

Article

Not peer-reviewed version

Textile Wastewater COD Forecasting Using a Shock-Aware Explainable Gated Hybrid Framework

Syrin Jahan Ritu , Alamin Howlader , Rayhanul Islam Sony , Atique Ahammad Zawad ,
[Shaharior Islam Chowdhury](#) *

Posted Date: 17 March 2026

doi: 10.20944/preprints202603.1384.v1

Keywords: textile wastewater treatment; COD prediction; SAGE-GBTCN model; shock-aware modeling; water quality management; clean water and sanitation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Textile Wastewater COD Forecasting Using a Shock-Aware Explainable Gated Hybrid Framework

Syrin Jahan Ritu ¹, Alamin Howlader ², Rayhanul Islam Sony ³, Atique Ahammad Zawad ⁴ and Shaharior Islam Chowdhury ^{5,*}

¹ Department of Textile Engineering, BGMEA University of Fashion and Technology, Dhaka1230, Bangladesh

² Department of Chemical Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh

³ Trine University, Angola, IN 46703, Indiana, USA

⁵ Department of Architecture, Khulna University of Engineering & Technology, Khulna, 9203, Bangladesh

* Correspondence: isharior@gmail.com

Abstract

Textile dyeing industry is a significant contributor of complicated and extremely polluting wastewater. This wastewater has intermittent loads of chemical oxygen demand (COD), stains and other pollutants which puts dangerous effects on the sustainability of the environment and human beings in general. The traditional operation of wastewater treatment plants is reactive and rule-based to a large extent. These methods are ineffective in dealing with the non-linear dynamic character of the effluent of the textile business, resulting in low efficacy and recurring regulatory breach. To overcome these shortcomings, this paper will suggest a new hybrid architecture SAGE-GBTCN (Shock-Aware Gated Ensemble with Gradient Boosting and Temporal Correction Network) to be used in the effective prediction of wastewater pollution. This model combines a gradient boosting ensemble to produce baseline predictions and a parallel temporal network with a residual correction. A shock-sensitive gating system is used to dynamically modify the correction process to consider any sudden, non-stationary changes in the nature of the effluents. This design makes the model very useful in capturing the long-term trends as well as abrupt disruptions within textile wastewater. The suggested SAGE-GBTCN model was tested with the help of data on a full-scale wastewater treatment facility. The findings are shown to be more accurate in prediction and better resistant to abnormal operating condition. The model also demonstrates high possibilities to facilitate active and energy saving management of textile wastewater treatment processes, which will result in an R^2 predictive value of 0.942 and a RMSE of 30.30 of COD. Although validated on full-scale industrial WWTP data, the proposed framework targets operational characteristics typical of textile effluent treatment plants, including batch-wise COD loading, abrupt shock events, and chemically driven variability.

Keywords: textile wastewater treatment; COD prediction; SAGE-GBTCN model; shock-aware modeling; water quality management; clean water and sanitation

1. Introduction

The textile dyeing industry is one of the most water-intensive and environmentally problematic industrial sectors in the world [1], where huge amounts of complex wastewater containing high concentrations of chemical oxygen demand (COD), biological oxygen demand (BOD), total suspended solids (TSS), colorants, salts, surfactants, and auxiliary chemicals are produced [2]. Rapid industrialization, growing consumption of dyed fabrics by the end users, accelerated production cycles, etc, have increased the pollutant load released from the textile effluent treatment plants (ETPs) [3]. In many developing and industrializing regions, textile wastewater has become a key issue in degrading the waterway system and causing toxicity to water ecosystems and compliance restrictions, and posing long-term threats to environmental sustainability and public health [4]. Consequently, effective monitoring, prediction, and control of wastewater pollution in textile dyeing processes is becoming a critical

priority in the minds of both the regulatory and industry stakeholders [5]. Conventional wastewater treatment approaches in textile ETPs are largely reactive because they are based on fixed operational rules, on-time, and on laboratory and operator experience [6]. Such methods are in many cases unable to respond adequately to the highly dynamic and non-linear nature of textile wastewater, in which the quality can change markedly as a result of changing dye batches, chemical dosing, production timings, and hydraulic loading conditions. These fluctuating changes are often the cause of treatment inefficiencies, energy overconsumption, chemical wastage and the occasional regulatory exceedance. With the increasing importance of sustainable manufacturing practices, energy efficiency, and circular water infrastructure, it is crucial to have intelligent data-driven frameworks capable of anticipating the behavior of the pollutant and enabling proactive decision-making in the management of textile wastewater rather than its retrospective correction [7].

In recent years, artificial intelligence (AI) and machine learning (ML) technologies have attracted much attention to the prediction of wastewater quality and process optimization [8]. Data-driven models have been used to predict key pollution parameters like COD, BOD, and TSS, making it possible to have early warning systems for treatment plants [9]. Traditional machine learning approaches, such as linear regression, support vector machines, random forests, and gradient boosting models, have proven to be skilled at prediction in comparison with mechanistic models, especially if there is enough historical data [10]. The latter methods have the ability to capture non-linear relationships among process variables without any need to prescribe explicit physical equations, which makes them attractive for complex industrial wastewater systems such as textile dyeing ETPs [11]. However, although they are becoming more popular, current AI-enabled wastewater prediction methods have some significant drawbacks when applied to real-world cases of textile wastewater scenarios [12]. First, several models provide only point-wise prediction accuracy and do not account for the temporal structuring and regime shifts of industrial operations [13]. Textile dyeing wastewater is by no means stationary, but is influenced by batch-wise production cycles, sudden pollutant shocks, and intervalological operational disturbances [14]. Models that assume stable statistics often have a hard time dealing with such conditions. Second, deep learning models, although powerful, are often task-intensive in terms of the data (i.e., high dimensionality and homogeneity) required and high computational resources, and may not be practical for many industrial ETPs. Moreover, purely deep architectures are not always better in performance compared to well-designed ensemble learning methods when the level of data is limited or the noise level is high [15]. Third, many studies are still limited to laboratory-scale or single-plant case studies, so their adaptability and generalization to other textile facilities with different operational characteristics are limited [16].

Beyond the issues of modeling, there is an extra dimension of complexity due to issues of sustainability. Energy used for aeration, pumping, and chemical dosing takes a substantial part of the operational cost of textile ETPs and directly contributes to the environmental footprint of the facilities [17]. Predictive models that lack process variability will also potentially lead to energy consumption because these models will drive conservative or excessive treatment design [18]. Therefore, accurate pollutant forecasting needs to be combined with an understanding of operational dynamics in order to support energy-efficient and environmentally responsible wastewater management [19]. These challenges pose a fundamental research gap because of the absence of robust, adaptive, and process-aware predictive frameworks for reliable probabilistic projection of textile wastewater pollution under dynamic operating conditions, and at the same time computationally efficient and practically deployable. Specifically, there is a need for hybrid modeling strategies, which can merge both the qualities of good ensemble learning methods incorporating temporal awareness and shock-sensitive behavior to handle both gradual trends and abrupt disturbance commonly seen in the textile dyeing effluents [20].

Motivated by this gap, the current research aims to explore an artificial intelligence (AI) based wastewater pollution forecasting framework that is generic to the textile dyeing ETP, and that is formulated with sustainability as the main theme. Rather than depending on a single modeling

paradigm, the proposed approach diverges by a hybrid-based architecture, i.e., Shock Aware Gated Ensemble with Gradient Boosting with Temporal Correction Network, or SAGE-GBTCN. It is designed to gather the power of ensemble learning with temporal residual correction mechanisms. The core notion behind this proposal is to utilize the solid generalization ability of the gradient boosting models for a polluting baseline prediction while introducing temporal correction for regime change, pollutant shocks and for short-term dynamics, which are an intrinsic feature in textile wastewater systems. Importantly, the model is designed to be interpretable and computationally feasible in order to ease its adaptation to actual industrial situations. The development of this framework is based on full-scale wastewater treatment plant data that encompasses the important physicochemical, operational and environmental variables of relevance to industrial effluent treatment. While the basic data set constitutes a generalized wastewater treatment scenario, the modeling approach is specifically developed to be transferable to textile dyeing ETPs by AD, about predicted pollutant dynamics as closely as possible with textile-specific operational routines (e.g., batch processing, variable chemical dosing, ebb and flow of hydraulic factors). By concentrating on COD along with assigned indicators as major targets, the framework responds to one of the most important regulatory and operational parameters in the textile wastewater field. From a practical perspective, the proposed approach aims to enable proactive forms of control in textile ETPs, including the early detection of pollutant load escalation, informed adjustments to aeration intensity, and optimized chemical consumption. From a scientific perspective, this work contributes to the growing body of knowledge on hybrid AI models for environmental systems, specifically in the context of ensemble learning, where temporal intelligence can enhance robustness in non-stationary conditions. Unlike pure deep learning methods, in the proposed framework, reliability, adaptability, and sustainability relevance, instead of model complexity alone, are emphasized.

The key contributions and insights of this study are summarized as follows:

- A hybrid artificial intelligence framework, SAGE-GBTCN, is proposed for pollutant forecasting of wastewater, which integrates ensemble learning based on gradient boosting and temporal correction for residual, enabling the robust prediction effect under non-stationary conditions for industrial operating conditions.
- A process-aware modeling strategy is presented in order to capture the pollutant variability, temporal short-term dependencies, and shock-like behavior which are typically observed in textile dyeing wastewater treatment plants.
- A thorough empirical assessment of state-of-the-art machine learning and deep learning models is carried out, showing that well-designed ensemble learning models could reach near-ceiling predictive performance in COD forecasting, whereas more complex deep or hybrid residual designs were not necessarily more accurate in data-limited industrial wastewater applications.
- A sustainability-oriented perspective is incorporated into the forecasting framework through the interdependence of reliable pollutant prediction and possibilities of enhanced treatment efficiency, decreased over-aeration, and energy-conscious operational decision-making in textile ETPs.
- An adaptable and transferable methodology is presented, which, despite being developed with full-scale wastewater treatment data, is clearly allied to the operational characteristics of textile dyeing ETPs, which supports practical deployment in diverse industrial facilities.

The remainder of this paper is organized as follows. Section 2 reviews related work on AI-based wastewater pollution prediction and sustainability-driven treatment optimization, with a focus on industrial and textile applications. Section 3 describes the dataset, preprocessing strategy, and the proposed SAGE-GBTCN modeling framework. Section 4 discusses the results and their implications for textile wastewater sustainability, and Section 5 concludes the paper with key findings and directions for future research.

2. Related Work

As the process of wastewater treatment becomes more complicated, the quality of the effluent water can no longer be predicted with accuracy, which in its turn would ensure the stable operation and control over the process. Conventional modelling mechanisms may not be able to represent nonlinear and dynamic nature of wastewater systems, which has created increased enthusiasm over the data-driven machine learning and deep learning based modelling techniques. More recent works investigated an extensive variety of AI-driven models, such as LSTM, GRU, CNN and hybrid networks to forecast the quality of effluents, sensor uncertainty, and optimization of treatment performance. In this section, the current literature is examined by combining the studies sharing similar modeling strategies, stating the way various researchers face similar prediction problems with the help of similar model architectures and data-driven methods.

2.1. Time-Series Deep Learning Models for Effluent Quality Prediction

The wastewater effluent quality prediction has been widely carried out using time-series deep learning models because they are effective in addressing the nonlinear temporal relationships. [21] suggested the hybrid Attention-LSTM-XGBoost models to forecast COD, NH₃-N, TN, and TP, and they showed higher performance than independent LSTM and XGBoost models. The research found the significance of adaptive feature weighting in the challenging WWTP dynamics by combining mechanisms of attention with ensemble learning. Similarly, [22,23] used the LSTM-based models to predict the quality of effluents, but included SHAP-based explainability to examine the contribution of time and multivariate features. Their results back up the notion that LSTM structures are capable of being very predictive with well-represented long-term dependencies, even though model performance has remained reliant on feature selection and data quality. With this line of direction, [24] proposed a dual-attention LSTM model that concurrently focuses on the significant input variables and crucial time steps, which further enhance the prediction accuracy and generalization. In order to handle extreme conditions during the operation, [25] suggested a probabilistic encoder decoder LSTM that can predict the effluent quality during shock loading conditions and provides better early warning at the cost of more complex computations.

2.2. Hybrid and Ensemble Learning Frameworks for Wastewater Prediction

In an effort to make various frameworks stronger in a wide range of operating conditions, a number of studies discussed hybrid and ensemble learning. [26] proposed a unified system of control utilizing CNN, LSTM, and GRU models with a QR-RF meta-learner and achieved significant enhancement in the accuracy of effluent prediction and performance. In this paper, it was established that ensemble learning has the ability to combine complementary deep learning models in WWTP control. Likewise, [27] suggested a hybrid model combining signal decomposition and deep learning and dynamic features selection and obtained accurate predictions of the influent COD and BOD₅ at various WWTPs. [28] also found that feature decomposition became more useful in predictive stability by using LSTM- and GRU-based data fusion using multi-source data, which entails combining water quality data, process data, energy data, and flow data. In a similar attempt, [29] applied multi-source fusion methods to urban sewer systems and demonstrated that GRU-based models are always better than traditional machine learning and linear techniques in cases when heterogeneous data sources are successfully integrated.

2.3. Soft Sensor and Fault-Tolerant Prediction Approaches

The issue of sensor unreliability and missing data is still a serious concern in WWTP monitoring which encourages the creation of soft sensor models. A dual-stage attention-fusion deep learning model was introduced by [30], which is able to reconstruct the defective sensor signal and forecast the quality of effluents. Their findings showed better prediction and energy saving with the importance of attention mechanisms in fault-tolerant models. When considering missing data in a different light, [31] explored machine learning frameworks that use KNN-based imputation approaches to forecast effluent.

They saw that while appropriate missing-data handling are able to clearly reduce the performance decline, realistic predictive precision has been limited by the amount of data and sensor sensitivity.

2.4. Flow Forecasting and Operational Optimization Machine Learning Model.

Other places of machine learning use include wastewater flow prediction and optimization of operation besides prediction of effluent quality. [32] introduced a hybrid AquaFlowNet that is a regression tree, LS-SVM and PSO that were used to predict wastewater flow and quality in real-time based on SCADA data. The experiment showed that the efficiency of the treatment and overflow could be greatly increased by optimization with the help of the ML, but the performance could not be guaranteed in the extreme cases of the flow. On the same note, [33] used XGBoost to predict wastewater inflow during rainy seasons, which was found to be better than either a Random Forest or LSTM model. Their findings not only emphasized the effectiveness of gradient boosting techniques in responding to the nonlinear inflow patterns, but also on the reliance of good forecasts of rainfall and the calibration of the techniques to the sites.

2.5. Image-Based and Multimodal Prediction Models.

The recent literature has investigated the concept of multimodal learning to alleviate sensor dependency and improve predictive robustness. [34] introduced a brain-inspired CNN-LSTM model, which combines wastewater surface image and temporal water quality data; this model will provide better prediction accuracy when even without some sensor inputs. This method revealed the possibilities of visual information as an auxiliary source of data, but it was only tested on a pilot-scale level.

2.6. Machine Learning Intelligent Wastewater Treatment of Textiles and Industry.

The use of AI to optimize the treatment of textile and industrial wastewater has also been applied in modelling. [35] have conducted a review of machine learning and deep learning tools used in the process of photocatalytic dye degradation, with a special focus being on the high efficacy of hybrid kinetic-ML models to optimize the process. [36,37] applied AI-aided optimization of the treatment and electrocoagulation processes of nano-zerovalent aluminum in experimental research, respectively, and obtained high pollutant removal efficiencies in controlled conditions, respectively. In addition to these methods, [38–41] used regression, gradient boosting, ANN, and neural time-series models to estimate dye removal and adsorption efficiency. Although these studies always showed high levels of predictive accuracy, the majority of them were limited to laboratory-scale data, which casts doubt on the question of scalability and real-world use. [42] also emphasized the importance of supervised AI models in the maximization of photocatalytic wastewater treatment in terms of the key operational parameters determined by explainable learning methods.

2.7. Review and Bibliometric Analyses.

In order to put these developments into perspective, [43] presented an extensive literature review of AI uses in wastewater treatment, noting developments in prediction, monitoring and fault detection. [44] added to this point of view with a bibliometric analysis to reveal the trends that characterize the research and the rising demand to have explainable, scalable, and real-time AI solutions.

Table 1. Summary of AI-Based Wastewater Treatment and Effluent Quality Prediction Studies

Year	Reference	Model Name	Key Results	Limitations
2025	[21]	Attention-LSTM-XGBoost	Test R^2 up to 0.986 for COD, $\text{NH}_3\text{-N}$, TN, TP	Prediction delay under shock loads, scalability issues, low interpretability
2025	[26]	CNN-LSTM-GRU + QR-RF	RMSE = 4.76 mg/L (24 h), MAE = 0.85 mg/L (1 h)	Increased system complexity, scalability challenges
2025	[27]	Signal Decomposition + DL + DFS	$R^2 = 0.96$ (COD), 0.93 (BOD_5)	High computational cost, real-time data dependency
2025	[28]	LSTM / GRU Multi-Source Fusion	R^2 improvement up to 47.75% over RF	Requires complete multi-source monitoring
2025	[30]	AFAE-ACNN Soft Sensor	$R^2 = 0.9909$, 57.2% error reduction	Sensitive to unseen sensor faults
2025	[32]	AquaFlowNet (RT + LS-SVM + PSO)	~99% accuracy, major energy savings	Sensitive to noise and extreme flow events
2025	[31]	ML + KNN Imputation	R^2 up to 0.874	Moderate real-time accuracy
2024	[22]	LSTM + SHAP	$R^2 = 0.96$ (COD), 0.95 (TP)	Feature selection and tuning complexity
2024	[24]	Dual-Attention LSTM	Accuracy improvement of 5–14%	High computational demand
2025	[25]	Probabilistic ED-LSTM	49.7% accuracy improvement, 6-h early warning	Complex probabilistic calibration
2024	[43]	AI Review Framework	Reported R^2 range: 0.64–1.00	High implementation cost
2024	[44]	Bibliometric Analysis	ML/ANN reported R^2 up to 0.99	No primary predictive modeling
2025	[33]	XGBoost	$R^2 = 0.79\text{--}0.83$ (rain inflow)	Rainfall forecast dependency
2024	[34]	CNN-LSTM (Image + Data)	>23% error reduction	Pilot-scale validation only
2021	[29]	GRU / LSTM / RNN Fusion	R^2 improvement up to 43%	City-specific calibration
2025	[35]	Hybrid ML-DL Models	Up to 99% dye removal	Limited full-scale data
2025	[36]	nZVAL + ML	$R^2 = 0.974$, high COD removal	Controlled experimental conditions
2025	[37]	EC / EC-DAF	>95–98% dye removal	Synthetic wastewater only
2025	[38]	Linear Regression	$R^2 = 0.9265$	Laboratory-scale dependency
2024	[39]	Gradient Boosting + BO	~90.47% adsorption efficiency	Batch experimental setup
2025	[40]	ANN / GPR	ANN $R^2 = 0.85$	Controlled effluent conditions
2024	[45]	TCN-LSTM	33–63% accuracy improvement	Site-specific calibration
2024	[23]	LSTM + SHAP	$R^2 = 0.96$ (COD)	High data quality requirement
2025	[46]	ANN + RF	$R^2 = 0.89$, 38% energy saving	Pilot/simulation-based validation
2025	[41]	NNTS	Correlation coefficient = 0.99992	Column-scale experiment
2025	[42]	ANN / SVM / Tree Models	$R^2 > 0.95$	Limited real-effluent datasets

3. Methodology

This section outlines the entire methodological procedure that will be used in this study in the prediction of COD as shown in Figure 1. The proposed pipeline will guarantee reasonable comparison, high learning and good utilisation of time dynamics in homogenous and shocking operating circumstances.

It starts with the creation of a complete dataset comprising of past COD measurements and corresponding auxiliary variables. Preprocessing of raw data is carried out by normalization, missing value treatment, and development of new features such as lagged variables, and rolling statistical features. The resulting processed data is subsequently divided into training (70%), validation (15%), and testing (15%) data to allow the development and testing of the models without bias.

To set reference levels of performance, a number of baseline machine learning models, i.e., Gradient Boosting based on Histogram, CatBoost, Random Forest, Support Vector Regression, and Multilayer Perceptron are trained on the same data splits and feature representations. These models offer valuable standards of evaluating the advantages of the offered hybrid solution.

The main contribution of this study will be the SAGE-GBTCN framework that combines statistical learning with deep temporal modeling. As depicted by the central block of 1, the given model first uses gradient boosting predictor to reveal the main international tendencies and creates a basic COD forecast. This difference between the observed and this base prediction is then explicitly modeled-in terms of a parallel residual prediction learning and shock aware gating mechanism, so we are able to model and actually correct these under volatile regimes.

Lastly, model performance is determined by the various error and goodness of fit measures, RMSE, MAE, MAPE, and R^2 to provide a holistic measure of accuracy and strength also provides the Shap feature importance for explainability of the proposed model. Such a methodological approach offers a clear and repeated basis of the evaluation of the efficiency of the suggested hybrid framework.

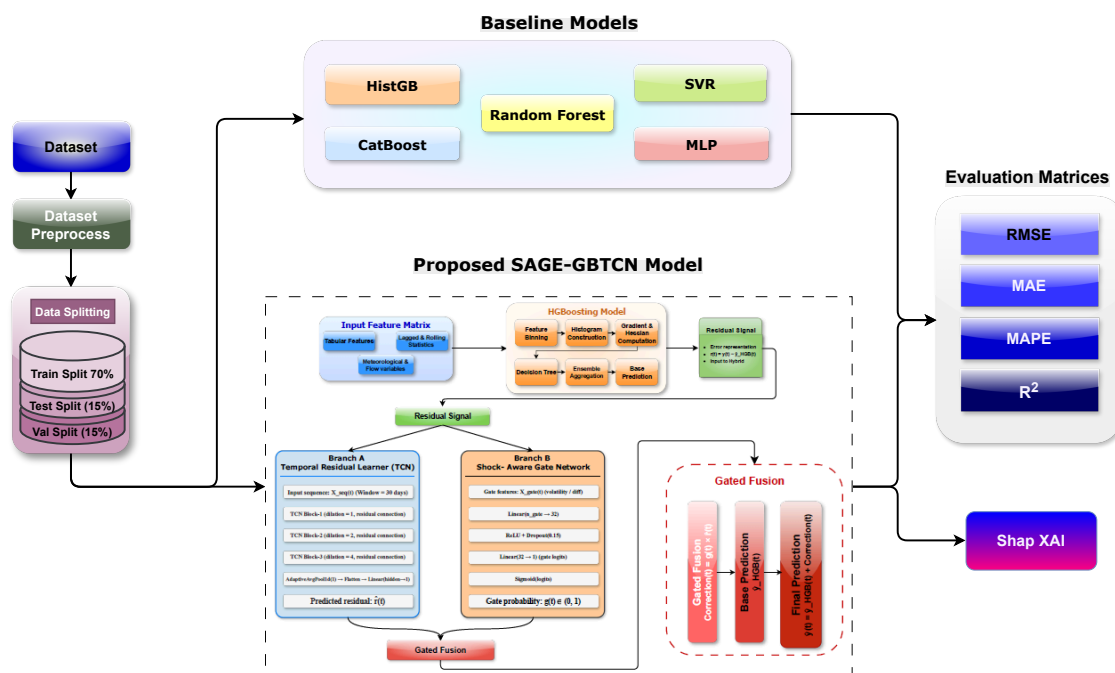


Figure 1. Overall methodological workflow for COD prediction. The pipeline illustrates data preprocessing and feature engineering, baseline model training, and the proposed SAGE-GBTCN framework integrating histogram-based gradient boosting with parallel residual learning and shock-aware gating, followed by comprehensive performance evaluation and shap xai.

3.1. Dataset Details

Full Scale Waste Water Treatment Plant Data ([?]) is a multivariate time-series dataset that records daily operation measurement of a full-scale wastewater treatment plant from January 1, 2014 to June 27, 2019, more than five years. The data set contains 1,382 records, 20 columns, which include not only variables characterizing the wastewater process (e.g., Average Inflow / Outflow, Energy Consumption, Ammonia, Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Total Nitrogen) but also meteorological data (temperature, humidity, rainfall, visibility and wind speeds). It also has direct time elements (Year, Month, and Day) to analyse time. The wide range of hydraulic, biological, and climatic variables makes it an appropriate dataset to be used in exploratory analytics, time-series modeling, or machine learning to address performance analysis, fault detection, or processes optimization in the context of wastewater treatment.

The dataset is especially useful in environmental engineering and data science research due to the fact that it includes both process dynamics and external weather conditions to study how factors of the climate affect the efficiency of the treatment process and the consumption of resources, such as energy consumption. Having all of the numeric columns standardized to facilitate modeling, researchers are able to develop predictive models on key targets like energy consumption, water quality performances, or system performance against different operational and environmental conditions. The broad time span gives it the opportunity to perform a seasonal trend analysis and create forecasting models that would facilitate decision making in the operation of treatment plants and sustainability evaluation. Textile dyeing effluent treatment plants represent a particularly challenging subclass of industrial WWTPs due to their highly variable influent composition, batch-based discharge patterns, and frequent chemical dosing operations, all of which result in pronounced COD shocks. The selected full-scale WWTP dataset exhibits these same operational characteristics, making it a representative benchmark for evaluating shock-aware COD forecasting frameworks intended for textile wastewater applications.

Table 2. Descriptive Statistics of Key Wastewater and Operational Parameters

Variable	Mean	Std. Dev.	Min	Max
COD	845.96	145.42	360.00	1700.00
BOD	382.06	86.00	140.00	850.00
Energy	275159.09	44640.53	116638.00	398328.00
TN	62.74	3.57	40.00	92.00
NH ₃	39.22	7.76	13.00	93.00
Inflow	4.51	1.44	2.59	18.97
Outflow	3.93	1.23	0.00	7.92

3.2. Preprocessing Pipeline

To ensure robustness and reproducibility of the predictive models, a systematic preprocessing pipeline was implemented. This pipeline transforms raw wastewater and operational measurements into a structured representation suitable for high-dimensional time-series learning.

3.2.1. Variable Standardization and Statistical Summary

To maintain consistency with scientific nomenclature, raw sensor variables were standardized into concise chemical and operational abbreviations. Core wastewater indicator such as Chemical Oxygen Demand (COD), Biological Oxygen Demand (BOD), Total Nitrogen (TN), and Ammonia (NH₃) were retained along with operational variables including energy consumption, inflow, and outflow rates. Table 2 summarizes the descriptive statistics of the selected features, illustrating their operational ranges and variability.

3.2.2. Temporal Characterization and Exploratory Analysis

Temporal behavior of wastewater quality and operational energy consumption was examined over the 2014-2019 period to identify trends, variance shifts, and non-stationary patterns. Figure 2 and 3 presents the pollutant and energy time-series profiles.

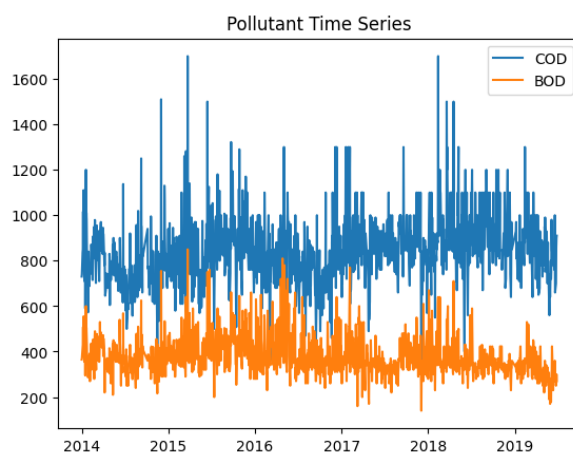


Figure 2. Time-series representation of key wastewater pollutant concentrations, including COD and BOD, illustrating temporal variability and long-term trends in textile wastewater characteristics.

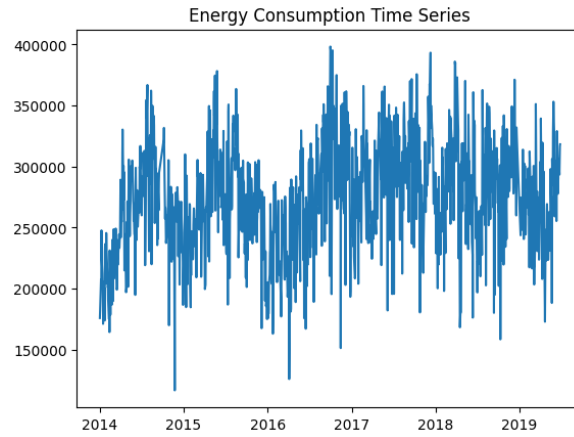


Figure 3. Temporal profile of operational energy consumption associated with textile wastewater treatment processes, highlighting variability in system demand over time.

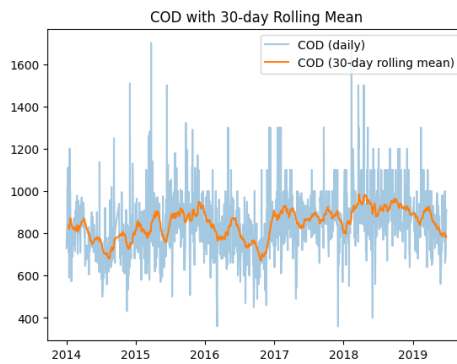


Figure 4. Daily COD concentration time series overlaid with a 30-day rolling mean, highlighting long-term trends while preserving short-term variability in textile wastewater pollution dynamics.

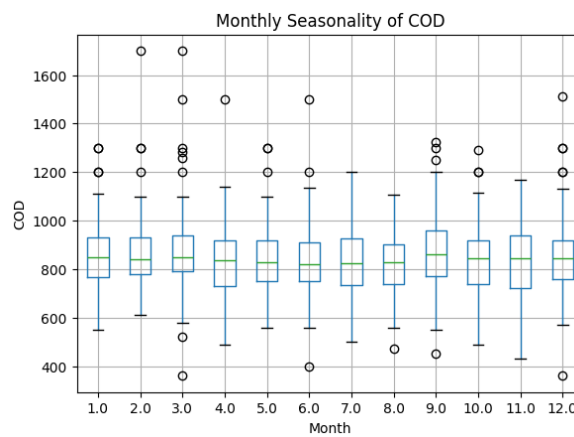


Figure 5. Monthly aggregated COD concentrations illustrating seasonal patterns in textile wastewater pollution, highlighting periodic variability and long-term seasonal trends.

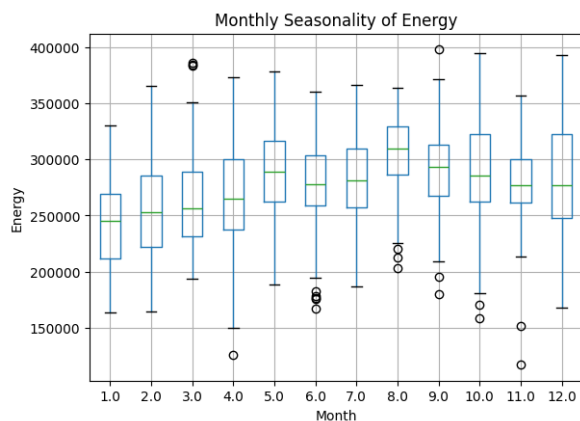


Figure 6. Monthly energy consumption profile of the wastewater treatment process, revealing seasonal fluctuations associated with operational and load variability.

To suppress high-frequency noise and highlight underlying temporal structures, a 30-day rolling mean was applied to COD concentrations, as shown in Fig. 4. This smoothing reveals cyclical patterns otherwise obscured by daily variability.

3.2.3. Seasonal and Periodicity Assessment

Seasonal variability driven by industrial activity and climatic conditions was evaluated using monthly distributions of COD and energy consumption. Figure 5 and 6 illustrates distinct seasonal shifts, particularly in energy demand.

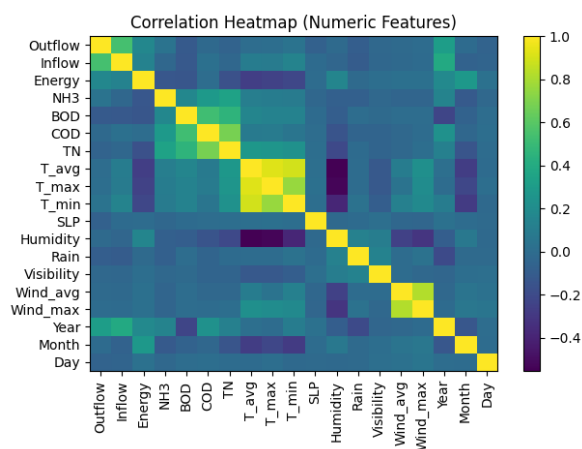


Figure 7. Pearson correlation heatmap illustrating linear relationships among numerical wastewater quality indicators and operational variables. Strong correlations highlight interdependencies between pollutant concentrations and process-related features, providing insight into potential predictive drivers for model development.

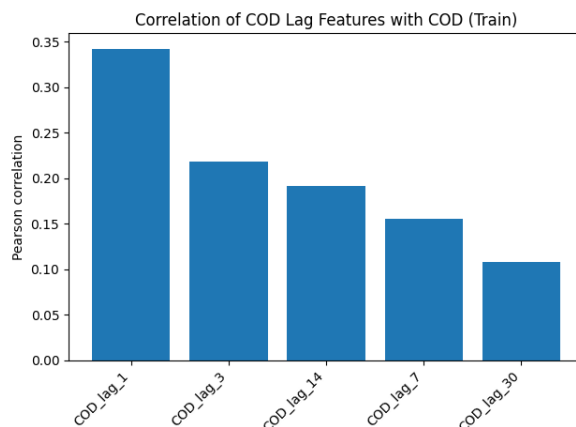


Figure 8. Correlation structure of lagged COD features, illustrating the temporal dependency between current COD levels and past observations. Strong correlations at short lags highlight the autoregressive nature of textile wastewater pollution dynamics.

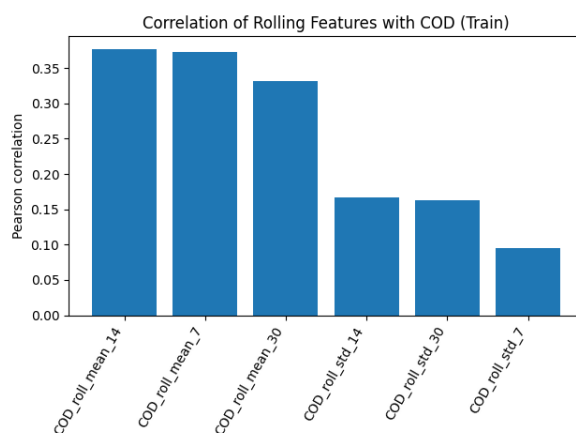


Figure 9. Correlation analysis of rolling-window COD features, demonstrating how aggregated temporal statistics capture medium-term trends and smooth short-term variability in wastewater pollutant concentrations.

3.2.4. Multivariate Correlation Analysis

Linear relationships among wastewater and operational variables were quantified using a Pearson correlation matrix. This analysis aids in diagnosing multicollinearity and identifying dominant drivers influencing effluent quality (Figure 7).

3.2.5. Temporal Feature Engineering: Lags and Rolling Windows

Given the autoregressive nature of wastewater processes, temporal dependencies were explicitly modeled using lagged and rolling statistical features. The correlation structure of these engineered features is illustrated in Figure 8 and 9.

3.2.6. Data Integrity and Gap Diagnostics

To account for sensor downtime and irregular sampling, gap diagnostics were performed. Figure 10 illustrates the distribution of COD values relative to elapsed time since the last valid observation.

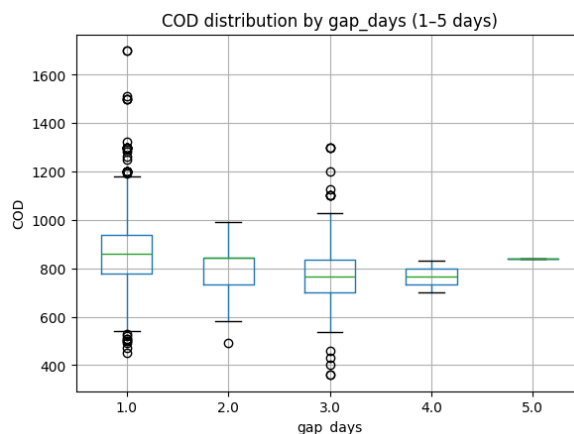


Figure 10. Distribution of COD concentrations stratified by temporal data gaps. The variation across gap categories highlights the impact of missing observations on pollutant value distributions and motivates robust preprocessing and imputation strategies for reliable wastewater forecasting.

3.2.7. Shock Indicator Generation (PSI)

To enhance the model's responsiveness to abrupt system disturbances, a Point Shock Indicator (PSI) was introduced. A shock event is defined when the absolute first-order difference, $|COD_t - COD_{t-1}|$, exceeds the 90th percentile threshold. Figure 11 illustrates the PSI formulation.

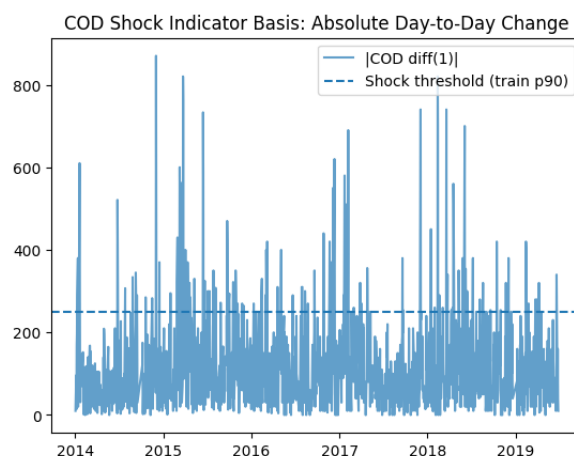


Figure 11. Point Shock Indicator (PSI) derived from daily COD fluctuations, highlighting abrupt changes in pollutant concentration associated with abnormal operating conditions or external disturbances. This indicator forms the basis for identifying shock events and guiding regime-aware model learning.

3.3. Train, Validation and Test Splitting

To preserve temporal consistency and avoid data leakage, the dataset was partitioned using a chronological split into training (70%), validation (15%), and testing (15%) subsets. The split was performed prior to model training and maintained strict time-ordering, ensuring that future observations were never used during training or validation. Table 3 summarizes the sample counts and date ranges for each subset after preprocessing. Figure 12 visually illustrates the chronological train-validation-test partitioning of the COD time series along with its 30-day rolling mean, confirming the absence of temporal overlap between subsets. A strict chronological train-validation-test split was adopted to emulate realistic deployment conditions, where forecasting models are trained on historical observations and applied to unseen future periods. Given the multi-year temporal span of the dataset, this strategy implicitly exposes the model to seasonal variability, operational regime shifts, and gradual data drift, serving as a deployment-oriented approximation of rolling-origin evaluation while avoiding temporal leakage.

Table 3. Chronological train-validation-test split of the wastewater treatment dataset before and after preprocessing.

Subset	Original Samples	Original Features	After Preprocessing	Features	(%)
Training	967	19	937	33	70
Validation	207	19	207	33	15
Testing	208	19	208	33	15
Total	1382	19	1352	33	100.0

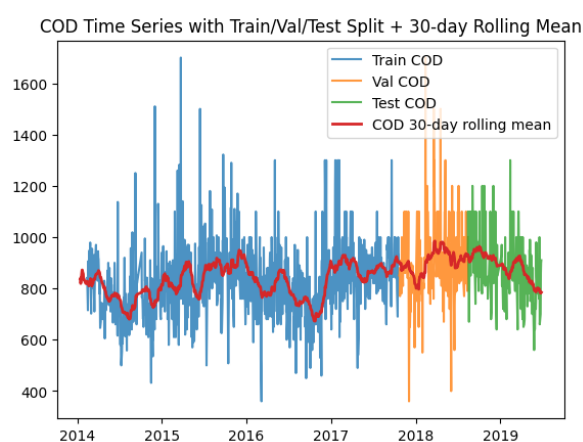


Figure 12. Time-series representation of Chemical Oxygen Demand (COD) with chronological training, validation, and test splits. A 30-day rolling mean is shown to highlight long-term trends while maintaining strict temporal separation between subsets, thereby preventing information leakage in model evaluation.

3.4. Proposed Hybrid Model - SAGE-GBTCN

In this work, SAGE-GBTCN, a novel hybrid model, has been suggested to aim at improving the quality and strength of the chemical oxygen demand (COD) forecast in normal and sudden shocking environments. The fundamental idea of the suggested strategy is to explicitly separate the prediction problem into a global trend learning phase and an adaptive correction phase and, thus, overcome the weaknesses of a single monolithic model to be applied when dealing with highly nonstationary wastewater processes. Figure 13 below represents the layout of the entire architecture of SAGE-GBTCN, and it outlines the entire workflow of input feature building through to the final corrected prediction.

SAGE-GBTCN works in three steps, which are sequential and conceptually different as illustrated in Figure 13. The first step uses a histogram-based gradient boosting model (HGB) as a powerful base predictor to estimate the overall temporal gradient of COD with the help of engineered tabular predictors, such as lagged values, rolling statistics, and environmental-related variables. To do this, rather than using this base prediction directly, or prior to mainstream modeling, the model used explicitly tries to extract the residual signal, which comprises unexplained dynamics and irregular fluctuations, that are often correlated with regime changes or shock events.

The second step is that the residual signal is extracted and then fed to parallel hybrid network that contains a temporal residual learner and shock-aware gating mechanism. The temporal branch is concerned with the modelling of structured temporal dependence within the residual sequence, whereas the gating branch predicts the relevance of shock conditions on the basis of volatility sensitive features. This parallel design allows for freedom of learning about residual magnitude and shock importance, and provides complementary learning. Lastly, during the third step, the learned gate modulates the predicted residual with the aid of the base HGB prediction and integrates with it to generate the final COD forecast.

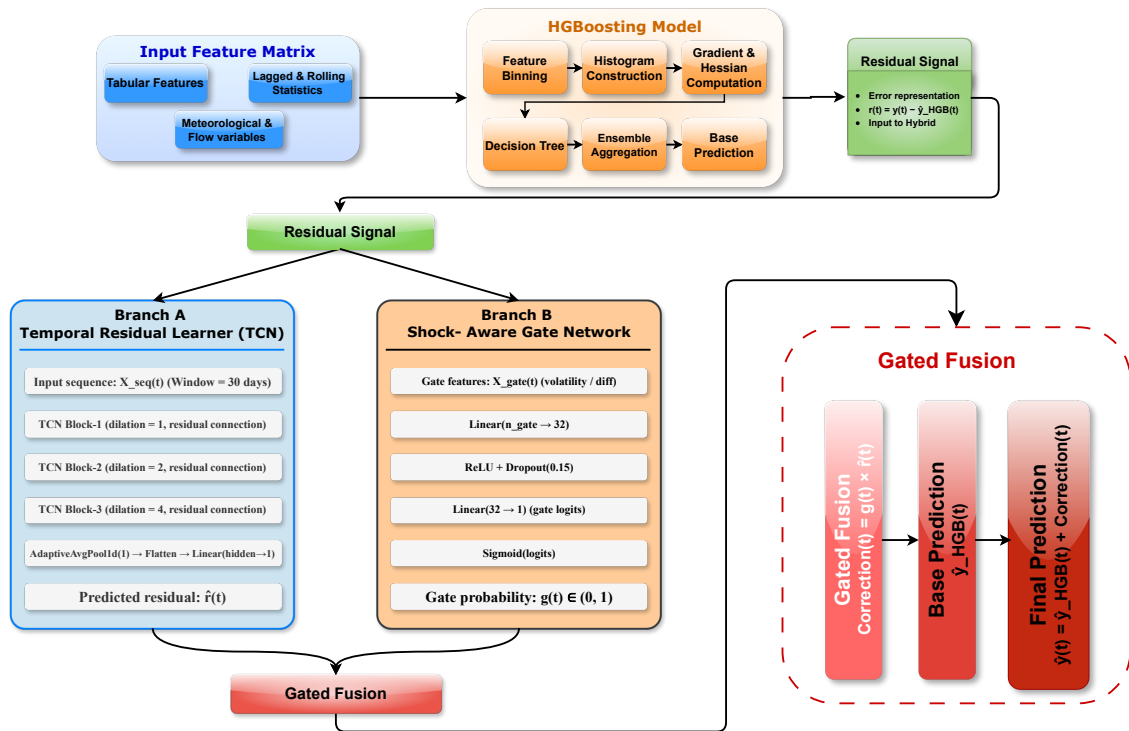


Figure 13. Overall architecture of the proposed hybrid SAGE-GBTCN framework for COD prediction using HGB base forecasting and gated residual correction.

With such a systematic decomposition and gated correction approach, SAGE-GBTCN attains both interpretability and adaptation capability thus being especially applicable in complex wastewater treatment systems that are typified by sudden perturbations and nonlinear time dynamics.

3.4.1. HGB Base Prediction and Residual Construction

In the proposed SAGE-GBTCN model, the modeling process will start with determining a solid base prediction mechanism that can help in the formation of the major time behavior of the target variable. As demonstrated in Figure 14, this part is responsible for transforming the engineered input features into an original COD forecast and thereafter by deriving a residual signal for adaptive correction in the subsequent stages.

Let the input feature representation at time step t be represented by a d dimensional vector consisting of heterogeneous tabular attributes, e.g. historical lagged values, statistical measures for rolling window values and exogenous environmental variables. These attributes are combined together in a structured input vector in form of.

$$\mathbf{x}_t = [x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(d)}], \quad (1)$$

where d denotes the total number of selected features.

In order to simulate the world trend of COD dynamics, a Histogram-based Gradient Boosting (HGB) model is utilized as the basic predictor. HGB creates an ensemble of weak decision trees, solving a loop of minimizing empirical loss by solving a gradient-based optimization process. The model prediction is updated at boosting iteration m as.

$$\hat{y}_t^{(m)} = \hat{y}_t^{(m-1)} + \eta h_m(\mathbf{x}_t), \quad (2)$$

where $h_m(\cdot)$ represents the m -th decision tree and η is the learning rate controlling the contribution of individual learners.

After completing M boosting iterations, the resulting base prediction is obtained as

$$\hat{y}_t^{\text{HGB}} = \sum_{m=1}^M \eta h_m(\mathbf{x}_t), \quad (3)$$

which is effective in capturing smooth and large scale temporal trends in the COD time series. Although this formulation has a good predictive performance when operating conditions are stable, it might fail to capture sharp changes and switching of regimes that are common in the wastewater treatment processes.

To specifically isolate such unexplained dynamics, a residual signal is calculated whereby a discrepancy between the observed COD value y_t and the base prediction is calculated. The remaining value at time t can be defined as

$$r_t = y_t - \hat{y}_t^{\text{HGB}}, \quad (4)$$

Thus, high-frequency variations, anomalies, and shock-induced deviations are captured in the base learner, which are not modelable by the base learner.

To reduce the effect of extreme residual values caused by measurement noise or the occurrence of rare events, the residual signal is again leveled by quantile based clipping. The tipped over leftover can be formulated as

$$\tilde{r}_t = \min(\max(r_t, q_{0.01}), q_{0.99}), \quad (5)$$

where $q_{0.01}$ and $q_{0.99}$ correspond to the lower and upper residual quantiles estimated from the training data.

The stabilised residual sequence \tilde{r}_t gives a smooth approximation of the dynamics that are not understood but gross noise effects are inhibited. The temporal residual learner and shock-conscious gating mechanism then makes use of this residual signal to carry out adaptive correction. The general signal decomposition of the given framework can be outlined as.

$$y_t = \hat{y}_t^{\text{HGB}} + \tilde{r}_t, \quad (6)$$

which directly motivates the hybrid design philosophy of SAGE-GBTCN.

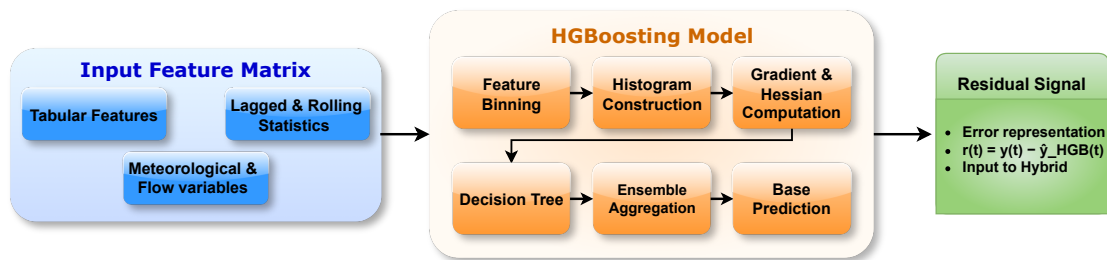


Figure 14. HGB base prediction and residual construction module of the proposed SAGE-GBTCN framework.

3.4.2. Parallel Residual Learning and Shock-aware Gating

Following the building up of the stabilized residual signal, the proposed scheme of SAGE-GBTCN proposes a parallel hybrid learning mechanism that simultaneously models temporal residual dynamics and the shock relevance. As shown in Figure 15, this component is composed of two complementary branches that take advantage of different feature space and allow the base prediction to be corrected adaptively and based on the situation.

The former branch is concerned with the learning of structured temporal dependence in the residual sequence. At a fixed time index, t , a fixed window of the stabilized residual-related feature matrix is tapped to obtain short run and medium run temporal features. This input of sequence is defined as

$$\mathbf{x}_t^{\text{seq}} = [\mathbf{x}_{t-W}, \mathbf{x}_{t-W+1}, \dots, \mathbf{x}_{t-1}], \quad (7)$$

where W is the length of a sequence. In order to capture long-range temporal dependencies in data with computational efficiency, the Temporal Convolutional Network (TCN) using dilated convolutions in the causal order is used. Each TCN block has the form of applying convolutional transformation followed by residual connection which can be represented by

$$\mathbf{H}_l = \mathcal{F}_l(\mathbf{H}_{l-1}) + \mathbf{H}_{l-1}, \quad (8)$$

where $\mathcal{F}_l(\cdot)$ denotes the dilated convolutional operation at layer l . By stacking multiple TCN blocks with increasing dilation factors, the model captures residual dependencies across multiple temporal scales.

The final temporal representation is aggregated by using a global pooling operation and mapped to scalar output and the predicted residual magnitude is

$$\hat{r}_t = f_{\text{TCN}}(\mathbf{X}_t^{\text{seq}}), \quad (9)$$

which estimates the expected correction required at time t .

Simultaneously, the relevance of sudden changes in the system is measured by introducing a shock-conscious gating branch. The volatility-sensitive features operated by this branch are represented as being on a small scale.

$$\mathbf{x}_t^{\text{gate}} = [g_t^{(1)}, g_t^{(2)}, \dots, g_t^{(G)}], \quad (10)$$

where the features are capturing sudden changes, variability and regime instability indicators. A lightweight fully connected network is used to process these inputs to generate a gating score which is further normalised using a sigmoid activation function:

$$\gamma_t = \sigma\left(f_{\text{gate}}(\mathbf{x}_t^{\text{gate}})\right), \quad (11)$$

where $\gamma_t \in (0, 1)$ represents the adaptive gate value reflecting shock intensity.

The time-dependent residual learner and shock-aware gate output is then multiplied to create the gated residual correction, which is given as:

$$\Delta_t = \gamma_t \cdot \hat{r}_t. \quad (12)$$

This construction allows the model to selectively boost or cut off the residual corrections depending upon the inferred shock conditions so that strong residual corrections are used in the volatile regimes but stability in normal operating conditions.

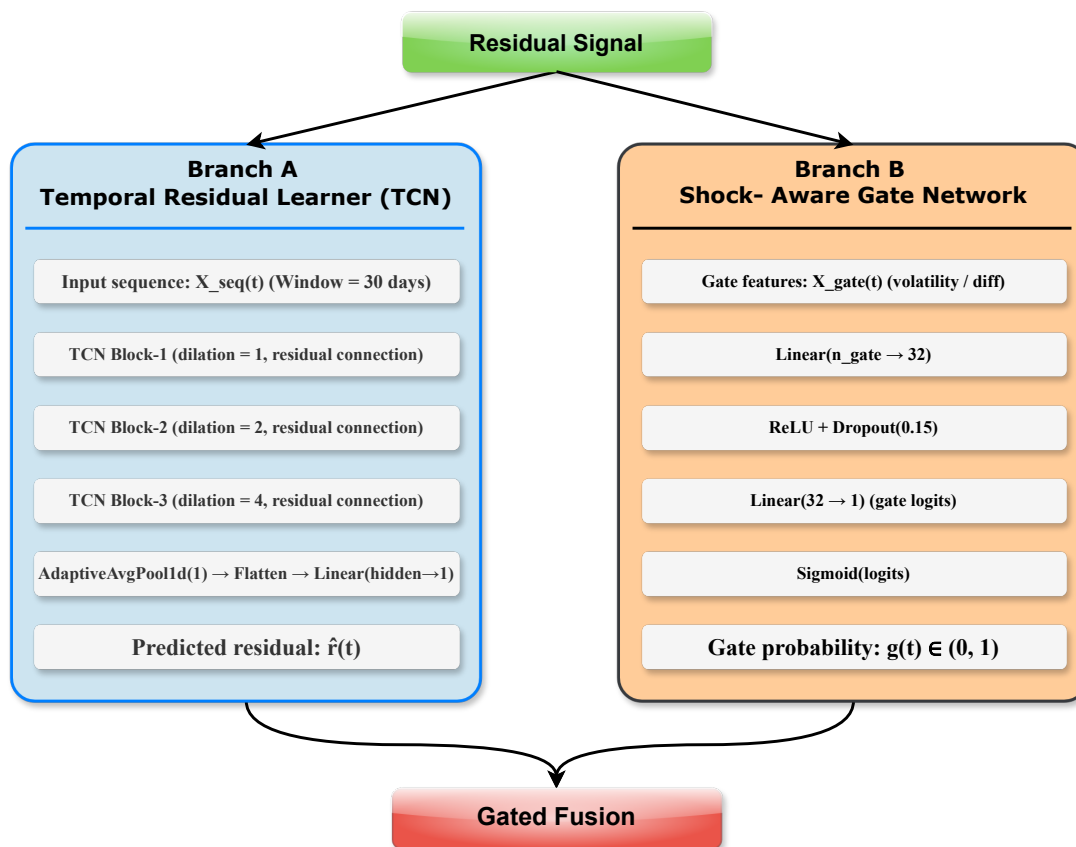


Figure 15. Parallel residual learning and shock-aware gating: TCN residual estimate and gate probability combined via gated fusion to form the correction term.

3.4.3. Final Gated Correction and Output Prediction

Once the magnitude of a residual is estimated and the shock-sensitive gate value is quickly determined, the residual correction with an adaptively modulated one is added to the base prediction to give the final corrected COD forecast in the proposed SAGE-GBTCN framework. The step is a realization of the central hybrid concept of the framework: the gradient boosting model is a good global predictor, and the residual correction learned by the model makes the prediction more precise as per the assumed system state.

Let the base prediction at time t obtained from the gradient boosting component be denoted by \hat{y}_t^{HGB} . From the temporal residual learner, the predicted residual magnitude is \hat{r}_t , while the shock-aware gating network outputs a gate value $\gamma_t \in (0, 1)$. The gated correction term is defined as

$$\Delta_t = \gamma_t \cdot \hat{r}_t, \quad (13)$$

which makes individual adjustments be activated selectively in response to the estimates of the relevance of the shocks.

Have to limit the correction term to avoid being dominated by extreme residual predictions so as to prevent very large training correction terms which will not help performance. The truncated residual forecast is given as

$$\hat{r}_t^c = \min(\max(\hat{r}_t, q_{0.01}), q_{0.99}), \quad (14)$$

where $q_{0.01}$ and $q_{0.99}$ are the lower and upper quantile thresholds. Incorporating this constraint into the gated correction yields

$$\Delta_t^c = \gamma_t \cdot \hat{r}_t^c. \quad (15)$$

The final COD prediction is then arrived at by adding gated residual correction to the base prediction:

$$\hat{y}_t = \hat{y}_t^{\text{HGB}} + \Delta_t^c. \quad (16)$$

This additive formulation is transparent to breaking down the forecast into a global component of the forecast and a context-dependent refinement term.

In the view of learning, the correction mechanism is conditioned to reduce a residual regression goal and, at the same time, promote the accurate estimation of gates. In the case of the residual branch, weighted mean-squared error is used to focus on shock samples:

$$\mathcal{L}_{\text{res}} = \mathbb{E} \left[w_t (\hat{r}_t - r_t)^2 \right], \quad (17)$$

where w_t assigns larger weights to shock-associated instances. For the gate branch, a binary cross-entropy objective is used to align gate logits with the shock indicator:

$$\mathcal{L}_{\text{gate}} = -\mathbb{E} [\psi_t \log(\gamma_t) + (1 - \psi_t) \log(1 - \gamma_t)]. \quad (18)$$

The combination of both terms in the overall objective allows the learning of the magnitude of correction and relevance of shock together in a final prediction which is robust to the stable regimes and responsive to abrupt disturbance.

Table 4. Algorithmic workflow of the proposed SAGE-GBTCN model for COD prediction.

Step	Procedure
1	Input: engineered tabular features $\mathbf{x}_t \in \mathbb{R}^d$, target COD y_t , and a time index set for train/validation/test.
2	Base predictor (HGB): train a histogram-based gradient boosting model on the training set and generate base predictions \hat{y}_t^{HGB} .
3	Residual construction: compute residual labels $r_t = y_t - \hat{y}_t^{\text{HGB}}$ and stabilize residuals using train-based quantile clipping $\tilde{r}_t = \text{clip}(r_t; q_{0.01}, q_{0.99})$.
4	Shock indicator (PSI): compute $\Delta y_t = y_t - y_{t-1} $ and define a train-derived shock threshold $\tau = Q_{0.90}(\Delta y)$; obtain shock labels $\psi_t = \mathbb{I}(\Delta y_t > \tau)$.
5	Sequence construction: for each valid t , form a windowed sequence $\mathbf{X}_t^{\text{seq}} \in \mathbb{R}^{W \times F}$ from selected sequential features over the last W timesteps.
6	Gate features: form a compact gate vector $\mathbf{x}_t^{\text{gate}} \in \mathbb{R}^G$ from volatility/diff indicators (e.g., rolling std, abs diff).
7	Residual learner (TCN): predict residual magnitude $\hat{r}_t = f_{\text{TCN}}(\mathbf{X}_t^{\text{seq}})$ using stacked dilated TCN blocks.
8	Shock-aware gate: estimate gate probability $\gamma_t = \sigma(f_{\text{gate}}(\mathbf{x}_t^{\text{gate}}))$ using a lightweight MLP.
9	Gated correction and final forecast: apply gated correction $\Delta_t = \gamma_t \cdot \text{clip}(\hat{r}_t; q_{0.01}, q_{0.99})$ and obtain the final prediction $\hat{y}_t = \hat{y}_t^{\text{HGB}} + \Delta_t$.
10	Training objective: optimize a joint loss $\mathcal{L} = \mathcal{L}_{\text{res}} + \lambda \mathcal{L}_{\text{gate}}$, where \mathcal{L}_{res} is a shock-weighted MSE on residuals and $\mathcal{L}_{\text{gate}}$ is BCE on shock labels.

This section presented the SAGE-GBTCN which is a structured hybrid modeling framework aimed at solving the problem of COD forecast in extremely dynamic wastewater systems. The suggested method breaks the forecasting problem down into complementary elements in order to balance global trend learning and local adaptive correction. First, a histogram-based gradient boosting model is used in order to derive major temporal trends and produce an effective base prediction. The rest of the unexplained dynamics are directly separated as residual signals which are then further refined by a parallel learning process. Temporal residual dependencies are formulated with help of a dilated convolutional architecture while a shock-aware gating network adjusts the impact of the residual corrections according to the volatility-sensitive indicators. The end result is the final prediction which

is the combination of the base forecast and a gated residual correction which is stabilizing when the situation is normal and responsive when the situation is suddenly disrupted. This interpolable and modular design will enable the SAGE-GBTCN to combine statistical learning and deep temporal modeling to offer a strong and flexible solution to the complicated forecasting of wastewater process.

Table 5. Hyperparameter and training configuration of SAGE-GBTCN (code-aligned settings).

Category	Setting	Value / Description
Target	Variable	COD
Sequence length	Window size W	30 timesteps
Base model	Predictor	Histogram-based Gradient Boosting (HGB)
Base output	Forecast	$\hat{y}_t^{\text{HGB}} \in \mathbb{R}$
Residual label	Definition	$r_t = y_t - \hat{y}_t^{\text{HGB}}$
Residual stabilization	Clipping	Train quantiles: $q_{0.01}$ and $q_{0.99}$
Shock labeling	Threshold quantile	$\tau = Q_{0.90}(\Delta y)$ (train-based)
Shock indicator	PSI label	$\psi_t = \mathbb{I}(\Delta y_t > \tau)$
TCN residual learner	Architecture	3 TCN blocks with dilation $\{1, 2, 4\}$
TCN hidden size	Channels	48
TCN kernel size	k	3
TCN regularization	Dropout	0.15 in TCN blocks
TCN output head	Aggregation	AdaptiveAvgPool1d(1) \rightarrow Flatten \rightarrow Linear(48 \rightarrow 1)
Gate network	MLP structure	Linear($G \rightarrow 32$) + ReLU + Dropout(0.15) + Linear(32 \rightarrow 1 logits)
Gate activation	Probability	$\gamma_t = \sigma(\text{logits})$
Optimization	Optimizer	AdamW
Learning rate	η	1×10^{-3}
Weight decay	L2	1×10^{-4}
Gradient clipping	Max norm	1.0
Loss design	Residual loss	shock-weighted MSE (weight: SHOCK_W=5.0; BASE_W=1.0)
Loss design	Gate loss	BCEWithLogitsLoss
Loss weighting	Gate weight λ	GATE_LOSS_W = 1.0
Batching	Batch size	Train: 64; Validation/Test: 128
Early stopping	Strategy	Minimum epochs = 40; patience = 20; improvement delta = 10^{-3}
Epoch budget	Max epochs	140
Model selection	Criterion	Best validation residual RMSE
Output fusion	Final prediction	$\hat{y}_t = \hat{y}_t^{\text{HGB}} + \gamma_t \cdot \text{clip}(\hat{r}_t; q_{0.01}, q_{0.99})$

3.5. Baseline Models

To determine the credible reference performance and be able to make consistent comparisons with the advanced learning methods, some popular regression models were used as they served as a baseline. These models are various paradigms of learning and some of the commonly used models are ensemble learning, kernel-based learning, neural networks, and gradient boosting and they have been widely tested in environmental and time-series prediction tasks.

3.5.1. Random Forest Regressor

Random Forest (RF) is an ensemble learning method that constructs multiple decision trees during training and outputs the mean prediction of individual trees for regression tasks ([47]). The algorithm introduces randomness through bootstrap sampling of training data and random feature selection at each split.

For a Random Forest with T trees, the prediction for input \mathbf{x} is:

$$\hat{f}_{\text{RF}}(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^T h_t(\mathbf{x}) \quad (1)$$

where $h_t(\mathbf{x})$ is the prediction of the t -th tree. Each tree h_t is trained on a bootstrap sample \mathcal{D}_t from the original training data \mathcal{D} , and at each split, a random subset of m features from the total p features is considered for splitting.

3.5.2. CatBoost Regressor

CatBoost is a gradient boosting algorithm specifically designed to handle categorical features efficiently and combat prediction shift through ordered boosting ([48]). The algorithm minimizes the loss function \mathcal{L} using gradient boosting with ordered boosting scheme.

The model at iteration m is:

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \eta \cdot h_m(\mathbf{x}) \quad (2)$$

where $h_m(\mathbf{x})$ is the decision tree at iteration m , and η is the learning rate. For categorical features, CatBoost uses ordered target statistics:

$$\hat{x}_k^i = \frac{\sum_{j=1}^{p-1} [x_{\sigma_j, k} = x_{\sigma_p, k}] \cdot Y_{\sigma_j} + a \cdot P}{\sum_{j=1}^{p-1} [x_{\sigma_j, k} = x_{\sigma_p, k}] + a} \quad (3)$$

where σ is a random permutation, a is a prior, and P is the target average.

3.5.3. Support Vector Regressor (SVR)

Support Vector Regression applies the principles of Support Vector Machines to regression problems by finding a function that deviates from actual values by no more than ϵ while remaining as flat as possible ([49]). SVR solves the optimization problem:

$$\min_{\mathbf{w}, b, \zeta_i, \zeta_i^*} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\zeta_i + \zeta_i^*) \quad (4)$$

subject to:

$$y_i - \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle - b \leq \epsilon + \zeta_i \quad (5)$$

$$\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle + b - y_i \leq \epsilon + \zeta_i^* \quad (6)$$

$$\zeta_i, \zeta_i^* \geq 0 \quad (7)$$

where $\phi(\mathbf{x})$ is the feature mapping, ζ_i, ζ_i^* are slack variables, and C controls the trade-off between margin width and error tolerance. Using the kernel trick, the solution becomes:

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \quad (8)$$

where $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ is the kernel function.

3.5.4. Multi-Layer Perceptron (MLP) Regressor

MLP is a class of feedforward artificial neural network that learns non-linear relationships through backpropagation and gradient descent optimization ([50]). For a network with L layers, the forward propagation for layer l is:

$$\mathbf{a}^{(l)} = f^{(l)}(\mathbf{W}^{(l)} \mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}) \quad (9)$$

where $\mathbf{a}^{(l)}$ is the activation at layer l , $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ are weights and biases, and $f^{(l)}$ is the activation function. The output for regression is:

$$\hat{y} = \mathbf{a}^{(L)} \quad (10)$$

The network is trained by minimizing the loss function $\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}})$ using gradient descent:

$$\mathbf{W}^{(l)} \leftarrow \mathbf{W}^{(l)} - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(l)}}, \quad \mathbf{b}^{(l)} \leftarrow \mathbf{b}^{(l)} - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{b}^{(l)}} \quad (11)$$

where η is the learning rate.

3.5.5. Histogram-Based Gradient Boosting Regressor

HistGradientBoosting is an optimized gradient boosting implementation that uses histogram-based algorithms for efficiency ([51]). The algorithm bins continuous features into discrete histograms, reducing computational complexity from $O(n \cdot f)$ to $O(b \cdot f)$, where n is the number of samples, f is the number of features, and b is the number of bins ($b \ll n$).

The gradient boosting update with histogram approximation is:

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \eta \cdot h_m(\mathbf{x}; \Theta_m) \quad (12)$$

where Θ_m represents the histogram-based split points. For each feature, the histogram H_k with b bins is constructed:

$$H_k = \{(b_1, G_1, H_1), (b_2, G_2, H_2), \dots, (b_b, G_b, H_b)\} \quad (13)$$

where $G_b = \sum_{i \in \text{bin}_b} g_i$ and $H_b = \sum_{i \in \text{bin}_b} h_i$ are the sums of gradients and Hessians within each bin.

4. Results and Discussion

4.1. Experimental Setup

All experiments were conducted using Google Colab Pro, an online cloud-based platform providing access to high-performance computational resources, including GPU acceleration, which is essential for efficiently training deep learning models on multivariate time-series data. The entire implementation was carried out in Python, ensuring reproducibility and flexibility. The deep learning components, including the Temporal Convolutional Network (TCN) and gated hybrid architecture, were implemented using PyTorch, while Scikit-learn was used for data preprocessing, evaluation metrics, and auxiliary utilities. In addition, a tree-based gradient boosting model (Histogram-based Gradient Boosting) was employed as the baseline forecasting model, enabling a robust hybrid learning framework that combines traditional machine learning with deep neural networks.

The dataset was chronologically divided into training, validation, and testing sets using a 70%–15%–15% split, ensuring that future observations were never used during training, thus preventing data leakage. This temporal split reflects real-world deployment conditions in wastewater monitoring systems, where predictions must be made strictly on unseen future data. The validation set was used exclusively for hyperparameter tuning and early stopping, while the test set was reserved for final performance evaluation.

A consistent and robust preprocessing pipeline was applied across all data splits. Continuous input variables were standardised to have zero mean and unit variance, facilitating stable and efficient neural network training. Time-series inputs were transformed into fixed-length sliding windows, allowing the model to capture temporal dependencies and delayed effects commonly observed in wastewater dynamics. In addition to sequential inputs, a separate set of gating features was constructed to provide contextual information relevant to regime changes and abnormal pollution events.

To explicitly model rare but critical events, a binary shock indicator was introduced to label abnormal pollution regimes, such as sudden discharge events or system disturbances. Since such shock events occur infrequently but have high environmental impact, a weighted loss strategy was adopted during training. Residual prediction errors corresponding to shock periods were penalised more

heavily than those from normal conditions, encouraging the model to focus on high-risk scenarios without degrading overall stability.

Model training was performed using the AdamW optimiser with gradient clipping to ensure numerical stability. Early stopping was applied based on validation residual RMSE, with a minimum number of training epochs enforced to allow sufficient learning of both the residual correction network and the shock-gating mechanism. This training strategy reduced overfitting while ensuring robust convergence.

Overall, this experimental setup provides a realistic, challenging, and deployment-oriented evaluation framework for wastewater pollution forecasting. By combining chronological data splitting, regime-aware learning, and hybrid modelling, the proposed approach is well-suited for sustainable environmental monitoring and decision-support systems.

4.2. Comparative Performance Evaluation of COD Prediction Models

The sub-section contains a comparative analysis of different machine learning models to estimate chemical oxygen demand (COD) in an industrial wastewater treatment problem analysis. This analysis is aimed at testing the predictive performance, stability, and extrapolation ability of common baseline models against prediction efficiency of the given hybrid framework of SAGE-GBTCN. In order to evaluate a more objective and comprehensive comparison of all the models, they were all trained and tested on the data according to the same parameters in data handling, feature engineering and validation. Model performance is measured against widely accepted performance regression measures including root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and coefficient of determination (R^2). Through this assessment, one is able to obtain a clue on the weaknesses and strength of various modeling paradigms in case they are employed to handle dynamic and non-linear data focusing on wastewater pollution.

Table 6. Comparative performance of baseline models and the proposed SAGE-GBTCN framework for COD prediction (lower RMSE/MAE/MAPE and higher R^2 indicate better performance).

Model	RMSE	MAE	MAPE (%)	R^2
HistGradientBoosting (HistGB)	33.973	25.277	2.883	0.927
CatBoost	37.172	25.108	2.722	0.913
Random Forest	53.581	39.708	4.506	0.819
Support Vector Regression (SVR)	58.346	41.256	4.524	0.785
Multilayer Perceptron (MLP)	61.361	47.219	5.586	0.762
Proposed SAGE-GBTCN	30.303	22.536	2.603	0.942

Table 6 presents the comparison of the predictive power of the predictive models based on the baseline machine learning models and the proposed SAGE-GBTCN algorithm of COD predictions. The analysis is done based on a number of error-based and goodness-of-fit measures such as RMSE, MAE, MAPE and the R^2 which are taken to provide a moderate estimate and assessment of error and effectiveness, as well as strong performance under dynamic wastewater settings.

HistGradientBoosting (HistGB) is the best baseline model with an RMSE of 33.97 and an R^2 of 0.927. This result suggests that the use of ensemble models with trees can be significant to model non-linear organizational connections between procedures variable of wastewater action and the degree of concentration of COD. CatBoost also places relatively high in error values with small disparity but similar generalization performance implying two things: it is strong to feature interaction and noise that is usually strong in industrial wastewater data. Random Forest in comparison exhibits much greater prediction errors and reduced explanatory power; i.e. fewer changes to the dynamics of time and changes in pollutants. Support Vector Regression (SVR) and Multilayer Perceptron (MLP) models follow a kind of poor result; the RMSE, MAE are significantly greater, and the R^2 is lower. Such findings suggest that both the kernel-based and deep learning methods were not good at generalizing under condition of the data limits and non-stationary scenarios when processing wastewater to reduce

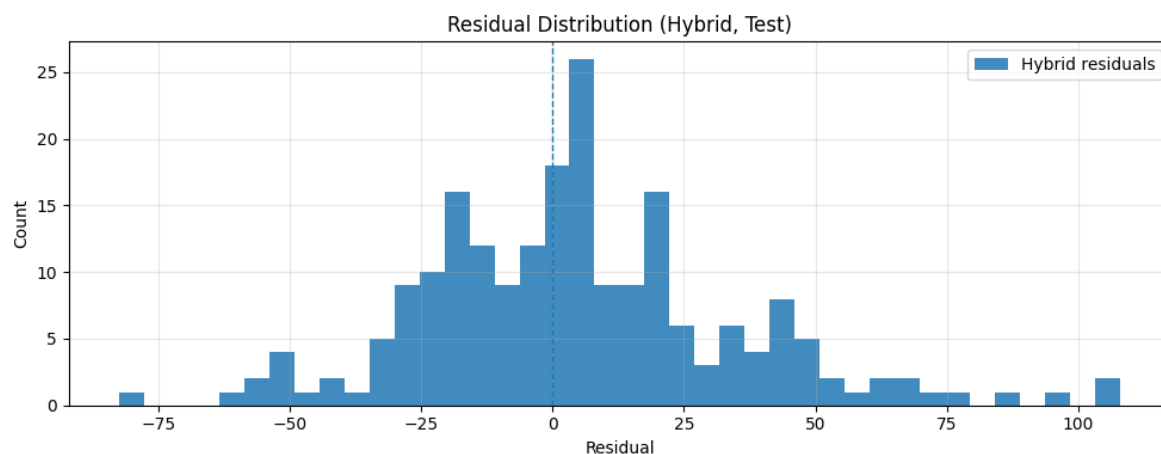


Figure 16. Distribution of prediction residuals for the proposed hybrid model on the test dataset. The near-zero centring indicates minimal systematic bias, while heavier tails correspond to rare extreme pollution events.

pollution, the variation of pollutants will quickly increase, and the regimes of operation are likely to transform. This finding reinforces the point that even a complex model structure may not result in an improved model predictive control of the actual wastewater system.

Interestingly, the proposed SAGE-GBTCN model outperforms all the baseline models by all the evaluation measures; the RMSE and MAE are the lowest, that is, 30.30 and 22.54, respectively, whereas the highest value of the R^2 , 0.942 and MAPE of 2.60%. The performance increase is explainable by the hybrid nature of SAGE-GBTCN, including the good forecasting capability of the gradient boosting and high-residual capabilities of the temporal prediction to address the pollutant shocks and short-term dynamics more effectively. These findings demonstrate that through integrating this process conscious temporal intelligence into the system, the resilience to make predictions about the framework, and at the same time not compromising the computational efficiency, which makes this proposed framework particularly applicable in industrial and textile wastewater treatment settings.

4.3. Residual Distribution and Statistical Normality

The residual distribution and the distribution of normality should be evaluated to comprehend the statistical credibility of the proposed hybrid model. The residual histogram ensures that the model has no systematic bias since the prediction errors are found to be centred around zero. The close-to-symmetric distribution proves that the hybrid residual correction mechanism is very efficient in the process of the structural errors that statues of the hybrid predictors propagate.

In order to explore some more, a Q-Q plot of residues is studied. Most of the residual values are close to the theoretical normal reference line especially at the central quantile range. This implies that the residuals under normal operating conditions follow the generic of Gaussian behaviour. Also the linear deviation in the tail areas is mostly related to exceptional and outliers of the pollution which inject heavier tailed features in the error.

This behaviour of tail is normal in the wastewater systems of the real world and does not invalidate the model. Rather, it is a manifestation of inherent uncertainty when it comes to abnormal regimes, which maximizes the significance of shock-conscious consideration.

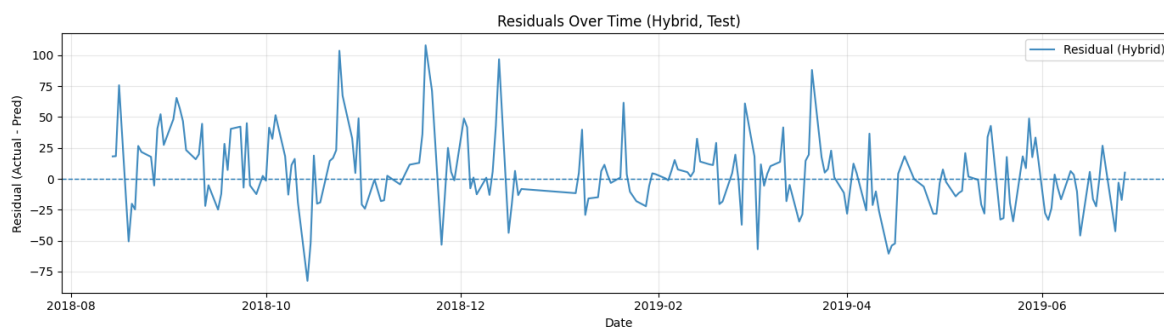


Figure 18. Temporal evolution of prediction residuals on the test set. Residuals oscillate around zero without persistent drift, indicating stable generalisation over time.

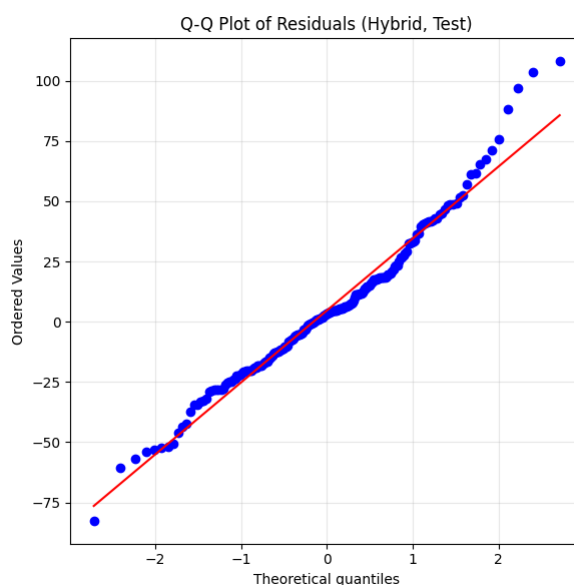


Figure 17. Q-Q plot of residuals on the test set. Close alignment with the theoretical normal line in the central region indicates approximate normality, while tail deviations reflect the influence of rare shock events.

4.4. Temporal and Regime-Wise Error Characteristics

Residual behaviour is studied in time and across operating regimes to provide an analysis of when and under what conditions errors in prediction are possible. The temporal plot of the residual indicates that there were no accumulation or drifts of errors over the test period. Such means that the generalisation through the seasonal and operational changes is stable in the proposed hybrid model.

The instantaneous residual spikes are seen where there is the change in pollution, which is associated with a shock. These deviations are stabilised by rapid mechanisms implying that migrates of the gated residual mechanism stabilise fast without creating instability in the long run.

This behaviour is also pointed out by further comparison of absolute errors regime wise. In a normal working day, the model has a somewhat small median absolute error with a rather narrow interquartile range. Shock days, on the other hand, have larger median errors and more dispersion indicating the inherent inability to predict extreme events. Critically, the magnitude of the errors in periods of shock are limited which explains why even in unfavorable circumstances the model does not result in catastrophic breakdowns.

Collectively, the above findings show that the proposed framework is able to balance its affectedness to time stability and its resilience to abnormal regimes.

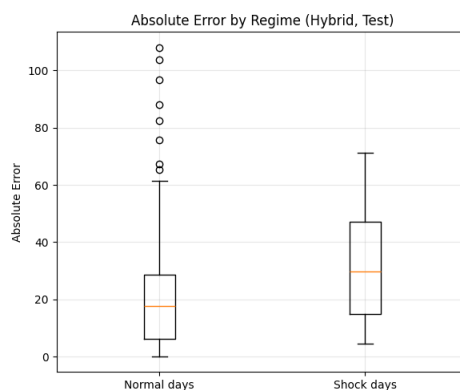


Figure 19. Comparison of absolute prediction errors across normal and shock regimes on the test set. Increased dispersion during shock days reflects higher uncertainty under abnormal conditions, while bounded error ranges indicate robust behaviour.

4.5. Predicted-Actual Agreement and Bias Analysis of COD Forecasting Models

This subsection allows a visual evaluation of the agreement between the predicted values and observed values of chemical oxygen demand (COD) for the test data set for the baseline models and the proposed SAGE-GBTCN framework. Predicted–actual scatter plots are used to analyze model behavior over the range of COD values, and attention is paid to systematic bias, dispersion plots, and deviations from the ideal $y = x$ reference line. Such visual diagnostics help to complement numerical diagnostics of performance, by showing tendencies to over- or under-estimate that may not be completely recorded by aggregate measures of error. The analysis is done by focusing on the ability to predict new data (test set), which is important to assess generalization ability, i.e., the ability to generalize under previously unseen conditions, which will be important in the case of real-world wastewater treatment applications. This comparison allows for a better understanding of the modeling approaches with respect to non-linearity in variability and dynamics of pollutants in industrial wastewater systems.

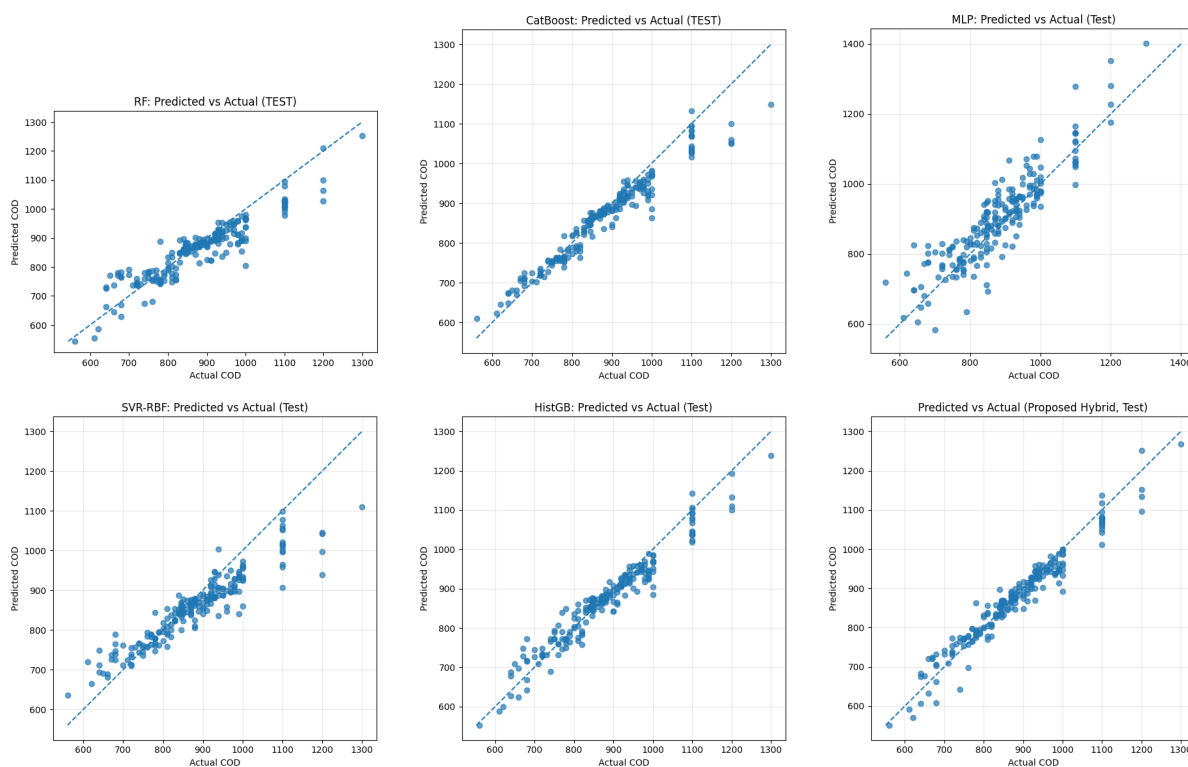


Table 7. Predicted versus actual COD values on the test set for baseline models (RF, CatBoost, MLP, SVR-RBF, and HistGB) and the proposed SAGE-GBTCN framework. Points closer to the $y = x$ line indicate improved agreement and reduced prediction bias across the COD range.

Figure 7 shows the correlation between predicted and observed values of the chemical oxygen demand (COD) in the test data group for the five baseline models (Random Forest (RF), CatBoost, Multilayer Perceptron (MLP), Support Vector Regression with the RBF kernel (SVR-RBF), HistGradientBoosting (HistGB), and our proposed SAGE-GBTCN framework. The predicted-actual scatter plots are one way of providing an intuitive visual assessment of model agreement, dispersion, and systematic bias over the full range of COD, with the dashed $y = x$ line representing ideal prediction behavior.

Among the baseline models, RF has an evident dispersion around the reference line, especially for higher COD values: this result suggests a greater variance and a tendency to underestimate in conditions of elevated pollutants. CatBoost shows better alignment with the $y = x$ line at most loads of the COD range, although some moderate deviations are observed when the instance is of high load, which could mean sensitivity to abrupt fluctuations of the pollutant. The MLP model exhibits relatively greater scatter and irregular spread, indicating a decrease in stability and poor generalization when non-linear and non-stationary wastewater dynamics are considered. SVR-RBF shows a better clustering than RF and MLP in the middle range of clearly defined COD levels; however, apparent bias is evident at the higher concentrations, where the predictions have a systematic deviation from the observed values. This behavior is used to point out the limitations in capturing sharp regime changes using kernel-based approaches. HistGB produces a consistent and increased agreement with lowered dispersion over low and moderate COD ranges and highlights the effectiveness of ensemble-based tree models in dealing with non-linear interactions between wastewater process variables.

Notably, the interesting fact is that the proposed SAGE-GBTCN framework shows the least deviation from the reference line of $y = x$ in the whole COD spectrum. The predicted values exhibit lower scatter as well as minimal systematic bias, especially at higher COD conditions in which baseline models are prone to divergence. This improved agreement can be traced back to the hybrid design

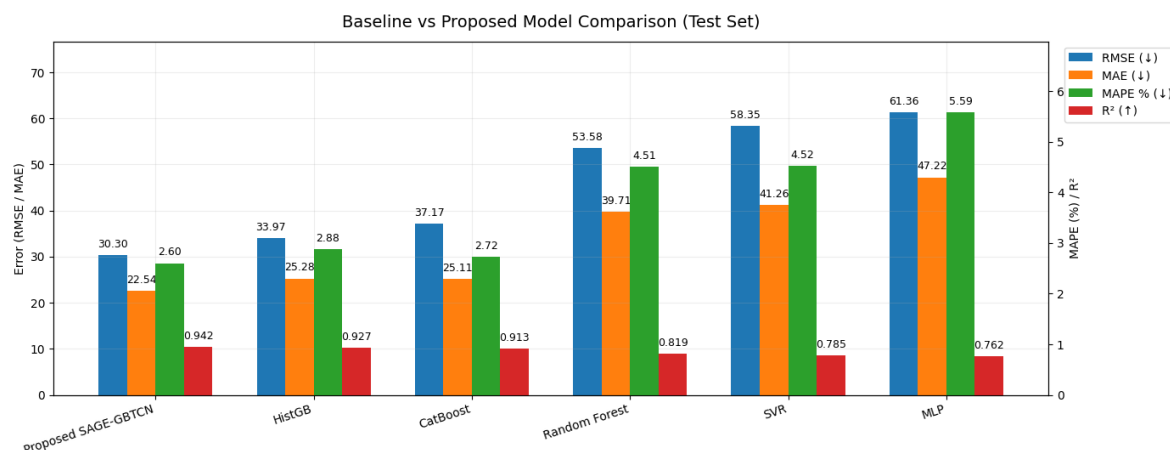


Figure 20. Quantitative comparison between baseline models and the proposed hybrid framework on the test set using RMSE, MAE, MAPE, and R^2 metrics. Lower error values and higher R^2 indicate superior performance.

of SAGE-GBTCN that combines good ensemble-based baseline prediction and temporal residual correction to better deal with short-term dynamics and pollutant shock events. Overall, the quantitative performance metrics, show that the proposed framework can achieve better generalization and robustness in COD forecasting in dynamic industrial wastewater treatment scenarios. The proposed SAGE-GBTCN framework is formulated as a deterministic, regime-aware regression model rather than a probabilistic exceedance predictor or binary early-warning classifier. Consequently, the focus of evaluation is placed on forecast robustness and accuracy under shock-induced regime changes, rather than precision–recall analysis or calibration of exceedance probabilities. The proposed framework exhibits a modular prediction behavior that can be interpreted at three complementary levels. The gradient boosting base predictor primarily captures global COD trends and slowly varying process dynamics under stable operating conditions. The temporal convolutional residual learner focuses on structured short-term deviations that arise from localized process disturbances and transient operational fluctuations. The shock-aware gating mechanism dynamically regulates the influence of residual correction by attenuating its contribution during stable regimes and amplifying adaptive correction during periods of elevated volatility, as indicated by the Point Shock Indicator. This hierarchical interaction enables robust forecast stabilization without overreaction to noise.

4.6. Comparison Between Baseline Models and the Proposed Hybrid Framework

To conduct a comprehensive study of the performance of the proposed hybrid model, its performance is assessed in relation to a number of commonly used base machine learning algorithms, such as HistGB, CatBoost, Random Forest, Support Vector Regression (SVR), and Multilayer Perceptron (MLP). Using a variety of complementary performance measures, the comparison on the test data is to be done, which will guarantee the fair and robust assessment.

As shown in the bar-chart comparison in Figure 20, the quantitative performance of the proposed hybrid framework has evidently improved. The proposed model also has the lowest values of error-based measures, including RMSE, MAE, and MAPE, among the rest of the compared methods, which implies that the model is better at predicting. Simultaneously, it has the largest coefficient of determination (R^2), demonstrating a greater explanatory capacity and closer correspondence to the observed COD dynamics. Conversely, classical machine learning models progressively increase the magnitude of the errors, especially in the occurrence of nonlinear and highly-complicated patterns of pollution.

A radar-based visualisation is also used to conduct a holistic and scale-free comparison, with each of the metrics normalised through a ratio-based scoring system (the best model score is one, higher

scores indicate better performance). This representation allows the assessment of several metrics simultaneously, and one should not be biased towards a specific criterion.

The radar plot in the appendix of Figure 21 unequivocally indicates that the proposed hybrid model always prevails in all computed dimensions, as the outermost curve of the radar chart. Baseline models have skewed performance and they demonstrate superior performance in specific metrics and poorly in others. This contradiction highlights the well-balanced and healthy nature of the offered framework that combines the benefits of gradient-based learning and temporal residual correction in the single framework.

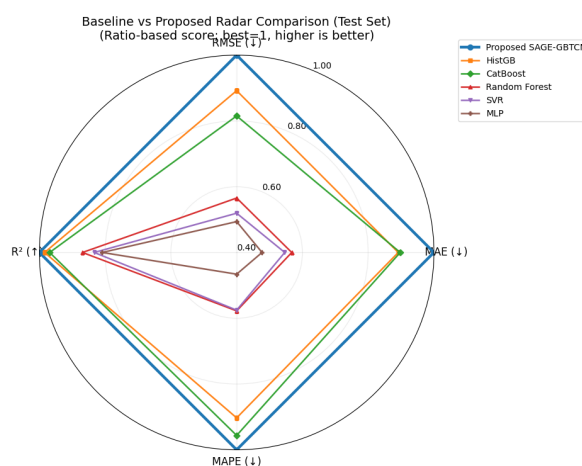


Figure 21. Radar-based comparison of baseline and proposed models on the test set using normalised performance scores. The proposed hybrid model consistently achieves the highest scores across all metrics, indicating balanced and superior performance.

Comprehensively, both the quantitative and visual results demonstrate that the proposed hybrid model would outperform standard baseline models in terms of accuracy and robustness significantly. The steady achievements in the various measures used demonstrate the aptness of the model to real-waste water pollution prediction where a stable working in the regular and unusual circumstances is highly needed.

4.7. Ablation Study and Component-Wise Contribution Analysis

A quantitative analysis of the contribution made by every component in the proposed hybrid scheme is carried out through an ablation study applied to the test dataset. Four variants of model have been tested: (i) the baseline model, HistGB, (ii) Augmented Residual Network HistGB + TCN Residual, (iii) Gated Residual Correction Model HistGB, and (iv) Full Proposed SAGE-GBTCN. This step-by-step evaluation can allow a satisfactory evaluation of the way each architectural element enhances the predictive functionality.

The quantitative analysis employing real metric values illustrates that there is an increasing performance augmentation with the incorporation of more components. Learning a TCN-based residual correction has lower RMSE and MAE than the baseline model, which demonstrates that Temporal learned residuals provide a more quantitative representation of the systematic errors that not even HistGB are able to predict. Additional enhancement of the gating mechanism also increases robustness, via selective activation of residual corrections, resulting in better performance in a complicated operating environment. The complete proposed SAGE-GBTCN model has the smallest error values and the biggest coefficient of determination, which proves the additional positive effect of residual learning and shock-conscious gating.

To compare all the main performance measures in an unbiased and holistic manner and to alleviate the scale effect, a radar-based visualisation is used with ratio-normalised scores, with the highest performing model in each performance measure receiving a one score. Such representation points to

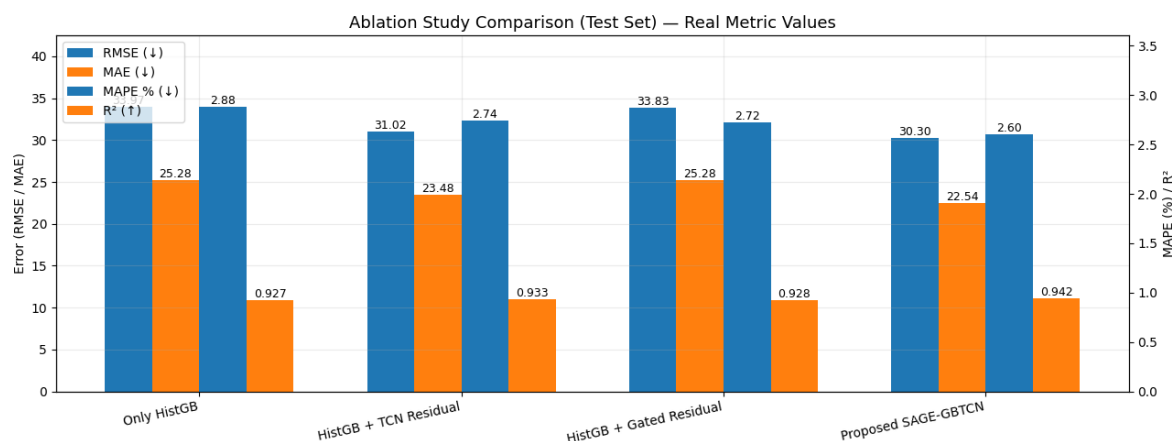


Figure 22. Ablation study comparison on the test set using RMSE, MAE, MAPE, and R^2 metrics. Progressive performance improvements are observed as residual learning and gating mechanisms are incorporated, with the proposed SAGE-GBTCN model achieving the best overall results.

the strengths and weaknesses of each model variation in comparison, but does not favor either of the metrics.

The radar chart makes it obvious that the suggested scheme of SAGE-GBTCN dominates in all the considered dimensions, which creates the outermost envelope of the comparison. Conversely, partial variations present equal opportunity performance gains, thus performing low in certain measures and high in others. This finding supports the importance of a collective incorporation of the temporal residual learning and gated correction to the technologies in order to attain a balanced and vigorous performance in wastewater pollution prediction.

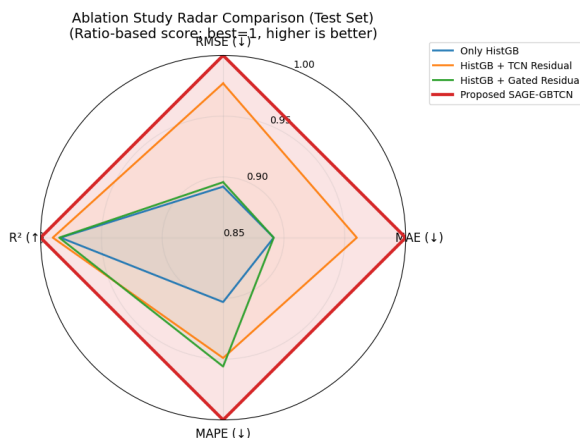


Figure 23. Radar-based ablation study comparison on the test set using ratio-normalised performance scores. The proposed SAGE-GBTCN model consistently achieves the highest scores across all metrics, indicating balanced and superior performance.

4.8. Model Interpretability Using SHAP Analysis

In order to make the proposed SAGE-GBTCN framework more transparent and interpretable, SHapley Additive exPlanations (SHAP) are utilized to measure the contribution of individual input features to the model prediction. SHAP is based on cooperative game theory and offers the consistent and additive feature attributions, which is why it is especially applicable to explaining complex hybrid architectures. The purpose of the current analysis is to determine the prevailing forces behind the COD forecasting and to confirm the existence of the learned relationships as compared to the known facts about the wastewater processes.

The importance of global features is analysed at the first stage with the mean absolute SHAP values taking the average magnitude of the influence of each feature on the model output. As

demonstrated in Figure 24, in the short term, COD dynamics prevails in the prediction. Especially, the first-order COD difference and one-day lagged COD become the most significant predictors. This observation indicates the powerful autoregressive effect with memories of COD development in wastewater systems. Variables associated with nutrients, e.g. Total Nitrogen (TN) and Biochemical Oxygen Demand (BOD), are also an important source of useful information and symptoms characterise the organic load, as well as the treatment performance.

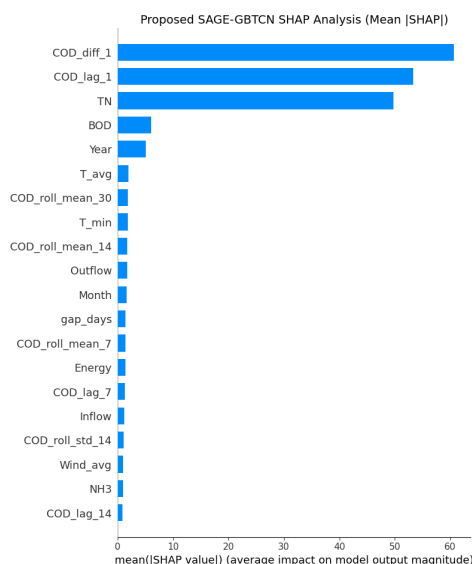


Figure 24. Global feature importance based on mean absolute SHAP values for the proposed SAGE-GBTCN model. Short-term COD dynamics, particularly first-order differences and lagged values, dominate the prediction process, indicating strong temporal dependence in wastewater pollution behaviour.

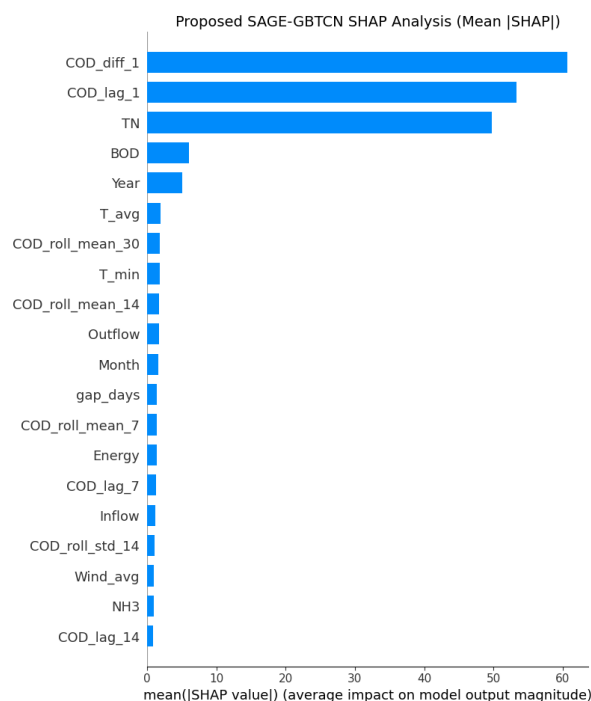


Figure 25. Alternative global SHAP feature importance representation for the proposed model, confirming the robustness of the dominant temporal COD-related predictors across different aggregation perspectives.

A consistent importance pattern is observed when an alternative global SHAP aggregation view is considered. As illustrated in Fig. 25, the dominance of short-horizon COD-related features persists across different importance representations, confirming the robustness of the identified feature

rankings. In contrast, meteorological variables, flow-related indicators, and longer-term rolling statistics exhibit comparatively lower importance, suggesting that short-term temporal dependencies are more critical than slowly varying exogenous factors for COD prediction in this setting.

An alternative global SHAP aggregation perspective suggests a constant pattern of importance. The predominance of short-horizon COD-related features is observed in all the importance representations which supports the strength of the obtained feature ranking as shown in Figure 24. Conversely, meteorological variables, flow-based pointers, and more long-term rolling statistics are showing even lower significance, implying that short term temporal relationships are more fundamental than those that vary slowly but exogenously in predicting of COD in this context.

In addition to the global significance, a SHAP beeswarm plot is used to analyze the directionality and the distribution of effects of the features. To illustrate in Figure 26, the higher values of the COD predictive coefficients, namely COD_diff_1 and COD_lag_1, are mostly associated with higher SHAP values, therefore, they influence more model outputs. Weaker patterns are observed with TN and BOD. The fact that the concentration of SHAP values around zero on less influential features is confirming the fact that the model is not overly dependent on irrelevant inputs.

In general, the interpretability analysis using the SHAP revealed that the proposed SAGE-GBTCN framework can learn physically intuitive and temporally consistent relations. This correspondence of data-generated feature importance with domain knowledge increases the trust in the predictive validity of the model and contributes to its further usage as a tool of wastewater monitoring and decision-making in the real-world.

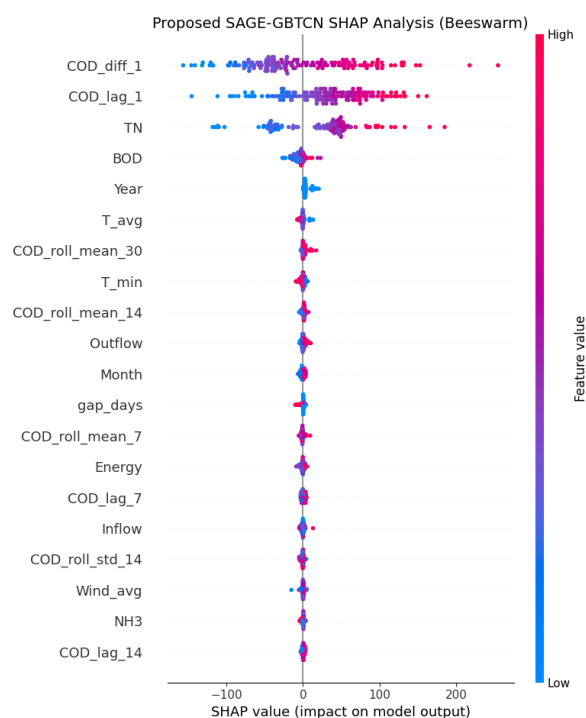


Figure 26. SHAP beeswarm plot illustrating the distribution and direction of feature contributions to COD predictions. Colour intensity represents feature value magnitude, revealing how high and low values of key predictors influence the model output.

4.9. Comparison with Representative AI-Based Wastewater Prediction Models

To contextualise the proposed SAGE-GBTCN framework within recent advances in AI-based wastewater quality prediction, a comparative assessment is conducted against representative state-of-the-art approaches, as summarised in Table 8. The selected studies span attention-based recurrent architectures [21], interpretable LSTM models with post-hoc explanation [22], and hybrid temporal convolutional frameworks [45], which collectively reflect prevailing methodological trends in the literature.

As shown in Table 8, existing approaches primarily emphasise improving average predictive accuracy through increasingly complex deep learning architectures. While these models demonstrate strong performance under normal operating conditions, most lack explicit mechanisms for residual correction or regime-aware adaptation. In particular, neither attention-based recurrent models [21] nor SHAP-enhanced LSTM frameworks [22] incorporate selective correction strategies, resulting in limited robustness when confronted with abrupt disturbances or shock loads. Similarly, hybrid TCN–LSTM approaches [45], although effective at capturing temporal dependencies, typically rely on site-specific calibration and do not explicitly address extreme-event behaviour.

Table 8. Compact comparison between the proposed model and representative AI-based wastewater prediction approaches

Aspect	[21]	[22]	[45]	Proposed
Base Architecture	Attn–LSTM + XGB	LSTM + SHAP	TCN–LSTM	GB + Gated TCN
Residual Learning	No	No	No	Yes
Shock Awareness	Limited	No	No	Yes
Selective Correction	No	No	No	Yes
Interpretability	Limited	SHAP	Limited	SHAP + Ablation
Extreme Event Evaluation	Partial	Not reported	Not reported	Explicit
Deployment Readiness	Complex	High tuning	Site-specific	Modular

In contrast, the proposed SAGE-GBTCN framework introduces an explicit residual learning strategy on top of a strong gradient boosting baseline and employs a learned gating mechanism to selectively activate corrections during abnormal regimes. This design enables the model to maintain stability under normal operating conditions while improving responsiveness during extreme pollution events, a capability that is largely absent in existing studies. Moreover, the integration of SHAP-based interpretability and ablation analysis provides transparent insight into feature contributions and component-wise effects, addressing a common limitation of prior deep learning approaches.

Overall, the comparison highlights that the proposed framework advances the state of the art not merely through architectural complexity, but by explicitly addressing regime shifts, robustness, and deployment-oriented reliability. These characteristics are particularly important for real-world wastewater monitoring and decision-support systems, where non-stationarity and rare but high-impact events are unavoidable.

5. Conclusions

A framework of theoretically motivated hybrid AI called the SAGE-GBTCN framework was created in this study by breaking down the complicated problem of forecasting wastewater pollutants into complementary and understandable parts. Its main theoretical conjecture is that the predictability of non-stationary industrial wastewater can be made more efficient by dividing the globally learnable trends and localized shock-driven dynamics. This has been operationalized by the three stage methodology in which: First, a Histogram-based Gradient Boosting (HGB) model is used to capture generalized temporal relationships and deliver a robust baseline COD prediction; Second, the residual unexplained by the HGB model is processed in parallel by a Temporal Convolutional Network (TCN) to model structured temporal relationships and neural gating network that provides a lightweight and compact neural gating network that predicts the relevance of shocks based on volatility sensitive factors; and Third a gated fusion mechanism is used to

The rigorous preprocessing that was done during the methodological implementation included the temporal feature engineering, shock indicator (PSI) generation, and strict chronological dataset splitting to avoid data leakage. Nonetheless, the study has a number of limitations. This model was built and tested with the data of a general municipal wastewater treatment plant and although it was constructed in mind of the textile related feature of operation (e.g., batch processing, shock events) its actual operation on real-time data of a textile dyeing ETP needs to be validated. Moreover, the input

of quality historical data and continuous sensor input required in the framework can be a weakness in facilities with a weak data infrastructure. Future work will explore rolling-origin retraining and cross-plant validation to further assess long-term robustness under evolving wastewater treatment regimes.

The concept of this research has a multi-dimensional outlook in the future. The next major steps would be to facilitate and optimize the SAGE-GBTCN model using specific, high-resolution data of operational textile dyeing effluent treatment facilities. The development of future work should also center on combining this predictive model with real-time decision-support systems to think of the automated optimization of aeration, chemical dosing and energy consumption in direct relation to prediction and sustainable operation. The further extension of the model to multi-target prediction (e.g., BOD, color, toxicity) along with the explanation of artificial intelligence (XAI) methods to increase operator confidence and familiarity in the industrial domain would make it even more useful in practice and adopted by customers in the industry.

Author Contributions: Syrin Jahan Ritu: Conceptualization, Methodology, Writing - Original Draft. Alamin Howlader: Resources, Data Curation. Rayhanul Islam Sony: Data Curation, Validation. Atique Ahammad Zawad: Writing - Original Draft. Shaharior Islam Chowdhury: Writing - Review & Editing, Validation. Declaration of Generative AI: During the preparation of this work, the authors used ChatGPT and Grammarly AI to enhance writing clarity and language quality. After using these tools, the authors carefully reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding: The authors received no specific funding for this work.

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request. Materials Availability: Not applicable. Code Availability: The code used to support the findings of this study is available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Ethics Approval and Consent to Participate: Not applicable. This study does not involve human participants, animal subjects, or any personally identifiable data requiring ethical approval.

References

1. Deshmukh, S.M.; Dhokpande, S.R.; Sankhe, A.; Khandekar, A. Effluent wastewater technologies for textile industry: a review. *Reviews in Inorganic Chemistry* **2025**, *45*, 21–40.
2. Sayam, S.; Das, N.; Akter, S.; Sarker, A.; Sen, A.; Habibullah, H.; Sajib, G.A.; Haque, I.; Mia, M.B.; Faruk, M.O.; et al. Water and electricity consumption patterns with effluent quality in the textile processing industry of Bangladesh. *RSC advances* **2025**, *15*, 46627–46648.
3. Abid, Z.; Sarwar, Z.; Munir, N.; Safi, S.Z.; Arshad, M. Effluent Treatment in Textile Industry to Achieve SDGs. In *Enzymes in Textile Processing: A Climate Changes Mitigation Approach: Textile Industry, Enzymes, and SDGs*; Springer, 2025; pp. 363–390.
4. Masum, M.M.H.; Islam, M.B.; Anowar, M.R.; Arsh, K.; Alam, S. Assessment of textile industry effluents and their impact on local water bodies in an urban setting of Bangladesh. *Discover Environment* **2025**, *3*, 59.
5. Sahu, D.K.; Rath, G. Innovative Approaches to Recover Waste in the Textile Sector. *Nanotechnology-Assisted Recycling of Textile Waste: Sustainable Tools for Textiles* **2025**, pp. 129–169.
6. Bari, A.H.; Akolkar, H.N.; Hatvate, N.T.; Haghi, A. Eco-Friendly Textile Processes, 2025.
7. Ojadi, J.O.; Owulade, O.A.; Odionu, C.S.; Onukwulu, E.C. AI-Driven Optimization of Water Usage and Waste Management in Smart Cities for Environmental Sustainability. *Engineering and Technology Journal* **2025**, *10*, 4284–4306.
8. Sun, W.; Gao, Y.; Zhou, J.; Shah, K.J.; Sun, Y. An Overview of the Latest Developments and Potential Paths for Artificial Intelligence in Wastewater Treatment Systems. *Water* **2025**, *17*, 2432.
9. Dai, W.; Pang, J.W.; Ding, J.; Wang, J.h.; Xu, C.; Zhang, L.Y.; Ren, N.Q.; Yang, S.S. Integrated real-time intelligent control for wastewater treatment plants: data-driven modeling for enhanced prediction and regulatory strategies. *Water Research* **2025**, *274*, 123099.
10. Alnemari, A.M.; Elmessery, W.M.; Qazaq, A.S.; Moustapha, M.E.; Rakhimgaliyeva, S.; Abuhussein, M.F.; Alhag, S.K.; Al-Shuraym, L.A.; Moghanm, F.S.; Szűcs, P.; et al. Developing highly accurate machine learning

- models for optimizing water quality management decisions in tilapia aquaculture. *Scientific Reports* **2025**, *15*, 35600.
11. Dutta, S.; Chhabra, A.; Banerjee, A.; Mukherjee, S.; Mohanty, S. The curious case of parabolic encounters: gravitational waves with linear & non-linear memory. *arXiv preprint arXiv:2511.01968* **2025**.
 12. Satyam, S.; Patra, S. The evolving landscape of advanced oxidation processes in wastewater treatment: Challenges and recent innovations. *Processes* **2025**, *13*, 987.
 13. Kang, I.; Kim, E.; Ryu, W.; Shin, J.; Yu, S.; Kang, Y.H.; Jeong, S.; Kim, E.; Kim, S.; Shim, H. Real-Time Long Horizon Air Quality Forecasting via Group-Relative Policy Optimization. *arXiv preprint arXiv:2511.22169* **2025**.
 14. Sigonya, S.; Mothudi, B.M.; Fakayode, O.J.; Mokhena, T.C.; Mayer, P.; Mokhothu, T.H.; Makhanya, T.R.; Shingange, K. Electrospun molecularly imprinted polymers for environmental remediation: a mini review. *Polymers* **2025**, *17*, 2082.
 15. Pan, C.; Qu, B.; Miao, R.; Wang, X. Cost-Efficient Fall Risk Assessment With Attention Augmented Vision Machine Learning on Sit-to-Stand Test Videos. *IEEE Access* **2025**.
 16. Göbbels, L.; Feil, A.; Raulf, K.; Greiff, K. Current State of the Art and Potential for Construction and Demolition Waste Processing: A Scoping Review of Sensor-Based Quality Monitoring and Control for In-and Online Implementation in Production Processes. *Sensors* **2025**, *25*, 4401.
 17. Shakeel, M.; Zaidi, S.; Ahmad, A.; Abahussain, A.; Nazir, M. Benchmarking of key performance factors in textile industry effluent treatment processes for enhanced environmental remediation. *Scientific Reports* **2024**, *14*, 26629.
 18. Capodaglio, A.G.; Callegari, A. Use, potential, needs, and limits of AI in wastewater treatment applications. *Water* **2025**, *17*, 170.
 19. Song, Y.; Wang, Y.; Xu, T.; Shi, X.; Wang, A.J.; Wang, H.C. Data-driven management strategies for carbon emissions and emerging contaminants control in wastewater treatment plants. In *Water Security: Big Data-Driven Risk Identification, Assessment and Control of Emerging Contaminants*; Elsevier, 2024; pp. 537–549.
 20. Sakib, M.; Mustajab, S.; Alam, M. Ensemble deep learning techniques for time series analysis: a comprehensive review, applications, open issues, challenges, and future directions. *Cluster Computing* **2025**, *28*, 73.
 21. Xiong, Z.; Liu, X.; Igou, T.; Li, Z.; Chen, Y. Using hybrid machine learning to predict wastewater effluent quality and ensure treatment plant stability. *Water* **2025**, *17*, 1851.
 22. Li, R.; Feng, K.; An, T.; Cheng, P.; Wei, L.; Zhao, Z.; Xu, X.; Zhu, L. 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. <https://doi.org/10.1021/acsestwater.4c00040>.
 23. Li, R.; Feng, K.; An, T.; Cheng, P.; Wei, L.; Zhao, Z.; Zhu, L. 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.
 24. An, T.; Feng, K.; Cheng, P.; Li, R.; Zhao, Z.; Xu, X.; Zhu, L. Adaptive prediction for effluent quality of wastewater treatment plant: Improvement with a dual-stage attention-based LSTM network. *Journal of Environmental Management* **2024**, *359*, 120887. <https://doi.org/10.1016/j.jenvman.2024.120887>.
 25. Yin, H.; Chen, Y.; Zhou, J.; Xie, Y.; Wei, Q.; Xu, Z. A probabilistic deep learning approach to enhance the prediction of wastewater treatment plant effluent quality under shocking load events. *Water Research X* **2025**, *26*, 100291.
 26. Lai, J. Research on prediction algorithm of effluent quality and development of integrated control system for waste-water treatment. *Scientific Reports* **2025**, *15*, 19257.
 27. Chen, Y.; Zhang, H.; You, Y.; Zhang, J.; Tang, L. A hybrid deep learning model based on signal decomposition and dynamic feature selection for forecasting the influent parameters of wastewater treatment plants. *Environmental Research* **2025**, *266*, 120615.
 28. Zhang, S.; Cao, J.; Gao, Y.; Sun, F.; Yang, Y. A Deep Learning Algorithm for Multi-Source Data Fusion to Predict Effluent Quality of Wastewater Treatment Plant. *Toxics* **2025**, *13*, 349.
 29. Jiang, Y.; Li, C.; Sun, L.; Guo, D.; Zhang, Y.; Wang, W. A deep learning algorithm for multi-source data fusion to predict water quality of urban sewer networks. *Journal of Cleaner Production* **2021**, *318*, 128533.
 30. Ba-Alawi, A.H.; Kim, J. Dual-stage soft sensor-based fault reconstruction and effluent prediction toward a sustainable wastewater treatment plant using attention fusion deep learning model. *Journal of Environmental Chemical Engineering* **2025**, *13*, 116221.

31. Deng, Z.; Wan, J.; Ye, G.; Wang, Y. Data-driven prediction of effluent quality in wastewater treatment processes: Model performance optimization and missing-data handling. *Journal of Water Process Engineering* **2025**, *71*, 107352.
32. Prabu, P.; Alluhaidan, A.S.; Aziz, R.; Basheer, S. AquaFlowNet: A machine learning based framework for real time wastewater flow management and optimization. *Scientific Reports* **2025**, *15*, 19182.
33. González Barberá, A.; Iserte, S.; Castillo, M.; Luis-Gómez, J.; Martínez-Cuenca, R.; Monrós-Andreu, G.; Chiva, S. Machine Learning-Based Forecasting of Wastewater Inflow During Rain Events at a Spanish Mediterranean Coastal WWTPs. *Water* **2025**, *17*, 3225.
34. Li, J.; Lin, S.; Zhang, L.; Liu, Y.; Peng, Y.; Hu, Q. Brain-inspired multimodal approach for effluent quality prediction using wastewater surface images and water quality data. *Frontiers of Environmental Science & Engineering* **2024**, *18*, 31.
35. El Messaoudi, N.; Miyah, Y.; Benjelloun, M.; Georjin, J.; Franco, D.S.; Kaur, P.; Knani, S. The role of artificial intelligence in optimizing photocatalytic degradation technologies of dyes in textile wastewater: Recent advances, challenges, and prospects. *Journal of Water Process Engineering* **2025**, *77*, 108457.
36. Mahmoud, A.S.; Peters, R.W.; Mostafa, M.K.; Hassan, R.G. Sustainable textile wastewater remediation using nano zerovalent aluminum for organic removal and pathogen inactivation. *Scientific Reports* **2025**, *15*, 37784.
37. Cruz, F.F.; Góes, M.C.; Santana, C.G.; Santos, T.G.; Boscolo, M.; Luz, R.C.; Bezerra, C.W. Optimizing Electrocoagulation for Textile Effluent Treatment: Operational Efficiency and Environmental Assessment of Remazol Red Dye Removal. *ACS Omega* **2025**.
38. Yakut, Ş.M.; Atasever, S. Modeling textile dye removal by ultrasound-assisted Fenton process using machine learning approaches. *Environmental Progress & Sustainable Energy* **2025**, p. e70089.
39. Gamboa, D.M.P.; Abatal, M.; Lima, E.; Franseschi, F.A.; Ucán, C.A.; Tariq, R.; Vargas, J. Sorption behavior of azo dye Congo red onto activated biochar from *Haematoxylum campechianum* waste: Gradient boosting machine learning-assisted Bayesian optimization for improved adsorption process. *International Journal of Molecular Sciences* **2024**, *25*, 4771.
40. Noor, S.; Al Awadh, M.; Usman Jan, Q.M.; Afaq, M.; Danish, M.; Ahmed, K.U.; Akhtar, R. Integrating experimental design and machine learning for efficient dye adsorption in textile wastewater treatment. *Journal of Industrial Textiles* **2025**, *55*, 15280837251349315.
41. Karishma, S.; Kamalesh, R.; Saravanan, A.; Yaashikaa, P.R.; Vickram, A.S. AI-driven neural time series network forecasting and cost analysis for dye removal prediction in packed bed adsorption using ultrasonic biomass composites for sustainable wastewater management. *Environmental Research* **2025**, p. 123396.
42. Rehman, A.; Iqbal, M.A.; Haider, M.T.; Majeed, A. Artificial intelligence-guided supervised learning models for photocatalysis in wastewater treatment. *AI* **2025**, *6*, 258.
43. Nagpal, M.; Siddique, M.A.; Sharma, K.; Sharma, N.; Mittal, A. Optimizing wastewater treatment through artificial intelligence: recent advances and future prospects. *Water Science & Technology* **2024**, *90*, 731–757.
44. Baarimah, A.O.; Bazel, M.A.; Alaloul, W.S.; Alazaiza, M.Y.; Al-Zghoul, T.M.; Almuhaya, B.; Mushtaha, A.W. Artificial intelligence in wastewater treatment: Research trends and future perspectives through bibliometric analysis. *Case Studies in Chemical and Environmental Engineering* **2024**, *10*, 100926.
45. Xie, Y.; Chen, Y.; Wei, Q.; Yin, H. A hybrid deep learning approach to improve real-time effluent quality prediction in wastewater treatment plant. *Water Research* **2024**, *250*, 121092.
46. Islam, M.R. AI-driven water purification model implementation in smart cities: Real-time solar desalination and effluent treatment. *International Journal of Business and Economics Insights* **2025**, *5*, 270–293.
47. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.
48. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: unbiased boosting with categorical features. In Proceedings of the Advances in neural information processing systems, 2018, pp. 6638–6648.
49. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Statistics and computing* **2004**, *14*, 199–222.
50. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural networks* **1989**, *2*, 359–366.
51. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A highly efficient gradient boosting decision tree. In Proceedings of the Advances in neural information processing systems, 2017, pp. 3146–3154.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s)

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.