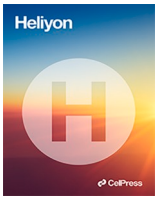
Heliyon 9 (2023) e21351

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



# Predicting biochemical oxygen demand in wastewater treatment plant using advance extreme learning machine optimized by Bat algorithm

Hayat Mekaoussi [a,b], Salim Heddam [c,*], Nouri Bouslimanni [d], Sungwon Kim [e], Mohammad Zounemat-Kermani [f]

[a] *Institute of veterinary and agronomic sciences, Agronomy Department, Hydraulics Division, University Batna 1-Hadj Lakhdar- Allées 19 mai, Route de Biskra Batna, 05000 Algeria*
[b] *Laboratory of Research in Biodiversity Interaction Ecosystem and Biotechnology (LRIBEB) University 20 Août 1955 Skikda, Algeria*
[c] *Laboratory of Research in Biodiversity Interaction Ecosystem and Biotechnology (LRIBEB), Faculty of Science, Agronomy Department, University 20 Août 1955-Skikda, Route El Hadaik, BP 26, Skikda, Algeria*
[d] *Institute of veterinary and agronomic sciences, Agronomy Department, Chemical Division, University Batna 1-Hadj Lakhdar- Allées 19 mai, Route de Biskra Batna, 05000 Algeria*
[e] *Department of Railroad Construction and Safety Engineering, Dongyang University, Yeongju 36040, Republic of Korea*
[f] *Department of Water Engineering, Shahid Bahonar University of Kerman, Kerman, Iran*

A R T I C L E I N F O

*Keywords:*
Modelling
WWTP
$BOD_5$
ELM
Bat algorithm
RVFL
RFR
GPR
MLPNN

A B S T R A C T

Wastewater quality modelling plays a vital role in planning and management of wastewater treatment plants (*WWTP*). This paper develops a new hybrid machine learning model based on extreme learning machine (*ELM*) optimized by Bat algorithm (*ELM-Bat*) for modelling five day effluent biochemical oxygen demand ($BOD_5$). Specifically, this hybrid model combines the Bat algorithm for model parameters optimization and the standalone ELM. The proposed model was developed using historical measured effluents wastewater quality variables, i.e., the chemical oxygen demand (*COD*), temperature, pH, total suspended solid (*TSS*), specific conductance (*SC*) and the wastewater flow (*Q*). The performances of the hybrid ELM-Bat were compared with those of the multilayer perceptron neural network (*MLPNN*), the random forest regression (*RFR*), the Gaussian process regression (GPR), the random vector functional link network (RVFL), and the multiple linear regression (MLR) models. By comparing several input variables combination, the improvement achieved in the accuracy of prediction through the hybrid ELM-Bat was quantified. All models were first calibrated using training dataset and later tested using validation and based on four performances metrics namely, root mean square error (RMSE), mean absolute error (MAE), the correlation coefficient (R), and the Nash-Sutcliffe model efficiency (NSE). In all, it is concluded that the ELM-Bat is the most accurate model when all the six input were included as input variables, and it outperforms all other benchmark models in terms of predictive accuracy, exhibiting RMSE, MAE, R and NSE values of approximately, 0.885, 0.781, 2.621, and 1.989, respectively.

* Corresponding author
*E-mail addresses:* hayet.mekaoussi@univ-batna.dz (H. Mekaoussi), heddamsalim@yahoo.fr, s.heddam@univ-skikda.dz (S. Heddam), nouri.bouslimani@univ-batna.dz (N. Bouslimanni), swkim1968@dyu.ac.kr (S. Kim), zounemat@uk.ac.ir (M. Zounemat-Kermani).

https://doi.org/10.1016/j.heliyon.2023.e21351
Received 16 May 2023; Received in revised form 8 October 2023; Accepted 19 October 2023
Available online 21 October 2023

## 1. Introduction

Wastewater treatment plants (WWTPs) are an important aspect of cities and towns' infrastructure because they improve societal life and health by treating municipal and industrial sewage and releasing the treated and cleaned wastewater effluent into different sectors of industry, agriculture, and society. Releasing and exposing raw wastewater to surface and groundwater resources has a significant negative and even hazardous environmental impact owing to the consumption of dissolved oxygen by microbes. Because of its biological nature and the infinite components that may be identified, generating raw wastewater is exceedingly difficult in terms of chemical composition. Consequently, biological characteristic of wastewater must be adequately controlled during the all treatment process, and continuously monitored in order to evaluate the WWTPs' efficiency. Thus, the five day biochemical oxygen demand ($BOD_5$), along with other wastewater variables such as ammoniacal nitrogen ($NH_3$), chemical oxygen demand (COD), and several organic compounds, is known as one of the important and critical biological wastewater variables. $BOD_5$ measures the biodegradable content of wastewater and dictates the amount of aeration, which is the most energy-intensive stage in WWTPs. On the other hand, $BOD_5$ is one of the key indicators for evaluating surface water quality. It is commonly utilized in the monitoring of organic contamination in water bodies [1–5].

**Nomenclature**

| | |
|---|---|
| AdaBoost | adaptive boosting |
| ANFIS | adaptive neuro-fuzzy inference system |
| $NH_3$ | Ammonia |
| ANN | artificial neural network |
| $BOD_5$ | biochemical oxygen demand |
| DT | decision tree |
| EC | electrical conductivity |
| XGBoost | eXtreme Gradient Boosting |
| ELM | Extreme Learning Machine |
| GRNN | generalized regression neural network |
| GB | Gradient Boosting |
| KNN | K-Nearest Neighbor |
| ML | Machine Learning |
| OrgN | Organic nitrogenous compounds |
| OrgP | Organic phosphorous compounds |
| OP | Orthophosphates |
| pH | Potential of Hydrogen |
| RBF | Radial Basis Function |
| RFR | Random Forest regression |
| RNN | Recurrent Neural Network |
| SS | Suspended Sediment |
| TP | Total Phosphorus |
| TSS | Total Suspended Solids |
| WWTP | Wastewater Treatment Plant |
| M5Tree | M5 model tree |
| LSSVM | least squares support vector machine |
| MARS | Multivariate adaptive regression Spline |

Assessing all of the influent characteristics takes time and necessitates performing difficult experiments and consuming hazardous materials, as thoroughly stated in the standard methodology for water and wastewater assessment. To solve this issue, various electrical sensors have recently been created to provide real-time measurements of the quality characteristics of the influent [2,6]. Nevertheless, several critical metrics, such as $BOD_5$ and COD, are difficult and expensive to measure using sensors, necessitating the creation of mathematical prediction models for calculating their values based on previous data. Without explicitly specifying the treatment process using mathematical or chemical formulae, machine learning (ML) is capable of modeling complicated nonlinear connections. It provides the opportunity to explore new knowledge of wastewater behavior, which is difficult to detect, in contrast to traditional models. ML integration with wastewater treatment processes has been successfully used as a capable soft computing tool to boost environmental preservation, optimize plant performance, and improve the treatment process [7,8].

Over the course of the past two decades, several studies have applied ML models for modeling and predicting $BOD_5$ in water bodies and rivers [9], and some research has dealt with modeling water quality indices in WWTPs [2,10]. For example, Qiao et al. [11] employed a fuzzy neural network to account for the nonlinear inaccuracy of the mechanism model of the sewage system. They applied a gradient descent approach to adjust the tuning parameters of the network. According to the experimental findings, the applied fuzzy-based ML method had better estimation accuracy than the conventional radial basis function (RBF) artificial neural network

(ANN). Heddam et al. [12] developed a model based on a generalized regression neural network (GRNN) to predict the concentration of effluent $BOD_5$ at a WWTP, located in the east of Algeria. The constructed GRNN model was built on a function of five effluent wastewater quality variables, such as pH, temperature, COD, electrical conductivity (EC) and total suspended solids (TSS). It was reported that the developed GRNN model produced consistent and acceptable results for estimating BOD at the WWTP. Yu et al. [13] used an extreme learning machine (ELM), which is considered a fast and reliable ANN model, to estimate the $BOD_5$ value at WWTP using real-time data. The findings of the study demonstrated the satisfactory capability of the ELM model.

Alsulaili and Refaie [14] constructed an ANN model to forecast wastewater influent $BOD_5$ at a WWTP located in Kuwait. The

**Table 1**
Summary of the conventional and modern ML models in modeling BOD in WWTPs.

| References | Case study/Location | Applied datasets | Models employed | Remarks |
|---|---|---|---|---|
| [16] | Wastewater treatment plant at RIPASA, Brazil | $BOD_5$, COD, Inlet Discharge, Outlet Discharge | Artificial Neural Network (ANN) | The results demonstrated the benefit of ANNs in representing highly nonlinear interactions, even in a system with operational data restrictions. |
| [17] | El-Asfar WWTP in the Greater Cairo district, Egypt | $BOD_5$, Suspended sediment (SS) | ANN | The generated models continuously outperform in the face of variable precision and amount of input data. |
| [18] | Not disclosed due to confidentiality | sewage sample odours | ANN | Overall, the results show that ANNs may be utilized to categorize sewage samples collected from various locations of a wastewater treatment facility. |
| [19] | A local WWTP in Turkey | 364 daily records of the year 2005 consists of COD, Discharge, suspended sediment (SS), total nitrogen, and phosphorus | ANN, MLR | The ANN model was proven to be effective in calculating the daily $BOD_5$ in the input of wastewater biochemical treatment facilities. |
| [1] | WWTP in south Iran | WWTP parameters measured over the span of almost two years | ANN | It was discovered that filtering the data is critical for developing better ANNs models. Moreover, adopting a multiple input-single output technique yields a superior ANN. |
| [11] | A small scale wastewater treatment plant in Beijing | COD, SS, pH, DO | K-means clustering method and Fuzzy Neural Network | The simulation results revealed that the TSFNN with K-means clustering outperforms the other two techniques in terms of approximation performance in predicting $BOD_5$ levels. |
| [12] | Sidi Marouane Wastewater Treatment Plant (WWTP), Algeria | 691 measured data points based on COD, TSS, electrical conductivity (EC), temperature, and pH | Generalized regression neural network (GRNN) and MLR | Based on the findings of this investigation, the created GRNN model can be utilized to correctly estimate $BOD_5$ at WWTP. |
| [13] | Benchmark Simulation Model no. 1 (BSM1) | Discharge, SS, TSS, and other quality parameters | extreme learning machine (ELM) based on an improved cuckoo search algorithm (ICS) | According to simulation findings, the soft sensor model has superior real-time performance, high prediction accuracy, and better generalization performance for $BOD_5$ measurement. |
| [14] | WWTP in Kuwait | Dataset covers seven years of operation, containing 2397 observations, including influent temperature, pH, conductivity, $BOD_5$, COD, and TSS | ANN | The authors discovered that the COD characteristic had the most impact in predicting $BOD_5$ outcomes. |
| [3] | Wastewater treatment plant in Hong Kong | Input data consists of COD, TSS, pH, Discharge, Zinc, OP-P, Cond, NH3, and sediment | XGBoost, ANN, SVM | XGBoost calculated a broad variety of BOD5 values, demonstrating consistent performance across several test sets. |
| [20] | Zargandeh WWTP, Iran | A dataset that contains 265 observations for the prediction of COD, BOD5, and TSS | ANN, ANN optimized by GA: (MLP-GA) | The most accurate results were obtained by monitoring the prediction model performance by modifying the ANN design using the GA. |
| [21] | Benchmark Simulation Model no. 1 (BSM1) | BSM1 model consists of five reaction tanks and one settling tank. Thirteen auxiliary variables like soluble and insoluble organic matters, DO, NH3, and alkalinity | Broad learning system (BLS), SVM, RNN, and ANNs | The overall prediction accuracy of SVM and RNN was lower than that of OBLS; the performance of SVM and RNN was influenced by the environment, resulting in lower prediction accuracy and a weaker overall prediction impact than OBLS. |
| [2] | Madinat Salman WWTP, Bahrain | Datasets from four WWTPs, including pH, conductivity, TSS, NH3, COD, TP, TN | RF, GB, DT, and AdaBoost | The proposed BOD prediction model was a decision-supporting tool to aid WWTP operators in acquiring the essential information. In general, the RF and GB outperformed the other versions. |
| [15] | Influents of 7 WWTPs in Hong Kong | Monthly data collected from the inflow of 7 WWTPs over a three-year period. Datasets include concentrations of five independent variables of TSS, NH3, OrgN, InorgP and OrgP | GEP, MLPNN, MLR, GB, RT | The goal parameters were chosen to be $BOD_5$ and COD. The most effective indicators for predicting BOD and COD were TSS and $NH_3$. |

collection contains 2397 observations from seven years of operation. Many parameters were assessed, including influent temperature, pH, EC, COD, and TSS. The authors discovered that the COD characteristic had the most impact on predicting $BOD_5$ outcomes. Ching et al. [3] developed and upgraded a soft sensor by combining ANN, SVM, and eXtreme Gradient Boosting (XGBoost) ML models to measure the $BOD_5$ of two distinct WWTPs. The models findings demonstrated that XGBoost could detect the high values better than conventional soft sensors. In addition, analyzing the results demonstrated that XGBoost was more accurate than the SVM in some specific tests. In another study, Qambar and Al Khalidy [2] applied integrated ML models combined with remote sensing techniques to predict $BOD_5$ in four WWTPs. The ML models include tree-based models, such as RF, DT, AdaBoost, and gradient boost (GB) algorithms. They applied the k-fold cross-validation technique for developing and executing the ML models. Based on the sensitivity analysis, it was found that pH, $NH_3$, and conductivity are among the most influential parameters in modeling $BOD_5$. It was reported that the RF and GB outperformed the other models in general. Recently, Aghdam et al. [15] trained several ML models, including Gene expression programming (GEP), multilayer perception neural networks (MLPNN), multi-linear regression (MLR), k-nearest neighbors, GB, and regression trees (RT)-based models, for the prediction of monthly $BOD_5$ and COD. Based on the tree-year datasets, the GEP modeling findings were shown to be consistent with the underlying chemistry of the wastewater quality parameters.

In order to gain a general insight into the use of conventional and modern ML models for modeling $BOD_5$ at WWTPs, a summary of the research conducted in this field is compiled in Table 1. According to the information in Table 1, the initial studies (up to 2012) on the application of ML models in modeling $BOD_5$ was confined to the use of ANN, namely feedforward ANN. Nevertheless, the adoption of various types of ML models, including SVRs, tree-based (e.g., RT), and ensemble (e.g., XGBoost), has increased. Furthermore, as ML models get more complicated, it has been possible in recent years to utilize additional input parameters in models. For example, in the research published in 2022 and 2023, more than seven input factors were used to predict $BOD_5$ increases, which led to the improvement of the ML outcomes. Therefore, in the present study, we propose a new modelling strategy for better prediction of $BOD_5$ in WWTPs using an extreme learning machine model optimized using the Bat algorithm (ELM-Bat). The proposed ELM-Bat was developed using effluent wastewater quality collected at the Sidi Marouane WWTP, Algeria. The performances of the ELM-Bat were compared with those of the multilayer perceptron neural network (MLPNN), random forest regression (RFR),Gaussian process regression (GPR), random vector functional link (RVFL), and the standalone multiple linear regression (MLR). The innovation and scientific contribution of the present study can be summarized as follow: (*i*) a novel hybrid ML model was introduced which takes advantages of ML and metaheuristic optimization algorithm for improving the prediction of the effluent $BOD_5$, and (*ii*) the effect of several water quality variables on the estimation of $BOD_5$ has been analyzed. The remainder of the present paper is as follow: Section 2 provides a brief description of the WWTP plant and the dataset used for models development. Section 3 provides the theoretical description of the ML models used in the present study. Section 4 was reserved to the experimental results and discussion. In Section 5 we provide some conclusions and future recommendations.

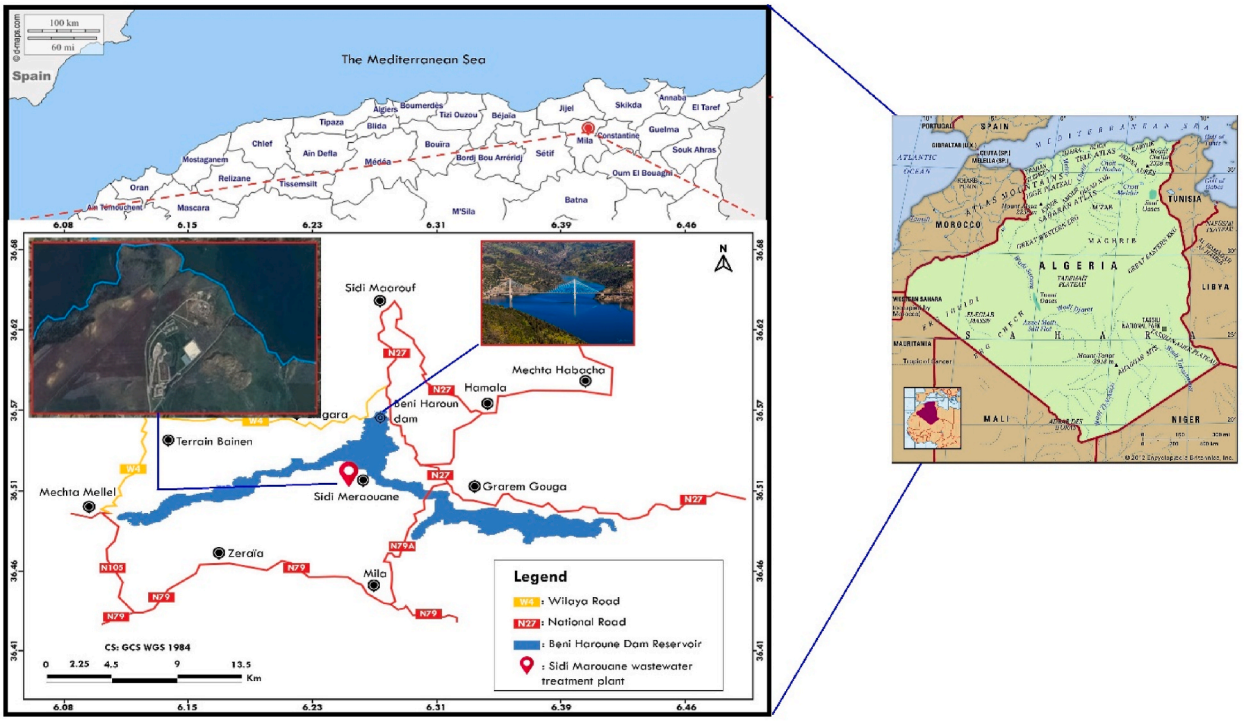



**Fig. 1.** Map showing the location of Sidi Merouane Wastewater Treatment Plant.

## 2. Study area and data used

### 2.1. Description of the wastewater treatment plant process

In the present study, the modelling framework was developed using data collected from the effluent of municipal wastewater treatment plant (WWTP) located at Sidi Marouane town, Mila province, Algeria (Fig. 1). The WWTP is realized for treating more than 20657 $m^3$/day. The WWTP receives wastewater from several sewage stations located near its surface. The plant employs traditional wastewater treatment processes consisting of coarse and fins screens, grit, scum, primary sedimentation tanks, activated sludge aeration tanks, secondary sedimentation tanks, and final clarification and chlorination facilities [12]. A total of 1235 patterns covering the period from August 01, 2009 to July 11, 2013 were collected and divided into training (865) and validation (370) with respect to the ratios of (70 %/30 %). Data collected consists of daily effluent chemical oxygen demand (COD), wastewater temperature ($T_w$), specific conductance (SC), wastewater flow rate ($Q$), total suspended solids (TSS), wastewater pH, and the five-day biochemical oxygen demand ($BOD_5$) (Table 2). In Table 2, we report au summary statistic for all variables used in the present study. Hence, the $BOD_5$ was used as the predicted variable, i.e., the variable to be modelled; while the COD, pH, $T_w$, TSS, SC, and the $Q$, were used as the input variables combined together with respect to 09 input combination (Table 3).

### 2.2. Performance assessment of the models

In the present study, four performances metrics were selected for models comparison and evaluation: the root mean square error (RMSE), mean absolute error (MAE), correlation coefficient (R), and Nash-Sutcliffe efficiency (NSE) applying Equations (1)–(4).

$$MAE = \frac{\sum_{i=1}^{N} \left| BOD_{5pre,i} - BOD_{5obs,i} \right|}{N} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left( BOD_{5obs,i} - BOD_{5pre,i} \right)^2}{N}} \tag{2}$$

$$NSE = 1 - \left[ \frac{\sum_{i=1}^{N} \left( BOD_{5obs,i} - BOD_{5pre,i} \right)^2}{\sum_{i=1}^{N} \left( BOD_{5obs,i} - \overline{BOD_{5obs}} \right)^2} \right] \tag{3}$$

**Table 2**
Summary statistics of the effluent Sidi Merouane municipal WWTP.

| Variables | Subset | Unit | $X_{mean}$ | $X_{max}$ | $X_{min}$ | $S_x$ | $C_v$ | $R$ |
|---|---|---|---|---|---|---|---|---|
| ***Sidi Merouane Wastewater Treatment Plant*** | | | | | | | | |
| $Q$ | Training | $m^3$/day | 3872.72 | 14736.00 | 128.00 | 2308.88 | 0.60 | −0.11 |
| | Validation | $m^3$/day | 3828.70 | 13221.00 | 86.00 | 2266.30 | 0.59 | −0.09 |
| | All data | $m^3$/day | 3859.53 | 14736.00 | 86.00 | 2280.18 | 0.59 | −0.11 |
| $Tw$ | Training | °C | 19.94 | 28.70 | 11.10 | 4.54 | 0.23 | −0.15 |
| | Validation | °C | 19.93 | 27.90 | 9.50 | 4.71 | 0.24 | −0.03 |
| | All data | °C | 19.93 | 28.70 | 9.50 | 4.58 | 0.23 | −0.11 |
| pH | Training | // | 7.66 | 8.50 | 7.00 | 0.30 | 0.04 | 0.25 |
| | Validation | // | 7.68 | 8.52 | 7.01 | 0.31 | 0.04 | 0.28 |
| | All data | // | 7.67 | 8.52 | 7.00 | 0.30 | 0.04 | 0.26 |
| $COD$ | Training | mg/L | 20.69 | 56.30 | 3.50 | 10.84 | 0.52 | 0.63 |
| | Validation | mg/L | 21.86 | 59.50 | 5.30 | 11.44 | 0.52 | 0.69 |
| | All data | mg/L | 21.04 | 59.50 | 3.50 | 10.99 | 0.52 | 0.65 |
| $TSS$ | Training | mg/L | 10.24 | 92.80 | 0.20 | 10.25 | 1.00 | 0.13 |
| | Validation | mg/L | 9.98 | 63.20 | 0.20 | 9.04 | 0.91 | 0.13 |
| | All data | mg/L | 10.16 | 92.80 | 0.20 | 9.79 | 0.96 | 0.13 |
| $SC$ | Training | $\mu$.s/cm | 1673.08 | 2290.00 | 1210.00 | 170.99 | 0.10 | −0.24 |
| | Validation | $\mu$.s/cm | 1665.28 | 2270.00 | 1210.00 | 175.54 | 0.11 | −0.26 |
| | All data | $\mu$.s/cm | 1670.75 | 2290.00 | 1210.00 | 171.20 | 0.10 | −0.24 |
| $BOD_5$ | Training | mg/L | 8.05 | 26.00 | 0.60 | 5.28 | 0.66 | 1.00 |
| | Validation | mg/L | 8.96 | 27.00 | 1.00 | 5.68 | 0.63 | 1.00 |
| | All data | mg/L | 8.32 | 27.00 | 0.60 | 5.40 | 0.65 | 1.00 |

[Abbreviations: $X_{mean}$, mean; $X_{max}$, maximum; $X_{min}$, minimum; $S_x$, standard deviation; $C_v$, coefficient of variation; R, coefficient of correlation with $BOD_5$, $T_w$: wastewater temperature, $SC$: specific conductance, $COD$: chemical oxygen demand, $BOD_5$: five-day biochemical oxygen demand, $Q$: wastewater flow rate, $TSS$: total suspended solids, mg/l: milligrams per liter, $\mu$.s/cm: microsiemens per centimeter].

**Table 3**
The input combinations of different extreme learning machine models.

| MLR | ELM_Bat | MLPNN | RVFL | RFR | GPR | Input combination | Output |
|---|---|---|---|---|---|---|---|
| MLR1 | ELM_Bat1 | MLPNN1 | RVFL1 | RFR1 | GPR1 | Q, Tw, pH, COD, TSS, SC | $BOD_5$ |
| MLR2 | ELM_Bat2 | MLPNN2 | RVFL2 | RFR2 | GPR2 | Tw, pH, COD, TSS, SC | $BOD_5$ |
| MLR3 | ELM_Bat3 | MLPNN3 | RVFL3 | RFR3 | GPR3 | Q, Tw, pH, COD, TSS | $BOD_5$ |
| MLR4 | ELM_Bat 4 | MLPNN4 | RVFL4 | RFR4 | GPR4 | Q, Tw, pH, COD | $BOD_5$ |
| MLR5 | ELM_Bat 5 | MLPNN5 | RVFL5 | RFR5 | GPR5 | Tw, pH, COD, TSS | $BOD_5$ |
| MLR6 | ELM_Bat6 | MLPNN6 | RVFL6 | RFR6 | GPR6 | Q, COD, SC | $BOD_5$ |
| MLR7 | ELM_Bat7 | MLPNN7 | RVFL7 | RFR7 | GPR7 | Q, Tw, COD | $BOD_5$ |
| MLR8 | ELM_Bat8 | MLPNN8 | RVFL8 | RFR8 | GPR8 | Q, SC | $BOD_5$ |
| MLR9 | ELM_Bat9 | MLPNN9 | RVFL9 | RFR9 | GPR9 | Tw, COD | $BOD_5$ |

$$R = \frac{\sum_{i=1}^{N}\left(BOD_{5obs,i} - \overline{BOD_{5obs}}\right)\left(BOD_{5pre,i} - \overline{BOD_{5pre}}\right)}{\sqrt{\sum_{i=1}^{N}\left(BOD_{5obs,i} - \overline{BOD_{5obs}}\right)^2 \sum_{i=1}^{N}\left(BOD_{5pre,i} - \overline{BOD_{5pre}}\right)^2}} \tag{4}$$

$\overline{BOD}_{5obs}$ and $\overline{BOD}_{5pre}$ are the mean measured, and mean predicted five days biochemical oxygen demand, respectively, $BOD_{obs}$ and $BOD_{pre}$ specifies the observed and predicted five days biochemical oxygen demand, and N shows the number of data points.

## 3. Methodology

### 3.1. Bat algorithm optimized extreme learning machine (ELM-bat)

The extreme learning machine (ELM), a successful classification of feedforward neural networks (FFNN), was originally suggested and employed by Huang et al. [22,23] document. It can outline utilizing one hidden layer composed from $L$ hidden neurons, $N$ input variables, and it can be formulated as follow (Equation (5)):

$$f(x) = \sum_{j=1}^{N}\sum_{i=1}^{L}\beta_i g_i\left(\mathrm{w}_i x_j + b\right) \tag{5}$$

Were, $g(.)$ = the activation function used by each hidden neuron in the single hidden layer, and $\beta_i$ is the output connection weights linking the hidden layer and the output layer. The most employed activation functions are the Gaussian and sigmoid functions in the ELM classification. In addition, the Gaussian activation function can be formulated as (Equation (6)):

$$g(x_i) = h(a, c, x_i) = \exp\left(-a\|x_i - c\|^2\right) \tag{6}$$

Where $a$ and $c$ = the activation functions parameters. In addition, the output target variable can be formulated as (Equation (7)).

$$y = \sum_{j=1}^{N}\sum_{i=1}^{L}\beta_i g_i\left(\mathrm{w}_i x_j + b\right) = t + \varepsilon \tag{7}$$

Where $\varepsilon$ = the error. During the training process, the connection weights are stabilized in the ELM classification. Namely, random values are authorized to activation function of neurons directly instead of applying an iterative process for renewing them. The least square method (LSM) can manage the connection weights of output neuron continuously. That is, the approximation error can be reduced by calculating the $\|\mathbf{H\beta} - \mathbf{T}\|^2$ for the connection weight ($\boldsymbol{\beta}$), where $\mathbf{T}$ = the target matrix and $\mathbf{H}$ = the randomized matrix corresponding to the hidden layer.

ELM trains FFNN with one hidden layer in two steps including random feature mapping and linear parameters solving. In the first step, it starts the hidden layer to map the input data into a feature space utilizing the mapping functions, which can be any nonlinear continuous functions. In the second step, the connection weights between hidden and output layers can be solved by minimizing approximation error [22,23]. The previous articles [22–27] suggested and investigated the special description on the ELM model's application. Fig. 2 shows the schematic diagram of ELM in this study.

With the high level of advancement gained during the last few years in the area of metaheuristic optimization, several algorithms have been developed and successfully used for optimizing ML models. These algorithms have helped in solving several complex problems and the improvement of the ML performances was underpinned by a substantial decrease in the running time of the models and the error generated during the training of ML models. Among the metaheuristic optimization algorithms, we use in the present study the Bat algorithm proposed by Yang [28]. Thus, the Bat algorithm is combined with the ELM model and a new hybrid ML model was proposed and called as the ELM-Bat (Fig. 2). The Bat algorithm can be summarized as follow and more details can be found in Refs. [28,29]. The Bats are creatures that have the ability to fly anywhere with an exact and fixed objective: the search of preys. However, this research of preys is based on the echolocation for localization, and an update of the position and speed of each Bat individual in a

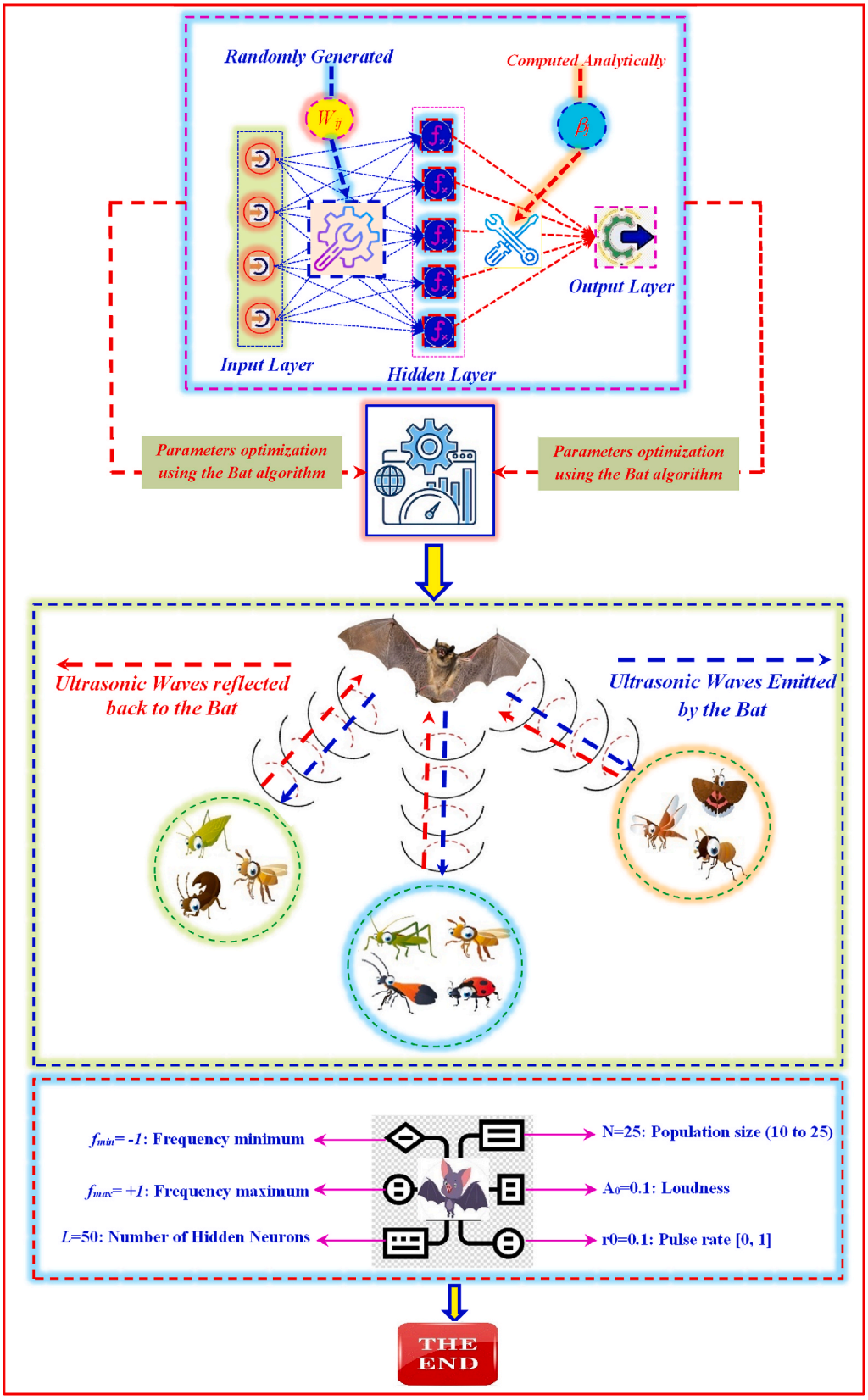



**Fig. 2.** The ELM_Bat architecture.

multidimensional confined escape. Furthermore, the Bats can successfully achieve the hunting operation in total darkness using the ration between the emitting sound and receiving sound despite the distances that separate them, which helped them for recognizing and differencing the food and the barriers. The Bat algorithm is based on the combination of two parts: (*i*) the position of the Bat and (*ii*) the traveling velocity of the Bat. Using these two components, Yang [28] provided a mathematical algorithm for problems optimization, which can be briefly described as follow [30–33]. Using a velocity $Ve_i$ and a location $X_i$, a frequency $Fr$, we can write the following equations (Equations (8)–(10)):

$$Fr_i = Fr_i^{min} + \beta\left(Fr_i^{max} - Fr_i^{min}\right) \tag{8}$$

$$Ve_i^{t+1} = Ve_i^t + Fr_i\left[X_i^t - X_{best}\right] \tag{9}$$

$$X_i^{t+1} = Ve_i^{t+1} + X_i^t \tag{10}$$

In the above equations, the ''$\beta$'' is a random vector in the range of [0, 1]. In addition to this, the $X_{best}$ is the best or optimum position obtained during the iterative process. Therefore, for each iteration, the optimum solution is updated based on randomly step as follow (Equation (11)):

$$X_i^{new} = X_i^{old} + \delta \times A_t^{mean} \tag{11}$$

In the above equation, δ is a random number in the range of [-1, +1], and the $A_t^{mean}$ is the mean loudness calculated for all Bat population. More details about the Bat algorithm can be found in Refs. [30–33]. The Bat was used for hybridizing the ELM model and for better determination of the weights and biases parameters. In the present study, we use the MatLab code of the Bat algorithm is available at: *https://www.mathworks.com/matlabcentral/fileexchange/74768-the-standard-bat-algorithm-ba*.

### 3.2. Multilayer perceptron neural network (MLPNN)

The multilayer perceptron neural networks (MLPNN) keeps an input and output layers including one or additional hidden layers [34,35]. The classification of ML based model involve one hidden layer in the model configuration, while two or more hidden layers combined in the configuration are included in the category of deep learning-based model, respectively [36]. The MLPNN model uses the training dataset for determining the necessary model parameters, i.e., the connection weights and biases from the input to the hidden layers, and from the hidden to the output layers. In addition, the MLPNN uses the backpropagation (BP) training algorithms for obtaining these optimal parameters [37]. From the perspective of gradient calculation and successive adjustment for training parameters, the conjugate gradient backpropagation algorithm (CGBP) [38] was different from the traditional BP algorithm. The route of gradient descent flows down along a course, which is conjugate to the prior stage. The improvement in gradient is considered as orthogonal with function minimization compared with the prior stage [39]. Detailed information on the MLPNN's application and implementation can be found in Refs. [36,40,41]. Fig. 3 represents the schematic diagram of MLPNN in this study.

### 3.3. Gaussian process regression (GPR)

In general, Gaussian process (GP) supposes that Gaussian is applied for joint probability distribution of model output. Since GP can be trained utilizing a matrix access without prior information of function and dataset, GP feature is suitable to provide the solution of complex and nonstationary problems in nature. The hyperparameters in the Bayesian approaches and maximum likelihood can manage the scheme of GP, which brings on training performance of an automatic and pertinent choice [42]. The Gaussian process

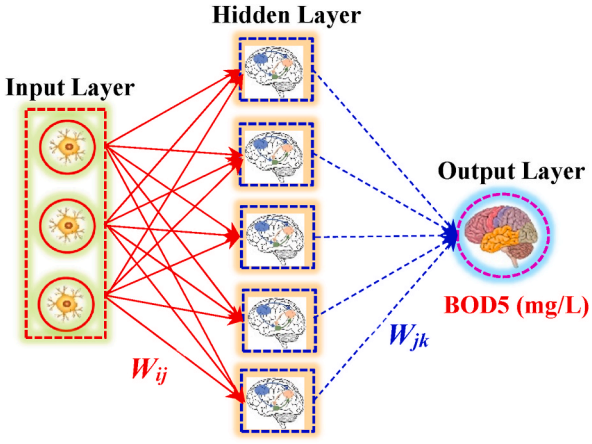



**Fig. 3.** Schematic diagram of MLPNN.

regression (GPR) can be explained as a nonparametric probabilistic approach utilizing the idea of spatial smoothing [43]. GPR has been implemented in the divergent majors including medicine, engineering, and neuroscience fields [44–46]. In addition, GPR can be applied for estimated, forecasted, and predicted questions, and can provide the confidence interval for the matching points, which calculate the predictive uncertainty. In addition, GPR calculates a vector-based mean and covariance instead of occupying a scalar-based mean and variance values based on GP application [47]. Therefore, the benefit of GPR access is invincible to find the missing input data because the training data are merged to calculate the hyperparameters of covariance function. However, since GPR is not conscious for diverse choices of covariance, functions in the modeling process [48]; the implementation of covariance function can make the performance of GPR stronger than other machine learning models. Readers for this article can find the specific description on the development and application of GPR from [25,46,49]. Fig. 4 illustrates the schematic diagram of GPR.

### 3.4. Random forest regression (RFR)

The article of Breiman [50] implemented the random forest based on the solution of regression and classification problems. The random forest regression (RFR) which consists of regression tree (RT) and bootstrap aggregation (Bagging), is a perfect model for solving actual time problems in diverse geophysical fields including environment [24,36,51], hydrometeorology [52], and hydrology [53,54]. RFR carries out the in-built cross-validation processes and training feature utilizing the out-of-bag samples. The training error calculated from the mean square error (MSE) could provide a predictive estimation of RFR's efficiency [55] (Were et al., 2015). In RFR modeling procedure, three parameters including $n_{tree}$ (the number of tree to grow), $m_{try}$ (randomly selected predictor variables), and *nodesize* (the minimal number of observations) are employed. In addition, when the predictors are ignored one by one from RFR, the importance of each predictor is calculated utilizing increased MSE. The relative importance of each predictor can be selected from the specific operation of RFR. The researches of RFR can be found in Refs. [24,50,54] with detailed description. Fig. 5 provides the schematic diagram of RFR.

### 3.5. Random vector functional link (RVFL)

The article of Pao et al. [56] proposed the random vector functional link (RVFL), and the generalization and training of RVFL were discussed in Ref. [57]. Therefore, RVFL has been employed to resolve the nonlinear issues in different fields. Tyukin and Prokhorov [58] explored the background of simulation and modeling, and connected the unsupervised adjustment of neurons to the input variables with subsequent supervised training feature of successive variables. Chi and Ersoy [59] described that RVFL connected enhancement neurons with the statistical hypothesis to develop statistical training feature. Based on the RVFL's modeling and simulation, the activation functions cannot be determined completely when the connection weights are generated randomly relying on the connection weights between the input and enhancement neurons. The research of Alhamdoosh and Wang [60] investigated that all of connection weights were produced utilizing a uniform distribution with [-$S$, +$S$], where $S$ = a scale factor. RVFL can be almost divided into two classes depending on the approach for calculating the output connection weights. In the first class, it can be written as the iterative RVFL, which calculates the connection weights with an iterative feature utilizing the error gradient function. In the second class, it can be described as the closed-pattern RVFL, which calculates the connection weights utilizing a single-stage [61]. Fig. 6 displays the schematic diagram of random vector functional link in this study. The researches of [62–64] described the RVFL's application and employment in detail.

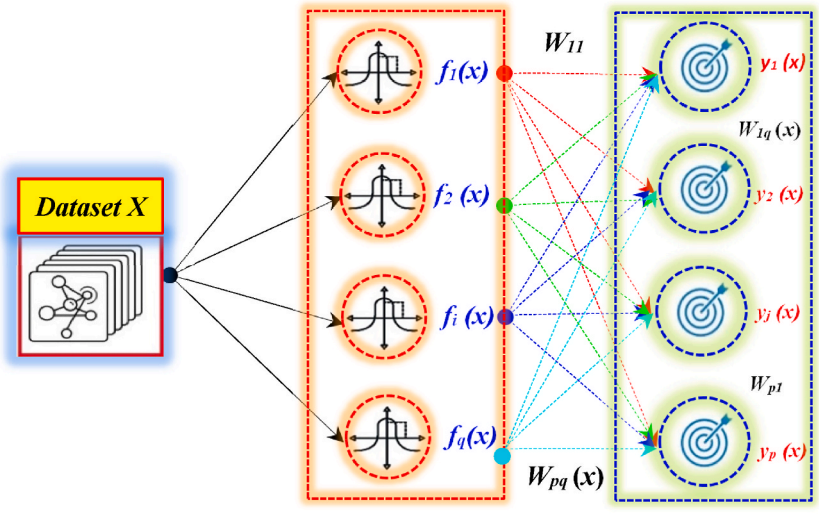



**Fig. 4.** Schematic diagram of GPR.

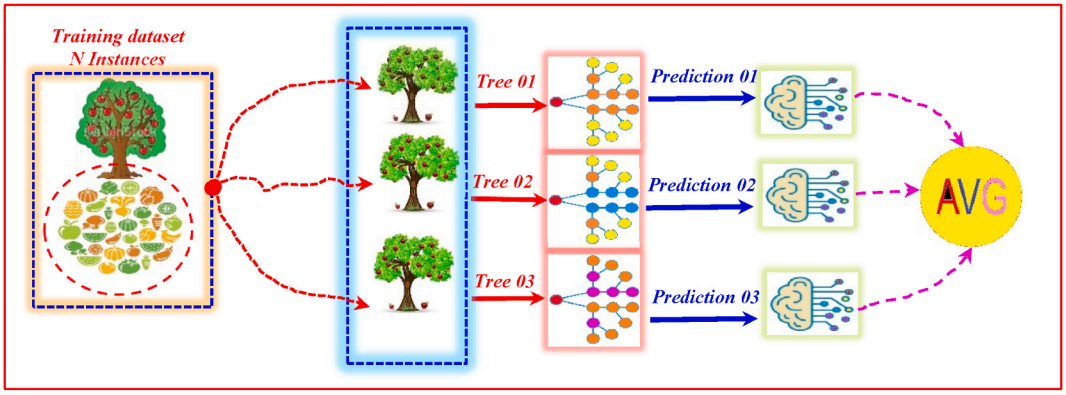



**Fig. 5.** Schematic diagram of RFR.

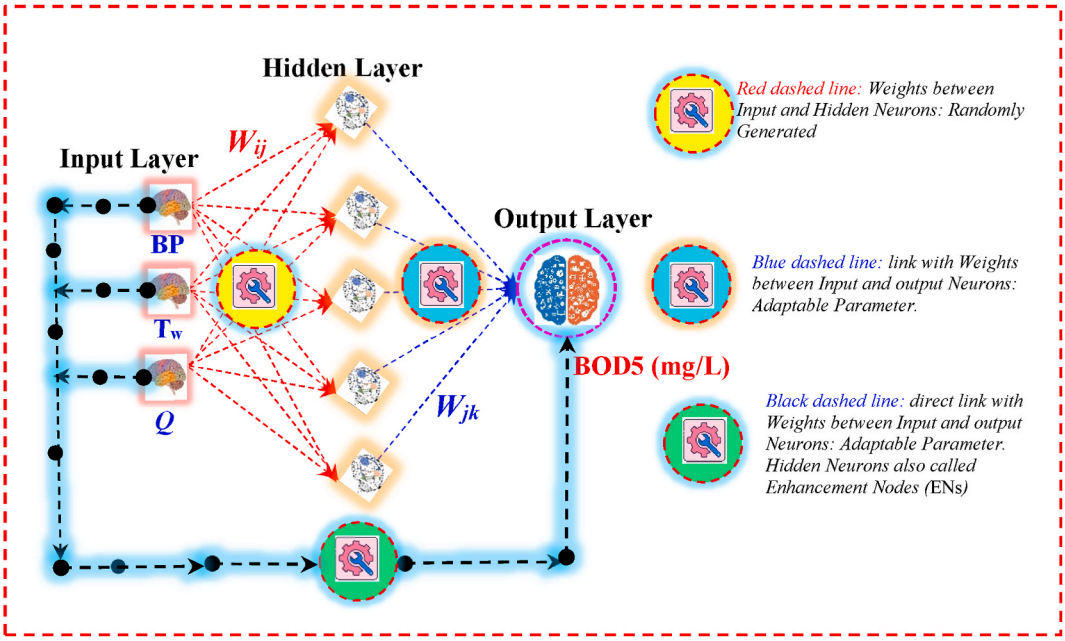



**Fig. 6.** Schematic diagram of Random Vector Functional Link (RVFL).

## 4. Results and discussion

### 4.1. Models evaluation and comparison

In the present study, a new hybrid model based on the combination of extreme learning machine and Bat algortihm (ELM-Bat) was proposed for modelling effluent five days biochemical oxygen demand ($BOD_5$). In order to show the superiority of the proposed ELM-Bat method, five standalone ML models were also tested and used to compare their performances with those of the ELM-Bat. The flowchart of the proposed modelling framework is depicted in Fig. 7. The Matlab code of the ML model is given in the Supplemental Files (Text S1.). They are respectively, the MLPNN, RVFL, GPR, RFR, and the standalone multiple linear regression model (MLR). Furthermore, in order to perform sensitivity analysis faster, the results are experimentally performed taking into account several input combination and in total 09 combinations having different input variables were evaluated and examined in the present study as stated in Table 4, and only the results in the validation stage were analyzed hereafter. It can be seen from Table 4 that when the models use the all six input variables, i.e., the Q, $T_w$, pH, COD, TSS, and SC, the best predictive accuracies were obtained using the ELM_Bat1. More precisely, the ELM_Bat1 exhibited less prediction error and large fitting capability: the RMSE and MAE values were approximately equal the values of ≈2.621 mg/L and 1.989 mg/L, respectively, while the R and NSE values were approximately ≈0.885 and ≈ 0.781, respectively. It was found that, the R-values, i.e., the correlation coefficients, exhibited by the ELM-Bat models are scattered in the range of ≈0.452 (ELM_Bat8) to ≈0.885 (ELM_Bat1) with a mean value of approximately ≈0.802. For the lower R-value likely below

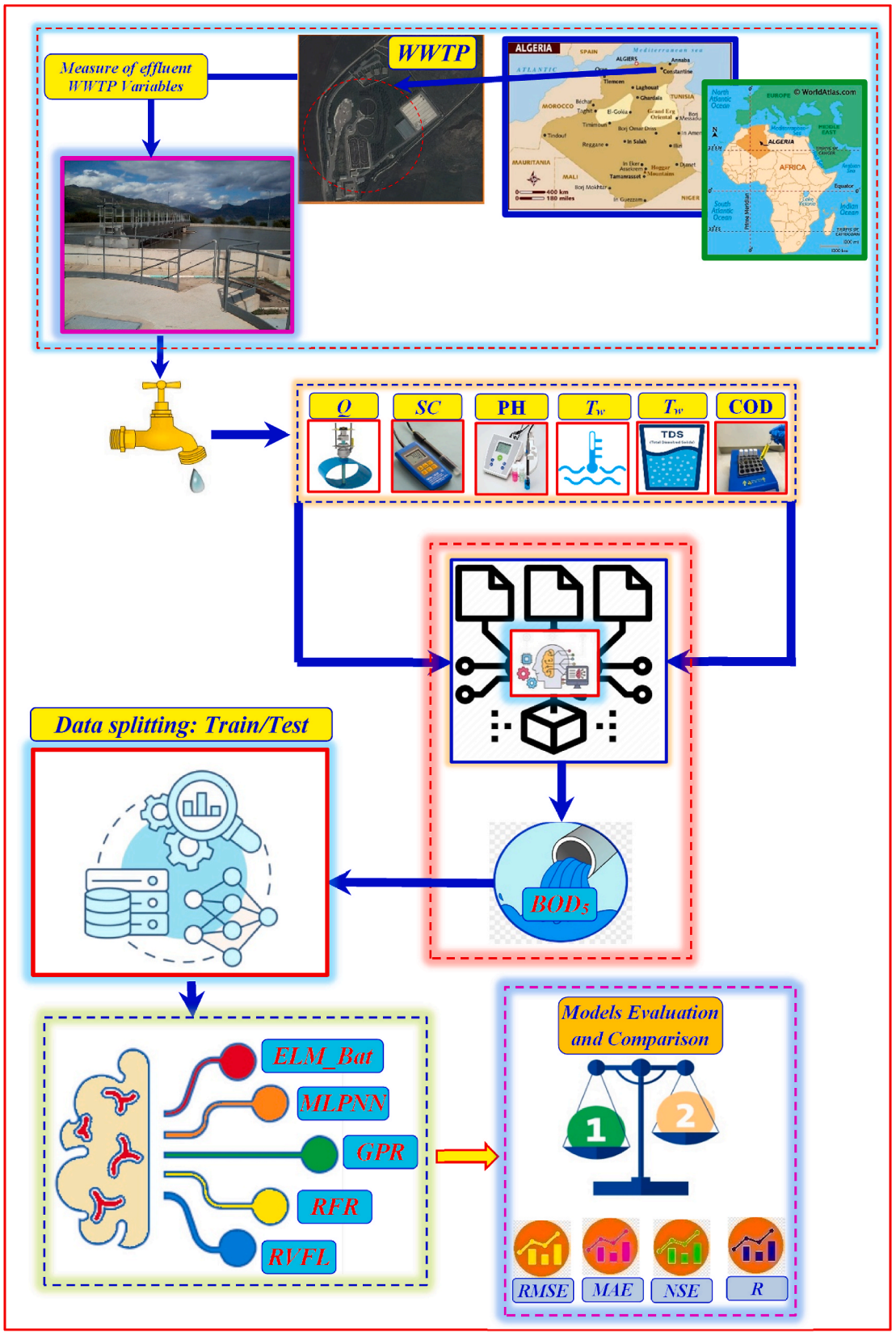



**Fig. 7.** Flowchart of the proposed modelling framework for five days biochemical oxygen demand ($BOD_5$).

**Table 4**
Performances of different machine learning models for $BOD_5$ prediction.

| Models | Training | | | | Validation | | | |
|---|---|---|---|---|---|---|---|---|
| | R | NSE | RMSE | MAE | R | NSE | RMSE | MAE |
| MLR1 | 0.678 | 0.459 | 3.856 | 2.853 | 0.707 | 0.492 | 3.996 | 2.981 |
| MLR2 | 0.661 | 0.436 | 3.937 | 2.913 | 0.709 | 0.492 | 3.997 | 2.990 |
| MLR3 | 0.659 | 0.435 | 3.943 | 2.940 | 0.704 | 0.485 | 4.025 | 2.990 |
| MLR4 | 0.657 | 0.432 | 3.954 | 2.946 | 0.702 | 0.481 | 4.040 | 3.013 |
| MLR5 | 0.646 | 0.418 | 4.001 | 2.979 | 0.707 | 0.486 | 4.019 | 3.001 |
| MLR6 | 0.664 | 0.441 | 3.922 | 2.900 | 0.684 | 0.460 | 4.120 | 3.055 |
| MLR7 | 0.646 | 0.417 | 4.003 | 2.956 | 0.688 | 0.461 | 4.117 | 3.042 |
| MLR8 | 0.277 | 0.077 | 5.038 | 4.050 | 0.288 | 0.102 | 5.437 | 4.203 |
| MLR9 | 0.625 | 0.391 | 4.092 | 3.024 | 0.689 | 0.460 | 4.121 | 3.058 |
| ELM_Bat1 | 0.824 | 0.679 | 2.971 | 2.246 | 0.885 | 0.781 | 2.621 | 1.989 |
| ELM_Bat2 | 0.821 | 0.675 | 2.991 | 2.194 | 0.856 | 0.732 | 2.903 | 2.197 |
| ELM_Bat3 | 0.823 | 0.677 | 2.982 | 2.236 | 0.865 | 0.746 | 2.828 | 2.181 |
| ELM_Bat 4 | 0.813 | 0.661 | 3.053 | 2.286 | 0.829 | 0.687 | 3.138 | 2.392 |
| ELM_Bat 5 | 0.806 | 0.649 | 3.105 | 2.265 | 0.838 | 0.701 | 3.066 | 2.312 |
| ELM_Bat6 | 0.781 | 0.610 | 3.274 | 2.441 | 0.822 | 0.673 | 3.205 | 2.465 |
| ELM_Bat7 | 0.795 | 0.631 | 3.184 | 2.373 | 0.850 | 0.720 | 2.964 | 2.211 |
| ELM_Bat8 | 0.554 | 0.307 | 4.364 | 3.270 | 0.452 | 0.183 | 5.067 | 3.876 |
| ELM_Bat9 | 0.768 | 0.590 | 3.360 | 2.395 | 0.820 | 0.669 | 3.228 | 2.467 |
| MLPNN1 | 0.872 | 0.759 | 2.574 | 1.896 | 0.763 | 0.560 | 3.721 | 2.820 |
| MLPNN2 | 0.865 | 0.748 | 2.634 | 1.928 | 0.746 | 0.541 | 3.797 | 2.795 |
| MLPNN3 | 0.857 | 0.734 | 2.707 | 1.985 | 0.770 | 0.573 | 3.664 | 2.710 |
| MLPNN4 | 0.842 | 0.710 | 2.826 | 2.102 | 0.768 | 0.586 | 3.608 | 2.750 |
| MLPNN5 | 0.831 | 0.690 | 2.921 | 2.111 | 0.770 | 0.587 | 3.603 | 2.724 |
| MLPNN6 | 0.778 | 0.606 | 3.294 | 2.433 | 0.656 | 0.417 | 4.280 | 3.162 |
| MLPNN7 | 0.794 | 0.631 | 3.185 | 2.340 | 0.692 | 0.454 | 4.142 | 2.986 |
| MLPNN8 | 0.507 | 0.256 | 4.522 | 3.441 | 0.370 | 0.100 | 5.318 | 4.085 |
| MLPNN9 | 0.739 | 0.546 | 3.533 | 2.508 | 0.757 | 0.567 | 3.688 | 2.773 |
| RVFL1 | 0.817 | 0.668 | 3.020 | 2.297 | 0.765 | 0.563 | 3.707 | 2.830 |
| RVFL2 | 0.797 | 0.635 | 3.170 | 2.350 | 0.744 | 0.543 | 3.790 | 2.883 |
| RVFL3 | 0.807 | 0.651 | 3.098 | 2.338 | 0.752 | 0.559 | 3.722 | 2.807 |
| RVFL4 | 0.803 | 0.645 | 3.124 | 2.277 | 0.742 | 0.502 | 3.957 | 2.841 |
| RVFL5 | 0.798 | 0.636 | 3.162 | 2.316 | 0.692 | 0.438 | 4.204 | 2.960 |
| RVFL6 | 0.755 | 0.570 | 3.437 | 2.574 | 0.650 | 0.368 | 4.458 | 3.290 |
| RVFL7 | 0.661 | 0.436 | 3.938 | 2.896 | 0.699 | 0.486 | 4.021 | 3.053 |
| RVFL8 | 0.328 | 0.106 | 4.959 | 3.957 | 0.342 | 0.114 | 5.277 | 4.217 |
| RVFL9 | 0.666 | 0.443 | 3.915 | 2.852 | 0.727 | 0.522 | 3.876 | 2.963 |
| RFR1 | 0.914 | 0.813 | 2.269 | 1.592 | 0.790 | 0.614 | 3.484 | 2.673 |
| RFR2 | 0.910 | 0.807 | 2.301 | 1.610 | 0.784 | 0.609 | 3.505 | 2.644 |
| RFR3 | 0.908 | 0.805 | 2.315 | 1.620 | 0.795 | 0.621 | 3.451 | 2.656 |
| RFR4 | 0.905 | 0.802 | 2.334 | 1.623 | 0.793 | 0.621 | 3.453 | 2.644 |
| RFR5 | 0.902 | 0.796 | 2.367 | 1.629 | 0.782 | 0.606 | 3.520 | 2.662 |
| RFR6 | 0.870 | 0.734 | 2.706 | 1.941 | 0.695 | 0.483 | 4.033 | 3.082 |
| RFR7 | 0.872 | 0.739 | 2.679 | 1.898 | 0.750 | 0.556 | 3.736 | 2.812 |
| RFR8 | 0.770 | 0.535 | 3.577 | 2.693 | 0.375 | 0.122 | 5.255 | 4.064 |
| RFR9 | 0.846 | 0.705 | 2.850 | 1.997 | 0.739 | 0.545 | 3.781 | 2.800 |
| GPR1 | 0.823 | 0.676 | 2.987 | 2.245 | 0.795 | 0.629 | 3.416 | 2.613 |
| GPR2 | 0.814 | 0.660 | 3.059 | 2.239 | 0.784 | 0.612 | 3.493 | 2.645 |
| GPR3 | 0.804 | 0.644 | 3.129 | 2.330 | 0.795 | 0.631 | 3.406 | 2.595 |
| GPR4 | 0.784 | 0.613 | 3.262 | 2.428 | 0.786 | 0.616 | 3.476 | 2.686 |
| GPR5 | 0.810 | 0.653 | 3.090 | 2.244 | 0.784 | 0.614 | 3.486 | 2.627 |
| GPR6 | 0.726 | 0.527 | 3.608 | 2.647 | 0.705 | 0.497 | 3.979 | 3.050 |
| GPR7 | 0.761 | 0.577 | 3.410 | 2.515 | 0.734 | 0.538 | 3.813 | 2.869 |
| GPR8 | 0.414 | 0.170 | 4.778 | 3.712 | 0.368 | 0.135 | 5.215 | 4.124 |
| GPR9 | 0.705 | 0.496 | 3.722 | 2.660 | 0.751 | 0.561 | 3.717 | 2.814 |

≈0.450 obtained using the ELM_Bat8, it might be whether the training is unsuccessful or the input variables does not helped in a better simulation of the $BOD_5$. When we have a look to the all developed algorithms in the present study, the lowest R-value was achieved using the same input combination, i.e., using the effluent flow (*Q*) and the specific conductance (*SC*) for which the values of ≈0.288, ≈0.452, ≈0.370, ≈0.342, ≈0.375, and ≈0.368 were obtained using the MLR8, ELM_Bat8, MLPNN8, RVFL8, RFR8 and GPR8, respectively. Therefore, we can conclude that, the combination of the Q and SC is not a good solution for modelling $BOD_5$. Finally, overall comparison between the models based on the four numerical performances reported in Tables 4 and it is obvious that the number of input variables increase leads to a significant improvement with an increase of the R and NSE values, and a decrease of the RMSE and MAE values of the ELM_Bat1. More precisely, when the ELM_Bat1 is compared to the others benchmarks models, the RMSE and MAE values were decreased by 34.40 % and 33.27 %, 29.56 % and 29.46 %, 29.29 % and 29.71 %, 24.77 % and 25.58 %, 23.27 % and 23.88 %, respectively. It is clear from the above results that, the GPR1 was ranked in the second rank after the ELM_Bat1, while the

MLR1 was ranked in the last rank with the highest poorest performances.

As shown in Table 4, for scenarios 2 and 3, for which only five input variables were included in the models, the ELM_Bat2 and ELM_Bat3 exhibited the lowest RMSE and MAE, and the highest R and NSE values compared to the all other algorithms. For numerical comparison, the ELM_Bat3 exhibited R, NSE, RMSE, and MAE values of approximately ≈0.865, ≈0.746, ≈2.828 mg/L, and ≈2.181 mg/L, respectively, while the RVFL3 was found to be the less accurate ML model having R, NSE, RMSE, and MAE values of approximately ≈0.752, ≈0.559, ≈3.722 mg/L, and ≈2.807 mg/L, respectively, ranked just after the MLR3 who exhibited the lowest numerical performances. It is demonstrated in Table 4 that, the numerical values of the four performances metrics obtained using the ELM-Bat are higher that other models regardless of the input variables and especially, the models prediction errors are the smallest. However, an important point to be noted is that, the configuration of the RFR, MLPNN, GPR and RVFL models takes less computational time compared to the ELM-Bat models.

For the models having only four input variables and taking into account the four metrics, i.e., the R, NSE, RMSE and MAE on the validation dataset are illustrated in Table 4, we can see these that the RMSE and MAE are the lowest for the ELM-Bat 4 and ELM-Bat 5, while the R and NSE are the highest among all developed models. As shown in Table 4, the RMSE and MAE values of the ELM_Bat5 are 3.066 mg/L and 2.312 mg/L, respectively, which are all the lowest of the six prediction models (i.e., the models having the same input variables). The RMSE value of the ELM_Bat5 model decreases by more than 23.71 %, 14.90 %, 22.51 %, 11.20 %, and 11.795 % compared to the MLR5, MLPNN5, RVFL4, RFR4 and GPR4, respectively. In addition to this, the MAE value of the ELM_Bat5 model has particularly large decrease from 3.001 mg/L to 2.312 mg/L (22.96 %) compared to the MLR5, from 2.724 mg/L to 2.312 mg/L (15.125

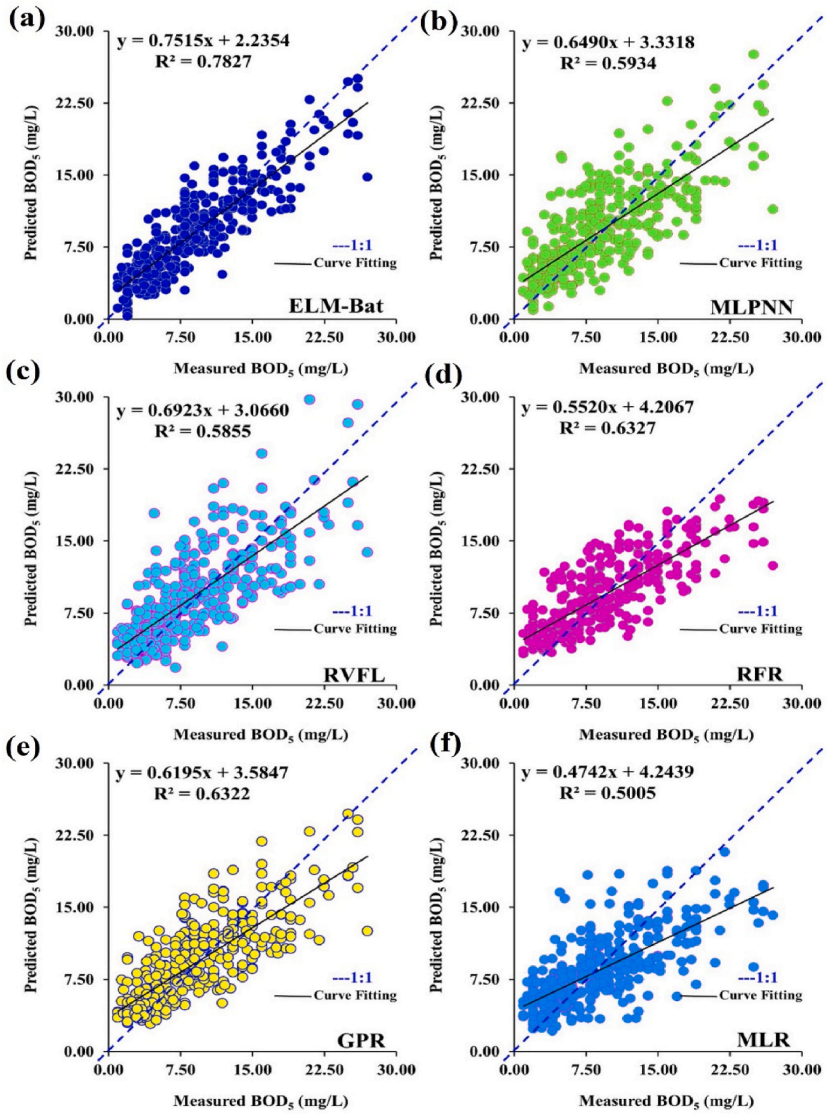



**Fig. 8.** Scatterplot of measured against predicted five days biochemical oxygen demand ($BOD_5$) using the best machine learning models for the validation stage: (a) ELM-bat, (b) MLPNN, (c) RVFL, (d) RFR, (e) GPR, and (f) MLR.

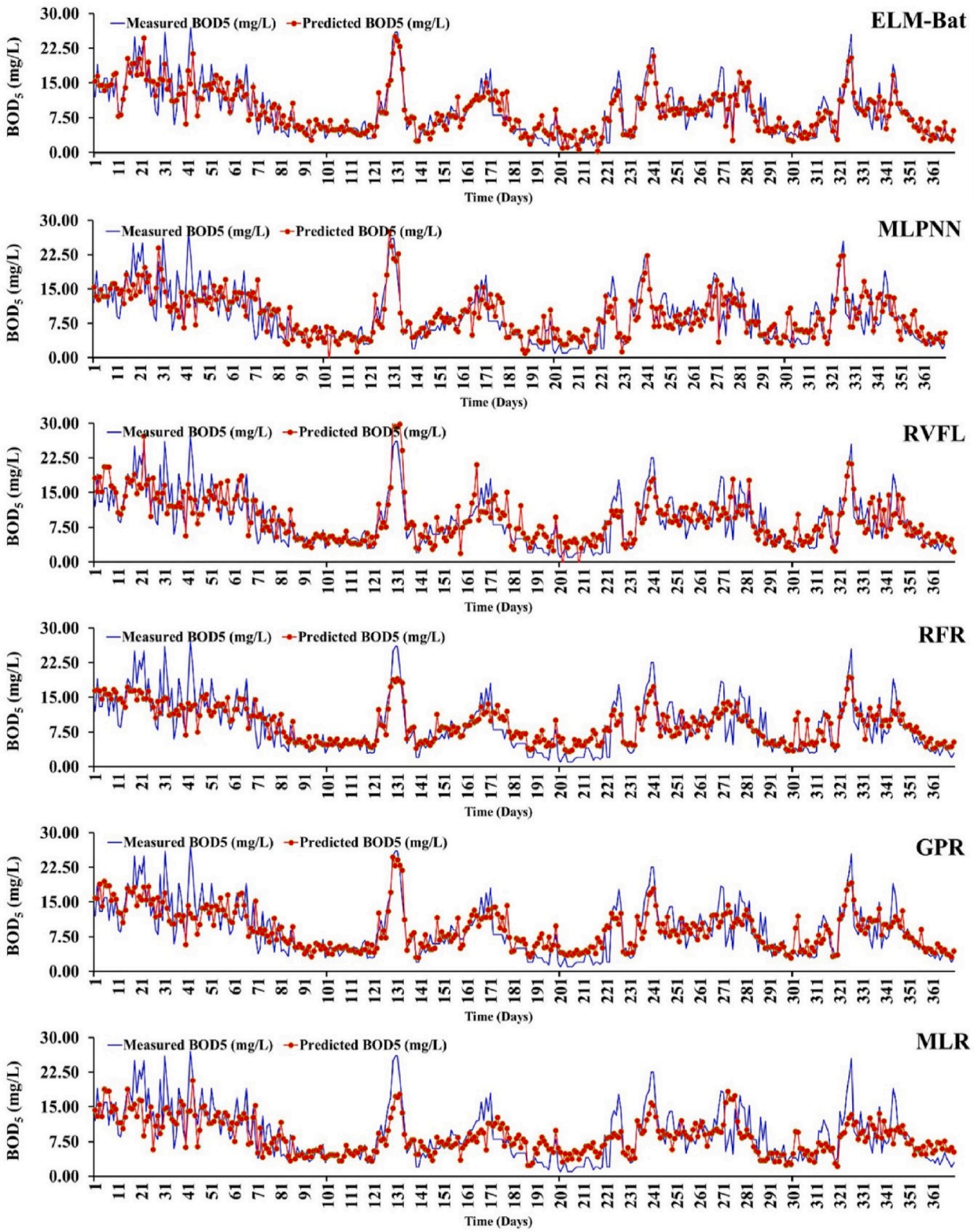



**Fig. 9.** Comparison between measured and predicted five days biochemical oxygen demand ($BOD_5$) using the best machine learning models: validation stage.

%) compared to the MLPNN5, from 2.841 mg/L to 2.312 mg/L (18.62 %) compared to the RVFL5, from 2.644 mg/L to 2.312 mg/L (12.55 %) compared to the RFR5, and from 2.686 mg/L to 2.312 mg/L (13.92 %) compared to the GPR5. Further comparison between the models as reported in Table 4 revealed that, based only on three input variables (i.e., Q, $T_w$ and COD), the R and NSE values of the ELM_Bat7 model (R = 0.850, NSE = 0.720) are higher than other models, showing that its model fitting performance is the best. In Table 4, the numerical values of RMSE and MAE of ELM_Bat7 model (RMSE = 2.96, MAE = 2.21) are generally less than the benchmark methods showing it high capability in reducing the error between the measured and predicted $BOD_5$. In conclusion, the ELM_Bat7 model has the best predictive performance and the relatively better fitting capability over the six models.

It can be seen from Fig. 8 that, based on the scatterplot of measured and predicted data, the ELM_Bat model was characterized by less scattered data compared to the all other models, while the MLR model was the model for which the data were highly scattered exhibiting the poorest predictive accuracies. Furthermore, it can be seen from Fig. 9 that the fluctuation trend of the red line (i.e., the calculated data) is relatively close to the fluctuation trend of the blue line (i.e., the measured data) than all other curves. Thus, it is more remarkable that, the blue and red lines have the biggest degree of superposition and the closet fluctuation trend.

### 4.2. Discussion

In the present study, the most effective ELM_Bat model is chosen through a selection procedure among several machine-learning models developed and compared to improve the accuracy of the $BOD_5$ prediction at the municipal wastewater treatment plant (WWTP) located at Sidi Marouane, Algeria. The most significant contribution of the present study is the hybridization of the ELM using the Bat algorithm, and as a result, the performances of the ELM_Bat was improved. However, although the superiority of the ELM_Bat was clearly demonstrated, the following questions will immediately arise: (*i*) How will this impact the future of $BOD_5$ modelling? (*ii*) At each level of success our approach can be compared to what is already published in the literature? (*iii*) How many input variables are necessary in the model? (*iv*) What combination of the water quality variables is most appropriate?

First, it is clear from the obtained results that the inclusion of the all input variables helped in obtaining the best predictive accuracies, i.e., the use of the Q, $T_w$, pH, COD, TSS, and SC. However, it will more suitable if the ELM_Bat model was evaluated using other measured effluent water quality variables, not taken into account in the present study, which certainly impacted the accuracy of $BOD_5$ modelling. Second, regarding the number of variables necessary for obtaining high predictive accuracy, we can argue as follow. From the best model having six input variables (the ELM_Bat1) to the model having five input variables (ELM_Bat3), for which the SC was excluded from the input variables, the overall performances were slightly decreased: (*i*) the R and NSE were dropped from 0.885 to 0.781 to 0.865 and 0.746, showing a slightly decrease, while the RMSE and MAE values have started to rise again. The RMSE saw their level climb to 2.828 mg/L (+7.32 %) while the level for MAE metric edged up to 2.181 mg/L (+7.32 %). In the other hand, regarding the models having only three input variables (*Q, $T_w$, COD*), the decrease in the model performances was more pronounced for which

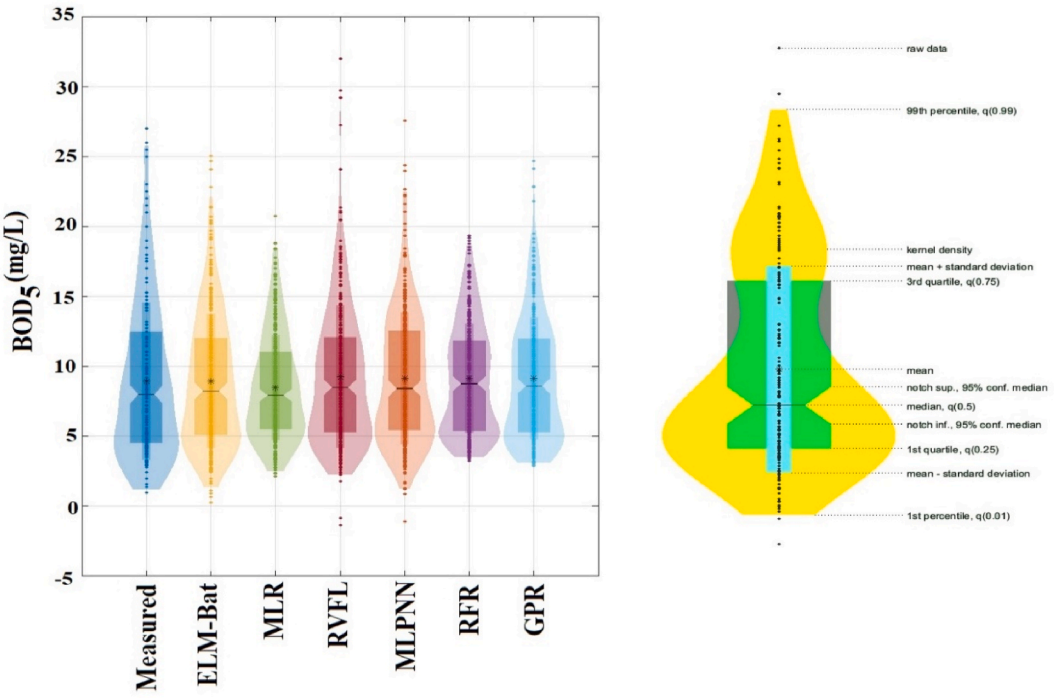



**Fig. 10.** Violin plots of the measured and predicted five days biochemical oxygen demand ($BOD_5$) during validation stage for all machine leaning models.

the RMSE and MAE of the ELM_Bat1 were increased by approximately ≈11.57 % and ≈10.041 %, respectively, but if we consider that, a percentage of approximately 11 % is acceptable, we can declare that the ELM_Bat1 is an excellent model with only fewer input variables and combining the *Q, $T_w$, COD* is an excellent choice. The superiority of the ELM_Bat1 was further highlighted using the boxplot and violinplot depicted in Figs. 10 and 11.

Finally, we compare briefly our results with those reported in the literature. For example, the R-value obtained in the present study using the ELM_Bat1 (i.e., R = 0.885) is less than the value obtained using the decision tree (DT) model (i.e., R = 0.970) and less than the value obtained using RFR (i.e., R = 0.959) reported in the study of Qambar and Al Khalidy [2,65], and higher than the value obtained using the adaptive boosting algorithm (i.e., R = 0.707), and higher than the value obtained using the gradient boosting (GB) (i.e., R = 0.812). By comparison to another investigation conducted by Aghdam et al. [15], for which the GEP (R ~ 0.865), RFR (R ~ 0.841), KNN (R ~ 0.822), MLR (R ~ 0.818), MLPNN (R ~ 0.809), GB (R ~ 0.799), and RT (R ~ 0.752) based models were applied and compared for modelling $BOD_5$, we can see that our ELM_Bat1 model was more accurate compared to the all previous machine learning models. The ELM_Bat1 developed in the present study was found to be more accurate compared to the ANFIS model applied by Ahmed et al. [66], who obtained an R-value of approximately (R ~ 0.830). In another study, Solgi et al. [67] have obtained high numerical performances using the SVR and ANFIS models for modelling $BOD_5$ with R values of approximately (R ~ 0.918) and (R ~ 0.910) slightly higher than the R value obtained in our study (i.e., R ~ 0.885). Further comparison with already published works revealed that, the ELM_Bat1 we more accurate compared to the M5Tree and RFR models developed by Golabi et al. [68], i.e. R ~ 0.885 compared to the value of R ~ 0.751 and R ~ 0.872, respectively. In a study conducted by Najafzadeh and Ghaemi [69], good predictive accuracies were obtained using the LSSVM (R ~ 0.850), MARS (R ~ 0.790), MLPNN (R ~ 0.740), ANFIS (R ~ 0.810), and the MLR (R ~ 0.780) which were all lower than the ELM_Bat1 (i.e., R ~ 0.885) developed in our present study. The obtained results were further highlighted using the violinplot (Fig. 10) [70] and the boxplot (Fig. 11), for which the superiority of the ELM_Bat was obvious.

## 5. Conclusion

In this work, a new method for predicting five days biochemical oxygen demand ($BOD_5$) based on extreme learning machine and Bat algorithm (ELM-Bat) has been developed. Using in situ measured data, it was found that the proposed method presents some advantages relative to other tested machine learning models, i.e., the MLPNN, RFR, GPR, RVFL, and the MLR. The $BOD_5$ time's series can be very accurately simulated and the nonlinearity between the wastewater quality variables and the $BOD_5$ can be easily captured. With respect to the other methods, the ELM-Bat correctly and accurately predicts the $BOD_5$ taking into account various input combination of water quality variables. Moreover, it was found that, the ELM-Bat is flexible and provides sufficient predictive accuracies using only two input variables where the others models have failed. For example, using only wastewater temperature ($T_w$) and chemical oxygen demand (COD), the ELM-Bat exhibited R, NSE, RMSE, and MAE values of approximately 0.820, 0.669, 3.228 and 2.467, respectively. Among all analyzed cases, the proposed hybrid ELM-Bat algorithm identified and extracted accurately and properly the $BOD_5$. However, an exception for one case was registered for which the predictive accuracies were found to be very low, i.e., using only the effluent flow (*Q*) and the specific conductance (SC); the R, NSE, RMSE, and MAE values were approximately equal to 0.452, 0.183, 5.067 and 3.876, respectively. Finally, we can argue that, the better performances of the ELM-Bat method in modelling effluent $BOD_5$ indicates it's considerably improved reliability and robustness. Future research may be focused on exploring some others optimization algorithms and then develop the predictive models based on the combination of the standalone machine learning and metaheuristic algorithms.

### Ethical approval

Not applicable.

### Consent to participate

Not applicable.

### Consent to publish

All the authors have declared their consent to publish the manuscript.

### Funding

Not applicable.

### Institutional review board statement

Not applicable.

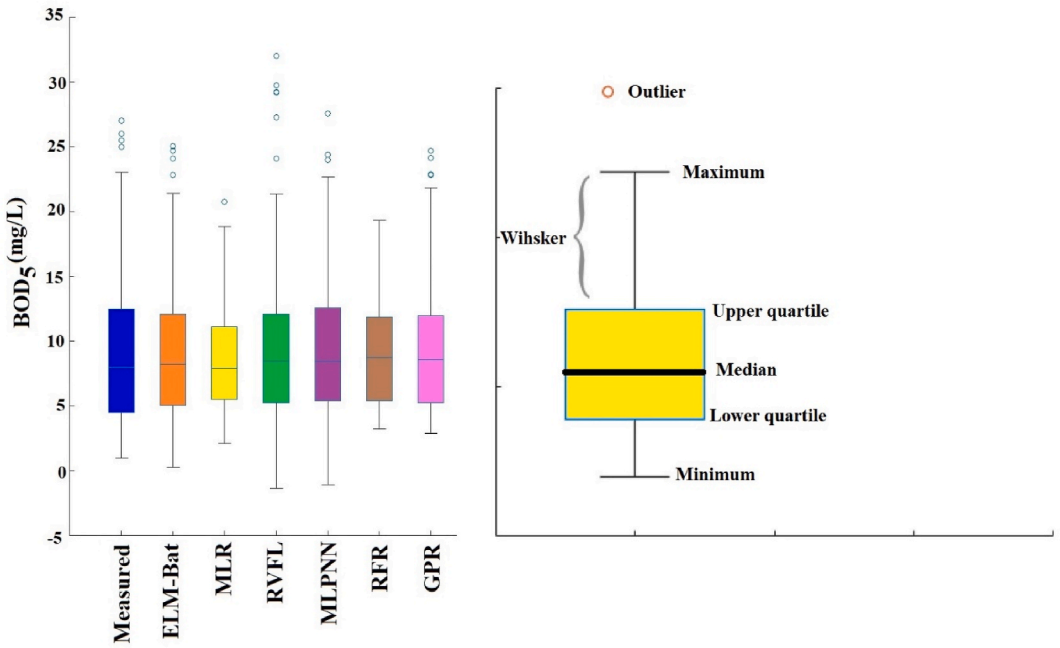



**Fig. 11.** Boxplots illustrating the overall accuracies of the hybrid ELM-Bat and the standalone machine models for the validation stage dataset.

## 9. Informed consent statement

Not applicable.

## Data availability statement

The authors do not have permission to share data.

## CRediT authorship contribution statement

**Hayat Mekaoussi:** Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Writing – original draft. **Salim Heddam:** Conceptualization, Data curation, Formal analysis, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Nouri Bouslimanni:** Formal analysis, Writing – original draft. **Sungwon Kim:** Supervision, Validation, Writing – original draft. **Mohammad Zounemat-Kermani:** Supervision, Validation, Visualization, Writing – original draft.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e21351.

## References

[1] M.S. Jami, I.A. Husain, N.A. Kabashi, N. Abdullah, Multiple inputs artificial neural network model for the prediction of wastewater treatment plant performance, Australian Journal of Basic and Applied Sciences 6 (1) (2012) 62–69, https://doi.org/10.2316/P.2011.736-050.
[2] A.S. Qambar, M.M. Al Khalidy, Optimizing dissolved oxygen requirement and energy consumption in wastewater treatment plant aeration tanks using machine learning, J. Water Proc. Eng. 50 (2022), 103237, https://doi.org/10.1016/j.jwpe.2022.103237.
[3] P.M.L. Ching, X. Zou, D. Wu, R.H.Y. So, G.H. Chen, Development of a wide-range soft sensor for predicting wastewater BOD5 using an eXtreme gradient boosting (XGBoost) machine, Environ. Res. 210 (2022), 112953, https://doi.org/10.1016/j.envres.2022.112953.
[4] A.E.D. Mahmoud, N.A. Khan, Y.T. Hung, Advances in artificial intelligence applications in sustainable water remediation, in: Artificial Intelligence and Modeling for Water Sustainability, Global Challenges, 2023, pp. 53–69, https://doi.org/10.1201/9781003260455-4.

[5] M. Nasr, A.E.D. Mahmoud, M. Fawzy, A. Radwan, Artificial intelligence modeling of cadmium (II) biosorption using rice straw, Appl. Water Sci. 7 (2017) 823–831, https://doi.org/10.1007/s13201-015-0295-x.

[6] M. Kheimi, M.A. Almadani, M. Zounemat-Kermani, Simulating wastewater treatment plants for heavy metals using machine learning models, Arabian J. Geosci. 15 (17) (2022) 1458, https://doi.org/10.1007/s12517-022-10736-9.

[7] M. Zounemat-Kermani, D. Stephan, R. Hinkelmann, Multivariate NARX neural network in prediction gaseous emissions within the influent chamber of wastewater treatment plants, Atmos. Pollut. Res. 10 (6) (2019) 1812–1822, https://doi.org/10.1016/j.apr.2019.07.013.

[8] N.D. Viet, A. Jang, Machine learning-based real-time prediction of micropollutant behaviour in forward osmosis membrane (waste) water treatment, J. Clean. Prod. (2023), 136023, https://doi.org/10.1016/j.jclepro.2023.136023.

[9] R. Noori, S. Safavi, S.A.N. Shahrokni, A reduced-order adaptive neuro-fuzzy inference system model as a software sensor for rapid estimation of five-day biochemical oxygen demand, J. Hydrol. 495 (2013) 175–185, https://doi.org/10.1016/j.jhydrol.2013.04.052.

[10] M. Zounemat-Kermani, M. Alizamir, B. Keshtegar, O. Batelaan, R. Hinkelmann, Prediction of effluent arsenic concentration of wastewater treatment plants using machine learning and kriging-based models, Environ. Sci. Pollut. Control Ser. 29 (14) (2022) 20556–20570, https://doi.org/10.1007/s11356-021-16916-6.

[11] J. Qiao, W. Li, H. Han, Soft computing of biochemical oxygen demand using an improved T-S fuzzy neural network, Chin. J. Chem. Eng. 22 (11–12) (2014) 1254–1259, https://doi.org/10.1016/j.cjche.2014.09.023.

[12] S. Heddam, H. Lamda, S. Filali, Predicting effluent biochemical oxygen demand in a wastewater treatment plant using generalized regression neural network based approach: a comparative study, Environmental Processes 3 (2016) 153–165, https://doi.org/10.1007/s11356-021-16916-6.

[13] P. Yu, J. Cao, V. Jegatheesan, X. Du, A real-time BOD estimation method in wastewater treatment process based on an optimized extreme learning machine, Appl. Sci. 9 (3) (2019) 523, https://doi.org/10.3390/app9030523.

[14] A. Alsulaili, A. Refaie, Artificial neural network modeling approach for the prediction of five-day biological oxygen demand and wastewater treatment plant performance, Water Supply 21 (5) (2021) 1861–1877, https://doi.org/10.2166/ws.2020.199.

[15] E. Aghdam, S.R. Mohandes, P. Manu, C.M. Cheung, A. Yunusa-Kaltungo, T. Zayed, Predicting quality parameters of wastewater treatment plants using artificial intelligence techniques, J. Clean. Prod. (2023), 137019, https://doi.org/10.1016/j.jclepro.2023.137019.

[16] K.P. Oliveira-Esquerre, M. Mori, R.E. Bruns, Simulation of an industrial wastewater treatment plant using artificial neural networks and principal components analysis, Braz. J. Chem. Eng. 19 (2002) 365–370, https://doi.org/10.1590/S0104-66322002000400002.

[17] M.M. Hamed, M.G. Khalafallah, E.A. Hassanien, Prediction of wastewater treatment plant performance using artificial neural networks, Environ. Model. Software 19 (10) (2004) 919–928, https://doi.org/10.1016/j.envsoft.2003.10.005.

[18] G. Onkal-Engin, I. Demir, S.N. Engin, Determination of the relationship between sewage odour and BOD by neural networks, Environ. Model. Software 20 (7) (2005) 843–850, https://doi.org/10.1016/j.envsoft.2004.04.012.

[19] E. Dogan, A. Ates, E.C. Yilmaz, B. Eren, Application of artificial neural networks to estimate wastewater treatment plant inlet biochemical oxygen demand, Environ. Prog. 27 (4) (2008) 439–446, https://doi.org/10.1002/ep.10295.

[20] Y. Azimi, M. Talaeian, H. Sarkheil, R. Hashemi, R. Shirdam, Developing an evolving multi-layer perceptron network by genetic algorithm to predict full-scale municipal wastewater treatment plant effluent, J. Environ. Chem. Eng. 10 (5) (2022), 108398, https://doi.org/10.1016/j.jece.2022.108398.

[21] P. Chang, L. Zhao, F. Meng, Y. Xu, Soft measurement of effluent index in sewage treatment process based on overcomplete broad learning system, Appl. Soft Comput. 115 (2022), 108235, https://doi.org/10.1016/j.asoc.2021.108235.

[22] G.B. Huang, L. Chen, C.K. Siew, Universal approximation using incremental constructive feedforward networks with random hidden nodes, IEEE Trans. Neural Network. 17 (4) (2006) 879–892, https://doi.org/10.1109/TNN.2006.875977.

[23] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1–3) (2006) 489–501, https://doi.org/10.1016/j.neucom.2005.12.126.

[24] M. Alizamir, S. Heddam, S. Kim, A.D. Mehr, On the implementation of a novel data-intelligence model based on extreme learning machine optimized by bat algorithm for estimating daily chlorophyll-a concentration: case studies of river and lake in USA, J. Clean. Prod. 285 (2021), 124868, https://doi.org/10.1016/j.jclepro.2020.124868.

[25] S. Difi, Y. Elmeddahi, A. Hebal, V.P. Singh, S. Heddam, S. Kim, O. Kisi, Monthly streamflow prediction using hybrid extreme learning machine optimized by bat algorithm: a case study of Cheliff watershed, Algeria, Hydrol. Sci. J. 68 (2) (2023) 189–208, https://doi.org/10.1080/02626667.2022.2149334.

[26] Y.D. Zhang, G. Zhao, J. Sun, X. Wu, Z.H. Wang, H.M. Liu, J. Li, Smart pathological brain detection by synthetic minority oversampling technique, extreme learning machine, and Jaya algorithm, Multimed. Tool. Appl. 77 (2018) 22629–22648, https://doi.org/10.1007/s11042-017-5023-0.

[27] S. Heddam, S. Kim, A.D. Mehr, M. Zounemat-Kermani, M. Ptak, A. Elbeltagi, A. Malik, Y. Tikhamarine, Bat algorithm optimised extreme learning machine (Bat-ELM): a novel approach for Daily River water temperature modelling, Geogr. J. 189 (1) (2023) 78–89, https://doi.org/10.1111/geoj.12478.

[28] X.S. Yang, A new metaheuristic bat-inspired algorithm, in: *Nature inspired Cooperative* 1086 *Strategies for Optimization (NICSO 2010)*, Springer, Berlin, Heidelberg, 2010, pp. 65–74, https://doi.org/10.1007/978-3-642-12538-6_6.

[29] X.S. Yang, X. He, Bat algorithm: literature review and applications, Int. J. Bio-Inspired Comput. 5 (3) (2013) 141–149, https://doi.org/10.1504/IJBIC.2013.055093.

[30] R. Chawla, S.M. Beram, C.R. Murthy, T. Thiruvenkadam, N.P.G. Bhavani, R. Saravanakumar, P.J. Sathishkumar, Brain tumor recognition using an integrated bat algorithm with a convolutional neural network approach, Measurement: Sensors 24 (2022), 100426, https://doi.org/10.1016/j.measen.2022.100426.

[31] T. Vu-Huu, S. Pham-Van, Q.H. Pham, T. Cuong-Le, An improved bat algorithms for optimization design of truss structures, in: Structures (Vol. 47, Pp. 2240-2258), Elsevier, 2023, https://doi.org/10.1016/j.istruc.2022.12.033.

[32] X. Hu, W. Jiang, X. Ying, M. Eslami, The application of a new design of bat optimizer for energy efficiency enhancement in PEMFCs based on fractional order theory, Sustain. Energy Technol. Assessments 55 (2023), 102904, https://doi.org/10.1016/j.seta.2022.102904.

[33] Z.M. Ali, T. Alquthami, S. Alkhalaf, H. Norouzi, S. Dadfar, K. Suzuki, Novel hybrid improved bat algorithm and fuzzy system based MPPT for photovoltaic under variable atmospheric conditions, Sustain. Energy Technol. Assessments 52 (2022), 102156, https://doi.org/10.1016/j.seta.2022.102156.

[34] S. Kim, V.P. Singh, Y. Seo, Evaluation of pan evaporation modeling with two different neural networks and weather station data, Theor. Appl. Climatol. 117 (1) (2014) 1–13, https://doi.org/10.1007/s00704-013-0985-y.

[35] S.H. Wang, Y. Zhang, Y.J. Li, W.J. Jia, F.Y. Liu, M.M. Yang, Y.D. Zhang, Single slice based detection for Alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization, Multimed. Tool. Appl. 77 (2018) 10393–10417, https://doi.org/10.1007/s11042-016-4222-4.

[36] S. Kim, M. Alizamir, M. Zounemat-Kermani, O. Kisi, V.P. Singh, Assessing the biochemical oxygen demand using neural networks and ensemble tree approaches in South Korea, J. Environ. Manag. 270 (2020), 110834, https://doi.org/10.1016/j.jenvman.2020.110834.

[37] H. Adeli, S.L. Hung, Machine Learning: Neural Networks, Genetic Algorithms, and Fuzzy Systems, John Wiley & Sons, Inc, 1994.

[38] R. Fletcher, C.M. Reeves, Function minimization by conjugate gradients, Comput. J. 7 (2) (1964) 149–154, https://doi.org/10.1093/comjnl/7.2.149.

[39] J.P. Fitch, S.K. Lehman, F.U. Dowla, S.Y. Lu, E.M. Johansson, D.M. Goodman, Ship wake-detection procedure using conjugate gradient trained artificial neural networks, IEEE Trans. Geosci. Rem. Sens. 29 (5) (1991) 718–726, https://doi.org/10.1109/36.83986.

[40] S. Kim, J. Shiri, O. Kisi, V.P. Singh, Estimating daily pan evaporation using different data-driven methods and lag-time patterns, Water Resour. Manag. 27 (2013) 2267–2286, https://doi.org/10.1007/s11269-013-0287-2.

[41] M. Alizamir, S. Kim, O. Kisi, M. Zounemat-Kermani, A comparative study of several machine learning based non-linear regression methods in estimating solar radiation: case studies of the USA and Turkey regions, Energy 197 (2020), 117239, https://doi.org/10.1016/j.energy.2020.117239.

[42] S. Zhu, M. Ptak, Z.M. Yaseen, J. Dai, B. Sivakumar, Forecasting surface water temperature in lakes: a comparison of approaches, J. Hydrol. 585 (2020), 124809, https://doi.org/10.1016/j.jhydrol.2020.124809.

[43] C.E. Rasmussen, H. Nickisch, Gaussian processes for machine learning (GPML) toolbox, J. Mach. Learn. Res. 11 (2010) 3011–3015. https://dl.acm.org/doi/10.5555/1756006.1953029.

[44] P. Sihag, N.K. Tiwari, S. Ranjan, Modelling of infiltration of sandy soil using Gaussian process regression, Modeling Earth Systems and Environment 3 (3) (2017) 1091–1100, https://doi.org/10.1007/s40808-017-0357-1.

[45] Y. Li, Q. Zhang, S.W. Yoon, Gaussian process regression-based learning rate optimization in convolutional neural networks for medical images classification, Expert Syst. Appl. 184 (2021), 115357, https://doi.org/10.1016/j.eswa.2021.115357.
[46] Z.M. Yaseen, P. Sihag, B. Yusuf, A.M.S. Al-Janabi, Modelling infiltration rates in permeable stormwater channels using soft computing techniques, Irrigat. Drain. 70 (1) (2021) 117–130, https://doi.org/10.1002/ird.2530.
[47] W. Bukhari, S.M. Hong, Real-time prediction and gating of respiratory motion in 3D space using extended Kalman filters and Gaussian process regression network, Phys. Med. Biol. 61 (5) (2016) 1947, https://doi.org/10.1088/0031-9155/61/5/1947.
[48] J.Q. Shi, B. Wang, R. Murray-Smith, D.M. Titterington, Gaussian process functional regression modeling for batch data, Biometrics 63 (3) (2007) 714–723. https://www.jstor.org/stable/4541403.
[49] H. Tao, M.M. Hameed, H.A. Marhoon, M. Zounemat-Kermani, S. Heddam, S. Kim, S.O. Sulaiman, M.L. Tan, Z. Sa'adi, A.D. Mehr, M.F. Allawi, S.I. Abba, J. M. Zain, M.W. Falah, M. Jamei, N.D. Bokde, M. Bayatvarkeshi, M. Al-Mukhtar, S.K. Bhagat, T. Tiyasha, K.M. Khedher, N. Al-Ansari, S. Shahid, Z.M. Yaseen, Groundwater level prediction using machine learning models: a comprehensive review, Neurocomputing 489 (2022) 271–308, https://doi.org/10.1016/j.neucom.2022.03.014.
[50] L. Breiman, Bagging predictors, Mach. Learn. 24 (1996) 123–140.
[51] A.M. Melesse, K. Khosravi, J.P. Tiefenbacher, S. Heddam, S. Kim, A. Mosavi, B.T. Pham, River water salinity prediction using hybrid machine learning models, Water 12 (10) (2020) 2951, https://doi.org/10.3390/w12102951.
[52] M. Alizamir, S. Kim, M. Zounemat-Kermani, S. Heddam, A.H. Shahrabadi, B. Gharabaghi, Modelling daily soil temperature by hydro-meteorological data at different depths using a novel data-intelligence model: deep echo state network model, Artif. Intell. Rev. 54 (2021) 2863–2890, https://doi.org/10.1007/s10462-020-09915-5.
[53] M.A. Ghorbani, R. Khatibi, V.P. Singh, E. Kahya, H. Ruskeepää, M.K. Saggi, B. Sivakumar, S. Kim, F. Salmasi, M. Hasanpour Kashani, S. Samadianfard, Continuous monitoring of suspended sediment concentrations using image analytics and deriving inherent correlations by machine learning, Sci. Rep. 10 (1) (2020) 8589, https://doi.org/10.1038/s41598-020-64707-9.
[54] Z.M. Yaseen, S. Naghshara, S.Q. Salih, S. Kim, A. Malik, M.A. Ghorbani, Lake water level modeling using newly developed hybrid data intelligence model, Theor. Appl. Climatol. 141 (2020) 1285–1300, https://doi.org/10.1007/s00704-020-03263-8.
[55] K. Were, D.T. Bui, Ø.B. Dick, B.R. Singh, A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape, Ecol. Indicat. 52 (2015) 394–403, https://doi.org/10.1016/j.ecolind.2014.12.028.
[56] Y.H. Pao, S.M. Phillips, D.J. Sobajic, Neural-net computing and the intelligent control of systems, Int. J. Control 56 (2) (1992) 263–289, https://doi.org/10.1080/00207179208934315.
[57] Y.H. Pao, G.H. Park, D.J. Sobajic, Learning and generalization characteristics of the random vector functional-link net, Neurocomputing 6 (2) (1994) 163–180, https://doi.org/10.1016/0925-2312(94)90053-1.
[58] I.Y. Tyukin, D.V. Prokhorov, Feasibility of random basis function approximators for modeling and control, in: IEEE Control Applications and Intelligent Control, IEEE, 2009, pp. 1391–1396, https://doi.org/10.1109/CCA.2009.5281061.
[59] H.M. Chi, O.K. Ersoy, A statistical self-organizing learning system for remote sensing classification, IEEE Trans. Geosci. Rem. Sens. 43 (8) (2005) 1890–1900, https://doi.org/10.1109/TGRS.2005.851188.
[60] M. Alhamdoosh, D. Wang, Fast decorrelated neural network ensembles with random weights, Inf. Sci. 264 (2014) 104–117, https://doi.org/10.1016/j.ins.2013.12.016.
[61] L. Zhang, P.N. Suganthan, A comprehensive evaluation of random vector functional link networks, Inf. Sci. 367 (2016) 1094–1105, https://doi.org/10.1016/j.ins.2015.09.025.
[62] A.M. Hussein, M. Abd Elaziz, M.S.A. Wahed, M. Sillanpää, A new approach to predict the missing values of algae during water quality monitoring programs based on a hybrid moth search algorithm and the random vector functional link network, J. Hydrol. 575 (2019) 852–863, https://doi.org/10.1016/j.jhydrol.2019.05.073.
[63] R.M. Adnan, R.R. Mostafa, A.R.M.T. Islam, A.D. Gorgij, A. Kuriqi, O. Kisi, Improving drought modeling using hybrid random vector functional link methods, Water 13 (23) (2021) 3379, https://doi.org/10.3390/w13233379.
[64] S. Heddam, S. Kim, A. Elbeltagi, O. Kisi, Random vector functional link network based on variational mode decomposition for predicting river water turbidity, Current Directions in Water Scarcity Research 7 (2022) 245–264, https://doi.org/10.1016/B978-0-323-91910-4.00015-7.
[65] A.S. Qambar, M.M.M. Al Khalidy, Development of local and global wastewater biochemical oxygen demand real-time prediction models using supervised machine learning algorithms, Eng. Appl. Artif. Intell. 118 (2023), 105709, https://doi.org/10.1016/j.engappai.2022.105709.
[66] A.M. Ahmed, S.M.A. Shah, Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River, Journal of King Saud University-Engineering Sciences 29 (3) (2017) 237–243, https://doi.org/10.1016/j.jksues.2015.02.001.
[67] 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 (2) (2017) 164–175, https://doi.org/10.1016/j.ecohyd.2017.02.002.
[68] M.R. Golabi, S. Farzi, F. Khodabakhshi, F. Sohrabi Geshnigani, F. Nazdane, F. Radmanesh, Biochemical oxygen demand prediction: development of hybrid wavelet-random forest and M5 model tree approach using feature selection algorithms, Environ. Sci. Pollut. Control Ser. 27 (2020) 34322–34336, https://doi.org/10.1007/s11356-020-09457-x.
[69] M. Najafzadeh, A. Ghaemi, Prediction of the five-day biochemical oxygen demand and chemical oxygen demand in natural streams using machine learning methods, Environ. Monit. Assess. 191 (2019) 1–21, https://doi.org/10.1007/s10661-019-7446-8.
[70] A. Legouhy, al_goodplot - boxblot & violin plot. https://www.mathworks.com/matlabcentral/fileexchange/91790-al_goodplot-boxblot-violin-plot, 2022. (Accessed 22 September 2022). MATLAB Central File Exchange. Retrieved.