




Data-Driven Dam Outflow Prediction Using Deep Learning with Simultaneous Selection of Input Predictors and Hyperparameters Using the Bayesian Optimization Algorithm

Vinh Ngoc Tran¹ · Duc Dang Dinh² · Binh Duy Huy Pham² · Kha Dinh Dang³ · Tran Ngoc Anh^{2,3} · Ha Nguyen Ngoc⁴ · Giang Tien Nguyen³ 

Received: 18 July 2022 / Accepted: 17 November 2023 / Published online: 1 December 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract

Reservoirs and dams are critical infrastructures that play essential roles in flood control, hydropower generation, water supply, and navigation. Accurate and reliable dam outflow prediction models are important for managing water resources effectively. In this study, we explore the application of three deep learning (DL) algorithms, i.e., gated recurrent unit (GRU), long short-term memory (LSTM), and bidirectional LSTM (BiLSTM), to predict outflows for the Buon Tua Srah and Hua Na reservoirs located in Vietnam. An advanced optimization framework, named the Bayesian optimization algorithm with a Gaussian process, is introduced to simultaneously select the input predictors and hyperparameters of DLs. A comprehensive investigation into the performance of three DLs in multistep-ahead prediction of outflow of two dams shows that all three models can predict the reservoir outflow accurately, especially for short lead-time predictions. The analysis results based on the root mean square error, Nash–Sutcliffe efficiency, and Kling–Gupta efficiency indicate that BiLSTM and GRU are the most suitable models to diagnose the outflow of Buon Tua Srah and Hua Na reservoirs, respectively. Conversely, the results of the similarity assessment of 11 hydrological signatures show that LSTM outperforms BiLSTM and GRU in both case studies. This result emphasizes the importance of determining the purpose and objective function when choosing the best model for each case study. Ultimately, these results strengthen the potential of DL for efficient and effective reservoir outflow predictions to help policymakers and operators manage their water resource system operations better.

Keywords Dam outflow prediction · Long short-term memory · Input predictor selection · Hyperparameter optimization

1 Introduction

Reservoirs and dams serve critical functions in mitigating natural disasters including droughts and floods, providing potable and irrigation water, and generating electricity. As anthropogenic structures requiring human operation, in addition to beneficial impacts,

these reservoirs also induce alterations of natural regimes pertaining to flow, sediment transport, and the ambient environment. Specifically, dams primarily influence the hydrologic regime by changing the magnitude and timing of the discharges downstream, often with the intent to mitigate hydrologic extremes (i.e., floods and droughts) (Beiranvand and Ashofteh 2023; Döll et al. 2009). Dams reduce peak discharges by roughly a third on average while dampening the daily streamflow by a similar amount (Graf 2006). The optimization of reservoir/dam operations is rendered more complex when accounting for numerous influencing factors, including operational expertise, site-specific rule curves, intended reservoir uses, human manipulation, precipitation patterns, reservoir inflow, water levels, downstream river regimes, and localized release decision-making practices distinct to each infrastructure project. Consequently, the streamflows downstream of dams and reservoirs constitute anthropogenically engineered fluxes divergent from natural flow regimes. This underscores a principal challenge in accurately modeling dam discharge dynamics, which bears critical import for administering water resources upstream and downstream of reservoirs.

Dam outflow is typically a nonlinear and complex process driven by anthropogenic and environmental influences that make predicting dam outflow difficult, particularly in the context of predicting dam-induced hydrologic responses at diurnal or subdiurnal time steps (El-Shafie et al. 2006; Jothiprakash and Magar 2012; Seo et al. 2015). Gutenson et al. (2020) classified two approaches for dam outflow prediction, nondata- and data-driven. The non-data-driven approach is based on conceptualizing reservoir responses using available information (e.g., dam water level, inflow, reservoir storage, and outflow) (Beiranvand and Ashofteh 2023; Döll et al. 2009; Gutenson et al. 2020; Hanasaki et al. 2006). This approach was mainly developed to present the operation of natural reservoirs (Gutenson et al. 2020; Han et al. 2020). Conversely, the data-driven approach, also known as machine learning or artificial intelligence, can be effectively applied to dynamic nonlinear systems, particularly when the governing influence on the system does not follow any deterministic model (Coerver et al. 2018; Ehsani et al. 2016; Mohan and Ramsundram 2016; Zhang et al. 2019). These approaches involve reservoir-related data or specific parameters to build a model that can be used as a predictive model.

Recently, data-driven approaches have attracted attention owing to their strong learning capabilities and suitability for modeling complex nonlinear processes (Aksoy and Dahamshah 2018; Mohandes et al. 2004; Nourani et al. 2014; Shi et al. 2015; Yaseen et al. 2015). Many techniques can provide satisfactory results in earth science applications, such as artificial neural network (ANN), recurrent neural network (RNN), support vector regression, genetic programming, multilayer perceptron (MLP), and long short-term memory (LSTM) and its variants (e.g., gated recurrent unit, GRU; and bidirectional LSTM, BiLSTM), have been proven. The latter techniques, i.e., GRU, LSTM, and BiLSTM (called deep learning, DL, hereinafter), overcome the notorious problem of vanilla RNN in addressing the difficulty in long-range dependency learning (Greff et al. 2017). Therefore, DL models can learn the nonlinearity of input variables with an arbitrary length, effectively capture long-term time dependencies, and provide predictions more accurately than other methods such as ANN, RNN, or MLP (Hu et al. 2018; Le et al. 2019; Ni et al. 2020; Xiang et al. 2020). Although the effectiveness of DL has been demonstrated, few studies have used it to ensure that it can provide reliable predictions in the case of dam outflow.

Although DL models perform well, their complex topology design and hyperparameter configuration pose a challenge in building a well-performing DL (Khosravi et al. 2022; Kratzert et al. 2018). Common methods, such as trial-and-error, grid search, and random search, are often used, but they have a slow convergence rate and do not specifically

consider the effects of the interaction between hyperparameters (Bergstra and Bengio 2012). Recently, the Bayesian optimization algorithm (BOA) has received attention owing to its higher efficiency compared with other algorithms (grid or random searches) as it can acquire satisfactory results with fewer iterations and is more suitable for computationally expensive optimization problems (Alizadeh et al. 2021). Another issue in DL applications is selecting proper input variables and their sequence lengths (also known as the lookback periods or lengths of the time lag) (Adamowski and Sun 2010). These input variables and their respective sequence lengths are collectively termed “input predictors” because they are fed into DLs as predictors to predict the target outputs. Inappropriate inputs lead to nonconvergence in model training and poor reliability of the trained model predictions (Bowden et al. 2005; Latif and Ahmed 2023). This highlights the need for a thorough understanding of the underlying physical processes from available data and the effect of such data on dam outflow. The most common approach in previous studies was to use trial-and-error based on multiple scenarios of input combinations, *ad hoc* selections, or statistical analyses for critical inputs (Bozorg-Haddad et al. 2016; Sauhats et al. 2016; Yang et al. 2017a, b). In most previous studies, the selection of principal inputs and hyperparameters was performed stepwise separately. Specifically, the hyperparameters were fixed while selecting the principal input variables and their sequence lengths. The new hyperparameters are then optimized with the selected principal inputs (Ahmad and Hossain 2019; Tran et al. 2021). However, the optimization of the selection of principal inputs is often closely related to the hyperparameters of DL models. The number of input predictors used affects the model configuration, and vice versa. Thus, such an independent optimization may not produce the best-performing model (Alizadeh et al. 2021).

This study aimed to investigate the potential of DL models for predicting dam outflow and to develop a unique framework to simultaneously optimize hyperparameters and select principal input variables and their sequence lengths for DL models using the BOA. For these purposes, three DL models, LSTM, BiLSTM, and GRU, were implemented to identify which of these models best produced accurate dam outflow. All experiments were conducted using a dataset from two case studies of Buon Tua Srah and Hua Na dams located in Vietnam. The rest of this study is organized as follows: Section 2 describes the methodologies of the three DL models, the BOA, a DL modeling framework, and evaluation metrics. Section 3 describes the study area, dataset, and experimental setup. Section 4 presents the experimental results and discussion, and a conclusion follows in Section 5.

2 Materials and Methods

2.1 Deep Learning Methods

2.1.1 Long Short-Term Memory Network (LSTM)

Long Short-Term Memory (LSTM) is a variant of recurrent neural networks (RNNs) that mitigates the vanishing gradient problem through a specialized memory cell termed the LSTM cell. LSTM cells can retain information over extended time lags and regulate information propagation to subsequent cells. This enables LSTM networks to learn long-term dependencies inherent in sequential data (Hochreiter and Schmidhuber 1997). The LSTM equations are expressed in Section S.1 in supplementary material (SM) file.

2.1.2 Gated Recurrent Unit (GRU)

GRU is a type of gating mechanism used in RNNs with a memory neuron that can address the issue of vanishing or exploding gradients (Cho et al. 2014). By simplifying the structure of LSTM, the GRU architecture has two gates: an update (z_t) and a reset (r_t). The update gate determines how much information will be retained from the state of the previous step h_{t-1} and flow to the neuron, whereas the reset gate determines whether to ignore the previous state and upset the current state. The GRU equations are presented in Section S.2 in the SM file.

2.1.3 Bidirectional Long Short-Term Memory Networks (BiLSTM)

BiLSTM is a deformation structure of LSTM that contains forward and backward LSTM layers (Schuster and Paliwal 1997). It can analyze data forward and backward simultaneously. Therefore, BiLSTM is better in capturing the future and past information of the input sequence compared to LSTM. This type of process is helpful in time-series data when we want to understand the data at each timestep (Salehinejad et al. 2017).

2.2 Bayesian Optimization with Gaussian Process

In this section, an optimization framework is presented to simultaneously determine the optimal input variables, their sequence lengths, and model hyperparameters. Specifically, the sequence lengths of candidate inputs are assumed to be hyperparameters with values varying between 0 and 30 (days). The value of 0 indicates that the candidate input will not be selected as the model input, whereas a value > 0 denotes the sequence length of the selected input. Hyperparameter optimization can be considered a black-box problem, where the objective function of optimization is a black-box model. The hyperparameter optimization problem can be expressed as follows:

$$X^* = \arg_{X \in U} \max f(X) \quad (1)$$

where X^* is the set of optimal hyperparameters and U is the feasible search space.

The BOA was used to optimize the hyperparameters and is summarized as follows:

1. Initialize the hyperparameters randomly from their feasible space and evaluate them in the true objective function.
2. Build a surrogate model of the objective function/model $f(X)$ based on the initial hyperparameters using a Gaussian process.
3. Estimate the next hyperparameters based on a Gaussian process by optimizing an acquisition function.
4. Update the surrogate model with new hyperparameters.
5. Repeat steps 2–4 for N iterations.

In this study, the expected improvement (EI) acquisition function (Eq. (2)) is applied to select samples that are expected to have an improvement over the present best observation.

$$EI(X_i) = \begin{cases} (\mu(X_i) - f(X^*))\Phi\left(\frac{\mu(X_i)-f(X^*)}{\sigma(X_i)}\right) + \sigma(X)\phi\left(\frac{\mu(X_i)-f(X^*)}{\sigma(X_i)}\right), & \text{if } \sigma(X_i) > 0 \\ 0, & \text{if } \sigma(X_i) = 0 \end{cases} \quad (2)$$

where X^* is the current selected hyperparameters; Φ and ϕ are the cumulative distribution and probability density functions of $\frac{\mu(X_i)-f(X^*)}{\sigma(X_i)}$, respectively; and $\mu(X)$ and $\sigma(X)$ are the expected prediction and variance, respectively. Further details on the BOA can be found in Shahriari et al. (2015)

2.3 Summary of the DL Modeling Framework

The methodology for dam outflow prediction implemented in this study follows the schematic outlined in Fig. 1:

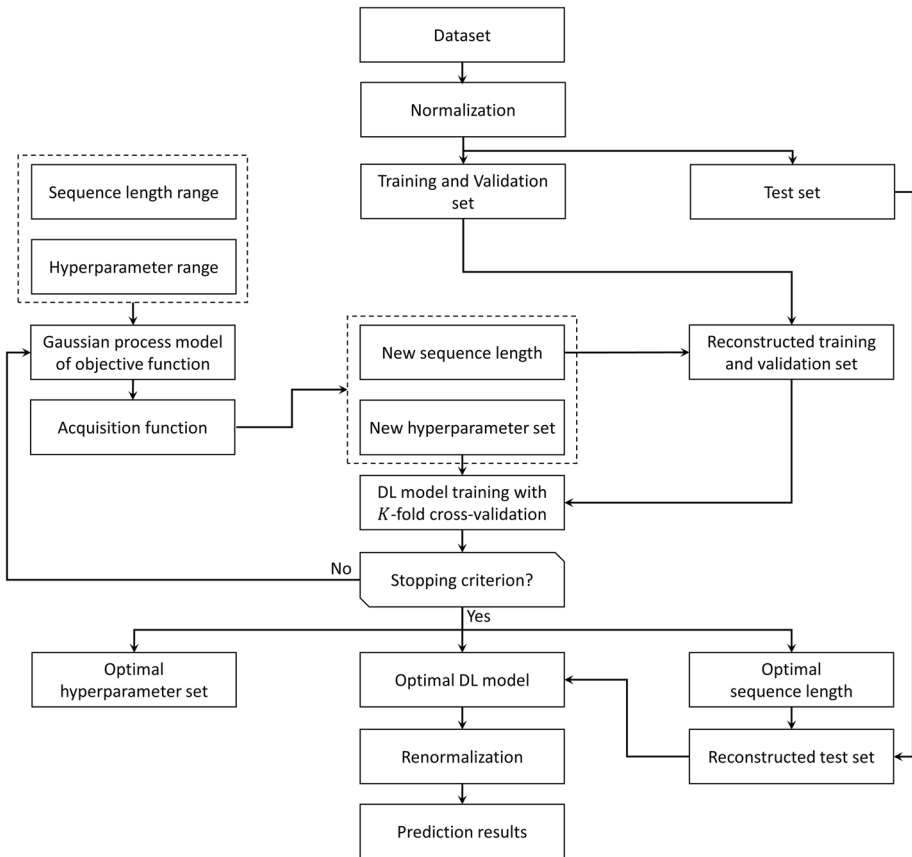


Fig. 1 Overview of a modeling framework for dam outflow prediction using Deep learning and Bayesian Optimization algorithm to simultaneously optimize hyperparameters, principal input variables and their sequence lengths for the DL models

1. Collect the dataset, including the target output (i.e., dam outflow) and candidate input variables (i.e., dam outflow, dam inflow, water level, and precipitation). Any inappropriate or missing values in the collected data should be reviewed carefully.
2. Then, the dataset is normalized to values between 0 and 1 and divided into two sets, including “training and validation” and “test” sets. In this study, we set 80% and 20% of the total data length for training, validation, and testing. The training and validation set was used for searching for the optimal set of model input predictors and hyperparameters and the testing set was used to evaluate the performance of the trained model with optimal inputs and hyperparameters.
3. Given the range of hyperparameters and the training and validation set, BOA and K-fold cross-validation are used to optimize the hyperparameters. A K-fold of 10 is selected as preferred in previous studies (Jung et al. 2020; Singh and Panda 2011; Yadav and Shukla 2016). Specifically, the training and validation set is partitioned into $K = 10$ distinct, equitable subsets, or “folds.” Then, an DL model is trained on $K - 1$ folds of the data and subsequently validated on the leftover fold. This approach is cycled K times with K models, with each fold being used in turn as the validation dataset.

At each iteration, different sequence lengths for the candidate input variables are generated and can be used to reconstruct the training and validation set to train and validate the DL model. For the stopping criteria of the hyperparameter optimization, we fixed the number of iterations in the BOA to 100. That is, the optimization of hyperparameters is stopped once the number of BOA iterations reaches 100. The results of this step are the optimal values for five hyperparameters, the optimal sequence lengths of four candidate input predictors, and an optimal DL model for outflow prediction. This number (100) was ad hoc selected to prove the feasibility of the proposed framework and based on the experimental design for the BOA of Snoek et al. (2012)

4. The model input in the test set is reconstructed using the optimal sequence lengths and used as input to the optimal DL models. Then, the model results are renormalized and evaluated with observed dam outflow.

2.4 Study Area and Dataset

This study considers two case studies, including Buon Tua Srah and Hua Na dams located in the North Central and South Central regions of Vietnam, respectively (Fig. 2a, b). These two dams belong to two intercountry river systems, namely, the Srepok (Fig. 2a) and Chu-Ma (Fig. 2b) river systems, with controlled areas of 2930 and 5345 km², respectively. Both Buon Tua Srah and Hua Na are multipurpose reservoirs that have roles in generating electricity, controlling flood downstream, supplying water for irrigation, and regulating against drought. The mean annual discharge, design flood discharge, normal water level, and total storage of Buon Tua Srah and Hua Na dams are 102 and 94.63 m³/s, 4267 and 5703 m³/s, 487.5 and 240 m, and 786.9 and 569.36 × 10⁶ m³, respectively. Buon Tua Srah and Hua Na dams started operating in 2011 and 2013, respectively.

In this study, the data used to forecast dam outflow included the previous dam outflow, dam inflow, water level, and precipitation. These data are favored by most relevant studies (Gutenson et al. 2020; Han et al. 2020; Zhang et al. 2018, 2019). The dam operation data were obtained from the official website of the Vietnam Electricity Corporation (<https://hochuathuydien.evn.com.vn>). The dataset spans a period of approximately 9 (01/01/2012–12/31/2020) and 5 (12/01/2015–12/31/2020) years for the Buon Tua Srah and Hua Da dams, respectively. The data were partitioned into two sets: 80% and 20% were

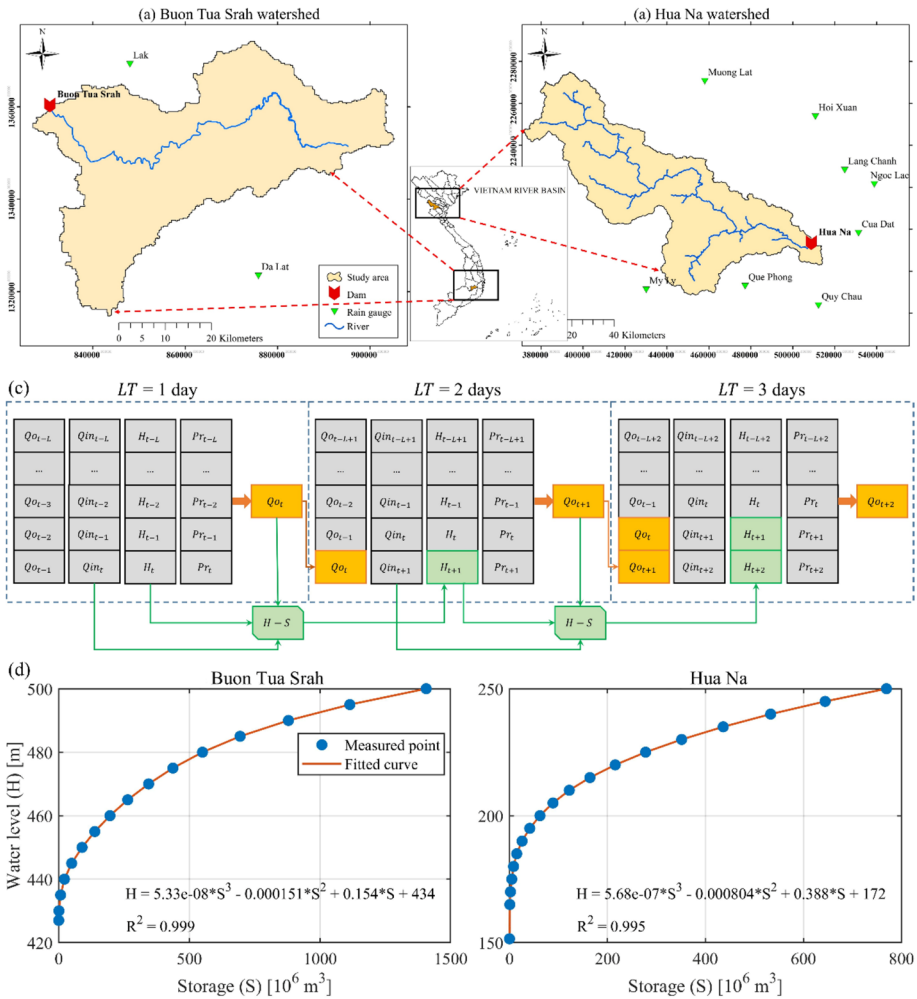


Fig. 2 Location of the two case studies: **a** Buon Tua Srah and **b** Hua Na watersheds located in Central Vietnam. Subplot **c** illustrates a recursive procedure for multi-step-ahead prediction. The prediction results are used continuously as input predictors to predict the next-step-ahead target outputs. Q_o , Q_{in} , H , and Pr represent dam outflow, dam inflow, dam water level, and precipitation, respectively. Gray boxes denote input predictors to the DL model, while yellow boxes denote DL model's output. Green box denotes a correlation function between water level (H) and reservoir storage (S). This function is used to compute H for the next time step based on the water balance. Subplot **d** shows the fitted curves and fitted equations representing the relationship between water level (H) and reservoir storage (S) for the two reservoirs using the third-degree polynomial functions

allocated to the training and validation and testing sets, respectively. The dam operation dataset included the daily inflow and outflow of the reservoir and the water level upstream of the dam. Daily precipitation data were provided by the National Center for Hydrometeorological Forecasting, Vietnam Meteorological, and Hydrological Administration (<http://www.nchmf.gov.vn>). Precipitation data were obtained from two and eight rain gauges near the study areas of Buon Tua Srah and Hua Na dams, respectively (Fig. 2a, b).

Although the Buon Tua Srah and Hua Na reservoirs were commissioned in 2011 and 2013, respectively, data acquisition and storage systems were not established until 2012 and 2015, respectively. As a result, the data used to train the model (especially for the Hua Na reservoir) are limited. Clearly, using a short data series affected the performance of the model. For DL approaches, 5 years of daily data is considered sufficient to apply such models, as has been demonstrated in a previous study (Tang et al. 2023). Therefore, in this study, a 5-year daily data series can be considered suitable for applying DL methods.

2.5 Model Configurations

2.5.1 Hyperparameter Setting

In this study, nine hyperparameters were optimized for three DL models using BOA. Four hyperparameters have a value range from 0 to 30 and denote the sequence lengths for four candidate inputs, including dam outflow (Qo), dam inflow (Qin), water level (H), and precipitation (Pr). The remaining five hyperparameters are for DL configurations, including the numbers of hidden layers (N_L), hidden units (N_U), and epochs (N_E); dropout rate (N_D); and batch size (N_B). The value ranges of these five hyperparameters were [1–3], [64–256], [10–300], [0–1], and [64–512]. Additionally, three benchmarking DL models were built with fixed hyperparameters that were preferred in previous studies, $N_L = 1$; $N_U = 256$; $N_E = 30$; $N_D = 0.4$; $N_B = 512$ (Frame et al. 2021; Kratzert et al. 2018, 2019). These benchmarking models were used to evaluate whether the optimized DL models performed well in forecasting the dam outflows.

2.5.2 Modeling Setup for Multistep-Ahead Outflow Prediction

To predict multistep-ahead (1–6 days ahead) dam outflow, we adopted a recursive procedure to perform the simulation from all models, as shown in Fig. 2c. Specifically, for 1-day-ahead prediction, the observed Qo at $t-1$, Qin , H , and Pr at t will be used to predict Qo at t . For longer day-ahead predictions, the previously predicted Qo will be used to predict Qo at the next time step, and the input H for the next step prediction will be updated using the water level (H)–reservoir storage (S) curve represented by Eqs. (3) and (4) and equations in Fig. 2d.

$$H_{t+1} = g(S_{t+1}) \quad (3)$$

$$S_{t+1} = S_t + (Qin_t - Qo_t) \times 86400 \quad (4)$$

where g denotes the ‘relation equation’ between H and S detailed in Fig. 2d. The relation equation was formed based the third-degree polynomial function. Equation (4) is in the form of a water balance equation that can be used to calculate the reservoir storage (in cubic meters) for the next step (the next day) based on the current reservoir storage and dam inflow and outflow.

2.6 Evaluation Metrics

To assess the modeling performance, the accuracy metrics Nash–Sutcliffe efficiency (NSE), root mean square error (RMSE), and Kling–Gupta efficiency (KGE) were chosen.

NSE is traditionally used to evaluate the accuracy and power of deterministic models (Pushpalatha et al. 2012). RMSE is one of the most commonly used measures for evaluating the quality of predictions. It shows how far the predictions fall from the true measured values using the Euclidean distance. KGE provides a diagnostically interesting decomposition of the NSE (and thus the mean square error), which facilitates the analysis of the relative importance of its different components (correlation, bias, and variability). The formulas for NSE, RMSE, and KGE can be found in Section S.3 of the SM.

Additionally, an attempt was made to evaluate the prediction results more comprehensively by analyzing 11 hydrological signatures based on observations and simulations from three DL models. This investigation can be used to confirm the effectiveness of the model in providing simulations that accurately represent hydrological characteristics and assess the physical understandability of each DL model. Hydrological signatures are specific characteristics or metrics used to describe and quantify various aspects of hydrological processes and conditions in watersheds, rivers, or other water-related systems (McMillan 2020). These signatures are valuable for understanding and analyzing the behavior of water resources and the effects of environmental changes, including climate variability, land use, and human activities. Hydrological signatures represent various characteristics of hydrological time series, including magnitude, timing, frequency, duration, and rate of change. Eleven hydrological signatures were selected (McMillan 2020), including base-flow index (BFI), flow autocorrelation (QAC), overall flow variability (QCV), high-flow event duration (QHD), high-flow event frequency (QHF), high-flow variability (QHV), low-flow event duration (QLD), low-flow event frequency (QLF) (Pushpalatha et al. 2011), low-flow variability (QLV), mean flow (QMEAN), and slope of the normalized flow duration curve (SFDC).

3 Results

3.1 Optimization of the Principal Inputs and Hyperparameters

Hyperparameters and input predictors must be predetermined to construct DL models; however, their optimal values to maximize the performance of the trained model are unknown. Here, the results of the BOA scheme are presented, which can be a guideline for other studies to simultaneously tune the hyperparameters and select the sequence lengths of candidate input variables. The results that signify the convergence criteria (RMSE) used for determining the performance of models are shown in Fig. 3a, b. Figure 4 presents the optimal values of five hyperparameters and sequence lengths of input variables. These results are subject to variation depending on the DL proposed but are significantly different between case studies.

Figure 3a, b shows that the RMSE for each of the three models decreased as the number of iterations increased and changed very slightly when the number of iterations was larger than 25. In other words, if a larger number of iterations is used for optimization, the overall accuracy increases; however, at a certain point, the RMSE becomes stably. Figure 3a, b confirmed that the hyperparameters and sequence lengths of input variables determined with 25 iterations are suitable for training three DL models. Additionally, GRU and LSTM outperformed BiLSTM in providing lower RMSE values for two case studies. Specifically, at iteration of 25, the RMSEs obtained using GRU and LSTM are lower than those using BiLSTM approximately 2 and 3 times for the case studies of

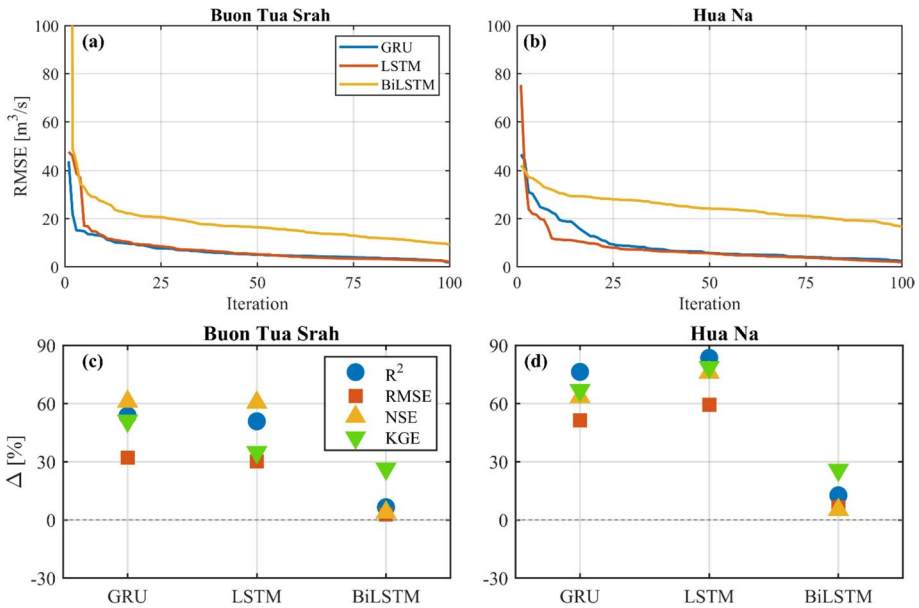


Fig. 3 **a** and **b** present Trace plots of model performances (i.e., RMSE) for two case studies of the three DL models (GRU, LSTM, and BiLSTM) trained by BOA with 100 iterations. **c** and **d** show results of a percentage 'difference' metric (Δ) in Eq. (5) computed for four evaluation metrics between three optimal DL models and three benchmark DL models

Buon Tua Srah and Hua Na, respectively. For both case studies, the RMSEs of GRU and LSTM with 25 iterations are equal to or even lower than those of BiLSTM with N of 100. GRU and LSTM can build an efficient model with an accurate degree even if they use a BOA iteration that is four times smaller than that of BiLSTM.

The optimal results of the hyperparameters and sequence lengths of the input variables are shown in Fig. 4. Generally, the optimal results vary depending on the model type and specific case study. Specifically, the optimal hyperparameters of GRU and LSTM seem to be more similar when compared with those of BiLSTM (Fig. 4a–e). For example, for both case studies, GRU and LSTM involve two layers, whereas BiLSTM needs one layer; the number of epochs for both GRU and LSTM is also higher than that for BiLSTM (i.e., ~ 255 – 280 versus 100 for Buon Tua Srah and 280 – 300 versus 200 for Hua Na); the dropout rates for GRU and LSTM are smaller than 10^{-3} and 5×10^{-3} , respectively, for Buon Tua Srah and Hua Na, whereas for BiLSTM, the N_D that is higher than 10^{-2} is required. Regarding the optimal input predictors for the three models, Fig. 4f–i shows that the optimal results are less similar between models. This result confirms that the selection of the input variables and the determination of their sequence lengths must be optimized concurrently with the corresponding model configuration and independently for each different model type. In previous studies, the input variables and their sequence lengths were selected mainly from statistical analysis methods. Then, a trial-and-error method or an optimization procedure is conducted to obtain the hyperparameters (model configuration) of a DL model. These procedures do not assure an optimal model because changing the structure afterward will dramatically affect the performance of the model and make the previously optimized input dataset no longer optimal.

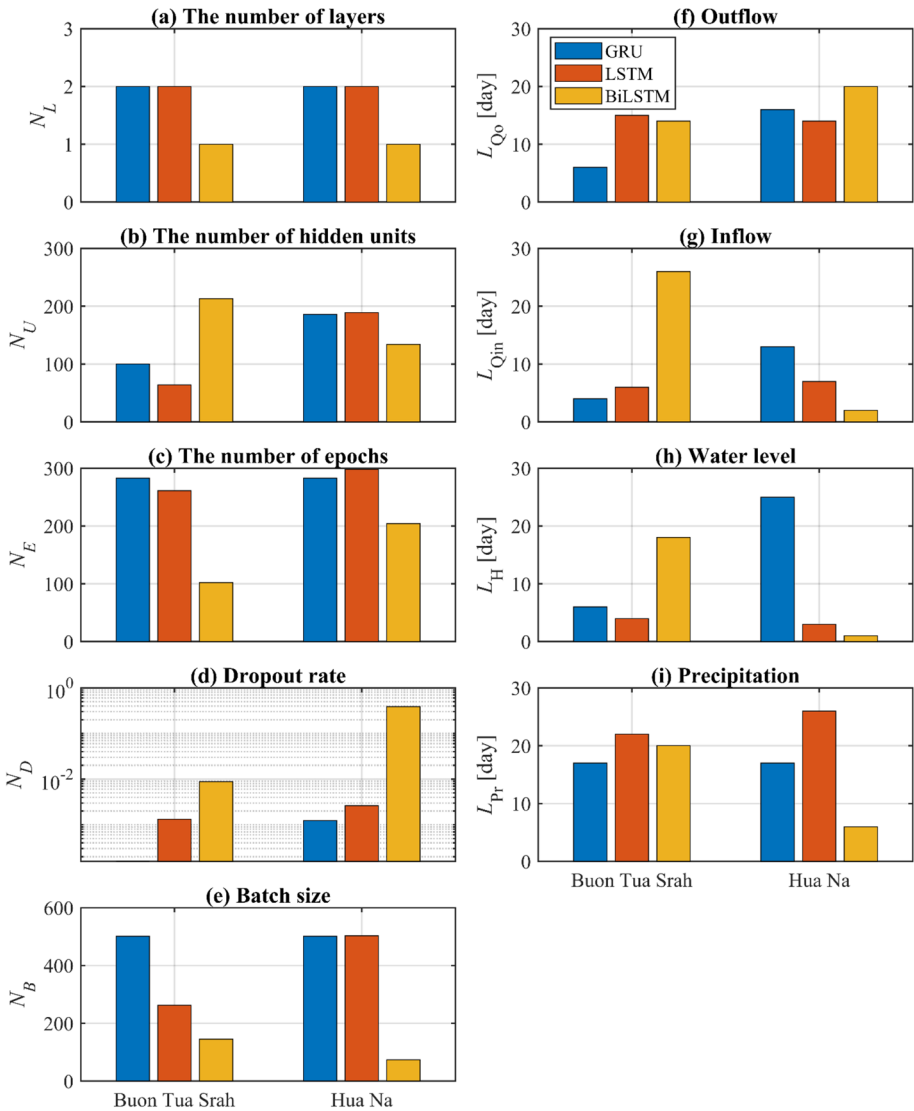


Fig. 4 Results of the optimal hyperparameters and principal input predictors for the three DL models (GRU, LSTM, and BiLSTM) for two case studies

To compare the degree of performance deterioration between optimal DL and benchmarking models, another percentage “difference” metric (Δ) is computed as

$$\Delta = - \frac{|\text{Metric}_{\text{BOA}} - \text{Metric}_{\text{ideal}}| - |\text{Metric}_{\text{Bench}} - \text{Metric}_{\text{ideal}}|}{|\text{Metric}_{\text{Bench}} - \text{Metric}_{\text{ideal}}|} \times 100 \quad (5)$$

where $\text{Metric}_{\text{BOA}}$ and $\text{Metric}_{\text{Bench}}$ denote the evaluation metrics (R^2 , RMSE, NSE, and KGE) of the optimal and benchmarking DL models. $\text{Metric}_{\text{ideal}}$ represents the ideal (perfect) values of the metrics of R^2 , RMSE, NSE, and KGE, that is 1, 0, 1, and 1, respectively.

The positive (or negative) values of Δ indicate that the prediction results of the optimal DL model are more (or less) accurate than those computed using the benchmarking model. The results of Δ for the comparisons between optimal and benchmarking DL models are illustrated in Fig. 3c, d. First, the results of Δ between all comparison pairs are mostly positive, revealing that the optimal DL models outperform the benchmarking models by up to 60% and 90% for both Buon Tua Srah and Hua Na case studies, respectively. For both case studies, the four metrics of GRU and LSTM perform better than those of their benchmarking models by up to 30%–60% and 50%–90%, respectively. The results obtained from the optimal BiLSTM model are more accurate than those obtained from the benchmarking BiLSTM model at approximately 0%–30% for four metrics over both case studies. In summary, the optimal DL models using the proposed approach have proven to be superior to the benchmarking DL models in providing accurate forecasting results.

3.2 Dam Outflow Predictions

Three DL models trained using optimal hyperparameters and input predictors were applied to the test set to predict 1- to 6-day-ahead outflows of the Buon Tua Srah and Hua Da reservoirs. The multistep-ahead prediction scheme is described in Section 2.5.2. Overall, the performance of three DLs are different regarding the lead times and case studies. In this section, the prediction skills of the three DL models are comparatively analyzed, and conclusions are drawn from the following two perspectives: the predictive performances with different lead times and the ability to replicate the hydrological signatures of the DL models.

3.2.1 Predictability Skills According to Lead-Time Predictions

The predictions of two dam outflows with two different lead times of 1 and 6 days are presented in Fig. 5. As expected, the forecasting performances of the three models decrease with increasing lead times. It is noted that the increasing of prediction error for longer time ahead is inevitable. Specifically, in the case study of Buon Tua Srah, the ranges of degradation of RMSE, NSE, and KGE reported in Fig. 6 at a lead time of 6 days compared with those at a lead time of 1 day are approximately 2 to 3, 6–9, and 6–12 times, respectively. In the case study of Buon Tua Srah, these ranges are 2 to 3, 2 to 3, and 3 to 4 times. Interestingly, three DL models show comparable results for 1-day predictions with consistent hydrograph patterns and with the R^2 values that are higher than 0.8 (Fig. 5). Performance differences between models are evident with longer lead times predictions.

Comparing the simulated hydrographs with observations, especially for long lead-time predictions, the overall variation and magnitude of the predicted outflow using BiLSTM agree more closely with observations than the results produced by other models for the case study of Buon Tua Srah. Conversely, for Hua Na outflow, GRU outperforms both LSTM and BiLSTM. Quantitatively, at a lead-time prediction of 6 days, BiLSTM has an R^2 of 0.48 for the Buon Tua Srah case study, which is higher than the R^2 of 0.31 and 0.11 produced by GRU and LSTM, respectively. Conversely, for the Hua Na case study, GRU has an R^2 of 0.33, which is higher than the R^2 of 0.26 and 0.18 produced by LSTM and BiLSTM, respectively. The results produced by RMSE, NSE, and KGE reported in Fig. 6 confirm that the predictions from BiLSTM and GRU are closest to the observations for Buon Tua Srah and Hua Na case studies, respectively. For the first case study, BiLSTM has RMSE, NSE, and KGE of $29 \text{ m}^3/\text{s}$, -0.15 , and 0.73 , respectively; all metrics

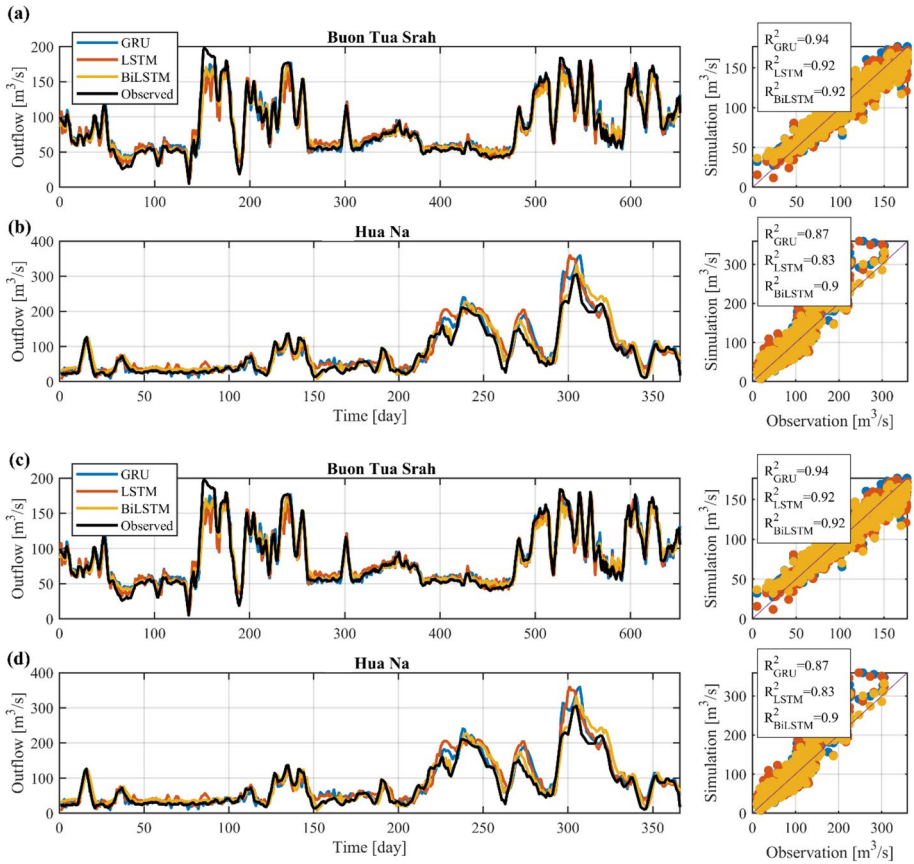


Fig. 5 Comparisons of outflow predictions at $LT = 1$ -day (a-b) and 6-day (c-d) of three DL models (GRU, LSTM, and BiLSTM) with observations (black lines) for two case studies using test set (20% of total data)

were significantly improved to an RMSE of 35 and 40 m^3/s , NSE of -0.14 and -1.2 , and KGE of 0.65 and 0.47 obtained from GRU and LSTM, respectively (Fig. 6a). Conversely, for the second case study of Hua Na dam, the prediction results of GRU are more accurate (approximately 5% and 8% of RMSE, 20% and 23% of NSE, and 23% and 25% of GRE) compared to those of LSTM and BiLSTM, respectively.

3.2.2 Predictability Skills According to Hydrological Signature Replication

Evaluation metrics such as RMSE, NSE, and KGE are applied to assess the general trend and similarity of the forecast results with observed data, but they fail to describe the hydrological characteristics. In optimizing the operation of dams, one of the important objectives is to retain the basic hydrological signatures in relation to the natural environment and aquatic ecosystem. In this study, 11 important and well-known hydrological signatures suggested by McMillan (2020) were used for a standard assessment of the ability of the DL model to replicate hydrological characteristics. These 11 signatures represent the characteristics of streamflow, including magnitude, timing, frequency, duration, and rate of change.

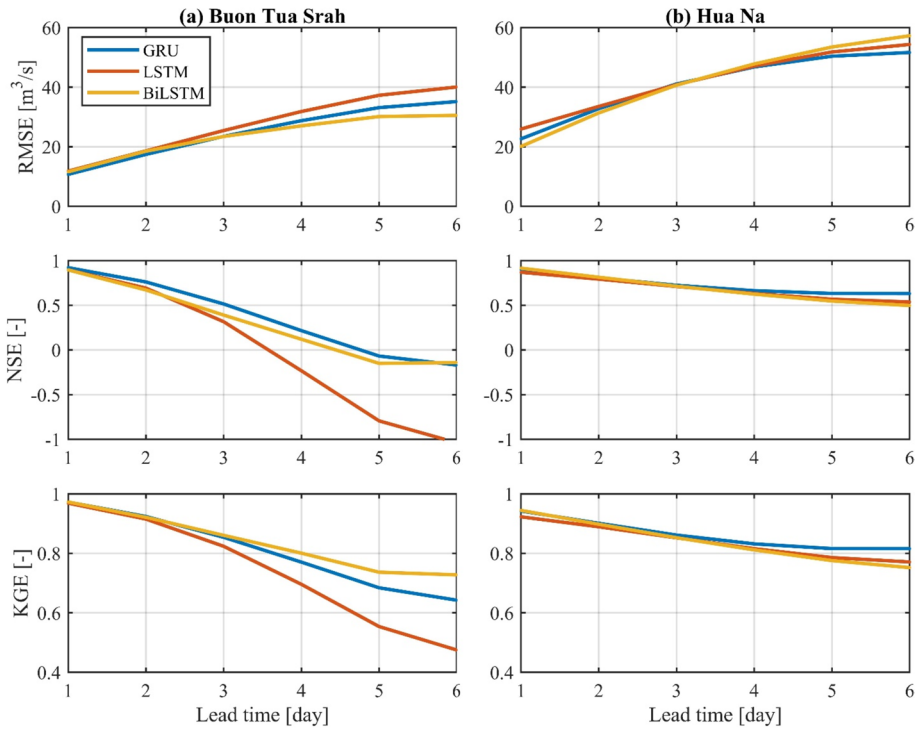


Fig. 6 Evaluation metrics (RMSE, NSE, and KGE) of three DL models for 1 to 6-day ahead outflow predictions of the two case studies

The outflow simulation results of reservoirs with lead times of 1 and 6 days are used to calculate 11 hydrological signatures and are compared with those computed from observations. The detailed results of 11 signatures are shown in Tables S.1 and S.2 in the SM file. The results of evaluating the similarity and difference of these signatures between simulations and observations are shown in Fig. 7 via a relative difference (RD) metric computed as follows:

$$RD = \frac{|HS_{SIM} - HS_{OBS}|}{HS_{OBS}} \times 100 \tag{6}$$

where HS_{SIM} and HS_{OBS} denote the hydrological signatures computed from the simulations of three models and observations, respectively. The ideal value of RD is 0, which denotes the similarity between results from DL models and observations.

The RD results reported in Fig. 7 show that the simulation results of three models with a lead time of 1 day can replicate nine hydrological signatures quite well with RDs that are mostly close to 0 and <20% for both case studies, except for QLF and QLD. Conversely, for 6 days of lead-time predictions, due to the less accurate predictions as mentioned in Section 3.2.1, the hydrological signatures compared with those computed from observations are less similar with larger RD values, e.g., QCV, QHD, SFDC, QLF, and QLD. For both lead-time predictions, the simulation results of three models can present six hydrological signatures that are close to those from the observations, with RDs <5%, including

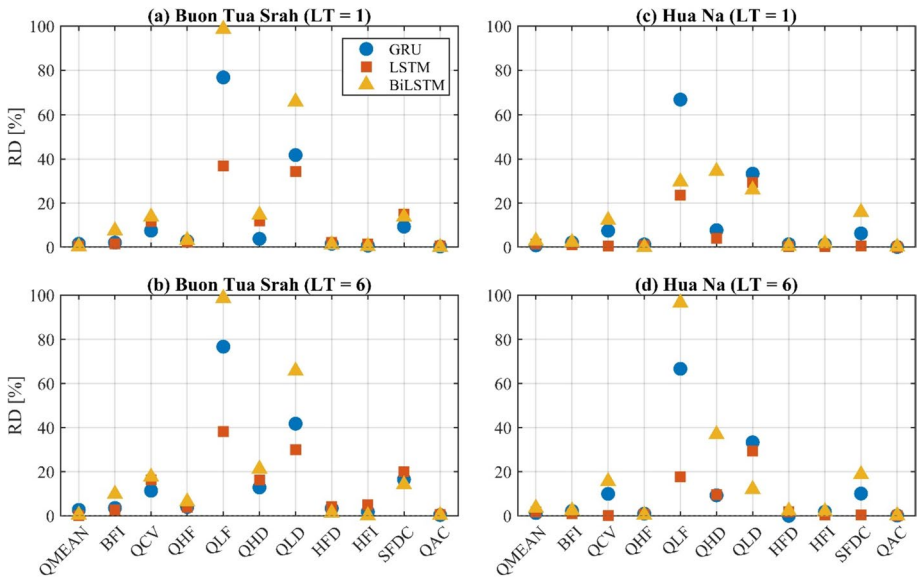


Fig. 7 Relative difference of 11 hydrological signatures computed from observation and simulations of GRU, LSTM, and BiLSTM with one (a-c) and six (b-d) -day ahead predictions over the test set for two case studies

QMEAN, BFI, QHF, HFD, HFI, and QAC, whereas the ability to replicate QLF and QLD is the worst with RDs that varied between 20 and 100%. These two signatures represent the frequency and duration of low flows that are greatly influenced by the dam operating regimes and are extremely elusive with various uncertainties.

Interestingly, different from Section 3.2.1, where GRU was concluded to be superior to both LSTM and BiLSTM, here LSTM provides more accurate hydrological signatures than GRU and BiLSTM. Specifically, for Buon Tua Srah dam, the RD values of QLF and QLD from LSTM are smaller than those computed from GRU and BiLSTM, i.e., 40% versus 78% and 99% (for QLF) and 30% versus 41% and 66% (for QLD) (Fig. 7a). Conversely, for the Hua Na case study, for most indicators, the RD values of QVC, QLF, and SFDC from LSTM are significantly smaller than those computed from GRU and BiLSTM, and other indicators have almost comparable values, except for QLD. The aforementioned results highlight that using hydrological signatures as evaluation metrics can serve as an effective approach for selecting appropriate DL models tailored to specific objectives. The findings in Section 3.2 demonstrate that the optimal model is contingent on the specific case study and intended model application, for example, to achieve high overall accuracy or reliably capture pertinent hydrological characteristics.

4 Discussions

4.1 Is the Proposed Framework Necessary in Constructing the DL Model?

This study proposes an optimization framework that uses the BOA to determine the optimal inputs and hyperparameters of DL models. The framework and investigation of model performance in Section 3 revealed that optimizing the inputs and hyperparameters of DL models separately is inappropriate. Consequently, the imperative of this proposed framework is underscored. First, it provides a global optimization solution that considers all interactions between potential inputs and hyperparameters to build the best DL model instead of just examining each component as in previous studies (Ali-zadeh et al. 2021; Tran et al. 2021; Zhang et al. 2018). Second, the proposed framework improves efficiency using the BOA to help the optimization converge faster than that using trial-and-error or grid search methods. Additionally, it reduces the amount of work required by eliminating the need for tasks such as data correlation analysis and providing criteria for selecting inputs.

Additionally, this automatic optimization framework can solve the difficulty of selecting inputs when the data have low correlation. Correlation analysis (e.g., cross-correlation function, CCF, or partial autocorrelation function) is often preferred to identify the most correlated data that are selected for the input of data-driven models (Ahmad and Hossain 2019; Tran et al. 2021; Yang et al. 2017a, b). However, there are many candidate input predictors that have very low linear correlations with the target data. It is difficult for modelers to choose the right one, for example, water level or precipitation in Fig. S.1 in the SM file with a CCF factor < 0.2 . The low correlation does not mean there is no correlation, and nonlinear and nonmonotonic relationships are hardly detected with available statistical techniques, especially for dam outflow application (Altman and Krzywinski 2015; Goodwin and Leech 2006). The proposed framework eliminates correlation analysis steps, and all candidate inputs can be fed into the model and optimized through the BOA. The optimized inputs will be based on the performance of the trained model and not on the correlation between the inputs and target outputs, like in traditional analysis methods.

4.2 Challenges of DL Applications for Dam Outflow Prediction

Although the implementation and analysis of experiments are valid for the presented scope of the experimental design, modelers must proceed with caution when this approach is extended to more case studies of the dam outflow prediction. In this section, we discuss the challenges that should be addressed in DL applications related to dam outflow forecasting in future studies. A primary concern is the growing anthropogenic influence on dam operations, which proves more arduous to comprehend and foresee than natural hydrological forcings, particularly under extreme conditions such as flooding or drought. While previous studies have demonstrated superior performance of data-driven approaches, including DL, compared to conventional non-data-driven methods (Gutenson et al. 2020; Zhang et al. 2018), reservations persist regarding the efficacy of DLs in furnishing realistic forecasts under the aforementioned conditions. This concern would be ameliorated given sufficient data pertaining to dam operations for model training purposes. However, the perennial issue of data paucity and limited data sharing in reservoir operations persists, attributable to numerous constraints of a political nature.

Secondly, a well-established characteristic and challenge of DL and data-driven models is the inability to extrapolate beyond the domain encompassed by the training data (Frame et al. 2021; Kratzert et al. 2019; Tran et al. 2023a, b; Zhao et al. 2019). Fundamentally, DLs possess uniqueness to the particular training data space. Theoretically, this limitation would be surmounted given sufficient training data encompassing even rare extreme events. However, comprehensively collecting observational data of exceptional phenomena presents difficulties. Recently, Tran and Kim (2022) proposed three strategies to augment the predictive capability for extreme events where adequate data is lacking and such events deviate substantially from the training distribution. Firstly, high-fidelity samples informed by physical relationships or operating guidelines contingent on relevant factors should be leveraged to ensure robust learning when training samples are sparse. Data generated per governing equations codifying dam operating rules can potentiate physical process comprehension in deep learning. Secondly, extrapolation aptitude may be enhanced by expanding the prediction space through incorporating input noise and parameter uncertainty. Finally, hybrid models combining deep learning with techniques exhibiting extrapolation capabilities warrant exploration.

Finally, it is certain that uncertainties intrinsically persist within all predictions, including DL-based dam outflow predictions. Despite substantial progress in DL for hydrological modeling, prediction uncertainties from DL architectures have garnered significant attention in contemporary literature (Fang et al. 2020; Kasiviswanathan and Sudheer 2012; Srivastav et al. 2007; Tran et al. 2023). These uncertainties primarily stem from learnable model parameters and inputs (Fang et al. 2020). A prevalent technique to represent input uncertainty involves injecting noise adhering to prescribed distributions, generating an ensemble of perturbed inputs to derive ensemble predictions (Fang et al. 2020; Tran and Kim 2022). Conversely, Monte Carlo dropout is preferred for evaluating uncertainties from learnable parameters, randomly omitting neural network units to construct an ensemble of models with diverse parameters, also used for ensemble prediction (Gal and Ghahramani 2016).

5 Conclusions

This study investigated the efficacy of three deep learning architectures for daily discharge prediction at the Buon Tua Srah and Hua Na dams in Vietnam. The deep learning models were coupled with Bayesian optimization to enable efficient hyperparameter tuning and input variable selection. Notably, Bayesian optimization simultaneously optimized five hyperparameters, input variables, and sequence lengths, expediting model training. The key conclusions regarding the utility of Bayesian optimization and performance of the deep learning models are summarized as follows.

An optimization framework based on Bayesian optimization with Gaussian processes was proposed to concurrently optimize hyperparameters, input variables, and sequence lengths for the deep learning models. This framework holistically accounts for interactions between potential inputs and deep learning architectures, as parameterized by hyperparameters, thereby determining the optimal input variables, lags, and hyperparameters. Compact objective function values were achieved, circumventing discrete optimization of individual factors as in prior works and obviating exhaustive trial-and-error. Moreover, the framework automatically selects input variables and lags from the provided candidate set, absolving manual data analysis and input screening.

A comprehensive assessment of three deep learning architectures (GRU, LSTM, and BiLSTM) was conducted for multi-step dam discharge prediction. Overall, the results demonstrated that all models furnish accurate simulations, corroborating the capability for multi-step ahead forecasting. However, model rankings depended on performance metrics and case studies. The BiLSTM and GRU models achieved the lowest RMSE, NSE, and KGE for the Buon Tua Srah and Hua Na dams, respectively. However, the LSTM replicated the most hydrological signatures accurately for both dams, underscoring the need to consider modeling objectives during model selection.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11269-023-03677-9>.

Acknowledgements This work was mainly supported by the Ministry of Science and Technology of Vietnam through the project No. NĐT.58.RU/19.

Funding Ministry of Science and Technology, NĐT.58.RU/19, Tran Ngoc Anh.

Data Availability The data that support the findings of this study are available from the corresponding author, [N.T.G], upon reasonable request.

References

- 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>
- Ahmad SK, Hossain F (2019) A generic data-driven technique for forecasting of reservoir inflow: Application for hydropower maximization. *Environ Model Softw* 119:147–165. <https://doi.org/10.1016/j.envsoft.2019.06.008>
- Aksoy H, Dahamsheh A (2018) Markov chain-incorporated and synthetic data-supported conditional artificial neural network models for forecasting monthly precipitation in arid regions. *J Hydrol* 562:758–779. <https://doi.org/10.1016/j.jhydrol.2018.05.030>
- Alizadeh B, Ghaderi Bafti A, Kamangir H, Zhang Y, Wright DB, Franz KJ (2021) A novel attention-based LSTM cell post-processor coupled with bayesian optimization for streamflow prediction. *J Hydrol* 601:126526. <https://doi.org/10.1016/j.jhydrol.2021.126526>
- Altman N, Krzywinski M (2015) Points of significance: Association, correlation and causation. *Nat Methods* 12(10)
- Beiranvand B, Ashofteh P-S (2023) A systematic review of optimization of dams reservoir operation using the meta-heuristic algorithms. *Water Resour Manag* 37(9):3457–3526. <https://doi.org/10.1007/s11269-023-03510-3>
- Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. *J Mach Learn Res* 13(1):281–305
- Bowden GJ, Maier HR, Dandy GC (2005) Input determination for neural network models in water resources applications. Part 2. Case study: forecasting salinity in a river. *J Hydrol* 301(1–4):93–107. <https://doi.org/10.1016/j.jhydrol.2004.06.020>
- Bozorg-Haddad O, Zarezadeh-Mehrizi M, Abdi-Dehkordi M, Loáiciga HA, Mariño MA (2016) A self-tuning ANN model for simulation and forecasting of surface flows. *Water Resour Manag* 30(9):2907–2929. <https://doi.org/10.1007/s11269-016-1301-2>
- Cho K, Van Merriënboer B, Bahdanau D, Bengio Y (2014) On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*
- Coerver HM, Rutten MM, van de Giesen NC (2018) Deduction of reservoir operating rules for application in global hydrological models. *Hydrol Earth Syst Sci* 22(1):831–851. <https://doi.org/10.5194/hess-22-831-2018>
- Döll P, Fiedler K, Zhang J (2009) Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. *Hydrol Earth Syst Sci* 13(12):2413–2432. <https://doi.org/10.5194/hess-13-2413-2009>

- Ehsani N, Fekete BM, Vörösmarty CJ, Tessler ZD (2016) A neural network based general reservoir operation scheme. *Stoch Env Res Risk Assess* 30(4):1151–1166. <https://doi.org/10.1007/s00477-015-1147-9>
- El-Shafie A, Taha MR, Noureldin A (2006) A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam. *Water Resour Manag* 21(3):533–556. <https://doi.org/10.1007/s11269-006-9027-1>
- Fang K, Kifer D, Lawson K, Shen C (2020) Evaluating the potential and challenges of an uncertainty quantification method for long short-term memory models for soil moisture predictions. *Water Resour Res*. <https://doi.org/10.1029/2020wr028095>
- Frame JM, Kratzert F, Klotz D, Gauch M, Shalev G, Gilon O, Qualls LM, Gupta HV, Nearing GS (2022) Deep learning rainfall–runoff predictions of extreme events. *Hydrol Earth Syst Sci* 26:3377–3392. <https://doi.org/10.5194/hess-26-3377-2022>
- Gal Y, Ghahramani Z (2016) Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning* (pp. 1050–1059). PMLR
- Goodwin LD, Leech NL (2006) Understanding correlation: Factors that affect the size of r . *J Exp Educ* 74(3):249–266
- Graf WL (2006) Downstream hydrologic and geomorphic effects of large dams on American rivers. *Geomorphology* 79(3):336–360. <https://doi.org/10.1016/j.geomorph.2006.06.022>
- Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J (2017) LSTM: A search space odyssey. *IEEE Trans Neural Netw Learn Syst* 28(10):2222–2232. <https://doi.org/10.1109/TNNLS.2016.2582924>
- Gutenson JL, Tavakoly AA, Wahl MD, Follum ML (2020) Comparison of generalized non-data-driven lake and reservoir routing models for global-scale hydrologic forecasting of reservoir outflow at diurnal time steps. *Hydrol Earth Syst Sci* 24(5):2711–2729. <https://doi.org/10.5194/hess-24-2711-2020>
- Han Z, Long D, Huang Q, Li X, Zhao F, Wang J (2020) Improving reservoir outflow estimation for ungauged basins using satellite observations and a hydrological model. *Water Resour Res* 56(9). <https://doi.org/10.1029/2020wr027590>
- Hanasaki N, Kanae S, Oki T (2006) A reservoir operation scheme for global river routing models. *J Hydrol* 327(1–2):22–41. <https://doi.org/10.1016/j.jhydrol.2005.11.011>
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput*. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hu C, Wu Q, Li H, Jian S, Li N, Lou Z (2018) Deep learning with a long short-term memory networks approach for rainfall–runoff simulation. *Water* 10(11):1543. <https://doi.org/10.3390/w10111543>
- Jothiprakash V, Magar RB (2012) Multi-time-step ahead daily and hourly intermittent reservoir inflow prediction by artificial intelligent techniques using lumped and distributed data. *J Hydrol* 450–451:293–307. <https://doi.org/10.1016/j.jhydrol.2012.04.045>
- Jung K, Bae D-H, Um M-J, Kim S, Jeon S, Park D (2020) Evaluation of nitrate load estimations using neural networks and canonical correlation analysis with k-fold cross-validation. *Sustainability* 12(1):400
- Kasiviswanathan KS, Sudheer KP (2012) Quantification of the predictive uncertainty of artificial neural network based river flow forecast models. *Stoch Env Res Risk Assess* 27(1):137–146. <https://doi.org/10.1007/s00477-012-0600-2>
- Khosravi K, Golkarian A, Tiefenbacher JP (2022) Using optimized deep learning to predict daily streamflow: A comparison to common machine learning algorithms. *Water Resour Manag* 36(2):699–716. <https://doi.org/10.1007/s11269-021-03051-7>
- Kratzert F, Klotz D, Brenner C, Schulz K, Herrnegger M (2018) Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrol Earth Syst Sci* 22(11):6005–6022. <https://doi.org/10.5194/hess-22-6005-2018>
- Kratzert F, Klotz D, Shalev G, Klambauer G, Hochreiter S, Nearing G (2019) Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrol Earth Syst Sci* 23(12):5089–5110. <https://doi.org/10.5194/hess-23-5089-2019>
- Latif SD, Ahmed AN (2023) Streamflow prediction utilizing deep learning and machine learning algorithms for sustainable water supply management. *Water Resour Manag* 37(8):3227–3241. <https://doi.org/10.1007/s11269-023-03499-9>
- Le XH, Ho HV, Lee G, Jung S (2019) Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water* 11(7):1387. <https://doi.org/10.3390/w11071387>
- McMillan HK (2020) A review of hydrologic signatures and their applications. *WIREs Water* 8(1). <https://doi.org/10.1002/wat2.1499>
- Mohan S, Ramsundram N (2016) Predictive temporal data-mining approach for evolving knowledge based reservoir operation rules. *Water Resour Manag* 30(10):3315–3330. <https://doi.org/10.1007/s11269-016-1351-5>


- Mohandes MA, Halawani TO, Rehman S, Hussain AA (2004) Support vector machines for wind speed prediction. *Renew Energy* 29(6):939–947. <https://doi.org/10.1016/j.renene.2003.11.009>
- Ni L, Wang D, Singh VP, Wu J, Wang Y, Tao Y, Zhang J (2020) Streamflow and rainfall forecasting by two long short-term memory-based models. *J Hydrol* 583:124296. <https://doi.org/10.1016/j.jhydrol.2019.124296>
- Nourani V, Hosseini Baghanam A, Adamowski J, Kisi O (2014) Applications of hybrid wavelet–Artificial Intelligence models in hydrology: A review. *J Hydrol* 514:358–377. <https://doi.org/10.1016/j.jhydrol.2014.03.057>
- Pushpalatha R, Perrin C, Le Moine N, Mathevet T, Andréassian V (2011) A downward structural sensitivity analysis of hydrological models to improve low-flow simulation. *J Hydrol* 411(1):66–76. <https://doi.org/10.1016/j.jhydrol.2011.09.034>
- Pushpalatha R, Perrin C, Moine NL, Andréassian V (2012) A review of efficiency criteria suitable for evaluating low-flow simulations. *J Hydrol* 420–421:171–182. <https://doi.org/10.1016/j.jhydrol.2011.11.055>
- Salehinejad H, Sankar S, Barfett J, Colak E, Valaee S (2017) Recent advances in recurrent neural networks. *arXiv preprint arXiv:1801.01078*
- Sauhats A, Petrichenko R, Broka Z, Baltputnis K, Sobolevskis D (2016) ANN-based forecasting of hydro-power reservoir inflow. In 2016 57th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON) (pp. 1–6). IEEE. <https://doi.org/10.1109/rtucon.2016.7763129>
- Schuster M, Paliwal KK (1997) Bidirectional recurrent neural networks. *IEEE Trans Signal Process* 45(11):2673–2681
- Seo Y, Kim S, Kisi O, Singh VP (2015) Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *J Hydrol* 520:224–243. <https://doi.org/10.1016/j.jhydrol.2014.11.050>
- Shahriari B, Swersky K, Wang Z, Adams RP, De Freitas N (2015) Taking the human out of the loop: A review of Bayesian optimization. *Proc IEEE* 104(1):148–175
- Shi X, Chen Z, Wang H, Yeung D-Y, Wong W-K, Woo W-C (2015) Convolutional LSTM Network: A machine learning approach for precipitation nowcasting. *arXiv:1506.04214*. <https://ui.adsabs.harvard.edu/abs/2015arXiv150604214S> Retrieved from <https://ui.adsabs.harvard.edu/abs/2015arXiv150604214S>
- Singh G, Panda RK (2011) Daily sediment yield modeling with artificial neural network using 10-fold cross validation method: a small agricultural watershed, Kapgari, India. *Int J Earth Sci Eng* 4(6):443–450
- Snoek J, Larochelle H, Adams RP (2012) Practical Bayesian optimization of machine learning algorithms. *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2*, 2951–2959. <https://doi.org/10.5555/2999325.2999464>
- Srivastav RK, Sudheer KP, Chaubey I (2007) A simplified approach to quantifying predictive and parametric uncertainty in artificial neural network hydrologic models. *Water Resour Res* 43(10). <https://doi.org/10.1029/2006wr005352>
- Tang S, Sun F, Liu W, Wang H, Feng Y, Li Z (2023) Optimal postprocessing strategies with LSTM for global streamflow prediction in ungauged basins. *Water Resour Res* e2022WR034352
- Tran TD, Tran VN, Kim J (2021) Improving the accuracy of dam inflow predictions using a long short-term memory network coupled with wavelet transform and predictor selection. *Mathematics* 9(5):551. <https://doi.org/10.3390/math9050551>
- Tran VN, Ivanov VY, Kim J (2023a) Data reformation – A novel data processing technique enhancing machine learning applicability for predicting streamflow extremes. *Adv Water Resour* 182:104569. <https://doi.org/10.1016/j.advwatres.2023.104569>
- Tran VN, Ivanov VY, Xu D, Kim J (2023) Closing in on hydrologic predictive accuracy: combining the strengths of high-fidelity and physics-agnostic models. *Geophys Res Lett* 50(17):e2023GL104464. <https://doi.org/10.1029/2023GL104464>
- Tran VN, Kim J (2022) Robust and efficient uncertainty quantification for extreme events that deviate significantly from the training dataset using polynomial chaos-kriging. *J Hydrol* 127716. <https://doi.org/10.1016/j.jhydrol.2022.127716>
- Xiang Z, Yan J, Demir I (2020) A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resour Res* 56(1). <https://doi.org/10.1029/2019wr025326>
- Yadav S, Shukla S (2016) Analysis of k-Fold Cross-Validation over Hold-Out validation on colossal datasets for quality classification. 2016 IEEE 6th International Conference on Advanced Computing (IACC) 78–83. <https://doi.org/10.1109/IACC.2016.25>
- Yang G, Guo S, Liu P, Li L, Xu C (2017a) Multiobjective reservoir operating rules based on cascade reservoir input variable selection method. *Water Resour Res* 53(4):3446–3463. <https://doi.org/10.1002/2016wr020301>

- Yang T, Asanjan AA, Welles E, Gao X, Sorooshian S, Liu X (2017b) Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. *Water Resour Res* 53(4):2786–2812. <https://doi.org/10.1002/2017wr020482>
- Yaseen ZM, El-shafie A, Jaafar O, Afan HA, Sayl KN (2015) Artificial intelligence based models for stream-flow forecasting: 2000–2015. *J Hydrol* 530:829–844. <https://doi.org/10.1016/j.jhydrol.2015.10.038>
- Zhang D, Lin J, Peng Q, Wang D, Yang T, Sorooshian S, ... Zhuang J (2018) Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *J Hydrol* 565:720–736. <https://doi.org/10.1016/j.jhydrol.2018.08.050>
- Zhang D, Peng Q, Lin J, Wang D, Liu X, Zhuang J (2019) Simulating reservoir operation using a recurrent neural network algorithm. *Water* 11(4):865. <https://doi.org/10.3390/w11040865>
- Zhao WL, Gentine P, Reichstein M, Zhang Y, Zhou S, Wen Y, ... Qiu GY (2019) Physics-constrained machine learning of evapotranspiration. *Geophys Res Lett* 46(24):14496–14507. <https://doi.org/10.1029/2019gl085291>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Vinh Ngoc Tran¹ · Duc Dang Dinh² · Binh Duy Huy Pham² · Kha Dinh Dang³ ·
Tran Ngoc Anh^{2,3} · Ha Nguyen Ngoc⁴ · Giang Tien Nguyen³ 

✉ Giang Tien Nguyen
giangnt@vnu.edu.vn

¹ Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, USA

² Center for Environmental Fluid Dynamics, VNU University of Science, Vietnam National University, Thanh Xuan, 334 Nguyen Trai, Hanoi, Vietnam

³ Department of Hydrology and Water Resources, VNU University of Science, Vietnam National University, Thanh Xuan, 334 Nguyen Trai, Hanoi, Vietnam

⁴ National Center for Water Resources Planning and Investigation (NAWAPI), Long Bien District, No. 93, Lane 95, Vu Xuan Thieu Street Sai Dong Ward, Hanoi, Vietnam