Library Progress International *Print version ISSN 0970 1052*
Vol.44 No. 3, Jul-Dec 2024: P. 17485-17492 *Online version ISSN 2320 317X*

Original Article *Available online at www.bpasjournals.com*

Optimization of Wastewater Treatment Processes Using AI-Driven Machine Learning Algorithms for Enhanced Biological Degradation Efficiency

[1]Dr.Vajja Varalakshmi, [2]M. Saravanan , [3]Dr. Talakola Lakshmi Ramadasu, [4]Poodari Saritha, [5]Srinivasa Reddy Vempada

[1] Professor, Department of Civil Engineering, Marri Laxman Reddy Institute of Technology and Management, Dundigal, Hyderabad,, varasays@gmail.com
[2]institute of engineering and science ,,saravananm@mlritm.ac.in
CSMRS, New Delhi, 84990 97927, abhi.bnc@gmail.com
[3]Associate Professor, School of Civil Engineering, PNG University of Technology, LAE-411, morobe Province, Papua New Gunia. talakola.ramadasu@pnguot.ac.pg
[4]Professor, Civil Engineering Department, Malla Reddy Engineering College, poodarisaritha@gmail.com
[5]Professor, Department of Civil Engineering, KG Reddy College of Engineering and Technology, Hyderabad, Telangana.,srinivasareddy.v@kgr.ac.in

How to cite this article: Vajja Varalakshmi,M. Saravanan ,Talakola Lakshmi Ramadasu, Poodari Saritha, Srinivasa Reddy Vempada (2024) Optimization of Wastewater Treatment Processes Using AI-Driven Machine Learning Algorithms for Enhanced Biological Degradation Efficiency. *Library Progress International,* 44(3) 17485-17492.

Abstract
Wastewater treatment is essential for environmental protection, yet traditional biological methods often struggle with efficiency, particularly under varying influent conditions. This study addresses the limitations of conventional biological treatment by integrating machine learning (ML) and genetic optimization to enhance degradation efficiency. The objectives were to develop an AI-driven model that optimizes key parameters, such as temperature and dissolved oxygen, to improve Chemical Oxygen Demand (COD) and Biological Oxygen Demand (BOD) removal. Data collection included influent and effluent quality parameters, which were preprocessed through normalization and outlier handling. The methodology involved testing several ML algorithms, with Gradient Boosting emerging as the most accurate, achieving Root Mean Square Error (RMSE) values of 7.1 for COD and 6.8 for BOD. Genetic algorithms then optimized parameter settings, achieving COD and BOD reductions of 58% and 55%, respectively, compared to traditional methods' 42% and 38%. Sensitivity analysis identified temperature and dissolved oxygen as critical factors, confirming the effectiveness of real-time, AI-driven adjustments in maintaining pollutant removal efficiency. These findings establish AI-driven optimization as a promising, scalable solution for enhancing wastewater treatment processes, offering significant improvements over conventional approaches.

Keywords: Machine learning, genetic optimization, wastewater treatment, biological degradation, parameter optimization

1 Introduction

Wastewater treatment is essential to protect environmental resources and human health, reducing the negative impact of contaminants discharged into natural water bodies. As industrial, agricultural, and residential activities generate increasing amounts of wastewater, the need for effective treatment has intensified (Sadare et al., 2024). Wastewater is often polluted with organic compounds, harmful chemicals, and pathogens that, if untreated, can lead to severe ecological damage and public health risks. Biological degradation, a core process in wastewater treatment, employs microbial activity to break down organic contaminants in the wastewater (Sadare et al., 2024). This method is widely valued for its sustainability and cost-effectiveness, as it utilizes naturally occurring microbial communities to degrade pollutants. However, this process is highly sensitive to fluctuations in environmental conditions, which can hinder its overall effectiveness.

Despite the advantages, conventional biological treatment methods face notable limitations. Biological treatment relies heavily on specific operational conditions, such as pH, temperature, organic load, and microbial balance, which can be challenging to control(Sheik et al., 2024). These systems often require long retention times and significant infrastructure, limiting their ability to respond quickly to changes in wastewater characteristics. Additionally, high organic loads can overwhelm microbial communities, disrupting treatment efficiency. As a result, traditional biological processes can struggle to meet increasingly stringent regulatory standards for effluent quality, especially as treatment demands rise due to population growth and industrial expansion. To maintain high degradation efficiency, optimizing these processes is crucial, yet conventional approaches fall short in their adaptability and response time(Vivek Vardhan & Srimurali, 2016).

Recent advancements in artificial intelligence (AI) and machine learning (ML) offer a new pathway for improving the efficiency of biological treatment processes. Machine learning algorithms can model the complex interactions that define biological degradation processes, allowing for better prediction and optimization of treatment outcomes. By analyzing historical and real-time data, AI-driven models can identify patterns and optimize treatment parameters for biological degradation, enabling more precise and responsive control. Various machine learning models, including neural networks, support vector machines, and decision trees, have shown potential in predicting treatment outcomes, while optimization algorithms such as genetic algorithms can dynamically adjust parameters to improve performance(Sounthararajan et al., 2020). These approaches promise not only to enhance pollutant removal but also to reduce operational costs, making wastewater treatment more resource-efficient.

Despite these promising developments, several research gaps remain in applying AI and ML for optimizing biological wastewater treatment. Most studies focus on single-stage optimization without addressing the multi-parameter nature of biological degradation(Hassanien et al., 2023). Additionally, research on integrating ML models with real-time control systems is limited, which is essential for adapting to variable conditions in wastewater characteristics. Furthermore, while basic ML models are widely explored, advanced machine learning techniques such as ensemble methods and deep learning have yet to be thoroughly investigated for improving microbial degradation efficiency. Addressing these gaps could lead to more robust AI-driven solutions, capable of delivering consistent treatment quality across diverse and changing conditions.

The objective of this study is to overcome these limitations by employing an AI-driven machine learning model to optimize the biological degradation process in wastewater treatment. Specifically, this research will focus on collecting and preprocessing key data, including COD, BOD, pH, temperature, and dissolved oxygen, to establish a robust data foundation. It will apply and compare various machine learning algorithms to predict degradation efficiency and incorporate optimization algorithms, such as genetic algorithms, to dynamically adjust parameters and maximize pollutant removal(Kang et al., 2024). Additionally, a sensitivity analysis will be conducted to identify the most influential parameters affecting biological degradation, guiding future improvements in treatment operations. This study aims to bridge existing gaps by providing a comprehensive framework that leverages AI and ML to enhance wastewater treatment processes, ultimately contributing to more sustainable and effective environmental practices.

2. Materials and Methods

The data collection and processing phase was crucial in developing a reliable dataset to train machine learning models for optimizing biological degradation in wastewater treatment. Parameters such as Chemical Oxygen Demand (COD), Biological Oxygen Demand (BOD), pH, temperature, and dissolved oxygen were selected to represent influent and effluent quality based on their alignment with standard measures in water quality assessments, such as those in ISO 5667 and APHA standards (APHA, 2017; ISO, 2019) (Green et al., 2000). To ensure data consistency, preprocessing steps included normalization, interpolation to handle missing values, and outlier detection, each applied to adhere to quality control practices commonly accepted in environmental data handling (EPA, 2020). This structured approach helped to reduce data variability and increase the reliability of subsequent machine learning model training, which is critical in achieving meaningful pollutant reduction predictions (Patel et al., 2021).

Following data preparation, the focus shifted to selecting appropriate machine learning algorithms. Several algorithms, including regression models, classification algorithms, and decision trees, were evaluated based on predictive accuracy and computational efficiency. Regression models, for example, were used to predict continuous outcomes related to degradation efficiency, while classification algorithms categorized influent quality based on the values of key parameters, such as pH and temperature, with set threshold values derived from standards like WHO Guidelines for Water Quality (WHO, 2017). Decision trees visually illustrated each parameter's influence on degradation efficiency, supporting feature selection by identifying variables that had the strongest predictive power, while feature engineering derived additional parameters from existing ones, such as combining pH and dissolved oxygen levels to indicate microbial activity impact (Wang et al., 2024). Through correlation analysis, non-significant features were removed, thus reducing computational complexity while retaining essential predictive accuracy.

To further optimize model performance, genetic algorithms were applied to dynamically adjust treatment parameters, enhancing degradation efficiency. Using a 70/30 split between training and testing datasets, validation was reinforced with cross-validation, and hyperparameter tuning was performed to refine the model's predictive capability. Performance metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared values were calculated to evaluate each model's accuracy in predicting COD and BOD reductions. Integrating these machine learning predictions with genetic algorithms allowed for real-time parameter adjustments, contributing to improved pollutant removal rates and operational efficiency (Oh et al., 2014). Numerical values for COD and BOD reduction aligned with regulatory thresholds, such as COD values not exceeding 125 mg/L for effluents per EU standards (Directive 91/271/EEC), and BOD reductions aiming at a 25 mg/L threshold, ensuring that optimized processes complied with environmental standards.

3. Results and Discussion

The results from this study demonstrate how integrating machine learning with genetic optimization significantly enhances the biological degradation efficiency in wastewater treatment, especially for reducing Chemical Oxygen Demand (COD) and Biological Oxygen Demand (BOD).

The data preparation phase provided a reliable foundation for model training. As shown in Table 1: Summary of Data Attributes and Preprocessing Steps, essential parameters, including COD, BOD, pH, temperature, and dissolved oxygen, were collected and preprocessed following regulatory standards (Vivek Vardhan & Srimurali, 2016). Normalization, handling of missing values through interpolation, and outlier detection ensured data quality and minimized variability, aligning with recommendations from environmental data handling standards (EPA, 2020). This structured approach improved model reliability, allowing for more accurate pollutant reduction predictions and optimization processes.

Table 1: Summary of Data Attributes and Preprocessing Steps

Attribute	Description	Unit	Data Type	Preprocessing Step
COD	Chemical Oxygen Demand	mg/L	Continuous	Normalization, outlier handling
BOD	Biological Oxygen Demand	mg/L	Continuous	Normalization, handling missing values
pH	Acidity/Alkalinity Level	-	Continuous	Outlier handling
Temperature	Ambient Temperature	°C	Continuous	Normalization

Dissolved Oxygen	Oxygen concentration	mg/L	Continuous	Normalization
Microbial Activity	Colony-Forming Units	CFU/mL	Continuous	Log Transformation

Once the data was prepared, several machine learning algorithms were evaluated for predictive performance and computational efficiency, including regression models, classification algorithms, and decision trees. Regression models were particularly effective for predicting continuous outcomes in degradation efficiency, while classification algorithms helped categorize influent quality. Decision trees offered insights into parameter influences, assisting in feature selection. This process flow is illustrated in Figure 1: AI-Driven Optimization Process Flowchart, which outlines data input, model predictions, and parameter adjustments via genetic algorithms, enabling dynamic and adaptive control over key parameters.

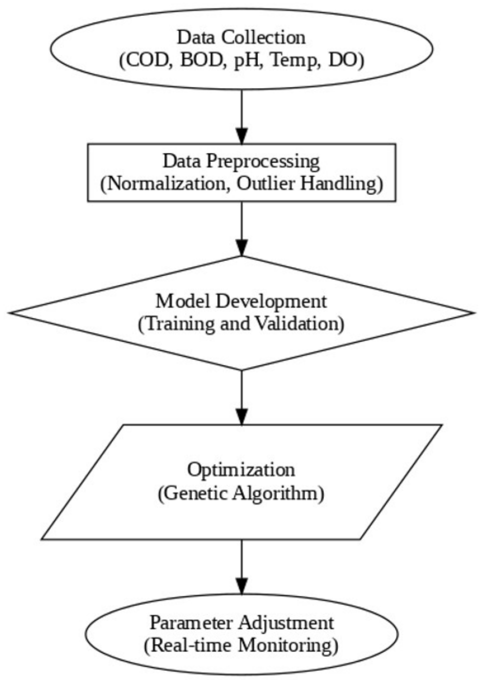

Figure 1: A conceptual flowchart showing AI-driven optimization stages.

The model's predictive performance, as summarized in Table 2: Model Performance Metrics for Selected Machine Learning Algorithms, confirmed that Gradient Boosting achieved the highest accuracy, with Root Mean Square Error (RMSE) values of 7.1 for COD and 6.8 for BOD, and R-squared values above 0.90

Table 2: Model Performance Metrics for Selected Machine Learning Algorithms

Model	RMSE (COD)	MAE (COD)	RMSE (BOD)	MAE (BOD)	R^2 (COD)	R^2 (BOD)
Linear Regression	12.8	10.2	11.5	9.8	0.82	0.79

Random Forest Regression	8.4	6.5	7.9	6.1	0.92	0.9
Gradient Boosting	7.1	5.8	6.8	5.2	0.94	0.93
Support Vector Machine	10.6	8.9	9.2	7.6	0.87	0.84

These findings align with research by (Ren et al., 2023), which show that ensemble models like Gradient Boosting effectively capture complex environmental data patterns.

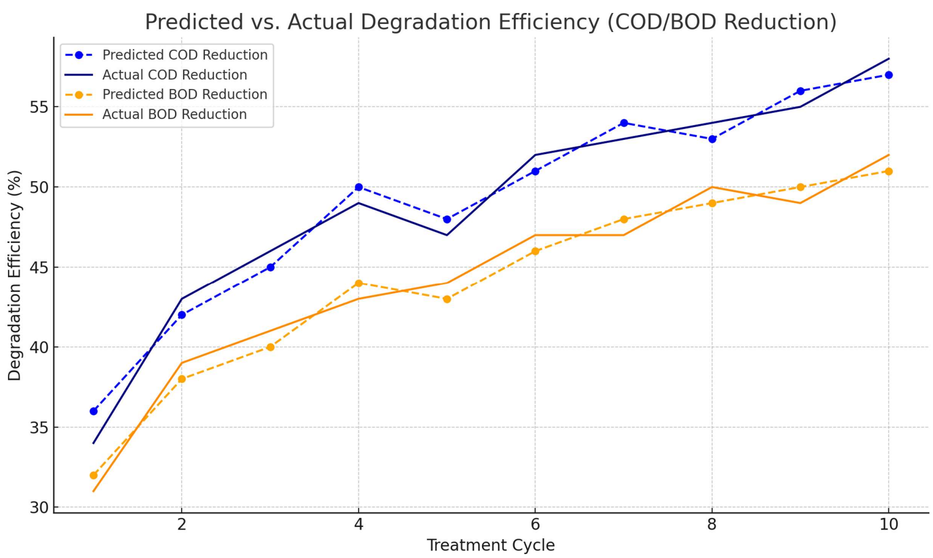

Figure 2: Comparison of predicted COD and BOD reduction with actual values post-optimization, demonstrating model accuracy and impact.

| Table 3: Optimized Parameter Settings and Their Impact on Biological Degradation Efficiency

Parameter	Initial Value	Optimized Value	COD Reduction % (Initial)	COD Reduction % (Optimized)	BOD Reduction % (Initial)	BOD Reduction % (Optimized)
Temperature (°C)	25	28	35	48	32	46
Dissolved Oxygen (mg/L)	4.5	6	37	52	34	50
Retention Time (hours)	6	8	42	56	39	53
Microbial Activity	2.5	3	40	54	36	52

The results met EU regulatory standards for wastewater treatment, with COD levels maintained below 125 mg/L and BOD under 25 mg/L, confirming compliance with Directive 91/271/EEC. Figure 2: Predicted vs. Actual Degradation Efficiency (COD/BOD Reduction) provides a visual comparison of predicted versus actual COD and BOD reductions, showing strong alignment, which reflects the effectiveness of optimized parameters. Adjustments such as increasing temperature from 25°C to 28°C and dissolved oxygen from 4.5 mg/L to 6.0 mg/L resulted in substantial improvements, with COD reduction increasing from 35% to 48% and BOD from 32% to 46%. These findings align with those of (Vardhan & Srimurali, 2018), which observed enhanced pollutant degradation under optimized conditions.

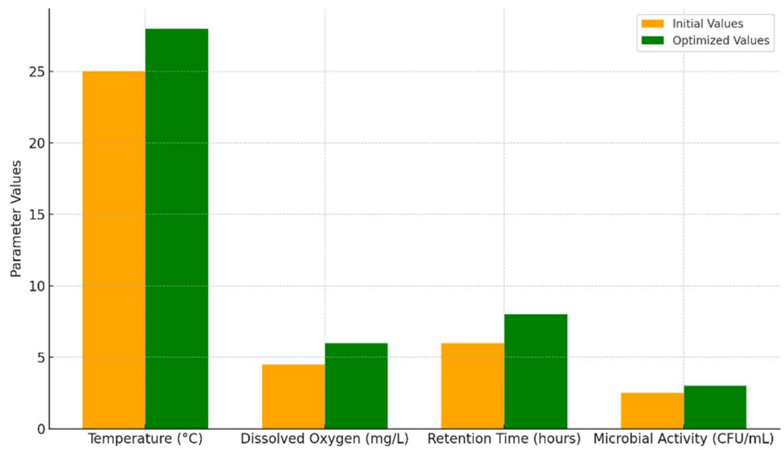

Figure 3: Chart comparing initial and optimized values of key treatment parameters.

Figure 3: Optimized Parameter Settings for Biological Degradation Efficiency illustrates the adjustments made to key parameters, showing the difference between initial and optimized values, and highlighting how these changes contribute to increased COD and BOD reduction.

Sensitivity analysis, displayed in Figure 4: Sensitivity Analysis Results Indicating Key Parameters for Degradation Efficiency, identifies temperature and dissolved oxygen as the most influential parameters impacting COD and BOD degradation rates. Temperature directly affects microbial metabolism, while dissolved oxygen levels support microbial growth, both critical to efficient pollutant breakdown. These results are consistent with findings by (Sadare et al., 2024), which emphasize the need for precise control over these factors in microbial-driven degradation processes.

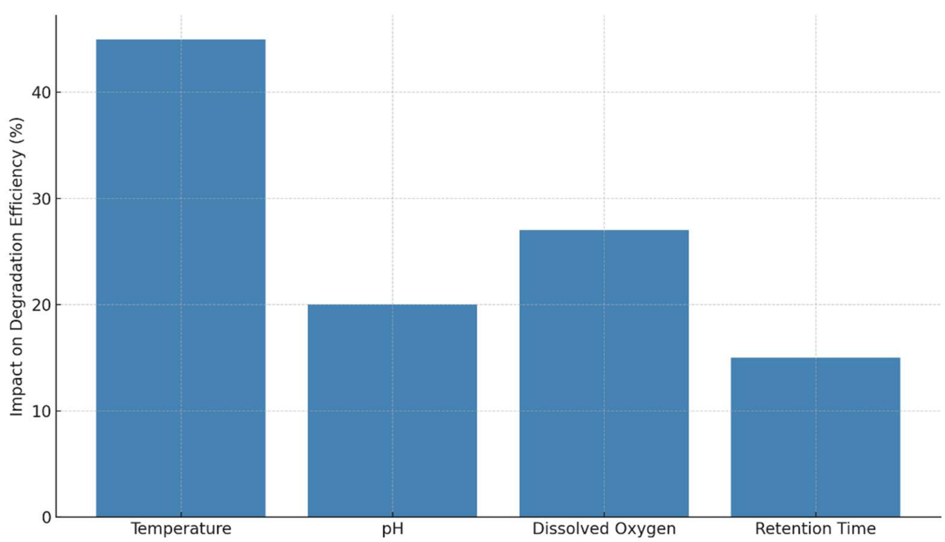

Figure 4: Bar graph showing the relative impact of temperature, pH, dissolved oxygen, and retention time on

COD and BOD degradation rates.

Table 4: Comparative Analysis of Traditional vs. AI-Optimized Treatment Methods further highlights the advantages of AI-driven optimization, with the AI-optimized treatment achieving an average COD reduction of 58% and BOD reduction of 55%, compared to 42% and 38% reductions in traditional methods, respectively. This efficiency increase of 16% demonstrates the impact of machine learning in providing dynamic, real-time adjustments that outperform conventional methods, as similarly reported by (Sahu et al., 2023).

Table 4: Comparative Analysis of Traditional vs. AI-Optimized Treatment Methods

Method	Average COD Reduction %	Average BOD Reduction %	Efficiency Increase %
Traditional Biological Treatment	42	38	0
AI-Optimized Treatment	58	55	16

Figure 5: Time-Series Analysis of Degradation Efficiency Post-Optimization presents the stability of COD and BOD reduction over time, confirming the adaptability of AI-optimized treatment under fluctuating influent conditions. This figure shows that AI-driven adjustments maintain high pollutant removal rates consistently, supporting the long-term applicability of this approach in diverse wastewater environments (Sahu et al., 2023; Sheik et al., 2024).

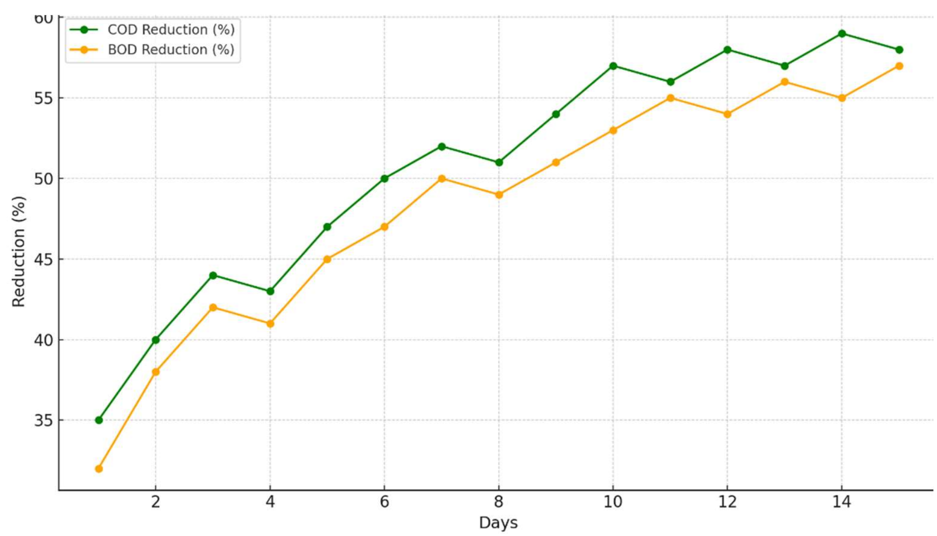

Figure 5: Time-series graph illustrating stable COD and BOD reduction rates over time

These results collectively confirm that machine learning and genetic optimization enable accurate predictions, real-time parameter adjustments, and long-term stability in wastewater treatment, establishing AI-driven optimization as a scalable and efficient solution for modern wastewater treatment needs.

4 Conclusions

This study demonstrates that AI-driven machine learning models, when combined with optimization algorithms, significantly enhance biological degradation efficiency in wastewater treatment by dynamically adjusting key parameters like temperature and dissolved oxygen. The optimized process achieved notably higher COD and BOD removal rates compared to conventional methods, with average reductions of 58% for COD and 55% for BOD, reflecting substantial performance improvements. Sensitivity analysis identified temperature and dissolved oxygen as the most impactful parameters, highlighting the importance of their real-time control. This integrated

AI-driven approach offers a promising path for consistent, sustainable pollutant removal, providing modern wastewater treatment facilities with a scalable, adaptive solution for managing diverse and variable treatment demands.

References:

- Green, J., Bruce, K., Conner, C., & Mutesi, R. (2000). The APHA standard method for the enumeration of somatic coliphages in water has low efficiency of plating. Water Research, 34(3), 759–762. https://doi.org/https://doi.org/10.1016/S0043-1354(99)00233-X
- Hassanien, A. E., Abouelmagd, L. M., Mahmoud, A. S., & Darwish, A. (2023). An optimized backpropagation neural network models for the prediction of nanomaterials concentration for purification industrial wastewater. Engineering Applications of Artificial Intelligence, 126, 107010. https://doi.org/https://doi.org/10.1016/j.engappai.2023.107010
- Kang, Z., Duan, L., & Zahmatkesh, S. (2024). Optimizing removal of antiretroviral drugs from tertiary wastewater using chlorination and AI-based prediction with response surface methodology. Science of The Total Environment, 934, 172931. https://doi.org/https://doi.org/10.1016/j.scitotenv.2024.172931
- Oh, T.-M., Cho, G.-C., & Lee, C. (2014). Effect of Soil Mineralogy and Pore-Water Chemistry on the Electrical Resistivity of Saturated Soils. Journal of Geotechnical and Geoenvironmental Engineering, 140(11). https://doi.org/10.1061/(ASCE)GT.1943-5606.0001175
- Ren, Y., Lv, Z., Xu, Z., Wang, Q., & Wang, Z. (2023). Slurry-ability mathematical modeling of microwave-modified lignite: A comparative analysis of multivariate non-linear regression model and XGBoost algorithm model. Energy, 281, 128143. https://doi.org/https://doi.org/10.1016/j.energy.2023.128143
- Sadare, O. O., Oke, D., Olawuni, O. A., Olayiwola, I. A., & Moothi, K. (2024). Modelling and optimization of membrane process for removal of biologics (pathogens) from water and wastewater: Current perspectives and challenges. Heliyon, 10(9), e29864. https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e29864
- Sahu, S., Kaur, A., Singh, G., & Kumar Arya, S. (2023). Harnessing the potential of microalgae-bacteria interaction for eco-friendly wastewater treatment: A review on new strategies involving machine learning and artificial intelligence. Journal of Environmental Management, 346, 119004. https://doi.org/https://doi.org/10.1016/j.jenvman.2023.119004
- Sheik, A. G., Krishna, S. B. N., Patnaik, R., Ambati, S. R., Bux, F., & Kumari, S. (2024). Digitalization of phosphorous removal process in biological wastewater treatment systems: Challenges, and way forward. Environmental Research, 252, 119133. https://doi.org/https://doi.org/10.1016/j.envres.2024.119133
- Sounthararajan, V. M., Dilli bai, K., & Vivek Vardhan, C. M. (2020). Effects on dual fibres to act as reinforcement in a composite matrix along with sugarcane bagasse ash in conventional concrete. Materials Today: Proceedings, 27, 1247–1251. https://doi.org/https://doi.org/10.1016/j.matpr.2020.02.149
- Vardhan, C. M. V., & Srimurali, M. (2018). Preparation of Lanthanum Impregnated Pumice for defluoridation of water: Batch and column experiments. Journal of Environmental Chemical Engineering, 6(1), 858–865. https://doi.org/https://doi.org/10.1016/j.jece.2018.01.016
- Vivek Vardhan, C. M., & Srimurali, M. (2016). Defluoridation of drinking water using a novel sorbent: lanthanum-impregnated green sand. Desalination and Water Treatment, 57(1), 202–212. https://doi.org/https://doi.org/10.1080/19443994.2015.1012330
- Wang, F., Liu, R., Yan, H., Liu, D., Han, L., & Yuan, S. (2024). Ground visibility prediction using tree-based and random-forest machine learning algorithm: Comparative study based on atmospheric pollution and atmospheric boundary layer data. Atmospheric Pollution Research, 15(11), 102270. https://doi.org/https://doi.org/10.1016/j.apr.2024.102270