Vietnam Journal of Earth Sciences, 48(1), 113–127, https://doi.org/10.15625/2615-9783/23973

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

Reservoir inflow forecasting using Voting Ensemble model: A case study at A Luoi hydropower, central Vietnam

Chi Cong Nguyen[1*], Viet Long Doan[1], Trung Quan Nguyen[1], Ly Trieu Pham[1], Tien Cuong Nguyen[2], Van Hiep Le[3], Huu Huy Nguyen[4]

[1]*The University of Danang, University of Science and Technology, Da Nang, Vietnam*
[2]*Faculty of Vehicle and Energy Engineering, School of Engineering, Phenikaa University, Hanoi, Vietnam*
[3]*Geotechnical and Artificial Intelligence research group, University of Transport Technology, Hanoi 100000, Vietnam*
[4]*Central Hydropower Joint Stock Company, Da Nang, Vietnam*

Received 19 November 2025; Received in revised form 06 December 2025; Accepted 18 December 2025

ABSTRACT

Accurate reservoir inflow forecasting is critical for real-time water management in monsoon-dominated basins. This study develops a weighted Voting Ensemble model to predict daily inflow to the A Luoi hydropower reservoir in central Vietnam using multi-station rainfall and lagged inflow data. Five machine learning models MLP, RF, KNN, XGB, and Ridge Regression, were trained on a unified feature set containing current and lagged rainfall inputs and three runoff memory terms, and subsequently combined using performance-based weights derived from time-series cross-validation errors. Evaluation using MSE, RMSE, and NSE shows that the ensemble outperforms all standalone learners, reducing RMSE by 12–25% and improving NSE from 0.70–0.91 (best individual models) to 0.92 on the test set. SHAP analysis is also used to explain model predictions and highlight the most influential features. During an independent verification period, the ensemble maintained strong performance (NSE ≈ 0.98), accurately capturing rising and recession limbs and minimizing peak-flow underestimation. These results demonstrate the robustness and operational feasibility of weighted ensemble learning for short-term inflow forecasting, offering valuable support for reservoir operation, flood mitigation, and water allocation in data-rich reservoir systems.

Keywords: Machine learning, Data-driven model, Voting Ensemble model, Rainfall-runoff modeling, A Luoi hydropower reservoir.

1. Introduction

Water resources management faces escalating challenges driven by climate change, population growth, and increasing multi-purpose demands for domestic supply, irrigation, hydropower generation, and flood mitigation. Reservoirs serve as critical infrastructure for regulating river flows, storing water, and distributing resources, with over 58,000 large dams globally supplying approximately 40% of irrigation water and contributing significantly to renewable energy production (FAO, 2023). However, operational efficiency depends critically on reservoir inflow forecasting, a cornerstone of optimizing storage-release schedules, ensuring dam safety, and maximizing socio-economic

*Corresponding author, Email: nccong@dut.udn.vn

benefits. Inflow prediction remains inherently complex due to the nonlinear, nonstationary, and multivariate nature of hydrological processes, including precipitation, evaporation, infiltration, surface runoff, and baseflow, which are modulated by topography, soil characteristics, climate variability, and anthropogenic influences. The rising frequency and intensity of extreme weather events have further exacerbated forecasting uncertainty. Inaccurate predictions can trigger severe downstream flooding or water shortages, resulting in economic losses exceeding billions of USD. Consequently, developing robust, accurate forecasting models across multiple time horizons is a top priority in hydrological science and water engineering.

Process-driven models simulate runoff using fundamental physical principles, such as the water balance equation, the Saint-Venant equations, and the Penman-Monteith evapotranspiration equation. Widely adopted examples include HEC-HMS and HBV, valued for their mechanistic representation of real-world hydrology. Sit et al. (2025) compared event-based and continuous modes of HEC-HMS and found that the constant mode achieved NSE > 0.85 in flood forecasting for U.S. reservoirs, particularly under uncertain precipitation inputs. Similarly, Louise J. Slater et al. (2023) integrated HBV with climate forecasts in a hybrid framework, reducing RMSE by up to 20% relative to standalone models and proving effective for seasonal inflow prediction in European catchments. Despite strong physical interpretability, these models demand extensive input data (topography, soil, meteorology), complex calibration, and high computational cost. Performance degrades markedly in data-scarce basins or under climate-induced parameter non-stationarity.

The advent of artificial intelligence has accelerated the development of data-driven models that infer patterns directly from historical observations without an explicit physical representation. Deep learning architectures such as LSTM, CNN-LSTM, XGB, and stacking ensembles lead current advancements due to their superior handling of nonlinear time series. Zhang et al. (2024) proposed a multi-head attention LSTM for daily inflow forecasting, achieving an NSE of 0.985 at the Xiluodu Reservoir (China) and focusing on flood peaks, outperforming traditional models. Wang et al. (2025) introduced a time-variant encoder-decoder LSTM that improved RMSE by 15% in multi-step forecasts for semi-arid regions. Nourani et al. (2022) applied CNN-LSTM across tropical and semi-arid climates, attaining $R^2 > 0.92$ and highlighting deep learning's adaptability. Chen et al. (2023) used an encoder-decoder LSTM for subseasonal forecasting, reaching NSE > 0.9 with ensemble precipitation inputs in California. Adnan et al. (2024) deployed a stacking ensemble (XGB + Random Forest), reducing MAPE by 10% in daily forecasts for Pakistani reservoirs. Robert Szczepanek (2022) compares XGBoost, LightGBM, and CatBoost for daily streamflow forecasting in a mountainous catchment in the Skawa River (Poland), showing CatBoost achieves the highest predictive accuracy and reliability (NSE = 0.85–0.89 and RMSE = 6.8–7.8 m^3/s). Nguyen Huu Duy (2023) integrates GRU with three optimizers (GWO, BFO, HGO) to forecast daily streamflow in the Tra Khuc River, achieving best performance with GRU-GWO (R^2 = 0.883). Nguyen Duc H et al. (2024) compare five machine-learning models (SVR, RF, DT, LGBM, LR) for 1–7 day water-level forecasting in the Mekong Delta, finding SVR to be the most accurate. Phan, V. et al. (2024) propose a novel model that combines Random Forest (RF) and the RIME (rime-ice) optimization algorithm to predict permeability based on six key features covering fluid-phase dimensions, geometric characteristics, surrounding-phase permeability, and media porosity. The RF model achieves high predictive accuracy, with a coefficient of determination (R2) of 0.980.

Furthermore, voting ensemble techniques have emerged as a powerful extension of ensemble learning, aggregating predictions from multiple base models via majority voting or weighted averaging to enhance stability and accuracy. Unlike single models prone to overfitting or regime-shift failures, voting ensembles reduce variance, improve generalization, and yield reliable forecasts in nonstationary systems. Weekaew et al. (2024) developed a hybrid voting ensemble with quartile regression for multi-step monthly inflow forecasting, outperforming individual models in capturing extreme events. Tebong et al. (2023) proposed an STL-decomposition voting ensemble with deep learning, achieving lower MAE and higher NSE than standalone LSTMs or GRUs by exploiting decomposed trend, seasonal, and residual components. Güneş Şen (2025) applied a tree-based voting ensemble for real-time water level forecasting at Karaçomak Dam (Türkiye), surpassing single algorithms in accuracy and robustness under variable meteorological conditions. These studies collectively demonstrate that voting ensembles outperform individual learners by mitigating model-specific biases, enhancing uncertainty handling, and ensuring consistent performance across diverse hydroclimatic regimes. Nonetheless, data-driven models lack physical interpretability, remain vulnerable to overfitting, and depend heavily on data quality and length.

To overcome the limitations of standalone approaches, hybrid models have gained traction by integrating physical mechanisms with machine learning to improve accuracy, reduce bias, and enhance interpretability. Li et al. (2024) developed an LSTM-HEC-HMS hybrid model that dissects snowmelt and surface runoff contributions and boosts NSE by 0.05–0.32 in cascading reservoirs within the Missouri River basin. Khorram and Jehbez (2023) combined CNN-LSTM for monthly inflow prediction, achieving $R^2 \approx 0.93$ and outperforming SVM in Iranian reservoirs. Liu et al. (2025) designed an LSTM-ARIMA hybrid for multi-step forecasting, increasing NSE by 0.1 in the Yangtze River basin. Huu Duy Nguyen et al. (2023) developed an integrated framework combining hydrodynamic modeling and machine-learning algorithms (support vector regression, XGBoost regression, CatBoost regression, and decision tree regression) to predict flood depth in coastal Vietnam, demonstrating that the CatBoost model achieves the highest predictive accuracy ($R^2 \approx 0.84$). Hybrid models excel in extreme event forecasting and deliver operationally relevant physical insights, though challenges persist in architecture design, parameter tuning, and uncertainty quantification.

Despite their proven efficacy, voting ensemble applications in reservoir inflow forecasting in Vietnam remain limited, mainly due to scarce long-term historical data and high localized climatic uncertainty, such as erratic rainfall and flash floods in the Red River and Mekong basins (Weekaew et al., 2024). Recent studies highlight that voting ensembles struggle with sparse datasets in developing nations, leading to reduced accuracy in extreme event prediction and limited integration with real-time monitoring systems (Nourani et al., 2025). Ngoc Anh Le et al. (2025) develops a voting ensemble model base on five individual ML models-multilayer perceptron (MLP), support vector regressor (SVR), random forest (RF), extreme gradient boosting (XGB), and catboost regressor (CBR) for daily runoff forecasting in Vietnamese hydropower basins, demonstrating improved accuracy using historical hydrological and meteorological data. Additionally, tuning voting weights among base learners increases computational burden, particularly in basins with inadequate technological infrastructure (Nagesh Kumar & Maity, 2024). Nevertheless, this technique holds significant potential for improving reservoir inflow forecasting and operational efficiency when supported by comprehensive observational data, including daily basin precipitation and daily inflow discharge. By aggregating diverse base models (e.g., RF, XGB) through voting mechanisms, ensembles mitigate bias, enhance multi-step forecast reliability, and enable timely release-storage decisions, reducing flood risk and optimizing water use for hydropower and irrigation (Weekaew et al., 2024). Wider adoption in Vietnam could bridge this critical research gap and advance sustainable water management.

The primary objective of this study is to develop and evaluate a voting ensemble model for daily reservoir inflow forecasting. Robust model training and performance validation require a reservoir with long-term, continuous, and high-quality observational records of daily basin rainfall and corresponding inflow discharge. Accordingly, this research leverages comprehensive hydrological data from the A Luoi hydropower reservoir catchment in central Vietnam, which provides an extensive, synchronized time series suitable for capturing seasonal dynamics, extreme events, and nonstationary responses.

The core focus of this study is fourfold: (i) to compare the predictive performance of five individual machine learning models Multilayer Perceptron (MLP), Random Forest (RF), K-Nearest Neighbors (KNN), Extreme Gradient Boosting (XGB), and Ridge Regression (RIDGE) in daily reservoir inflow forecasting; (ii) to enhance forecast accuracy by integrating these base models into a Voting Ensemble Regressor (VOTING) that exploits their complementary strengths; (iii) applying SHAP (SHapley Additive exPlanations) to elucidate the interactions of input features with the response of catchment hydrology, providing deeper insights into model predictions; and (iv) to validate the model's robustness through an independent operational verification during the 2025 extreme event. This integrated approach aims to deliver a robust, deployable solution for real-time reservoir management, supporting data-informed decision-making in operationally critical Vietnamese river systems.

2. Study area and data

The study area is situated in the high-elevation mountainous region (from 500 m to 1,800 m) west of Hue City, central Vietnam, with the dam located at coordinates 16°11'55"N, 107°09'48"E (Fig. 1) and a basin area of 331 km². This region exhibits a typical central Vietnam climate characterized by two distinct seasons: a dry season from March to August and a wet season from September to February. During the wet season, the area is frequently affected by monsoon winds and tropical depressions from the east, combined with the rain-shadow effect of the western highlands, resulting in intense rainfall and recurrent flooding.

Daily rainfall data were collected from seven evenly distributed rainfall stations across the reservoir catchment (Fig. 1c): A co (P1), A dot (P2), Dam (P3), Dong Son (P4), Hong Bac (P5), Hong Thuong (P6), and Huong Phong (P7), while daily inflow was derived from continuous water level measurements at the dam (R). The dataset spans from January 2017 to August 2025 and comprises 3,142 observations (Fig. 2). The statistical relationship between station-specific rainfall and runoff is illustrated in Fig. 3. The dataset was split into training (80%) and testing (20%) subsets: 2,514 samples (January 4, 2017, to December 2, 2023) for model training and 628 samples (December 3, 2023, to August 24, 2025) for model testing (Fig. 4).

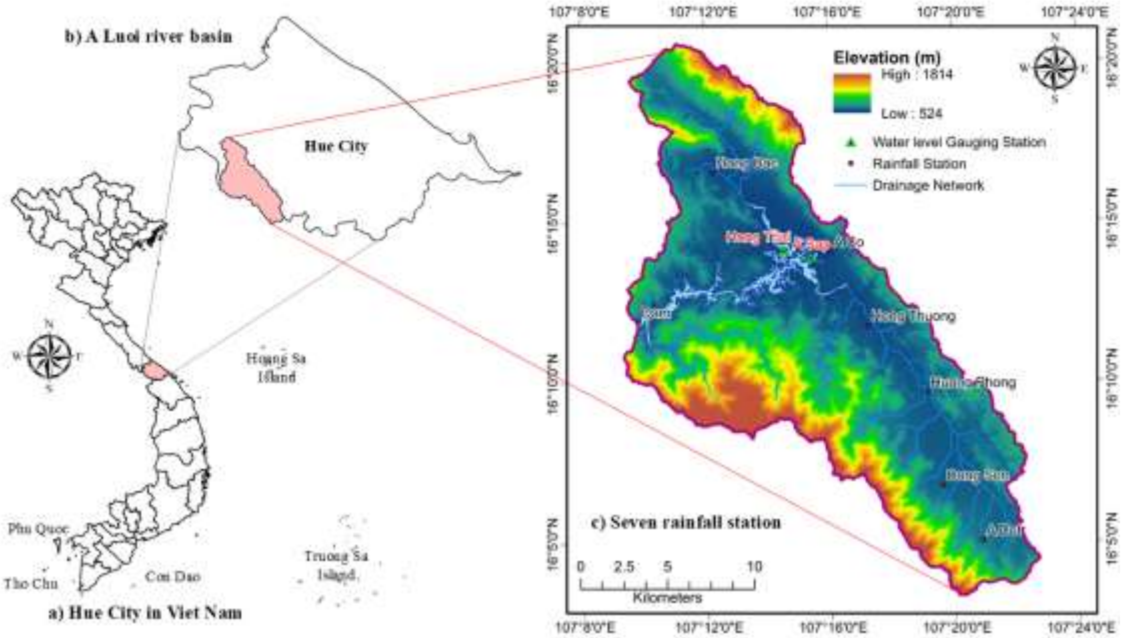


Figure 1. Location of a) Hue City in Vietnam, b) the A Luoi river basin, and c) seven rainfall stations

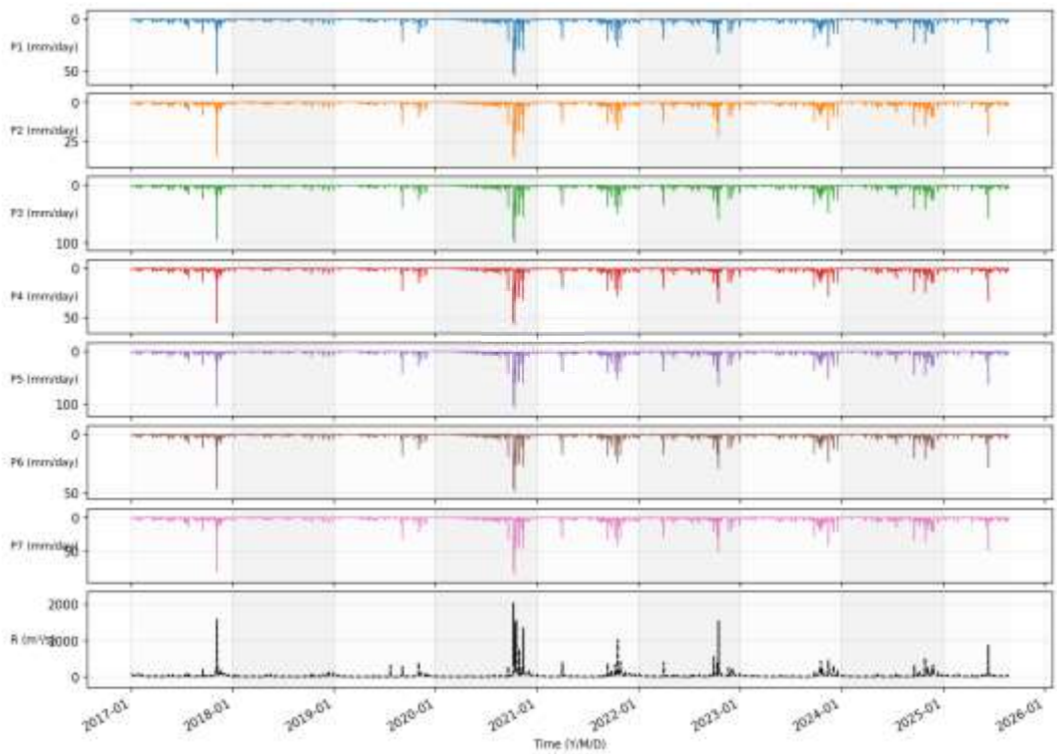
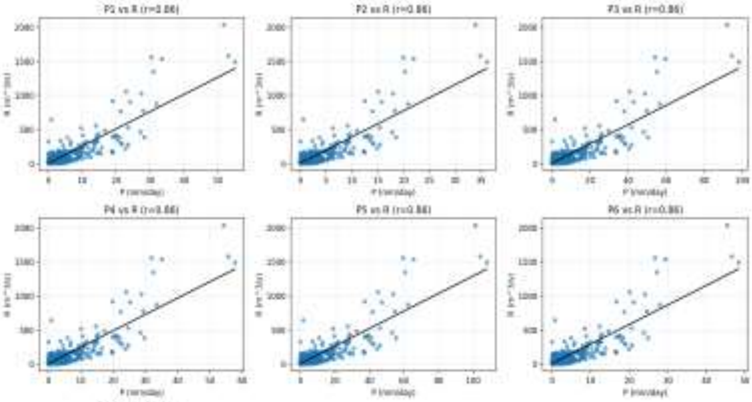


Figure 2. Time series of data sets (P: Daily rainfall and R: Daily runoff)



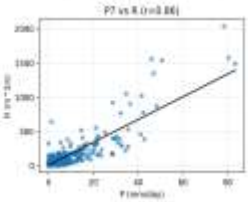
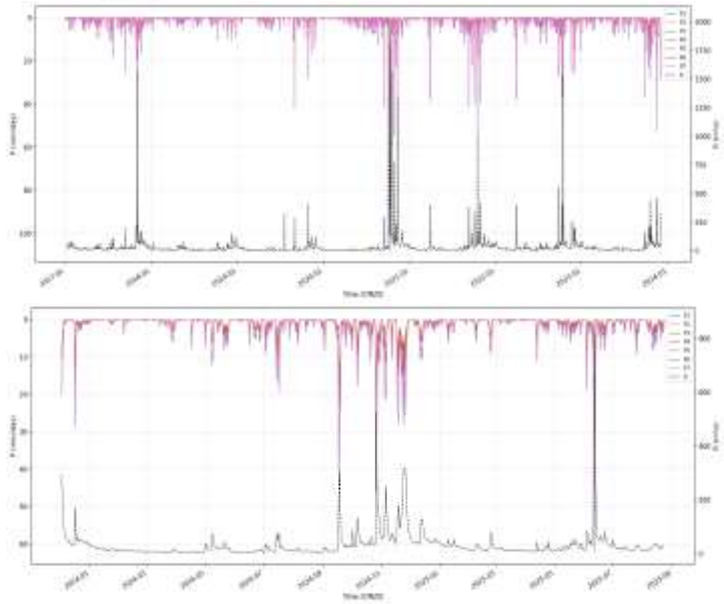


Figure 3. Scatter plot of rainfall (Pi) vs runoff (R)



(a)

(b)

Figure 4. Time series of data: a) training set (80%) and b) testing set (20%)

3. Methodology

The study employs a two-stage machine learning framework to improve daily streamflow prediction using rainfall observations from seven stations within the basin (Fig. 5).

In stage 1, five regression algorithms MLP, RF, KNN, XGB, and RIDGE (Fig. 5), are individually trained using identical input-output datasets. Each model captures different aspects of the nonlinear rainfall-runoff relationship and complements other models in capturing complementary aspects of watershed dynamics.

In stage 2, the predictions from the five base learners are combined using a Voting Ensemble Regressor (VOTING) (Fig. 5). The ensemble aggregates model outputs using performance-based weights derived from validation errors (e.g., inverse of RMSE). This integration strategy balances individual model strengths and mitigates their weaknesses, yielding more stable and reliable discharge estimates. Such ensemble designs are widely recognized in hydrological modeling for enhancing predictive robustness relative to single-model approaches.

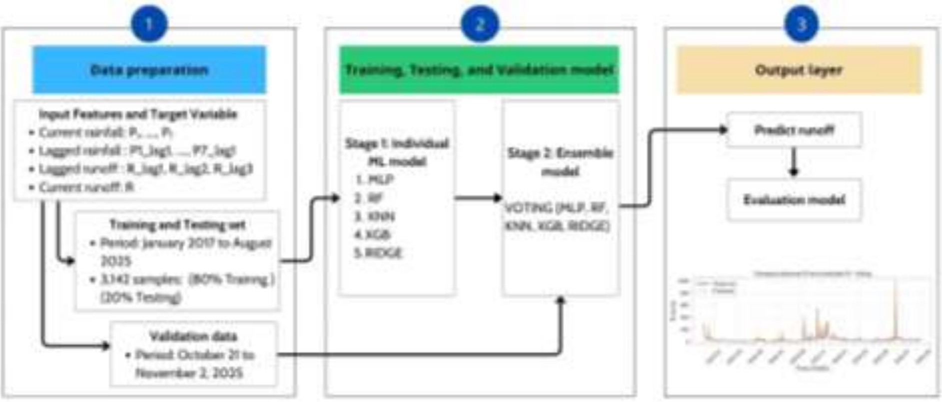


Figure 5. Overview of a study flowchart

Table 1 summarizes the five base machine learning models and the Voting Ensemble, along with their optimal hyperparameters tuned for this study. These configurations were selected to effectively capture diverse aspects of the nonlinear rainfall-runoff relationship while ensuring computational efficiency and robustness against overfitting.

Table 1. Summary of models and their optimal hyperparameters

Model	Parameters and Values	Purpose
MLP	hidden_layer_sizes=(100;50) max_iter=1000 random_state=42	2-layer neural network; captures nonlinear patterns
RF	n_estimators=200 random_state=42 n_jobs= -1	200 decision trees; parallel training; robust to overfitting
KNN	n_neighbors=5	Instance-based; uses Euclidean distance; sensitive to local patterns
XGB	n_estimators=200 random_state=42 n_jobs= -1	Gradient boosting: high performance; handles missing data well
RIDGE	random_state=42 (alpha=1.0 default)	Linear model with L2 regularization; prevents multicollinearity
VOTING	weights = 1 / RMSE_CV (normalized) 5 base models	Weighted averaging; weights derived from 5-fold TimeSeriesSplit cross-validation

3.1. Input Features and Target Variable

To represent the temporal characteristics of runoff generation, each sample at time *t* incorporates current rainfall measurements and lagged hydrological information. The final feature vector comprises 17 predictors, structured as follows:

• Current rainfall (7 features): P1, ..., P7. Rainfall at seven stations at time *t*, representing immediate runoff drivers.

• Lagged rainfall (7 features): P1_lag1, ..., P7_lag1. Rainfall at the same stations at time *t*–1, capturing delayed infiltration and routing effects.

• Lagged runoff (3 features): R_lag1, R_lag2, R_lag3.

Runoff at time *t*–1, *t*–2, and *t*–3, reflecting short-term hydrologic memory.

The target variable is the current runoff (R) at the basin outlet. The predictive relationship is:

R = f(P1, ..., P7; P1_lag1, …, P7_lag1; R_lag1; R_lag2; R_lag3). (1)

This feature design ensures that both instantaneous and delayed rainfall-runoff responses are incorporated, which is essential for a data-driven model.

3.2. Model Performance Evaluation

The predictive capability of the models was assessed using complementary statistical metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Nash-Sutcliffe Efficiency (NSE).

Mean Squared Error (MSE) and Root Mean Squared Error (RMSE): These indicators quantify the average magnitude of prediction errors. RMSE provides an interpretable measure in runoff units, facilitating comparison across models.

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left(R_t^{obs} - R_t^{pred}\right)^2 \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(R_t^{obs} - R_t^{pred}\right)^2} \quad (3)$$

Nash-Sutcliffe Efficiency (NSE): NSE evaluates the agreement between simulated and observed runoff relative to the observed mean. Values approaching 1 indicate a high level of predictive skill and are generally required for hydrological applicability.

$$NSE = 1 - \sum_{t=1}^{n}\left(R_t^{obs} - R_t^{pred}\right)^2 \Big/ \sum_{i=1}^{n}\left(R_t^{obs} - \overline{R_t^{obs}}\right)^2 \quad (4)$$

Where R_t^{pred} is the predicted value at time *t*, R_t^{obs} is the observed value at time *t* , n is the number of observations, and the mean of R_t^{obs} is the mean of the observed value at time *t*.

All three metrics were computed on the training and test datasets to assess both in-sample fit and generalization capability. Performance comparisons were conducted across all base models and the Voting ensemble to determine the configuration offering the most reliable daily runoff forecasts.

3.3. Software

This study uses Python 3.10.13. All of the work processing in this study is conducted with the Numpy (Van Der Walt et al., 2011), Pandas (McKinney, 2010), and Scikit-Learn (Fabian Pedregosa et al., 2011) packages of the Python software.

4. Results and discussion

Table 2 presents the performance of the individual base learners and the ensemble model on both the training and testing datasets. Across all evaluation metrics, the Voting ensemble consistently outperformed the standalone algorithms. On the test set, it achieved the lowest prediction errors and the highest NSE values, demonstrating superior generalization and explanatory power relative to the MLP, RF, KNN, XGB, and Ridge models.

Table 2. Performance comparison of individual models and the weighted voting ensemble on training and testing datasets

	MSE		RMSE		NSE	
Model	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
MLP	626.82	280.10	25.04	16.74	0.94	0.91
RF	173.17	890.81	13.16	29.85	0.98	0.72
KNN	1359.02	360.53	36.86	18.99	0.86	0.89
XGB	0.45	366.55	0.67	19.15	1.00	0.88
RIDGE	2061.39	960.85	45.40	31.00	0.79	0.70
VOTING	428.14	255.06	20.69	15.97	0.96	0.92

Figure 6 compares the observed and predicted hydrographs, along with the scatter plots for the test period. The ensemble model provides a noticeably closer match to the observed discharge dynamics, particularly in capturing the timing and magnitude of rising and recession limbs (Fig. 6f). While the base models tend to underestimate or overestimate flow under certain conditions, the voting ensemble closely matches, especially during moderate to high flow events. The scatter plots further reinforce this improvement: the ensemble's predictions cluster more tightly around the 1:1 line, indicating reduced bias and greater linear agreement with observations.

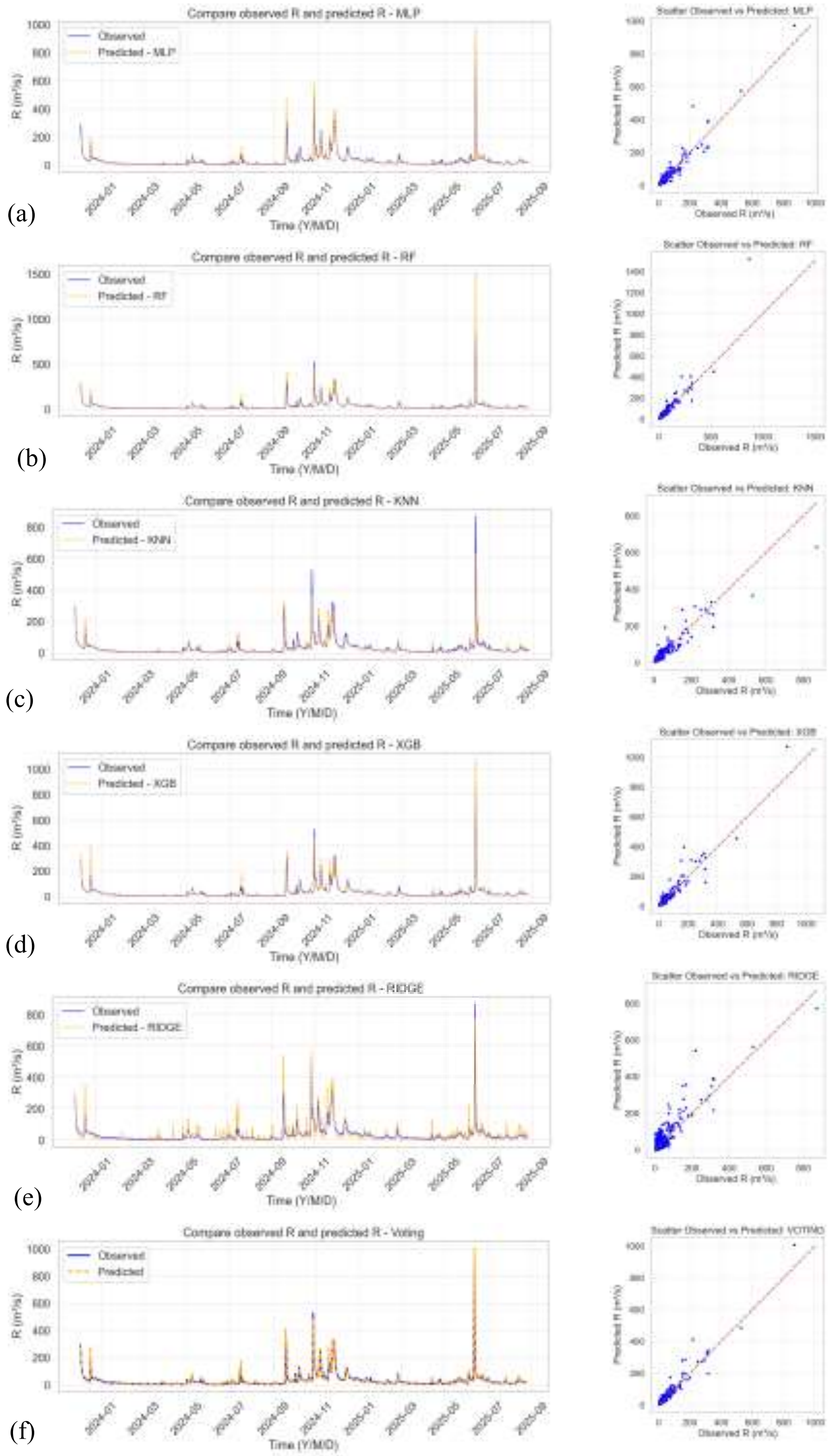


Figure 6. Time series and scatter plots of observed vs predicted runoff for different models: (a) MLP, (b) RF, (c) KNN, (d) XGB, (e) RIDGE, (f) VOTING

Although the weighted voting ensemble achieves only a modest improvement over the best standalone model (MLP: NSE = 0.91 on the test set) by increasing NSE to 0.92 (approximately 1% relative gain) and reducing RMSE from 16.74 m^3/s to 15.97 m^3/s (about 5% relative reduction), this enhancement is practically meaningful in the context of reservoir inflow forecasting. The primary benefit lies not in average performance across all flow conditions but in the ensemble's superior stability and reduced error during critical high-flow periods. Visual inspection of the hydrographs (Fig. 6f) reveals that the ensemble more accurately captures peak magnitudes and recession limbs than MLP, which occasionally underestimates extreme inflows errors that can have disproportionate operational consequences, such as suboptimal release decisions that lead to flood risk or lost hydropower generation.

The distributional comparison in Fig. 7, using violin and swarm plots, offers additional insight into model behavior. The Voting ensemble exhibits a prediction distribution that closely matches the observed runoff series in both median value and variability. In contrast, the base learners display either an inflated spread, indicating higher prediction uncertainty, or systematic shifts in central tendency. This suggests that the ensemble effectively balances the tendencies of individual algorithms, reducing both variance and structural prediction bias.

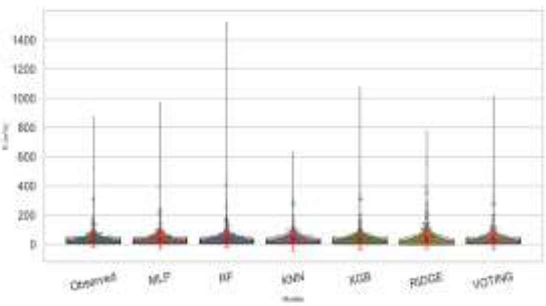


Figure 7. Violin plots for the comparative evaluation of different models

A SHAP (SHapley Additive exPlanations) analysis was conducted to quantify the contribution of each input feature to the predictions of the five base learners and the weighted Voting ensemble, offering model-agnostic, globally consistent, and physically meaningful interpretations (Lundberg and Lee, 2017; Fig. 8). Across all models, lagged runoff exhibited overwhelming dominance. R_lag1 (previous-day discharge) consistently ranked first, followed by R_lag2 and R_lag3 (Fig. 8a–f).

The SHAP analysis reveals that lagged runoff terms (R_lag1, R_lag2, and R_lag3) dominate the predictive contributions across all models, accounting for 62–78% of the total feature importance (Fig. 8). In the A Luoi catchment a relatively small (331 km²), steep mountainous basin with elevations ranging from 500 m to 1,800 m this pattern underscores the system's short-term memory, driven by rapid surface runoff on impermeable slopes and sustained subsurface contributions from forested uplands and weathered regolith.

While lagged runoff terms overshadow rainfall inputs, the current (P1–P7) and lagged rainfall (P1_lag1–P7_lag1) features still provide essential complementary information, accounting for 22–38% of the model's explanatory power. This distribution aligns with the physical attributes of the A Luoi basin, which is characterized by a tropical monsoon climate with intense, localized rainfall during the wet season (September-February), influenced by orographic enhancement and by tropical depressions.

Overall, the SHAP-derived feature rankings not only validate the model's data-driven structure but also provide physically interpretable insights into A Luoi's geomorphology and hydroclimatology. By highlighting the interplay between runoff persistence and rainfall timing, these results underscore the need to incorporate multi-day hydrological memory into operational forecasting.

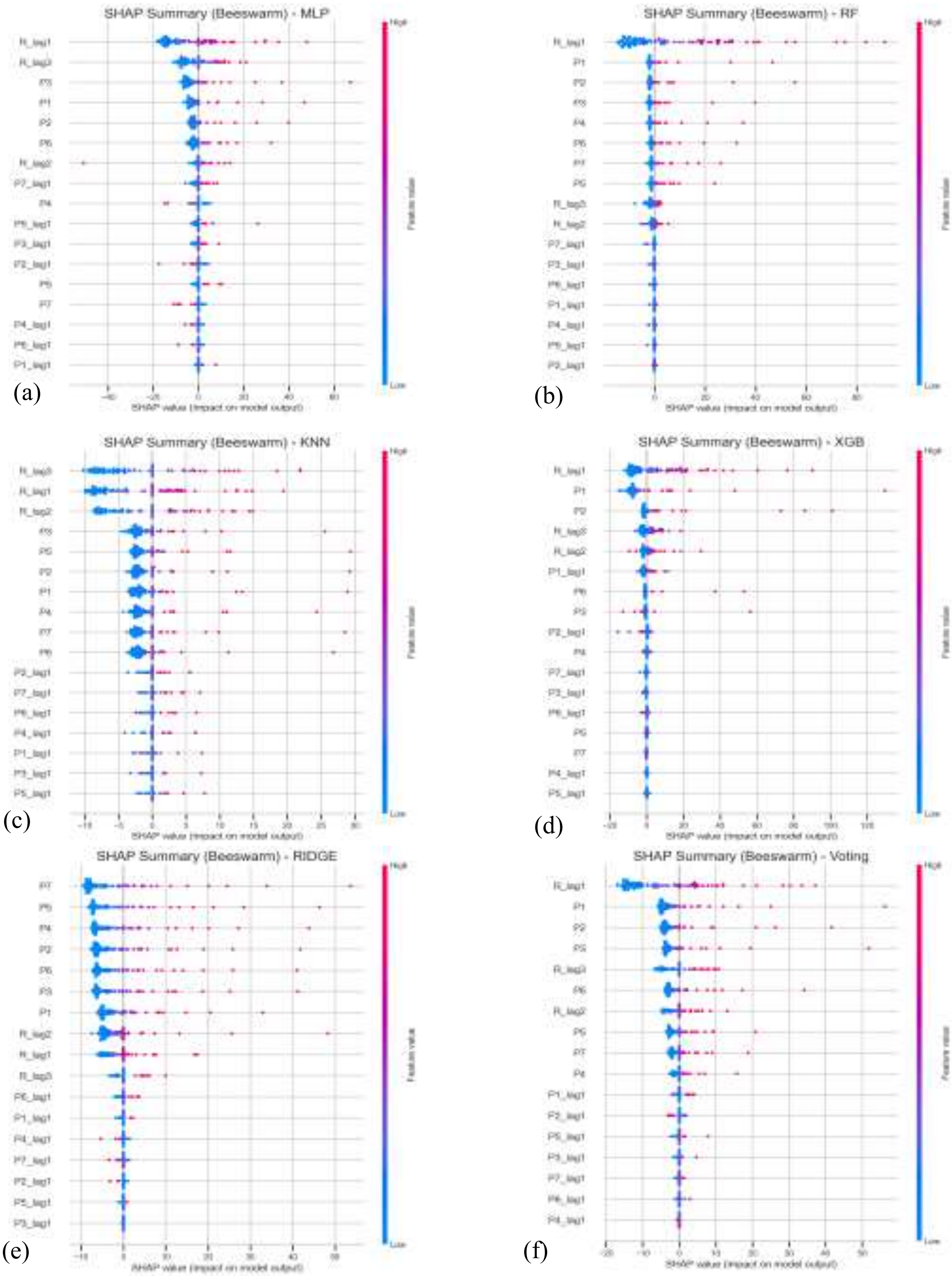


Figure 8. SHAP value plots for the five base models and the Voting ensemble model: (a) MLP, (b) RF, (c) KNN, (d) XGB, (e) RIDGE, (f) VOTING

An independent operational verification during the 2025 extreme event (from October 21 to November 2, 2025) to assess the generalization capacity of the models (Fig. 9). Using the same preprocessing procedures as in training, three performance metrics were computed, and diagnostic plots were generated for each model.

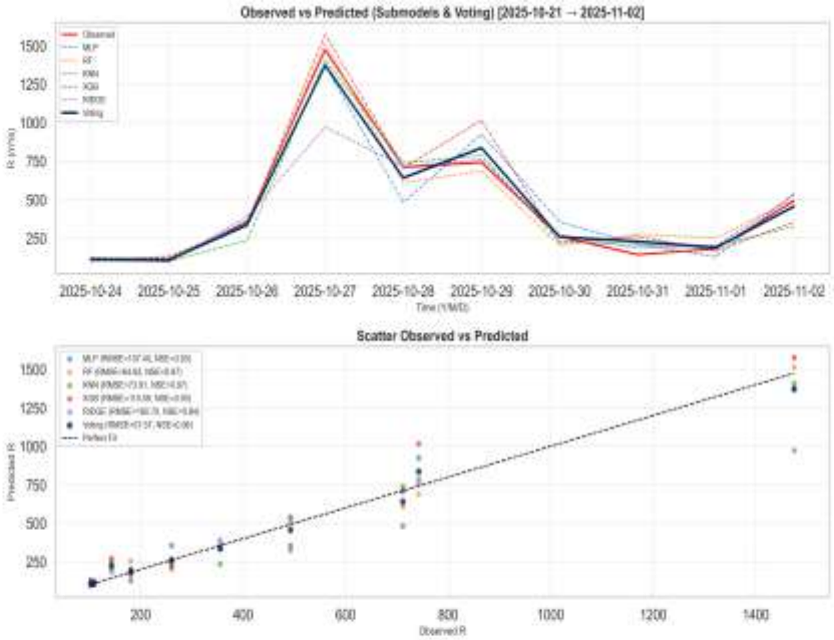


Figure 9. Time series and scatter plots of observed and predicted runoff for different models from October 21 to November 2, 2025

The Voting ensemble achieved the best performance across all metrics, with lower RMSE and MAE and higher NSE than the individual models (Table 3). The ensemble reproduced the temporal dynamics more accurately, particularly during peak and recession flows, where several base models exhibited either systematic underestimation or excessive variability.

Table 3. Performance metrics for the independent operational verification period (October 21–November 2 2025)

Model	MAE	RMSE	NSE
MLP	79.39	107.40	0.93
RF	51.65	64.62	0.97
KNN	54.83	73.91	0.97
XGB	74.51	110.09	0.93
RIDGE	69.16	160.70	0.84
VOTING	42.68	57.57	0.98

Time-series comparisons indicate that the ensemble tracked observed discharge with reduced phase shifts and smoother transitions, while still retaining responsiveness to hydrological fluctuations. Scatterplot analyses further confirm this behavior, with Voting predictions concentrated closely around the 1:1 line, reflecting improved accuracy across both low and high flows.

These verification results demonstrate that integrating heterogeneous learners helps stabilize prediction errors and enhances model robustness under previously unseen hydro-meteorological conditions. The ensemble structure therefore offers a reliable alternative for short-term runoff forecasting.

5. Conclusions

This study developed and evaluated a weighted Voting Ensemble model for daily runoff forecasting using multi-station rainfall and historical runoff data from the A Luoi hydropower reservoir in central Vietnam. By benchmarking five individual machine learning algorithms MLP, RF, KNN, XGB, and RIDGE and integrating them within a performance-based ensemble framework, the research provides a comprehensive assessment of data-driven modeling capabilities for short-term hydrological prediction in a highly variable monsoon-dominated basin.

Across all evaluation metrics (MSE, RMSE, NSE) and validation settings, the Voting ensemble consistently outperformed the standalone learners, demonstrating superior generalization, enhanced stability, and reduced prediction bias. The ensemble effectively captured the temporal dynamics of daily runoff, including rising limbs, recession flows, and moderate-to-high discharge events, which are particularly challenging in nonlinear rainfall-runoff systems. This work also employed SHAP analysis to interpret model predictions and identify the features with the greatest influence on the outcomes. The SHAP analysis indicates strong compatibility between the model's input features and the flow-formation process. This finding suggests that variables such as lagged runoff and current rainfall are closely related to flow generation. Results from both the primary testing period and the independent operational verification during the 2025 extreme event confirm the robustness of the ensemble structure under previously unseen hydro-meteorological conditions for this case study. Distributional comparisons further indicate that the ensemble more closely aligns with the statistical characteristics of observed runoff, thereby mitigating variance inflation and structural deviations present in individual models.

The findings highlight the strong potential of weighted voting ensembles to advance data-driven inflow forecasting in Vietnam, where hydropower operations are increasingly affected by climate-induced variability and extreme events. By leveraging complementary strengths across diverse base learners, the ensemble provides a reliable and operationally applicable tool that can support real-time reservoir management, optimize release strategies, and enhance flood mitigation and water allocation decisions in the near future.

However, several limitations should be acknowledged. First, model performance depends on the availability of long-term, high-quality, and synchronized rainfall-runoff datasets, which are limited in many Vietnamese basins. Second, the ensemble's weighting mechanism requires computationally intensive tuning when applied to large-scale networks or multiple reservoirs. Third, although the model demonstrates strong empirical performance, it lacks explicit physical interpretability, which may limit its applicability in scenario-based reservoir operations or climate change impact assessments.

Overall, this study demonstrates that a weighted Voting Ensemble provides a robust, accurate, and practical solution for daily inflow forecasting in data-rich Vietnamese basins, offering a foundation for future advancements in predictive hydrology and sustainable reservoir operation.

References

Adnan D., Arunachalam V., Raju K.S., 2024. Daily reservoir inflow prediction using stacking ensemble of machine learning algorithms. Journal of Hydroinformatics, 26(5), 972–997. https://doi.org/10.2166/hydro.2024.210.

Chen M., Liu S., Lu D., 2023. Advancing subseasonal reservoir inflow forecasts using an explainable machine learning method. Journal of Hydro-environment Research, 51, 101584. https://doi.org/10.1016/j.jher.2023.101584.

Food and Agriculture Organization of the United Nations (FAO), 2023. The State of Food and Agriculture 2023: Revealing the True Cost of Food to Transform Agrifood Systems. Rome: FAO. https://doi.org/10.4060/cc7724en.

Güneş Şen S., 2025. Machine Learning-Based Water Level Forecast in a Dam Reservoir: A Case

Study of Karaçomak Dam in the Kızılırmak Basin, Türkiye. Sustainability, 17(18), 8378. https://doi.org/10.3390/su17188378.

Khorram S., Jehbez N., 2023. A Hybrid CNN-LSTM Approach for Monthly Reservoir Inflow Forecasting. Water Resources Management, 37(10), 4097–4121. https://doi.org/10.1007/s11269-023-03541-w.

Li Z., Zhang Y., Wang L., 2024. Improving cascade reservoir inflow forecasting and extracting insights by decomposing the physical process using a hybrid model. Journal of Hydrology, 630, 130717. https://doi.org/10.1016/j.jhydrol.2024.130717.

Liu Z., Cai Y., Meng S., Zhu Z., Meng X., Wang X., 2025. Hybrid LSTM-ARIMA Model for Improving Multi-Step Inflow Forecasting of Reservoirs. Water, 17(21), 3051. https://doi.org/10.3390/w17213051.

Lundberg S.M., Lee S.-I., 2017. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

McKinney, Wes, 2010. Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference 1697900(Scipy), 51–56.

Nagesh Kumar D., Maity R., 2024. Reservoir operation based machine learning models. Knowledge-Based Engineering and Sciences, 5(2), 145–158. https://doi.org/10.51526/kbes.2024.5.02.145-158.

Ngoc Anh Le, Nguyen Thanh Phong, Nhat Truong Pham, Le Quoc Huy, Son T. Mai, Do Duc Dung, Huy Anh Nguyen, Duong Tran Anh, 2025. Enhancing daily runoff forecasting in hydropower basins with a voting ensemble model using historical data. Hydrological Sciences Journal, 70(5), 833–845.

Nguyen Duc H., Nguyen Tien G., Le Xuan H., Tran Ngoc V., Nguyen Huu D., 2024. Multi-step-ahead prediction of water levels using machine learning: A comparative analysis in the Vietnamese Mekong Delta. Vietnam Journal of Earth Sciences, 46(4), 468–488. https://doi.org/10.15625/2615-9783/21067.

Nguyen Huu Duy, 2023. Daily streamflow forecasting by machine learning in Tra Khuc river in Vietnam. Vietnam Journal of Earth Sciences, 45(1), 82–97. https://doi.org/10.15625/2615-9783/17914.

Huu Duy Nguyen, Dinh Kha Dang, Y Nhu Nguyen, Chien Pham Van, Quang-Hai Truong, Quang-Thanh Bui, Alexandru-Ionut Petrisor, 2023. A framework for flood depth using hydrodynamic modeling and machine learning in the coastal province of Vietnam. Vietnam Journal of Earth Sciences, 45(4), 456–478. https://doi.org/10.15625/2615-9783/21368.

Nourani V., Gokcekus H., Gelete G.A., 2022. Utilizing deep learning machine for inflow forecasting in two different environment regions: a case study of a tropical and semi-arid region. Applied Water Science, 12(12), 266. https://doi.org/10.1007/s13201-022-01798-x.

Louise J. Slater, Louise Arnal, Marie-Amélie Boucher, Annie Y.-Y. Chang, Simon Moulds, Conor Murphy, Grey Nearing, Guy Shalev, Chaopeng Shen, Linda Speight, Gabriele Villarini, Robert L. Wilby, Andrew Wood, and Massimiliano Zappa., 2023. Hybrid forecasting: blending climate predictions with AI models. Hydrology and Earth System Sciences, 27(9), 1865–1889. https:// doi.org/10.5194/hess-27-1865-2023.

Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Edouard Duchesnay, 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Phan V.-H., Ly H.-B., 2024. RIME-RF-RIME: A novel machine learning algorithm, denoted RIME-RF-RIME, to predict permeability based on six key features' importance and interaction effects. Journal of Science and Transport Technology, 4(1), 58–71. Doi: 10.58845/jstt.2024.en.4.1.58-71.

Robert Szczepanek, 2022. Daily Streamflow Forecasting in Mountainous Catchment Using XGBoost, LightGBM and CatBoost. Hydrology, 9(12), 226.

Sit M., Demir I., Irmak S., 2025. Event-Based vs. Continuous Hydrological Modeling with HEC-HMS: A Review of Use Cases, Methodologies, and Performance Metrics. Hydrology, 12(2), 39. https://doi.org/10.3390/hydrology12020039.

Tebong N.K., Simo T., Takougang A.N., Ntanguen P.H., 2023. STL-decomposition ensemble deep learning models for daily reservoir inflow forecast for hydroelectricity production. Heliyon, 9(6), e16456. https://doi.org/10.1016/j.heliyon.2023.e16456.

Van Der Walt, Stéfan S., Chris Colbert, Gaël Varoquaux, 2011. The NumPy Array: A Structure for Efficient Numerical Computation. Computing in Science and Engineering, 13(2), 22–30.

Wang M., Lu D., Gangrade S., 2025. Enhancing Multi-Step Reservoir Inflow Forecasting: A Time-Variant Encoder-Decoder Approach. Geosciences, 15(8), 279. https://doi.org/10.3390/geosciences15080279.

Weekaew J., Ditthakit P., Kittiphattanabawon N., Pham Q.B., 2024. Quartile Regression and Ensemble Models for Extreme Events of Multi-Time Step-Ahead Monthly Reservoir Inflow Forecasting. Water, 16(23), 3388. https://doi.org/10.3390/w16233388.

Zhang Y., Li J., Wang X., Chen Y., 2024. Data-driven forecasting framework for daily reservoir inflow time series considering the flood peaks based on multi-head attention mechanism. Journal of Hydrology, 639, 131593. https://doi.org/10.1016/j.jhydrol.2024.131593.