Open Access

Advances in Environmental and Engineering Research

LIDSEN Publishing Inc.

Research Article

Comparative Analysis of Machine Learning Models and Explainable Artificial Intelligence for Predicting Wastewater Treatment Plant Variables

Fuad Bin Nasir [†], Jin Li [†, *]

Department of Civil and Environmental Engineering, University of Wisconsin-Milwaukee, WI 53211, USA; E-Mails: fnasir@uwm.edu; li@uwm.edu

† These authors contributed equally to this work.

* **Correspondence:** Jin Li; E-Mail: li@uwm.edu

Academic Editor: José L. Segovia-Juárez

Special Issue: Artificial Intelligence in Environmental Research

Adv Environ Eng Res
2024, volume 5, issue 4
doi:10.21926/aeer.2404020

Received: July 30, 2024
Accepted: October 13, 2024
Published: October 17, 2024

Abstract

Increasing urban wastewater and rigorous discharge regulations pose significant challenges for wastewater treatment plants (WWTP) to meet regulatory compliance while minimizing operational costs. This study explores the application of several machine learning (ML) models specifically, Artificial Neural Networks (ANN), Gradient Boosting Machines (GBM), Random Forests (RF), eXtreme Gradient Boosting (XGBoost), and hybrid RF-GBM models in predicting important WWTP variables such as Biochemical Oxygen Demand (BOD), Total Suspended Solids (TSS), Ammonia (NH_3), and Phosphorus (P). Several feature selection (FS) methods were employed to identify the most influential WWTP variables. To enhance ML models' interpretability and to understand the impact of variables on prediction, two widely used explainable artificial intelligence (XAI) methods-Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) were investigated in the study. Results derived from FS and XAI methods were compared to explore their reliability. The ML model performance results revealed that ANN, GBM, XGBoost, and RF-GBM have great potential for variable prediction with low error rates and strong correlation coefficients such as R^2 value of 1 on the training set and 0.98 on the test set. The study also revealed that XAI methods identify common influential variables in each model's prediction. This is a novel attempt to get an overview of both LIME and SHAP explanations on ML models for a WWTP variable prediction.

Keywords

Machine learning; wastewater; explainable artificial intelligence; local interpretable model-agnostic explanations; Shapley additive explanations

1. Introduction

Wastewater treatment plants (WWTPs) play an essential role in safeguarding the aquatic environment by processing municipal and industrial sewage. Increasing amount of urban wastewater and demands for clean water present substantial challenges to WWTP operators in meeting regulatory effluent standards and reducing operating costs [1-4]. Moreover, the complexity of the treatment process demands a high level of precision to achieve the desired standard limits of various variables. To enhance effluent quality and comply with regulatory standards at WWTP while minimizing operation and maintenance cost, the implementation of advanced technologies is crucial. There is a potential for WWTPs to improve decision-making process and to optimize resource allocation by utilizing machine learning (ML), a subfield of artificial intelligence (AI) that can ultimately assist in achieving sustainable treatment system. The application of ML in predicting WWTP variables has been effective [4-18]. ML models were also used to regulate WWTP operation that resulted in notable amount of energy savings [19]. According to studies [20, 21], ML can process substantial datasets with impressive precision.

As WWTPs are complex and comprise several concurrent nonlinear mechanisms, researchers investigated a wide range of variables, such as water quality, water quantity, and meteorological data, in predicting WWTP variables using various ML models [14]. Biochemical Oxygen Demand (BOD) and Total Suspended Solids (TSS) are among the most influential variables in a WWTP. They were commonly investigated together because they share many similarities, including their hardness to measure, lack of information that may be obtained, the potential for complex model nonlinearity, and importance in prediction models [10]. Other common pollutants in wastewater are ammonia (NH_3) and phosphorus (P), both need to be reduced to the required level before being released into the environment [22]. A thorough understanding of influent and effluent nutrient characteristics is essential for the optimization of treatment operations [23, 24]. Therefore, accurate influent and effluent variable (BOD, NH_3, P, and TSS) prediction through ML can facilitate efficient adjustment of operational parameters such as aeration rates or chemical dosages to effectively meet effluent quality standards.

ML-based approaches are specifically being employed for the monitoring and design of complex non-linear issues at WWTPs [25]. Traditional methods of variable measurements in WWTP involve time-consuming laboratory analysis. Advancements in sensor technologies and online monitoring systems have introduced real-time and alternative approaches. The difficulty of measuring BOD online and the length of time required for laboratory measurements highlight the significance of developing predictive models that can eliminate the requirement for measurements performed by humans. ML methods can rely on the connection created between the input and output datasets by extracting correlations between variables from historical data. Previous studies on various ML models to predict WWTP variables have a large variability in results, with R^2 for BOD ranging from 0.48 to 0.99, TSS ranging from 0.63 to 0.98, (NH_3) ranging from 0.32-0.84, and P ranging from 0.28-0.93 [2, 6, 13, 16, 26-34]. Moreover, relying solely on ML models without an understanding of the contexts of the predictions is not ideal. Recent trend towards the practice of ML models in variable prediction requires explainability in addition to prediction accuracy. This is especially important in WWTPs where operators need to understand the reasons behind model predictions to increase their confidence in real-world application. Questions on rationale behind ML predictions, the basis for trust in these predictions, and methods for error correction are some of the concerns especially relevant in WWTP, where the reliability of ML practices is critical. While many studies have focused on predicting variables in WWTP using ML, research on implementing explainable artificial intelligence (XAI) is still developing. Some recent studies integrated XAI to interpret ML output [14, 35, 36]. However, investigation of XAI methods with various ML models is lacking. Therefore, it is a novel attempt to investigate multiple XAI approaches to enhance the interpretability of ML models applications in WWTP.

This study applied XAI methods to improve the interpretability of ML models in predicting influential variables of a WWTP. Various feature selections and XAI methods were employed to identify the importance of input variables in ML models performance. We collected a broad range of WWTP variables, encompassing water quality, water quantity, and electrical data. Several standalone ML models i.e. artificial neural network (ANN), gradient boosting machine (GBM), random forest (RF), eXtreme gradient boosting (XGBoost), and hybrid model RF-GBM performance were tested and compared with historical datasets in predicting influent and effluent BOD, NH_3, P, and TSS. This study provides a better understanding of ML model performance in predicting WWTP variables with the help of XAI, which aids in making informed decisions to optimize treatment plant performance.

2. Materials and Methods

2.1 Data Collection

The data were collected from a WWTP in Milwaukee, Wisconsin, USA that treats wastewater from industrial, municipal, and domestic sources. Water quality, water quantity, and electrical data (daily and hourly) were collected from 1st January 2019 to 31st December 2023. After data processing, the following variables were considered in the study: Influent BOD (BOD_i), Effluent BOD (BOD_e), Influent Flow ($Flow_i$), Effluent Flow ($Flow_e$), Influent Ammonia $(NH_3)_i$, Effluent Ammonia $(NH_3)_e$, Influent TSS (TSS_i), Effluent TSS (TSS_e), Influent Phosphorus (P_i), Effluent Phosphorus (P_e), TSS_e Removed, BOD_e Removed, Primary Sludge, Iron Dose, Detention Time, Aeration (Aer) Basin Temp, DO Set Pt, Sludge Volume Index (SVI), Mean Cell Residence Time (MCRT), Waste Activated Sludge (WAS), WAS Flow, pH_e, $Temp_e$, Total Residual Chlorine (TRC), Gravity Belt Thickening (GBT) Polymer Used, Fecal Coliforms, E.coli, Total Electricity (Elec) Generated, and Total Blower Elec Used. Time series of variables can be found in Figure 1. A list of abbreviations is shown in Table S1 of additional materials.

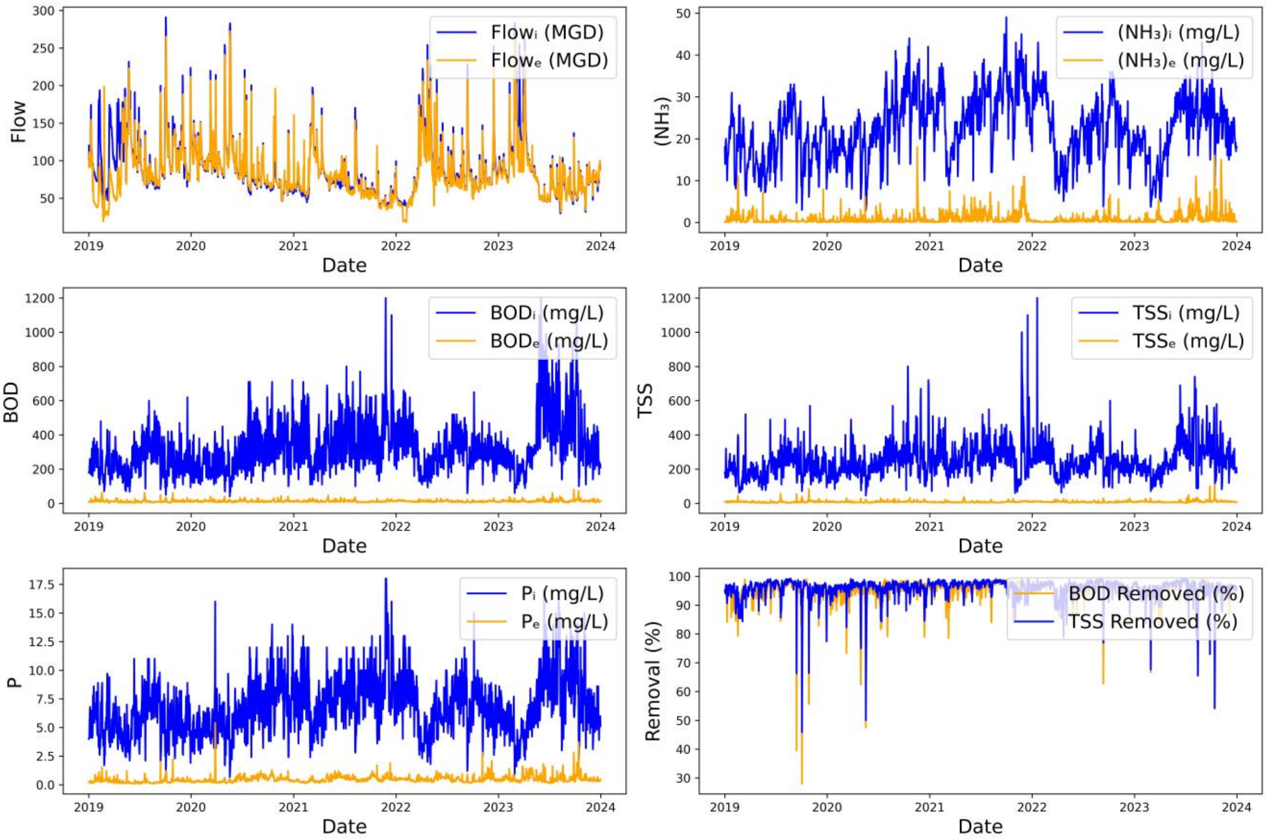

Figure 1 Time series of variables (top left: Flow; top right: (NH_3); middle left: BOD; middle right: TSS; bottom left: P; bottom right: BOD and TSS removed (%)).

2.2 Data Pre-Processing

Typically, sensor-collected data contain anomalies related to the recording process. During examination of the dataset for missing or inaccurate data, several anomalies were identified through human observation and subsequently replaced with average values. Additionally, any missing values in the dataset were filled in using the average value of the respective variable. We converted hourly variables to daily variables. The variables Flow$_i$ and Flow$_e$ exhibited a high correlation (0.9). To minimize multicollinearity, only Flow$_i$ was included in the study. Consequently, 28 out of the 29 collected variables consisting of 51128 data entries were considered for the analysis. Statistical properties of data are presented in Table 1. When eliminating redundant or irrelevant features that do not significantly affect the prediction, lowers noise, and enhances model performance [28], it is crucial to consider the context in which the model is used. Variables such as DO set points controlled by blowers may have a more indirect impact on the prediction accuracy, as they influence the performance of the overall treatment process rather than directly correlating with target variables. Moreover, we did not identify or remove outliers in the dataset to understand the whole picture of the analysis as suggested by other studies [37]. Therefore, in the study, we considered the full dataset of WWTP that includes the most common input variables found in relevant papers to run the ML models [38].

Table 1 Data set statistical properties.

Variables	Units	Min	Max	Mean	Std
$Flow_i$	(MGD)	29.58	290.95	89.18	39.32
$(NH_3)_i$	(mg/L)	2.90	49.00	22.36	7.73
$(NH_3)_e$	(mg/L)	0.02	18.00	0.95	1.56
BOD_i	(mg/L)	40.00	1200.00	331.86	155.69
BOD_e	(mg/L)	2.00	80.00	12.96	6.69
TSS_i	(mg/L)	46.00	1200.00	252.72	104.33
TSS_e	(mg/L)	1.90	110.00	9.23	6.21
P_i	(mg/L)	0.67	18.00	6.92	2.59
P_e	(mg/L)	0.09	5.40	0.47	0.32
TSS Removed	(%)	45.83	99.21	95.70	3.87
BOD Removed	(%)	28.00	99.14	95.16	4.36
Primary sludge	(TPD)	1.04	192.20	54.25	20.69
Iron Dose	(mg/L)	0.00	29.81	11.06	4.65
Detention Time	(min)	25.07	265.87	94.62	34.67
Aer Basin Temp	(F)	45.30	83.50	59.61	5.67
DO Set Pt	(mg/L)	3.00	5.00	3.63	0.36
SVI	(mL/g)	39.25	332.50	114.12	39.58
MCRT	(Days)	4.05	28.14	10.27	2.59
WAS	(TPD)	0.00	77.04	37.79	13.13
WAS Flow	(MGD)	0.00	2.88	1.61	0.45
pH_e	-	6.75	7.71	7.17	0.09
$Temp_e$	(F)	48.33	228.27	158.57	83.24
TRC	(mg/L)	0.00	0.04	0.01	0.01
GBT Polymer Used	(lbs/day)	0.00	86402.40	5107.26	5687.10
Fecal Coliforms	(CFU/100ml)	2.00	30000.00	306.74	1818.53
E coli.	(MPN/100ml)	1.00	24000.00	1017.97	8734.47
Total Elec Generated	(MW)	0.00	5.09	3.34	0.82
Total Blower Elec Used	(KW)	1371.38	3406.56	2859.10	331.63

2.3 Feature Selection

Several feature selection (FS) methods were employed to identify the most significant variables for predicting target variables, including analysis of variance (ANOVA), least absolute shrinkage and selection operator (LASSO), mutual information (MI), random forest (RF), and Pearson correlation (PC) [14, 22]. ANOVA F-values are non-negative and can theoretically range from 0 to infinity. LASSO scores can be negative or positive, whereas PC scores range from -1 to 1. The MI scores range from 0, indicating no shared information, to positive values. RF generates a feature importance score from 0 to 1, where 0 means the feature was not used in the prediction, and 1 means the feature perfectly predicted the output. Traditional FS methods were chosen to compare their derived results with XAI method outputs.

2.4 SHapley Additive exPlanations

SHapley Additive exPlanations (SHAP) analysis is a recently developed XAI method based on game theory that interprets the behavior of ML models [14, 38-40]. It explains the models' predictions by showcasing the relative influence of input variables on model performance [35]. Using Shapley values from game theory, each feature is attributed values, as described by [41-43] as follows:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|s|!\,(n-|s|-1)!}{n!}\left[f_x(sU\{i\}) - f_x(s)\right] \tag{1}$$

Where ϕ_i is the SHAP value of *i*th input feature, *n* is the number of all input features, *s* is the subset of feature subsets, $|s|$ is the feature subsets element number, $f_x(sU\{i\})$ is trained with that feature present, and $f_x(s)$ is trained with feature withheld.

SHAP values at higher positions signify a greater importance of input variables on the models' performance. A positive weight indicates that increasing the feature's value typically boosts the models' prediction, whereas a negative weight implies that increasing the feature's value tends to reduce the model's prediction. SHAP summary plots are being used in WWTP to interpret models' output [14]. In this study, we chose the commonly used function, SHAP summary plot, to investigate how the top features in a dataset impact the models' output.

2.5 Local Interpretable Model-Agnostic Explanation

Local Interpretable Model-Agnostic Explanation (LIME) is an XAI tool that interprets black-box ML models by using a local, interpretable model to clarify each prediction [36]. LIME is obtained by following equation:

$$\xi(x) = {}^{\operatorname{argmin}}_{g \in G}\mathcal{L}(f, g, \pi_x) + \Omega(g) \tag{2}$$

Where, $\mathcal{L}$ indicates fidelity function, *G* indicates explanation families, and Ω indicates complexity measure. The explanation model for instance *x* is the model *g*, π_x indicates proximity measure and *f* indicates original model.

LIME identifies the top features contributing most to the model's predictions, associating each feature with a weight that indicates its impact on the prediction. Features with positive weights have a positive effect on the prediction, while those with negative weights have a negative effect. The magnitude of the weight reflects the strength of the feature's influence on the prediction. Features are ranked by their importance, with the most influential ones listed first. A detailed explanation of LIME is provided by [44].

2.6 ML Models

To predict BOD_i, BOD_e, $(NH_3)_i$, $(NH_3)_e$, P_i, P_e, TSS_i, and TSS_e, several ML models, i.e., ANN, GBM, RF, RF-GBM, and XGBoost were applied. These models were chosen because of their widespread application in water quality variable prediction. ANN consists of layered networks of interconnected nodes, with multiple hidden layers that allow the identification of intricate relationships and patterns in the data [45, 46]. However, it requires substantial data and careful hyperparameter tuning. In this study, we explored different configurations of hidden layers, activation functions, and optimization strategies to train ANN model. A boosting approach called GBM combines several weak prediction models, typically decision trees, to produce a powerful predictive model [47]. To fix the errors created by the previous trees, GBM iteratively adds new models. In this research, we used different values for learning rate, number of trees, tree depth, min sample leaf, and minimum sample split to identify the best combination that optimize model performance on the training data. An ensemble learning technique called RF uses several decision trees to produce predictions [48-52]. In this study, we used various values for number of trees, tree depth, min sample leaf, and minimum sample split to identify the best combination that optimizes model performance on the training data. RF-GBM combines the principles of RF and GBM. This hybrid model combines the advantages of RF and GBM to improve prediction performance. A newly developed version of the gradient boosting decision tree algorithm called XGBoost has the potential to reduce overfitting and increase robustness [38]. In XGBoost, several hyperparameters were also tuned to find that optimal configuration. In all ML models', the GridSearchCV is employed to identify the best combination of hyperparameters by testing multiple combinations using cross-validation.

2.7 Model Training and Evaluation

The dataset was divided into training and testing sets to ensure that the models were trained on a representative subset and evaluated on unseen data, providing a reliable measure of their generalization capability [53]. Two commonly recommended splits of training and test set ratios (90:10 and 80:20) were used as suggested by other relevant studies [14, 37, 52]. The testing set acts as an independent dataset to assess the performance of the models, while the training set was utilized to train multiple ML models. Validation is a crucial step of the model development process to ensure that the developed model is accurate enough for the intended use [54-56]. For validation purposes, splitting the data guarantees that the models are trained on a representative subset of the data and evaluated on unseen data, giving a trustworthy assessment of their generalization ability [53]. A 5-fold cross-validation was implemented to confirm the model's accuracy for its intended application by dividing the dataset into five equal parts [57]. In each iteration, a different fold was used as the test set, while the remaining folds constituted the training set. The model's performance is dependent on the hyperparameters used during training. To identify the best hyperparameter configuration, a grid search method was employed to find the optimal set that delivered the best performance.

To evaluate the regression model's performance, several model metrics can be used depending on the specific tasks, data characteristics, and circumstances [58-60]. In this regression study, three widely used assessment metrics-R-squared (R^2), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE)-were used to evaluate the performances of the ML models. MAE measures the average magnitude of the errors between predicted and actual values (eq 3). R^2 quantifies the percentage of variance that is explained by the models (eq 4), whereas RMSE denotes the average size of the residuals (eq 5). These metrics reveal information about the produced ML models' precision, goodness-of-fit, and accuracy. Higher values of R^2 and lower values of the error measures indicate better prediction performance and accuracy [61].

$$MAE = \frac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{n} \quad (3)$$

$$R^2 = 1 - \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad (5)$$

3. Results

3.1 ML Model Performance

The performance of ML models, including ANN, GBM, RF, XGBoost, and a hybrid RF-GBM, was evaluated using 90:10 and 80:20 train-test splits. The comparison between training and test performance helps to evaluate the models' generalization ability. An exceptionally high training performance relative to the test performance could be a sign of overfitting. Table 2 shows the model performance metrics for BOD prediction. The performance metrics for all target variables for 90:10 and 80:20 train-test splits are shown in Table S2 of additional materials.

Table 2 Model performance metrices for 90:10 and 80:20 train-test splits.

			(90:10)			(80:20)		
Target Variable	Model	Set	MAE	R^2	RMSE	MAE	R^2	RMSE
BOD_i	ANN	Training	7.47	0.99	11.25	5.70	1.00	8.61
		Test	13.58	0.93	39.10	12.21	0.97	25.73
	GBM	Training	4.64	1.00	6.16	4.40	1.00	5.84
		Test	15.99	0.97	23.28	15.46	0.98	23.77
	RF	Training	13.22	0.98	20.74	14.18	0.98	21.54
		Test	32.20	0.89	48.95	38.83	0.86	57.84
	XGBoost	Training	4.84	1.00	6.37	4.24	1.00	5.47
		Test	14.43	0.98	22.52	15.40	0.98	23.01
	RF-GBM	Training	6.10	1.00	7.69	2.87	1.00	3.71
		Test	14.21	0.98	19.91	15.88	0.98	24.22
BOD_e	ANN	Training	0.09	1.00	0.14	0.09	1.00	0.14
		Test	0.60	0.96	1.25	0.60	0.96	1.25
	GBM	Training	0.28	1.00	0.37	0.27	1.00	0.37
		Test	0.72	0.95	1.35	0.68	0.94	1.35
	RF	Training	0.45	0.99	0.77	0.48	0.99	0.79
		Test	1.27	0.86	2.36	1.18	0.86	2.07
	XGBoost	Training	0.33	1.00	0.45	0.29	1.00	0.39
		Test	0.74	0.95	1.36	0.69	0.94	1.36
	RF-GBM	Training	0.21	1.00	0.28	0.30	1.00	0.38
		Test	0.68	0.96	1.32	0.69	0.95	1.28

3.1.1 Train-Test Split (90:10)

ANN model showed good performance for BOD_i on the training set but higher errors on the test set. The GBM model exhibited excellent training and test performance. The RF model demonstrated good training performance but had higher test set errors (MAE of 35.51, R^2 of 0.88, RMSE of 51.07). XGBoost and RF-GBM showed good training and test performance. ANN model achieved nearly perfect results for BOD_e on the training set (MAE of 0.09, R^2 of 1.00, and RMSE of 0.14) and maintained strong performance on the test set (MAE of 0.60, R^2 of 0.96, and RMSE of 1.25). The GBM model had slight errors on the test set compared to ANN. The RF model showed higher errors on the test set compared to other models. XGBoost and RF-GBM maintained good performance on the test set.

ANN model had reasonable training performance for $(NH_3)_i$ but higher test set errors (MAE of 10.22, R^2 of 0.82, and RMSE of 3.20). The GBM model showed excellent training performance and moderate test set errors (MAE of 7.52, R^2 of 0.87, and RMSE of 2.74). The RF model had higher test set errors compared to GBM. XGBoost and RF-GBM had lower test set errors compared to other models. For $(NH_3)_e$, the ANN model had good training performance but poor test set performance (MAE of 0.53, R^2 of 0.33, and RMSE of 1.06). The GBM model showed better set results (MAE of 0.48, R^2 of 0.60, and RMSE of 0.82). The RF, XGBoost and RF-GBM showed similar results as GBM.

For P_i, the ANN model had good training results (MAE of 0.45, R^2 of 0.94, and RMSE of 0.62) and moderate test set errors (MAE of 0.64, R^2 of 0.84, and RMSE of 0.50). The GBM model maintained good performance on both sets (test MAE of 0.57, R^2 of 0.86, and RMSE of 0.89). RF, XGBoost, and RF-GBM showed moderate test set performance. For P_e, the ANN model had reasonable training performance, but poor test set performance (test MAE of 0.14, R^2 of 0.42, and RMSE of 0.18). The GBM model had better test set results (MAE of 0.10, R^2 of 0.65, and RMSE of 0.14). The RF model had moderate performance with slightly higher test set errors (MAE of 0.11, R^2 of 0.61, and RMSE of 0.15). XGBoost, and RF-GBM showed good performance on both sets.

For TSS_i, the ANN model showed good performance on both sets (test MAE of 5.75, R^2 of 0.99, and RMSE of 9.65). The GBM model had higher test set errors (MAE of 12.20, R^2 of 0.96, and RMSE of 20.43). The RF model showed significantly higher errors on the test set (MAE of 26.71, R^2 of 0.82, and RMSE of 41.55). For TSS_e, the ANN model showed excellent training performance and strong test set results (test MAE of 0.52, R^2 of 0.95, and RMSE of 0.91). The GBM model had better test set results (MAE of 0.38, R^2 of 0.97, and RMSE of 0.70). The RF model showed moderate performance with higher test set errors (MAE of 0.60, R^2 of 0.90, and RMSE of 1.27). XGBoost and RF-GBM maintained good performance for both TSS_i and TSS_e.

3.1.2 Train-Test Split (80:20)

For BOD_i, ANN model performed well on the training set with an MAE of 5.70, R^2 of 1.00, and RMSE of 8.61, but exhibited higher errors on the test set with an MAE of 12.21, R^2 of 0.97, and RMSE of 25.73. The GBM model also showed higher errors on the test set. The RF model demonstrated good training performance but significantly higher errors on the test set (MAE of 38.46, R^2 of 0.86, and RMSE of 56.95). The XGBoost and RF-GBM models, similar to GBM, had excellent training performance but higher errors on test set results. For BOD_e, ANN model achieved almost perfect results on the training set (MAE of 0.09, R^2 of 1.00, and RMSE of 0.14) and maintained strong performance on the test set (MAE of 0.60, R^2 of 0.96, and RMSE of 1.25). The GBM model exhibited slight errors on the test set (MAE of 0.68, R^2 of 0.94, and RMSE of 1.35). The RF model had higher errors on the test set compared to other models. The XGBoost and RF-GBM models maintained good performance on the test set.

For $(NH_3)_i$, ANN model had reasonable training performance but higher test set errors (MAE of 2.30, R^2 of 0.81, and RMSE of 3.16). The GBM model showed excellent training performance and moderate test set errors (MAE of 1.99, R^2 of 0.85, and RMSE of 2.82). The RF model had moderate performance with higher test set errors (MAE of 2.09, R^2 of 0.84, and RMSE of 2.91). The XGBoost and RF-GBM models, similar to GBM, had lower test set errors. For $(NH_3)_e$, ANN model had good training performance (MAE of 0.11, R^2 of 0.99, and RMSE of 0.16) but poor test set performance (MAE of 0.46, R^2 of 0.35, and RMSE of 0.89). The GBM model had better test set results (MAE of 0.47, R^2 of 0.45, and RMSE of 0.82). The RF, XGBoost, and RF-GBM models had good performance on the test set.

For P_i, ANN model had good training results (MAE of 0.45, R^2 of 0.95, and RMSE of 0.62) and moderate test set errors (MAE of 0.62, R^2 of 0.86, and RMSE of 0.89). The GBM model maintained good performance on both sets (test MAE of 0.57, R^2 of 0.87, and RMSE of 0.85). The RF, GBM, XGBoost, and RF-GBM models showed performance similar to GBM. For P_e, ANN model had reasonable training and test performance. The GBM model had better test set results (test MAE of 0.11, R^2 of 0.55, and RMSE of 0.17). The RF, XGBoost and RF-GBM models had good performance on both sets.

For TSS_i, ANN model had good performance on both sets. The GBM model had higher test set errors (MAE of 11.97, R^2 of 0.95, and RMSE of 20.17). The RF model had significantly higher errors on the test set (MAE of 23.06, R^2 of 0.85, and RMSE of 36.11). The XGBoost model maintained good performance (test MAE of 10.82, R^2 of 0.96, and RMSE of 18.31). The RF-GBM model showed balanced results with similar errors to GBM and XGBoost. For TSS_e, ANN model had excellent training performance and strong test set results (MAE of 0.56, R^2 of 0.94, and RMSE of 0.96). The GBM model had better test set results (MAE of 0.33, R^2 of 0.97, and RMSE of 0.71). The RF model had moderate performance with higher test set errors (MAE of 0.54, R^2 of 0.91, and RMSE of 1.24). The XGBoost and RF-GBM models maintained good performance on both sets (MAE of 0.42, R^2 of 0.96, and RMSE of 0.82).

3.2 Feature Selection Methods

Various FS methods were employed to identify the most significant variables impacting the concentrations of BOD_i, BOD_e, $(NH_3)_i$, $(NH_3)_e$, P_i, P_e, TSS_i, and TSS_e in WWTP. Table 3 shows common features shared by FS methods for various target variables. For BOD_i, P_i consistently emerges as the most influential variable across various methods. Other important variables include BOD Removed (%), TSS_i, and $(NH_3)_i$, which are consistently identified in multiple methods. For BOD_e, TSS_e is identified as the most significant variable across multiple methods, with BOD Removed (%) frequently highlighted as important. For $(NH_3)_e$, $Flow_i$ is most significant across methods and BOD_e is constantly significant in all methods. For P_i, BOD_i and TSS_i, are the topmost and secondmost across all the FS methods. For P_e, TSS_e and BOD_e are topmost and secondmost across all the methods. For TSS_i, P_i is most significant in multiple methods and BOD_i is consistently identified in all methods. For TSS_e, BOD and TSS Removed (%) are most significant in all methods. The consistency across different FS methods strengthens the reliability of these findings providing a robust basis for further research and practical applications. The top five strongly correlated variables related to target variables based on five FS methods are shown in Figure S1 of additional materials.

Table 3 Common features selected by FS methods.

No. of features	Name of features	Target variable
1	P_i	BOD_i
3	BOD Removed (%), P_e, $(NH_3)_e$	BOD_e
3	Detention Time, $Flow_i$, P_i	$(NH_3)_i$
2	BOD_e, P_e	$(NH_3)_e$
3	BOD_i, TSS_i, $(NH_3)_i$	P_i
2	BOD_e, TSS_e	P_e
2	BOD_i, P_i	TSS_i
3	BOD Removed (%), TSS Removed (%)	TSS_e

3.3 XAI

The results of the LIME and SHAP analyses for various target variables revealed the order of feature influence and their effects on ML models. Figure 2 shows one of the LIME plots. Figure 2(i) and Figure 2(ii) show the variables and their contributions (blue as negative, orange as positive) to BOD_i and BOD_e respectively for RF-GBM model for 50th instance. Predicted values of 50th instance for BOD_i and BOD_e are 305.95 mg/L and 15.29 mg/L respectively. According to the figure, BOD_e and TSS_e show strongest positive effect on BOD_i and BOD_e prediction respectively. Figure 3 shows one of the SHAP summary plots. Figure 3(i) and Figure 3(ii) show the variables and their contributions to BOD_i and BOD_e respectively for RF-GBM model. The figure shows that higher values of TSS_i (red dots) tend to contribute positively to the BOD_i prediction. In comparison, the lower values (blue dots) have negative contributions. While BOD_e and $(NH_3)_e$ have highest importance on BOD_i and BOD_e prediction respectively, SHAP summary plots also show it is indecisive for multiple variable regarding direction of prediction.

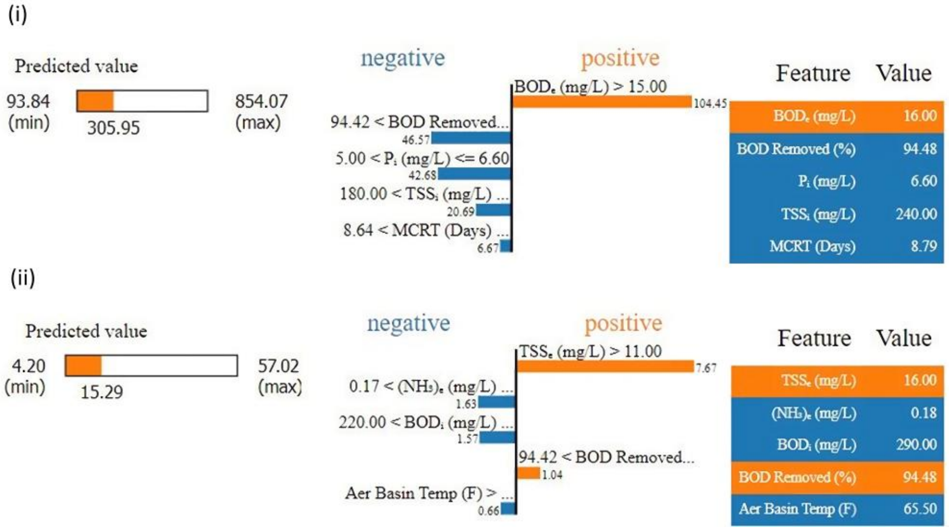

Figure 2 LIME explanation for RF-GBM model (i) BOD_i; LIME predicted value is 305.95 with a range between 93.84 and 854.07. BOD_e with a value of 16.00 mg/L significantly

contributes to the predicted BOD_i. BOD Removed % (94.48), P_i (6.60 mg/L), and TSS_i (240.00 mg/L) all play a role in reducing the predicted BOD_i value. (ii) BOD_e; LIME predicted 15.29 with a range between 4.20 and 57.02. TSS_e with a value of 16.00 mg/L is the most significant feature positively influencing the BOD_e prediction. $(NH_3)_e$, BOD_i, and Aer Basin Temp negatively affect the prediction.

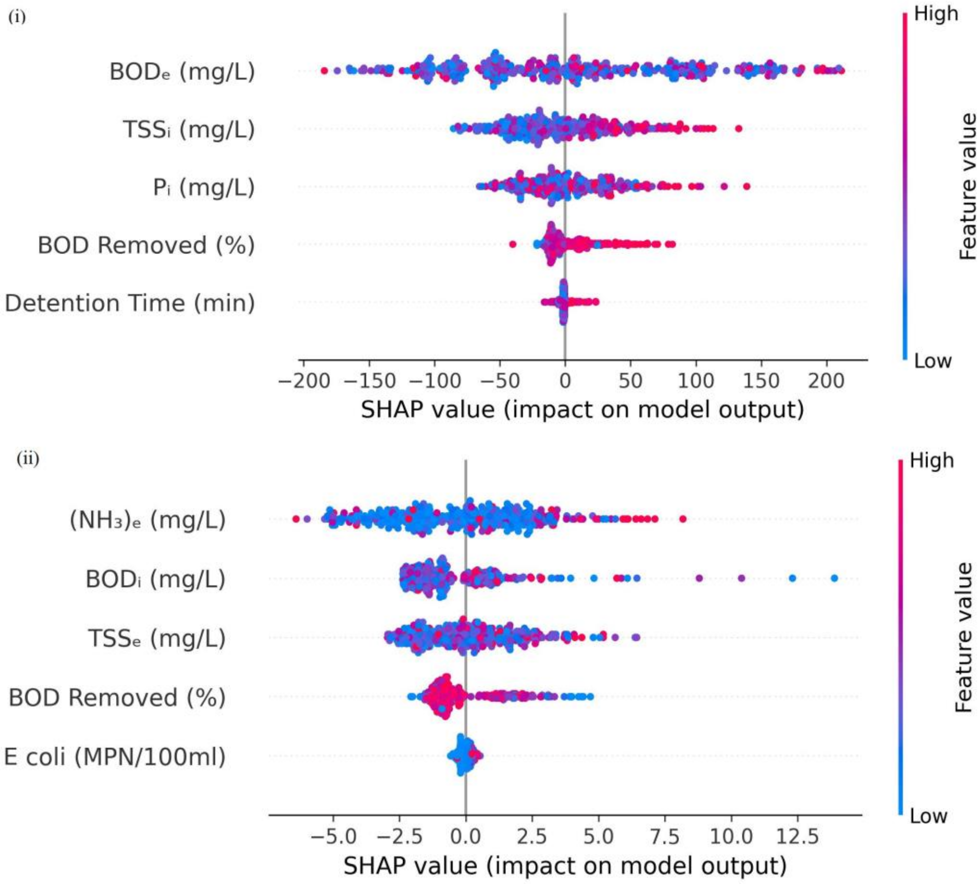

Figure 3 SHAP explanation for RF-GBM model. (i) BOD_i (ii) BOD_e.

Multiple variables were shared by both LIME and SHAP. Full variable list for LIME and SHAP in order of influence shown in Figure S1 of additional materials. Since LIME provided positive and negative impacts of variables explicitly, signs (positive or negative) were provided next to the variable's name. The SHAP summary plot was indecisive regarding positive or negative influence in many cases. Therefore, no signs were provided next to the SHAP identified variables. For predicting BOD_i, LIME and SHAP both identified BOD_e, BOD Removed (%), P_i, and TSS_i as key variables in all models, although the specific order varied slightly. For BOD_e, LIME and SHAP analyses consistently identified TSS_e and BOD_e Removed (%) as influential variables across all models, with some discrepancies in the order of influence. In the case of TSS_i, LIME and SHAP analyses identified similar key variables but with differences in the order of influence. For TSS_e, both LIME and SHAP analyses indicated TSS Removed (%), BOD_e, and P_e as significant variables. For $(NH_3)_i$, both LIME and SHAP identified $Flow_i$, P_i, and GBT Polymer Used as influential variables. For $(NH_3)_e$, LIME consistently highlighted BOD_e and *E. Coli* as significant variables, with varying impact directions across the models. SHAP's results were less consistent, with BOD_e and *E. Coli* appearing in differing orders of importance. In the prediction of P_i, both LIME and SHAP analyses identified BOD_i, $(NH_3)_i$, and TSS Removed (%) as the key variables, though the order and direction of influence differed. For P_e, LIME and SHAP both identified TSS_e, BOD_e, $Temp_e$ and Aer Basin Temp as important features, with varying order or influence.

4. Discussion

The study investigated the performance of multiple ML models, i.e., ANN, GBM, RF, XGBoost, and RF-GBM, in predicting several influential influent and effluent water quality variables in a WWTP. ANN, GBM, and XGBoost demonstrated significant potential for variable prediction as they produced low error rates and strong correlation coefficients (R^2). Table 4 shows the models with the highest R^2 for each target variable, including cases where multiple models achieved the same performance.

Table 4 Model(s) with the highest R^2 for each target variable.

Target variable	Highest (R^2)	Model(s)
BOD_i	0.98	XGBoost, GBM, RF-GBM
BOD_e	0.96	ANN, RF-GBM
$(NH_3)_i$	0.87	XGBoost, RF-GBM
$(NH_3)_e$	0.60	GBM
P_i	0.87	GBM, RF, RF-GBM
P_e	0.65	GBM, RF-GBM
TSS_i	0.97	XGBoost
TSS_e	0.97	GBM, RF-GBM

Based on our findings, the complex interactions among various WWTP variables can be captured by GBM. For example, GBM performed particularly well in predicting variables such as BOD and NH_3. This agrees with other study that GBM performed better than ANN in WWTP variable prediction [22]. Although RF performed very well on training data, overfitting caused poor performance on the test set (unseen data). As an alternative to RF model, hybrid RF-GBM model was able to increase the models' accuracy particularly for predicting BOD and P levels, by utilizing the advantages of both models. Overall, hybrid RF-GBM model provided a flexible approach that can be tailored to specific prediction challenges within WWTPs. ANN provided a competitive alternative, while GBM, XGBoost, and RF-GBM stood out as superior performers. The performance of XGBoost is consistent with other researchers' findings [38, 62]. XGBoost utilizes gradient-boosting methods to sequentially create an ensemble of weak prediction models and fix errors, leading to greater overall performance [63].

The LIME and SHAP analyses produced strong agreement with the FS results. Table 5 compares the shared feature(s) chosen by the FS methods with the features chosen by LIME and SHAP for the ML models. Traditionally FS methods are used in various studies to identify the most suitable input data from a dataset to increase model accuracy [22]. While FS methods does not consider ML models in selecting influential variables for target variables, XAI tools i.e. LIME and SHAP show influential variables significance on each models' prediction. Our study revealed that FS and XAI have identified several common influential variables regardless of choice of model or FS methods in predicting target variables.

Table 5 Comparison of the shared feature(s) chosen by the FS methods with the features chosen by LIME and SHAP.

Target variable	Common feature by FS methods	LIME	SHAP
BOD_i	P_i	ANN, GBM, RF, RF-GBM, XGBoost	ANN, GBM, RF, RF-GBM, XGBoost
BOD_e	BOD Removed (%)	ANN, GBM, RF, RF-GBM	GBM, RF, RF-GBM, XGBoost
	$(NH_3)_e$	GBM, RF, RF-GBM, XGBoost	ANN, GBM, RF, RF-GBM, XGBoost
	P_e	RF	RF
$(NH_3)_i$	Detention Time	-	RF-GBM, XGBoost
	$Flow_i$	ANN, GBM, RF, RF-GBM, XGBoost	GBM, RF, RF-GBM, XGBoost
	P_i	GBM, RF, RF-GBM, XGBoost	GBM, RF, RF-GBM, XGBoost
$(NH_3)_e$	BOD_e	ANN, GBM, RF, RF-GBM, XGBoost	GBM, RF, RF-GBM, XGBoost
	P_e	ANN, GBM, RF-GBM, XGBoost	-
P_i	BOD_i	ANN, GBM, RF, RF-GBM, XGBoost	GBM, RF, RF-GBM, XGBoost
	TSS_i	ANN, GBM, RF, RF-GBM, XGBoost	GBM, RF, RF-GBM, XGBoost
	$(NH_3)_i$	ANN, GBM, RF, RF-GBM, XGBoost	GBM, RF, RF-GBM, XGBoost
P_e	BOD_e	ANN, GBM, RF, RF-GBM, XGBoost	GBM, RF, RF-GBM, XGBoost
	TSS_e	ANN, GBM, RF, RF-GBM, XGBoost	GBM, RF, RF-GBM, XGBoost
TSS_i	BOD_i	GBM, RF, RF-GBM, XGBoost	GBM, RF, RF-GBM, XGBoost
	P_i	ANN, GBM, RF, RF-GBM, XGBoost	GBM, RF, RF-GBM, XGBoost
TSS_e	BOD Removed (%)	-	XGBoost
	TSS Removed (%)	ANN, GBM, RF, RF-GBM, XGBoost	ANN, GBM, RF, RF-GBM, XGBoost

It is also interesting to find that although ML models perform without knowledge of real-world impact of input variables on target variable, some of the common variables significantly impact certain models according to XAI. For instance, BOD_e, TSS_i, and P_i were all shown to be significant to BOD_i predictions by both LIME and SHAP. LIME explicitly reported positive and negative impacts while SHAP summary plot displayed varying importance without an apparent direction of influence. Based on our findings, LIME and SHAP can help in understanding the variables' importance in ML-based prediction, thereby can support targeted interventions in WWTP operation.

5. Conclusions

This study compared several XAI tools in predicting key WWTP variables using various ML models. Based on the findings of this study, the following conclusions are reached:

- ML models, ANN, GBM, XGBoost, and RF-GBM consistently outperform the others, exhibiting strong prediction abilities with reduced errors and higher R^2 values.
- The use of SHAP and LIME enhances the interpretability of ML models by providing the impact of input variables on the model outputs.
- The reliability of XAI tools in identifying important WWTP factors is supported by the agreement of results between FS approaches and XAI tools.

Future research should focus on incorporating diverse case studies from various WWTPs and operational conditions to enhance the adaptability and generalization of the models. The effects of various variable sets on model performance or dimension reduction strategies can also be further investigated. WWTP can optimize operations and reduce costs while mitigating environmental impacts by leveraging the interpretation provided by XAI and using robust ML models.

Acknowledgments

The authors acknowledge the Milwaukee Metropolitan Sewage District (MMSD) for providing data for the study.

Author Contributions

The original concept and supervision of the research and editing of the article by Jin Li; figures, data analysis, and writing of the article by Fuad Bin Nasir.

Competing Interests

The authors have declared that no competing interests exist.

Additional Materials

The following additional materials are uploaded at the page of this paper.

1. Table S1: List of abbreviations.
2. Table S2: Model performance metrices for 90:10 and 80:20 train-test splits.
3. Figure S1: (i, ii, iii, iv) Top five strongly correlated variables related to target variables.

4. Table S3: LIME and SHAP selected top five influential variables in descending order of influence.

References

1. Torregrossa D, Schutz G, Cornelissen A, Hernández-Sancho F, Hansen J. Energy saving in WWTP: Daily benchmarking under uncertainty and data availability limitations. Environ Res. 2016; 148: 330-337.
2. Abba SI, Elkiran G. Effluent prediction of chemical oxygen demand from the astewater treatment plant using artificial neural network application. Procedia Comput Sci. 2017; 120: 156-163.
3. Bernardelli A, Marsili-Libelli S, Manzini A, Stancari S, Tardini G, Montanari D, et al. Real-time model predictive control of a wastewater treatment plant based on machine learning. Water Sci Technol. 2020; 81: 2391-2400.
4. Zhang S, Wang H, Keller AA. Novel machine learning-based energy consumption model of wastewater treatment plants. ACS ES T Water. 2021; 1: 2531-2540.
5. Guo H, Jeong K, Lim J, Jo J, Kim YM, Park JP, et al. Prediction of effluent concentration in a wastewater treatment plant using machine learning models. J Environ Sci. 2015; 32: 90-101.
6. Wang D, Thunéll S, Lindberg U, Jiang L, Trygg J, Tysklind M, et al. A machine learning framework to improve effluent quality control in wastewater treatment plants. Sci Total Environ. 2021; 784: 147138.
7. El-Rawy M, Abd-Ellah MK, Fathi H, Ahmed AK. Forecasting effluent and performance of wastewater treatment plant using different machine learning techniques. J Water Process Eng. 2021; 44: 102380.
8. Li G, Ji J, Ni J, Wang S, Guo Y, Hu Y, et al. Application of deep learning for predicting the treatment performance of real municipal wastewater based on one-year operation of two anaerobic membrane bioreactors. Sci Total Environ. 2022; 813: 151920.
9. Zhu J, Jiang Z, Feng L. Improved neural network with least square support vector machine for wastewater treatment process. Chemosphere. 2022; 308: 136116.
10. Zhu JJ, Borzooei S, Sun J, Ren ZJ. Deep learning optimization for soft sensing of hard-to-measure wastewater key variables. ACS ES T Eng. 2022; 2: 1341-1355.
11. Aghdam E, Mohandes SR, Manu P, Cheung C, Yunusa-Kaltungo A, Zayed T. Predicting quality parameters of wastewater treatment plants using artificial intelligence techniques. J Clean Prod. 2023; 405: 137019.
12. Shyu HY, Castro CJ, Bair RA, Lu Q, Yeh DH. Development of a soft sensor using machine learning algorithms for predicting the water quality of an onsite wastewater treatment system. ACS Environ Au. 2023; 3: 308-318.
13. Wei X, Yu J, Tian Y, Ben Y, Cai Z, Zheng C. Comparative performance of three machine learning models in predicting influent flow rates and nutrient loads at wastewater treatment plants. ACS ES T Water. 2023; 4: 1024-1035.
14. Xu Y, Wang Z, Nairat S, Zhou J, He Z. Artificial intelligence-assisted prediction of effluent phosphorus in a full-scale wastewater treatment plant with missing phosphorus input and removal data. ACS ES T Water. 2023; 4: 880-889.

15. Yu J, Tian Y, Jing H, Sun T, Wang X, Andrews CB, et al. Predicting regional wastewater treatment plant discharges using machine learning and population migration big data. ACS ES T Water. 2023; 3: 1314-1328.
16. Alsulaili A, Refaie A. Artificial neural network modeling approach for the prediction of five-day biological oxygen demand and wastewater treatment plant performance. Water Supply. 2021; 21: 1861-1877.
17. Nasir FB, Li J. Understanding machine learning predictions of wastewater treatment plant sludge with explainable artificial intelligence. Water Environ Res. 2024; 96: e11136.
18. Fan M, Hu J, Cao R, Ruan W, Wei X. A review on experimental design for pollutants removal in water treatment with the aid of artificial intelligence. Chemosphere. 2018; 200: 330-343.
19. Adibimanesh B, Polesek-Karczewska S, Bagherzadeh F, Szczuko P, Shafighfard T. Energy consumption optimization in wastewater treatment plants: Machine learning for monitoring incineration of sewage sludge. Sustain Energy Technol Assess. 2023; 56: 103040.
20. Keerio HA, Shah SA, Ali Z, Panhwar S, Solangi GS, Ali A, et al. A fascinating exploration into nitrite accumulation into low concentration reactors using cutting-edge machine learning techniques. Process Biochem. 2024; 146: 160-168.
21. Solangi GS, Ali Z, Bilal M, Junaid M, Panhwar S, Keerio HA, et al. Machine learning, water quality index, and GIS-based analysis of groundwater quality. Water Pract Technol. 2024; 19: 384-400.
22. Bagherzadeh F, Mehrani MJ, Basirifard M, Roostaei J. Comparative study on total nitrogen prediction in wastewater treatment plant and effect of various feature selection methods on machine learning algorithms performance. J Water Process Eng. 2021; 41: 102033.
23. Wu Z, Duan H, Li K, Ye L. A comprehensive carbon footprint analysis of different wastewater treatment plant configurations. Environ Res. 2022; 214: 113818.
24. Keerio HA, Bae W, Park J, Kim M. Substrate uptake, loss, and reserve in ammonia-oxidizing bacteria (AOB) under different substrate availabilities. Process Biochem. 2020; 91: 303-310.
25. Singh NK, Yadav M, Singh V, Padhiyar H, Kumar V, Bhatia SK, et al. Artificial intelligence and machine learning-based monitoring and design of biological wastewater treatment systems. Bioresour Technol. 2023; 369: 128486.
26. Zhao LJ, Chai TY, Yuan DC. Selective ensemble extreme learning machine modeling of effluent quality in wastewater treatment plants. Int J Autom. Comput. 2012; 9: 627-633.
27. Bagheri M, Mirbagheri SA, Ehteshami M, Bagheri Z, Kamarkhani AM. Analysis of variables affecting mixed liquor volatile suspended solids and prediction of effluent quality parameters in a real wastewater treatment plant. Desalin Water Treat. 2016; 57: 21377-21390.
28. Sharghi E, Nourani V, AliAshrafi A, Gökçekuş H. Monitoring effluent quality of wastewater treatment plant by clustering based artificial neural network method. Desalin Water Treat. 2019; 164: 86-97.
29. Khatri N, Khatri KK, Sharma A. Prediction of effluent quality in ICEAS-sequential batch reactor using feedforward artificial neural network. Water Sci Technol. 2019; 80: 213-222.
30. Al-Ghazawi Z, Alawneh R. Use of artificial neural network for predicting effluent quality parameters and enabling wastewater reuse for climate change resilience-A case from Jordan. J Water Process Eng. 2021; 44: 102423.
31. Elmaadawy K, Abd Elaziz M, Elsheikh AH, Moawad A, Liu B, Lu S. Utilization of random vector functional link integrated with manta ray foraging optimization for effluent prediction of wastewater treatment plant. J Environ Manage. 2021; 298: 113520.

32. Nourani V, Asghari P, Sharghi E. Artificial intelligence based ensemble modeling of wastewater treatment plant using jittered data. J Clean Prod. 2021; 291: 125772.
33. Ly QV, Truong VH, Ji B, Nguyen XC, Cho KH, Ngo HH, et al. Exploring potential machine learning application based on big data for prediction of wastewater quality from different full-scale wastewater treatment plants. Sci Total Environ. 2022; 832: 154930.
34. Dantas MS, Christofaro C, Oliveira SC. Artificial neural networks for performance prediction of full-scale wastewater treatment plants: A systematic review. Water Sci Technol. 2023; 88: 1447-1470.
35. Mahanna H, El-Rashidy N, Kaloop MR, El-Sapakh S, Alluqmani A, Hassan R. Prediction of wastewater treatment plant performance through machine learning techniques. Desalin Water Treat. 2024; 319: 100524.
36. Park J, Lee WH, Kim KT, Park CY, Lee S, Heo TY. Interpretation of ensemble learning to predict water quality using explainable artificial intelligence. Sci Total Environ. 2022; 832: 155070.
37. Hu Y, Wei R, Yu K, Liu Z, Zhou Q, Zhang M, et al. Exploring sludge yield patterns through interpretable machine learning models in China's municipal wastewater treatment plants. Resour Conserv Recycl. 2024; 204: 107467.
38. Shao S, Fu D, Yang T, Mu H, Gao Q, Zhang Y. Analysis of machine learning models for wastewater treatment plant sludge output prediction. Sustainability. 2023; 15: 13380.
39. Shafighfard T, Kazemi F, Asgarkhani N, Yoo DY. Machine-learning methods for estimating compressive strength of high-performance alkali-activated concrete. Eng Appl Artif Intell. 2024; 136: 109053.
40. Shafighfard T, Kazemi F, Bagherzadeh F, Mieloszyk M, Yoo DY. Chained machine learning model for predicting load capacity and ductility of steel fiber-reinforced concrete beams. Comput Aided Civ Infrastruct Eng. 2024. doi: 10.1111/mice.13164.
41. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017. doi: 10.48550/arXiv.1705.07874.
42. Lundberg SM, Erion GG, Lee SI. Consistent individualized feature attribution for tree ensembles. ArXiv. 2018. doi: 10.48550/arXiv.1802.03888.
43. Li R, Feng K, An T, Cheng P, Wei L, Zhao Z, et al. Enhanced insights into effluent prediction in wastewater treatment plants: Comprehensive deep learning model explanation based on SHAP. ACS ES T Water. 2024; 4: 1904-1915.
44. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining; 2016 August 13-17; San Francisco, CA, USA. New York, NY: Association for Computing Machinery. pp. 1135-1144.
45. Ye Z, Yang J, Zhong N, Tu X, Jia J, Wang J. Tackling environmental challenges in pollution controls using artificial intelligence: A review. Sci Total Environ. 2020; 699: 134279.
46. Matheri AN, Ntuli F, Ngila JC, Seodigeng T, Zvinowanda C. Performance prediction of trace metals and cod in wastewater treatment using artificial neural network. Comput Chem Eng. 2021; 149: 107308.
47. Konstantinov AV, Utkin LV. Interpretable machine learning with an ensemble of gradient boosting machines. Knowl Based Syst. 2021; 222: 106993.
48. Tyralis H, Papacharalampous G, Langousis A. A brief review of random forests for water scientists and practitioners and their recent history in water resources. Water. 2019; 11: 910.

49. Nafsin N, Li J. Prediction of total organic carbon and E. coli in rivers within the Milwaukee River basin using machine learning methods. Environ Sci Adv. 2023; 2: 278-293.
50. Jiang M, Wang J, Hu L, He Z. Random forest clustering for discrete sequences. Pattern Recognit Lett. 2023; 174: 145-151.
51. Szomolányi O, Clement A. Use of random forest for assessing the effect of water quality parameters on the biological status of surface waters. GEM. 2023; 14: 20.
52. Sun Z, Wang G, Li P, Wang H, Zhang M, Liang X. An improved random forest based on the classification accuracy and correlation measurement of decision trees. Expert Syst Appl. 2024; 237: 121549.
53. Yadav P, Chandra M, Fatima N, Sarwar S, Chaudhary A, Saurabh K, et al. Predicting influent and effluent quality parameters for a UASB-based wastewater treatment plant in Asia covering data variations during COVID-19: A machine learning approach. Water. 2023; 15: 710.
54. Xie Y, Chen Y, Lian Q, Yin H, Peng J, Sheng M, et al. Enhancing real-time prediction of effluent water quality of wastewater treatment plant based on improved feedforward neural network coupled with optimization algorithm. Water. 2022; 14: 1053.
55. Sargent RG. Verification and validation of simulation models. Proceedings of the 2010 Winter Simulation Conference; 2010 December 05-08; Baltimore, MD, USA. Piscataway, NJ: IEEE. pp. 166-183.
56. Tsioptsias N, Tako A, Robinson S. Model validation and testing in simulation: A literature review. Proceedings of the 5th Student Conference on Operational Research (SCOR 2016); 2016 April 08-10; Nottingham, UK. Wadern, Germany: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
57. Zhang X, Liu CA. Model averaging prediction by K-fold cross-validation. J Econom. 2023; 235: 280-301.
58. Kazemi F, Asgarkhani N, Shafighfard T, Jankowski R, Yoo DY. Machine-learning methods for estimating performance of structural concrete members reinforced with fiber-reinforced polymers. Arch Comput Methods Eng. 2024. doi: 10.1007/s11831-024-10143-1.
59. Bagherzadeh F, Shafighfard T, Khan RM, Szczuko P, Mieloszyk M. Prediction of maximum tensile stress in plain-weave composite laminates with interacting holes via stacked machine learning algorithms: A comparative study. Mech Syst Signal Process. 2023; 195: 110315.
60. Shafighfard T, Bagherzadeh F, Rizi RA, Yoo DY. Data-driven compressive strength prediction of steel fiber reinforced concrete (SFRC) subjected to elevated temperatures using stacked machine learning algorithms. J Mater Res Technol. 2022; 21: 3777-3794.
61. Safder U, Kim J, Pak G, Rhee G, You K. Investigating machine learning applications for effective real-time water quality parameter monitoring in full-scale wastewater treatment plants. Water. 2022; 14: 3147.
62. Zhang Y, Wu H, Xu R, Wang Y, Chen L, Wei C. Machine learning modeling for the prediction of phosphorus and nitrogen removal efficiency and screening of crucial microorganisms in wastewater treatment plants. Sci Total Environ. 2024; 907: 167730.
63. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 August 13-17; San Francisco, CA, USA. New York, NY: Association for Computing Machinery. pp. 785-794.