Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2021.26903

215 (2021) 414–426
March

Improving novel extreme learning machine using PCA algorithms for multi-parametric modeling of the municipal wastewater treatment plant

S.I. Abba[a], Gozen Elkiran[b,*], Vahid Nourani[c]

[a]*Department of Civil Engineering Baze University, Abuja, Nigeria, email: saniisaabba86@gmail.com*
[b]*Department of Civil Engineering, Faculty of Engineering, Near East University, Near East Boulevard 99138, North Cyprus, Via Mersin 10, Turkey, email: gozen.elkiran@neu.edu.tr*
[c]*Department of Water Resources Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran, email: vnourani@yahoo.com*

Received 21 March 2020; Accepted 7 December 2020

ABSTRACT

In order to develop a tool for modeling the efficiency of municipal wastewater treatment plants (MWWTP), a reliable prediction tool is essential. In this research, two scenarios (I and II) were investigated for modeling the performance of Nicosia MWWTP. The extreme learning machine (ELM), which is a newly developed black-box model, combined with principal component analysis was developed in scenario I and two principal components (PCs) variables were generated, while in scenario II, traditional multi-layer perceptron (MLP) neural network and multiple linear regression (MLR) models were established for comparison. The daily measured data obtained from new Nicosia MWWTP includes (pH_{inf}, $Conductivity_{inf}$, BOD_{inf}, COD_{inf}, $Total\text{-}N_{inf}$, $Total\text{-}P_{inf}$, $NH_4\text{–}N_{inf}$, SS_{inf} and TSS_{inf}) as the input variables and (BOD_{eff}, COD_{eff}, $Total\text{-}N_{eff}$, $Total\text{-}P_{eff}$) as the corresponding outputs. Taylor diagrams, box and whisker were also utilized to examine the similarities and comparisons between the observed and predicted values for both the ELM and PCs-ELM in scenario I. The obtained results based on the performance indices showed that the PCs-ELM model has higher performance accuracy than the novel ELM model. The results also showed increases of the PCs-ELM of about 12%, 2%, 20% and 6% for BOD_{eff}, COD_{eff}, TN_{eff} (total nitrogen) and TP_{eff} (total phosphorite) with regard to the ELM model. Also, the comparison results demonstrated that ELM and MLP revealed higher prediction accuracy than the MLR model, and the ELM model comparably outperformed the MLP model. The overall results indicated that both the PCs-ELM and two scenarios could be alternative reliable tools for modeling the performance of Nicosia MWWTP.

Keywords: Wastewater treatment plant; Extreme learning machine; Principal component analysis; Multi-layer perceptron neural network; Nicosia-Cyprus

1. Introduction

Water is essential for sustaining life; therefore, an affordable and adequate supply of water must be available [1]. Municipal wastewater treatment plants (MWWTP) are processes that remove contaminants from untreated domestic wastewater with the goal of safeguarding public health and the natural environment [2]. A satisfactory treatment plant is vital in order to avoid the discharge of high pollutants and to meet the required standards by law. A combination of parameters from physical, chemical and biological characteristics are often the major factors affecting the operation and control of MWWTP [3]. Due to various compositions and characteristics of wastewater treatment plant (WWTP) variables, the performance can be assessed by considering certain sensitive variables such as total nitrogen (TN), biological oxygen demand

* Corresponding author.

Presented at the 2nd International Conference on the Environment Survival and Sustainability (ESS 2019), 7–11 October 2019, Nicosia, North Cyprus

(BOD), total suspended solids (TSS) and chemical oxygen demand (COD). The available literature and published studies for predicting the MWWTP have used these parameters [3]. The quality of untreated and treated sewage has a significant effect on the operation and efficiency of any MWWTP. However, MWWTPs comprise a large number of parameters and operations which are complex in terms of measurement and evaluation [4]. Hence, modeling this system is considered difficult due to the nature of the process and most of the available traditional models are based on assumptions, estimations and require too much time and money; as such, a reliable and appropriate tool is indispensable in predicting the performance of MWWTPs [5,6].

On the other hand, artificial intelligence (AI) models play a vital role and have created significant variations for forecasting several environmental and hydrological phenomena [7–13]. Meanwhile, recent studies have shown that black-box models like support vector machine (SVM), artificial neural networks (ANN) and adaptive neuro-fuzzy inference system (ANFIS) could be suitable alternatives for the performance analysis of WWTPs. For example, Maleki et al. [14] predicted the influent parameters in a WWTP using an auto-regressive integrated moving average (ARIMA) and neural network auto-regression (NNAR) models. Despite the acceptable performance of the ARIMA model, the results showed better prediction performance for NNAR with regard to ARIMA. Chen et al. [15] developed a combination of ANN, genetic algorithm and fuzzy logic as a new method for modeling industrial WWTPs in Taiwan. The proposed new method served as a control strategy for the successful management of the WWTP. Granata et al. [16] proposed two different machine learning models (SVM and decision tree) for WWTPs using different input combinations to predict the treated effluents (COD, BOD, TSS and total dissolved solids (TDS)). The obtained results indicated that both SVM and decision tree produced predictive skills; however, comparatively, SVM was slightly better than the decision tree. Verma et al. [17] demonstrated the ability of five different data mining approaches including multi-layer perceptron (MLP), K-nearest neighbor, SVM, random forest and multi-variate adaptive regression spline to estimate the TSS in a WWTP using different input parameters. The obtained results depicted that MLP outperformed all the other models.

In addition, [18–20] conducted studies (i.e., ANFIS, ANN, SVM) and demonstrated the superiority of the AI models over the classical models in terms of performance accuracy. Due to the problems of overfitting, local minima, and slow learning speed by some of the AI models such as ANN, a novel, new and emerging algorithm known as the extreme learning machine (ELM) model was proposed by Huang et al. [21] to overcome the disadvantages of the traditional feed-forward backpropagation. On the other hand, principal component analysis (PCA) has been applied successfully in the analysis of environmental engineering problems [11,22–24].

More recent state-of-art studies were carried in the field of WWTP for instance, Yaqub et al. [25], adopted long short-term memory (LSTM) for modeling and removal of ammonium (NH_4–N), TN, and total phosphorus (TP) in an anaerobic membrane bioreactor using various influent parameters. Based on the evaluation criteria the proposed deep learning state of art model (LSTM) displayed promising ability with regards to the removal and the prediction of NH_4–N, TN, and TP. Kang et al. [26] proposed an ANN model for the simulation of odor concentration in WWTP using different input variables (BOD, pH, dissolved oxygen). The results indicated that odor concentration can be successfully predicted using the most utilized AI model (ANN). Ansari et al. [27] employed a neuro-fuzzy logic model coupled with GA and optimization of particle swam algorithms (PSO), the study served as the multi-parametric modeling including BOD, COD, NH_3–N, pH, oil and grease, and SS. The outcomes indicated that both GA-FIS and POS-FIS outperformed the ANFIS model in terms of error estimate. Anter et al. [28] proposed a new algorithm based on an updated version of whale optimization algorithms integrated with feature input selection called chaos theory and fuzzy logic (CF-BWOA). Similarly, Patel et al. [29] obtained more than 80% accuracy for the prediction of effluent TSS present in the sedimentation tank of clariflocculator. The model proved the capability of detecting sensor faults in WWTP with promising accuracy. Sharafati et al. [30] proposed a predictive approach using a novel ensemble learning model namely; AdaBoost regression (ABR) gradient boost regression (GBR) and random forest regression for the estimation of TDS, BOD, and COD in Qom industrial wastewater treatment plant Iran. Based on the performance matrix the obtained results indicated the prediction skill of the ABR model for TDS, GBR model for both COD and BOD.

However, no studies on Nicosia MWWTP that include the novel ELM integrated with PCA algorithms have been conducted to the best of the author's knowledge Therefore, this study employed two different scenarios: (i) scenario I was aimed at developing the potential of ELM with PCA to predict the performance of new Nicosia MWWTP based on the multi-parametric effluent modeling of BOD_{eff}, COD_{eff}, total nitrogen (TN_{eff}) and total phosphorite (TP_{eff}). The advantage of introducing PCA is for choosing the proper inputs of the models and to understand whether it is feasible to enhance the prediction accuracy of the ELM model. (ii) In scenario II, the traditional MLP neural network and multiple linear regression (MLR) models were established for comparison using the same input combinations as the scenario I.

2. Materials and methods

2.1. Plant description

Nicosia's new MWWTP will benefit the environment as well as human health by protecting drinking and bathing waters from contamination. In addition, it opens up the potential to allow wastewater to be reused in irrigation. The recently established technology known as membrane bioreactor is being used and approximately 270,000 residents from both communities will be served. For sustainable development and recycling purposes, more than 300 tons/y will be generated. A total of about 10 million m^3 of quality effluent can be reused for different agricultural purposes [4,31]. In new Nicosia MWWTP, the line of treatment comprises 11 stages from the raw to treated sewage.

Firstly, the sewage is separated into liquid and solid waste and passes through the first chamber called the

screening chamber (1), in which the solids larger than 6 mm are removed. The inflow then flows down slowly so that the heavy solids (grit, sand) can fall to the bottom and oil and grease float to the surface in the grit and grease chamber (2). The pump station (3) pumps up the water to the next unit called the fine sieve (4), which removes solids larger than 2 mm. The next step is the biological treatment of waste, which is the stage that creates the condition to encourage bacteria to consume the waste comprising three units (5, 6 and 7). Stage (8) involves the separation and treatment of the by-products of the whole process into clean water, fertilizer and biogas in a process called membrane treatment. After that, the water is disinfected in the chlorine contact tank (10). Tank (11) treats the sewage effluent before it is discharged into the river. The concentration of the effluent BOD_{eff}, COD_{eff}, TN_{eff} and TP_{eff} all in (mg/L) at the plant exit is shown in Fig. 1. It can be seen from the figure that the time series and box and whiskers plots indicate the profiles, range and extent of the outliers in each parameter. Some of the measured daily data were outliers for all the parameters. This explains the inconsistency between the numbers of samples even though the distribution of outliers is relatively close to normality.

2.2. Model development and data used

For the development of the model for the current study, Fig. 2 shows the flowchart of a used model. From the flowchart, it can be observed that the input and output data (392 instances) are collected, pre-processed and normalized based on Eq. (1). For the purpose of this study, two different scenarios were employed for modeling the performance of Nicosia MWWTP as mentioned above. In scenario I, data is directly imposed on the ELM model and modeling is performed using all the input variables. If the attained error is acceptable based on performance criteria, then the best models are selected; however, if they are not acceptable, then the modeling is repeated by adjusting the model parameters. In the second stage of the flowchart, the PCA algorithm was employed for the appropriate selection of the input variable and to improve the ELM model by using the new principal component (PCs) variables as the new input variables of the ELM models (Eq. (2). Finally, the same procedure is repeated for the selected PCs as that of the ELM models. For scenario II, the two most commonly used linear and non-linear models (i.e., MLR and MLP) were also introduced for comparison with the novel ELM using the same input combination and PCs variable.

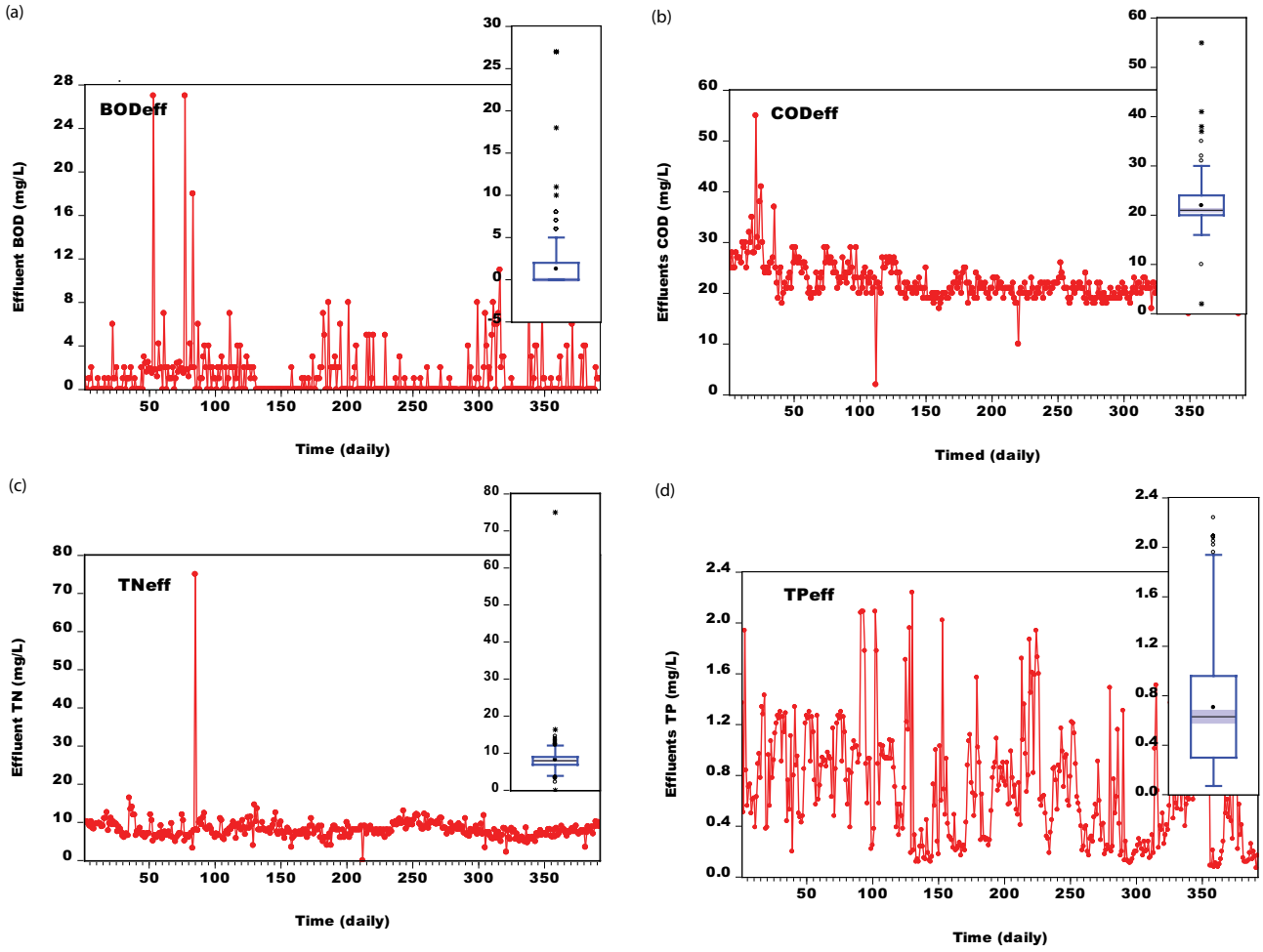

Fig. 1. Effluent (a) BOD_{eff}, (b) COD_{eff}, (c) TN_{eff} and (d) TP_{eff} concentration at the plant exit, including a box plot of the data.

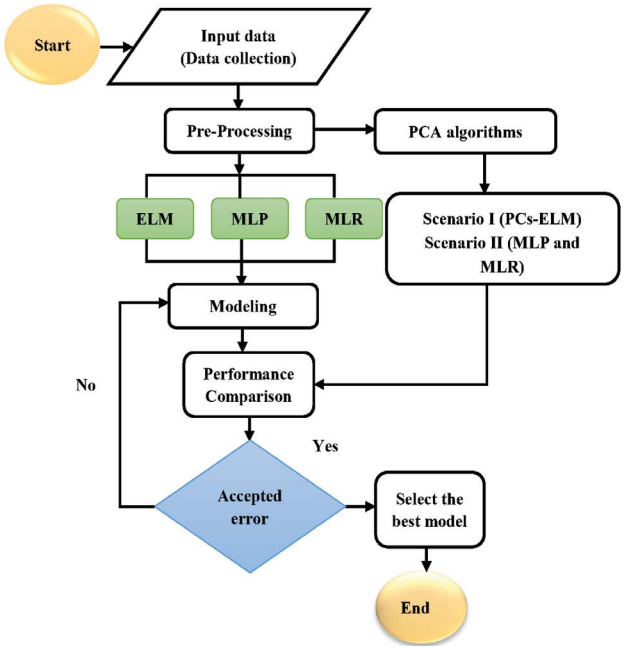

Fig. 2. Flow chart of the proposed models employed in this paper.

$$y = 0.05 + \left(0.95 \times \left(\frac{x - x_{\min}}{x_{\max} - x_{\min}} \right) \right) \tag{1}$$

where y is the normalized data, x is the measured data, $x_{\max}$ and $x_{\min}$ are the maximum and minimum value of the measured data, respectively.

$$\left.\begin{aligned} &4\text{PCs} - \text{ELM} = \text{pH}_{\text{inf}}, \text{BOD}_{\text{inf}}, \text{COD}_{\text{inf}}, \text{TN}_{\text{inf}} \\ &6\text{PCs} - \text{ELM} = \text{pH}_{\text{inf}}, \text{BOD}_{\text{inf}}, \text{COD}_{\text{inf}}, \text{TN}_{\text{inf}}, \text{TP}_{\text{inf}}, \text{TSS}_{\text{inf}} \\ &\text{All}\left(9 \text{ inputs}\right) = \text{pH}_{\text{inf}}, \text{BOD}_{\text{inf}}, \text{COD}_{\text{inf}}, \text{TN}_{\text{inf}}, \text{TP}_{\text{inf}}, \\ &\text{TSS}_{\text{inf}}, \text{SS}_{\text{inf}}, \text{NH}_4\text{N}_{\text{inf}}, \text{Cond}_{\text{inf}} \end{aligned}\right\} \tag{2}$$

The daily measured data obtained from new Nicosia MWWTP includes (pH_{inf}, $\text{Conductivity}_{\text{inf}}$, BOD_{inf}, COD_{inf}, Total-N_{inf}, Total-P_{inf}, $\text{NH}_4\text{–N}_{\text{inf}}$, SS_{inf} and TSS_{inf}) as the input variables and (BOD_{eff}, COD_{eff}, Total-N_{eff}, Total-P_{eff}) as the corresponding outputs. The normalized data were separated into 75% and 25% for calibration and verification, respectively.

2.3. Extreme learning machine

As a newly emerging black-box data-driven algorithm, the ELM was first proposed by [21] and comprises a single hidden layer feed-forward network (SLFN). The ELM is different from the traditional feed-forward neural network as it overcomes the problems of slow learning speed, local minima and overfitting [32–36]. It is notable that the potential of the ELM could be attributed to its generalization ability and fast learning speed. Due to it is promising performance ability, ELM has been applied in various fields of hydro-environmental studies [37]. For more details about ELM, models refer to [36,38–40]. The structure of the ELM network used in this study is presented in Fig. 3.

In this study, an ELM model was developed using calibration and validation data sets as mentioned above. For the set of N training samples (i.e., t = 1,2,…,N) in which $x_t \in \mathbb{R}^d$ and $y_t \in \mathbb{R}$, a SLFN with H hidden nodes is mathematically expressed as [21]:

$$\sum_{i=1}^{H} B_i g_i\left(\alpha_i \cdot x_t + \beta_i\right) = z_t \tag{3}$$

where $B \in \mathbb{R}^H$, $Z\left(z_t \in \mathbb{R}\right)$ and $G(\alpha,\beta,x)$ represents the predicted weights in the output layer, model output and activation function of the hidden layer, respectively. Additionally, α_i, β_i, i and d indicate the weights of the randomized layers, biases of these randomized layers, the index of the specific node in the hidden layer and the number of inputs, respectively.

As mentioned above, the study the employed activation function as:

$$G\left(x\right) = \frac{1}{1 + \exp\left(-x\right)} \tag{4}$$

In an ELM model, a suitable number of hidden neurons, randomized input layer weights (α), and randomized hidden layer biases (β) can lead to a zero error, which therefore produced the weights of the output layer that can be obtained analytically for any training [21]:

$$\sum_{t=1}^{N} \left\| z_t - y_t \right\| = 0 \tag{5}$$

The system of linear equations can be used to obtain the value of B for any input-output training samples:

$$Y = GB \tag{6}$$

in which

$$G\left(\alpha,\beta,x\right) = \begin{bmatrix} g\left(x_1\right) \\ \vdots \\ g\left(x_N\right) \end{bmatrix} = \begin{bmatrix} g_1\left(\alpha_1 \cdot x_1 + \beta_1\right) & \cdots & g_L\left(w_H \cdot x_1 + \beta_H\right) \\ \vdots & \cdots & \vdots \\ g_1\left(\alpha_N \cdot x_N + \beta_1\right) & \cdots & g_L\left(w_H \cdot x_N + \beta_F\right) \end{bmatrix}_{N \times H} \tag{7}$$

and

$$B = \begin{bmatrix} B_1^T \\ \vdots \\ B_H^T \end{bmatrix}_{H \times 1} \tag{8}$$

and

$$Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times 1} \tag{9}$$

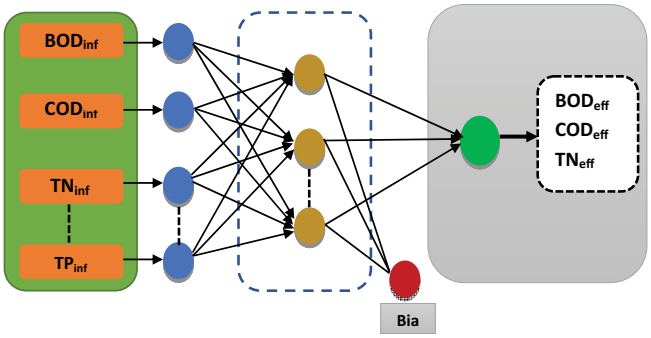

Fig. 3. Structure of the extreme learning machine network.

where G represents the hidden layer output and T is the transpose of the matrix. The output weights $\hat{B}$ can be estimated by inverting the matrix of the hidden layer using the Moore–Penrose generalized inverse function (+):

$$\hat{B} = G^{+}Y \tag{10}$$

Ultimately, the estimated values $\hat{y}$ (i.e., represents a predicted value of BOD_{eff}, COD_{eff}, TP_{eff} and TN_{eff}) can be determined by:

$$\hat{y} = \sum_{i=1}^{H} \hat{B}_i g_i \left(\alpha_i \cdot x_t + \beta_i\right) \tag{11}$$

2.4. Multi-layer perceptron neural network

MLP neural network, as one of the most commonly used types of ANN, has the capability to handle non-linear systems and has been described in several studies as a universal approximator among the different categories of ANNs [41]. As with the other traditional ANNs, MLP consists of an input layer, and one or more hidden and output layers in its architecture (Fig. 4) [42,43]. The nodes of the input layer are connected to those in the hidden layer and subsequently the output layer. The information and signals are processed and transmitted from the input to the output layer with the help of weights and biases through sequential mathematical operations. The Levenberg–Marquardt algorithm is used as a learning algorithm to optimize the error between the measured and computed values [44–46]. The training algorithm is iteratively repeated until the desired outcomes are achieved.

2.5. Multi-linear regression analysis

MLR is the most well-known and utilized linear regression due to it is multi predictor combination ability. In MLR, the linear correlation between one response (dependent variable or output) can be estimated with two or more predictors (independent variable or input) [46,47]. In the present study, MLR is employed for comparison with non-linear models and also as a benchmarking model. The most common form of MLR is shown as:

$$\bar{Y} = a_0 + \sum_{j=1} a_j + X_j \tag{12}$$

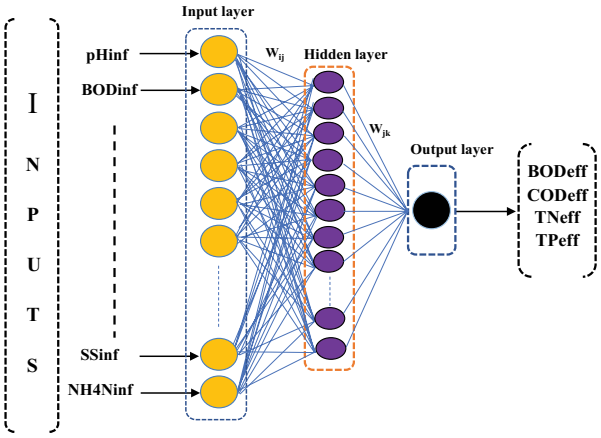

Fig. 4. Multilayer perceptron neural network (MLPNN) architecture.

where $\hat{Y}$ is the model's output, X_j's are the independent input variables to the model, and $a_0, a_1, a_2, \ldots, a_m$ are partial regression coefficients.

2.6. Principal component analysis

PCA is one of the common multi-variate statistical techniques for reducing the dimensions of high volumes of data. The dimensionality reduction is normally achieved by randomly identifying the linear correlation between the variables [48]. By applying this method, input variables are changed and be used as independent PCs variable [49]. Kaiser–Meyer–Olkin (KMO) is among the most common statistics used to assess the suitability of data in any factor analysis (FA) [24]. The classification of the KMO coefficient is demonstrated in Table 1 and the KMO index is presented in Eq. (4). For more explanation of PCA, refer to the studies of [23,48,50–52].

$$\text{KMO} = \frac{\sum\sum r_{ij}^2}{\sum\sum r_{ij}^2 + \sum\sum r_{ij}^2} \tag{13}$$

where r_{ij} is the correlation coefficient between the variable of i and j, and a_{ij} is the partial correlation coefficient between them.

2.7. Model performance indicators

Various evaluation criteria can be used to determine the comparative accuracy of the predictive models; as such, a multi-criteria indicator for measuring the model's performance was employed in this study. The determination coefficient of (R^2 or DC) as a goodness-of-fit and two statistical error including mean-squared error (RMSE) and mean absolute percentage error (MAPE) were used for the evaluation of the models [41,44]:

$$R^2 = \text{DC} = 1 - \frac{\sum_{j=1}^{N}\left[\left(Y\right)_{\text{obs},j} - \left(Y\right)_{\text{com},j}\right]^2}{\sum_{j=1}^{N}\left[\left(Y\right)_{\text{obs},j} - \left(\bar{Y}\right)_{\text{obs},j}\right]^2} \tag{14}$$

Table 1
Classification of KMO coefficients

Relation of data with FA	KMO coefficient
Excellent	≥0.9
Very well	0.8–0.89
Well	0.7–0.79
Mediocre	0.6–0.69
Poor	0.5–0.59
Unacceptable	<0.5

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\left(Y_{\mathrm{obsi}} - Y_{\mathrm{comi}}\right)^2}{N}} \tag{15}$$

$$\mathrm{MAPE} = \frac{1}{N}\left[\sum_{i=1}^{N}\left|\frac{Y_{\mathrm{obsi}} - Y_{\mathrm{comi}}}{Y_{\mathrm{obsi}}}\right|\right] \tag{16}$$

where N, Y_{obsi}, $\hat{Y}$ and Y_{comi} are data number, observed data, average value of the observed data and computed values, respectively.

3. Results and discussion

3.1. Implementation of scenario I

Various structures for ELM, 4PCs-ELM and 6PCs-ELM were used to obtain the best structure of the model. The optimum number of hidden neurons was identified as the best optimal ELM structure for all the combinations. PCA was employed for choosing the input variable in order to enhance the ELM prediction [22]. According to the obtained KMO value of 0.735, the PCA is suitable for all the output variables (Table 1). In PCA, different approaches are used for deciding which factors can noticeably affect the resulting pattern of the data. As such, this research employed the approach of selecting the factors with eigenvalues equal to or greater than 1.00 (Table 2). According to Holland, [50], in any correlation matrix, eigenvalues are used to condense the variance where the highest eigenvalues (1 and above) are traditionally considered for any analysis by eigenvectors ranking. Fig. 5a shows the specific values and the percentage variance of each factor as a graph, which demonstrates 9 input variables with the corresponding 9 eigenvectors and eigenvalues. Similarly, Table 2 shows the value of each factor and its percentage of separation from the primary variable. It can be seen from the table that more than 80% of factors were explained by the first 6 variables. Likewise, the results indicated that, for up to 8 factors, there exists a significant percentage of about 95%, which can be proved as in Fig. 5a of the obtained results.

Fig. 5b examined the orthonormal loadings biplot relationship between the variables, where the horizontal axis is the first PCA dimension representing 23.9% and the vertical axis is the second PCA dimension. The long or short red vector lines indicate the suitability of the presentation or otherwise. From both Figs. 5a and b we can extract both the 4PCs and 6PCs accordingly. According to Table 3, the best results for $\mathrm{BOD_{eff}}$, $\mathrm{COD_{eff}}$, $\mathrm{TN_{eff}}$ and $\mathrm{TP_{eff}}$ were obtained using 4PCs-ELM, 6PCs-ELM, 6PCs-ELM and 6PCs-ELM, respectively. This can be proved by comparing the values of R^2, RMSE and MAPE. Although the PCs-ELM combination generates the most accurate results in all cases, using the single ELM model was also observed to reliable for the prediction. This is due to its promising ability to handle highly complex and non-linear processes.

A close examination shows that both ELM and PCs-ELM produced different performance accuracies, which signifies that the individual model type responds in a different way to the same or different input parameters. Table 3 also confirms that in both calibration and verification, the PCs-ELM model achieved the lowest RMSE and MAPE for $\mathrm{BOD_{eff}}$, $\mathrm{COD_{eff}}$, $\mathrm{TN_{eff}}$ and $\mathrm{TP_{eff}}$ modeling. The result also shows increases in the PCs-ELM of about 12%, 2%, 20% and 6% for $\mathrm{BOD_{eff}}$, $\mathrm{COD_{eff}}$, $\mathrm{TN_{eff}}$ and $\mathrm{TP_{eff}}$ with regard to the novel ELM model. Box plots for observed data and predicted models are shown in Fig. 6. From the Fig, the PCs-ELM model was clearly found to obtain the best fit line between the observed and estimated values; hence, this demonstrates the high prediction ability for Nicosia MWWTPs and it may be therefore considered a valuable and reliable tool for conducting its performance analysis. The plot

Table 2
Eigenvalue and percentage of data explained by each factor

Number	Eigenvalue	Difference	Proportion	Cumulative Value	Cumulative Proportion
1	2.153019	0.646156	0.2392	2.153019	0.2392
2	1.506864	0.139714	0.1674	3.659883	0.4067
3	1.36715	0.358621	0.1519	5.027033	0.5586
4	1.008529	0.102154	0.1121	6.035561	0.6706
5	0.906374	0.213251	0.1007	6.941935	0.7713
6	0.693123	0.135041	0.077	7.635059	0.8483
7	0.558082	0.132987	0.062	8.193141	0.9103
8	0.425095	0.043332	0.0472	8.618236	0.9576
9	0.381764	–	0.0424	9	1

also demonstrates the closeness of all the models with the observed values; the plot contained (box and whisker median, mean and staples). According to the plot, the extent of spread of the values between the observed and predicted models indicates the superiority of the PCs-ELM models.

In the same way, the results of RMSE and MAPE depict the performance indicator for the best model. It was reported that the smaller the values of RMSE and MAPE, the more accurate the prediction results [53,54]. Further examination of performance accuracy was also investigated using a two-dimensional graphical diagram (i.e., Taylor diagram), as depicted in Figs. 7a–d. The Taylor diagram is a graphical representation method that exhibits how closely a model or different model matches the observed and corresponding computed values. Moreover, the computed models and the observed data are described quantitatively in terms of their correlation coefficient (R) and standard deviations (SD). Fig. 7a shows that the best predictive BOD_{eff} model is far from the actual (observed) data, which signifies less performance accuracy; this could be attributed to the small value of R and high dispersion between the observed and predictive model. Similarly, Figs. 7b–d proves the results in Table 3, the developed model (COD_{eff}, TN_{eff} and TP_{eff}) showed outstanding performance in determining the performance of Nicosia MWWTP. According to the value of R and SD for Figs. 7b–d, the best models depicted the extent and degree of the prediction skills.

Moreover, a scatter diagram of the best-computed model is shown in Fig. 8. The plots indicate a close agreement between the observed and computed values for COD_{eff}, TN_{eff} and TP_{eff}, while there is a fair agreement for BOD_{eff}. This conclusion is in line with that of Nourani et al. [3]. Note that data pairs closer to the 45° line represent better prediction cases in any scatter plots.

3.2. Implementation of scenario II

As stated above, different scenarios were constructed for the multi-parametric prediction of MWWTP performance. In scenario II, the MLP and MLR models were addressed according to the input variables stated in Eq. (2). As with any AI modeling, finding the optimal architecture is the main problem due to the fact that there is no standard pattern for selecting the desired architecture prior to the calibration phase [55]. As such, different numbers of hidden neurons ranging from 1 to 30 were observed in MLP by a trial and error procedure. Three different models were trained based on scenario I in section 2.2; the model types were defined as MLP-M1 (4-6-1), MLP-M2 (6-6-1) and MLP-M3 (9-10-1) indicating the three-input combination set in Eq. (2). In MLP-(4-6-1), 4 stands for number of inputs imposed to the model, 6 indicates the hidden neuron and 1 stands for the target output of the model. Similarly, for MLR, the models were defined as MLR-M1 (4-1). MLR-M2 (6-1) and MLR-M3 (9-1) indicating the model type, input and output of the models. The performance indices of MLP and MLR are shown in Table 4. It can clearly be observed that MLP-M2 (6-6-1) and MLR-M2 (6-1) outperformed other models for modeling the BOD_{eff}. Whereas for modeling

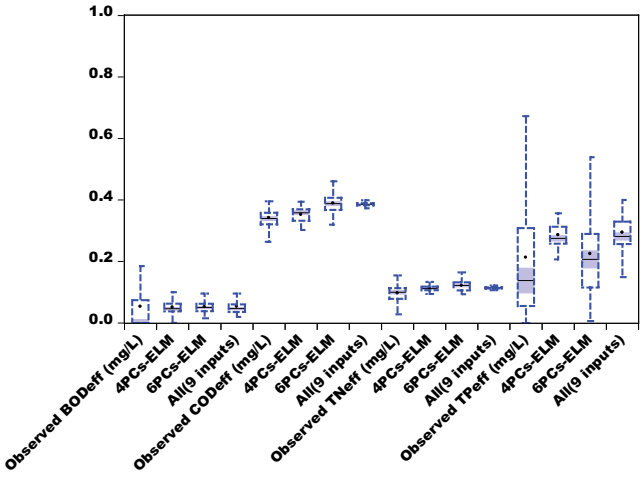

Fig. 6. Comparison box-plot of the observed data and all the predicted models.

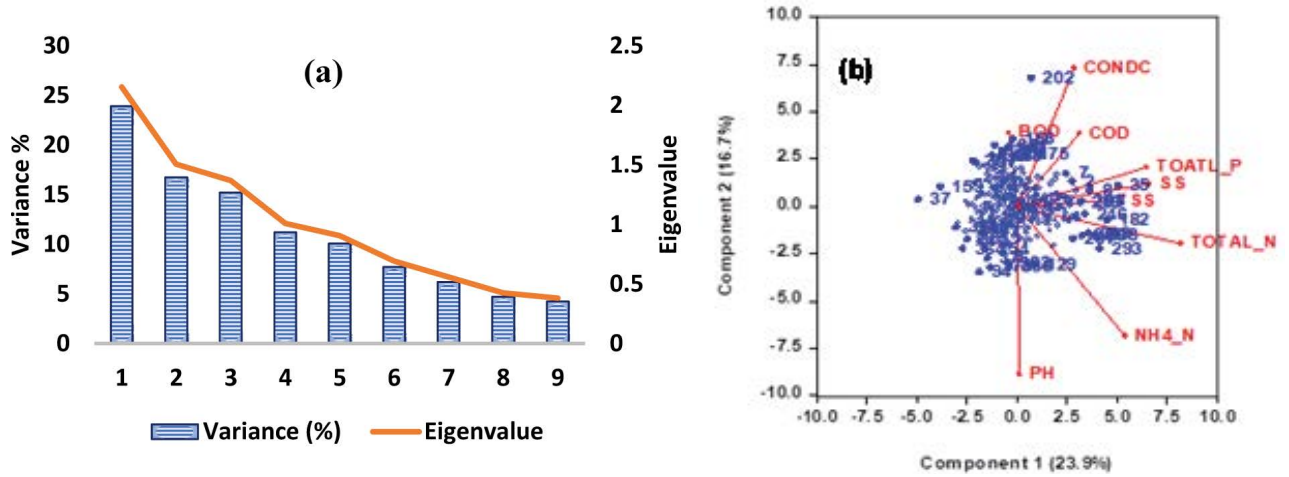

Fig. 5. (a) Shows the percentage variance vs. a number of factors and eigenvalue vs. a number of factors and (b) orthonormal loadings bi-plot of the first two components of the PCA model.

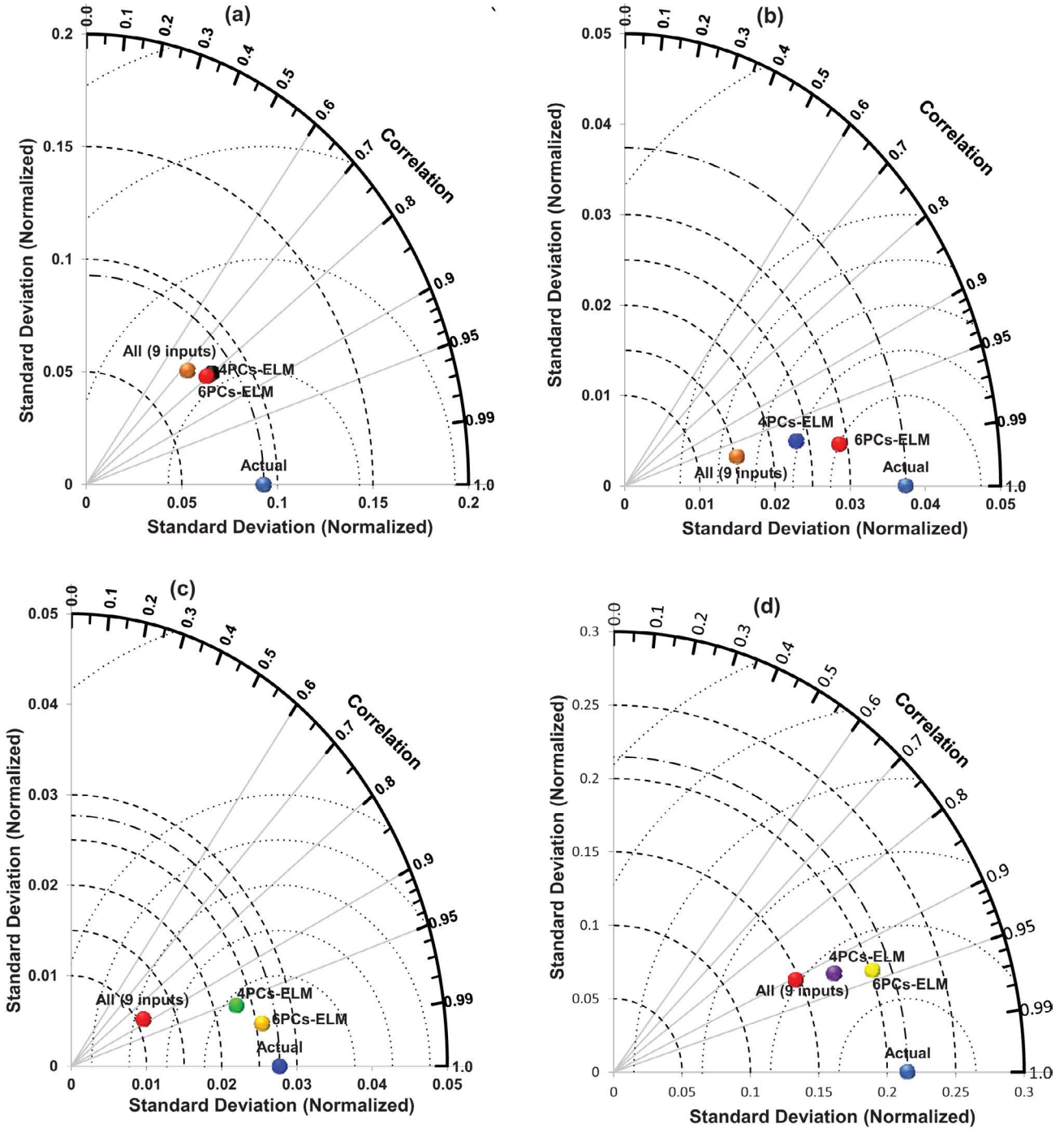

Fig. 7. Taylor diagram showing the degree of prediction in terms of R and SD for (a) BOD_{eff}, (b) COD_{eff}, (c) TN_{eff}, and (d) TP_{eff}.

the COD_{eff}, TN_{eff} and TP_{eff}, MLP-M3 (9-10-1) and MLR-M3 (9-1) models types emerged as the best combinations. The time series plots showing the relationship between the observed and computed values for the best MLP and MLR models are shown in Fig. 9.

According to Table 4, the presented results indicate the improved performance accuracy of MLP in comparison to MLR of up to 8%, 3%, 10% and 16% for BOD_{eff}, COD_{eff}, TN_{eff} and TP_{eff}, respectively. A similar conclusion was drawn by Nourani et al. [3] based on a comparison of SVM and MLR models. Based on the employed performance indices, it is apparent that MLP demonstrated better predictive skills than the MLR models despite the promising ability of MLR to predict COD_{eff}, TN_{eff} and TP_{eff}. This finding was also in line with that of Zhu et al. [43], who reported a slight performance increase of MLP over MLR model. According to the proposed scenarios (I and II), the comparative results between Tables 3 and 4 revealed that the best performance

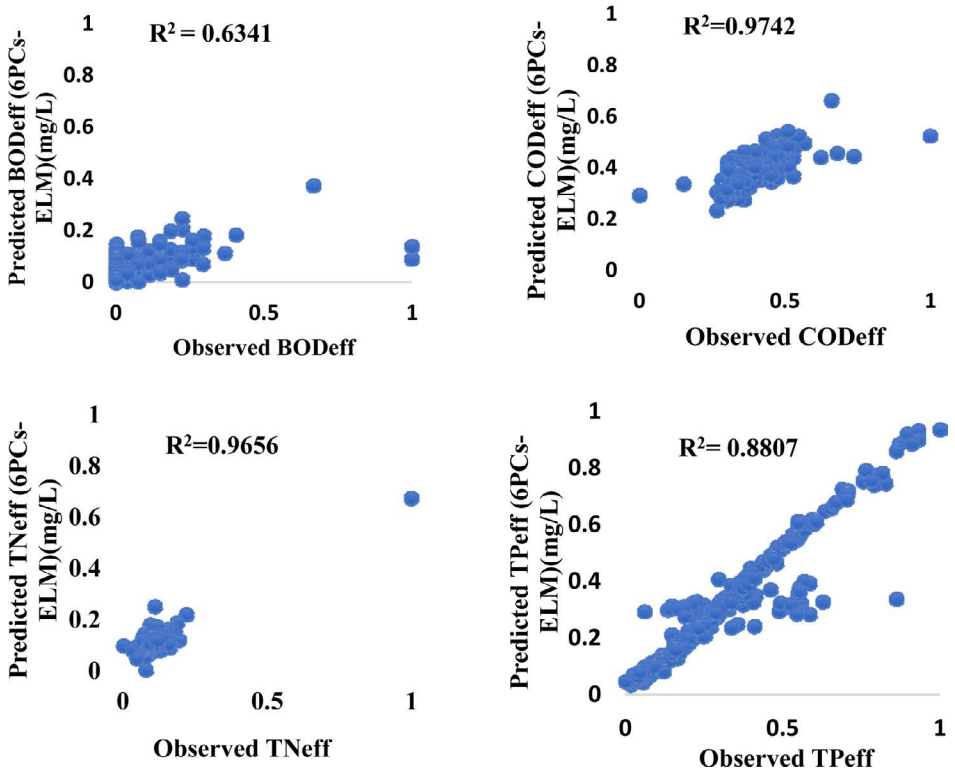

Fig. 8. Scatter plots of observed and computed values for the best model of (a) BOD_{eff}, (b) COD_{eff}, (c) TN_{eff} and (d) TP_{eff}.

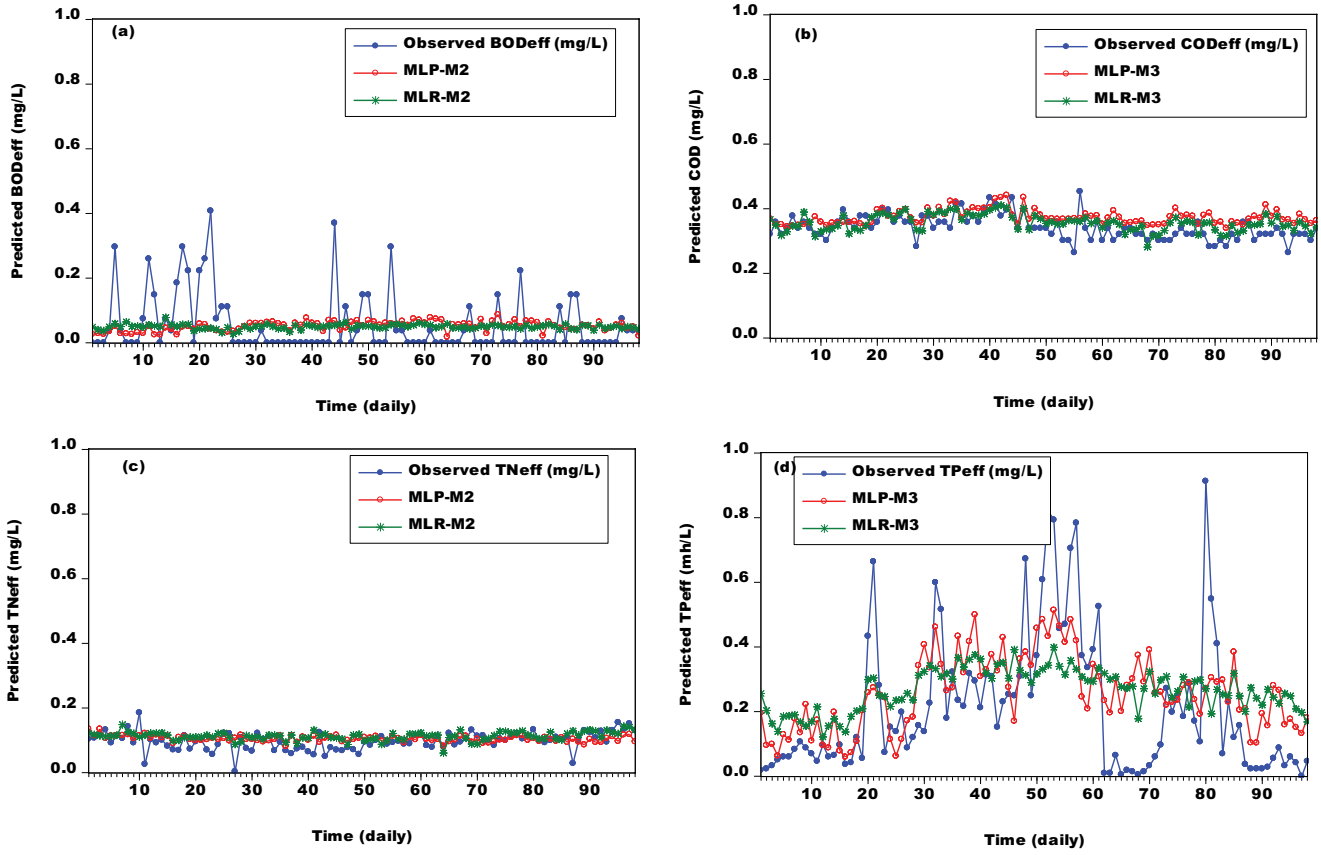

Fig. 9. Time series between observed and computed values for the best model of (a) BOD_{eff}, (b) COD_{eff}, (c) TN_{eff} and (d) TP_{eff}.

Table 3
Results of ELM, and PCs-ELM for BOD_{eff}, COD_{eff}, TN_{eff} and TP_{eff}

		Calibration			Verification		
Parameter	Model	R^2	RMSE	MAPE	R^2	RMSE	MAPE
BOD_{eff}	All (9 inputs)	0.5439	0.0749	0.0126	0.5168	0.0803	0.0482
	4PCs-ELM	0.5711	0.0714	0.0042	0.6341	0.0562	0.0143
	6PCs-ELM	0.5618	0.0727	0.0088	0.6285	0.0902	0.2009
COD_{eff}	All (9 inputs)	0.9632	0.0101	0.0051	0.9541	0.0399	0.0191
	4PCs-ELM	0.9522	0.0268	0.0003	0.9545	0.0534	0.0452
	6PCs-ELM	0.9757	0.0208	0.0103	0.9742	0.0515	0.0403
TN_{eff}	All (9 inputs)	0.8643	0.0424	0.0081	0.7651	0.0347	0.0837
	4PCs-ELM	0.9169	0.0387	0.0238	0.9128	0.0336	0.0561
	6PCs-ELM	0.9457	0.0983	0.0098	0.9656	0.0335	0.0522
TP_{eff}	All (9 inputs)	0.8803	0.0819	0.0112	0.8159	0.0718	0.1019
	4PCs-ELM	0.8629	0.0191	0.0335	0.8509	0.0450	0.2542
	6PCs-ELM	0.9629	0.0312	0.0205	0.8807	0.0491	0.1303

Table 4
Results of MLP and MLR models for BOD_{eff}, COD_{eff}, TN_{eff} and TP_{eff}

		Calibration			Verification		
	Model types	R^2	RMSE	MAPE	R^2	RMSE	MAPE
BOD_{eff}	MLP-M1 (4-6-1)	0.5473	0.1043	0.0564	0.4651	0.1093	0.0341
	MLP-M2 (6-6-1)	**0.5786**	**0.1024**	**0.0239**	**0.5776**	**0.1095**	**0.0468**
	MLP-M3 (9-10-1)	0.5331	0.1066	0.1445	0.5035	0.1091	0.1494
	MLR-M1 (4-1)	0.4775	0.1035	0.0093	0.4531	0.1093	0.0703
	MLR-M2 (6-1)	**0.5062**	**0.1034**	**0.0101**	**0.5020**	**0.1093**	**0.0757**
	MLR-M3 (9-1)	0.5005	0.1035	0.0291	0.4991	0.1091	0.2187
COD_{eff}	MLP-M1 (4-6-1)	0.9516	0.0774	0.0116	0.9756	0.0646	0.0774
	MLP-M2 (6-6-1)	0.9599	0.0705	0.0051	0.9747	0.0747	0.0857
	MLP-M3 (9-10-1)	**0.9617**	**0.0689**	**0.0027**	**0.9555**	**0.0648**	**0.0960**
	MLR-M1 (4-1)	0.9505	0.0783	0.0094	0.9419	0.0549	0.0955
	MLR-M2 (6-1)	0.9505	0.0734	0.0088	0.9242	0.0547	0.0893
	MLR-M3 (9-1)	**0.9574**	**0.0727**	**0.0043**	**0.9552**	**0.0536**	**0.0437**
TN_{eff}	MLP-M1 (4-6-1)	0.64026	0.08839	0.08961	0.63755	0.07613	0.55037
	MLP-M2 (6-6-1)	0.86359	0.08196	0.02441	0.81611	0.08324	0.16344
	MLP-M3 (9-10-1)	**0.87072**	**0.08028**	**0.03004**	**0.86662**	**0.08302**	**0.0799**
	MLR-M1 (4-1)	0.61499	0.08096	0.0187	0.52407	0.07361	0.21512
	MLR-M2 (6-1)	0.74987	0.08097	0.01867	0.75299	0.07358	0.21477
	MLR-M3 (9-1)	**0.76505**	**0.08044**	**0.01478**	**0.76181**	**0.07323**	**0.17002**
TP_{eff}	MLP-M1 (4-6-1)	0.74479	0.20443	0.03042	0.73283	0.02207	0.39568
	MLP-M2 (6-6-1)	0.72657	0.20961	0.03789	0.73995	0.02087	0.30032
	MLP-M3 (9-10-1)	**0.74923**	**0.19557**	**0.00829**	**0.72544**	**0.01761**	**0.21034**
	MLR-M1 (4-1)	0.63768	0.20647	0.02897	0.29993	0.02253	0.39024
	MLR-M2 (6-1)	0.64319	0.20489	0.02841	0.35973	0.02155	0.38279
	MLR-M3 (9-1)	**0.63421**	**0.20746**	**0.0193**	**0.56072**	**0.01978**	**0.26007**

accuracy was obtained with the ELM model. Hence, ELM yielded the best accuracy among all the models (MLP and MLR) in terms of predictive skills.

Further examination of the models proved that the ELM predicted values attained a high level of precision (Fig. 8). The PCs-ELM increased the prediction accuracy of BOD_{eff} up 5% and 13%, COD_{eff} up to an average of 2%, TN_{eff} up to 10% and 20% and TP_{eff} up to 15% and 32% with regard to the MLP and MLR models, respectively. This serves as additional evidence regarding the capability of PCs-ELM for

modeling the complex and uncertain systems in MWWTPs. Similarly, with larger R^2 and smaller values of RMSE and MAPE, ELM ranked as the best followed by MLP and lastly, the MLR model. With regards ELM with all 9 variables the prediction accuracy (RMSE) decreases by 3%, 2%, and 5% for BOD, COD, and TN and increases by 5% for TP.

However, there are several factors that affect the model's performance such as overfitting in the case of AI (ANN), class in balance, systematic noise associated with the data, pre-processing, model types and randomness of the data. According to [8,56–59] for a good analysis of any data intelligence model, the efficiency performance should include at least one goodness-of-fit (e.g., R^2) and at least one absolute error measure (e.g., RMSE). In addition, several studies have already shown that even for the same type of data set, the performance results may deviate from one model performance to another. For example, R^2 does not take into consideration any biases that might be present in the data. Therefore, a good model might have a low R^2 value, or a model that does not fit the data might have a high R^2 value. Hence, combining the goodness-of-fit, the error measure and biases measure could lead to promising and reliable simulation [60,61].

4. Conclusion

In this research, two scenarios (I and II) were investigated for modeling the performance of Nicosia MWWTP in terms of the effluents BOD_{eff}, COD_{eff}, TN_{eff} and TP_{eff} using three different model input combinations. The ELM, as a newly emerged black-box model combined with PCA was developed in scenario I, while in scenario II, traditional MLP neural network and MLR models were established for comparison.

In scenario I, PCA was employed in this study to understand whether it is feasible for improving the accuracy of the emerging ELM algorithm. The PCA technique helps the ELM mapping by its orthogonal transformation of variables and the reduction of system dimensionality. The obtained results showed an increase for PCs-ELM of about 12%, 2%, 20% and 6% for BOD_{eff}, COD_{eff}, TN_{eff} and TP_{eff}, respectively, with regard to the novel ELM model. Nevertheless, the ELM model demonstrated accurate prediction capability and can also serve as a reliable tool. On the other hand, PCA algorithms can be employed to reduce the dimensionality of the input vectors, which may lead to the achievement of highly accurate prediction.

For scenario II, MLP and MRL models were addressed according to the same input variables of the first scenario and the results indicated the improved performance accuracy of MLP with regard to MLR up to 8%, 3%, 10% and 16% for BOD_{eff}, COD_{eff}, TN_{eff} and TP_{eff}, respectively. According to the two scenarios, the comparative results revealed that the best performance accuracy was obtained using ELM model. Hence, ELM yielded the best accuracy among all the models (MLP and MLR) in terms of predictive skills. The outcomes of the current study may contribute to the mentioned multi-parametric modeling of the treated effluents and provide a reference benchmark for wastewater management and control. It is suggested that other algorithms may be applied with a combination of PCs in order to develop a new model that could produce higher accuracy and more reliable estimates.

Acknowledgment

The authors wish to thank the Management of Near East University and the staff of Nicosia municipal wastewater treatment plant, TRNC for providing the data used to carry out this research.

References

[1] WHO, UNCF (UNICEF), Progress on Sanitation and Drinking-Water: Joint Monitoring Programme 2010 Update, World Health Organization, United Nations Children's Fund, WHO Press, Geneva, Switzerland, 2012.
[2] T. Gómez, G. Gémar, M. Molinos-Senante, R. Sala-Garrido, R. Caballero, Assessing the efficiency of wastewater treatment plants: a double-bootstrap approach, J. Cleaner Prod., 164 (2017) 315–324.
[3] V. Nourani, G. Elkiran, S.I. Abba, Wastewater treatment plant performance analysis using artificial intelligence – an ensemble approach, Water Sci. Technol., 78 (2018) 2064–2076.
[4] S.I. Abba, G. Elkiran, Effluent prediction of chemical oxygen demand from the wastewater treatment plant using artificial neural network application, Procedia Comput. Sci., 120 (2017) 156–163.
[5] N. Bekkari, A. Zeddouri, Using artificial neural network for predicting and controlling the effluent chemical oxygen demand in wastewater treatment plant, Manage. Environ. Qual. Int. J., 30 (2019) 593–608.
[6] X.D. Wang, K. Kvaal, H. Ratnaweera, Explicit and interpretable nonlinear soft sensor models for influent surveillance at a full-scale wastewater treatment plant, J. Process Control, 77 (2019) 1–6.
[7] S.R. Naganna, P.C. Deka, M.A. Ghorbani, S.M. Biazar, N. Al-Ansari, Z.M. Yaseen, Dew Point temperature estimation: application of artificial intelligence model integrated with nature-inspired optimization algorithms, Water (Switzerland), 11 (2019) 1–17.
[8] S.I. Abba, Q.B. Pham, G. Saini, N.T. Linh, A.N. Ahmed, M. Mohajane, M. Khaledian, R.A. Abdulkadir, Q.-V. Bach, Implementation of data intelligence models coupled with ensemble machine learning for prediction of water quality index, Environ. Sci. Pollut. Res., 27 (2020) 41524–41539.
[9] G. Elkiran, V. Nourani, S.I. Abba, J. Abdullahi, Artificial intelligence-based approaches for multi-station modeling of dissolve oxygen in river, Global J. Environ. Sci. Manage., 4 (2018) 439–450.
[10] V. Nourani, G. Andalib, F. Sadikoglu, Multi-station streamflow forecasting using wavelet denoising and artificial intelligence models, Procedia Comput. Sci., 120 (2017) 617–624.
[11] Y.Y. Zhang, X. Gao, K. Smith, G. Inial, S.M. Liu, L.B. Conil, B.C. Pan, Integrating water quality and operation into prediction of water production in drinking water treatment plants by genetic algorithm enhanced artificial neural network, Water Res., 164 (2019) 114888, https://doi.org/10.1016/j.watres.2019.114888.
[12] G. Elkiran, V. Nourani, S.I. Abba, Multi-step ahead modeling of river water quality parameters using ensemble artificial intelligence-based approach, J. Hydrol., 577 (2019) 123962, https://doi.org/10.1016/j.jhydrol.2019.123962.
[13] V. Nourani, N. Farboudfam, Rainfall time series disaggregation in mountainous regions using hybrid wavelet-artificial intelligence methods, Environ. Res., 168 (2019) 306–318.
[14] A. Maleki, S. Nasseri, M.S. Aminabad, M. Hadi, Comparison of ARIMA and NNAR models for forecasting water treatment plant's influent characteristics, KSCE J. Civ. Eng., 22 (2018) 3233–3245.
[15] W.C. Chen, N.-B. Chang, W.K. Shieh, Advanced hybrid fuzzy-neural controller for industrial wastewater treatment, J. Environ. Eng., 127 (2001) 1048–1059.

[16] F. Granata, S. Papirio, G. Esposito, R. Gargano, G. de Marinis, Machine learning algorithms for the forecasting of wastewater quality indicators, Water (Switzerland), 9 (2017) 1–12.
[17] A.K. Verma, T.N. Singh, Prediction of water quality from simple field parameters, Environ. Earth Sci., 69 (2013) 821–829.
[18] H. Guo, K.H. Jeong, J.Y. Lim, J.W. Jo, Y.M. Kim, J.-P. Park, J.H. Kim, K.H. Cho, Prediction of effluent concentration in a wastewater treatment plant using machine learning models, J. Environ. Sci., 32 (2015) 90–101.
[19] S.I. Abba, V. Nourani, G. Elkiran, Multi-parametric modeling of water treatment plant using AI-based non-linear ensemble, J. Water Supply Res. Technol. AQUA, 68 (2019) 547–561.
[20] S.I. Abba, G. Elkiran, V. Nourani, Non-linear Ensemble Modeling for Multi-step Ahead Prediction of Treated COD in Wastewater Treatment Plant, R.A. Aliev, J. Kacprzyk, W. Pedrycz, M. Jamshidi, M.B. Babanli, F.M. Sadikoglu, Eds., 10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions – ICSCCW-2019, Advances in Intelligent Systems and Computing, Vol. 1095, Springer, Cham, 2020, pp. 683–689.
[21] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing, 70 (2006) 489–501.
[22] A. Solgi, A. Pourhaghi, R. Bahmani, H. Zarei, Improving SVR and ANFIS performance using wavelet transform and PCA algorithm for modeling and predicting biochemical oxygen demand (BOD), Ecohydrol. Hydrobiol., 17 (2017) 164–175.
[23] Q. Wang, Kernel principal component analysis and its applications in face recognition and active shape models, Comput. Vision Pattern Recognit., (2012) arXiv:1207.3538.
[24] S.I. Abba, Q.B. Pham, A.G. Usman, N.T. Thuy Linh, D.S. Aliyu, Q. Nguyen, Q.-V. Bach, Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant, J. Water Process Eng., 33 (2020) 101081, https://doi.org/10.1016/j.jwpe.2019.101081.
[25] M. Yaqub, H. Asif, S.B. Kim, W.T. Lee, Modeling of a full-scale sewage treatment plant to predict the nutrient removal efficiency using a long short-term memory (LSTM) neural network, J. Water Process Eng., 37 (2020) 101388, https://doi.org/10.1016/j.jwpe.2020.101388.
[26] J.-H. Kang, J.H. Song, S.S. Yoo, B.-J. Lee, H.W. Ji, Prediction of odor concentration emitted from wastewater treatment plant using an artificial neural network (ANN), Atmosphere (Basel), 11 (2020) 784, https://doi.org/10.3390/atmos11080784.
[27] M. Ansari, F. Othman, A. El-Shafie, Optimized fuzzy inference system to enhance prediction accuracy for influent characteristics of a sewage treatment plant, Sci. Total Environ., 722 (2020) 137878, doi: 10.1016/j.scitotenv.2020.137878.
[28] A.M. Anter, D. Gupta, O. Castillo, A novel parameter estimation in dynamic model via fuzzy swarm intelligence and chaos theory for faults in wastewater treatment plant, Soft Comput., 24 (2020) 111–129.
[29] N. Patel, J. Ruparelia, J. Barve, Prediction of total suspended solids present in effluent of primary clarifier of industrial common effluent treatment plant: mechanistic and fuzzy approach, J. Water Process Eng., 34 (2020) 101146, https://doi.org/10.1016/j.jwpe.2020.101146.
[30] A. Sharafati, S.B.H.S. Asadollah, M. Hosseinzadeh, The potential of new ensemble machine learning models for effluent quality parameters prediction and related uncertainty, Process Saf. Environ. Prot., 140 (2020) 68–78.
[31] UNDP, New Nicosia Waste Water Treatment Plant, United Nations Development Programme, Nicosia, Northern Part of Cyprus, 2014.
[32] P. Shi, G.H. Li, Y.M. Yuan, G.Y. Huang, L. Kuang, Prediction of dissolved oxygen content in aquaculture using clustering-based softplus extreme learning machine, Comput. Electron. Agric., 157 (2019) 329–338.
[33] G. Huang, G.-B. Huang, S.J. Song, K.Y. You, Trends in extreme learning machines: a review, Neural Networks, 61 (2015) 32–48.
[34] Z.M. Yaseen, S.O. Sulaiman, R.C. Deo, K.-W. Chau, An enhanced extreme learning machine model for river flow forecasting: state-of-the-art, practical applications in water resource engineering area and future research direction, J. Hydrol., 569 (2018) 387–408.
[35] S.J. Hadi, S.I. Abba, S.S. Sammen, S.Q. Salih, N. Al-Ansari, Z.M. Yaseen, Non-linear input variable selection approach integrated with non-tuned data intelligence model for stream-flow pattern simulation, IEEE Access, 7 (2019) 141533–141548.
[36] S. Zhu, S. Heddam, Prediction of dissolved oxygen in urban rivers at the three Gorges reservoir, China: extreme learning machines (ELM) versus artificial neural network (ANN), Water Qual. Res. J. Canada, 55 (2020) 106–118.
[37] H. Chen, Q. Zhang, J. Luo, Y. Xu, X. Zhang, An enhanced Bacterial Foraging Optimization and its application for training kernel extreme learning machine, Appl. Soft Comput., 86 (2020) 105884.
[38] S. Heddam, O. Kisi, Extreme learning machines: a new approach for modeling dissolved oxygen (DO) concentration with and without water quality variables as predictors, Environ. Sci. Pollut. Res., 24 (2017) 16702–16724.
[39] J. Jin, P. Jiang, L. Li, H. Xu, G. Lin, Water quality monitoring at a virtual watershed monitoring station using a modified deep extreme learning machine, Hydrol. Sci. J., 65 (2020) 415–426.
[40] Z.M. Yaseen, H. Faris, N. Al-Ansari, Hybridized extreme learning machine model with salp swarm algorithm: a novel predictive model for hydrological application, Complexity, 2020 (2020), doi: 10.1155/2020/8206245.
[41] Q.B. Pham, S.I. Abba, A.G. Usman, N.T.T. Linh, V. Gupta, A. Malik, R. Costache, N.D. Vo, D.Q. Tri, Potential of hybrid data-intelligence algorithms for multi-station modeling of rainfall, Water Resour. Manage., 33 (2019) 5067–5087.
[42] M.A. Ghorbani, R.C. Deo, Z.M. Yaseen, M.H. Kashani, B. Mohammadi, Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran, Theor. Appl. Climatol., 133 (2018) 1119–1131.
[43] S.L. Zhu, S. Heddam, E.K. Nyarko, M. Hadzima-Nyarko, S. Piccolroaz, S.Q. Wu, Modeling daily water temperature for rivers: comparison between adaptive neuro-fuzzy inference systems and artificial neural networks models, Environ. Sci. Pollut. Res., 26 (2019) 402–420.
[44] A.G. Usman, S. Işik, S.I. Abba, A novel multi-model data-driven ensemble technique for the prediction of retention factor in HPLC method development, Chromatographia, 83 (2020) 933–945.
[45] S.I. Abba, A.G. Usman, S. Işik, Simulation for response surface in the HPLC optimization method development using artificial intelligence models: a data-driven approach, Chemom. Intell. Lab. Syst., 201 (2020) 104007, https://doi.org/10.1016/j.chemolab.2020.104007.
[46] H.U. Abdullahi, A.G. Usman, S.I. Abba, Modeling the absorbance of a bioactive compound in HPLC method using artificial neural network and multilinear regression methods, Dutse J. Pure Appl. Sci., 6 (2020) 362–371.
[47] S.I. Abba, S.J. Hadi, S.S. Sammen, S.Q. Salih, R.A. Abdulkadir, Q.B. Pham, Z.M. Yaseen, Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination, J. Hydrol., 587 (2020) 124974, doi: 10.1016/j.jhydrol.2020.124974.
[48] T.T. Yu, S. Yang, Y. Bai, X. Gao, C. Li, Inlet water quality forecasting of wastewater treatment based on kernel principal component analysis and an extreme learning machine, Water (Switzerland), 10 (2018) 873, doi: 10.3390/w10070873.
[49] M. Noori, R. Abdoli, M.A. Ghasrodashti, A.A. Ghasrodashti, J. Ghazizade, Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis: a case study of Mashhad, Environ. Prog. Sustainable Energy, 28 (2009) 249–258.
[50] S.M. Holland, Principal components analysis (PCA), Dep. Geol. Univ. Georg. Athens, GA, 2008, pp. 30602–32501.
[51] J. Yang, G.W. Xu, H.W. Kong, Y.F. Zheng, T. Pang, Q. Yang, Artificial neural network classification based on high-performance liquid chromatography of urinary and serum

nucleosides for the clinical diagnosis of cancer, J. Chromatogr. B, 780 (2002) 27–33.
[52] W.-Z. Lu, W.-J. Wang, X.-K. Wang, S.-H. Yan, J.C. Lam, Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong, Environ. Res., 96 (2004) 79–87.
[53] M.S. Gaya, M.U. Zango, L.A. Yusuf, M. Mustapha, B. Muhammad, A. Sani, A. Tijjani, N.A. Wahab, M.T. Khairi, Estimation of turbidity in water treatment plant using Hammerstein-Wiener and neural network technique, Indonesian J. Electr. Eng. Comput. Sci., 5 (2017) 666–672.
[54] M.S. Gaya, S.I. Abba, A.M. Abdu, A.I. Tukur, M.A. Saleh, P. Esmaili, N.A. Wahab, Estimation of water quality index using artificial intelligence approaches and multi-linear regression, IAES Int. J. Artif. Intell., (2020) 8938, doi: 10.11591/ijai.v9.i1.pp126-134.
[55] S.W. Kim, V.P. Singh, Modeling daily soil temperature using data-driven models and spatial distribution, Theor. Appl. Climatol., 118 (2014) 465–479.
[56] B. Mohammadi, N.T.T. Linh, Q.B. Pham, A.N. Ahmed, J. Vojteková, Y. Guan, A. El-Shafie, Adaptive neuro-fuzzy inference system coupled with shuffled frog leaping algorithm for predicting river streamflow time series, Hydrol. Sci. J., (2020) (In Press).
[57] D.R. Legates, G.J. McCabe Jr., Evaluating the use of "goodness of fit" measures in hydrologic and hydroclimatic model validation, Water Resour. Res., 35 (1999) 233–241.
[58] M. Alas, S.I.A. Ali, Y. Abdulhadi, S.I. Abba, Experimental evaluation and modeling of polymer nanocomposite modified asphalt binder using ANN and ANFIS, J. Mater. Civ. Eng., 32 (2020) 04020305.
[59] R.A. Abdulkadir, S.I.A. Ali, S.I. Abba, P. Esmaili, Forecasting of daily rainfall at Ercan Airport Northern Cyprus: a comparison of linear and non-linear models, Desal. Water Treat., 177 (2020) 297–305.
[60] S.I. Abba, N.T. Linh, J. Abdullahi, S.I. Ali, Q.B. Pham, R.A. Abdulkadir, R. Costache, D.T. Anh, Hybrid machine learning ensemble techniques for modeling dissolved oxygen concentration, IEEE Access, 8 (2020) 157218–157237.
[61] Q.B. Pham, M.S. Gaya, S.I. Abba, R.A. Abdulkadir, P. Esmaili, N.T. Linh, C. Sharma, A. Malik, D.N. Khoi, Modeling of Bunus regional sewage treatment plant using machine learning approaches, Desal. Water Treat., 203 (2020) 80–90.