Applied Water Science (2025) 15:174
https://doi.org/10.1007/s13201-025-02535-w

ORIGINAL ARTICLE

Developing a novel layer network structure for a LSTM model to predict mean monthly river streamflow

Amin Gharehbaghi[1] · Redvan Ghasemlounia[2] · Shahaboddin Daneshvar[3] · Farshad Ahmadi[4]

Received: 30 January 2025 / Accepted: 1 June 2025 / Published online: 25 June 2025

Abstract

In this research, novel innovative DDN layer network structures by hybridizing double-LSTM model with an addition layer (+) (*i.e.,* 2LSTM and 2LSTM + layer network models) are developed purposefully to enhance prediction performance of the mean monthly Maroon River streamflow ($MRSF_m$) in Iran from October 1987 to September 2017. For doing so, to select the most effective parameters on $MRSF_m$, the Pearson's correlation coefficient (PCC) and Cosine amplitude sensitivity (CAS) as features selection process are carried out for potential meteorological variables in the study area (*i.e.,* average monthly temperature (T_m), evaporation (ET_m), and precipitation (P_m)) and target ($MRSF_m$). The results show that T_m and ET_m have an insignificant influence on $MRSF_m$, thus, only P_m is used as the most effective input variable in predicting $MRSF_m$. Due to a well-balanced network model's structural outline in the suggested novel hybrid 2LSTM + model, it accordingly yields to a suitable total learnable parameter (*TLP*) compared to ordinary standalone LSTM and GRU as the benchmark models developed in the similar meta-parameters. This model under the optimal meant meta-parameters tuned *i.e.,* state activation functions (*SAF*) = *tanh-softsign*, numbers of hidden neurons (*NHN*) = 75, dropout rate (*P-rate*) = 0.5, performs best among the models with an R^2 of 0.68, *NSE* of 0.63, *PBIAS* of 41%, *KGE* of 0.79, and *RMSE* of 19.24 m^3/s. Comparatively, a standard gated recurrent units (GRU) and LSTM as benchmark models using the optimal scenario generate the following results: R^2 are 0.57 and 0.67, *NSE* are 0.53 and 0.61, *PBIAS* are 109 and 49%, *KGE* are 0.63 and 0.79, and *RMSE* are 21.11 and 19.32 m^3/s, respectively. Generally, in all models, in the equal *NHN*, rising *P-rate* value reduces convergence time.

Keywords LSTM model · GRU model · Novel hybrid 2LSTM + layer network model · TLP parameter · Maroon River streamflow

List of symbols

MLMs	Machine learning models
ANNs	Artificial neural networks
DNNs	Deep neural networks
RNN	Recurrent neural network
LSTM	Long short-term memory
GRU	Gated recurrent units
$MRSF_m$	Maroon river streamflow (m^3/s)
f	Nonlinear function
PCC	Pearson's correlation coefficient
P-rate	Learning dropout rate
SAF	State activation functions
MSO	Model structural outlines
RMSE	Root mean square error (m^3/s)
STDV	Standard deviation
R^2	Determination coefficient [-]
CV	Coefficient of variation
TLP	Total learnable parameters
N	Number of datasets
T_m	Average monthly temperature
ET_m	Average monthly evaporation
P_m	Average monthly Precipitation
σ_x	Standard deviation of observed $MRSF_m$ at time i
σ_y	Standard deviation of estimated $MRSF_m$ at time i
y_i	Predicted $MRSF_m$ at time i
x_i	Observed $MRSF_m$ at time i
μ_x	Average of observed $MRSF_m$ at time i
μ_y	Average of estimated $MRSF_m$ at time i
NHN	Number of hidden neurons
HMSE	Half mean squared error
PBIAS	Percent bias

Extended author information available on the last page of the article

Introduction

Background and literature review

Accurately predicting a river's streamflow is noteworthy for consistent management and sustainable usage of water resources. Hydrological prediction requires the assessment of the hydrological conditions of a region. Precise prediction of streamflow rates in both long and short timeframes can support decision making around water engineering problems like the design of flood prevention structures, provision of municipal and agricultural water resources, operation of hydroelectric facilities, and river basin management (Ghorbani et al. 2016; Band et al. 2020). The task requires realizing a complex and intricate process influenced by extremely stochastic, non-stationary behaviors, and nonlinear interactions of hydro-climatic parameters like temperature, evaporation, and precipitation with river flows (Bayazit 2015). Robust methods are needed to achieve this objective.

Several forecasting methods have recently been developed to estimate the time-series variables in complex problems. They can be categorized as machine learning (ML) and physical-based models. Physical models necessitate mathematical equations, relevant physical data, and extensive experience and knowledge (Gharehbaghi et al., 2016; Fang et al. 2019).

ML modes are generally more instinctive and do not entail characterized physical progress (Chen et al. 2022). To date, numerous conventional standalone ML models have been adopted to forecast streamflow in diverse regions (e.g., Fathian et al. (2019), Mohammadi et al. (2020), and Huang et al. (2021a, b)). Since streamflow is not static and long-term historical records reflect the variability of rivers, these models do not perform well due to overfitting and a lack of ability to reflect temporal relationships (Fang et al. 2008). In contrast, traditional standalone data-driven and spatial–temporal models revolve around approximation deterministic and stochastic relationships, respectively (Mohammadi et al. 2020).

Recent tests of deep-learning (DL) procedures have been inspired by artificial neural network (ANN) models that imitate the knowledge realization and structural talents of the human mind by using data processing and disseminated-communiqué-nodes analysis systems (Band et al. 2020). The term "deep" in DNN models refers to the number of layers in the networks. DNNs are highly regarded as they could assess elaborate nonlinear relationships better than conventional neural networks. Too, these models reduce the number of variables used through iterative processes (Ghasemlounia et al. 2021). Investigations have verified that DNNs perform with greater precision and they show incredible potential as innovative and resourceful ML models (Chen et al. 2020).

Among DNNs, LSTM neural networks perform very well due to their capacity to adapt to sequential real-world problems in many domains (Sherstinsky 2020). Their cutting-edge chain-form structure with internal iterative self-looped cells provides a high capacity to recall patterns over time for analysis (Hochreiter and Schmidhuber 1997). Because of their evolved and more dynamic interaction through distributed attribute representations, LSTM can identify intricate nonlinear relationships in a time series of input-target and can successfully address the gradient problem (An et al. 2020). As their performances depend on the values of some meta-parameters, careful selection of the appropriate architectural structure of the DNNs is vital (Gharehbaghi et al. 2022).

Currently, LSTM has been used to estimate the streamflow of different rivers (Yuan et al. (2018), Le et al. (2019), An et al. (2020), Lin et al. (2021), and Ahmed et al. (2021)). Bai et al. (2021) compared the strength of an LSTM network for runoff forecasting under fluctuating climate conditions with the strength of two hydrologic models applied to 278 model parameter estimation experiment (MOPEX) basins. They claimed that if adequate calibration data existed, an LSTM would be more useful than hydrologic models. Gao et al. (2020) predicted interim runoff with LSTM and GRU networks unaffected by time-stage optimization under sample generation in South-east China. The consequences demonstrated that GRU model executes as well as LSTM model. Xiang and Demir (2020) estimated long-term hourly streamflow by using a neural runoff model (NRM) to forecast 120 h into the future. Their model outperformed ridge regression and random forest regression on several measures. Abbas et al. (2020) simultaneously estimated sub-surface and surface streams in a sub-basin at a 6-min time step using DNNs. The results confirmed that the surface runoff estimation of a simple LSTM exceeded those of the other models tested. Yang et al. (2023) developed the LSTM-prototypical network fusion model on the basis of few-shot learning to predict monthly runoff in the source region of the Yellow River basin (SRYRB) and the Lancang River basin (LRB), China. They stated that the proposed model outdone in respect of other MLMs used including SimpleRNN, GRU, LSTM, BiLSTM, ANN, SVR, RF, and ARMA models. The suggested model in the LRB resulted in an *NSE* of 0.8 and 0.83 as the ratio of calibration data (*K*) was 45 and 20%, enhanced by 0.22 and 0.52 corresponding to the average *NSE* of other approaches used, respectively. Also, in the SRYRB, as *K* was 40%, *NSE* got 0.83 which increased by 0.35. A summary of the different DNN-based models used for forecasting streamflow in diverse regions is presented in Table 1.

Table 1 DNN-based models for forecasting streamflow over different time scales

Location	[a]Models Used (**Best**)	Modeling Results	Reference
Eight stations in the Tibetan Plateau	Hydro-LSTM	*NSE*, *CC*, and *KGE* range between 0.71–0.96, 0.84–0.98, and 0.78–0.96, respectively	Hou et al. (2025)
Six different basins in the United States	**VMDI-LSTM-ED,** LSTM-ED, VMD-LSTM	Average *NSE* of 0.88 for 1-day ahead forecasts over the six basins	Liu et al. (2025)
Qilishan station, Yangtze River Basin of China	EB-GRU, TSF, MLP, SVM	1-day lead time $NSE = 0.96$, $RRMSE = 0.11$	Tao et al. (2025)
Heihe River basin, China	DA-BiGRU-RED	For1-7 day ahead forecasting NSE values exceeding 0.69 in nearly all prediction scenarios	Huang et al. (2024)
Klang River Basin, Malaysia	KNN, SVM, **RF**, ANN, LSTM	*SMAPE* and *MdAPE* values of 0.36 and 0.37, respectively	Soo et al. (2024)
Wei River basin in China	Hybrid BPX with DL models (RNN, GRU, LSTM and **TCN**)	$KGE = 0.91$	Tang et al. (2024)
CAMELS dataset	**EFS-KESVREMA**, EFS-ESVREMA and EFS-LSTMEMA	$NSE = 0.73–0.94$	Xu et al. (2024)
LSJ catchment in the province of Quebec, Canada	LSTM	$MAE = 25\ m^3/s$ (for the first day of lead time)	Sabzipour et al. (2023)
Soyangho Lake in South Korea	**WRF-Hydro-LSTM,** WRF-Hydro-only, and LSTM-only	$NSE = 0.95$ and $R \geq 0.96$	Cho and Kim, (2022)
Maozhou River (MZR), China	ANFIS, ANN, **LSTM**, SOBEK	$R^2 = 0.97$, $RMSE = 7.72\ (m^3/s)$, $NSE = 0.97$	Huang et al. (2021a, b)

[a]*KNN* K-Nearest Neighbors, *SVM* Support Vector Machines, *RF* Random Forests, *ANN* Artificial Neural Network, *SMAPE* symmetric mean absolute percentage error, *MdAPE* median absolute percentage error, *DA* dual attention, *BiGRU* Bidirectional Gated Recurrent Unit, *RED* recursive encoder-decoder, *WRF-Hydro* weather research and forecasting hydrological modeling system, *TCN* temporal convolutional network, *CAMELS* Catchment Attributes and Meteorology for Large-sample Studies, *ESVR* Extreme support vector regression, *EMA* Evolutionary mating algorithm, *VMDI* improved decomposition ensemble model based on DI (data integration), *ED* Encoder-Decoder, *EFS* embedded feature selection method

Motivation for this study

In this study, a novel and all-inclusive analytical structure of LSTM, ML-based, termed 2LSTM and 2LSTM + models is developed to overcome the limits and difficulties of forecasting mean monthly Maroon River streamflow ($MRSF_m$). In these models, the sequence-to-sequence regression forecasting module is applied to cope with long-period nonlinear and non-stationary procedures in $MRSF_m$. Moreover, the standard structure of GRU and LSTM neural networks as the benchmark time-series models is developed for comparison.

There have been few studies that use optimum design architectures of LSTM algorithms for time-series forecasting of streamflow. This research is novel in that it introduces a groundbreaking optimized layer network structure into an LSTM-based termed 2LSTM and 2LSTM + layer network model to estimate streamflow rate, a complicated hydrological process. To the best of our knowledge, among the present computational intelligence-system literature focusing on streamflow prediction, only simple and commonplace DNN architectural structures have been emphasized. No studies have probed the performance of novel introduced evolutionary neural architecture models.

Research objectives

The fundamental contributions of present investigation are that it develops the different layer structures of an LSTM-based *i.e.,* 2LSTM and 2LSTM + layer network models, deep neuro-evolution time-series paradigm to accurately predict $MRSF_m$ using observation data describing hydro-meteorological parameters; it adopts an algorithm-tuning process to determine the optimal spectrum of applied meta-parameters and the optimal architecture of resulting models for well-configuration of designed models to aid learning and lower overfitting issue; and it appraises and compares the experimental results of LSTM-based structures with the standard GRU time-series to differentiate the attributes of the optimum method through statistical assessments.

Study area and data description

The Maroon River stems from Nil and Sadat ranges of the Zagros Mountains and flows across the provinces of Kohgilueye and Behbahan. The Maroon River watershed (Fig. 1) with a semi-arid and cold climate covers an area of

Fig. 1 Location map of the study region

2747 km^2 in southwestern Iran. The Maroon Dam is located 120 km downstream, after which the river flows through Takab Canyon and irrigates Jayezan and Behbahan plains. The river may reach the Persian Gulf during periods of flood. The only hydrometeorological station in the watershed is the Idanak Station at 560 m above sea level upstream of the dam (Fig. 2). The average monthly streamflow of the Maroon River between October 1987 and September 2017 was 48.59 m^3/s. The minimum and maximum recorded monthly flows occurred in November (0.68 m^3/s) and March (377.17 m^3/s).

In this research, a temporal dataset of 360 monthly hydrometeorological observations covering the period from October 1987 to September 2017 is used to forecast mean

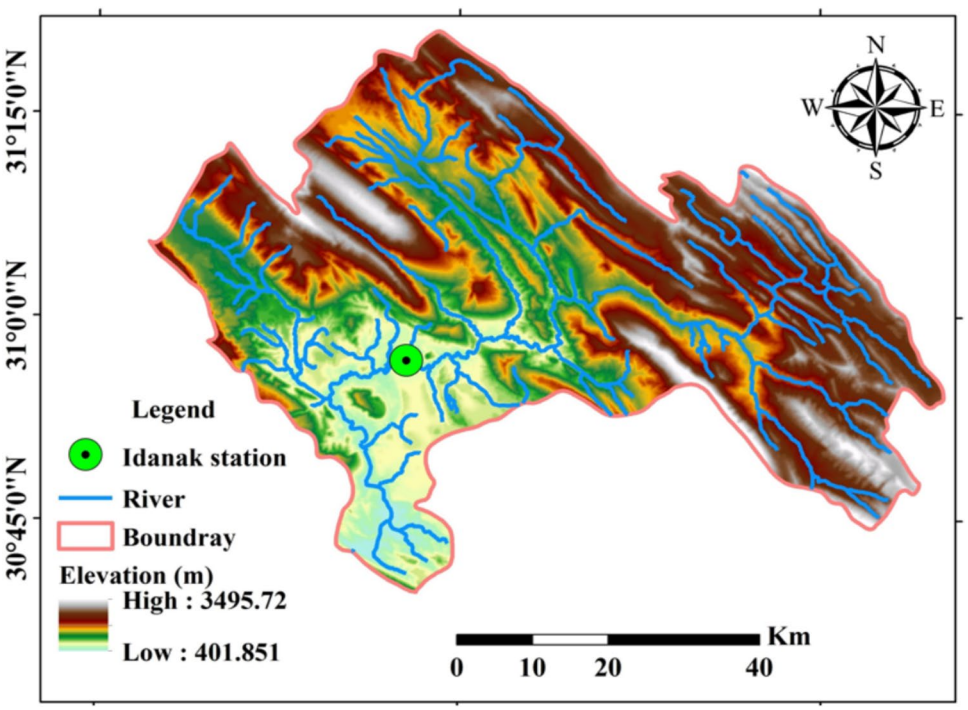

Fig. 2 Maroon River basin and Idanak Station

monthly streamflow in the Maroon River ($MRSF_m$). In this context, first of all, measured datasets are normalized to unit variance and zero mean as advised by Lawrence et al. (1997). Then, they are separated into two subgroups. One limited 70% of the data (252 monthly measured) are utilized to calibrate the estimating models. The lasting 30% (108 samples) are set aside to be used for validation. This process ensured that the data are on a uniform scale, so discrete variable sensitivity did not confuse the results.

The datasets are comprised of four variables, mean monthly evapotranspiration (ET_m), temperature (T_m), precipitation (P_m), and $MRSF_m$ recorded by the Idanak station collected from the Iran Meteorological Organization (IMO). Descriptive statistics indices of the variables are provided in Table 2. The range of $MRSF_m$ for the entire period is 376.49 m^3/s.

Features-selection process

Since the performance of any simulation mainly depends on the appropriate choice of forecasters for precise prediction, unfit choices could negatively affect the effectiveness of any approach. So, in this section, available large-dimensional potential atmospheric hydro-meteorological datasets recorded in studied region are assessed to identify the most effective parameters on predicting $MRSF_m$ as the model predictor variables. To do so, these variables of greatest 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 analysis for the variables employed in Table 2 is performed by changing each predictor variable at a constant ratio and holding the other predictor variables constant (Momeni et al. 2014) as follows:

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

where I_i and O_j are predictor and target parameters, respectively, and n is number of datasets. The R_{ij} value [0, 1] determines the strength of the relationship amid the predictor and target parameters in Table 2. The value of PCC and CAS methods is given in Table 3.

In relation to Table 3, because of the insignificant amount of PCC and CAS data analyses methods of ET_m and T_m, their effects on modeling $MRSF_m$ by proposed models are ignored. So, only P_m is proved to be an important input parameter. Finally, the equation for the prediction of $MRSF_m$ can be expressed as follows:

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

The data from the entire dataset are graphed to reveal the seasonality of streamflow (Fig. 3A, B). 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. 1, only precise and robust approaches can enable to forecast $MRSF_m$. As aforesaid, four LSTM-based network

Table 3 PCC and CAS values versus $MRSF_m$

Analysis Method	Target	T_m	P_m	ET_m
PCC	$MRSF_m$	−0.46	0.62	−0.42
CAS	$MRSF_m$	−0.24	0.71	−0.21

Table 2 Statistical indices of variables operated in the study stations

Variable	Data period	Min	Max	Mean	STDV	Skewness	CV
ET_m (mm)	Whole	24.2	672.3	231.67	164.63	0.441	0.71
	Training	24.2	672.3	240.4	173.99	0.4	0.72
	Testing	36.7	465.6	211.31	139	0.376	0.657
T_m (°C)	Whole	9.1	36.8	23.21	8.68	0.005	0.374
	Training	9.1	36.8	23.23	8.73	0.011	0.376
	Testing	9.9	36.8	23.17	8.61	−0.009	0.371
P_m (mm)	Whole	0	522.2	49.29	78.21	2.36	1.58
	Training	0	522.2	52.74	84.71	2.34	1.6
	Testing	0	263	41.25	59.96	1.72	1.45
$MRSF_m$ (m^3/s)	Whole	0.68	377.17	48.79	57.29	2.33	1.17
	Training	0.68	377.17	55.48	64.26	2.04	1.15
	Testing	4.24	194.07	33.19	31.22	1.97	0.94

*CV and $STDV$ illustrate the coefficient of variation and standard deviation, respectively

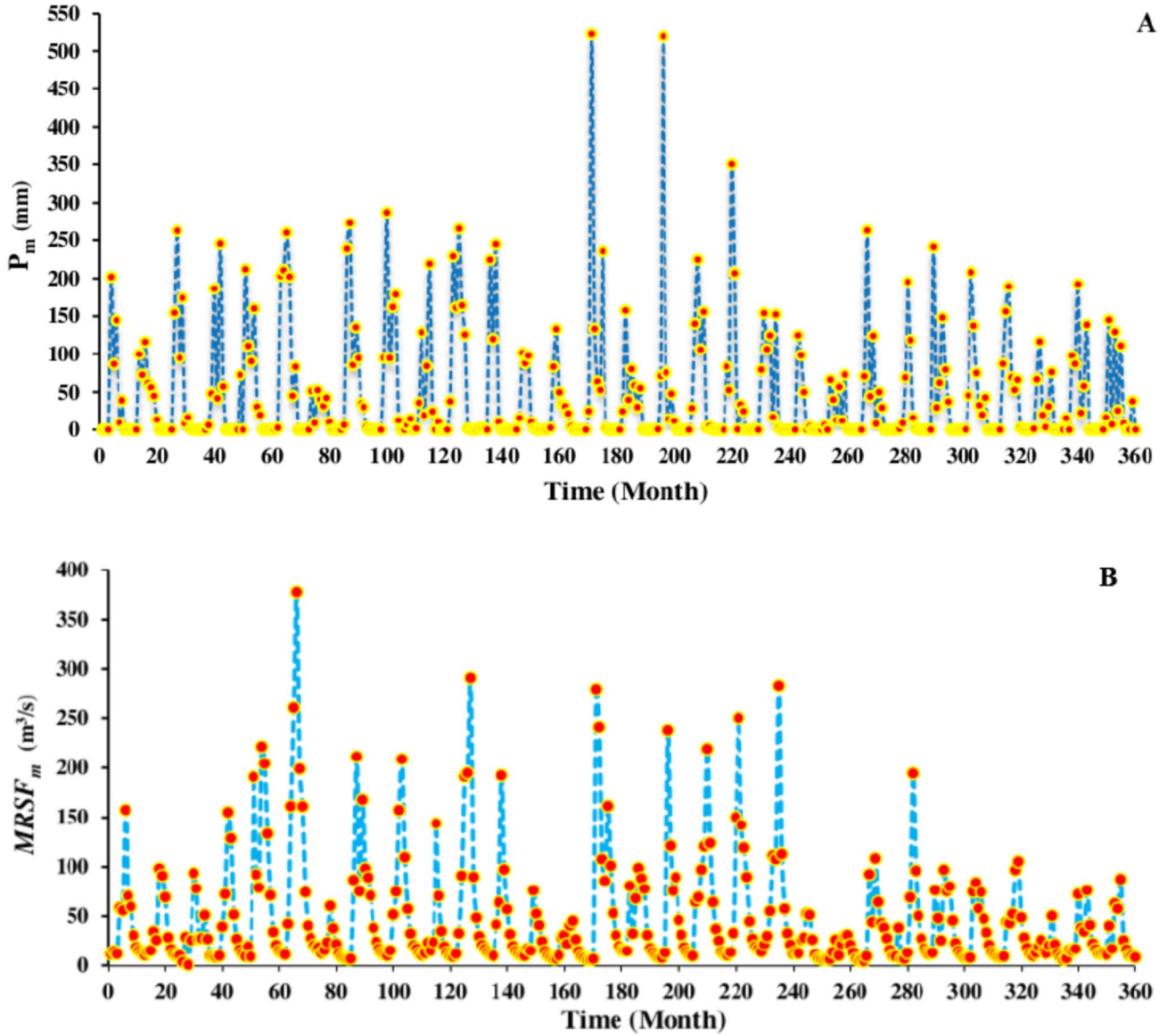

Fig. 3 Time-series graphs of variables used in Eq. 2 between October 1987 and September 2017 (360 months) in the Maroon River watershed: **A** Precipitation and **B** $MRSF_m$

structures with different layers are designed for their capacities to estimate $MRSF_m$.

LSTM and GRU neural networks

RNN is an NN that assimilates data from the past to estimate conditions of the future by enforcing data with definite dependencies. A disadvantage of conventional RNN in modeling river streamflow is its restricted memory volume, typically fewer than ten time steps (Bengio et al. 1994). Consequently, it causes gradients to disappear in the back-propagation of learning over long periods of study (Graves 2013). In other words, conventional RNN can only use data from the last 10 months to estimate streamflow for the next month. However, 10 months is much too short for streamflow prediction and responses to rainfall events are inadequate for longer periods (Millares et al. 2009). LSTM was designed by Hochreiter and Schmidhuber (1997). The model improved the learning ability of RNNs by hidden units added to the memory cell.

DNNs contain coupled time-based memory blocks to empower the use of greater amounts of data covering longer periods (Hochreiter and Schmidhuber 1997). Technically, a memory block contains four components: a constant-error carousel (CEC) cell combined with three specific multiplicative elements (gates) that prevent the gradient from

vanishing during back-propagation while training the LSTM (Hochreiter and Schmidhuber 1997; Kratzert et al. 2019). The internal mechanism of an LSTM in each memory block at time t is presented in Fig. 4. In a block, x_t is the input data in time t, the signs of "×" and "+" symbolize point-wise multiplication and addition, respectively, and *tanh* is a hyperbolic tangent function.

GRU is a common type of DNN developed to eliminate the dominant problems in RNNs caused by particular sub-structures. The internal structure of a GRU memory cell is portrayed in Fig. 5.

where h_{t-1}, h_t, h'_t are the hidden states in time $t-1$, t, and $t+1$ (renewed state at time t). Like the LSTM, the hidden layer comprises memory cells that account for the major functions. GRU combines the i, f, and o gates into r_t and z_t gates (control gates) for remembering and keeping posted of sequential information (z_t specifies the dimensions of the overlooked previous data and freshly data added and r_t redacts the dimensions of information transferred to the new state) (Cho et al. 2014). Counter to LSTM, GRU has no isolated memory cells, but it employs an individual h_t to allocate data at time steps. Basically, LSTM and GRU act approximately in the comparable mode, nevertheless, because the GRU has less learnable factors with compressed construction, it operates speedily, assuming that the dimensions of information are not too massive (Cho et al. 2014).

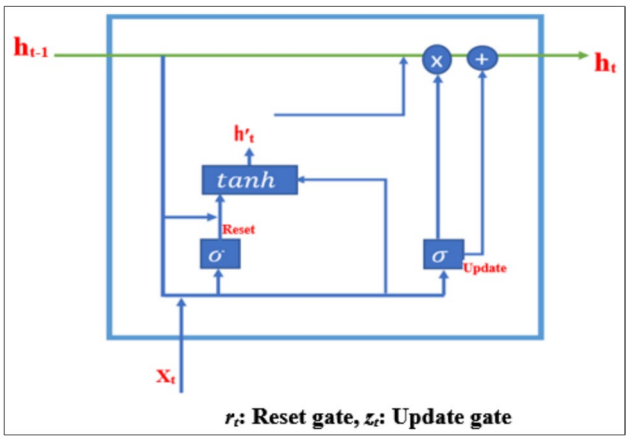

Fig. 5 Interior structure of a GRU memory cell (Lin et al. 2022)

Models development

The architecture of a DNN model must be consistent with the research question. In this study, to tune the models' structural outlines (MSO) as a meta-parameter, three different layer structures of LSTM-based models and standard GRU as the benchmark model are designed using MATLAB (2021) (Fig. 6). The innovative structures of models 3 and 4 differ from routine DL models and are not often used in hydrology to estimate the river streamflow.

An algorithm-tuning procedure is used to appropriately set up and improve these models by reducing overfitting. In this sense, the *SAF*, *P-rate*, and MSO are tuned meta-parameters. Because there is no method to determine the appropriate meta-parameters for a model with an assumed dataset, this procedure is a challenging and timewasting undertaking. Hence, several tests are done to determine the proper meta-parameters.

In the structures of these models, the first layers (*i.e.*, Input Layer) take the rainfall and streamflow time series into the models with a dimension equal to 1. Addition Layer (+) with a sole output adds the input datasets from the LSTM layers so that the predictors are of the same size, which is determined automatically by the software (MathWorks, Inc., 2021a). To prevent overfitting, dropout layer is used to routinely discount some neurons with a specific possibility frequency of P in allotted layers at every training iteration (Srivastava et al. 2014). In these models, 0.3, 0.5, and 0.7 are values tested to tune *P-rate*. Fully-connected layer is incorporated into the networks to address restrictions to fitting in scrutinizing long-period temporal datasets. This layer multiplied input datasets

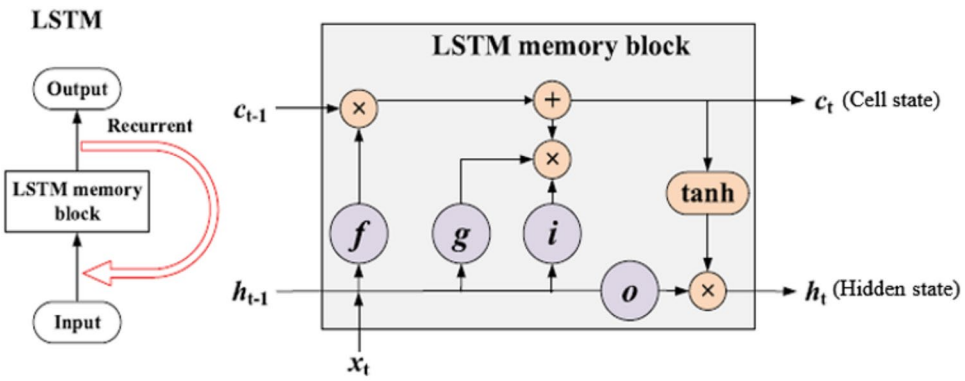

Fig. 4 Interior building and mechanisms of LSTM memory block (Ghasemlounia et al. 2022)

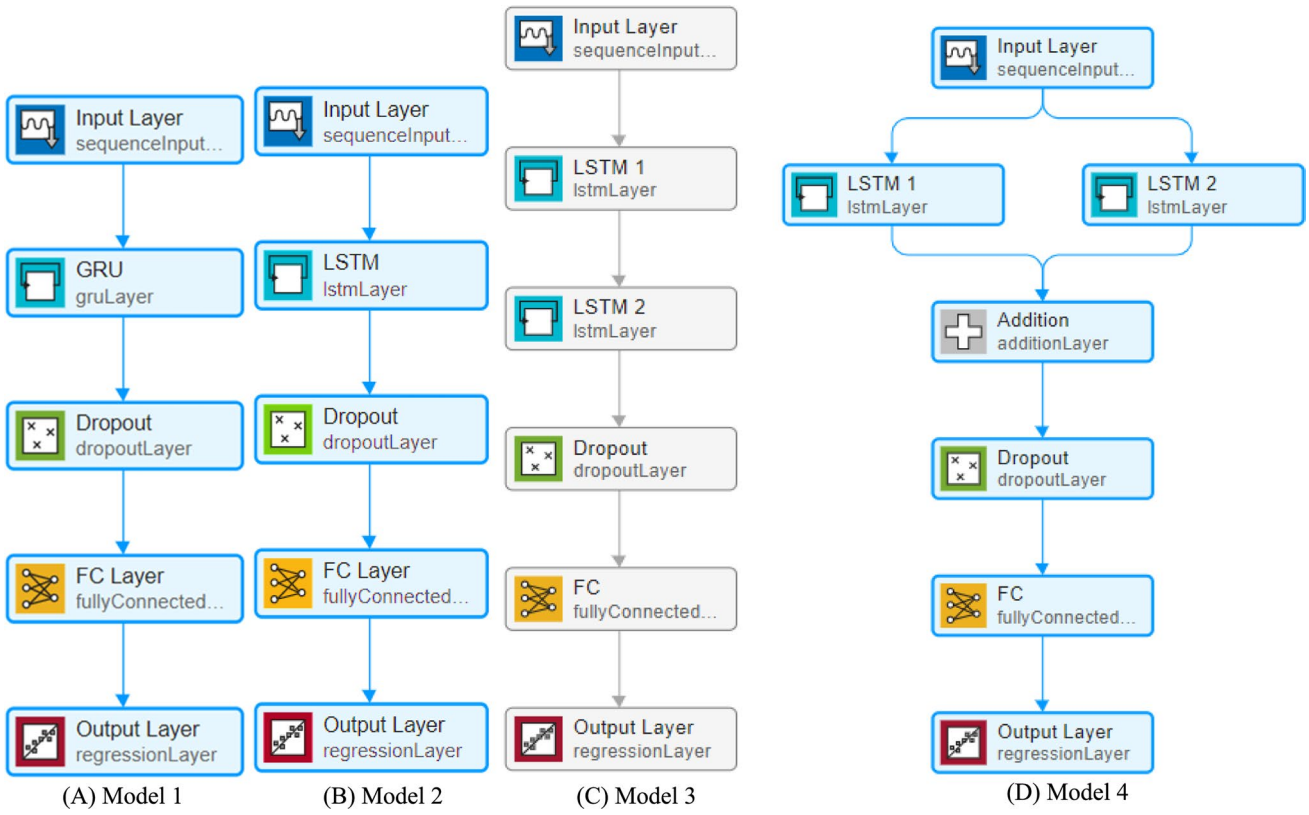

Fig. 6 The architectures of GRU and LSTM-based layer network models: **A** General single-GRU layer network model (1), **B** General single-LSTM layer network model (2), **C** Simple double-LSTM layer network (2LSTM) model (3), **D** The proposed double-LSTM layer network model coupled with Addition Layer (+) (*i.e.*, 2LSTM + layer network model (4))

using a weighting matrix and added a bias vector to recover the fitting capacity (The MathWorks, Inc., 2021a). Its input and output sizes are set to "auto" and 1, respectively, to allow the software to determine the optimal input size. As the ultimate layer, regression output layer is employed to determine the HMSE using the loss function.

Different combinations of *tanh* and *softsign* accompanied by *sigmoid* gate activation functions are distinctly employed in the LSTM layers of coupled models 3 and 4 to tune *SAF*. Different values are tested in all models to tune *NHN*. To upgrade the network bias and weights in all structures, the calibration option is set to employ the *Adam* optimization algorithm with MaxEpochs of 1000. To preclude the gradients from waning and lessen the negative influence of padding drawbacks, training features are characterized as recommended by Lin et al. (2022) and Ghasemlounia et al. (2021).

Performance evaluation criterions

R^2 (Eq. 3), *RMSE* (Eq. 4), *NSE* (Eq. 5), *PBIAS* (Eq. 6), and Kling-Gupta efficiency (*KGE*) (Eq. 7) are used to appraise the forecasting accuracies and efficacy of the models.

$$R^2 = \left(\frac{\sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y)}{N \sigma_x \sigma_y} \right)^2 \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (x_i - y_i)^2}{N}} \tag{4}$$

$$NSE = 1 - \frac{\sum_{i=1}^{N} (x_i - y_i)^2}{\sum_{i=1}^{N} (x_i - \mu_x)^2} \tag{5}$$

$$PBIAS = 100 \frac{\sum_{i=1}^{N} (y_i - x_i)}{\sum_{i=1}^{N} (x_i)} \tag{6}$$

$$KGE = 1 - \sqrt{(CC - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \tag{7}$$

where y_i and x_i show the predicted and measured $MRSF_m$ in time step i, N represents the number of datasets, σ_y and σ_x show standard deviations of the predicted and measured $MRSF_m$, μ_y and μ_x display the mean predicted and measured $MRSF_m$. CC is the correlation coefficient, α is the standard

deviation ration, and β is the mean ratio for the predicted and measured $MRSF_m$. The optimal values of Eqs. 3–7 are 1,0,1,0 and 1, respectively.

Results and discussion

Validation of the models

The features and outcomes of DNN models by the optimal scenario in the validation phase are shown (Table 4). The models in their training stages are more precise than in their testing stages. The *tanh-softsign* pairing in the hidden layers of the 2LSTM and 2LSTM + models cause to learn nonlinear and more intricate functions and it makes the model less vulnerable to overfitting. *Softsign* increases the speed of training and *tanh* can better estimate complicated relationships in long-term temporal data (Yin et al., 2020).

RMSE variations in coupled 2LSTM + model 4 over the range of *NHN* used under the optimal hyperparameter in the testing phase are graphed (Fig. 7). High *NHN* causes *RMSE* to rise due to overfitting, but low *NHN* diminishes network learning as a consequence of underfitting.

Comparing performance of the models

The effectiveness of DNN-based models is assessed using a substantial measure, total learnable parameters (*TLP*). *TLP* considers the primary determinant of physical capacity (Table 4). Forecasting with the optimum MSO and *NHN* resulted in good-adjusted *TLP* which reduced the impacts of overfitting (Gharehbaghi et al. 2022).

In the current modeling, the coupled 2LSTM model 3 is expected to be the most accurate model owing to the extreme *TLP* value; however, the coupled 2LSTM + model 4 outperformed model 3 and had less error. The chief reason for this is the large *TLP* in model 3. Due to its extra capacity, model 3 to a degree has failed to recall the training dataset and lead it to overfit and lose optimization. The amount of *TLP* in model 4 showed that insertion of addition layer creates a more suitable MSO and a more balanced *TLP*. While it is not certain that growing the number of LSTM layers and *NHN* improves the exactness and effectiveness of LSTM-based models, an appropriate MSO with optimal *NHN*, a well-balanced *TLP*, and enforcement of an algorithm-tuning practice will produce the most useful DL-based models.

Concerning convergence time and repetitions in the training phase, model 4 is more effective than model 3 thanks to the rational *TLP* quantity that enables the model to get the ideal weight sets more quickly—1000 iterations

Table 4 Features and results of developed DNN models under the ideal scenario in the validation phase

Model	NHN	SAF	P-rate	RMSE (m^3/s)	R^2	NSE	PBIAS (%)	KGE	Convergence time (s)	*TLP*
Model 1	100	tanh	0.5	21.11	0.57	0.53	109	0.63	68	30,701
Model 2	125	tanh	0.5	19.32	0.67	0.61	49	0.76	89	64,626
Model 3	150	tanh-softsign	0.7	19.8	0.65	0.57	72	0.73	366	271,951
Model 4	75	tanh-softsign	0.5	19.24	0.68	0.63	41	0.79	131	46,276

*Quantities that are shown in bold are the optimal model attributes

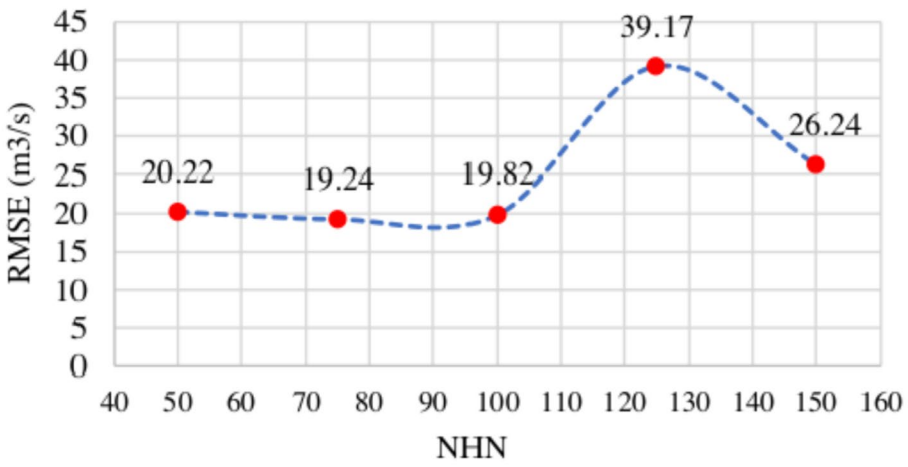

Fig. 7 *RMSE* variations in the model 4 over the range of *NHN* used under the optimal hyperparameter in the testing phase

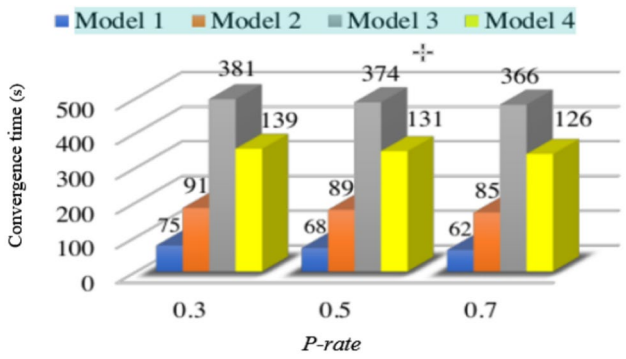

Fig. 8 Convergence time versus *P-rate* used of all models by optimal hyper-parameters in the training phase

in 131 s. Model 3, having the highest *TLP*, requires more time to train—1000 iterations in 366 s. In the training phase, as *NHN* is held fixed, rising *P-rate* decreased convergence time in all models. When both *P-rate* and *NHN* are held fixed, the *softsign* scenarios are relatively faster than those using *tanh*. Similarly, when P-*rate* remained fixed, rising *NHN* values increased convergence time (Fig. 8).

A violin plot is created to graphically depict both the divergence of the models' performances and the distribution of data by the ideal scenario in the testing stage to highlight the model that provided the best forecast of $MRSF_m$ (Fig. 9). The proposed innovative model 4 appears to fit approximately for the observed $MRSF_m$ best among the models.

The experimental and estimated chronological $MRSF_m$ by model 4 in the validation stage is shown (Fig. 10). Model 4 remembers $MRSF_m$ of previous time-series observations and captures the variations in the trends fairly well, particularly in the peak and low streamflow, which are fundamental factors in water resource management and practical flood prevention. In this regard, the peak measured streamflow (194.07 m^3/s) is overestimated by 40.1% by 2LSTM + model 4, and the lowest one (4.24 m^3/s) is overestimated 39.8%.

Conclusion

In this study, different layer structures of LSTM model *i.e.,* 2LSTM and 2LSTM + models were developed to estimate the mean monthly streamflow of the Maroon River ($MRSF_m$). To the best of author knowledge, no studies have

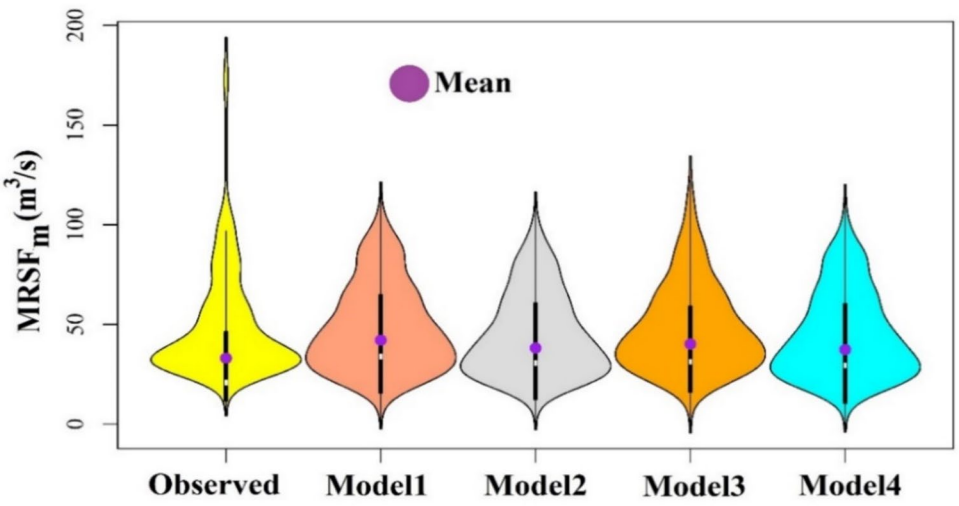

Fig. 9 Violin plot of the observed compared to the predicted time-series $MRSF_m$ by the optimal scenario in the testing stage

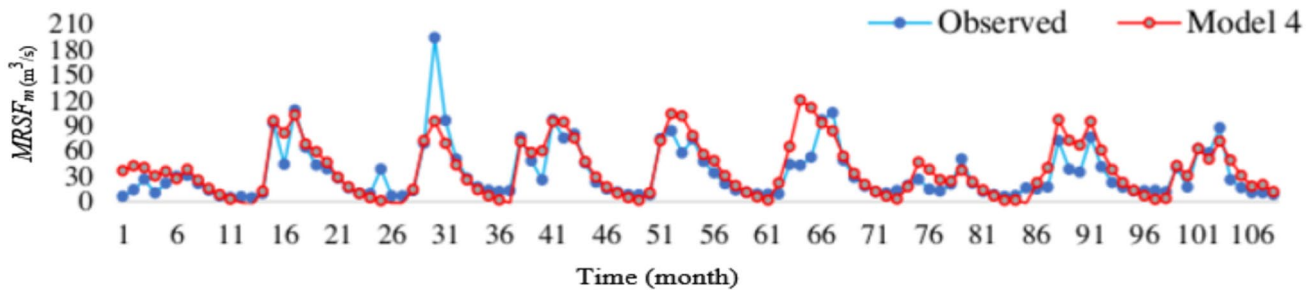

Fig. 10 Contrasting the estimated and observed $MRSF_m$ under the ideal scenario of models developed in the validation phase (108 months between Oct 2008 and Sep 2017)

probed the performance of novel introduced evolutionary neural architecture models to estimate streamflow rate as a complicated hydrological process.

Preprocessing by PCC analysis was performed between potential inputs (*i.e.,* average monthly temperature (T_m), evaporation (ET_m), and precipitation (P_m)) and output ($MRSF_m$) variables from October 1987 to September 2017. The *NHN*, *P-rate*, *SAF*, and MSO as meta-parameters were tuned to minimize the impacts of overfitting. The most worth-mentioning outcomes of the modeling process are:

1. The PCC and CAS data analysis methods confirmed that the T_m and ET_m have an inconsequential influence on $MRSF_m$, so they were thrown out in modeling process and only P_m was used as the input variable in the models developed.
2. The training-stage forms of all models were more accurate than their validation counterparts. When *NHN* was held constant, increased *P-rate* reduced convergence time. However, when *P-rate* was held fixed, mounting *NHN* dramatically increased the convergence time.
3. After several trials, the suggested coupled 2LSTM + layer network model 4 was confirmed as the best model by performance evaluation criteria to forecast $MRSF_m$. The optimal *P-rate*, *NHN,* and *SAF* for this model were determined to be 0.5, 75, and *tanh-softsign*, respectively.
4. Regardless of the high *TLP* value in the coupled 2LSTM layer network model 3, model 4 performed better. This consequence substantiates that to achieve effective LSTM-based models, the optimum tuned MSO, *NHN*, and *TLP* should be applied.

The model 4 structure is a leading-edge method as demonstrated by its admirable performance (verified statistically). It can therefore be utilized as a smart intellectual model for forecasting time-series river streamflow under varying water availability circumstances.

Model 4 is a cheap-to-run, straightforward, and robust process. Its well-formed MSO, induced an opposite response to *TLP* and generated more accuracy than the standard LSTM and benchmark model in the similar meta-parameters; yet, it required too much time to train. Despite the benefits of model 4, it has some restrictions: it requires, for instance, a very long period of detailed (*i.e.,* frequent measurements) rainfall data to generate a forecast.

Even though this work assessed the influences of different structures of LSTM-based model on forecasting $MRSF_m$, future studies could analyze other approaches. For instance, modeling could combine DNN models with optimization algorithms, stochastic gradient descent (SGD), Harris Hawks optimization (HHO), and others. The results should be compared to the outcomes of this study to move toward the most effective model.

This study demonstrated that *TLP* may be a fundamental measure for the discerning predicting capacities and performances of DNN-based models. It is useful for the assessment of the practical capabilities of such models and can also evaluate tendencies toward over and underfitting observation data used for training models.

Acknowledgements We are extremely grateful to Hamid Saadatnejadgharahassanlou for the association.

Data availability Computations and analysis minutiae for the approaches used are available from the corresponding authors on request. All datasets and information used in this study, however, are official and are the property of the Iran Water Organization. Its use may be permitted only with their approval and this must be formally requested.

Declarations

Conflict of interest The authors state that there are no competing financial interests or personal relationships and all work was performed in compliance with ethical standards. Additionally, we state that this study has not been published nor is it under consideration for publication elsewhere.

References

Abbas A, Baek S, Kim M, Ligaray M, Ribolzi O, Silvera N, Min JH, Boithias L, Cho KH (2020) Surface and sub-surface flow estimation at high temporal resolution using deep neural networks. J Hydrol 590:125370

Ahmed AM, Deo RC, Feng Q, Ghahramani A, Raj N, Yin Z, Yang L (2021) Deep learning hybrid model with Boruta-Random Forest optimizer algorithm for streamflow forecasting with climate mode indices, rainfall, and periodicity. J Hydrol 599:126350

An L, Hao Y, Yeh TCJ, Liu Y, Liu W, Zhang B (2020) Simulation of karst spring discharge using a combination of time–frequency analysis methods and long short-term memory neural networks. J Hydrol 589:125320

Bai P, Liu X, Xie J (2021) Simulating runoff under changing climatic conditions: a comparison of the long short-term memory network with two conceptual hydrologic models. J Hydrol 592:125779

Band SS, Janizadeh S, Chandra Pal S, Saha A, Chakrabortty R, Melesse AM, Mosavi A (2020) Flash flood susceptibility modeling using new approaches of hybrid and ensemble tree-based machine learning algorithms. Remote Sens 12(21):3568

Bayazit M (2015) Nonstationary of hydrological records and recent trends in trend analysis: a state-of-the-art review. Environ Process 2(3):527–542

Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Networks 5(2):157–166

Chen H, Chen A, Xu L, Xie H, Qiao H, Lin Q, Cai K (2020) A deep learning CNN architecture applied in smart near-infrared analysis of water pollution for agricultural irrigation resources. Agric Water Manag 240:106303

Chen C, Zhang Q, Kashani MH, Jun C, Bateni SM, Band SS, Chau KW (2022) Forecast of rainfall distribution based on fixed sliding window long short-term memory. Eng Appl Comput Fluid Mech 16(1):248–261

Cho K, Kim Y (2022) Improving streamflow prediction in the WRF-hydro model with LSTM networks. J Hydrol 605:127297

Cho K, Kim Y (2022) Improving streamflow prediction in the WRF-Hydro model with LSTM networks. J Hydrol 605:127297

Cho K, Van Merri¨enboer 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.

Choi J, Lee J, Kim S (2022) Utilization of the long short-term memory network for predicting streamflow in ungauged basins in Korea. Ecol Eng 182:106699

Fang S, Kang S, Huo Z, Chen S, Mao X (2008) Neural networks to simulate regional groundwater levels affected by human activities. Ground Water 46(1):80–90

Fang W, Huang S, Ren K, Huang Q, Huang G, Cheng G, Li K (2019) Examining the applicability of different sampling techniques in the development of decomposition-based streamflow forecasting models. J Hydrol 568:534–550

Fathian F, Mehdizadeh S, Sales AK, Safari MJS (2019) Hybrid models to improve the monthly river flow prediction: Integrating artificial intelligence and non-linear time series models. J Hydrol 575:1200–1213

Gao S, Huang Y, Zhang S, Han J, Wang G, Zhang M, Lin Q (2020) Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. J Hydrol 589:125188

Gharehbaghi A (2016) Explicit and implicit forms of differential quadrature method for advection—diffusion equation with variable coefficients in semi-infinite domain. J Hydrol 541:935–940

Gharehbaghi A, Ghasemlounia R, Ahmadi F, Albaji M (2022) Groundwater level prediction with meteorologically sensitive gated recurrent unit (GRU) neural networks. J Hydrol 612:128262

Ghasemlounia R, Gharehbaghi A, Ahmadi F, Saadatnejadgharahassanlou H (2021) 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

Ghorbani MA, Ahmad Zadeh H, Isazadeh M, Terzi O (2016) A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. Environ Earth Sci. https://doi.org/10.1007/s12665-015-5096-x

Graves A (2013) Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850

Hadiyan PP, Moeini R, Ehsanzadeh E (2020) Application of static and dynamic artificial neural networks for forecasting inflow discharges, case study: Sefidroud Dam reservoir. Sustain Comput Inf Syst 27:100401

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.10.1109/72.279181

Hou S, Wei J, Hou M, Xu J, Han L (2025) 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

Huang P, Song J, Cheng D, Sun H, Kong F, Jing K, Wu Q (2021a) Understanding the intra-annual variability of streamflow by incorporating terrestrial water storage from GRACE into the Budyko framework in the Qinba Mountains. J Hydrol 603:126988

Huang X, Li Y, Tian Z, Ye Q, Ke Q, Fan D, Liu J (2021b) Evaluation of short-term streamflow prediction methods in Urban river basins. Phys Chem Earth Parts a/b/c 123:103027

Huang C, Zhou T, Li W, Yu H, Li R, Fang J (2024) A coupled model integrating dual attention mechanism into BiGRU-RED for multi-step-ahead streamflow forecasting. J Hydrol 645:132137

Kratzert F, Klotz D, Herrnegger M, Sampson AK, Hochreiter S, Nearing GS (2019) Toward improved predictions in ungauged basins: exploiting the power of machine learning. Water Resour Res 55(12):11344–11354. https://doi.org/10.1029/2019wr026065

Lawrence S, Back AD, Tsoi AC, Giles CL (1997) On the distribution of performance from multiple neural-network trials. IEEE Trans Neural Networks 8(6):1507–1517

Le X-H, Ho HV, Lee G, Jung S (2019) Application of long short-term memory (LSTM) neural network for flood forecasting. Water 11(7):1387

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

Lin H, Gharehbaghi A, Zhang Q, Band SS, Pai HT, Chau K-W, Mosavi A (2022) Time series-based groundwater level forecasting using gated recurrent unit deep neural networks. Eng Appl Comput Fluid Mech 16(1):1655–1672. https://doi.org/10.1080/19942060.2022.2104928

Lin Hsu K, Gupta HV, Sorooshian S (1997) Application of a recurrent neural network to rainfall-runoff modelling. In: Proceedings of the 1997 24th Annual Water Resources Planning and Management Conference. ASCE, pp. 68–73

Liu J, Xu T, Lu C (2025) VMDI-LSTM-ED: A novel enhanced decomposition ensemble model incorporating data integration for accurate non-stationary daily streamflow forecasting. J Hydrol. https://doi.org/10.1016/j.jhydrol.2025.132769

MATLAB User's Guide 2021a, The MathWorks Inc. (Deep Learning Toolbox). Natick, Massachusetts, United State; (2021). Computer Software. www.mathworks.com/.

Millares A, Polo M, Losada M (2009) The hydrological response of baseflow in fractured mountain areas. Hydrol Earth Syst Sci Discuss 6(2):3359–3384

Mohammadi B, Ahmadi F, Mehdizadeh S, Guan Y, Pham QB, Linh NTT, Tri DQ (2020) Developing novel robust models to improve the accuracy of daily streamflow modeling. Water Resour Manage 34(10):3387–3409

Momeni E, Nazir R, Armaghani DJ, Maizir H (2014) Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN. Measurement 57(4):122–131

Reimers N, Gurevych I (2017) Optimal hyperparameters for deep LSTM-networks for sequence labeling tasks. arXiv preprint arXiv:1707.06799.

Sabzipour B, Arsenault R, Troin M, Martel JL, Brissette F, Brunet F, Mai J (2023) Comparing a long short-term memory (LSTM) neural network with a physically-based hydrological model for streamflow forecasting over a Canadian catchment. J Hydrol 627:130380

Sherstinsky A (2020) Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Phys D 404:132306

Song X, Liu Y, Xue L, Wang J, Zhang J, Wang J, Jiang L, Cheng Z (2020) Time-series well performance prediction based on long short-term memory (LSTM) neural network model. J Petrol Sci Eng 186:106682

Soo EZX, Chin RJ, Ling L, Huang YF, Lee JL, Lee FW (2024) Streamflow simulation and forecasting using remote sensing and machine learning techniques. Ain Shams Eng J 15(12):103099

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15:1929–1958

Tang Z, Zhang J, Hu M, Ning Z, Shi J, Zhai R, Wang G (2024) Improving streamflow forecasting in semi-arid basins by combining data segmentation and attention-based deep learning. J Hydrol 643:131923

Tao L, Nan Y, Cui Z, Wang L, Yang D (2025) An explainable Bayesian gated recurrent unit model for multi-step streamflow forecasting. J Hydrol Reg Stud 57:102141

Thomas HA, Fiering MB (1962) Mathematical synthesis of streamflow sequences for the analysis of river basin by simulation. Harvard University Press, Cambridge, p 751

Xiang Z, Demir I (2020) Distributed long-term hourly streamflow predictions using deep learning–a case study for State of Iowa. Environ Model Softw 131:104761

Xu L, Shi P, Wu H, Qu S, Li Q, Sun Y, Qiu C (2024) 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

Yang M, Yang Q, Shao J, Wang G, Zhang W (2023) A new few-shot learning model for runoff prediction: demonstration in two data scarce regions. Environ Model Softw 162:105659

Yuan X, Chen C, Lei X, Yuan Y, Muhammad Adnan R (2018) Monthly runoff forecasting based on LSTM–ALO model. Stoch Env Res Risk Assess 32(8):2199–2212. https://doi.org/10.1007/s00477-018-1560-y

Authors and Affiliations

Amin Gharehbaghi[1] · Redvan Ghasemlounia[2] · Shahaboddin Daneshvar[3] · Farshad Ahmadi[4]

✉ Farshad Ahmadi
f.ahmadi@scu.ac.ir

Amin Gharehbaghi
amin.gharehbaghi@hku.edu.tr;
gharehbaghi.amin@gmail.com

Redvan Ghasemlounia
redvan.ghasemlounia@gedik.edu.tr

Shahaboddin Daneshvar
shahaboddin.daneshvar@hku.edu.tr;
daneshvar.shahab@yahoo.com

[1] Department of Civil Engineering, Faculty of Engineering, Hasan Kalyoncu University, 27110 Şahinbey, Gaziantep, Turkey

[2] Department of Civil Engineering, Faculty of Engineering, Istanbul Gedik University, 34876 Istanbul, Turkey

[3] Department of Computer Engineering, Faculty of Engineering, Hasan Kalyoncu University, 27110 Şahinbey, Gaziantep, Turkey

[4] Department of Hydrology and Water Resources Engineering, Shahid Chamran University of Ahvaz, Ahvaz 6135783151, Iran