hidden layer, h_m and '$(m-1)^{th}$' hidden layer, h_{m-1}, is afforded as follows,

$$h_m = A(w_m h_{m-1} + bias_m) \quad (3)$$

$$\tilde{o} = R(h_M) \quad (4)$$

At the output layer, the estimation of the network output is obtained as,

$$\tilde{o} = \xi^* = \arg\min_{\xi}\{P(o, \tilde{o}; d, \xi) + \kappa \cdot \gamma(w) + \chi \cdot \phi(s)\} \quad (5)$$

$$\gamma(w) = \sum_{m} \|W_m\|_T^2 \tag{6}$$

The activation parameter is defined as

$$A_f = \sum_{b=1}^{\square} \alpha_b * \left(\frac{1}{1+exp(-\sum_{a=1}^{N} M_a \omega_{ab})}\right) \tag{7}$$

$$O_l = \sum_{i=1}^{n} \alpha_i \, \sigma(F_{i(optimal)}) \tag{8}$$

$$E_i = \sqrt{\frac{\sum_{i=1}^{ND}(D_i - P_i)^2}{ND}} \tag{9}$$

Where, ND is the total number of data used. D is the original range and P is the expected range of output. Thus, to improve the efficiency of DNN we proposing search and rescue optimization for updating those weight.

3.2 Search and Rescue Optimization Algorithm (S-ROA)

The group members collect clues of information during the search operation. A few of these clues remain a group to determine more significant clues but the searching operation is improved via their information. The matrix *N* dimension is similar to matrix *Y*. The number of group member *N* with the problem dimension *d* is denoted as $N \times d$ matrices. Based on the clue's matrix, each new solution in social and individual stages are generated.

$$c = \begin{bmatrix} Y \\ N \end{bmatrix} = \begin{bmatrix} Y_{11} & \cdots & Y_{1d} \\ \vdots & \ddots & \vdots \\ Y_{M1} & \cdots & Y_{Md} \\ N_{11} & \cdots & N_{1d} \\ \vdots & \cdots & \vdots \\ N_{M1} & \cdots & N_{Md} \end{bmatrix} \tag{10}$$

$$Sd_j = (Y_j - CL_i) \qquad i \neq j \tag{11}$$

Here, the j^{th} human position and k^{th} clue positions are denoted as Y_j and CL_i. The search direction of j^{th} human is Sd_j.

$$Y^{`}_{j,k} = \begin{cases} \begin{cases} CL_{i,k} + R_1 \times (Y_{i,k} - CL_{i,k}) & if\ F(CL_i) > F(Y_j) \\ Y_{j,k} + R_2 \times (Y_{j,k} - CL_{i,k}) & Otherwise \end{cases} & if\ R_2 < AE\ or\ k = k_{random},\ k = 1,2,..,d \\ Y_{j,k} & Otherwise \end{cases} \tag{12}$$

For j^{th} human, the k^{th} dimension with its new position is $Y^{`}_{j,k}$. For i^{th} clue, the k^{th} dimension with its position is $CL_{j,k}$. The objective function values of CL_i and Y_j are expressed as $F(CL_i)$ and $F(Y_j)$. The random number R_1 randomly distributed to [-1, 1] interval. Similarly, the random number R_2 randomly distributed to [0, 1] interval.

The third stage is the individual stage where the humans search their current position in the individual stage.

$$Y^{`}_j = Y_j + R_3 \times (CL_i - CL_n), \qquad j \neq i \neq n \tag{13}$$

Hence, the random integer *i* and *n* were distributed to the interval [1, 2M]. The *i* and *n* are selected in such a way that $j \neq i \neq n$ to prevent movement along with other clues. The random integer R_3 randomly tends to the interval [0, 1]. The individual and social stage located in the solution space obtain the solutions.

$$Y^{`}_{j,k} = \begin{cases} (Y_{j,k} + Y_k^{\max})/2 & if\ \ Y^{`}_{j,k} > Y_k^{\max} \\ (Y_{j,k} + Y_k^{\min})/2 & if\ \ Y^{`}_{j,k} > Y_k^{\min} \end{cases} \qquad k = 1,....,d \tag{14}$$

For the k^{th} dimension, the maximum and minimum threshold values are $Y_k^{\max}$ and $Y_k^{\min}$. After each stage, the group member will search in each iteration based on these two stages.

$$N_m = \begin{cases} Y_j & if\ \ F(Y^{`}_j) > F(Y_j) \\ N_m & Otherwise \end{cases} \tag{15}$$

$$Y_j = \begin{cases} Y^{`}_j & if\ \ F(Y^{`}_j) > F(Y_j) \\ Y_j & Otherwise \end{cases} \tag{16}$$

$$U_j = \begin{cases} U_j^{`} + 1 & if \quad F(Y_j^{`}) < F(Y_j) \\ 0 & Otherwise \end{cases} \tag{17}$$

$$Y_{j,k} = Y_k^{\min} + R_4 \times (Y_k^{\max} - Y_k^{\min}), \quad k = 1, \ldots, d \tag{18}$$

Hence, the random number R_4 is distributed to [0, 1] interval and it is varied for all dimensions. In the possible regions, the population is converged in the local optimum. The similarities between them are excessive and the entire population is possible. The solution is possible when the constraint violation degree with a standard deviation is less than to predefined value (β). The restart mechanism is used thereby randomly generates the matrices such as human and memory. Finally, the optimal solution is updated to improve the prediction accuracy of proposed DNN.

4. RESULTS AND DISCUSSION

In this paper we have compared our proposed S-ROA based DNN with other existing methods like SVM, ANN and Optimal ANN to show the monthly inflow prediction. The proposed model is implemented using dataset collected from Egypt [17]. We have used 60% of the data for training and remaining 40% of the data for testing the proposed model. To evaluate the performance of the proposed system metrics such as RMSE, MAE, NSE and correlation coefficient are used.

Correlation Coefficient:

$$C_c = \frac{\sum_{x=1}^{M}(K_{S,x} - K'_S)(K_{T,x} - K^-_T)}{\sqrt{\sum_{x=1}^{M}(K_{S,x} - K'_S)^2 \sum_{x=1}^{M}(K_{T,x} - K^-_T)^2}} (- - 1 < C_c < 1)$$

Root Mean square Error:

$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{x=1}^{M}(K_{S,x} - K_{T,x})^2}\ (0 < RMSE < \infty)$$

Nash Sutcliffe Coefficient:

$$\text{NSE} = \left[\frac{\sum_{x=1}^{M}(K_{S,x} - K_{T,x})^2}{\sum_{x=1}^{M}(K_{S,x} - K'_S)^2}\right] (- - \infty < NSE < 1)$$

Mean Absolute Error:

$$\text{MAE} = \frac{1}{M}\sum_{x=1}^{M} | K_{T,x} - K_{S,x} |\ (0 < MAE < \infty)$$

Where, $K_{S,x}$ is the observed streamflow range. $K_{T,x}$ is the forecasted value. K'_S is the average observed range. K^-_T is the average range of forecasting and M is the total number of data.

Table 2: Performance Evaluation of RMSE, CC, MAE and NSE for proposed S-ROA based DNN

	S-ROA based DNN			
Models	**RMSE**	**CC**	**MAE**	**NSE**
S1	2.8103	0.8821	1.8652	0.7751
S2	2.7269	0.9123	1.1452	0.8693
S3	2.0125	0.9412	1.2128	0.8829
S4	2.2498	0.9389	1.2954	0.8656
S5	2.1681	0.9275	1.3352	0.8748

Table 3: Performance Evaluation of RMSE, CC, MAE and NSE for Existing optimal ANN model

	Optimal ANN			
Models	**RMSE**	**CC**	**MAE**	**NSE**
S1	3.2124	0.8425	2.6742	0.7159
S2	2.2315	0.9364	1.3125	0.8563
S3	2.1589	0.9174	1.3000	0.8505
S4	2.4298	0.9253	1.4012	0.8428
S5	2.3965	0.9197	1.3498	0.8142

Table 4: Performance Evaluation of RMSE, CC, MAE and NSE for Existing ANN model

	ANN			
Models	**RMSE**	**CC**	**MAE**	**NSE**
S1	3.1289	0.8695	2.2956	0.6719
S2	2.5128	0.9022	1.4331	0.8821
S3	2.4396	0.9138	1.7153	0.8382
S4	2.5878	0.9100	1.6891	0.7085
S5	3.3368	0.9178	1.6729	0.7278

Table 5: Performance Evaluation of RMSE, CC, MAE and NSE for Existing SVM model

	SVM			
Models	**RMSE**	**CC**	**MAE**	**NSE**
S1	3.8145	0.7699	2.7352	0.5489
S2	2.6289	0.9178	1.4689	0.8162
S3	2.3947	0.9145	1.3645	0.8311
S4	25825	0.9054	1.3329	0.8277
S5	2.9417	0.8612	1.6895	0.7596

The table 2 to table 5 shows the comparison of performance evaluation of RMSE, CC, MAE and NSE for S-ROA based DNN, OANN, ANN and SVM respectively. The collected data were trained and tested using presented model for all set of combinations. The SROA based DNN outperforms all the existing models in terms of accuracy for both training and testing stage which is followed by OANN, ANN and SVM. The figure 1 (a-d) shows the graphical representation of those performance evaluation metrics and the improvements over the presented model.

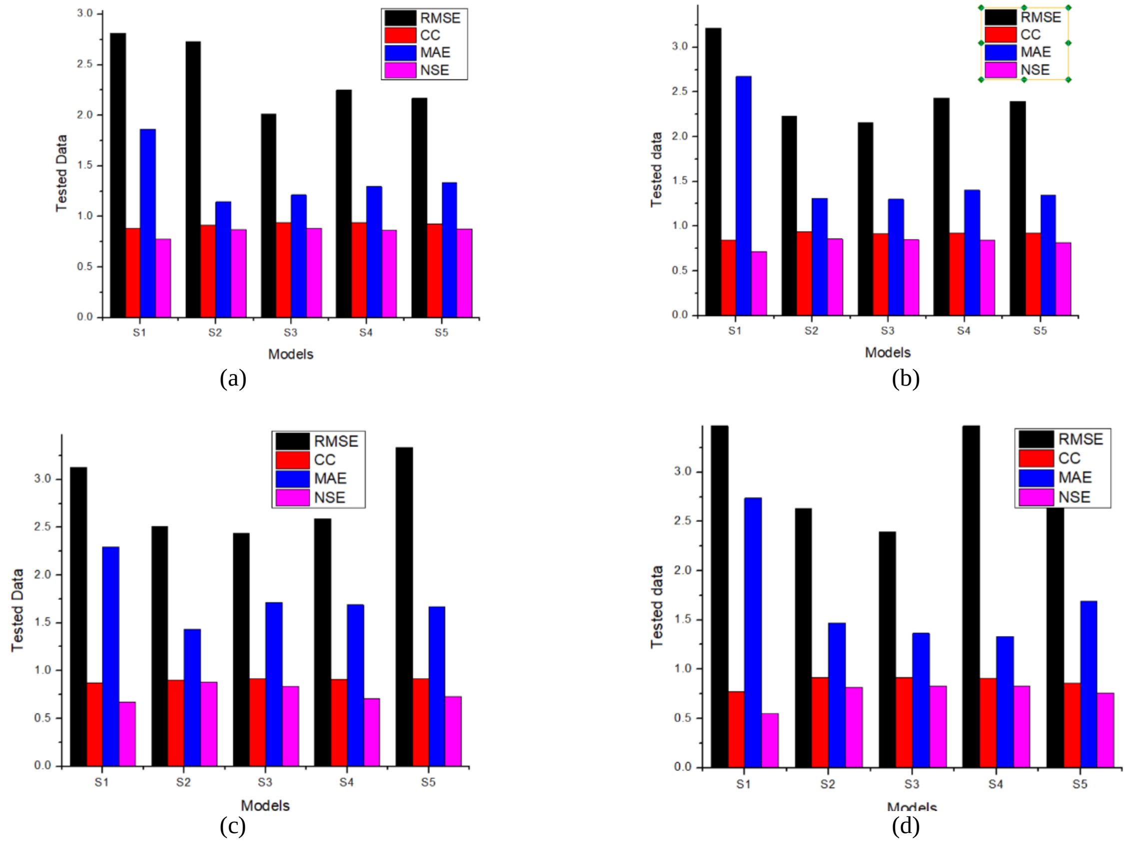

Figure 1: Performance Evaluation of RMSE, CC, MAE and NSE (a) Proposed S-ROA based DNN (b) Optimal ANN (c) ANN Model (d) SVM model

The performance of the S-ROA based DNN and the existing models were summarized in the table. Also, these existing methods providing better accuracy when there is a smaller number of data as the input and our method outperforms those models by showing better accuracy in terms of RMSE, MAE, NSE and CC.

5. CONCLUSION

The proposed S-ROA based DNN was implemented to predict the monthly streamflow. S-ROA model is used to optimize the DNN parameters for improving the prediction result and then compared with OANN, SVM and ANN. The outcome shows that the proposed S-ROA based DNN outperforms all the models in terms of efficiency for the evaluated parameters such as RMSE, MAE, NSE and correlation coefficient. Moreover, search and rescue optimization prove to better algorithm for optimizing the parameters which improves the prediction accuracy. In future research a hybrid model should be presented to improve the efficiency for the prediction of any number of data. In addition, a real-time forecasting model should be implemented by evaluating the time complexity and accuracy of data prediction model.

6. REFERENCES

[1] Adnan RM, Liang Z, Heddam S, Zounemat-Kermani M, Kisi O, Li B. Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. Journal of Hydrology. 2020 Jul 1; 586:124371.

[2] Anusree K, Varghese KO. Streamflow prediction of Karuvannur River Basin using ANFIS, ANN and MNLR models. Procedia Technology. 2016 Jan 1; 24:101-8.

[3] Adnan RM, Liang Z, Trajkovic S, Zounemat-Kermani M, Li B, Kisi O. Daily streamflow prediction using optimally pruned extreme learning machine. Journal of Hydrology. 2019 Oct 1; 577:123981.

[4] Rasouli K, Hsieh WW, Cannon AJ. Daily streamflow forecasting by machine learning methods with weather and climate inputs. Journal of Hydrology. 2012 Jan 11; 414:284-93.

[5] Li Y, Liang Z, Hu Y, Li B, Xu B, Wang D. A multi-model integration method for monthly streamflow prediction: modified stacking ensemble strategy. Journal of Hydroinformatics. 2020 Mar; 22(2):310-26.

[6] Córdoba-Machado S, Palomino-Lemus R, Gámiz-Fortis SR, Castro-Díez Y, Esteban-Parra MJ. Seasonal streamflow prediction in Colombia using atmospheric and oceanic patterns. Journal of Hydrology. 2016 Jul 1; 538:1-2.

[7] Chiang YM, Hao RN, Zhang JQ, Lin YT, Tsai WP. Identifying the sensitivity of ensemble streamflow prediction by artificial intelligence. Water. 2018 Oct;10(10):1341.

[8] Noori N, Kalin L. Coupling SWAT and ANN models for enhanced daily streamflow prediction. Journal of Hydrology. 2016 Feb 1; 533:141-51.

[9] Khatibi R, Ghorbani MA, Pourhosseini FA. Stream flow predictions using nature-inspired firefly algorithms and a multiple model strategy–directions of innovation towards next generation practices. Advanced Engineering Informatics. 2017 Oct 1; 34:80-9.

[10] Hassan M, Zaffar H, Mehmood I, Khitab A. Development of streamflow prediction models for a weir using ANN and step-wise regression. Modeling Earth Systems and Environment. 2018 Sep 1; 4(3):1021-8.

[11] Jaeger KL, Sando R, McShane RR, Dunham JB, Hockman-Wert DP, Kaiser KE, Hafen K, Risley JC, Blasch KW. Probability of Streamflow Permanence Model (PROSPER): A spatially continuous model of annual streamflow permanence throughout the Pacific Northwest. Journal of Hydrology X. 2019 Jan 1; 2:100005.

[12] Keteklahijani VK, Alimohammadi S, Fattahi E. Predicting changes in monthly streamflow to Karaj dam reservoir, Iran, in climate change condition and assessing its uncertainty. Ain Shams Engineering Journal. 2019 Dec 1;10(4):669-79.

[13] Chu H, Wei J, Wu W. Streamflow prediction using LASSO-FCM-DBN approach based on hydro-meteorological condition classification. Journal of Hydrology. 2020 Jan 1;580:124253.

[14] Ghumman AR, Ahmad S, Hashmi HN. Performance assessment of artificial neural networks and support vector regression models for stream flow predictions. Environmental monitoring and assessment. 2018 Dec 1;190(12):704.

[15] Tikhamarine Y, Souag-Gamane D, Kisi O. A new intelligent method for monthly streamflow prediction: hybrid wavelet support vector regression based on grey wolf optimizer (WSVR–GWO). Arabian Journal of Geosciences. 2019 Sep 1;12(17):540.

[16] Hao Z, Hao F, Singh VP. A general framework for multivariate multi-index drought prediction based on Multivariate Ensemble Streamflow Prediction (MESP). Journal of hydrology. 2016 Aug 1; 539:1-0.

[17] Tikhamarine, Y., Souag-Gamane, D., Ahmed, A. N., Kisi, O., & El-Shafie, A. (2020). Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. Journal of Hydrology, 582, 124435.