



NATURE'S SYMPHONY

ISSN-E: 3007-2034

ISSN-P: 3007-2026

Website link: www.tsfn.com

Machine Learning for Smarter Industrial Wastewater Treatment: Innovations and Insights

Ayesha Asghar, Burhan Khalid, Nadeem Ghafoor, Muhammad Ikram, Muhammad Atiq Ashraf, Nazma Khan, Musrat Shaheen, Jakkampudi Sowjanya, Hafsa Fatima, Muhammad Mohsin Mumtaz and Talha Riaz

To cite: Asghar, A., Khalid, B., Ghafoor, N., Ikram, M., Ashraf, MA., Khan, N., Shaheen, M., Sowjanva, J., Fatima, H., Mumtaz, MM., and Riaz, T. 2025. Machine Learning for Smarter Industrial Wastewater Treatment: Innovations and Insights. *Nature's Symphony*, 3(2):39-63.

Article QR



Published online: 23 September 2025

Submit your article to this journal: <https://tsfjb.com/ojs/index.php/tsfn/about/submissions>

View archive: <https://tsfjb.com/ojs/index.php/tsfn/issue/archive/>



Machine Learning for Smarter Industrial Wastewater Treatment: Innovations and Insights

Ayesha Asghar¹ | Burhan Khalid² | Nadeem Ghafoor³ | Muhammad Ikram² | Muhammad Atiq Ashraf² | Nazma Khan⁴ | Musrat Shaheen⁵ | Jakkampudi Sowjanya⁶ | Hafsa Fatima⁷ | Muhammad Mohsin Mumtaz⁴ | Talha Riaz^{2*}

¹Harbin Institute of Technology, Shenzhen, China

²Huazhong Agricultural University, Wuhan, China

³Southwest University of Science and Technology, Mianyang, China

⁴University of Agriculture, Faisalabad, Pakistan

⁵Government College University, Faisalabad, Pakistan

⁶Central Research Institute on Dryland Agriculture, Hyderabad, India

⁷MNS University of Agriculture, Multan, Pakistan

*Correspondence

Talha Riaz

Email:

talhariaz2844@gmail.com

Keywords:

Industrial wastewater; wastewater treatment; machine learning; innovation; artificial intelligence

ABSTRACT

Industrial wastewater treatment is an important challenge for the environment. It requires a sustainable and efficient approach to decrease the pollutants in wastewater to meet the regulatory standards. Traditional methods for the wastewater treatment are common but face with a lot of challenges such as high-costs, limited efficiency and poor adaptability in handling the complex wastewater. This review paper demonstrates how machine learning (ML) is utilized to address the complex wastewater of industries by enabling predictive maintenance and by using optimized processes. Machine learning (ML) techniques are covered, including supervised learning for pollutant prediction, unsupervised learning for anomaly detection, and reinforcement learning for dynamic process control. These techniques not only improve the treatment precision by reducing chemicals and energy use by roughly 30% and 25% but also support real-time decision-making. Various case studies in chemical, textile and pharmaceutical industries have shown that ML is beneficial for the facilities and the environmental regulations. The presence of problems such as inadequate data, model transparency concerns, and high computing needs reduced its application on a large scale for wastewater treatment. We evaluate emerging solutions, including hybrid ML-mechanistic models, IoT-enhanced sensor networks, and explainable AI (XAI) for regulatory transparency. The paper also outlines future directions, such as autonomous treatment systems and ML's potential to achieve Sustainable Development Goal #6 (Clean Water and Sanitation). Cooperation between researchers, businesses, and legislators is essential to achieving this promise.

1. INTRODUCTION

With the growth of industries worldwide, one of the basic elements of environmental management is effective treatment of industrial waste. Possibly, the release of untreated wastewater is dangerous to human health, the ecosystem and water supply resources.

Industrial effluents are made up of hazardous materials like organic substances, heavy metals, and toxic chemicals. These pollutants, produced by the industries, may damage the environment by polluting the water, biodiversity loss, and leading to long-term

environmental degradation (Chowdhary *et al.*, 2020). Industrial effluents should be treated to encourage sustainability of industrial practice, and this should conform to the rules that are being established for the global economy.

Traditional approaches to wastewater treatment, which include physical, chemical and biological treatment methods, have been in use for many years. Nonetheless, these technologies are limited in numerous ways such as high operating costs, inability to control in real-time and the inability to process complex effluents (Wang *et al.*, 2024). The characteristics of industrial wastewater streams which change according to production cycles and seasonality cannot be addressed using traditional treatment methods because of the diverse and dynamic nature of industrial wastewater. These issues demonstrate the necessity to find another and more innovative method to enhance the efficiency and sustainability of treatment procedures.

Recently, machine learning (ML) has become a disruptive process to the constraints of conventional wastewater treatments. Machine learning (ML) algorithms facilitate real-time decision making, predictive performance and optimization of treatment processes by analyzing large volumes of data and their intricate patterns (Fatima, 2024). Machine learning (ML) models have been applied to forecast level of pollutants, the most suitable chemical doses in various conditions, and the potential malfunctions of the system. The applications have contributed to enhancing the performance of treatment at reduced operating costs (Aparna and Swarnalatha, 2024). ML is even more useful when it is integrated with contemporary technologies, such as big data analytics (BDA) and Internet of Things (IoT) technologies, to manage industrial wastewater.

This review aims to provide an overall review of the advanced technologies in treatment of industrial wastewater. The general aims are to:

- (1) explain the main principles, definitions, and scope of these technologies and their use in wastewater treatment;
- (2) consider the ways to maximize the treatment processes;
- (3) analyze the problems and limitations which may be faced in the course of the implementation; and
- (4) provide recommendations for possible directions of the future research.

In addition to adding to the existing body of knowledge on the application of machine learning in environmental engineering, the purpose of this review is to encourage and inspire additional creative thinking in the research community by connecting case studies and contemporary advances in the field.

However, despite their promising performance, machine learning models face significant challenges in real-world industrial wastewater applications, including data scarcity, limited generalizability across treatment plants, interpretability concerns, and integration with existing process control systems.

2. OVERVIEW OF INDUSTRIAL WASTEWATER TREATMENT

Industrial wastewater is treated for the removal of pollutants in a combined process to remove pollutants. Treatment is typically a combination of physical, chemical, and biological treatment to remove specific types of contaminants. The typical physical methods for particulate and solid suspended particle removal consist of flotation, filtration, and sedimentation (Wang *et al.*, 2024). For dissolved organic and inorganic pollutants, these include chemical processes that involve oxidation, flocculation, coagulation, etc., and advanced oxidation processes (AOPs) (Ricardo *et al.*, 2021). Microorganisms are also employed to degrade organic substances, namely the nutrients phosphorus and nitrogen, in biofilm reactors, activated sludge systems, and anaerobic digestion (Sabliy *et al.*, 2019).



Industrial wastewater is often characterized by the simultaneous presence of a multitude of pollutants such as organic compounds (phenols/ perfumes and pesticides), heavy metals (lead, cadmium and mercury) as well as the newer pollutants are microplastics and Pharmaceuticals (Razzak *et al.*, 2022; Kumar *et al.*, 2025). These contaminants are extremely harmful to life in the aquatic system as well as human health. Heavy metals, for instance, are bio-persistent and pass through the food chain where they could bio-accumulate and bio-magnify (Ray and Vashishth 2024). Lake blooms and oxygen depletion in waterbodies are due to eutrophication from high levels of phosphorus and nitrogen. There are many organic pollutants -polycyclic aromatic hydrocarbons (PAHs)- are induced cancer and mutagenesis (Zhao *et al.*, 2025). Although progress has been made, existing wastewater

treatment technology has numerous drawbacks and inefficiencies. The composition of wastewater may differ greatly according to the industry and production process, and even operating conditions, which poses an intractable problem (Wu *et al.*, 2015).

These fluctuations can lead to less successful conventional remediation techniques which were designed for particular contaminant profiles and can have difficulty adjusting to changing inputs. Moreover, traditional techniques generally involve high energy and chemical utilization, which result in secondary pollution and high operating costs (Kopac, 2021). For instance, chemical coagulation will produce a sludge in need of additional treatment, while biological methods can be slow and vulnerable to shock damage.

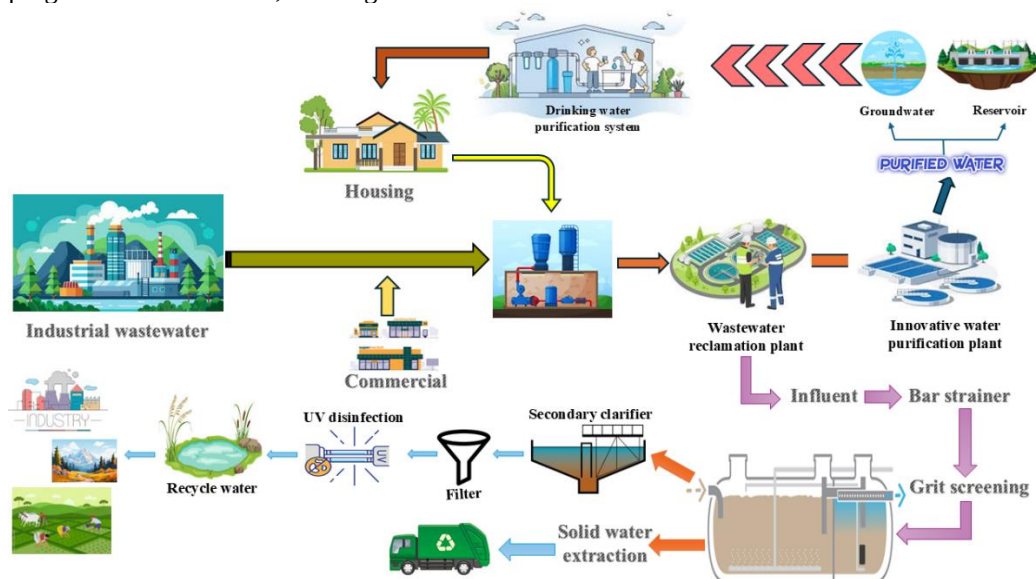


Figure 1. Stages and technologies are noted for a schematic diagram of an industrial wastewater treatment process. The picture illustrates the main procedures of industrial wastewater treatment-influent (raw industrial wastewater entry), primary (bar screens, grit chamber), secondary (clarifier, solid waste removal), and tertiary treatment (UV disinfection, sand filtration, wastewater recycling). The end product is clean water that can be safely discharged or reused. Key stages and technologies are highlighted in this schematic diagram of an industrial wastewater treatment process. The picture shows the basic process of treating industrial wastewater, starting with the

influent stage (raw wastewater entry) and progressing through the primary (bar strainer, grit screening), secondary (clarifier, solid extraction), and advanced treatment stages (UV disinfection, filtration, recycling). Clean water that can be safely released or utilized again is the final result. Other essential features, such as the original (and groundbreaking) purification works and reservoirs, emphasize the physical, chemical, and biological aspects.

Another shortcoming of previous systems is that they did not have real time monitoring and control. The lack of a direct inline, real-time response without the need of repeating laboratory testing and sampling, direct inline monitoring such as operated in the traditional treatment may be slow in responding to changes in quality of the influent or system upset (Du *et al.*, 2014). This inability to continuously monitor the sample is likely to reduce the effectiveness of remediation and result in breaches of environmental rules in place to protect water quality. Additionally, the increasing complexity of industrial wastewater, and particularly the emergence of unregulated pollutants, has highlighted the shortcomings of current treatment technologies to remove all pollutants (Shah *et al.*, 2020).

While emerging treatment technologies such as membrane bioreactors, advanced oxidation processes, and anaerobic digestion offer enhanced removal efficiencies, their large-scale deployment is often constrained by high capital and operational costs, energy demands, and sensitivity to fluctuating influent composition. In industrial settings, variability in organic loads, toxic compounds, and hydraulic conditions can significantly reduce process stability, thereby limiting the scalability and adaptability of these technologies when compared to more robust conventional systems.

Management of industrial wastewater is still a big challenge, and the methods often don't work well. Due to this, there is increasingly an interest in experimenting with newer approaches. Automated systems, smart sensors, and machine learning are also beginning to be applied in treatment plants.

Such techniques provide information to the operators and can even provide a warning about problems before they escalate. These approaches will be able to assist treatment facilities to operate more effectively, generate less waste, and leave a reduced environmental impact.

3. FUNDAMENTS OF MACHINE LEARNING IN WASTEWATER TREATMENT

Machine learning (ML), a subset of artificial intelligence (AI), is currently being implemented in such fields as wastewater treatment and environmental engineering. ML is able to learn on its own, unlike conventional computer programs that require strict instructions. It stores patterns and associations. That is why it is applicable to predicting outcomes, enhancing performance, and assisting individuals in making decisions. In the case of wastewater treatment, it might reduce costs, enhance their efficiency, and address some of the challenges that the conventional methods have experienced (Sakkaravarthy *et al.*, 2024).

Machine learning enhances data analysis and decision-making in industrial wastewater treatment by integrating heterogeneous data sources, including online sensor measurements, laboratory analyses, and historical operational records. By learning nonlinear relationships among process variables, ML models enable predictive monitoring, early fault detection, optimization of treatment parameters, and adaptive process control, thereby improving operational efficiency and regulatory compliance under



dynamic industrial conditions.

The wastewater treatment challenge is the enormous volume of information that is generated daily. Plants employ sensors, surveillance devices, and laboratory analyses that generate big and complex data. It can only be partially explained using standard statistical tools. That's where ML is different. It has the ability to filter through the data, identify the underlying trends, and translate them into workable knowledge. It allows creating predictive models, optimizing systems, and even being able to monitor things in real-time (Zhang *et al.*, 2020; Abualigah, 2025). In industrial wastewater, especially where conditions vary and older approaches often fall short, ML can make a real difference.

Industrial wastewater treatment challenges can be formally expressed as machine learning tasks:

- Regression tasks, such as predicting COD, BOD, nutrient concentrations, or energy consumption
- Classification tasks, including compliance/non-compliance detection and fault classification
- Anomaly detection tasks, where unsupervised models identify deviations from normal operational states
- Sequential decision-making tasks, addressed through reinforcement learning for aeration, dosing, and control

Table 1. Comparison of Machine Learning Techniques in Wastewater Treatment

ML Technique	Applications	Advantages	Limitations	References
Supervised Learning	Predicting pollutant levels, COD/BOD	High-accuracy, interpretable models	Requires labeled data, sensitive to outliers	(Makumbura <i>et al.</i> , 2024; Duan <i>et al.</i> , 2025; Irwan <i>et al.</i> , 2025)
Unsupervised Learning	Clustering, anomaly detection	No need for labeled data, identifies patterns	Less interpretable, requires preprocessing	(Pinto <i>et al.</i> , 2024)
Reinforcement Learning	Process optimization, autonomous systems	Adapts to dynamic conditions, self-learning	High computational cost, complex implementation	(Myakala <i>et al.</i> ; Qi <i>et al.</i> , 2019; Espinosa-Leal <i>et al.</i> , 2020; Kalusivalingam <i>et al.</i> , 2022; Pintos Gómez de las Heras <i>et al.</i> , 2023)
Deep Learning	Image analysis, time-series forecasting	Handles complex data, with high predictive accuracy	Requires large datasets, computationally expensive	(Brownlee, 2018; Torres <i>et al.</i> , 2018; Ismail Fawaz <i>et al.</i> , 2019)

ML model performance in wastewater treatment is commonly evaluated using metrics

such as RMSE, MAE, R^2 , and MAPE for regression tasks, while precision, recall, F1-



score, and ROC-AUC are used for classification. Reinforcement learning performance is typically assessed using cumulative reward, energy savings, and constraint satisfaction. Standardized evaluation metrics are essential for fair comparison across studies and industrial deployment.

3.1 Types of Machine Learning Techniques

The techniques of machine learning are generally divided into three main types: supervised learning, unsupervised learning, and reinforcement learning. All of these work in different ways and are suited to different kinds of problems. In wastewater treatment, the choice of method usually depends on what kind of data is available and what the treatment system is trying to achieve.

Supervised learning is most commonly applied to predicting pollutant concentrations and treatment efficiency due to the availability of labeled historical data. Unsupervised learning is particularly useful for anomaly detection, process clustering, and identification of unknown operational patterns without prior labeling. Reinforcement learning offers unique advantages for real-time process control and energy optimization by learning optimal control policies through interaction with the treatment system.

3.1.1 Supervised Learning

Supervised learning is a machine learning model trained in data that already has the answers. In other words, the inputs are given along with the expected output, and the model learns the connection between these. This kind of approach is quite easy in wastewater treatment. For example, it can be used to predict how much energy a process might need, how efficient the treatment will be or what the level of a certain pollutant could look like (Aparna and Swarnalatha, 2024). Common techniques include methods such as support vector machines (SVMs), decision trees, random forests, and linear regression. SVMs

have been used to estimate wastewater quality indicators such as biochemical oxygen demand (BOD) and chemical oxygen demand (COD) (Qambar and Al Khalidi, 2022).

3.1.2 Unsupervised Learning

Unsupervised learning is applied in cases in which the inputs do not form labelled outputs. Its objective is to make a specific decision of a specific answer but also to identify patterns that are not evident in the data. This can be helpful in activities such as identifying trends within massive datasets, identifying atypical values, or grouping similar wastewater samples with each other (Navato and Mueller, 2021). Among the popular methods, there are principal component analysis (PCA), hierarchical clustering, and k-means clustering. For example, PCA has been used in wastewater (and other) studies to dimensionally reduce the data, which can then be more easily analysed and visualised (Marín Celestino *et al.*, 2018).

3.1.3 Reinforcement Learning

Reinforcement learning (RL) is simply trial and error learning. An agent tries different actions, gets rewarded when it does the right thing, and penalised when it doesn't. Over time, it figures out which decisions work best. In wastewater treatment, RL is starting to be used to improve how processes are managed and how resources are allocated. A good example is in activated sludge plants, where RL has been applied to control aeration. By adjusting aeration in this way, operators have been able to cut down on energy use while still keeping treatment performance at a high level (Gu *et al.*, 2023). In wastewater treatment, RL formulations typically define the state as process variables (e.g., DO, ammonia, flow rate), actions as control decisions (e.g., aeration rate, chemical dosage), and reward functions based on treatment efficiency, energy consumption, and regulatory compliance.

3.2 Key Machine Learning Algorithms



Numerous machine learning methods, each with distinct benefits based on the application, have been effectively used in wastewater treatment. Among the most popular algorithms are:

3.2.1 Neural Networks:

Complex, non-proportional interactions in wastewater data are easily handled by neural networks, particularly deep learning algorithms. Time-series forecasting of water quality metrics and image analysis of sludge settling have both made use of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) (Alvi *et al.*, 2023).

Decision trees are simple, interpretable models that use feature values to separate data into branches. An ensemble of decision trees called a random forest improves prediction accuracy and decreases overfitting. In wastewater treatment facilities, these algorithms have been used to optimise chemical dosage and forecast effluent quality (Fatima, 2024). SVMs perform exceptionally well on tasks involving regression and classification, particularly in high-dimensional domains. They have been used to forecast the effectiveness of treatment and to categorize wastewater samples according to contaminant levels (Ekinici *et al.*, 2023). GBMs, like LightGBM and XGBoost, are sophisticated ensemble methods. That constructs models one after the other in order to fix mistakes from earlier interactions. They have been used in identifying faults in the treatment system and precise forecasting of the water quality indicators.

The possibility to enhance the effectiveness of operations, to reduce costs, and to ensure environmental needs are significant. These are then succeeded by the integration of machine learning technology within wastewater treatment systems. Nonetheless, data quality,

model transparency, and computing resources should all be taken into consideration effectively to implement it. The following sections discuss these issues.

Although deep learning models offer high predictive power, their effectiveness is constrained by data availability. In many industrial settings, limited labeled datasets favor tree-based models or hybrid ML–mechanistic approaches, which often outperform deep networks under data-scarce conditions.

While deep learning models offer high predictive accuracy, their increased complexity often reduces interpretability, which can limit trust and adoption in operational wastewater treatment settings. In contrast, decision trees and ensemble methods such as random forests provide a balance between predictive performance and interpretability, making them more suitable for regulatory compliance and operator-guided decision-making in industrial environments.

4. APPLICATIONS FOR MACHINE LEARNING IN INDUSTRIAL WASTEWATER TREATMENT

Numerous important problems such as resource management, defect detection, process optimization, predictive modelling, and real-time monitoring are being addressed using machine learning (ML) in the context of industrial wastewater treatment. The applications enhance wastewater treatment systems in terms of effectiveness, reliability, and sustainability, as the abilities of machine learning algorithms to analyze complex data, identify trends, and provide valuable information.



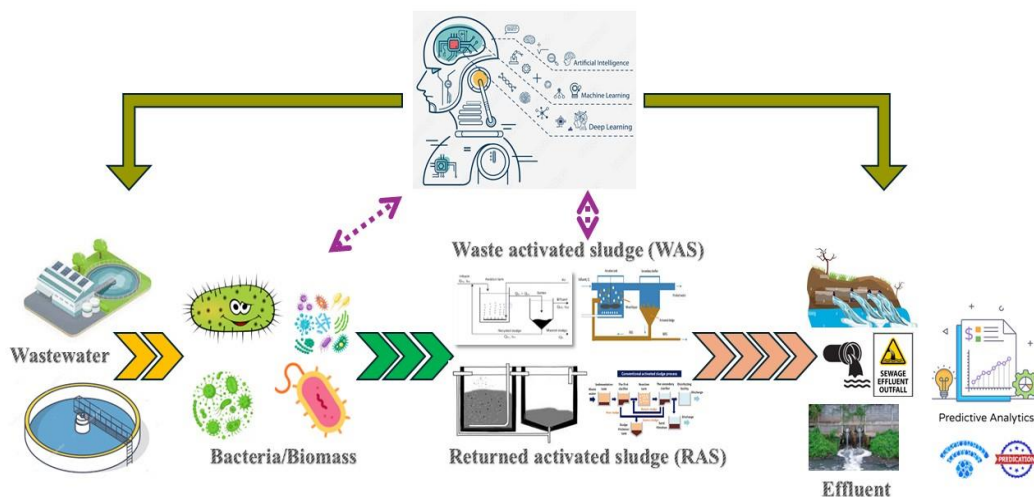


Figure 2. Flow of Activated Sludge Process in Wastewater Treatment. The activated sludge process, which is a crucial biological wastewater treatment technique, is shown in the image. It displays sewage inflow combined with bacteria/biomass and waste-activated sludge (WAS). This breaks down organic contaminants, while some of the sludge (Returned Activated Sludge, or RAS) is recycled to preserve microbial activity. The effluent which has been cleared is segregated to be disposed of or further treatment. To achieve effective treatment, the focus of this procedure should be to find a balance between biomass retention and removing trash (Baskar *et al.*, 2024).

4.1 Predictive Modeling: Forecasting Pollutant Levels and Treatment Efficiency

The most prevalent use of machine learning in wastewater treatment is predictive modelling.

ML models are efficient in forecasting the level of pollutants based on available historical and real-time information, such as chemical oxygen demand (COD), biochemical oxygen demand (BOD) and concentrations of heavy metals (Qambar and Al Khalidy, 2022). One such example is the prediction of COD removal efficiency in the anaerobic digestion processes through the application of artificial neural networks (ANNs). Those that have attained correlation coefficients of above 0.95 (Jeong *et al.*, 2021). Equally, ensemble techniques have been used to predict effluent quality parameters to make proactive changes to treatment

processes such as random forests and gradient boosting machines (GBMs) (Ekinici *et al.*, 2023). These predictive capacities are essential in reduction of chances of the removal of untreated effluents and for ensuring compliance with environmental standards.

4.2 Process Optimization: Enhancing Treatment Processes

The wastewater treatment processes that have been shown to be helpful and effective as process optimization tools encompass aeration, chemical dosing, and sludge control, to mention but a few. As an example, reinforcement learning (RL) algorithms have been used to optimise aeration rates in the activated sludge processes. They have been demonstrated to save up to 25% of the expenditure on energy and maintain the effectiveness of the treatment (Monday *et al.*,

2024). Examples of machine learning (ML) models include support vectors machines (SVMs) and decision trees. These are applied in chemical dosing to identify the best coagulant dosages, lowering chemical usage and operating expenses (Achite *et al.*, 2024). Furthermore, it has been demonstrated that the process optimization of ML-driven sludge dewatering increases solids capture rates and decreases sludge volume, which lowers disposal costs (Farghali and Osman, 2024).

ML-based predictive models utilize historical datasets combined with real-time sensor data to forecast pollutant concentrations and treatment performance indicators. Time-series models capture temporal dependencies, while regression and ensemble learning techniques model nonlinear relationships between influent characteristics, operational parameters, and effluent quality, enabling proactive process adjustments.

4.3 Fault detection and diagnosis: identifying system anomalies and failures

Fault diagnosis and detection are necessary to ensure the reliability and effectiveness of wastewater treatment systems. Anomalies in sensor data and potential system failures have been studied using machine learning algorithms especially in unsupervised learning algorithms such as principal component analysis (PCA) and k-means clustering (Alvi *et al.*, 2023; Murugaiah, 2024). As an example, deep learning models like autoencoders have been applied to detect outliers in standard operating parameters to identify membrane bioreactor (MBR) issues (Frontistis *et al.*, 2023). Early detection of defects makes it possible to take timely corrective measures that minimize downtime and expensive repair.

4.4 Real-time monitoring: using IoT and ML for continuous water quality monitoring

The combined effect of internet of things (IoT) and machine learning (ML) has emerged for

the treatment of wastewater through real-time monitoring. Machine learning models process and demonstrate the constant stream of information and give real-time information about the performance of the system. IoT sensors record data regarding such parameters as pH, turbidity, and dissolved oxygen (Islam *et al.*, 2023). CNNs have been utilized to analyze settle sludge pictures by allowing real-time identification of the efficiency of the sedimentation. Similarly, time-series have also been subjected to recurrent neural networks (RNNs) to predict water quality patterns and enable fast adjustments in the treatment process (Wongburi and Park, 2023). Moreover, they can be monitored in realtime to confirm their adherence to regulations to improve operational efficiency.

4.5 Resource management: Reducing energy and chemical consumption

ML are the tools that play an important role in the treatment of wastewater in a sustainable way within the available resources. ML models help to optimize the usage of energy for treatment plants by employing a high amount of energy in aeration system (Qambar and Al Khalidy, 2022). Optimal operating conditions that enhance treatment performance and reduce energy have been identified with the help of genetic and ML algorithms. ML-based predictive models are also developed to assist in the management of chemicals and to optimize the utilization of coagulants and disinfectants. This practice also reduced the expenses involved in the utilization of different chemicals (Abass *et al.*, 2025). The above discussion demonstrates how one can get significant environmental and financial benefits by utilizing the ML.



5. CASE STUDIES AND SUCCESS STORIES: MACHINE LEARNING IN ACTION

Chemicals, pharmaceuticals, and textiles are just a few of the industries that have demonstrated success in using machine learning (ML) for the treatment of industrial wastewater. These case studies demonstrate how machine learning (ML) may improve the treatment process, change industry-related problems, and have a major positive impact on the economy and environment.

5.1 Textile industry: optimizing dye removal and reducing chemical usage

The textile industry is one of the famous industries for wastewater production. Traditional techniques for the treatment of textile wastewater are very difficult because of the presence of very high levels of organic pollutants and dyes. From the study of many case studies in plant of India, it was investigated that, an ml-based system was used to extract the azo dye from wastewater (Sharmila *et al.*, 2024). This system used to estimate the optimal dose of the oxidant and coagulants based on the combination of supervised learning techniques, including the support vector machines (SVMs) and random forests. By the use of this method, results showed that usage of chemicals reduced by 30% and dye removal efficiency increased by 25 percent. This method not only increases the benefits for improving the regulations of environment but saves the cost used for that purpose.

Another successful example of the application of deep learning models to the textile industry is in effluent quality monitoring in real-time. In one of the plants in Bangladesh, colour and turbidity parameters were monitored in real-time using convolutional neural networks (CNNs) and Internet of Things (IoT) sensors (Mohsin *et al.*, 2024). This method minimized

the chances of non-compliance. The ecological effects of wastewater release were also minimized through the ability to make rapid changes to the treatment processes.

5.2 Chemical industry: enhancing heavy metal removal and energy efficiency

The treatment of wastewater containing hazardous organic compounds and heavy metals presents difficulties for the chemical sector. A German chemical plant used machine learning (ML) to optimise its electrocoagulation process for heavy metal removal (Sharmila *et al.*, 2024). The facility maintained excellent removal efficiency for lead and cadmium and reduced energy usage by 20% through the use of reinforcement learning (RL) algorithms. The ML model continuously adapted to provide optimal performance under a range of influent circumstances of operating parameters like voltage and electrode spacing.

In a separate case, an ML-based fault detection system was implemented by a US chemical manufacturing firm. This aimed at identifying anomalies in their wastewater treatment processes (Imen *et al.*, 2023). The technique reduced downtime by 15% and extended the lifetime of membranes by 20%. Unsupervised learning techniques including principal component analysis (PCA) and k-means clustering were used to detect early indications of membrane fouling in reverse osmosis units.

5.3 Pharmaceutical industry: addressing emerging contaminants and improving process control

Wastewater from the pharmaceutical sector contains complex organic molecules, such as active pharmaceutical ingredients (APIs). These are challenging for traditional treatment techniques to break down. A Swiss pharmaceutical facility included machine learning (ML) models to improve API



Table 2. Case Studies of ML Applications in Industrial Wastewater Treatment

Industry	Application	ML Technique Used	Results/Improvements	References
Textile	Dye removal optimization	Random Forests, SVMs	30% reduction in chemical usage	(Zaharia <i>et al.</i> , 2021; Ganthavee <i>et al.</i> , 2024; Liu <i>et al.</i> , 2024)
Chemical	Heavy metal removal	Reinforcement Learning	20% energy savings, high removal efficiency	(Lakshmi <i>et al.</i> , 2021; Barkhordari <i>et al.</i> , 2024)
Pharmaceutical	API degradation	Gradient Boosting Machines	40% improvement in API removal	(Mizera <i>et al.</i> ; Jiang <i>et al.</i> , 2023; Vora <i>et al.</i> , 2023)
Food Processing	Biogas production optimization	Ensemble Learning	30% increase in biogas yield	(Olatunji <i>et al.</i> , 2024; Sidi Habib <i>et al.</i> , 2024; Duan <i>et al.</i> , 2025)

degradation into its advanced oxidation process (AOP) (Thakur *et al.*, 2023). A 40% increase in API removal efficiency was attained by the ML system that was based on gradient boosting machines (GBMs) and predicted the ideal oxidant and reaction time combination.

A Japanese pharmaceutical company used machine learning (ML) to operate its biological treatment system in real time (Wasalathanthri *et al.*, 2020). The company optimised aeration rates and nutrient dosing to analyse time-series data from bioreactors by using recurrent neural networks (RNNs). This resulted in a 15% increase in treatment capacity and a 25% reduction in energy usage.

5.4 Cross-industry success: resource recovery and circular economy

Machine learning has helped wastewater treatment transition to a circular economy and recover resources. An ML-based system was

used in the Netherlands in which extraction of phosphorus from sludge was done at a municipal wastewater treatment facility (Sheikh *et al.*, 2023). The system achieved 85% recovery efficiency by optimising precipitation conditions. This also predicted phosphorus recovery rates using supervised learning techniques. This produced a lucrative by-product for agricultural use in addition to lessening the environmental impact of phosphorus release. In a similar vein, ML was utilised by a Brazilian food processing facility to maximise the production of biogas through anaerobic digestion (Cruz *et al.*, 2022). By predicting ideal feedstock compositions and digestion times, the machine learning model—which was based on ensemble learning techniques—was able to reduce greenhouse gas emissions and increase biogas generation by 30%.

Table 3. Challenges and Solutions for ML Implementation in Wastewater Treatment

Challenge	Description	Potential Solutions	References
Data Quality and Availability	Inconsistent monitoring, missing data, noise	Advanced sensors, data preprocessing, collaborative data sharing	(Liu <i>et al.</i> , 2021; Bertrand <i>et al.</i> , 2022; Bernardi <i>et al.</i> , 2023; Zhao <i>et al.</i> , 2024)
Model Interpretability	Black-box nature of advanced ML models	Explainable AI (XAI) techniques (e.g., SHAP, LIME)	(Zhong <i>et al.</i> , 2022; Hassija <i>et al.</i> , 2024)
Computational Costs	High resource requirements for training and deployment	Lightweight algorithms, edge computing, cloud-based platforms	(Andriulo <i>et al.</i> , 2024; Cao <i>et al.</i> , 2024; Jouini <i>et al.</i> , 2024)
Integration with Infrastructure	Traditional systems not designed for ML	Retrofitting with IoT, modular ML systems, workforce training	(Elahi <i>et al.</i> , 2023; Rehman <i>et al.</i> , 2024; Khan <i>et al.</i> , 2025; Mendonca <i>et al.</i> , 2025)
Regulatory and Ethical Issues	Compliance with standards, algorithmic bias, data privacy	Robust regulatory frameworks, transparent decision-making, ethical guidelines	(Hanna <i>et al.</i> , 2024; Tilala <i>et al.</i> , 2024; Radanliev, 2025)

6. CHALLENGES AND CONSTRAINTS OF MACHINE LEARNING IN INDUSTRIAL WASTEWATER TREATMENT

The possibilities of machine learning (ML) in revolutionizing wastewater treatment in industries are immense, yet there are challenges associated with its implementation. These constraints that cross the technological, operational, and infrastructure boundaries should be mitigated to ensure that industries realize the full benefits of machine learning.

6.1 Data quality and availability issues

The success of any data-driven model depends on both the quality and amount of data to be used in training and validation. Industrial wastewater treatment may pose a challenge in data collection because of sensor breakdowns, missing data, and inconsistencies in monitoring (Ba-Alawi *et al.*, 2022). Poorly-quality data,

including that containing noise, outliers, and biased information, may cause inaccurate output and unreliable performance (Karkouch *et al.* 2016).

Different industries have no specific format of data and exchange protocols, which become more difficult for the development of practical and standardized analytical systems (Lwakatare *et al.*, 2020). These problems need a huge investment in better data processing techniques, high-quality sensor technologies and collaborative structures to share the data.

Poor-quality data, including sensor noise, missing values, inconsistent sampling frequencies, and lack of standardization across facilities, can significantly degrade the accuracy and reliability of ML models. These issues may lead to biased predictions, reduced generalizability, and unstable model performance, highlighting the need for robust data preprocessing, validation, and quality



assurance protocols.

6.2 Model interpretability and transparency

The complicated and dense nature of many computational models, like deep neural networks and ensemble models, poses a significant barrier for their use in the treatment of wastewater (Imen *et al.*, 2023). These decision-making models give great accuracy of prediction, but sometimes these models become very difficult to understand because of their complex structure. Complex computational systems made it very difficult for workers to use in industries especially where legal and safety compliance are major considerations (Emeihe *et al.*, 2024). Therefore, researchers are working on these complex models to address this problem by creating strategies like Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP) which are very useful to improve the understanding level of model outputs. For example, SHAP analysis has been used to quantify the influence of influent COD, temperature, and hydraulic retention time on effluent quality predictions, enabling operators to trust ML-based recommendations.

6.3 High computational costs and resource requirements

The creation and use of powerful computational models may require the use of high-performance technology, including graphics processing units (GPUs), and extensive computing power (Murugaiah, 2024). Nonetheless, the requirements may be prohibitively expensive to small and medium-sized businesses (SMEs) that have minimal funds or technology infrastructure (Matsoso *et al.*, 2021). Additionally, large and complex datasets may consume a lot of time and delay real-time solutions (Lwakatare *et al.*, 2020). One way of overcoming these problems is by researchers exploring distributed systems, cloud platforms, and lightweight methods of

computing to improve accessibility, efficiency and to reduce the processing requirements.

6.4 Incorporation with existing infrastructure

There are various practical and technical challenges to integrating modern data-driven technology into the current wastewater treatment infrastructure. The facilities remain mostly equipped with outdated technology that was not designed to support the processing of complex data and real-time monitoring (Chow *et al.*, 2018). Therefore, it may be difficult to add Internet of Things (IoT) sensors, data collection tools, and automated control systems to these systems (Bhadauria, 2024). The lack of skilled specialists to use and manage such innovative technologies poses an extra barrier for their adoption in a successful way (Sun and Jung, 2024). To overcome these problems, researchers, business leaders, and legislators should collaborate to come up with practical applications with solutions that should be easy to implement for the treatment of wastewater.

6.5 Regulatory and ethical concern

Ethical and regulatory concerns connected to the use of machine learning to treat wastewater are also very serious. When making decisions by using these methods, industries must comply with the prescribed rules and regulations that govern the environment (Singh *et al.*, 2024a; Singh *et al.*, 2024b). Ethical concerns can be a problem when false or prejudiced data can result in unreliable results, particularly where the well-being of people and the environment is at stake (Albahri *et al.*, 2023). This is why it is necessary to establish effective ethical standards and robust legal systems to promote the safe and transparent use of machine learning in wastewater management.



6.6 Transferability across treatment plants

A key limitation of current ML applications is limited generalizability across facilities due to plant-specific influent characteristics and operational strategies. Emerging approaches such as transfer learning, federated learning, and domain adaptation offer promising solutions by allowing models trained in one context to be adapted to new plants with minimal retraining.

7. FUTURE DIRECTIONS AND OPPORTUNITIES

The use of machine learning in wastewater treatment is also associated with several ethical and regulatory issues. In the case when such systems are used to facilitate the decision-making process during the important treatment processes, it is relevant that industries can adhere to all the appropriate environmental laws and standards of operations (Singh *et al.*, 2024a; Singh *et al.*, 2024b). Ethical issues arise when inaccurate or discriminatory data results, inaccurate predictions, particularly when human health or the environment may be impacted (Albahri *et al.*, 2023). To mitigate these dangers, the responsible use of machine learning in wastewater treatment should be guided by clear ethical principles and robust regulatory action.

7.1 Emerging trends: deep learning and hybrid models

A subbranch of machine learning known as deep learning can process big data with non-linear relationships. It has become a topic of great interest within the wastewater treatment industry. The methods that are most frequently used include Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to perform different tasks, including image analysis of sludge-settling and prediction of water quality parameters across time (Borzooei *et al.*, 2024). Moreover, cross-hybrid models that combine methods based on

data with ideas of physical or mechanistic models are gaining relevance to enhance model interpretability and prediction quality (Wang *et al.*, 2022). These types of hybrid models can be utilized to optimise reactor design and overall increase the performance of treatment; it involves combining Computational Fluid Dynamics (CFD) with machine learning (Alotaibi *et al.*, 2024).

7.2 Potential for AI-driven autonomous systems

This is the future of wastewater management through the development of AI-powered self-learning devices, which are able to draw conclusions and make improvements in real time. This is motivated by reinforcement learning (RL) and deep reinforcement learning (DRL). They allow the systems to trial and judge the most appropriate control schemes (Al-Saadi *et al.*, 2023). As an example, autonomous systems that are DRL-optimized can save a huge sum of money and energy by streamlining the chemical dosing and aeration phases (Nam *et al.*, 2020). When combined with new technologies in computing and IoT, these solutions will become much more versatile and adaptable in numerous industrial environments.

7.3 Role of ML in attaining sustainable development goals (SDGs)

Machine learning is directly relevant to the realization of the United Nations Sustainable Development Goals (SDGs), in particular, Goal 6 (Clean Water and Sanitation) and Goal 9 (Industry, Innovation, and Infrastructure). Such data-based solutions can contribute to the reduction of the environmental impact of industrial operations and the process optimization of resources utilization and quality control of water (Kumar *et al.*, 2025). Machine learning models have been used to safeguard biodiversity too by anticipating and mitigating the effect of wastewater discharge over aquatic habitats (Yang and Xiong, 2024).



Moreover, the techniques aid the recovery of valuable resources from wastewater, such as energy and nutrients, and, by extension, the transition to a sustainable and circular industrial economy (Soo *et al.*, 2024).

Machine learning contributes to achieving Sustainable Development Goals by enhancing water quality monitoring and treatment efficiency (SDG 6), reducing energy consumption through process optimization (SDG 7), and enabling resource recovery and waste minimization strategies that support circular economy principles (SDG 12).

7.4 Collaboration between researchers, industries, and policymakers

Machine learning could be highly instrumental in wastewater treatment, and close cooperation between researchers, industries, and policymakers is required. Academic efforts should focus on developing models which are scalable, interpretable, and energy-efficient to address real-world treatment issues (Ahmad *et al.*, 2022). Industrial stakeholders will be able to obtain the full potential of these technologies by investing in trustworthy data systems, staff development, and modern sensor devices. Policymakers, in their turn, have a significant part to play in the establishment of the legal and institutional framework, which will facilitate data sharing and provide incentives for adopting technology-based solutions (Chy and Buadi, 2024). Also, there are the public-private alliances and industrial partnerships that are a good medium of sharing knowledge and enhancing rapid innovation in wastewater control.

7.5 Ethical and social considerations

The growing use of machine learning in the process of wastewater treatment is associated with a number of ethical and social concerns. To develop trust in the community and promote greater utilization, the privacy of data needs to be secured, bias in analysis models should be

avoided, and fair access to technological solutions should be ensured (Oladosu *et al.*, 2024). The sustainability of such technologies will also rely on the long-term decision-making processes that are transparent, inclusive, and guided by sound ethics.

8. FUTURE DIRECTIONS

The integration of machine learning in industrial wastewater treatment is a significant change in the process of managing and enhancing the treatment systems by industries. This review has demonstrated that machine learning can be useful in resource optimization, fault detection, real-time process monitoring, predictive control, and overall system improvement in a number of important aspects. These approaches can ensure that wastewater treatment processes become more effective, cost-efficient, and sustainable when used adequately, as they will ensure a higher level of compliance with environmental regulations and enhanced resource recovery.

Applications in various industries, such as textiles, chemicals, and medicines exhibit quantitative changes in the treatment effectiveness, energy efficiency, and chemical efficiency. Although these benefits exist, several challenges still hinder large-scale adoption. The quality of data, the transparency of the model, the requirements of the computation and the compatibility with existing infrastructure will be of concern. Their solution will require further investment in technology and training and a more concerted effort, between the researchers, engineers, and industry specialists.

Beyond improving operational efficiency, machine learning may also lead to new opportunities. It encourages the creation of intelligent systems that are capable of autonomous decision-making and ongoing process optimization, supporting the broader objectives of sustainability and the circular



economy. Its strategic importance in the development of sustainable industries is highlighted by their relevance to global initiatives such as the Sustainable Development Goals of the UN, particularly SDG 6 (Clean Water and Sanitation) and SDG 9 (Industry, Innovation, and Infrastructure).

Future advancements will be predicated on cooperation among scientists, decision-makers, and business executives. The development of transparent, energy-efficient, and real-world adaptable models should be the aim of researchers. In order to guarantee implementation, industries should simultaneously invest in modern sensors, data systems, and capacity building initiatives. By creating mechanisms that encourage creativity and responsible data sharing, policymakers may accelerate this progress.

8. CONCLUSION

The present review paper provides an in-depth and recent discussion on the role of machine learning in increasing the aspect of industrial wastewater treatment. It introduces details of recent developments, real world applications, and the difficulties of implementing these methods to enhance efficiency of treatment, reduce environmental effects, and facilitate sustainable growth. There is a good chance that the increasing application of machine learning will revolutionize the industrial wastewater treatment sector. Practice-based innovation, multidisciplinary research, and long-term collaboration will be necessary to realize this potential.

Acknowledgements

We would like to express our deep gratitude to Dr. Rizwan Maqbool of the Department of Agronomy, University of Agriculture, Faisalabad, Pakistan, who gave us precious advice and other resources in order to prepare this review paper.

Author's Contribution

The review was conceived and designed by AA, MMN, BK, and MS. The manuscript was written by MI, MAA, HF, and NG. TR and JS critically revised it.

Funding

No specific grant from a public, private, or nonprofit funding organization was obtained for this work.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Conflict of interest

The authors have declared no conflict of interest.

REFERENCES

- Abass, N., Khan, S.A., Jillani, S.M.S., Ali, S. and Mansha, M. 2025. Advancements in polymeric membranes: Fabrication techniques and innovative solutions for enhanced performance in reverse osmosis. *Asian Journal of Organic Chemistry*, 14: e202400678.
- Abualigah, L. 2025. Enhancing real-time data analysis through advanced machine learning and data analytics algorithms. *International Journal of Online & Biomedical Engineering*. 21: 4-25.
- Achite, M., Farzin, S., Elshaboury, N., Anaraki, M.V., Amamra, M. and Toubal, A.K. 2024. Modeling the optimal dosage of coagulants in water treatment plants using various machine learning models. *Environment, Development and Sustain*, 26:3395–3421.
- Ahmad, T., Madonski, R., Zhang, D., Huang, C. and Mujeeb, A. 2022. Data-driven probabilistic machine learning in sustainable smart energy/smart energy



- systems: Key developments, challenges, and future research opportunities in the context of smart grid paradigm. *Renewable and Sustainable Energy Reviews*, 160:112128.
- Albahri, A.S., Duhaime, A.M., Fadhel, M.A., Alnoor, A., Baqer, N.S., Alzubaidi, L. Albahri, O.S., Alamoodi, A.H., Bai, J. and Salhi, A. 2023. A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information Fusion*, 96:156–191.
- Alotaibi, F.N., Berrouk, A.S. and Salim, I.M. 2024. Scaling up dry methane reforming: Integrating computational fluid dynamics and machine learning for enhanced hydrogen production in industrial-scale fluidized bed reactors. *Fuel*, 376:132673
- Al-Saadi, M., Al-Greer, M. and Short, M. 2023. Reinforcement learning-based intelligent control strategies for optimal power management in advanced power distribution systems: A survey. *Energies (Basel)*, 16:1608.
- Alvi, M., Batstone, D., Mbamba, C.K., Keymer, P., French, T., Ward, A., Dwyer, J. and Cardell-Oliver, R. 2023. Deep learning in wastewater treatment: a critical review. *Water Research*, 245:120518.
- Andriulo, F.C., Fiore, M., Mongiello, M., Traversa, E. and Zizzo, V. 2024. Edge computing and cloud computing for internet of things: A review. *Informatics*, 11:71.
- Aparna, K.G. and Swarnalatha, R. 2024. Optimizing wastewater treatment plant operational efficiency through integrating machine learning predictive models and advanced control strategies. *Process Safety and Environmental Protection*, 188:995–1008.
- Ba-Alawi, A.H., Loy-Benitez, J., Kim, S. and Yoo, C. 2022. Missing data imputation and sensor self-validation towards a sustainable operation of wastewater treatment plants via deep variational residual autoencoders. *Chemosphere*, 288:132647.
- Barkhordari, M.S., Zhou, N., Li, K. and Qi, C. 2024. Interpretable machine learning for predicting heavy metal removal efficiency in electrokinetic soil remediation. *Journal of Environmental Chemical Engineering*, 12:114330.
- Baskar, G., Omer, S.N., Saravanan, P., Rajeshkannan, R., Saravanan, V., Rajasimman, M. and Shanmugam, V. 2024. Status and future trends in wastewater management strategies using artificial intelligence and machine learning techniques. *Chemosphere*, 362:142477.
- Bernardi, F.A., Alves, D., Crepaldi, N., Yamada, D.B., Lima, V.C. and Rijo, R. 2023. Data quality in health research: integrative literature review. *Journal of Medical Internet Research*, 25:e41446.
- Bertrand, Y., Van Belle, R., De Weerd, J. and Serral, E. 2022. Defining data quality issues in process mining with IoT data. In: *International Conference on Process Mining, Springer*, 422–434.
- Bhadauria, P.K.S. 2024. Comprehensive review of AI and ML tools for earthquake damage assessment and retrofitting strategies. *Earth Science Informatics*, 17:3945–3962.
- Borzooei, S., Scabini, L., Miranda, G., Daneshgar, S., Deblieck, L., Bruno, O., De Langhe, P., De Baets, B., Nopens, I. and Torfs, E. 2024. Evaluation of activated sludge settling characteristics from microscopy images with deep convolutional neural networks and transfer



- learning. *Journal of Water Process Engineering*, 64:105692.
- Brownlee, J. 2018. Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in Python. *Machine Learning Mastery*.
- Cao, L., Huo, T., Li, S., Zhang, X., Chen, Y., Lin, G., Wu, F., Ling, Y., Zhou, Y. and Xie, Q. 2024. Cost optimization in edge computing: a survey. *Artificial Intelligence Review*, 57:312.
- Chow, C.W.K., Liu, J., Li, J., Swain, N., Reid, K. and Saint, C.P. 2018. Development of smart data analytics tools to support wastewater treatment plant operation. *Chemometrics and Intelligent Laboratory Systems*, 177:140–150.
- Chowdhary, P., Bharagava, R.N., Mishra, S. and Khan, N. 2020. Role of industries in water scarcity and its adverse effects on environment and human health, In: Shukla, V., Kumar, N. (eds) *Environmental Concerns and Sustainable Development*. Springer, Singapore.
- Chy, M.K.H. and Buadi, O.N. 2024. Role of machine learning in policy making and evaluation. *International Journal of Innovative Science and Research Technology*, 9:456-463.
- Cruz, I.A., Chuenchart, W., Long, F., Surendra, K.C., Andrade, L.R.S., Bilal, M., Liu, H., Figueiredo, R.T., Khanal, S.K. and L.F.R. Ferreira. 2022. Application of machine learning in anaerobic digestion: Perspectives and challenges. *Bioresource Technology*, 345:126433.
- Du, B., Price, A.E., Scott, W.C., Kristofco, L.A., Ramirez, A.J., Chambliss, C.K., Yelderman, J.C. and Brooks, B.W. 2014. Comparison of contaminants of emerging concern removal, discharge, and water quality hazards among centralized and on-site wastewater treatment system effluents receiving common wastewater influent. *Science of the Total Environment*, 466:976–984.
- Duan, J., Cao, G., Ma, G. and Yazdani, B. 2025. Boosting biogas production through innovative data-driven modeling and optimization methods at NJWTP. *Scientific Reports*, 15:4814.
- Ekinci, E., Özbay, B., Omurca, S.İ., Sayın, F.E. and Özbay, İ. 2023. Application of machine learning algorithms and feature selection methods for better prediction of sludge production in a real advanced biological wastewater treatment plant. *Journal of Environmental Management*, 348:119448.
- Elahi, M., Afolaranmi, S.O., Lastra, J.L.M and Garcia, J.A.P. 2023. A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment. *Discover Artificial Intelligence*, 3:43.
- Emeihe, E.V., Nwankwo, E.I., Ajegbile, M.D., Olaboye, J.A. and Maha, C.C. 2024. The impact of artificial intelligence on regulatory compliance in the oil and gas industry. *International Journal of Life Science Research Archive*, 7:28–39.
- Espinosa-Leal, L., Chapman, A. and Westerlund, M. 2020. Autonomous industrial management via reinforcement learning: Towards self-learning agents for decision-making. *Journal of Intelligent & Fuzzy Systems*, 39:8427–8439.
- Farghali, M. and Osman, A.I. 2024. Revolutionizing waste management: unleashing the power of artificial intelligence and machine learning. In: *Advances in Energy from Waste*, Elsevier, 225–279.



- Fatima, S. 2024. Improving healthcare outcomes through machine learning: applications and challenges in big data analytics. *International Journal of Advanced Research in Engineering Technology*, 11:2349-2819
- Frontistis, Z., Lykogiannis, G. and Sarmpanis, A. 2023. Machine learning implementation in membrane bioreactor systems: progress, challenges, and future perspectives: a review. *Environments*, 10:127.
- Ganthavee, V., Fernando, M.M.R. and Trzcinski, A.P. 2024. Monte carlo simulation, artificial intelligence and machine learning-based modelling and optimization of three-dimensional electrochemical treatment of Xenobiotic Dye wastewater. *Environmental Processes*, 11:41.
- Gu, Y., Li, Y., Yuan, F. and Yang, Q. 2023. Optimization and control strategies of aeration in WWTPs: A review. *Journal of Cleaner Production*, 418:138008.
- Hanna, M., Pantanowitz, L., Jackson, B., Palmer, O., Visweswaran, S., Pantanowitz, J., Deebajah, M. and H. Rashidi. 2024. Ethical and bias considerations in artificial intelligence (AI)/machine learning. *Modern Pathology*, 38:100686.
- Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M. and Hussain, A. 2024. Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation*, 16:45–74.
- Imen, S., Croll, H.C., McLellan, N.L., Bartlett, M., Lehman, G. and Jacangelo, J.G. 2023. Application of machine learning at wastewater treatment facilities: A review of the science, challenges and barriers by level of implementation. *Environmental Technology Reviews*, 12:493–516.
- Irwan, D., Ibrahim, S.L., Latif, S.D., Winston, C.A., Ahmed, A.N., Sherif, M., El-Shafie, A.H. and El-Shafie, A. 2025. River water quality monitoring using machine learning with multiple possible *in-situ* scenarios. *Environmental and Sustainability Indicators*, 26:100620.
- Islam, M.M., Kashem, M.A., Alyami, S.A. and Moni, M.A. 2023. Monitoring water quality metrics of ponds with IoT sensors and machine learning to predict fish species survival. *Microprocessors and Microsystems*, 102:104930.
- Fawaz, H.I., Forestier, G., Weber, J., Idoumghar, L. and Muller, P.A. 2019. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33:917–963.
- Jeong, K., Abbas, A., Shin, J., Son, M., Kim, Y.M. and Cho, K.H. 2021. Prediction of biogas production in anaerobic co-digestion of organic wastes using deep learning models. *Water Research*, 205:117697.
- Jiang, J., Lu, A., Ma, X., Ouyang, D. and Williams III, R.O. 2023. The applications of machine learning to predict the forming of chemically stable amorphous solid dispersions prepared by hot-melt extrusion. *International Journal of Pharmaceutics: X*, 5:100164.
- Jouini, O., Sethom, K., Namoun, K., Aljohani, N., Alanazi, M.H. and Alanazi, M.N. 2024. A survey of machine learning in edge computing: Techniques, frameworks, applications, issues, and research directions. *Technologies (Basel)*, 12:81.
- Kalusivalingam, A.K., Sharma, A., Patel, N. and Singh, V. 2022. Optimizing autonomous factory operations using



- reinforcement learning and deep neural networks. *International Journal of AI and ML*, 3.
- Karkouch, A., Mousannif, H., Al Moatassime, H. and Noel, T. 2016. Data quality in internet of things: A state-of-the-art survey. *Journal of Network and Computer Applications*, 73:57–81.
- Khan, M.I., Yasmeen, T., Khan, M., Hadi, N.U., Asif, M., Farooq, M. and Al-Ghamdi, S.G. 2025. Integrating industry 4.0 for enhanced sustainability: Pathways and prospects. *Sustainable Production and Consumption*, 54:149–189.
- Kopac, T. 2021. Emerging applications of process intensification for enhanced separation and energy efficiency, environmentally friendly sustainable adsorptive separations: A review. *International Journal of Energy Research*, 45:15839–15856.
- Kumar, A., Sheik, A.G., Bux, F. and Kumari, S. 2025. Machine learning and artificial intelligence application in automotive water quality monitoring, analysis, and management In: Dubey, A. K., Srivastav, A. L., Kumar, A., Garcia Marquez, F. P. and Giannakoudakis, D. A. [Eds.] *Computational Automation for Water Security*. Elsevier. pp. 89-109.
- Lakshmi, D., Akhil, D., Kartik, A., Gopinath, K.P., Arun, J., Bhatnagar, A., Rinklebe, J., Kim, W. and Muthusamy, G. 2021. Artificial intelligence (AI) applications in adsorption of heavy metals using modified biochar. *Science of the Total Environment*, 801:149623.
- Liu, C., Peng, G., Kong, Y., Li, S. and Chen, S. 2021. Data quality affecting Big Data analytics in smart factories: research themes, issues and methods. *Symmetry (Basel)*, 13:1440.
- Liu, C., Balasubramanian, P., Li, F. and Huang, H. 2024. Machine learning prediction of dye adsorption by hydrochar: Parameter optimization and experimental validation. *Journal of Hazardous Materials*, 480:135853.
- Lwakatare, L.E., Raj, A., Crnkovic, I., Bosch, J. and Olsson, H.H. 2020. Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions. *Information and Software Technology*, 127:106368.
- Makumbura, R.K., Mampitiya, L., Rathnayake, N., Meddage, D.P.P., Henna S., Dang, T.L., Hoshino, Y. and Rathnayake, U. 2024. Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence (XAI) techniques like shapley additive explanations (SHAP) for interpreting the black-box nature. *Results in Engineering*, 23:102831.
- Celestino, A.E.M., Cruz, D.A.M., Sánchez, E.M.O., Reyes, F.G. Soto, D.V. 2018. Groundwater quality assessment: An improved approach to K-means clustering, principal component analysis and spatial analysis: A case study. *Water (Basel)*, 10:437.
- Matsoso, M.L., Nyathi, M. and Nakpodia, F.A. 2021. An assessment of budgeting and budgetary controls among small and medium-sized enterprises: evidence from a developing economy. *Journal of Accounting in Emerging Economies*, 11:552–577.
- Mendonca, R.S., da Silva, M., Ayres Jr, F.A.C., Bessa, I.V., Medeiros, R.L.P. and Lucena Jr, V.F. 2025. Development of a novel retrofit framework considering industry 4.0 concepts: a case study of a modular production system. *Processes*, 13:136.



- Mizera, M., Muratov, E.N., Alves, V.M., Tropsha, A. and Cielecka-Piontek, J. Computer-aided discovery of new solubility-enhancing drug delivery system. *Biomolecules*, 2020; 10: 913.
- Mohsin, A.S.M., Choudhury, S.H. and Das, B. 2024. Deployment of real time effluent treatment plant monitoring and future prediction using machine learning. *International Journal of Environmental Science and Technology*, 22: 5987–6004.
- Monday, C., Zaghloul, M.S., Krishnamurthy, D. and Achari, G. 2024. A review of AI-driven control strategies in the activated sludge process with emphasis on aeration control. *Water (Basel)*, 16:305.
- Murugaiah, M. 2024. Application of machine learning and deep learning in high performance computing. In: *High Performance Computing in Biomimetics: Modeling, Architecture and Applications*, Springer, 271–286.
- Myakala, P.K., Bura, C., Jonnalagadda, A.K. and Jakku, P.C. A unified framework for self-learning ai: reinforcement learning, neural search, and adaptive evolution. *International Journal of Computer Applications*, 975:8887.
- Nam, K., Heo, S., Loy-Benitez, J., Ifaei, P. and Yoo, C. 2020. An autonomous operational trajectory searching system for an economic and environmental membrane bioreactor plant using deep reinforcement learning. *Water Science and Technology*, 81:1578–1587.
- Navato, A.P. and Mueller, A.V. 2021. Enabling automatic detection of anomalies in wastewater: A highly simplified approach to defining “normal” in complex chemical mixtures. *Frontiers in Water*, 3:734361.
- Oladosu, S.A., Ike, C.C., Adepoju, P.A., Afolabi, A.I., Ige, A.B. and Amoo, O.O. 2024. Frameworks for ethical data governance in machine learning: Privacy, fairness, and business optimization. *Magna Scientia Advanced Research Reviews*, 7:96–106.
- Olatunji, O.O., Adedeji, P.A., Madushele, N., Rasmeni, Z.Z. and van Rensburg, N.J. 2024. Evolutionary optimization of biogas production from food, fruit, and vegetable (FFV) waste. *Biomass Conversion and Biorefinery*, 14:12113–12125.
- Parisineni, S.R.A. and Pal, M. 2024. Enhancing trust and interpretability of complex machine learning models using local interpretable model agnostic shap explanations. *International Journal of Data Science and Analytics*, 18:457–466.
- Pinto, A., Herrera, L.-C., Donoso, Y. and Gutierrez, J.A. 2024. Enhancing critical infrastructure security: unsupervised learning approaches for anomaly detection. *International Journal of Computational Intelligence Systems*, 17:236.
- Pintos Gómez de las Heras, B., Martínez-Tomás, R. and Troncoso, J.M.C. 2023. Self-learning robot autonomous navigation with deep reinforcement learning techniques. *Applied Sciences*, 14:366.
- Qambar, A.S. and Al Khalidy, M.M. 2022. Optimizing dissolved oxygen requirement and energy consumption in wastewater treatment plant aeration tanks using machine learning. *Journal of Water Process Engineering*, 50:103237.
- Qi, X., Luo, Y., Wu, G., Boriboonsomsin, K. and Barth, M. 2019. Deep reinforcement learning enabled self-learning control for energy efficient driving. *Transportation Research Part C: Emerging Technologies*, 99:67–81.



- Radanliev, P. 2025. AI ethics: integrating transparency, fairness, and privacy in AI development. *Applied Artificial Intelligence*, 39:2463722.
- Ray, S. and Vashishth, R. 2024. From water to plate: Reviewing the bioaccumulation of heavy metals in fish and unraveling human health risks in the food chain. *Emerging Contaminants*, 10:100358.
- Razzak, S.A., Faruque, M.O., Alsheikh, Z., Alsheikhmohamad, L., Alkuroud, D., Alfayez, A., Hossain, S.M.Z. and Hossain, M.M. 2022. A comprehensive review on conventional and biological-driven heavy metals removal from industrial wastewater. *Environmental Advances*, 7:100168.
- Rehman, Z., Tariq, N., Moqurrah, S.A., Yoo, J. and Srivastava, G. 2024. Machine learning and internet of things applications in enterprise architectures: Solutions, challenges, and open issues. *Expert Systems*, 41:e13467.
- Ricardo, I.A., Alberto, E.A., Júnior, A.H.S., Macuvele, D.L.P., Padoin, N., Soares, C., Riella, H.G., Starling, M.C.V.M. and Trovo, A.G. 2021. A critical review on microplastics, interaction with organic and inorganic pollutants, impacts and effectiveness of advanced oxidation processes applied for their removal from aqueous matrices. *Chemical Engineering Journal*, 424:130282.
- Sabliy, L., Kuzminskiy, Y., Zhukova, V., Kozar, M. and Sobczuk, H. 2019. New approaches in biological wastewater treatment aimed at removal of organic matter and nutrients. *Ecological Chemistry and Engineering*, 26:331–343.
- Sakkaravarthy, S., Jano, N.A. and Vijayakumar, A. 2024. Overcoming challenges in traditional wastewater treatment through ai-driven innovation. In: *The AI cleanse: transforming wastewater treatment through artificial intelligence: harnessing data-driven solutions*, Springer, 53–81.
- Shah, A.I., Dar, M.U.D., Bhat, R.A., Singh, J.P., Singh, K. and Bhat, S.A. 2020. Prospectives and challenges of wastewater treatment technologies to combat contaminants of emerging concerns. *Ecological Engineering*, 152:105882.
- Sharmila, V.G., Kumar, M.D. and Tamilarasan, K. 2024. Machine learning-driven advances in metal-organic framework nanomaterials for wastewater treatment: developments and challenges. *Separation & Purification Reviews*, 55:33–55.0
- Sheik, A.G., Tejaswini, E.S.S., Seepana, M.M. and Ambati, S.R. 2023. Control of anaerobic-anoxic-aerobic (A²/O) processes in wastewater treatment: a detailed review. *Environmental Technology Reviews*, 12:420–440.
- Habib, S.S., Torii, S., S-Moi, K. and Nair, A.C.A. 2024. Optimization of the factors affecting biogas production using the Taguchi design of experiment method. *Biomass*, 4:687–703.
- Singh, S., Kaur, R. and Agrawal, N.K. 2024a. Ethical considerations in AI-driven wastewater treatment. In: *The AI Cleanse: Transforming Wastewater Treatment Through Artificial Intelligence: Harnessing Data-Driven Solutions*, Springer, 313–335.
- Singh, S., Singh, R., Singh, K., Katoch, K., Zaeen, A.A., Birhan, D.A., Singh, A., Sandhu, H.S., Singh, H. and Sahrma, L.K. 2024b. Smart fertilizer technologies: An environmental impact assessment for sustainable agriculture. *Smart Agricultural Technology*, 8:100504.



- Soo, A., Gao, L. and Shon, H.K. 2024. Machine learning framework for wastewater circular economy—Towards smarter nutrient recoveries. *Desalination*, 592:118092.
- Sun, Y. and Jung, H. 2024. Machine learning (ML) modeling, IoT, and optimizing organizational operations through integrated strategies: The role of technology and human resource management. *Sustainability*, 16:6751.
- Thakur, A.K., Kumar, R., Kumar, A., Shankar, R., Khan, N.A., Gupta, K.N., Ram, M. and Arya, R.K. 2023. Pharmaceutical wastewater treatment via advanced oxidation based integrated processes: An engineering and economic perspective. *Journal of Water Process Engineering*, 54:103977.
- Tilala, M.H., Chenchala, P.K., Choppadandi, A., Kaur, J., Naguri, S., Saoji, R., Devaguptapu, B. and Tilala, M. 2024. Ethical considerations in the use of artificial intelligence and machine learning in health care: a comprehensive review. *Cureus*, 16: e62443.
- Torres, J.F., Galicia, A., Troncoso, A. and Martínez-Álvarez, F. 2018. A scalable approach based on deep learning for big data time series forecasting. *Integrated Computer Aided Engineering*, 25:335–348.
- Vora, L.K., Gholap, A.D., Jetha, K., Thakur, R.R.S., Solanki, H.K. and Chavda, V.P. 2023. Artificial intelligence in pharmaceutical technology and drug delivery design. *Pharmaceutics*, 15:1916.
- Wang, J., Li, Y., Gao, R.X. and Zhang, F. 2022. Hybrid physics-based and data-driven models for smart manufacturing: Modelling, simulation, and explainability. *Journal of Manufacturing Systems*, 63:381–391.
- Wang, Q., Ren, Y., Wang, H., Wang, J., Yang, Y., Zhang, Q. and Zhou, G. 2024. Wind-induced response of rapeseed seedling stage and lodging prediction based on UAV imagery and machine learning methods. *Computers and Electronics in Agriculture*, 217:108637.
- Wasalathanthri, D.P., Ding, J. and Li, Z.J. 2020. Real time process monitoring in biologics development. *American Pharmaceutical Review*, 23:72–75.
- Wongburi, P. and Park, J.K. 2023. Prediction of wastewater treatment plant effluent water quality using recurrent neural network (RNN) models. *Water (Basel)*, 15:3325.
- Wu, S., Wallace, S., Brix, H., Kuschik, P., Kirui, W.K., Masi, F. and Dong, R. 2015. Treatment of industrial effluents in constructed wetlands: challenges, operational strategies and overall performance. *Environmental Pollution*, 201:107–120.
- Yang, F. and Xiong, X. 2024. Carbon emissions, wastewater treatment and aquatic ecosystems. *Science of the Total Environment*, 921:171138.
- Zaharia, C., Leon, F., Curteanu, S. and Iacob-Tudose, E.T. 2021. Textile wastewater treatment in a spinning disc reactor: Improved performances—experimental, modeling and SVM optimization. *Processes*, 9:2003.
- Zhang, W., Tooker, N.B. and Mueller, A.V. 2020. Enabling wastewater treatment process automation: leveraging innovations in real-time sensing, data analysis, and online controls. *Environmental Science: Water Research & Technology*, 6:2973–2992.
- Zhao, Y.F., Xie, J. and Sun, L. 2024. On the data quality and imbalance in machine



learning-based design and manufacturing—A systematic review. *Engineering*, 45:105-131.

Zhao, Z., Gong, X., Yao, X., Deng, J. and Zhang, L. 2025. Effects of PAHs on nitrogen uptake by phytoplankton species: Implications for environmental risk assessment of micropollutants and algal bloom formation. *Journal of Environmental Sciences*, 154:820–832.

Zhong, X., Gallagher, B., Liu, S., Kailkhura, B., Hiszpanski, A. and Han, T.Y.-J. 2022. Explainable machine learning in materials science. *NPJ Computational Materials*, 8:204.

