Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 46 No. 12
December 2025

Comparative Evaluation of Machine Learning Models for Multi-Lead-Time Water-Level Forecasting in the Gandak River Basin

Nishant Kumar [1], Sushant Kumar [2], Vikash Kumar [3], Sumit Kumar [4], Avinash Kumar [5*], Anand Kumar [6]

[1]Assistant Professor, Department of Computer Science and Engineering, Sitamarhi Institute of Technology, Sitamarhi, India

[2]Assistant Professor, Department of Civil Engineering, Sitamarhi Institute of Technology, Sitamarhi, India

[3]Assistant Professor, Department of Computer Science and Engineering, Sitamarhi Institute of Technology, Sitamarhi, India

[4]Assistant Professor, Department of Civil Engineering, Sitamarhi Institute of Technology, Sitamarhi, India

[5]Assistant Professor, Department of Civil Engineering, Sitamarhi Institute of Technology, Sitamarhi, India

[6]Assistant Professor, Department of Civil Engineering, GEC Sheikhpura, Sheikhpura, Bihar, India

***Corresponding Author Email**: [5] avinash.ce@sitsitamarhi.ac.in

Abstract

Accurate short- and medium-range water-level forecasting is vital for effective flood management and reservoir operation in monsoon-driven basins. This study evaluates the performance of three machine-learning models: Random Forest (RF), Support Vector Regression (SVR), and K-Nearest Neighbors (KNN) for one-, five-, and ten-day-ahead water-level prediction at the Triveni gauging station in the Gandak River Basin, India. Daily water level and discharge data from the Central Water Commission (CWC) and rainfall data from three Indian Meteorological Department (IMD) stations (Balmiki, Bagaha, and Ramnagar) for the period 2003–2023 were used as inputs. Model performance was assessed using the coefficient of determination (R^2), root-mean-square error (RMSE), and mean absolute error (MAE). Results showed that all models performed well for one-day forecasting, with RF achieving the highest accuracy (R^2 = 0.941, RMSE = 0.296 m) and SVR yielding the smallest average error (MAE = 0.145 m). As forecast lead time increased, performance declined gradually; however, both RF and SVR maintained R^2 values above 0.90 even at ten days, indicating strong temporal persistence. KNN performed satisfactorily for short horizons but showed higher dispersion for longer leads. Overall, RF demonstrated the best balance between accuracy and stability, while SVR provided superior average error control. These findings highlight the potential of ensemble and kernel-based models for improving real-time flood forecasting and water-level management in data-scarce river basins.

Keywords: Machine learning, RF, SVR, KNN, Water level forecasting.

1. Introduction

Accurate forecasting of water levels in rivers, reservoirs and monitoring stations is a critical component of hydrological management, flood early-warning, and sustainable water-resources planning. The behavior of stage (water-level) fluctuations is subject to antecedent conditions, rainfall–runoff responses, upstream inflows, channel geometry, and basin-scale hydrometeorological variability. In flood-prone basins such as the Gandak Basin of northern India, where monsoonal influxes, upstream Himalayan contributions, and rapid channel responses coexist, the ability to reliably forecast water level is especially valuable for reservoir operation, flood mitigation and community resilience.

Historically, physically-based hydrological models (e.g., rainfall–runoff models, conceptual stage models) have been applied for water-level forecasting. However, such models demand detailed data (e.g., soil moisture, full precipitation networks, channel storage), extensive calibration, and often struggle to capture rapid nonlinear responses in data-scarce or highly dynamic basins. Consequently, data-driven approaches, particularly machine-learning (ML) algorithms have gained popularity because they can learn nonlinear relationships directly from hydrometeorological and

hydrological time-series data, require fewer explicit assumptions, and often yield competitive predictive performance [1, 2]. For instance, review studies have shown that ML and deep-learning techniques have delivered significant improvements in hydrological forecasting tasks such as streamflow, groundwater level and water-level prediction [3-6].

Among ML techniques, certain algorithms stand out for their regular usage, interpretability and performance in hydrological contexts. Random Forest (RF) stands out as a robust ensemble learning method built by aggregating multiple decision trees, each trained on bootstrapped samples and splitting features randomly, thereby reducing variance and overfitting. Since its introduction by Breiman (2001) [7], RF has been extensively applied in hydrological forecasting tasks, including flood susceptibility modelling [8], runoff forecasting, and reservoir level prediction. Its ability to handle high-dimensional input, nonlinear interactions, and noisy data render it well-suited for water-level prediction. For example, in a sea-level forecasting study along the Mediterranean coast, Karsavran (2021) [9], found that RF achieved R^2 values of 0.80 for one-day ahead and 0.63 for two-day ahead predictions, outperforming SVR and KNN in that setting. Similarly, Bargam et al. (2024) [10], reported that RF improved streamflow predictions when compared with multiple linear regression (MLR) and SVR in a watershed context. These findings underscore RF's adaptability and relative reliability in hydrological prediction.

Support Vector Regression (SVR) is a kernel-based machine-learning algorithm derived from support vector machines, extended to regression tasks by minimizing a margin of error in a high-dimensional feature-space. SVR is widely used in applications where the underlying relationship is nonlinear but smooth, and where data may be limited. In hydrology, SVR has been used for streamflow forecasting, groundwater level prediction, and water-level estimation. For instance, a recent study by Bargam et al. (2024) compared SVR, RF and MLR for streamflow prediction, and found that while RF delivered better MAE, SVR achieved lower RMSE in certain datasets [10]. In water-level forecasting, Onay (2021) applied SVR among other ML methods and documented its suitability for short-lead horizons and small datasets [11]. SVR's strengths include generalization capability and resistance to overfitting if properly tuned; however, it may require careful kernel selection and hyperparameter optimization.

K-Nearest Neighbors (KNN) is a simple, non-parametric, instance-based learning algorithm, which predicts outputs by averaging or weighting the outputs of the K most similar historical instances in feature space. While less frequently used in hydrology than RF or SVR, KNN offers advantages of conceptual simplicity and transparency. In groundwater modelling, Kombo et al. (2020) used a hybrid KNN-RF scheme for seasonal groundwater level prediction and demonstrated that the combination leveraged the local neighbouring information of KNN with the ensemble strength of RF, achieving high R^2 and low RMSE in a data-scarce environment. In water-level prediction, Onay (2021) found KNN performed comparably to RF and SVR for very short lead times in coastal settings [11]. However, KNN may struggle with high dimensional inputs or rapidly shifting distributions (for example during flood peaks) because local neighbours may not represent extreme conditions adequately.

While each of RF, SVR and KNN has been applied individually across various hydrological forecasting tasks (water level, groundwater level, streamflow), comparative studies that benchmark them specifically for one-day-ahead water-level forecasting are relatively limited. A particularly noteworthy gap emerges for river-basin stage prediction in Indian monsoonal environments. For example, the review by "Developments & Trends in Water Level Forecasting" (2023) [12], emphasised that most research to date focusses on longer lead times or on deep learning methods (e.g., LSTM, CNN) rather than comparing classical ML algorithms side-by-side. Moreover, studies frequently evaluate discharge or runoff rather than stage (water level) itself, despite the fact that stage is a more operationally relevant variable for flood management. For instance, Yeh et al. (2021) and other reviews report abundant research on groundwater and streamflow forecasting, but fewer on river water level forecasting using RF, SVR and KNN [13]. Furthermore, many past studies have not clearly separated wet and dry seasons, or focused on peak capture (flood peaks) which is crucial for early warning [14].

In the context of the Gandak Basin—an area characterized by rapid monsoonal inflows, upstream snowmelt contributions, and congested channel dynamics—these gaps are especially significant. Given the practical need for short-lead (one-day) forecasting

of water level at the key gauging station (Triveni), a dedicated comparative evaluation of RF, SVR and KNN, integrating antecedent water level, discharge, and rainfall features, is warranted. In this light, the present study aims to bridge this knowledge gap by developing and comparing predictive frameworks based on RF, SVR and KNN for one-day ahead water-level forecasting at the Triveni station in the Gandak Basin. The research investigates the predictive skill of each algorithm, examines their ability to capture day-to-day fluctuations (including peak events), and identifies the best performing method based on standard metrics—coefficient of determination (R^2), root-mean-square error (RMSE) and mean-absolute error (MAE). By systematically assessing these models on the same dataset, the work aims to provide actionable insights into model selection for operational water-level forecasting and strengthen early-warning capacity in flood-prone basins.

In summary, while the application of ML models for hydrological forecasting has matured substantially, and algorithms like RF, SVR and KNN are well-established, their comparative performance for short-lead water-level forecasting in data-driven, monsoon-dominated river basins is under-explored. This research addresses this gap and provides evidence-based guidance for model selection, thereby contributing to improved flood preparedness, basin management and resilience strategies in the Gandak Basin.

2. Study Area and Data

The Gandak River Basin covers an approximate area of 6624 km², extending between 27°26′10″N, 83°53′53″E and 25°38′18″N, 85°11′33″E in eastern India. The river originates at an elevation of about 7620 meters near the Nepal–Tibet border, north of the Dhaulagiri Mountain range. Flowing southward through steep Himalayan valleys, the Gandak enters Indian territory in Champaran district, Bihar, where it serves as a vital source of irrigation, drinking water, and sustenance for the local population [15]. Within India, the river flows southwest and then southeast along the Uttar Pradesh–Bihar border, traversing the fertile Gangetic plains before joining the Ganga River near Sonepur, close to Patna [16]. The basin elevation declines sharply from the Himalayan foothills to approximately 31 meters above mean sea level near its confluence with the Ganga. Its six principal tributaries—Kali, Seti, Marshandi, Budhi, Trisuli, and Rapti—drain distinct physiographic zones, contributing to pronounced hydrological variability across the basin [17].

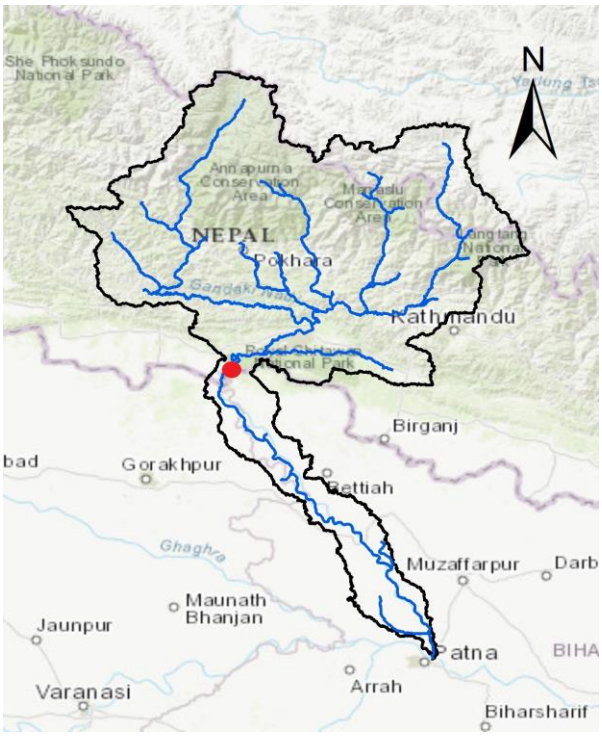

Figure 1. Study area showing Gandak basin with river stream

Climatically, the basin exhibits considerable spatial and temporal heterogeneity. The northern Himalayan region experiences a cold alpine to tundra climate, while the southern plains have a humid subtropical regime [18]. Annual rainfall varies from over 2500 mm in the northern mountainous catchments to 1200–1500 mm in the plains, with nearly 80% of the precipitation occurring during the southwest monsoon (June–September) [19]. The hydrological response of the basin is thus governed by monsoon-driven rainfall, snowmelt, and upstream inflow variability from Nepal's sub-catchments. Figure 1 presents the location and physiographic extent of the Gandak Basin.

For this study, daily discharge and water level data for the Triveni gauging station were obtained from the Central Water Commission (CWC) for the period 2003–2023. These long-term datasets capture the basin's hydrodynamic variability under diverse climatic conditions and flow regimes. Corresponding daily rainfall data for the same period were collected from three nearby Indian Meteorological Department (IMD) stations—Balmiki, Bagaha, and Ramnagar—representing spatial rainfall gradients across the basin (IMD, 2023) [20]. The collected datasets underwent rigorous quality control and preprocessing, including

chronological alignment, interpolation of missing values, and outlier detection, to ensure consistency and continuity for model development. The integrated hydro-meteorological dataset thus served as the foundation for building and validating the machine-learning models for short- to medium-term water-level forecasting in the Gandak River Basin.

3. Methodology

The methodological framework adopted in this study integrates multiple data-driven algorithms to forecast daily water levels at the Triveni gauging station of the Gandak River Basin for three forecast horizons: one-day, five-day, and ten-day ahead. The approach combines the strengths of ensemble-based, kernel-based, and instance-based learning techniques—Random Forest (RF), Support Vector Regression (SVR), and K-Nearest Neighbors (KNN)—to provide a comprehensive comparison of their predictive capability and generalization performance. The overall procedure involved sequential stages of data preprocessing, feature derivation, target formulation for multiple lead times, model development, validation, and comparative evaluation. Each stage was designed to ensure that the forecasting system accurately captured the nonlinear and temporal dependencies of hydrological variables that control water-level dynamics in a monsoon-driven catchment such as the Gandak Basin.

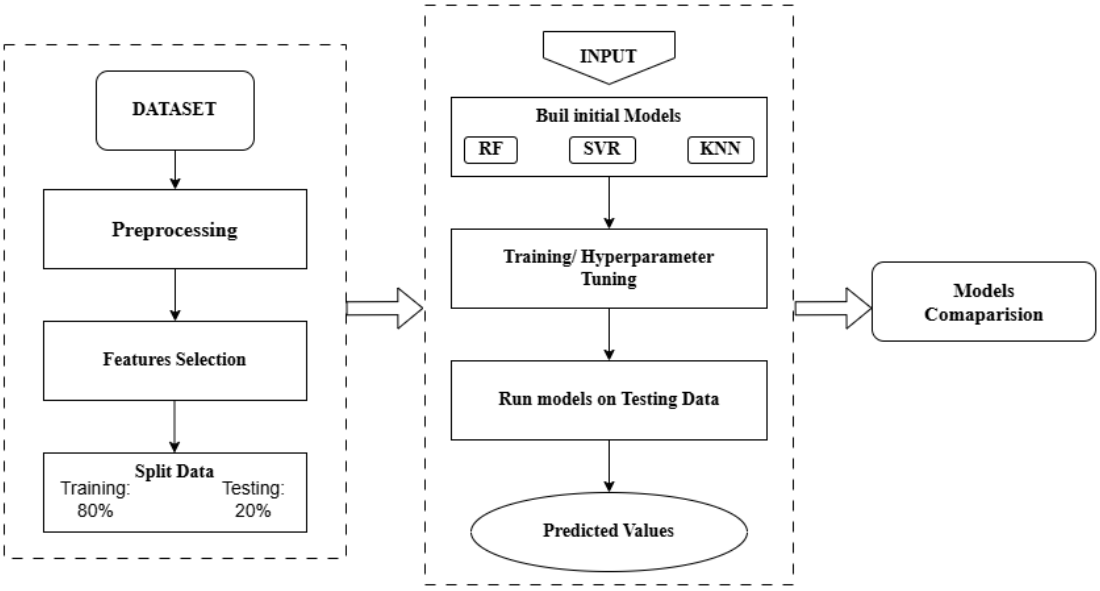

Figure 2. Methodology Flow chart

Hydro-meteorological datasets often contain inconsistencies due to sensor malfunctions, recording errors, or gaps in observation. Hence, the raw data consisting of daily records of water level, discharge, and rainfall from three India Meteorological Department (IMD) stations—Balmi, Bagaha, and Ramnagar—were subjected to quality control before modelling. Missing observations, which constituted less than one percent of the total dataset, were linearly interpolated to maintain temporal continuity, while apparent outliers were detected through the interquartile-range criterion and replaced by median values. The variables were then temporally synchronized so that each observation represented the same day across all stations, ensuring the consistency of the predictor–response pairs. After cleaning, exploratory analysis revealed strong monsoonal seasonality and significant one-day and two-day autocorrelation in the water-level and discharge series, confirming the necessity of incorporating lagged features to capture short-term hydrological memory.

Feature engineering was performed to derive predictors that could represent both immediate and delayed catchment responses. Antecedent values of discharge, rainfall, and water level were generated as lag variables to incorporate temporal dependencies, while short-term averages were calculated to represent hydrological storage and infiltration processes. Specifically, previous-day discharge, rainfall, and water level were included to account for local memory effects, while rolling three-day means of discharge and rainfall were computed to smooth high-frequency fluctuations and represent basin-scale retention. In addition, rainfall from the three stations was averaged

to generate a spatially representative rainfall index for the catchment. The inclusion of these derived variables ensured that the model could capture not only instantaneous runoff contributions but also delayed groundwater and surface-water interactions influencing the next-day or next-week stage variation. All predictor variables were standardized using z-score normalization to achieve zero mean and unit variance, which prevents any single variable from dominating the model training due to differences in measurement scales.

To assess the temporal persistence and degradation of model accuracy with increasing forecast horizon, three separate predictive scenarios were constructed corresponding to one-day, five-day, and ten-day lead times. For each case, the target variable was defined as the water level shifted forward by the respective number of days, i.e., WL_{t+1}, WL_{t+5}, and WL_{t+10}. This setup enabled the evaluation of each algorithm's ability to learn from antecedent hydrological conditions and generalize over different temporal windows. The dataset was divided chronologically into 80 percent for training and 20 percent for testing to preserve the sequential nature of time-series data and to avoid information leakage. All three models were trained independently for each lead time, resulting in a total of nine forecasting models. Each model was developed and optimized using the same input features and evaluation criteria to ensure a fair basis for comparison.

3.1. Models

3.1.1. Random Forest

The Random Forest algorithm, introduced by Breiman (2001), was implemented as the ensemble-based model in this study. RF operates by constructing multiple decorrelated decision trees using random subsets of both samples and predictors, and then averaging the individual tree predictions to obtain a final output. This process effectively reduces model variance, mitigates overfitting, and provides a stable estimate of nonlinear relationships. For the present analysis, the number of trees, the maximum depth of each tree, and the minimum samples per node were optimized using a grid-search procedure combined with five-fold cross-validation. The criterion for split selection was mean squared error minimization, and the random seed was fixed to ensure reproducibility. The final RF model not only generated predictions but also provided feature importance rankings, which were subsequently analyzed to interpret the relative influence of discharge, rainfall, and antecedent water level on future stage variations. Owing to its ensemble structure and capability to handle multicollinearity, RF has been widely applied in hydrological prediction problems such as flood susceptibility mapping and streamflow forecasting [21, 22].

3.1.2. Support Vector Regression

Support Vector Regression (SVR), is based on the structural-risk-minimization principle of the support-vector machine framework [23]. SVR seeks to determine a regression function that balances the trade-off between model flatness and prediction accuracy by minimizing a cost function defined by an ε-insensitive loss. Given the nonlinear characteristics of hydrological processes, the radial-basis-function (RBF) kernel was adopted to project input variables into a higher-dimensional feature space where linear regression could be effectively performed. The main parameters influencing SVR performance—the penalty coefficient (C), kernel width (γ), and the ε margin—were tuned through grid search within empirically justified ranges. Parameter optimization aimed at minimizing the root-mean-square error (RMSE) on the validation folds. SVR was trained separately for the three lead times, thereby allowing assessment of its performance stability as the physical relationship between predictors and target weakened with increasing temporal distance. Owing to its strong generalization capability and ability to model smooth nonlinear dependencies, SVR has been widely used in water-level, streamflow, and sediment-transport forecasting [24, 25].

3.1.3. K-Nearest Neighbors

K-Nearest Neighbors (KNN), represents a non-parametric, instance-based approach where predictions are generated by averaging the outcomes of the *k* closest samples in feature space. The closeness between observations was measured using Euclidean distance, and both uniform and distance-weighted schemes were tested. The optimal number of neighbors was determined through cross-validation using RMSE as the evaluation criterion. Although relatively simple, KNN provides an interpretable benchmark and is particularly useful for exploring local similarities between antecedent and subsequent hydrological states. Its performance in this study was expected to decrease with longer lead times as the

similarity between present and future conditions diminishes, a trend commonly observed in data-driven hydrological applications [11, 26].

All models were implemented in the Python programming environment (version 3.12) using the *scikit-learn* library, with additional support from *pandas* and *numpy* for data handling. Computations were performed in Google Colab, ensuring consistent runtime settings and accessibility. Each experiment was repeated ten times with the same random seed to ensure reproducibility, and the mean performance statistics were reported. The training of RF, SVR, and KNN models was carried out independently for each lead-time scenario, with their parameters tuned through grid search using five-fold cross-validation to avoid overfitting. The models were then evaluated on the independent testing subset spanning 2019–2023, which was not used during any phase of model calibration.

The performance of all models was evaluated using three statistical measures: the coefficient of determination (R^2), root-mean-square error (RMSE), and mean absolute error (MAE). These indices were selected because they jointly capture the explanatory power, precision, and reliability of the forecasts. R^2 quantifies the proportion of observed variance explained by the model, RMSE penalizes larger errors more severely and is sensitive to peak deviations, while MAE provides an average measure of absolute deviation, reflecting overall forecast reliability [27]. The mathematical expressions for these metrics follow the standard formulations adopted in hydrological literature [28]. For visual inspection, predicted and observed water levels were plotted over time, along with 1:1 scatter plots, to detect systematic biases and assess model behavior under varying flow regimes.

4. Result

The comparative performance of the Random Forest (RF), Support Vector Regression (SVR), and K-Nearest Neighbors (KNN) models for one-day, five-day, and ten-day ahead water-level forecasts at the Triveni gauging station is summarized in Table 1 and visualized in Figures 3–5. Each figure contains two components: the forecasting hydrographs showing observed and predicted water levels through time, and the corresponding scatter plots depicting the relationship between observed and simulated values. Together, these provide a comprehensive depiction of the models' predictive strength, bias structure, and generalization capability across increasing lead times.

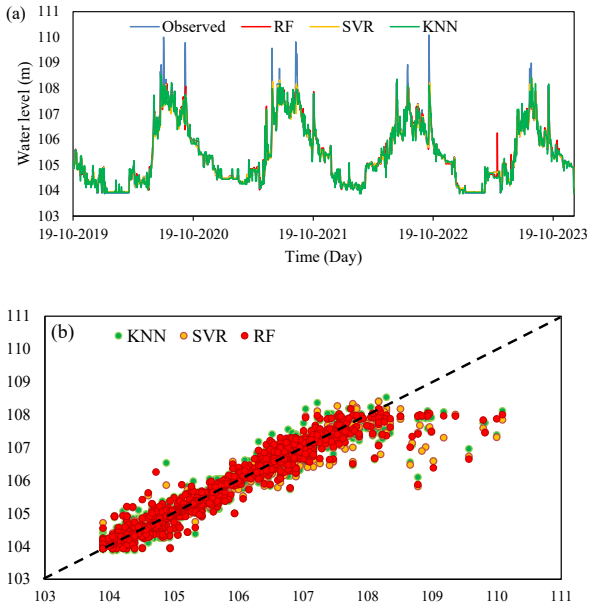

Figure 3: Forecasting line plots (a) and scattered plot (b) for lead times of 1-day

For the one-day-ahead forecasts, all three models demonstrated excellent agreement between predicted and observed water levels. The hydrograph (Fig. 3a) reveals that each model accurately tracked daily fluctuations, capturing both rising and recession limbs of the hydrograph with negligible lag. The RF model achieved the highest coefficient of determination (R^2 = 0.941) and the lowest root-mean-square error (RMSE = 0.296 m), indicating strong fidelity in reproducing daily variations. SVR followed closely (R^2 = 0.938; RMSE = 0.304 m), but it yielded the smallest mean absolute error (MAE = 0.145 m), reflecting more precise alignment in the central flow regime. KNN also performed robustly (R^2 = 0.930; RMSE = 0.322 m), though its predictions were slightly smoother during sharp rises. The corresponding scatter plots (Fig. 3b) display tight clustering of data points around the 1:1 reference line, confirming the models' near-perfect correlation with observations. RF exhibits the narrowest spread, implying the least variance in residuals, while SVR's points are symmetrically distributed along the regression line, indicating minimal systematic bias. KNN's scatter pattern remains compact but reveals mild underestimation at higher water levels, consistent with the smoothing effect inherent in neighbour-based regression. Collectively, these results affirm the models' strong capacity for short-term stage prediction and validate the selected predictor set's

ability to encapsulate the basin's immediate hydrological response.

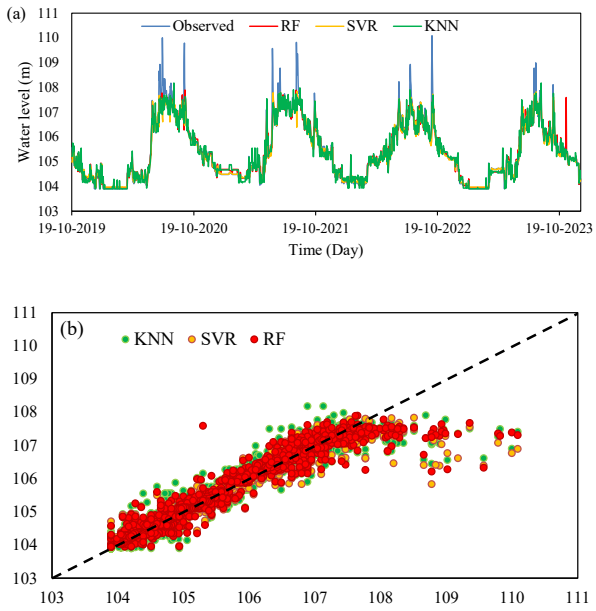

Figure 4: Forecasting line plots (a) and scattered plot (b) for lead times of 5-days

As the forecast horizon extended to five days, model accuracy declined modestly but the relative ranking among models remained consistent. The hydrographs in Figure 4a show that RF continued to reproduce both trend and magnitude effectively, with R^2 = 0.920, RMSE = 0.346 m, and MAE = 0.191 m. SVR maintained nearly equivalent skill (R^2 = 0.917; RMSE = 0.350 m) and again delivered the lowest MAE (0.168 m), demonstrating its advantage in minimizing typical daily deviations. KNN's performance decreased more noticeably (R^2 = 0.902; RMSE = 0.381 m; MAE = 0.235 m), and its predicted hydrograph displayed a tendency to lag and smooth peak values. In the scatter plots (Fig. 4b), both RF and SVR retained dense linear clusters near the 1:1 line, though the point dispersion widened slightly compared with the one-day case. RF's scatter distribution remained tightly concentrated for both low and moderate stages but exhibited a few scattered under-predictions at extreme peaks. SVR's scatter retained a nearly uniform spread across the full range, while KNN displayed a broader, fan-shaped dispersion, reflecting increasing bias and variance with lead time. These graphical patterns corroborate the statistical metrics and underscore each model's differing ability to preserve predictive precision as temporal memory decays.

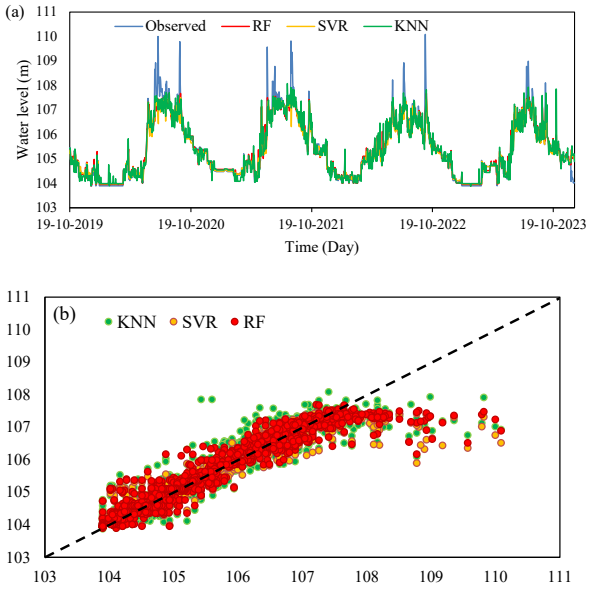

Figure 5: Forecasting line plots (a) and scattered plot (b) for lead times of 10-days

At the ten-day-ahead horizon, the deterioration in model performance became more pronounced yet remained hydrologically coherent. RF achieved R^2 = 0.910, RMSE = 0.364 m, and MAE = 0.203 m, maintaining the strongest overall explanatory power. SVR followed closely (R^2 = 0.906; RMSE = 0.372 m; MAE = 0.184 m), preserving its characteristic advantage in minimizing median errors. KNN exhibited the greatest decline (R^2 = 0.883; RMSE = 0.416 m; MAE = 0.259 m), with visible smoothing of the hydrograph peaks and delayed responses (Fig. 5a). The scatter plots (Fig. 5b) clearly reflect these trends. RF's points remained closely aligned along the 1:1 line, though their spread widened slightly toward higher stages, indicating under-prediction of extreme events. SVR maintained a well-distributed point cloud with smaller overall bias but marginally higher variance in the upper range, while KNN's points scattered more widely, showing systematic underestimation at both high and transitional levels. Despite this expected degradation, both RF and SVR preserved R^2 values above 0.90, confirming their ability to sustain predictive coherence even when antecedent hydrological memory weakens over longer horizons.

The collective interpretation of hydrographs and scatter plots highlights distinctive modelling behaviours. The RF model consistently exhibited the best balance between precision and dispersion control, capturing both mean and extreme water-level fluctuations. Its ensemble structure efficiently averaged out random noise while maintaining

sensitivity to significant hydrological events, which explains the narrow clustering around the 1:1 line across all lead times. SVR's scatter distributions were similarly tight but displayed slightly more symmetry around the regression line, resulting in smaller MAE values—a manifestation of the ε-insensitive loss function that prioritizes minimizing moderate deviations over fitting extremes. KNN, though adequate at short horizons, produced wider scatter distributions and smoother hydrographs at extended leads, confirming its limited ability to generalize temporal relationships in a dynamic monsoon-dominated system.

Table 1

Models	Lead time	R2	RMSE	MAE
RF	1	0.941	0.296	0.16
	5	0.92	0.346	0.191
	10	0.91	0.364	0.203
SVR	1	0.938	0.304	0.145
	5	0.917	0.35	0.168
	10	0.906	0.372	0.184
KNN	1	0.93	0.322	0.189
	5	0.902	0.381	0.235
	10	0.883	0.416	0.259

The divergence between RMSE and MAE patterns provides additional diagnostic insight. RF's consistently lower RMSE across all horizons implies stronger capacity to reproduce large-magnitude fluctuations, particularly during monsoon peaks, while SVR's lower MAE suggests superior stability during routine daily flows. The higher RMSE and MAE of KNN beyond five days result from its reliance on spatial proximity in feature space, which becomes less meaningful when antecedent conditions are less correlated with future states. This interpretation is supported by the hydrographs, where KNN's predictions show progressively dampened peaks, and by the scatter plots, where its residual variance expands markedly from one to ten days.

From an operational forecasting perspective, these results imply distinct use cases for each model. RF's superior RMSE and high R^2 values make it ideal for flood early-warning applications, where accurate representation of extreme water-level events and threshold exceedances is critical. SVR's minimal MAE and stable scatter alignment indicate high reliability for daily water-management decisions and reservoir operations, where smaller but consistent deviations are more consequential. KNN, while computationally simple and interpretable, is more suitable as a benchmark model or a supplemental component within an ensemble framework rather than as a standalone forecasting tool for extended lead times.

The comparative scatter-plot behaviour also reveals information about residual bias structures. RF's residuals were symmetrically distributed around zero, suggesting limited systematic bias and balanced prediction across flow regimes. SVR's residuals displayed slightly heavier tails at the upper range, implying a minor underestimation during rapid rises, whereas KNN's residuals skewed negatively at high stages, reflecting its under-prediction tendency. These patterns are consistent with the physical hydrological dynamics of the Gandak Basin, where sudden inflow surges from the upper catchment introduce nonlinear responses that challenge purely data-driven models lacking physical routing components. Nonetheless, the persistence of strong correlations and narrow scatter distributions across all lead times confirms that the machine-learning framework effectively captured the dominant statistical structure of the stage–discharge–rainfall relationships.

In summary, Figures 3–5 and Table 1 collectively demonstrate that the Random Forest model achieved the highest overall accuracy and stability across all forecast horizons, sustaining R^2 values above 0.90 and the lowest RMSE at each lead time. SVR consistently provided the smallest MAE, reflecting high precision in median flow regimes, while KNN showed the greatest sensitivity to lead-time extension, with its accuracy declining notably at five and ten days. The close alignment of observed and predicted values in both hydrographs and scatter plots reinforces the robustness of the RF and SVR models for operational forecasting at Triveni.

5. Conclusion

This study evaluated the performance of three machine-learning models—Random Forest (RF), Support Vector Regression (SVR), and K-Nearest Neighbors (KNN)—for forecasting water levels at the Triveni gauging station in the Gandak River Basin for lead times of 1, 5, and 10 days. The models used antecedent water level, discharge, and rainfall as

predictors to capture short-term hydrological dynamics.

Results showed that all models performed well for the one-day forecast, with RF achieving the highest accuracy (R^2 = 0.941, RMSE = 0.296 m) and SVR recording the lowest MAE (0.145 m). As the lead time increased, model performance gradually declined, reflecting reduced correlation between antecedent and future states. However, RF and SVR maintained R^2 values above 0.90 even at the 10-day horizon, highlighting their robustness for extended forecasts. KNN performed reasonably well for short horizons but exhibited larger deviations at longer leads due to its sensitivity to dynamic basin conditions.

Hydrographs and scatter plots confirmed that RF provided the best overall fit, particularly for capturing peaks, while SVR produced the most stable average predictions. Hence, RF is best suited for flood-warning applications, and SVR for routine operational management. RF and SVR offer reliable and complementary frameworks for short- to medium-term water-level forecasting in monsoon-dominated basins. Future research should integrate hybrid ensemble models and additional hydro-meteorological predictors to further enhance forecast accuracy and extend applicability to real-time flood management.

Reference

[1]. Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, *10*(11), 1536. https://doi.org/10.3390/w10111536

[2]. Shen, C. (2017). A trans-disciplinary review of deep-learning research for water-resources scientists. *EarthArXiv Preprint*. https://arxiv.org/abs/1712.02162

[3]. Willard, J. D., Varadharajan, C., Jia, X., & Kumar, V. (2023). Time-series predictions in unmonitored sites: A survey of machine-learning techniques in water resources. *arXiv Preprint*. https://arxiv.org/abs/2308.09766

[4]. Keisler, R. (2024). Do data-driven models beat numerical models in forecasting extreme weather?. Geoscientific Model Development, 17, 7915–7932. https://gmd.copernicus.org/articles/17/7915/2024/ .

[5]. Bi, K., Chen, K., & Wu, T. (2024). FuXi-S2S: A data-driven subseasonal-to-seasonal forecasting system. Nature Communications, 15, 50714. https://www.nature.com/articles/s41467-024-50714-1 .

[6]. Wu, Y., & Xue, W. (2024). Data-Driven Weather Forecasting and Climate Modeling from the Perspective of Development. Atmosphere, 15(6), 689. https://www.mdpi.com/2073-4433/15/6/689.

[7]. Breiman, L. (2001). Random forests. Machine learning, 45, 5-32, https://doi.org/10.1023/A:1010933404324.

[8]. Zhao, G., Pang, B., Xu, Z., Yue, J., & Tu, T. (2018). Mapping flood susceptibility in mountainous areas on a national scale in China. Science of the Total Environment, 615, 1133-1142, https://doi.org/10.1016/j.scitotenv.2017.10.037.

[9]. Karsavran, Y. (2024). Comparison of random forest, SVR and KNN based models in sea level prediction for Erdemli coast of Mersin. *Celal Bayar University Journal of Science*, *20*(2), 14-18. https://doi.org/10.18466/cbayarfbe.1384547

[10]. Bargam, B., Boudhar, A., Kinnard, C., Bouamri, H., Nifa, K., & Chehbouni, A. (2024). Evaluation of the support vector regression (SVR) and the random forest (RF) models accuracy for streamflow prediction under a data-scarce basin in Morocco. *Discover Applied Sciences*, *6*(6), 306. https://doi.org/10.1007/s42452-024-05994-z

[11]. Onay, B. (2021). *Development of machine learning models for short-term water level forecasting* (Master's thesis, Universitat Politècnica de Catalunya). https://upcommons.upc.edu/bitstreams/7a70a2dc-e310-4899-974a-aeeda1c2e6a2/download

[12]. Azad, A. S., Islam, N., Nabi, M. N., Khurshid, H., & Siddique, M. A. (2025). Developments and Trends in Water Level Forecasting Using Machine Learning Models–A Review. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3557910

[13]. Yeh, C. H., Hsu, K., & Chen, S. T. (2021). Applications of machine learning techniques in hydrological forecasting: A review. *Journal of Hydrology, 603*, 127015. https://doi.org/10.1016/j.jhydrol.2021.127015

[14]. Willard, J. D., Varadharajan, C., Jia, X., & Kumar, V. (2023). Time-series predictions in unmonitored sites: A survey of machine-learning techniques in water resources. *Environmental Modelling & Software, 167*, 105711. https://doi.org/10.1016/j.envsoft.2023.105711

[15]. Sinha, R., Srivastava, S., & Jain, V. (2019). River systems of Bihar plains: Morphology, dynamics, and flood risk. Geomorphology, 331, 31–49. https://doi.org/10.1016/j.geomorph.2018.12.025

[16]. Singh, S. K., Prakash, R., & Tripathi, S. (2020). Assessment of hydrological variability and flood characteristics in the Gandak River Basin using satellite and ground-based observations. *Water Resources Management, 34*(12), 3947–3965. https://doi.org/10.1007/s11269-020-02639-9

[17]. Mishra, V., Ojha, C. S. P., & Kumar, R. (2021). Modeling the hydro-climatic characteristics of Himalayan river basins under changing climate scenarios. *Journal of Hydrology: Regional Studies, 36*, 100847. https://doi.org/10.1016/j.ejrh.2021.100847

[18]. Bharati, L., Sharma, B. R., Gurung, P., & Pant, D. (2011). Hydrologic impacts of climate change on the water balance of the Gandak River Basin, India. *Hydrological Processes, 25*(21), 3325–3336. https://doi.org/10.1002/hyp.8238

[19]. Rao, G. S., Singh, R., & Kumar, A. (2018). Spatial variability of rainfall and runoff in the Gandak River Basin, India. *Journal of Earth System Science, 127*(7), 1–15. https://doi.org/10.1007/s12040-018-0976-5

[20]. IMD. (2023). *Daily Rainfall Data Archives (2003–2023): Bihar and Uttar Pradesh stations*. Indian Meteorological Department, Government of India. https://mausam.imd.gov.in

[21]. Wahba, M., Essam, R., El-Rawy, M., Al-Arifi, N., Abdalla, F., & Elsadek, W. M. (2024). Forecasting of flash flood susceptibility mapping using random forest regression model and geographic information systems. *Heliyon*, *10*(13). https://doi.org/10.1016/j.heliyon.2024.e33982

[22]. Chen, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H., & Ahmad, B. B. (2020). Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods. *Science of the Total Environment, 701*, 134979. https://doi.org/10.1016/j.scitotenv.2019.134979

[23]. Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

[24]. Velasco, L. C., Estose, A. J., Opon, M., Tabanao, E., & Apdian, F. (2024). Performance evaluation of support vector regression machine models in water-level forecasting. *Procedia Computer Science, 234*, 436–447. https://doi.org/10.1016/j.procs.2024.03.025

[25]. Hao, R., & Bai, Z. (2023). Comparative study for daily streamflow simulation with different machine learning methods. *Water, 15*(6), 1179. https://doi.org/10.3390/w15061179

[26]. Kombo, O. H., et al. (2020). Long-term groundwater level prediction model based on hybrid KNN-RF technique. *Water, 7*(3), 59. https://doi.org/10.3390/hydrology7030059

[27]. Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research, 30*(1), 79–82. https://doi.org/10.3354/cr030079

[28]. Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology, 10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6