Monthly Streamflow Forecasting Using Convolutional Neural Network

Xingsheng Shu1 · Wei Ding1 · Yong Peng1 · Ziru Wang1 · Jian Wu1 · Min Li1

Received: 29 November 2020 / Accepted: 31 August 2021 / Published online: 1 November 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract
Monthly streamflow forecasting is vital for managing water resources. Recently, numerous studies have explored and evidenced the potential of artificial intelligence (AI) models in hydrological forecasting. In this study, the feasibility of the convolutional neural network (CNN), a deep learning method, is explored for monthly streamflow forecasting. CNN can automatically extract critical features from numerous inputs with its convolution–pooling mechanism, which is a distinct advantage compared with other AI models. Hydrological and large-scale atmospheric circulation variables, including rainfall, streamflow, and atmospheric circulation factors are used to establish models and forecast streamflow for Huanren Reservoir and Xiangjiaba Hydropower Station, China. The artificial neural network (ANN) and extreme learning machine (ELM) with inputs identified based on cross-correlation and mutual information analyses are established for comparative analyses. The performances of these models are assessed with several statistical metrics and graphical evaluation methods. The results show that CNN outperforms ANN and ELM in all statistical measures. Moreover, CNN shows better stability in forecasting accuracy.

Keywords Discharge prediction · Atmospheric circulation factors · Input variable selection · Data-driven model · Feature extraction

1 Introduction
Long-term streamflow forecasting is an essential basis for managing water conservancy and hydropower projects. This type of forecasting has a long forecast period, which gives water managers sufficient time to allocate water to different sectors. However, the complex

Highlights
• CNN is investigated for monthly streamflow forecasting.
• The input selection process can be automatically completed by CNN.
• The performance of CNN is superior to ANN and ELM, with smaller errors and better stability.

* Wei Ding
  weiding@dlut.edu.cn

* Yong Peng
  pengyong@dlut.edu.cn

1 Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian 116023, Liaoning, China
nature of the processes, such as the mechanism of runoff generation, climatic variation, and the effect of human activities, make accurate forecasting difficult. Therefore, favorable long-term streamflow forecasting is challenging (Yaseen et al. 2016).

During the past few decades, numerous methods have been developed for long-term streamflow forecasting (Zhang et al. 2015). These models can be classified into physical and data-driven models (Sahay and Srivastava 2014). Typically, physical models have the advantage of assisting the physical understanding of hydrological processes; however, complex physical equations, parametric assumptions, and the variables involved in the modeling process make their design and implementation difficult (Wu and Chau 2006; Yaseen et al. 2016). Meanwhile, data-driven models find the relationships between system state variables without explicit knowledge of their physical behavior using statistical or machine learning algorithms (Ghorbani et al. 2016). Owing to their advantages, including design and implementation simplicity, minimum information requirements, and relatively high accuracy, they are becoming increasingly popular in hydrological forecasting (Adamowski and Sun 2010; Yaseen et al. 2016; Deo and Şahin 2016).

In the early stages, statistical methods, such as the autoregressive model, moving average model, and autoregressive moving average model, were extensively employed for streamflow forecasting (Yu and Tseng 1996; Abrahart and See 2000; Montanari et al. 2000). At present, artificial intelligence (AI) models, such as artificial neural network (ANN), extreme learning machine (ELM), and support vector machine (SVM), have been vastly developed to solve this problem (Kişi 2004; Samsudin et al. 2011; Yaseen et al. 2016; Li et al. 2019). In addition, many hybrid methods have been proposed to improve the streamflow forecasting accuracy (Zhang et al. 2015; Ghaith et al. 2020; Ibrahim et al. 2021). However, all these methods require a common procedure before training, i.e., input variable selection. Selecting appropriate input variables is a crucial step for AI-based forecasting models (Wang et al. 2009). Unfortunately, there is no set methodology to complete such a procedure (Babel et al. 2015). Although there are some techniques available for determining input variables, such as cross-correlation (CC), mutual information (MI), and sensitivity analyses, choosing a suitable input selection technique for a specific model still relies on prior knowledge. Further, the input variables selected for a certain model are only suitable for specific forecasting cases, and this has to be repeated when data or basin changes. Therefore, for most AI-based models, the selection of appropriate input variables is also a complicated procedure.

Among many AI models, the convolutional neural network (CNN) model stands out due to its feature extraction ability. Essentially, CNN is a feedforward neural network similar to ANN. However, more advanced than ANN, CNN has convolution and pooling layers. The addition of such layers contributes to the metric of CNN to automatically extract critical features from the input layer. Therefore, CNN application in streamflow forecasting can avoid the cumbersome step of input variable selection. According to the form of the input layer, CNNs currently used in the field of hydrology can be designed into one-dimensional (1D) or two-dimensional (2D), namely, 1D-CNN and 2D-CNN. 1D-CNN employs vectors as inputs, whereas 2D-CNN employs 2D matrices as inputs. With the same network depth, 2D-CNN can extract features from more candidate variables. During the past three years, CNN has been gradually applied in hydrological forecasting (Haidar and Verma 2018; Wang et al. 2019; Barino et al. 2020; Chong et al. 2020; Huang et al. 2020; Hussain et al. 2020; Chen et al. 2021). However, the studies mainly focused on 1D-CNN applications. For example, Haidar and Verma (2018) applied a 1D-CNN model to forecast monthly rainfall for a suburb in eastern Australia. The results evidenced its good capacity in monthly rainfall forecasting. Wang et al. (2019) employed 1D-CNN for water-level forecasts. The
results indicated that 1D-CNN is superior to SVM and multilayer perceptron. Chong et al. (2020) developed a hybrid 1D-CNN to forecast daily and monthly rainfall over the Langat River Basin in Malaysia. Hussain et al. (2020) employed 1D-CNN to forecast the streamflow of four rivers in the UK. Barino et al. (2020) investigated the availability of 1D-CNN in multiple days ahead river discharge forecasting. Compared with 1D-CNN, literature on the use of 2D-CNN in streamflow forecasting is limited: Huang et al. (2020) applied 2D-CNN to forecast daily streamflow, and the 2D-CNN forecasting accuracy was much better than that of comparative models. Chen et al. (2021) designed a 2D-CNN model to realize flood process forecasting.

Motivated by its feature extracting ability, outperformance in short-term streamflow forecasting, and limited applications in long-term streamflow forecasting, we investigate the potential of the 2D-CNN to accurately forecast monthly streamflow in this study. Antecedent rainfall, streamflow, and atmospheric circulation factor (ACF) data are converted into 2D matrices and then used to drive CNNs to forecast one-month-ahead streamflow, which, to the best of our knowledge, has not been considered in existing studies. ANN and ELM are chosen as comparative methods. Two cases, Huanren Reservoir and Xiangjiaba Hydropower Station of China, are used to verify the feasibility of 2D-CNN.

2 Theoretical Overview

2.1 Convolutional Neural Network

Generally, CNN refers to 2D-CNN, the inputs of which are 2D matrices (Hussain et al. 2020). However, there are other types of CNN, such as 1D- and three-dimensional (3D)-CNN. All types of CNN have the same characteristics but differ in the dimension of the input matrix. The CNN employed in this study is the 2D-CNN. Two salient characteristics contribute to the uniqueness of CNNs. First, a neuron is only connected to their local nearby neurons in the previous layer. Second, the pooling mechanism is introduced to significantly reduce the number of coefficients in the network (Mozo et al. 2018). A standard CNN generally comprises five types of layers: input, convolution, pooling, fully connected, and output layers (Fig. 1).

The input layer provides input information for the entire model. In this study, antecedent data with time and space dimensions are employed to forecast streamflow. Let $x$- and $y$-axis denote the time and space of a matrix, respectively. The elements in the matrix are

---

**Fig. 1** Typical convolutional network architecture for forecasting
observations of different input variables at different times. In the time dimension, time ranges from the past to the present, and the time interval depends on the type of forecast. For the one-month-ahead streamflow forecasting, the time interval is 1 month. In the space dimension, different variables are viewed as a sequence of dots. Suppose the length and width of the input matrix are \(m\) and \(n\), respectively, the input matrix \(X\) can be written as follows:

\[
X = \begin{bmatrix}
x_{11} & x_{12} & \cdots & x_{1n} \\
x_{21} & x_{22} & \cdots & x_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
x_{m1} & x_{m2} & \cdots & x_{mn}
\end{bmatrix}
\]  

(1)

where \(i\) denotes different timeseries, \(j\) denotes the periods, \(m\) represents the number of timeseries employed, \(n\) is the length of timeseries, and \(x_{ij}\) represents the value of \(i\)th timeseries at the \(j\)th period.

The convolution layer comprises several convolution kernels and learns feature representations from the input layer. Denote the parameters, input, and output of the \(j\)th kernel in the \(l\)th convolution layer by \((a_j^l, b_j^l), x_i\) and \(o_j^l\), respectively, the output of the \(j\)th kernel can be written as follows:

\[
o_j^l = f\left(\sum_{k=1}^{M_l-1} a_j^l x_i^k + b_j^l\right), \quad j \in [1, M_l]
\]

(2)

where \(j\) is the index of the convolution kernel in the convolution layer, \(x_i^k\) is the \(i\)th channel of the \(x_i\), \(M_l\) is the number of kernels in the \(l\)th layer, \(f(\cdot)\) is the activation function, and \(b_j^l\) represents a bias item.

The pooling layer is used to decrease the size of the outputs from convolution layers with pooling methods. In this study, the average-pooling operation is employed. Since only one value (average value), in each scanned region is selected, the number of CNN parameters after the pooling operation significantly reduces. The alternation of the convolution and pooling layers not only reduces the network scale of a CNN but also identifies the most prominent features of the input layer.

In the fully connected layer, different features learned by the convolution and pooling layers are converted into a dense vector. The output layer is used to establish the relation, \(y = Wo + b\), among the dense vector \(o\), bias item \(b\), and the forecasted value \(y\). Once these parameters (the vector \(W\) and \(b\)) are determined, a final forecasted result can be obtained when giving a dense vector \(o\).

### 2.2 Artificial Neural Network

ANN is a nonlinear dynamic system and can approximate any nonlinear mathematical input–output relations (Cichocki and Unbehauen 1993; Pashova and Popova 2011). There are many ANN variants (Yilmaz and Yuksek 2008). In this study, we employ the feedforward backpropagation (BP) network as a comparative model. The training of a BP neural network is an optimization process (Kuang and Xu 2018), which comprises two parts: the forward and backward passes. In the forward pass, the input is processed through the ANN, and then the forecasted results are obtained. If the deviation between the forecast and observation is large, the backward pass will be performed to modify the parameters.
of each layer. After sufficient iterations of the forward and backward passes, the ANN can forecast accurately.

### 2.3 Extreme Learning Machine

ELM is a type of single hidden layer feedforward neural network (Huang et al. 2006; Hadi et al. 2020). We present a brief description of ELM.

Considering a set of training samples \(\{(X_1, y_1), (X_2, y_2), \ldots, (X_i, y_i), \ldots, (X_N, y_N)\}\), where \(X_i \subseteq \mathbb{R}^l\) is the input vector of the \(i\)th sample, \(y_i\) is the corresponding observation, and \(N\) is the size of training samples. The ELM with \(P\) hidden neurons can be modeled as follows:

\[
\sum_{j=1}^{P} \beta_j f(\omega_j X_i + b_j) = y_i, \quad i = 1, 2, \ldots, N
\]

(3)

where \(\omega_j\) is the weight vector connecting the input variables to the \(j\)th hidden neuron, \(\beta_j\) is the weight vector connecting the \(j\)th hidden neuron to the output neuron, and \(b_j\) is the bias of the \(j\)th hidden neuron. Equation (3) can be expressed in matrix form as \(y = H\beta\), where \(y = (y_1, y_2, \ldots, y_N)^T\). \(\beta = (\beta_1, \beta_2, \ldots, \beta_P)\) and \(H\) defined as follows:

\[
H = \begin{bmatrix}
    f(\omega_1 X_1 + b_1) & \cdots & f(\omega_P X_1 + b_P) \\
    \vdots & \ddots & \vdots \\
    f(\omega_1 X_N + b_1) & \cdots & f(\omega_P X_N + b_P)
\end{bmatrix}
\]

(4)

The solution of ELM is \(\hat{\beta} = H^+ y\), where \(H^+ = (H^T H)^{-1} H\) is called the Moore–Penrose generalized inverse of \(H\) (Huang et al. 2004). Finally, the forecasted value is expressed as follows:

\[
\hat{y} = \sum_{j=1}^{P} \hat{\beta}_j f(\omega_j X_{test} + b_j)
\]

(5)

where \(X_{test}\) is the input vector in the testing period.

### 2.4 Input selection methods for artificial neural network and extreme learning machine

The CC function measures how two random variables \(X\) and \(Y\) covary linearly by calculating the linear correlation coefficient \(r_{XY}\):

\[
r_{XY} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{N} (y_i - \bar{y})^2}}
\]

(6)

where \(x_i\) and \(y_i\) are sample values of \(X\) and \(Y\), respectively; \(N\) is the sample size; \(\bar{x}\) and \(\bar{y}\) are the mean value of the samples. \(r_{XY} = 1\) implies a perfect linear correlation, whereas an intermediate value corresponds to partial correlations, and \(r_{XY} = 0\) implies \(X\) is uncorrelated with \(Y\).
MI quantifies the stochastic dependency between two random variables without assuming the nature of their relation (Babel et al. 2015). For two discrete variables, $X$ and $Y$, $MI$ can be expressed as follows (Siqueira et al. 2018):

$$MI(X, Y) = \sum \sum p(X, Y) \log \left( \frac{p(X, Y)}{p(X)p(Y)} \right)$$ (7)

where the $p(X, Y)$ is the joint probability density function; $p(X)$ and $p(Y)$ are the marginal distribution function of $X$ and $Y$, respectively. $MI$ values range between 0 and infinity ($\infty$). $MI(X, Y) = 0$ implies that $X$ and $Y$ are independent of each other. If $X$ and $Y$ are dependent, the $MI$ value will be greater than 0, and the larger the value, the stronger the dependence.

### 2.5 Model Performance Evaluation

The following three statistical indices are used to evaluate the performance of the models developed in this study (Hadi et al. 2019).

I. Mean absolute error ($MAE$):

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |Q_{sim}^i - Q_{obs}^i|$$ (8)

II. Root-mean-square error ($RMSE$):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (Q_{sim}^i - Q_{obs}^i)^2}$$ (9)

III. Nash–Sutcliffe efficiency coefficient ($NSE$):

$$NSE = 1 - \frac{\sum_{i=1}^{N} (Q_{obs}^i - Q_{sim}^i)^2}{\sum_{i=1}^{N} (Q_{obs}^i - \bar{Q}_{obs})^2}$$ (10)

where $Q_{sim}^i$ and $Q_{obs}^i$ are the forecasted and observed $i$th value of the streamflow, respectively; $\bar{Q}_{obs}$ is the average of observed streamflow, $N$ is the number of observations.

### 3 Case Study

#### 3.1 Description of Catchments and Data

Huanren Reservoir and Xiangjiaba Hydropower Station are taken as case studies (Fig. 2). The Huanren Reservoir basin, located in the upper reaches of the Hunjiang River in China with a drainage area of 10,400 km², is characterized by a mountainous environment with good vegetation. The average annual rainfall is approximately 860 mm with an average annual streamflow of 142 m³/s, 70% of which are from May to October. The Xiangjiaba Hydropower Station, located in the Jinshajiang River, is China’s third-biggest hydropower
station, following the Three Gorges Hydropower Station and Xiluodu Hydropower Station. The annual average streamflow in the river is approximately 3810 m³/s, and the annual average rainfall of the entire basin is approximately 756.4 mm, 90% of which concentrates in summer.

The ACF data, including 66 variables is provided by the National Climate Center, China. Together with rainfall and streamflow, 68 kinds of variables are taken as candidate inputs. The dataset for Huanren covers 600 months (1967–2016) and is partitioned into three parts: the training dataset (1967–1997), the cross-validation dataset (1998–2005), and the testing dataset (2006–2016). The dataset for Xiangjiaba covers a period of 528 months (1967–2010), among which, 1967–1993, 1994–2000, and 2001–2010 are partitioned as the training, cross-validation, and testing periods, respectively.

### 3.2 Model Development

In this subsection, CNN, ANN, and ELM models are established to forecast the monthly streamflow in flood season (May–October) for Huanren and the entire year for Xiangjiaba.

#### 3.2.1 Convolutional Neural Network Model

For CNN, 68 variables during the past 36 months are employed as predictors. The structure of CNN used in this study includes two convolution layers, two pooling layers, and one fully connected layer. Each convolution layer has only one channel, and the hyper-parameters including...
learning rate, batch size, and the number of epochs, are set as 1.0, 3, and 80, respectively. The kernel size in the convolution and pooling layers is determined by trial and error as follows.

(a) Generate parameter combinations of the model structure, which include kernel sizes in convolution and pooling layers. The parameter combinations are sampled with kernel sizes in the convolution and pooling layers, ranging from 1 to 20 and 1 to 8, respectively.

(b) Exclude the infeasible parameter combinations of which the dimension of the output matrix in the convolution or pooling layer violates the integer constraint.

(c) Model training with feasible parameter combinations. In this study, all CNN models are trained in MATLAB 8.6 with the deep learning toolbox.

(d) Eliminate relatively poor parameter combinations. When different parameter combinations lead to the same dense vector dimension, we only reserve the parameter combination that achieves the smallest MAE in the training period.

To avoid overfitting in CNN (also in ANN and ELM), the training process is subjective to the stopping criteria, where the cross-validation error increases for a specified number of iterations, or the training reaches the scheduled maximum iterations. After the operation from step a to d, there are still many feasible structures. Among these different feasible structures, nine parameter combinations, representing different structures (Table 1) were finally chosen for forecasting.

3.2.2 ANN and ELM Models

For the ANN and ELM, their inputs are determined via CC and MI analyses, and four comparative models are established: CC-ANN, CC-ELM, MI-ANN, and MI-ELM. The number of inputs for the ANN and ELM models is designed to agree with the dimension of dense vectors in the fully connected layers listed in Table 1.

To identify the best architecture for ANN, we follow the approach in Zhang et al. (2015). A three-layer BP network model structure is used, and the optimal number of neurons in the hidden layer is determined using a heuristic method. Specifically, different numbers of neurons from 1 to 10 are tried 20 times, and the architecture that acquires the smallest MAE value in the training period is considered the optimal model. For the ELM, different

| Structure No | Size of filters | The dimension of dense vector in the fully connected layer |
|--------------|----------------|----------------------------------------------------------|
|              | Convolution layer 1 | Pooling layer 1 | Convolution layer 2 | Pooling layer 2 | |
| 1            | 13              | 4              | 5              | 2              | 5         |
| 2            | 13              | 4              | 3              | 2              | 12        |
| 3            | 17              | 1              | 19             | 2              | 17        |
| 4            | 7               | 2              | 12             | 2              | 20        |
| 5            | 15              | 1              | 19             | 2              | 36        |
| 6            | 5               | 2              | 14             | 1              | 57        |
| 7            | 13              | 4              | 2              | 1              | 65        |
| 8            | 7               | 2              | 12             | 1              | 80        |
| 9            | 19              | 1              | 16             | 1              | 105       |

| Table 1 Structure parameters of CNN models selected for monthly streamflow forecasting |
numbers of hidden neurons (from 1 to 30) and different types of activation functions (sig, sine, hardline, tribas, and radbas) are trained 20 times, and the model that achieves the smallest $\text{MAE}$ value in the training period is considered the optimal model.

4 Results and Discussion

The 1 month-ahead streamflow forecast performances in the testing period of the three models for Huanren and Xiangjiaba are listed in Table 2. CNN outperformed ANN and ELM, achieving the highest $\text{NSE}$, lowest $\text{MAE}$, and lowest $\text{RMSE}$ with an equal number of inputs, and ELM outperformed ANN in most situations. Considering the Xiangjiaba as an example, when the number of inputs is 5, the error of CNN reduced by 15.7% and 25.5% in $\text{MAE}$ and 19.7% and 37.9% in $\text{RMSE}$ compared with CC-ELM and CC-ANN, respectively ($\text{MAE}_{\text{CNN}} = 891.21 \text{ m}^3/\text{s}$, $\text{MAE}_{\text{CC-ANN}} = 1118.96 \text{ m}^3/\text{s}$, $\text{MAE}_{\text{CC-ELM}} = 1386.78 \text{ m}^3/\text{s}$, $\text{RMSE}_{\text{CC-ANN}} = 1913.02 \text{ m}^3/\text{s}$). When the number of inputs is 20, the CNN achieved better forecasting accuracy than MI-ELM and MI-ANN ($\text{MAE}_{\text{CNN}} = 904.52 \text{ m}^3/\text{s}$, $\text{MAE}_{\text{MI-ANN}} = 962.03 \text{ m}^3/\text{s}$, $\text{MAE}_{\text{MI-ELM}} = 1152.26 \text{ m}^3/\text{s}$, $\text{RMSE}_{\text{CNN}} = 1375.02 \text{ m}^3/\text{s}$, $\text{RMSE}_{\text{MI-ANN}} = 1555.57 \text{ m}^3/\text{s}$, $\text{RMSE}_{\text{MI-ELM}} = 1940.13 \text{ m}^3/\text{s}$). The results also showed that there was no strict increasing or decreasing trend for each performance index for any model, which implied that the inclusion of more inputs did not necessarily improve forecast results. Notably, the forecasting for Xiangjiaba was more accurate than that for Huanren with $\text{NSE}$ varying from 0.13 to 0.84 and −0.55 to 0.37, respectively.

Figure 3a details the best forecast results from each model in the testing period for Xiangjiaba, including the observed and forecasted data, absolute errors (forecasted value−observed value), and relative errors (RE). The forecasts from the three models all fit well with the observations, with an average RE smaller than 10%. However, CNN underestimated the peaks of streamflow, whereas ANN and ELM were more likely to overestimate the peaks. For medium flow, the forecast results of the models were close, and all fit well with the observations. In terms of low flow, all models underestimated the streamflow.

Figure 3b exhibits the best forecast results from each model in the testing period for Huanren. From the figure, the peaks of streamflow in this region were mostly overestimated by each model, except for some extremely high peaks. For the medium and low flows, the average REs for each model were approximately 50%, and the maximum RE was up to 500%. Compared with Xiangjiaba, the forecast performances by each model in Huanren were relatively poor, which might be attributed to the Huanren Reservoir basin characteristics. The Huanren is located in northern China, where the formation of rainfall in the flood season is strongly influenced by the mutual effects of Siberian cold air and the summer monsoon. The interaction between the Siberian cold air and the summer monsoon is complicated and varies significantly, thus leading to the difficult forecasting of long-term streamflow. Besides, streamflow in Huanren has a greater variation, which might increase the difficulty in forecasts.

The scatter plots and linear correlation coefficients ($r$) for the best forecast results from each model are shown in Fig. 4. In Fig. 4a, the data points of each model were scattered, which confirmed the conclusion that the forecast results for Huanren were relatively poor. Nevertheless, CNN still performed slightly better than ANN and ELM with a relatively high $r$-value of 0.615. In Fig. 4b, the data points of each model concentrated in the vicinity of the diagonal $y = x$, and the correlation coefficients of all models were above 0.9.
| Station | Evaluation metric | Model  | Number of inputs | 5  | 12  | 17  | 20  | 36  | 57  | 65  | 80  | 105 |
|---------|------------------|--------|------------------|----|-----|-----|-----|-----|-----|-----|-----|-----|
|         |                  |        |                  |    |     |     |     |     |     |     |     |     |
| Huanren | MAE              | CNN    | 119.02           | 122.02 | 110.75 | 108.80 | 111.83 | 117.99 | 117.73 | 123.21 | 129.28 |
|         |                  | CC-ANN | 138.85           | 157.77 | 137.31 | 141.33 | 160.44 | 172.59 | 186.60 | 194.20 | 194.77 |
|         |                  | CC-ELM | 145.69           | 132.20 | 122.87 | 132.62 | 129.17 | 130.96 | 143.59 | 136.88 | 133.77 |
|         |                  | MI-ANN | 125.88           | 160.57 | 183.69 | 157.55 | 188.56 | 137.84 | 191.13 | 179.43 | 176.60 |
|         |                  | MI-ELM | 120.79           | 146.48 | 144.22 | 135.66 | 138.60 | 128.37 | 140.94 | 133.94 | 135.25 |
|         | RMSE             | CNN    | 201.05           | 207.57 | 195.82 | 196.25 | 211.01 | 188.94 | 208.66 | 214.27 | 215.93 |
|         |                  | CC-ANN | 203.18           | 246.93 | 210.73 | 212.04 | 253.42 | 295.83 | 292.43 | 258.29 | 257.66 |
|         |                  | CC-ELM | 223.54           | 214.19 | 208.60 | 215.52 | 216.30 | 216.01 | 211.34 | 223.48 | 216.62 |
|         |                  | MI-ANN | 210.87           | 291.60 | 290.01 | 261.07 | 285.76 | 227.34 | 282.06 | 255.87 | 249.40 |
|         |                  | MI-ELM | 211.14           | 237.67 | 237.76 | 241.84 | 263.58 | 229.79 | 231.81 | 231.93 | 220.13 |
|         | NSE              | CNN    | 0.28             | 0.24   | 0.32   | 0.32   | 0.21   | 0.37   | 0.23   | 0.19   | 0.17   |
|         |                  | CC-ANN | 0.27             | −0.08  | 0.21   | 0.20   | −0.14  | −0.55  | −0.52  | −0.18  | −0.18  |
|         |                  | CC-ELM | 0.11             | 0.19   | 0.33   | 0.18   | 0.17   | 0.17   | 0.21   | 0.11   | 0.17   |
|         |                  | MI-ANN | 0.21             | −0.51  | −0.49  | −0.21  | −0.45  | 0.08   | −0.41  | −0.16  | −0.10  |
|         |                  | MI-ELM | 0.21             | 0.00   | 0.00   | −0.04  | −0.23  | 0.06   | 0.05   | 0.05   | 0.14   |
| Xiangjiaba | MAE            | CNN    | 891.21           | 889.42 | 947.59 | 904.52 | 945.9  | 947.38 | 952.75 | 915.25 | 988.88 |
|         |                  | CC-ANN | 1118.96          | 1072.87 | 1035.89 | 1180.75 | 1207.65 | 1521.14 | 2058.99 | 2049.68 | 2037.72 |
|         |                  | CC-ELM | 1030.85          | 976.47 | 1016.09 | 956.84 | 1042.07 | 1025.97 | 1099.43 | 1119.85 | 1114.16 |
| Station | Evaluation metric | Model       | Number of inputs |
|---------|------------------|-------------|------------------|
|         |                  |             | 5    | 12    | 17    | 20    | 36    | 57    | 65    | 80    | 105   |
|         |                  | MI-ANN      | 1189.31 | 1043.08 | 1718.60 | 1152.26 | 1553.53 | 1887.29 | 1671.96 | 1812.35 | 2138.22 |
|         |                  | MI-ELM      | 932.70  | 1052.34 | 983.96  | 962.03  | 1013.46 | 1002.38 | 1136.28 | 1080.84 | 1179.70 |
|         | RMSE             | CNN         | 1386.78 | 1454.04 | 1521.66 | 1375.02 | 1455.98 | 1428.12 | 1437.22 | 1406.16 | 1470.65 |
|         |                  | CC-ANN      | 1913.02 | 1842.85 | 1671.74 | 2006.48 | 2106.44 | 2361.55 | 3391.41 | 3043.03 | 2974.89 |
|         |                  | CC-ELM      | 1659.43 | 1687.11 | 1518.36 | 1504.38 | 1708.49 | 1527.15 | 1656.86 | 1632.24 | 1706.22 |
|         |                  | MI-ANN      | 1917.73 | 1813.27 | 3112.44 | 1940.13 | 2622.58 | 2962.95 | 2817.73 | 3070.36 | 3251.38 |
|         |                  | MI-ELM      | 1727.11 | 1861.02 | 1626.48 | 1555.57 | 1579.54 | 1616.53 | 1689.03 | 1631.86 | 1750.68 |
|         | NSE              | CNN         | 0.84    | 0.83    | 0.81    | 0.84    | 0.83    | 0.83    | 0.83    | 0.84    | 0.82    |
|         |                  | CC-ANN      | 0.70    | 0.72    | 0.77    | 0.67    | 0.64    | 0.54    | 0.06    | 0.24    | 0.27    |
|         |                  | CC-ELM      | 0.77    | 0.77    | 0.81    | 0.81    | 0.76    | 0.81    | 0.77    | 0.78    | 0.76    |
|         |                  | MI-ANN      | 0.70    | 0.73    | 0.20    | 0.69    | 0.43    | 0.28    | 0.35    | 0.23    | 0.13    |
|         |                  | MI-ELM      | 0.76    | 0.72    | 0.78    | 0.80    | 0.80    | 0.79    | 0.77    | 0.78    | 0.75    |

Table 2 (continued)
Although the \( r \)-value of CNN was the lowest (\( r_{CNN} = 0.910, r_{ANN} = 0.929, r_{ELM} = 0.938 \)), it was only a 2.04\% and 2.99\% reduction compared with ANN and ELM, respectively. Table 3 presents the metric improvement of CNN compared with ANN and ELM in terms of their best forecast results. In Table 3, the metrics of CNN were better than those of ANN and ELM for the Huanren. For the Xiangjiaba, all metrics, except \( r \), outperformed those of ANN and ELM.

In the process of establishing forecast models, it is not easy to obtain the optimal input combination for a specific model. Therefore, the sensitivity of the model performance to input combinations is also a key indicator and worthy of attention. A lower sensitivity can make forecasts closer to the optimal value, although the optimal input combination is not employed. Figure 5 displays the overall distribution of the performance metrics. Notably, for \( MAE \) and \( RMSE \), smaller median and shorter box height represent higher accuracy and stability, whereas, for \( NSE \), a larger median and shorter box height represent more satisfying forecasts and lower sensitivity. For the Huanren, ANN had a high sensitivity to the input combinations with a large box height of each performance metric. CNN and ELM had nearly the same but better stability than ANN. CNN outperformed ELM because of its smaller box median of \( MAE \) and \( RMSE \) and larger median of \( NSE \). For the Xiangjiaba, a similar result was realized.
In this study, CNN was applied to long-term streamflow forecasting for the first time. In the application of Huanren, CNN outperformed its comparative models with better forecasting results and stable forecasting ability. In the application of Xiangjiaba, similar results were obtained. Therefore, we can assert that CNN can be used for monthly streamflow forecasting, and its feature extraction ability is effective. For most data-driven models (e.g., ANN and ELM), a careful selection of the input variables is required, whereas, for CNN, such work can be completed by its feature extraction module automatically, and only constructing the candidate variables (e.g., rainfall, streamflow, and ACFs) into input matrices is required. In fact, we cannot underestimate this advantage, which can save us a lot of work, especially for predictions from numerous catchments. Further, incorporating this advantage into other data-driven models is meaningful and relevant for further study, which will make them also have the feature extraction ability.

Although encouraging forecasting results have been achieved in Huanren and Xiangjiaba, CNN has some shortcomings that require mitigation. For example, it tends to underestimate or overestimate the peaks of streamflow and yields relatively poor forecasting results for the streamflow with greater irregularity in some areas. In addition, the forecasting effect in some small watersheds (e.g., a watershed with an area less than 200 km²) requires further verification.

![Fig. 4 Scatter plots of observed and forecasted values within the testing period from optimal CNN, ANN, and ELM for (a) Huanren and (b) Xiangjiaba](image)

Table 3  The performance improvement achieved by CNN comparing the optimal forecast results of each model (%)

| Station   | Compared model | MAE  | RMSE | NSE   | r    |
|-----------|----------------|------|------|-------|------|
| Huanren   | ELM            | 3.97 | 9.42 | 12.12 | 21.54|
|           | ANN            | 14.07| 10.34| 76.19 | 1.15 |
| Xiangjiaba| ELM            | 5.47 | 8.60 | 3.75  | −2.99|
|           | ANN            | 12.68| 17.75| 9.62  | −2.05|
Conclusions

In this article, we explored the potential of CNN for monthly streamflow forecasting. The monthly streamflow and rainfall data from Huanren Reservoir and Xiangjiaba Hydropower Station and ACF data from the National Climate Center, China were employed for model training, validation, and testing. Nine cases of different structures of CNN were designed with ACFs, rainfall, and streamflow delayed for 1 to 36 months as inputs. ANN and ELM were employed as comparison models.

The results demonstrated that CNN could be successfully applied to forecast the monthly streamflow. Comparing the results of CNN, ANN, and ELM, CNN outperformed ANN and ELM with smaller MAE and RMSE and higher NSE. With the change of the model structure or input combinations, CNN showed better stability in forecasting accuracy than ANN and ELM. The results also demonstrated that CNN, ANN, and ELM yielded satisfactory forecasts for Xiangjiaba but were unable to maintain their accuracy for Huanren. Overall, the results and analysis presented in this study demonstrate that CNN is a superior and an alternative to ANN and ELM in monthly streamflow forecasting.

Author Contributions YP designed the study. XS performed the research and wrote the initial draft of manuscript. WD analyzed the data and made revisions to the draft. ZW contributed to the revisions. JW contributed to the revisions. ML contributed to the revisions.

Funding This work was supported by the National Key Research and Development Program of China [2016YFC0400906, 2017YFC0406005]; and the National Natural Science Foundation of China [91547111].

Declarations

Ethical Approval Not applicable.

Consent to Participate Not applicable.
Consent to Publish  Not applicable.

Conflict of Interest  Authors declare no conflict of interest.

References

Abrahart RJ, See L (2000) Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. Hydrol Process 14(11–12):2157–2172. https://doi.org/10.1002/1099-1085(20000815)14:11<2157::AID-HYP573>3.0.CO;2-S

Adamowski J, Sun K (2010) Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. J Hydrol 390(1–2):85–91. https://doi.org/10.1016/j.jhydrol.2010.06.033

Babel MS, Badgujar GB, Shinde VR (2015) Using the mutual information technique to select explanatory variables in artificial neural networks for rainfall forecasting. Meteorol Appl 22(3):610–616. https://doi.org/10.1002/met.1495

Barino FO, Silva VNH, Lopez-Barbero AP, De Mello HL, Santos ABD (2020) Correlated time-series in multi-day-ahead streamflow forecasting using convolutional networks. IEEE Access 8:215748–215757. https://doi.org/10.1109/ACCESS.2020.3040942

Chen C, Hui Q, Xie W, Wan S, Zhou Y, Pei Q (2021) Convolutional neural networks for forecasting flood process in internet-of-things enabled smart city. Comput Netw 16107744. https://doi.org/10.1016/j.comnet.2020.107744

Chong KL, Lai SH, Yao Y, Ahmed AN, Jaafar WZW, El-Shafie A (2020) Performance enhancement model for rainfall forecasting utilizing integrated wavelet-convolutional neural network. Water Resour Manag 34(8):2371–2387. https://doi.org/10.1007/s11269-020-02554-z

Cichocki A, Unbehauen R (1993) Neural networks for optimization and signal processing. Wiley, Chichester

Deo RC, Sahin M (2016) An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. Environ Monit Assess. https://doi.org/10.1007/s10661-016-5094-9

Ghaith M, Siam A, Li Z, El-Dakhakhni W (2020) Hybrid hydrological data-driven approach for daily streamflow forecasting. J Hydrol Eng. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001866

Ghorbani MA, Zadeh HA, 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

Hadi SJ, Abba SI, Sammen SS, Salih SQ, Al-Ansari N, Yaseen ZM (2019) Non-linear input variable selection approach integrated with non-tuned data intelligence model for streamflow pattern simulation. IEEE Access 7(7):141533–141548. https://doi.org/10.1109/ACCESS.2019.2943515

Hadi SJ, Tombul M, Salih SQ, Al-Ansari N, Yaseen ZM (2020) The capacity of the hybridizing wavelet transformation approach with data-driven models for modeling monthly-scale streamflow. IEEE Access 8:101993–102006. https://doi.org/10.1109/ACCESS.2020.2998437

Haidar A, Verma B (2018) Monthly rainfall forecasting using one-dimensional deep convolutional neural network. IEEE Access 6:69053–69063. https://doi.org/10.1109/ACCESS.2018.2880044

Huang C, Zhang J, Cao L, Wang L, Luo X, Wang J, Bensoussan A (2020) Robust forecasting of riverflow based on convolutional neural network. IEEE Trans Sustainable Computing. https://doi.org/10.1109/TSUSC.2020.2983097

Huang G, Zhu Q, Siew C (2004) Extreme learning machine: a new learning scheme of feedforward neural networks, 2004. IEEE. https://doi.org/10.1109/IJCNN.2004.1380068

Huang G, Zhu Q, Siew C (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1–3):489–501. https://doi.org/10.1016/j.neucom.2005.12.126

Hussain D, Hussain T, Khan AA, Naqvi SAA, Jamil A (2020) A deep learning approach for hydrological time-series prediction: a case study of Gilgit river basin. Earth Sci Inform. https://doi.org/10.1007/s12145-020-00477-2

Ibrahim KSMH, Huang YF, Ahmed AN, Koo CH, El-Shafie A (2021) A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. Alex Eng J. https://doi.org/10.1016/j.ajej.2021.04.100

Kuang D, Xu B (2018) Predicting kinetic triplets using a 1d convolutional neural network. Thermochim Acta 669:8–15. https://doi.org/10.1016/j.tca.2018.08.024
Li X, Sha J, Wang Z (2019) Comparison of daily streamflow forecasts using extreme learning machines and the random forest method. Hydrol Sci J 64(15):1857–1866. https:// doi.org/10.1080/02626667.2019.1680846
Montanari A, Rosso R, Taqqu MS (2000) A seasonal fractional ARIMA Model applied to the Nile River monthly flows at Aswan. Water Resour Res 36(5):1249–1259. https:// doi.org/10.1029/2000wr900012
Mozo A, Ordozgoiti B, Gómez-Canaval S (2018) Forecasting short-term data center network traffic load with convolutional neural networks. PLoS ONE 13(2):e191939. https:// doi.org/10.1371/journal.pone.0191939
Kişi Ö (2004) River flow modeling using artificial neural networks. J Hydrol Eng 9(60):60–63
Pashova L, Popova S (2011) Daily sea level forecast at tide gauge Burgas, Bulgaria using artificial neural networks. J Sea Res 66(2):154–161. https:// doi.org/10.1016/j.seares.2011.05.012
Sahay RR, Srivastava A (2014) Predicting monsoon floods in rivers embedding wavelet transform, genetic algorithm and neural network. Water Resour Manag 28(2):301–317. https:// doi.org/10.1007/s11269-013-0446-5
Samsudin R, Saad P, Shabri A (2011) River flow time series using least squares support vector machines. Hydrol Earth Syst Sci 15(6):1835–1852. https:// doi.org/10.5194/hess-15-1835-2011
Siqueira H, Boccato L, Luna I, Attux R, Lyra C (2018) Performance analysis of unorganized machines in streamflow forecasting of Brazilian plants. Appl Soft Comput 68:494–506. https:// doi.org/10.1016/j.asoc.2018.04.007
Wang W, Chau K, Cheng C, Qiu L (2009) A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. J Hydrol 374(3–4):294–306. https:// doi.org/10.1016/j.jhydr ol.2009.06.019
Wang J, Lin G, Chang M, Huang I, Chen Y (2019) Real-time water-level forecasting using dilated causal convolutional neural networks. Water Resour Manag 33(11):3759–3780. https:// doi.org/10.1007/s11269-019-02342-4
Wu CL, Chau KW (2006) Evaluation of several algorithms in forecasting flood. Adv Appl Artif Intell 4031:111–116
Yaseen ZM, Jaafar O, Deo RC, Kisi O, Adamowski J, Quilty J, El-Shafie A (2016) Stream-flow forecasting using extreme learning machines: a case study in a semi-arid region in Iraq. J Hydrol 542:603–614. https:// doi.org/10.1016/j.jhydr ol.2016.09.035
Yılmaz I, Yüksel AG (2008) An example of artificial neural network (ANN) application for indirect estimation of rock parameters. Rock Mech Rock Eng 41(5):781–795. https:// doi.org/10.1007/s00603-007-0138-7
Yu P, Tseng T (1996) A model to forecast flow with uncertainty analysis. Hydrol Sci J 41(3):327–344. https://doi.org/10.1080/02626669609491506
Zhang X, Peng Y, Zhang C, Wang B (2015) Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. J Hydrol 530:137–152. https://doi.org/10.1016/j.jhydrol.2015.09.047

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.