Vietnam Journal of Earth Sciences, 45(1), 82-97, https://doi.org/10.15625/2615-9783/17914

Vietnam Journal of Earth Sciences
https://vjs.ac.vn/index.php/jse

Daily streamflow forecasting by machine learning in Tra Khuc river in Vietnam

Huu Duy Nguyen

Faculty of Geography, University of Science, Vietnam National University, Hanoi, Vietnam

Received 26 August 2022; Received in revised form 10 November 2022; Accepted 02 December 2022

ABSTRACT

Precise streamflow prediction is crucial in the optimization of the distribution of water resources. This study develops the machine learning models by integrating recurrent gate unit (GRU) with bacterial foraging optimization (BFO), gray wolf optimizer (GWO), and human group optimization (HGO) to forecast the streamflow in the Tra Khuc River, Vietnam. For this purpose, the time series of daily rainfall and river flow at Son Giang station from 2000 to 2020 were employed to forecast the streamflow. The statistical indices, namely the root mean square error, the mean absolute error, and the coefficient of determination (R^2), was utilized to evaluate the performance of the proposed models. The results showed that the three optimization algorithms (HGO, GWO, and BFO) effectively enhanced the performance of the GRU model.

Moreover, among the four models (GRU, GRU-HGO, GRU-GWO, and GRU-BFO), the GRU-GWO model outperformed the other models with $R^2 = 0.883$. GRU-HGO achieved $R^2 = 0.879$, and GRU-BFO achieved R^2=0.878. The results of this study showed that GRU combined with optimization algorithms is a reliable modeling approach in short-term flow forecasting.

Keywords: machine learning, streamflow, gate recurrent unit, bacterial foraging optimization, gray wolf optimizer, human group optimization.

1. Introduction

The streamflow process is exceptionally complex and essential in the hydrological cycle (Dehghani et al., 2020; Ahmed et al., 2021; Samanataray and Sahoo, 2021). It has been affected by various elements such as Precipitation, evaporation, and anthropogenic activities (Parisouj et al., 2020). Accurate streamflow forecasting plays a vital role in agricultural development, irrigation system layout, hydropower generation, flood control, and drought management (Ghimire et al., 2021; Cho and Kim, 2022). Streamflow forecasting models can be regrouped into physics-based and data-based. The first group includes such models as soil and water assessment tool (Easton et al., 2008; Bieger et al., 2014), Mike Nam (Ghosh et al., 2022; Nannawo et al., 2022), variable infiltration capacity (Tesemma et al., 2015; Wang et al., 2019), the topography-based hydrological model (Gumindoga et al., 2011; Gumindoga et al., 2014). These models simulates the physical process of runoff formation, configured with various parameters physically. They can generate predictions

*Corresponding author, Email: nguyenhuuduy@hus.edu.vn

through the evaluation and simulation of the hydrological cycle. The use of physical parameters allows us to comprehend the different hydrological processes with relatively high spatial and temporal resolution (Lane et al., 2019; Khosravi et al., 2021). Although considerable effort has been made to improve the precision of physics-based models, they have been restricted by uncertainties in datasets, parameter heterogeneity, and non-linearities in generating streamflow. Furthermore, these models require large amounts of reliable data (Adnan et al., 2021; Rahimzad et al., 2021; Hunt et al., 2022). This presents difficulties when applying the models in areas with limited data. The establishment of these models is also complicated and time-consuming. Therefore, these approaches must be replaced by data-based models, particularly in global warming.

Data-based models in streamflow prediction have been widely applied in recent years. These models can present the mathematical relationships (linear and nonlinear) between the streamflow and its explanatory factors to predict streamflow effectively (Seo et al., 2015; Samanataray and Sahoo, 2021). These models can be regrouped into two categories: statistical and machine learning. Statistical models are based on the dataset's structure, including a long-term trend, random or cyclical variation, or seasonal changes. The development of these models assumes stability in the data set. Therefore, most statistical models are limited in predicting non-linearity in hydrological time series (Adnan et al., 2021).

Machine learning models include artificial neural networks (Dolling and Varas, 2002), support vector regression (SVR) (Kisi and Cimen, 2011), random forest (Peng et al., 2020), extreme learning machine (ELM) (Adnan et al., 2019), long short-term memory (LSTM) (Ghimire et al., 2021), recurrent gate unit (GRU) (Wang et al., 2021), random subspace (Nhu et al., 2022), radial basis function classifiers (Luu et al., 2022) and multilayer perceptron (MLP) (Hosseinzadeh Talaee, 2014; Panahi et al., 2021). They can effectively simulate the nonlinear features of streamflow. This has increased machine learning applications in hydrology and water resource management. Parisouj et al., (2020) applied three machine learning algorithms, namely SVR, artificial neural network (ANN), and extreme machine learning, to predict streamflow in four rivers in the USAThe authors pointed out that SVR was better than the models used. Rahimzad et al. (2021) used four algorithms, namely linear regression (LR), multilayer perceptron (MLP), support vector machine (SVM), and LSTM, to predict daytime streamflow for the Kentucky River in eastern Kentucky, USA. The results indicated that the LSTM was better at streamflow prediction than the other models. Siddiqi et al. (2021) developed a hybrid model by combining ELM and ANN with a wavelet to predict the average monthly streamflow of the Tarbale Dam on the Indus River. The hybrid models outperformed the individual models in predicting streamflow. Meshram et al. (2022) predicted the streamflow into the Shakkar watershed in India using an adaptive neuro-fuzzy inference system (ANFIS), genetic programming (GP), and ANN. The performance of ANFIS was superior to the other models; GP came second, and ANN was last. Although machine learning models has the ability to address non-linearity and non-stationarity issues in the hydrological process (Adnan et al., 2021). However, their implementation is not consistent, and there is not yet a universal conclusion for the superior methods (Ghimire et al., 2021). Furthermore, machine learning models still have limitations in the generalization performance problem, trapping in local optimization and overfitting problems (Mosavi et al., 2018; Zhao et al., 2021). To surpass these limitations, various studies have developed hybrid models. These models have been regrouped into five approaches: ensemble framework, a hybrid evolutionary algorithm, a swarm-based algorithm, a physics-based algorithm, and an ensemble of statistics and machine learning. Adnan et al. (2021) integrated the support vector machine model (SVM) with the simulated annealing algorithm (SA)-mayfly optimization algorithm (MOA) to predict the streamflow in the Helium River Basin. Kilinc et al. (2022) combined a Gated recurrent unit (GRU) with a gray wolf algorithm (GWO) to forecast the streamflow in the Seyhan River Basin of Turkey. The advantage of hybrid models is that they can reduce the weak points of individual models (Nguyen, 2022). Tran and Kim (2022) have shown that hybrid models solve generalization performance problems.

Currently, nature-inspired meta-heuristic algorithms like particle swarm optimization (Ch et al., 2013), genetic algorithm (Nguyen et al., 2022), ant colony optimization (Adnan et al., 2022), differential evolution (Tao et al., 2017), artificial bee colony (Kisi et al., 2012), and gray wolf optimization (GWO) (Tikhamarine et al., 2020), combined with machine learning, have been successfully utilized to predict streamflow in a variety of regions around the globe. The structure of any metaheuristic algorithm consists of two main processes: exploration and exploitation. Exploration is the ability to expand the search space, while exploitation is the potential to find optimization solutions (Chakraborty and Kar, 2017; Brezonik et al., 2018). Abdel-Basset et al. (2018) pointed out that the key to a good research process is balancing the two processes. Although these methods have proven effective in predicting streamflow in multiple locations, difficulties persist, especially in river engineering applications. However, due to the complexity of streamflow data, the accurate prediction of streamflow has been considered an important problem for several decades. The literature shows that selecting appropriate models and methods in the hundreds of models is challenging and that these models depend on each region's location and characteristics. Due to the complexity of streamflow, especially in the context of global warming and urban growth, it is necessary to develop new models by combining machine learning models with optimization algorithms.

GRU is considered one of the most influential and efficient models. In addition to quickly solving nonlinear relationships, GRU has a remarkable ability to analyze relationships between input and output data. In addition, the simplicity of application of this model is one of the most appropriate characteristics for water resource managers. However, GRU has the disadvantages of slow convergence and low learning efficiency. Therefore, the integrations of the GRU model with the optimization algorithms are essential to predict the daily streamflow with high accuracy (Muhammad et al., 2019).

Furthermore, several previous studies have pointed out that there are no universal conclusions on the best models to predict streamflow with high accuracy (Parisouj et al., 2020). Therefore, developing new models is a convenient and scientific tool. New algorithms based on data mining are being developed, and they are receiving attention from some researchers in the world thanks to their ability to solve the weak points of traditional machine learning algorithms. Their performance exceeds the performance of traditional machine learning.

This study looks to develop models by integrating GRU with BFO, GWO, and HGO to forecast streamflow for one and six day ahead in the Tra Khuc River. One and six days ahead were selected as the output data of the models, which are similar to the previous studies (Alizadeh et al., 2021; Sharma et al., 2021). The comprehensive comparison between popular models can make a more challenging task in selecting algorithms to simulate and predict the natural process more accessible. In other words, this study tried to identify the best algorithms that can predict streamflow with high accuracy to provide guides for managers for water resource management actions. In recent years, the river's streamflow has been highly modified due to climate change and human activity, such as dam construction; the results in this article are essential for developing policies and strategies for water resource management. This study's results significantly affect streamflow simulation in watersheds where data quality and availability are serious challenges.

2. Data and methods

2.1. Study area and observational data

The Tra Khuc River basin is situated in the South-Central region of Vietnam (Fig. 1). The study area has an area of about 3703 km² with elevations ranging from 0 to 1442 m and an average slope of approximately 23.9%. It includes four types of terrain: coastal sandy areas, plains, high mountains, and plateaus. With a tropical monsoon climate region, the study area has two seasons: the rainy and dry seasons. The rainy season starts from September to January, and the dry season starts from February to August. The average annual rainfall in the basin is about 2,960 mm. The rainy season accounts for 70–75% of annual rainfall.

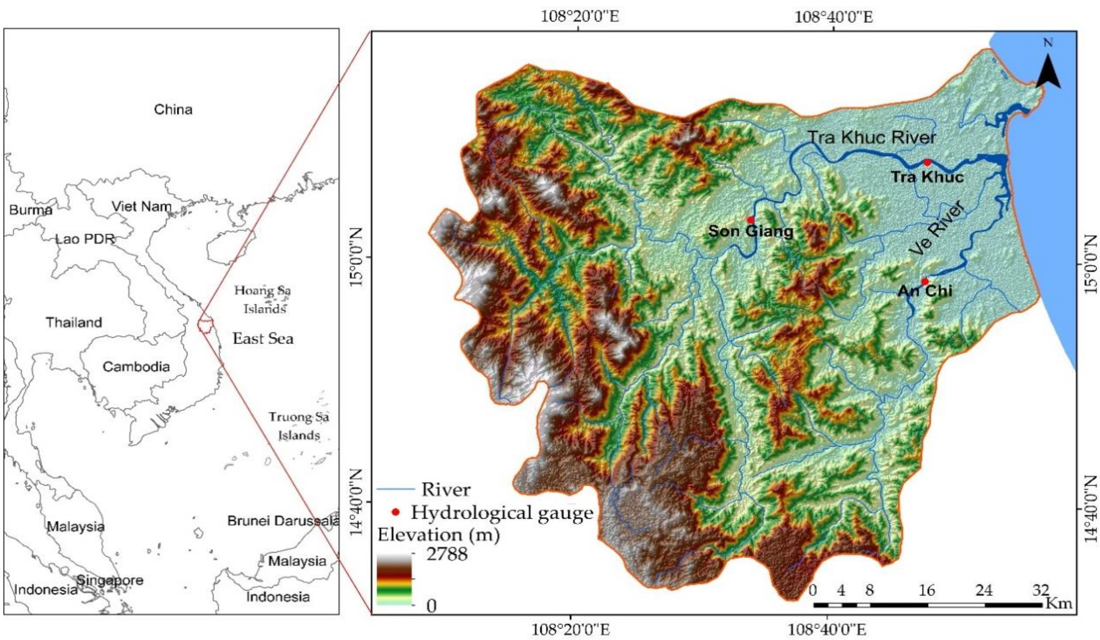

Figure 1. Location of study area

Tra Khuc River has a total length of 195 km with an average annual flow of about 176 m^3/s. The uneven distribution of water flows between the wet and dry seasons poses fundamental challenges in water resource management in this region. In the rainy season, the basin is often affected by floods, with an average of 5-7 floods per year. Meanwhile, drought and saline intrusion in the coastal area severely affect the dry season. Therefore, flow forecasting plays an essential part in managing and allocating water resources for agricultural and industrial development in the basin.

Figure 2 shows the observed rainfall and discharge at Son Giang station between 2000 and 2020, upstream of the Tra Khuc River. These data were available at the Vietnam Hydrometeorological Data and Information Center. These data were used to build the streamflow prediction model to forecast the streamflow before one day and six days (Fig. 2). Specifically, daytime precipitation data from 2000 to 2020 were used to simulate daytime river flows from 2000 to 2020 and predict the streamflow before one day and six days.

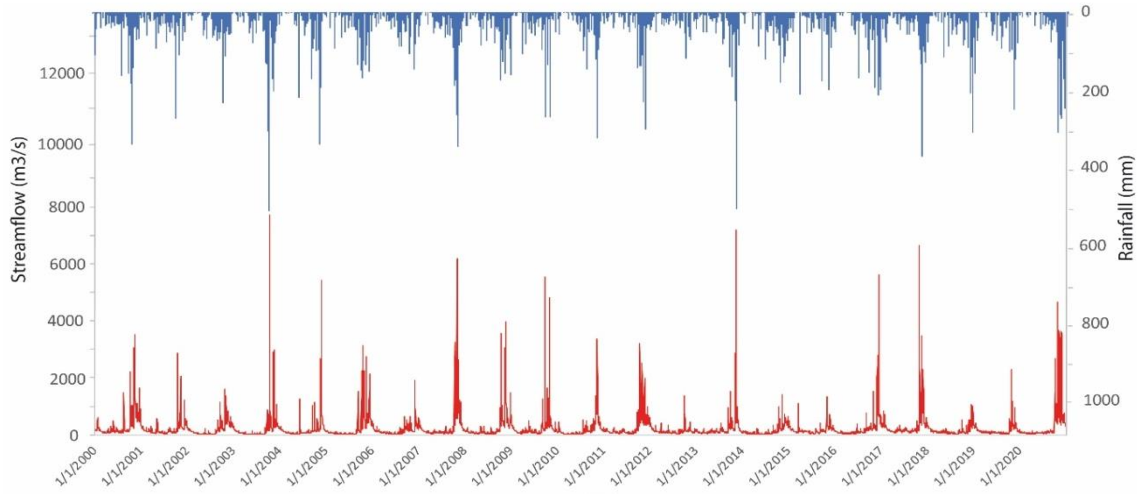

Figure 2. The rainfall and streamflow at Son Giang station on the Tra Khuc river from 2000 to 2020

In general, the model training process using the machine learning approach encounters several difficulties because the raw streamflow data has nonlinear characteristics, which strongly influence the model if we directly use these data in the model (Khosravi et al., 2021). It is, therefore, necessary to normalize these data. Streamflow prediction studies have proposed several methods, such as nominal, ordinal, ratio, min-max, and interval types. The selection of the appropriate method depends on the available dataset and the algorithm. GRU takes into account the original values of the input data, so min-max normalization was used.

2.2. Methodology

2.2.1. GRU

GRU was first developed in 2014 by Cho et al. (2014) and is another improved version of RNN model, which enhances the memory and clustering performance of ML tasks (Ha et al., 2021). This model allows altering input weights in a neural network to avoid the trailing gradient problem, which is represented in RNN frequently (Zhao et al., 2021). Similar LSTM model, GRU also uses the gating mechanism. However, it only uses two gates, including the update and reset gates (Ha et al., 2021). The update gate is used to reform the information in the preceding step, while the reset gate identifies the information that needs to remove in the antecedent state. This allows neurons to remove unnecessary information to predict the streamflow and save long-term information by reducing loads. The hidden layer candidate represents the memory being created at the current time (Ha et al., 2021). The following equation presents the structure of GRU:

$$z_t = \sigma(W_z[h_{t-1}, X_t])$$
$$r_t = \sigma(W_r[h_{t-1}, X_t])$$
$$h_t = (1 - z_t) \otimes h_{t-1} + h_{t-1} \otimes \widetilde{h_t}$$
$$\widetilde{h_t} = \tanh(W[r_t \otimes h_{t-1}, x_t])$$

Where σ is the sigmoid function as the activation function for the hidden layer, which germinates values in the range [0, 1], and the tanh function is the activation function for the output layer, which has values between the range [-1, 1]. Wz, Wr, and W are the weight matrixes.

2.2.2. BFO

BFO was first developed by Passino (2010).It is inspired by the foraging behavior of groups of bacteria, such as E. coli and M. Xanthus. The bacteria sense chemical gradients in the environment and move toward or away from specific cues. Bacteria always try to find areas of high food resources and avoid harmful substances. According to the optimization points of view, the value of places where the food is higher is the optimization value. Bacteria can be placed in predetermined locations or dispersed throughout the nutrient medium. The bacteria move to places of high nutritional value by applying random walks with a constant step size. After performing these walks, the health of the bacteria is assessed based on the nutritional value they received. Healthy bacteria receive the most nutrients; they are selected to participate in the reproduction process to take the place of their mother. During this time, low-nutrient bacteria die off (Passino, 2010).

2.2.3. GWO

GWO was presented for the first time in 2014 by Mirjalili. It mimics the predatory abilities and social hierarchy of the gray wolf. Gray wolves generally live in packs of between 5 and 12 and are classified into four main groups: alpha (α), beta (β), delta (δ), and omega (ω); (Tikhamarine et al., 2020). Alpha is considered the pack's leader and is responsible for making decisions during the hunt. Beta supports alpha to help make more accurate decisions and is a backup candidate when alpha is absent. Deltas act as scouts or hunters. Omega watches arguments within the group. The order of the dominance hierarchy descends from alpha to omega (Tikhamarine et al., 2020). GWO works by dividing the solutions to the optimization problem into four main groups: α, β, δ, and ω, of which three solutions - α, β, and δ - are considered the best. To implement this mechanism, the hierarchy is updated according to the three best solutions at each iteration. The GWO algorithm works on four main processes: finding, circling, hunting, and attacking prey (Tikhamarine et al., 2019). GWOs have the advantage of being easy to implement because they have few tuning parameters and a fast convergence speed (Hao and Sobhani, 2021; Zhang et al., 2021).

2.2.4. HGO

HGO was first developed by Carbone and Giannoccaro (2015) and is inspired by how humans solve self-interest problems in groups and seek consensus. The group is conceived as the whole of the individual who makes selections based on rational calculations and self-interested motivations (De Vincenzo et al., 2016). However, any individual choice is influenced by social relationships, which lead individuals to change their initial choices. It is human nature to seek consensus and avoid conflict with others. Therefore, the most accurate group decisions are the choices made by the many individuals interacting in the group (DiMaggio and Powell, 1983). HGO is based on the statistical process of individual choices in the continuous-time Markov process. The continuous-time Markov model is proposed to describe the time evolution of the decision-making process. Like other optimization algorithms, the model parameters are continuously adjusted throughout the optimization process to ensure convergence and overall optimization. However, in HGO, the parameters are determined using the NK model (N is the decision, and K is the interaction between them). The individual's idea transition rate is like the product of the Ising-Glauber rate to achieve consensus building and model the individual's autonomous behavior. The driving force of this system is the phase transition from low value to high value.

2.2.5. Performance assessment

The precision of the proposed models was quantified applying various statistical indices: RMSE, MAE and R^2. These indices have been extensively applied in previous studies (Siddiqi et al., 2021; Adnan et al., 2022). They were computed by the following equations:

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

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

2.2.6. Basic step of modeling by GRU-BFO, GRU-HGO, GRU-GWO

The methodology employed to forecast the streamflow before one day and six days were separated into four steps: (i) data collection and preparation, (ii) building the models, (iii) model validation (Fig. 3).

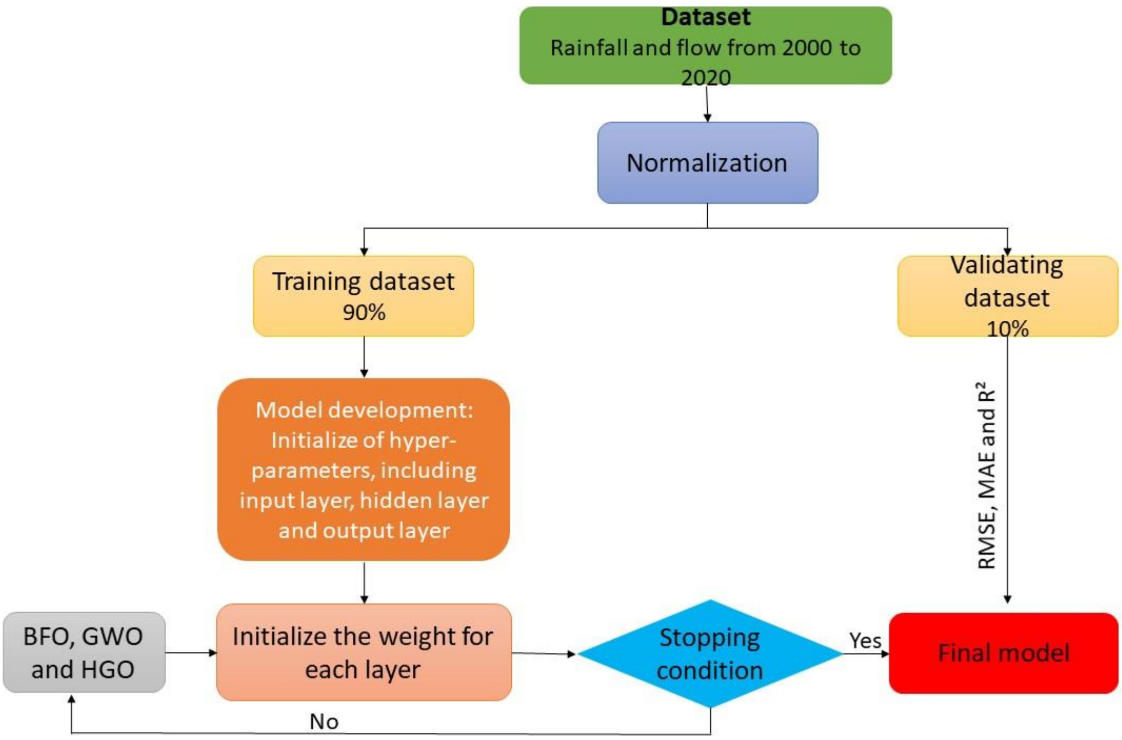

Figure 3. Flowchart of the proposed hybrid GRU models (GRU-BFO, GRU-GWO, and GRU-HGO)

(i) Data collection and preparation. The input data from 2000 to 2020 at Son Giang station were Precipitation and discharge data. In this study, 90% of the data (from 2000 to 2018) were used to train the models, and 10% (from 2019 to 2020) for validate the models. Several rates (60/40, 70/30, 80/20) were tested in this study. However, the models had more performance with a rate of 90/10. Additionally, Precipitation and discharge data have been normalized over a range of 0 to 1 to ensure data consistency and reduce the complexity of the models using the mean.

(ii) Building the models. Three optimization algorithms - BFO, HGO, and GWO - were used to optimize the hyper-parameters of the GRU model. The model hyper-parameter initialization was duplicated, including the GRU model hyper-parameter initialization and the parameters of the three optimization algorithms. The structure of the GRU model consists of an input layer, a hidden layer, and an output layer. Unlike the LSTM-based model, the architecture of the GRU-based model has only two gates, including reset and update, as mentioned in 2.2.1. The GRU-based model's performance (including accuracy and computing speed) depended on various hyper-parameters, including window size, number of neurons in the hidden state, number of hidden layers, batch size, epoch size, and initial learning rate. In which the most complex problem is to determine the number of layers. This study chose the number of hidden layers and epoch hyper-parameters to optimize the GRU-based model. In this study, we implemented one GRU layer as a hidden layer with 32 neurons. This configuration of GRU model was used in many studies to avoid the overfitting problem (Kilinc and Yurtsever, 2022; Li et al., 2020).

After determining the parameters, the proposed models were established in the Python environment with two preprocessed datasets read separately to train the model.

(iii) Model validation. RMSE, MAE, and R^2 were used to predict the performance of the proposed models.

3. Results

3.1. Comparison of models

Table 1 shows the precision of the proposed models. All proposed models performed well in predicting streamflow in the Tra Khuc River for one day and six days. For the one-day-ahead prediction, GRU-GWO was strongest, with the highest R^2 value and lowest RMSE and MAE values (R^2=0.883, RMSE=52.675, MAE=25.006). Second was GRU-HGO (R^2=0.879, RMSE=53.521, MAE=28.112), then came GRU-BFO (R^2=0.878, RMSE=53.791, MAE=28.226), and GRU was least successful (R^2=0.865, RMSE=56.488, MAE=30.088). Six days ahead, the picture was similar: GRU-GWO was more accurate than GRU-HGO, GRU-BFO and GRU with R^2=0.706, RMSE=83.486, and MAE=39.946. Second was GRU-HGO (R^2=0.703, RMSE=83.839, MAE=40.957), then GRU-BFO (R^2=0.697, RMSE=84.692, MAE=43.174), and finally GRU (R^2=0.683, RMSE=86.622, MAE=46.926) (Table 1).

Table 1. Model performance and comparison

	For one day ahead			**For six days ahead**		
	RMSE	MAE	R^2	RMSE	MAE	R^2
GRU	56.488	30.088	0.865	86.622	46.926	0.683
GRU-BFO	53.791	28.226	0.878	84.692	43.174	0.697
GRU-HGO	53.521	28.112	0.879	83.839	40.957	0.703
GRU-GWO	52.675	25.006	0.883	83.486	39.946	0.706

Note: (*) - GRU: gate recurrent unit (GRU); BFO: bacterial foraging optimization; GWO: gray wolf optimizer; HGO: human group optimization; RMSE: root-mean-square error; MAE: mean absolute error; R^2: the coefficient of determination

3.2. Assessment of the one-day, seven-day ahead

In this study, the other ahead outflows (one and six days ahead) were used to analyze the usefulness of the proposed models. In general, when the forecast moved from one day ahead to six days ahead, the performance of all the models (GRU, GRU-BFO, GRU-HGO, and GRU-GWO) decreased (Fig. 4 and Fig. 5). For the GRU model, the values of RMSE and MAE augmented from 56.488 to 86.622 and from 30.088 to 43.174, respectively. While the

R^2 value reduced from 0.865 to 0.683. For the GRU-GWO model, the trend was the same: RMSE and MAE augmented from 52.675 to 83.486 and from 25.006 to 39.946, respectively. The R^2 value reduced from 0.883 to 0.706. For the GRU-HGO model, the values of RMSE and MAE augmented sharply from 53.521 to 83.839 and from 28.112 to 40.957. The R^2 reduced from 0.879 to 0.703. For the GRU-BFO model, RMSE and MAE also augmented from 53.791 to 84.692 and from 28.226 to 43.174, and R^2 reduced from 0.878 to 0.697. In conclusion, the prediction accuracy is reduced when the prediction step increases.

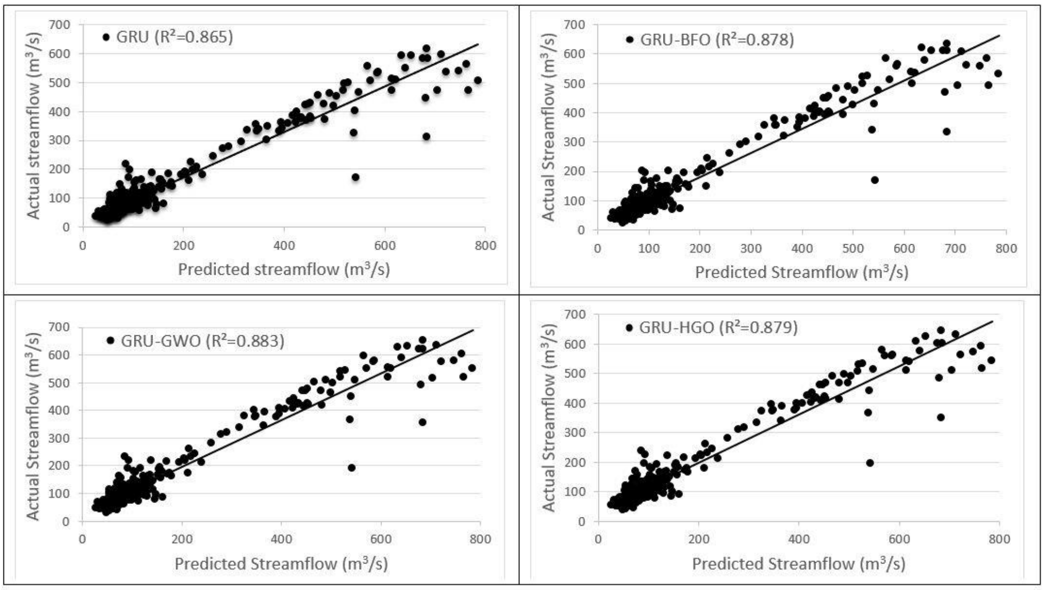

Figure 4. Scatterplots of the observed and predicted streamflow for one-day-ahead predictions

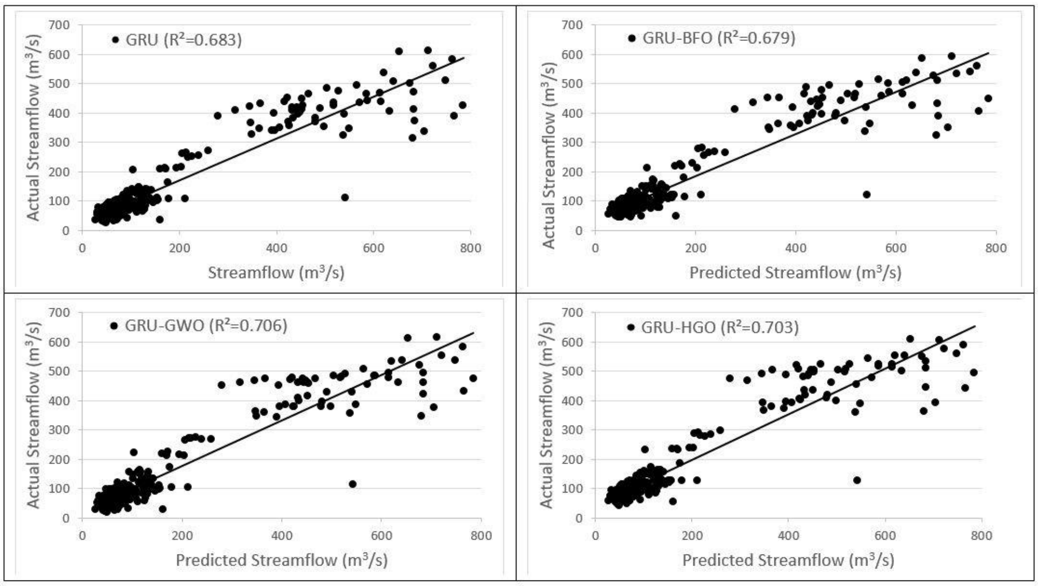

Figure 5. Scatterplots of the observed and predicted streamflow for six-day-ahead predictions

Figure 6 presents examples of the streamflow results before one day and six days of forecasting after validating the proposed models for the rainy season in 2020. The results generally show that the predicted streamflow closely follows the observed streamflow for both cases of one- and six-day forecasting. However, the streamflow during flooding predicted by the proposed models tends to be underestimated compared to the observed streamflow. Among the proposed models, the streamflow prediction results of the GRU-GWO model are closer to the observed streamflow value than the remaining models, followed by the GRU-HGO, GRU-BFO, and GRU models.

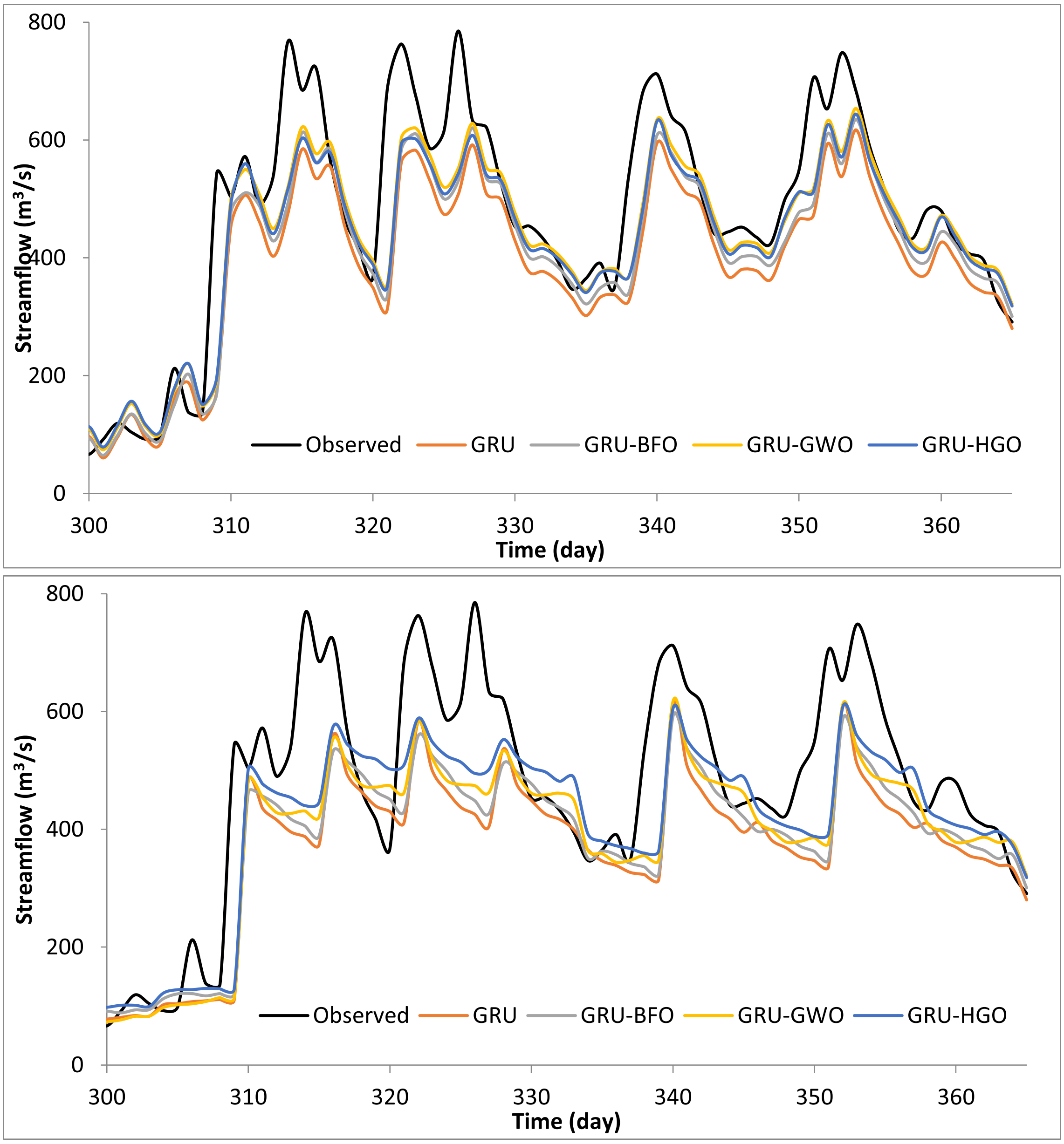

Figure 6. Streamflow prediction for one day ahead (top) and six days ahead (below)

4. Discussions

This study develops machine learning models based on GRU, BFO, GWO, and HGO, to forecast the streamflow in the Tra Khuc River, where problems in the management of water resources are regularly encountered, particularly in the context of climate change.

The results confirmed that the three optimization algorithms successfully improved the GRU model's precision. Of the proposed models, GRU-GWO was the strongest, with R^2 = 0.883 (before one day) and 0.706 (before six days), followed by GRU-HGO (R^2 = 0.879, 0.703), GRU-BFO (R^2 = 0.878, 0.697), and finally GRU (R^2 = 0.865, 0.683). Although GRU has the advantage of improving the memory capacity of a recurrent neural network which helps to facilitate training, the models and the hidden units can be used to solve vanishing gradient problems in the recurrent neural network (Wang et al., 2019). However, GRU also has convergence speeds and low learning capacity disadvantages. This is why it is necessary to use optimization algorithms. GRU-GWO was the strongest because GWO has the power exploration capability to avoid local optimization problems. In addition, GWO balances the processes of exploration and exploitation (Wang and Li, 2019), which makes it effective in solving complex problems like streamflow prediction. The GRU-HGO model was the second most powerful. De Vincenzo et al. (2016) explained how HGO could solve complex problems, and the algorithm has been effective in giving the right solutions in the case of missing data. Of the hybrid models, GRU-BFO was least effective because despite the advantages of being insensitive to initial values, fast convergence, and global optimization, the BFO algorithm also tends to get stuck in local optimization (Hernández-Ocana et al., 2013; Sahib et al., 2018), leading to reduced accuracy.

The proposed models (GRU, GRU-BFO, GRU-GWO, GRU-HGO) do not accurately predict the peak flow in the case of the six-day (long-term) forecast, but they perform well in short-term flow forecasting. This is found in previous studies: Khosravi et al. (2022) integrated the random forest model, support vector machine, multilayer perceptron, adaptive neuro-fuzzy inference system, and convolutional neural network with Bat metaheuristic algorithm to predict the daily streamflow in the Korkorsar catchment in Iran. Most proposed models had the prediction results of under-predicting or over-predicting maximum values. Adnan et al. (2020) built four models, namely optimally pruned extreme learning machine (OP-ELM), least square support vector machine (LSSVM), multivariate adaptive regression splines (MARS), and M5 model tree (M5Tree) to predict monthly stream flows in Swat River basin of Pakistan. The results showed that these models simulated the trend of hydrological processes well. However, they cannot accurately predict the maximum values. The main reasons are that the maximum value data in the training dataset does not cover all the maximum values in the validation dataset. Reis et al. (2021) pointed out that it is essential to provide the model with sufficiently long data on extreme events and the variables that influence these events. This is underlined by Cheng et al. (2020).

The global optimization problem is considered one of the crucial challenges encountered in the machine learning approach. Models perform less efficiently when the dataset is not in the range of the training data. Several authors have proposed solutions to these problems: Bui et al. (2020) demonstrated that more training data can mean greater accuracy and improved ability to solve the global optimization problem. However, collecting comprehensive and inclusive data on all events is very challenging, especially in the context of global warming. Therefore, in such a case, many authors extended the spatial prediction range by dealing with input noises or built hybrid models by integrating models like LSTM, Deep learning with optimization algorithms (Tran and Kim, 2022).

The potential of nature-inspired meta-heuristic algorithms is seen as the adaptation to improve the GRU model for the implemented application. In an actual application, the consistency of the proposed models can be studied for more of the river. Although Precipitation is the primary source of river flow in the catchment area, it is clear that using Precipitation as input to the flow prediction model is correct. However, in rivers like the Tra Khuc, the flow depends on the upstream reservoir regulation policy. Therefore, for future research, adding the flow at the reservoirs as input data to the forecasting model is necessary. In addition, future models proposed in this study should predict other processes, such as groundwater or flood volume prediction. In addition, applying the proposed models in the different rivers will be necessary to obtain more conclusive evidence. In the end, this study is limited in predicting maximum value. Therefore, it is necessary to provide models with more data on this characteristic. The exploitations of machine learning models are necessary to find the best models that can predict streamflow with high precision.

In recent years, the scientific community has tried to develop machine learning algorithms, to replace physics-based models due to their efficiency. This study provides essential references related to streamflow prediction for future studies. In the context of global warming and the increased dependence on hydroelectricity, the hydrology regime has been modified so strongly that the findings of this study can provide important information to decision-makers or water managers to build more strategies and policies clearly, for the management of water resources. Although this study applies to one river in Vietnam, this method can be generalized to apply to rivers all over the world.

5. Conclusions

Accurate streamflow prediction is crucial in optimization for agriculture development, industry, and power center operation. Therefore, the study aims to develop machine learning models, namely GRU-BFO, GRU-HGO, and GRU-GWO, to predict the streamflow before one day and six days in the Tra Khuc River in Vietnam.

All three optimization algorithms effectively improved the GRU model's performance, with $R^2 > 0.8$. The GRU-GWO model performed best, with $R^2 = 0.883$, followed by GRU-HGO (R^2=0.879) and GRU-BFO (R^2=0.878), respectively. The models proposed can be applied to the streamflow in the other river in Vietnam to support developers in building water resource policies and management.

For future research, the proposed models can be used in various applications related to hydrology, such as groundwater volume prediction, natural hazard prediction, and in several areas of water resources. In addition, the proposed models can be applied in more rivers to obtain more conclusive evidence.

The results of the current study show that data-driven approaches can be considered as an effective tool to analyze and build policies and strategies of water resource management, especially in data-limited regions.

References

Abdel-Basset M., Abdel-Fatah L., Sangaiah A.K., 2018. Metaheuristic algorithms: A comprehensive review. Computational intelligence for multimedia big data on the cloud with engineering applications, 185-231.

Adnan R.M., Liang Z., Trajkovic S., Zounemat-Kermani M., Li B., Kisi O., 2019. Daily streamflow prediction using optimally pruned extreme learning machine. Journal of Hydrology, 577, 123981.

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., 2021. 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., Kisi O., Mostafa R.R., Ahmed A.N., El-Shafie A., 2021. The potential of a novel support vector machine trained with modified mayfly optimization algorithm for streamflow prediction. Hydrological Sciences Journal, 67(2), 161-174.

Adnan R.M., Liang Z., Heddam S., Zounemat-Kermani M., Kisi O., Li B., 2020. 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, 586, 124371.

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.

Alizadeh B., Bafti A.G., Kamangir H., Zhang Y., Wright B.D., Franz K. J., 2021. A novel attention-based LSTM cell post-processor coupled with bayesian optimization for streamflow prediction. Journal of Hydrology, 601, 126526.

Bieger K., Hörmann G., Fohrer N., 2014. Simulation of streamflow and sediment with the soil and water assessment tool in a data scarce catchment in the three Gorges region, China. Journal of Environmental Quality, 43(1), 37-45.

Brezočnik L., Fister I., Podgorelec V., 2018. Swarm intelligence algorithms for feature selection: a review. Applied Sciences, 8(9), 1521.

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.

Carbone G., I. Giannoccaro, 2015. Model of human collective decision-making in complex environments. The European Physical Journal, B88(12), 1-10.

Ch S., Anand N., Panigrahi B.K., Mathur S., 2013. Streamflow forecasting by SVM with quantum behaved particle swarm optimization. Neurocomputing, 101, 18-23.

Chakraborty A., A.K. Kar, 2017. Swarm intelligence: A review of algorithms. Nature-inspired computing and optimization, 475-494.

Cheng M., Fang F., Kinouchi T., Navon I., Pain C., 2020. Long lead-time daily and monthly streamflow forecasting using machine learning methods. Journal of Hydrology, 590, 125376.

Cho K., Y. Kim, 2022. Improving streamflow prediction in the WRF-Hydro model with LSTM networks. Journal of Hydrology, 605, 127297.

Cho K., Van Merriënboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv, 1406.1078.

De Vincenzo I., Giannoccaro I., Carbone G., 2016. The Human Group Optimizer (HGO): Mimicking the collective intelligence of human groups as an optimization tool for combinatorial problems. Preprint at https://arxiv.org/abs/1608.01495.

De Vincenzo I., Giannoccaro I., Carbone G., 2016. The Human Group Optimizer (HGO): Mimicking the collective intelligence of human groups as an optimization tool for combinatorial problems arXiv preprint arXiv, 1608.01495.

Dehghani R., Torabi Poudeh H., Younesi H., Shahinejad B., 2020. Daily streamflow prediction using support vector machine-artificial flora (SVM-AF) hybrid model. Acta Geophysica, 68(6), 1763-1778.

DiMaggio P.J., W.W. Powell, 1983. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. American sociological review, 147-160.

Dolling O.R., E.A. Varas, 2002. Artificial neural networks for streamflow prediction. Journal of Hydraulic Research, 40(5), 547-554.

Easton Z.M., Fuka D.R., Walter M.T., Cowan D.M., Schneiderman E.M., Steenhuis T.S., 2008. Re-conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas. Journal of Hydrology, 348(3-4), 279-291.

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.

Ghosh A., Roy M.B., Roy P.K., 2022. Evaluating the performance of MIKE NAM model on rainfall-runoff in lower Gangetic floodplain, West Bengal, India. Modeling Earth Systems and Environment, 1-17.

Gumindoga W., Rientjes T., Haile A., Dube T., 2014. Predicting streamflow for land cover changes in the Upper Gilgel Abay River Basin, Ethiopia: A TOPMODEL based approach. Physics and Chemistry of the Earth, Parts A/B/C, 76, 3-15.

Gumindoga W., Rwasoka D., Murwira A., 2011. Simulation of streamflow using TOPMODEL in the Upper Save River catchment of Zimbabwe. Physics and Chemistry of the Earth, Parts A/B/C, 36(14-15), 806-813.

Ha S., Liu D., Mu L., 2021. Prediction of Yangtze River streamflow based on deep learning neural network with El Niño-Southern Oscillation. Scientific Reports, 11(1), 1-23.

Hao P., B. Sobhani, 2021. Application of the improved chaotic grey wolf optimization algorithm as a novel and efficient method for parameter estimation of solid oxide fuel cells model. International Journal of Hydrogen Energy, 46(73), 36454-36465.

Hernández-Ocana B., Mezura-Montes E., Pozos-Parra P., 2013. A review of the bacterial foraging algorithm in constrained numerical optimization. In Proceedings of the 2013 IEEE Congress on Evolutionary Computation, Cancun, Mexico, 20-23, 2695-2702.

Hosseinzadeh Talaee P., 2014. Multilayer perceptron with different training algorithms for streamflow forecasting. Neural Computing and Applications, 24(3), 695-703.

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.

Hussain D., Hussain T., Khan A.A., Naqvi S.A.A., Jamil A., 2020. A deep learning approach for hydrological time-series prediction: A case study of Gilgit river basin. Earth Science Informatics, 13(3), 915-927.

Hussain D., A.A. Khan, 2020. Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan. Earth Science Informatics, 13(3), 939-949.

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.

Kosravi K., Golkarian A., Tiefenbacher J.P., 2022. Using Optimized Deep Learning to Predict Daily Streamflow: A Comparison to Common Machine Learning Algorithms. Water Resour Manage, 36, 699-716.

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

Kisi O., Ozkan C., Akay B., 2012. Modeling discharge-sediment relationship using neural networks with artificial bee colony algorithm. Journal of Hydrology, 428, 94-103.

Kilinc H.C., Yurtsever A., 2019. Short-Term Streamflow Forecasting Using Hybrid Deep Learning Model Based on Grey Wolf Algorithm for Hydrological Time Series. Sustainability, 14(6), 3352.

Lane R., Coxon G., Freer J., Wagener T., Johnes P., Bloomfield J., Greene S., Macleod C., Reaney S., 2019. Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain. Hydrology and Earth System Sciences, 23, 4011-4032.

Li Y., Yang J., Wang H., Cui J., Ma Y., Huang S., 2020. Dynamic equivalent modeling for microgrid based on GRU. Energy Reports, 6, 1291-1297.

Luu C., Nguyen D.D., Amiri M., Tran V.P., Bui D.Q., Prakash I., Pham T.B., 2022. Flood susceptibility modeling using Radial Basis Function Classifier and Fisher's linear discriminant function. Science of the Earth, 44(1), 55-72.

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

Mirjalili S., Mirjalili S.M., Lewis A., 2014. Grey wolf optimizer. Advances in Engineering Software, 69, 46-61.

Muhammad A.U., Li X., Feng J., 2019. Using LSTM GRU and Hybrid Models for Streamflow Forecasting. In International Conference on Machine Learning and Intelligent Communications. Springer, Cham, 510-524.

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

Nannawo A.S., Lohani T.K., Eshete A.A., Ayana M.T., 2022. Evaluating the dynamics of hydroclimate and streamflow for data-scarce areas using MIKE11-NAM model in Bilate river basin, Ethiopia. Modeling Earth Systems and Environment, 1-16.

Nguyen D.H., Le X.H., Anh D.T., Kim S.-H., Bae D. -H., 2022. Hourly streamflow forecasting using a Bayesian additive regression tree model hybridized with a genetic algorithm. Journal of Hydrology, 606, 127445.

Nguyen H.D., 2022. 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., Dang D.K., Nguyen Q.-H., Bui Q.-T., Petrisor A.-I., 2022. Evaluating the effects of climate and land use change on the future flood susceptibility in the central region of Vietnam by integrating land change modeler, machine learning methods. Geocarto International, 1-36.

Nhu V.-H., Bui T.T., My L.N., Vuong H., Duc H.N., 2022. A new approach based on integration of random subspace and C4.5 decision tree learning method for spatial prediction of shallow landslides. Vietnam Journal of Earth Sciences, 44(3), 327-342.

Panahi F., Ehteram M., Ahmed A.N., Huang Y.F., Mosavi A., El-Shafie A., 2021. Streamflow prediction with large climate indices using several hybrid multilayer perceptrons and copula Bayesian model averaging. Ecological Indicators, 133, 108285.

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.

Passino K.M., 2010. Bacterial foraging optimization. International Journal of Swarm Intelligence Research (IJSIR), 1(1), 1-16.

Peng F., Wen J., Zhang Y., Jin J., 2020. Monthly streamflow prediction based on random forest algorithm and phase space reconstruction theory. IOP Journal of Physics: Conference Series, 1637(1), 1-6.

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

Reis G.B., Silva D.D., Filho E.I.F., Moreira M.C., Veloso G.V., Fraga M.S., Pinheiro S.A.R., 2021. Effect of environmental covariable selection in the hydrological modeling using machine learning models to predict daily streamflow. Journal of Environmental Management, 290, 112625.

Sahib M.A., Abdulnabi A.R., Mohammed M.A., 2018. Improving bacterial foraging algorithm using non-uniform elimination-dispersal probability distribution. Alexandria Engineering Journal, 57(4), 3341-3349.

Samanataray S., A. Sahoo, 2021. A Comparative study on prediction of monthly streamflow using hybrid ANFIS-PSO approaches. KSCE Journal of Civil Engineering, 25(10), 4032-4043.

Saraiva S.V., de Oliveira Carvalho F., Santos C.A.G., Barreto L.C., Freire P.K.D.M.M., 2021. Daily streamflow forecasting in Sobradinho Reservoir using machine learning models coupled with wavelet transform and bootstrapping. Applied Soft Computing, 102, 107081.

Seo Y., Kim S., Kisi O., Singh V.P., 2015. Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. Journal of Hydrology, 520, 224-243.

Sharma P.J., Patel P., Jothiprakash V., 2021. Data-driven modelling framework for streamflow prediction in a physio-climatically heterogeneous river basin. Soft Computing, 25(8), 5951-5978.

Siddiqi T.A., Ashraf S., Khan S.A., Iqbal M.J., 2021. Estimation of data-driven streamflow predicting models using machine learning methods. Arabian Journal of Geosciences, 14(11), 1-9.

Tao L., He X., Wang R., 2017. A hybrid LSSVM model with empirical mode decomposition and differential evolution for forecasting monthly Precipitation. Journal of Hydrometeorology, 18(1), 159-176.

Tesemma Z., Wei Y., Peel M., Western A., 2015. The effect of year-to-year variability of leaf area index on Variable Infiltration Capacity model performance and simulation of runoff. Advances in Water Resources, 83, 310-322.

Tian D., He X., Srivastava P., Kalin P., 2021. A hybrid framework for forecasting monthly reservoir inflow based on machine learning techniques with dynamic climate forecasts, satellite-based data, and climate phenomenon information. Stochastic Environmental Research and Risk Assessment, 1-23.

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.

Tikhamarine Y., Souag-Gamane D., Kisi O., 2019. A new intelligent method for monthly streamflow prediction: hybrid wavelet support vector regression based on grey wolf optimizer (WSVR-GWO). Arabian Journal of Geosciences, 12(17), 1-20.

Tran V.N., J. Kim, 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.

Tyralis H., Papacharalampous G., Langousis A., 2021. Super ensemble learning for daily streamflow forecasting: Large-scale demonstration and comparison with multiple machine learning algorithms. Neural Computing and Applications, 33(8), 3053-3068.

Wang J.-S., Li S.-X., 2019. An improved grey wolf optimizer based on differential evolution and elimination mechanism. Scientific Reports, 9(1), 1-21.

Wang J., Cui Q., Sun X., 2021. A novel framework for carbon price prediction using comprehensive feature screening, bidirectional gate recurrent unit and Gaussian process regression. Journal of Cleaner Production, 314, 128024.

Wang K., Shi H., Chen J., Li T., 2019. An improved operation-based reservoir scheme integrated with Variable Infiltration Capacity model for multiyear and multipurpose reservoirs. Journal of Hydrology, 571, 365-375.

Wang X., Xu J., Shi W., Liu J., 2019. OGRU: An optimized gated recurrent unit neural network. Journal of Physics: Conference Series, 1325, 012089.

Zakhrouf M., Hamid B., Kim S., Madani S., 2021. Novel insights for streamflow forecasting based on deep learning models combined the evolutionary optimization algorithm. Physical Geography, 1-24.

Zhang X., Lin Q., Mao W., Liu S., Dou Z., Liu G., 2021. Hybrid Particle Swarm and Grey Wolf Optimizer and its application to clustering optimization. Applied Soft Computing, 101, 107061.

Zhao X., Lv H., Sang Y., Wei Y., Zhu X., 2021. Enhancing robustness of monthly streamflow forecasting model using gated recurrent unit based on improved grey wolf optimizer. Journal of Hydrology, 601, 126607.

Zhao X., Lv H., Wei Y., Lv S., Zhu X., 2021. Streamflow forecasting via two types of predictive structure-based gated recurrent unit models. Water, 13(1), 91.