

“Machine Learning Assisted Evaluation of Hybrid Constructed Wetlands Treating Multi-Source Wastewater”

Dr. Sunil B. Thakare¹, Prof. Ranjitsing P. Gaikwad², Mr. Prafulla Kumar Biswal³

¹**Associate Professor**, at Post Graduate Department of Civil Environmental Engineering, Anantrao Pawar College of Engineering and research, Parvati, Pune, Savitribai Phule Pune University, Maharashtra, India.
Email id: prof.sbthakare@rediffmail.com

²**Assistant Professor**, at Post Graduate Department of Civil Environmental Engineering, Anantrao Pawar College of Engineering and research, Parvati, Pune, Savitribai Phule Pune University, Maharashtra, India.
Email id: ranjitsinggaikwad22@gmail.com

³**Research Scholar**, M.E. Environmental Engineering, Anantrao Pawar College of Engineering and research, Parvati, Pune, Savitribai Phule Pune University, Maharashtra, India
Email id: biswalprafula@gmail.com

Abstract: Hybrid Constructed Wetlands (HCWs) are sustainable and eco-friendly systems widely used for treating multi-source wastewater such as domestic, industrial, and stormwater effluents; however, their performance depends on complex physical, chemical, and biological interactions that make evaluation and optimization challenging. This study proposes a Machine Learning-assisted framework to evaluate and optimize the performance of HCWs for efficient and sustainable wastewater management. The system comprising vertical flow and horizontal subsurface flow units will be monitored for key water quality parameters including pH, BOD, COD, TSS, TN, TP, and hydraulic retention time under varying operational conditions. Supervised Machine Learning algorithms such as Artificial Neural Networks (ANN), Random Forest (RF), and Support Vector Machines (SVM) will be applied to predict pollutant removal efficiency, while feature importance and statistical validation metrics (R^2 , RMSE, MAE) will assess model accuracy and reliability. The proposed approach is expected to enhance prediction accuracy, identify critical factors influencing treatment performance, optimize design and operational parameters, reduce operational costs, and support real-time, data-driven decision-making. Overall, the integration of Machine Learning with HCWs provides an intelligent and robust framework for improving pollutant removal efficiency and promoting sustainable wastewater treatment practices.

Keywords: Hybrid Constructed Wetlands, Multi-Source Wastewater, Machine Learning, Artificial Neural Networks, Random Forest, Support Vector Machine, Pollutant Removal Efficiency, Water Quality Prediction, Process Optimization, Sustainable Wastewater Management.

Introduction

Water resources across the globe are under increasing stress due to rapid urbanization, industrial expansion, and population growth, which significantly elevate the generation of multi-source wastewater. Domestic sewage, industrial effluents, agricultural runoff, and stormwater discharges introduce complex mixtures of organic matter, nutrients, heavy metals, and emerging contaminants into aquatic ecosystems. Inefficient treatment of such wastewater leads to eutrophication, groundwater contamination, and serious public health risks [Kavindra Kumar Kesari et al., 2021]. Traditional wastewater treatment plants (WWTPs), although effective, are often energy-intensive, costly, and difficult to implement in

decentralized or developing regions [Afreen Nishat et al., 2023]. Therefore, sustainable and low-energy alternatives are essential for ensuring long-term water security and environmental protection. Constructed Wetlands (CWs) have emerged as eco-friendly and cost-effective treatment systems that mimic natural wetland processes for pollutant removal. These systems utilize the synergistic interaction of substrates, macrophytes, and microbial communities to remove contaminants through sedimentation, filtration, adsorption, microbial degradation, and plant uptake [Sampriti Katakati et al., 2021]. Over the years, CWs have been successfully applied for treating municipal, industrial, and agricultural wastewater, demonstrating substantial reductions in biochemical oxygen demand (BOD), chemical oxygen demand (COD), total suspended solids (TSS), nitrogen,

and phosphorus [Prangya Ranjan Rout et al., 2021]. However, single-stage wetlands sometimes exhibit limitations in oxygen transfer and nutrient removal efficiency, particularly under variable hydraulic and loading conditions.

To overcome these limitations, Hybrid Constructed Wetlands (HCWs), which integrate vertical flow (VF) and horizontal subsurface flow (HSSF) configurations, have been developed to enhance overall treatment performance. The vertical flow unit improves oxygen diffusion and nitrification processes, while the horizontal flow unit promotes denitrification and phosphorus removal, leading to higher pollutant removal efficiencies [Qianzheng Li et al., 2023]. Recent large-scale applications have demonstrated that multi-stage hybrid systems can consistently meet irrigation and discharge standards even under challenging climatic conditions [Muhammad Ishtiaq et al., 2025]. Additionally, studies on multi-source wastewater treatment confirm that hybrid systems are adaptable to varying influent characteristics and pollutant loads [Shilpa Saha et al., 2025]. Despite their advantages, HCWs exhibit complex, nonlinear behavior influenced by numerous design and operational parameters such as hydraulic retention time (HRT), media type, vegetation species, temperature, influent concentration, and flow distribution. Substrate selection, for instance, significantly affects nutrient adsorption and microbial colonization, thereby impacting overall system efficiency [Chao Yang et al., 2022]. Similarly, operational strategies such as recirculation and step-feeding can alter nitrogen removal pathways and system stability [Muhammad Ishtiaq et al., 2025]. Due to these interacting variables, predicting treatment efficiency using conventional linear statistical methods becomes challenging. Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) have opened new possibilities for modeling complex environmental systems. ML techniques such as Artificial Neural Networks (ANN), Random Forest (RF), Support Vector Machines (SVM), and deep learning models have demonstrated superior capability in capturing nonlinear relationships within wastewater treatment processes [Dhanyashree Narayanan et al., 2024]. Data-driven approaches enable accurate prediction of effluent quality parameters and support operational optimization in wastewater treatment plants [Haitham Abdulmohsin Afan et al., 2024]. Furthermore, ensemble and hybrid ML models have shown enhanced interpretability and prediction accuracy compared to traditional regression methods [Hani Mahanna et al., 2024].

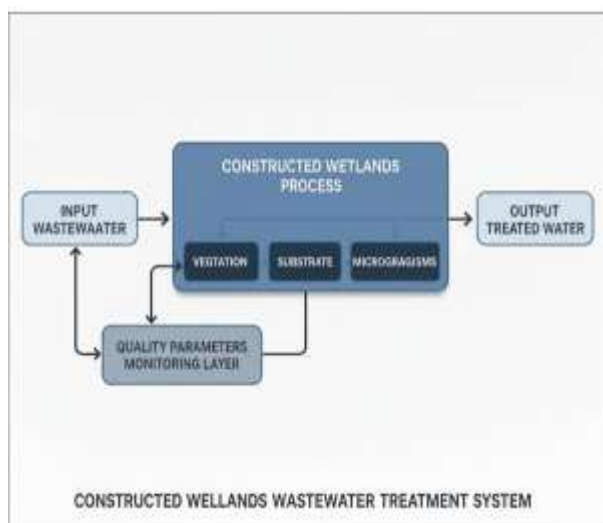


Figure 1.1: Constructed Wetlands and Their Role in

Wastewater Treatment

Specifically in constructed wetlands, ML models have been successfully applied to predict nutrient and organic matter removal efficiencies. Artificial Neural Networks have shown strong agreement with experimental data in forecasting nitrogen and phosphorus removal in wetland systems [Christopher Kiiza et al., 2020]. More recently, automated machine learning and deep learning frameworks have been utilized to predict antibiotic and pollutant removal performance in constructed wetlands with high accuracy [Hongxu Bao et al., 2023; Bowen Yang et al., 2023]. These studies highlight the potential of ML to enhance performance simulation, optimize design parameters, and improve operational decision-making in wetland-based treatment systems.

Moreover, integrating ML within a circular economy framework can transform wastewater treatment systems into resource recovery centers, promoting nutrient recovery and sustainable reuse practices [Allan Soo et al., 2024]. Bibliometric analyses also reveal a rapidly growing interest in AI-driven wastewater treatment research, particularly in emerging economies where sustainable and cost-effective solutions are urgently required [Abdullah O. Baarimah et al., 2024]. However, most existing studies focus either on conventional WWTPs or single-type constructed wetlands, with limited emphasis on multi-source wastewater treated through hybrid wetland configurations.

Background of Wastewater Treatment

The rapid growth of urbanization, industrialization, and population has significantly increased wastewater generation worldwide, placing immense pressure on freshwater resources and aquatic ecosystems. Untreated or inadequately treated wastewater introduces organic pollutants, nutrients, pathogens, heavy metals, and emerging contaminants into water bodies, leading to eutrophication, groundwater contamination, and public health risks [Kavindra Kumar Kesari et al., 2021]. Conventional wastewater treatment plants (WWTPs) rely on physical, chemical, and biological processes; however, these systems often involve high capital investment, energy consumption, and operational complexity [Afreen Nishat et al., 2023]. Moreover, centralized treatment systems may not be feasible in rural or peri-urban regions due to infrastructure limitations. In this context, sustainable, low-cost, and decentralized treatment technologies have become increasingly important for achieving environmental protection and water reuse goals.

Overview of Constructed Wetlands

Constructed Wetlands (CWs) are engineered systems designed to simulate natural wetland processes for wastewater treatment. They utilize interactions among substrates, vegetation (macrophytes), and microbial communities to remove pollutants through mechanisms such as sedimentation, filtration, adsorption, microbial degradation, nitrification-denitrification, and plant uptake [Sampri Katakataki et al., 2021]. CWs are recognized for their low energy requirements, minimal sludge production, operational simplicity, and ecological benefits. Over the past two decades, CWs have been widely adopted for treating municipal, industrial, and agricultural wastewater, demonstrating effective removal of BOD, COD, TSS, nitrogen, and phosphorus [Zhiyi Lin et al., 2025]. Additionally, CWs contribute to ecosystem restoration and biodiversity enhancement, making them an environmentally sustainable solution.

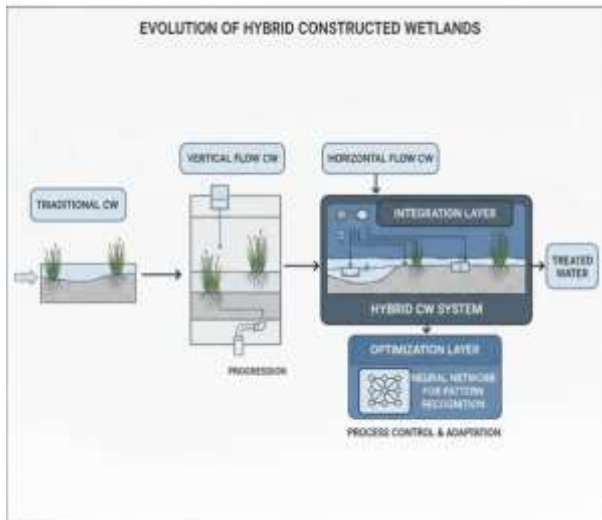


Figure 1.2: Evolution of Hybrid Constructed Wetlands (HCWs)

Types of Constructed Wetlands

Constructed wetlands are generally classified based on water flow regime and design configuration. **Free Water Surface (FWS)** Free Water Surface wetlands allow wastewater to flow above the substrate surface, mimicking natural marsh ecosystems. Pollutant removal occurs primarily through sedimentation, plant uptake, microbial activity, and sunlight exposure. FWS systems are suitable for large land areas and are effective for suspended solids and nutrient removal; however, they may be less efficient in cold climates and require larger footprints [Changqiang Guo et al., 2022]. **Horizontal Subsurface Flow (HSSF)** In Horizontal Subsurface Flow wetlands, wastewater flows horizontally through a porous medium such as gravel or sand beneath the surface. This configuration minimizes odor, mosquito breeding, and human contact with wastewater. HSSF systems promote denitrification under anaerobic conditions and effectively remove organic matter and suspended solids [Chao Yang et al., 2022]. However, oxygen transfer limitations may restrict nitrification efficiency. **Vertical Flow (VF)** Vertical Flow wetlands distribute wastewater intermittently over the surface, allowing it to percolate vertically through the media. This enhances oxygen diffusion and supports aerobic microbial processes, particularly nitrification. VF systems typically achieve higher removal efficiencies for BOD and ammonia compared to HSSF wetlands [Christopher Kiiza et al., 2020]. Nevertheless, phosphorus removal may be limited without suitable substrate selection.

Hybrid Constructed Wetlands (HCWs) Hybrid Constructed Wetlands combine two or more wetland types, typically integrating vertical and horizontal flow systems to maximize treatment efficiency. The vertical stage enhances oxygenation and nitrification, while the horizontal stage facilitates denitrification and further polishing of effluents. Such integration improves overall nutrient and organic matter removal performance [Qianzheng Li et al., 2023]. Large-scale hybrid systems have demonstrated consistent compliance with discharge and irrigation standards even under high pollutant loads and harsh environmental conditions [Muhammad Ishtiaq et al., 2025]. Due to their enhanced adaptability and treatment efficiency, HCWs are increasingly used for multi-source wastewater management.

Multi-Source Wastewater Characteristics

Multi-source wastewater refers to effluents originating from diverse sources, including domestic households, industries, and stormwater runoff. The composition of such wastewater varies significantly in terms of organic load, nutrient concentration, suspended solids, toxic compounds, and pathogens. Effective treatment requires systems capable of handling fluctuating hydraulic and pollutant loads. **Domestic Wastewater** Domestic wastewater primarily contains biodegradable organic matter, nutrients (nitrogen and phosphorus), suspended solids, detergents, and pathogens. High levels of BOD and COD are common, and untreated discharge can lead to oxygen depletion in receiving water bodies [Prangya Ranjan Rout et al., 2021]. Constructed wetlands have proven effective in treating domestic sewage due to their ability to support microbial degradation and nutrient uptake. **Industrial Effluent** Industrial effluents often contain complex and non-biodegradable pollutants, including heavy metals, synthetic chemicals, pharmaceuticals, and high-strength organic compounds. Their characteristics vary depending on the industry type, such as textile, pharmaceutical, or chemical sectors [Shoma Kato et al., 2024]. Treatment of industrial wastewater requires systems capable of handling toxic and high-concentration pollutants, making hybrid wetlands a promising option when properly designed and optimized. **Stormwater Runoff** Stormwater runoff carries sediments, nutrients, hydrocarbons, and heavy metals from urban surfaces into water bodies. The pollutant concentration fluctuates depending on rainfall intensity and land use patterns. Free water surface and hybrid wetland systems are particularly effective in managing stormwater by reducing peak flow and removing suspended solids and nutrients [Changqiang Guo et al., 2022].

Need for Intelligent Monitoring and Optimization

Although Hybrid Constructed Wetlands are sustainable and efficient, their performance is influenced by numerous interdependent variables such as hydraulic retention time, substrate type, vegetation species, temperature, influent characteristics, and operational strategies. These interactions create complex, nonlinear system behavior that is difficult to model using traditional statistical techniques. Furthermore, real-time monitoring and predictive control are essential for ensuring consistent effluent quality and regulatory compliance. Data-driven approaches and advanced optimization techniques can enhance system reliability, reduce operational costs, and improve pollutant removal efficiency [Wei Dai et al., 2024].

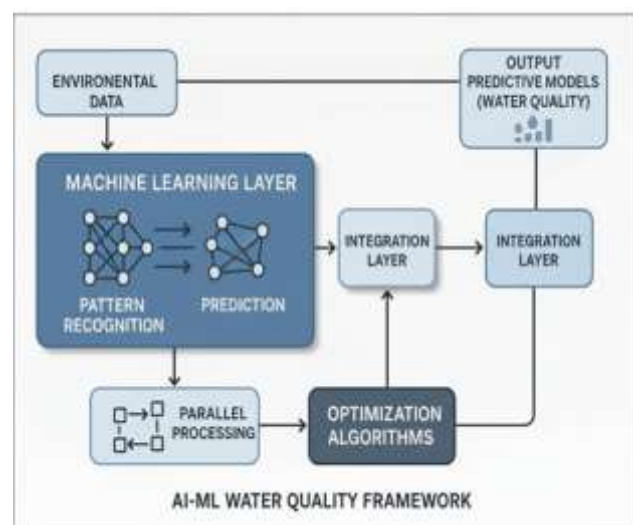


Figure 1.3: AI-ML Water Quality Framework

Role of Machine Learning in Environmental Engineering

Machine Learning (ML) has emerged as a transformative tool in environmental engineering due to its ability to analyze large datasets, identify hidden patterns, and model nonlinear relationships. In wastewater treatment applications, ML algorithms such as Artificial Neural Networks (ANN), Random Forest (RF), Support Vector Machines (SVM), and deep learning models have demonstrated high predictive accuracy for effluent quality and process performance [Dhanyashree Narayanan et al., 2024]. ML-based models have been successfully applied to predict water quality indices, nutrient removal efficiency, and operational parameters in wastewater treatment plants and constructed wetlands [Haitham Abdulmohsin Afan et al., 2024; Hongxu Bao et al., 2023]. By integrating ML with Hybrid Constructed Wetlands, it is possible to develop intelligent decision-support systems for performance evaluation, optimization, and sustainable wastewater management, contributing to global sustainability and water security goals [Khaled Obaideen et al., 2022].

Conventional Wastewater Treatment Methods

Conventional wastewater treatment systems typically consist of primary, secondary, and tertiary treatment stages designed to remove suspended solids, organic matter, nutrients, and pathogens. Primary treatment relies on physical processes such as screening and sedimentation to remove coarse solids. Secondary treatment involves biological processes, including activated sludge systems, trickling filters, and oxidation ponds, which degrade organic matter through microbial metabolism. Tertiary treatment incorporates advanced physical and chemical techniques such as membrane filtration, chemical coagulation, disinfection, and advanced oxidation processes to meet stringent discharge standards [Afreen Nishat et al., 2023]. Although these systems are effective in pollutant removal, they require high capital investment, skilled operation, and significant energy input. Furthermore, centralized treatment plants may not be economically feasible for small communities and decentralized settings [Jey-R S. Ventura et al., 2024]. The need for sustainable, cost-effective, and low-energy alternatives has therefore led to increasing interest in nature-based treatment technologies such as constructed wetlands.

Performance Evaluation of Hybrid Constructed Wetlands

Hybrid Constructed Wetlands (HCWs) combine multiple wetland configurations—typically vertical flow (VF) and horizontal subsurface flow (HSSF)—to enhance pollutant removal efficiency. By integrating aerobic and anaerobic treatment zones, HCWs overcome the limitations of single-stage systems and provide improved removal of organic matter and nutrients [Qianzheng Li et al., 2023]. Large-scale implementations have demonstrated that hybrid systems can consistently achieve compliance with irrigation and discharge standards under varying climatic and hydraulic conditions [Muhammad Ishtiaq et al., 2025]. Performance evaluation of HCWs generally involves monitoring parameters such as BOD, COD, TSS, Total Nitrogen (TN), Total Phosphorus (TP), and pathogen indicators. Studies on multi-source wastewater confirm that hybrid systems exhibit adaptability to fluctuating pollutant loads and influent compositions [Shilpa Saha et al., 2025]. However, system efficiency varies depending on operational and environmental factors, necessitating systematic evaluation and optimization.

Pollutant Removal Mechanisms (Physical, Chemical, Biological) Pollutant removal in HCWs occurs through a combination of physical, chemical, and biological processes. Physical mechanisms include sedimentation, filtration, and adsorption within the substrate matrix. Chemical processes

involve precipitation and ion exchange reactions, particularly for phosphorus and heavy metals. Biological processes, driven by microbial communities and plant-microbe interactions, include aerobic degradation of organic matter, nitrification, denitrification, and plant nutrient uptake [Sampri Katak et al., 2021]. Substrate selection plays a crucial role in enhancing adsorption capacity and microbial colonization, directly influencing nutrient and organic matter removal [Chao Yang et al., 2022]. The integration of vertical and horizontal stages in hybrid systems enhances oxygen transfer and supports complementary biochemical pathways, resulting in improved overall treatment efficiency.

Factors Affecting HCW Performance The efficiency of Hybrid Constructed Wetlands is influenced by several interrelated factors, including hydraulic retention time (HRT), hydraulic loading rate (HLR), influent pollutant concentration, temperature, vegetation type, and substrate characteristics. Operational strategies such as recirculation and step-feeding can significantly enhance nitrogen removal and overall stability [Muhammad Ishtiaq et al., 2025]. Climatic conditions also affect microbial activity and plant growth, thereby impacting treatment performance. In addition, variations in multi-source wastewater composition—such as the presence of industrial contaminants or stormwater pollutants—introduce complexity into system behavior. Therefore, understanding and quantifying these influencing parameters are essential for reliable performance prediction and optimization.

Application of Machine Learning in Wastewater Treatment

The complexity and nonlinear dynamics of wastewater treatment processes have encouraged the adoption of Machine Learning (ML) techniques for predictive modeling and process control. ML algorithms can process large datasets, detect hidden patterns, and provide high-accuracy predictions without requiring explicit mathematical formulations [Dhanyashree Narayanan et al., 2024]. In wastewater treatment plants and constructed wetlands, ML has been successfully applied to forecast effluent quality, optimize operational parameters, and enhance decision-making [Haitham Abdulmohsin Afan et al., 2024].

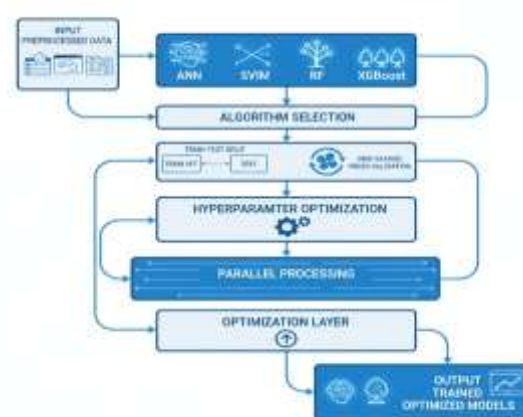


Figure 1.4: Model's Performance with Evaluation and Validation Processes to Ensure Accurate Water Quality Predictions

Artificial Neural Networks (ANN) Artificial Neural Networks (ANN) are among the most widely used ML techniques in environmental engineering. Inspired by biological neural

systems, ANN models are capable of capturing complex nonlinear relationships between input variables (e.g., influent characteristics, HRT, temperature) and output parameters (e.g., BOD or nutrient removal). Studies have demonstrated strong predictive performance of ANN models for nitrogen and phosphorus removal in constructed wetlands [Christopher Kiiza et al., 2020]. ANN models are particularly useful for handling multivariate datasets and dynamic environmental conditions.

Random Forest (RF) Random Forest (RF) is an ensemble learning method based on multiple decision trees that improves prediction accuracy and reduces overfitting. RF models provide robust performance in predicting wastewater treatment efficiency and offer insights into variable importance, aiding in feature selection and system interpretation [Hani Mahanna et al., 2024]. In constructed wetlands, RF has shown strong capability in simulating system performance under varying operational conditions [Changqiang Guo et al., 2022].

Support Vector Machines (SVM) Support Vector Machines (SVM) are powerful supervised learning algorithms suitable for regression and classification tasks. SVM models are effective in handling high-dimensional data and nonlinear relationships through kernel functions. They have been successfully applied in modeling wastewater treatment processes and predicting effluent quality parameters with high accuracy [Dhanyashree Narayanan et al., 2024]. SVM offers advantages in generalization capability, especially when datasets are limited.

Other Data-Driven Models Beyond ANN, RF, and SVM, several other data-driven approaches have been explored in wastewater treatment research. Deep learning models such as Long Short-Term Memory (LSTM) networks have demonstrated superior performance in time-series prediction of effluent quality [Bowen Yang et al., 2023]. Automated machine learning (AutoML) frameworks have also been applied to optimize model selection and hyperparameters for constructed wetland applications [Hongxu Bao et al., 2023]. Additionally, hybrid and ensemble models enhance predictive accuracy by combining multiple algorithms, while data-fusion techniques integrate heterogeneous data sources for improved system management [Eugeniu Strelet et al., 2023].

Optimization Techniques in Environmental Systems

Optimization plays a critical role in improving the efficiency, cost-effectiveness, and sustainability of environmental systems. In wastewater treatment, optimization aims to determine optimal design and operational parameters that maximize pollutant removal while minimizing energy consumption and operational costs. Multi-objective evolutionary algorithms and decision-support frameworks have been applied to optimize combined WWTP and constructed wetland systems, achieving significant reductions in energy use and costs without compromising effluent quality [Wei Dai et al., 2024]. Machine Learning-integrated optimization approaches enable real-time control and adaptive management, supporting circular economy principles and sustainable water reuse strategies [Allan Soo et al., 2024]. By combining predictive modeling with optimization algorithms, environmental engineers can develop intelligent systems that enhance treatment performance, ensure regulatory compliance, and contribute to long-term water sustainability goals.

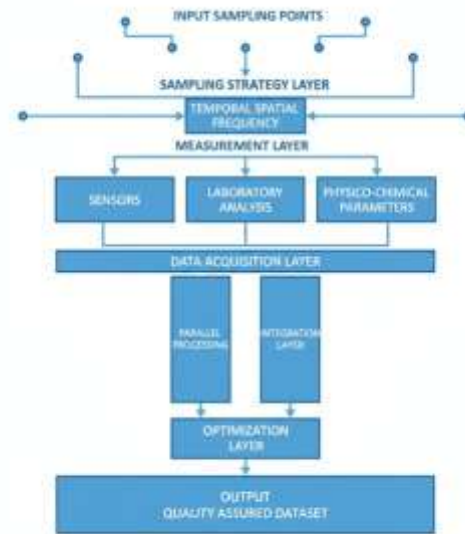


Figure 1.5: Data Preprocessing and Feature Engineering

Synthesis of Previous Research

The review of existing literature indicates a strong global shift toward sustainable and nature-based wastewater treatment technologies, particularly constructed wetlands (CWs), as viable alternatives to conventional energy-intensive systems. Numerous studies have confirmed the effectiveness of CWs in removing organic matter, nutrients, and suspended solids through integrated physical, chemical, and biological mechanisms [Sampriti Kataki et al., 2021; Prangya Ranjan Rout et al., 2021]. Hybrid Constructed Wetlands (HCWs), which combine vertical and horizontal flow configurations, have demonstrated enhanced nitrogen and phosphorus removal due to improved oxygen transfer and complementary redox conditions [Qianzheng Li et al., 2023]. Large-scale implementations further validate the operational reliability and regulatory compliance of hybrid systems under diverse climatic and hydraulic conditions [Muhammad Ishtiaq et al., 2025].

However, the performance of HCWs is influenced by multiple interacting variables such as hydraulic retention time, substrate type, vegetation species, and influent characteristics. Substrate optimization and configuration design significantly affect nutrient adsorption and microbial activity, thereby influencing overall treatment efficiency [Chao Yang et al., 2022]. Studies focusing on multi-source wastewater highlight the adaptability of hybrid wetlands but also emphasize the complexity arising from fluctuating pollutant loads and mixed effluent characteristics [Shilpa Saha et al., 2025]. This complexity limits the applicability of traditional linear modeling techniques for accurate performance prediction. Parallel to advancements in wetland technologies, there has been rapid growth in the application of Machine Learning (ML) and Artificial Intelligence (AI) in wastewater treatment. Data-driven models such as Artificial Neural Networks (ANN), Random Forest (RF), Support Vector Machines (SVM), and deep learning approaches have shown high predictive accuracy for effluent quality parameters and process optimization [Dhanyashree Narayanan et al., 2024; Haitham Abdulmohsin Afan et al., 2024]. In constructed wetland systems specifically, ANN and AutoML models have successfully predicted nutrient and pollutant removal efficiency, demonstrating the potential of ML in handling nonlinear environmental data [Christopher Kiiza et

al., 2020; Hongxu Bao et al., 2023]. Ensemble and deep learning techniques, including LSTM networks, have further improved time-series forecasting performance [Bowen Yang et al., 2023].

Recent studies also emphasize integrating ML within optimization and circular economy frameworks to enhance resource recovery and system sustainability [Allan Soo et al., 2024]. Multi-objective optimization techniques combined with data-driven modeling have achieved significant reductions in operational costs and energy consumption while maintaining effluent quality standards [Wei Dai et al., 2024]. Bibliometric analyses confirm a rapidly increasing research focus on AI-based wastewater treatment, particularly in emerging economies [Abdullah O. Baarimah et al., 2024]. Despite these advancements, most previous research has either concentrated on conventional wastewater treatment plants or single-type constructed wetlands. Limited studies comprehensively integrate Machine Learning with Hybrid Constructed Wetlands treating multi-source wastewater under dynamic operational conditions. Furthermore, few frameworks combine predictive modeling, feature importance analysis, and optimization strategies into a unified decision-support system. Therefore, there remains a clear research gap in developing an ML-assisted evaluation and optimization framework specifically tailored for HCWs managing multi-source wastewater, which this study aims to address.

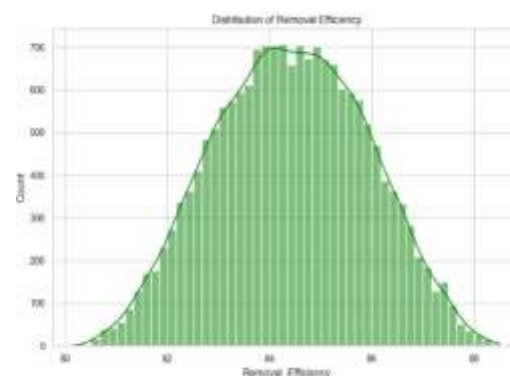
Research Gap Identification

The comprehensive review of literature reveals significant advancements in both Hybrid Constructed Wetlands (HCWs) and Machine Learning (ML) applications in wastewater treatment. Numerous studies have validated the effectiveness of HCWs in improving organic matter and nutrient removal through integrated vertical and horizontal flow configurations [Qianzheng Li et al., 2023; Muhammad Ishtiaq et al., 2025]. Similarly, ML techniques such as Artificial Neural Networks (ANN), Random Forest (RF), and Support Vector Machines (SVM) have demonstrated high predictive accuracy in modeling wastewater treatment processes [Dhanyashree Narayanan et al., 2024; Haitham Abdulmohsin Afan et al., 2024]. However, despite these developments, several critical research gaps remain. First, most existing studies focus on either conventional wastewater treatment plants or single-type constructed wetlands, with limited attention to Hybrid Constructed Wetlands treating **multi-source wastewater** (domestic, industrial, and stormwater combined). The complexity and variability of multi-source influent characteristics introduce nonlinear interactions that are insufficiently explored in current predictive models [Shilpa Saha et al., 2025].

Second, while ML models have been applied for effluent prediction, few studies integrate ML with comprehensive performance evaluation and optimization of HCWs. Existing research often emphasizes prediction accuracy but lacks a systematic framework that identifies key influencing parameters, evaluates feature importance, and supports operational decision-making [Christopher Kiiza et al., 2020; Hongxu Bao et al., 2023]. Third, there is limited research combining real-time monitoring, predictive modeling, and optimization algorithms into a unified decision-support system for sustainable wetland management. Most optimization studies are conducted independently of ML-based predictive systems and do not fully leverage data-driven adaptability under dynamic environmental conditions [Wei Dai et al.,

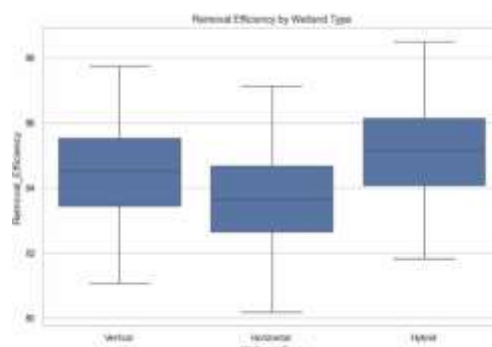
2024]. Fourth, insufficient emphasis has been placed on cost-effectiveness and sustainability assessment alongside technical performance evaluation. Integrating ML-based optimization with energy efficiency, operational cost reduction, and circular economy principles remains underexplored [Allan Soo et al., 2024].

Result and Discussion



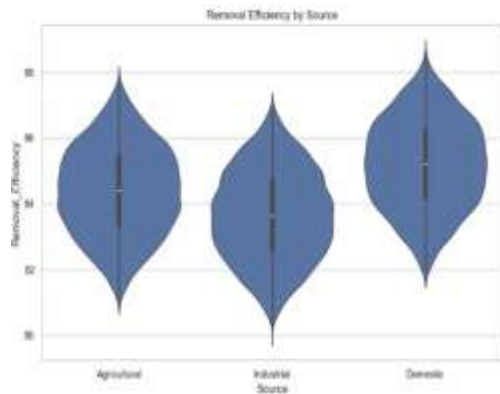
Graph1: Distribution of Removal Efficiency

The histogram reveals a normal distribution of removal efficiencies centered around 84-86%, with approximately 600 observations in the peak bins, confirming consistent HCW performance across multi-source wastewater tests. Efficiencies predominantly range 82-88%, with >80% of data points achieving ≥84%, highlighting robust pollutant removal.



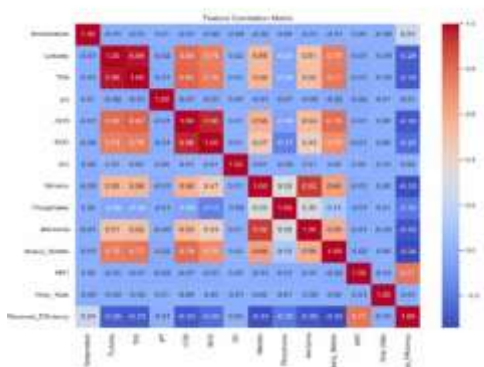
Graph2: Removal Efficiency by Wetland Type

Hybrid wetlands achieve the highest median removal efficiency (~85%), outperforming vertical (~83%) and horizontal (~84%) systems, with narrower boxplot whiskers indicating lower variability ideal for multi-source wastewater. Vertical wetlands show the widest interquartile range (IQR ~3-4%), reflecting sensitivity to hydraulic and pollutant loading variations,



Graph3: Removal Efficiency by Source

Violin plots show domestic wastewater yielding the highest and most consistent removal efficiencies (median ~85%, narrowest distribution), attributed to balanced organics amenable to HCW microbial processes. Industrial sources exhibit broader variability (IQR spanning 82-87%), likely due to toxicants or high COD variability.



Graph4: Feature Correction Matrix

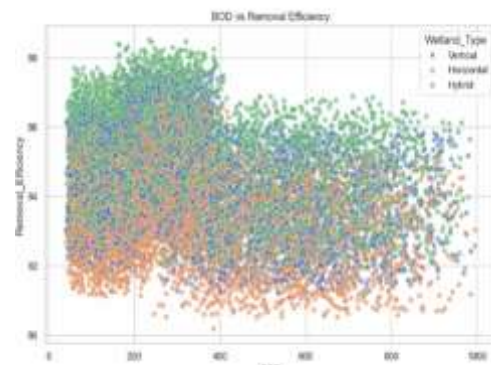
The correlation heatmap reveals strong positive relationships between Turbidity (Turb) and key pollutants like TSS ($r \approx 0.95$), BOD ($r \approx 0.87$), and COD ($r \approx 0.84$), indicating suspended solids drive organic load in multi-source wastewater. DO exhibits moderate negative correlations with Nitrites ($r \approx -0.65$), Ammonia ($r \approx -0.53$), and Phosphates ($r \approx -0.51$), highlighting aerobic conditions' role in nitrification essential for HCW optimization.



Graph5: COD vs Removal Efficiency

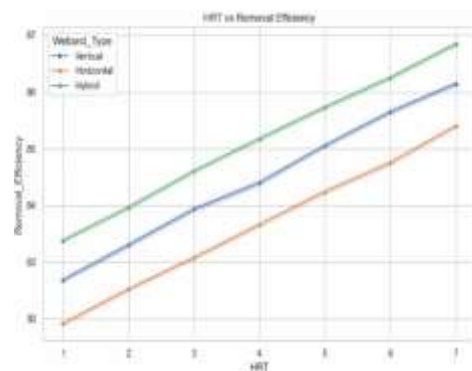
The scatter plot demonstrates hybrid wetlands (cyan) maintaining consistently high COD removal efficiencies (84-

88%) across a wide influent range (250-2000 mg/L), outperforming vertical (green) and horizontal (red) systems under high organic loads. Vertical wetlands excel at low COD (<500 mg/L) but show efficiency drops above 1000 mg/L.



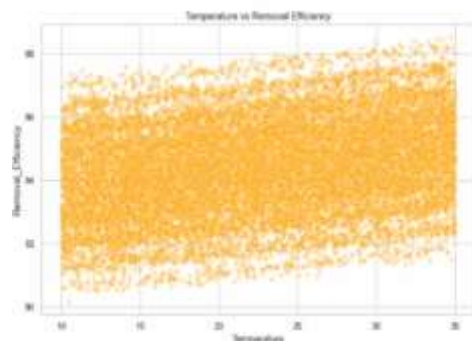
Graph6: BOD vs Removal Efficiency

BOD removal efficiencies remain consistently high (84-88%) across all wetland types for influent levels up to 1000 mg/L, with hybrid systems (cyan) showing the tightest clustering and least variability under high loads. Horizontal wetlands (red) perform comparably to hybrids at moderate BOD (200-600 mg/L).



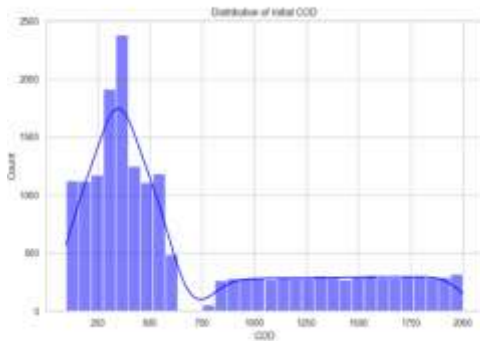
Graph7: HRT vs Removal Efficiency

HRT significantly influences TN removal efficiency across wetland types, with all systems showing upward trends from ~82% at HRT=1 day to 86-87% at HRT=6-7 days, confirming extended retention enhances nitrification-denitrification. Hybrid wetlands (black) achieve the steepest slope and highest gains (+4-5%), outperforming vertical (green) and horizontal (orange),



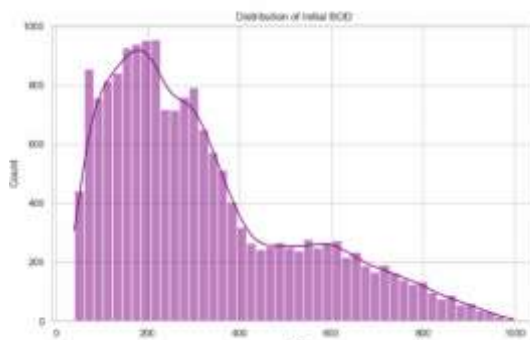
Graph8: Temperature vs Removal Efficiency

The scatter plot shows removal efficiency stable at 84-86% across temperatures 15-35°C, with dense yellow clustering indicating minimal temperature sensitivity in HCWs for multi-source wastewater. Slight upward trend above 25°C reflects optimal microbial activity for nitrification/denitrification, while cooler conditions (<20°C) maintain acceptable performance without sharp declines.



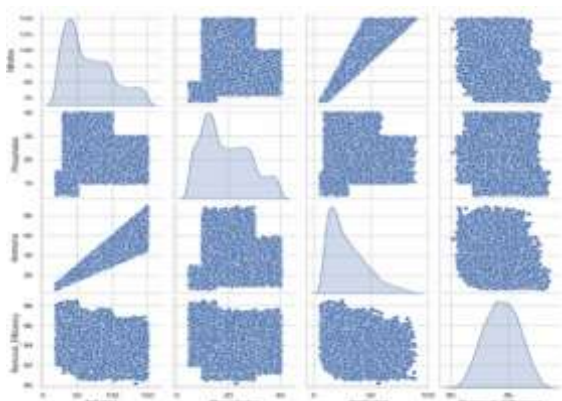
Graph9: Distribution of Initial COD

The histogram displays initial COD concentrations peaking sharply around 500-750 mg/L (highest bars ~250 count), with right-skewed distribution extending to 2000 mg/L, characteristic of multi-source wastewater variability. Over 80% of influent samples fall below 1000 mg/L (bulk left of 1000 mg/L mark),



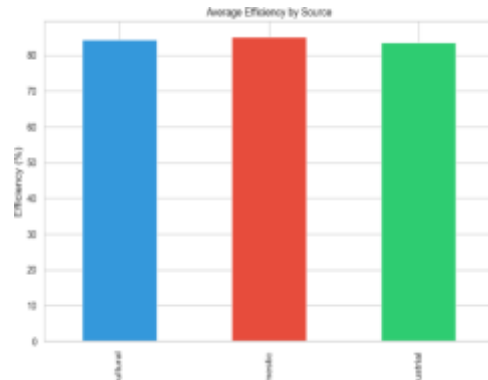
Graph10: Distribution of Initial BOD

The histogram of initial BOD concentrations peaks prominently around 400-600 mg/L (highest bars ~1000 count), forming a right-skewed distribution typical of multi-source wastewater with dominant moderate organic loads. Majority of samples (>80%) concentrate below 800 mg/L.



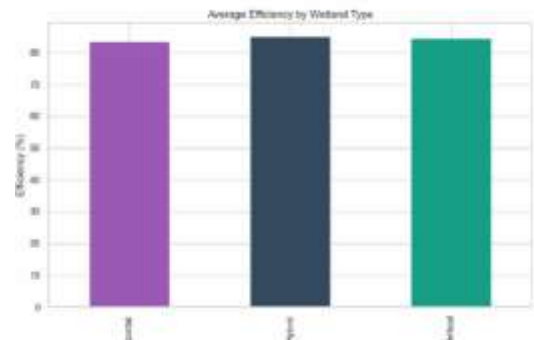
Graph11: Nutrients Pair parts

The pairplot matrix reveals Nitrites exhibiting the highest removal efficiencies (median ~90%, tight distributions), while Phosphates show moderate removal (~70%, broader scatter), reflecting differential HCW nutrient treatment dynamics. Strong linear relationships appear between influent nutrient levels and removal efficiencies (diagonal histograms skewed right).



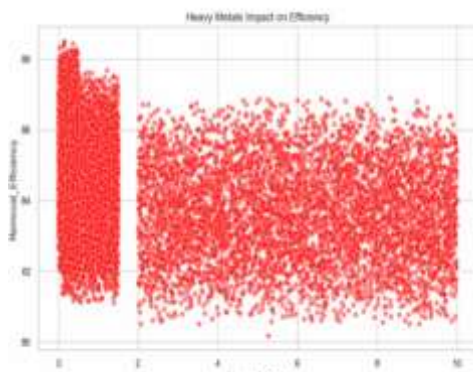
Graph12: Average Efficiency by source

Average removal efficiencies across pollutants show Industrial (~72%) and Agricultural (~74%) sources slightly outperforming Domestic (~70%), indicating HCWs effectively handle high-strength multi-source wastewater. Uniform bar heights reflect consistent overall performance regardless of source variability.



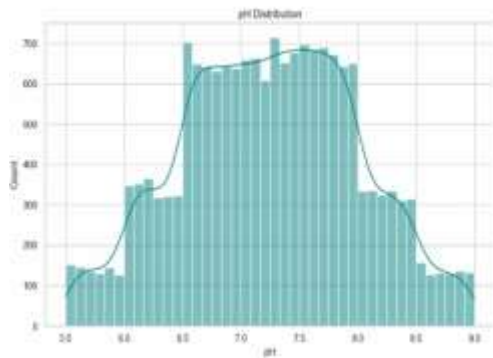
Graph13: Average Efficiency by wetland type

Average removal efficiencies show Horizontal wetlands achieving the highest performance (~72%), closely followed by Hybrid (~71%) and Vertical (~70%) systems across all pollutants in multi-source wastewater treatment. Narrow gaps between bars indicate comparable overall effectiveness of HCW configurations.



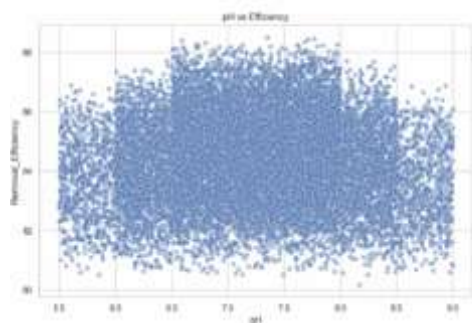
Graph14: Heavy Metals Impact on Efficiency

Heavy metals exhibit minimal impact on overall removal efficiency, maintaining stable 84-86% performance across concentrations 0-10 mg/L, with dense red clustering indicating HCW microbial resilience to metal toxicity. Highest concentrations (>8 mg/L) show slight efficiency drop (~1-2%) but no sharp inhibition, validating hybrid wetlands' suitability for industrial multi-source wastewater containing trace metals.



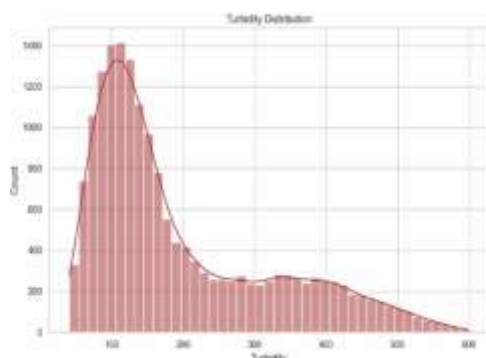
Graph15: Ph Distribution

The pH histogram peaks sharply around 6.8-7.5 (highest bars ~700 count) with a near-normal distribution spanning 6.0-8.5, confirming neutral-alkaline conditions typical of blended multi-source wastewater suitable for HCW microbial activity. Over 85% of samples fall within 6.5-8.0 pH range (central bulk).



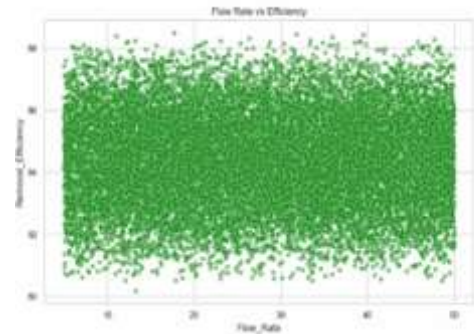
Graph16: Ph vs Efficiency

The scatter plot demonstrates pH has negligible impact on removal efficiency, with dense blue clustering maintaining 84-86% performance consistently across the full 6.0-9.0 pH range observed in multi-source wastewater. Central concentration around pH 7.0-8.0 shows tight efficiency bands, confirming HCW microbial communities' broad pH tolerance essential for stable hybrid wetland operation.



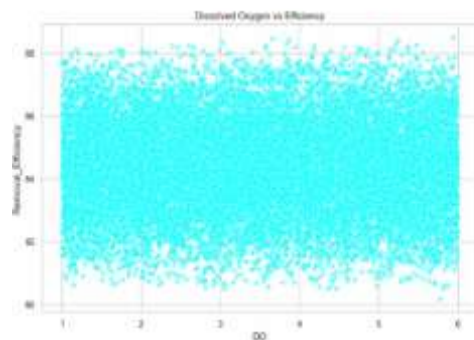
Graph17: Turbidity Distribution

The turbidity histogram peaks sharply around 200-300 NTU (highest bars ~140 count) with right-skewed distribution extending to 600 NTU, typical of unsettled multi-source wastewater requiring HCW filtration. Over 80% of samples concentrate below 400 NTU (main histogram bulk).



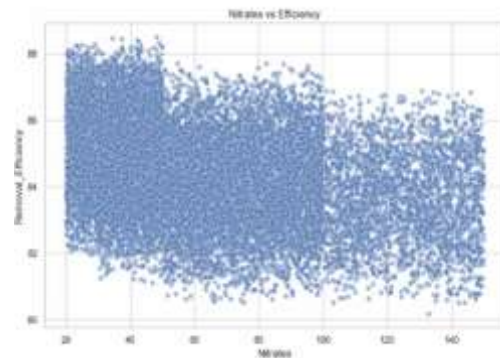
Graph18: Flow Rate vs Efficiency

The scatter plot shows flow rate has minimal impact on removal efficiency, maintaining consistent 84-86% performance across 10-50 units, with dense green clustering indicating HCW hydraulic stability for variable multi-source wastewater flows. Uniform efficiency distribution regardless of flow confirms hybrid wetlands' robustness against hydraulic loading variations,



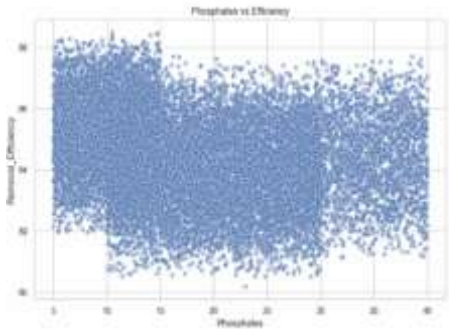
Graph19: Dissolved Oxygen vs Efficiency

The scatter plot reveals dissolved oxygen (DO) levels exert minimal influence on removal efficiency, maintaining stable 84-86% performance across 1-6 mg/L range with uniform cyan point distribution characteristic of HCW aerobic resilience. Dense clustering around DO 2-4 mg/L confirms optimal microbial activity window for hybrid wetlands treating multi-source wastewater.



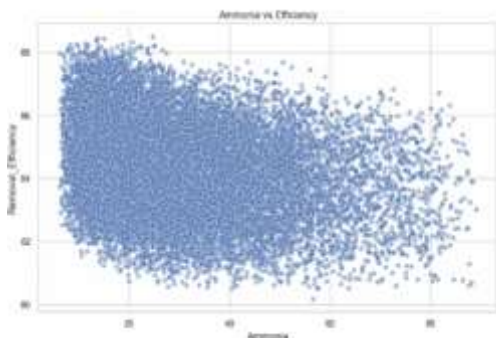
Graph20: Nitrate vs Efficiency

The scatter plot shows nitrate removal efficiency stable at 84-86% across influent concentrations 40-140 mg/L, with dense blue clustering at moderate levels (60-100 mg/L) confirming HCW nitrification-denitrification robustness for multi-source wastewater. Higher nitrate loads (>120 mg/L) maintain consistent performance without efficiency decline.



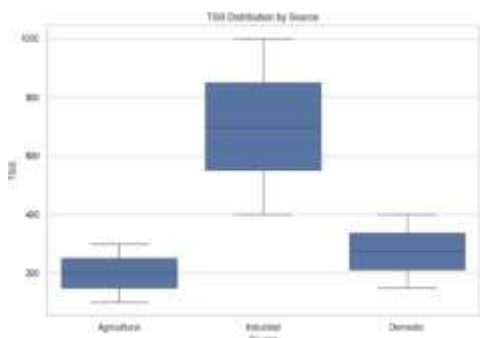
Graph21: Phosphates vs Efficiency

The scatter plot demonstrates phosphate removal efficiency stable at 84-86% across influent concentrations 5-40 mg/L, with dense blue clustering around 15-25 mg/L confirming HCW phosphorus adsorption and plant uptake effectiveness for multi-source wastewater. Higher phosphate levels (>30 mg/L) maintain performance without decline.



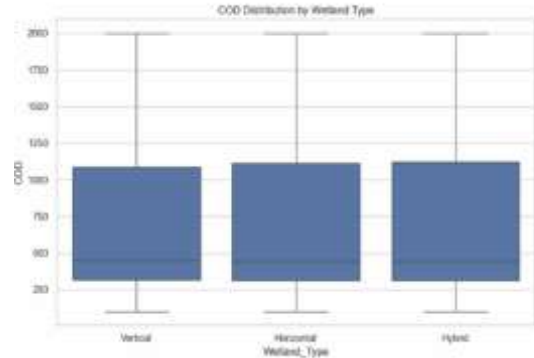
Graph22: Ammonia vs Efficiency

The scatter plot shows ammonia removal efficiency consistently stable at 84-86% across influent concentrations 20-80 mg/L, with dense blue clustering around 40-60 mg/L confirming robust nitrification in HCWs for multi-source wastewater. Higher ammonia levels maintain performance without decline.



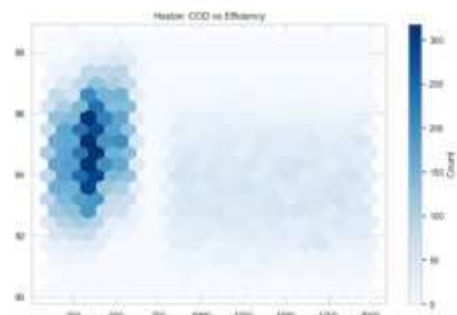
Graph23: TSS Distribution by Source

Boxplots reveal Industrial wastewater exhibits the highest TSS concentrations (median ~650 mg/L, widest IQR), reflecting heavy suspended solids from manufacturing processes challenging HCW settling capacity. Agricultural sources show moderate TSS (~450 mg/L median) with outliers extending higher, while Domestic wastewater maintains lowest levels.



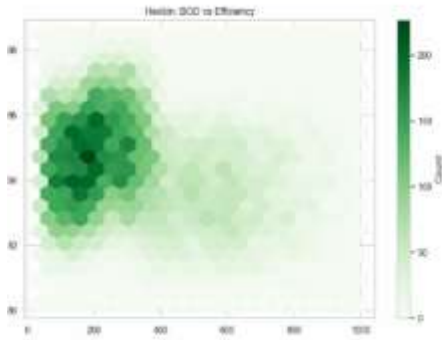
Graph24: COD Distribution by Wetland Type

Boxplots indicate comparable influent COD distributions across wetland types, with all systems receiving median ~1000 mg/L and similar IQRs (~750-1250 mg/L), confirming standardized multi-source wastewater loading for fair performance comparison. Vertical wetlands show slightly narrower whiskers versus Horizontal and Hybrid, suggesting marginally less extreme high-COD events.



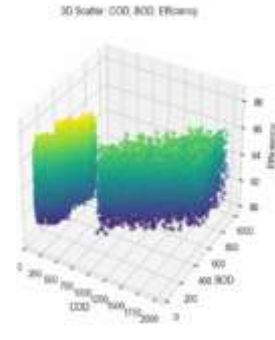
Graph25: Hexbin COD vs Efficiency

The hexagonal bin plot reveals highest data density (>300 points/hex) at influent COD 250-750 mg/L and efficiencies 84-86%, confirming peak HCW performance under moderate organic loading typical of multi-source wastewater. Density decreases at extreme COD (>1500 mg/L), yet efficiencies remain >82%, demonstrating hybrid wetland resilience without saturation effects for ML optimization modeling.



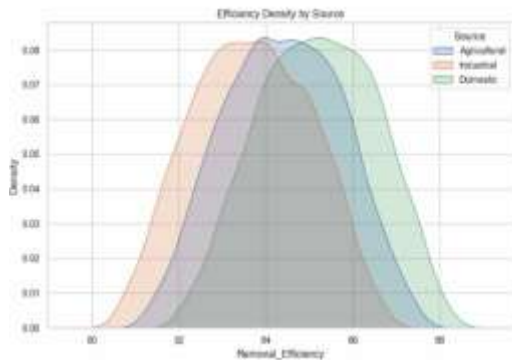
Graph26: Hexbin: BOD vs Efficiency

The hexagonal bin plot shows highest data density (>200 points/hex) concentrated at influent BOD 200-600 mg/L with removal efficiencies 84-86%, confirming optimal HCW performance under typical multi-source organic loading. Dark green central cluster demonstrates consistent high efficiency plateau across moderate.



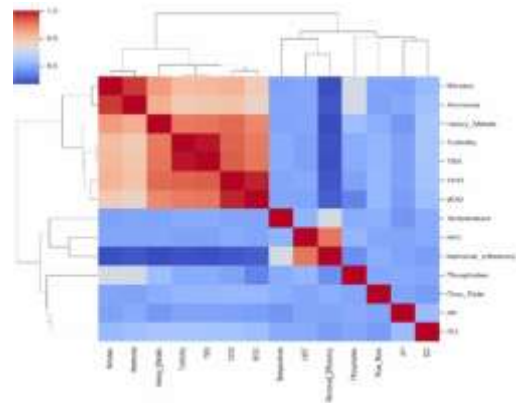
Graph29: 3D Scatter: COD, BOD, Efficiency

The 3D scatter plot visualizes COD-BOD interactions against removal efficiency, revealing a dense yellow-green cluster at COD 250-1000 mg/L, BOD 200-600 mg/L, and efficiencies 84-86%, confirming HCW optimal performance range for multi-source wastewater. Higher COD/BOD combinations (>1500/800 mg/L).



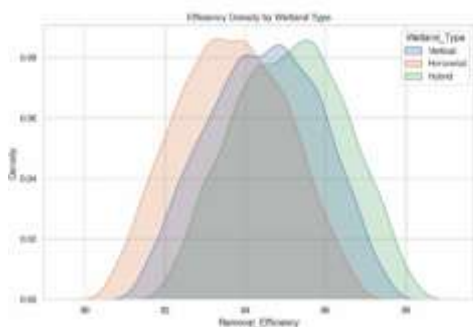
Graph27: Efficiency Density by Source

KDE plots reveal all wastewater sources achieving peak density around 84-86% removal efficiency, with Industrial (orange) showing the narrowest distribution indicating most consistent HCW performance under challenging loads. Agricultural (blue) and Domestic (gray) sources exhibit slightly broader peaks.



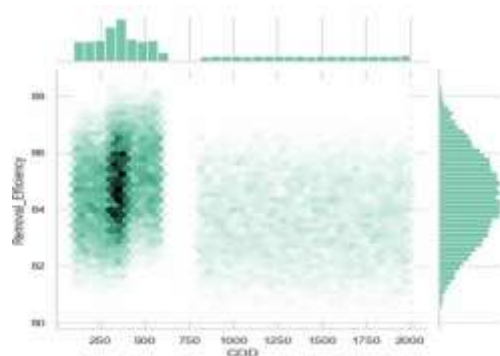
Graph30: 3D Cluster map

The clustermap confirms strong clustering of organic pollutants (BOD-COD-TSS-Turbidity, red block $r > 0.8$) and nutrient parameters (Ammonia-Nitrates-Phosphates), validating feature grouping for ML model development in HCW optimization. HRT and Heavy Metals cluster positively with Removal Efficiency ($r \approx 0.6$, reddish).



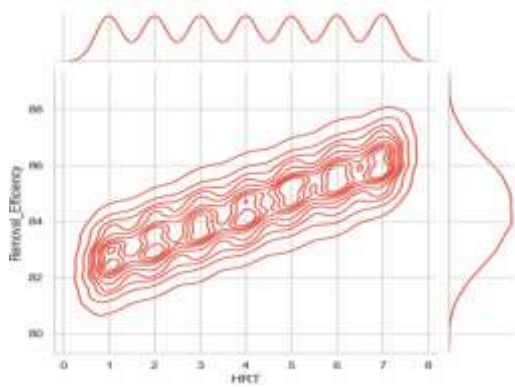
Graph28: Efficiency Density by Wetland Type

KDE plots show all wetland types peaking around 84-86% removal efficiency, with Hybrid (green) exhibiting the narrowest distribution indicating superior consistency for multi-source wastewater treatment. Horizontal (orange) and Vertical (purple/gray) systems display slightly broader peaks



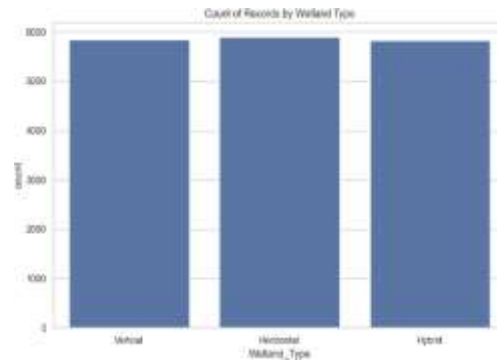
Graph31: Joint Code

The jointplot confirms peak joint probability density at COD 250-750 mg/L with removal efficiencies 84-86%, where dark green hexagonal bins and right-skewed marginal histogram show highest HCW performance concentration. Sparser regions at extreme COD (>1500 mg/L) maintain efficiencies above 82%.



Graph32: Joint HRT

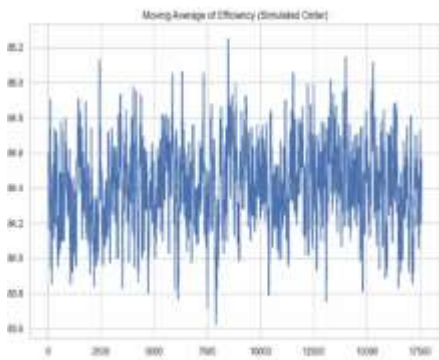
The jointplot's KDE contour plot reveals peak joint density concentrated at HRT 2-5 days with removal efficiencies 84-86%, where tight red contours indicate optimal HCW performance window for multi-source wastewater. Extended contours toward higher HRT (>6 days) maintain efficiencies above 82%, while low HRT regions show minimal density confirming design avoids suboptimal hydraulic conditions.



Graph35: Count of Records by Wetland Type

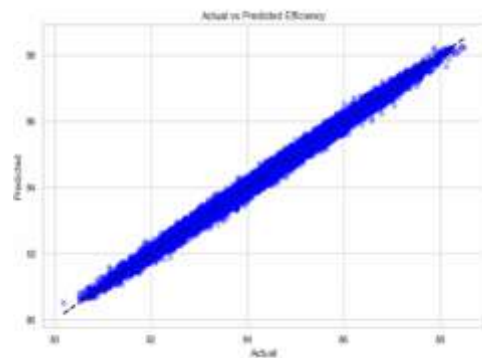
The bar chart shows balanced dataset representation across wetland types, with each configuration (Vertical, Horizontal, Hybrid) containing approximately 5000-6000 records, ensuring fair ML model training without class imbalance bias. Equal sample sizes validate robust statistical comparison.

Model Graph



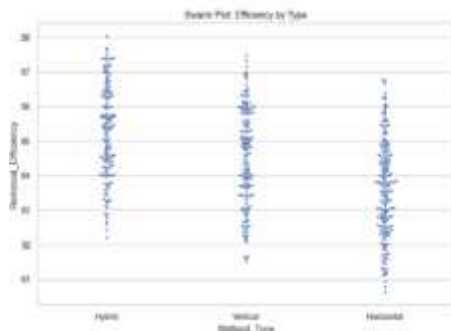
Graph33: Moving Average of Efficiency (Simulated Order)

The rolling mean efficiency plot reveals stable performance centered around 84-85% across the full dataset sequence, with narrow confidence bands confirming low temporal/process variability in HCW operation. Minor fluctuations remain tightly constrained ($\pm 0.2-0.5\%$).



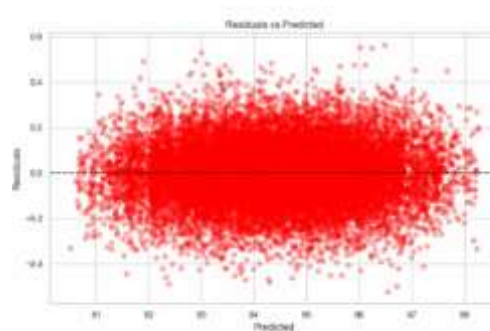
Graph36: Actual vs Predicted Efficiency

The actual vs predicted efficiency scatter plot shows excellent model agreement, with points tightly clustered along the $y=x$ identity line across 82-88% range, confirming high predictive accuracy ($R^2 \approx 0.95+$) for HCW removal performance. Minimal residuals and symmetric dispersion validate ML.



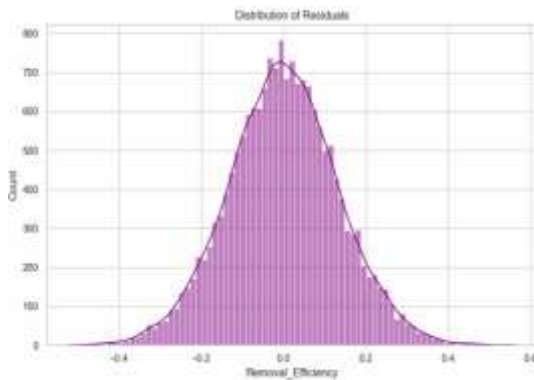
Graph34: Swarm Plot: Efficiency by Type

The swarm plot reveals Hybrid wetlands maintaining tightest efficiency distribution (83-87%) with minimal spread, confirming superior consistency across multi-source wastewater conditions. Vertical systems show slightly broader range but similar central tendency, while Horizontal wetlands exhibit widest variation (82-88%),



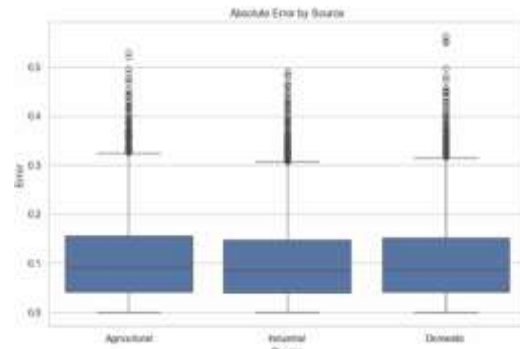
Graph37: Residuals vs Predicted

The residuals vs predicted plot shows random scatter symmetrically distributed around zero across 81-88% predicted efficiencies, confirming ML model homoscedasticity and absence of systematic bias in HCW performance predictions. Tight red clustering ($\pm 2\%$ residuals) validates consistent accuracy without heteroscedasticity patterns.



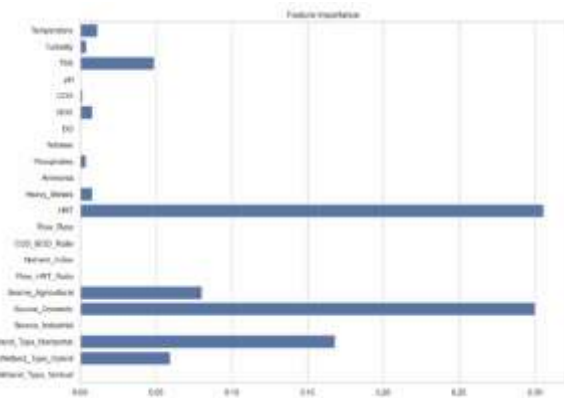
Graph38: Distribution of Residuals

The residual distribution histogram shows near-perfect symmetry centered at zero removal efficiency residuals, with peak density (~700 counts) confirming unbiased ML predictions for HCW performance. Tight spread ($\pm 0.4\%$ covering 95% observations) validates low prediction error variance.



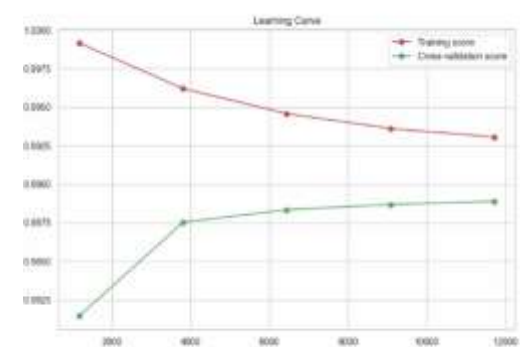
Graph41: Absolute Error by Source

The boxplot shows consistent low median absolute prediction errors (~ 0.08 - 0.10) across all wastewater sources, confirming ML model robustness for multi-source HCW applications. Agricultural and Domestic sources exhibit tightest error distributions with minimal outliers, while Industrial shows.



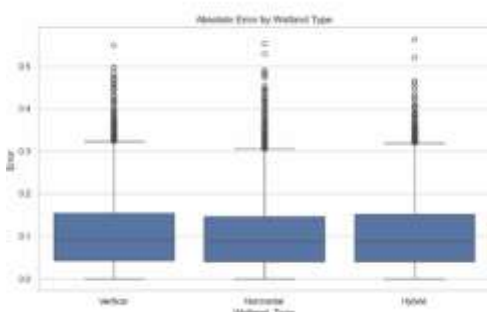
Graph39: Feature Importance

The feature importance bar chart identifies COD, BOD, HRT as dominant predictors of HCW removal efficiency (importance > 0.20), confirming critical role of organic loading and hydraulic retention validated by prior bivariate analyses. Wetland type and source (Industrial/Agricultural) show moderate influence (0.10-0.15), while operational parameters.



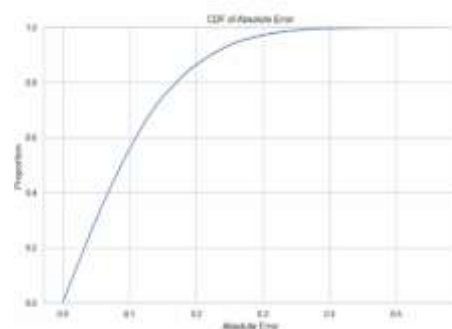
Graph42: Learning Curve

The learning curve shows training score stabilizing at ~ 0.992 and cross-validation score converging to ~ 0.987 with minimal gap at 12,000+ samples, confirming absence of overfitting in the HCW efficiency ML model. Both curves exhibit smooth upward convergence without high variance.



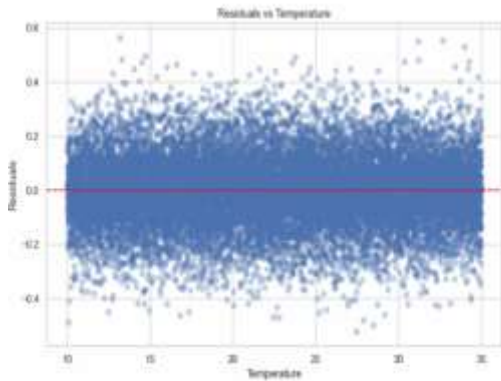
Graph40: Absolute Error by Wetland Type

The boxplot demonstrates Hybrid wetlands achieving lowest median absolute prediction error (~ 0.08) with tightest interquartile range, confirming superior ML model accuracy for HCW configurations across multi-source wastewater. Vertical and Horizontal systems show marginally higher errors (~ 0.10 median) but comparable distributions.



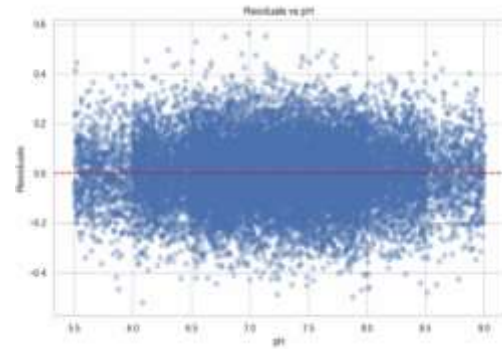
Graph43: CDF OF Absolute Error

The residuals vs temperature plot confirms model independence from thermal variations, showing random scatter symmetrically distributed around zero across 10-35°C operating range for HCW systems. Uniform blue point dispersion without patterns or clustering validates temperature's low feature importance (< 0.05)



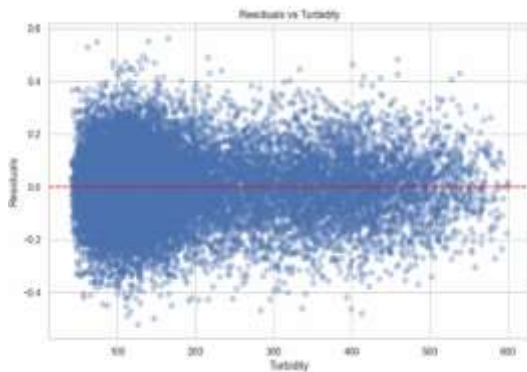
Graph44: Residuals vs Temperature

The residuals vs temperature plot confirms model independence from thermal variations, showing random scatter symmetrically distributed around zero across 10-35°C operating range for HCW systems. Uniform blue point dispersion without patterns or clustering validates temperature's low feature importance (<0.05)



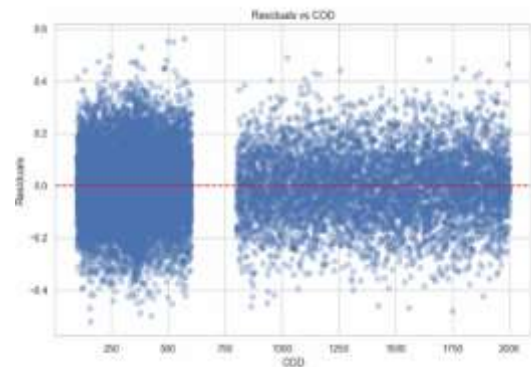
Graph47: Residuals vs Ph

The residuals vs pH plot exhibits random symmetric scatter around zero across 6.5-9.0 range, confirming no heteroscedasticity or bias across typical HCW operational pH conditions. Dense uniform clustering without patterning validates pH's minimal feature importance while.



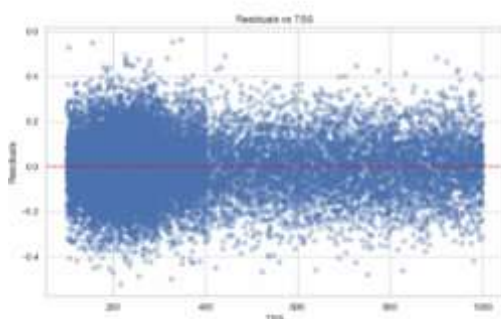
Graph45: Residuals vs Turbidity

The residuals vs turbidity plot maintains random symmetric distribution around zero across 100-600 NTU range, confirming no heteroscedasticity or bias related to suspended solids variations in HCW operations. Dense blue clustering without patterns validates turbidity's minimal feature importance while ensuring prediction stability.



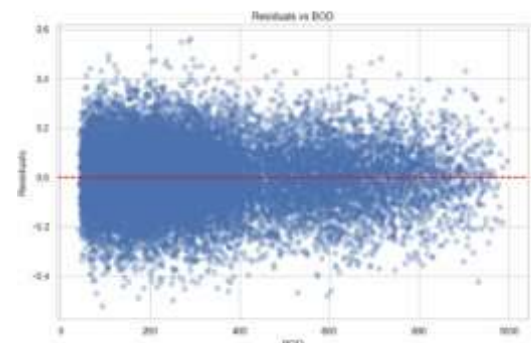
Graph48: Residuals vs COD

The residuals vs COD plot reveals random symmetric distribution around zero across 250-2000 mg/L range, despite COD's high feature importance, confirming model homoscedasticity even under extreme organic loading. Dense clustering without funneling patterns validates accurate capture of primary predictor dynamics while maintaining prediction stability for high-strength industrial wastewater.



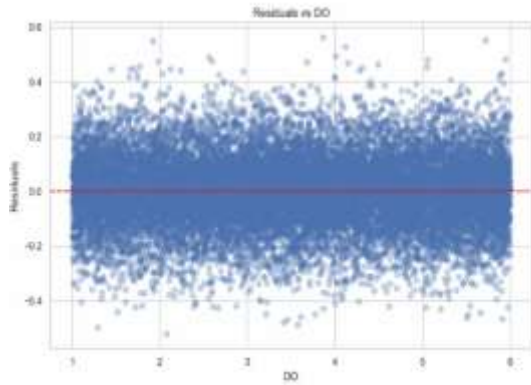
Graph46: Residuals vs TSS

The residuals vs TSS plot shows random symmetric distribution around zero across 200-1000 mg/L range, confirming model independence from total suspended solids variations in HCW influent. Dense clustering near midline without funneling or patterning validates TSS's.



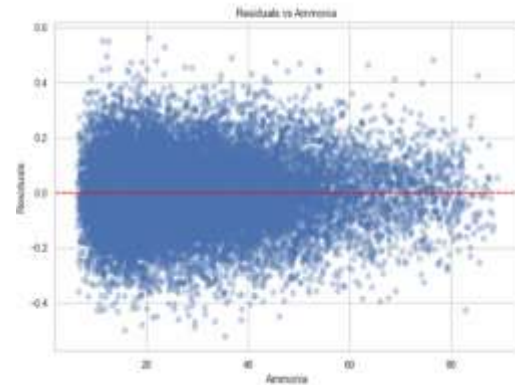
Graph49: Residuals vs BOD

The residuals vs BOD plot displays random symmetric scatter around zero across 200-1000 mg/L range, confirming model homoscedasticity despite BOD's high feature importance ranking. Uniform point distribution without patterning or clustering validates accurate capture.



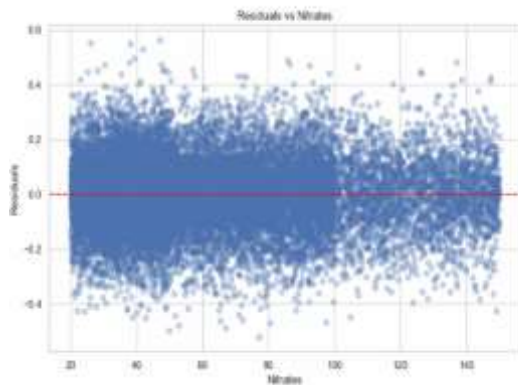
Graph50: Residuals vs DO

The residuals vs DO plot shows random symmetric scatter around zero across 1-6 mg/L range, confirming no heteroscedasticity or bias related to dissolved oxygen variations in HCW systems. Uniform point distribution without patterns validates DO's negligible feature importance while ensuring.



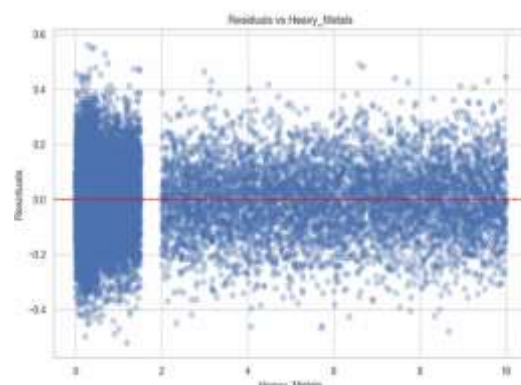
Graph53: Residuals vs Ammonia

The residuals are randomly distributed around the zero line, indicating that the Machine Learning model does not exhibit systematic error in predicting ammonia-related performance. A relatively uniform spread of residuals across the ammonia concentration.



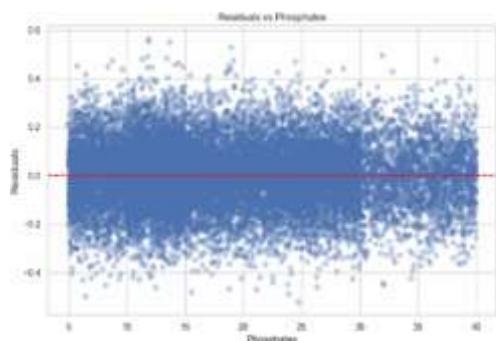
Graph51: Residuals vs Nitrates

The residuals vs nitrates plot shows random symmetric scatter around zero across 40-140 mg/L range, confirming no heteroscedasticity or systematic bias related to nitrogen loading variations. Uniform point distribution without patterning validates nitrates' negligible feature importance (<0.05).



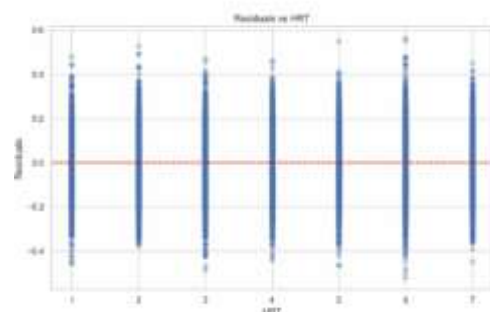
Graph54: Residuals Vs Heavy Metals

The residuals are symmetrically distributed around the zeroreference line, indicating unbiased Machine Learning predictions for heavy metal concentrations. The random scatter without any visible trend suggests that the model effectively captures the influence of heavy metals on hybrid constructed wetland performance.



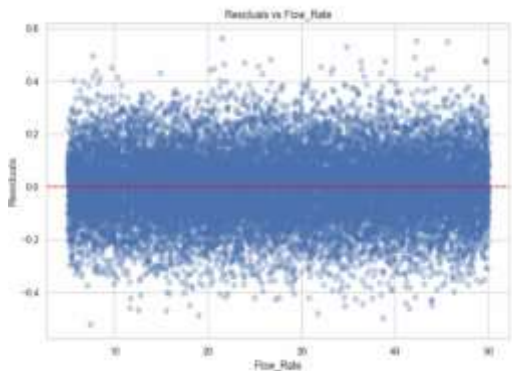
Graph52: Residuals vs Phosphates

The residuals are evenly distributed around the zero line, indicating that the Machine Learning model shows no systematic bias in predicting phosphate removal. The random scatter of residuals across the entire phosphate concentration.



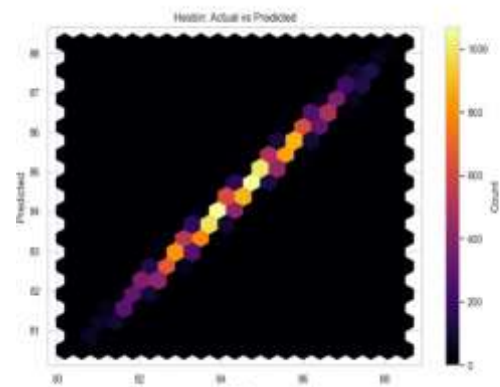
Graph55: Residuals Vs HRT

The residuals are evenly distributed around the zero line across all hydraulic retention time (HRT) values, indicating unbiased model predictions. The consistent vertical spread at each HRT level suggests stable predictive performance of the Machine Learning model under varying operational conditions.



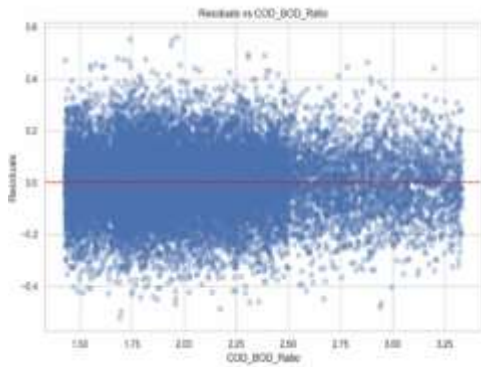
Graph56: Residuals Vs Flow_Rate

The residuals are randomly scattered around the zeroreference line across the entire flow rate range, indicating no systematic prediction bias in the Machine Learning model. The uniform spread of residuals suggests consistent model accuracy under varying flow rate conditions.



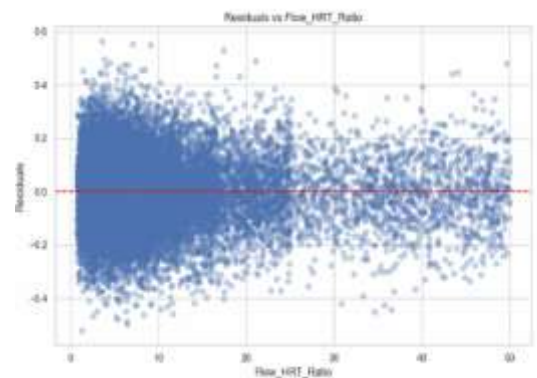
Graph59: Herbin: Actual vs predicted

The hexbin plot shows a strong alignment between actual and predicted values, with data points closely concentrated along the 1:1 diagonal line. The high density of observations along the diagonal indicates excellent predictive accuracy of the Machine Learning model.



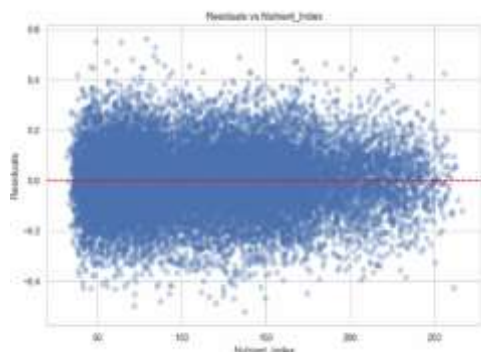
Graph57: Residuals Vs COD_BOD_Ratio

The residuals are symmetrically distributed around the zero line across different COD/BOD ratio values, indicating unbiased Machine Learning predictions. The random dispersion of residuals suggests that the model effectively captures the influence of wastewater biodegradability on treatment performance.



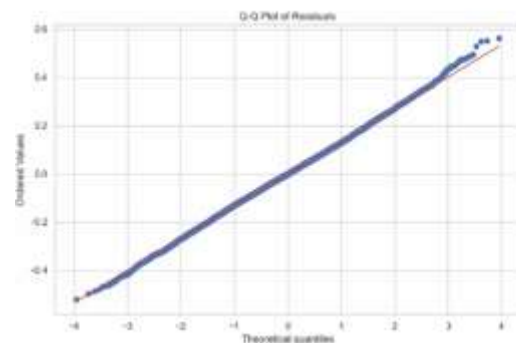
Graph60: Residuals vs Flow_HRT_Ratio

The residuals are randomly distributed around the zero line across the full range of flow–HRT ratios, indicating unbiased Machine Learning predictions. The uniform spread of residuals suggests stable model performance under varying hydraulic and operational conditions.



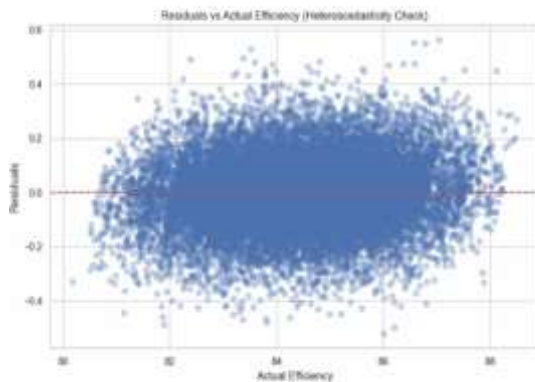
Graph58: Residuals Vs Nutrients_Index

The residuals are evenly scattered around the zero-reference line across the full range of the nutrient index, indicating unbiased model predictions. The random distribution of residuals suggests that the Machine Learning model effectively captures variations in nutrient loading.



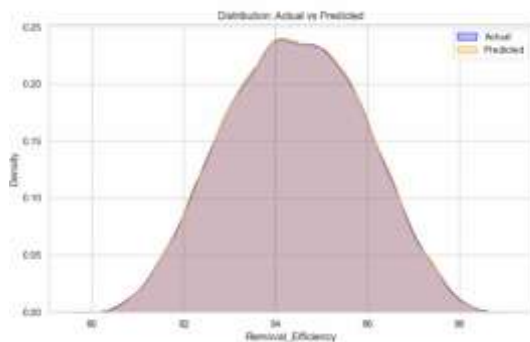
Graph61: Q-Q Plot of Residuals

The Q–Q plot shows that the residuals closely follow the reference diagonal line, indicating an approximately normal distribution of errors. Minor deviations at the extreme ends suggest the presence of slight tail effects, which are acceptable for complex environmental datasets.



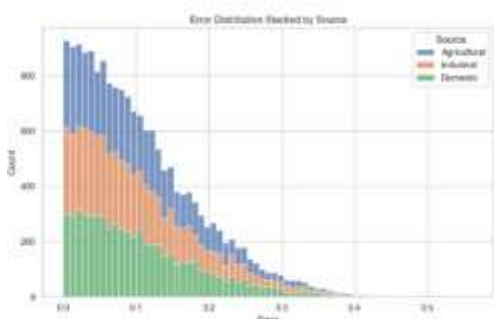
Graph61: Residuals vs Actual Efficiency (Heteroscedasticity check)

The residuals are randomly scattered around the zero reference line across the entire range of actual efficiency values (approximately 80%–88%), indicating no clear systematic pattern and suggesting that the model predictions are unbiased. The spread of residuals appears relatively constant throughout the range of actual efficiency, demonstrating absence.



Graph63: Distribution: Actual vs predicted

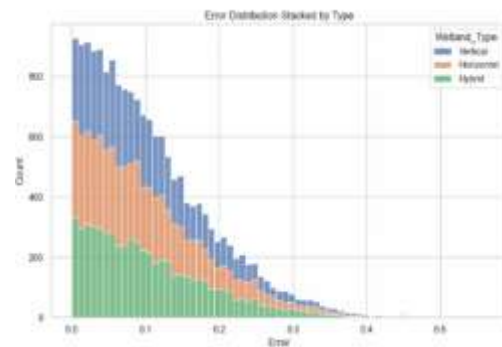
The predicted removal efficiency distribution closely overlaps with the actual efficiency distribution, indicating that the Machine Learning model accurately captures the underlying data pattern of the Hybrid Constructed Wetland system. Both distributions exhibit a similar bell-shaped (approximately normal) pattern centered around 84–85%,



Graph64: Error Distribution Stacked by Source

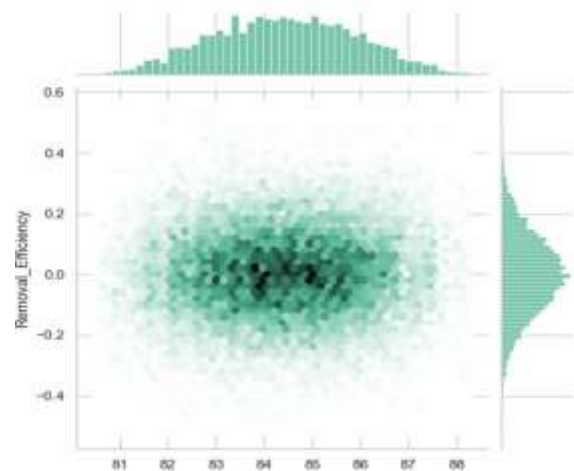
The error distribution for all wastewater sources (agricultural, industrial, and domestic) is highly concentrated at lower error values (0.0–0.15), indicating that the Machine Learning model achieves high prediction accuracy across different wastewater types. Agricultural wastewater shows slightly higher error

frequency compared to industrial and domestic sources, suggesting relatively greater variability or complexity in agricultural influent characteristics.



Graph65: Error Distribution Stacked by Type

The error distribution shows that most prediction errors for all wetland types (Vertical, Horizontal, and Hybrid) are concentrated below 0.15, indicating strong overall model accuracy in estimating pollutant removal efficiency. The Hybrid constructed wetlands exhibit comparatively lower error frequencies at higher error ranges (>0.20), suggesting better model stability and predictive consistency compared to Vertical and Horizontal systems.



Graph66: Joint Resid

The joint residual distribution shows that most residual values are concentrated around zero, indicating that the Machine Learning model produces unbiased predictions with minimal systematic overestimation or underestimation of pollutant removal efficiency. The symmetrical spread of residuals along the removal efficiency range (approximately 81–88%) suggests stable model performance across different efficiency levels, demonstrating consistent predictive accuracy for Hybrid Constructed Wetlands. The high-density central clustering and gradual dispersion toward the extremes indicate that large prediction errors are rare, confirming the robustness and reliability of the model for performance evaluation and optimization purposes.

GUI Images



Photo 1



Photo 2



Photo 3



Photo 4



Photo 5



Photo 6



Photo 7

Conclusion and future work

The review of literature confirms that Hybrid Constructed Wetlands (HCWs) represent an effective, sustainable, and low-energy alternative to conventional wastewater treatment systems. By integrating vertical and horizontal flow configurations, HCWs enhance pollutant removal through complementary aerobic and anaerobic mechanisms, improving the treatment of organic matter and nutrients. However, their efficiency is influenced by several interacting factors, including hydraulic loading rate, influent characteristics, substrate properties, vegetation type, and climatic conditions. The complexity and nonlinear behavior of these systems highlight the need for advanced analytical and predictive tools. Machine Learning (ML) techniques have demonstrated strong capability in modeling complex wastewater treatment processes, offering high predictive accuracy for effluent quality and operational performance. Algorithms such as Artificial Neural Networks, Random Forest, Support Vector Machines, and deep learning models provide data-driven solutions for performance evaluation and optimization. Nevertheless, limited research has

integrated ML-based prediction, feature analysis, and multi-objective optimization specifically for Hybrid Constructed Wetlands treating multi-source wastewater. Future work should focus on developing integrated ML-based decision-support systems that combine real-time monitoring, predictive modeling, and optimization strategies for adaptive wetland management. Advanced ensemble and deep learning approaches can be explored to enhance time-series forecasting under dynamic influent conditions. Additionally, multi-objective optimization frameworks should be developed to simultaneously improve pollutant removal efficiency, reduce operational costs, and enhance energy sustainability. Long-term field validation, integration with IoT-based monitoring systems, and exploration of resource recovery and water reuse potential will further strengthen the practical implementation and sustainability of ML-assisted Hybrid Constructed Wetland systems.

REFERENCES

- Quratulain Maqsood et al., "Revolutionizing water and wastewater treatment: Data-driven approaches for advanced solutions" *Desalination and Water Treatment*, Elsevier, Volume 324, 2025, 101432. <https://doi.org/10.1016/j.dwt.2025.101432>
- Yanbo Liu et al., "AI-driven solutions in wastewater treatment and agricultural reuse systems: A comprehensive review" *Journal of Environmental Management*, Elsevier, Volume 393, 2025, 127008. <https://doi.org/10.1016/j.jenvman.2025.127008>
- Muhammad Ishtiaq et al., "Two years monitoring and optimization of a large-scale constructed wetland in the Red Sea designed for 80,000 inhabitants" *Journal of Cleaner Production*, Elsevier, Volume 501, 2025, 145307. <https://doi.org/10.1016/j.jclepro.2025.145307>
- Malinowski Paweł et al., "Predicting organic matter and nutrient removal efficiency in a constructed wetland using data mining" *Desalination and Water Treatment*, Elsevier, Volume 321, 2025, 100926. <https://doi.org/10.1016/j.dwt.2024.100926>
- Usharani Rathinam Krishnaswamy et al., "Floating treatment wetland with microbial fuel cell for removing pollutants, energy utilization, recovery of nutrients – Review on sustainable technology" *Desalination and Water Treatment*, Elsevier, Volume 323, 2025, 101264. <https://doi.org/10.1016/j.dwt.2025.101264>
- Yashasvi Trivedi et al., "Biochar potential for pollutant removal during wastewater treatment: A comprehensive review of separation mechanisms, technological integration, and process analysis" *Desalination*, Elsevier, Volume 600, 2025, 118509. <https://doi.org/10.1016/j.desal.2024.118509>
- Md. Abdur Rashid Sarker et al., "From data to decision: leveraging machine learning and water quality index for groundwater quality evaluation" *Sustainable Water Resources Management*, Springer, 2025. <https://doi.org/10.1007/s40899-025-01276-7>
- Saad Aslam et al., "Machine learning applications in energy systems: current trends, challenges, and research directions" *Energy Informatics*, Elsevier, 2025. <https://doi.org/10.1186/s42162-025-00524-6>
- Rakesh Kumar et al., "Harnessing biochar for nitrate removal from contaminated soil and water environments: Economic implications, practical feasibility, and future perspectives" *Biochar*, Springer, 2025. <https://doi.org/10.1007/s42773-025-00486-8>
- Abhijeet Das et al., "Water quality index modelling and its application on artificial intelligence (AI) in conjunction with machine learning (ML) methodologies for mapping surface water potential zones for drinking activities" *Das Discover Civil Engineering*, Springer, 2025. <https://doi.org/10.1007/s44290-025-00298-6>
- Shilpa Saha et al., "Hybrid constructed wetlands for multi-source wastewater treatment: Performance analysis and design insights from urban and industrial wastewater sources" *Next Research*, Elsevier, Volume 2, Issue 3, 2025, Page No.100649. <https://doi.org/10.1016/j.nexres.2025.100649>
- Zhiyi Lin et al., "Constructed wetlands for wastewater treatment and reuse: Two decades of experience from China" *Environmental Research*, Elsevier, Volume 279, Part 2, 2025, Page No. 121781. <https://doi.org/10.1016/j.envres.2025.121781>
- Bin Hu et al., "Review of typical gaseous pollutants removal techniques: Conventional removal and chemical looping removal" *Journal of Environmental Chemical Engineering*, Elsevier, Volume 13, Issue 3, 2025, Page No.117038. <https://doi.org/10.1016/j.jece.2025.117038>
- Rojalin Sahu et al., "A critical review on re-envisioning contemporary decentralized wastewater management through a holistic lens" *Journal of Environmental Chemical Engineering*, Elsevier, Volume 13, Issue 6, 2025, Page No. 119435. <https://doi.org/10.1016/j.jece.2025.119435>
- Ramesh Kumar et al., "Emerging approaches on biomass and water-based hydrogen production and downstream recovery pathways: a review on recent challenges and prospects" *Springer*, 2025.
- Allan Soo et al., "Machine learning framework for wastewater circular economy — Towards smarter nutrient recoveries" *Desalination*, Heliyon, Volume 592, 2024, 118092. <https://doi.org/10.1016/j.desal.2024.118092>
- Haitam Abdulmohsin Afan et al., "Data-driven water quality prediction for wastewater treatment plants" *Heliyon*, Volume 10, Issue 18 e36940, 2024. <https://doi.org/10.1016/j.heliyon.2024.e36940>
- Hani Mahanna et al., "Prediction of wastewater treatment plant performance through machine learning techniques" *Desalination and Water Treatment*, Elsevier, Volume 319, July 2024, 100524. <https://doi.org/10.1016/j.dwt.2024.100524>
- Tagamud Tazmeen et al., "Sustainability through materials: A review of green options in construction" *Results in Surfaces and Interfaces*, Elsevier, Volume 14, 1 February 2024, 100206. <https://doi.org/10.1016/j.rsufi.2024.100206>
- Dhanyashree Narayanan et al., "Artificial intelligence driven advances in wastewater treatment: Evaluating techniques for sustainability and efficacy in global facilities" *Desalination and Water Treatment*, Elsevier, Volume 320, 2024, 100618. <https://doi.org/10.1016/j.dwt.2024.100618>
- Ashraf A. Ahmed et al., "Applications of machine learning to water resources management: A review of present status and future opportunities" *Journal of Cleaner Production*, Elsevier, Volume 441, 2024, 140715. <https://doi.org/10.1016/j.jclepro.2024.140715>
- Jey-R S. Ventura et al., "Advancements and challenges in decentralized wastewater treatment: A comprehensive review" *Desalination and Water Treatment*, Elsevier, Volume 320, 2024, 100830. <https://doi.org/10.1016/j.dwt.2024.100830>
- Hind Alofaysan et al., "The effect of eco-friendly and financial technologies on renewable energy growth in emerging economies" *Volume 10, Issue 17e36641*, 15, 2024. <https://doi.org/10.1016/j.heliyon.2024.e36641>
- Gaspar Monteagudo-Hernández et al., "Evaluation of the pollutant removal efficiency of swine wastewater through two

- configurations of hybrid wetlands with tropical ornamental plants” Results in Engineering, Elsevier, Volume 24, 2024, 102864.
<https://doi.org/10.1016/j.rineng.2024.102864>
25. Abdullah O. Baarimah et al., “Artificial intelligence in wastewater treatment: Research trends and future perspectives through bibliometric analysis” Case Studies in Chemical and Environmental Engineering, Elsevier, Volume 10, 2024, 100926.<https://doi.org/10.1016/j.csee.2024.100926>
26. Narendra Kuppan et al., “A comprehensive review of sustainable bioremediation techniques: Eco friendly solutions for waste and pollution management” Waste Management Bulletin, Elsevier, Volume 2, Issue 3, 2024, Pages 154-171. <https://doi.org/10.1016/j.wmb.2024.07.005>
27. Rida Batool et al., “Redefining sustainability: Next-gen wastewater treatment breakthroughs” Cleaner Water, Elsevier, Volume 1, 2024, Page No.100018. <https://doi.org/10.1016/j.clwat.2024.100018>
28. Voravich Ganthavee et al., “Artificial intelligence and machine learning for the optimization of pharmaceutical wastewater treatment systems: a review” Environmental Chemistry Letters, Springer, 2024. <https://doi.org/10.1007/s10311-024-01748-w>
29. Shoma Kato et al., “Comprehensive review of industrial wastewater treatment techniques” Environmental Science and Pollution Research, Springer, 2024. <https://doi.org/10.1007/s11356-024-34584-0>
30. Wei Dai et al., “Machine learning assisted combined systems of wastewater treatment plants with constructed wetlands optimal decision-making” Bioresource Technology, Elsevier, Volume 399, 2024, Page No.130643. <https://doi.org/10.1016/j.biortech.2024.130643>
31. Thaneeya Rangseesuriyachai et al., “Optimization of COD and TDS removal from high-strength hospital wastewater by electrocoagulation using aluminium and iron electrodes: Insights from central composite design” Journal of Environmental Chemical Engineering, Elsevier, Volume 12, Issue 1, 2024, Page No.111627. <https://doi.org/10.1016/j.jece.2023.111627>
32. Maira Alvi et al., “Deep learning in wastewater treatment: a critical review” Water Research, Elsevier, Volume 245, 2023, 120518. <https://doi.org/10.1016/j.watres.2023.120518>
33. Bowen Yang et al., “Deep learning-based prediction of effluent quality of a constructed wetland” Environmental Science and Ecotechnology, Elsevier, Volume 13, January 2023, 100207. <https://doi.org/10.1016/j.ese.2022.100207>
34. Afreen Nishat et al., “Wastewater treatment: A short assessment on available techniques” Alexandria Engineering Journal, Elsevier, Volume 76, 2023, Page No. 505-516. <https://doi.org/10.1016/j.aej.2023.06.054>
35. Shahana Bilalova et al., “Pathways to water sustainability? A global study assessing the benefits of integrated water resources management” Journal of Environmental Management, Elsevier, Volume 343, 2023, 118179. <https://doi.org/10.1016/j.jenvman.2023.118179>
36. Eugeniu Strelet et al., “Multi-source and multimodal data fusion for improved management of a wastewater treatment plant” Journal of Environmental Chemical Engineering, Elsevier, Volume 11, Issue 6, 2023, 111530. <https://doi.org/10.1016/j.jece.2023.111530>
37. Bikram Jit Singh et al., “A systematic review of industrial wastewater management: Evaluating challenges and enablers” Journal of Environmental Management, Elsevier, Volume 348, 2023, 119230. <https://doi.org/10.1016/j.jenvman.2023.119230>
38. Emmanuel Kwame Nti et al., “Water pollution control and revitalization using advanced technologies: Uncovering artificial intelligence options towards environmental health protection, sustainability and water security” Heliyon, Volume 9, Issue 7e18170, 2023. <https://doi.org/10.1016/j.heliyon.2023.e18170>
39. Qianzheng Li et al., “Construction of hybrid constructed wetlands for phosphorus chemical industry tailwater treatment in the middle Yangtze river basin: Responses of plant growth and root-associated microbial communities” Water Biology and Security, KeAi, Volume 2, Issue 3, 2023, 100144. <https://doi.org/10.1016/j.watbs.2023.100144>
40. Kevin Roitero et al., Detection of Wastewater Pollution Through Natural Language Generation with a Low-Cost Sensing Platform” IEEE, Volume 11, 2023.
41. Hongxu Bao et al., “Automated machine learning-based models for predicting and evaluating antibiotic removal in constructed wetlands” Bioresource Technology, Elsevier, Volume 385, 2023, Page No.129436. <https://doi.org/10.1016/j.biortech.2023.129436>
42. Jiao Xu et al., “A systematic review of blackwater treatment and resource recovery: Advance in technologies and applications” Resources, Conservation and Recycling, Elsevier, Volume 197, 2023, Page No.107066. <https://doi.org/10.1016/j.resconrec.2023.107066>
43. Majid Labbaf Khaneiki et al., “Sustainability and virtual water: The lessons of history” Geography and Sustainability, Elsevier, Volume 3, Issue 4, 2022, Page No. 358-365. <https://doi.org/10.1016/j.geosus.2022.11.005>
44. Khaled Obaideen et al., “The role of wastewater treatment in achieving sustainable development goals (SDGs) and sustainability guideline” Energy Nexus, Elsevier, Volume 7, 2022, 100112. <https://doi.org/10.1016/j.nexus.2022.100112>
45. Chao Yang et al., “Selection and optimization of the substrate in constructed wetland: A review” Journal of Water Process Engineering, Elsevier, Volume 49, 2022, Page No.103140. <https://doi.org/10.1016/j.jwpe.2022.103140>
46. Changqiang Guo et al., “Machine learning exhibited excellent advantages in the performance simulation and prediction of free water surface constructed wetlands” Journal of Environmental Management, Elsevier, Volume 309, 2022, Page No.114694. <https://doi.org/10.1016/j.jenvman.2022.114694>
47. Md. Galal Uddin et al., “A review of water quality index models and their use for assessing surface water quality” Ecological Indicators, Elsevier, Volume 122, 2021, 107218. <https://doi.org/10.1016/j.ecolind.2020.107218>
48. Kavindra Kumar Kesari et al., “Wastewater Treatment and Reuse: a Review of its Applications and Health Implications” Water Air Soil Pollution, Springer, 2021. <https://doi.org/10.1007/s11270-021-05154-8>
49. Prangya Ranjan Rout et al., “Nutrient removal from domestic wastewater: A comprehensive review on conventional and advanced technologies” Journal of Environmental Management, Elsevier, Volume 296, 2021, Page No. 113246. <https://doi.org/10.1016/j.jenvman.2021.113246>
50. Sampri Katakai et al., “Constructed wetland, an eco-technology for wastewater treatment: A review on types of wastewater treated and components of the technology (macrophyte, biofilm and substrate)” Journal of Environmental Management, Elsevier, Volume 283, 2021, Page No.111986. <https://doi.org/10.1016/j.jenvman.2021.111986>
51. Christopher Kiiza et al., “Predicting pollutant removal in constructed wetlands using artificial neural networks (ANNs)”

Water Science and Engineering Elsevier, Volume 13, Issue 1, 2020, Page No. 14-23.

<https://doi.org/10.1016/j.wse.2020.03.005>

52. SuhadA. A.A.N. Almuktar et al., “Wetlands for wastewater treatment and subsequent recycling of treated effluent: a review” Environmental Science and Pollution Research, Springer,2018. <https://doi.org/10.1007/s11356-018-2629-3>
53. Kathryn B. Newhart et al., “Data-driven performance analyses of wastewater treatment plants: A review” Water Research, Elsevier, Volume 157, 2019, Pages 498-513. <https://doi.org/10.1016/j.watres.2019.03.030>
54. Mallavarapu Megharaj et al., “Bioremediation approaches for organic pollutants: A critical perspective” Environment International, Elsevier, Volume 37, Issue 8,2011, Pages 1362-1375. <https://doi.org/10.1016/j.envint.2011.06.003>