*Geographia Technica, Vol. 17, Issue, 2, pp 148 to 163*

# DAILY STREAMFLOW FORECASTING USING EXTREME LEARNING MACHINE AND OPTIMIZATION ALGORITHM. CASE STUDY: A RIVER IN VIETNAM

***Huu Duy NGUYEN*** [1]

DOI : 10.21163/GT_ 2022.172.13

**ABSTRACT:**
Accurate prediction of streamflow plays an important role in water resource management and sustainability. Recent years have seen increased interest in data-based models, compared to the more established physics-based models, due to the accuracy of their predictions. Better results mean greater support for those who are tasked with formulating strategies and writing policy around water resource management. The objective of this study is the development of a state-of-the-art streamflow prediction method based on extreme learning machine (ELM), optimized by both hunger games search (HGS) and social spider optimization (SSO) to make accurate predictions for the Tra Khuc River in Vietnam. Rainfall and flow from 2000 to 2020 at Son Giang station on the Tra Khuc River were used to build the streamflow prediction model. The statistical indices root-mean-square error, mean absolute error, and the coefficient of determination ($R^2$) were applied to assess the predictive ability of the proposed models. The results showed that both optimization algorithms successfully improved the ELM model to predict the streamflow for one day and six days ahead by using data from one day and three days before the day in question. Of the proposed models, the ELM-SSO model scored highest, with $R^2$=0.891 for the one-day-ahead prediction and $R^2$=0.701 for six days ahead. Second was ELM-HGS ($R^2$=0.889 and $R^2$=0.699 for one day and six days respectively), and third was ELM ($R^2$=0.883, $R^2$=0.696). The results demonstrate ELM to be a robust data-driven method for simulating time series regimes that is appropriate for various hydrological applications. The models proposed in this study can be generalized to predict streamflow in rivers around the world.

***Key-words:*** *ELM-SSO, ELM-HGS, Streamflow, Machine learning, Tra Khuc river.*

## 1. INTRODUCTION

Streamflow prediction plays an important role in water resource management. It is required in the optimization of water resource distribution, water quality assessment, and agriculture and industrial development (Adnan et al., 2022; Lin et al., 2021; Parisouj et al., 2020). The streamflow process is very complicated because it is influenced by multiple parameters, such as precipitation, temperature, evaporation, and land use. It is also characterized by a nonlinear relationship between flow rates and characteristics of the watershed. Therefore, accurate prediction of streamflow is difficult (Ahmed et al., 2021; Parisouj et al., 2020).

The literature broadly consists of two sets of streamflow prediction models: physics-based and data-based. Physics-based models are developed only using real-life streamflow data (Khosravi et al., 2021; Rahimzad et al., 2021). Although this method has been proven effective in predicting the streamflow of rivers around the world, the development of such models is very complicated and time-consuming. In addition, physical-based models require detailed data like topography, precipitation, and land use/land cover to calibrate model parameters, and these models can also be negatively affected when watershed data do not respond well to water balance constraints (Khosravi et al., 2021). The uncertainty of precipitation and hydrology data also greatly influences streamflow prediction,

[1] *Faculty of Geography, University of Science, Vietnam National University, Hanoi, Vietnam, nguyenhuuduy@hus.edu.vn*

and physics-based models suffer in data-limited regions (Krzysztofowicz, 2002). Such models may be replaced by more robust automated techniques.

In recent years, data-driven models have received increasing attention from the worldwide scientific community. They enjoy low input-data requirements and are easy to use. There are two main groups: statistical models and machine learning models. Statistical models include logistic regression (Long et al., 2022), fuzzy logic (Kothari and Gharde, 2015), autoregressive integrated moving average model (Ab Razak et al., 2018; Singh et al., 2020), and autoregressive model (Özgür, 2005; Terzi and Ergin, 2014). These models assume a linear relationship between input and output data, so they cannot explain the nonlinear relationship of hydrological processes. This limits performance.

Machine learning has proven more effective than statistical models in solving the problem of nonlinearity when predicting streamflow. Such models include support vector machine (Guo et al., 2011; Kisi and Cimen, 2011), random forest (Wang et al., 2021a), extreme learning machine (ELM; Yaseen et al., 2016; Yaseen et al., 2019), artificial neural network (Ali and Shahbaz, 2020; Ghimire et al., 2021), and long short-term memory network (LSTM; Ghimire et al., 2021; Hunt et al., 2022). However, traditional models use samples discretely, including input and output data, so the performance of these models often deteriorates if the input data contains some degree of error (Rahimzad et al., 2021; Tikhamarine et al., 2020). Additionally, several studies have pointed out that data-driven models have limited effectiveness in streamflow prediction because they require time-series information and produce temporal dependence on the data (Adnan et al., 2021a; Ahmed et al., 2021). To overcome these shortcomings, several studies have demonstrated how prediction accuracy can be improved by developing hybrid models, eliminating the weak points of individual models (Bui et al., 2020; Nguyen, 2022a; Pham et al., 2020).

Adnan et al. (2022) integrated the adaptive neuro-fuzzy inference system (ANFIS) model with five optimization algorithms, namely gradient-based optimization (GBO), particle swarm optimization (PSO), genetic algorithm (GA), differential evolution, ant colony optimization, and gray wolf optimization (GWO) to predict streamflow in Pakistan. ANFIS-GBO was shown to perform better than the other models in predicting streamflow. Adnan et al. (2021b) combined the ELM model with PSO, GWO, and gravitational search algorithm (GSA) to predict the streamflow in the Mangla watershed of northern Pakistan. The ELM-PSOGWO model was superior to the other models (ELM, ELM-PSO, ELM-GWO, and ELM-PSOGSA). Kilinc and Haznedar (2022) used LSTM network and GA to predict the streamflow in the Euphrates River in western Asia. The LSTM-GA model outperformed the LSTM model. Al-Juboori (2021) combined K-nearest neighbor (KNN) with random tree (RT) to predict the monthly streamflow in three rivers in Iraq. The results saw the hybrid KNN-RT model outperform the individual KNN and RT models.

These models can be divided into three main approaches: ensemble framework, evolutionary algorithm, and swarm-based algorithm. Of the three approaches, swarm-based algorithms are the most popular and have been shown to be effective in predicting the streamflow in previous studies (Nguyen et al., 2021). These approaches can solve global optimization problems through the process of exploration and exploitation. However, according to the no-free-lunch theory, there are no methods that can solve all problems in all regions, due to differences in climatic, hydrological, and environmental conditions and in human activities (Bui et al., 2020). Because of this, the selection process is always subject to significant bias. Moreover, the overfitting problem looms large when using machine learning (although models can perform well in the training process because they learn the targets based on the samples in the past to predict streamflow in the future, in many cases, if the data is limited, these models may not perform well in the validation process (Mosavi et al., 2018).

The objective of this study is the development of state-of-the-art models based on the ELM, hunger games search (HGS), and social spider optimization (SSO) algorithms to predict streamflow in the Tra Khuc River. This study is different from previous studies because it is the first time the ELM model has been combined with HGS and SSO to predict streamflow. In recent years, streamflow in the study area has been strongly influenced both by climate change and human activity, particularly in the dry season.

## 2. STUDY AREA AND DATA USE

The Tra Khuc River is over 160 km long and is located in the Central region of Vietnam (**Fig. 1**). Its source is in the east of the Truong Son Mountain range and it flows into the sea at Quang Ngai. The basin has an area of about 3703 $km^2$. The basin's complex topography generally lowers from west to east. There are four main terrain types: plateau, high mountains, plains, and sandy coast. The plateau has an elevation of between 1100 and 1300 m and accounts for about 5% of the basin area. The high mountains have an average elevation of 500 to 700 m and account for about 70% of the study area. The plain runs from north to south, close to the sea; it has an altitude of 20–10 m and covers about 20% of the study area. The sandy coastal area consists of sand dunes distributed in a narrow strip, running along the coast, with an average width of about 2m.

The Tra Khuc River basin is located in a tropical monsoon region and sees an average annual rainfall of about 2960 mm. The climate of the study area is characterized by two main seasons: a rainy season from September to January – which accounts for 70–75% of total annual rainfall – and a dry season from February to August, where severe drought is not uncommon.

The hydrological system in the Tra Khuc River is characterized by short rivers and steep slopes. The river has nine main branches: Daclang, Nuoc Lac, Dacseco, Tam Dinh, Xa Dieu, Tam Rao, Song Giang, Song Phuoc, and tributary number 9. Their combined length is 195 km. The average flow rate of the Tra Khuc River (calculated over several years) is 176 m3/s. The flood season figure is 13 L/s/$km^2$. The uneven distribution of flows between the rainy and dry seasons is considered to be one of local authorities' greatest challenges. During each rainy season, the basin floods an average of between five and seven times, while in the dry season, the river flow is depleted enough to cause severe drought, with significant effects on economic development.

Water resource management is considered one of the major challenges of local government in the study area. In recent years, reservoirs have been built upstream of the Tra Khuc river, which cause negative effects on the water resource downstream: the water level of the river downstream tends to become drier and drier in dry season and increase rapidly in the rainy season. Therefore, streamflow prediction is an important task to build appropriate strategies for water resource management and agriculture development.

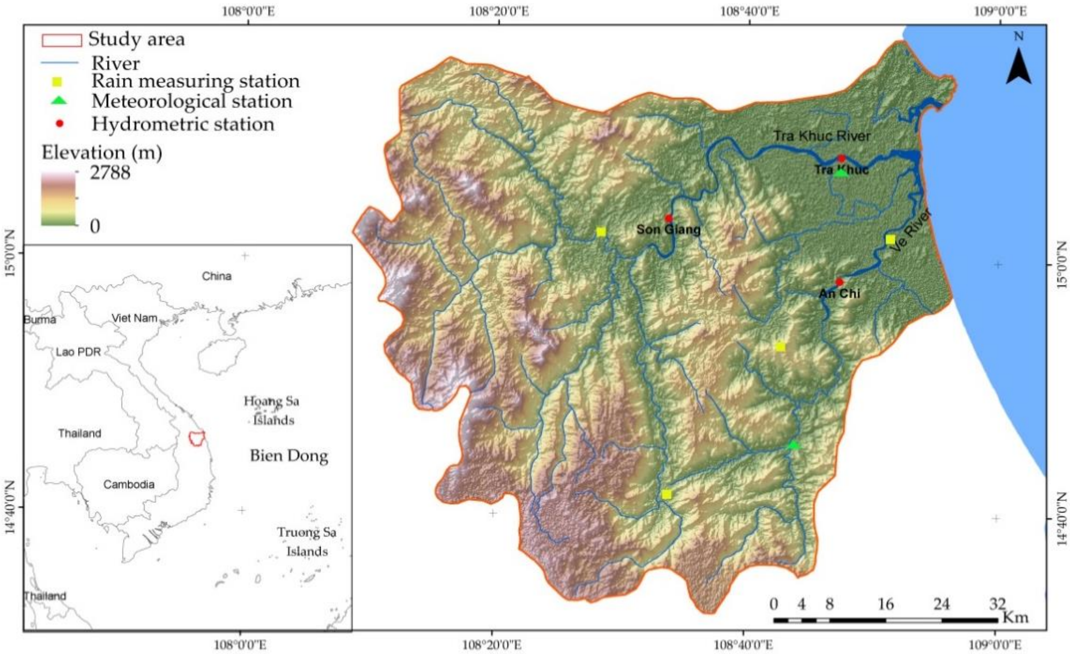


**Fig. 1**. The location of Tra Khuc River in Vietnam.

In this study, daily rainfall and streamflow data at Son Giang station on the Tra Khuc River from the period 2000–2020 were used to construct streamflow predictions for one day and six days ahead. These data were divided into two parts: 90% were used to calibrate the model parameters, while the remaining 10% were used to evaluate model performance (**Fig.2**).

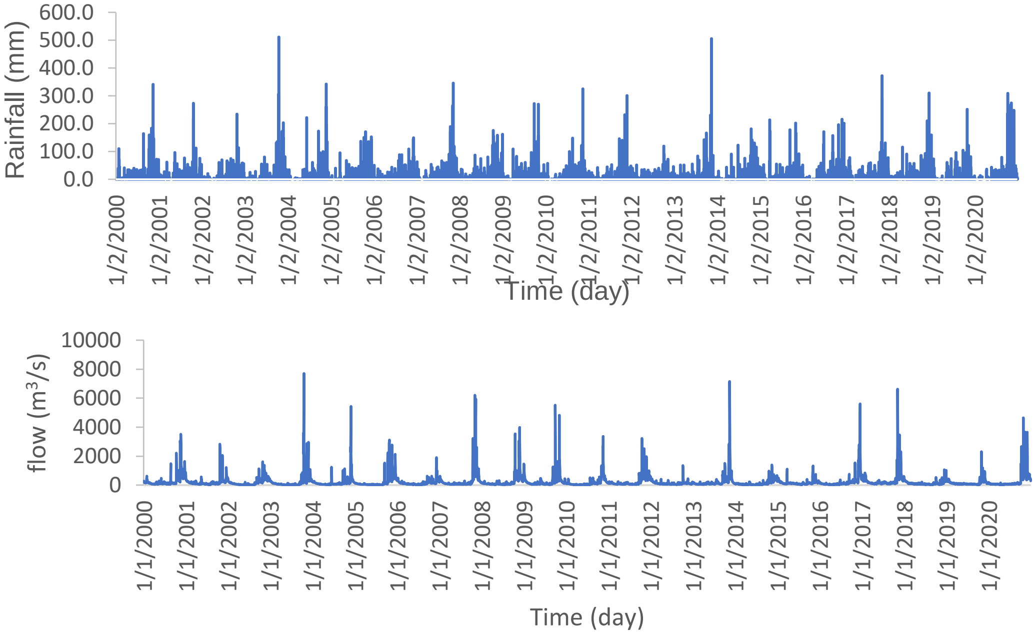


**Fig. 2.** Rainfall (top) and flow (bottom) at Son Giang, 2000–2020.

To make the model input data consistent and increase convergence in the model training process, precipitation and flow were normalized before being input into the model. There are several normalization techniques, such as nominal, ordinal, ratio, and min-max normalization. In this study, min-max normalization was used to prepare the data. In this technique, the original values of the precipitation and flow data were kept, but the input database was normalized using the ranges of similar measurements. The min-max was calculated by the following equation:

$$D_{normalised} = \frac{(D_i - D\text{min})}{(\text{Dmax} - D\text{min})}$$

## 3. METHODS

This study can be divided into four main steps: i) data collection, ii) model building, (iii) model validation, and iv) streamflow prediction (**Fig. 3**).

i) Data collection

Precipitation and flow data at Son Giang station were collected to build the streamflow prediction models. Precipitation and flow data from the past one and three days were used to predict the streamflow for one day and six days ahead. Although multiple hydro-meteorological factors affect streamflow – including temperature, evaporation, and change in land cover – we considered only precipitation, as the development of an accurate prediction model using limited information is very useful, especially in regions where data is not available, and because the study region is located in a mountainous region with low average temperature, so temperature and evaporation may not play as important a role in streamflow prediction as they usually do.

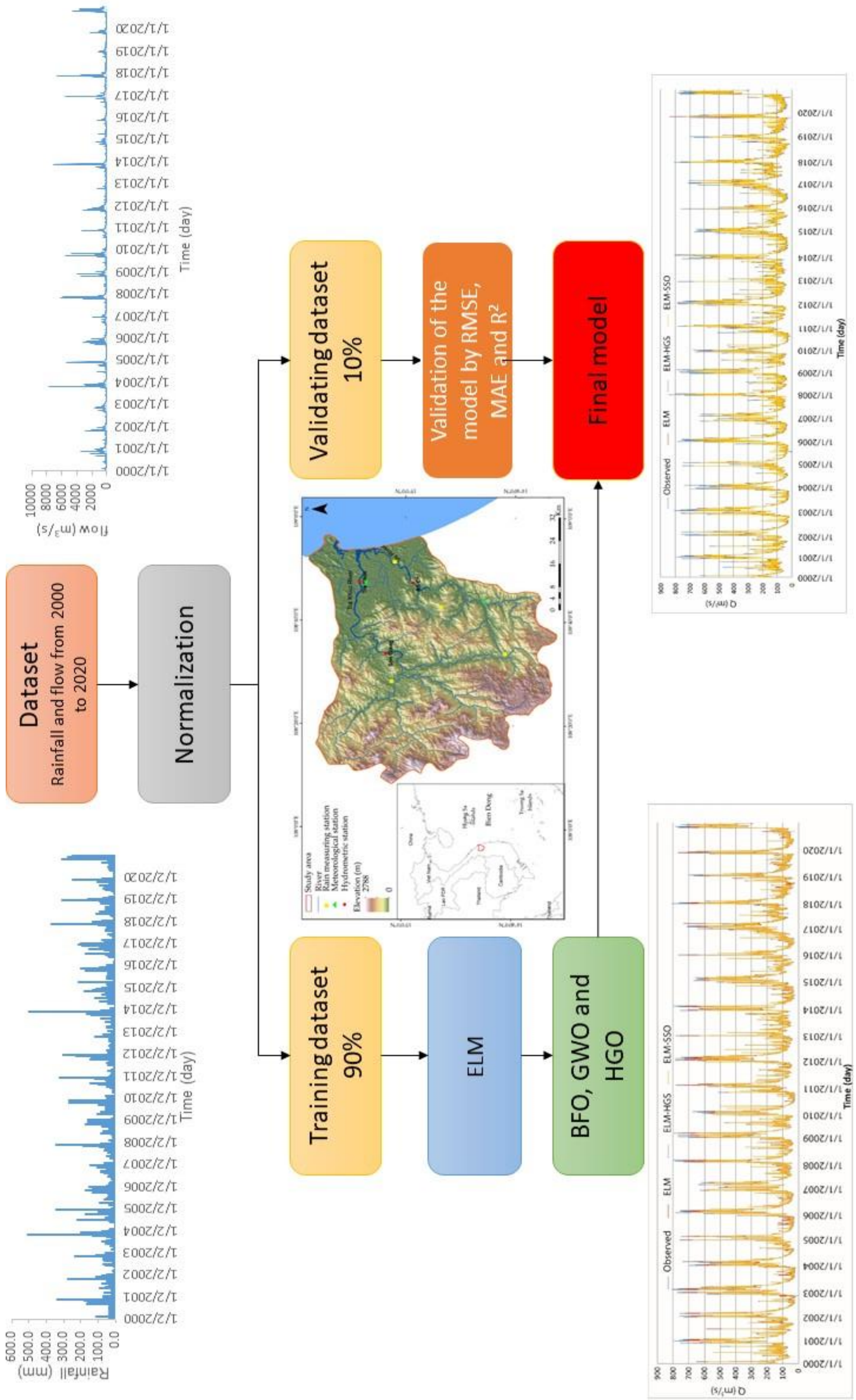


**Fig. 3**. Flowchart of the models ELM, ELM-HGS, and ELM-SSO.

The model used in this study is based on neural networks, which have the advantage of describing non-linear relationships between the input and output variables of a system. Initially, an analysis of the data obtained made it possible to highlight the information necessary to define the input variables of the network. Then, we identified the weights or parameters assigned to network connections. For this, the data was divided into two parts. As there are no universal guidelines for splitting data, we tried several different rates. The best results were obtained by using 90% of the data to build the models and 10% of the data to validate.

Precipitation and flow data were not normally distributed; therefore, before use, these data were normalized to improve the predictive ability of the models.

ii) Model construction

HGS and SSO were used to improve the prediction ability of the ELM model. The model construction process was divided into two main steps: the first was the initialization of the ELM model parameters using the trial-and-error method. The second was the determination of the optimal parameters of the hybrid models.

ELM is a feed-forward neural network for predicting the streamflow with a single layer of hidden nodes, where the weights connecting the inputs to the hidden nodes are randomly distributed and never updated. The weights between the hidden nodes and the outputs are learned in a single step, which is essentially the same as learning a linear model. In this study, the initialization of the neural network parameters included 256 weight values. 256 represents – for example – the position of the spider in the web for SSO or the number of animals in the hungry state for HGS.

iii) Model validation

The statistical indices RMSE, MAE, and $R^2$ were used to validate the streamflow model. These indices have been shown to be effective in previous studies.

iv) Use model

Predict streamflow at Son Giang station on the Tra Khuc River.

## 3.1. ELM

ELM is a feed-forward neural network, which means that the data only traverses the series of layers in one direction (Ding et al., 2014). The structure of ELM includes three layers of neurons: the input layer, the hidden layer, and the output layer (Wang et al., 2021b). The input layer gets the information from the input data, while the output layer gets a linear one without any transformation function. The hidden layer plays a role of linking the input layer and the output layer. Extreme learning machines use the concept of random projection and early perceptron patterns to do specific types of problem solving (Huang et al., 2006). In the ELM model, rather than adjusting all the weights of a neural network to emulate a function, the neural network is made up of a large number of neurons in the inner layer. Input weights are randomly initialized once and stay with that value. The adaptation, which is also done in one go, therefore only concerns the weights of the output layer (Qu et al., 2021; Wang et al., 2021b).

The output function of the ELM model is computed by the following equation:

$$Y = \sum_{i=1}^{m} B_i \; f\left(w_i x_j + b_i\right), j \in [1, n]$$

where $m$ is the number of hidden nodes, $B$ is the vector of output weights, $x$ is the output vector of the hidden layer, and f is the activation function.

Compared with conventional models, the hidden node parameters in ELM are not only independent from the training data, but also from each other. Additionally, ELM can generate node parameters before considering formation

### 3.2. HGS

HGS is a swarm-based optimizer algorithm, first introduced by Yang et al. (2021). This algorithm is inspired by the behavior of animals in a state of starvation. In order to find food and improve their survival, animals tend to cooperate with each other. Stronger animals have a greater heart capacity to obtain food than weak animals (Yang et al., 2021).

In nature, animal behavior is influenced by many different factors, a primary one being hunger. When the food source is limited, it leads to competition between animals – a "hunger game." The HGS algorithm is divided into two stages: the first stage simulates the process of cooperation between animals to find a food source; the second step describes the animals' activities in a state of starvation (Abu Shanab et al., 2021; Yang et al., 2021). The HGS algorithm has proven effective in the technical assessment and analysis of natural hazards (Nguyen, 2022b).

### 3.3. SSO

SSO is a swarm-based optimization algorithm first developed in 2015 (Cuevas et al., 2013). It is inspired by the foraging behavior of spiders (Bui et al., 2020). In nature, spiders often live in groups and cooperate with each other in search of food. The spider web is considered as an n-dimensional search space where each node is a solution to the optimization problem. Spiders are agents that search by moving through the area (web) from node to node in search of the optimal solution (Humaidi et al., 2021). Spiders are divided into male and female groups. Each spider receives a weight value that corresponds to the solution it represents. The movement of each spider generates vibrations that propagate through the search space. Other spiders receive vibrations and rely on capacitive dynamics to determine the size of the spider and how far away it is. The spider's current position is affected by the current positions of all other spiders in the colony and their previous positions (Ochoa et al., 2017). The SSO algorithm solves the optimization problem by performing the following steps (Klein et al., 2015; Luque-Chang et al., 2018):

i) The optimization process starts by collecting information from random locations on the spider web.
ii) Spiders are divided into two groups (60-90% of spiders in the colony are female; the rest are male).
iii) The weight of each spider is determined by the objective function.
iv) Identify the best spider in the herd, the best female, and the spider in the center position.
v) The position of each spider is continuously updated after each loop.
vi) Male and female spiders within the mating radius will mate to produce new spiders.
vii) New spiders will replace weak spiders if they have better weight.

### 3.4. Performance assessment

In this study, to assess the accuracy of machine learning technique in the training and validation process, the statistical indices RMSE, $R^2$, and MAE were used. These indices have been shown to be effective in previous studies (Parisouj et al., 2020; Rasouli et al., 2012).

RMSE and MAE are used to measure the errors between observation values and prediction values (Chicco et al., 2021). They are calculated by the following equations:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_{predicted} - Y_{observed})^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_{predicted} - Y_{observed}|$$

where $n$ is the number of samples, $Y_{predicted}$ is the prediction value at sample $i$, and $Y_{observed}$ is the observation value at sample $i$.

$R^2$ is a statistical indicator that measures the strength and weakness of the relationship between the observation and prediction streamflow value (Meshram et al., 2022; Pirzado et al., 2021).

The value of $R^2$ varies from 0 to 1. The closer $R^2$ is to 1, the more accurate the streamflow prediction model is.

## 4. RESULTS

### 4.1. Evaluation of the number of previous days

**Table 1** presents the accuracy of the proposed models in predicting the streamflow for one day and six days ahead. In general, when data from three days ago was used, the performance of the models increased. In the case of the one-day-ahead prediction, for the ELM model, the value of RMSE decreased from 52.707 to 51.373, MAE decreased from 25.261 to 24.995, and $R^2$ increased from 0.883 to 0.891. For the ELM-HGS model, the value of RMSE decreased from 51.165 to 50.301. MAE value decreased from 24.721 to 23.206. $R^2$ value increased from 0.889 to 0.896. For the ELM-SSO model, the value of RMSE decreased from 50.74 to 50.233. MAE value decreased from 24.667 to 23.199. While the $R^2$ value increased from 0.891 to 0.901.

In the case of six days ahead, for the ELM model, the RMSE value decreased from 84.763 to 83.815. MAE value decreased from 41.141 to 40.742. While the $R^2$ value increased from 0.696 to 0.699. For the ELM-HGS model, the RMSE value decreased from 84.337 to 39.333. MAE value decreased from 40.274 to 39.333. $R^2$ value increased from 0.699 to 0.705. For the ELM-SSO model, the value of RMSE decreased from 84.17 to 83.289. MAE value decreased from 40.161 to 39.245. While the $R^2$ value increased from 0.701 to 0.707.

In general, the ELM-SSO model performed better than the other models in predicting one-six days ahead using the one and three previous days, followed by ELM-HGS, ELM, respectively. The results also showed that two optimization algorithms were successfully improved to predict the streamflow in the Tra Khuc river.

**Table 1.**

**Performance of the models for one-six days ahead using one and three previous days.**

| One previous day | | | | | | |
|---|---|---|---|---|---|---|
| | For one day ahead | | | For six days ahead | | |
| | RMSE | MAE | $R^2$ | RMSE | MAE | $R^2$ |
| ELM | 52.707 | 25.261 | 0.883 | 84.763 | 41.141 | 0.696 |
| ELM-HGS | 51.165 | 24.721 | 0.889 | 84.337 | 40.274 | 0.699 |
| ELM-SSO | 50.74 | 24.667 | 0.891 | 84.17 | 40.161 | 0.701 |
| Three previous days | | | | | | |
| ELM | 51.373 | 24.995 | 0.891 | 83.815 | 40.742 | 0.699 |
| ELM-HGS | 50.301 | 23.206 | 0.896 | 83.302 | 39.333 | 0.705 |
| ELM-SSO | 50.233 | 23.199 | 0.901 | 83.289 | 39.245 | 0.707 |

### 4.2. Evaluation of the one-day and seven-day ahead

To assess the prediction capacity of the proposed models, this study used two prediction scenarios (for one day ahead and six days ahead). In general, as the number of prediction days increases, the accuracy of the models is decreased in both cases of using one previous day and three previous days. **Fig. 4** and **5** showed the value of $R^2$ for the proposed models for one and six days ahead using one and three previous days.

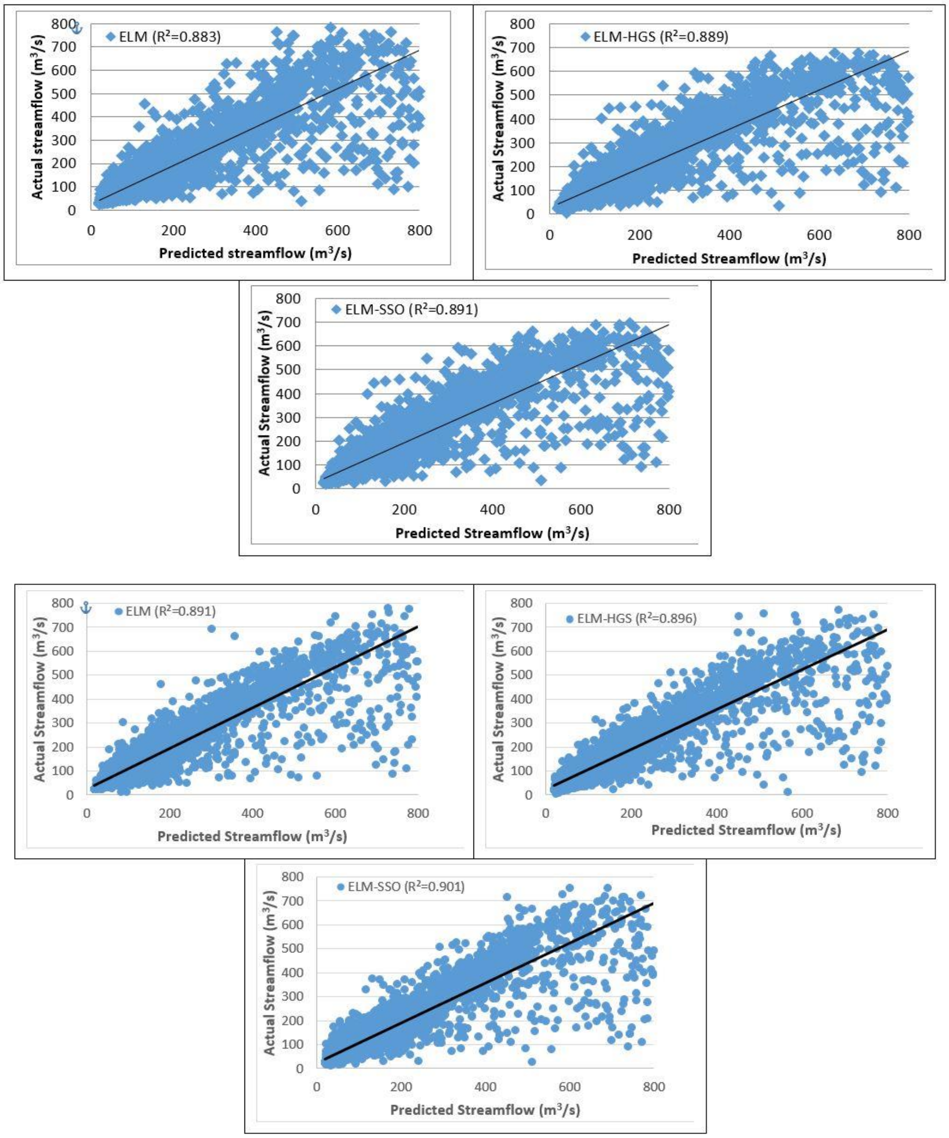


**Fig. 4.** $R^2$ value for one day ahead using one (top) and three (down) previous days.

For one previous day, for the ELM model, the value of RMSE increased from 52.707 to 84.763 when the day ahead increased from one to six days. The same, the value of MAE increased from 25.261 to 41.141. While the $R^2$ value decreased from 0.883 to 0.696. For the ELM-HGS model, the value of RMSE increased from 51.165 to 84.337 and from 24.721 to 40.274 for MAE. The $R^2$ value decreased from 0.889 to 0.699. For the ELM-SSO model, the RMSE value increased from 50.74 to 84.17 and from 24.667 for MAE. While the $R^2$ value decreased from 0.891 to 0.701.

For three previous days, the value of RMSE increased from 51.373 to 83.815 and from 24.995 to 40.742 for the value of MAE. While the value of $R^2$ decreased from 0.891 to 0.699 for the ELM model. For the ELM-HGS model, the value of RMSE increased from 50.301 to 83.302 and from 23.206 to 39.333 for the value of MAE. While the $R^2$ value decreased from 0.896 to 0.705. For the ELM-SSO model, the value of RMSE increased from 50.233 to 83.289 and from 23.199 to 39.245 for the value of MAE. The $R^2$ value decreased from 0.901 to 0.707.

It can be seen on the above figures that by increasing the number of previous days to forecast the flow, the observed and predicted values tend to concentrate along the regression line, and to approach each other. Therefore, the accuracy of the model increases.

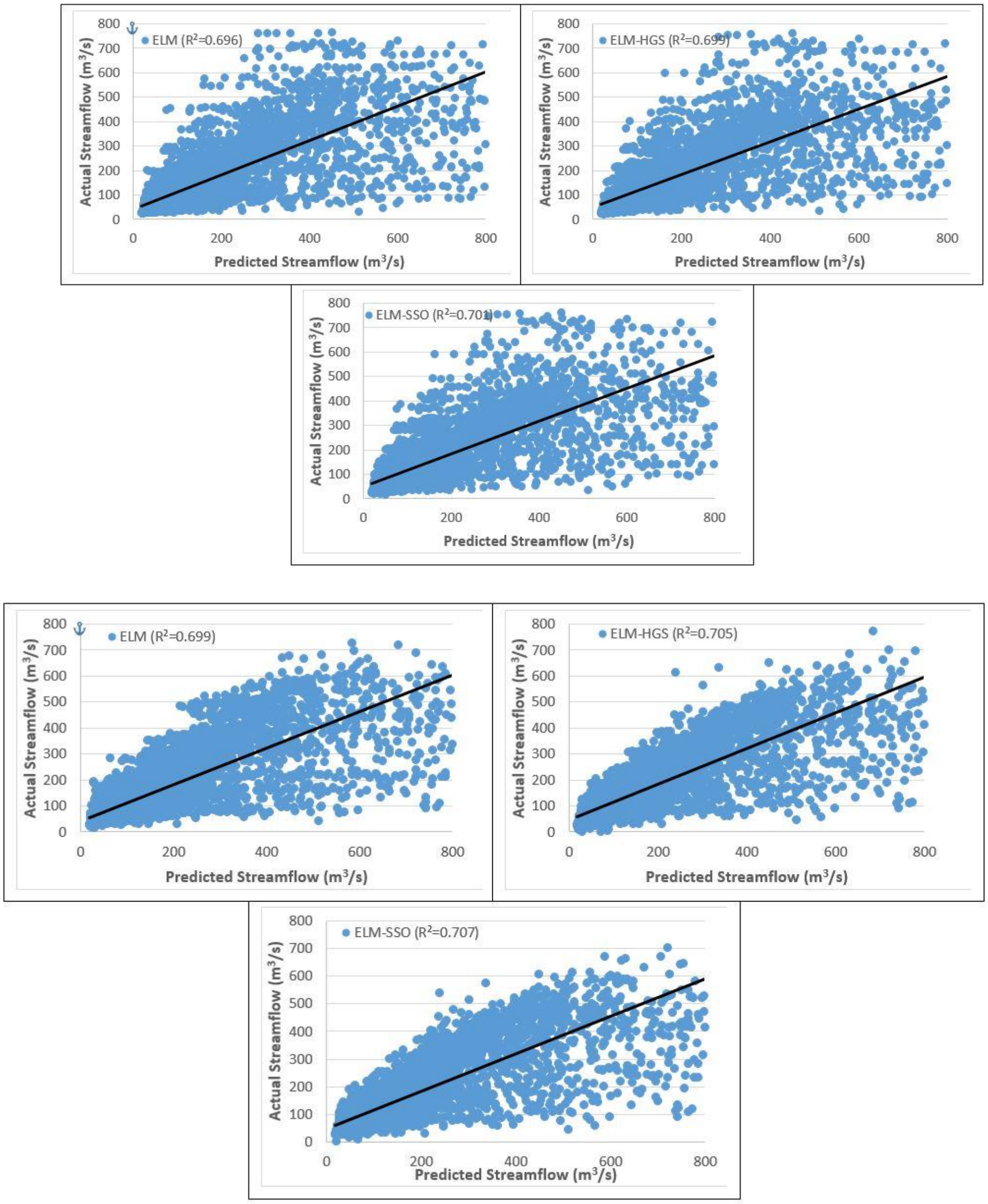


**Fig. 5.** $R^2$ value for six days ahead using one and three previous days.

On the contrary, by increasing the number of forecast days from 1 to 6 days, the values of observations and forecasts tend to be further from the regression line and further out. As a result, the accuracy of the model decreases.

**Fig. 6** and **7** show the throughput value for one and six days ahead using one and three previous days. For one day ahead using on and three previous days, in general, the predicted streamflow value follows the observed streamflow value. However, the predicted streamflow value during major flooding tends to be lower than the observed streamflow value. Meanwhile, for six days ahead using one and three previous days, not only the streamflow value at large floods is lower than the observed value, but also at small and medium floods.

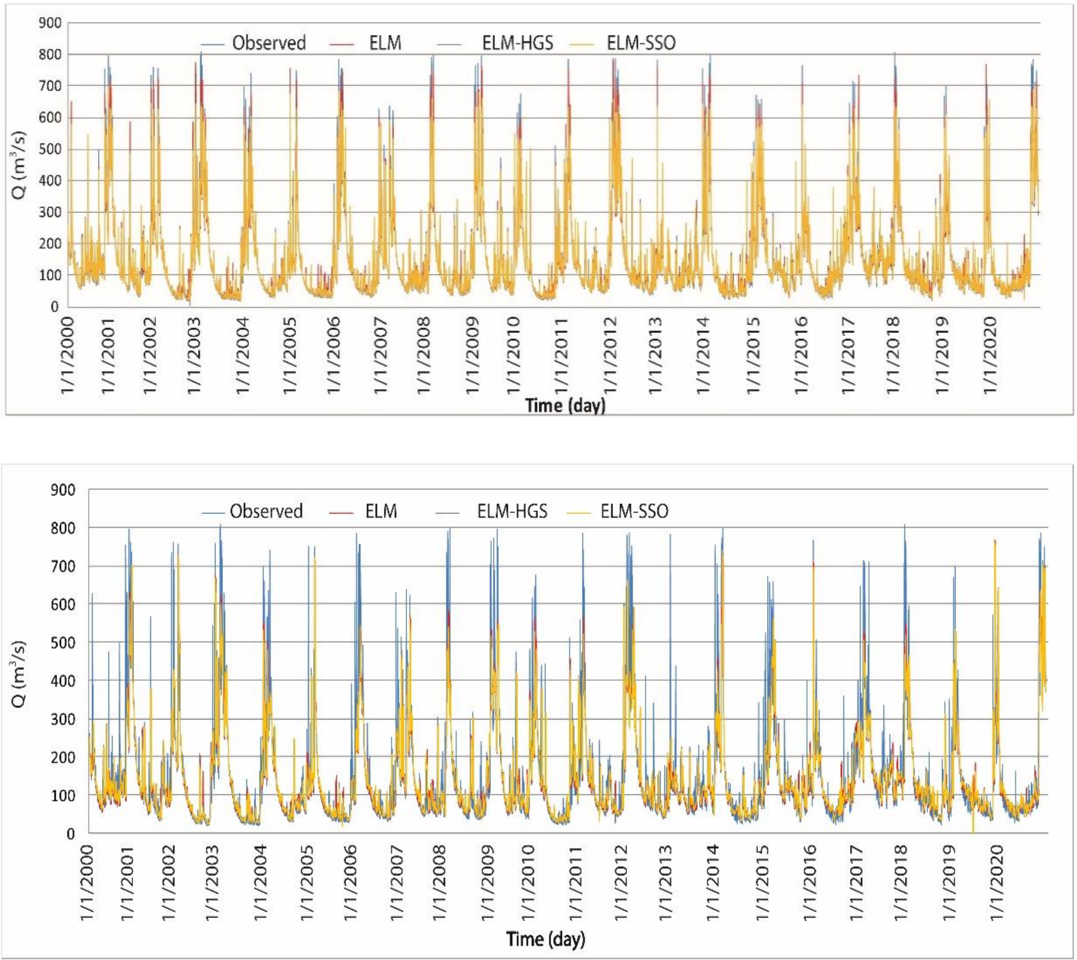


**Fig. 6.** The streamflow for one (top) and six (down) days ahead using one previous day.

**Fig. 6** and **7** show the throughput value for one and six days ahead using one and three previous days. For one day ahead using on and three previous days, in general, the predicted streamflow value follows the observed streamflow value. However, the predicted streamflow value during major flooding tends to be lower than the observed streamflow value. Meanwhile, for six days ahead using one and three previous days, not only the streamflow value at large floods is lower than the observed value, but also at small and medium floods.

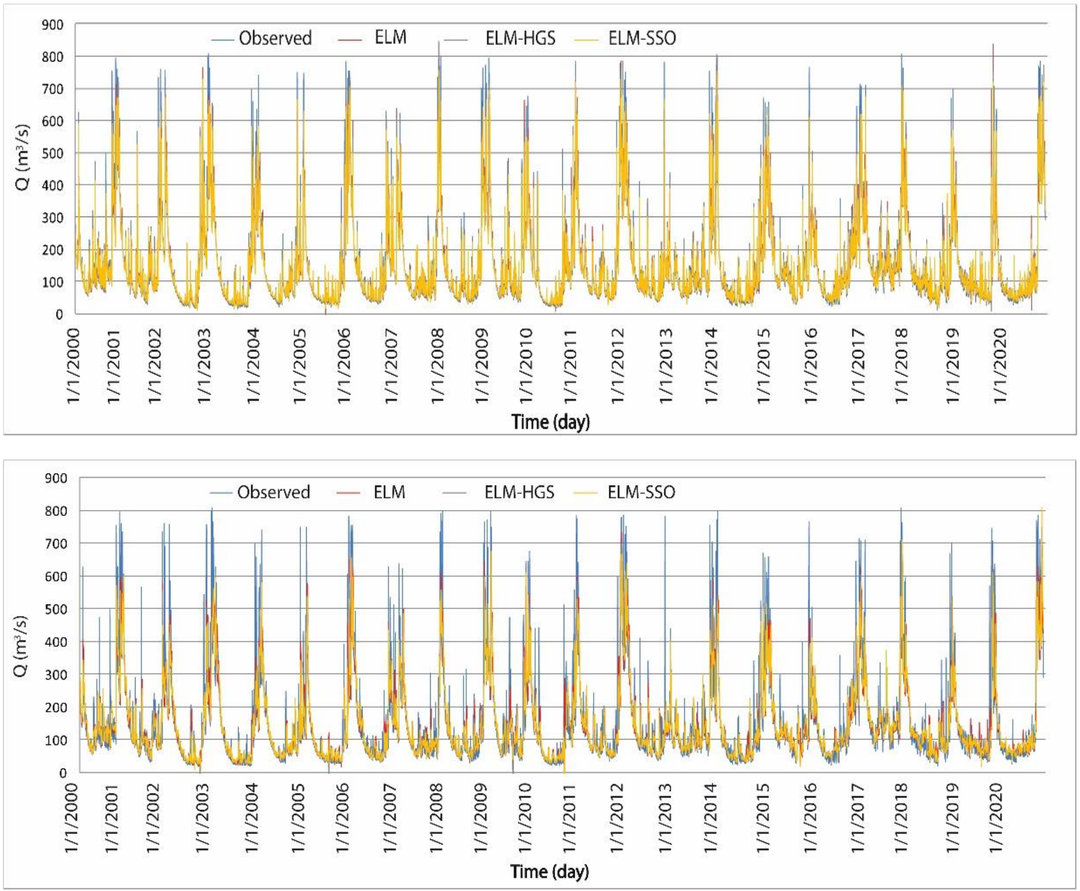


**Fig. 7.** The streamflow for one and six days ahead using three previous days.

## 5. DISCUSSION

The accurate prediction of short- and long-term streamflow plays an important role in water-resource management, operation, and planning. While short-term prediction (on hourly, multi-hour, or daily timescales) is used to provide information on flood forecasting, long-term prediction (over weeks, months, seasons, or years) provides important information for planning reservoir operations and for water resource management in general (Liu et al., 2022; Mohammadi, 2021; Rahimzad et al., 2021). However, both are difficult, due to the stochastic and nonlinear characteristics of the flow process at long time scales (Mendoza et al., 2017). Therefore, in recent years, great effort has been made in the development of automated control and monitoring methods. Using artificial intelligence techniques as a basic tool for decision support can generate more detailed answers.

ELM stands out from other tools due to its capacity and its speed of learning. In addition, it has the advantage of being able to intrinsically describe non-linear relationships between the input variables and the output variables of a system (Adnan et al., 2021a; Anmala and Turuganti, 2021; Gholizadeh et al., 2022).

The results confirmed that two optimization algorithms (HGS and SSO) successfully improved the prediction ability of the ELM model. Among the proposed models, ELM-SSO performed best, with $R^2$ = 0.891 (one day ahead) and 0.701 (six days ahead), followed by ELM-HGS with $R^2$ = 0.889 (one day ahead) and 0.699 (six days ahead), and then ELM with $R^2$ = 0.883 (one day ahead) and 0.696 (six days ahead).

SSO has advantages regarding the balance between the processes of exploration and exploitation. Compared with traditional methods, the learning and convergence speed of SSO is faster (Luque-Chang et al., 2018; Mirjalili et al., 2015). HGS is based on a population with a factor that transforms randomly, so it can enrich the ability of exploration and exploitation in the process of foraging. Moreover, HGS has advantages on the adaptive and time-varying mechanism, so this algorithm can solve the local optimization problem (Yang et al., 2021).

In recent years, streamflow has been strongly influenced by human activities such as dam construction and land-cover change. This causes difficulties in streamflow prediction, especially in extreme conditions such as flooding. Although several previous studies have shown the data-based approach to be better than other approaches (Tran et al., 2021; Tran and Kim, 2022), there is still debate around the effectiveness of machine learning and deep learning in streamflow prediction under the aforementioned conditions. This may be resolved if the amount of data to train the data is sufficient.

This study encountered general limitations. Several studies have pointed out that the water level at the station is considered an important factor in predicting streamflow; however, in the study area, the water level is very difficult to measure. So, in future research, we will try to collect more data to improve the predictive ability of the proposed models. Additionally, as mentioned above, streamflow has been influenced by human and climatic activities. This topic needs to be explored in greater depth.

## 6. CONCLUSIONS

Streamflow prediction with high precision can play a crucial role in optimizing the distribution of water resources, the development of agriculture, and in industry more generally. Therefore, the objective of this study is the development of state-of-the-art method based on ELM and two optimization algorithms, namely HGS and SSO to predict the daily streamflow in the Tra Khuc river of Vietnam.

In this study, two optimization algorithms were successfully proven to improve the prediction ability of the ELM model to predict the streamflow in the Tra Khuc River. The complete proposed models can be generalized to predict the streamflow in the other river in Vietnam, especially in data-limited regions. The use of machine learning can support decision makers in building appropriate strategies and policies for water resource management.

The hydrological regime of the Tra Khuc River was recorded during the process of ELM, ELM-HGS, ELM-SSO model formation, with an accuracy of +0.8. Of the proposed models, the ELM-SSO model performed best. The results highlighted that the prediction of the extreme discharge values of the proposed models is still limited and the accuracy of prediction results decrease when increasing the number of prediction days. The results of this study can be an effective tool to analyze and develop water resource management strategies in Vietnam in particular and in the whole world in general. The methodology used in this study can also be developed to predict the natural hazard such as salinity prediction.

## REFERENCES

Ab Razak, N., Aris, A., Ramli, M., Looi, L. & Juahir, H. (2018) Temporal flood incidence forecasting for segamat river (malaysia) using autoregressive integrated moving average modelling. *Journal of Flood Risk Management*, 11, S794-S804.

Abushanab, W.S., Abd Elaziz, M., Ghandourah, E.I., Moustafa, E.B. & Elsheikh, A.H. (2021) A new fine-tuned random vector functional link model using hunger games search optimizer for modeling friction stir welding process of polymeric materials. *Journal of materials research and technology*, 14, 1482-1493.

Adnan, R.M., Mostafa, R.R., Elbeltagi, A., Yaseen, Z.M., Shahid, S. & Kisi, O. (2022) Development of new machine learning model for streamflow prediction: Case studies in pakistan. *Stochastic Environmental Research and Risk Assessment*, 36 (4), 999-1033.

Adnan, R.M., Mostafa, R.R., Kisi, O., Yaseen, Z.M., Shahid, S. & Zounemat-Kermani, M. (2021a) Improving streamflow prediction using a new hybrid elm model combined with hybrid particle swarm optimization and grey wolf optimization. *Knowledge-Based Systems*, 230, 107379.

Adnan, R.M., R. Mostafa, R., Kisi, O., Yaseen, Z.M., Shahid, S. & Zounemat-Kermani, M. (2021b) Improving streamflow prediction using a new hybrid elm model combined with hybrid particle swarm optimization and grey wolf optimization. *Knowledge-Based Systems*, 230, 107379.

Ahmed, A.M., Deo, R.C., Feng, Q., Ghahramani, A., Raj, N., Yin, Z. & Yang, L. (2021) Deep learning hybrid model with boruta-random forest optimiser algorithm for streamflow forecasting with climate mode indices, rainfall, and periodicity. *Journal of Hydrology*, 599, 126350.

Al-Juboori, A.M. (2021) A hybrid model to predict monthly streamflow using neighboring rivers annual flows. *Water Resources Management*, 35 (2), 729-743.

Ali, S. & Shahbaz, M., (2020) Streamflow forecasting by modeling the rainfall–streamflow relationship using artificial neural networks. *Modeling Earth Systems and Environment*, 6 (3), 1645-1656.

Anmala, J. & Turuganti, V. (2021) Comparison of the performance of decision tree (dt) algorithms and extreme learning machine (elm) model in the prediction of water quality of the upper green river watershed. *Water Environment Research*, 93 (11), 2360-2373.

Bui, Q.-T., Nguyen, Q.-H., Nguyen, X.L., Pham, V.D., Nguyen, H.D. & Pham, V.-M. (2020) Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping. *Journal of Hydrology*, 581, 124379.

Chicco, D., Warrens, M.J. & Jurman, G. (2021) The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, 7, e623.

Cuevas, E., Cienfuegos, M., Zaldívar, D. & Pérez-Cisneros, M. (2013) A swarm optimization algorithm inspired in the behavior of the social-spider. *Expert Systems with Applications*, 40 (16), 6374-6384.

Ghimire, S., Yaseen, Z.M., Farooque, A.A., Deo, R.C., Zhang, J. & Tao, X. (2021) Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. *Scientific Reports*, 11 (1), 1-26.

Gholizadeh, R., Yılmaz, H. & Danandeh Mehr, A. (2022) Multitemporal meteorological drought forecasting using bat-elm. *Acta Geophysica*, 70 (2), 917-927.

Guo, J., Zhou, J., Qin, H., Zou, Q. & Li, Q. (2011) Monthly streamflow forecasting based on improved support vector machine model. *Expert Systems with Applications*, 38 (10), 13073-13081.

Humaidi, A.J., Najem, H.T., Al-Dujaili, A.Q., Pereira, D.A., Ibraheem, I.K. & Azar, A.T. (2021) Social spider optimization algorithm for tuning parameters in pd-like interval type-2 fuzzy logic controller applied to a parallel robot. *Measurement and Control*, 54 (3-4), 303-323.

Hunt, K.M., Matthews, G.R., Pappenberger, F. & Prudhomme, C. (2022) Using a long short-term memory (lstm) neural network to boost river streamflow forecasts over the western united states. *Hydrology and Earth System Sciences Discussions*, 1-30.

Khosravi, K., Golkarian, A., Booij, M.J., Barzegar, R., Sun, W., Yaseen, Z.M. & Mosavi, A. (2021) Improving daily stochastic streamflow prediction: Comparison of novel hybrid data-mining algorithms. *Hydrological sciences journal*, 66 (9), 1457-1474.

Kilinc, H.C. & Haznedar, B. (2022) A hybrid model for streamflow forecasting in the basin of euphrates. *Water*, 14 (1), 80.

Kisi, O. & Cimen, M. (2011) A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *Journal of Hydrology*, 399 (1-2), 132-140.

Klein, C.E., Segundo, E.H., Mariani, V.C. & Coelho, L.D.S. (2015) Modified social-spider optimization algorithm applied to electromagnetic optimization. *IEEE Transactions on Magnetics*, 52 (3), 1-4.

Kothari, M. & Gharde, K. (2015) Application of ann and fuzzy logic algorithms for streamflow modelling of savitri catchment. *Journal of Earth System Science*, 124 (5), 933-943.

Krzysztofowicz, R. (2002) Bayesian system for probabilistic river stage forecasting. *Journal of hydrology*, 268 (1-4), 16-40.

Lin, Y., Wang, D., Wang, G., Qiu, J., Long, K., Du, Y., Xie, H., Wei, Z., Shangguan, W. & Dai, Y. (2021) A hybrid deep learning algorithm and its application to streamflow prediction. *Journal of Hydrology*, 601, 126636.

Liu, J., Yuan, X., Zeng, J., Jiao, Y., Li, Y., Zhong, L. & Yao, L. (2022) Ensemble streamflow forecasting over a cascade reservoir catchment with integrated hydrometeorological modeling and machine learning. *Hydrology and Earth System Sciences*, 26 (2), 265-278.

Long, Y., Lv, Q., Wen, X. & Yan, S. (2022) Bayesian logistic regression in providing categorical streamflow forecasts using precipitation output from climate models. *Stochastic Environmental Research and Risk Assessment*, 1-12.

Luque-Chang, A., Cuevas, E., Fausto, F., Zaldívar, D. & Pérez, M. (2018) Social spider optimization algorithm: Modifications, applications, and perspectives. *Mathematical Problems in Engineering*, 2018.

Mendoza, P.A., Wood, A.W., Clark, E., Rothwell, E., Clark, M.P., Nijssen, B., Brekke, L.D. & Arnold, J.R. (2017) An intercomparison of approaches for improving operational seasonal streamflow forecasts. *Hydrology and Earth System Sciences*, 21 (7), 3915-3935.

Meshram, S.G., Meshram, C., Santos, C.a.G., Benzougagh, B. & Khedher, K.M. (2022) Streamflow prediction based on artificial intelligence techniques. I*ranian Journal of Science and Technology, Transactions of Civil Engineering*, 46 (3), 2393-2403.

Mirjalili, S.Z., Saremi, S. & Mirjalili, S.M. (2015) Designing evolutionary feedforward neural networks using social spider optimization algorithm. *Neural Computing and Applications*, 26 (8), 1919-1928.

Mohammadi, B. (2021) A review on the applications of machine learning for runoff modeling. *Sustainable Water Resources Management*, 7 (6), 1-11.

Mosavi, A., Ozturk, P. & Chau, K.-W. (2018) Flood prediction using machine learning models: Literature review. *Water*, 10 (11), 1536.

Nguyen, H.D. (2022a) Gis-based hybrid machine learning for flood susceptibility prediction in the nhat le–kien giang watershed, vietnam. *Earth Science Informatics*, 1-18.

Nguyen, H.D. (2022b) Hybrid models based on deep learning neural network and optimization algorithms for the spatial prediction of tropical forest fire susceptibility in nghe an province, vietnam. *Geocarto International*, 1-25.

Nguyen, H.D., Nguyen, Q.-H., Du, Q.V.V., Nguyen, T.H.T., Nguyen, T.G. & Bui, Q.-T. (2021) A novel combination of deep neural network and manta ray foraging optimization for flood susceptibility mapping in quang ngai province, vietnam. *Geocarto International*, 1-25.

Ochoa, A., Juárez-Casimiro, K., Olivier, T., Camarena, R. & Vázquez, I. (2017) Social spider algorithm to improve intelligent drones used in humanitarian disasters related to floods. *Nature-inspired design of hybrid intelligent systems*, 457-476.

Özgür, K. (2005) Daily River flow forecasting using artificial neural networks and auto-regressive models. *Turkish Journal of Engineering and Environmental Sciences*, 29 (1), 9-20.

Parisouj, P., Mohebzadeh, H. & Lee, T. (2020) Employing machine learning algorithms for streamflow prediction: A case study of four river basins with different climatic zones in the united states. *Water* Resources Management, 34 (13), 4113-4131.

Pham, Q.B., Afan, H.A., Mohammadi, B., Ahmed, A.N., Linh, N.T.T., Vo, N.D., Moazenzadeh, R., Yu, P.-S. & El-Shafie, A. (2020) Hybrid model to improve the river streamflow forecasting utilizing multi-layer perceptron-based intelligent water drop optimization algorithm. *Soft Computing*, 24 (23), 18039-18056.

Pirzado, A.A., Sutahar, S.K., Sutahar, V., Qurashi, N.A., Jatoi, I.K., Khaskeli, S.A., Chandio, R.A., Aarain, K. & Peerzado, M.B. (2021) Checking the significance of correlation coefficient from the regression analysis using wheat yield. *Journal of Applied Research in Plant Sciences*, 2 (2), 132-141.

Rahimzad, M., Moghaddam Nia, A., Zolfonoon, H., Soltani, J., Danandeh Mehr, A. & Kwon, H.-H. (2021) Performance comparison of a lstm-based deep learning model versus conventional machine learning algorithms for streamflow forecasting. *Water Resources Management*, 35 (12), 4167-4187.

Rasouli, K., Hsieh, W.W. & Cannon, A.J. (2012) Daily streamflow forecasting by machine learning methods with weather and climate inputs. *Journal of Hydrology*, 414, 284-293.

Singh, S., Parmar, K.S., Kumar, J. & Makkhan, S.J.S. (2020) Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties' cases of covid-19. *Chaos, Solitons & Fractals*, 135, 109866.

Terzi, Ö. & Ergin, G. (2014) Forecasting of monthly river flow with autoregressive modeling and data-driven techniques. *Neural Computing and Applications*, 25 (1), 179-188.

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.

Tran, T., Tran, V. & Kim, J. (2021) Improving the accuracy of dam inflow predictions using a long short-term memory network coupled with wavelet transform and predictor selection. *Mathematics*, 9, 551.

Tran, V. & Kim, J. (2022) Robust and efficient uncertainty quantification for extreme events that deviate significantly from the training dataset using polynomial chaos-kriging. *Journal of Hydrology*, 609, 127716.

Wang, J., Bao, W., Gao, Q., Si, W. & Sun, Y. (2021) Coupling the xinanjiang model and wavelet-based random forests method for improved daily streamflow simulation. *Journal of Hydroinformatics*, 23 (3), 589-604.

Yang, Y., Chen, H., Heidari, A.A. & Gandomi, A.H. (2021) Hunger games search: Visions, conception, implementation, deep analysis, perspectives, and towards performance shifts. *Expert Systems with Applications*, 177, 114864.

Yaseen, Z.M., Jaafar, O., Deo, R.C., Kisi, O., Adamowski, J., Quilty, J. & El-Shafie, A. (2016) Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in iraq. *Journal of Hydrology*, 542, 603-614.

Yaseen, Z.M., Sulaiman, S.O., Deo, R.C. & Chau, K.-W. (2019) An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. *Journal of Hydrology*, 569, 387-408.