

Article

Optimization of Wastewater Treatment Through Machine Learning-Enhanced Supervisory Control and Data Acquisition: A Case Study of Granular Sludge Process Stability and Predictive Control

Igor Gulshin ^{1,*}  and Olga Kuzina ²

¹ Research and Education Centre “Water Supply and Wastewater Treatment”, Moscow State University of Civil Engineering, 26, Yaroslavskoye Highway, 129337 Moscow, Russia

² Department of Information Systems, Technologies and Automation in Construction, Moscow State University of Civil Engineering, 26, Yaroslavskoye Highway, 129337 Moscow, Russia; kuzinaon@mgsu.ru

* Correspondence: gulshinia@mgsu.ru

Abstract: This study presents an automated control system for wastewater treatment, developed using machine learning (ML) models integrated into a Supervisory Control and Data Acquisition (SCADA) framework. The experimental setup focused on a laboratory-scale Aerobic Granular Sludge (AGS) reactor, which utilized synthetic wastewater to model real-world conditions. The machine learning models, specifically N-BEATS and Temporal Fusion Transformers (TFTs), were trained to predict Biological Oxygen Demand (BOD₅) values using historical data and real-time influent contaminant concentrations obtained from online sensors. This predictive approach proved essential due to the absence of direct online BOD₅ measurements and an inconsistent relationship between BOD₅ and Chemical Oxygen Demand (COD), with a correlation of approximately 0.4. Evaluation results showed that the N-BEATS model demonstrated the highest accuracy, achieving a Mean Absolute Error (MAE) of 0.988 and an R² of 0.901. The integration of the N-BEATS model into the SCADA system enabled precise, real-time adjustments to reactor parameters, including sludge dose and aeration intensity, leading to significant improvements in granulation stability. The system effectively reduced the standard deviation of organic load fluctuations by 2.6 times, from 0.024 to 0.006, thereby stabilizing the granulation process within the AGS reactor. Residual analysis suggested a minor bias, likely due to the limited number of features in the model, indicating potential improvements through additional data inputs. This research demonstrates the value of machine learning-driven predictive control for wastewater treatment, offering a resilient solution for dynamic environments. By facilitating proactive management, this approach supports the scalability of wastewater treatment technologies while enhancing treatment efficiency and operational sustainability.

Keywords: machine learning; SCADA systems; wastewater treatment; automated control; sustainability; predictive analytics; environmental impact



Academic Editor: Ahmed Abu-Siada

Received: 8 November 2024

Revised: 11 December 2024

Accepted: 15 December 2024

Published: 27 December 2024

Citation: Gulshin, I.; Kuzina, O. Optimization of Wastewater Treatment Through Machine Learning-Enhanced Supervisory Control and Data Acquisition: A Case Study of Granular Sludge Process Stability and Predictive Control. *Automation* **2025**, *6*, 2. <https://doi.org/10.3390/automation6010002>

Automation **2025**, *6*, 2. <https://doi.org/10.3390/automation6010002>

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The integration of machine learning (ML) into SCADA systems for wastewater treatment represents a notable advancement in environmental engineering and industrial automation. SCADA systems play a critical role in managing complex processes, and ML algorithms further enhance operational efficiency, predictive maintenance, and decision-making in real-time. This development is especially significant as industries face increased

pressure to manage wastewater under tightening environmental regulations [1,2]. Techniques such as Artificial Neural Networks (ANNs), Random Forest (RF), and Support Vector Machines (SVMs) have proven effective in optimizing energy consumption, chemical dosing, and compliance with environmental standards [3,4]. These advanced methods offer adaptability and efficiency, addressing the unique challenges of complex treatment systems [5]. However, the integration of ML brings challenges in terms of data quality, cybersecurity, and system complexity, which can affect commissioning and operations [1,6].

Cloud-based infrastructures are considered key for enabling remote monitoring, reducing lifecycle costs, and enhancing scalability [2]. In this rapidly evolving sector, research and practical case studies are critical for developing best practices in data integration and system maintenance for wastewater treatment [7,8]. Intelligent automation fosters sustainable water resource management, aligning with environmental objectives [7,8]. The integration of machine learning (ML) and SCADA systems into wastewater treatment represents the industry's response to contemporary environmental challenges, offering improved efficiency, reliability, and sustainability [9,10].

The primary objective of treatment facilities is to minimize the discharge of pollutants, such as nitrogen and phosphorus, which contribute to eutrophication. Stricter regulations and the pursuit of sustainability are driving the adoption of new technologies, including membrane filtration, activated sludge, biofilm-based treatments, and innovative biological processes such as Anammox [11–13]. Treatment plants are increasingly viewed not only as purification stations but also as sources of secondary resources (e.g., biomass, energy), contributing to closed-loop production systems. Effective monitoring and control are essential for the economic viability of advanced treatment methods. Both large-scale facilities (e.g., the Kuryanovo treatment plant in Moscow with a capacity of 3,000,000 m³/day) and decentralized systems (10–1000 m³/day) benefit from enhanced software and hardware capabilities. An optimal balance between complexity and reliability is achieved through a combination of traditional and innovative approaches; ML and big data analytics are becoming indispensable elements of automation [2]. When automating, it is crucial to consider the composition of wastewater. In Russia, wastewater is categorized as surface, industrial, domestic, or mixed (urban). Urban facilities primarily handle mixed flows (domestic and pre-treated industrial), whereas surface and untreated industrial waters are treated separately.

This study focuses on treatment facilities for domestic and similar urban wastewaters. The core technology is bioreactors, where microbial communities consume pollutants under controlled conditions, producing treated effluent, excess biomass (for processing), and gasses (CO₂, nitrogen). SCADA systems act as intermediaries between remote devices and visualization and control centers [1,6], encompassing data collection, processing, network connectivity, and commissioning [11]. A critical issue is the reliance on local control units, which are functionally and connectivity-limited. In the event of internet outages, these units revert to simplified operational logic [1].

To improve reliability and scalability, the proposed architecture applies control algorithms and logic in the cloud. This enables updates to control functions without on-site interventions, reducing lifecycle costs by simplifying local programmable logic controllers (PLCs) [2]. The cloud-based approach facilitates functionality adjustments and remote commissioning, decreasing the need for expert personnel on site and reducing capital expenditures [1]. Increasing wastewater volumes necessitate precise measurement and control [2]. Integrating ML into SCADA enhances adaptability and the intelligence of automation [5]. The choice of ML models—such as ANN, RF, SVM, or KNN—depends on process requirements and data quality. Since 2018, ML applications have expanded significantly.

ML decreases the need for manual intervention and increases operational efficiency, allowing operators to focus on critical tasks. Real-time analytics improve monitoring,

especially for complex systems [6]. Predictive maintenance can anticipate equipment failures, reducing downtime and maintenance costs [9]. Real-time ML analytics detect anomalies and critical events, facilitating timely decision-making [3,9]. However, effective ML integration into SCADA requires addressing data quality, cybersecurity, and system compatibility issues, as well as collaboration among domain experts, data scientists, and system integrators [3,6,9]. As ML in SCADA systems becomes more autonomous, effective human–machine interaction and operator training are essential [6,9]. The six-step methodology proposed for SCADA component analysis emphasizes outcomes rather than procedural details. The cloud-based architecture integrates control logic with analytics, correlating various data sources (e.g., meteorological data) to enhance efficiency, including proactive responses to weather changes [1].

In the event of lost cloud connectivity, local control logic maintains essential operations to ensure uninterrupted service. This lifecycle-oriented approach incorporates ML algorithms for preventive maintenance and data-driven insights, supporting both immediate improvements and future system evolution in line with technological advancements [1]. Case studies confirm ML's effectiveness: one facility optimized chemical dosing, reduced aeration costs, and maintained compliance [4]; another shifted from flow-paced chemical dosing to ML-based precise dosing, significantly cutting chemical use while maintaining the required microbial balance [4]. SCADA system development faces integration challenges due to multi-vendor architecture and compliance requirements, which demand a resilient and distributed structure for reliable data delivery. Rashad's study [14] highlights the success of SCADA systems incorporating PLCs, HMI, and ML classifiers, underscoring the benefits and limitations of such frameworks:

- **PLC and HMI as Key Components:** PLCs and HMIs are fundamental components of SCADA systems. PLCs perform the automatic control of technological processes, while HMIs provide operators with an interface for interacting with the system.
- **Implementation of Add-On Instruction (AOI):** Incorporating Add-On Instructions (AOIs) in PLC programming enables the creation of modular and reusable code. This simplifies the programming of complex control tasks, enhances processor efficiency, and facilitates system maintenance. These improvements contribute to the overall enhancement of SCADA system performance.
- **Integration of ML Classifiers:** Incorporating ML classifiers into the system allows for the analysis of large volumes of data collected by the SCADA system. This facilitates the prediction of system behavior, anomaly detection, and operational optimization. Consequently, the intellectual capabilities of the SCADA system are augmented, making it more adaptive and efficient.
- **Real-Time Monitoring and Control:** SCADA systems are designed for the real-time monitoring and management of industrial processes. The integration of ML and enhanced PLC programming (using AOI) enables the automation of tank-level monitoring, the maintenance of optimal operating conditions, and the protection of valuable resources. This reduces the need for manual intervention and increases the system's responsiveness.
- **Advanced Data Analysis:** SCADA systems collect and process data from various sensors and devices. The integration of ML algorithms allows for more sophisticated data analysis, the identification of hidden patterns, and informed decision-making based on the obtained information.

These findings underscore the potential of machine learning (ML) in SCADA systems to improve functionality, efficiency, and reliability in complex industrial processes. In Houston Water's wastewater treatment plants, traditional reactive maintenance creates challenges for operators, who often rely on intuition due to information overload. Shifting to a proactive approach using predictive ML models allows real-time adjustments, enhanc-

ing efficiency and ensuring regulatory compliance [15]. Kaittan's study utilizes PLCs and SCADA to optimize biochemical processing, while Rajhans [16,17] highlights SCADA's role in preventing abnormal events during biological treatment.

Fuzzy logic principles are increasingly integrated into wastewater control systems, as shown in Abdel-Basset's [18] risk assessment framework for cyber-physical systems, combining DEMATEL and TOPSIS methods with neurisophic theory to enhance control accuracy. IoT-enabled monitoring in water treatment, investigated in [19], allows remote access, simplifies maintenance, and improves efficiency. Comparisons demonstrate that fuzzy logic controllers outperform traditional PID systems in stability and energy efficiency.

In Europe, the EU Water Framework Directive (WFD) stresses integrated water management within drainage basins, emphasizing system interaction understanding, optimal sensor placement, and Instrumentation, Control, and Automation (ICA) implementation for holistic water resource management [20]. Recent research focuses on integrating AI into wastewater treatment using regression and classification models, including linear regression, decision trees, random forests, and neural networks (e.g., RNN, LSTM, Transformer Networks) [21–35]. The primary modeling challenges involve data selection and preparation, with feature identification being essential for effective ML deployment in SCADA-based wastewater monitoring. Figure 1 presents a block diagram of the automated control system for wastewater treatment processes based on machine learning methods.

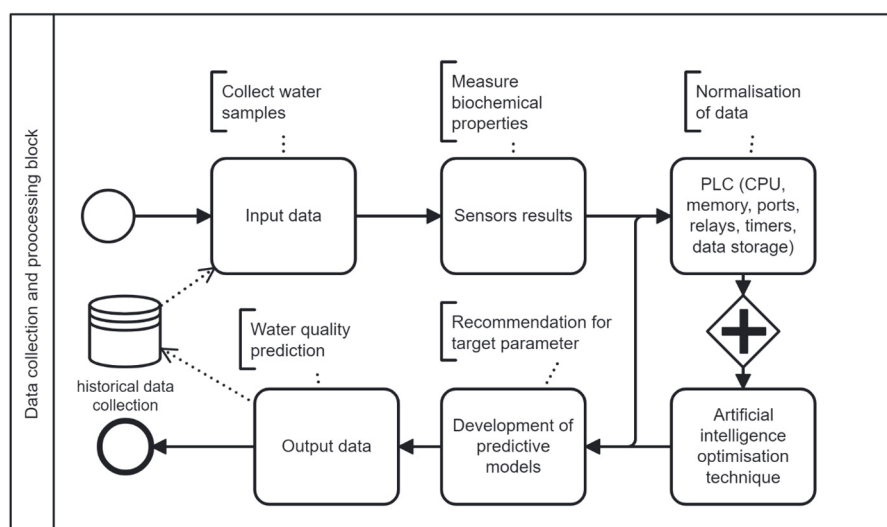


Figure 1. Block diagram of automated control system for wastewater treatment processes based on machine learning methods.

In the present study, we examine the operation of a laboratory-scale Aerobic Granular Sludge (AGS) reactor with the objective of creating and enhancing machine learning models for integration into Supervisory Control and Data Acquisition (SCADA) systems. AGS reactors represent a significant advancement in wastewater treatment technology due to their high biomass retention, superior settling characteristics, and ability to withstand variable loading conditions [36]. Despite these advantages, the complex and dynamic nature of the biological processes within AGS systems poses challenges for traditional control methodologies. By leveraging advanced machine learning algorithms, we aim to develop predictive models that can accurately capture the nonlinear relationships between operational parameters and reactor performance, thus facilitating real-time optimization and control within the SCADA framework.

The relevance of this research lies in addressing the critical need for intelligent control strategies tailored to AGS reactors, which remain underexplored in the context of automation and process optimization. Recent studies address the modeling of activated sludge granulation

processes using machine learning algorithms [37–40]. These models focus particularly on the mechanisms of sludge settling, aimed at optimizing the granulation process.

The scientific novelty emerges from the application of state-of-the-art machine learning techniques, such as deep learning and reinforcement learning, to model the intricate interactions within the microbial communities of the AGS system. This approach departs from conventional methods by focusing on the development of algorithms that can adapt to the highly variable and complex conditions of biological wastewater treatment processes. In practical terms, integrating these sophisticated models into SCADA systems promises to enhance process stability, improve effluent quality, reduce energy consumption, and lower operational costs. The outcomes of this research have the potential to significantly contribute to the field of wastewater treatment by providing a robust framework for the intelligent control of AGS reactors, thereby promoting sustainable and efficient wastewater management solutions.

This comprehensive study focuses on optimizing activated sludge granulation processes for subsequent application in various biological wastewater treatment technologies. This article presents a segment of the research dedicated to machine learning methods applied in modeling the granulation process and implemented within an automated control system. This phase involved data collection during laboratory setup operations, statistical evaluation and the processing of the data, model preparation, training and validation, performance assessment, and its subsequent integration into an SCADA-based system.

2. Materials and Methods

The experiment was performed in a laboratory bioreactor–fermenter equipped with a suite of analytical equipment. The main system, developed by Yocell Biotechnology, comprises a bioreactor vessel with a working volume of 11 L, constructed from borosilicate glass. The vessel is outfitted with a controllable electromechanical agitator with adjustable rotation speed, a pneumatic aeration system with adjustable aeration intensity, and a heating and cooling system via an external circuit. The bioreactor operated in a sequencing batch reactor (SBR) mode with a flow rate of 90 L per day. Figure 2 presents a photograph of the SBR used in the study.



Figure 2. Laboratory bioreactor used in the experiment.

Table 1 presents the operational parameters of the reactor.

Table 1. Reactor operation parameters.

Parameter	Values
Fill Time [s]	600
Idle Time [s]	60
Aeration Time [s]	9000
Settling Time [s]	300–600
Decanting Time [s]	600–1200

The study was conducted using synthetic wastewater resembling the composition of wastewater from the Moscow region in Russia. The synthetic wastewater was prepared based on dry peptone, Ammonium chloride (NH_4Cl), Sodium acetate (NaCH_3COO), and Potassium dihydrogen phosphate (KH_2PO_4). During the study, the concentrations in the wastewater were varied according to the experimental program and are presented within the ranges shown in Table 2. The historical data collected included the following wastewater characteristics:

- Biochemical Oxygen Demand (BOD_5), mgO_2/L —target indicator, also available in the model as a predictor with a 5-day temporal lag.
- Chemical Oxygen Demand (COD), mgO_2/L —Chemical Oxygen Demand, often considered as a predictor with a linear correlation with BOD_5 . In this case, however, no consistent linear relationship between BOD_5 and COD was observed, highlighting the need for modeling.
- Ammonium Nitrogen ($\text{NH}_4\text{-N}$), mg/L —a critical component of wastewater, which, due to its origin, may exhibit certain dependencies with BOD_5 .
- Orthophosphate Phosphorus ($\text{PO}_4\text{-P}$), mg/L —similar to ammonium nitrogen in relevance.
- Total Suspended Solids (TSSs), mg/L —a wastewater indicator with a nonlinear association with BOD_5 .

Table 2. Concentrations of synthetic wastewater.

Parameter	Mean	Median	Std
BOD_5 , mgO_2/L	364	365	30.5
$\text{NH}_4\text{-N}$, mg/L	41.99	42.37	4.04
$\text{PO}_4\text{-P}$, mg/L	4.60	4.57	0.61
COD, mgO/L	498.2	495.5	68.3
TSS, mg/L	557.34	544.44	162.55

Data accumulation for model training commenced post-granulation setup, with system operation conducted in a quasi-static mode. The primary regulated parameters included:

- Dissolved Oxygen Concentration: Essential for the active growth of granular sludge and for maintaining aerobic conditions conducive to granule stability, maintained at 3 mg/L during the aeration phase.
- Settling Time: Critical for the separation of granular sludge and treated water. Optimized to prevent granule washout, and used to control MLVSS and organic load.
- Carbon–Nitrogen–Phosphorus Ratio: A balanced ratio of 100:10:1 was maintained to ensure granule growth and stability, preventing structural disintegration.
- Active Sludge Dose (MLVSS): To maintain required granule density, MLVSS was kept at 5–6 g/L depending on organic load.
- pH: Controlled within the range of 7.0–8.0 to prevent granule disintegration and ensure effective biological activity.

- Aeration Intensity: Kept at 1.2–2.5 L/min per liter of reactor volume, based on influent BOD₅ concentration, to avoid excessive turbulence or oxygen deficiency.
- Mixing of Sludge Mixture: Mixing speed was maintained at 30–60 rpm, adjusted inversely with aeration intensity.
- Organic Loading Rate (OLR): In a laboratory bioreactor setting, OLR was regulated via the F/M ratio, fixed at 0.3 kg BOD₅/kg MLVSS per day. MLVSS adjustments were made to account for influent BOD₅ fluctuations, using settling and decanting times to control MLVSS.

The bioreactor was further enhanced with a set of analytical sensors corresponding to those used in actual wastewater treatment facilities: the Hamilton VisiFerm DO sensor for dissolved oxygen (Hamilton Company, Reno, NV, USA), the Hamilton Polilyte Plus pH ARC sensor for pH measurement (Hamilton Company, Reno, NV, USA), the HACH A-ISE system for nitrogen compound analysis (Hach, Loveland, CO, USA), and the CarboVis 701/705 IQ system for analyzing COD (Xylem Analytics, San Diego, CA, USA).

Control and calibration measurements were conducted using the HACH Lange DR6000 UV–vis spectrophotometer (Hach, Loveland, CO, USA), the WTW OxiTOP-IDS system (Xylem Analytics, San Diego, CA, USA), the WTW Oxi3310 equipped with the CelLOX 325 sensor (Xylem Analytics, San Diego, CA, USA), and the WTW pH 3310 analyzer (Xylem Analytics, San Diego, CA, USA). Control tests were performed using standard methods. All equipment underwent verification and calibration. Real-time operating sensors were calibrated prior to testing. The setup was controlled using an integrated controller as well as a personal computer through Siemens Simatic software (Siemens SIMATIC WinCC 7.4). In addition to automating the control of the main components of the setup, data collection from all analyzers was integrated into the control program. The collected data were exported to a csv file for further analysis.

In this study, obtaining BOD₅ values is essential for control purposes. However, there is no direct method for measuring BOD₅ using online sensors. The only feasible solution involves measuring Chemical Oxygen Demand (COD) and subsequently converting it to BOD₅. Nevertheless, this approach carries significant risks due to the possibility of unregulated COD discharges, leading to unstable COD/BOD₅ ratios. The machine learning task in this study aims to forecast BOD₅ values based on historical data and current concentrations of wastewater contaminants, measured in real-time by corresponding analytical sensors. These predictive solutions are then integrated into the SCADA system, as detailed in the Results and Discussion Section.

The study employed Neural Basis Expansion Analysis for Interpretable Time Series Forecasting (N-BEATS) and Temporal Fusion Transformers (TFTs) as machine learning models.

N-BEATS (Neural Basis Expansion Analysis for Interpretable Time Series Forecasting) is a state-of-the-art deep neural network specifically designed for time series forecasting. Introduced in 2020, the model has demonstrated outstanding results, outperforming many existing methods, including statistical models and other neural networks [41]. The model architecture consists of the Stack, Block, and Basis expansion layer components. Initially, N-BEATS was developed for univariate forecasting; however, its architecture can be adapted for multiple predictors through the following steps:

- Combining historical target values with current predictor values into a single input vector;
- Updating the model's input layer to accept the extended input vector;
- Training the model on a dataset that includes predictors.

The N-BEATS model, while effective for time series forecasting, has limitations. It requires large datasets to learn temporal patterns effectively, making it less suitable for sparse or small datasets. The model's reliance on high-quality data and extensive preprocessing increases the risk of poor performance with noisy or incomplete data. N-BEATS is

computationally intensive, requiring significant resources for training, and its architecture may overfit without careful tuning. Additionally, it is less adept at handling multivariate time series or complex temporal dependencies compared to models like TFT. Lastly, its “black box” nature limits interpretability, which may be a concern for certain applications. The modifications to the N-BEATS architecture effectively addressed its limitations by enabling it to handle multivariate time series and complex dependencies through a single block design with additional features. By incorporating residual learning, the model improved predictive accuracy and reduced overfitting, while preprocessing steps, such as lagged features and standardization, ensured robustness against noisy and incomplete data. These enhancements allowed the model to provide reliable BOD₅ predictions in wastewater management applications.

The Temporal Fusion Transformer (TFT) is a modern deep learning model specifically designed for multivariate time series forecasting with complex temporal dependencies. Introduced by Google in 2019, TFT combines the benefits of recurrent neural networks and attention mechanisms, providing high forecasting accuracy and model interpretability [42,43]. The model architecture consists of the following components: Variable Selection Network, LSTM Encoder–Decoder, Static Covariate Encoders, Gated Residual Network, and Interpretable Multi-Head Attention.

The Temporal Fusion Transformer (TFT) is a powerful model for time series forecasting but has limitations. It requires large datasets, making it less effective with sparse or limited data. High computational complexity and sensitivity to hyperparameter tuning can lead to long training times and potential overfitting. While it includes interpretability features, understanding its outputs may still be challenging. TFT struggles with highly non-stationary data or varying seasonal patterns without extensive preprocessing. Additionally, its performance heavily relies on high-quality feature engineering. Lastly, the model’s complexity and inference speed may hinder its use in real-time applications or scenarios with sudden data shifts.

To evaluate the accuracy of the predicted values, the metrics Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) were used, calculated according to Formulas (1)–(3).

$$R^2 = 1 - \frac{D[y|x]}{D[y]} = 1 - \frac{\sigma^2}{\sigma_y^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2} \quad (1)$$

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

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

where $D[y] = \sigma_y^2$ represents the variance of the random variable, $D[y|x] = \sigma^2$ denotes the variance of the model error, n is the number of observations, y_i is the target value, \hat{y}_i is the predicted value, and \bar{y}_i is the mean value.

A correlation assessment using chi-squared values was employed. The chi-squared test is a statistical method used to examine the dependence between categorical, discrete, and continuous numerical variables. It can be applied to test hypotheses about data distribution within a sample, as well as to analyze contingency tables (for example, to assess the independence of variables).

The general algorithm for calculating the chi-squared statistic is as follows:

1. Formulation of Hypotheses:

- Null Hypothesis (H_0): The variables are independent.

- Alternative Hypothesis (H_1): The variables are dependent.
2. Data Collection and Organization:
 - A contingency table is created, displaying the frequency of observations for each category (in the case of assessing categorical relationships).
 3. Calculation of Expected Values:
 - For each cell in the contingency table, the expected value is calculated using Formula (4):

$$E_{ij} = \frac{(R_i \times C_j)}{N} \quad (4)$$

where E_{ij} is the expected frequency, R_i is the total frequency for row i , C_j is the total frequency for column j , and N is the overall total number of observations.

1. Calculation of Chi-Squared Statistic:
 - Formula (5) is used:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (5)$$

where O_{ij} —represents the observed frequency in a cell.

2. Determination of Degrees of Freedom:
 - For a contingency table, the degrees of freedom (df) are calculated using Formula (6):

$$df = (r - 1) \times (c - 1) \quad (6)$$

where r is the number of rows and c is the number of columns.

3. Comparison with Critical Value:
 - The critical chi-squared χ^2 value is determined for the specified significance level (in this study 0.05);
 - If the calculated χ^2 value exceeds the critical value obtained from the chi-squared distribution table, the null hypothesis is rejected. This rejection indicates that there is a statistically significant association between the variables under investigation.

The determination of degrees of freedom and the subsequent comparison with the critical chi-squared value are essential steps in the chi-squared test. By performing these calculations, researchers can objectively assess the independence or dependence of variables within SCADA systems used in wastewater treatment. This statistical validation supports the identification of meaningful relationships between operational parameters, thereby enhancing the monitoring and optimization processes within the system.

3. Results and Discussion

A dataset consisting of 366 rows and five features was obtained in the course of the study. Table 3 presents an example of the structure of the compiled table with the initial data.

Table 3. A portion of the dataset used in the study.

Date	BOD_5	COD	NH_4	PO_4	TSS
16 August 2022	302.63	441.32	28.45	2.65	157.21
17 August 2022	330.4	338.6	32.21	3.3	264.92
18 August 2022	363.06	560.15	36.69	3.49	299.1
...
...
...
16 August 2024	421.5157994	621.18	53.82	6.25	1196.8

Figure 3 presents a histogram of the target feature, BOD₅, showing a normal distribution of values. The dataframe contains no missing values, anomalies, or excessive outliers.

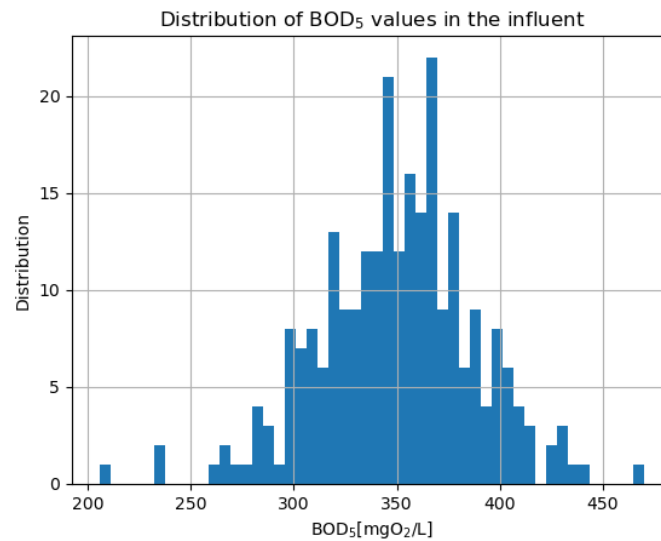


Figure 3. Histogram of the target feature BOD₅.

Figure 4 presents the concentration trends of the main contaminants in the influent wastewater (characteristics), shown after weekly resampling.

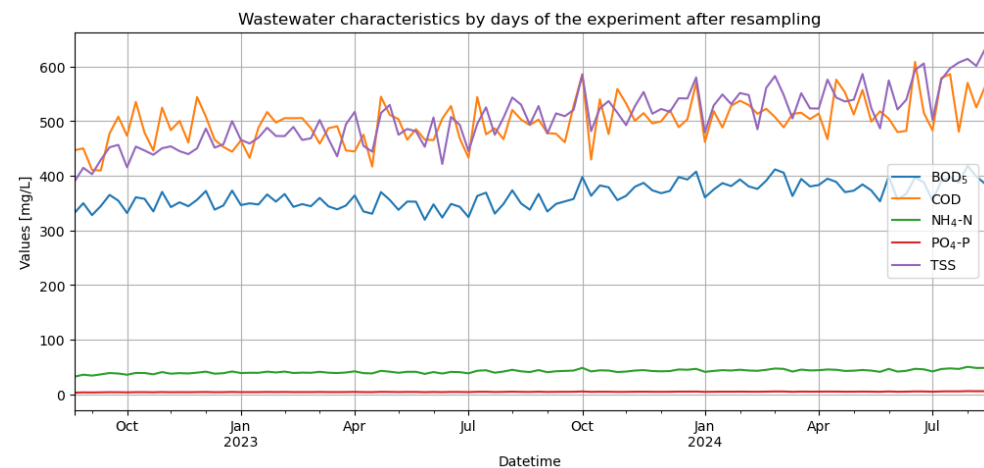


Figure 4. Wastewater characteristics by days of the experiment after resampling (by week).

The target characteristic exhibits a pronounced upward trend, which is most clearly revealed after decomposition. Figure 5 presents the trend and seasonal component plots of the target feature BOD₅.

The seasonal component is not pronounced, which is also associated with the absence of explicit seasonality in the preparation of synthetic wastewater. This study aimed to enable system operation under conditions lacking explicit seasonal dependencies, which are generally present in municipal wastewater to varying degrees. The strength of correlation between the target feature and measurable predictors was assessed using chi-squared values, as described in Section 2. The correlation matrix with calculated values is presented in Figure 6.

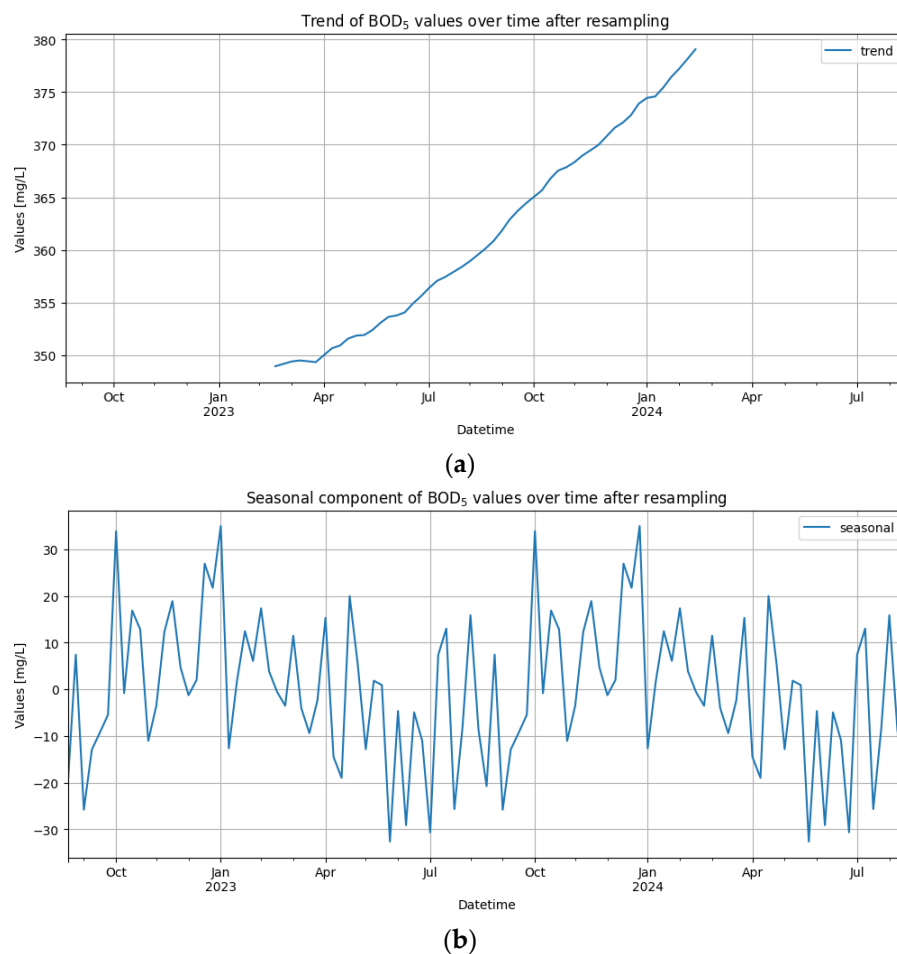


Figure 5. Characteristics of the timeline: (a) trend of BOD₅ values over time after resampling; (b) seasonal component of BOD₅ values over time after resampling.

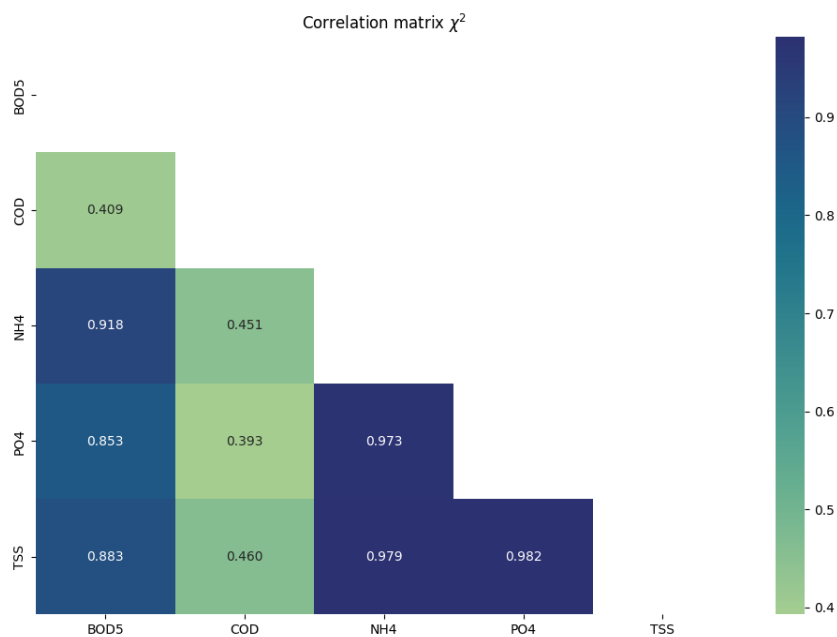


Figure 6. Correlation matrix χ^2 .

The correlation matrix reveals relationships between the target feature and other characteristics (ranging from 0.85 to 0.92) as well as among the predictors themselves. In this project, there is an understanding of the relationship between the concentrations of biogenic

substances in the influent water and TSS due to the characteristics of synthetic wastewater preparation. The allowable multicollinearity threshold was set at 0.99; thus, predictor exclusion is not required. As specified in the task, the relationship between COD and BOD₅ is minimal (0.4), which may be associated with the nature of wastewater formation. Additionally, to assess the correlation significance of time lags, an autocorrelation function (ACF) plot for the target feature time series was generated. The ACF plot is presented in Figure 7.

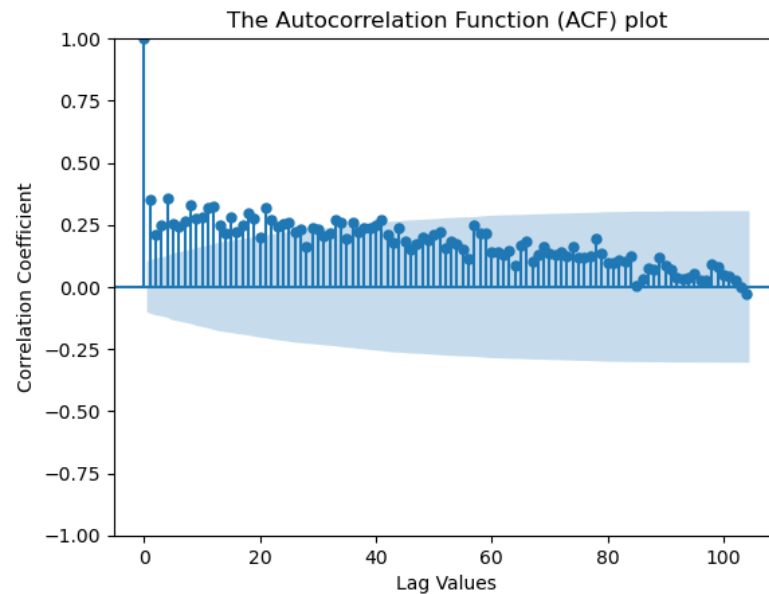


Figure 7. The Autocorrelation Function for BOD₅.

As shown in the plot, the maximum time lag for this time series is 20 at a significance level of 0.05. Thus, a corresponding number of input neurons will be used in model construction. It can be observed that the time series is characterized by low dependency, which necessitates considering the influence of additional predictors when building the model. Based on the identified correlation patterns, neural networks following the N-BEATS and TFT architectures were constructed.

Figure 8 presents a simplified diagram of the N-BEATS architecture. A modification has been introduced to the network structure.

The N-BEATS model architecture begins with an input layer that accepts sequences of length 20, each containing multiple features corresponding to the lagged values and additional parameters. The core of the model consists of three stacks, each containing three N-BEATS blocks, designed to handle multivariate time series data. Each block begins by flattening the input tensor to a one-dimensional vector. The flattened vector is then processed through four fully connected layers, each with 256 units and ReLU activation functions, to capture complex nonlinear relationships in the data. This stacked architecture allows for a hierarchical decomposition of the input data, with residuals from each block passed to subsequent blocks within the same stack, and residuals from each stack forwarded to the next stack for further refinement of the forecast and backcast outputs.

Following the dense layers, the model computes a set of parameters (thetas) with a dimension of 128 through a linear activation layer. These thetas are used to generate both the backcast and forecast outputs. The backcast is produced by a dense layer that reconstructs the input sequence, aiming to model the residuals between the input and its approximation. It reshapes the output to match the original input dimensions. The forecast output, representing the BOD₅ prediction five days ahead, is generated through another dense layer with linear activation. A residual connection is implemented by subtracting

the backcast from the input layer, allowing the model to focus on learning the differences between the input and its approximation. This residual learning strategy helps in refining the predictive performance of the model.

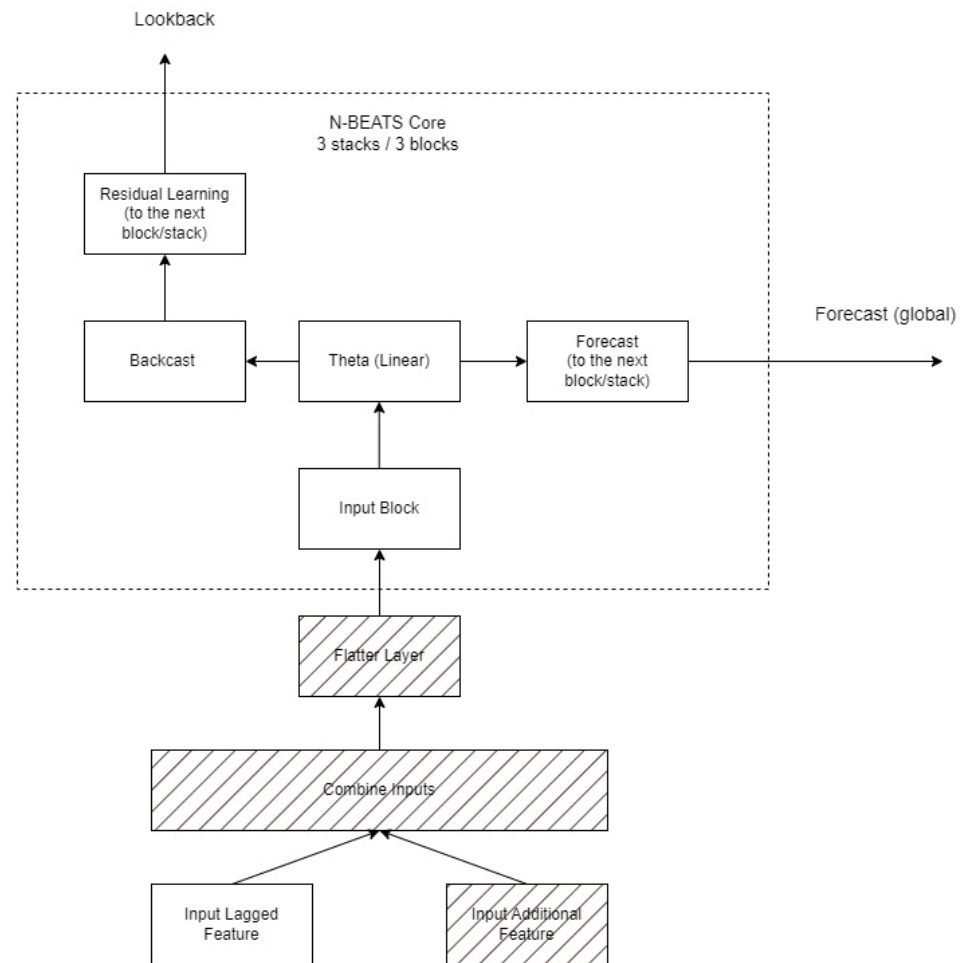


Figure 8. N-BEATS Architecture. The modified N-BEATS architecture differs from the standard version in several ways. The input layer accepts multivariate time series, combining lagged target values and additional predictors, whereas the standard model processes only univariate time series. The modified architecture includes a flatten layer to convert inputs into a one-dimensional vector, a step not required in the standard version. Each N-BEATS block incorporates four fully connected layers (256 neurons, ReLU activation) and a linear layer for parameter (theta) computation, with explicit generation of backcast (reconstruction) and forecast (prediction). Additionally, a residual connection subtracts the backcast from the input, improving model learning, which is less emphasized in the standard model.

For training, we used the Adam optimizer with a mean squared error (MSE) loss function. The model was trained over 100 epochs with a batch size of 16. Data preprocessing involved sorting the data chronologically, creating lagged features, handling missing values resulting from lagging, and standardizing both the input features and the target variable using the StandardScaler. The dataset was split into training and testing sets with an 80/20 ratio. This modified N-BEATS model effectively leverages historical BOD₅ values and additional water quality parameters to provide accurate predictions. By integrating both temporal lags and multivariate features, the model captures essential patterns and dependencies crucial for forecasting BOD₅ levels in wastewater management applications.

The architecture of the TFT model comprises an encoder–decoder framework tailored for multivariate time series forecasting. The encoder ingests sequences of past observations,

including both the target BOD₅ values and supplementary features, capturing historical patterns and temporal dynamics. The decoder processes future-known features over the prediction horizon, excluding the target variable, enabling the model to incorporate any available future information. As “future-known features”, previously predicted BOD₅ values are employed, since the forecast is made 5 days ahead, allowing the accumulation of the necessary data. However, it should be noted that these data are derived from the predictions of the same model.

A distinctive aspect of our model is the incorporation of an additive attention mechanism between the encoder and decoder. This attention layer allows the decoder to selectively focus on relevant encoder time steps, enhancing the model’s ability to capture long-term dependencies and complex feature interactions. By combining the context vectors from attention with decoder outputs, the model effectively fuses historical and future information. To address the challenge of limited data, we employed extensive preprocessing, including scaling and careful sequence generation, ensuring the model effectively learns from the available observations. Training was conducted using the Adam optimizer and mean squared error loss function over 50 epochs, with early stopping based on validation performance to prevent overfitting. The scientific novelty lies in the tailored adaptation of the TFT model for wastewater quality prediction, specifically BOD₅ forecasting, by integrating lagged features and multiple parameters within an attention-based sequence-to-sequence framework. This approach not only enhances predictive accuracy but also offers interpretability through the attention weights, potentially revealing critical periods and features influencing BOD₅ levels. Such insights are invaluable for environmental monitoring and decision-making in wastewater management. Figure 9 presents a graphical representation of the TFT model architecture.

Table 4 presents the key metric values for the models under consideration.

Table 4. The key metric values.

Model	R ²	RMSE	MAE
TFT	0.898	1.115	1.010
N-BEATS	0.901	1.085	0.988

According to the modeling results, the N-BEATS model demonstrated the best metric performance, with an MAE of 0.988 and a coefficient of determination of 0.901, and was subsequently incorporated into the SCADA system.

The SCADA system in this study is built on Siemens SIMATIC WinCC 7.4, developed by Siemens AG in Munich, Germany, and integrated with Siemens S7 series programmable logic controllers (PLCs) for seamless communication and reliable control. Analytical sensors measure key wastewater parameters such as dissolved oxygen, pH, nitrogen compounds, and COD, with signal converters transforming analog outputs into digital data streams for real-time monitoring and control. Historical data, collected between 16 August 2022 and 16 August 2024, was stored locally using PostgreSQL, structured into time series tables with timestamps, variable names, and values. This setup supports a retention policy of up to three years for detailed data, with optional aggregation for long-term storage. Data from WinCC were exported to PostgreSQL using native tools, forming the basis for subsequent machine learning (ML) tasks. Apache Airflow automated data extraction at scheduled intervals, performing preprocessing steps such as scaling, generating lagged features, and addressing missing values. The processed data were then sent to the ML model via a RESTful API built with FAST API, with predictions (BOD₅ values) returned and stored in PostgreSQL for integration into SCADA workflows and further analysis. To ensure data reliability and system integrity, the database was backed up daily with incremental copies

Based on the predicted values (BOD_5) provided by the ML model, the SCADA system dynamically adjusts the sludge dosing rate and modifies the duration of settling and decanting stages. Higher predicted BOD_5 values prompt the system to increase the sludge dosage accordingly and extend the settling and decanting phases. These parameters—sludge dosing rates and process stage durations—are set via SCADA, ensuring the treatment process remains proactive and responsive to anticipated wastewater loading conditions. Figure 10 presents the SCADA system flowchart.

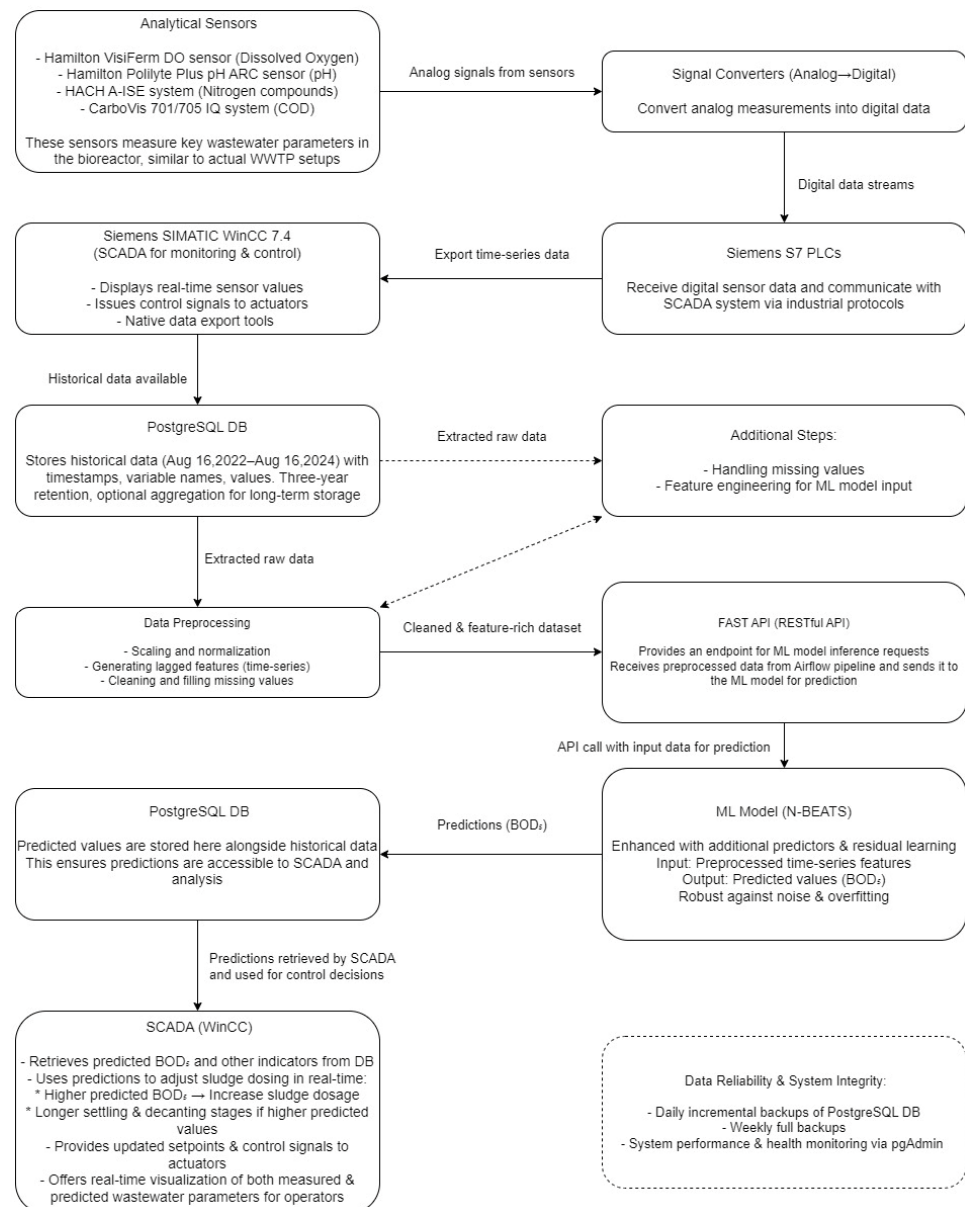


Figure 10. The SCADA system flowchart.

The operation of the SCADA-based automated system stabilized the organic load on the activated sludge, thus contributing to the stabilization of the sludge granulation process. Figure 11 presents the predicted and actual BOD_5 values, illustrating the model's performance.

The prediction accuracy is relatively high (as confirmed by the metrics in Table 4); however, a residual bias is noticeable. To verify the presence of bias, residual analysis was performed. Figure 12 presents the residual plot, which shows that the model errors are biased (the model underestimates the predicted values).

The observed bias in values is likely due to an insufficient number of features used in model training, given the time series has low autocorrelation [39]. To improve the quality of model predictions, additional features could be incorporated in the future (which would require further analytical instrumentation).

Nevertheless, intelligent predictive control of the sludge dose allowed for the leveling of organic loads on the activated sludge. This is shown in Figure 13, which presents the actual F:M ratio values, as well as the F:M values at a constant MLVSS (5 g/L).

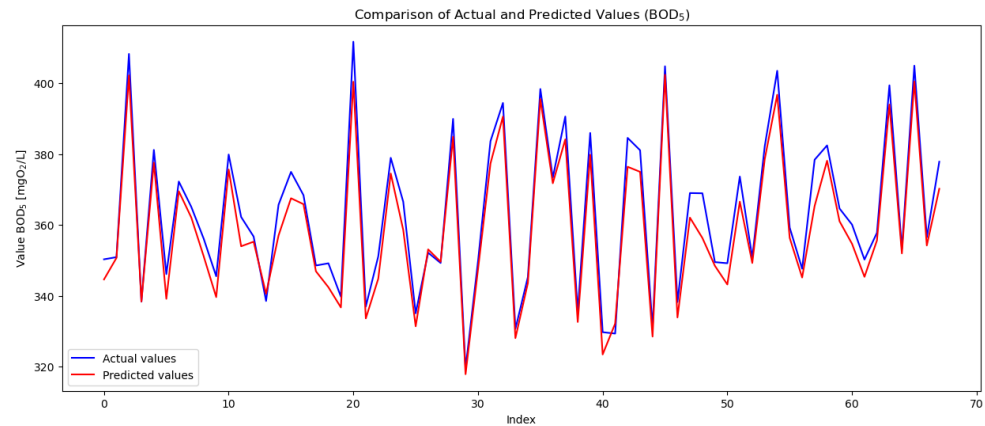


Figure 11. Actual vs. Predicted Values (BOD₅) on Test Set (N-BEATS).

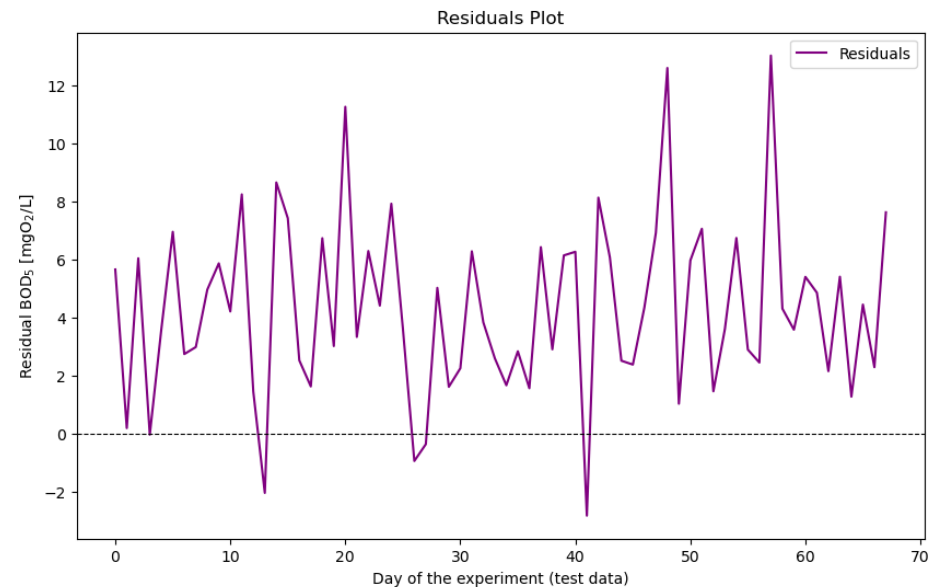


Figure 12. Residuals Plot for BOD₅ Prediction.

As shown in the graph, preventive regulation of the activated sludge dose was able to partially smooth out fluctuations in the organic load on the activated sludge, which, in turn, stabilized the activated sludge granulation process. The standard deviation was reduced by 2.6 times (from 0.024 before regulation to 0.006 after regulation). Such predictive modeling with feedback through SCADA systems can be applied for various functional purposes. However, granular activated sludge technologies are the most vulnerable in terms of maintaining a quasi-static system state, making control methods for such systems of particular interest. The integration of modern ML algorithms enables the use of time series data resembling white noise; however, further research on the possibility of connecting various predictor sets is needed to improve model quality.

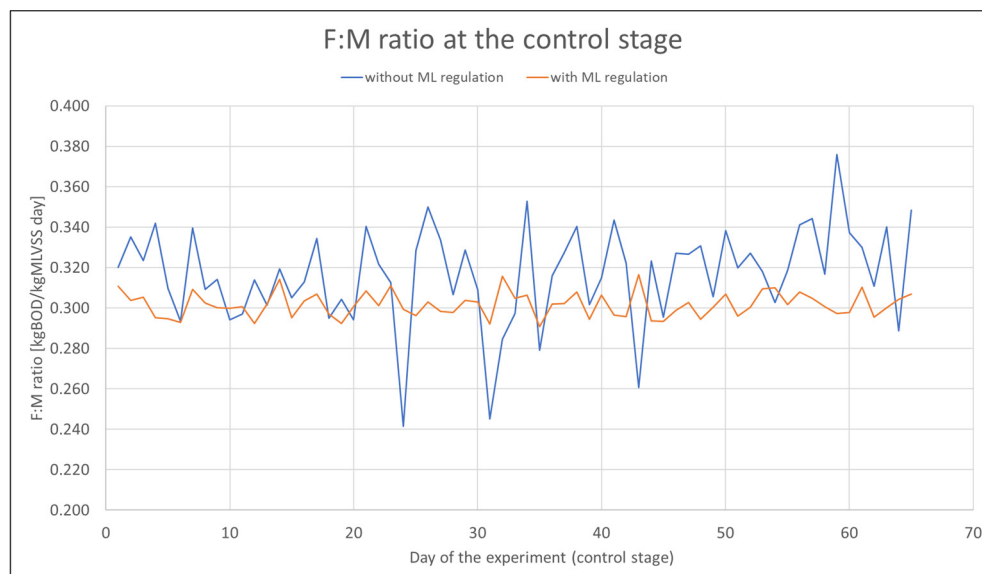


Figure 13. F:M ratio at the control stage.

4. Conclusions

This study presented the development and implementation of an automated control system for wastewater treatment processes based on machine learning (ML) models, specifically focusing on a laboratory-scale Aerobic Granular Sludge (AGS) reactor. By leveraging advanced predictive models integrated into a SCADA (Supervisory Control and Data Acquisition) framework, the system demonstrated significant improvements in stabilizing and optimizing activated sludge processes.

The experimental setup utilized an AGS reactor with a working volume of 11 L, where synthetic wastewater was treated under quasi-static operational conditions. Key process parameters, including dissolved oxygen concentration, settling time, carbon–nitrogen–phosphorus ratio, and active sludge dose (MLVSS), were regulated to maintain the reactor’s performance. Dissolved oxygen was held at 3 mg/L during aeration, while the C:N ratio was maintained at an optimal 100:10:1, ensuring steady microbial growth and sludge granulation stability. Notably, the MLVSS concentration was kept at 5–6 g/L to support granule density and bioreactor efficiency.

To address the absence of direct online measurement methods for Biological Oxygen Demand (BOD₅), a predictive machine learning approach was adopted. Given that Chemical Oxygen Demand (COD) often lacks a stable linear correlation with BOD₅ in synthetic wastewater (with a correlation value of approximately 0.4), alternative data-driven methods were necessary. Machine learning models were therefore implemented to predict BOD₅ values, using real-time COD measurements and other influent contaminant concentrations from online sensors as input features. The study employed N-BEATS (Neural Basis Expansion Analysis for Interpretable Time Series Forecasting) and Temporal Fusion Transformers (TFTs) as the primary ML models for this predictive task, ultimately selecting the N-BEATS model based on superior performance.

Quantitative evaluation of the models revealed that N-BEATS achieved a Mean Absolute Error (MAE) of 0.988 and a coefficient of determination (R^2) of 0.901, as indicated in Table 3. These metrics confirmed the model’s high predictive accuracy, making it suitable for integration into the SCADA control system. The SCADA system was equipped with sensors, a signal converter, a PLC (programmable logic controller), and a communication network (using the Modbus protocol) to enable data acquisition and automated control. The Siemens SIMATIC WinCC Unified platform served as the SCADA server, where ML model

predictions were used to adjust control parameters dynamically. For example, the system managed aeration, settling, and decanting times to achieve optimal reactor performance.

One notable outcome of the system's operation was the stabilization of the organic load on activated sludge. By implementing intelligent control over the sludge dose, the system successfully leveled organic load fluctuations, as illustrated in Figure 8. Preventive regulation of the sludge dose led to a 2.6-fold reduction in standard deviation, from 0.024 before regulation to 0.006 after regulation. This stabilization directly contributed to the improvement of granule formation within the activated sludge, thereby enhancing the overall efficiency of the treatment process. Additionally, this adaptive control reduced the need for manual adjustments, supporting continuous, unattended operation.

Residual analysis, shown in Figure 7, highlighted that while the model achieved high accuracy, a minor bias was observed, with the model tending to underestimate BOD₅ values slightly. This bias is potentially attributable to the limited number of features available for training the model, particularly given the low autocorrelation observed in the time series data (with a maximum significant time lag of 20). Future improvements in predictive accuracy may be achievable by incorporating additional features into the model, which would require further analytical instrumentation.

The integration of machine learning algorithms into SCADA-based automated systems demonstrated several important benefits for wastewater treatment. Firstly, predictive models such as N-BEATS and TFT enabled accurate, real-time forecasting of critical parameters, facilitating proactive control of reactor conditions. The inclusion of machine learning algorithms allowed the control system to handle time series data with low correlation, effectively using data that resemble white noise. This ability is especially relevant for granular sludge technologies, which are sensitive to maintaining quasi-static operational states and therefore require robust, adaptive control strategies.

While the SCADA-based system performed successfully within a controlled laboratory environment, further research is needed to evaluate its scalability and adaptability in full-scale wastewater treatment facilities. The ML models used in this study, specifically N-BEATS and TFT, proved effective in managing complex nonlinear relationships inherent in biological wastewater treatment processes. However, given the limitations observed in residual bias, future studies could explore the incorporation of a broader range of predictor variables to enhance model robustness and accuracy further.

In conclusion, this research highlights the potential of combining machine learning models with SCADA systems to improve the control and efficiency of wastewater treatment processes. The success of the N-BEATS model, which reduced the MAE to 0.988 and achieved an R² of 0.901, demonstrates the feasibility of using advanced ML techniques for predictive control in dynamic, data-intensive environments. By stabilizing the organic load on activated sludge and reducing the standard deviation by 2.6 times, the automated control system showcased here holds promise for advancing wastewater treatment technology. The integration of intelligent algorithms into SCADA systems may pave the way for more resilient and adaptive solutions in the field of environmental engineering, ultimately contributing to more sustainable wastewater management practices.

Author Contributions: Conceptualization, O.K. and I.G.; methodology, O.K.; software, I.G.; validation, I.G. and O.K.; formal analysis, I.G.; investigation, I.G.; resources, O.K.; data curation, I.G.; writing—original draft preparation; writing—review and editing, I.G.; visualization, I.G.; supervision, O.K.; project administration, I.G.; funding acquisition, I.G. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by the National Research Moscow State University of Civil Engineering.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Brad, S.; Murar, M.; Vlad, G.; Brad, E.; Popanton, M. Lifecycle Design of Disruptive SCADA Systems for Waste-Water Treatment Installations. *Sustainability* **2021**, *13*, 4950. [[CrossRef](#)]
2. Alzahrani, A.; Aldhyani, T.H.H. Design of Efficient Based Artificial Intelligence Approaches for Sustainable of Cyber Security in Smart Industrial Control System. *Sustainability* **2023**, *15*, 8076. [[CrossRef](#)]
3. Velásquez, D.; Vallejo, P.; Toro, M.; Odriozola, J.; Moreno, A.; Naveran, G.; Giraldo, M.; Maiza, M.; Sierra, B. EDAR 4.0: Machine Learning and Visual Analytics for Wastewater Management. *Sustainability* **2024**, *16*, 3578. [[CrossRef](#)]
4. Kaittan, K.H.; Mohammed, S.J. PLC-SCADA automation of inlet wastewater treatment processes: Design, implementation, and evaluation. *J. Eur. Syst. Autom.* **2024**, *57*, 787–796. [[CrossRef](#)]
5. El Alaoui El Fels, A.; Mandi, L.; Kammoun, A.; Ouazzani, N.; Monga, O.; Hbid, M.L. Artificial Intelligence and Wastewater Treatment: A Global Scientific Perspective through Text Mining. *Water* **2023**, *15*, 3487. [[CrossRef](#)]
6. Lowe, M.; Qin, R.; Mao, X. A Review on Machine Learning, Artificial Intelligence, and Smart Technology in Water Treatment and Monitoring. *Water* **2022**, *14*, 1384. [[CrossRef](#)]
7. Coito, T.; Viegas, J.L.; Martins, M.S.E.; Cunha, M.M.; Figueiredo, J.; Vieira, S.M.; Sousa, J.M.C. A Novel Framework for Intelligent Automation. *IFAC-PapersOnLine* **2019**, *52*, 1825–1830. [[CrossRef](#)]
8. Phuyal, S.; Bista, D.; Izykowski, J.; Bista, R. Design and implementation of cost efficient SCADA system for industrial automation. *Int. J. Manuf. Eng.* **2020**, *10*, 15. [[CrossRef](#)]
9. Wang, Y.; Cheng, Y.; Liu, H.; Guo, Q.; Dai, C.; Zhao, M.; Liu, D. A Review on Applications of Artificial Intelligence in Wastewater Treatment. *Sustainability* **2023**, *15*, 13557. [[CrossRef](#)]
10. Alprol, A.E.; Mansour, A.T.; Ibrahim, M.E.E.-D.; Ashour, M. Artificial Intelligence Technologies Revolutionizing Wastewater Treatment: Current Trends and Future Prospective. *Water* **2024**, *16*, 314. [[CrossRef](#)]
11. Makisha, N. Application of Biofilm Carrier in Aerobic Reactors as a Method to Improve Quality of Wastewater Treatment. *Hydrology* **2021**, *8*, 77. [[CrossRef](#)]
12. Makisha, N. Advanced Research on Polymer Floating Carrier Application in Activated Sludge Reactors. *Polymers* **2022**, *14*, 2604. [[CrossRef](#)] [[PubMed](#)]
13. Makisha, N. Preliminary Design Analysis of Membrane Bioreactors Application in Treatment Sequences for Modernization of Wastewater Treatment Plants. *Membranes* **2022**, *12*, 819. [[CrossRef](#)] [[PubMed](#)]
14. Zhang, W.; Tooker, N.B.; Mueller, A.V. Enabling wastewater treatment process automation: Leveraging innovations in real-time sensing, data analysis, and online controls. *Environ. Sci. Water Res. Technol.* **2020**, *6*, 2973–2992. [[CrossRef](#)]
15. Cao, B. Data Centric “Smart” Digital Platform for Wastewater Treatment Plant Proactive Optimization and Maintenance. In Proceedings of the Water Environment Federation; WEFTEC 2024, New Orleans, LA, USA, 5–9 October 2024; Water Environment Federation: Alexandria, VA, USA, 2024; pp. 100–110. [[CrossRef](#)]
16. Rajhans, A.R.; More, S.S.; Gambhir, S.V.; Deshmukh, V.H. Review paper on wastewater treatment plant using PLC & SCADA. *Int. J. Eng. Appl. Sci. Technol.* **2020**, *4*, 667–671.
17. Rajhans, A.R.; More, S.S.; Gambhir, S.V.; Deshmukh, V.H. Research Paper on Wastewater Treatment Plant Using PLC & SCADA. *Int. J. Eng. Appl. Sci. Technol.* **2020**, *4*, 651–657.
18. Abdel-Basset, M.; Gamal, A.; Moustafa, N.; Askar, S.S.; Abouhawwash, M. A risk assessment model for cyber-physical water and wastewater systems: Towards sustainable development. *Sustainability* **2022**, *14*, 4480. [[CrossRef](#)]
19. Ali, M.; Miry, A.; Salman, T. Implementation of intelligent industrial controller based on fuzzy logic and PLC. *Al-Qadisiyah J. Eng. Sci.* **2020**, *13*, 55–60. [[CrossRef](#)]
20. Olsson, G. Water and Wastewater Operation: Instrumentation, Monitoring, Control, and Automation. In *Water Sustainability*; Springer: New York, NY, USA, 2023; pp. 393–408. [[CrossRef](#)]
21. Zhang, Y.; Wu, H.; Xu, R.; Wang, Y.; Chen, L.; Wei, C. Machine learning modeling for the prediction of phosphorus and nitrogen removal efficiency and screening of crucial microorganisms in wastewater treatment plants. *Sci. Total Environ.* **2024**, *907*, 167730. [[CrossRef](#)]
22. Wang, D.; Thunell, S.; Lindberg, U.; Jiang, L.; Trygg, J.; Tysklind, M.; Souihi, N. A machine learning framework to improve effluent quality control in wastewater treatment plants. *Sci. Total Environ.* **2021**, *784*, 147138. [[CrossRef](#)]
23. Xu, B.; Pooi, C.K.; Tan, K.M.; Huang, S.; Shi, X.; Ng, H.Y. A novel long short-term memory artificial neural network (LSTM)-based soft-sensor to monitor and forecast wastewater treatment performance. *J. Water Process Eng.* **2023**, *54*, 104041. [[CrossRef](#)]

24. Abouzari, M.; Pahlavani, P.; Izaditame, F.; Bigdeli, B. Estimating the chemical oxygen demand of petrochemical wastewater treatment plants using linear and nonlinear statistical models—A case study. *Chemosphere* **2021**, *270*, 129465. [[CrossRef](#)] [[PubMed](#)]
25. Recio-Colmenares, R.; Becerril, E.L.; Tun, K.J.G.; Conchas, R.F. Design of a Soft Sensor Based on Long Short-Term Memory Artificial Neural Network (LSTM) for Wastewater Treatment Plants. *Sensors* **2023**, *23*, 9236. [[CrossRef](#)] [[PubMed](#)]
26. El-Rawy, M.; Abd-Ellah, M.K.; Fathi, H.; Ahmed, A.K.A. Forecasting effluent and performance of wastewater treatment plant using different machine learning techniques. *J. Water Process Eng.* **2021**, *44*, 102380. [[CrossRef](#)]
27. Puasa, S.W.; Ismail, K.N.; Mahadi, M.A.A.; Zainuddin, N.A.; Mukelas, M.N.M. Polynomial Regression Analysis for Removal of Heavy Metal Mixtures in Coagulation/Flocculation of Electroplating Wastewater. *Indones. J. Chem.* **2021**, *21*, 46–56. [[CrossRef](#)]
28. Zaghoul, M.S.; Hamza, R.A.; Iorhemen, O.T.; Tay, J.H. Comparison of adaptive neuro-fuzzy inference systems (ANFIS) and support vector regression (SVR) for data-driven modelling of aerobic granular sludge reactors. *J. Environ. Chem. Eng.* **2020**, *8*, 103742. [[CrossRef](#)]
29. Alrbai, M.; Al-Dahidi, S.; Alahmer, H.; Al-Ghussain, L.; Hayajneh, H.; Shboul, B.; Abusorra, M.; Alahmer, A. Utilizing waste heat in wastewater treatment plants for water desalination: Modeling and multi-objective optimization of a multi-effect desalination system using decision tree regression and Pelican optimization algorithm. *Therm. Sci. Eng. Prog.* **2024**, *54*, 102784. [[CrossRef](#)]
30. Al Nuaimi, H.; Abdelmagid, M.; Bouabid, A.; Chrysikopoulos, C.V.; Maalouf, M. Classification of WatSan Technologies using machine learning techniques. *Water* **2023**, *15*, 2829. [[CrossRef](#)]
31. Wang, Q.; Li, Z.; Cai, J.; Zhang, M.; Liu, Z.; Xu, Y.; Li, R. Spatially adaptive machine learning models for predicting water quality in Hong Kong. *J. Hydrol.* **2023**, *622*, 129649. [[CrossRef](#)]
32. Halalsheh, N.; Alshboul, O.; Shehadeh, A.; Al Mamlook, R.E.; Al-Othman, A.; Tawalbeh, M.; Almuflih, A.S.; Papelis, C. Breakthrough curves prediction of selenite adsorption on chemically modified zeolite using boosted decision tree algorithms for water treatment applications. *Water* **2022**, *14*, 2519. [[CrossRef](#)]
33. Ridha, H.M.; Hizam, H.; Mirjalili, S.; Othman, M.L.; Ya'acob, M.E.; Ahmadipour, M.; Ismaeel, N.Q. On the problem formulation for parameter extraction of the photovoltaic model: Novel integration of hybrid evolutionary algorithm and Levenberg Marquardt based on adaptive damping parameter formula. *Energy Convers. Manag.* **2022**, *256*, 115403. [[CrossRef](#)]
34. Ranade, N.V.; Nagarajan, S.; Sarvothaman, V.; Ranade, V.V. ANN based modelling of hydrodynamic cavitation processes: Biomass pre-treatment and wastewater treatment. *Ultrason. Sonochem.* **2021**, *72*, 105428. [[CrossRef](#)] [[PubMed](#)]
35. Sibiya, N.P.; Amo-Duodu, G.; Tetteh, E.K.; Rathilal, S. Model prediction of coagulation by magnetised rice starch for wastewater treatment using response surface methodology (RSM) with artificial neural network (ANN). *Sci. Afr.* **2022**, *17*, e01282. [[CrossRef](#)]
36. Purba, L.D.A.; Ibiyeye, H.T.; Yuzir, A.; Mohamad, S.E.; Iwamoto, K.; Zamyadi, A.; Abdullah, N. Various applications of aerobic granular sludge: A review. *Environ. Technol. Innov.* **2020**, *20*, 101045. [[CrossRef](#)]
37. Zaghoul, M.S.; Iorhemen, O.T.; Hamza, R.A.; Tay, J.H.; Achari, G. Development of an Ensemble of Machine Learning Algorithms to Model Aerobic Granular Sludge Reactors. *Water Res.* **2021**, *189*, 116657. [[CrossRef](#)]
38. Zaghoul, M.S.; Achari, G. A Review of Mechanistic and Data-Driven Models of Aerobic Granular Sludge. *J. Environ. Chem. Eng.* **2022**, *10*, 107500. [[CrossRef](#)]
39. Liu, Z.; Lei, J.; Cheng, L.; Yang, R.; Yang, Z.; Shi, B.; Wang, J.; Zhang, A.; Liu, Y. Intelligent Optimal Control Model of Selection Pressure for Rapid Culture of Aerobic Granular Sludge Based on Machine Learning and Simulated Annealing Algorithm. *Bioresour. Technol.* **2024**, *413*, 131509. [[CrossRef](#)]
40. Li, Z.H.; Wang, R.L.; Lu, M.; Wang, X.; Huang, Y.P.; Yang, J.W.; Zhang, T.Y. A Novel Method for Identifying Aerobic Granular Sludge State Using Sorting, Densification and Clarification Dynamics during the Settling Process. *Water Res.* **2024**, *253*, 121336. [[CrossRef](#)]
41. Wang, L.; An, W.; Li, F.T. Text-based corn futures price forecasting using improved neural basis expansion network. *J. Forecast.* **2024**, *43*, 2042–2063. [[CrossRef](#)]
42. Koya, S.R.; Roy, T. Temporal Fusion Transformers for streamflow Prediction: Value of combining attention with recurrence. *J. Hydrol.* **2024**, *637*, 131301. [[CrossRef](#)]
43. Lim, B.; Arik, S.Ö.; Loeff, N.; Pfister, T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [[CrossRef](#)]

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.