KSÜ Mühendislik Bilimleri Dergisi, 28(1), 2025 KSU J Eng Sci, 28(1), 2025
Araştırma Makalesi Research Article

Kahramanmaras Sutcu Imam University
Journal of Engineering Sciences

Geliş Tarihi : 13.11.2024 Received Date : 13.11.2024
Kabul Tarihi :20.12.2024 Accepted Date : 20.12.2024

A GRAPHICAL USER INTERFACE DESIGN FOR FORECASTING NUTRIENT CONCENTRATIONS IN WWTP

ATIKSU ARITIM TESİSİNDE ORGANİK MADDE ÖNGÖRÜSÜ İÇİN GRAFİKSEL ARAYÜZ GELİŞTİRİLMESİ

Eda GÖZ[1] (ORCID: 0000-0002-3111-9042)

[1] Ankara University, Department of Chemical Engineering, Ankara, Türkiye

*Sorumlu Yazar / Corresponding Author: Eda Göz, esemizer@eng.ankara.edu.tr

ABSTRACT

Wastewater management poses a global challenge. Integrating data-driven models has significantly enhanced treatment facilities' design and operational efficiency. In this study, an Artificial Neural Network (ANN) algorithm was adapted as a time-series forecasting model to predict effluent TN (Total Nitrogen) and TP (Total Phosphorus) concentrations in a real municipal wastewater treatment plant (WWTP). For this purpose, six independent TN and TP models were developed and evaluated using Mean Absolute Percentage Error (MAPE, %) and Root Mean Square Error (RMSE) metrics. Based on these criteria, all models demonstrated similar performance, with MAPE and RMSE values for TN forecasting at approximately 12% and 1.4, respectively, in the test phase. The MAPE was approximately 30% for TP forecasting, and RMSE was 0.25. Upon completing the modeling studies, one model was integrated into a user-friendly graphical user interface (GUI) and tested with actual data, allowing users to obtain results with a single click.

Keywords: Wastewater treatment plant, total nitrogen, total phosphorus, modelling, GUI design

ÖZET

Atıksu yönetimi, dünya genelinde önemli bir zorluk teşkil etmektedir. Veriye dayalı modellerin entegrasyonu, arıtma tesislerinin tasarım ve işletim verimliliğini artırmıştır. Bu çalışmada, yapay sinir ağı algoritması (ANN) evsel atıksu arıtım tesisinde çıkış akımı toplam azot (TN) ve toplam fosfor (TP) parametrelerinin tahmini için zaman serisi öngörüsü yapacak şekilde modifiye edilmiştir. Bu amaçla, 6 farklı ve bağımsız TN ve TP modelleri geliştirilmiştir. Model performansı, Ortalama mutlak yüzde hatası (MAPE, %) ve Hataların Karesinin Ortalamasının Karekökü (RMSE) ile değerlendirilmiştir. Bu kriterlere göre tüm alternatif modeller benzer performans sergilemiştir. Çıkış akımı toplam azot (TN) tahmin modellerinin test fazında MAPE ve RMSE değerleri sırasıyla %12 ve 1.4 civarında elde edilmiştir. Çıkış akımı toplam fosfor (TP) için MAPE değeri yaklaşık %30, RMSE ise 0.25 olarak hesaplanmıştır. Modelleme çalışmaları tamamlandıktan sonra, bir model kullanıcı dostu bir grafik kullanıcı arayüzüne (GUI) entegre edilmiştir ve gerçek verilerle test edilmiştir. Bu, kullanıcıya tek tıklama ile sonuç alma imkânı sunmaktadır.

Anahtar Kelimeler: Atıksu Arıtım Tesisi, toplam azot, toplam fosfor, modelleme, grafiksel arayüz tasarımı

INTRODUCTION

Water is essential for sustaining life, yet water pollution has dramatically increased in recent years due to industrialization, technological advancements, rapid population growth, and anthropogenic activities (Akhtar et al., 2021). Wastewater treatment has become crucial to prevent pollution, maintain water quality, and ensure water resource sustainability. However, wastewater treatment is a complex chemical, physical, and biological process (Sravan et al., 2024), and the operation of treatment plants is challenging due to these processes' nonlinear and

ToCite: GÖZ, E., (2025). A GRAPHICAL USER INTERFACE DESIGN FOR FORECASTING NUTRIENT CONCENTRATIONS IN WWTP. *Kahramanmaras Sutcu Imam University Journal of Engineering Sciences, 28(1)*, 479-486.

intricate nature. Modeling techniques are widely employed to better understand these processes, with mechanistic and empirical models providing foundational insights. Nonetheless, these models often involve complex equations and require in-depth process knowledge.

The rise of data mining has spurred the application of machine learning methods to wastewater treatment data. These methods identify input and output parameters and train models using historical data. In this context, artificial neural networks (ANN), support vector machines (SVM), extreme learning machines (ELM), fuzzy logic (FL), deep learning (DL), and random forest (RF) algorithms have been utilized to predict parameters such as biological oxygen demand (BOD), chemical oxygen demand (COD), and energy efficiency (Zhao et al., 2020; Safeer et al., 2022).

Wastewater treatment plants also need to monitor critical parameters, such as total nitrogen (TN) and total phosphorus (TP), which contribute to eutrophication and biodiversity loss due to algae growth. Measurement, monitoring, and prediction of these parameters are essential, although fewer TN and TP models exist for full-scale plants compared to BOD and COD models. For example, Manu and Thalla (2017) predicted total Kjeldahl nitrogen (TKN) using Support Vector Machines (SVM) and Adaptive-Network-Based Fuzzy Inference Systems (ANFIS) with data from a full-scale aerobic treatment plant. Model inputs included pH, COD, total solids (TS), free ammonia, and Kjeldahl nitrogen, with SVM outperforming ANFIS, based on correlation coefficient (R), root mean squared error (RMSE), and Nash-Sutcliffe (NS) coefficient.

Other studies have explored diverse modeling approaches. Yu and Bai (2018) applied least-squares support vector machine (LS-SVM), ELM, and ANN to predict TP, using inputs like COD, BOD, NH_3-N, suspended solid (SS), TP, and TN, with ELM showing the best results. Similarly, Abba et al. (2021) used ELM with principal component analysis and combined ANN with multiple linear regression to predict TN and TP values, with ELM yielding superior performance based on R^2, RMSE, and MAPE. Additional studies (Mohammadi et al., 2022; Wang et al., 2022) have tested a range of algorithms—such as K-nearest neighbours (KNN), decision trees (DT), gradient boosting decision trees (GBDT), Ridge, Lasso, and SVR—on effluent TN and TP prediction, with KNN and GBDT achieving strong performance.

Early warning or alarm systems based on time series forecasting are invaluable for real-time wastewater treatment applications. The Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) model has been widely used (Ly et al., 2022), often alongside recurrent neural network (RNN) models like Long Short-Term Memory (LSTM) networks, which are highly effective but require substantial datasets (Xu et al., 2021; Oliveira et al., 2021; Safeer et al., 2022). El-Rawy et al. (2021) also developed deep learning models for time series forecasting and Artificial Neural Network (ANN)--based prediction models targeting total suspended solid (TSS), COD, BOD_5, ammonia, and sulfur removal. These studies emphasize the importance of time series forecasting models in optimizing wastewater treatment plant operations, focusing on enhancing system efficiency and preventing operational issues. In a related study by Hansen et al. (2022), two distinct LSTM models were designed to predict TP concentration: the first model integrated environmental conditions with process operation data, while the second relied on previous TP values. The R^2 score was chosen as the performance metric to evaluate the models.

Despite the variety of models available, integrating these models into accessible systems for operators remains essential for streamlining operations in wastewater treatment plants. However, such user-friendly applications remain limited in the literature. Six independent ANN models with distinct input combinations were developed to address this gap and predict effluent TN and TP concentrations for the next day in a domestic wastewater treatment plant. Model performances were compared using RMSE and MAPE (%), and a model with optimal performance was incorporated into an interactive user interface in MATLAB 2023b. This interface simplifies model use, enabling users to obtain results with a single click, significantly reducing workload and laboratory costs.

MODELING STUDIES

Data Collection and Preprocessing

Data for this study were obtained from a wastewater treatment plant in South Carolina, USA, and include daily records from 2014 to 2018, comprising 2,168 data points. The dataset was converted into a time series format with mean imputation for any missing values. All data were normalized to a 0-1 range for consistency, as shown in Equation 1.

$$x_{nor} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Modelling Studies

This study developed independent models for predicting total nitrogen (TN) and total phosphorus (TP) concentrations, designated as Models 1.1 to 1.6 for TN and Models 2.1 to 2.6 for TP. Each model utilized six different input combinations, selected based on literature review and expert insights. Table 1 presents the specific input and output parameters for each model.

Table 1. Inputs and Outputs for Total Nitrogen (TN) and Total Phosphorus (TP) Models

Input Parameters	Output Parameters	Model 1.1/2.1	Model 1.2/2.2	Model 1.3/2.3	Model 1.4/2.4	Model 1.5/2.5	Model 1.6/2.6
Influent BOD (t)		√	√	√	√		
Influent TKN (t)		√		√		√	
Influent NH_3-N (t)			√		√		√
Influent TP (t)				√	√	√	√
Influent TSS (t)		√	√	√	√	√	√
Aeration basin pH(t)	Effluent TN (t+1)	√	√	√	√	√	√
Aeration basin alkalinity (t)		√	√	√	√	√	√
Aeration basin DO (t)	Effluent TP (t+1)	√	√	√	√	√	√
MLSS(t)		√	√	√	√	√	√
SRT (t)		√	√	√	√	√	√
Effluent temperature (t)		√	√	√	√	√	√
Effluent Flow rate (t)		√	√	√	√	√	√
Effluent TN(t)		√	√	√	√	√	√
Effluent TP (t)		√	√	√	√	√	√

DO(t) Dissolved oxygen concentration, MLSS: Mixed liquor suspended solids (MLSS), SRT: Sludge retention time

Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) are computational models that emulate the human brain's learning processes to generate new insights from data (Haykin, 1999). ANNs are fundamentally mathematical models trained on historical datasets and consist of interconnected units called neurons. A typical ANN architecture includes three main layers: the input layer (where data is received), the hidden layer (where data processing occurs), and the output layer (where results are generated). Neurons in these layers are connected through synapses, and each connection is assigned a weight. Transfer functions adjust signals passing through neurons, and this process is repeated until the network generates the desired output. The number of neurons in the hidden layer is generally problem-specific and often determined through trial and error.

This study employed a three-layer ANN structure, with separate models corresponding to each input combination in Table 1. The models were designed to predict TN and TP concentrations one day in advance. The dataset was divided into 60% for training, 20% for testing, and 20% for validation. During hyperparameter optimization, various activation functions and neuron quantities in the hidden layer were evaluated to identify optimal settings. Among tested activation functions, tangent sigmoid (‘*tansig*’), linear (‘*purelin*’), and logarithmic sigmoid (‘*logsig*’) provided the best results in the hidden and output layers. Training algorithms, including Levenberg-Marquardt (‘*trainlm*’), Bayesian regularization (‘*trainbr*’), and BFGS quasi-Newton (‘*trainbfg*’), were applied, with Bayesian regularization yielding the highest accuracy.

Model performances were compared using RMSE and MAPE (%) metrics, which offer a more reliable assessment for nonlinear models than R^2 or R. R^2, while widely used, has limitations in nonlinear models as it does not always encapsulate simpler models with single parameters and fails to account for the number of parameters in complex models (Archontoulis and Miguez, 2015; Manav-Demir et al., 2022). This highlights the need for composite metrics, such as RMSE and MAPE (%), to provide a more accurate evaluation. The formulas for the metrics are provided in Equation 2 and Equation 3:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - x_i)^2}{N}} \quad (2)$$

$$MAPE(\%) = \frac{1}{N}\sum_{i=1}^{N}\left(\left|\frac{y_i - x_i}{x_i}\right|\right)100 \tag{3}$$

N is the total number of data points in this equation, and y_i and x_i represent the observed and predicted values, respectively.

Graphical User Interface Design

A user-interactive interface for predicting effluent TN and TP concentrations was developed in the MATLAB environment. This software incorporates the ANN models and enables users to predict future TN and TP values with a single click. The interface is designed for ease of use, allowing operators to select from various TN and TP models as needed. An illustration of the graphical user interface, which includes options for selecting different models, is provided in Figure 1.

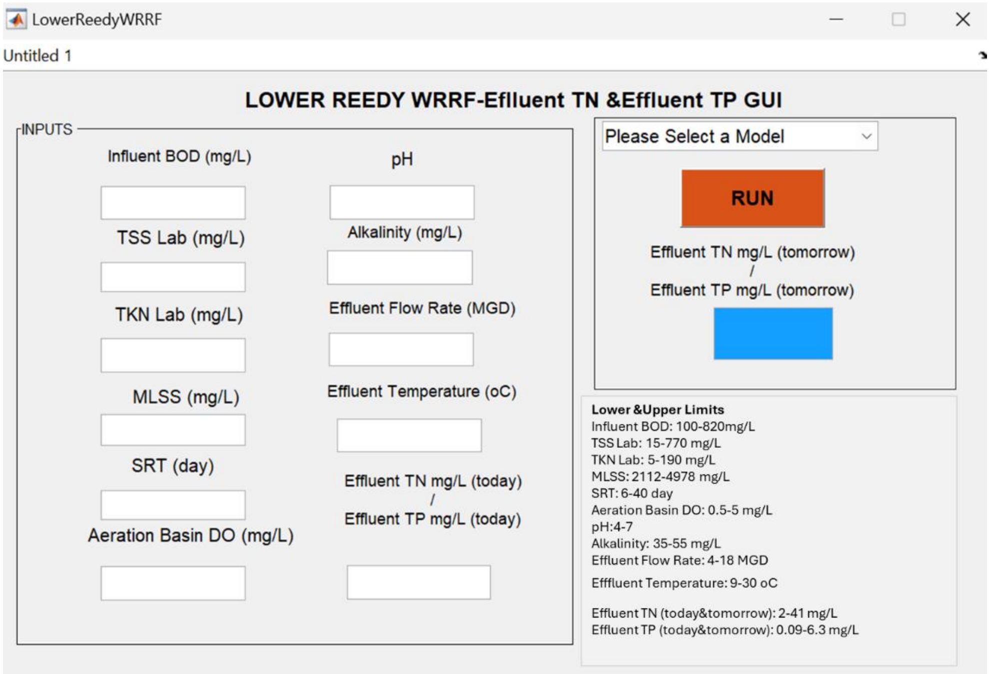

Figure 1. Overview of the Graphical Interface

The interface shown in Figure 1 provides a user-friendly area for inputting data, selecting models, and running predictions. In the input section, users can enter specific data values, reset the form, choose among different TN and TP models, execute the selected model, and print the output data. The design emphasizes simplicity and usability, making it an efficient tool for operators.

The steps to use the software are as follows:

- Launch MATLAB.
- Use the "Browse Folder" option to specify the folder location.
- Enter "LowerReedyWRRF" in the command window and press ENTER.
- The LowerReedy WRRF software page will open (as illustrated in Figure 1).
- Enter the model parameters in the appropriate fields on the interface.
- Select the desired model and click the RUN button to initiate the prediction.

RESULTS AND DISCUSSION

Total Nitrogen Forecasting Results

This study proposed six different models for predicting total nitrogen (TN) concentration in the effluent stream of a wastewater treatment plant. Evaluating influent TKN and NH_3-N data was essential for accurately predicting effluent TN concentrations. Test phase results for these models are presented in Table 2.

Table 2. Total Nitrogen (TN) Forecasting Results

		Test Phase		Train Phase		Validation Phase	
Models	**Algorithm**	**MAPE %**	**RMSE**	**MAPE %**	**RMSE**	**MAPE %**	**RMSE**
Model 1.1	ANN	12.71	1.44	11.01	1.97	15.11	1.7
Model 1.2	ANN	12.15	1.4	10.8	1.96	14.73	1.68
Model 1.3	ANN	12.77	1.45	10.97	1.97	15.24	1.71
Model 1.4	ANN	12.61	1.44	10.77	1.95	15.05	1.69
Model 1.5	ANN	12.16	1.41	10.78	1.96	14.81	1.68
Model 1.6	ANN	12.13	1.4	10.75	1.96	14.89	1.68

Upon examining the test phase results in Table 2, it is observed that the model incorporating NH_3-N (Model 1.2) performs slightly better in predicting effluent TN concentration than the model using TKN (Model 1.1). This can be attributed to the greater influence of influent NH_3-N on effluent TN values compared to TKN. Pearson correlation analysis supports this finding, with correlation coefficients of R = 0.14 for TKN and R = 0.22 for NH_3-N. However, the minor differences in RMSE and MAPE values between Models 1.1 and 1.2 may stem from the high NH_3-N/TKN ratio of 0.56 in the dataset, indicating that the organic nitrogen portion of TKN, in addition to NH_3-N, has a limited effect on model performance.

The limited impact of other selected parameters on model performance may be due to the inherent characteristics of the time series models, where the previous day's total nitrogen concentration is the most significant predictor. This is supported by a Pearson correlation coefficient of R = 0.44 between the previous day's TN value and the effluent TN concentration. Prediction and actual fit graphs for total nitrogen concentration are displayed in Figure 2.

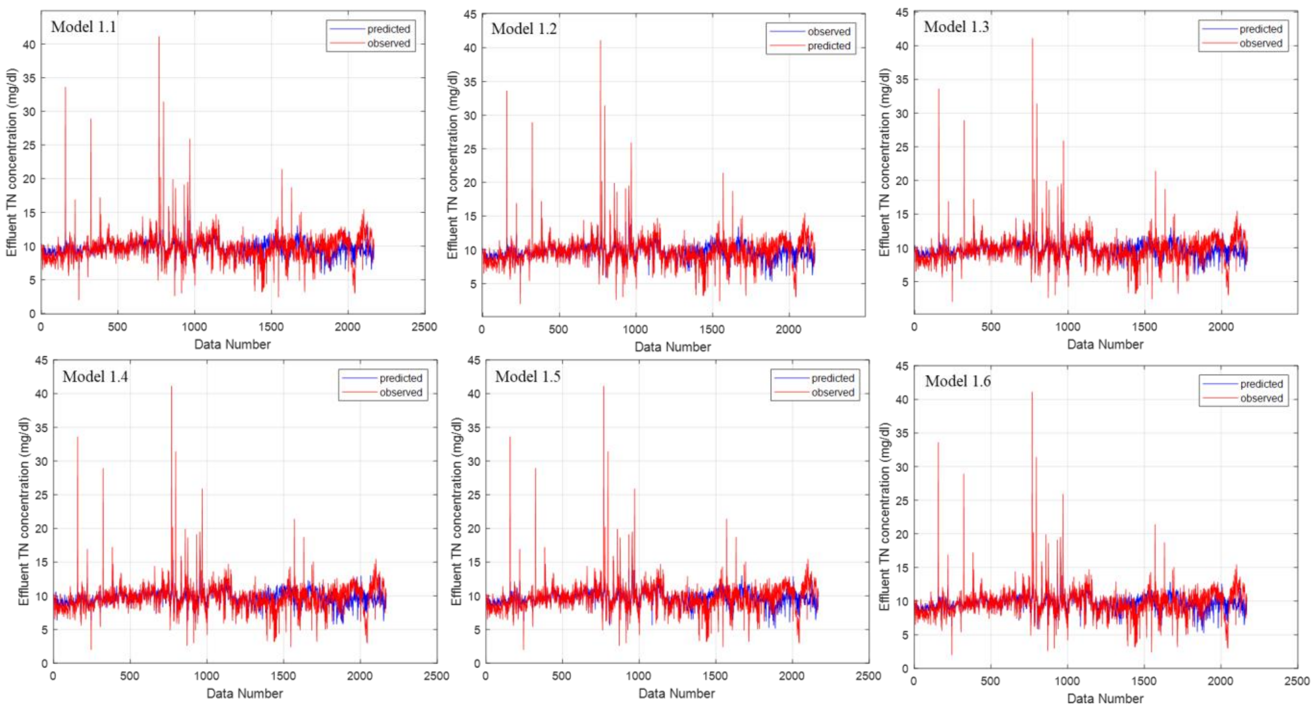

Figure 2. TN Forecasting Results

Model 1.1 was selected from the developed models for integration into the GUI due to its comparable performance and use of influent TKN data. This selection is advantageous as NH_3-N is already included within the Total Kjeldahl Nitrogen (TKN) parameter.

Total Phosphorus Forecasting Results

This study also proposed six models to predict total phosphorus (TP) concentration in the effluent stream of a wastewater treatment plant. In contrast to the TN model, TP predictions were based on the previous day's TP data rather than TN. The test phase performance results for effluent TP predictions are provided in Table 3.

Table 3. Effluent TP Forecasting Results

		Test Phase		Train Phase		Validation Phase	
Models	**Algorithm**	**MAPE %**	**RMSE**	**MAPE %**	**RMSE**	**MAPE %**	**RMSE**
Model 1.1	ANN	30.34	0.24	56.81	0.41	43.97	0.42
Model 1.2	ANN	30.08	0.26	55.45	0.41	44.26	0.42
Model 1.3	ANN	31.23	0.25	56.52	0.41	42.73	0.42
Model 1.4	ANN	30.46	0.26	55.03	0.41	44.01	0.42
Model 1.5	ANN	30.15	0.24	56.35	0.41	43.54	0.42
Model 1.6	ANN	29.65	0.26	54.8	0.41	44.66	0.42

Upon examining Table 3, no significant differences in performance were observed among the models. Correlation coefficient analysis between the effluent TP value and influent parameters revealed that the TP value from the previous day had the highest correlation (R = 0.55), which is typical for time series models. Among other parameters, temperature showed the next highest correlation with effluent TP (R = 0.11). The fit between experimental and predicted TP values from the test phase is illustrated in Figure 3.

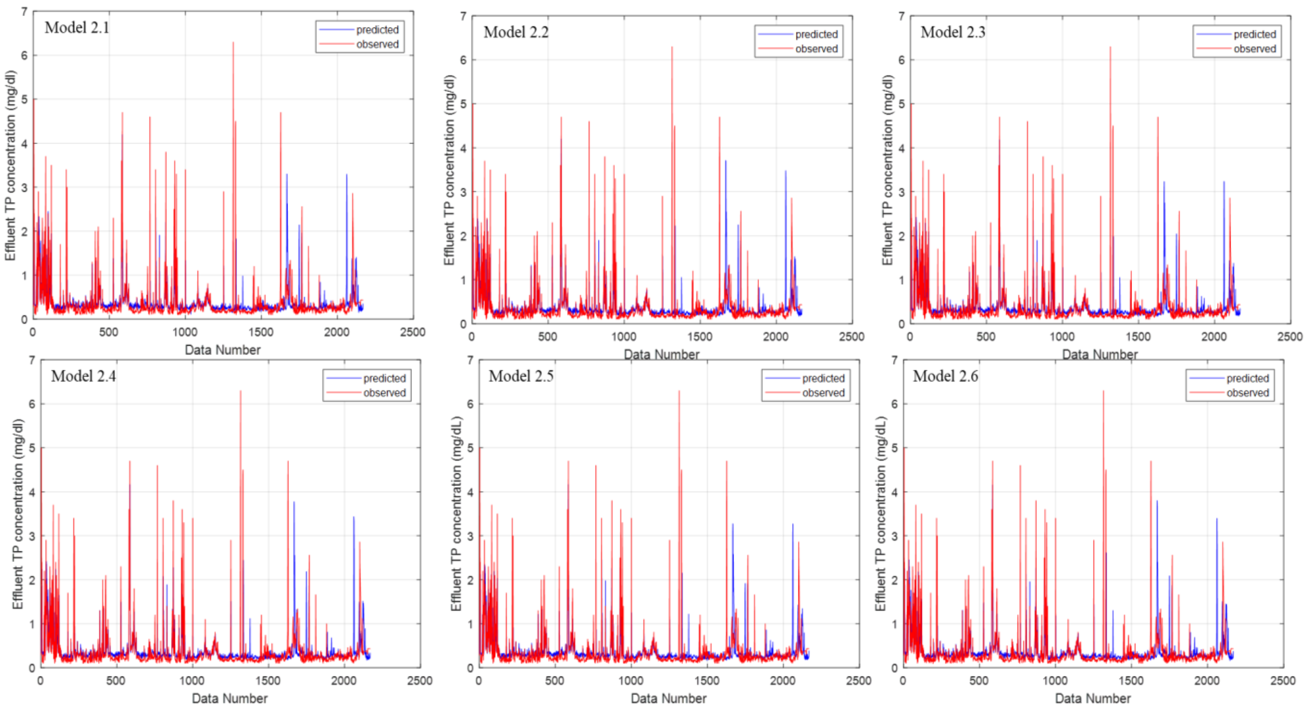

Figure 3. Effluent TP Forecasting Results

Graphical User Interface Results

The following results for total nitrogen (TN) were obtained from experiments using actual measurement values collected at the wastewater treatment plant, demonstrating the practical effectiveness of the graphical user interface (GUI)-integrated model.

Figure 4 displays model predictions based on actual measurement values obtained from WWTP data, demonstrating the model's effectiveness in forecasting TN concentrations for the following day.

A similar graphical user interface has also been implemented to predict TP concentrations, as shown in Figure 5.

As shown in Figure 5, the graphical user interface (GUI) effectively predicts TP concentrations for the following day, demonstrating its reliability for effluent TP forecasting.

CONCLUSION

This study developed artificial neural network (ANN) models with six different input combinations to predict total nitrogen (TN) and total phosphorus (TP) concentrations one day in advance at a municipal wastewater treatment plant. The highest-performing models were integrated into a user-friendly graphical interface, offering operators an accessible tool for monitoring TN and TP levels. Managing TN and TP concentrations is essential, as elevated levels can result in significant environmental issues. Machine learning techniques, such as ANN, can provide substantial benefits in full-scale wastewater treatment's complex and cost-intensive environment. The implemented ANN models offer a straightforward and efficient solution, enabling users to obtain predictions with a single click. Future phases of this project will explore more advanced deep learning algorithms to enhance model accuracy.

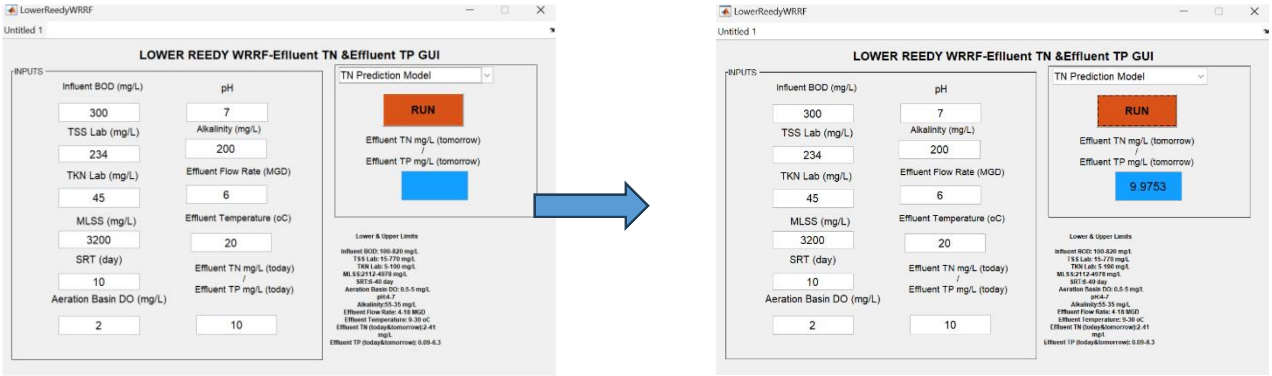

Figure 4. GUI Performance Results for Effluent TN

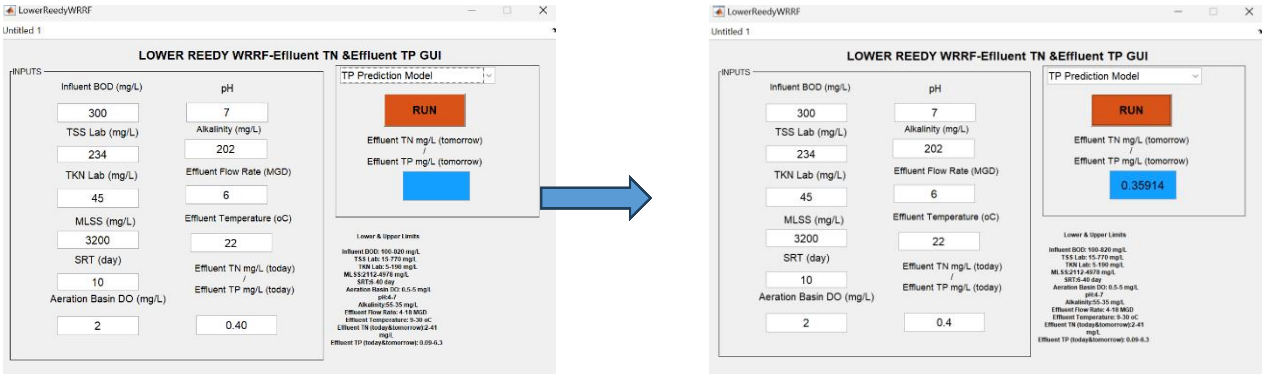

Figure 5. GUI Performance Results for Effluent TP

ACKNOWLEDGMENT

I would like to thank TÜBİTAK (The Scientific and Technological Research Council of Turkey) for financial support through the 2219 International Postdoctoral Research Fellowship Program for Turkish Citizens. I also acknowledge water utility in the USA for providing the data and express my gratitude to my supervisor, Prof. Dr. Tanju Karanfil, for his valuable insights and encouragement.

REFERENCES

Abba, S., Elkiran, G., & Nourani, V. (2021). Improving novel extreme learning machine using PCA algorithms for multi-parametric modeling of the municipal wastewater treatment plant. *Desalination and Water Treatment*, *215*, 414-426. https://doi.org/10.5004/dwt.2021.26903

Akhtar, N., Ishak, M. I. S., Bhawani, S. A., & Umar, K. (2021). Various natural and anthropogenic factors responsible for water quality degradation: A review. *Water*, *13*, 2660. https://doi.org/10.3390/w13192660

Archontoulis, S. V., & Miguez, F. E. (2015). Nonlinear regression models and applications in agricultural research. *Agronomy Journal*, *107*(3), 786–798. https://doi.org/10.2134/agronj2012.0506

El-Rawy, M., Abd-Ellah, M. K., Fathi, H., & Ahmed, A. K. A. (2021). Forecasting effluent and performance of wastewater treatment plant using different machine learning techniques. *Journal of Water Process Engineering*, *44*, 102380. https://doi.org/10.1016/j.jwpe.2021.102380

Hansen, L. D., Stokholm-Bjerregaard, M., & Durdevic, P. (2022). Modeling phosphorous dynamics in a wastewater treatment process using Bayesian optimized LSTM. *Computers and Chemical Engineering*, *160*, 107738. https://doi.org/10.1016/j.compchemeng.2022.107738

Haykin, S. (1999). Neural networks: A comprehensive foundation (2nd ed.). Prentice Hall.

Ly, Q. V., Truong, H., Ji, B., Nguyen, X. C., Cho, K. H., Ngo, H. H., & Zhang, Z. (2022). Exploring potential machine learning application based on big data for prediction of wastewater quality from different full-scale wastewater treatment plants. *Science of the Total Environment, 832*, 154930. https://doi.org/10.1016/j.scitotenv.2022.154930

Manav-Demir, N., Gelgor, H. B., Oz, E., Ilhan, F., Ulucan-Altuntaş, K., Tiwary, A., & Debik, E. (2022). Effluent parameters prediction of a biological nutrient removal (BNR) process using different machine learning methods: A case study. *Journal of Environmental Management, 351*, 119899. https://doi.org/10.1016/j.jenvman.2023.119899

Manu, D. S., & Thalla, A. K. (2017). Artificial intelligence models for predicting the performance of biological wastewater treatment plant in the removal of Kjeldahl nitrogen from wastewater. *Applied Water Science, 7*, 3783-3791. https://doi.org/10.1007/s13201-017-0526-4

Mohammadi, F., Yavari, Z., Mohammadi, F., & Rahimi, S. (2022). Prediction the performance of full scale wastewater treatment plant with A-B process using artificial neural network and genetic algorithm. *International Journal of Environmental Health Engineering, 11*(1), 1-7. https://doi.org/10.4103/ijehe.ijehe_52_20

Oliveira, P., Fernandes, B., Analide, C., & Novais, P. (2021). Forecasting energy consumption of wastewater treatment plants with a transfer learning approach for sustainable cities. *Electronics, 10*, 1149. https://doi.org/10.3390/electronics10101149

Safeer, S., Pandey, R. P., Rehman, B., Hasan, S. W., & Ullah, A. (2022). A review of artificial intelligence in water purification and wastewater treatment: Recent advancements. *Journal of Water Process Engineering, 49*, 102974. https://doi.org/10.1016/j.jwpe.2022.102974

Sravan, J. S., Matsakas, L., & Sarkar, O. (2024). Advances in biological wastewater treatment processes: Focus on low-carbon energy and resource recovery in biorefinery context. *Bioengineering, 11*(3), 1-15. https://doi.org/10.3390/bioengineering11030281

Wang, R., Yu, Y., Chen, Y., Pan, Z., Li, X., Tan, Z., & Zhang, J. (2022). Model construction and application for effluent prediction in wastewater treatment plant: Data processing method optimization and process parameters integration. *Journal of Environmental Management, 302*, 114020. https://doi.org/10.1016/j.jenvman.2021.114020

Xu, Y., Zeng, X., Bernard, S., & He, Z. (2021). Data-driven prediction of neutralizer pH and valve position towards precise control of chemical dosage in a wastewater treatment plant. *Journal of Cleaner Production, 348*, 131360. https://doi.org/10.1016/j.jclepro.2022.131360

Yu, T., & Bai, Y. (2018). A comparative study of extreme learning machine, least squares support vector machine, and back propagation neural network for outlet total phosphorous prediction. In Prognostics and System Health Management Conference (pp. 717-722).

Zhao, L., Dai, T., Qiao, Z., Sun, P., Hao, J., & Yang, Y. (2020). Application of artificial intelligence to wastewater treatment: A bibliometric analysis and systematic review of technology, economy, management, and wastewater reuse. *Process Safety and Environmental Protection, 133*, 169-182. https://doi.org/10.1016/j.psep.2019.11.014