



Explainable AI and Ensemble Machine Learning Analysis of River Flow Dynamics: Influence of Key Climatic Variables (Temperature, Humidity, Precipitation)

Mustafa Çakır¹, Gizem Nazlı Ural², Mükerrerem Oral³, Okan Oral⁴, Mesut Yılmaz^{2*}

¹ Iskenderun Vocational School of Higher Education, Iskenderun Technical University, Iskenderun, Türkiye.

² Faculty of Fisheries, Akdeniz University, Antalya, Türkiye.

³ Kemer Faculty of Maritime, Akdeniz University, Antalya, Türkiye.

⁴ Faculty of Engineering, Akdeniz University, Antalya, Türkiye.

* Corresponding author. E-mail: myilmaz@akdeniz.edu.tr

Abstract: Accurate short-term river flowrate forecasting is essential for flood risk mitigation and sustainable water management. However, many machine learning (ML) applications in hydrology lack strict temporal validation and interpretability, limiting operational reliability. This study develops a reproducible and explainable framework for one-day-ahead daily flowrate forecasting in the Eşen River Basin (Türkiye) using hydro-meteorological data (2017–2022) from one flowrate station and four meteorological monitoring stations. The workflow integrates the Box-Cox transformation, lag-based feature engineering (up to 3 days, reflecting short-term hydrological memory), Boruta feature selection, and strictly time-aware rolling validation (2017–2021 training; 2022 independent test). Classical time-series models (ARIMA, TBATS), interpretable baselines (DT, LR), and advanced ML algorithms (RF, GBM, XGBoost, SVM, ANN) were benchmarked using RMSE, MAE, R^2 , NSE, and KGE. Ensemble tree-based models consistently outperformed classical and baseline approaches in magnitude-sensitive metrics. XGBoost achieved the highest predictive accuracy ($R^2 = 0.864$; NSE = 0.864; RMSE = 6590 dm³ s⁻¹). Although TBATS yielded the highest KGE (0.865), ensemble models better captured nonlinear dynamics and flowrate variability. SHAP and LIME analyses revealed that short-term flow lags dominate predictive structure, while precipitation and temperature exert regime-dependent influence. The complete workflow is openly deployed via a reproducible *R-Shiny* environment. The results demonstrate that explainable ensemble learning, combined with strict temporal validation, provides a reliable and transparent framework for operational hydrological forecasting.

Keywords: Daily flowrate forecasting; Ensemble learning; Explainable AI; SHAP; LIME; Time-aware validation; Hydrology.

1 INTRODUCTION

Accurate flowrate forecasting plays a fundamental role in water resources management, supporting decision-making processes related to flood risk mitigation (Brenner et al., 2016; Sadeghi-Pouya et al., 2017) hydropower operation (Berger, 2018; Hečková et al., 2022), irrigation planning (Moore et al., 2010; Schacht and Marschner, 2015; Zhang et al., 2018), and long-term freshwater system sustainability.

Under increasing hydro-climatic variability and growing anthropogenic pressure (Najafabadi et al., 2015; Mishra and Kaushal, 2025), reliable short-term flowrate predictions are essential for anticipating extreme events, optimizing reservoir management (Vyshnevskiy and Shevchuk, 2021; Kozel and Sary, 2022), and reducing socio-economic risks (Chiew et al., 2003; Cheng et al., 2020; Naganna et al., 2023; Ahmed et al., 2024).

Traditional hydrological forecasting has largely relied on statistical time-series approaches, particularly autoregressive integrated moving average (ARIMA) models and exponential smoothing variants such as Trigonometric Box-Cox ARMA Trend Seasonal (TBATS) (Çakır et al., 2026). Owing to their structural simplicity and interpretability, these methods remain widely adopted in operational settings (Hasan et al., 2023;

Abderrahmane et al., 2025). However, river flowrate systems are often governed by nonlinear interactions, temporal persistence, and complex coupling with meteorological drivers, conditions under which classical linear frameworks may exhibit reduced predictive flexibility.

In response, machine learning (ML) techniques have increasingly been applied to hydrological forecasting. Algorithms including Random Forest (RF), Gradient Boosting Machines (GBM), Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Extreme Gradient Boosting (XGBoost) have shown enhanced capability in modeling nonlinear relationships and high-dimensional dependencies in flowrate data (Desai and Ouarda, 2021; Song et al., 2022; Feng et al., 2022; Szczepanek, 2022; Tareke & Awoke, 2023). Despite demonstrated improvements in predictive accuracy, broader operational adoption of ML in hydrology remains constrained by methodological and practical limitations.

Two interrelated challenges are particularly evident in the literature. First, many ML-based hydrological studies employ random train-test splits or conventional k-fold cross-validation, overlooking the temporal structure of flowrate series and potentially introducing information leakage that inflates performance estimates (Bergmeir & Benítez, 2012). Second, ML models are frequently criticized for limited interpretability, as

their internal decision structures often obscure the relative and conditional influence of predictors. This lack of transparency restricts trust and practical integration into water management processes (Núñez et al., 2023; Aderemi et al., 2025).

Within this framework, the Eşen River basin provides a relevant case study due to its hydrological variability and socio-economic importance. The basin supports domestic water supply (Shumova, 2009; Orfanus et al., 2017), agriculture (Saci et al., 2025), hydropower production (Mantovani et al., 2014), and aquaculture activities, including intensive rainbow trout farming that contributes significantly to regional and international markets.

As part of an interconnected water-energy-food nexus, the Eşen River system represents a strategic setting in which accurate and interpretable flowrate forecasting is essential for resource security and regional resilience (ten Veen, 2004; Karakoca & Ünver, 2024; Elítez & Yaltırak, 2024; Tatar, 2025).

In light of these considerations, this study develops an integrated and reproducible framework for daily flowrate forecasting that simultaneously addresses methodological rigor and interpretability. The proposed workflow combines the Box-Cox transformation, lag-based feature engineering, and Boruta-based feature selection to construct a structured predictor set. Multiple ML algorithms (RF, GBM, XGBoost, SVM, ANN, Decision Tree (DT), and Linear Regression (LR)) are systematically benchmarked against ARIMA and TBATS under a strictly time-aware rolling validation design to prevent temporal leakage. To enhance interpretability, global and local explainability analyses are conducted using SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME), enabling evaluation of both overall feature influence and event-specific predictor contributions. The entire analytical pipeline is implemented within a transparent *Shiny* and *R Markdown* environment to support reproducibility and potential transferability to other hydrological settings (Molnar, 2023).

By integrating strict temporal validation, comparative benchmarking, explainability diagnostics, and computational reproducibility within a unified framework, the study advances the structured application of ML in hydrological forecasting. Rather than focusing solely on predictive performance, the proposed approach positions ML models as interpretable and

operationally deployable tools for informed water resource management.

2 MATERIALS AND METHODS

2.1 Study area and data

Daily flowrate data were obtained from the State Hydraulic Works (DSİ), while corresponding meteorological variables—precipitation, air temperature, and relative humidity—were provided by the Turkish State Meteorological Service (MGM). The dataset spans the period from January 2017 to December 2022, yielding a total of 2,191 consecutive daily observations.

Daily mean flowrate (FR, $\text{dm}^3 \text{s}^{-1}$) was defined as the response variable. Explanatory variables consisted of contemporaneous and lagged hydro-meteorological predictors derived from precipitation, temperature, and humidity records. Meteorological data were collected from four monitoring stations (stations 1, 2, 3 and 4). These stations capture spatial variability within the watershed and represent distinct hydrological and topographic sub-regions.

The geographic setting of the study area with digital elevation model, river network structure, and the location of meteorological stations is illustrated in the study-area map (Fig. 1).

2.2 Experimental design

Forecasting performance was compared across three model categories: classical time-series approaches (ARIMA, TBATS), interpretable baseline models (DT; LR), and advanced ML algorithms (RF, GBM, XGBoost, SVM, ANN). This structured comparison enables evaluation of predictive gains relative to both traditional statistical frameworks and transparent baseline learners.

The regression tree structure derived from the DT model (Figure 2) illustrates the hierarchical decision process, where lagged FR variables form the primary splitting criteria and meteorological predictors contribute at deeper nodes. This configuration reflects the dominance of short-term temporal persistence while retaining conditional adjustments from hydro-meteorological inputs. For the LR model, FR was expressed as a function of selected lagged FR and meteorological predictors. The general formulation is given by:

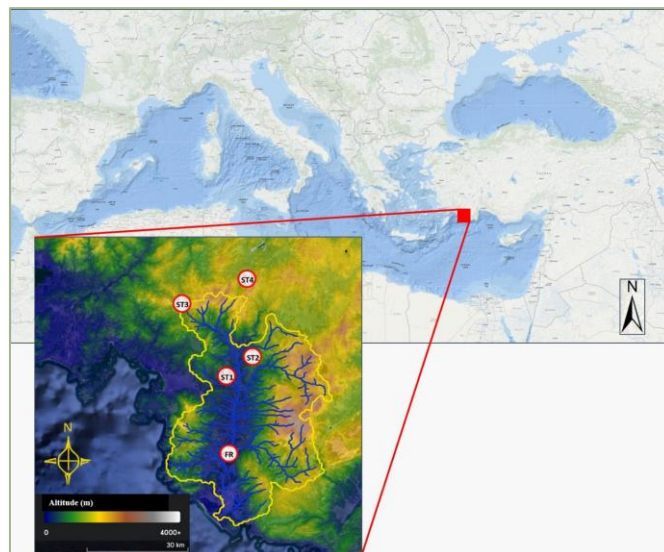


Fig. 1. Eşen River Basin and stations (ST1: Station 1, ST2: Station 2, ST3: Station 3, ST4: Station 4 and FR: Flowrate Station).

$$FR_t = \beta_0 + \sum_{k \in S_{FR}} \alpha_k FR_{t-k} + \sum_{j \in J_P} \sum_{k \in S_{P_j}} \gamma_{j,k} P_{j,t-k} + \sum_{j \in J_T} \sum_{k \in S_{T_j}} \delta_{j,k} T_{j,t-k} + \sum_{j \in J_H} \sum_{k \in S_{H_j}} \eta_{j,k} H_{j,t-k} + \epsilon_t \quad (1)$$

where: FR_t : observed flowrate at time t (dependent variable). β_0 : regression intercept.

FR_{t-k} : flowrate lagged by k days; α_k is the corresponding regression coefficient. $P_{j,t-k}$: precipitation at station j , lagged by k days; $\gamma_{j,k}$ is the corresponding coefficient. $T_{j,t-k}$: temperature at station j , lagged by k days; $\delta_{j,k}$ is the coefficient. $H_{j,t-k}$: humidity at station j , lagged by k days; $\eta_{j,k}$ is the coefficient. S_{FR} , S_{P_j} , S_{T_j} , S_{H_j} : selected sets of lag indices for flowrate and each meteorological variable, determined by the feature selection method. J_P , J_T , J_H : index sets of available precipitation, temperature, and humidity stations/variables. ϵ_t : model residual (error term), assumed to be normally distributed with zero mean.

This formulation explicitly incorporates temporal persistence through lagged FR terms while integrating meteorological lags as exogenous predictors. Although fully interpretable and structurally transparent, the LR model is inherently limited to linear relationships and therefore serves as a benchmark for assessing the additional flexibility provided by nonlinear ensemble methods.

Similarly, DT provides rule-based interpretability but lacks the representational depth of ensemble learners. Consequently, both LR and DT function as reference baselines against which the performance of more flexible ML models can be evaluated.

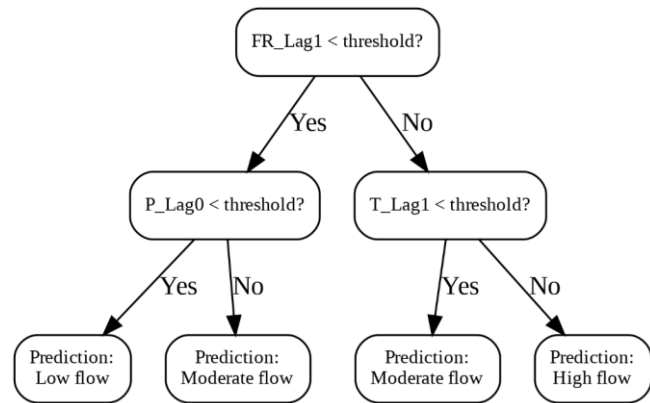


Fig. 2. Regression tree structure for FR forecasting with splits on lagged predictors and terminal nodes indicating flow regimes.

2.2.1 Preprocessing

Data preprocessing was conducted to ensure statistical consistency and modeling robustness. Missing observations were first examined for frequency and temporal clustering; isolated gaps were addressed using linear interpolation to preserve continuity in the daily time series.

Exploratory analysis indicated that the raw flowrate series exhibited positive skewness and variance heterogeneity. Accordingly, a Box-Cox transformation was applied to stabilize variance and approximate normality prior to model development. This transformation supports improved convergence behavior and error stability across both statistical and ML models.

To represent temporal persistence and short-term hydrological memory, lagged predictors were generated for flowrate and meteorological variables up to a maximum of 3 days (Lag1-Lag3). The selection of a 3-day lag window was guided by hydrological and statistical considerations.

Autocorrelation (ACF) and partial autocorrelation (PACF) diagnostics indicated that the strongest statistically significant persistence occurred within the first three days, after which ACF decayed rapidly.

This behavior is consistent with the fast rainfall-runoff response and relatively short concentration time characteristic of Mediterranean-type basins with steep topography and limited storage buffering. Extending the lag window beyond three days did not provide additional predictive gain but increased dimensionality and the risk of over-parameterization. Therefore, a 3-day lag structure was adopted to balance hydrological realism, statistical parsimony, and computational efficiency. Lag construction was performed strictly within the training window to preserve chronological integrity and prevent temporal information leakage.

Potential multicollinearity among predictors was assessed through pairwise correlation analysis. Feature selection was subsequently performed using the Boruta algorithm, which iteratively evaluates variable relevance against randomized shadow features. The procedure identified lagged FR variables among the most influential predictors while retaining selected meteorological lags contributing to flowrate variability.

2.2.2. Forecasting Models

Two categories of forecasting models were evaluated. The first comprised classical time-series approaches, including ARIMA and TBATS models, which are widely used in hydrological applications due to their structural transparency and established performance in seasonal and trend-driven series.

The second category included ML models: RF, GBM, XGBoost, SVM, ANN, DT, and LR. Ensemble tree-based learners (RF, GBM, XGBoost) were selected for their ability to model nonlinear interactions and complex feature dependencies, whereas DT and LR served as interpretable baseline models to provide transparent structural comparisons.

Hyperparameter tuning was systematically performed for ensemble-based learners (RF, GBM, and XGBoost) using grid search embedded within the time-aware validation framework. This approach ensured that parameter selection respected chronological order and avoided temporal leakage. For ANN and SVM models, standard baseline configurations were retained to reflect commonly adopted operational settings and to maintain methodological consistency across model classes. Rather than pursuing exhaustive model-specific optimization, the benchmarking strategy was designed to emphasize comparative structural learning capacity while preserving reproducibility and evaluation fairness.

2.2.3. Evaluation Strategy and Metrics

The forecasting objective was one-day-ahead daily mean flowrate (FR_{t+1}), expressed in $\text{dm}^3 \text{s}^{-1}$. To ensure chronological integrity and eliminate temporal information leakage, a strictly time-aware validation strategy was implemented. Data from 2017 to 2021 were used for model development and hyperparameter tuning via rolling time-slice cross-validation, preserving the sequential structure of the flowrate series. The year 2022 was held out entirely as an independent test set to provide an unbiased out-of-sample evaluation.

Model performance was assessed using both magnitude-sensitive and hydrology-specific metrics. Root mean square error (RMSE) and coefficient of determination (R^2) were used as primary indicators of predictive accuracy, while mean absolute

error (MAE) served as a complementary measure of average deviation.

To evaluate hydrological realism beyond aggregated error magnitude, the Nash-Sutcliffe Efficiency (NSE) and the Kling-Gupta Efficiency (KGE) were also computed. NSE quantifies relative predictive skill compared to the observed mean and is defined as:

$$NSE = 1 - \frac{\sum(Q_{obs} - Q_{sim})^2}{\sum(Q_{obs} - \bar{Q}_{obs})^2} \quad (2)$$

where Q_{obs} denotes observed flowrate, Q_{sim} simulated flowrate, and $\bar{Q}_{obs}$ the mean observed flowrate. NSE ranges from $-\infty$ to 1, with values closer to 1 indicating improved agreement.

KGE evaluates model performance through decomposition into correlation, variability, and bias components:

$$KGE = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad (3)$$

where r represents the Pearson correlation coefficient between observed and simulated flowrate, β the bias ratio, and γ the variability ratio. By jointly assessing these components, KGE provides a balanced evaluation of predictive consistency, variability reproduction, and systematic bias.

The combined use of RMSE, R^2 , MAE, NSE, and KGE allows for a multi-dimensional assessment of model behavior, ensuring that both numerical accuracy and hydrological consistency are considered in performance comparison.

2.2.4. Explainability and Reproducibility

Model interpretability was systematically incorporated through complementary explainable artificial intelligence (XAI) techniques. Global feature attribution was quantified using SHAP, which decomposes model predictions into additive contributions of individual predictors based on cooperative game theory principles (Lundberg and Lee, 2017). In parallel, LIME were employed to generate case-specific explanations, enabling examination of predictor influence under distinct hydrological regimes. The integration of global and local interpretability approaches facilitates both aggregate understanding of model behavior and event-level diagnostic analysis.

Table 1. Descriptive statistics of daily mean flowrate and meteorological variables (2017–2022).

Variables	N	Mean	SD	Min	Max	NA
P1	2191	1.965	6.634	0.000	80.900	0
P2	2191	1.859	6.716	0.000	92.200	0
P3	2191	3.321	11.199	0.000	138.100	0
P4	2191	2.823	9.659	0.000	189.300	0
T1	2191	17.413	7.708	−0.800	33.600	0
T2	2191	18.434	7.739	2.000	37.800	0
T3	2191	14.087	7.509	−4.000	31.100	0
T4	2191	8.882	7.594	−12.300	24.000	0
H1	2191	59.858	20.933	16.100	100.000	0
H2	2191	65.137	17.759	16.500	100.000	0
H3	2191	67.149	19.549	16.100	100.000	0
H4	2191	76.873	18.096	17.700	100.000	0
FR	2191	27834.690	18694.409	8940.000	306000.000	0

All statistical analyses were conducted within the R statistical computing environment (R Core Team, 2024). Model training, hyperparameter tuning, and time-aware resampling were implemented using the caret package (Kuhn, 2008). Ensemble learning algorithms were developed via the randomForest for RF (Liaw & Wiener, 2002), the xgboost package for XGBoost (Chen & Guestrin, 2016), and the gbm package for GBM (Greenwell et al., 2023). SVM and ANN were implemented using e1071 (Meyer et al., 2023) and nnet (Venables & Ripley, 2002) packages, respectively. Classical time-series models, including ARIMA and TBATS, were estimated using the forecast package (Hyndman & Khandakar, 2008). Feature selection was performed using the Boruta package (Kursa & Rudnicki, 2010). Model interpretability analyses were conducted using the iml, SHAPforxgboost, and lime packages (Ribeiro et al., 2016), whereas data visualization was performed with ggplot2 (Wickham, 2016). The interactive analytical interface was developed using the shiny package (Chang et al., 2023).

To enhance methodological transparency and computational reproducibility, the complete *Shiny* application, source code, and supporting documentation are archived in a permanent Zenodo repository (Çakır, 2025) and are publicly accessible at <https://doi.org/10.5281/zenodo.17114915>. This open deployment enables independent replication, scrutiny, and potential adaptation of the proposed framework to other hydrological contexts.

3 RESULTS

3.1 Data Characteristics

Descriptive statistics of daily flowrate and meteorological predictors for the 2017–2022 period are summarized in Table 1. The flowrate series exhibits pronounced positive skewness and substantial dispersion, reflecting the intermittent occurrence of high-magnitude flowrate events. Precipitation variables display strong variability across monitoring stations, while temperature and humidity series show comparatively smoother seasonal oscillations.

3.2 Preprocessing Outcomes

The application of the Box-Cox transformation improved variance homogeneity and reduced distributional asymmetry in the flowrate series, as demonstrated by histogram comparisons and Q-Q diagnostics (Figure 3). The transformed series exhibited more stable dispersion across flowrate magnitudes, supporting improved model convergence and residual stability.

Exploratory analyses of FR-meteorological relationships revealed hydrologically consistent patterns (Figure 4). Scatterplots with fitted regression trends indicated a positive association between relative humidity and flowrate, reflecting moisture availability effects on sustained flow conditions. Precipitation variables showed stronger and more immediate positive relationships with flowrate particularly at higher rainfall magnitudes, consistent with rainfall-runoff response behavior. In contrast, temperature displayed a predominantly negative association with flowrate, suggesting the role of evapotranspiration and seasonal energy balance in modulating flowrate during warmer periods. Collectively, these patterns are consistent with typical hydro-meteorological variability observed in Mediterranean-type basins.

ACF and PACF diagnostics confirmed strong short-term persistence in the flowrate series, supporting the inclusion of lagged predictors (Figure 5). Correlation heatmap analysis indicated multicollinearity among temperature variables, necessitating a structured feature-selection approach (Figure 6).

Boruta-based importance ranking identified lagged FR variables (FR_Lag1-3) among the most influential predictors, reflecting the temporal memory structure of the basin (Figure 7). Selected meteorological variables also contributed to flowrate variability, albeit with spatial and temporal heterogeneity. Temperature variables across stations and their short lags (Lag 0-3) were retained as relevant predictors, while precipitation effects were primarily associated with contemporaneous values at specific stations. Notably, rainfall measured at Stations 1 and 2 exhibited stronger associations with flowrate compared to other locations, whereas precipitation at Station 4 showed comparatively lower importance. This spatial differentiation likely reflects localized meteorological variability across the monitoring stations.

Humidity effects were similarly station-dependent. In particular, humidity recorded at Station 3 (Lag 0-1) displayed higher relative importance compared to other stations. This pattern suggests localized meteorological variability influencing the statistical relationship with flowrate.

Overall, feature-selection results reinforce the dual control of temporal persistence and meteorological forcing in shaping flowrate variability within the basin.

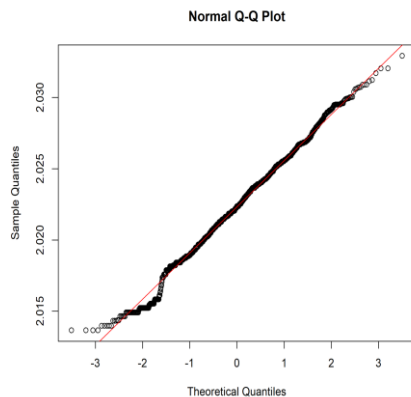


Fig. 3. The Q-Q plot presents the quantiles of the Box-Cox transformed flowrate series against theoretical normal quantiles.

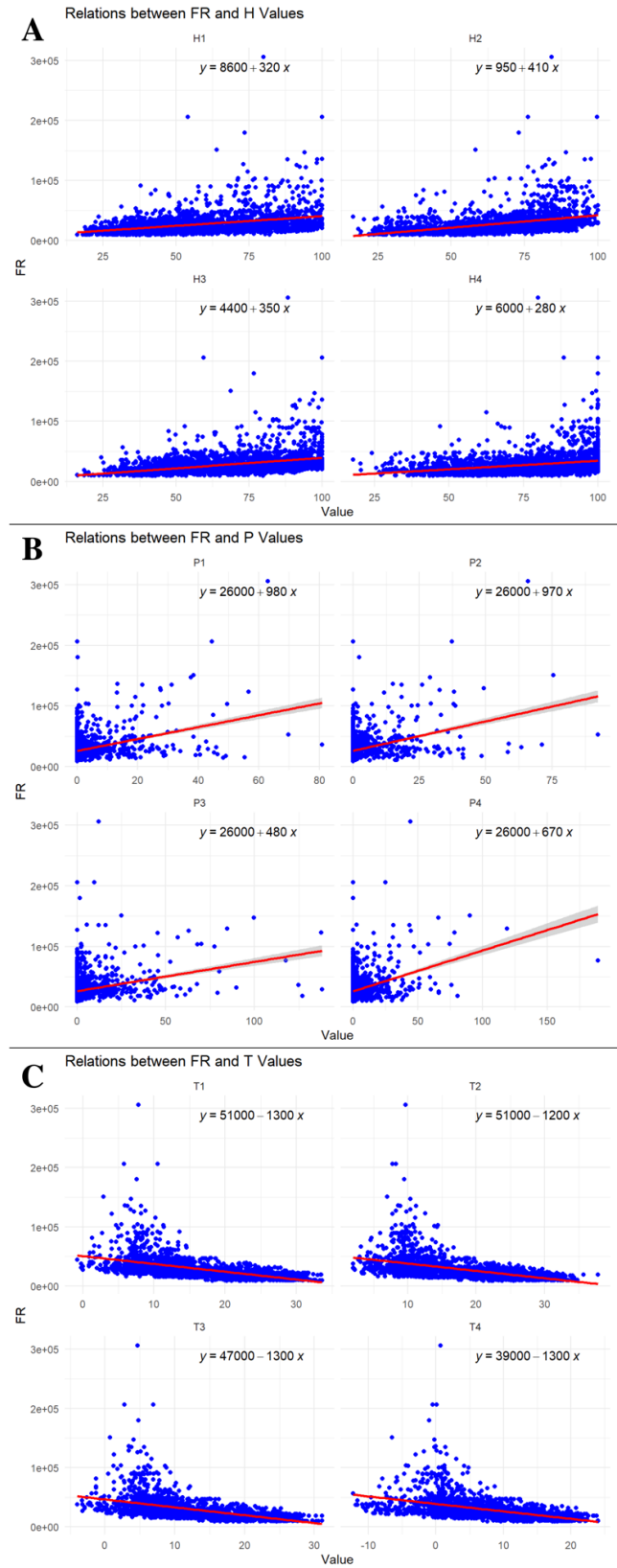


Fig. 4. Relationships between FR (flowrate: $\text{dm}^3 \text{ s}^{-1}$) and meteorological variables. (A) Relationships between FR and humidity variables H1-H4 (humidity: %) with fitted linear regression trends. (B) Relationships between FR and precipitation variables P1-P4 (precipitation: mm); the shaded areas represent the 95% confidence intervals of the fitted linear models. (C) Relationships between FR and temperature variables T1-T4 (temperature: $^{\circ}\text{C}$), showing negative correlations.

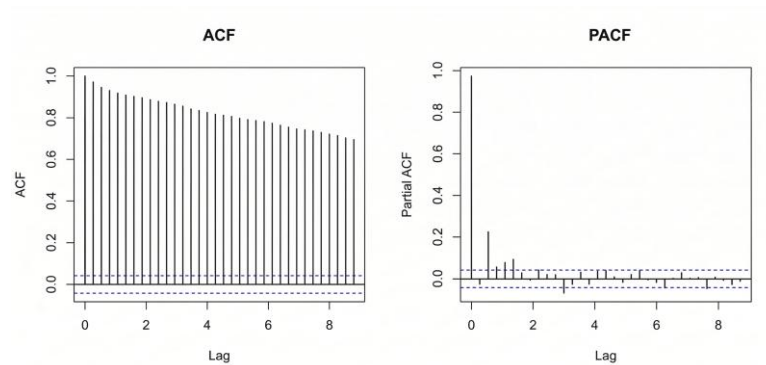


Fig. 5. Autocorrelation (ACF) and partial autocorrelation (PACF) plots of FR with lag expressed in days.

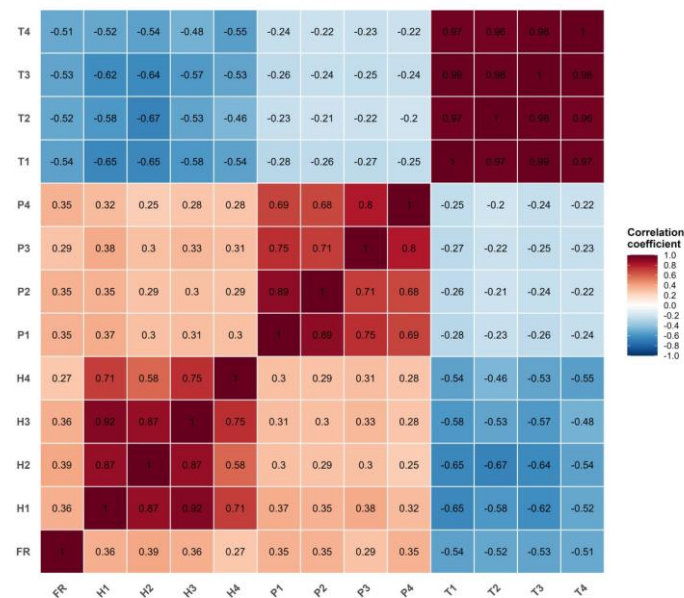


Fig. 6. Correlation heatmap of FR predictors showing strong multicollinearity among temperature variables (T1-T4). *Red colours represent positive (+) correlations, while blue colours indicate negative (-) correlations.

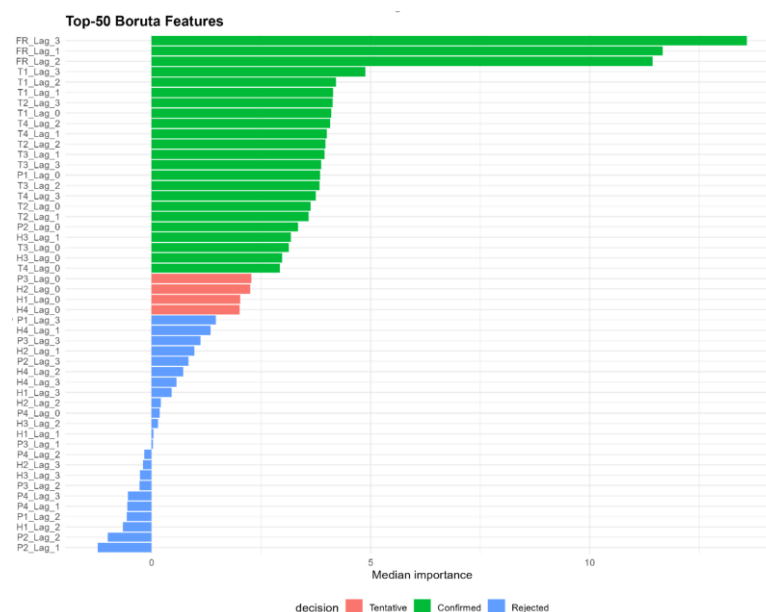
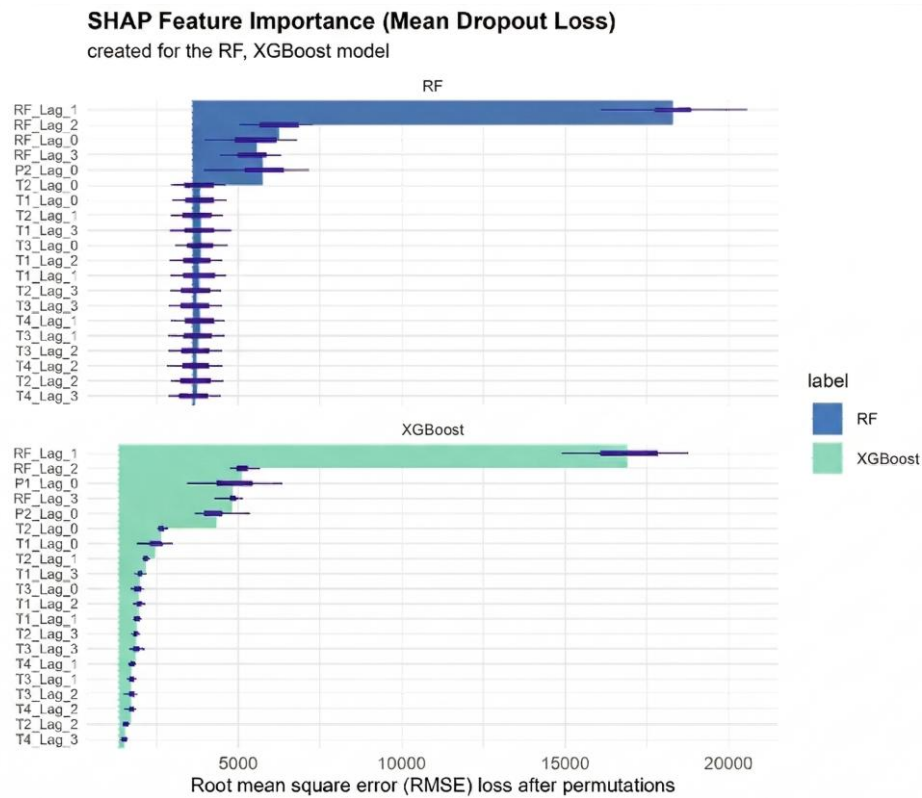


Fig. 7. Boruta-based feature importance ranking showing lagged FR (FR_Lag1-3) as dominant predictors, followed by temperature and precipitation lags.

Table 2. Comparative forecasting performance (2022 hold-out dataset).

Model	Test_RMSE	Test_MAE	Test_R2	Test_NSE	Test_KGE	ModelType
XGBoost	6590.053	2483.552	0.864	0.864	0.814	ML
GBM	6715.247	2826.931	0.859	0.859	0.826	ML
RF	7011.452	2307.629	0.846	0.846	0.784	ML
LR	8598.156	3220.693	0.769	0.769	0.685	ML
ANN	8611.892	3261.744	0.768	0.768	0.684	ML
TBATS	9400.251	4930.901	0.747	0.747	0.865	Time Series
SVM	10361.698	2922.821	0.664	0.664	0.633	ML
DT	11366.752	7261.125	0.596	0.596	0.609	ML
ARIMA	12293.180	5033.816	0.568	0.568	0.781	Time Series

**Fig. 8.** Global feature importance from SHAP values showing lagged FR (FR_Lag1-3) as the most influential predictors, followed by temperature and precipitation lags.

* Boxplots represent the distribution of SHAP values across all observations, where the central line indicates the median, boxes represent interquartile ranges, and whiskers denote variability beyond the upper and lower quartiles.

3.3. Forecasting Performance

Model performance was evaluated using the independent 2022 hold-out dataset to ensure unbiased out-of-sample assessment. Table 2 summarizes RMSE, MAE, R^2 , NSE, and KGE values for all model configurations.

Ensemble tree-based algorithms (XGBoost, GBM, and RF) consistently achieved lower RMSE and MAE values and higher R^2 and NSE scores compared to classical time-series and baseline models. Among these, XGBoost produced the lowest RMSE and the highest R^2 and NSE values, indicating strong predictive precision and improved reproduction of flowrate magnitude.

Classical models (ARIMA and TBATS) exhibited comparatively lower performance in magnitude-sensitive metrics. Interpretable baseline models (DT and LR) provided

transparent structural representations but showed reduced flexibility in capturing nonlinear interactions and complex temporal dependencies.

Notably, TBATS yielded the highest KGE value among all models. Given that KGE decomposes predictive skill into correlation, variability, and bias components, this result suggests that TBATS effectively preserved overall flowrate variability and bias consistency during the evaluation period. In contrast, magnitude-sensitive metrics such as RMSE and NSE more strongly penalize deviations in high-flowrate events. This contrast highlights the complementary insights provided by multi-metric evaluation: while ensemble learners demonstrate superior predictive accuracy in reproducing flowrate magnitude, TBATS maintains competitive skill in capturing overall variability structure.

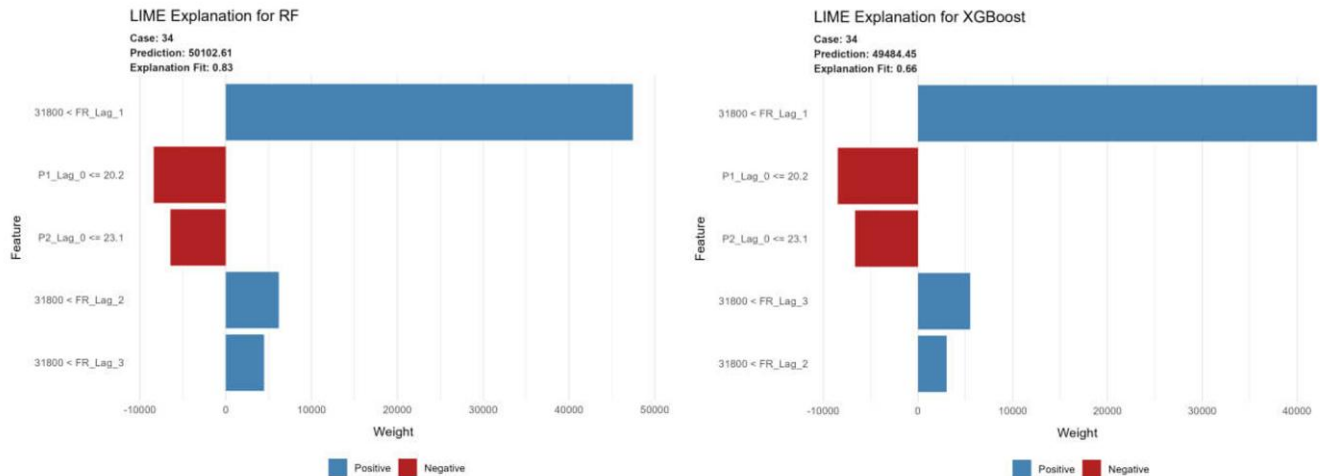


Fig. 9. Local LIME explanations for (a) low-flow conditions and (b) high-flow conditions. Positive contributions increase predicted flowrate, while negative contributions decrease it. Weight indicates the magnitude and direction of each feature's local contribution to the predicted flowrate; positive values increase the prediction, whereas negative values decrease it.

3.4. Explainability of Forecasts

Global SHAP analyses consistently identified lagged FR variables (FR_Lag1-3) as the most influential predictors across ensemble models, followed by selected precipitation and temperature lags (Figure 8). These findings were congruent with Boruta-based feature-selection results, reinforcing the structural importance of temporal persistence in flowrate dynamics.

Local interpretability obtained through LIME provided event-specific diagnostic insights (Figure 9). During high-flow conditions, precipitation and recent flowrate lags exhibited stronger positive contributions to predicted flowrate, consistent with rainfall-driven runoff amplification. In contrast, under low-flow regimes, temperature and humidity variables displayed relatively greater influence, reflecting the role of evapotranspiration and seasonal atmospheric controls in modulating flowrate. The complementary use of global (SHAP) and local (LIME) explanations enables a multi-scale understanding of model behavior, linking predictive performance to hydrological process interpretation.

The pronounced importance of lagged FR predictors should be interpreted within the context of intrinsic flowrate persistence rather than as evidence of model over-reliance on autoregressive structure. Meteorological predictors operate within this persistence-modulated system, contributing to regime shifts and deviations from baseline dynamics, particularly during transitional and high-intensity rainfall periods. Accordingly, the observed lag dominance reflects the statistical temporal structure of the observed flowrate series rather than methodological bias.

3.5. Baseline Interpretable Models

The regression tree derived from the DT model exhibited clear hierarchical partitioning, with FR_Lag1 serving as the primary splitting variable and subsequent nodes incorporating selected meteorological predictors (Figure 10). This structure delineated distinct flowrate regimes while preserving rule-based interpretability. However, the limited representational capacity of a single-tree structure constrained its ability to capture complex nonlinear interactions and higher-order dependencies, which may explain its comparatively lower predictive performance relative to ensemble models.

The linear regression (LR) model, formulated using selected lagged FR and meteorological predictors, provided fully transparent coefficient-based relationships. While structurally interpretable, its assumption of linear additivity restricts flexibility in modeling nonlinear flowrate responses and interaction effects. Consequently, its predictive performance remained moderate when compared with more flexible ML approaches.

Together, DT and LR functioned as interpretable reference baselines, enabling structured comparison between transparent parametric or rule-based models and more adaptive ensemble learners.

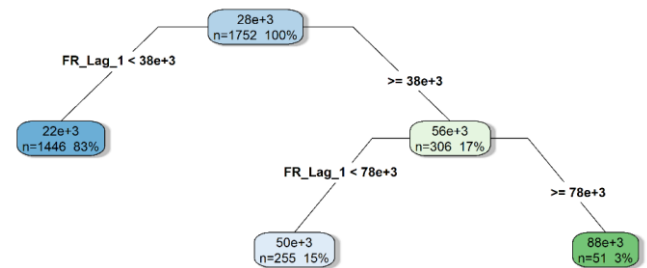


Fig. 10. Decision tree (DT) structure for daily flowrate forecasting, with primary splits on FR_Lag1 and terminal nodes representing low-, medium-, and high-flow conditions.

4 DISCUSSION

The results indicate that ensemble ML approaches offer measurable improvements in magnitude-sensitive performance metrics, particularly RMSE, NSE, and R^2 , when compared with classical time-series models and simpler baseline learners. However, the relatively high KGE value observed for TBATS suggests that classical approaches may retain competitive skill in preserving overall flowrate variability and bias structure. This contrast underscores the importance of multi-metric evaluation in hydrological forecasting, as different indices emphasize distinct aspects of predictive behavior, including magnitude accuracy, variability reproduction, and systematic bias.

Among the evaluated models, XGBoost, GBM, and RF consistently outperformed the classical and baseline approaches in reproducing flow magnitude and temporal dynamics. This

pattern is broadly consistent with prior hydrological forecasting studies conducted across contrasting hydroclimatic settings. For example, in the glaciated Lötschental basin in Switzerland, Hameed et al. (2025) reported that XGBoost achieved the best one month ahead monthly flow forecasting performance, with $R^2 = 0.904$, $RMSE = 1.554 \text{ m}^3 \text{ s}^{-1}$, and $NSE = 0.797$, outperforming RF and LSTM. Likewise, in the Cauvery River basin in India, Naganna et al. (2023) showed that advanced ensemble and deep learning models provided stronger daily flow forecasting skill across multiple stream orders than conventional multilayer perceptron based approaches. In contrast, Szczepanek (2022), working in the mountainous Skawa catchment in Poland, found that although XGBoost was highly competitive for daily flow forecasting, LightGBM and CatBoost outperformed it under some configurations, indicating that model superiority remains catchment specific and sensitive to tuning strategy. Within this context, the present results are both consistent and appropriately bounded: XGBoost delivered the best hold out performance in the Eşen River basin ($R^2 = 0.864$, $NSE = 0.864$, $RMSE = 6590 \text{ dm}^3 \text{ s}^{-1}$), followed closely by GBM ($R^2 = 0.859$, $NSE = 0.859$, $RMSE = 6715 \text{ dm}^3 \text{ s}^{-1}$) and RF ($R^2 = 0.846$, $NSE = 0.846$, $RMSE = 7011 \text{ dm}^3 \text{ s}^{-1}$), whereas ARIMA and TBATS remained less accurate in magnitude sensitive metrics. These comparisons indicate that advanced ensemble learners can provide robust advantages for short term flow forecasting in nonlinear hydrological systems, while also confirming that their relative superiority should be interpreted in relation to basin characteristics, forecast horizon, and model configuration.

A key contribution of the present framework lies in its structured integration of XAI. Global SHAP analyses confirmed the structural importance of lagged FR predictors, while meteorological variables provided regime-dependent contributions.

Local explanations derived from LIME further demonstrated that predictor importance varies across hydrological regimes. Precipitation exerts stronger influence during high-flow episodes, consistent with rainfall-runoff amplification, whereas temperature and humidity contribute more substantially under low-flow conditions, reflecting evapotranspiration controls. The combined use of global and local interpretability techniques therefore enhances understanding of model behavior by linking predictive structure to hydrological process representation.

The inclusion of interpretable baselines (DT and LR) strengthens benchmarking rigor by situating ensemble performance within a transparent modeling context. While these models provide rule-based or coefficient-based interpretability, their structural constraints limit their ability to represent higher-order nonlinear interactions. Consequently, ensemble methods extend predictive flexibility while maintaining interpretability through post-hoc explanation frameworks.

Reproducibility constitutes another structural component of the proposed framework. By implementing preprocessing, feature engineering, model development, validation, and explainability analyses within an openly accessible *Shiny*-based environment, the study promotes computational transparency and facilitates independent verification and adaptation across hydrological settings.

Several limitations should be acknowledged. The analysis is confined to a single basin and a six-year observational record, which restricts direct generalization to other hydro-climatic regimes. The adopted 3-day lag structure captures short-term hydrological memory associated with rainfall-runoff processes. However, longer-term hydrological dynamics may require alternative modeling strategies (e.g., state-space or hybrid approaches) (Bramm et al., 2025) which were beyond the scope

of the present short-term forecasting objective. Additionally, alternative feature-selection techniques (e.g., recursive feature elimination, LASSO, PCA) could provide complementary insights into predictor relevance (Hejduk and Kasprzak, 2010; Pavelková et al., 2012; Balashov et al., 2021; Považanová et al., 2023; Karol, 2025).

Although ensemble models were tuned to optimize predictive capacity, ANN and SVM were evaluated under baseline configurations to reflect operationally realistic conditions. Future investigations may explore uniform hyperparameter optimization across all model classes to further assess comparative performance boundaries.

An additional consideration concerns model behavior under extreme flowrate conditions. While aggregated metrics such as RMSE and NSE capture overall predictive accuracy, they may underrepresent localized errors during peak-flow events. Given the episodic high-intensity rainfall typical of Mediterranean basins, reliable reproduction of rapid flowrate escalation remains critical for operational flood preparedness. Future work should therefore incorporate event-specific or quantile-based evaluation strategies to further examine predictive robustness under extreme hydrological conditions.

5. CONCLUSION

This study presented an integrated and reproducible ML framework for daily flowrate forecasting, combining lag-based feature engineering, Boruta feature selection, ensemble learners, and XAI diagnostics. Ensemble tree-based algorithms (XGBoost, GBM, and RF) demonstrated improved performance in magnitude-sensitive metrics compared with classical time-series models and simpler baselines, highlighting their capacity to capture nonlinear dependencies and short-term temporal persistence.

The integration of SHAP and LIME enabled both global and event-level interpretability, linking predictive behavior to hydrological process representation. The prominence of lagged FR variables reflects intrinsic temporal persistence, while meteorological predictors provide regime-dependent adjustments under varying flow conditions. This combination of predictive precision and interpretability supports the structured application of ML models in hydrological forecasting.

By implementing the complete analytical workflow within an openly accessible *Shiny*-based environment, the study promotes computational transparency and facilitates independent replication and adaptation. Rather than replacing classical approaches, the proposed framework complements existing hydrological tools by enhancing predictive flexibility while maintaining interpretability.

Several avenues for future research remain. Extension to multiple basins, longer observation records, and alternative feature-selection strategies would strengthen generalizability. Future work may also explore additional process-based analyses to complement data-driven forecasting approaches. Additionally, incorporating event-specific or quantile-based evaluation metrics would allow more detailed assessment of model behavior under extreme flowrate conditions. Overall, the findings indicate that ensemble ML models, when embedded within explainable and reproducible workflows, can provide a structured and adaptable approach for improving short-term flowrate forecasting under increasing hydro-climatic variability.

Beyond methodological advancement, the proposed framework has direct operational and policy relevance. In aquaculture-dominated basins such as the Eşen River, short-term flowrate predictability supports proactive stock management,

water intake regulation, and risk mitigation under both high-flow and low-flow conditions. For policymakers, the integration of strictly time-validated and explainable ensemble models provides transparent, evidence-based decision support for water allocation, environmental flow regulation, and basin-scale planning. By combining predictive accuracy with interpretability, the framework strengthens institutional trust and facilitates the responsible integration of data-driven tools into adaptive water governance.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process: The authors used generative AI-assisted tools only for language editing, readability improvement, and stylistic refinement during manuscript preparation. No AI tool was used to generate scientific content, perform data analysis, or contribute to the interpretation of the results. The authors reviewed and verified all revised text and remain fully responsible for the content of the manuscript.”

REFERENCES

- Abderrahmane, B., Chahid, M., Aqnouy, M., Milewski, A. M., Lahcen, B., 2025. Evaluating time series models for monthly rainfall forecasting in arid regions: Insights from Tamanghasset (1953-2021), southern Algeria. *Geosciences*, 15(7), 273. <https://doi.org/10.3390/geosciences15070273>
- Aderemi, I. A., Kehinde, T. O., Ugochukwu, D. O., Ahmad, K. H., Adjei, K. Y., Chijioke, C. E., 2025. Beyond the black box: A systematic review of explainable AI for transparent and trustworthy water quality monitoring. *IEEE Sensors Reviews*. <https://doi.org/10.1109/SR.2025.3595500>
- Ahmed, A. A., Sayed, S., Abdoulhalik, A., Moutari, S., Oyedele, L., 2024. Applications of machine learning to water resources management: A review of present status and future opportunities. *Journal of Cleaner Production*, 441, 140715. <https://doi.org/10.1016/j.jclepro.2024.140715>
- Balashov, E., Buchkina, N., Šimanský, V., Horák, J., 2021. Effects of slow and fast pyrolysis biochar on N2O emissions and water availability of two soils with high water-filled pore space. *Journal of Hydrology and Hydromechanics*, 69(4), 467-474. <https://doi.org/10.2478/johh-2021-0024>
- Berger, K., 2018. Operational validation of HYDRUS (2D/3D) for capillary barriers using data of a 10-m tipping trough. *Journal of Hydrology and Hydromechanics*, 66(2), 153. <https://doi.org/10.1515/johh-2017-0059>
- Bergmeir, C., Benítez, J. M., 2012. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192-213. <https://doi.org/10.1016/j.ins.2011.12.028>
- Bramm, A. M., Matrenin, P. V., Khalyasmaa, A. I., 2025. A review of XAI methods applications in forecasting runoff and water level hydrological tasks. *Mathematics*, 13(17), 2830. <https://doi.org/10.3390/math13172830>
- Brenner, C., Meisch, C., Apperl, B., Schulz, K., 2016. Towards periodic and time-referenced flood risk assessment using airborne remote sensing. *Journal of Hydrology and Hydromechanics*, 64(4), 438-447. <https://doi.org/10.1515/johh-2016-0034>
- Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B., 2023. Shiny: Web Application Framework for R (R package version1.8.0). <https://doi.org/10.32614/CRAN.package.shiny>
- Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. <https://doi.org/10.1145/2939672.2939785>
- Cheng, M., Fang, F., Kinouchi, T., Navon, I. M., Pain, C. C., 2020. Long lead-time daily and monthly flowrate forecasting using machine learning methods. *Journal of Hydrology*, 590, 125376. <https://doi.org/10.1016/j.jhydrol.2020.125376>
- Chiew, F. H. S., Zhou, S. L., McMahon, T. A., 2003. Use of seasonal flowrate forecasts in water resources management. *Journal of Hydrology*, 270(1-2), 135-144. [https://doi.org/10.1016/S0022-1694\(02\)00292-5](https://doi.org/10.1016/S0022-1694(02)00292-5)
- Çakır, M., Yılmaz, M., Ural, G. N., Oral, M. A., Oral, O., 2026. Flowrate forecasting and ecohydrodynamic insights from the Eşen Stream (Southwestern Türkiye): Time series perspectives. *Teknik vjesnik*, 33(4) (in press). <https://doi.org/10.17559/TV-20251119003143>
- Desai, S., Ouarda, T. B., 2021. Regional hydrological frequency analysis at ungauged sites with random forest regression. *Journal of Hydrology*, 594, 125861. <https://doi.org/10.1016/j.jhydrol.2020.125861>
- Elitez, I., Yaltrak, C., 2024. Tectonic geomorphology of the south-eastern section of the Burdur-Fethiye Shear Zone, Western Taurides: morphotectonic framework of the Eşen Basin. *Turkish Journal of Earth Sciences*, 33(5), 503-529. <https://doi.org/10.55730/1300-0985.1927>
- Feng, Z. K., Niu, W. J., Wan, X. Y., Xu, B., Zhu, F. L., Chen, J., 2022. Hydrological time series forecasting via signal decomposition and twin support vector machine using cooperation search algorithm for parameter identification. *Journal of Hydrology*, 612, 128213. <https://doi.org/10.1016/j.jhydrol.2022.128213>
- Greenwell, B. M., Dahlmann, A., Dhoble, S., 2023. Explainable Boosting Machines with Sparsity--Maintaining Explainability in High-Dimensional Settings. *arXiv preprint arXiv:2311.07452*. <https://doi.org/10.48550/arXiv.2311.07452>
- Hameed, M. M., Masood, A., Hamid, A., Elbeltagi, A., Razali, S. F. M., Salem, A., 2025. Forecasting monthly runoff in a glacierized catchment: a comparison of extreme gradient boosting (XGBoost) and deep learning models. *Plos one*, 20(5), e0321008. <https://doi.org/10.1371/journal.pone.0321008>
- Hasan, N. A., Dongkai, Y., Al-Shibli, F., 2023. SPI and SPEI drought assessment and prediction using TBATS and ARIMA models. *Jordan Water*, 15(20), 3598. <https://doi.org/10.3390/w15203598>
- Hečková, P., Bareš, V., Stránský, D., Sněhota, M., 2022. Performance of experimental bioretention cells during the first year of operation. *Journal of Hydrology and Hydromechanics*, 70(1), 42-61. <https://doi.org/10.2478/johh-2021-0038>
- Hejduk, S., Kasprzak, K., 2010. Specific features of water infiltration into soil with different management in winter and early spring period. *Journal of Hydrology and Hydromechanics*, 58(3), 175. <https://doi.org/10.2478/v10098-010-0016-y>
- Hyndman, R. J., Khandakar, Y., 2008. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 1-22. <https://doi.org/10.18637/jss.v027.i03>
- Karakoca, E., Ünver, A., 2024. Analitik hiyerarşi süreci ve coğrafi bilgi sistemleri kullanarak Eşen Çayı Havzası'nda taşkın riski değerlendirmesi ve haritalandırılması. *Geomatik*, 10(1), 127-143. <https://doi.org/10.29128/geomatik.1542251>
- Karol, K., 2025. The average and closure problem of turbulence theory resolved in random space. *Journal of Hydrology and*

- Hydromechanics, 73(4), 333-353. <https://doi.org/10.2478/johh-2025-0026>
- Kozel, T., Stary, M., 2022. Adaptive stochastic management of the storage function for a large, open reservoir using learned fuzzy models. *Journal of Hydrology and Hydromechanics*, 70(2), 213-221. <https://doi.org/10.2478/johh-2022-0010>
- Kuhn, M., 2008. Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1-26. <https://doi.org/10.18637/jss.v028.i05>
- Kursa, M. B., Rudnicki, W. R., 2010. Feature selection with the Boruta package. *Journal of Statistical Software*, 36(11), 1-13. <https://doi.org/10.18637/jss.v036.i11>
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R News*, 2(3), 18-22. <https://doi.org/10.32614/RN-2002-022>
- Lundberg, S. M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765-4774. <https://doi.org/10.48550/arXiv.1705.07874>
- Mantovani, D., Veste, M., Gypser, S., Halke, C., Koning, L., Freese, D., Lebzien, S., 2014. Transpiration and biomass production of the bioenergy crop Giant Knotweed *Igniscum* under various supplies of water and nutrients. *Journal of Hydrology and Hydromechanics*, 62(4), 316. <https://doi.org/10.2478/johh-2014-0028>
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., 2023. e1071: Misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien (R package version 1.7-13). <https://doi.org/10.32614/CRAN.package.e1071>
- Mishra, S., Kaushal, D. R., 2025. Highly concentrated iron ore slurry flow through pipeline with and without chemical additive; part I: Experimental investigations and proposed model for the prediction of pressure drop. *Journal of Hydrology and Hydromechanics*, 73(2), 143-154. <https://doi.org/10.2478/johh-2025-0011>
- Molnar, C., 2023. Interpretable machine learning (2nd ed.). <https://doi.org/10.5281/zenodo.8213525>
- Moore, D., Kostka, S., Boerth, T., Franklin, M., Ritsema, C., Dekker, L., Oostindie, K., Stoof, C., Wesseling, J., 2010. The effect of soil surfactants on soil hydrological behavior, the plant growth environment, irrigation efficiency and water conservation. *Journal of Hydrology and Hydromechanics*, 58(3), 142. <https://doi.org/10.2478/v10098-010-0013-1>
- Naganna, S. R., Marulasiddappa, S. B., Balreddy, M. S., Yaseen, Z. M., 2023. Daily scale flowrate forecasting in multiple stream orders of Cauvery River, India: Application of advanced ensemble and deep learning models. *Journal of Hydrology*, 626, 130320. <https://doi.org/10.1016/j.jhydrol.2023.130320>
- Najafabadi, E. F., Afzalimehr, H., Sui, J., 2015. Turbulence characteristics of favorable pressure gradient flows in gravel-bed channel with vegetated walls. *Journal of Hydrology and Hydromechanics*, 63(2), 154. <https://doi.org/10.1515/johh-2015-0019>
- Núñez, J., Cortés, C. B., Yáñez, M. A., 2023. Explainable artificial intelligence in hydrology: Interpreting black-box snowmelt-driven flowrate predictions in an arid Andean basin of north-central Chile. *Water*, 15(19), 3369. <https://doi.org/10.3390/w15193369>
- Orfanus, T., Amer, A. M. M., Jozefaciuk, G., Fulajtar, E., Čelková, A., 2017. Water vapour adsorption on water repellent sandy soils. *Journal of Hydrology and Hydromechanics*, 65(4), 395. <https://doi.org/10.1515/johh-2017-0030>
- Pavelková, H., Dohnal, M., Vogel, T., 2012. Hillslope runoff generation-comparing different modeling approaches. *Journal of Hydrology and Hydromechanics*, 60(2), 73-86. <https://doi.org/10.2478/v10098-012-0007-2>
- Považanová, B., Čistý, M., Bajtek, Z., 2023. Using feature engineering and machine learning in FAO reference evapotranspiration estimation. *Journal of Hydrology and Hydromechanics*, 71(4), 425-438. <https://doi.org/10.2478/johh-2023-0032>
- R Core Team., 2024. R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://doi.org/10.32614/R-Core>
- Ribeiro, M. T., Singh, S., Guestrin, C., 2016. Why should I trust you? Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144. <https://doi.org/10.1145/2939672.2939778>
- Saci, R., Keblouti, M., Katipoğlu, O. M., Durin, B., Majour, H., Sayad, L., Bouzahar, F., Benchaiba, L., 2025. Assessing the efficacy of various predictive models in simulating monthly reference evapotranspiration patterns and its impact on water resource management for agriculture in the Kebir-West watershed, North-East of Algeria. *Journal of Hydrology and Hydromechanics*, 73(3), 284-294. <https://doi.org/10.2478/johh-2025-0022>
- Sadeghi-Pouya, A., Nouri, J., Mansouri, N., Kia-Lashaki, A., 2017. Developing an index model for flood risk assessment in the western coastal region of Mazandaran, Iran. *Journal of Hydrology and Hydromechanics*, 65(2), 134-145. <https://doi.org/10.1515/johh-2017-0007>
- Schacht, K., Marschner, B., 2015. Treated wastewater irrigation effects on soil hydraulic conductivity and aggregate stability of loamy soils in Israel. *Journal of Hydrology and hydromechanics*, (1), 47-54. <https://doi.org/10.1515/johh-2015-0010>
- Shumova, N., 2009. Crop water supply and its relation to yield of spring wheat in the south of Russian plain. *Journal of Hydrology and Hydromechanics*, 57(1), 26. <https://doi.org/10.2478/v10098-009-0003-3>
- Song, Z., Xia, J., Wang, G., She, D., Hu, C., Hong, S., 2022. Regionalization of hydrological model parameters using gradient boosting machine. *Hydrology and Earth System Sciences*, 26(2), 505-524. <https://doi.org/10.5194/hess-26-505-2022>
- Szczepanek, R., 2022. Daily flowrate forecasting in mountainous catchment using XGB, LightGBM and CatBoost. *Hydrology*, 9(12), 226. <https://doi.org/10.3390/hydrology9120226>
- Tareke, K. A., Awoke, A. G., 2023. Hydrological drought forecasting and monitoring system development using artificial neural network (ANN) in Ethiopia. *Heliyon*, 9(2). <https://doi.org/10.1016/j.heliyon.2023.e13287>
- Tatar, O., 2025. Editorial Note to the article “Tectonic geomorphology of the south-eastern section of the Burdur-Fethiye Shear Zone, Western Taurides: morphotectonic framework of the Eşen Basin” [Turkish Journal of Earth Sciences 33 (5) 2024 503-529]. *Turkish Journal of Earth Sciences*, 34(5), 685-685. <https://doi.org/10.55730/1300-0985.1983>
- ten Veen, J. H., 2004. Extension of Hellenic forearc shear zones in SW Turkey: the Pliocene–Quaternary deformation of the Eşen Çay Basin. *Journal of Geodynamics*, 37(2), 181-204. <https://doi.org/10.1016/j.jog.2004.02.001>

- Venables, W. N., Ripley, B. D., 2002. Modern applied statistics with S (4th ed.). Springer. <https://doi.org/10.1007/978-0-387-21706-2>
- Vyshnevskiy, V., Shevchuk, S., 2021. Thermal regime of the Dnipro Reservoirs. Journal of Hydrology and Hydromechanics, 69(3), 300-310. <https://doi.org/10.2478/johh-2021-0016>
- Wickham, H., 2016. ggplot2: Elegant graphics for data analysis. Springer. <https://doi.org/10.1007/978-3-319-24277-4>
- Yaseen, Z. M., 2023. A new benchmark on machine learning methodologies for hydrological processes modelling: A comprehensive review for limitations and future research directions. Knowledge-Based Engineering and Sciences, 4(3), 65-103. <https://doi.org/10.51526/kbes.2023.4.3.65-103>
- Zhang, J., Li, J., Shi, X., 2018. Encounter probability analysis of irrigation water and reference crop evapotranspiration in irrigation district. Journal of Hydrology and Hydromechanics, 66(3), 279-284. <https://doi.org/10.2478/johh-2018-0015>

Received 26 December 2025

Accepted 26 April 2026