6th Conference on Sustainability in Civil Engineering (CSCE'24)
Department of Civil Engineering
Capital University of Science and Technology, Islamabad Pakistan

OPTIMIZED PREDICTION MODELING FOR TiO_2-CATALYZED PHOTO-DEGRADATION RATE CONSTANTS OF WATER CONTAMINANTS USING MACHINE LEARNING

[a] Ayesha Noreen and [b] Majid Khan*

a: Departement of Civil Engineering, UET, Peshawar, Pakistan. 18pwciv5186@uetpeshawar.edu.pk
b: Departement of Civil Engineering, COMSATS University, Abbottabad, Pakistan. 18pwciv4988@uetpeshawar.edu.pk
* Corresponding author

Abstract- Titanium dioxide (TiO_2) is widely recognized as a photocatalyst in water contaminant treatment applications. Generally, obtaining the kinetics of photo-degradation rates requires extensive experimentation, involving significant labor and experimental resources. This study presents a unique approach to employing multi-expression programming (MEP) and artificial neural networks (ANN) to forecast the photo-degradation of water contaminants using TiO_2. The collected dataset for model development consists of 446 data points with six input variables. The MEP model exhibited higher prediction accuracy for the TiO_2-photocatalytic degradation of organic water contaminants. The MEP model exhibited predictions with R values of 0.946, 0.862, and 0.869 for training, testing, and validation, respectively. While the ANN model exhibited good accuracy during the training phase, its performance in testing and validation was poor, with the R-value significantly lower than the recommended threshold of 0.80. Among both developed models, MEP is a better choice to forecast the degradation of organic water contaminants using the TiO_2-photocatalytic process.

Keywords- Machine learning; Titanium dioxide; Photo-degradation of water contaminants; water treatment

1 Introduction

Recognizing the significance of environmental preservation, researchers are diligently pursuing eco-conscious alternative technologies across various aspects of daily living. Variations in water quality worldwide due to industrial and geographical factors, mean that a single solution cannot address all water contamination issues. Consequently, nanotechnology will always be essential in developing effective water treatment technologies [1]. In this context, photocatalysis using TiO_2 is a widely used technique for wastewater treatment. Titanium dioxide stands out among semiconductor photocatalysts due to its remarkable chemical and biological inertness, resistance to photoanodic corrosion, and cost-effective production. The use of TiO_2 in photocatalytic water and air purification is widely preferred for its efficacy and environmentally friendly nature, making it a favored advanced oxidation process [2]. Similarly, the photo-Fenton approach is also an effective advanced oxidation process for oxidizing water pollutants. Moreover, the non-selective nature of the photocatalytic process enables TiO_2 to degrade contaminants [4]. Over recent decades, extensive research has focused on experimentally examining the photo-catalytic degradation of various water pollutants using TiO_2. The recent research has also delved into exploring the underlying principles and theoretical aspects of TiO_2 performance. However, experimental testing requires significant time and financial investment, demanding considerable resources in terms of labor and equipment. To address the challenges associated with experimental testing, machine learning (ML) approaches have recently been widely utilized to predict outcomes, optimizing both time and resources. By leveraging ML, researchers can efficiently model complex systems and identify key variables, significantly reducing the need for extensive and costly experimentation. This integration of ML into water treatment technology development, particularly in the context of nanotechnology, offers a promising pathway to tailor solutions for better wastewater treatment [5,6].

This study introduces two predictive models utilizing MEP and ANN to estimate the photo-degradation rate constants of contaminants exposed to TiO_2 and ultraviolet irradiation in aqueous solutions. The database for these predictive models comprises 446 data points gathered from published experimental reports. Six input variables are considered for the model development. Different statistical performance metrics were utilized to assess the efficacy of the developed models.

2 Research Methodology

2.1. Data collection

The database for model development was collected from 26 published experimental studies, acquired by Jiang et al. [7]. The dataset consists of 446 data points. Six experimental variables were considered which include, experimental temperature (T), type of organic contaminant (OC), TiO_2 dosage (D), ultraviolet light intensity (I), initial water contaminant concentration (IC), and solution pH (pH). The response parameter is the photo-degradation rate constant (k), which is converted to its base 10 logarithm -log(k) for better understanding. In a typical experimental procedure, a specified content of Degussa TiO_2 was mixed with polluted water with a modified pH for about 45 minutes to achieve adsorption-desorption equilibrium before UV light irradiation. The statistical description of the input variables is given in **Table 1**. The suspension was exposed to UV light. At specified intervals, small samples were extracted and filtered to remove the TiO_2. The first order of k was determined using Eq. 1.

$$\frac{dC}{dt} = kC \quad (1)$$

Table 1 Statistical analysis of the collected dataset

Statistics	I (mW cm^{-2})	T (°C)	D (g L^{-1})	IC (mg L^{-1})	pH	k (min^{-1})
Mean	3.90	27.78	1.12	43.29	5.24	1.21
Kurtosis	6.01	1.67	8.33	4.71	0.56	1.90
Skewness	2.16	1.33	2.03	1.85	1.01	0.70
Mode	0.18	25.00	1.00	10.00	3.60	1.30
Standard Deviation	12.34	7.54	1.01	50.99	1.96	0.55
Minimum	0.18	20.00	0.00	0.13	2.00	0.02
Range	74.82	40.00	7.50	342.34	9.00	3.98

2.2. Model development

The ANN model was implemented in MATLAB 2021a. The hyperparameters of the ANN model were optimized using random search, where the parameters of the model are selected arbitrarily from a predefined space. This method enables extensive exploration of the parameter range. Different random combinations of hyperparameters are evaluated based on a predetermined metric to assess model performance. The goal is to find the combination that yields the best results. For optimizing the ANN model, the number of neurons is adjusted to improve accuracy. The Levenberg-Marquardt was utilized for the training process. The optimized parameters of the ANN model are given in **Table 2**. Certain parameters for the MEP approach need to be adjusted to build a good predictive model. These parameters are selected based on an iterative process. The population size defines the established programs; a larger population can result in a more intricate and reliable model but takes longer to converge and risks overfitting if too large. **Table 2** details the setup variables used in this work. The function set includes basic mathematical operations (-, ×, ÷, +, sqr, exp) to establish a simple model. The number of generations affects the method's precision, with more generations leading to fewer errors. Various variable combinations were tested to hyper-tune the model. The optimized prediction model with the lowest error was selected as presented in Table 2.

Table 2. Hyperparameters optimization of the suggested models

MEP		ANN	
Parameter	Optimized value	Parameter	Optimized value
Sub-population size	50	Number of epochs	100
Mutation probability	0.01	Activation function	purelin, logsig, tansig
Generations	250	Learning rate	0.26
Operators	sqr, exp, ×, ÷, +, −	Momentum rate	0.9
Tournament size	2	Hidden neurons	40

Length of code	0.5		
Fitness function	MAE		
Crossover prob.	0.9		

3 Results

3.1 ANN Model Performance

The performance of the ANN is depicted in Figures 1-3, demonstrating a consistent decrease in error, indicative of improved model accuracy with more training cycles. Figure 1 highlights the MSE across validation, training, and testing stages, showing a significant optimization after seven epochs. This trend is corroborated by Figures 2 and 3, underscoring the model's effectiveness in reducing error throughout various stages. This hyper-tuning process provides a better understanding of the ANN performance and convergence. The training of the ANN utilized R and MSE as an ultimate termination approach. Specifically, adjustments to the neural network architecture were made when the accuracy metric (R) showed no significant improvement over several epochs, or when the R-value on the validation set plateaued and the MSE stopped decreasing. The ANN model with a 40-neuron architecture demonstrated good prediction records. To further provide a performance assessment of the ANN model, regression plots of the ANN approach are illustrated in Figure 4. The ANN model showed excellent precision with an R-value of 0.843 in training and an MSE value of 0.156. Moreover, while the model exhibited good accuracy during the training phase, its performance in testing and validation was poor, with the R-value significantly lower than the recommended threshold of 0.80. This disparity suggests that the model may be overfitting to the training data, failing to generalize well to unseen data.

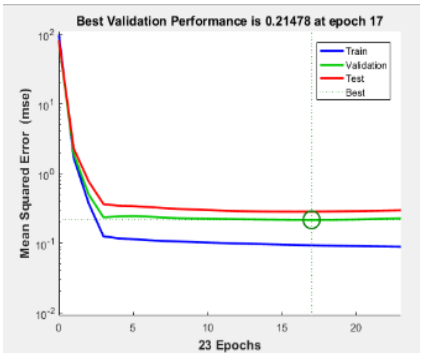

Figure 1: ANN model Performance

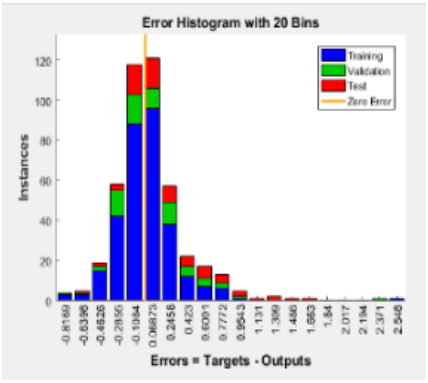

Figure 2: ANN model. Training state

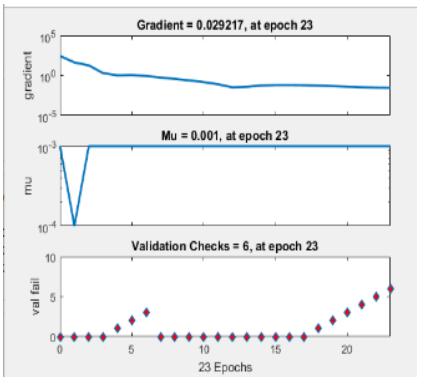

Figure 3: ANN model Error Limits

3.2 MEP Model Performance

The performance of the established MEP model is illustrated in Figure 5**.** It is generally recommended that the regression slope should be higher than 0.80 for an accurate prediction model. The MEP approach showed good predictions with a regression slope value of 0.996 for training, 0.901 for testing, and 0.856 for validation. These values are significantly higher than the recommended criteria of 0.80, suggesting a good alignment between the model and experimental data. Furthermore, the fitting lines for all three subsets are well-aligned with the ideal fit line, indicating that the MEP model provided comparable performance for all three subsets, thereby overcoming the issue of overfitting. Moreover, the MEP model demonstrated good accuracy in predicting the output, with R and MSE values of 0.946 and 0.064 for training, 0.862 and 0.184 for testing, and 0.869 and 0.083 for validation. The R-values for the three subsets are higher than 0.80 and close to 1, indicating the MEP model's potential to provide high prediction accuracy. The MEP model exhibited superior efficacy compared to the developed ANN model, making it the best choice for estimating TiO2-catalyzed photo-degradation. Additionally, the lower MSE values across training, testing, and validation further reinforce the reliability and robustness of the MEP model in practical applications. In addition, the MEP model provides better accuracy than the ML-based prediction model developed by Jiang et al. [7].

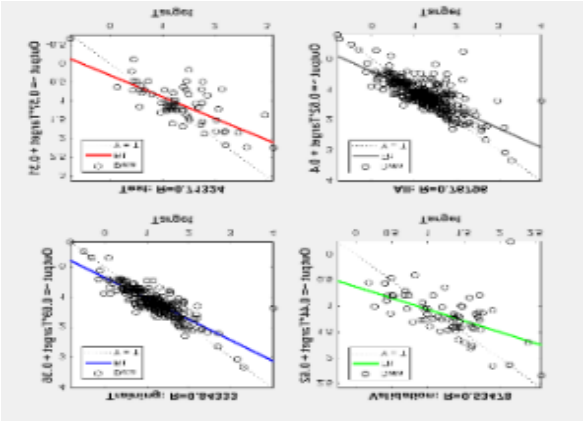

Figure 4: ANN model Regression analysis

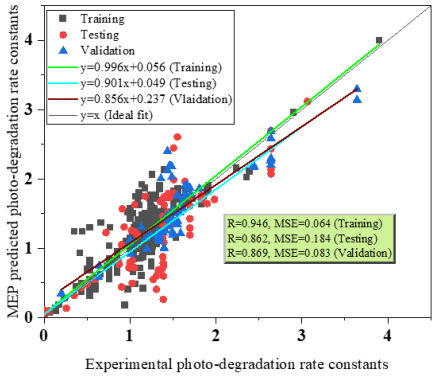

Figure 5: Performance of the developed MEP model

4 Conclusion

In the present study, ANN and MEP models were developed to predict the TiO_2-assisted photo-degradation efficiency for the removal of water contaminants. The models incorporated data from extensive studies with six input features and considered 78 organic contaminants, covering most experimental setups for the photocatalytic degradation process. SHAP interpretations were provided to enhance model interpretability. The major findings of the study are summarized herein:

1. The MEP model exhibited higher prediction accuracy for the TiO_2- photocatalytic degradation of organic water contaminants. The MEP model exhibited predictions with R values of 0.946, 0.862, and 0.869 for training, testing, and validation, respectively.
2. ANN model showed excellent accuracy with an R-value of 0.843 in training and an MSE value of 0.156. Moreover, while the model exhibited good accuracy during the training phase, its performance in testing and validation was poor, with the R-value significantly lower than the recommended threshold of 0.80.
3. Among both developed models, MEP is a better choice to forecast the TiO_2- photocatalytic degradation of organic water contaminants.

Reference

[1] M.M. Mahlambi, C.J. Ngila, B.B. Mamba, Recent Developments in Environmental Photocatalytic Degradation of Organic Pollutants: The Case of Titanium Dioxide Nanoparticles—A Review, J. Nanomater. 2015 (2015) 1–29. https://doi.org/10.1155/2015/790173.

[2] Z. Guo, X. Zhang, X. Li, C. Cui, Z. Zhang, H. Li, D. Zhang, J. Li, X. Xu, J. Zhang, Enhanced charge separation by incomplete calcination modified co-doped TiO2 nanoparticle for isothiazolinone photocatalytic degradation, Nano Res. 17 (2024) 4834–4843. https://doi.org/10.1007/s12274-024-6453-4.

[3] V.J.P. Vilar, L.X. Pinho, A.M.A. Pintor, R.A.R. Boaventura, Treatment of textile wastewaters by solar-driven advanced oxidation processes, Sol. Energy. 85 (2011) 1927–1934. https://doi.org/10.1016/j.solener.2011.04.033.

[4] A.J. Haider, Z.N. Jameel, I.H.M. Al-Hussaini, Review on: Titanium Dioxide Applications, Energy Procedia. 157 (2019) 17–29. https://doi.org/10.1016/j.egypro.2018.11.159.

[5] J.C. Serna-Carrizales, A.I. Zárate-Guzmán, R. Flores-Ramírez, L. Díaz de León-Martínez, A. Aguilar-Aguilar, W.M. Warren- Vega, E. Bailón-García, R. Ocampo-Pérez, Application of artificial intelligence for the optimization of advanced oxidation processes to improve the water quality polluted with pharmaceutical compounds, Chemosphere. 351 (2024) 141216. https://doi.org/10.1016/j.chemosphere.2024.141216.

[6] I. Salahshoori, M. Namayandeh Jorabchi, A. Baghban, H.A. Khonakdar, Integrative analysis of multi machine learning models for tetracycline photocatalytic degradation with MOFs in wastewater treatment, Chemosphere. 350 (2024) 141010. https://doi.org/10.1016/j.chemosphere.2023.141010.

[7] Z. Jiang, J. Hu, X. Zhang, Y. Zhao, X. Fan, S. Zhong, H. Zhang, X. Yu, A generalized predictive model for TiO2–Catalyzed photo-degradation rate constants of water contaminants through artificial neural network, Environ. Res. 187 (2020) 109697. https://doi.org/10.1016/j.envres.2020.109697.