Paper:

Streamflow prediction using Cascaded-ANFIS algorithm in Kelani River, Sri Lanka

Namal Rathnayake[1], Upaka Rathnayake[2], Tuan Linh Dang[3], and Yukinobu Hoshino[4]

Kochi University of Technology, 185 Miyanokuchi, Tosayamada, Kami, 782-8502, Kochi, Japan. [1,4]

Department of Civil Engineering, Faculty of Engineering, Sri Lanka Institute of Information Technology, Malab, 10115, Sri Lanka.[2]

School of Information and Communications Technology, Hanoi University of Science and Technology, No. 1, Dai Co Viet Road, Hanoi, 100000, Vietnam. [3]

E-mail: namalhappy@gmail.com[1], 246007h@gs.kochi-tech.ac.jp [1],upaka.r@sliit.lk[2], upakasanjeewa@gmail.com[2], linhdt@soict.hust.edu.vn[3], linh.dangtuan@hust.edu.vn[3], hoshino.yukinobu@kochi-tech.ac.jp[4]

[Received 00/00/00; accepted 00/00/00]

Abstract. Sri Lanka is rich in rainfall throughout the entire year, and river overflow incidents occur continuously. Heavy rainfall and poor infrastructure management cause the runoff of rivers. Therefore, this study proposes an accurate rainfall-runoff prediction model development using the Cascaded-ANFIS algorithm. The performance of the proposed method is compared with five algorithms, namely, Linear Regression, Lasso Regression, Ridge Regression, Gated Recurrent Unit, and Long Short-Term Memory. Moreover, the evaluation was done using six parameters: correlation coefficient, root mean squared error, percent bias, Nash-Sutcliffe Efficiency, and Kling-Gupta Efficiency. The proposed algorithm shows 0.9139, 1.15×10^{-10}, 0.2100, 0.8351, and 0.8776 results, respectively. The results indicate that the Cascaded-ANFIS algorithm outperforms other algorithms in every aspect.

Keywords: Cascaded-ANFIS, Regression, GRU, LSTM, RNN, Flood Forecasting.

1. Introduction

Streamflow warning is one of the vital roles in disaster management systems. Due to the heavy rainfall and poor river management system in Sri Lanka, several river runoff events occur yearly. As of 11th June 2021, 2113 people have been affected, and 112 properties have been damaged in the central province of Sri Lanka [2, 10]. In 2021, nine people lost their lives because of the river runoff events. Most of these incidents occur in flood-prone areas of the Kalani, Kalu, and Mahaweli rivers.

An undelayed, accurate, and informative flood warning system can prevent most damage caused by river runoffs. There has been numerous research on preventing the damage of these disasters during the last few decades. However, these methods are still not sophisticated to apply in real-time. Although some models are developed in the literature with acceptable accuracy, the models do not have the capability of real-time usage due to the lack of physical information.

Stream-flow prediction is a trending topic in the scientific world. There are several different approaches to accomplishing this task. The Streamflow and the rainfall possess a direct connection. However, a few other methods are used to develop models for streamflow prediction. In general, researchers tend to use computer-aided tools for the simulation. Some of the frequently used models are the HBV-Light model [19], the J2000 model [13], the soil and water assessment tool [3], and the variable infiltration capacity model [14].

Nevertheless, because of the technological development and easy use of higher grade computer systems, Artificial Intelligence (AI) based hydrological model development shows promising results. Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy models are ubiquitous within these past AI-based approaches [16]. ANN-based methods are used in a vast world area for streamflow forecasting [5, 7, 8, 11].

The main objective of this study is to design and implement a novel model of rainfall-runoff forecasting system based on fuzzy logic. It is well-known that precipitation is high in uncertainty, and rainfall is a highly uncertain variable [1]. Fuzzy has a good reputation for handling uncertainty variables. Moreover, according to work done by Rathnayake et al. [18], the cascaded-ANFIS algorithm shows promising results when dealing with rainfall data. They have proved that the Cascaded-ANFIS outperform LSTM and GRU results when forecasting the hydropower generation.

Therefore, taking these points into account, a newly introduced Cascaded-ANFIS algorithm is used in this study for rainfall-based streamflow forecasting model development [17].

2. Methodology

2.1. Dataset

In Sri Lanka, Kelani River is considered as one of the longest rivers. The river basin of Kelani is around 2230 km^2. The latitude and longitude of the Kelani River are

Table 1. : Descriptive Analysis of Kelani River dataset

	Kitulala	Holombuwa	Deraniyagala	Hanwella	Colombo	Rathmalana	Water Level
count	5201	5201	5201	5201	5201	5201	5201
mean	11.89	8.01	11.55	8.26	6.60	6.90	1.34
std	22.26	17.58	22.33	17.54	17.14	17.16	0.52
min	0	0	0	0	0	0	0.42
25%	0	0	0	0	0	0	1.06
50%	1.9	0.3	1.6	0	0	0.2	1.24
75%	14.6	7.7	14	8.2	5.1045	5.8	1.47
max	336.4	248.4	355.6	289.6	440.2	382.8	7.44

$6°47'$ and $7°05'$ respectively [4]. The yearly discharge and the perception of the river are 7860 MCM, and 3450mm [6]. It is reported that most of the high rainfall is obtained between April to June [6]. As mentioned in the introduction, rainfall-runoff is a significant issue in Sri Lanka. The Kelani River is the most disastrous because it overflows every year, and it is located in a highly-populated area.

The dataset used in the current study was collected from six rain gauges in the Kelani River basin, namely, Kitulgala, Holombuwa, Deraniyagala, Hanwella, Colombo, and Rathmalana. The output water level is measured and documented in the Nagalingam Streat station, which is very close to the end of the Kelani River. A descriptive analysis of the dataset is shown in Table 1. The rainfall data is measured in millimeters (mm), and the water level is measured in centimeters (cm). The range of the data collection period is from 2003/01/01 to 2017/03/01.

2.2. The Cascaded-ANFIS algorithm

Roger introduced the Adaptive Network-based Fuzzy Inference System (ANFIS) in 1993 [12]. ANFIS possesses a vast range of advantages when compared with conventional algorithms. The Cascaded-ANFIS can be introduced as an extended version of ANFIS due to the repetitive usage of the minor structure of the ANFIS algorithm (two-input one-output structure). ANFIS generally is a structure having five layers. The first layer is used for the inputs, and in the next layer, fuzzy membership functions are generated for each of the inputs used. Here, the shape of the membership function and the amount of the membership functions can vary depending on the problem statement and the performance. The third layer calculates the cumulative product and converges the output to the fourth layer. The fourth layer normalizes, and at the fifth layer, the mean of the past layer outputs is calculated and considered the solution. The Equations (1-5) show each layer's calculations respectively.

$$\mu_{Ai} = max(min(\frac{x-a_i}{b_i-a_i}, \frac{c_i-x}{c_i-b_i}), 0), i = 1,2 \quad . \quad (1)$$

$$O_{2,j} = w_j = \mu_{Ai}(x_1) * \mu_{Bi}(x_2) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$O_{3,j} = \bar{w_j} = \frac{w_j}{\sum w_i}, j = 1,2,...,n \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$O_{4,j} = \bar{w_j} f_j = \bar{w_j}(p_j x_1 + q_j x_2 + r_j) \quad . \quad . \quad . \quad . \quad (4)$$

$$O_{5,1} = \sum_{j=1}^{n} \bar{w_j} f_j \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Where, μ_{Ai} is the A^{th} triangular membership function for the input $i = 1,2$ and x, a, b, and c are input values and the premise parameters in the triangular membership function, respectively. $O_{(k,j)}$ is the output of the respective layer k and j^{th} rule. w_j is the firing strength of the j^{th} rule and p, q, and r are the consequent parameters for the defuzzification.

There are several parameters that train during the ANFIS training process. However, depending on the membership function used in the algorithm, the parameter number is changed. As shown in Equation 1, the triangular membership functions have three trainable parameters named a, b, and c.

The construction of the Cascaded-ANFIS can be shown in Figure 1. The Cascaded-ANFIS has two main steps: selecting pairs and training the model using the selected pairs. In every step, a simple two-input one-output ANFIS structure is used. As shown in Figure 1, once the input is fed into the Cascaded-ANFIS algorithm, pairing is executed to find the best match for each input. Then the training process starts with the use of the best matching pairs. At the first iteration, mean accuracy and error are calculated. If the error and accuracy are not in the range of expected values, the algorithm propagates to the second iteration. The specialty in this propagation is that the output from the first iteration is used as the input of the second iteration. Therefore increasing the number of iteration lower the error and increases the accuracy of individual pairs.

According to the current study, there were six inputs and one output. Therefore, at the first iteration, the most suitable match was selected in the pair selection module in the Cascaded-ANFIS. Then at the end of the first iteration, the mean water level was predicted. The accuracy and the root mean squared error were then calculated using Equation 6,7. Propagation of the next iteration was decided according to the calculated results. The output water levels for each selected pair were used in the next iteration. The process continues until the expected accuracy is acquired.

$$Accuracy = \frac{|ObservedValue - ActualValue|}{ActualValue} \times 100 (6)$$

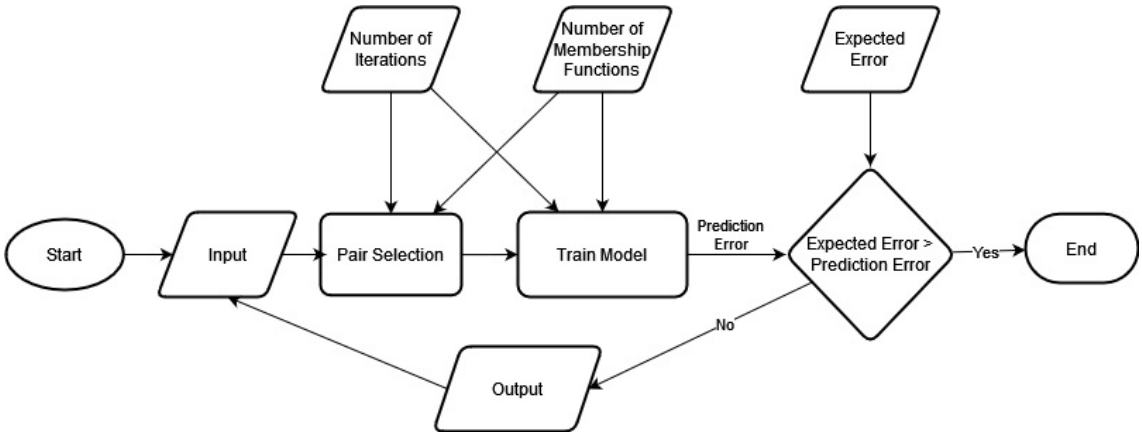


Fig. 1. : Flowchart of the Cascaded-ANFIS algorithm

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Predicted_i - Actual_i)^2}{n}} \quad . \; . \quad (7)$$

2.3. Performance Evaluation

The evaluation of the study is accomplished using several parameters such as Pearson Correlation Coefficient (*R*), Percent Bias (Bias), Root Mean Squared Error (RMSE), Nash-Sutcliffe Efficiency (NSE), and Kling-Gupta Efficiency (KGE).

R represents the goodness of the fit. It varies from -1 to 1, and the best is when it becomes 1. The calculation of the *R* is shown in Equation 8. Here, x_i and y_i are the x variable and y variable values in a sample. $\bar{x}$ and $\bar{y}$ are the mean values of x and y variables. Bias is the trend of how large or small the actual value. The ideal bias value is 0, and 1 becomes the worst. NSE calculates the perfectness of the match between real and prediction. The results of the NSE can vary between minus infinity being the worst and 1 being the ideal [15]. KGE is a combined calculation of three main parameters: NSE, bias, and coefficient of variation. Recently it has been used more in hydrological model performance calculations [9].

$$R = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad . \; . \; . \; . \; . \; . \; . \quad (8)$$

3. Results and Discussion

The proposed model performance is compared with the state-of-the-art five algorithms as mentioned below.

1. Gated Recurrent Unit (GRU)
2. Lasso Regression (LARE)
3. Long Short-Term Memory (LSTM)
4. Linear Regression (LIRE)
5. Ridge Regression (RIRE)

Figures 2 shows the *R* of each algorithm for the used rainfall dataset. It is noticeable that the *R* of the Cascaded-ANFIS algorithm is higher when compared with the other five algorithms. The Cascaded-ANFIS model provides 0.9139 *R* while LSTM, the second-best result, shows 0.9017. Here, the number of iterations used in the experiment is ten for the Cascaded-ANFIS. RIRE shows the lowest *R* value.

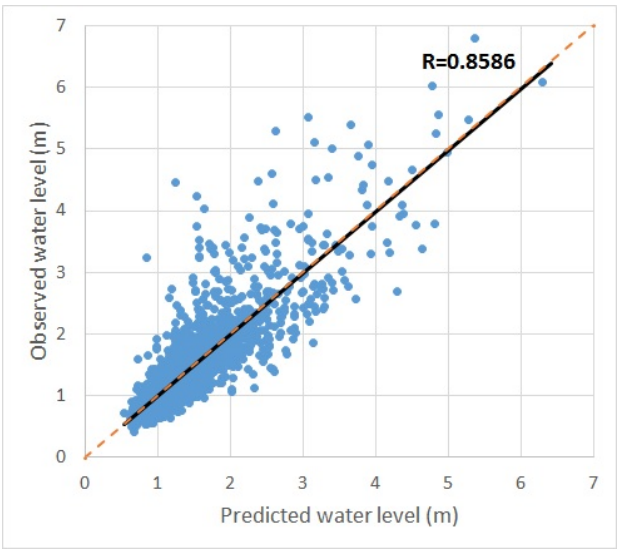
Figure 3a presents the calculation of bias for the considered algorithms. The proposed algorithm shows the most negligible bias having 1.15×10^{-10}, almost close to zero. GRU places the highest bias. LIRE and RIRE show similar bias levels. The comparison of RMSE is presented in Figure 3c. The Cascaded-ANFIS shows the lowest RMSE (0.2100). LARE, LIRE, and RIRE tend to have similar RMSE. However, LSTM has the second-lowest RMSE having 0.2239.

NSE and KGE performances are presented in Figures 3b and 3, respectively. The overall results appear similar in both parameters. The Cascaded-ANFIS algorithm shows the best results having 0.8351 and 0.8776, respectively.

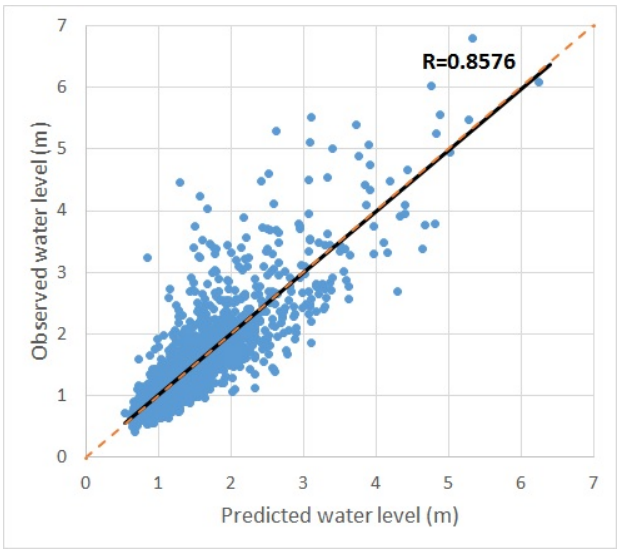
Therefore, it is safe to note that the Cascaded-ANFIS model outperforms other algorithms used in this study,

4. Conclusion

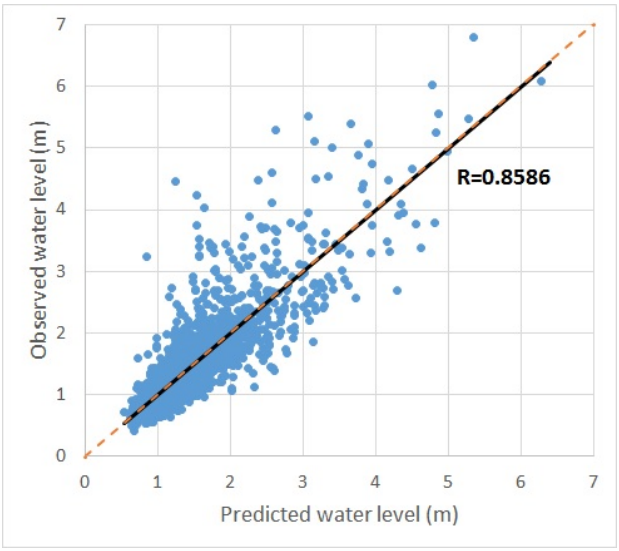
This study was conducted to design and implement an accurate, reliable rainfall-runoff prediction model. The experiment was conducted based on the Kelani River in Sri Lanka. The input data of the system is the rainfall measurements of six rain gauges in the Kelani River basin, namely, Kitulgala, Holombuwa, Deraniyagala, Hanwella, Colombo, and Rathmalana, and the output is the water level at Nagalingam Street. Due to the uncertainty factor of the rainfall, the Cascaded-ANFIS algorithm is used to develop the model because the base of the Cascaded-ANFIS is fuzzy logic. The proposed model is evaluated with five state-of-the-art algorithms: LSTM, GRU, Linear regression, Lasso regression, and ridge regression.



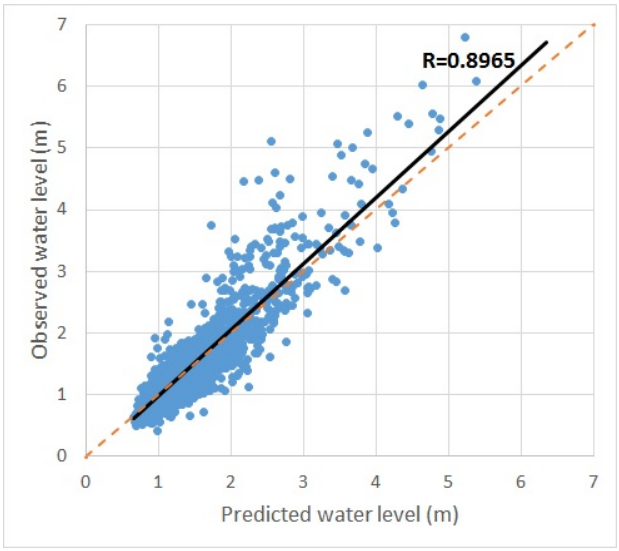
(a) Linear Regression



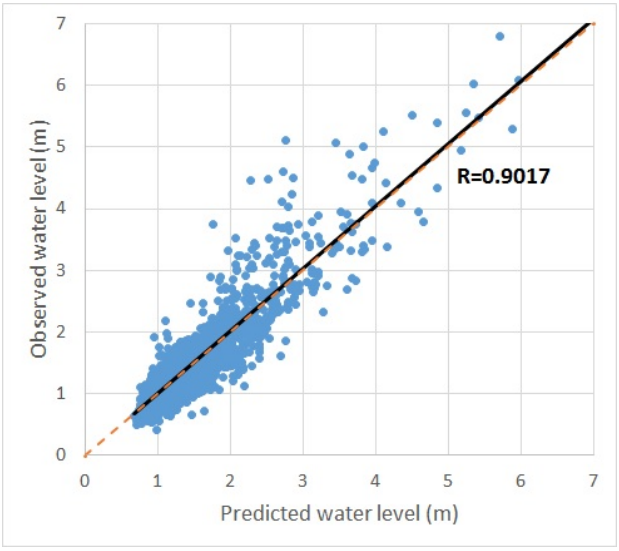
(b) Lasso Regression



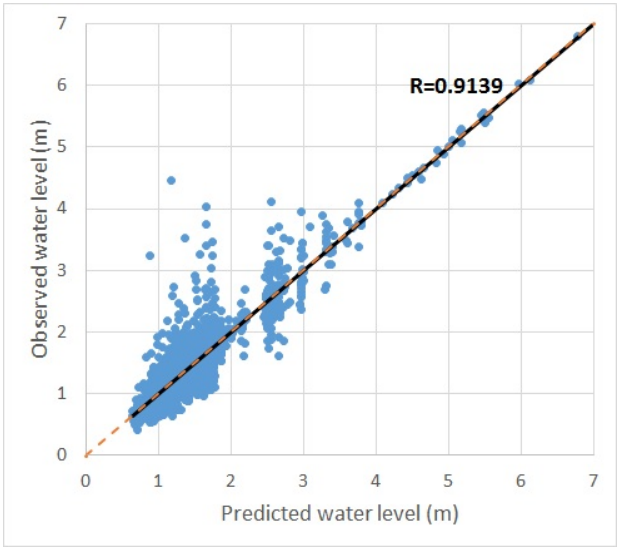
(c) Ridge Regression



(d) GRU

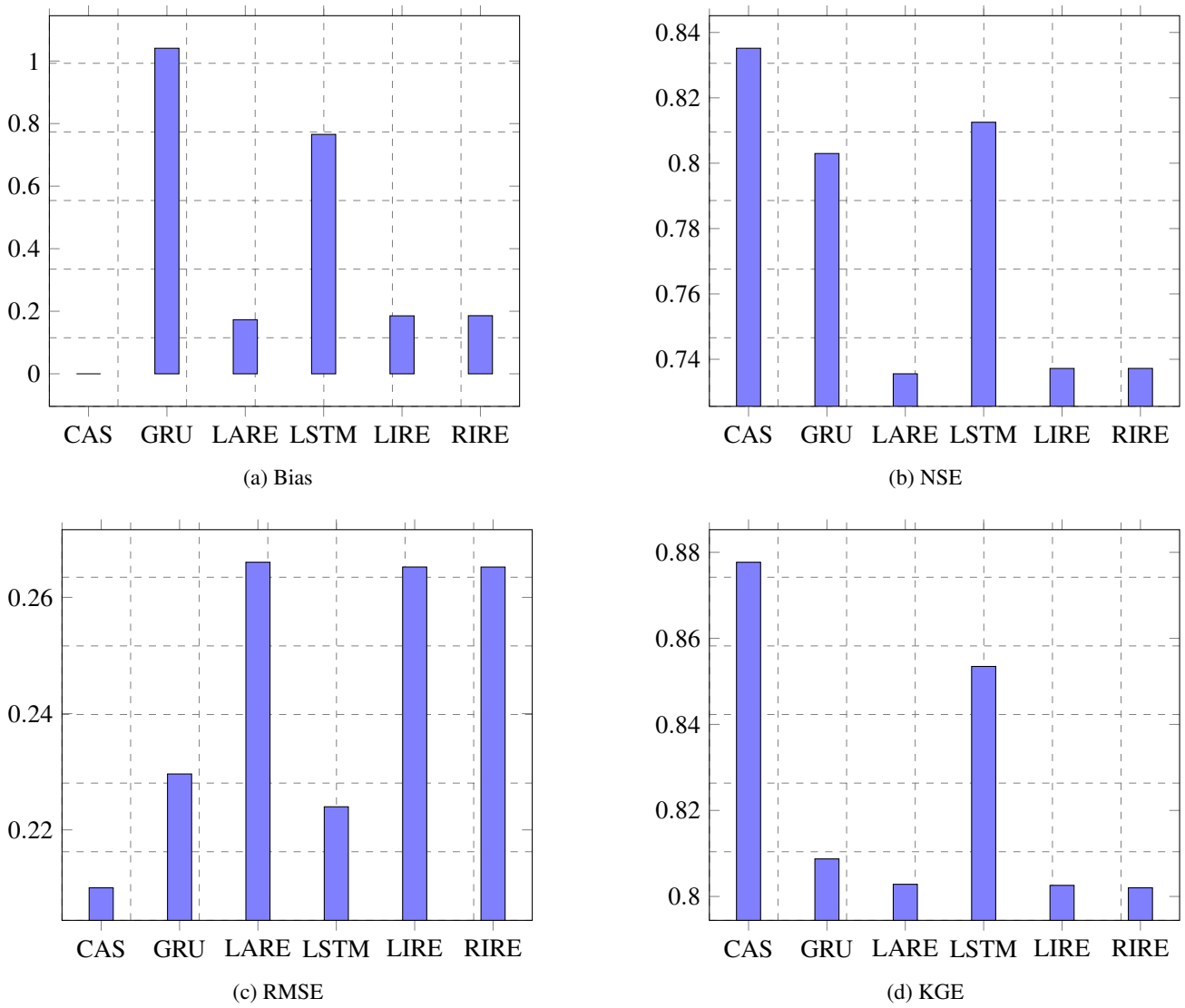


(e) LSTM



(f) Cascaded-ANFIS

Fig. 2. : Coefficient of Correlation (R)



(a) Bias

(b) NSE

(c) RMSE

(d) KGE

Fig. 3. : Evaluation parameters results.

The evaluation of the model was conducted using several parameters such as *R*, bias, RMSE, NSE, and KGE. The results prove that the Cascaded-ANFIS algorithm outperforms every aspect of the other algorithms. The Cascaded-ANFIS model achieved 0.9139, 1.15×10^{-10}, 0.2100, 0.8351, and 0.8776 respectively as the evaluation parameters. Due to the increase of natural disasters by rainfall in Sri Lanka, a design and development of a model using the Cascaded-ANFIS algorithm for streamflow forecasting can be stated as a future objective.

References:

[1] J. S. Alagha, M. A. M. Said, and Y. Mogheir, "Artificial intelligence based modelling of hydrological processes", In 4th International Engineering Conference–Towards Engineering of 21st Century, 2012.

[2] A. Ameen, "Sri Lanka floods: Nearly 500,000 displaced as death toll rises", BBC, May 2017.

[3] J. G. Arnold, R. Srinivasan, R. S. Muttiah, and J. R. Williams, "Large area hydrologic modeling and assessment part I: model development 1", JAWRA Journal of the American Water Resources Association, 34(1):73–89, 1998.

[4] M. De Silva, U. Ratnyake, S. Mahanama, S. Herath, and S. Weerakoon, "Flood inundation mapping along the lower reach of Kelani river basin under the impact of climatic change", 2012.

[5] W. Dilini, D. Attygalle, L. L. Hansen, and K. Nandalal, "Ensemble forecast for monthly reservoir inflow; a dynamic neural network approach", In Proceedings of the 4th Annual International Conference on Operations Research and Statistics (ORS 2016), Global Science and Technology Forum, Singapore, pp. 18–19, 2016.

[6] K. Dissanayaka and R. Rajapakse, "Long-term precipitation trends and climate extremes in the Kelani River basin, Sri Lanka, and their impact on streamflow variability under climate change", Paddy and Water Environment, 17(2):281–289, 2019.

[7] A. W. FLÜGEL, "Application of models with different types of modelling methodologies for river flow forecasting", In Weather Radar Information and Distributed Hydrological Modelling: Proceedings of an International Symposium (Symposium HS03) Held During IUGG 2003, the XXIII General Assembly of the International Union of Geodesy and Geophysics: at Sapporo, Japan, from 30 June to 11 July, 2003, number 282, pp. 218. International Assn of Hydrological Sciences, 2003.

[8] M. B. Gunathilake, C. Karunanayake, A. S. Gunathilake, N. Marasingha, J. T. Samarasinghe, I. M. Bandara, and U. Rathnayake, "Hydrological models and Artificial Neural Networks (ANNs) to simulate streamflow in a tropical catchment of Sri Lanka", Applied Computational Intelligence and Soft Computing, 2021, 2021.

[9] H. V. Gupta, H. Kling, K. K. Yilmaz, and G. F. Martinez, "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling", Journal of hydrology, 377(1-2):80–91, 2009.

[10] P. Hettiarachchi, "Hydrological report on Kelani River flood in May 2016", Hydrological annual, irrigation department, 2016.

[11] P. Hunukumbura and S. Weerakoon, "Comparison of two hydrological model applications for stream flow predictions in the upper Kotmale basin", 2003.

[12] J.-S. Jang, "ANFIS: adaptive-network-based fuzzy inference system", IEEE transactions on systems, man, and cybernetics, 23(3):665–685, 1993.

[13] P. Krause, "Quantifying the impact of land use changes on the water balance of large catchments using the J2000 model", Physics and Chemistry of the Earth, Parts A/B/C, 27(9-10):663–673, 2002.

[14] X. Liang, D. P. Lettenmaier, E. F. Wood, and S. J. Burges, "A simple hydrologically based model of land surface water and energy fluxes for general circulation models", Journal of Geophysical Research: Atmospheres, 99(D7):14415–14428, 1994.

[15] J. E. Nash and J. V. Sutcliffe, "River flow forecasting through conceptual models part I—A discussion of principles", Journal of hydrology, 10(3):282–290, 1970.

[16] E. Olyaie, H. Banejad, K.-W. Chau, and A. M. Melesse, "A comparison of various artificial intelligence approaches performance for estimating suspended sediment load of river systems: a case study in United States", Environmental monitoring and assessment, 187(4):1–22, 2015.

[17] N. Rathnayake, T. L. Dang, and Y. Hoshino, "A Novel Optimization Algorithm: Cascaded Adaptive Neuro-Fuzzy Inference System", International Journal of Fuzzy Systems, 23(7):1955–1971, 2021.

[18] N. Rathnayake, U. Rathnayake, T. L. Dang, and Y. Hoshino, "A Cascaded Adaptive Network-Based Fuzzy Inference System for Hydropower Forecasting", Sensors, 22(8):2905, 2022.

[19] J. Seibert, "HBV light version 2, user's manual", Department of Earth Sciences, Uppsala University, Uppsala, 2005.