



OPEN

## Developing a novel hybrid model based on GRU deep neural network and Whale optimization algorithm for precise forecasting of river's streamflow

Amin Gharehbaghi<sup>1</sup>, Redvan Ghasemlounia<sup>2</sup>, Farshad Ahmadi<sup>3</sup>, Rasoul Mirabbasi<sup>4</sup> & Ali Torabi Haghghi<sup>5</sup>✉

Streamflow contemplates a fundamental criterion to evaluate the impact of human activities and climate changes on the hydrological cycle. In this study, a novel innovative deep neural network (DNN) structure by integrating a double Gated Recurrent Units (GRU) neural network model with a multiplication layer and meta-heuristic whale optimization algorithm (WOA) (i.e., hybrid 2GRUx-WOA model) is developed to improve the prediction accuracy and performance of mean monthly Chehel-Chai River's streamflow ( $CCRSF_m$ ) in Iran. The Pearson's correlation coefficient (PCC) and Cosine Amplitude Sensitivity (CAS) as feature (input) selection process determine the only precipitation ( $P_m$ ) as the most effective input variable among a list of on-site potential climate time series parameters recorded in the study area. Thanks to a well-proportioned layer network structural framework in the suggested hybrid 2GRUx-WOA model, it leads to an appropriate total learnable parameter (TLP) compared to standard individual GRU and Bi-GRU as the benchmark models developed in the comparable meta-parameters. This hybrid model under the optimal meant meta-parameters tuned i.e., coupling a state activation functions (SAF) of *tanh-softsign*, dropout rate (P-rate) of 0.5, numbers of hidden neurons (NHN) of 70, outperforms with an  $R^2$  of 0.79,  $NSE$  of 0.76,  $MAE$  of 0.21 ( $m^3/s$ ),  $MBE$  of -0.11( $m^3/s$ ), and  $RMSE$  of 0.36 ( $m^3/s$ ). Hybridizing the 2GRUx model with WOA algorithm causes to increase in the value of  $R^2$  by 6.8% and reduce in the value of  $RMSE$  by 20.4%. Comparatively, standard individual GRU and Bi-GRU models result in an  $R^2$  of 0.59 and 0.66,  $NSE$  of 0.55 and 0.6,  $MAE$  of 0.91 and 0.53 ( $m^3/s$ ),  $MBE$  of 0.047 and -0.06 ( $m^3/s$ ),  $RMSE$  of 1.29 and 0.83 ( $m^3/s$ ), respectively.

**Keywords** GRU and Bi-GRU models, Meta-heuristic Whale optimization algorithm, Novel hybrid 2GRUx-WOA model, TLP parameter, Chehel-Chai river's streamflow

### Abbreviations

MLMs	Machine learning models
ANN	Artificial neural network
f	Non-linear function
LSTM	Long short-term memory
GRU	Gated recurrent units
Bi-GRU	Bi-directional GRU
WOA	Whale optimization algorithm
RMSE	Root mean square error ( $m^3/s$ )
STDV	Standard deviation

<sup>1</sup>Department of Civil Engineering, Faculty of Engineering, Hasan Kalyoncu University, Şahinbey, Gaziantep 27110, Turkey. <sup>2</sup>Department of Civil Engineering, Faculty of Engineering, Istanbul Gedik University, Istanbul 34876, Turkey. <sup>3</sup>Department of Hydrology & Water Resources Engineering, Shahid Chamran University of Ahvaz, Ahvaz 6135783151, Iran. <sup>4</sup>Department of Water Engineering, Shahrokh University, Shahrokh, Iran. <sup>5</sup>Water, Energy, and Environmental Engineering Research Unit, University of Oulu, Oulu, Finland. ✉email: ali.torabihaghghi@oulu.fi

$R^2$	Determination coefficient [-]
CV	Coefficient of variation
NSE	Nash-Sutcliffe efficiency [-]
R	Correlation coefficient [-]
$CCRSF_m$	Mean monthly Chehel-Chai River's streamflow ( $m^3/s$ )
$P_m$	Mean monthly precipitation (mm)
PCC	Pearson's correlation coefficient
CAS	Cosine amplitude sensitivity
SAF	State activation functions
NHN	Numbers of hidden neurons
TLP	Total learnable parameter
MAE	Mean absolute error ( $m^3/s$ )
MBE	Mean bias error ( $m^3/s$ )

## Background and literature review

During recent eras, due to quick populace growth, industrialization, urbanization, and increasing civic water demands, providing water is a crucial undertaking for lawmaking<sup>1-3</sup>. It necessitates a strong evaluation of available and future water supplies and the influences of climate and environmental changes on socio-hydrologic systems<sup>4</sup>. Devotion to this anxiety has been amplified lately because of water crises.

Climate changes and anthropogenic doings have brought on a perceptible intensification in periodic surface hydrologic extreme occasions such as an increase in frequency and intensity of universal temperature, rainfall, streamflow, droughts, and floods in the world in the twenty-first century<sup>5,6</sup>.

In the prior eras, hydrologists made many investigations to reply to the next question "What occurs to precipitations?". Streamflow has been regularly pondered as an inclusive response factor to evaluate watershed climatology, hydrology, and other catchment features. Precise spatiotemporal streamflow forecasting as a periodic feature of the atmospheric hydro-meteorological factors are very central matters for hydrologists in the water-related sectors such as regional cascade planning and managing of water resources, irrigation, municipal sustainable development, hydropower generation systems, optimum reservoir operation, agricultural planning, flooding control and risk scrutiny; social security and catastrophe hindrance<sup>7-9</sup>.

Streamflow forecasting for the long lead time is still a challenging mission as analyzing the river manners for operative objectives. For these purposes, hydrological modeling methods (HMMs) as the recognized worthy framework have been extensively utilized since the mid-1970s to scrutinize, comprehend, and forecast several complex natural procedures of periodic hydrology applications, for instance, streamflow<sup>10,11</sup>. In a wider perspective, HMMs dependent on offering solutions at diverse levels of computational intricacy are categorized into two chief kinds: (I) Process-Driven and (II) Data-Driven techniques<sup>3,12</sup>.

Adopting process-driven models for hydrological phases entails sophisticated intellectual mathematical formulas and a substantial quantity of geographical multi-source calibration data to assure a satisfactory rate of model exactness<sup>13,14</sup>. Accordingly, hydrologists have to exploit the data-driven models as an appropriate substitute to assess the intricate hydrological process. Data-driven models by applying artificial intelligence (AI) techniques try to attain a potential relation among various multidimensional dynamics predictors-target dynamics variables with exceptionally complex unbalanced trends without any previous hypothesis or information on the fundamental physical latent features and relationships among them in estimation catchments<sup>15,16</sup>. These flexible and robust approaches have been developed to obviate the troubles of numerical tactics application, costly and timewasting process of large-scale atmospheric hydro-meteorological data records in monitoring and assessing various periodic hydrological parameters in different complex geo-spatiotemporal environments and climatic regimes<sup>17</sup>. These techniques have shown admirable competence in assessing multivariate spatiotemporal byzantine and nonlinear univariate time series hydrological events and hydraulic variables in complex environs and climate change such as modeling the periodic groundwater level variations<sup>18</sup>, daily air temperature<sup>19</sup>, reservoir inflows discharge<sup>20</sup>, discharge coefficient of diverse weirs<sup>21,22</sup>, dimensions of flow separation zone<sup>23</sup>, drought forecasting<sup>24</sup>.

As far as this, to predict streamflow in different environs and hydro-climatic conditions, abundant water science engineers have developed diverse kinds of traditional single data-mining models e.g. Artificial Neural Networks (ANN), Genetic Programming, and Adaptive Neuro-Fuzzy Inference Systems (ANFIS)<sup>25-27</sup>, Non-Linear Autoregressive Moving Average with Exogenous Input Polynomial model<sup>28</sup>, Autoregressive-Moving Average, Autoregressive (AR) Moving, and Multivariate Adaptive Regression Splines (MARS) models<sup>29-31</sup>, Gaussian Process Regression model<sup>32,33</sup>, Functional Linear models<sup>34</sup>, Regression Tree models<sup>35</sup>, Online Sequential Extreme Learning Machine models<sup>36-38</sup>, and ensemble and stochastic conceptual data-driven methods<sup>39</sup>, Empirical Random Forest Family's model<sup>40</sup>, Bayesian Model Averaging<sup>41</sup>, Support Vector Regression (SVR) model<sup>42-44</sup>. Soo et al. compared the ability of five machine learning models (MLMs), including K-Nearest Neighbors, Support Vector Machine (SVM), Random Forest (RF), ANN, and Long Short-Term Memory (LSTM) in forecasting in Klang River Basin, Malaysia<sup>45</sup>. They announced that within the methods used, RF – III presented superior performance with Symmetric Mean Absolute Percentage Error and Median absolute percentage error amounts of 0.36 and 0.37, respectively.

Of data-driven models, Gated Recurrent Units (GRU) neural network was presented by Cho et al. as a reformed kind of LSTM, the most popular version of deep neural networks (DDNs)<sup>46</sup>. It can effectively satisfy the innate gradient vanishing problem in the usual neural networks and the temporal data using pertinent interior gates<sup>47,48</sup>. GRU is contemplated as the leading approach with noteworthy advancement in real-time forecasting of the geological nonlinear rainfall-runoff time series process more effectively with delay times of more than a few months of the watershed. Generally, DDNs can accomplish higher forecasting precision than

physical process-based approaches in most situations as they are able to analyze deeper structures and mine high-dimensional data. They have been utilized to assess streamflow process in different watersheds by the majority of researchers<sup>49–51</sup>.

The multivariate complex time series hydrological variables particularly streamflow, consist of naturally both perpetual stochastic and deterministic elements. Hence, an unimpeachable real-time appraisal contemplates a strenuous and time-consuming concern as a consequence of their extreme long-term non-stationarities and uncertainties latent in the spatiotemporal input-target data, randomization, human interventions, and highly indeterminate nonlinear characteristic accompanied by multipart interactions/forms within atmospheric elements<sup>52,53</sup>. For these reasons, the applicability of the traditionally used individual regression-based and data-mining tools is almost ineffectual and has generally bumped into serious predicaments such as high spatial-temporal fluctuations depending on severe uncertainties, nature resolutions, weights fit-tuning, etc<sup>54–56</sup>.

Taking all the together, to impede the all above-mentioned difficulties and interludes, different strong and leading-edge hybrid signal pre-processing (mode decomposition) and bio-inspired optimization-based data-mining models for large-scale data analytical as an appropriate alternative are being developed to improve noticeably predictions' talent and efficiency of ordinary models. The hybrid strategies employ the incorporation of two or further data integration and modelling modus operandi prompting feasible to prominently increase the exactness of forecasted streamflow data.

The nature-inspired metaheuristic optimization-based algorithms bring forth improve the ability of standalone predictive models by incorporating different optimization algorithms by realizing close-optimum results within a rational timeframe for the estimation parameter, while concurrently could diminish the computational convergence time period<sup>57</sup>. Outdated customary optimization algorithms have limitations including single-based solutions, complications in indefinite search spaces, and converging to local optima<sup>15,16,58</sup>. To date, numerous investigators have designed metaheuristic algorithms to address these limitations. On this point, the whale optimization algorithm (WOA) presented by Mirjalili & Lewis is a robust and reliable bio-inspired metaheuristic optimization algorithm motivated by the intelligence and social life manners of humpback whales<sup>59</sup>. This algorithm is characterized by the bubble-net hunting tactic and is applicable in global optimization machine learning and data mining problems by imitating physical or biological phenomena.

Hitherto, different hybrid algorithms have been developed by adjusting and optimizing the simulation factors and broader choice of the membership function to enhance the accuracy of predicting streamflow in different regions and hydro-climatic conditions. For example, different hybrid meta-heuristic optimization-based algorithms including the Shuffled Frog Leaping, Particle Swarm Optimization (PSO), Ant Colony Optimization, Gray Wolf Optimization (GWO) algorithms<sup>60</sup>, hybrid Genetic Algorithm with SVR and Bayesian Additive Regression Tree models<sup>61</sup>, hybrid Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise algorithm and GRU model with improved GWO algorithm<sup>62</sup>, hybrid Gravitational Search, PSO and GWO optimization algorithms with the extreme learning machine (ELM) model<sup>63</sup>, hybrid optimally pruned ELM (OP-ELM), least square support vector machine, seasonal auto regressive moving average, MARS, and M5 model tree<sup>64</sup>, hybrid SVR and generalized regression neural network models with seasonal and trend decomposition algorithms<sup>65</sup>, hybrid fuzzy information granulation with SVR model<sup>66</sup>. A summary of the different hybrid models developed for predicting streamflow in different areas is presented in Table 1.

### Motivation for this study

This research intends to estimate a long-term time series of mean monthly Chehel-Chai River's streamflow ( $CCRSF_m$ ) using climatic datasets from Sep 1990 to Aug 2020 by GRU deep learning, MLMs. To do so, first of all, the general single GRU layer network as the benchmark model, the general single Bi-directional GRU (Bi-GRU) layer network, and the double GRU coupled with a multiplication layer (i.e., 2GRU $\times$  model) network models are developed. Then, the ideal one of these initially designed meant models (based on performance evaluation metrics calculated) is intentionally hybridized with meta-heuristic whale optimization algorithm (WOA) (i.e., hybrid 2GRU $\times$ -WOA model) to further improve the prediction accuracy of  $CCRSF_m$ . Therefore, we do not limit our investigation only to the conventional deep learning (DL) network structure. Since the WOA algorithm mostly presents a steady and fast convergence rate and can identify optimum solutions in lower populations with the lesser opportunity of local trapping modes, it is used as nature-inspired optimization algorithm.

As the aptitude of these models relies on the kind and rate of some meta-parameters, realizing a fitting optimum pattern is a demanding and bewildering undertaking. Hence, various scenarios are adopted by tuning diverse meta-parameters in the construction of suggested models and the WOA algorithm.

The literature review shows that there has been little research operating different layer structures of the GRU model for time series streamflow forecasting. As far as authors know, amongst the current computational intelligence-system literature centering on streamflow prediction, only commonplace and wide-ranging simple DNN architectural structures have been focused on. The **novelty** of this research is the development of a leading-edge and robust, unique hybrid 2GRU $\times$ -WOA model with different analytical layer network structures, for the first time to predict more complex natural phenomena such as time series CCRSFm oscillations patterns.

### Study area and data description

Chehel-Chai River is one of the main branches of Gorganrood River and is located in Golestan Province, northern Iran. The Chehel-Chai watershed is situated inside the city border of Minoodasht with an area of 256,830 (m<sup>2</sup>), a mean slope of 46%, a maximum and minimum elevation of 2570 and 190 m above sea level, and a moist environment. The mean yearly precipitation and temperature were reported as 750 mm and 15.4 (°C), respectively. Forest (60%) and rain-fed lands (39%) shape the chief surrounding ground cover in this area<sup>76,77</sup>. Figure 1 shows Golestan Province and the location of the observation station generated using QGIS 3.40.

Location	<sup>a</sup> Models used (Best)	Modelling results	References
Three diverse watersheds in Germany	<b>AR-RBF</b> , MLP-RF-PR	$R^2 = 0.7-0.84$	Granata et al. <sup>67</sup>
Ten watersheds' data set gathered from CAMELS, US	<b>EFS-KESVR-EMA</b> , EFS-LSTM-EMA	$NSE = 0.73-0.94$	Xu et al. <sup>68</sup>
Two USGS stations, US	hybrid EMD-RFR, EMD-Bagging, EMD-AdaBoost, <b>EMD-ANN</b>	$R = 0.97, RMSE = 0.33, MAE = 0.17, NSE = 0.94$	Heddam et al. <sup>69</sup>
Dez River, Iran	SVMD-MLP-PSO	$R^2 = 0.89, RMSE = 13.91, NSE = 0.88$	Parsaie et al. <sup>70</sup>
Nile River at the High Aswan Dam, Egypt	MLP-EO, MLP-HGSO, <b>MLP-NRO</b>	$MAE = 1.35, RMSE = 2.35, R = 0.92$	Ahmed et al. <sup>71</sup>
Gaula barrage site in Uttarakhand state of India	ANFIS, ANN, <b>WANN</b>	$R = 0.99, RMSE = 5.51 (\text{ft}^3/\text{sec}), WI = 0.96, COE = 0.99$	Shukla et al. <sup>72</sup>
Yuelai New City, China	LightGBM-SSA	$NSE$ more than 0.9, peak flow forecasting error within 18%	Cui et al. <sup>73</sup>
Han River, China	VMD-DBN-IPSO	$NSE$ more than 0.8, peak flow forecasting error within 20%	Xie et al. <sup>74</sup>
Pahang River, Malaysia	<b>ANFIS-FFA</b> , ANFIS	$R = 1, RMSE = 0.98, MAE = 0.36, NSE = 1$	Yaseen et al. <sup>75</sup>

**Table 1.** Hybrid models for forecasting streamflow over different time scales. <sup>a</sup>Best model is highlighted in bold. AR-RBF: Additive Regression of Radial Basis Function Neural Networks. MLP-RF-PR: stacking with the Pace Regression of the Multilayer Perceptron with Random Forest. EFS-KESVR-EMA: Hybrid Embedded Feature Selection (EFS), Kernel Extreme Support Vector Regression (KESVR), and Evolutionary Mating Algorithm (EMA). EO: Equilibrium Optimization algorithm. HGSO: Henry Gases Solubility Optimization algorithm. NRO: Nuclear Reaction Optimization algorithm. ANFIS-FFA: hybrid ANFIS with the Firefly Optimizer algorithm (FFA). SVMD: successive variational mode decomposition. SWAT-MLP/MWOA: hybrid SWAT (soil and water assessment tool) and MLP neural network optimized by mutated whale optimization algorithm (MWOA). VMD-DBN-IPSO: hybrid VMD and deep belief network (DBN) model integrated with improved PSO (IPSO) algorithm. LightGBM-SSA: singular spectrum analysis (SSA) and light gradient boosting machine (LightGBM). WGEP: hybrid wavelet-gene expression programming. WSVM: wavelet-support vector machine. WANN: Wavelet-based artificial neural system. PCC: Pearson correlation coefficient. WI: Willmott index. COE: coefficient of efficiency. RFR: random forest regression. Bagging: bootstrap aggregating. AdaBoost: adaptive boosting, EMD: empirical mode decomposition.

In the current study, to forecast time series mean monthly Chehel-Chai River's streamflow ( $CCRSF_m$ ) ( $\text{m}^3/\text{s}$ ), 360 monthly atmospheric datasets documented from Sep 1990 to Aug 2020 by Lazoreh climatic observation station are exploited. The historic climatic parameters operated include monthly mean, maximum, minimum, absolute minimum, and absolute maximum air temperature ( $T_{ave}, T_{max}, T_{min}, T_{min}, T_{max}$ ), precipitation ( $P_m$ ), evapotranspiration ( $ET_m$ ) gotten from the IMO (Iran Meteorological Organization). Table 2 provides some descriptive statistics indices of the variables used in the study area and period.

## Research objectives

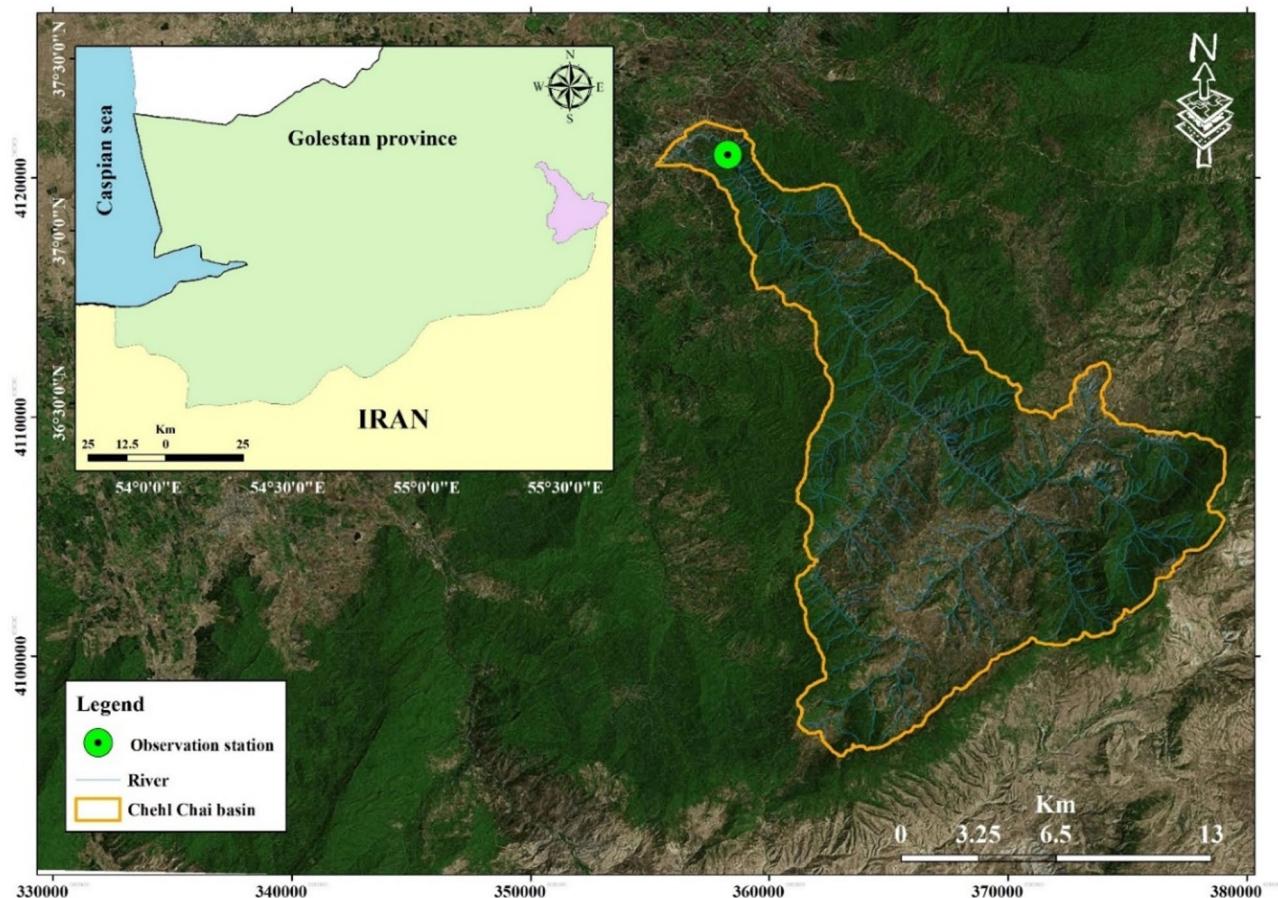
The main contributions of the present work are as follows:

1. Identify the most effective variables on  $CCRSF_m$ , among a list of on-site potential climate parameters recorded through feature selection techniques.
2. Development of standalone and hybrid GRU-based neuro-evolution time series paradigms optimized by WOA nature-inspired metaheuristic algorithm for precise forecasting of  $CCRSF_m$  oscillations rhythm.
3. Determine the optimal spectrum of aimed meta-parameters in GRU-based models developed and WOA optimization algorithm for better configuration and lessening the impact of overfitting/underfitting problems.
4. Assess and compare the accuracy of modeling with counterparts in the validation stage to differentiate the attributes of the best-developed model in offering better reliable and consistent performance using some comparison plots and statistical metrics.

## Feature selection process

Because the performance of any modeling is influenced mainly by an apt selection of input variables for the precise prediction of target, unfit selections could adversely affect the effectiveness of any methodology. So, in this section, existing large-dimensional potential hydro-climatic data sets recorded in the studied region are evaluated to recognize the most effective input variables for predicting  $CCRSF_m$  as the model target variable. In this context, the variables of extreme importance are selected using Pearson's correlation coefficient (PCC) and Cosine Amplitude Sensitivity (CAS) as linear and nonlinear representative data analysis methods.

The CAS data inquiry for the variables presented in Table 3 is done by altering each input variable at a fix ratio and holding the other input variables constant as follows<sup>78</sup>:



**Fig. 1.** Location map of Chehel-Chai basin in Golestan Province, Iran.

Variables	Max	Min	Mean	STDV	Skewness	CV
$T_{\min}$ (°C)	22.6	-1.1	11.5	6.51	0.057	0.56
$T_{\text{ave}}$ (°C)	29.7	4.2	17.9	7.04	-0.00087	0.39
$T_{\max}$ (°C)	38.5	9.3	24.2	7.71	-0.031	0.31
$ET_m$ (mm)	264.1	25	122.5	56.3	0.51	0.45
$P_m$ (mm)	324	0	71.4	51.1	1.31	0.71
$T_{\max a}$ (°C)	46	7	33.21	6.61	-0.56	0.19
$T_{\min a}$ (°C)	20	-10.6	7.11	7.19	0.13	1.01
$CCRSF_m$ ( $m^3/s$ )	18.7	0.02	1.91	2.06	3.41	1.08

**Table 2.** Statistical indices of variables recorded in the study area. \*CV and STDV illustrate the coefficient of variation and standard deviation, respectively.

Analysis Method	$T_{\min}$	$T_{\max}$	$T_{\text{ave}}$	$T_{\min a}$	$T_{\max a}$	$P_m$	$ET_m$
PCC	-0.41	-0.39	-0.41	-0.39	-0.37	0.32	-0.29
CAS	-0.36	-0.42	-0.31	-0.28	-0.22	0.41	-0.38

**Table 3.** PCC and CAS values versus  $CCRSF_m$ .

$$R_{ij} = \frac{\sum_{k=1}^n I_{ik} O_{jk}}{\sqrt{\sum_{k=1}^n I_{ik}^2} \cdot \sqrt{\sum_{k=1}^n O_{jk}^2}} \quad (1)$$

where,  $I_i$  and  $O_j$  are input and output parameters, respectively, and  $n$  is total number of datasets. The  $R_{ij}$  value [0,1] shows the strength of the relationship within the input and target parameters. The values of PCC and CAS methods are presented in Table 3.

According to Table 3, as a result of the insignificant amount of PCC and CAS data analysis methods for  $T_{min}$ ,  $T_{max}$ ,  $T_{ave}$ ,  $T_{mina}$ ,  $T_{maxa}$ ,  $ET_m$ , their effects on predicting  $CCRSF_m$  by suggested models can be disregarded. Hence, merely  $P_m$  can be considered as the most influential and important input parameter. To conclude, the equation for the prediction of  $CCRSF_m$  can be formulated as follows:

$$CCRSF_m = f(P_m) \quad (2)$$

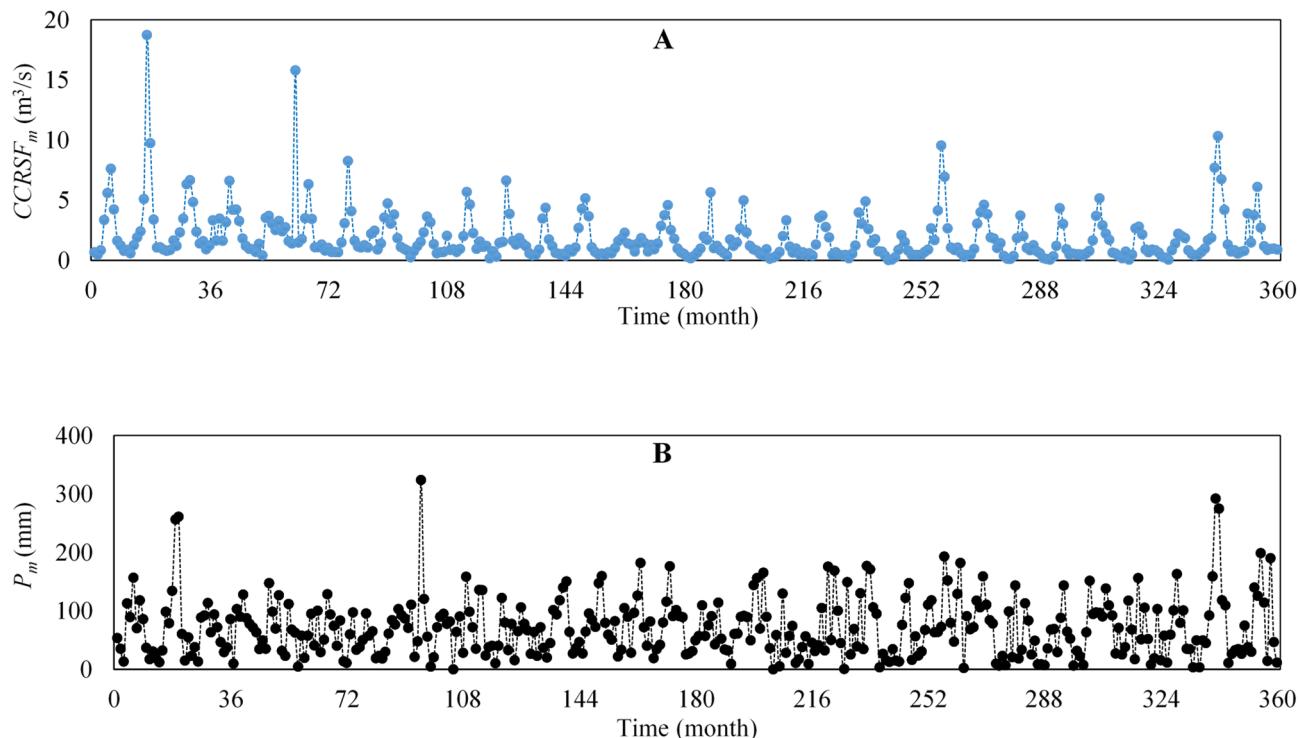
Figure 2 (A and B) shows the time series plots of  $CCRSF_m$  and  $P_m$  recorded by Lazoreh climatic observation station for the studied time that show the seasonality of data. Since the parameters have a temporal pattern, the monthly scale is used as a parameter.

### Methodology

Due to the unstable, intricate, and nonlinear relationship in Eq. 2, only precise and robust approaches are enabled to analyze  $CCRSF_m$ . In this context, a sequential dataset of 360 monthly hydro-climatic observations covering the period from September 1990 to August 2020 is used in modeling process. The datasets are normalized to zero mean and unit variance as advised by Lawrence et al.<sup>79</sup>. The normalized datasets are divided into two subclasses. One limited 70% of the data (252 monthly observations) are consecutively applied in calibrating the predictive models. And, the lasting 30% (108 samples) are set aside to be applied serially in validation, without randomization. This process warranted that the data be on a uniform scale, so discrete variable sensitivity did not complicate the results.

### GRU and Bi-GRU neural networks

Recurrent Neural Networks (RNNs) assimilate previous info to cross for forecasting the future state of a variable using input data with certain dependencies by enforcing a memory cell containing an unfolded loop cell. However, for large-scale data, its learning process meets with disappearing gradients in the backpropagation training algorithm over time<sup>80</sup>. To overcome this problem, LSTM adds intentionally hidden units to the memory cell of RNNs, so that it maintains information over long periods thanks to its unparalleled progressive structure named, Credit Assignment Paths (CAPs)<sup>80–83</sup>.



**Fig. 2.** Time series graphs of variables used in Eq. 2 between Sep 1990-Aug 2020 (360 months) in the Chehel-Chai River watershed: (A)  $CCRSF_m$ , (B) Precipitation ( $P_m$ ).

GRU neural network is capable of learning long-term relationships and assessing highly nonlinear historical information if the data scale is not too vast. Both GRU and LSTM neural network models act in an alike way with an analogous central framework. GRU is an adaptive neural network since it is quicker in computing, simpler in learning, less condensed in construction, and has fewer learnable factors with a distinguished inherent proficiency<sup>84,85</sup>. Figure 3 illustrates the working order of the system and the interior memory cell of GRU, where  $r_t$ ,  $x_t$ , and  $h_t$  are the reset gate of GRU, input variable, and hidden state at time  $t$ , respectively.  $W_r$  and  $U_r$  denote weight matrices for the input data and hidden state, respectively.

Against LSTM, GRU does not include isolated memory cells, as a substitute, it employs a separate  $h_t$  to dispense data over time steps. Furthermore, the input and forget gates are incorporated with an update gate ( $z$ ), and  $r_t$  is straightly applied to  $h_{t-1}$  to obtain  $h_t$  (the candidate state). In this system, the memory cell learns at time  $t$  by the input at time  $t$  and the output at the prior time step ( $t-1$ ). The instruction of GRU is defined by the subsequent computations<sup>46</sup>:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (3)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (4)$$

$$\tilde{h}_t = \tanh(W \tilde{h} x_t + U \tilde{h} (r_t \times \tilde{h}_{t-1}) + b \tilde{h}) \quad (5)$$

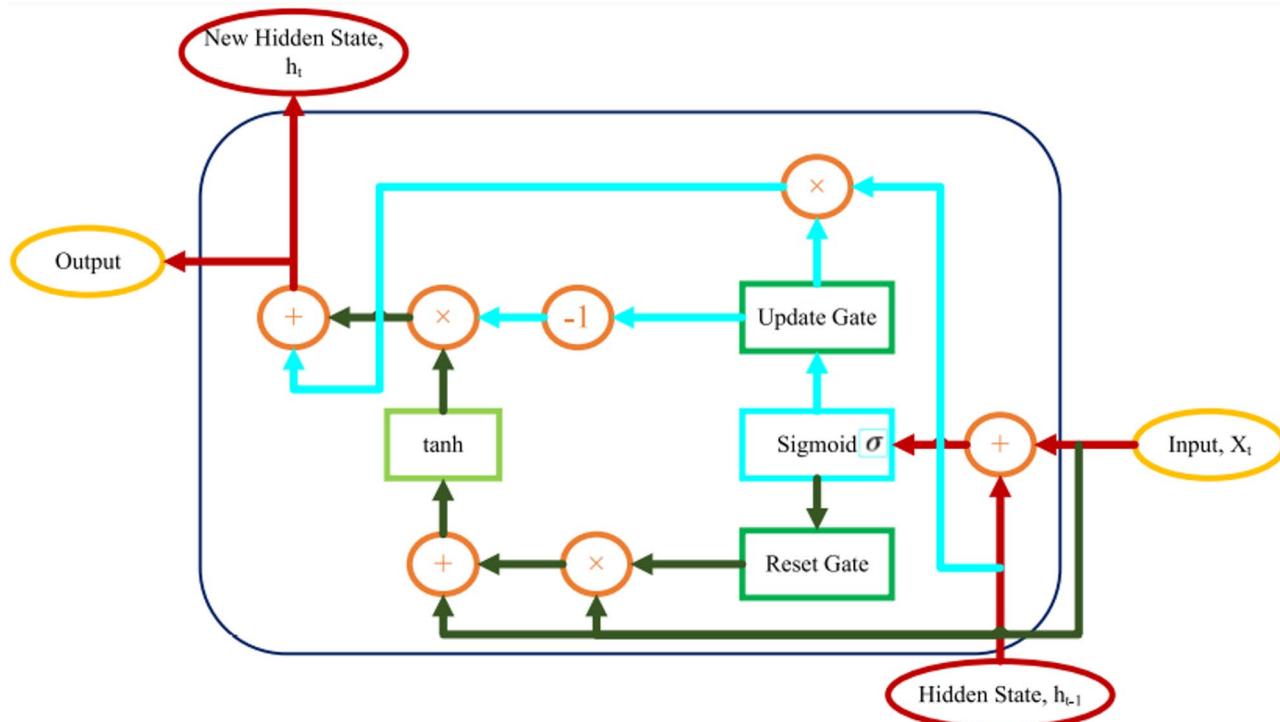
$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \quad (6)$$

In these equations,  $\tanh$  and  $\sigma$  are the hyperbolic tangent and logistic sigmoid functions, respectively. The sign “ $\times$ ” and  $b$  denotes the element-wise multiplication and bias vector, respectively. These factors are learnable sets. Attributable to the impact of  $\sigma$ , whole gates are a vector within (0, 1). When the  $r_t$  is locked, GRU is influenced just by  $x_t$  and  $z_t$  controls the information dimension of  $h_{t-1}$  can be passed into  $h_t$ <sup>46</sup>.

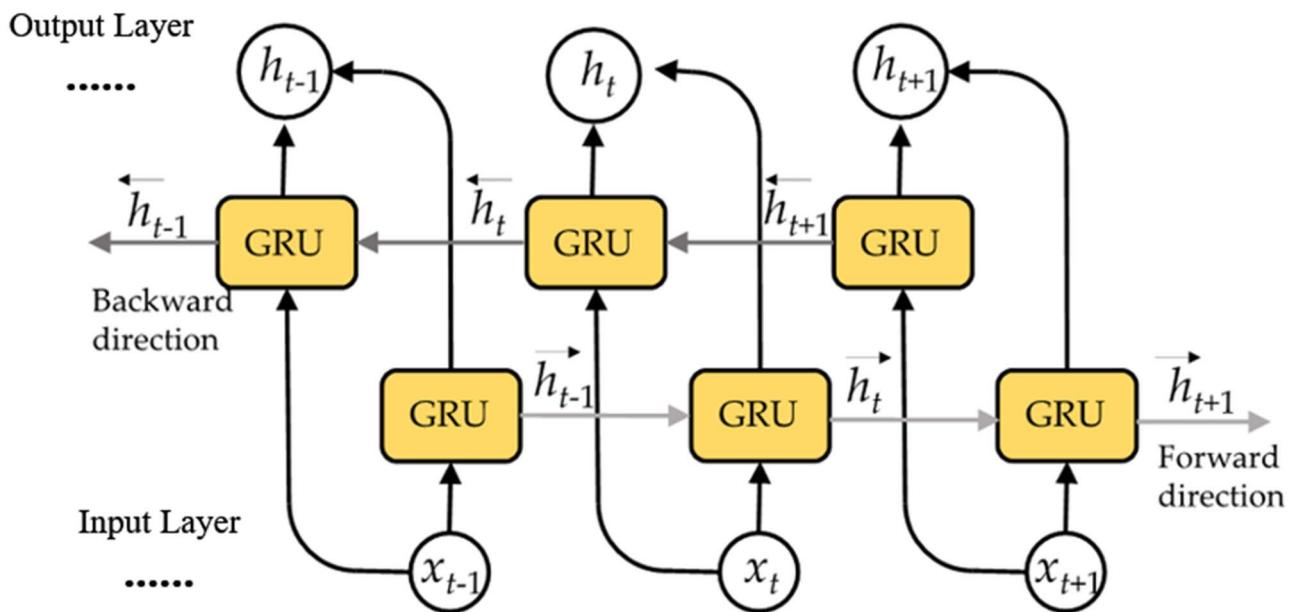
The GRU model only considers the effect of the prior information on the succeeding information without regarding the correlation sides in time series predicting<sup>86</sup>. The Bi-GRU network model with several gates in the memory cells is based on different forms of the general one-directional GRU; nonetheless, reiterating elements within the hidden layer are more intricate. It includes forward and backward GRU to manage the input-output current inside the network and extract features in-depth by forward and reverse historic sequence computations<sup>87,88</sup>. The model construction is displayed in Fig. 4. Last of all, the output is calculated by the following formulation<sup>86</sup>:

$$h_t = \vec{h}_t \odot \hat{h}_t \quad (7)$$

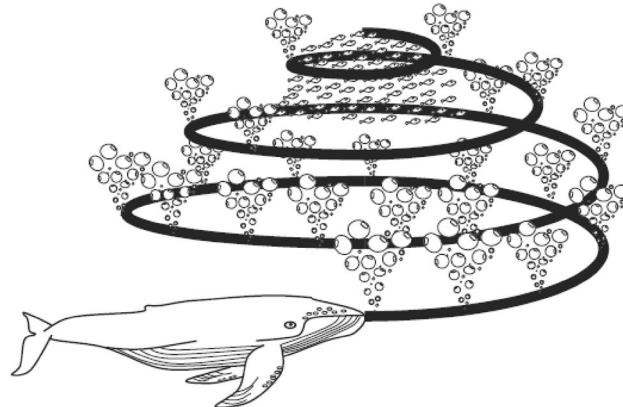
where,  $\vec{h}_t$  and  $\hat{h}_t$  are the outputs of the forward and backward GRU, respectively.



**Fig. 3.** Internal structure and mechanism of GRU memory cell.



**Fig. 4.** Structure of Bi-GRU neural network model.



**Fig. 5.** Bubble-net nourishing method of humpback whales<sup>59</sup>.

### Whale optimization algorithm (WOA)

Whales are considered the largest mammals with strong intellectual and emotional abilities, like humans, in the world. The whales lived habitually in groups, and the humpback whales have an exceptional hunting technique known as the bubble-net feeding technique. It enforces functional twisting movements to make a bubble-net raiding mechanism called bubble-net feeding. These bubbles are called double-loops and upward spirals. Humpback whales wish to forage for small fish or krill schools near the sea's surface. It has been perceived that this hunting mode is done by forming typical bubbles sideways a loop or '9'-shaped track as exposed in Fig. 5<sup>59,89,90</sup>.

The WOA was presented by Mirjalili & Lewis<sup>59</sup> and is considered a well-known swarm intelligence algorithm inspired by the real special chasing tactics of humpback whales in nature, done by the unsystematic or finest search agent to hunt the prey. It is executed in three stages: (i) Siege chasing; (ii) Operation stage: The process of raiding the net bubble; (iii) Exploration stage: Chasing search. Some unsystematic solutions initiated WOA. In each reiteration, the search agents bring their situation up to date by the three operators. The WOA supposes that the optimum solution at the moment (the optimum response) is prey; so, it identifies prey and then encircles prey. When the agent of the optimum search is recognized, other search agents inform their place to the optimum search agent<sup>59</sup>. The following formulas can describe this process:

$$\vec{D} = |\vec{C} \cdot \vec{X}^*(t) - \vec{X}(t)| \quad (8)$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A} \cdot \vec{D} \quad (9)$$

where,  $t$  signifies the present repetition,  $\vec{A}$  and  $\vec{C}$  are the coefficient vectors,  $X^*$  is the position vector of the optimum solution attained at present, and  $\vec{X}$  is the position vector.  $\vec{D}$  also displays the space amid the hunt and  $i$ th whale. The vectors  $\vec{A}$  and  $\vec{C}$  are computed as follows:

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \quad (10)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (11)$$

where,  $\vec{a}$  is linearly reduced from 2 to 0 over the progress of repetitions (in both exploration and exploitation stages) and  $\vec{r}$  is a random vector in  $[0,1]$ .

The bubble-net invading method contains two chief stages: (i) the encirclement process which whales drive to the water surface) denotes shrinking and includes decreasing  $\vec{a}$  in the Eq. (10). The quantity of  $\vec{A}$  reduces as  $\vec{a}$  declines; and (ii) spiral updating of the whales' locations is utilized to mimic the spiral activities of whales in the hunt boundary by computing the distance amid the hunt ( $X^*, Y^*$ ) and the hunter ( $X, Y$ )<sup>59</sup>:

$$\vec{X}(t+1) = \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) \quad (12)$$

$\vec{D}' = |\vec{X}^*(t) - \vec{X}(t)|$  describes the distance amid the  $i$ th whale and the hunt;  $b$  shows a coefficient defining the form of the logarithmic helix-formed movements;  $l$  shows a random quantity in  $[-1,1]$ . The movement of the whales near the prey happens alongside the spiral-formed routes by wincing the loops. The following equation is presented to describe this process<sup>59</sup>:

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - \vec{A} \cdot \vec{D} & \text{if } p < 0.5 \\ \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) & \text{if } p \geq 0.5 \end{cases} \quad (13)$$

where,  $p \in [0,1]$ ; this permits one to catch the likelihood of retaining the spin mode so as to bring up-to-date the positions of the whales. In the exploration (searching) stage, the humpback whales' quest for the prey is arbitrarily consistent with their position as matched to other whales<sup>59</sup>. Thus, the whales bring their situations up-to-date, compliant with randomly chosen search factors instead of the premium search factor<sup>59</sup>:

$$\vec{D} = \left| \vec{C} \cdot \overrightarrow{X_{rand}} - \vec{X} \right| \quad (14)$$

$$\vec{X}(t+1) = \overrightarrow{X_{rand}} - \vec{A} \cdot \vec{D} \quad (15)$$

where,  $\overrightarrow{X_{rand}}$  is a random site detected by the present population. Figure 6 offers pseudo-code of the WOA algorithm.

### Model development

As mentioned above, in this study, first, the general single GRU and Bi-GRU neural networks are used as benchmark models for performance comparison, and then a 2GRU $\times$  neural network model with sequence output mode is developed to estimate  $CCRSF_m$  in the study area. Figure 7A–C portrays the GRU-based layer network structure of models.

Finally, due to the smallest  $RMSE$  and highest  $R^2$  values in Model 3 compared to Models 1 and 2, it is preferred to hybridize with the WOA algorithm (i.e., hybrid 2GRU $\times$ -WOA model (4)) to improve the prediction accuracy of CCRSFm further. In this model, WOA trained the bias and weights within layers of the 2GRU $\times$  model so that they could be updated in keeping with a proportion recognized by the premium WOA. Then, the training set is applied to the renovated bias and weights of the 2GRU $\times$  model, and the WOA optimizes them in each repetition by randomly dispensing mode. This hybrid model terminates as a maximum number of iterations are gotten or as the best solution is obtained for a certain number of iterations; if not, it proceeds with the next generation operation.

The capability and computation complexity of the hybrid 2GRU $\times$ -WOA model generally relies on using the suitable main deterministic factors of the WOA algorithm, including population size (PS), maximum number of iterations, total load demand, up-coefficient vector, and down-coefficient vector. Selecting the optimal factors of bio-inspired WOA algorithm is very imperative so that the optimal key deterministic parameters and pattern for 2GRU $\times$  model be achieved using the optimization process. Figure 8 defines the forecasting process of the hybrid 2GRU $\times$ -WOA model.

In these models, the amount of  $P$ -rate and the kind of SAF as meta-parameters are exactly tuned to realize a proper pattern and augment the skill of the models designed. Nonetheless, since there is no formal pre-instruction to identify suitable meta-parameters for DDN models with a given dataset, this process is pondered as a time-wasting and demanding task<sup>18</sup>. For the sake of this aim, several scenarios are adopted to realize valid values.

In all models, the Input layer feeds the time series  $P_m$  into the layers' network structure. For a useful configuration and modification in the big dataset, the Dropout layer is operated to deter overfitting by passing over some hidden neurons with a prearranged option rate of  $P$ <sup>91</sup>. To tune the amount of  $P$ -rate as a meta-parameter, various values are tested. To strengthen fitting ability in learning long-term sequential datasets, the Fully Connected layer is utilized with input and output sizes of "auto" and 1, respectively. The Multiplication layer multiplies the inputs from various layers' neural network element-wise. As the ending layer, the Regression Output layer is utilized to compute the "half-mean-squared-error loss" for regression objectives. To tune SAF, a

```

Initialize the whales population  $X_i$  ( $i = 1, 2, \dots, n$ )
Initialize  $a, A, C, l$  and  $p$ 
Calculate the fitness of each search agent
 $X^* =$  the best search agent
while ( $it < \text{Maximum number of iterations}$ )
  for each search agent
    Update  $a, A, C, l$ , and  $p$ 
    if ( $p < 0.5$ )
      if ( $|A| < 1$ )
        Update the position of the current search agent by the equation (7)
      else if ( $|A| \geq 1$ )
        Select a random search agent ( $X_{\text{rand}}$ )
        Update the position of the current search agent by the equation (14)
      end
    else if ( $p \geq 0.5$ )
      Update the position of the current search agent by the equation (11)
    end
  end
  Calculate the fitness of each search agent
  Update  $X^*$  if there is a better solution
   $it = it + 1$ 
  Update  $a, A, C, l$  and  $p$ 
end while
return  $X^*$ 

```

Fig. 6. Pseudo-code of the WOA algorithm (Mirjalili and Lewis, 2016).

diverse combination of *tanh* and *softsign* are distinctly utilized, and for *NHU*, numerous amounts are tested. In order to preclude the gradients disappearing and lessen the negative influence of padding drawbacks, a training process with 1000 maximum repetitions is set as recommended by Lin et al.<sup>18</sup>. More details about the function of layers in the all models developed were provided by<sup>92</sup>.

#### Performance evaluation metrics

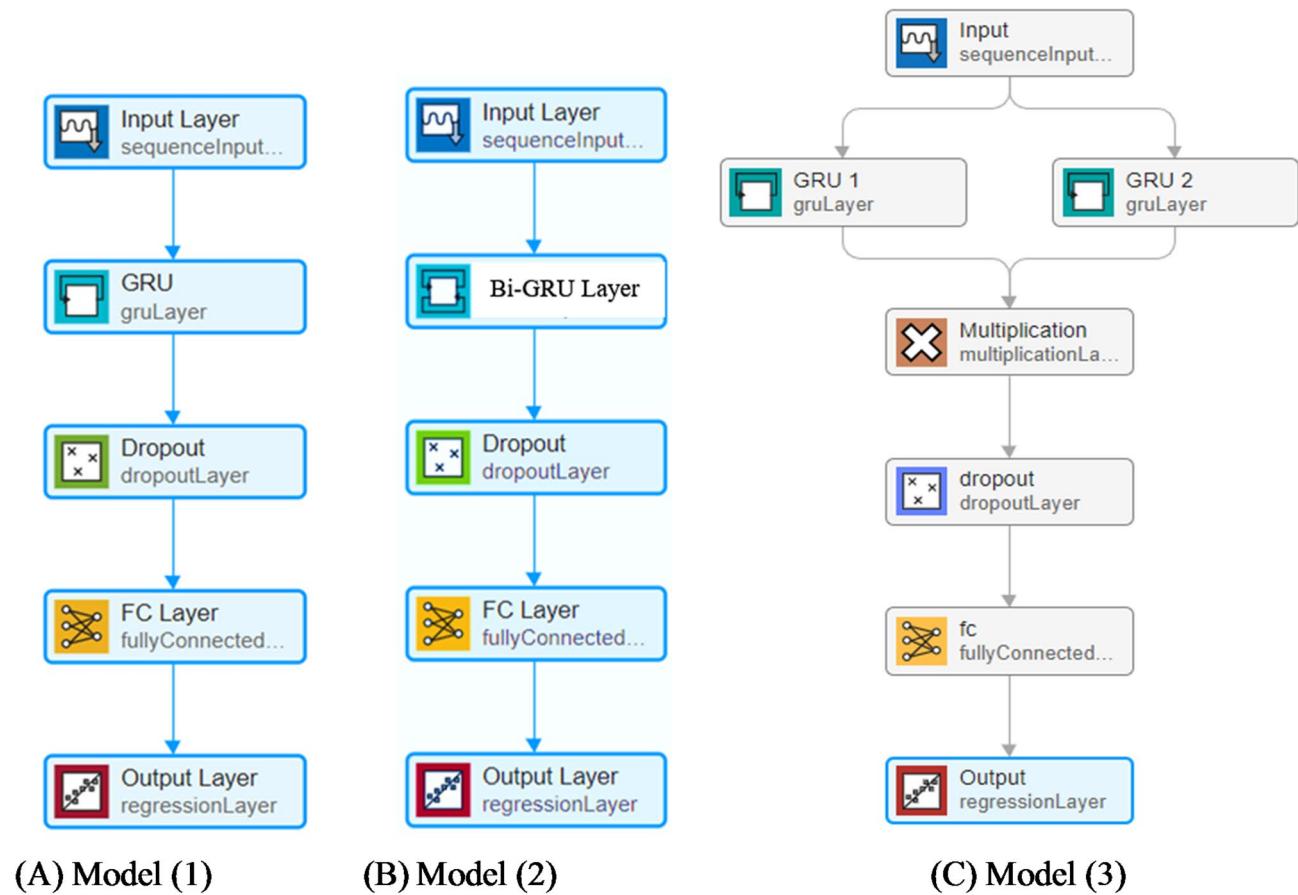
In this modelling, the following statistical metrics are applied to liken the capability and performance of all models used in predicting the time series  $CCRSF_m$ :

$$\text{Coefficient of Determination } (R^2); \\ R^2 = \frac{\sum_{i=1}^N (O_i - \mu_o)^2 - \sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \mu_o)^2} \quad (16)$$

$$\text{Root Mean Square Error } (RMSE); RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{N}} \quad (17)$$

$$\text{Mean Bias Error } (MBE); MBE = \frac{\sum_{i=1}^N (P_i - O_i)}{N} \quad (18)$$

$$\text{Nash -- Sutcliffe Efficiency } (NSE); NSE = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \mu_o)^2} \quad (19)$$



**Fig. 7.** Designed layer network structure of GRU-based models: (A) General single GRU model (1), (B) General single Bi-GRU model (2), (C) and 2GRU $\times$  model (3).

$$\text{Mean Absolute Error (MAE)} ; MAE = \frac{\sum_{i=1}^N |O_i - P_i|}{N} \quad (20)$$

where  $N$  is the number of datasets,  $P_i$  and  $O_i$  are the predicted and observed  $CCRSF_m$  at time  $i$ ,  $\sigma_o$  and  $\sigma_P$  are standard deviations of the observed and predicted the  $CCRSF_m$ . The  $\mu_p$  and  $\mu_o$  are the mean predicted and observed  $CCRSF_m$ . The best amount for Eqs. (16–20) are 1, 0, 0, 1, and 0, respectively.

## Results and discussion

### Validation of the models

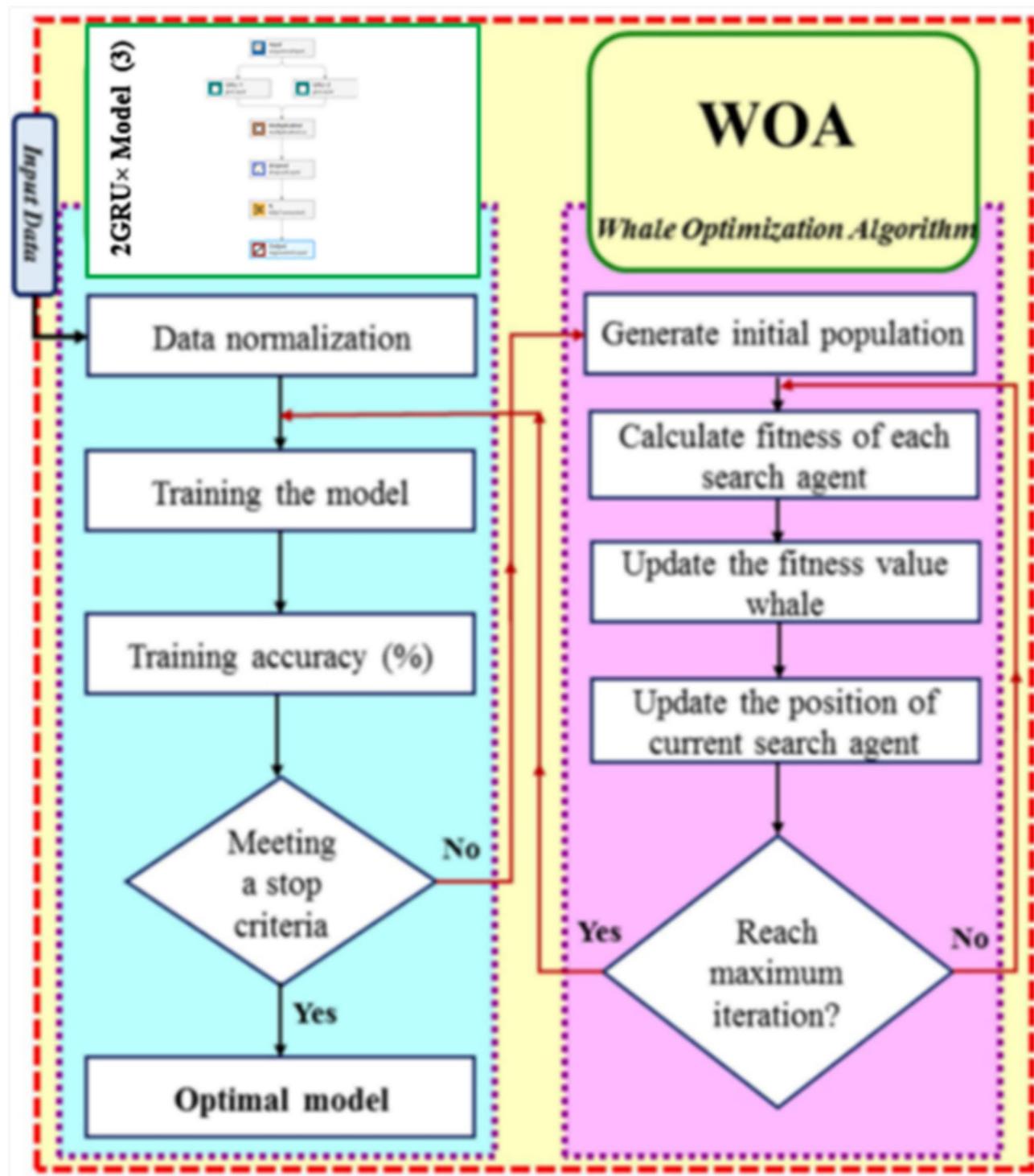
In this simulation, numerous experiments are conducted to determine the optimal value of the main deterministic factors in the models developed. The characteristics and statistical results of all models used under the optimal scenario in the validation phase in forecasting  $CCRSF_m$  are shown in Tables 4 and 5, respectively. The models in their training stages are more precise than in their testing stages. The *tanh-softsign* pairing in the hidden layers of the 2GRU $\times$  and hybrid 2GRU $\times$ -WOA models brings forth learning more complicated nonlinear functions, and accordingly, it causes the models not to be as much open to the overfitting dilemma. Besides, the ideal amount of main deterministic factors for WOA algorithm in the hybrid 2GRU $\times$ -WOA model for the best solution in forecasting  $CCRSF_m$  is achieved as an up-coefficient vector of 0.25, a down-coefficient vector of 0.1, a total load demand of 0.05, a maximum number of iterations of 500 and a population size of 30.

According to Table 5, it can be concluded that hybridizing with the WOA algorithm advances noticeably the performance and ability of the 2GRU $\times$  model. This optimization algorithm augments the 2GRU $\times$  model training phase and achieves better efficiency in the predicting  $CCRSF_m$ . Additionally, the value of *MBE* shows that all models except the model 1 underestimated the corresponding measured values at the validation phase.

*RMSE* variations in the model 4 over the range of *NHN* used under the optimal hyper-parameter in the testing phase is displayed (Fig. 9). High *NHN* causes *RMSE* to grow as a result of overfitting, nevertheless, small *NHN* reduces network learning skill because of underfitting.

### Performance comparison

In this modelling, to compare the skill and efficiency of models used in capturing the time series  $CCRSF_m$ , a substantial factor called *TLP* (Total Learnable Parameters) is exploited, as suggested by Lin et al.<sup>18</sup> (Table 4). The



**Fig. 8.** Flowchart of integrated 2GRU $\times$  with whale optimization algorithm (WOA) (i.e., hybrid 2GRU $\times$ -WOA model (4)).

TLP is considered a crucial criterion for the discriminating forecasting performances and practical capacities of DL-based models. Moreover, it can also assess tendencies toward under-/over-fitting effects.

Based on Tables 4 and 5, though Model 1 has the extreme TLP value, Model 4 results in the best performance and surpasses other models by capturing the time series CCRSF<sub>m</sub>. The dominant reason for this is explainable by the high quantity of TLP in the model, 1 which led to an extremely unnecessary network capacity and accordingly, it prompts overfitting and hinders the optimization process. On the contrary, due to the lesser amount of TLP in model 2, it has a lesser network capacity, accordingly, leading to underfit. Generally, models 1 and 2 are not capable of monitoring and predicting time series CCRSF<sub>m</sub> in the study region due to poor

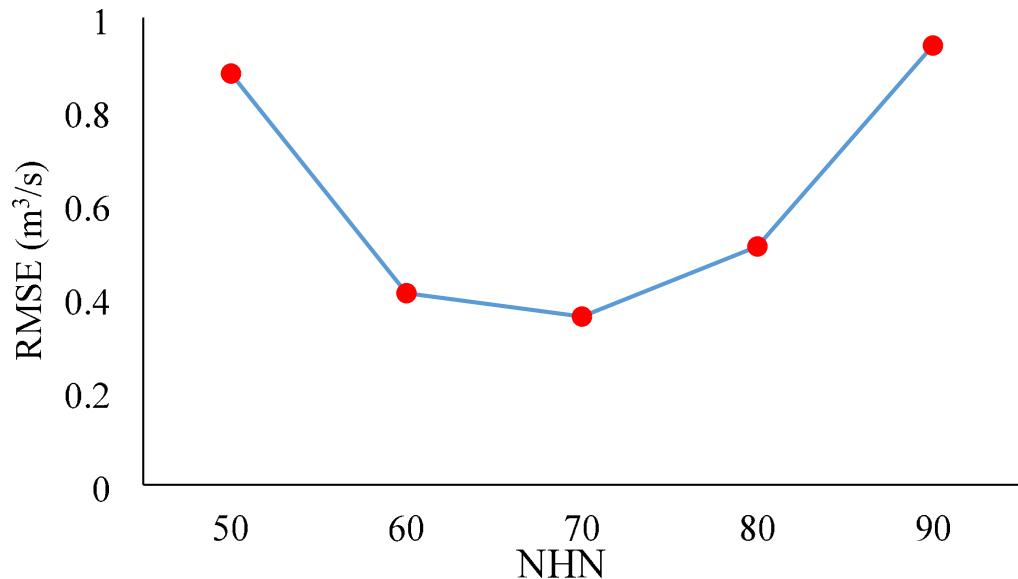
Main Deterministic Factors	GRU Model (1)	Bi-GRU Model (2)	2GRU $\times$ Model (3)	Hybrid 2GRU $\times$ -WOA Model (4)
NHN	60	50	60	70
SAF	<i>tanh</i>	<i>tanh</i>	<i>tanh-softsign</i>	<i>tanh-softsign</i>
P-rate	0.4	0.6	0.6	0.5
Optimization Algorithm	Adam	Adam	Adam	WOA
Learning Rate	4E-8	4E-8	3E-8	2E-8
Mini Batch Size	20	20	20	20
Initial Batch Size	128	128	64	32
Convergence Time (s)	48	24	28	33
TLP	11,160	4101	5791	7806

**Table 4.** The characteristics and ideal meta-parameters of all models developed in forecasting  $CCRSF_m$ .

Model	RMSE ( $m^3/s$ )	$R^2$	NSE	MBE ( $m^3/s$ )	MAE ( $m^3/s$ )
General single GRU model (1)	1.29	0.59	0.55	0.047	0.91
General single Bi-GRU model (2)	0.83	0.66	0.6	-0.06	0.53
2GRU $\times$ model (3)	0.49	0.74	0.71	-0.078	0.33
<b>Hybrid 2GRU<math>\times</math>-WOA model (4)</b>	<b>0.36</b>	<b>0.79</b>	<b>0.76</b>	<b>-0.01</b>	<b>0.21</b>

**Table 5.** The statistical results of models developed under the ideal meta-parameters in forecasting  $CCRSF_m$ .

\*Quantities that are shown in bold are the results of optimal model.

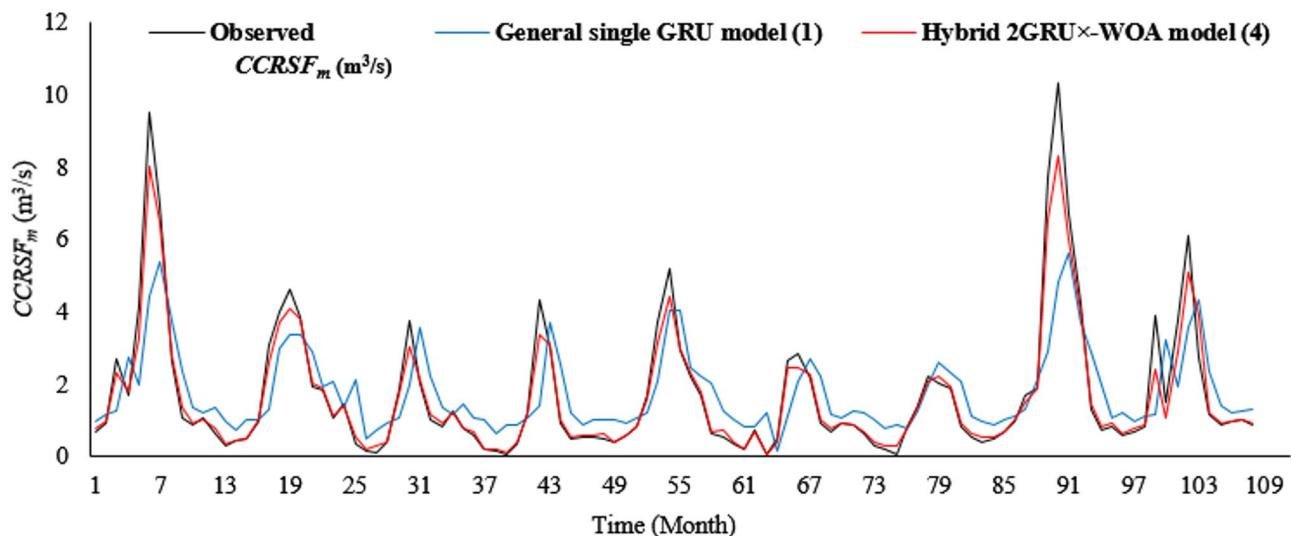


**Fig. 9.** RMSE variations in the model 4 over the range of NHN used under the optimal hyper-parameter in forecasting  $CCRSF_m$  in the testing phase.

performance. The model 4, thanks to the well-adjusted layer network structure and consequently *TLP* value, excelled over other models. It is owing to the well-proportioned *TLP* amount that it could hastily get the ideal weight sets – 1000 iterations in 33 s, while the model 1, for the uppermost *TLP* number, entails extra time to converge – 1000 iterations in 48 s.

Figure 10 matches pictographically the hydrograph plot for the measured and predicted temporal  $CCRSF_m$  by model 4 under the ideal meta-parameters in the validation phase. Along with this figure, model 4 owing to its inventive advanced layer's network structure can agreeably estimate the distribution of measured sequential  $CCRSF_m$  and fit relatively the vacillations trend mostly in the peak and deepest values of  $CCRSF_m$  that prove an acceptable unanimity.

In terms of distribution criteria, scatter diagrams for the measured and predicted time series  $CCRSF_m$  by the models used during the validation stage are presented in Fig. 11A–D. By a visual judgment, it is noticeable



**Fig. 10.** Hydrograph plot of the measured and predicted time series  $CCRSF_m$  by models 1 and 4 under the ideal meta-parameters during the validation phase (108 months between September 2011– August 2020).

that the predicted  $CCRSF_m$  by the hybrid 2GRU×-WOA model is generally near to the exact line (i.e., 1:1) for abundant data points with minor scattering compared to the GRU model. It validates a high steadiness and best-performing approach with a satisfactory  $R^2$  of 0.79.

A violin plot is presented to concomitantly match the performance and skill of all models and single out the best model used in the validation phase (Fig. 12). By a visual evaluation, it can be concluded that the hybrid 2GRU×-WOA model relatively better fits the distribution of the observed temporal  $CCRSF_m$  and could estimate to some extent exacter the peak and lowest values in comparison with the other models developed.

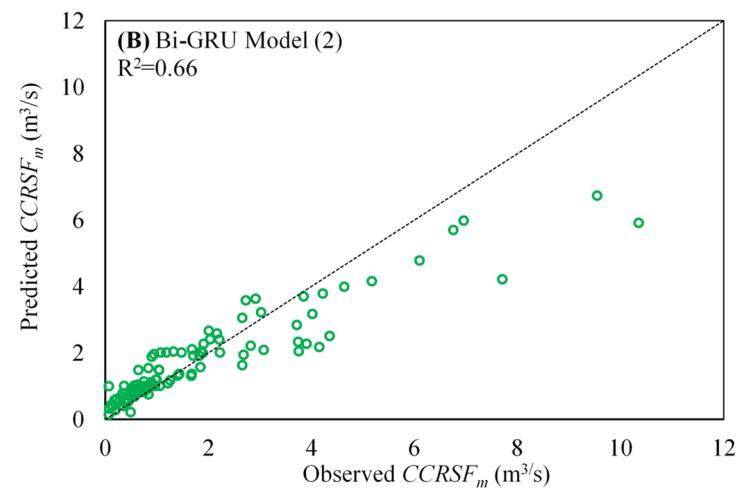
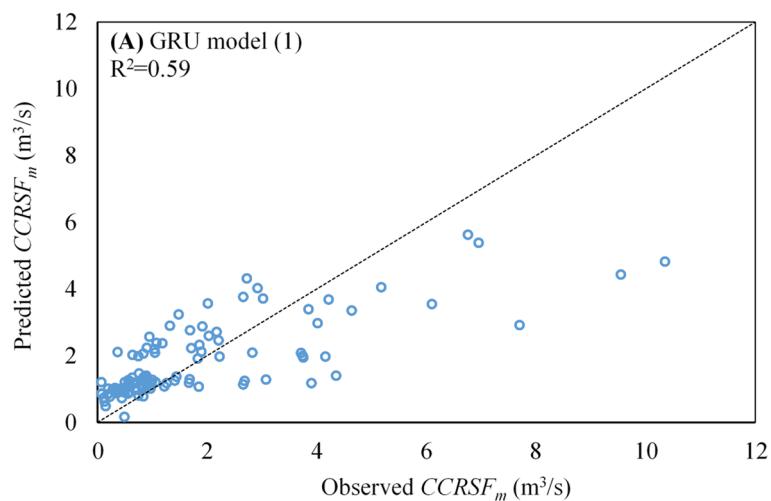
### Summary and conclusion

In this study, different layer structures of GRU-based deep learning framework were developed to estimate the forecasting  $CCRSF_m$  from Sep 1990 to Aug 2020 (360 months). In all models, to satisfy the long-period nonlinearity and non-stationary dilemmas, the seq2seq regression forecasting module is applied. The most worth mentioning outcomes of the modelling process are:

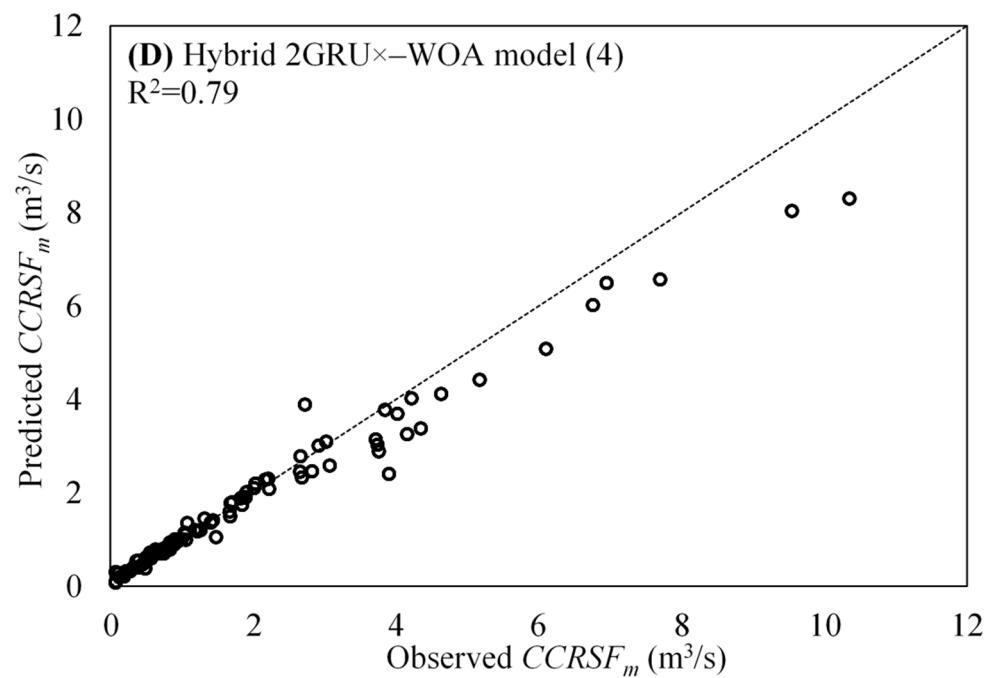
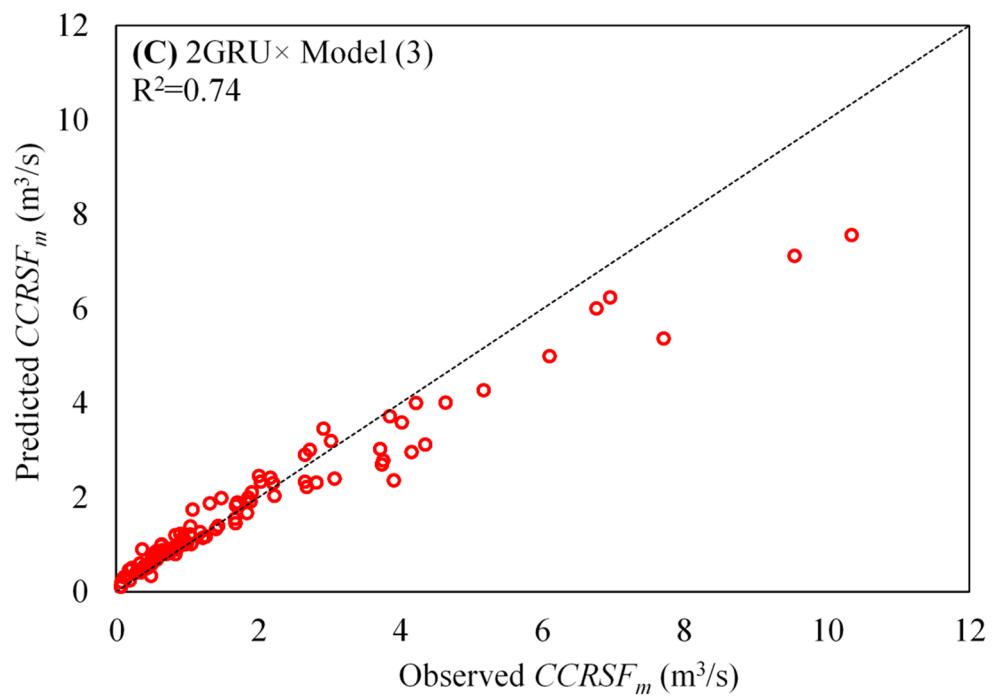
1. The PCC and CAS data analysis methods approved that the  $P_m$  was the most influential predictor variable on  $CCRSF_m$  to feed the models developed.
2. The training-stage forms of all models were more precise than their validation counterparts.
3. After several trials, the suggested hybrid 2GRU×-WOA model was accepted as the best-performing model by performance evaluation criteria to forecast  $CCRSF_m$ . The optimal  $P$ -rate,  $NHN$ , and  $SAF$  tuned for this model were obtained to be 0.5, 70, and *tanh-softsign*, respectively. Integrating the 2GRU× model with WOA algorithm caused to increase in the value of  $R^2$  by 6.8% and reduced in the value of  $RMSE$  by 20.4%.
4. By comparing the model structures developed and relevant  $TLP$  values, it can be concluded that inserting the Multiplication layer led to a more suitable layer network structure and well-adjusted  $TLP$ . So, for achieving effective DL-based models, an apt network structural,  $NHN$ , and well-balanced  $TLP$  value should be applied.
5. In all models, growing  $P$ -rate value lessens convergence time. The model 1 for the high  $TLP$  quantity, necessitated more time to train – 1000 repetitions in 48 s.

The hybrid 2GRU×-WOA structure is a cutting-edge method as verified by its commendable accuracy and performance (verified statistically). It can therefore be employed as an intelligent smart model for monitoring and predicting time series river streamflow under different climatic conditions. This hybrid model is an easy-to-implement, cost-effective, dependable, and time-saving process. Its well-formed layer network structure prompted an apt response to  $TLP$  and engendered more precision than the standard GRU and Bi-GRU as the benchmark models in the same meta-parameters. Despite the advantages of a hybrid model, it has some constraints: it entails an extremely long period of detailed (i.e., regular measurements) precipitation data to predict CCRSF<sub>m</sub> in a study area, as the seq2seq regression module of forecasting was employed.

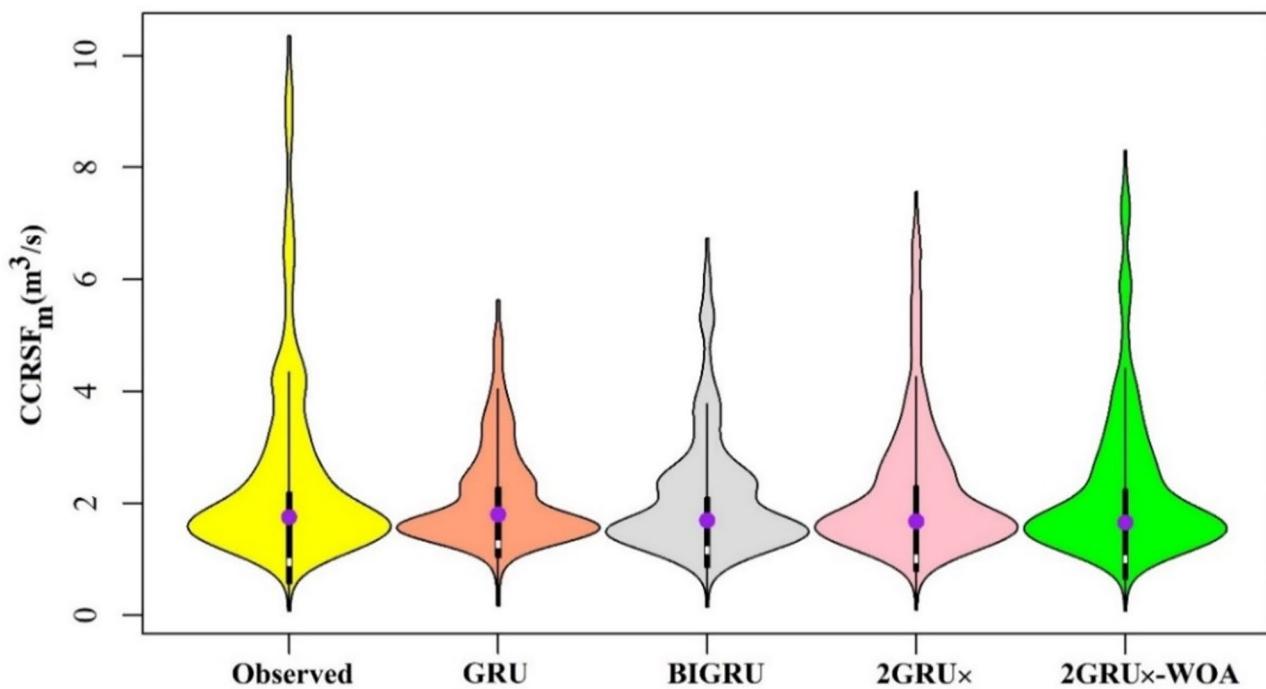
Even though this study evaluated the effects of different GRU-based model and WOA algorithm structures on forecasting  $CCRSF_m$ , upcoming studies could examine other methods. For instance, modeling could integrate DNN models with the most up-to-date optimization algorithms, such as the Puma Optimizer (PO) and Mountain-Gazelle Optimizer (MGO). The results should be equated to the outcomes of the current study to obtain the most effective technique.



**Fig. 11.** Scatter plot for the measured and predicted  $CCRSF_m$  ( $m^3/s$ ) by the (A) GRU model (1), (B) Bi-GRU model (2), (C) 2GRU $\times$  model (3), and (D) hybrid 2GRU $\times$ -WOA model (4) in the validation stage.



**Fig. 11.** (continued)



**Fig. 12.** Violin plot of the observed  $CCRSF_m$  against the predicted by the models developed under optimal meta-parameters in the validation phase.

### Data availability

Data used in this study are provided from the IMO (Iran Meteorological Organization) and will be available upon reasonable request.

Received: 16 February 2025; Accepted: 19 May 2025

Published online: 03 June 2025

### References

1. Papacharalampous, G., Koutsoyiannis, D. & Montanari, A. Advances in water resources quantification of predictive uncertainty in hydrological modelling by Harnessing the wisdom of the crowd: methodology development and investigation using toy models. *Adv. Water Resour.* **136**, 103471. <https://doi.org/10.1016/j.advwatres.2019.103471> (2020).
2. Moges, E., Demissie, Y., Larsen, L. & Yassin, F. Review: sources of hydrological model uncertainties and advances in their analysis. *Water* **13**, 1–23. <https://doi.org/10.3390/w13010028> (2021).
3. Horton, P., Schaeffer, B. & Kauzlaric, M. Why do we have so many different hydrological models? A review based on the case of Switzerland. *Wiley Interdiscip. Rev. Water* **9**, e1574. <https://doi.org/10.1002/wat2.1574> (2022).
4. Krysanova, V. et al. How the performance of hydrological models relates to credibility of projections under climate change. *Hydrolog. Sci. J.* **63**, 696–720. <https://doi.org/10.1080/02626667.2018.1446214> (2018).
5. Wang, S. et al. Extreme atmospheric rivers in a warming climate. *Nat. Commun.* **14**, 3219. <https://doi.org/10.1038/s41467-023-38980-x> (2023).
6. Masson-Delmotte, V. et al. Climate change 2021: the physical science basis. *Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*, 2(1), 2391 (2021). <https://doi.org/10.1017/9781009157896>, 2391 pp.
7. Burgan, H. I., Vaheddoost, B. & Aksay, H. Frequency analysis of monthly runoff in intermittent rivers. In *World Environmental and Water Resources Congress 2017* (pp. 327–334) (2017).
8. McDermott, T. K. J. Global exposure to flood risk and poverty. *Nat. Commun.* **13**, 3529. <https://doi.org/10.1038/s41467-022-30725-6> (2022).
9. Khorrami, B., Pirasteh, S., Ali, S., Sahin, O. G. & Vaheddoost, B. Statistical downscaling of GRACE TWSA estimates to a 1-km Spatial resolution for a local-scale surveillance of flooding potential. *J. Hydrol.* **624**, 129929 (2023).
10. Abbasi, M., Farokhnia, A., Bahreiniotagh, M. & Roozbahani, R. A hybrid of random forest and deep Auto-Encoder with support vector regression methods for accuracy improvement and uncertainty reduction of long-term streamflow prediction. *J. Hydrol.* **597**, 125717. <https://doi.org/10.1016/j.jhydrol.2020.125717> (2021).
11. Althoff, D., Rodrigues, L. N. & Bazame, C. H. Uncertainty quantification for hydrological models based on neural networks: the dropout ensemble. *Stoch. Environ. Res. Risk Assess.* **35**, 1051–1067. <https://doi.org/10.1007/s00477-021-01980-8> (2021).
12. Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E. & Rasmussen, J. An introduction to the European hydrological System—Système hydrologique Européen, SHE, 2: structure of a physically-based, distributed modelling system. *J. Hydrol.* **87** (1–2), 61–77. [https://doi.org/10.1016/0022-1694\(86\)90115-0](https://doi.org/10.1016/0022-1694(86)90115-0) (1986).
13. Anandharuban, P., La Rocca, M. & Elango, L. A box-model approach for reservoir operation during extreme rainfall events: a case study. *J. Earth Syst. Sci.* **128**, 229. <https://doi.org/10.1007/s12040-019-1258-7> (2019).
14. Silvestro, F., Ercolani, G., Gabellani, S., Giordano, P. & Falzacappa, M. Improving real-time operational streamflow simulations using discharge data to update state variables of a distributed hydrological model. *Hydrolog. Res.* **52**, 1239–1260. <https://doi.org/10.2166/nh.2021.162> (2021).
15. Fathian, F., Mehdizadeh, S., Sales, A. K. & Safari, M. J. S. Hybrid models to improve the monthly river flow prediction: integrating artificial intelligence and non-linear time series models. *J. Hydrol.* **575**, 1200–1213 (2019).

16. Vaheddoost, B., Safari, M. J. S. & Yilmaz, M. U. Rainfall-runoff simulation in ungauged tributary streams using drainage area ratio-based multivariate adaptive regression spline and random forest hybrid models. *Pure Appl. Geophys.* **180** (1), 365–382 (2023).
17. Rajib, A., Liu, Z., Merwade, V., Tavakoly, A. A. & Follum, M. L. Towards a large-scale locally relevant flood inundation modeling framework using SWAT and LISFLOODFP. *J. Hydrol.* **581**, 124406. <https://doi.org/10.1016/j.jhydrol.2019.124406> (2020).
18. Lin, H. et al. Time series-based groundwater level forecasting using gated recurrent unit deep neural networks. *Eng. Appl. Comput. Fluid Mech.* **16** (1), 1655–1672 (2022).
19. Ghasemlounia, R., Gharehbaghi, A., Ahmadi, F. & Albaji, M. Developing a novel hybrid model based on deep neural networks and discrete wavelet transform algorithm for prediction of daily air temperature. *Air Qual. Atmos. Health.* **17**, 2723–2737 (2024).
20. Ahmadi, F., Ghasemlounia, R. & Gharehbaghi, A. Machine learning approaches coupled with variational mode decomposition: a novel method for forecasting monthly reservoir inflows. *Earth Sci. Inf.* **17** (1), 745–760 (2024).
21. Gharehbaghi, A. & Ghasemlounia, R. Application of AI approaches to estimate discharge coefficient of novel kind of sharp-crested V-notch weirs. *J. Irrig. Drain. Eng.* **148** (3), 04022001 (2022).
22. Gharehbaghi, A. et al. A comparison of artificial intelligence approaches in predicting discharge coefficient of streamlined weirs. *J. Hydroinform.* **25** (4), 1513–1530 (2023).
23. Gharehbaghi, A., Ghasemlounia, R., Latif, S. D., Haghiahi, A. H. & Parsaei, A. Application of data-driven models to predict the dimensions of flow separation zone. *Environ. Sci. Pollut. Res.* **30** (24), 65572–65586 (2023).
24. Gharehbaghi, A., Ghasemlounia, R., Vaheddoost, B. & Ahmadi, F. Development of deep learning approaches for drought forecasting: a comparative study in a cold and semi-arid region. *Earth Sci. Inf.* **18** (1), 78 (2025).
25. Mehr, A. D., Kahya, E. & Yerdelen, C. Linear genetic programming application for successive-station monthly streamflow prediction. *Comput. Geosci.* **70**, 63–72 (2014).
26. Anusree, K. & Varghese, K. O. Streamflow prediction of Karuvannur river basin using ANFIS, ANN and MNLR models. *Proc. Technol.* **24**, 101–108 (2016).
27. Muhammad Adnan, R. et al. Application of soft computing models in streamflow forecasting. In *Proceedings of the Institution of Civil Engineers-Water Management*, vol. 172 123–134 (Thomas Telford Ltd, 2019).
28. Amisigo, B. A., Van de Giesen, N., Rogers, C., Andah, W. E. I. & Friesen, J. Monthly streamflow prediction in the Volta basin of West Africa: A SISO NARMAX polynomial modelling. *Phys. Chem. Earth (Pt AB C)* **33** (1–2), 141–150 (2008).
29. Yurekli, K., Kurunc, A. & Ozturk, F. Application of linear stochastic models to monthly flow data of Kelkit stream. *Ecol. Model.* **183** (1), 67–75 (2005).
30. Chen, Y. H. & Chang, F. J. Evolutionary artificial neural networks for hydrological systems forecasting. *J. Hydrol.* **367** (1–2), 125–137 (2009).
31. Wu, C. L. & Chau, K. W. Data-driven models for monthly streamflow time series prediction. *Eng. Appl. Artif. Intell.* **23** (8), 1350–1367 (2010).
32. Rasouli, K., Hsieh, W. W. & Cannon, A. J. Daily streamflow forecasting by machine learning methods with weather and climate inputs. *J. Hydrol.* **414**, 284–293 (2012).
33. Sun, A. Y., Wang, D. & Xu, X. Monthly streamflow forecasting using Gaussian process regression. *J. Hydrol.* **511**, 72–81 (2014).
34. Masselot, P., Dabo-Niang, S., Chebana, F. & Ouarda, T. B. Streamflow forecasting using functional regression. *J. Hydrol.* **538**, 754–766 (2016).
35. Erdal, H. I. & Karakurt, O. Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms. *J. Hydrol.* **477**, 119–128 (2013).
36. Lima, A. R., Cannon, A. J. & Hsieh, W. W. Forecasting daily streamflow using online sequential extreme learning machines. *J. Hydrol.* **537**, 431–443 (2016).
37. Atiquzzaman, M. & Kandasamy, J. Robustness of extreme learning machine in the prediction of hydrological flow series. *Comput. Geosci.* **120**, 105–114 (2018).
38. Feng, Z. K. et al. Parallel Cooperation search algorithm and artificial intelligence method for streamflow time series forecasting. *J. Hydrol.* **606**, 127434 (2022).
39. Hah, D., Quilty, J. M. & Sikorska-Senoner, A. E. Ensemble and stochastic conceptual data-driven approaches for improving streamflow simulations: exploring different hydrological and data-driven models and a diagnostic tool. *Environ. Model. Softw.* **157**, 105474 (2022).
40. Sa'adi, Z. et al. Comparative assessment of empirical random forest family's model in simulating future streamflow in different basin of Sarawak, Malaysia. *J. Atmos. Sol-Terr Phys.* **265**, 106381 (2024).
41. Moknatian, M. & Mukundan, R. Uncertainty analysis of streamflow simulations using multiple objective functions and bayesian model averaging. *J. Hydrol.* **617**, 128961 (2023).
42. Gizaw, M. S. & Gan, T. Y. Regional flood frequency analysis using support vector regression under historical and future climate. *J. Hydrol.* **538**, 387–398 (2016).
43. Tongal, H. & Booij, M. J. Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. *J. Hydrol.* **564**, 266–282 (2018).
44. Achieng, K. O. Averaging multi-climate model prediction of streamflow in the machine learning paradigm. In *Advances in Streamflow Forecasting* 239–262 (Elsevier, 2021).
45. Soo, E. Z. X. et al. Streamflow simulation and forecasting using remote sensing and machine learning techniques. *Ain Shams Eng. J.* **15** (12), 103099 (2024).
46. Cho, K. et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. ArXiv 1406.1078 (2014). <https://doi.org/10.48550/ArXiv.1406.1078>
47. Karamvand, A., Hosseini, S. A. & Azizi, S. A. Enhancing streamflow simulations with gated recurrent units deep learning models in the flood prone region with low-convergence streamflow data. *Phys. Chem. Earth (Pt AB C)* **136**, 103737 (2024).
48. Tao, L., Nan, Y., Cui, Z., Wang, L. & Yang, D. An explainable bayesian gated recurrent unit model for multi-step streamflow forecasting. *J. Hydrol. Reg. Stud.* **57**, 102141 (2025).
49. He, M. et al. Streamflow prediction in ungauged catchments through use of catchment classification and deep learning. *J. Hydrol.* **639**, 131638 (2024).
50. Hou, S., Wei, J., Hou, M., Xu, J. & Han, L. A hydrological knowledge-informed LSTM model for monthly streamflow reconstruction using distributed data: application to typical rivers across the Tibetan plateau. *J. Hydrol.* **649**, 132409 (2024).
51. Zhou, Y., Duan, Y., Yao, H., Li, X. & Li, S. Incorporating hydrological constraints with deep learning for streamflow prediction. *Expert Syst. Appl.* **259**, 125379 (2025).
52. Safari, M. J. S., Arashloo, S. R. & Mehr, A. D. Rainfall-runoff modeling through regression in the reproducing kernel hilbert space algorithm. *J. Hydrol.* **587**, 125014 (2020).
53. Mehr, A. D. & Safari, M. J. S. Genetic programming for streamflow forecasting: a concise review of univariate models with a case study. In *Advances in Streamflow Forecasting* 193–214 (Elsevier, 2021).
54. Nazeri Tahroudi, M., Mirabbasi, R., Ramezani, Y. & Ahmadi, F. Probabilistic assessment of monthly river discharge using copula and OSVR approaches. *Water Resour. Manag.* **36** (6), 2027–2043 (2022).
55. Ahmadi, F., Tohidi, M. & Sadrianzade, M. Streamflow prediction using a hybrid methodology based on variational mode decomposition (VMD) and machine learning approaches. *Appl. Water Sci.* **13** (6), 135 (2023).
56. Eslamitabar, V., Ahmadi, F., Sharafati, A. & Rezaverdinejad, V. River flow simulation based on empirical mode function signals and random forest algorithm. *Acta Geophys.* **73**, 1801–1817 (2025).

57. Eslamitabar, V., Ahmadi, F., Sharafati, A. & Rezaverdinejad, V. Bivariate simulation of river flow using hybrid intelligent models in sub-basins of lake urmia, Iran. *Acta Geophys.* **71**, 873–892 (2023).
58. Floudas, C. A. & Gounaris, C. E. A review of recent advances in global optimization. *J. Glob Optim.* **45**, 3–38 (2009).
59. Mirjalili, S. & Lewis, A. The Whale optimization algorithm. *Adv. Eng. Softw.* **95**, 51–67 (2016).
60. Jia, F., Zhu, Z. & Dai, W. Short-term forecasting of streamflow by integrating machine learning methods combined with metaheuristic algorithms. *Expert Syst. Appl.* **245**, 123076 (2024).
61. Nguyen, D. H., Le, X. H., Anh, D. T., Kim, S. H. & Bae, D. H. Hourly streamflow forecasting using a bayesian additive regression tree model hybridized with a genetic algorithm. *J. Hydrol.* **606**, 127445 (2022).
62. Zhao, X. et al. Enhancing robustness of monthly streamflow forecasting model using gated recurrent unit based on improved grey Wolf optimizer. *J. Hydrol.* **601**, 126607 (2021).
63. Adnan, R. M. et al. Improving streamflow prediction using a new hybrid ELM model combined with hybrid particle swarm optimization and grey Wolf optimization. *Knowl. -Based Syst.* **230**, 107379 (2021).
64. Adnan, R. M. et al. Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. *J. Hydrol.* **586**, 124371 (2020).
65. Luo, X. et al. J. A hybrid support vector regression framework for streamflow forecast. *J. Hydrol.* **568**, 184–193 (2019).
66. He, Y., Yan, Y., Wang, X. & Wang, C. Uncertainty forecasting for streamflow based on support vector regression method with fuzzy information granulation. *Energy Procedia.* **158**, 6189–6194 (2019).
67. Granata, F., Di Nunno, F. & Pham, Q. B. A novel additive regression model for streamflow forecasting in German rivers. *Results Eng.* **22**, 102104 (2024).
68. Xu, L. et al. Investigating the potential of EMA-embedded feature selection method for ESVR and LSTM to enhance the robustness of monthly streamflow forecasting from local meteorological information. *J. Hydrol.* **636**, 131230 (2024).
69. Heddam, S. et al. Hybrid river stage forecasting based on machine learning with empirical mode decomposition. *Appl. Water Sci.* **14** (3), 46 (2024).
70. Parsaei, A. et al. G. Novel hybrid intelligence predictive model based on successive variational mode decomposition algorithm for monthly runoff series. *J. Hydrol.* **634**, 131041 (2024).
71. Ahmed, A. N., Van Lam, T., Hung, N. D., Van Thieu, N. & Kisi, O. El-Shafie, A. A comprehensive comparison of recent developed meta-heuristic algorithms for streamflow time series forecasting problem. *Appl. Soft Comput.* **105**, 107282 (2021).
72. Shukla, R. et al. Modeling of stage-discharge using back propagation ANN-, ANFIS-, and WANN-based computing techniques. *Theor. Appl. Climatol.* **147**, 867–889 (2022).
73. Cui, Z. et al. Real-time rainfall-runoff prediction using light gradient boosting machine coupled with singular spectrum analysis. *J. Hydrol.* **603**, 127124 (2021).
74. Xie, T. et al. Hybrid forecasting model for non-stationary daily runoff series: a case study in the Han river basin, China. *J. Hydrol.* **577**, 123915 (2019).
75. Yaseen, Z. M. et al. Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model. *J. Hydrol.* **554**, 263–276 (2017).
76. Sobhani, A., Asgari, H. R., Noura, N., Ownegh, M. & Sakieh, Y. Application of land-use management scenarios to mitigate desertification risk in Northern Iran. *Environ. Earth Sci.* **76**, 1–15 (2017).
77. Shakib, H. S. & Shojarastegari, H. Climate change impacts and water resources management: Chehel-Chai basin, Iran. *Appl. Ecol. Environ. Res.* **15** (4), 741–754 (2017).
78. Momeni, E., Nazir, R., Armaghani, D. J. & Maizir, H. Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN. *Meas* **57**, 122–131 (2014).
79. Lawrence, S., Back, A. D., Tsoi, A. C. & Giles, C. L. On the distribution of performance from multiple neural network trials. *IEEE Trans. Neural Netw.* **8** (6), 1507–1517 (1997).
80. Graves, A. & Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* **18** (5–6), 602–610 (2005).
81. Hochreiter, S. & Schmidhuber, J. LSTM can solve hard long time lag problems. In *Proceedings of 1996 Neural Information Processing Systems* (pp. 473–479). Denver, United States: MIT Press (1997).
82. Bengio, Y. Learning deep architectures for al. *Found. Trends Mach. Learn.* **2** (1), 1–127 (2007).
83. Danesh, M., Gharehbaghi, A., Mehdizadeh, S. & Danesh, A. A comparative assessment of machine learning and deep learning models for the daily river streamflow forecasting. *Water Resour. Manag.* **39**, 1911–1930 (2025).
84. Gao, S. et al. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *J. Hydrol.* **589**, 125188 (2020).
85. Ghasemlounia, R., Gharehbaghi, A., Ahmadi, F. & Saadatnejadgharahassanlou, H. Developing a novel framework for forecasting groundwater level fluctuations using Bi-directional long Short-Term memory (BiLSTM) deep neural network. *Comput. Electron. Agric.* **191**, 106568 (2021).
86. Shen, C. A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resour. Res.* **54** (11), 8558–8593 (2018).
87. Adnan, R. M. et al. Harnessing deep learning and snow cover data for enhanced runoff prediction in snow-Dominated watersheds. *Atmos* **15** (12), 1407 (2024).
88. Jamei, M. & Et Quantitative improvement of streamflow forecasting accuracy in the Atlantic zones of Canada based on hydro-meteorological signals: A multi-level advanced intelligent expert framework. *Ecol. Inf.* **80**, 102455 (2024).
89. Watkins, W. A. & Schevill, W. E. Aerial observation of feeding behavior in four Baleen whales: *Eubalaena glacialis*, *Balaenoptera borealis*, *Megaptera Novaeangliae*, and *Balaenoptera physalus*. *J. Mammal.* **60** (1), 155–163 (1979).
90. Goldbogen, J. A., Friedlaender, A. S., Calambokidis, J., McKenna, M. F. & Simon, M. Nowacek, D. P. Integrative approaches to the study of Baleen Whale diving behavior, feeding performance, and foraging ecology. *BioScience* **63** (2), 90–100 (2013).
91. Zhang, J., Zhu, Y., Zhang, X., Ye, M. & Yang, J. Developing a long Short-Term memory (LSTM) based model for predicting water table depth in agricultural areas. *J. Hydrol.* **561**, 918–929 (2018).
92. Gharehbaghi, A., Ghasemlounia, R., Ahmadi, F. & Albaji, M. Groundwater level prediction with meteorologically sensitive gated recurrent unit (GRU) neural networks. *J. Hydrol.* **612**, 128262 (2022).

## Author contributions

All authors contributed to the study's conception and design. Farshad Ahmadi: writing–review, investigation & editing. Redvan Ghasemlounia: Conceptualization, Methodology, writing–original draft, Writing–review & editing. Amin Gharehbaghi: Methodology, Software, Writing – original draft preparation, Rasoul Mirabbassi: Supervision, Methodology, and Writing–review & editing. Ali Torabi Haghghi: Supervise, methodology, and writing–review & editing.

## Funding

This research received no external funding.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to A.T.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025, corrected publication 2025