Journal of Himalayan Earth Sciences Volume, No., 2026 pp. 1-24

Machine Learning-Based Prediction of Reservoir Inflows at Tarbela Dam under Historical and Future Climate Scenarios

Farjad Aziz[1], Muhammad Qasim[1], Mobeen Ahmed[1], and Alamgir Khalil[1*]

[1] Department of Civil Engineering, University of Engineering and Technology Peshawar, 25120 Pakistan

**Corresponding Author: alamgirkhalil@uetpeshawar.edu.pk*

Submitted Date:11/11/2025, Acceptance Date: 24/06/2026, Publication Date:---------

Abstract

This study focuses on predicting reservoir inflows at the Tarbela Dam using various machine learning algorithms under historical and projected climate scenarios. Daily hydrometeorological data from 14 stations across the upper Indus basin were used as inputs, including precipitation, maximum and minimum temperature, relative humidity, and solar radiation. Historical inflow data from 1981–2015 was used for training, and 2015-2020 was used for testing, while 2021 to 2024 was used for an independent validation set. Among the ML models, Catboost achieved the highest accuracy (R^2 = 0.92, RMSE = 291.42 m^3/s on training; R^2 = 0.87, RMSE = 811.00 m^3/s on testing), while XGBoost also performed well (R^2 = 0.92, RMSE = 317.77 m^3/s on training; R^2 = 0.88, RMSE = 817.64 m^3/s on testing). Future inflows were projected using the trained models under CMIP6 scenarios SSP2-4.5 and SSP5-8.5 for the period 2025–2100 in future simulations (2025–2100), Catboost remained the top performer under both SSP2-4.5 and SSP5-8.5 scenarios, achieving R^2 values above 0.90 and RMSE values below 750 m^3/s.

Keywords: Reservoir inflow, Machine learning, Upper Indus Basin, Tarbela Dam, XGBoost, Climate change

1. Introduction

Tarbela Dam, which is located on the Indus River, is important for Pakistan's water supply, electricity generation, and irrigation purposes. The successful implementation of such a reservoir largely relies on knowing the amount of supplying water. Inflow prediction has therefore been an important matter for longer-term planning, for operational decisions, and for managing risks. The Upper Indus Basin is characterized by a variety of climatic and geographic features, including variability in rainfall, snowmelt, and temperature changes. Prediction of reservoir inflows that may be variable poses a fundamental challenge for the management of Tarbela Dam, perhaps one of the most important dams for sustaining the water and power supply needs of Pakistan. The problems associated with variable inflows can be highlighted by the 2010 floods, where inadequate management of inflows and outflows led to flooding downstream of the Indus River. Inflow forecasting of the dam's reservoir is important for the real-time development of operation strategies of reservoirs so that water shortages can be managed due to increasing population and water demand. In recent years, changes in climatic patterns and changes in temperature and rainfall patterns have made relying on traditional forecasting techniques problematic because traditional hydrological models often require extensive input data. As a result of some of these challenges, data-driven techniques based on machine learning have received attention for their flexibility, and their improved predictive capability. Machine learning (ML) models have become an important research area with respect to reservoir inflow prediction, given the limitations of traditional statistical and physically based models. Traditional statistical models, such as linear regression

and time series models, are limited in their ability to represent the full variability and nonlinear dynamics of hydrologic systems (Bernardes et al., 2022). ML models can learn patterns from historical data without relying on significant physical information, while offering more adequate results than physically based models and flexibility in representing nonlinear relationships and spatial variabilities (Zarei et al., 2021).

In terms of inflow prediction, many ML models have become widely accepted and utilized. For example, artificial neural networks (ANNs) have been shown to be effective at representing nonlinear relationships in high-dimensional inputs (Dampage et al., 2020). The Random Forest (RF) approach is advantageous because it is based on the repeated construction of multiple decision trees from 5 datasets bootstrapped from the original dataset and able to account for noise and provides a measure of feature relevance/importance (Breiman, 2001). Ensemble machine learning techniques such as Artificial Neural Networks (ANN), Random Forests (RF), and Support Vector Machines (SVM) have increased in use in hydrology modelling fields due to their capacity to learn from data and adapt to changing conditions (Jan, 2025). Several studies have focused on reservoir inflow prediction for Tarbela Dam in UIB. Jan (2025) utilized Random Forest and KNN models for inflow forecast in the Upper Indus Basin and found that the RF better estimated nonlinear flow behaviour and peak inflows and produced lower prediction errors than KNN in operational and flood management applications. Khan and Koch (2020) employed Random Forest and Support Vector Regression for monthly flow forecasting in UIB sub-basins and documented that ensemble learning improved long-term predictive accuracy yet lacked calibration capability to estimate short-term flow fluctuations and variability. Jan et al. (2024) utilized XGBoost, Random Forest, and an XGBoost–RF ensemble model for inflow prediction at Tarbela Reservoir using 43 years of precipitation, temperature, and reservoir level data, revealing that XGBoost yielded the best accuracy by capturing nonlinear patterns and extreme inflow values efficiently, indicated by strong R^2 and low error metrics, across training and testing. Random Forest also gave acceptable results, but the ensemble version of the combination never outperformed the two individual models, meaning that in this instance combining the two models did not produce a better predictive model. Khan et al. (2025) examined inflow forecasting for rim-stations of the Indus Basin Irrigation System (i.e., Jhelum and Chenab Rivers class) using ANN-type models (GRNN & MLFN) and ARIMA, determining that the ANN approaches outperformed ARIMA with an average $R^2 \approx 0.92$ and RMSE ≈ 9.68, as compared to ARIMA's $R^2 \approx 0.88$ and RMSE $\approx$ of 10.17, demonstrating better skill for predictions of inflow at a one-year lead-time. While previous studies have explored machine learning for Tarbela reservoir inflow forecasting such as Jan et al. (2024) using XGBoost for short-term predictions and Khan et al. (2025) focusing on rim-station lead times, and significant gaps remain. This research extends the current literature in three distinct ways. Earlier studies that rely on limited gauging while we utilize a comprehensive 14 station dataset across the Upper Indus Basin. Beyond standard precipitation and temperature inputs, we incorporate relative humidity and solar radiation, which the SHAP analysis identifies as critical energy-balance drivers. We provide the first long-term (2025–2100) assessment using CMIP6 scenarios (SSP2-4.5 and SSP5-8.5) to evaluate how predictive skill evolves under future climate forcing. This provides a more robust and spatially explicit decision-support tool than previously available frameworks.

The main goal of the research is to create a high-performance and reliable machine learning framework for predicting reservoir inflow at Tarbela Dam in the Upper Indus Basin. The research evaluated the performance of multiple machine learning algorithms including AdaBoost, Catboost, Random Forest, Decision Tree, Support Vector Machine, Neural Network, and

XGBoost, as well as a composite stacked ensemble model (XGBoost × Random Forest), trained on a historical complete hydro-meteorological dataset. The specific objectives were to (i) conduct an evaluation and comparison of the forecasting capability of individual and composite machine learning models using a rigorous statistical evaluation framework (R^2, RMSE, MAE, MSE, and MAPE), (ii) identify the single or ensemble predictive machine learning model that best estimates non-linear variability and extreme flows, and (iii) increase the accuracy with respect to longer term forecasting to better support reservoir operation, hydropower generation, flood control and water resource management at Tarbela Dam.

2. Study area

Tarbela Dam, located in the Haripur district of Khyber Pakhtunkhwa, is one of the largest earth- and rock-filled dams in the world as shown in Figure 1. Completed in 1976, it is Pakistan's most important multipurpose water resource infrastructure. The dam's catchment area spans 169,600 km² and receives an average annual flow of 64 million acre-feet (MAF) (Munir et al., 2022). Tarbela alone contributes nearly 50% of the country's irrigation water, 30% of hydropower generation, and provides potable water to millions while helping prevent floods and supplying irrigation water. The upper Indus basin is located in the mountain ranges of Himalaya, Hindu Kush, and Karakoram covering parts of Afghanistan, Pakistan, India, and China as a transboundary river basin and covers a catchment area of 289,000 km^2 (Rauf et al., 2016).

3. Materials and methods

The daily data of hydrometeorological parameters were collected for 14 stations within the Upper Indus Basin over the extensive period from 1981 to 2024 as described in Table 1.

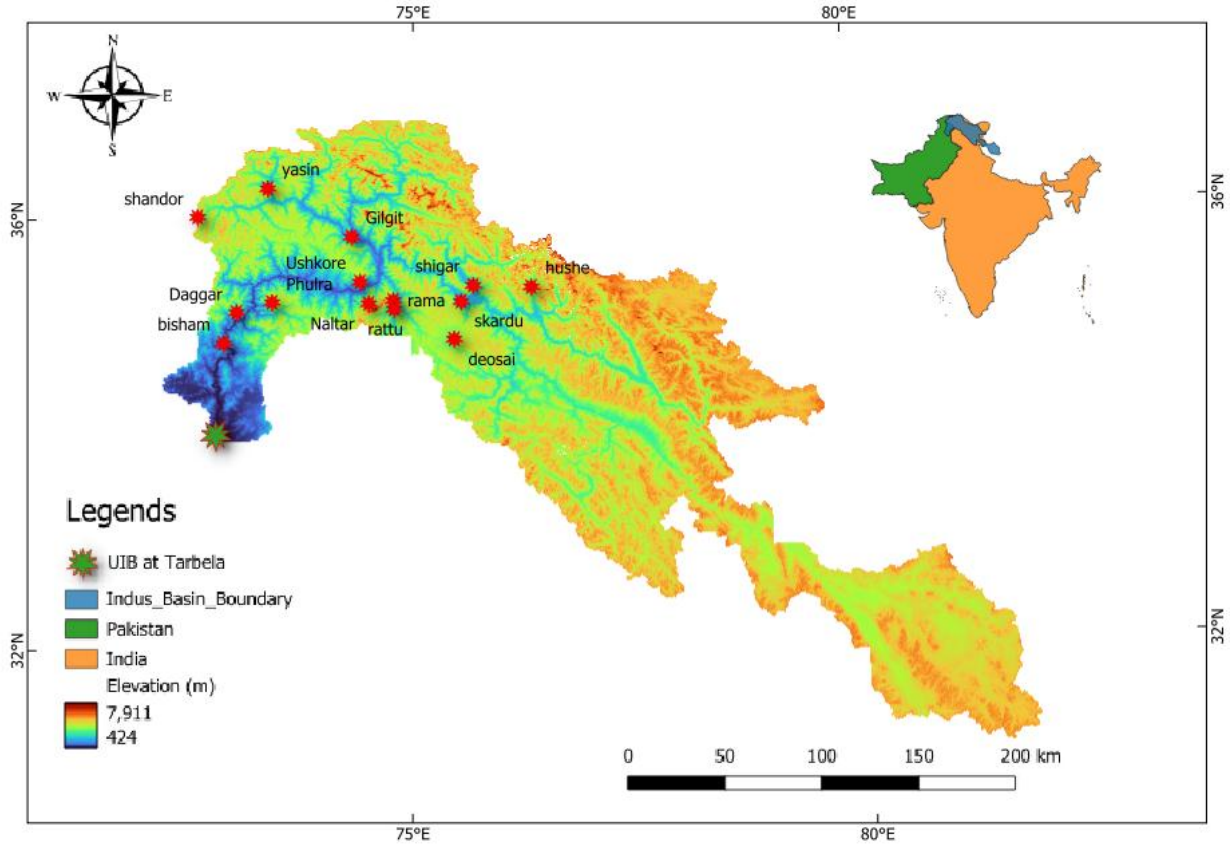


Fig. 1. Spatial distribution of 14 hydrometeorological stations in the upper Indus Basin.

The parameters included in our study are precipitation, maximum temperature (Tmax), minimum temperature (Tmin), relative humidity, solar radiation, and reservoir inflow. The spatial distribution of all these stations is shown in Figure 1. For the period from 1981 to 2003, the data was initially obtained from online archives. From 2003 to 2024, the data was acquired from the Water and Power Development Authority (WAPDA), providing a continuous and consistent dataset for the entire study period. We conducted monthly and seasonal trend analyses of the key hydrometeorological parameters for the 14 stations. Future climate projections for the period 2025 to 2100 were derived from five high-resolution experiments from the Coordinated Regional Downscaling Experiment for South Asia (CORDEX-SA). These experiments (including IPSL-CM5A-MR, MPI-ESM-LR, NorESM1-M, CanESM2, and GFDL-ESM2M) utilized Regional Climate Models (RCMs) such as RCA4 and RegCM4 to provide a finer spatial resolution of 0.44° compared to standard GCMs. The dataset was obtained in a statistically downscaled and bias-corrected format, where projections were adjusted to match the historical climatic properties of the Upper Indus Basin (UIB) from a study by Khan (2024). The RCP 4.5 and 8.5 scenarios from the CORDEX-SA dataset were mapped to the equivalent CMIP6 SSP2-4.5 and SSP5-8.5 pathways for consistency with current IPCC terminology. These included projected values of the same five meteorological variables used during the historical training phase. Trained ML models were applied to this GCM input data to forecast future reservoir inflows for the entire 2025–2100 period. This time series data has been divided into three parts. Near future from 2025-2049, Mid future from 2050-2074, and Far future from 2075-2100. Among the models evaluated on historical data, Catboost emerged as the best-performing algorithm. Therefore, the inflow predictions from Catboost were used as the reference baseline against which future inflow predictions from other ML models were compared. Standard statistical metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2) were used to assess consistency and performance during future forecasting.

Table 1: Description of the meteorological stations under study.

Stations	Latt.	Long.	Elevation (ft)	Temporal range
Ushkore	35.35	74.86	2391.5	1981-2024
Besham	34.91	72.86	780.8	
Daggar	35.66	74.6	1982.7	
Phulra	35.42	74.09	1186.6	
Naltar	35.91	74.38	1418.3	
Deosai	34.97	75.47	4137.9	
Gilgit	35.92	74.3	1504.1	
Hushey	35.45	76.35	3460.7	
Rama	35.33	74.78	3522.9	
Rattu	35.25	74.79	3528	
Shendure	36.08	72.53	3682.7	
Shiger	35.47	75.69	2253.8	
Skardu	35.32	75.55	2211.7	
Yasin	36.36	73.33	2512	

To mitigate the risk of overfitting, particularly in high-variance ensemble models, several regularization techniques were employed. For XGBoost, L1 and L2 regularization penalties were applied to the loss function. For Catboost, an Ordered Boosting approach was utilized, which helps prevent gradient bias and enhances the model's ability to generalize to unseen testing data. Hyperparameter optimization was conducted to ensure a balance between training depth and predictive stability. To make sure of the predictive integrity, the dataset was partitioned into three distinct phases: Training (1981 to 2015), Testing (2016 to 2020), and an Independent Validation period (2021 to 2024). The 2021 to 2024 data was intentionally excluded from the initial model development and hyperparameter tuning phases to serve as a blind evaluation set. This strategy was employed to verify whether the models specifically the top performing Catboost architecture could maintain accuracy against recent hydroclimatic non-stationarity and extreme events before being applied to long term CMIP6 future scenarios.

To ensure model convergence and prevent features with larger numerical ranges (e.g., reservoir inflow) from dominating those with smaller ranges (e.g., temperature), all input data were normalized using Min-Max Scaling. This transformed all variables into a uniform range of [0, 1], which is essential for the weight-optimization processes in the Artificial Neural Network (ANN) and Support Vector Machine (SVM) models.

3.1 Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a very efficient and scalable machine learning algorithm based on a kind of gradient boosting decision tree. XGBoost constructs an ensemble of weak learners (most commonly decision trees) in a sequential fashion, where each new tree corrects the residuals of all previously built trees (Chen and Guestrin, 2016) as shown in Figure 2. XGBoost incorporates regression-like regularization in terms of L1 and L2 penalty that provide increased protection against overfitting and are overall more conservative than boosting in a classic sense. XGBoost has been successfully utilized for flood forecasting, rainfall estimation, water demand forecasting, and reservoir inflow modeling, since it has a more nuanced ability to capture complex, nonlinear relationships in data (Gupta et al., 2022).



Fig. 2. XGBoost workflow (Gupta et al., 2022).

In hydrological applications, XGBoost has demonstrated strong predictive capacity due to its ability to model complex, non-linear interactions among meteorological predictors. Its robustness, speed, and interpretability make it a preferred algorithm for inflow forecasting, particularly when high-dimensional climatic datasets are involved. The loss function $\mathcal{L}(\phi)$ is calculated by Equation (1) (Chen and Guestrin, 2016):

$$\mathcal{L}(\phi) = \sum_\{i = 1\}\^\{n\} l(yi, \hat{y}i) + \sum_\{k = 1\}\^\{K\}\Omega(fk) \quad (1)$$

where $\Omega(f_k) = \gamma T + ½\lambda\|w\|^2$, and denotes the loss function, $\hat{y}i$ is the predicted output, fk represents the k^{th} regression tree, T is the number of leaves in the tree, w is the leaf weight vector, γ and λ are regularization parameters, and ϕ is the overall model.

3.2 Catboost

Catboost is a gradient boosting framework developed by Yandex that is specifically optimized for datasets containing both numerical and categorical features (Prokhorenkova et al., 2018). Unlike traditional boosting algorithms. Catboost processes residuals in a sequential manner using permutations of the dataset as visualized in Figure 3. This ordered training framework enables unbiased gradient estimation. It can handle missing values directly and uses automatic feature combination strategies to enhance learning from complex datasets.

The general form of Catboost's objective function as per Equation (2a) is:

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \lambda \sum_{k=1}^{k} \Omega(f_k) \quad (2a)$$

where $\mathcal{L}$ represents the loss function (e.g., squared error), $\hat{y}\boldsymbol{i}$ is the predicted output, Ω is the complexity penalty of the k^{th} tree, and λ is the regularization parameter that controls model complexity.

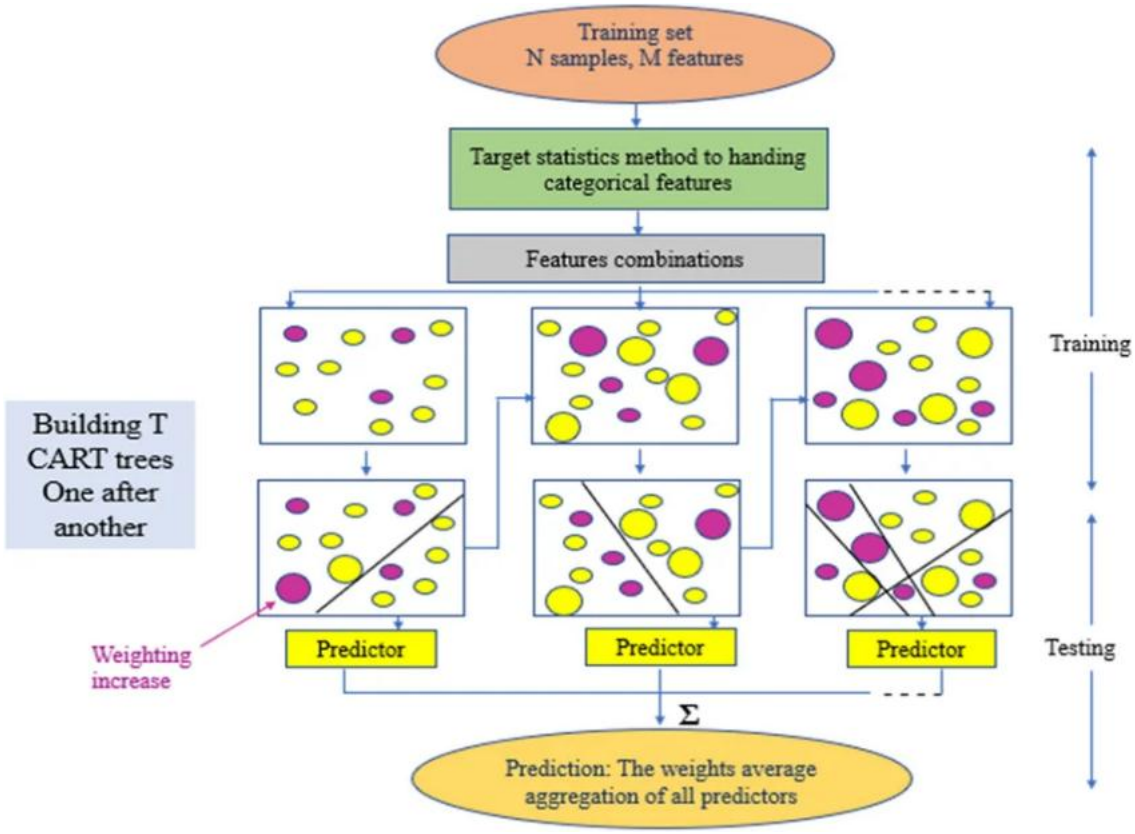


Fig. 3. Catboost workflow (Sireesha et al., 2023).

The general form of Catboost's objective function as per Equation (2b) is:

$$\mathcal{L} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \lambda \sum_{k=1}^{k} \Omega(f_k) \quad (2b)$$

where $\mathcal{L}$ represents the loss function (e.g., squared error), $\hat{y}_i$ is the predicted output, Ω is the complexity penalty of the k^{th} tree, and λ is the regularization parameter that controls model complexity.

In hydrological forecasting tasks such as streamflow prediction and reservoir inflow estimation Catboost has demonstrated strong performance, outperforming many traditional regression algorithms and ensemble learners (Wang et al., 2021).

3.3 Random Forest

Random Forest (RF) is an ensemble learning algorithm that aggregates the predictions of multiple decision trees to improve model accuracy and reduce overfitting (Breiman, 2001). Each decision tree in the ensemble is trained on a bootstrap sample drawn randomly from the original dataset, and at each split node, a random subset of features is selected to determine the optimal split. Random Forest is particularly well-suited for hydrological and environmental modeling due to its capacity to model complex, nonlinear relationships without requiring strong assumptions about the underlying data distributions. Typical structure of a Random forest workflow can be seen in Figure 4.

The final prediction $\hat{y}$ in a regression setting is computed as the average of predictions from all individual trees as shown in Equation 3.

$$\hat{y} = \left(\frac{1}{T}\right) \sum_{t=1} f_{t(x)} \quad (3)$$

where T is the total number of trees and $ft(x)$ is the output of the t^{th} decision tree for input x.

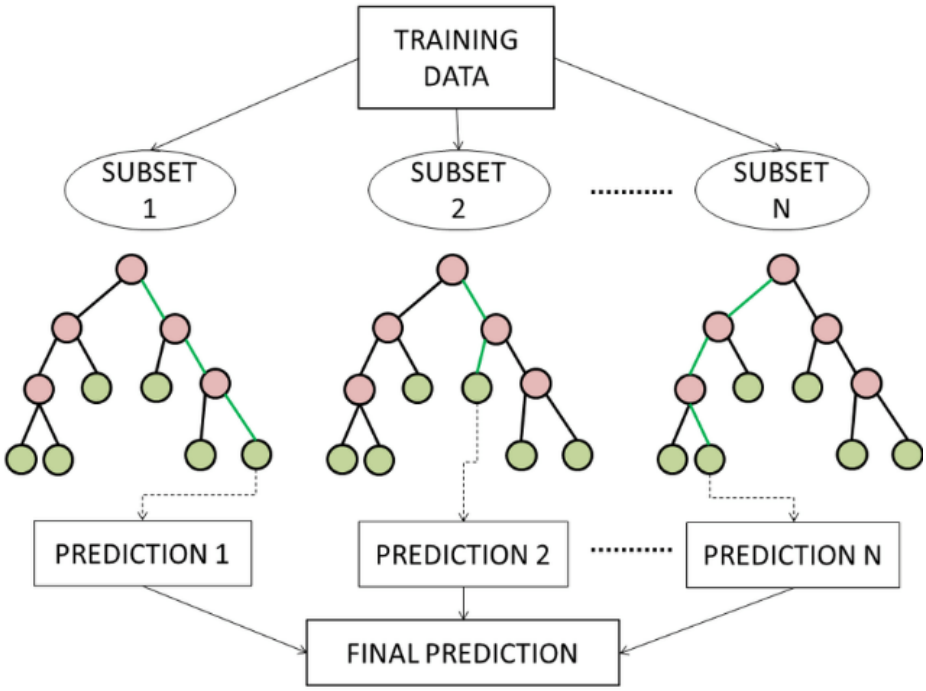


Fig. 4. Random Forest workflow.

Random Forest has been successfully applied to tasks such as streamflow forecasting and reservoir inflow prediction. Its resilience against overfitting and capacity to integrate diverse meteorological and hydrological inputs make it a reliable algorithm for inflow modeling under both historical and projected climate conditions (Mosaffa et al., 2022).

3.4 Artificial Neural Networks

Artificial Neural Networks (ANNs) are computer models, inspired by the structure and function of the brain's neural network. ANNs consist of layers of nodes (neurons) that process data through the weighted signaling and activation functions that produce an output (Haykin, 2009). In hydrology and water resource engineering, ANNs have a strong following for their applications with stream-flow prediction and predicting reservoir inflows because they can model complex and difficult-to-understand hydrological processes without developing explicit physical relationships (Maier et al., 2010).

A typical feed forward ANN consists of three types of layers as seen in Figure 5: an input layer, one or more hidden layers, and an output layer. Each neuron computes a weighted sum of its inputs, adds a bias term, and applies an activation function such as the sigmoid, tanh, or ReLU function. The general form of the neuron output is given by Equation 4:

$$z = \varphi\left(\sum_{i=1}^{n} wi\, xi + b\right) \tag{4}$$

where x_i represents the input features, w_i the corresponding weights, b the bias term, φ the activation function, and z the output of the neuron.

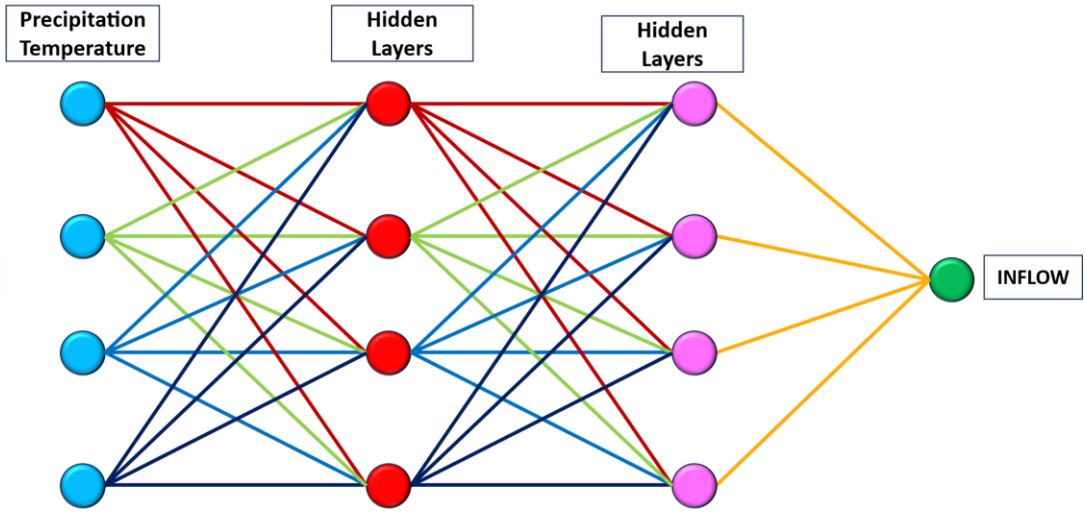


Fig. 5. ANN Workflow (Maier et al., 2010).

3.5 AdaBoost

Adaptive Boosting (AdaBoost) is one of the earliest and most widely used ensemble learning algorithms, introduced by Freund and Schapire (1997). It works by combining multiple weak learners typically decision stumps into a single strong predictor through an iterative process that adjusts the weight of training samples based on their classification or prediction error in previous iterations as pictured in Figure 6. This allows AdaBoost to adapt to bias with samples, and it is especially beneficial for advancing classification accuracy on clean datasets.

As an ensemble learning algorithm, AdaBoost will be influenced by noisy data and misclassified samples, but the method can be considered simple to implement. In a hydrological context, AdaBoost has been utilized for groundwater flooding prediction and assessing water quality viability, proving at least to be competitive with other ensemble learning approaches (Mosavi et al., 2018).

3.6 Support Vector Machine (SVM)

Support Vector Machines (SVM) are supervised machine learning models originally developed for classification tasks and later extended to regression problems through the Support Vector Regression (SVR) framework (Vapnik, 1995). SVM has been applied in various contexts in hydrology and environmental studies including rainfall-runoff modeling, drought forecasting, and reservoir inflow prediction for the very same reasons (Dibike et al., 2001). Figure 7 shows the working mechanism of SVM. SVM attempts to present the class structure of the data in the feature space using an optimal hyperplane and utilizes the kernel function to transform the input data where needed (Cortes and Vapnik, 1995).

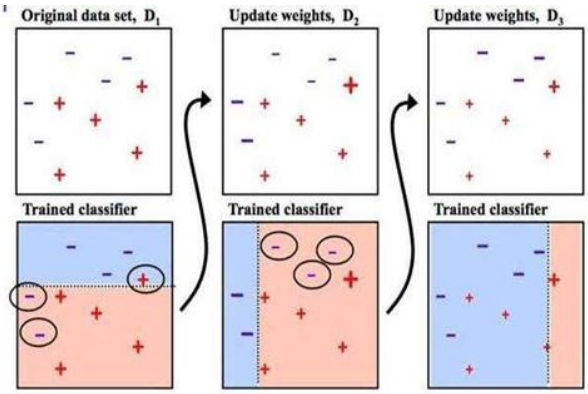


Fig. 6. AdaBoost workflow (Mosavi et al., 2018).

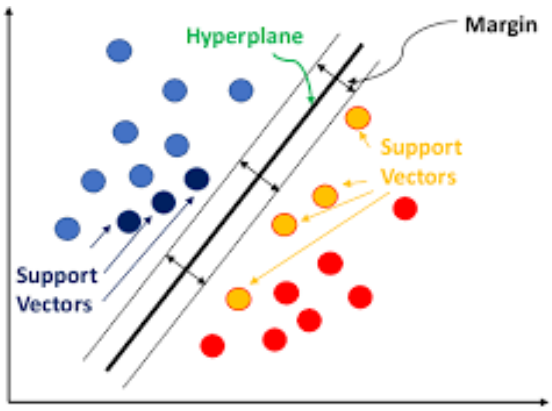


Fig. 7. SVM Workflow (Dibike et al., 2001).

3.7 Ensemble Model of XGBoost and RF

Hybrid models aim to combine the strengths of individual machine learning algorithms to enhance prediction performance, reduce generalization error, and capture diverse patterns present in complex datasets. In this study, a hybrid ensemble model combining Extreme Gradient Boosting (XGBoost) and Random Forest (RF) was employed to improve the accuracy of reservoir inflow prediction. To address differences, Random Forest uses a series of independent trees for each decision (this is achieved through majority voting) to minimize the level of variance (Breiman, 2001), while XGBoost builds upon the learning from the sequentially derived decision trees to minimize the extent of the residual errors through the process of gradient boosting (Chen and Guestrin, 2016).

The XGBoost × RF hybrid model, as shown in Figure 8, works by first training both base models independently using the same input features. The predictions from each base model are then fused, either through averaging, weighted aggregation, or stacked generalization, to produce the final inflow forecast. This architecture leverages the boosting capability of XGBoost to minimize bias, while Random Forest contributes through its variance reduction and feature robustness.

Hybrid ML models are increasingly used in hydrological modeling due to their ability to capture nonlinear interactions, enhance generalization capability, and reduce model specific limitations. Studies have demonstrated that combining tree-based learners in this way can outperform standalone models in applications such as runoff modeling, groundwater forecasting, and reservoir inflow prediction under variable climate scenarios.

4. Results

Eight different machine learning models were used for reservoir inflow prediction. First, the prediction was done for the time series from 1981 to 2025 and then for the time series of 2025 to 2100 under two future scenarios, SSP2-4.5 (moderate emission) and SSP5-8.5 (high emission).

4.1 Training and Testing results of machine learning models for time series 1981-2025

The result of the Inflow prediction is demonstrated by using both a line graph and a scatter plot for the training and testing period seen in Figure 9. The line graph shows the comparison between observed inflow and the predicted inflow for both the training and testing periods as the outcome of the machine learning model. These graphs help in understanding the temporal and seasonal patterns, variations, trends, fluctuations and the peaks of the inflow data. When both observed and predicted inflow lines are closer to each other, that shows the strong performance and accuracy of the model.

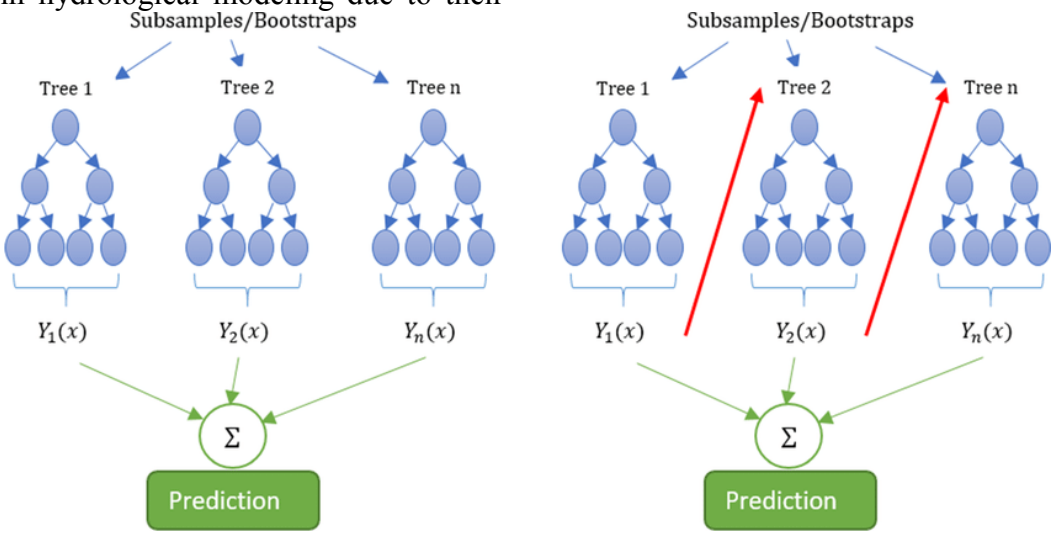


Fig. 8. Random Forest example (left), boosting/XG-boost example (right) (Gupta et al., 2022).

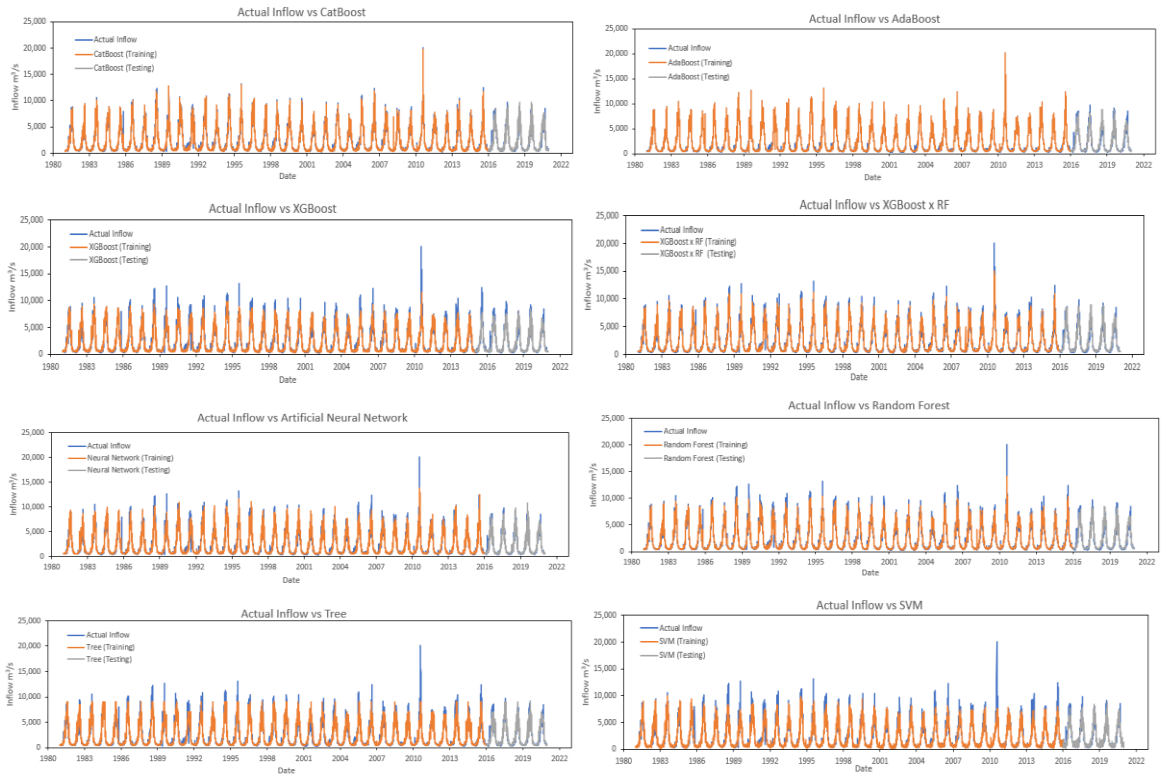


Fig. 9: Training and testing results of all machine learning models vs actual inflow for historical series.

In addition to the line graph, a scatter plot is also used to illustrate the training phase (1981 to 2015) in Figure 10 and the testing phase (2016 to 2020) in Figure 11. The scatter plot consists of data points that show a predicted value at the same time step. The x-axis shows observed Inflow values, while the y-axis shows the Inflow values predicted by ML models (Catboost, AdaBoost, ANN, RF, SVM, Tree, XGBoost, XGBoost x RF). A dashed 45-degree reference line (y = x) represents the agreement between observed and predicted values. If the values are clustered toward this line, then this shows the observed and predicted values are similar, and when the values are away from this reference line, that means the observed and predicted values have some error. So, by using a scatter plot, we can visually evaluate the accuracy, consistency and errors of the model-generated inflow.

A comparative analysis of training and testing metrics reveals distinct signals of overfitting in certain ensemble architectures. For instance, AdaBoost achieved a R^2 of 0.94 during training but dropped to 0.89 during testing. Similarly, Random Forest and the XGBoost x RF hybrid showed training performance (R^2 = 0.91) but exhibited a noticeable reduction in predictive skill during the validation phase (R^2 = 0.84). This indicates that these models captured noise within the historical 1981–2015 training set.

The performance of these eight machine learning models was analyzed during the testing and training phase to find their accuracy in predicting inflows and tabulated in Table 2. In all these eight models, Catboost stands out as the best-performing model, having an R^2 value of 0.92 for training and 0.87 for testing, indicating a strong correlation between observed and predicted inflows. The small difference between the training and testing value of R^2 shows the model generalizes well and the effectiveness of the model in preventing over- and under-fitting, resulting in better performance.

4.2 Inflow prediction for the time series 2025-2100

As seen, Catboost was the best performing model in historical phase. Therefore, Catboost serves as a robust

baseline for the future inflows under multiple SSP scenarios and it is important to acknowledge that these projections are model derived estimates rather than observed truths.

The use of a single reference model introduces potential inter-model dependency and uncertainty propagation from the underlying GCM inputs. Future studies may benefit from a multi-model ensemble approach to further characterize the envelope of uncertainty in long-term inflow assessments

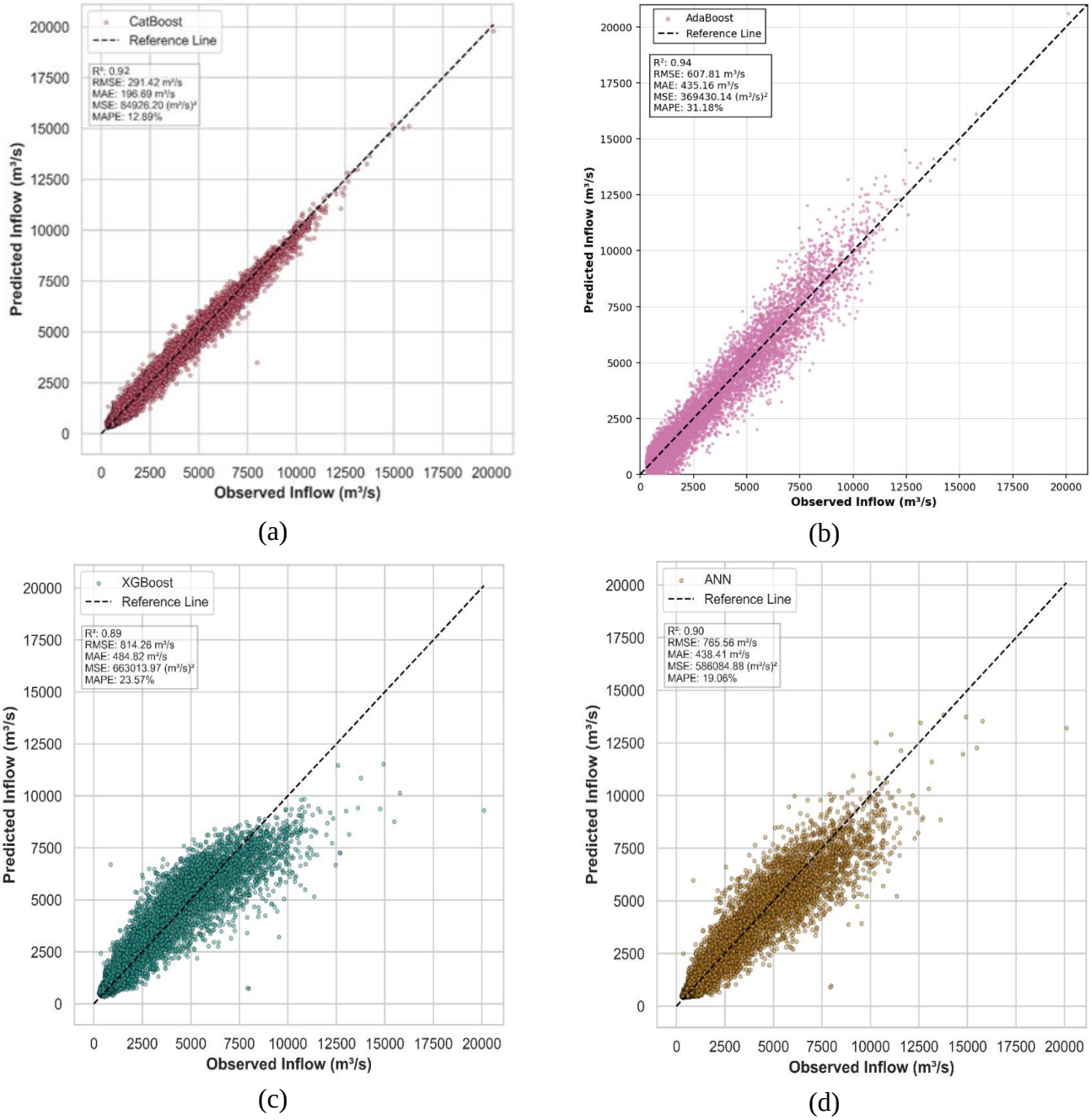


Fig. 10. Scatter plot of observed vs predicted inflows during the training phase (1981-2015).

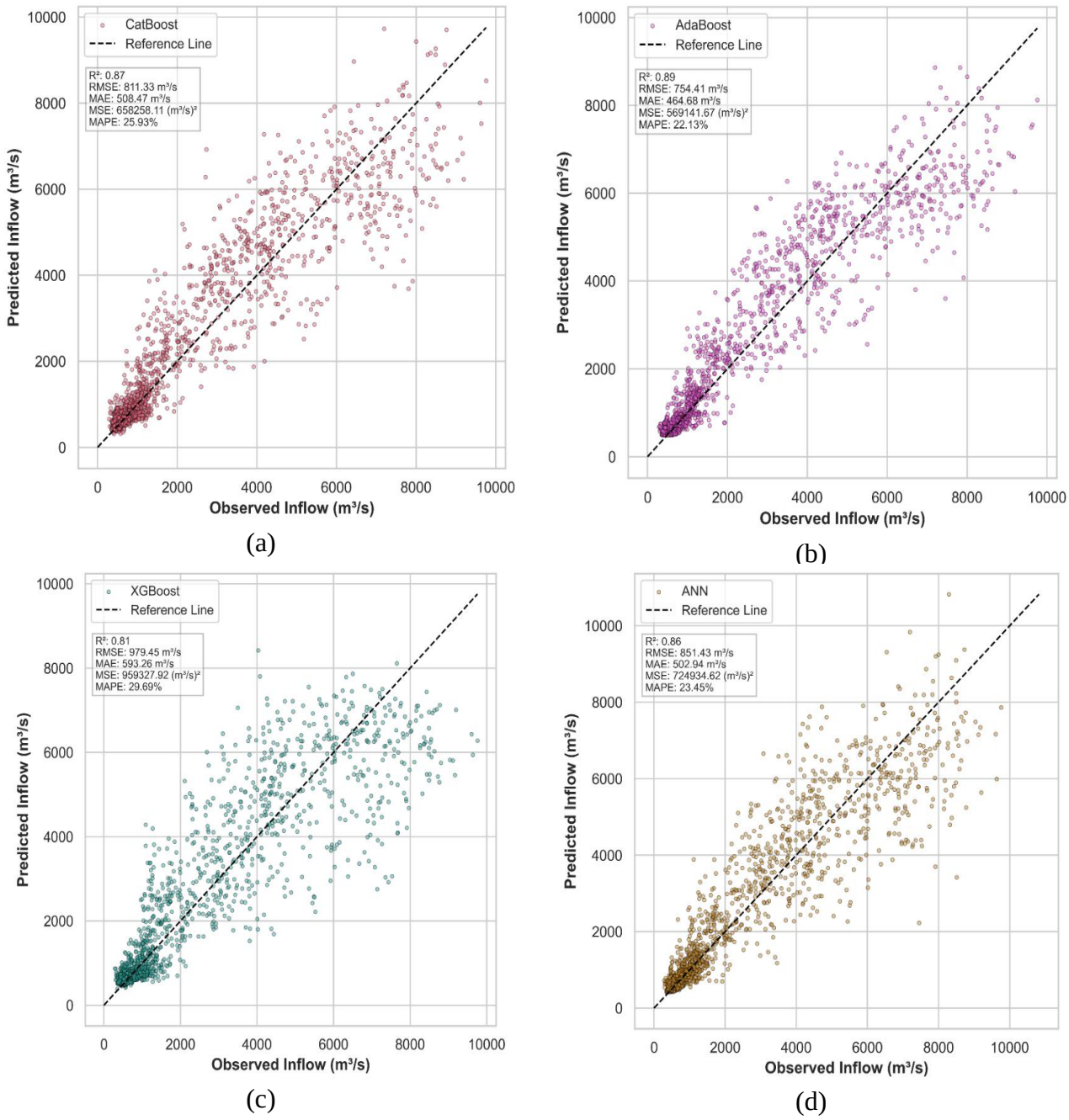


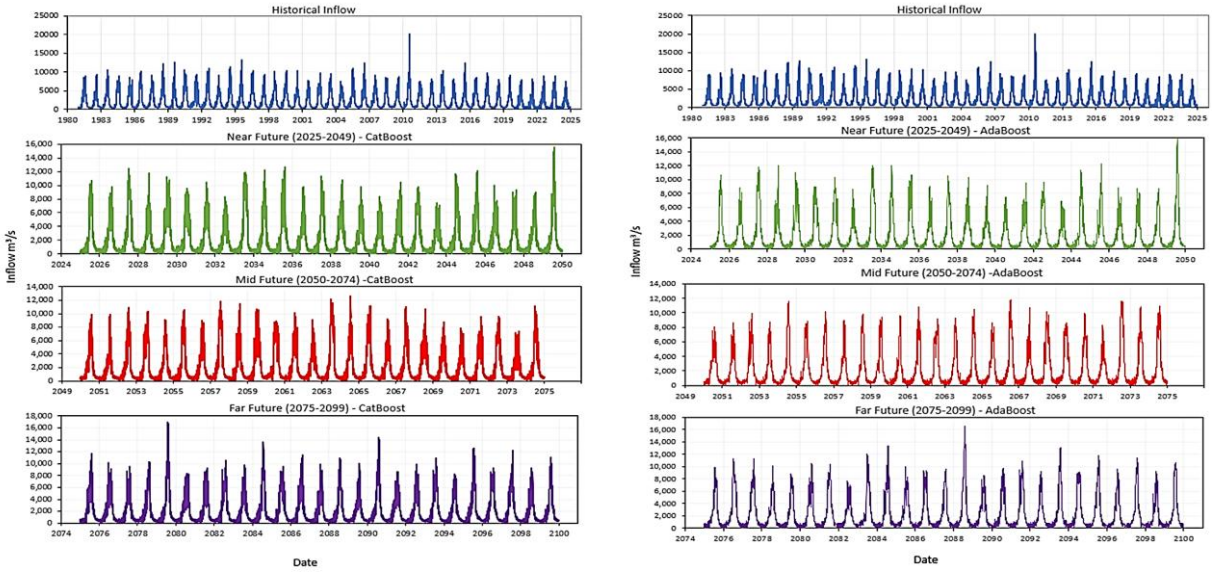
Fig. 11. Scatter plot of observed vs predicted inflows during the testing phase (2016-2020)

Table 2: Evaluation metrics of all models for training and testing phase during historical time series (1981-2020).

MODELS	TRAINING (1981-2015)					TESTING (2016-2020)				
	R^2	RMSE	MAE	MSE	MAPE	R^2	RMSE	MAE	MSE	MAPE
AdaBoost	0.94	607.81	435.16	369433	31.18	0.89	754.41	464.68	569134	22.13
Catboost	0.92	291.42	196.60	84926	12.89	0.87	811.00	508.42	657721	25.93
XGBoost x RF	0.91	744.41	528.74	554146	37.68	0.84	992.54	702.53	985136	48.84
Tree	0.86	945.53	527.45	894027	22.34	0.81	980.98	598.50	962322	30.38
Neural Network	0.91	765.56	438.41	586082	19.06	0.86	851.43	502.94	724933	23.45
Random Forest	0.98	347.13	187.17	120499	7.98	0.88	779.19	497.03	607137	25.67
XGBoost	0.89	814.26	484.82	663019	23.57	0.81	979.45	593.25	959322	29.69
SVM	0.86	927.73	502.90	860683	23.01	0.84	964.22	581.29	929720	35.43

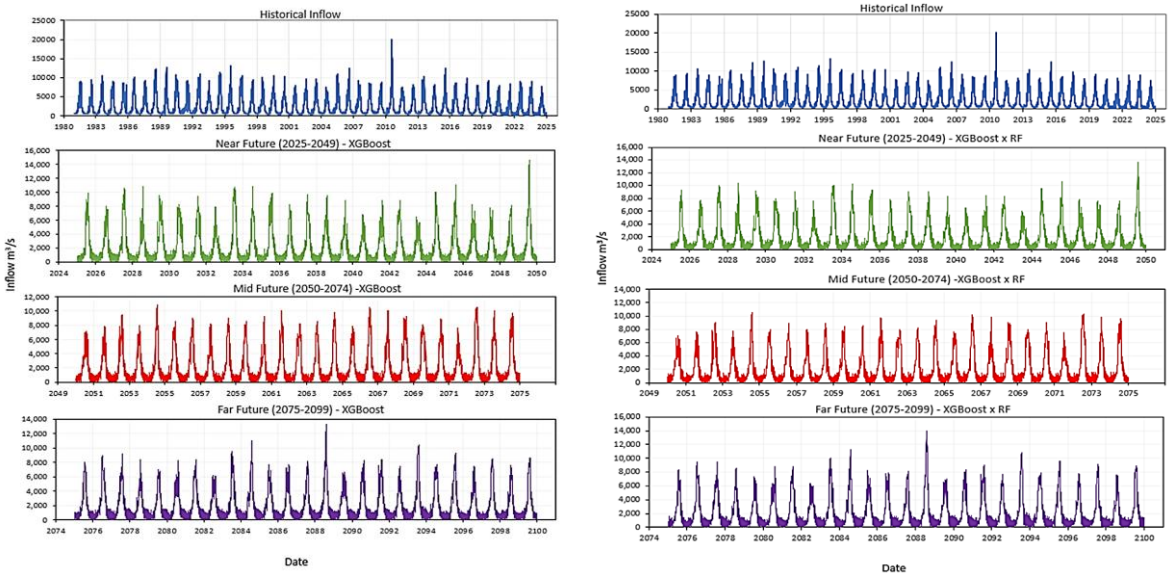
4.2.1 Results of future prediction (2025-2100) under SSP 2-4.5

Figure 12 show the results of Catboost and Adaboost, followed by Figure 13 which includes XGBoost and ensemble of XGBoost and Random Forest, then Figure 14 showcases ANN and Random forest, and lastly figure 15 shows SVM predicted inflows for near, mid and far future under SSP 2-4.5 scenario. Each graph consists of 4 subgraphs, including historical inflow from 1980 to 2025 and three future periods: the near future from 2025 to 2049, the mid future from 2050 to 2074, and the far future from 2075 to 2100.



(a) (b)

Fig. 12. Near, mid, and far future inflow prediction under SSP 2-4.5 for (a) Catboost and (b) AdaBoost.



(a) (b)

Fig. 13. Near, mid, and far future inflow prediction under SSP 2-4.5 for (a) XGBoost (b) XGBoost and RF Ensemble.

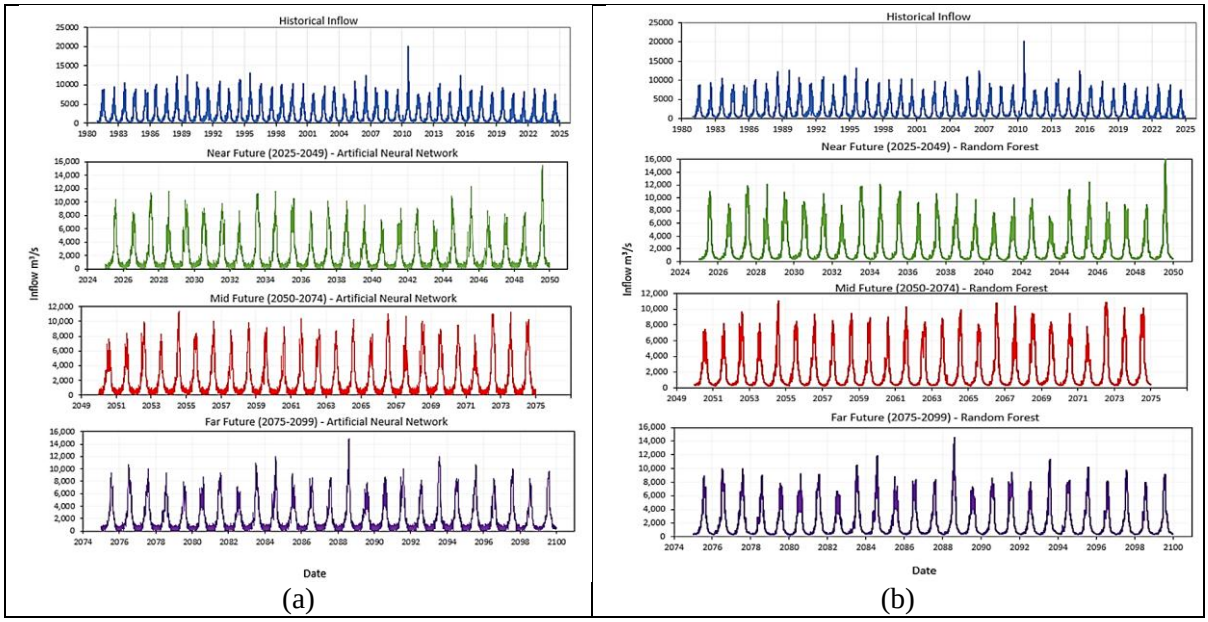


Fig. 14. Near, mid, and far future inflow prediction under SSP 2-4.5 for (a) ANN (b) Random Forest.

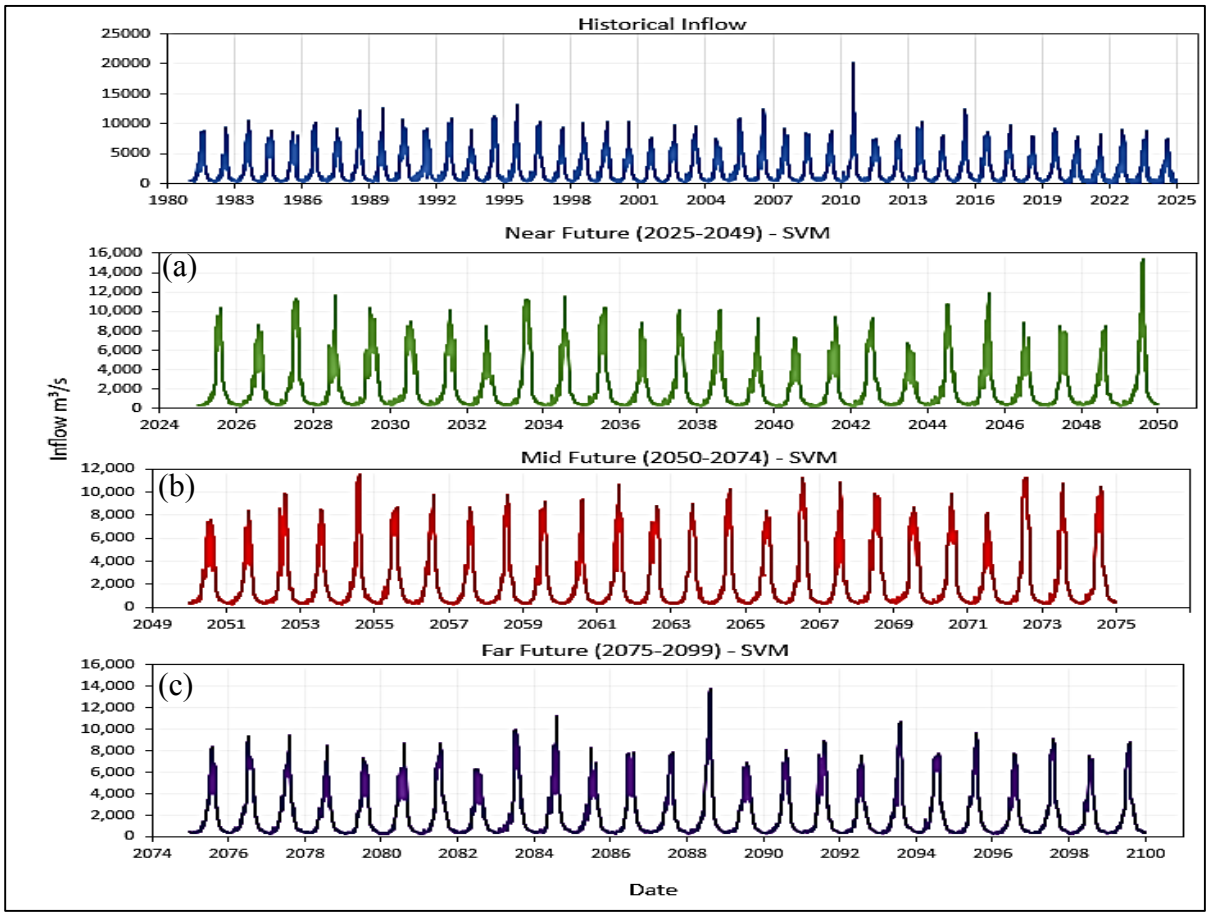


Fig. 15. SVM future inflow prediction under SSP 2-4.5 for a) near, b) mid, and c) far future.

In addition, Figure 16 shows the scatter plots between the actual inflow, in our case, that is, Catboost inflow, as it was the best performing model in historical time series and the predicted model inflow for all six models, which include (AdaBoost, XGBoost, XGBoost x RF, ANN, Random Forest, and SVM) under SSP 2-4.5 scenario upto year 2100. The scatter plot consists of data points that represent a predicted value at the same time step. The x-axis shows inflow values, which are predicted by the Catboost

model, while the y-axis shows the corresponding values predicted by other models. A dashed 45-degree reference line (y = x) represents the ideal case where both models' predictions become the same, so when the points are clustered close to this line, the agreement between the two models and if the points are far from this line, then the error will be larger in their prediction.

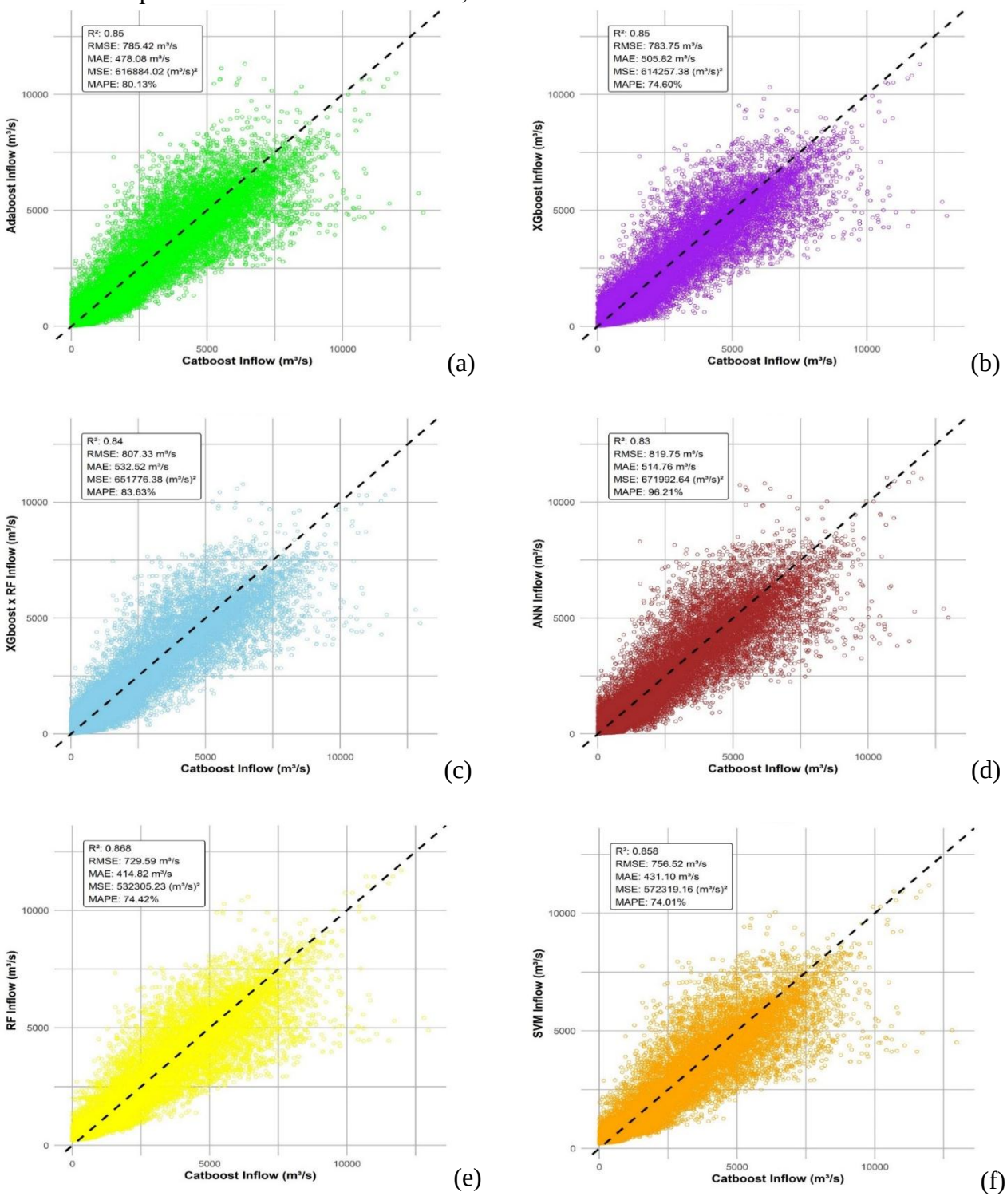


Fig. 16. Scatter plot comparing predicted inflows to Catboost predicted inflows upto year 2100 under SSP 2-4.5. a) AdaBoost, b) XGboost, c) XGboost x RF, d) ANN, e) RF, and f) SVM.

4.2.2 Results of future prediction (2025-2100) under SSP 5-8.5

Figures 17-20 show the results of Catboost, Adaboost, XGBoost, Random Forest, ensemble of XGBoost and Random Forest, ANN and finally SVM respectively for future inflow predictions under SSP 5-8.5 scenario. Each graph consists of 4 subgraphs, including historical inflow from 1980 to 2025 and three future periods: the near future from 2025 to 2049, the mid future from 2050 to 2074, and the far future from 2075 to 2100.

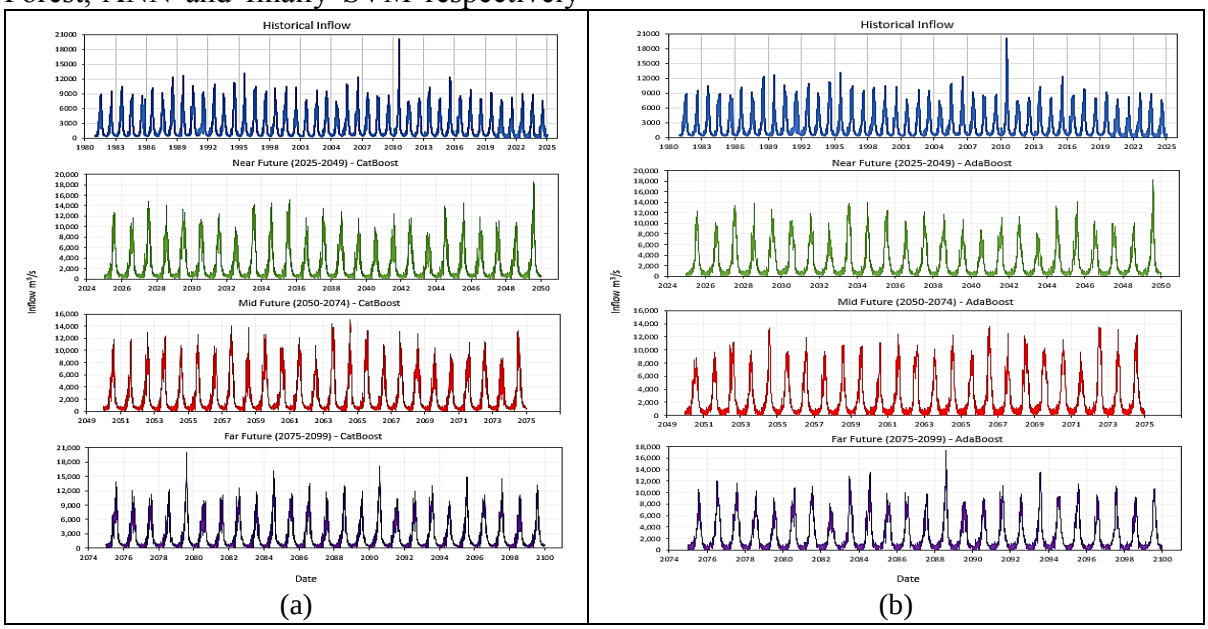


Fig. 17. Future inflow prediction under SSP 5-8.5 under near, mid, and far future for (a) Catboost (b) AdaBoost.

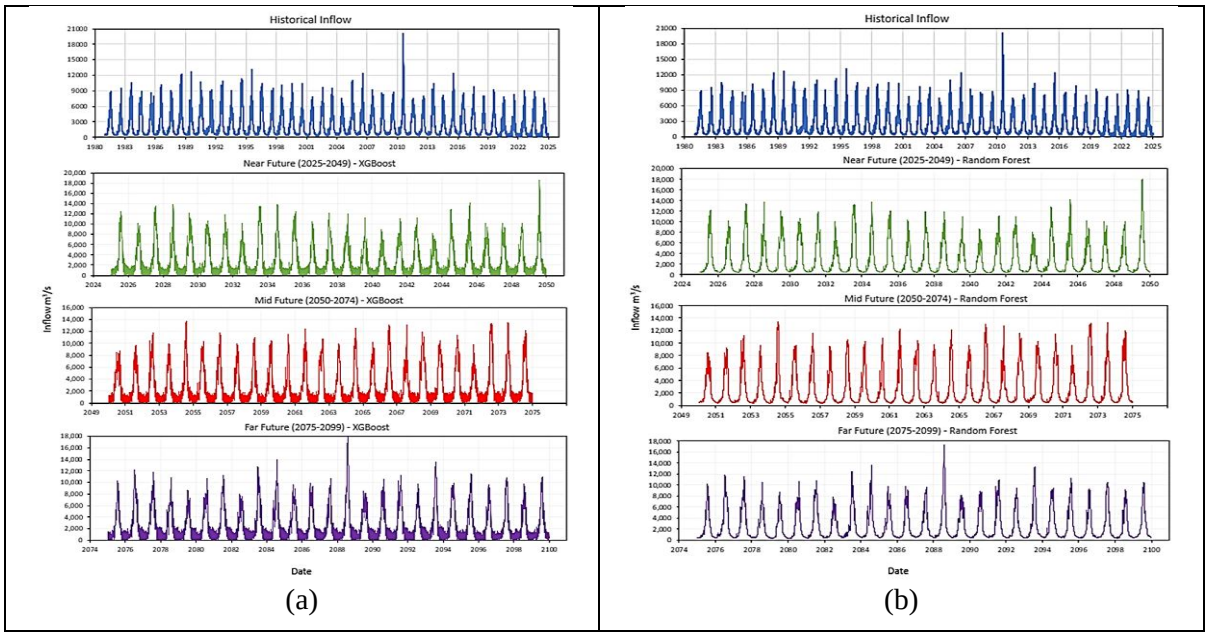


Fig. 18. Future inflow prediction under SSP 5-8.5 under near, mid, and far future for (a) XGBoost (b) Random Forest.

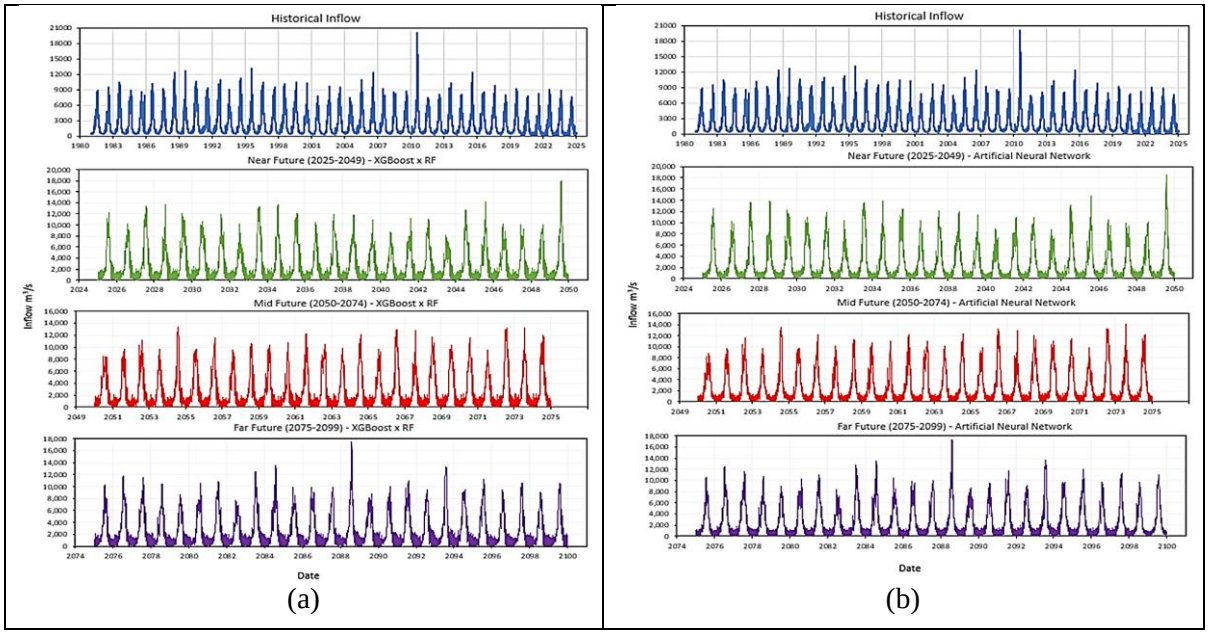


Fig. 19. Future inflow prediction under SSP 5-8.5 under near, mid, and far future for (a) XGBoost RF Ensemble (b) ANN.

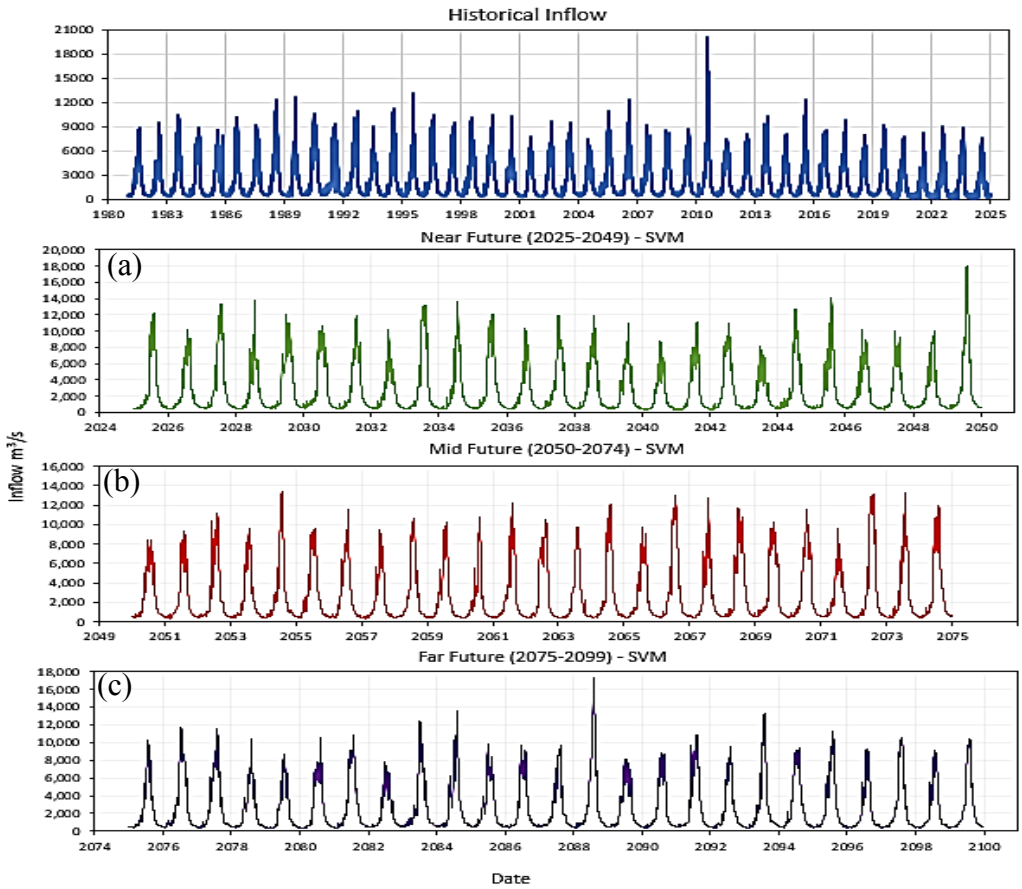


Fig. 20. SVM future inflow prediction under SSP 5-8.5 for (a) Near Future, (b) Mid Future, and (c) Far Future.

In addition, Figure 21 shows the scatter plot between the actual inflow, in our case, that is, Catboost inflow and the predicted inflow for all six models, which include (AdaBoost, XGBoost, XGBoost x RF, ANN, Random Forest, and SVM). The scatter plot includes data points that represent a predicted value at the same time step. The

x-axis shows the inflow values of Catboost prediction, while the y-axis shows the corresponding value predicted by other models. A dashed 45-degree reference line (y = x) represents the ideal case where both models' predictions become the same, so when the points are clustered close to this line higher will be the agreement between the two models and if the points are far from this line, then the error will be larger in their predictions.

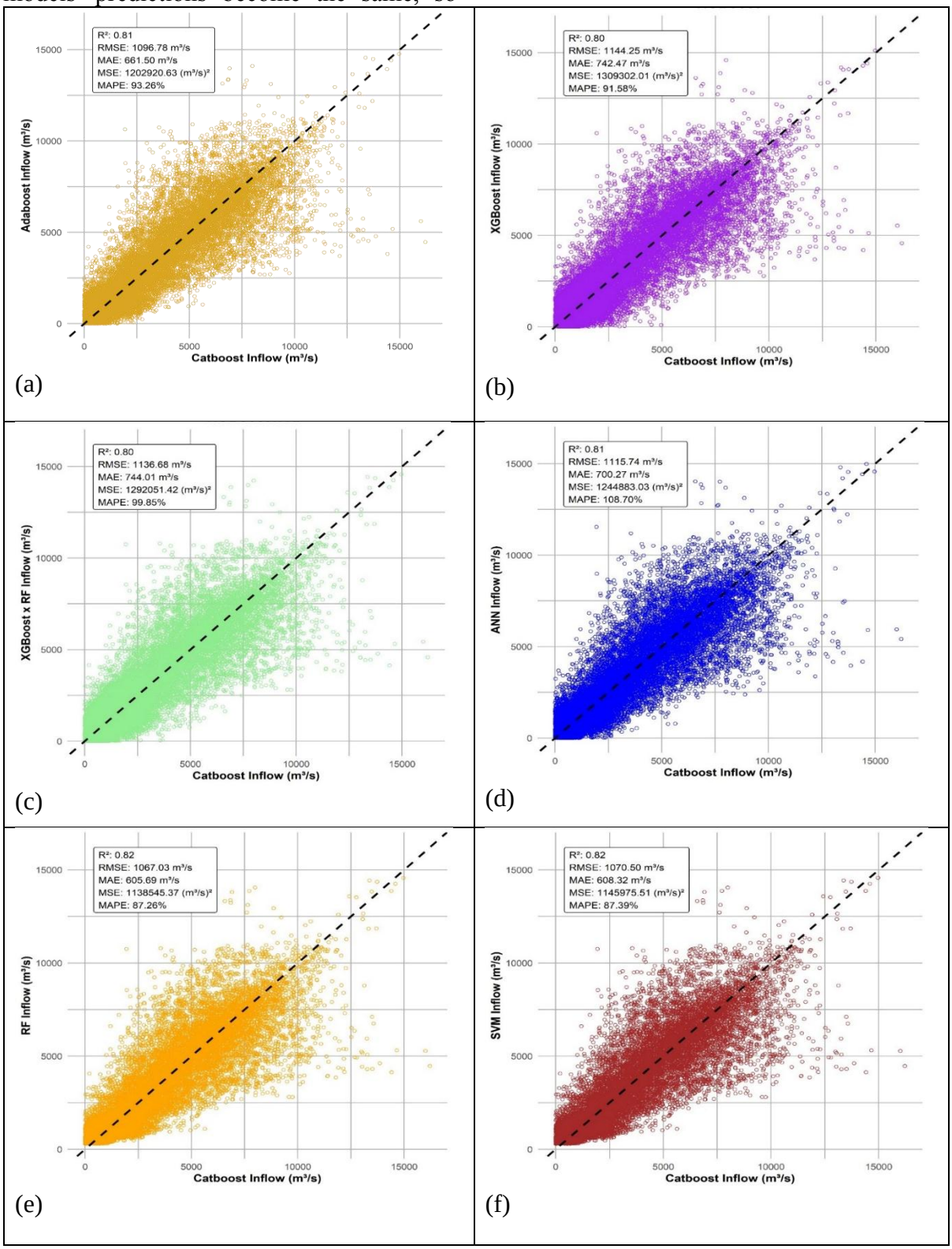


Fig. 21. Scatter plot comparing ML models predicted inflows to Catboost predicted inflows under SSP 5-8.5 upto year 2100. a) AdaBoost, b) XGboost, c) XGboost x RF, d) ANN, e) RF, and f) SVM.

4.3 Model evaluation metrics of all models under both SSP scenarios

Table 3 shows the values of R^2, RMSE, MAE, MSE, MAPE evaluation metrics for SSP2-4.5 and SSP5-8.5 future scenarios.

To further enhance physical interpretability, a SHAP analysis is conducted to evaluate the combined importance of meteorological parameters across the 14 stations as shown in Figure 22. The resulting heat map demonstrates that minimum temperature T_{min} is the primary driver of reservoir inflow across nearly all stations, with the highest importance observed at high-altitude stations such as Deosai (230.15), Rama (207.81), and Gilgit (205.42). This spatial dominance of T_{min} suggests that the model effectively prioritizes night-time thermal conditions, which regulate the continuous melting of the cryosphere in the Upper Indus Basin. Additionally, solar radiation at Deosai (110.40) and Besham (80.76) shows significant importance, indicating that the model captures the radiative energy balance necessary for triggering seasonal melt cycles. In contrast, precipitation and relative humidity show relatively lower, yet consistent, secondary importance across the basin.

5. Discussion

The analysis of eight different machine learning models for predicting monthly inflow showed a significant range in performance, with the Catboost displaying the best overall performance with an R^2 value of 0.92 during training and 0.87 during testing, as well as low RMSE values (291.42 and 811) further confirming a high degree of agreement between observed and predicted monthly inflows. Ensemble learners including Random Forest (RF) or hybrid XGBoost x RF were shown to perform well with perhaps too much training accuracy (R^2 = 0.98 and 0.91 as train and R^2 = 0.88 and 0.84 test accuracy for Random Forest and XGBoost x RF, respectively), but the combined models performed a bit worse in their generalization capability to the data set when testing their predictive capabilities resulting in R^2 variability of 0.88 or 0.87 respectively.

Table 3: Model evaluation metrics of all models under both scenarios.

MODELS	SSP2-4.5					SSP5-8.5				
	R^2	RMSE	MAE	MSE	MAPE	R^2	RMSE	MAE	MSE	MAPE
AdaBoost	0.85	785.42	478.08	616885	80.13	0.81	1,096.78	661.50	1202926	93.26
ANN	0.83	819.75	514.76	671990	96.21	0.81	1,115.74	700.27	1244876	108.07
XGBoost x RF	0.84	807.33	532.52	651782	83.63	0.80	1,136.68	744.01	1292041	99.85
Random Forest	0.86	729.59	414.82	532302	74.42	0.82	1,067.03	605.69	1138553	87.26
XGBoost	0.85	783.75	505.82	614264	74.60	0.80	1,144.25	742.47	1309308	91.58
SVM	0.86	756.52	431.10	572323	74.01	0.82	1,070.50	608.32	1145970	87.39



Fig. 22. Spatial meteorological feature importance heatmap for Catboost as the best predicting model.

Artificial Neural Network (ANN) performed reasonably well with an R^2 of 0.91 and an R^2 for testing of 0.86, but both RMSE (765.56 and 438.41 respectively during training) and MAE showed that it had difficulties encoding the more complex response of peak flows and other nonlinear responses compared to tree-based ensembles. The XGBoost and SVM models performed relatively similarly with moderate performance overall both with the same R^2 values for training (0.84 to 0.81 respectively), but significantly higher RMSE values were recorded in the range of 964 to 979.

The Decision Tree produced the weakest performance overall, again it had the lowest predictive stability with RMSE values >945 in training and nearly 981 in testing, and the model produced very high Mean Squares Error numbers above 9,623,000, again signaling high overfitting and poor predictive ability. While several models achieved higher training scores, Catboost demonstrated the most robust generalization capability. The relatively narrow gap between its training R^2 (0.92) and testing (0.87) suggests superior resistance to overfitting. This robustness is likely due to Catboost's ability to process categorical features and residuals in a sequential, unbiased manner, which is particularly effective for the non-linear, seasonal inflow patterns of the Tarbela Dam. When looking towards inflow projections, skill decreased in model predictions, regardless of emission pathways; but the decrease in performance diminished further in SSP5-8.5 level projections. Under SSP2-4.5, the SVM and the Random Forest models showed the most reliable skill of predictions with an $R^2 = 0.86$ prediction.

The AdaBoost and the XGBoost models showed a skill of $R^2 = 0.85$ and both the hybrid XGBoost × RF and the ANN models showed skills of $R^2 = 0.84$ and $R^2 = 0.83$ respectively. Catboost baseline values stayed consistent even though they did not present their skill with this metric table. All models showed decreased performance under SSP5-8.5: Random Forest and SVM dropped to $R^2 = 0.82$, AdaBoost and ANN dropped to $R^2 = 0.81$ and XGBoost and XGBoost × RF models dropped to $R^2 = 0.80$ suggesting a much greater uncertainty in the predictions under increased climate forced predictions. Increased RMSE scale experienced an increase from 729–820 under SSP2-4.5 to as high as 1,096–1,144 under SSP5-8.5 would further confirm the model difficulty to represent increased hydrological variability. To summarize, projections under SSP5-8.5 would suggest a more difficult forecast environment related to the acute climate impacts that higher variability of snowmelt and glacier melt driven flows would reduce confidence in long-term inflow estimates that are vital and used for operational purposes at the Tarbela Dam.

In a comparable study conducted by Jan et al. (2024) on UIB, XGBoost modelled the input flow into Tarbela Dam with the highest accuracy achieving an $R^2 = 0.92$ during training and $R^2 = 0.94$ during testing, followed closely by the Random Forest model with $R^2 = 0.89$ and hybrid XGBoost-Random Forest at $R^2 = 0.81$. Similarly, our findings indicated strong performance from ensemble learning models; however, Catboost yielded the highest predictability compared to all other algorithms, with $R^2 = 0.92$ in training data and $R^2 = 0.87$ in tests indicating a better capability of generalization than XGBoost. The primary driver for improvement is that Catboost built-in model structure works better to account for nonlinear interaction and seasonal variability in inflow. While Jan et al. (2024) only utilized data for short-term predictions up to 2023, we projected forward with changing climate influenced future conditions that indicated a decline in model skill that reached $R^2 = 0.80$-0.82 under the high emissions conformance scenario (SSP5-8.5) and greater uncertainty modeled for inflows throughout the study window.

Although the highest ranked model differed between the studies, Jan et al. (2024) conclude that both boosting based models and tree ensemble models are well suited for working with hydrology forecasts for the UIB, while those long-term climate uncertainties highlight the need for more rigorous approaches to assessing climate

adaptive practices for the long-term operation of Tarbela Dam. In a related study conducted by Khan et al. (2025) on the Upper Indus Basin, the authors also assessed machine learning performance for forecasting inflows to Tarbela and reported that some ensemble learners like Random Forest provided higher predictive accuracy relative to single tree or simpler models. The authors also noted that despite only a limited number of feature inputs, their model was able to establish strong correlation and robustness in predicting inflow variability.

These findings matched the conclusions of the present study, in which Random Forest and hybrid XGBoost × RF models had high accuracy (testing $R^2 \approx 0.87$–0.88) while also reliably capturing non-linearity. However, Khan et al. (2025) considered fewer climatic drivers than those considered in our modeling approach; in particular, humidity and sunshine were also considered as predictors in our models and were likely partly responsible for the better performance of Catboost and more stability in the ensembles throughout.

Jan (2025) similarly studied forecasting inflows to Tarbela and found that Random Forest resulted in the strongest performance with $R^2 \approx 0.89$ and lower RMSE compared to kNN models. The noted finding of kNN models underestimating high-flow events is consistent with the current work, as higher flow events were likely associated with non-linear processes driven by glacier melting and the kNN framework has limits on capturing non-linear processes. Tree-based ensemble models like Catboost and Random Forest both achieved higher testing accuracy (by $R^2 = 0.87$–0.88) and much lower RMSE ($\leq$ 811) compared to other methods. In contrast, even simpler models such as a kernel-based SVM and Decision Tree captured a lower range of testing R^2-value, 0.81–0.83, and a significantly higher RMSE (964-981), confirming that these methods could struggle to reproduce peak inflows. The results of this study emphasize the promising predictive capabilities of the more advanced machine learning approaches and could yield meaningful decision-support advantages for Tarbela Dam. The station-specific analysis provided by the SHAP heat map confirms that the Catboost model aligns with the glacio-hydrological reality of the Tarbela catchment. The concentration of high importance scores for T_{min} and solar radiation at high-elevation stations like Deosai (4,137.9 ft) highlights the model's reliance on thermal forcing for snow and glacier melt. By identifying specific stations as key contributors to predictive skill, this framework moves beyond a 'black-box' approach and offers actionable insights for dam operators, who can now identify which sub-basins are most critical for real-time inflow monitoring. Reliable pre-seasonal forecasts of inflow particularly during the spring to summer months when glacier and snowmelt dominates will support optimal reservoir storage for hydropower generation and irrigation supply in the agricultural system across Pakistan.

Similarly, the capacity to identify progression in reduced inflow during transitional months such as autumn could support early-stage planning for water allocation, sediment management, and minimum operational levels. In addition, the climate-scenario projections suggest that greater intensification of warming may also change future inflow variability of inflow even more; thus, combination of incorporating flexible operation rules and strategies to support proactive adaptation will be necessary. Overall, the strength of combining reliable inflow prediction modelling with possible future climate information in this study improves the adaptive capacity of those tasked with water management and operation of Tarbela Dam, and supports long-term sustainability, flood safe operation, and energy security while confronting changing hydrologic conditions.

6. Conclusions

An accurate inflow forecast is essential for the effective management of the Tarbela Dam and the broader Upper Indus Basin (UIB) in the context of a changing climate. The focus of this research was the evaluation of eight machine learning (ML)

models for monthly inflow forecasting using historical and projected climate data from the SSP2-4.5 and SSP5-8.5 scenarios. The study presents useful information on the prediction skill of advanced ensemble and boosting algorithms to capture the intricacies of UIB hydro-climatic dynamics. The key findings are:

1. Out of all the models, Catboost showed the best overall performance with R^2 = 0.92 in training and R^2 = 0.87 in testing, as well as the lowest RMSE (291.42 and 811), indicating a good amount of agreement between observed inflows versus predicted inflows. Ensemble models, such as Random Forest and XGBoost × RF also performed well in R^2 testing (0.87–0.88) but were slightly less robust in generalization due to minor overfitting in training.
2. The artificial neural network (ANN) model displayed moderate prediction skill (R^2 = 0.91 training, 0.86 testing), but higher RMSE and MAE values showed its ability to produce realistic predictions in hockey-stick-shaped nonlinear cases and events with peak flows was limited compared to ensemble learners. XGBoost and SVM performed similarly (R^2 = 0.81–0.84), while the Decision Tree had the weakest performance ($R^2 \approx 0.80$) and largest RMSE (>945), indicating limited capacity to adapt to dynamic inflow situations.
3. For future climate conditions, predictive skill for all models decreased. Acceptable predictive skill was still evident in mean annual cumulative snowmelt and glacier melt while using SSP2-4.5 (R^2 = 0.85–0.86; RMSE = 729–820), but under SSP5-8.5 it diminished (R^2 = 0.80–0.82; RMSE = 1,096–1,144). Predictive skill improvement determines how the influence of uncertainty operates under a stronger climate forcing and variability in dynamic contributions related to snowmelt and glacier melt to actual Tarbela inflows.
4. The results provide actionable value for the management of the Tarbela Dam in terms of improved reservoir regulation, hydropower scheduling, and irrigation water management. Using reliable pre-seasonal forecasts, storage can be optimized during periods of high inflows, while proactive pre-determination of decreases in inflows will allow decision-makers to act as a dry season develops.

The study highlights the increasing importance of data-driven intelligence in the practice of hydrological forecasting and water resources management. The linkage of more sophisticated machine learning models with climate change projections for future hydrologic modeling can greatly enhance the adaptive capacity of decision-makers for managing inflow variability, providing energy reliability, and assuring long-term water security in Pakistan. Future work should focus on improving inflow modeling that considers glacier mass balance, snow-cover extent, and satellite-based hydrological indices, providing improved and predictive accuracy to assist with sustainable basin-wide water management.

Acknowledgements

The authors are thankful to the Water and Power Development Authority (WAPDA) for providing data for this study.

Funding

No funding was received for this study.

Author contributions

Alamgir Khalil provided the study conception and design. Farjad Aziz contributed to the material preparation, data collection, and analysis, and wrote the manuscript. Muhammad Qasim and Mobeen Ahmed helped revise the manuscript.

Conflict of interest

The authors declare there is no conflict.

References

Bernardes, J., Jr., Santos, M., Abreu, T., Prado, L., Jr., Miranda, D., Julio, R., Viana, P., Fonseca, M., Bortoni, E., & Bastos, G. S. (2022). Hydropower operation optimization using machine learning: A systematic review. *AI, 3*(1), 78–99. https://doi.org/10.3390/ai3010006

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018

Dampage, S., Rathnayake, U., & Rajapaksha, R. (2020). Application of artificial neural networks in rainfall–runoff modeling: A case study in the Nilwala River Basin, Sri Lanka. *Journal of Water and Climate Change, 11*(3), 743–757. https://doi.org/10.2166/wcc.2019.091

Dibike, Y. B., Velickov, S., Solomatine, D., & Abbott, M. B. (2001). Model induction with support vector machines: Introduction and applications. *Journal of Computing in Civil Engineering, 15*(3), 208–216. https://doi.org/10.1061/(ASCE)0887-3801(2001)15:3(208

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences, 55*(1), 119–139. https://doi.org/10.1006/jcss.1997.1504

Gupta, A., Khare, D. & Shukla, R. 2022. Comparative analysis of machine learning models for streamflow prediction: Case study using XGBoost, Random Forest, and Support Vector Regression. *Journal of Hydrology*, 606, 127503. https://doi.org/10.1016/j.jhydrol.2022.127503

Haykin, S. S. (2009). *Neural networks and learning machines* (3rd ed.). Pearson.

Jan, S. (2025). Reservoir inflow prediction using machine learning techniques. *Journal Civil Engineering Study, 5*(2), 53–62. https://doi.org/10.34001/ces.v5i02.1225

Jan, S., Khan, U., Khalil, A., Khan, A. A., Jan, H. A., & Ullah, I. (2024). Use of machine learning techniques in predicting inflow in Tarbela reservoir of Upper Indus Basin. *Journal of Agrometeorology, 26*(4), 501–504. https://doi.org/10.54386/jam.v26i4.2676

Khan, A. J. (2024). *Bias corrected and downscaled climate projections over Upper Indus Basin (based on CORDEX-RCM's)* [Data set]. HydroShare. https://doi.org/10.4211/hs.bd2b1ec6ac94441db3acdfddb97c1bfa

Khan, A., & Koch, M. (2020). River flow forecasting with Random Forest and Support Vector Regression models in the Upper Indus Basin. *Water, 12*(4), 936. https://doi.org/10.3390/w12040936

Khan, M. M., Sarwar, M. K., Zafar, M. A., Rashid, M., Tariq, M. A. U. R., Haider, S., Okasha, A. M., Dewidar, A. Z., Mattar, M. A., & Salem, A. (2025). Comparative analysis of inflow forecasting using machine learning and statistical techniques: Case study of Mangla reservoir and Marala Headworks. *Frontiers in Environmental Science, 13*, 1590346. https://doi.org/10.3389/fenvs.2025.1590346

Maier, H. R., Jain, A., Dandy, G. C., & Sudheer, K. P. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environmental Modelling & Software, 25*(8), 891–909. https://doi.org/10.1016/j.envsoft.2010.02.003

Mosaffa, H., Sadeghi, M., Mallakpour, I., Naghdyzadegan Jahromi, M., & Pourghasemi, H. R. (2022). Application of machine learning algorithms in hydrology. In *Computers in Earth and Environmental Sciences* (pp. 585–591). Elsevier. https://doi.org/10.1016/B978-0-323-89861-4.00027-0

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

Munir, M. M., Shakir, A. S., Rehman, H.-U., Khan, N. M., Rashid, M. U., Tariq, M. A. U. R., & Sarwar, M. K. (2022). Simulation-optimization of Tarbela Reservoir operation to enhance multiple benefits and to achieve Sustainable Development Goals. *Water, 14*(16), 2512. https://doi.org/10.3390/w14162512

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems 31* (pp. 6638–6648).

Rauf, A.U., Rafi, M.S., Ali, I., & Muhammad, U.W. (2016). Temperature trend detection in Upper Indus Basin by using Mann-Kendall test. *Advances in Science, Technology and Engineering Systems Journal, 1*(4), 5–13. https://doi.org/10.25046/aj010402

Sireesha, M., Mahammood, V., & Rao, K. N. (2023). Prediction of soil salinity in the Upputeru river estuary catchment, India, using machine learning techniques. *Environmental Monitoring and Assessment, 195*, 1006. https://doi.org/10.1007/s10661-023-11613-y

Vapnik, V.N. 1995. *The Nature of Statistical Learning Theory*. Springer. https://doi.org/10.1007/978-1-4757-2440-0

Wang, L., Wu, J., Zhang, W., Wang, L., & Cui, W. (2021). Efficient Seismic Stability Analysis of Embankment Slopes Subjected to Water Level Changes Using Gradient Boosting Algorithms. *Frontiers in Earth Science*, 9, 807317. https://doi.org/10.3389/feart.2021.807317.

Zarei, M., Bozorg-Haddad, O., Baghban, S., Delpasand, M., Goharian, E., & Loáiciga, H. A. (2021). Machine-learning algorithms for forecast-informed reservoir operation (FIRO) to reduce flood damages. *Scientific Reports, 11*, 24295. https://doi.org/10.1038/s41598-021-03699-6