Artificial neural network modeling approach for the prediction of five-day biological oxygen demand and wastewater treatment plant performance

Abdalrahman Alsulaili and Abdelrahman Refaie

ABSTRACT

The measurement of the wastewater BOD5 level requires five days, and the use of a prediction model to estimate BOD5 saves time and enables the adoption of an online control system. This study investigates the application of artificial neural networks (ANNs) in predicting the influent BOD5 concentration and the performance of WWTPs. The WWTP performance was defined in terms of the COD, BOD, and TSS concentrations in the effluent. Sensitivity analysis was performed to identify the best-performing ANN network structure and configuration. The results showed that the ANN model developed to predict the BOD concentration performed the best among the three outputs. The top-performing ANN models yielded $R^2$ values of 0.752, 0.612, and 0.631 for the prediction of the BOD, COD, and TSS concentrations, respectively. The optimal performing models were obtained (three inputs – one output), which indicated that the influent temperature and conductivity greatly affect the WWTP performance as inputs in all models. The developed prediction model for the influent BOD5 concentration attained a high accuracy, i.e., $R^2 = 0.754$, which implies that the model is viable as a soft sensor for online control and management systems for WWTPs. Overall, the ANN model provides a simple approach for the prediction of the complex processes of WWTPs.

Key words | artificial neural networks, modeling, sensitivity analysis, wastewater plant, wastewater treatment

HIGHLIGHTS

- ANN model provides an assessment tool for WWTP design and performance.
- Increasing the number of model inputs beyond three inputs was not beneficial.
- Influent BOD and conductivity have the highest effect on the WWTP effluent.
- COD input parameter had the highest impact on BOD5 prediction model.
- BOD5 soft-sensor development is viable using ANN model.
INTRODUCTION

Rapid urban development leads to an increase in wastewater flow rates, which requires a high pace of advancement in treatment methods. Wastewater treatment plant (WWTP) effluent quality is a key factor in environmental and health concerns (Hamed et al. 2004). Thus, it is important to be able to model and simulate WWTPs to identify optimal conditions and to avoid future operation failures. However, the characteristics and variables involved in WWTPs are numerous and exhibit a high level of complexity, which leads to difficulties in modeling through linear regression models (Hamed et al. 2004). Moreover, physical and chemical simulations through software can be performed to simulate WWTPs. Ferrer et al. (2008) applied Design and Simulation of Activated Sludge Systems (DESASS) software to model and optimize WWTP performance in the steady or transient state. Other software includes anaerobic digestion models (Batstone et al. 2002) and the Activated Sludge Model (ASM) (Henze et al. 1987, 1995, 1999). However, these software programs require a large amount of input data regarding the process at each stage of the operation.

Data-driven simulation and modeling is an alternative approach in the modeling of complex systems (Elmolla et al. 2010). Some of these methods include fuzzy logic and artificial neural networks (ANNs). Data-driven modeling relies on advancements in the computing field and artificial intelligence (AI). ANNs can be applied to predict the performance of WWTPs. Due to their high prediction accuracy, ANNs were developed to model the water treatment process (Manu & Thalla 2017; Newhart et al. 2019; Alver & Kazan 2020). ANN modeling greatly relies on the quality of historical data. Poor historical data quality could lead to poor ANN modeling performance. However, ANN modeling requires a relatively small amount of data to provide acceptable prediction results. The use of neural network models to simulate wastewater plants provides a framework for plant operation monitoring (Mjalli et al. 2007). This monitoring framework leads to the minimization of the cost of operation and the determination of the quality of environmental stability. Mjalli et al. (2007) collected data over a span of one year (measurements were conducted every five days). It was determined that the outputs of the secondary treatment effluent (STE) were the inputs of their model. The overall data set was divided at a ratio of 4:2:1 into training, validation, and testing data sets. It should be noted that Mjalli et al. (2007) employed a feed-forward neural network (FFNN), which was developed using MATLAB software. In all single-input networks, the network structure consists of three layers (input, hidden, and output layers). In multi-input networks, the network structure consists of four layers (two hidden layers). The optimal network structure was found to be 1–40–1 with the COD as the input variable ($R^2 = 0.987$ and MSE = 0.021). The training output function was based on the mean square error (MSE) between the network prediction and observed values. Nasr et al. (2012) applied an ANN to predict the performance of wastewater treatment plants in
Alexandria. The plant performance was studied in terms of the total suspended solids (TSS), biological oxygen demand (BOD), and chemical oxygen demand (COD) over a one-year period. The gathered data were classified into four groups, where each group represents three months of the year. The feedforward network was developed and trained using back propagation. The predicted values were highly correlated with the measured values ($R^2 = 0.90317$).

Guo et al. (2015) applied machine learning through a feedforward ANN and support vector machine (SVM) to predict the total nitrogen concentration in the effluent of a WWTP in Ulsan, Korea. The inputs of the models were the total nitrogen, total phosphorus, and TSS of the influent flow. Based on the coefficient of determination ($R^2$), Nash–Sutcliffe efficiency (NSE) and relative efficiency ($d_{rel}$) criteria, the results showed that both models were effective in prediction ($R^2 = 0.55$, NSE = 0.56, and $d_{rel} = 0.8$). Even though the SVM model attained a higher prediction efficiency, the ANN model consisting of three layers (input, hidden, and output layers) was more effective in correlating the input values to the T-N concentrations. Nezhad et al. (2016) employed the MATLAB ANN toolbox to apply machine learning techniques to predict the effluent quality index (EQI) for wastewater treatment plants in Tehran. A feedforward back-propagated neural network was developed that consisted of three hidden layers. The inputs of the model included BOD, TDS, TSS, FC, PO₄, NH₄ and pH. The results indicated that the optimal network structure was 8–7–1, which led to a high EQI prediction efficiency ($R = 0.96$, and MSE = 0.1).

Moreover, Abba & Elkiran (2017) predicted the COD of the effluent of wastewater treatment plants using a feedforward neural network (FFNN). The COD of the effluent is considered a major parameter to assess the performance of WWTPs. An FFNN was applied to predict the COD, BOD, pH, T-P, T-N, TSS, and conductivity of the effluent at WWTP inlets. Several structures and input combinations were considered; however, the FFNN model with an 8–8–1 network structure, utilizing all four input parameters, yielded good performance and the highest accuracy in effluent COD prediction ($R^2 = 0.7$, and RMSE = 0.0108).

The use of neural network models to simulate wastewater plants provides a framework for plant operation monitoring (Mjalli et al. 2007). This monitoring framework enables minimizing the cost of operation and determining the quality of environmental stability. Moreover, previous studies did not present the method used for determining the ANN configuration. Thus, this study provides a simple approach to identify the optimal ANN configuration for WWTP plant performance prediction.

Moreover, several regression models and artificial neural networks were developed to predict the five-day biological oxygen demand (BOD₅) (Kasem et al. 2018; Baki et al. 2019; Najafzadeh & Ghaemi 2019). The importance of BOD₅ modeling stems from the extensive laboratory procedures performed to measure the BOD₅ concentration, as these tests require approximately five days. A regression model based on the relation between the characteristics of wastewater and BOD₅ was developed and proved to attain a high accuracy ($R^2$ up to 0.7966) (Baki et al. 2019). Furthermore, Kasem et al. (2018) developed a software sensor to monitor the BOD₅ concentration in the Sefidrood River in Iran using a feedforward ANN as a function of the dissolved oxygen level. The developed ANN performance indicated was suitable and a high $R^2$ value was generated (up to 0.89). In WWTPs, deep neural networks and genetic algorithms were also used to design a BOD₅ soft sensor (Qiu et al. 2016). The developed sensor was tested under three conditions, i.e., dry, rainy, and stormy weather conditions on the BSM1 simulation platform, and the results revealed good performance under extreme weather conditions. Although there have been a few studies that developed soft sensors for BOD monitoring, there is a lack of parameter analysis and its effect on predicting the inlet BOD concentration, which will be performed in this study.

Most of the literature utilized TSS, COD, and BOD to predict the performance of WWTPs using ANN-based models. These parameters are considered the controlling parameters of the effluent quality from WWTPs (Abba & Elkiran 2017). Furthermore, WWTPs commonly consist of a series of complex processes that cannot be modeled using simple regression techniques due to the large number of input parameters and data points required to capture the plant performance. Thus, the capacity of ANN modeling makes it a reliable solution in modeling the performance of wastewater treatment plants. The objective of this paper is to investigate the application of ANNs in WWTP modeling and to provide a simple algorithm to
determine the optimal ANN configuration capturing the complex behavior of treatment plants. In contrast to most of the previous literature, this study aims to provide a simple determination method for the optimal configuration, which is defined in this study as a simple and general configuration that can be extrapolated, in future studies, to multiple treatment plants with similar processes. Additionally, the generated model could be utilized in the design of control systems and the monitoring of plant performance and water quality parameters. This study also aims to develop a viable ANN to act as a soft sensor of the influent BOD$_5$ in WWTPs, as there is a lack of this specific approach in the previous literature. The above soft sensor is expected to decrease the five-day period of the BOD$_5$ measurement to several hours, which will allow the use of online control systems.

**METHODOLOGY AND MODEL DEVELOPMENT**

**Data collection and preparation**

The data analyzed in this study were collected from the Kabd WWTP, which is located in Kuwait, over a seven-year period (2013 to 2019). A total of six wastewater parameters were adopted as model inputs, including the influent temperature, pH, conductivity, and TSS, COD, and BOD concentrations. These chosen input parameters sufficiently describe the WWTP performance (Mjalli et al. 2007; Nasr et al. 2012). These parameters were employed to predict the WWTP performance, which is represented by the effluent BOD, COD, and TSS concentrations. The influent temperature, pH, conductivity, TSS, and COD were used to predict the influent BOD$_5$ concentration. The total number of data points was 2,397 points; however, they included missing values. Thus, IBM SPSS Statistics 26 was applied to conduct missing value analysis along with descriptive analysis on the data set, as indicated in Table 1. The BOD$_5$ concentration had the most missing values among all the other parameters, which could be explained by worker negligence in performing the tedious five-day experimental procedure to estimate the BOD$_5$ value. Thus, a BOD$_5$ prediction model is a valuable approach to calculate these missing values. Listwise deletion of missing data points was conducted using a developed MATLAB code. This code located the missing values of any parameter and deleted the whole data row. After removing the data points with missing values for all of the parameters, the total number of data points was 1,032 points.

Further data refining was performed by excluding any extreme values that were not within the 1.5 interquartile ranges (IQRs) of $(Q1 - 1.5*IQR)$ and $(Q3 + 1.5*IQR)$. Moreover, data normalization was performed according to Equation (1):

$$Y_{\text{norm}} = \frac{Y - Y_{\text{min}}}{Y_{\text{max}} - Y_{\text{min}}}$$

(Table 1) Descriptive and missing value analysis results on the data set used in this study

| Descriptive statistics | N   | Minimum | Maximum | Mean   | Standard deviation | Count | Percent | Low | High |
|------------------------|-----|---------|---------|--------|--------------------|-------|---------|-----|------|
| Temperature (°C)       | 1,992| 16.6    | 36.9    | 28.8   | 4.3                | 405   | 16.9    | 1   | 0    |
| pH                    | 1,973| 6.48    | 10.12   | 7.0    | 0.2                | 424   | 17.7    | 4   | 31   |
| Conductivity (μs/cm)  | 1,991| 728     | 2,100   | 1,504.1| 194.1             | 406   | 16.9    | 7   | 1    |
| TSS (mg/L)            | 1,990| 72      | 1,124   | 173.4  | 52.6               | 407   | 17.0    | 10  | 86   |
| COD (mg/L)            | 1,989| 196     | 1,570   | 613.9  | 143.8              | 408   | 17.0    | 16  | 126  |
| BOD$_5$ (mg/L)        | 1,365| 17      | 541     | 279.6  | 88.2               | 1,052 | 43.1    | 21  | 14   |
| TSS effluent (mg/L)   | 1,989| 0       | 92      | 7.1    | 5.1                | 408   | 17.0    | 5   | 50   |
| COD eff. (mg/L)       | 1,988| 0       | 217     | 25.3   | 12.3               | 409   | 17.1    | 2   | 70   |
| BOD$_5$ eff. (mg/L)   | 1,365| 1       | 41      | 6.2    | 3.4                | 1,052 | 43.1    | 0   | 14   |

*Number of cases outside the range of $(Q1 - 1.5*IQR, Q3 + 1.5*IQR)$. 

Downloaded from http://iwaponline.com/ws/article-pdf/21/5/1861/920644/ws021051861.pdf by guest
where $Y$ represents the values of each parameter studied and $Y_{\text{min}}$ and $Y_{\text{max}}$ are the minimum and maximum values, respectively, of the variable to be normalized. Furthermore, a correlation matrix was developed to ensure the independence of the input parameters. Data preprocessing and preparation were performed using MATLAB 2017b.

**Artificial neural network**

The idea of strong computing methods, known as neural networks, as an equivalent to the human brain, was developed in the late 1800s (Lippmann 1988). Artificial neural networks consist of artificial neurons in a connected layered structure to provide the desired output. The network is trained through the continuous addition of data. Then, generalization of the network is realized through the introduction of new unseen data. The principal advantages of ANNs lie in (a) their high learning speed and data processing and (b) their capacity to represent highly nonlinear systems. However, the major drawback is the black box aspect of ANNs (Lippmann 1988).

Figure 1 shows a common FFNN structure that consists of three main network layers. The first layer is called the input layer, which receives the input data and sends these data to the hidden layers. The hidden layers are the heart of the neural network and consist of multiple neurons. In a single network, there can be multiple hidden layers with numerous neurons. Usually, the number of hidden layers is proportional to the complexity of the system to be modeled (Dreyfus 2002). The working principle of ANN is basically the same for the many network types. The basic processing element, the neuron, receives input signals then processes them through an activation function and provides an output signal. Furthermore, the weight of each neuron and the transfer functions are responsible for passing signals from one layer to the next layer. The mathematical expression of the neural network working principle is given in Equation (2):

$$Y_i = f\left(\sum_{j=1}^{M} W_{ij} X_j + b_i\right)$$

where $Y_i$ is the value of predicted output $i$, $f$ is the activation function, $W_{ij}$ is the weight assigned to each input $j$, $M$ is the total number of inputs, and $b_i$ is the bias for each output.

The most common type of activation function is the sigmoid function (Haykin 2009). Examples of the sigmoid function are the tan-sigmoid function and the log-sigmoid function, expressed by Equations (3) and (4), respectively.
(Gangi Setti & Rao 2014):

\[
f(x) = \frac{2}{1 + \exp(-2x)} - 1
\]

\[
f(x) = \frac{1}{1 + \exp(-x)}
\]

where \( x \) is the input of the activation function. In this study, the MATLAB 2017b neural network tool is applied to develop and train the feedforward neural network with the tan-sigmoid transfer function (Equation (3)) for all layers. The ANN is trained using one input matrix containing up to six inputs for each of the three outputs. Table 2 summarizes the combinations used in this study. The input data are pretreated before training for better network training performance.

At present, to determine the optimal number of hidden layers and neurons in each layer, several network structures should be tested, and the performance of each network should be evaluated. Although a small number of hidden layers and neurons could lead to a suitable generalization, it may also lead to an underfitted network. In contrast, a large number of hidden layers could lower the training error but could increase network overfitting (Geman et al. 1992). Various configurations of the network structure were tested, and the optimal structure for each output was determined through trial and error. The configuration parameters included the number of inputs, hidden layers, and neurons in the hidden layer.

**RESULTS AND DISCUSSION**

**Correlation matrix**

The developed correlation matrix, summarized in Table 3, indicates the absence of any linear relationship among the parameters. However, this does not indicate the absence of any other types of relationships, as the small correlation coefficient indicates that conventional regression methods are unsuitable to predict such a complex system. Moreover, the input parameters lack any correlation among each other, which is required to train a reliable ANN. The presence of correlated input parameters biases the ANN towards the effect of these parameters.

**ANN modeling and prediction**

The development of the ANN model was realized using various structures to obtain the ANN with the optimal performance. First, each of the influential parameters was applied as input to predict each of the effluent parameters (one input to one output). Then, the number of inputs was gradually increased until all the inputs (six inputs) were applied to predict each output. The number of hidden...
layers and the number of neurons were also scanned for the optimal network parameters. The data set was divided into training, validation, and test sets at a ratio of 70:15:15.

The key performance indicator considered in this study is the coefficient of determination ($R^2$) (Equation (5)) and the index of agreement ($d$) (Equation (6)) between the predicted values and the measured values for each ANN developed:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (OBS_i - OBS)(P_i - P)^2}{\sum_{i=1}^{N} (OBS - OBS)^2 \sum_{i=1}^{N} (P_i - P)^2}$$  \hspace{1cm} (5)$$

$$d = 1 - \frac{\sum_{i=1}^{N} (OBS_i - P_i)^2}{\sum_{i=1}^{N} (P_i - OBS)^2 + (OBS_i - OBS)^2}$$  \hspace{1cm} (6)$$

where $N$ is the total number of points, $OBS$ is the measured value of the variable, $P$ is the corresponding predicted value of the variable, and the bar indicates the mean value of the variable.

The results of developing a one input to one output ANN for the estimation of the effluent TSS, BOD and COD concentrations from the WWTP plant are summarized in Table 4. The low $R^2$ value indicates the complexity of the system, as it cannot be modeled using only one input parameter. In comparison, Mjalli et al. (2007) reported high $R^2$ values while using a one input and one output ANN structure. However, the high performance occurred due to the large number of neurons used in their model, which led to overfitting and reduced model generalization.

The best input performances in predicting the effluent BOD were obtained for the influent BOD, pH, and temperature. The top three COD predictors were the temperature, pH, and conductivity. The effluent TSS was best predicted by the temperature, influent BOD, and influent TSS. The effect of the initial pH on the BOD rate and wastewater treatment process has been examined in previous literature.
The removal efficiency of COD and TSS is known to increase with the temperature (Ahsan et al. 2005). The next step was to train three input and one output ANN models. It should be noted that every possible combination of three input parameters was considered with different numbers of hidden layers.

Effluent BOD prediction

Figure 2 shows the $R^2$ values for predicting the effluent BOD for each input combination using the various numbers of hidden layers. It is found that the input-20 combination (temperature–conductivity–inf. BOD) outperforms the other input combinations. The results also show that for most of the input combinations, the prediction accuracy peaks when using three hidden layers. This explains the underfitting that occurs with a small number of hidden layers and the overfitting that occurs when using more than three hidden layers.

The input-20 combination was trained for various network structures, and the optimal ANN structure ($R^2 = 0.752$ and $d = 0.928$) contained three hidden layers with 17 neurons for each layer (1–17–17–17–1), as shown in Figure 3. Figures 4 and 5 show the performance of the three-input ANN model as a parity plot between the measured and predicted values and as a comparison of the prediction and measured values along the data sequence, respectively. Although the $R^2$ value is deemed inadequate, the modeled values match the measured values closely except for some extent values. These extent values might be caused by the noisy data set.

The resulting model attains a lower performance than that of Mjalli et al. (2007); however, they adopted a much smaller data set (80 data points) than the data set used in this study (1,365 data points). In addition, Mjalli et al. employed the secondary treatment effluent as the input of their model, which is strongly correlated to the effluent of the plant. Thus, their neural network only predicts the tertiary treatment process, while in our study, the artificial...
neural network is designed to predict the whole treatment process.

**Effluent COD prediction**

The accuracy results for each input combination are shown in Figure 6 using an ANN with different numbers of hidden layers. The input-26 combination (temperature–conductivity–pH) outperformed the other input combinations. It was also observed that the combinations containing temperature + conductivity, or temperature + pH (inputs 20 to 26) resulted in a higher prediction performance than that of the other combinations. The results also revealed performance spikes for the other combinations; however, these were not considered due to the lack of consistency.

The temperature–conductivity–pH combination was trained for various network structures, and the optimal ANN structure ($R^2 = 0.6115$ and $d = 0.877$) contained three hidden layers with 13 neurons for each layer (1–13–13–13), as shown in Figure 7. Figures 8 and 9 show the performance of the three-input ANN model as a parity
plot between the measured and predicted values and as a comparison of the prediction and measured values along the data sequence, respectively. Figure 9 indicates that the prediction accuracy is low at the extent values. Except for these extent values, the prediction values are nearly identical to the measured values.

Additionally, Mjalli et al. (2007) generated a better-performing ANN model with many fewer data points and
different input parameters. Moreover, the performance of the three-input ANN \( R^2 = 0.6115 \) is comparable to that of the ANN model developed with six inputs \( R^2 = 0.7034 \) (Abba & Elkiran 2017). This shows the reliability of our procedure of determining a simple, accurate ANN prediction model.

### Effluent TSS prediction

For the TSS ANN model, the input-20 combination (temperature–conductivity–BOD) yielded better results than the other combinations, as shown in Figure 10. Moreover, the input combinations containing the conductivity attained a higher prediction accuracy than the other combinations. In contrast to BOD and COD, optimum TSS modeling occurred with four hidden layers rather than with three hidden layers. This indicates the amount of complexity involved in TSS modeling over BOD and COD modeling.

The temperature–conductivity–BOD combination was trained for various network structures, and the optimal ANN structure \( R^2 = 0.6308 \) and \( d = 0.884 \) consisted of four hidden layers with 11 neurons for each layer.

![Figure 9](image9.png)

**Figure 9** | COD prediction values against measured values for all data sequences.

![Figure 10](image10.png)

**Figure 10** | TSS model prediction accuracy for all three-input combinations using various numbers of hidden layers.
(1–11–11–11–1), as shown in Figure 11. Figures 12 and 13 show the performance of the three-input ANN model as a parity plot between the measured and predicted values and as a comparison of the prediction and measured values along the data sequence, respectively. TSS was found to be the most challenging parameter to model. However, the developed ANN managed to predict TSS with a sufficient accuracy in regard to the measured values. As with

**Figure 11** | Optimal ANN structure for TSS prediction.

**Figure 12** | Parity plot for the TSS prediction model.

**Figure 13** | TSS prediction values against measured values for all data sequences.
Figure 14 | Residual plots of (a) the effluent BOD, (b) effluent COD, and (c) effluent TSS ANN models.
the BOD and COD predictions, with a smaller data set and a less generalized model, Mjalli et al. (2007) developed an ANN model with a high performance.

Furthermore, six inputs of ANN were also trained for BOD, COD and TSS; however, they yielded a decreased performance ($R^2 = 0.649$, 1–15–15–15–1; $R^2 = 0.47$, 1–8–8–8–1; and $R^2 = 0.6974$, 1–11–11–11–1; respectively). It is evident here that increasing the number of inputs is not always a benefit (Zare et al. 2011; Hamada et al. 2018). For all outputs in this study, using six inputs to develop the ANN failed to provide a higher performance (Wei 2015). However, the $R^2$ values are comparable to those reported by Abba & Elkiran (2007) when predicting the effluent COD concentration.

To study the cause of the low value of the performance indicator for the models developed in this paper, residual plots were generated, which depict the goodness of fit to the created ANN models (Guo et al. 2015). Figure 14 shows the residual plots for each model. The low relationship between the residual and predicted data indicates that the low model performance is mainly due to data noise. Hence, the ANN model would yield a higher prediction accuracy using smoother data. Moreover, ANN models provide sufficient accuracy for the prediction of the effluent concentration.

**Influent BOD$_5$ prediction**

Several ANN configurations (i.e., various input combinations, layers, and neurons) were developed and tested for the prediction of the BOD$_5$ concentration in the influent wastewater treatment plant. Figure 15 shows the best prediction results of the test data set for each ANN configuration with respect to the coefficient of determination ($R^2$). The difference between underfitting and overfitting is clearly demonstrated in Figure 15, where the $R^2$ values are lower using either a small (one layer) or large (five layers) number of layers, and the maximum $R^2$ value occurs between these extremes.

Moreover, sensitivity analysis of the input parameters was conducted utilizing the ANN prediction accuracy under the various input combinations. Table 5 summarizes the best results obtained under each number of inputs and corresponding ANN configuration. The table indicates that the COD concentration imposes the highest impact on the ANN prediction accuracy, followed by the temperature

| Table 5 | ANN configurations with the highest influent BOD$_5$ prediction accuracy for the different numbers of input parameters |
|---------|------------------------------------------------------------------------------------------------------------------|
|         | **Input parameters** | **ANN configuration** | **$R^2$** |
| ANN–V   | Temperature–pH–Conductivity–TSS–COD | (1–9–9–9–9–1) | 0.754 |
| ANN–IV  | Temperature–Conductivity–TSS–COD | (1–20–20–1) | 0.737 |
| ANN–III | Temperature–Conductivity–COD | (1–13–13–13–1) | 0.702 |
| ANN–II  | Temperature–COD | (1–10–10–10–1) | 0.464 |
| ANN–I   | COD | (1–19–1) | 0.176 |
and conductivity. From the results, the order of significance of the input parameters is as follows:

COD > temperature > conductivity > TSS > pH

Furthermore, the introduction of the conductivity parameter greatly increased the prediction accuracy. Thus, for an ANN to sufficiently predict the influent BOD$_5$, at least the influent COD, temperature, and conductivity are required as input parameters. Thus, decreasing the time to estimate BOD$_5$ from five days to several hours is the time required to measure the COD concentration of any sample. From the different configurations listed in Table 4, ANN-V was adopted to predict BOD$_5$, and the results are shown in Figure 16. The results demonstrate the high accuracy of the ANN in predicting BOD$_5$ with the exception of a few data points. The spikes deviating from the measurement values occur due to noise in the training data set similar to the previously developed ANN.

The $R^2$ values from the literature and previous models designed to predict BOD$_5$ in WWTPs are listed in Table 6. The previous regression models, in the literature, attain a slightly higher performance than the ANN developed in this study. However, regression models tend to require a large amount of data and rely on a large number of input parameters. For instance, Tables S1 to S6 (Supplementary Material) list the results of linear regression for the data used in this study and using the input parameters of the optimal ANN developed. The results reveal low $R^2$ values for the linear regression model. Moreover, the soft sensor of Qiu et al. (2016) yields a comparable or better performance in predicting BOD$_5$ than the model in this study. However, the model developed in this study is notably simpler, as it does not require the use of a genetic algorithm.

**CONCLUSION**

The developed ANN models suitably predicted the WWTP performance, which was defined based on the effluent TSS, BOD, and COD concentrations in the treatment plant, with a high degree of reliability. A low model performance mainly occurred due to noise in the data used to develop the ANN.

The results demonstrated that preliminary data analysis and preparation are essential for ANN training. It is recommended to gradually increase the complexity (number of inputs, hidden layers, and neurons) of the network configuration until no further improvement is recorded. In this study, a three input and one output configuration was sufficient, and a further increase in inputs resulted in overfitting of the system. The results also indicated that increasing the number of inputs was not always beneficial. Another finding from the developed models is the importance of the input parameters in the wastewater treatment process. The parameter significance could be interpreted from the model input combination that yielded the highest prediction accuracy. For instance, the influent temperature and conductivity greatly affected the WWTP performance as they were used as inputs in all models. However, the influent BOD concentration was important in the treatment process regarding the effluent
BOD and TSS concentrations, while the effluent COD concentration was indirectly dependent on the influent pH.

In this study, the developed prediction model for the influent BOD₃ concentration worked well and attained a high performance accuracy, namely, \( R^2 = 0.754 \). This result demonstrates the viability of using this model as a soft sensor for online control and management systems for WWTPs. A further recommendation is to implement a more robust control system entailing frequent measurements to further tune the model with data highly representative of real-time operation conditions.

**FUNDING**

This research did not receive a specific grant from any funding agencies in the public, commercial, or not-for-profit sectors.

**DATA AVAILABILITY STATEMENT**

Data cannot be made publicly available; readers should contact the corresponding author for details.

**REFERENCES**

Abba, S. I. & Elkiran, G. 2017 Effluent prediction of chemical oxygen demand from the wastewater treatment plant using artificial neural network application. *Procedia Computer Science* **120**, 156–163. https://doi.org/10.1016/j.procs.2017.11.223.

Ahsan, S., Rahman, M. A., Kaneco, S., Katsumata, H., Suzuki, T. & Ohta, K. 2005 Effect of temperature on wastewater treatment with natural and waste materials. *Clean Technologies and Environmental Policy* **7** (3), 198–202. https://doi.org/10.1007/s10098-005-0271-5.

Alver, A. & Kazan, Z. 2020 Prediction of full-scale filtration plant performance using artificial neural networks based on principal component analysis. *Separation and Purification Technology* **230**, 115868. https://doi.org/10.1016/j.seppur.2019.115868.

Baki, O. T., Aras, E., Akdemir, U. O. & Yılmaz, B. 2019 Biochemical oxygen demand prediction in wastewater treatment plant by using different regression analysis models. *Desalination and Water Treatment* **157**, 79–89.

Batstone, D. J., Keller, J., Angelidaki, I., Kalyuzhnyi, S. V., Pavlostathis, S. G., Rozzi, A., Sanders, W. T. M., Siegrist, H. & Vavilin, V. A. 2002 The IWA Anaerobic Digestion Model No 1 (ADM1). *Water Science and Technology* **45** (10), 65–73. https://doi.org/10.2166/wst.2002.0292.

Dreyfus, T. 2002 *Advanced mathematical thinking processes.* In: *Advanced Mathematical Thinking.* (D. Tall, ed.), Springer, Dordrecht, The Netherlands, pp. 25–41.

Ebrahimi, M., Gerber, E. L. & Rockaway, T. D. 2017 Temporal performance assessment of wastewater treatment plants by using multivariate statistical analysis. *Journal of Environmental Management* **193**, 234–246. https://doi.org/10.1016/j.jenvman.2017.02.027.

Elmolla, E. S., Chaudhuri, M. & Eltoukhy, M. M. 2010 The use of artificial neural network (ANN) for modeling of COD removal from antibiotic aqueous solution by the Fenton process. *Journal of Hazardous Materials* **179** (1–3), 127–134. https://doi.org/10.1016/j.jhazmat.2010.02.068.

Ferrer, J., Seco, A., Serralta, J., Ribes, J., Manga, J., Asensi, E., Morenilla, J. J. & Llavador, F. 2008 DESASS: A software tool for designing, simulating and optimising WWTPs. *Environmental Modelling & Software* **23** (1), 19–26. https://doi.org/10.1016/j.envsoft.2007.04.005.

Gangi Setti, S. & Rao, R. N. 2014 Artificial neural network approach for prediction of stress–strain curve of near β titanium alloy. *Rare Metals* **33** (3), 249–257. https://doi.org/10.1007/s12598-013-0182-2.

Geman, S., Bienenstock, E. & Doursat, R. 1992 Neural networks and the bias/variance dilemma. *Neural Computation* **4** (1), 1–58.

Guo, H., Jeong, K., Lim, J., Jo, J., Kim, Y. M., Park, J., Kim, J. H. & Cho, K. H. 2015 Prediction of effluent concentration in a wastewater treatment plant using machine learning models. *Journal of Environmental Sciences* **32**, 90–101. https://doi.org/10.1016/j.jes.2015.01.007.

Hamada, M., Adel Zaqoot, H. & Abu Jreiban, A. 2018 Application of artificial neural networks for the prediction of Gaza wastewater treatment plant performance – Gaza strip. *Journal of Applied Research in Water and Wastewater* **5** (1), 399–406. https://doi.org/10.22126/arww.2018.8874.

Hamed, M. M., Khalafallah, M. G. & Hassanien, E. A. 2004 Prediction of wastewater treatment plant performance using artificial neural networks. *Environmental Modelling & Software* **19** (10), 919–928. https://doi.org/10.1016/j.envsoft.2003.10.005.
Haykin, S. S. 2009 Neural Networks and Learning Machines, 3rd edn. Prentice Hall, Upper Saddle River, NJ, USA.
Henze, M., Grady, C. P. L. Jr, Gujer, W., Marais, G. v. R. & Matsuo, T. 1987 Activated Sludge Model No 1, IAWPRC Scientific and Technical Report No 1. IAWPRC, London, UK.
Henze, M., Gujer, W., Mino, T., Matsuo, T., Wentzel, M. C. & Marais, G. v. R. 1995 Activated Sludge Model No 2, IAWQ Scientific and Technical Report. IAWQ, London, UK.
Henze, M., Gujer, W., Mino, T., Matsuo, T., Wentzel, M. C., Marais, G. v. R. & Van Loosdrecht, M. C. M. 1999 Activated Sludge Model No 2d, ASM2D. Water Science & Technology 39 (1), 165–182.
Kasem, R., ALabdeh, D., Noori, R. & Karbassi, A. 2018 A software sensor for in-situ monitoring of the 5-day biochemical oxygen demand. Rudarsko-geološko-naftni Zbornik 33 (1), 15–23.
Lippmann, R. P. 1988 An introduction to computing with neural nets. In: Artificial Neural Networks: Theoretical Concepts (V. Vemuri, ed.), IEEE Computer Society Press, Washington, DC, USA, pp. 36–54.
Liu, Y., Chen, Y. & Zhou, Q. 2007 Effect of initial pH control on enhanced biological phosphorus removal from wastewater containing acetic and propionic acids. Chemosphere 66 (1), 123–129. https://doi.org/10.1016/j.chemosphere.2006.05.004.
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 (7), 3783–3791. https://doi.org/10.1007/s13201-017-0526-4.
Mjalli, F. S., Al-Asheh, S. & Alifadala, H. E. 2007 Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance. Journal of Environmental Management 83 (3), 329–338. https://doi.org/10.1016/j.jenvman.2006.03.004.
Mukherjee, S. K., Chatterji, A. K. & Saraswat, I. P. 1968 Effect of pH on the rate of BOD of wastewater. Journal (Water Pollution Control Federation) 40 (11), 1934–1939. JSTOR.
Najafažadeh, M. & Ghaemi, A. 2019 Prediction of the five-day biochemical oxygen demand and chemical oxygen demand in natural streams using machine learning methods. Environmental Monitoring and Assessment 191 (6), 380. https://doi.org/10.1007/s10661-019-7446-8.
Nasr, M. S., Moustafa, M. A. E., Seif, H. A. E. & El Kobra, G. 2012 Application of artificial neural network (ANN) for the prediction of EL-AGAMY wastewater treatment plant performance – Egypt. Alexandria Engineering Journal 51 (1), 37–43. https://doi.org/10.1016/j.aej.2012.07.005.
Newhart, K. B., Holloway, R. W., Hering, A. S. & Cath, T. Y. 2019 Data-driven performance analyses of wastewater treatment plants: a review. Water Research 157, 498–513. https://doi.org/10.1016/j.watres.2019.03.030.
Nezhad, M. F., Mehrdadi, N., Torabian, A. & Behboudian, S. 2016 Artificial neural network modeling of the effluent quality index for municipal wastewater treatment plants using quality variables: south of Tehran waste water treatment plant. Journal of Water Supply: Research and Technology – Aqua 65 (1), 18–27. https://doi.org/10.2166/aqua.2015.030.
Qiu, Y., Liu, Y. & Huang, D. 2016 Data-driven soft-sensor design for biological wastewater treatment using deep neural networks and genetic algorithms. Journal of Chemical Engineering of Japan 49 (10), 925–936. https://doi.org/10.1252/jcej.16we016.
Wei, X. 2013 Modeling and Optimization of Wastewater Treatment Process with a Data-Driven Approach. PhD thesis, University of Iowa, Iowa City, IA, USA. https://doi.org/10.17077/etd.wwzj01nf.
Zare, A. H., Bayat, V. M. & Daneshkare, A. P. 2011 Forecasting nitrate concentration in groundwater using artificial neural network and linear regression models. International Agrophysics 25 (2), 187–192.