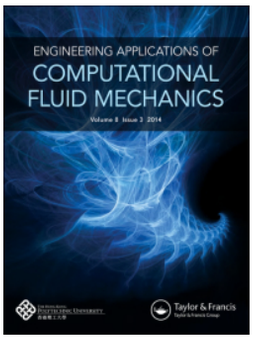

Engineering Applications of Computational Fluid Mechanics

ISSN: 1994-2060 (Print) 1997-003X (Online) Journal homepage: www.tandfonline.com/journals/tcfm20

Development of interpretable intelligent frameworks for estimating river water turbidity

Amin Gharehbaghi, Salim Heddam, Saeid Mehdizadeh & Sungwon Kim

To cite this article: Amin Gharehbaghi, Salim Heddam, Saeid Mehdizadeh & Sungwon Kim (2025) Development of interpretable intelligent frameworks for estimating river water turbidity, Engineering Applications of Computational Fluid Mechanics, 19:1, 2511886, DOI: 10.1080/19942060.2025.2511886

To link to this article: https://doi.org/10.1080/19942060.2025.2511886

Published online: 06 Jun 2025.

ENGINEERING APPLICATIONS OF COMPUTATIONAL FLUID MECHANICS
2025, VOL. 19, NO. 1, 2511886
https://doi.org/10.1080/19942060.2025.2511886

Taylor & Francis
Taylor & Francis Group

Development of interpretable intelligent frameworks for estimating river water turbidity

Amin Gharehbaghi [a], Salim Heddam [b], Saeid Mehdizadeh [c] and Sungwon Kim [d]

[a]Department of Civil Engineering, Faculty of Engineering, Hasan Kalyoncu University, Şahinbey, Gaziantep, Türkiye; [b]Faculty of Science, Agronomy Department, Hydraulics Division, University 20 Août 1955 Skikda, Skikda, Algeria; [c]Water Engineering Department, Urmia University, Urmia, Iran; [d]Department of Railroad Construction and Safety Engineering, Dongyang University, Yeongju, Republic of Korea

ABSTRACT

Turbidity (TU) is one of the paramount water quality indicators in rivers and streams. Therefore, knowledge of water TU plays a fundamental role in optimal managing and monitoring river water quality. This study aimed at developing four intelligent schemes including three boosting methods i.e. Categorical Boosting (CatBoost), Light Gradient-Boosting Machine (LightGBM), eXtreme Gradient Boosting (XGBoost), and a deep learning method named Convolutional Neural Networks (CNN). To evaluate the performance of proposed models, two gauging river stations situated in United States (i.e. USGS 14206950 and USGS 14211720) were selected as a case study. 70% and 30% of whole data were utilized as the training and validation datasets when developing the models, respectively. It is worthwhile to note that the development of boosting models and their performance companions with a deep learning model, as well as addressing the impacts of input features on the models' outputs using SHapley Additive exPlanations (SHAP) are the novel aspects of this study, which have been rarely considered in preceding studies for river water TU estimation. Based on the achieved results during the validation period, the CatBoost and XGBoost models were found to be generally best models for an accurate estimation of river water TU in the studied sites. During the validation period, the best-performing models were XGBoost (R = 0.951, NSE = 0.903, RMSE = 3.552 FNU, MAE = 1.816 FNU) at USGS 14206950, and CatBoost (R = 0.961, NSE = 0.920, RMSE = 2.502 FNU, MAE = 1.219 FNU) at USGS 14211720 both established using full-input estimators. An interpretability assessment of the developed models was finally conducted taking into account the SHAP. Analysis of the SHAP graphs in a global level during the validation phase illustrated that river discharge was the most important input variable affecting the output results of the best-performing implemented models.

ARTICLE HISTORY
Received 7 January 2025
Accepted 17 May 2025

KEYWORDS
Water turbidity; SHapley additive explanations; estimation; boosting models

1. Introduction

Rivers are among the most important water resources for meeting agricultural and industrial demands. In recent years, population growth and increased agricultural and industrial activities have led to a decrease in the quality of surface waters such as rivers. Therefore, knowledge of water quality seems to be an essential issue in the planning, development, and protection of water resources (Allawi et al., 2024; Yao et al., 2024). To understand the quality of river water, different variables of river water can be taken into consideration including water temperature (Gheisari et al., 2025; Sun et al., 2024), dissolved oxygen (Li, Qasem, et al., 2024), specific conductance (Saccotelli et al., 2024), turbidity (Rahaman et al., 2024), etc. Of them, turbidity (TU) is typically considered as one the major river water quality indicators influential on the sustainable development of aquatic ecosystems (Ding et al., 2021; Lin et al., 2023). The TU in river water is originated due to the high concentrations of the suspended solids in rivers (Heddam, 2023; Park et al., 2017). It can also be used as an efficient index to identify light scattering intensity (Gelda et al., 2009; Heddam, 2023).

The river water TU estimation can be conducted in two direct and indirect frameworks. Direct approach is based on an *in-situ* measurement or laboratory sampling (Mousavi, 2024). However, some problems like their high costs, the need for frequent maintenance, and being time-consuming have limited their applications in practice (Zounemat-Kermani et al., 2021). For this, indirect methods have been the interest of researchers globally for the river water TU estimation. With the fast development of computer science in recent years, artificial intelligence (AI) methods have received much more attention in solving engineering problems. The AI-based

CONTACT Saeid Mehdizadeh s.mehdizadeh@urmia.ac.ir Water Engineering Department, Urmia University, P.O. Box 165, Km 11 Sero Road, Urmia, Iran

models are based on developing data-driven methods e.g. machine learning and deep learning schemes using a series of independent input estimators influencing the target variable. As it is apparent, the river water TU could be affected and therefore estimated by other river water data consisting of the river discharge, rainfall, river water temperature, dissolved oxygen, PH, specific conductance, total dissolved solids. Hence, indirect frameworks like AI could be considered as dependable alternatives to the direct methods in river water TU estimation.

The concept of AI can be defined as the ability of computer-based affair to perform the various tasks (Arrieta et al., 2020; Tizhoosh & Pantanowitz, 2018). The classification of AI can involve various scientific fields including machine learning, deep learning, expert system, natural language processing, computer vision, and robotics, etc. (Mukhamediev et al., 2021). In addition, the most application of AI is based on machine learning (ML), which is a powerful weapon for data analysis (Usuga Cadavid et al., 2020). ML approaches can be classified depending on various learning algorithms including supervised learning (Kotsiantis et al., 2007), unsupervised learning (Mukhamediev et al., 2022), semi-supervised learning (Van Engelen & Hoos, 2020), reinforcement learning (Elavarasan & Vincent, 2020), and deep learning (LeCun et al., 2015), etc.

The useful advantages of machine and deep learning models can be explained as task and cost efficiency, improvement of decision making, and innovation ability and so on, while the disadvantages of machine and deep learning models are data dependency, complexity, and interpretability, etc. (Ahmed et al., 2023; Dahiya et al., 2022).

A review of previous literature clearly showed that the main focus of the preceding studies has been on estimating/forecasting other river water quality variables like water temperature (Philippus et al., 2024) and dissolved oxygen (Hu et al., 2024), while the river water turbidity has received less attention in this regard. Here, some of the previous studies are briefly reviewed. Gu et al. (2022) proposed a self-organizing multichannel deep learning system, and approved its dependable potential for river water TU modeling. A random forest (RF) was used by Lin et al. (2023), and concluded that it can be considered as a reliable tool to simulate river water TU. Park and Lee (2020) proposed a long short-term memory (LSTM) to overcome the shortcomings of artificial neural networks (ANN), and reported its good precision in predicting high TU values in rivers. Hu et al. (2023) applied an LSTM for the lake water TU time series prediction. The outcomes denoted that the LSTM can be considered as an alternative model for accurately TU forecasting. Mousavi (2024) proposed two wavelet-based hybrid methods utilizing the wavelet pre-processing technique coupled on ANN and adaptive neuro-fuzzy inference system (ANFIS). The outcomes illustrated that the data denoising via the wavelet could remarkably increase the performance of the ANN and ANFIS in forecasting river water TU. Rajaee and Jafari (2018) applied the wavelet theory to couple on the ANN, gene expression programming (GEP), and decision tree (DT) to predict river water TU. Their findings revealed that the wavelet could significantly increase the performance of standalone models. Zounemat-Kermani et al. (2021) proposed online sequential version of extreme learning machine (OS-ELM), and then compared its accuracy with four types of other ML models. It was concluded that the OS-ELM was the best-performing technique for the precise river water TU prediction. Teixeira et al. (2020) developed the ANN and FIS models for the river water TU prediction, and found out that the FIS performed better than the ANN. A nature-inspired-based hybrid method was proposed by Heddam (2023) through coupling Bat algorithm and ELM for forecasting the river water TU. The developed Bat-ELM was found to present superior results compared to feedforward neural network (FFNN) and dynamic evolving neural-fuzzy inference system (DENFIS).

As noted, river water turbidity (TU) is one of the important indicators to monitor and therefore manage river water quality. For this, TU forecasting tools need to be developed. In this research, four types of machine and deep learning paradigms including Categorical Boosting (CatBoost), Light Gradient-Boosting Machine (LightGBM), eXtreme Gradient Boosting (XGBoost), and a Convolutional Neural Network (CNN) were used to estimate daily water TU time series at two river stations located in the United States. The selected sites were USGS 14206950 Fanno Creek at Durham, and USGS 14211720 Willamette River at Portland, both located in Oregon State, United States.

Boosting ensemble models including CatBoost, LightGBM, and XGBoost and so on, which reduce the overfitting and underfitting problems utilizing diverse features of data, can supply boosted accuracy and efficiency for complex and difficult issues (Ganaie et al., 2022; Yang et al., 2023). CNN model includes fewer parameters compared to the fully connected neural networks. It can also reduce seriously the computational complexity of data compared to the fully connected neural networks, and employ the scanty connectivity arrangement and mutual weights (Lee et al., 2022; Yamashita et al., 2018).

During the development of models, eight different input configurations were defined using other available river data. For this, two different categories were considered (i.e. four scenarios using river water data and four

scenarios using the same inputs considering the periodicity terms). Some error evaluation metrics such as R, NSE, RMSE, MAE, and visual graphs including scatter, violin, and Taylor diagrams were employed to investigate the performance of all the developed models. SHapley Additive exPlanations (SHAP) analysis was eventually utilized to evaluate the interpretability of best-performing models. It is worth noting that applying boosting methods like CatBoost, LightGBM, and XGBoost, and comparing their performances with a deep learning-based method (i.e. CNN), as well as evaluating the interpretability of the applied methods are the novel perspectives of this study, which have not been used so far in estimating river water TU time series. Additionally, river water TU was considered as the target variable of this study, whereas previous literature has focused more on other river water quality variables like water temperature and dissolved oxygen. Hence, the issues raised above were the authors' motivations for conducting this study.

2. Materials and methods

2.1. Case study and data description

To investigate the accuracy of established models in this study, two river stations consisting of USGS 14206950 Fanno Creek at Durham, OR (Latitude 45°24′13″, Longitude 122°45′13″) and USGS 14211720 Willamette River at Portland, OR (Latitude 45°31′03″, Longitude 122°40′09″) were selected as the case study. The Fanno Creek is a 24-km tributary of the Tualatin River in the Oregon State of the United States. The Willamette River is also a main tributary of the Columbia River in the Oregon State, and it accounts for about 12–15 percent of the flow of the Columbia River. The required data to accomplish this study is composed of mean daily river water turbidity (TU) as the target variable, and other river data as the possible input estimators including discharge (Q), water temperature (T_w), specific conductance (SC), PH, and dissolved oxygen (DO) collecting by United States Geological Survey (USGS). The data gathering from the https://waterdata.usgs.gov/nwis/ website spans the time period from 1/1/2014 to 12/31/2022. The whole datasets were split into the training (70%) and validation (30%) subsets. Some of the statistical metrics for the used data such as mean (X_{mean}), maximum (X_{max}), minimum (X_{min}), standard deviation (S_x), and coefficient of variation (C_v) are summarized in Table 1. The Heatmaps showing the correlations between all water quality variables are depicted in Figure 1.

2.2. Models used

2.2.1. Categorical Boosting (CatBoost)

CatBoost (Categorical Boosting), which can be explained as a correct approach of categorical features in the predictive process, was recommended by Prokhorenkova et al. (2018). It lessens the complex topic associated with low bias and high variance and increases generalized performance. To lessen overfitting and underfitting problems embedded in CatBoost, users and developers provide a technique that merges a modified formation of an ordered boosting approach (Chen et al., 2024; Wang et al., 2023). The indicators employed in the training performance of the ordered boosting approach are arranged in accordance with a predefined order, creating various models relying on the sample cases of training performance (Hancock & Khoshgoftaar, 2020; Krishnan & Manikandan, 2024). CatBoost randomly creates all sampling dataset and organizes them into the identical class overall feature indicators. Also, when adapting the attitude of individual sampling dataset relying on numeric access, the output value of the sampling dataset is first calculated before the sampling process, and the comparable weights and arrangements are classified (Li et al., 2022; Prokhorenkova et al., 2018). It also applies a greedy approach to supply sequential measures via an iterative procedure for minimizing a loss function (Li, Yang, et al., 2024; Prokhorenkova et al., 2018) The new introduction and utilization of CatBoost for water quality prediction can be provided from various research articles including Li et al. (2022), Wang et al. (2023), and Krishnan and Manikandan (2024) and so on. Hyperparameters used to develop the CatBoost models were as follows: loss_function = 'RMSE', 'iterations': [80,80,80], 'learning_rate': [0.6,0.2], 'depth': [6,6,6,6], 'l2_leaf_reg': [0.05,0.01,0.1,0.3], 'optimization method': grid search.

2.2.2. Light Gradient-Boosting Machine (LightGBM)

LightGBM (Light Gradient Boosting Machine), one of the boosting ensemble approaches, can solve complicated and hard topics embedded in the universal boosting ensemble topics including time, small memory, and complexity (Shehadeh et al., 2021; Zhou et al., 2022). The decision classifiers enclosed on the boosting ensemble approach can be employed for managing and operating regression and classification topics (Bian et al., 2023; Gan et al., 2024). They can merge the gradient-based one-sided sampling and exclusive feature bundling sampling for creating regression and classification topics (Ke et al., 2017; Khoi et al., 2022). LightGBM manages a histogram method connected with a depth-constraint

Table 1. Summary statistics of variables for Water TU modelling.

Variables	Subset	Unit	X_{mean}	X_{max}	X_{min}	S_x	C_v	R
				USGS 14206950				
Q	Training	m^3/s	1.197	29.166	0.028	2.418	2.021	0.712
	Validation	m^3/s	1.353	24.579	0.032	2.539	1.877	0.731
	All data	m^3/s	1.243	29.166	0.028	2.456	1.975	0.716
Tw	Training	℃	13.424	27.200	0.100	5.531	0.412	−0.329
	Validation	℃	13.150	24.500	0.600	5.557	0.423	−0.392
	All data	℃	13.342	27.200	0.100	5.540	0.415	−0.347
SC	Training	µS/cm	226.14	747.00	65.000	85.23	0.377	−0.491
	Validation	µS/cm	220.81	637.00	76.000	85.30	0.386	−0.533
	All data	µS/cm	224.54	747.00	65.000	85.27	0.380	−0.503
pH	Training	/	7.285	7.900	6.500	0.173	0.024	−0.436
	Validation	/	7.284	7.800	6.600	0.176	0.024	−0.493
	All data	/	7.285	7.900	6.500	0.174	0.024	−0.452
DO	Training	mg/L	8.657	13.200	1.800	2.053	0.237	0.302
	Validation	mg/L	8.812	13.000	3.300	1.988	0.226	0.346
	All data	mg/L	8.704	13.200	1.800	2.035	0.234	0.314
TU	Training	FNU	10.389	170.00	0.300	12.27	1.182	1.000
	Validation	FNU	10.686	113.00	0.600	11.41	1.068	1.000
	All data	FNU	10.478	170.00	0.300	12.02	1.147	1.000
				USGS 14211720				
Q	Training	m^3/s	850.72	4983.7	52.103	820.3	0.964	0.862
	Validation	m^3/s	835.59	4898.8	94.295	825.3	0.988	0.895
	All data	m^3/s	846.18	4983.7	52.103	821.7	0.971	0.870
Tw	Training	℃	13.615	26.200	1.900	6.008	0.441	−0.443
	Validation	℃	13.861	26.100	2.100	6.200	0.447	−0.473
	All data	℃	13.688	26.200	1.900	6.067	0.443	−0.451
SC	Training	µS/cm	78.955	107.00	51.000	9.758	0.124	−0.467
	Validation	µS/cm	79.391	103.00	53.000	9.855	0.124	−0.533
	All data	µS/cm	79.086	107.00	51.000	9.788	0.124	−0.485
pH	Training	/	7.308	8.100	6.800	0.145	0.020	−0.451
	Validation	/	7.308	8.300	6.900	0.152	0.021	−0.444
	All data	/	7.308	8.300	6.800	0.147	0.020	−0.447
DO	Training	mg/L	10.996	14.700	6.000	1.838	0.167	0.565
	Validation	mg/L	10.913	14.700	7.000	1.903	0.174	0.602
	All data	mg/L	10.971	14.700	6.000	1.858	0.169	0.574
TU	Training	FNU	7.192	101.00	0.600	10.09	1.404	1.000
	Validation	FNU	6.874	73.000	0.800	8.866	1.290	1.000
	All data	FNU	7.096	101.00	0.600	9.745	1.373	1.000

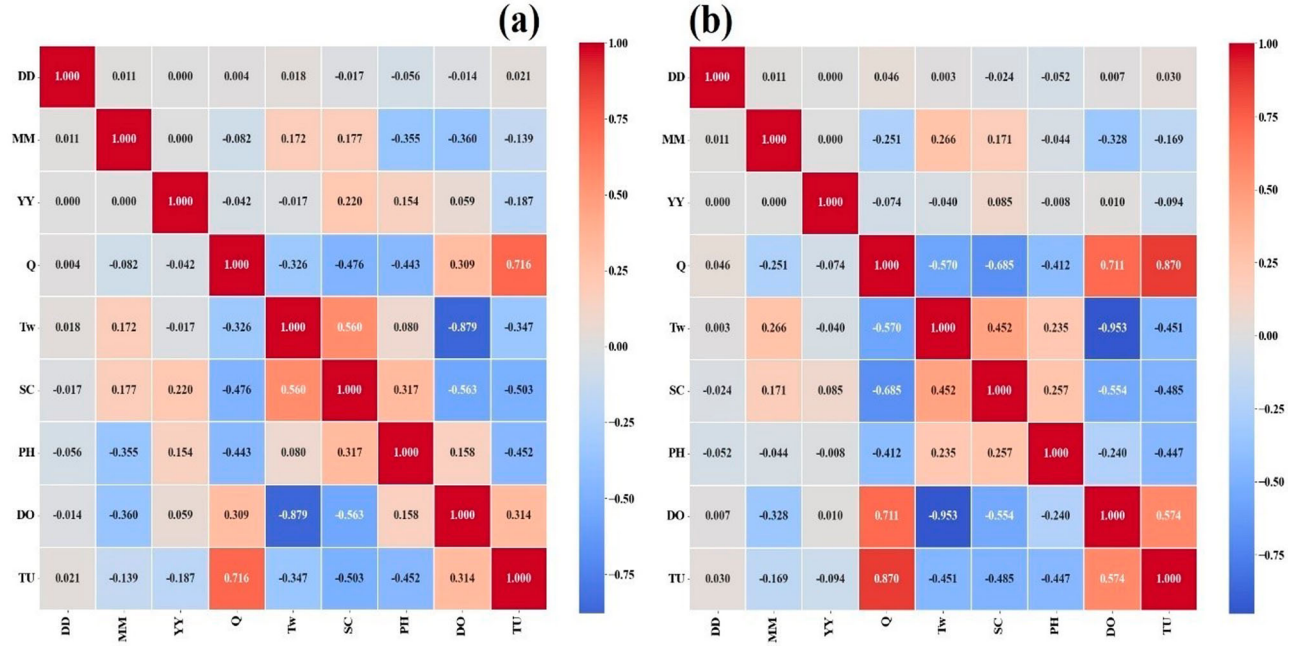

Figure 1. Heatmaps indicating correlations between the input estimators and TU: (a) USGS 14206950 and (b) USGS 14211720.

scheme for a leaf-wise growth system for minimizing memory consumption. It carries out larger accuracy with less computational cost while mitigating the risks associated with gradient boosting decision tree (GBDT). Also, it connects a maximum depth constraint approach and a design feature to increase accuracy while reducing the risk of overfitting and underfitting problems for leaf-wise growth system (Heydari et al., 2024; Shehadeh et al., 2021). LightGBM improves the direct feature indicator of classification. Most of the various boosting ensemble approaches need hot encoding to express the categorical feature (Ke et al., 2017; Liu et al., 2023). It also supplies fast speed and efficient memory utilization, and displays outstanding processes for big dataset units. Furthermore, it demonstrates high performance of nonlinear predictive topics for various kinds of dataset including numerical, categorical, and sparse indicators (Ahn et al., 2023). Excellent research on employing LightGBM in water quality prediction can be found in various articles including Li et al. (2022), Zhou et al. (2022), and Bian et al. (2023) etc. During the establishment of LightGBM models, the following hyperparameters were utilized: 'objective': regression, 'num_leaves': 10, 'learnnig_rage': 0.1, 'metric': 'l2','l1', 'verbose': -1, 'optimization method': grid search.

2.2.3. Extreme Gradient Boosting (XGBoost)

Chen and Guestrin (2016) developed XGBoost (Extreme Gradient Boosting) boosting ensemble approach, which includes a hypothesis of gradient boosting classifier and an important effect on gradient improvement, as an accurate solution by weighting classifiers for regression and classification topics. XGBoost improves predictive accuracy and efficiency by combining parallel computing and regularization approaches. Since XGBoost carries out better performance a random chosen model by merging different sole models into boosting ensemble model, users and developers can create a powerful model for predictive topics (Uddin et al., 2023; Wang et al., 2022). In prediction, forecasting, and modeling topics, the boosting ensemble model involves the population of individual feature indicators for estimating the same output indicator. The feature indicators are merged so that the output indicator is integrated to increase accuracy and perform more outstanding results (Chen & Guestrin, 2016; Lu & Ma, 2020). Since individual feature classifier is restricted on predictive accuracy of performance, XGBoost implements a classifier boosting ensemble to increase the predictive accuracy of performance relying on a classification and regression tree (CART). In the addressed approach, the value of output indicator for final predictive procedure can be computed by adding the values of output indicator estimated from individual CART (Li et al., 2022; Wu et al., 2022). Therefore, XGBoost can furnish powerful and efficient instrument that manage a broad range of regression and classification topics (Chen & Guestrin, 2016). Different researches for solving water quality prediction can be suggested from many previous literatures including Chen and Guestrin (2016), Lu and Ma (2020), and Makumbura et al. (2024) and so on. The hyperparameters employed when developing the XGBoost models were as follows: n_estimators $= 500$, max_depth $= 5$, eta $= 0.1$, subsample $= 0.7$, colsample_bytree $= 1.0$, optimization method: grid search.

2.2.4. Convolutional neural network (CNN)

LeCun et al. (1989) developed CNN (convolutional neural network), which is one of the world's popular models in deep-learning paradigm, to be employed for practical applications including image classification, speech recognition, and computational biology and so on. In addition, CNN is an effective deep-learning paradigm for automatic feature extraction and has performed remarkable achievements in medical image vision (Ismail et al., 2020). Also, CNN is the most outstanding deep-learning paradigm for extracting image features including two (2D) and three (3D) dimensional datasets relying on a high resolution (Alizamir et al., 2025; Ehteram et al., 2024). The architecture of CNN can be divided into convolutional and pooling layers. Recognizing the fully connected network (i.e. dense network), neurons are connected to all neurons of the following layers. However, neurons connected to convolutional and pooling layers are hardly connected, which is selected by kernel and pooling sizes (Barzegar et al., 2020; Tan et al., 2022). CNN, which features sparse connectivity and shared weights, can capture global patterns and reduce model capacity. In general, a pooling layer is employed after a convolutional layer. It can change the output of the network at a certain location with a brief statistic of neighbouring outputs (Mei et al., 2022; Zhang et al., 2024). There are abundant hyperparameters in predictive topics of CNN including pooling size, filter size, and the number of convolutional and pooling layers. The addressed hyperparameters must be tuned to perform the predictive topics with excellent accuracy (Baek et al., 2020; Yang et al., 2021). Various research for accomplishing water quality prediction utilizing CNN can be supplied from previous documents including Barzegar et al. (2020), Yang et al. (2021), and Haq and Harigovindan (2022) etc. During the development of CNN models, the following hyperparameters were used: activation = 'ReLU', loss = 'mse', optimizer = 'adam'.

2.3. SHapley Additive exPlanations (SHAP)

Lundberg and Lee (2017) introduced SHAP (SHapley Additive exPlanations) method to deliver a combined structure for interpreting outputs of model relying on the view of game theory concept. SHAP demonstrates as a powerful instrument for identifying important feature indicators, and admits consistent analysis of predictive topics in the fields of water quality prediction (Aldrees et al., 2024; Park et al., 2022). It can be demonstrated as a problem to describe the predictive accuracy of boosting ensemble models relying on sensitivity analysis (Baptista et al., 2022). It also allocates important values to individual feature indicator for predictive topics and adjusts the accuracy utilizing boosting ensemble, machine learning, and deep learning models (Nallakaruppan et al., 2024; Van den Broeck et al., 2022). In SHAP method, the predictive procedure can be evaluated as a game with the feature indicators of model for players to generate output indicator (Meng et al., 2020; Zheng et al., 2023). Therefore, this method presents an instrument for separating predictions among the participating feature indicators (Yao et al., 2024).

SHAP method also evaluates the effect of individual feature indicator, and computes both direction and strength (Hasani et al., 2024; Meng et al., 2020). In addition, SHAP method displays two schemes including the global and local interpretation. The global interpretation scheme applies the importance of feature indicator and summary plots as an access of illustrated plots (Majnooni et al., 2024; Wang et al., 2021). The local interpretation scheme, however, employs force plots to display the corresponding SHAP for sole model. It also catches the tasks of feature indicators at the sample classification to supply more efficient understanding of sole model (Majnooni et al., 2024; Zhang et al., 2023). The complicated knowledge for theorem and application of SHAP method for solving predictive topics can be perceived from previous literature including Lundberg and Lee (2017), Wang et al. (2021), and Yao et al. (2024) and so on.

2.4. Performance assessment metrics

Here, we used four commonly used error criteria including coefficient of correlation (R), Nash-Sutcliffe efficiency (NSE), root mean square error (RMSE), and mean absolute error (MAE) to assess the models' performances in estimating river water TUs of the studied stations. They can be written as follows:

$$R = \frac{\sum_{i=1}^{N}(TU_{o,i} - \overline{TU_o}) \cdot (TU_{e,i} - \overline{TU_e})}{\sqrt{\sum_{i=1}^{N}(TU_{o,i} - \overline{TU_o})^2 \cdot \sum_{i=1}^{N}(TU_{e,i} - \overline{TU_e})^2}} \quad (1)$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}(TU_{o,i} - TU_{e,i})^2}{\sum_{i=1}^{N}(TU_{o,i} - \overline{TU_o})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(TU_{o,i} - TU_{e,i})^2}{N}} \quad (3)$$

$$MAE = \frac{\sum_{i=1}^{N}|TU_{o,i} - TU_{e,i}|}{N} \quad (4)$$

where $TU_{o,i}$ and $TU_{e,i}$ respectively indicate the ith observed and estimated river water turbidites, $\overline{TU_o}$ and $\overline{TU_e}$ respectively illustrate the average values for the observed and estimated river water turbidites, and N is the number of observations. It is apparent that a lower value for the RMSE and MAE, as well as higher R and NSE denote the better capability of that model in estimating river water TU. Besides the error measures mentioned above, some visual comparative graphs were also taken into consideration comprising of the scatter, Taylor, and violin plots.

3. Results and discussion

In this study, the river water TU time series of two river stations located in United States including USGS 14206950 and USGS 14211720 were estimated. To achieve this aim, four types of intelligent schemes i.e. CatBoost, LightGBM, XGBoost, and CNN were established. In order to feed the applied models, different input configurations using the rivers water data such as Q, Tw, SC, PH, and DO were defined, which are indicated under the first to fourth scenarios in Table 2. To better define the input combinations, the heatmaps of correlations between the various variables are prepared that were shown in Figures 1(a–b), respectively for USGS 14206950 and USGS 14211720 stations. The outcomes indicated that the river discharge (Q) variable presents highest positive correlations with TU as the target variable at both the considered sites. After that, SC at USGS 14206950 site and DO at USGS 14211720 locations showed higher correlations. Therefore, due to the highest correlation of water TUs with Q data, this variable was used as a fixed variable when defining the input scenarios, and other variables were added to this variable.

The values of error statistics including the RMSE, MAE, R, and NSE computed for the developed models under scenarios #1-4 at USGS 14206950 and USGS 14211720 stations are tabulated in Tables 3 and 4, respectively. In general, using the least (scenario 1, inputs: Q, T_w) and highest (scenario 4, inputs: Q, Tw, SC, pH, DO) inputs led to better TU estimations at both the sites, except for some cases.

Table 2. Different proposed models.

	Input variables combination	Output
Scenario1	Q, T_w	TU
Scenario2	Q, T_w, SC	TU
Scenario3	Q, T_w, SC, pH	TU
Scenario4	Q, T_w, SC, pH, DO	TU
Scenario5	DD, MM, YY, Q, T_w	TU
Scenario6	DD, MM, YY, Q, T_w, SC	TU
Scenario7	DD, MM, YY, Q, T_w, SC, pH	TU
Scenario8	DD, MM, YY, Q, T_w, SC, pH, DO	TU

It is worthwhile to mention that two different strategies were taken into consideration in this study when defining the input configurations in Table 2. Four scenarios mentioned above were identified using the river water data. In addition, the same four input scenarios were redefined taking into account the periodicity. In fact, we used three periodicity terms including the number of the day (DD), number of the months (MM), and number of the years (YY) to investigate the effect of periodicity on the performance of developed models. So, scenarios #5 to 8 were defined as shown in Table 2. The results exhibited that considering the periodicity terms as the input estimators could lead to better river water TU estimates at both stations because of the enhanced error criteria of the RMSE, MAE, R, and NSE. This conclusion is more visible on the USGS 14206950 site.

Evaluating the accuracy of implemented models in terms of obtained error metrics tabulated in Tables 3 and 4 clearly demonstrated that the XGBoost model was found to perform the best among all the developed models at both the studied locations in the training period. On the other side, LightGBM at the USGS 14206950 and LightGBM and CNN models at USGS 14211720 site generally demonstrated better results than the other models in the validation period under scenarios #1-4 (i.e. without considering the periodicity). Besides, CatBoost and XGBoost methods at USGS 14206950 site and CatBoost models at USGS 14211720 performed better compared to others when considering the periodicity term in the validation stage of scenarios #5-8.

In addition to examining the performance of the developed models in terms of error statistics, some types of schematic diagrams including the scatter, violin, and Taylor graphs were also taken into consideration. To this end, the results of scenario #8 during the validation phase were utilized as all the models generally illustrated the best outcomes for this scenario (i.e. full-input combination). The scatter graphs shown in Figures 2 and 3 could present the potential of developed models considering the distribution of estimated data against the observed TU values. Figures 2 and 3 clearly demonstrate that XGBoost at USGS 14206950 and CatBoost at USGS 14211720 are the superior models due to the relatively lower dispersion of the data and consequently higher R^2 values.

Boxplots and violin graphs were then taken into consideration to visually assess the performances of best-performing models that are given in Figures 4 and 5. In fact, a boxplot demonstrates the spread of the data in terms of their quartiles, while a violin plot visualizing the data distribution is a combined form of a boxplot and a kernel density plot. It is apparent that CatBoost and XGBoost models at USGS 14206950 site and LightGBM at USGS 14211720 station illustrate superiority over the other models developed due to the better schematics of the data distribution.

The efficiency of developed models could also be evaluated by addressing the results of standard deviation and correlation coefficient in a single diagram named the Taylor diagram. Taylor graphs obtained at the studied stations are demonstrated in Figure 6. In this diagram, the results for each model are shown with a special point. The proximity and distance of the relevant point of each model from the point corresponding to the observational data indicate the better and worse performance of that model, respectively. Evaluation of the Taylor diagrams shows that XGBoost at USGS 14206950 and CatBoost at USGS 14211720 present the closest distance to the reference point of the observational data (i.e. red star) and therefore are the superior methods.

Finally, interpretability of models was done taking into account the SHAP analysis. SHAP interpretability was conducted in this study for both the global and local modes. Similar to the visual comparative graphs including the scatter, violin, and Taylor diagrams, the best-performing models during the validation stage for each method (i.e. full-input pattern) were selected. Figures 7 and 8 show the global Interpretability of the applied models using SHAP algorithms. These figures are used for features importance ranking and more precisely, the numerical values displayed in the bar chart plot (left panel) are calculated as the mean of the absolute SHAP values displayed in the force plot (right panel). More precisely, the dot plotted in the force plot corresponds each one to a sample and, it is a positively or negatively values. This corresponds to the distribution of the calculated SHAP values for each feature. Each dot corresponds to a single sample and each raw is used to represent one feature and they are sorted tacking into account their importance and indicated in the bar chart plot, which helped significantly in the quantification of the features impact on the final model's response. For USGS 14206950 location (Figure 7), Q variable is the important input feature in LightGBM, CatBoost, and XGBoost models, whereas

Table 3. Results of modelling TU at USGS 14206950.

	Training				Validation			
Models	*R*	NSE	RMSE (FNU)	MAE (FNU)	*R*	NSE	RMSE (FNU)	MAE (FNU)
CatBoost1	0.964	0.929	3.268	1.608	0.727	0.465	8.345	3.869
CatBoost2	0.972	0.943	2.921	1.848	0.802	0.610	7.126	3.422
CatBoost3	0.976	0.951	2.711	1.717	0.769	0.539	7.746	3.424
CatBoost4	0.996	0.992	1.092	0.788	0.732	0.457	8.412	3.651
CatBoost5	0.993	0.985	1.482	0.966	0.945	0.887	3.837	1.903
CatBoost6	0.993	0.987	1.417	0.923	0.943	0.886	3.858	1.981
CatBoost7	0.994	0.989	1.300	0.864	0.944	0.890	3.778	1.905
CatBoost8	0.994	0.988	1.355	0.898	0.944	0.889	3.795	1.976
LightGBM1	0.862	0.741	6.245	2.978	0.810	0.634	6.907	3.428
LightGBM2	0.876	0.766	5.937	2.835	0.805	0.620	7.032	3.495
LightGBM3	0.886	0.783	5.721	2.716	0.815	0.634	6.902	3.428
LightGBM4	0.893	0.794	5.567	2.656	0.817	0.640	6.842	3.350
LightGBM5	0.950	0.902	3.850	1.829	0.932	0.862	4.243	2.244
LightGBM6	0.953	0.907	3.740	1.759	0.935	0.869	4.136	2.191
LightGBM7	0.953	0.908	3.729	1.785	0.931	0.860	4.264	2.228
LightGBM8	0.956	0.912	3.638	1.734	0.934	0.866	4.174	2.205
XGBoost1	0.960	0.920	3.465	1.730	0.728	0.469	8.317	3.914
XGBoost2	0.992	0.984	1.550	1.052	0.733	0.450	8.463	3.984
XGBoost3	0.995	0.990	1.243	0.856	0.747	0.475	8.272	3.874
XGBoost4	0.997	0.993	1.040	0.718	0.780	0.559	7.574	3.579
XGBoost5	0.999	0.997	0.628	0.442	0.941	0.882	3.921	2.011
XGBoost6	0.999	0.999	0.419	0.305	0.951	0.902	3.564	1.828
XGBoost7	0.999	0.998	0.478	0.343	0.948	0.897	3.668	1.922
XGBoost8	0.999	0.999	0.255	0.191	0.951	0.903	3.552	1.816
CNN1	0.818	0.667	7.080	3.377	0.826	0.658	6.677	3.472
CNN2	0.828	0.679	6.953	3.230	0.805	0.610	7.129	3.371
CNN3	0.848	0.713	6.572	2.860	0.783	0.589	7.315	3.255
CNN4	0.870	0.757	6.047	2.778	0.764	0.501	8.059	3.371
CNN5	0.956	0.908	3.727	1.970	0.933	0.860	4.276	2.385
CNN6	0.970	0.936	3.096	1.855	0.928	0.846	4.475	2.550
CNN7	0.965	0.925	3.352	1.707	0.927	0.823	4.795	2.336
CNN8	0.975	0.950	2.732	1.415	0.930	0.857	4.322	2.225

Table 4. Results of modelling TU at USGS 14211720.

	Training				Validation			
Models	*R*	NSE	RMSE (FNU)	MAE (FNU)	*R*	NSE	RMSE (FNU)	MAE (FNU)
CatBoost1	0.970	0.960	2.475	1.478	0.880	0.732	4.591	2.231
CatBoost2	0.996	0.994	0.941	0.664	0.897	0.772	4.229	2.100
CatBoost3	0.988	0.984	1.563	1.012	0.923	0.845	3.489	1.791
CatBoost4	0.990	0.987	1.415	0.935	0.935	0.870	3.193	1.648
CatBoost5	0.993	0.990	1.231	0.770	0.951	0.898	2.833	1.387
CatBoost6	0.994	0.993	1.073	0.690	0.955	0.906	2.714	1.346
CatBoost7	0.999	0.999	0.370	0.270	0.945	0.869	3.202	1.481
CatBoost8	0.996	0.995	0.853	0.575	0.961	0.920	2.502	1.219
LightGBM1	0.928	0.907	3.775	1.902	0.903	0.798	3.979	2.101
LightGBM2	0.947	0.931	3.262	1.697	0.915	0.830	3.651	1.972
LightGBM3	0.955	0.941	3.012	1.594	0.925	0.848	3.453	1.893
LightGBM4	0.958	0.945	2.897	1.529	0.933	0.865	3.253	1.805
LightGBM5	0.961	0.949	2.796	1.398	0.934	0.861	3.305	1.685
LightGBM6	0.964	0.953	2.700	1.354	0.933	0.858	3.335	1.679
LightGBM7	0.968	0.957	2.570	1.322	0.936	0.866	3.245	1.662
LightGBM8	0.969	0.960	2.492	1.289	0.941	0.876	3.115	1.598
XGBoost1	0.984	0.979	1.799	1.064	0.860	0.683	4.993	2.335
XGBoost2	0.997	0.995	0.848	0.602	0.909	0.806	3.899	1.938
XGBoost3	0.998	0.997	0.711	0.506	0.921	0.828	3.677	1.817
XGBoost4	0.998	0.998	0.591	0.424	0.939	0.868	3.221	1.639
XGBoost5	0.999	0.999	0.435	0.311	0.936	0.855	3.375	1.494
XGBoost6	0.999	0.999	0.348	0.251	0.936	0.851	3.423	1.410
XGBoost7	0.999	0.999	0.421	0.297	0.942	0.863	3.276	1.337
XGBoost8	0.999	0.999	0.287	0.209	0.956	0.900	2.796	1.223
CNN1	0.896	0.868	4.503	2.265	0.915	0.822	3.741	2.136
CNN2	0.922	0.835	5.038	2.738	0.927	0.708	4.785	2.720
CNN3	0.948	0.920	3.502	1.808	0.935	0.869	3.213	1.769
CNN4	0.959	0.946	2.869	1.546	0.933	0.867	3.231	1.732
CNN5	0.956	0.935	3.154	1.671	0.929	0.810	3.862	2.002
CNN6	0.981	0.974	1.986	1.066	0.948	0.871	3.181	1.486
CNN7	0.965	0.954	2.669	1.506	0.947	0.877	3.110	1.651
CNN8	0.989	0.980	1.745	1.035	0.957	0.880	3.074	1.378

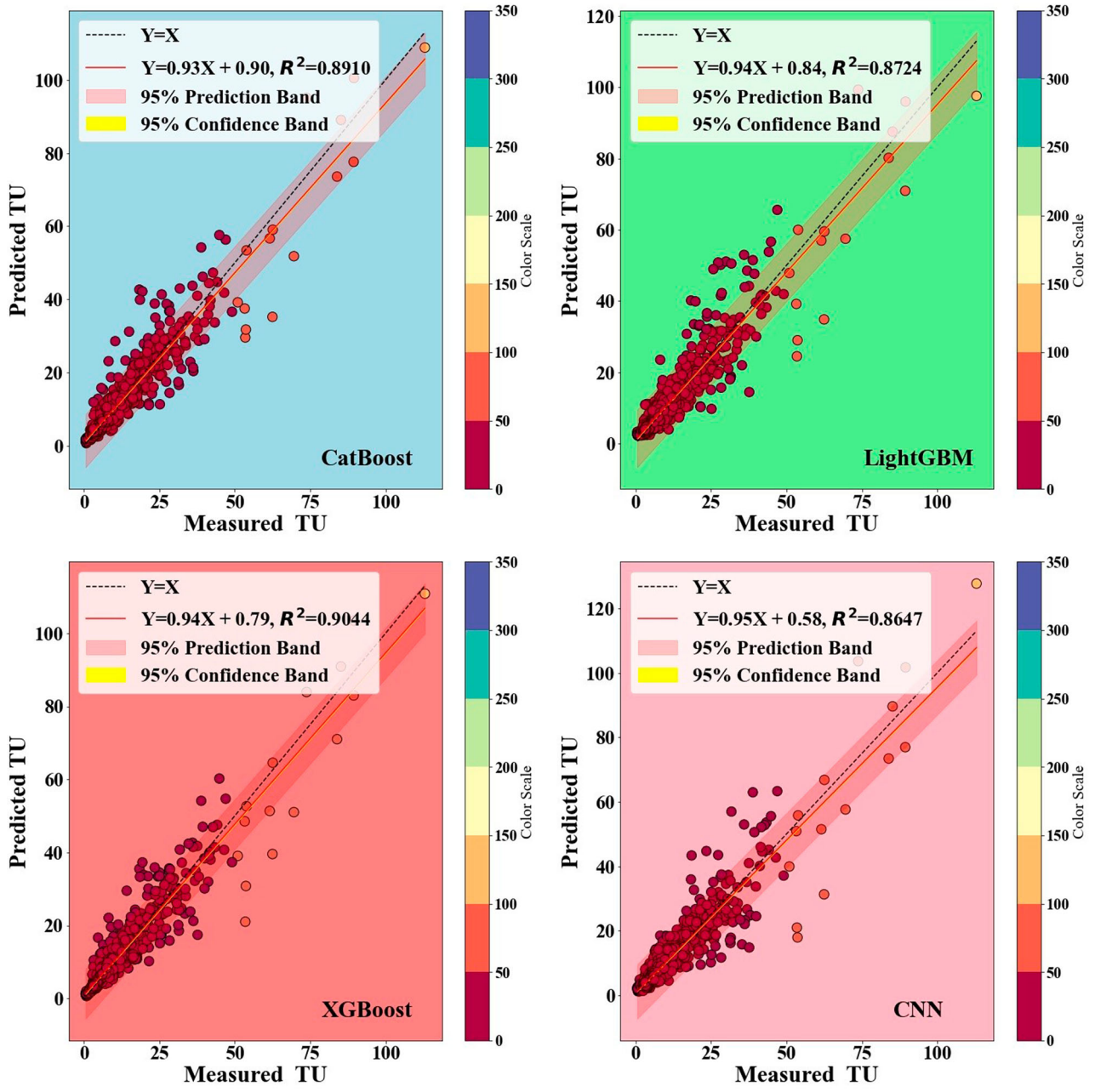

Figure 2. Scatter plots of the observed against the estimated TUs at USGS 14206950.

YY and then SC were the influential inputs in addressing the outputs of CNN model. Additionally, as it is apparent in Figure 8, river discharge (Q) is the most effective input variable affecting the outputs of all the four models at USGS 14211720 site.

As mentioned, in addition to the SHAP global interpretability, local interpretability was also performed. To that end, local interpretability is provided only for the best model among the four proposed techniques in validation period: i.e. XGBoost for USGS 14206950 and CatBoost for the USGS 14211720 in full-input scenario. In this context, randomly 6 samples as reported in Figures 9 and 10 were selected to address the SHAP local interpretability. In the left side of this Figures, variables indicated by red arrows demonstrate the input features with a positive impact on the models' outputs, whereas variables with a negative impact are shown in blue arrows. Investigating the SHAP local interpretability at both the stations clearly represents that different variables considered as the inputs to the models could have various impacts on the model's output results, whether positive or negative, as well as less or more effects. For example, at USGS 14211720 site, the inputs Q, DO, TW, MM, PH, YY in the sample number 100, and inputs Q, DO, MM

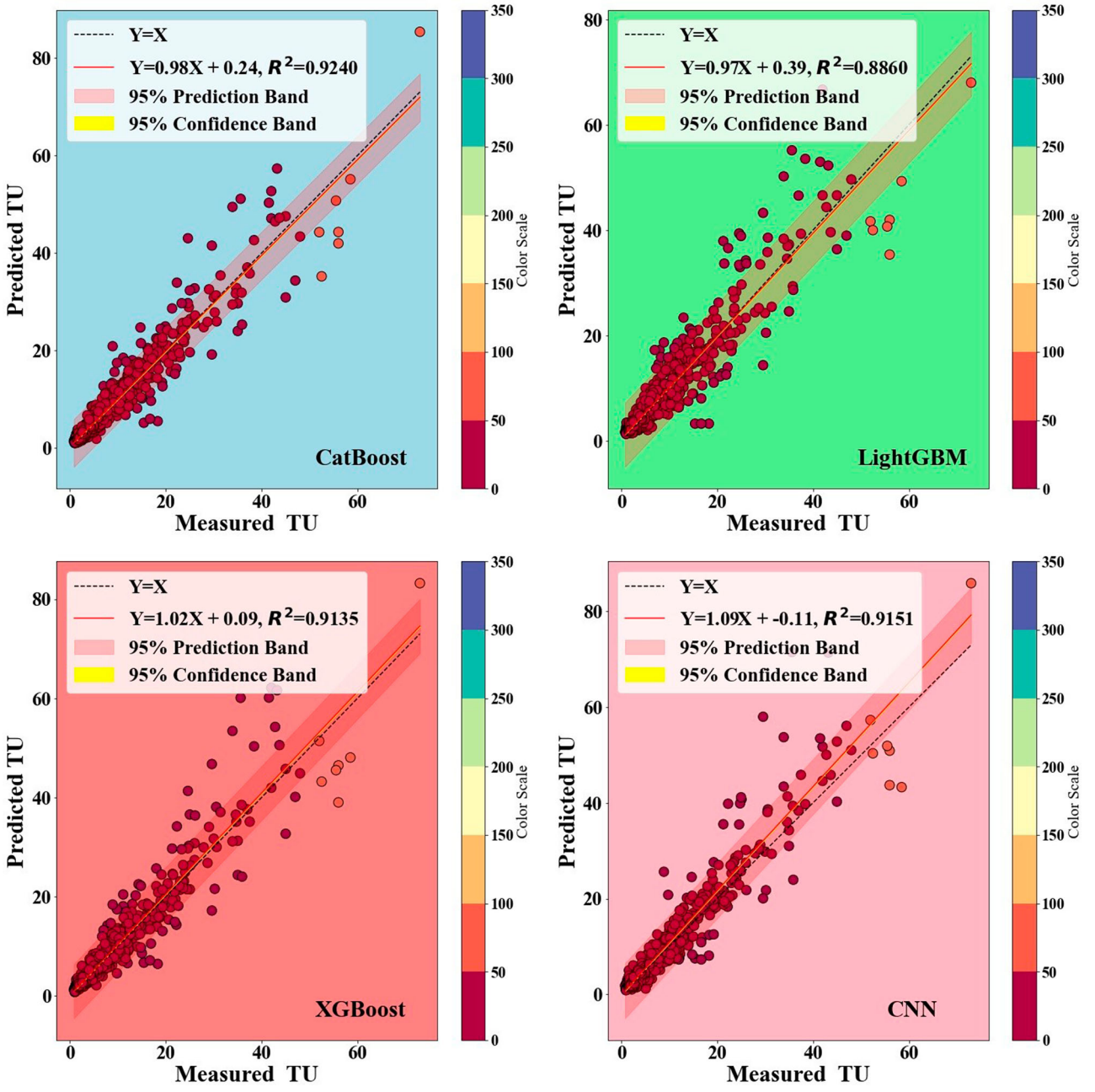

Figure 3. Scatter plots of the observed against the estimated TUs at USGS 14211720.

in sample number 300 were the influential inputs leading to an increased value of the target variable (i.e. TU). Conversely, the remaining variables in the mentioned samples have a decreasing effect on the output values of the models.

A performance comparison of whole the developed models under the different input combinations with error metrics listed in Tables 3 and 4 clearly illustrates that XGBoost8 (R = 0.951, NSE = 0.903, RMSE = 3.552 FNU, MAE = 1.816 FNU) at USGS 14206950 and CatBoost8 (R = 0.961, NSE = 0.920, RMSE = 2.502 FNU, MAE = 1.219 FNU) at USGS 14211720 were the superior models for river water TU estimation during the validation phase. The mentioned models were fed with full-input data including five river water data (Q, T_w, SC, pH, DO) and three periodicity terms (DD, MM, YY). The performance of best-performing models of the current study could be compared with the accuracy of developed models in previous works. In this regard, it has been focused on the values of RMSE metric as this is one of the most widely used criteria in evaluating the performance of intelligent frameworks in preceding

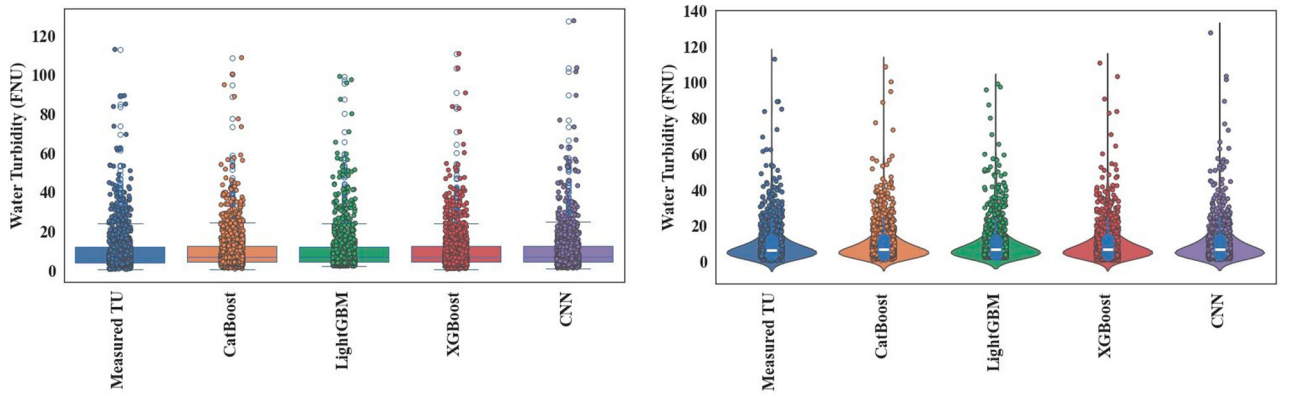

Figure 4. Box plots and violin plots of the observed against the estimated TUs at USGS 14206950.

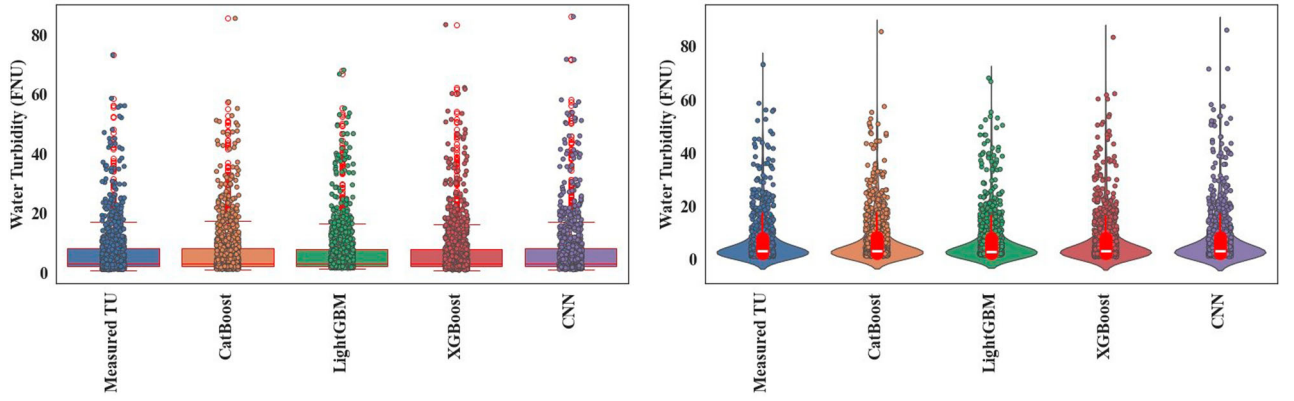

Figure 5. Box plots and violin plots of the observed against the estimated TUs at USGS 14211720.

studies when estimating the river water TU. Zounemat-Kermani et al. (2021) reported the least RMSE of 4.431 FNU for OS-ELM model fed with two inputs including river discharge and suspend sediment. Heddam (2023) forecasted the daily water TU time series at four river stations via developing four machine learning models. Their findings revealed that the RMSE values in the validation period varied from 1.731 FNU in Bat-ELM model with inputs of river discharge and periodicity to 6.657 FNU in DENFIS method fed with only river discharge. Santos et al. (2025) predicted the daily river water TUs at 12 river stations utilizing sential-2 remote sensing

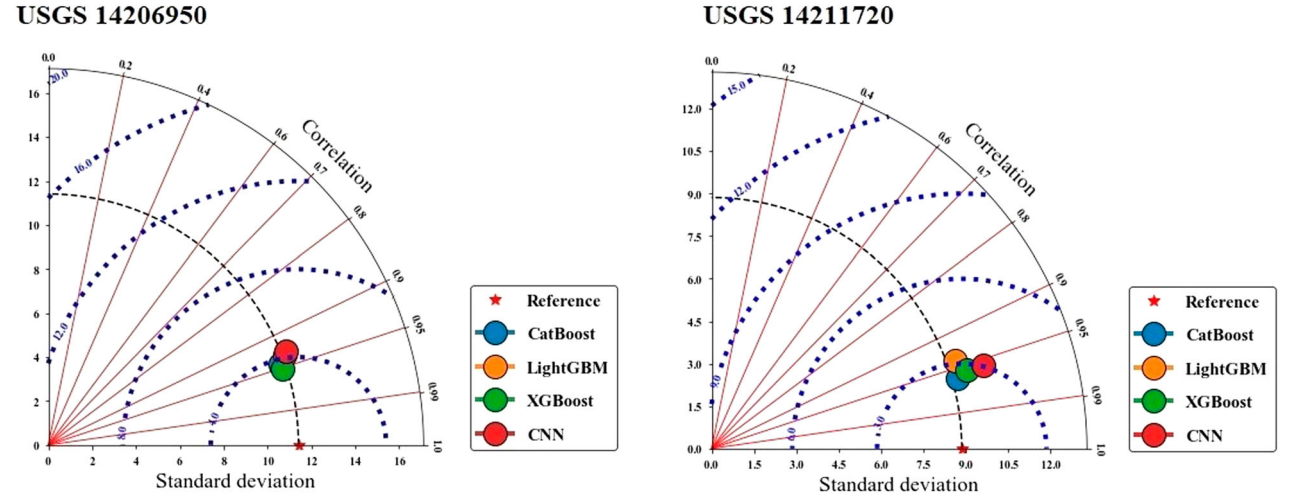

Figure 6. Taylor diagrams of the observed against the estimated TUs at the studied sites.

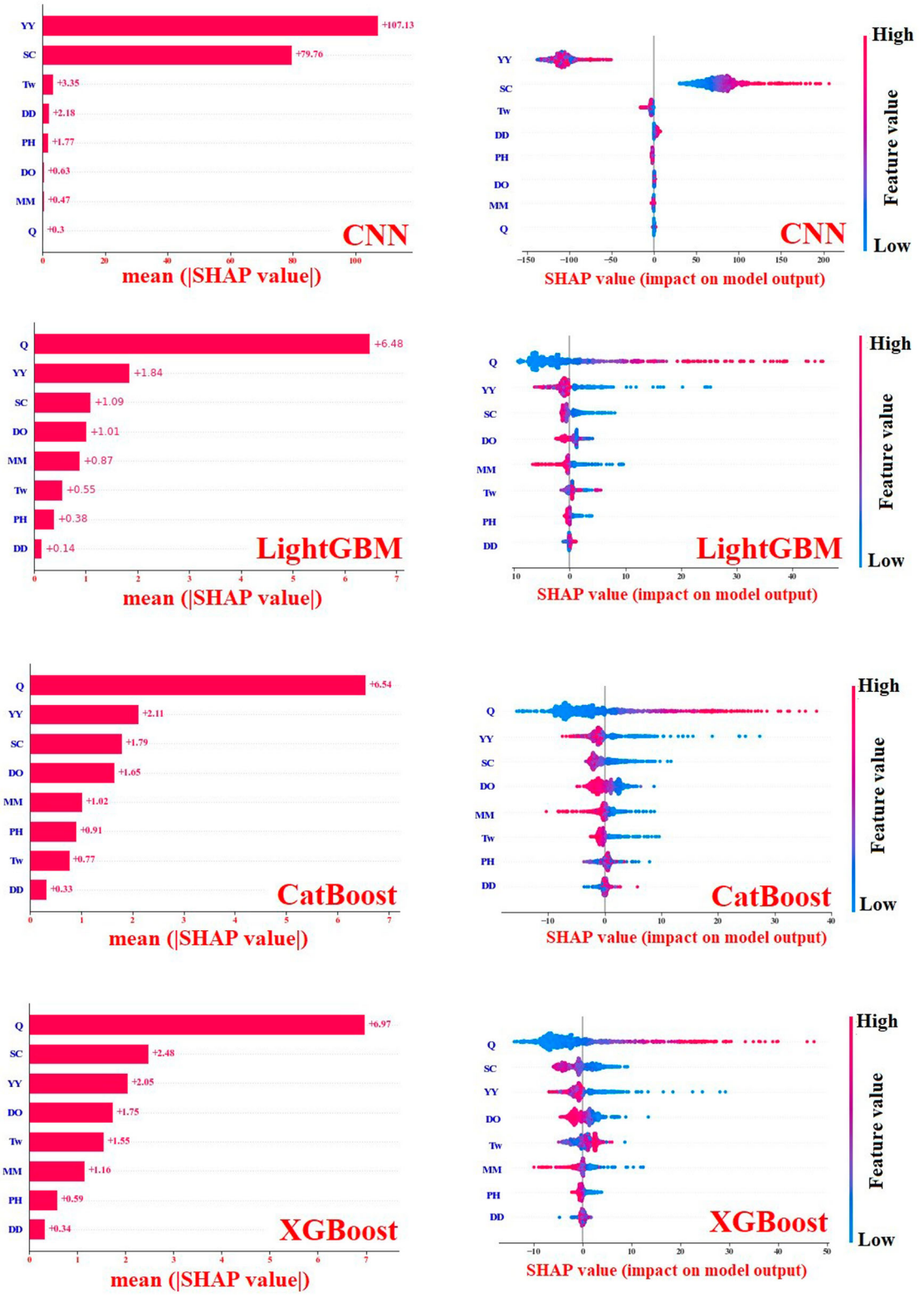

Figure 7. SHAP global interpretability plots at USGS 14206950: bar chart left panel and summary plot right panel.

data through developing six predictive models. On average, XGBoost and RF were found to be superior models with the RMSE of 27.53 and 35.25 FNU, respectively. Lu and Ma (2020) implemented some types of simple and hybrid machine learning models for river water TU prediction, and stated that the best model of complete ensemble empirical mode decomposition with adaptive noise-XGBoost (i.e. CEEMDAN-XGBoost) achieved the

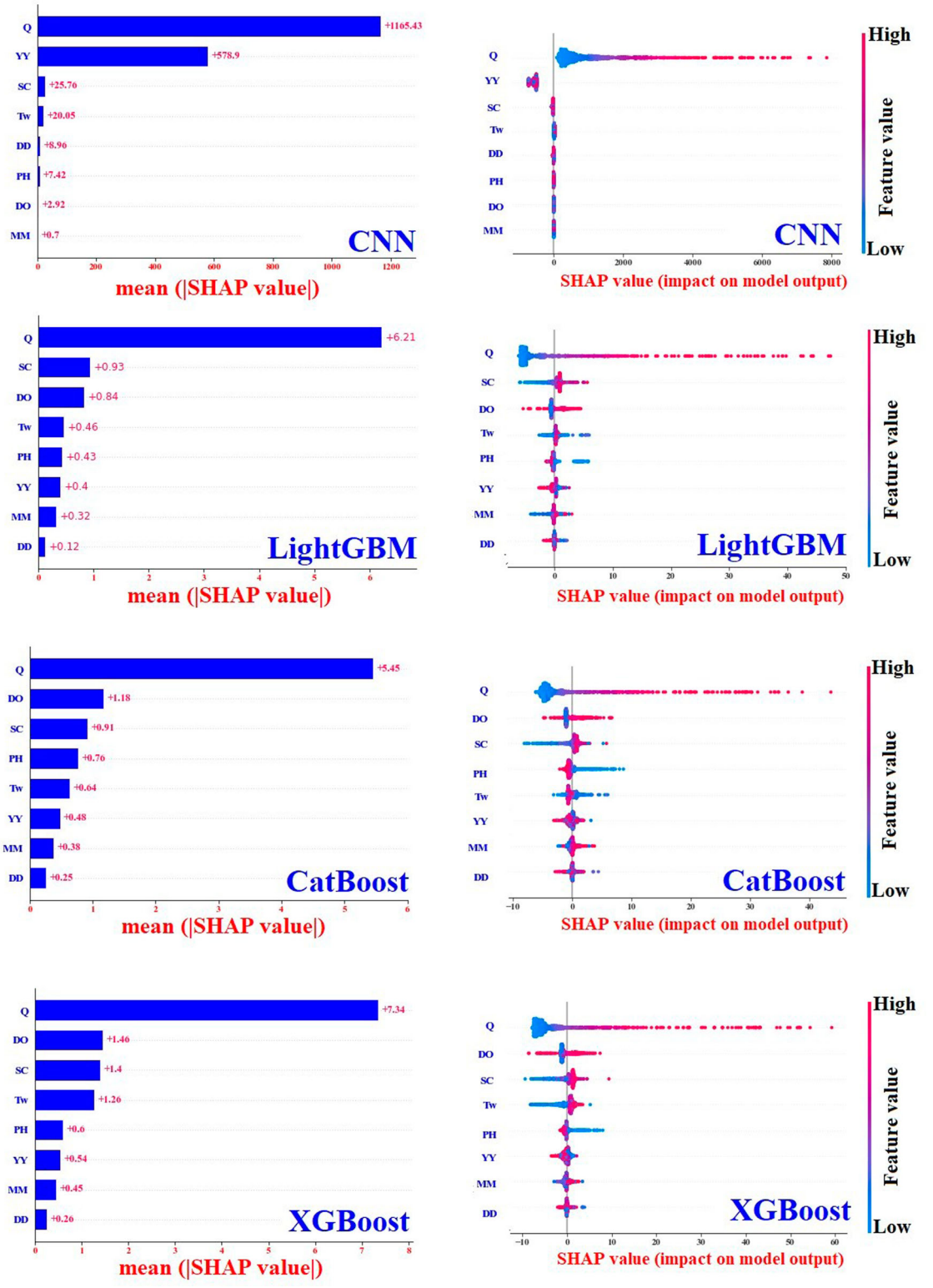

Figure 8. SHAP global interpretability plots at USGS 14211720: bar chart left panel and summary plot right panel.

lowest RMSE equal to 0.16 FNU. As a result, it can be concluded that the performance of models can depend on several factors including the values of observational river water TU data, the type of developed models, and the input estimators used to establish the models.

The difference between the models in terms of numerical performances can be interpreted in several ways. Even if we assume that the models have worked at very acceptable level; there are still multiple sources of differences. First, there could be even considerable differences

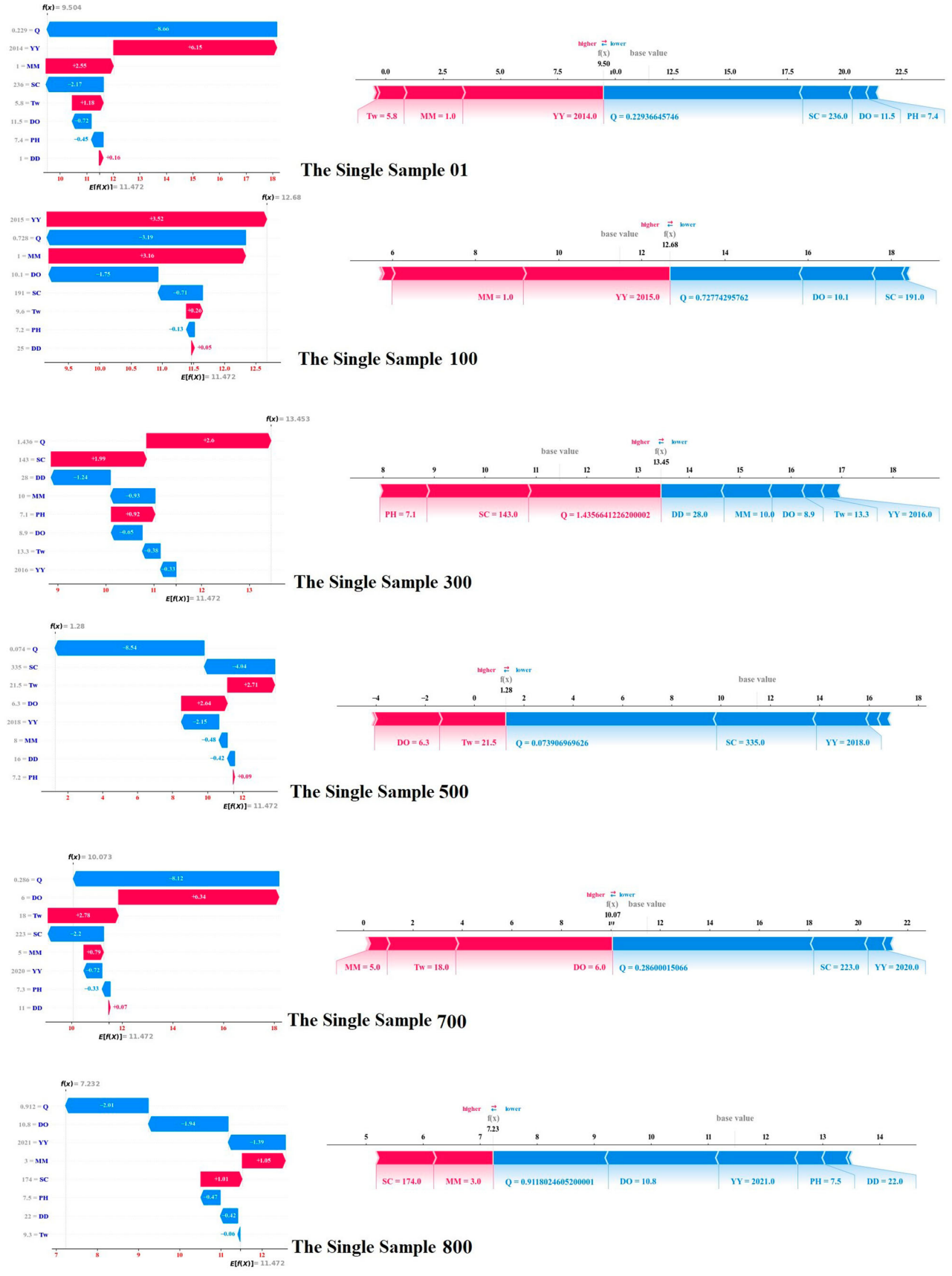

Figure 9. SHAP local interpretability plots at USGS 14206950: Waterfall plots left panel and SHAP force plot right panel.

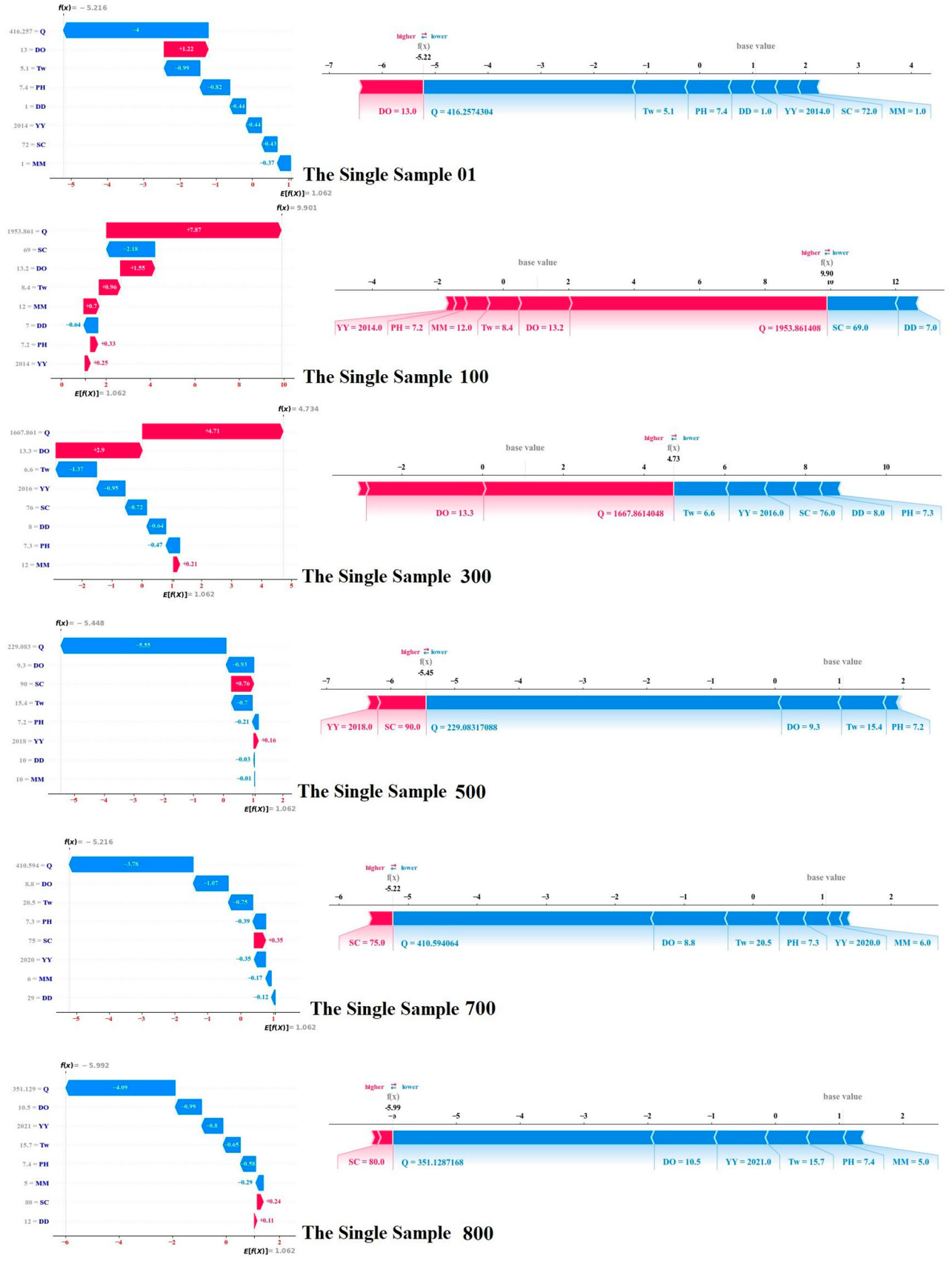

Figure 10. SHAP local interpretability plots at USGS 14211720: Waterfall plots left panel and SHAP force plot right panel.

in the physical nature of the parameters into these individual models, which leads to a difference in the training performance and the generalization ability. Each model has its own parameters and its own training algorithms, and one of these models, i.e. the CNN, possesses an extremely high number of parameters. Therefore, calibration with these different models will introduce different results in the final estimation of water TU as well. Another critical source of difference in TU estimation is the variability of data from one station to another. Furthermore, we can argue, the results obtained from the combination of two or three features change based on the used algorithm, and, one feature can have a high contribution in a particular model, whereas, its contribution, can be deteriorated when exposed to another model, which has been confirmed by the SHAP analysis.

4. Concluding remarks

The mean daily time series of water TU as one of the river water quality variables were estimated at two river stations including USGS 14206950 and USGS 14211720 located in the United states. To achieve this objective, four intelligent frameworks consisting of CatBoost, LightGBM, XGBoost, and CNN were established as the TU estimation tools. River water data such as Q, Tw, SC, PH, and DO were used as the possible input estimators when the development of the models used. Besides, two different strategies i.e. with and without considering the periodicity terms (DD, MM, YY) were utilized in order to evaluate their impact on the model's outputs. The outcomes illustrated that the accuracy of models could be generally increased via increasing the number of input variables, particularly at USGS 14211720 station. Moreover, the performance analysis of the models indicated that their accuracies were improved with the inclusion of the periodicity term in the model's output, which this conclusion was more obvious at USGS 14206950 site. A performance assessment of all the models developed in this study exhibited that CatBoost and XGBoost models generally performed better than the other two methods at both the stations. At the end of this research, SHAP interpretability was accomplished in two global and local modes. In SHAP global interpretability, river discharge (Q) variable was found to be the most important input affecting the output results of the selected best-performing models. Additionally, the analysis of SHAP local interpretability indicated that different inputs considered could provide various impacts on the models' outputs.

In this study, three types of boosting methods including CatBoost, LightGBM, and XGBoost, as well as a deep learning-based model namely CNN were utilized in estimating river water TUs. It is recommended that diverse versions of machine learning and deep-learning models be developed in the future for river water TU estimation. Furthermore, the current research focused on the development of standalone models, whereas future works could propose hybrid variants of the ML and DL frameworks. Finally, SHAP interpretability has recently opened a new window to further investigate the role of inputs in the output results of models. Therefore, it is suggested that it should be considered more in estimating hydrological variables like river water quality variables.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Ethical approval

This study does not involve any human participants and thus does not require ethical approval.

Data availability statement

The datasets used in this study are available from the corresponding author upon a reasonable request. The required data were compiled from the United States Geological Survey website (https://waterdata.usgs.gov/nwis/).

ORCID

Amin Gharehbaghi http://orcid.org/0000-0002-2898-3681
Salim Heddam http://orcid.org/0000-0002-8055-8463
Saeid Mehdizadeh http://orcid.org/0000-0002-3078-3689
Sungwon Kim http://orcid.org/0000-0002-9371-8884

References

Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, A. B. M., & Gandomi, A. H. (2023). Deep learning modelling techniques: Current progress, applications, advantages, and challenges. *Artificial Intelligence Review*, *56*(11), 13521–13617. https://doi.org/10.1007/s10462-023-10466-8

Ahn, J. M., Kim, J., & Kim, K. (2023). Ensemble machine learning of gradient boosting (XGBoost, LightGBM, CatBoost) and attention-based CNN-LSTM for harmful algal blooms forecasting. *Toxins*, *15*(10), 608. https://doi.org/10.3390/toxins15100608

Aldrees, A., Khan, M., Taha, A. T. B., & Ali, M. (2024). Evaluation of water quality indexes with novel machine learning and SHapley Additive ExPlanation (SHAP) approaches. *Journal of Water Process Engineering*, *58*, 104789. https://doi.org/10.1016/j.jwpe.2024.104789

Alizamir, M., Moradveisi, K., Ahmed, K. O., Bahrami, J., Kim, S., & Heddam, S. (2025). An efficient data fusion model based on Bayesian model averaging for robust water quality prediction using deep learning strategies. *Expert Systems*

with Applications, *261*, 125499. https://doi.org/10.1016/j.eswa.2024.125499

Allawi, M. F., Al-Ani, Y., Jalal, A. D., Ismael, Z. M., Sherif, M., & El-Shafie, A. (2024). Groundwater quality parameters prediction based on data-driven models. *Engineering Applications of Computational Fluid Mechanics*, *18*(1), 2364749. https://doi.org/10.1080/19942060.2024.2364749

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., & Chatila, R. (2020). Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Baek, S. S., Pyo, J., & Chun, J. A. (2020). Prediction of water level and water quality using a CNN-LSTM combined deep learning approach. *Water*, *12*(12), 3399. https://doi.org/10.3390/w12123399

Baptista, M. L., Goebel, K., & Henriques, E. M. (2022). Relation between prognostics predictor evaluation metrics and local interpretability SHAP values. *Artificial Intelligence*, *306*, 103667. https://doi.org/10.1016/j.artint.2022.103667

Barzegar, R., Aalami, M. T., & Adamowski, J. (2020). Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stochastic Environmental Research and Risk Assessment*, *34*(2), 415–433. https://doi.org/10.1007/s00477-020-01776-2

Bian, L., Qin, X., Zhang, C., Guo, P., & Wu, H. (2023). Application, interpretability and prediction of machine learning method combined with LSTM and LightGBM-a case study for runoff simulation in an arid area. *Journal of Hydrology*, *625*, 130091. https://doi.org/10.1016/j.jhydrol.2023.130091

Chen, B., Chen, Y., & Chen, H. (2024). An interpretable CatBoost model guided by spectral morphological features for the inversion of coastal water quality parameters. *Water*, *16*(24), 3615. https://doi.org/10.3390/w16243615

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*, 785–794.

Dahiya, N., Gupta, S., & Singh, S. (2022). A review paper on machine learning applications, advantages, and techniques. *ECS Transactions*, *107*(1), 6137–6150. https://doi.org/10.1149/10701.6137ecst

Ding, W., Zhao, J., Qin, B., Wu, T., Zhu, S., Li, Y., Xu, S., Ruan, S., & Wang, Y. (2021). Exploring and quantifying the relationship between instantaneous wind speed and turbidity in a large shallow lake: Case study of Lake Taihu in China. *Environmental Science and Pollution Research*, *28*(13), 16616–16632. https://doi.org/10.1007/s11356-020-11544-y

Ehteram, M., Ahmed, A. N., Sherif, M., & El-Shafie, A. (2024). An advanced deep learning model for predicting water quality index. *Ecological Indicators*, *160*, 111806. https://doi.org/10.1016/j.ecolind.2024.111806

Elavarasan, D., & Vincent, P. D. (2020). Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. *IEEE Access*, *8*, 86886–86901. https://doi.org/10.1109/ACCESS.2020.2992480

Gan, M., Lai, X., Guo, Y., Chen, Y., Pan, S., & Zhang, Y. (2024). Floodplain lake water level prediction with strong river-lake interaction using the ensemble learning LightGBM. *Water Resources Management*, *38*(13), 5305–5321. https://doi.org/10.1007/s11269-024-03915-8

Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, *115*, 105151. https://doi.org/10.1016/j.engappai.2022.105151

Gelda, R. K., Effler, S. W., Peng, F., Owens, E. M., & Pierson, D. C. (2009). Turbidity model for ashokan reservoir, New York: Case study. *Journal of Environmental Engineering*, *135*(9), 885–895. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000048

Gheisari, M., Shafi, J., Kosari, S., Amanabadi, S., Mehdizadeh, S., Fernandez Campusano, C., & Barzan Abdalla, H. (2025). Development of improved deep learning models for multi-step ahead forecasting of daily river water temperature. *Engineering Applications of Computational Fluid Mechanics*, *19*(1), 2450477. https://doi.org/10.1080/19942060.2025.2450477

Gu, K., Liu, J., Shi, S., Xie, S., Shi, T., & Qiao, J. (2022). Self-organizing multichannel deep learning system for river turbidity monitoring. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–13.

Hancock, J. T., & Khoshgoftaar, T. M. (2020). Catboost for big data: An interdisciplinary review. *Journal of Big Data*, *7*(1), 94. https://doi.org/10.1186/s40537-020-00369-8

Haq, K. R. A., & Harigovindan, V. P. (2022). Water quality prediction for smart aquaculture using hybrid deep learning models. *IEEE Access*, *10*, 60078–60098. https://doi.org/10.1109/ACCESS.2022.3180482

Hasani, S. S., Arias, M. E., Nguyen, H. Q., Tarabih, O. M., Welch, Z., & Zhang, Q. (2024). Leveraging explainable machine learning for enhanced management of lake water quality. *Journal of Environmental Management*, *370*, 122890. https://doi.org/10.1016/j.jenvman.2024.122890

Heddam, S. (2023). Bat algorithm optimized extreme learning machine: A new modeling strategy for predicting river water turbidity at the United States. In S. Eslamian & F. Eslamian (Eds.), *Handbook of hydroinformatics* (pp. 39–55). Elsevier.

Heydari, S., Nikoo, M. R., Mohammadi, A., & Barzegar, R. (2024). Two-stage meta-ensembling machine learning model for enhanced water quality forecasting. *Journal of Hydrology*, *641*, 131767. https://doi.org/10.1016/j.jhydrol.2024.131767

Hu, Y., Liu, C., & Wollheim, W. M. (2024). Prediction of riverine daily minimum dissolved oxygen concentrations using hybrid deep learning and routine hydrometeorological data. *Science of The Total Environment*, *918*, 170383. https://doi.org/10.1016/j.scitotenv.2024.170383

Hu, R., Xu, W., Yan, W., Wu, T., He, X., & Cheng, N. (2023). Comparison between machine-learning-based turbidity models developed for different lake zones in a large shallow lake. *Water*, *15*(3), 387. https://doi.org/10.3390/w15030387

Ismail, W. N., Hassan, M. M., Alsalamah, H. A., & Fortino, G. (2020). CNN-based health model for regular health factors analysis in internet-of-medical things environment. *IEEE Access*, *8*, 52541–52549. https://doi.org/10.1109/ACCESS.2020.2980938

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30*, 3149–3157.

Khoi, D. N., Quan, N. T., Linh, D. Q., Nhi, P. T. T., & Thuy, N. T. D. (2022). Using machine learning models for predicting the

water quality index in the La Buong River, Vietnam. *Water*, *14*(10), 1552. https://doi.org/10.3390/w14101552

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, *160*(1), 3–24.

Krishnan, S., & Manikandan, R. (2024). Water quality prediction: A data-driven approach exploiting advanced machine learning algorithms with data augmentation. *Journal of Water and Climate Change*, *15*(2), 431–452. https://doi.org/10.2166/wcc.2023.403

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. *Advances in Neural Information Processing Systems*, *2*, 396–404.

Lee, Y. H., Won, J. H., Kim, S., Auh, Q. S., & Noh, Y. K. (2022). Advantages of deep learning with convolutional neural network in detecting disc displacement of the temporomandibular joint in magnetic resonance imaging. *Scientific Reports*, *12*(1), 11352. https://doi.org/10.1038/s41598-022-15231-5

Li, S., Qasem, S. N., Band, S. S., Ameri, R., Pai, H. T., & Mehdizadeh, S. (2024). Explainable machine learning models for estimating daily dissolved oxygen concentration of the Tualatin River. *Engineering Applications of Computational Fluid Mechanics*, *18*(1), 2304094. https://doi.org/10.1080/19942060.2024.2304094

Li, L., Qiao, J., Yu, G., Wang, L., Li, H. Y., Liao, C., & Zhu, Z. (2022). Interpretable tree-based ensemble model for predicting beach water quality. *Water Research*, *211*, 118078. https://doi.org/10.1016/j.watres.2022.118078

Li, H., Yang, Z., Lv, H., Wang, M., Zhang, B., Yan, H., & Zhang, H. (2024). Research on mine water source classifications based on BO-CatBoost. *Environmental Monitoring and Assessment*, *196*(10), 876. https://doi.org/10.1007/s10661-024-13040-z

Lin, X., Wu, M., Shao, X., Li, G., & Hong, Y. (2023). Water turbidity dynamics using random forest in the Yangtze River Delta Region, China. *Science of the Total Environment*, *903*, 166511. https://doi.org/10.1016/j.scitotenv.2023.166511

Liu, W., Liu, S., Hassan, S. G., Cao, Y., Xu, L., Feng, D., Cao, L., Chen, W., Chen, Y., Guo, J., & Liu, T. (2023). A novel hybrid model to predict dissolved oxygen for efficient water quality in intensive aquaculture. *IEEE Access*, *11*, 29162–29174. https://doi.org/10.1109/ACCESS.2023.3260089

Lu, H., & Ma, X. (2020). Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere*, *249*, 126169. https://doi.org/10.1016/j.chemosphere.2020.126169

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874.

Majnooni, S., Fooladi, M., Nikoo, M. R., Al-Rawas, G., Haghighi, A. T., Nazari, R., Al-Wardy, M., & Gandomi, A. H. (2024). Smarter water quality monitoring in reservoirs using interpretable deep learning models and feature importance analysis. *Journal of Water Process Engineering*, *60*, 105187. https://doi.org/10.1016/j.jwpe.2024.105187

Makumbura, R. K., Mampitiya, L., Rathnayake, N., Meddage, D. P. P., Henna, S., Dang, T. L., Hoshino, Y., & Rathnayake, U. (2024). Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence (XAI) techniques like shapley additive explanations (SHAP) for interpreting the black-box nature. *Results in Engineering*, *23*, 102831. https://doi.org/10.1016/j.rineng.2024.102831

Mei, P., Li, M., Zhang, Q., & Li, G. (2022). Prediction model of drinking water source quality with potential industrial-agricultural pollution based on CNN-GRU-attention. *Journal of Hydrology*, *610*, 127934. https://doi.org/10.1016/j.jhydrol.2022.127934

Meng, Y., Yang, N., Qian, Z., & Zhang, G. (2020). What makes an online review more helpful: An interpretation framework using XGBoost and SHAP values. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(3), 466–490. https://doi.org/10.3390/jtaer16030029

Mousavi, S. (2024). Conjugation of deep learning and de noising data methods for short-term water turbidity forecasting. *Journal of Hydro-Environment Research*, *52*, 26–37. https://doi.org/10.1016/j.jher.2023.12.002

Mukhamediev, R. I., Popova, Y., Kuchin, Y., Zaitseva, E., Kalimoldayev, A., Symagulov, A., Levashenko, V., Abdoldina, F., Gopejenko, V., Yakunin, K., & Muhamedijeva, E. (2022). Review of artificial intelligence and machine learning technologies: Classification, restrictions, opportunities and challenges. *Mathematics*, *10*(15), 2552. https://doi.org/10.3390/math10152552

Mukhamediev, R. I., Symagulov, A., Kuchin, Y., Yakunin, K., & Yelis, M. (2021). From classical machine learning to deep neural networks: A simplified scientometric review. *Applied Sciences*, *11*(12), 5541. https://doi.org/10.3390/app11125541

Nallakaruppan, M. K., Gangadevi, E., Shri, M. L., Balusamy, B., Bhattacharya, S., & Selvarajan, S. (2024). Reliable water quality prediction and parametric analysis using explainable AI models. *Scientific Reports*, *14*(1), 7520. https://doi.org/10.1038/s41598-024-56775-y

Park, J., & Lee, H. (2020). Prediction of high turbidity in rivers using LSTM algorithm. *Journal of Korean Society of Water and Wastewater*, *34*(1), 35–43. https://doi.org/10.11001/jksww.2020.34.1.035

Park, J., Lee, W. H., Kim, K. T., Park, C. Y., Lee, S., & Heo, T. Y. (2022). Interpretation of ensemble learning to predict water quality using explainable artificial intelligence. *Science of the Total Environment*, *832*, 155070. https://doi.org/10.1016/j.scitotenv.2022.155070

Park, J. C., Um, M. J., Song, Y. I., Hwang, H. D., Kim, M. M., & Park, D. (2017). Modeling of turbidity variation in two reservoirs connected by a water transfer tunnel in South Korea. *Sustainability*, *9*(6), 993. https://doi.org/10.3390/su9060993

Philippus, D., Sytsma, A., Rust, A., & Hogue, T. S. (2024). A machine learning model for estimating the temperature of small rivers using satellite-based spatial data. *Remote Sensing of Environment*, *311*, 114271. https://doi.org/10.1016/j.rse.2024.114271

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). Catboost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, *31*, 6639–6649.

Rahaman, M. H., Sajjad, H., Hussain, S., Masroor, M., & Sharma, A. (2024). Surface water quality prediction in the lower thoubal river watershed, India: A hyper-tuned

machine learning approach and DNN-based sensitivity analysis. *Journal of Environmental Chemical Engineering*, *12*(3), 112915. https://doi.org/10.1016/j.jece.2024.112915

Rajaee, T., & Jafari, H. (2018). Utilization of WGEP and WDT models by wavelet denoising to predict water quality parameters in rivers. *Journal of Hydrologic Engineering*, *23*(12), 04018054. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001700

Saccotelli, L., Verri, G., De Lorenzis, A., Cherubini, C., Cacciopoli, R., Coppini, G., & Maglietta, R. (2024). Enhancing estuary salinity prediction: A machine learning and deep learning based approach. *Applied Computing and Geosciences*, *23*, 100173.

Santos, V., Rocha, P., & Gharabaghi, B. (2025). Evaluation of deep learning methods for forecasting turbidity in river networks using sentinel-2 remote sensing data. https://doi.org/10.2139/ssrn.5117066.

Shehadeh, A., Alshboul, O., Al Mamlook, R. E., & Hamedat, O. (2021). Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression. *Automation in Construction*, *129*, 103827. https://doi.org/10.1016/j.autcon.2021.103827

Sun, J., Di Nunno, F., Sojka, M., Ptak, M., Luo, Y., Xu, R., Zhu, S., & Granata, F. (2024). Prediction of daily river water temperatures using an optimized model based on NARX networks. *Ecological Indicators*, *161*, 111978. https://doi.org/10.1016/j.ecolind.2024.111978

Tan, W., Zhang, J., Wu, J., Lan, H., Liu, X., Xiao, K., Wang, L., Lin, H., Sun, G., & Guo, P. (2022). Application of CNN and long short-term memory network in water quality predicting. *Intelligent Automation and Soft Computing*, *34*(3), 1943–1958. https://doi.org/10.32604/iasc.2022.029660

Teixeira, L. C., Mariani, P. P., Pedrollo, O. C., dos Reis Castro, N. M., & Sari, V. (2020). Artificial neural network and fuzzy inference system models for forecasting suspended sediment and turbidity in basins at different scales. *Water Resources Management*, *34*(11), 3709–3723. https://doi.org/10.1007/s11269-020-02647-9

Tizhoosh, H. R., & Pantanowitz, L. (2018). Artificial intelligence and digital pathology: Challenges and opportunities. *Journal of Pathology Informatics*, *9*(1), 38. https://doi.org/10.4103/jpi.jpi_53_18

Uddin, M. G., Nash, S., Rahman, A., & Olbert, A. I. (2023). Performance analysis of the water quality index model for predicting water state using machine learning techniques. *Process Safety and Environmental Protection*, *169*, 808–828. https://doi.org/10.1016/j.psep.2022.11.073

Usuga Cadavid, J. P., Lamouri, S., Grabot, B., Pellerin, R., & Fortin, A. (2020). Machine learning applied in production planning and control: A state-of-the-art in the era of industry 4.0. *Journal of Intelligent Manufacturing*, *31*(6), 1531–1558. https://doi.org/10.1007/s10845-019-01531-7

Van den Broeck, G., Lykov, A., Schleich, M., & Suciu, D. (2022). On the tractability of SHAP explanations. *Journal of Artificial Intelligence Research*, *74*, 851–886. https://doi.org/10.1613/jair.1.13283

Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning*, *109*(2), 373–440. https://doi.org/10.1007/s10994-019-05855-6

Wang, Q., Li, Z., Cai, J., Zhang, M., Liu, Z., Xu, Y., & Li, R. (2023). Spatially adaptive machine learning models for predicting water quality in Hong Kong. *Journal of Hydrology*, *622*, 129649. https://doi.org/10.1016/j.jhydrol.2023.129649

Wang, S., Peng, H., & Liang, S. (2022). Prediction of estuarine water quality using interpretable machine learning approach. *Journal of Hydrology*, *605*, 127320. https://doi.org/10.1016/j.jhydrol.2021.127320

Wang, F., Wang, Y., Zhang, K., Hu, M., Weng, Q., & Zhang, H. (2021). Spatial heterogeneity modeling of water quality based on random forest regression and model interpretation. *Environmental Research*, *202*, 111660. https://doi.org/10.1016/j.envres.2021.111660

Wu, Y., Sun, L., Sun, X., & Wang, B. (2022). A hybrid XGBoost-ISSA-LSTM model for accurate short-term and long-term dissolved oxygen prediction in ponds. *Environmental Science and Pollution Research*, *29*(12), 18142–18159. https://doi.org/10.1007/s11356-021-17020-5

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights Into Imaging*, *9*(4), 611–629. https://doi.org/10.1007/s13244-018-0639-9

Yang, Y., Lv, H., & Chen, N. (2023). A survey on ensemble learning under the era of deep learning. *Artificial Intelligence Review*, *56*(6), 5545–5589. https://doi.org/10.1007/s10462-022-10283-5

Yang, Y., Xiong, Q., Wu, C., Zou, Q., Yu, Y., Yi, H., & Gao, M. (2021). A study on water quality prediction by a hybrid CNN-LSTM model with attention mechanism. *Environmental Science and Pollution Research*, *28*(39), 55129–55139. https://doi.org/10.1007/s11356-021-14687-8

Yao, Z., Wang, Z., Huang, J., Xu, N., Cui, X., & Wu, T. (2024). Interpretable prediction, classification and regulation of water quality: A case study of poyang lake, China. *Science of the Total Environment*, *951*, 175407. https://doi.org/10.1016/j.scitotenv.2024.175407

Zhang, Y., Li, C., Duan, H., Yan, K., Wang, J., & Wang, W. (2023). Deep learning-based data-driven model for detecting time-delay water quality indicators of wastewater treatment plant influent. *Chemical Engineering Journal*, *467*, 143483. https://doi.org/10.1016/j.cej.2023.143483

Zhang, M., Zhang, Z., Wang, X., Liao, Z., & Wang, L. (2024). The use of attention-enhanced CNN-LSTM models for multi-indicator and time-series predictions of surface water quality. *Water Resources Management*, *38*(15), 6103–6119. https://doi.org/10.1007/s11269-024-03946-1

Zheng, H., Liu, Y., Wan, W., Zhao, J., & Xie, G. (2023). Large-scale prediction of stream water quality using an interpretable deep learning approach. *Journal of Environmental Management*, *331*, 117309. https://doi.org/10.1016/j.jenvman.2023.117309

Zhou, S., Song, C., Zhang, J., Chang, W., Hou, W., & Yang, L. (2022). A hybrid prediction framework for water quality with integrated W-ARIMA-GRU and LightGBM methods. *Water*, *14*(9), 1322. https://doi.org/10.3390/w14091322

Zounemat-Kermani, M., Alizamir, M., Fadaee, M., Sankaran Namboothiri, A., & Shiri, J. (2021). Online sequential extreme learning machine in river water quality (turbidity) prediction: A comparative study on different data mining approaches. *Water and Environment Journal*, *35*(1), 335–348. https://doi.org/10.1111/wej.12630