



A new insight for real-time wastewater quality prediction using hybridized kernel-based extreme learning machines with advanced optimization algorithms

Javad Alavi¹ · Ahmed A. Ewees² · Sepideh Ansari¹ · Shamsuddin Shahid³ · Zaher Mundher Yaseen^{4,5,6}

Received: 22 May 2021 / Accepted: 21 October 2021 / Published online: 5 November 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Accurate prediction of inlet chemical oxygen demand (COD) is vital for better planning and management of wastewater treatment plants. The COD values at the inlet follow a complex nonstationary pattern, making its prediction challenging. This study compared the performance of several novel machine learning models developed through hybridizing kernel-based extreme learning machines (KELMs) with intelligent optimization algorithms for the reliable prediction of real-time COD values. The combined time-series learning method and consumer behaviours, estimated from water-use data (hour/day), were used as the supplementary inputs of the hybrid KELM models. Comparison of model performances for different input combinations revealed the best performance using up to 2-day lag values of COD with the other wastewater properties. The results also showed the best performance of the KELM-salp swarm algorithm (SSA) model among all the hybrid models with a minimum root mean square error of 0.058 and mean absolute error of 0.044.

Keywords Time-series learning · Consumer behaviour · Kernel-based extreme learning machine · Intelligent algorithms · Real-time water quality prediction · Wastewater

Introduction

Research background

Chemical oxygen demand (COD) is one of the most important indicators used to determine the efficiency of wastewater treatment plants (WWTPs) (Ibrahim 2019; Tung and Yaseen 2020). The COD is also required to assess the

improvement of automatic control of biological systems of a WWTP. Therefore, COD concentration measurement at the WWTP inlet is imperative for its performance enhancement (Revilla et al. 2016; Wei et al. 2019; Salih et al. 2020). The COD of municipal sewage usually varies with time. Therefore, WWTP operators commonly over-aerate the aeration tanks and over-add the chemicals to ensure the treatment quality of sewage (Man et al. 2019; Chen et al. 2020). The aeration unit accounts for 40–50% of the total power

Responsible Editor: Marcus Schulz

✉ Zaher Mundher Yaseen
zaheryaseen88@gmail.com; yaseen@alayan.edu.iq

Javad Alavi
alavi.sharif@gmail.com

Ahmed A. Ewees
a.eweess@hotmail.com

Sepideh Ansari
aysanansari1991@gmail.com

¹ Department of Environmental Sciences and Engineering, Kheradgarayan Motahar Institute of Higher Education, Mashhad, Iran

² Computer Department, Damietta University, Damietta, Egypt

³ School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia (UTM), 81310 Johor Bahru, Malaysia

⁴ Department of Urban Planning, Engineering Networks and Systems, Institute of Architecture and Construction, South Ural State University, 76, Lenin Prospect, 454080 Chelyabinsk, Russia

⁵ New Era and Development in Civil Engineering Research Group, Scientific Research Center, Al-Ayen University, Thi-Qar, Nasiriyah 64001, Iraq

⁶ College of Creative Design, Asia University, Taichung City, Taiwan

consumption in a municipal WWTP (Singh and Kansal 2018; Man et al. 2019). Therefore, over-aeration significantly increases operational expenditures. Over-aeration also causes poor settling of suspended solids and the quality reduction of effluent, leading to the destruction of the flocculating agent (Al doury and Al samerrai 2019; He et al. 2019). Therefore, rapid organic matter measurement is vital for optimum control of biological units (Yu and Bai 2019) and ensure the required effluent quality (Bolyard et al. 2019). The optimum operation of biological units is also needed to support downstream operations, maintain effluent quality, ensure minimum operational cost and provide an intelligent environmental solution (Dairi et al. 2019). Sophisticated instruments and hardware, generally used to measure COD concentrations, are highly expensive (Al-Doury and Alwan 2019). Besides, the hardware functions are often subjected to the limitations of the process conditions (He et al. 2015). A large delay in measurement is also a major challenge in process control (Wang et al. 2019). This emphasizes the need for optimum controlling of amount of aeration and chemical. The real-time prediction of the influent COD can help in controlling the amount of aeration and chemical. Correlation analysis of different quality parameters with influent COD values in wastewater is an inexpensive, reliable and continuous technique for assuming influent COD. Tommasen (2014) suggested that electrical conductivity (EC) and turbidity can be employed as a proxy for COD estimation. They showed that turbidity and conductivity could provide fitting formulas to estimate COD in wastewater with reasonable correlation ($r^2=0.65-0.69$). Yu et al. (2019) showed no correlation of COD with conductivity but with EC and total nitrogen/ NH_4^+ . Overall, the literature suggests that the correlation between wastewater quality parameters and COD is not as accurate as the industrial requirement.

Review of artificial intelligence-related literature

In recent years, artificial intelligence (AI) has been applied to overcome the limitation of hardware devices (Sharafati et al. 2020; Al-Sulttani et al. 2021; Tiyasha et al. 2021). The capability of AI to provide intelligent control and monitoring has attracted the attention of water treatment engineers in recent years to implement them in WWTPs (Abouzari et al. 2021; Wang and Man 2021). Mathematical or statistical models based on accessible historical observation of wastewater can provide an effective real-time prediction of current and future water quality parameters to control WWTPs (Bernardelli et al. 2020). Such models can be divided into four major categories: multivariate statistics, machine learning, fuzzy logic and hybrid (Zhu et al. 2018). Significant achievements have been made on wastewater quality forecasting using those models. However, the model accuracy drops significantly when the input data fluctuates sharply. For coping with such nonlinear and uncertain situations, heuristic

algorithms, such as an artificial neural network (ANN), have been used, considering their robust adaptability and learning ability. Abyaneh (2014) investigated the application of multiple linear regression (MLR) and ANN models for COD and BOD_5 prediction in WWTPs and showed a better performance of ANN compared to MLR ($r=0.83$ and 0.81 in the prediction of BOD and COD, respectively). Although ANN has been found feasible in predicting wastewater quality, it was not sufficient for industrial process control requirements (Wang et al. 2019). This is due to the shortcomings of classical ANN models to local trapping and long learning time (Man et al. 2019). According to the reported literature in the Scopus database, nearly 30 research articles discussed the feasibility of AI models in predicting COD at WWTP inlet. Figure 1 exhibits the occurrence of the major keywords of those studies. The figure indicates that the topic has received major attention in recent years. The determination of COD contributed to diverse environmental engineering aspects such as pollutant control, removal efficiency and environmental monitoring. In addition, it motivated exploring new intelligence models such as ensemble machine learning, decision tree models and others. Hence, the literature review studies emphasize the further investigation of the current research topic.

The extreme learning machine (ELM) is a novel neural algorithm with fast convergence capacity to global optima. The ELM showed outstanding predictive performance in different fields of engineering and natural sciences (Yu and Bai 2019; Hai et al. 2020; Yaseen et al. 2021). The paradigm of ELM is a biological learning technique that involves kernels, random neurons (with or without unknown modelling/shape) and optimization constraint. ELM is much more effective than the traditional ANN in practical applications in terms of ease of use, efficiency, generalization in performance, adaptability to several nonlinear kernel and activation functions, and convergence speed (Alaba et al. 2019). However, the structure of ELM is more complex for large data (like WWTP inlet wastewater quality values) as the random selection of hidden biases and input weights make the classical ELM dependent on more hidden nodes, which eventually influence the network generalization ability (Alaba et al. 2019). Two approaches are commonly used to improve the prediction performance of ELM: (i) the optimization of parameters and (ii) the changes in learning mode (Yu and Bai 2019). The combination of individual models with specific rules provides the capability of gathering comprehensive information about the data to improve the precision of predictive models. The hybrid algorithm-based prediction methods can reduce prediction error caused by the parameters or model misidentification (Wang et al. 2019). Therefore, recent studies attempted to enhance ELM performance through hybridization with optimization algorithms. The kernel extreme learning machine (KELM) can be used for speedy processing of large datasets (Li et al. 2019) and to avoid computational complexity caused by high-dimensional

simulated water consumer behaviours as supplementary inputs for the first time to improve the prediction accuracy of the basic KELM model and its hybridized versions, including the KELM-salp swarm algorithm (SSA), KELM-firefly algorithm (FA), KELM-particle swarm optimization (PSO), KELM-genetic algorithm (GA), KELM-grey wolf optimizer (GWO) and KELM-sine cosine algorithm (SCA).

Material and methods

Case study

The inlet wastewater quality data of a WWTP (No. 4) located in Mashhad, Iran, were collected and employed to construct the models. The system receives municipal wastewater with an average nominal capacity of $80,000 \text{ m}^3 \text{ day}^{-1}$ for 472,000 people in Mashhad, Iran (Fig. 2). The plant has been operating since 2016. Its biological process is modified Ludzak–Ettinger (MLE), one

of the most commonly used processes for nitrogen removal in municipal WWTPs. The online sensors (Endress + Hauser) equipped with a beam were installed at the plant's inlet to monitor the variations of real-time influent compositions. It measures chemical oxygen demand (COD), ammonium ion, temperature, EC and pH (Fig. 2). The Parshall flume also measured the flow rate with an Endress Hauser flow computer.

The integrated programmable logic controller (PLC) with control software installed deals with different operating conditions of the WWTP. The oxygen sensors installed in the aerobic reactors detect the COD values at the inlet, while the amount of aeration supplied by the blowers is controlled by a flowmeter. In addition, the amount of supplied air is necessary to be monitored and checked to prevent over-aeration, which leads to sludge settling problems in the final clarifiers. As the required air directly depends on the values of chemical and organic matter at the inlet, the accurate prediction of COD provides energy saving in the blower room. It can also avoid sludge settling problems due to the over-aeration in the plant.

Fig. 2 **a** An aerial view of Mashhad WWTP No. 4. **b** Sensors used for measuring wastewater quality and quantity



(a)



(b)

Brief description of kernel-based extreme learning machine

ELM randomly generates its initial weights and biases and subsequently fixes the network’s nonlinearities without looping (Huang et al. 2004). Assume that there are samples (x, y) with n instances, the output function of ELM $f(x)$ having L hidden neurons and $h(x)$ activation function can be expressed as

$$f(x) = \sum_{i=1}^L \beta_i h_i(x) = h(x)\beta \tag{1}$$

where $\beta = [\beta_1, \beta_2, \dots, \beta_L]$ denotes the weight vector that links the hidden nodes with the output nodes (Sanikhani et al. 2018). $H = \{h_{ij}\} (i = 1, \dots, N \text{ and } j = 1, \dots, L)$ denotes the output of the hidden layer. The function, $h(x)$, maps the data from d dimension to L -dimensional hidden nodes feature space H .

The minimal norm least square is applied to determine the output weights,

$$\beta = H^+ T \tag{2}$$

where H^+ indicates the Moore–Penrose generalized inverse of the output matrix H of the hidden layer.

Based on the concept of ELM, an improved version was proposed by Huang et al. (2012) using radial basis function (RBF) kernel, known as kernel-based extreme learning machine (KELM). The KELM adds a positive value, $1/c$, to calculate the output weights as

$$\beta = H^T \left(HH^T + \frac{I}{C} \right)^{-1} T \tag{3}$$

where the value of C is manually defended, and I denotes an identity matrix of N dimension. Accordingly, the $f(x)$ function can be expressed as,

$$f(x) = h\beta = h(x)H^T \left(HH^T + \frac{I}{C} \right)^{-1} T \tag{4}$$

When the $h(x)$ is unknown, the kernel matrix for KELM is applied in Eq. (5).

$$\Omega_{ELM} = HH^T : \Omega_{ELM_{ij}} = h(x_i) \cdot h(x_j) = K(x_i, x_j) \tag{5}$$

where $K(x_i, x_j)$ indicates the kernel function. The output function can be computed as,

$$f(x) = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}^T \left(\frac{I}{C} + \Omega_{ELM} \right)^{-1} T \tag{6}$$

KELM can use more than one type of kernel function, such as RBF, polynomial and linear (Liu et al. 2020). Among

them, the RBF is one of the popular kernel functions. It has some advantages, such as simplicity and fewer parameters. Equation (7) shows the mathematical model of the RBF function.

$$K(x, x_i) = \exp\left(-\frac{\|x, x_i\|^2}{\sigma^2}\right) \tag{7}$$

Intelligent optimization algorithms

Salp swarm algorithm (SSA)

SSA is a kind of optimization method developed to solve different types of optimization problems. SSA was developed based on the natural behaviour of the Salpidae’s family (Mirjalili et al. 2017). Salp follows a special style in moving and foraging called salp chain. This style can be considered as swarm conduct, as depicted in Fig. 3.

The first phase of the SSA algorithm is to split the population into two levels, namely leaders and followers. The leader is the front slap, and the followers are the rest of them. The salps change their positions frequently to search for the target. The change in the leader’s position can be expressed as

$$x_j^1 = \begin{cases} F_j + c_1 ((ub_j - lb_j)ub_j - lb_j) \times c_2 + lb_j & c_3 \leq 0 \\ F_j - c_1 ((ub_j - lb_j)ub_j - lb_j) \times c_2 + lb_j & c_3 > 0 \end{cases} \tag{8}$$

where x_j^1 indicates the position of the j th leader and F_j denotes the target in this dimension. The lower and upper boundaries are the ub_j and lb_j , respectively. The values of c_2 and c_3 are generated randomly in the range $[0, 1]$. The parameter c_1 is applied to balance between the exploitation and exploration phases using Eq. (9).

$$c_1 = 2e^{-\left(\frac{4t}{t_{max}}\right)^2} \tag{9}$$

where t denotes the current iteration, and t_{max} denotes the maximum number of iterations.

The changes in the positions of the followers can be represented using Eq. (10).

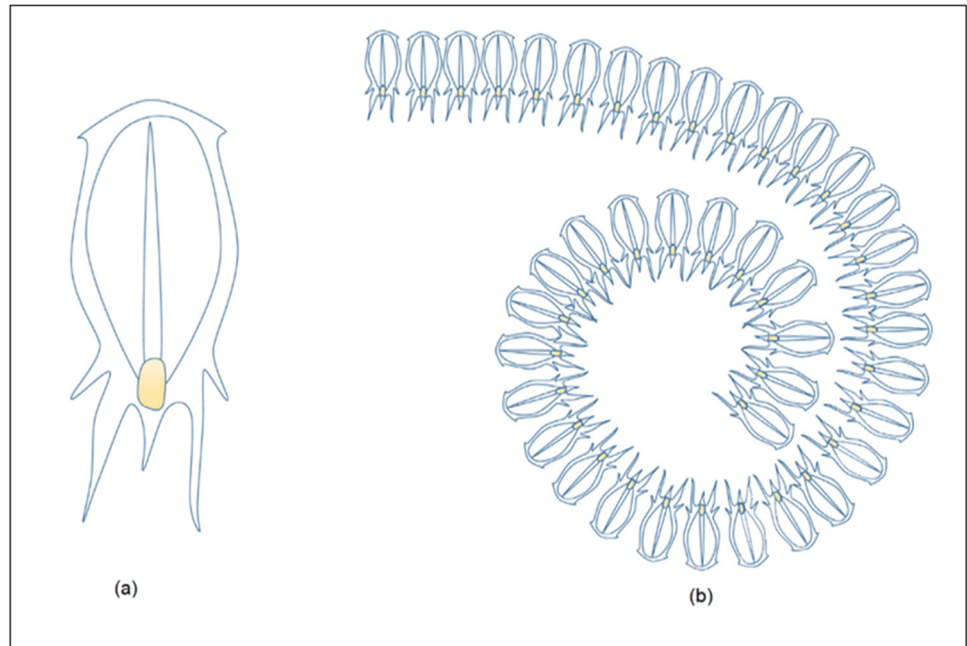
$$x_j^i = \frac{1}{2} (x_j^i + x_j^{i-1}) \tag{10}$$

where $i > 1$ and x_j^i represents the follower’s position in i th dimension.

Salp swarm algorithm (SSA)-ELM

Hybrid SSA-ELM (SSAELM), proposed in this study, was developed in two phases. The first phase optimizes the parameters of the KELM using the SSA algorithm. The second phase

Fig. 3 Example of salps: **a** one salp and **b** salp chain (swarm) (Mirjalili et al., 2017)



improves the performance of the KELM in prediction. The optimized parameters are C and σ . The SSAELM begins by splitting the input data into two groups to train and test the method besides initializing all parameter values. In the next step, the SSA algorithm optimizes the C and σ parameters of KELM by exploring the search domain. The KELM objective function and error evaluate each candidate parameter. The process is iterated until the smallest error is achieved or the maximum number of iterations is reached. The final optimized parameters are passed to optimize the kernel function of the KELM to evaluate prediction ability. Figure 4 shows the flow-chart of the proposed method.

Firefly algorithms (FA)

FA was developed by Yang (2010) based on the movement of the fireflies. When fireflies are attracted to the light, the swarm moves to the brightest firefly, which can be conceptually used to solve different optimization problems. The firefly’s attraction to the light is directly related to its brightness which relies on the intensity of the agent (Naganna et al. 2019). The attractiveness (β), the light intensity $I(r)$ and the Cartesian distance between any two fireflies x_i and x_j are the primary setup variables of the FA algorithm, which can be expressed using Eqs. (11) to (13).

$$I(r) = I_0 \exp(-\gamma r^2) \tag{11}$$

$$\beta(r) = \beta_0 \exp(-\gamma r^2) \tag{12}$$

$$r_{ij} = x_i + x_j = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2} \tag{13}$$

where I_0 and $I(r)$ are the light intensity from a firefly and the light intensity at distance r , $\beta(r)$ and β_0 are the attractiveness β at a distance r and $r=0$, $x_{i,k}$ indicates the k th component of the spatial coordinate x_i of the i th firefly, c presents the coefficient of the light absorption, and d denotes the dimensionality of the problem.

The next movement of the firefly (i) can be expressed using Eqs. (14) and (15).

$$x_i^{i+1} = x_i + \Delta x_i \tag{14}$$

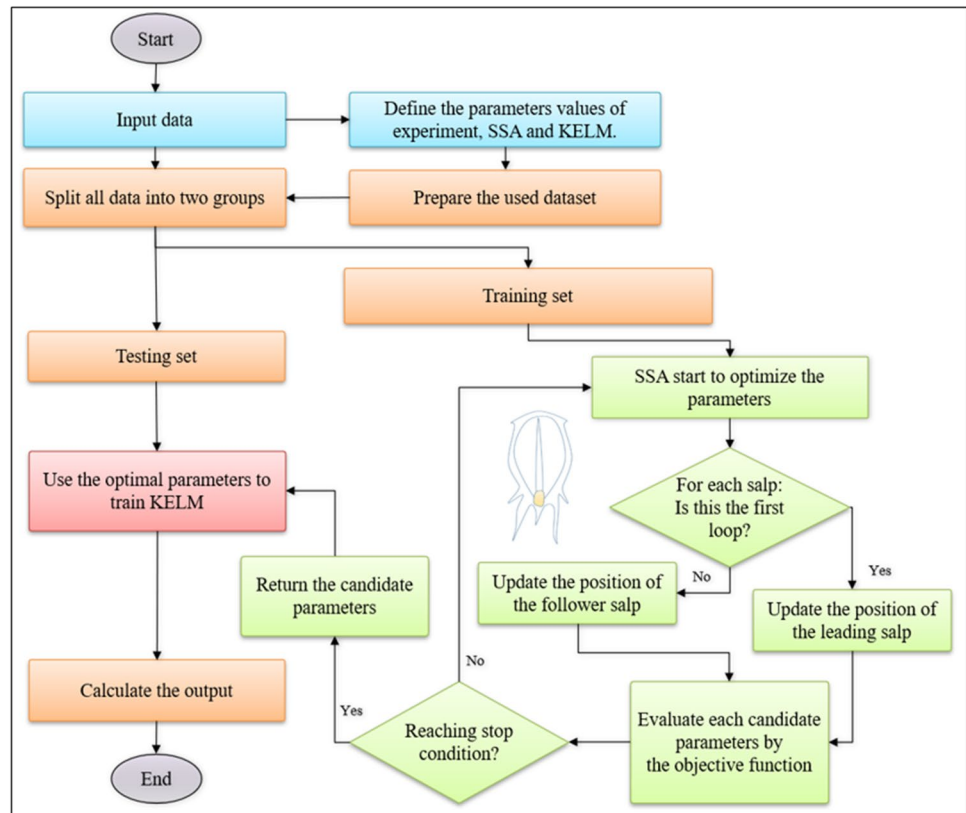
$$\Delta x_i = \beta_0 e^{-\gamma r^2} (x_j - x_i) + \alpha \epsilon_i \tag{15}$$

The first part of Eq. (15) determines the attraction and the second part represents the randomization process; α controls the randomization values in the range $[0,1]$, and ϵ_i presents the random number of the Gaussian distribution.

Particle swarm optimization (PSO)

PSO is an intelligent optimization technique inspired by the behaviour of birds flocking or fish schooling. Any individual in the PSO is known as a particle, and the population is called a swarm. Each particle in the PSO has a position and a velocity vector in the problem search domain, which memorizes the best position using a fitness function value. The particles adjust their positions and velocities according to the positions of other particles in the swarm. The current

Fig. 4 The flowchart of the method proposed in this study for inlet COD prediction at WWTP



position of the i th particle with D dimensions at t th iteration is determined as

$$X_i(t) = \{x_{i1}, x_{i2}, \dots, x_{id}\}^t \quad (16)$$

Earlier best position and velocity are determined by Eqs. (17) and (18).

$$P_i(t) = \{p_{i1}, p_{i2}, \dots, p_{id}\}^t \quad (17)$$

$$V_i(t) = \{v_{i1}, v_{i2}, \dots, v_{id}\}^t \quad (18)$$

Each location is updated by Eq. (19) during the search in space.

$$X_i(t) = X_i(t-1) + V_i(t) \quad (19)$$

The new velocity is updated by Eq. (20).

$$V_i(t) = \omega \times V_i(t-1) + c_1 \times R_1 \{P_i(t) - X_i(t-1)\} + c_2 \times R_2 \{P_g(t) - X_i(t-1)\} \quad (20)$$

where $i = 1, 2, \dots, n$ (n represents the population size), $t = 1, 2, \dots, t$ (t indicates the iterations number), R_1 and R_2 are randomly generated numbers in the range $[0, 1]$, ω = factor of inertia, c_1 and c_2 represent the learning

coefficients, $X_i(t)$ is computed by its previous location $X_i(t-1)$, and its current velocity, $V_i(t)$, denotes the particle's velocity used to change the location of the particles to a better one, $V_i(t-1)$ denotes the velocity from the previous iteration, P_i denotes the best position of each particle, and P_g denotes the best location determined by any particle.

The inertia weights, $\omega(t)$, and learning coefficients (c_1 and c_2) are determined by Eqs. (21) through (23).

$$\omega(t) = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{T} \times t \quad (21)$$

$$c_1 = c_{1,\max} - \frac{c_{1,\max} - c_{1,\min}}{T} \times t \quad (22)$$

$$c_2 = c_{2,\max} - \frac{c_{2,\max} - c_{2,\min}}{T} \times t \quad (23)$$

where T is the iterations numbers, ω_{\max} and ω_{\min} are the maximum and minimum inertia weights, $c_{1,\min}$ and $c_{2,\min}$ are the minimum learning factors, and $c_{1,\max}$ and $c_{2,\max}$ are the maximum learning factors.

Genetic algorithm (GA)

GA is a population-based stochastic optimization method developed based on the perceptions of natural evolution and the biological principles of natural selection and genetics. The method generates different solutions to optimize a problem by carrying out stochastic transformations (Al-Obaidi et al. 2017). Each solution is presented in this advanced optimization algorithm by a string of genetic factors known as chromosomes. The chromosome is generated through successive iterations, which are called generations. In each iteration, the population is generated using three genetic operations, selection, crossover and mutation. A new set of approximations at each generation of a GA is generated by selecting individuals according to their fitness in the problem domain and reproducing them using operators borrowed from natural genetics (Srinu Naik and Pydi Setty 2014). A fitness value, defined by a fitness function, is related to each individual in the population. The artificial evolution processes which mimic natural evolution are used to generate new candidate solutions. This process provides the evolution of individuals of a population better suited to their environment than the other individuals. The roulette wheel selection method was applied in this research which involves the random generation of positional design solutions to assess and refine the solutions until a stopping criterion is met (Aras et al. 2007). In this method, the parents are chosen when the chromosomes with higher fitness have a greater chance for selection. The probability for each chromosome i is computed using Eq. (24).

$$P_i = \frac{f_i}{\sum_{j=1}^n f_j} \tag{24}$$

where f_i denotes the fitness of chromosome i . N denotes the population number.

Grey wolf optimizer (GWO)

GWO is developed based on the hunting process of grey wolves (Mirjalili et al. 2014). The best solution is called alpha, while the second and third best solutions are beta and delta, respectively. Omega is the candidate solution. The algorithm uses the solutions alpha, beta and delta for hunting guidance, and omega follows the three solutions. These stages are represented by Eqs. (25) and (26) (Mirjalili et al. 2014):

$$\overline{D} = \left| \overline{C} \cdot \overline{X}_p(t) - \overline{X}(t) \right| \tag{25}$$

$$\overline{X}(t+1) = \overline{X}_p(t) - \overline{AD} \tag{26}$$

where A and C denote the vector of coefficients, \overline{X}_p and \overline{X} denote the position of the prey and grey wolves, \overline{D} denotes

a vector to determine the new location of GWO, t denotes the iteration value, and $\overline{X}(t+1)$ indicates the GWO position in the next iteration. \overline{A} and \overline{C} can be represented using Eqs. (27) and (29):

$$\overline{A} = 2\overline{ar}_1 - \overline{a} \tag{27}$$

$$\overline{a} = 2 - t \frac{2}{\text{Max}(t)} \tag{28}$$

$$\overline{C} = 2\overline{r}_2 \tag{29}$$

where \overline{a} is a value that increases linearly from 0 to 2 through iterations. It is 0 when t hits the max iterations number; \overline{r}_1 and \overline{r}_2 are random vectors in the range [0, 1].

The alpha, beta and delta (α , β and δ) are used to model the hunting behaviour of the GWO. Therefore, the α , β and δ are saved, and other search factors are forced to update their position based on optimal solutions following Eqs. (30) to (32):

$$\overline{X}(t+1) = \frac{\overline{X}_1 + \overline{X}_2 + \overline{X}_3}{3} \tag{30}$$

$$\overline{X}_1 = \left| \overline{X}_\alpha - \overline{A}_1 D_\alpha \right|, \overline{X}_2 = \left| \overline{X}_\beta - \overline{A}_2 D_\beta \right|, \overline{X}_3 = \left| \overline{X}_\delta - \overline{A}_3 D_\delta \right| \tag{31}$$

$$\overline{D}_\alpha = \left| \overline{C}_1 \overline{X}_\alpha - \overline{X} \right|, \overline{D}_\beta = \left| \overline{C}_2 \overline{X}_\beta - \overline{X} \right|, \overline{D}_\delta = \left| \overline{C}_3 \overline{X}_\delta - \overline{X} \right| \tag{32}$$

where \overline{X}_1 , \overline{X}_2 and \overline{X}_3 are the first three solutions in the GWO at t iteration. \overline{A} contains a random number between $-2a$ and $2a$, if $|\overline{A}| < 1$, the wolves are not attacking the target, and therefore, the wolves need to search for a better location.

Sine cosine algorithm (SCA)

SCA, introduced by Mirjalili et al. (2020), where sine and cosine functions are applied to update the solutions using Eqs. (33) and (34).

$$X_i = X_i + r_1 \times \sin(r_2) \times |r_3 \text{Best} - X_i| \tag{33}$$

$$X_i = X_i + r_1 \times \cos(r_2) \times |r_3 \text{Best} - X_i| \tag{34}$$

The equations are combined as

$$X_i = \begin{cases} X_i + r_1 \times \sin(r_2) \times |r_3 \text{Best} - X_i| & \text{if } r_4 < 0.5 \\ X_i + r_1 \times \cos(r_2) \times |r_3 \text{Best} - X_i| & \text{if } r_4 \geq 0.5 \end{cases} \tag{35}$$

where Best and X_i indicate the best and the current solutions, respectively; r_1 denotes a random variable responsible

for defining the search domain of the next solution. It is updated using Eq. (36) to add a balance between exploration and exploitation.

$$r_1 = a - t \frac{a}{t_{\max}} \quad (36)$$

where a and t present a constant and the current iteration, respectively, t_{\max} denotes the maximum number of iterations, r_2 is a randomly generated parameter used to define the movement of the next solution (for instance if it moves towards or outwards *Best*), r_3 indicates a random variable that determines a random weight for the best solution to stochastically emphasize ($r_3 > 1$) or deemphasize ($r_3 < 1$), and r_4 specifies if the solution will be updated by the sine or cosine equations as in Eq. (35).

Data learning methods and scenario description

The combined KELM network with intelligent optimization algorithms was applied in this study to predict real-time COD concentrations at the inlet of WWTP. The quality of influent exhibits nonlinear patterns and large temporal variability, which inherently reduce the model prediction accuracy in real time. Two approaches were used in this research to maximize the prediction accuracy of hybrid KELM models: (i) consideration of record time of measured data as input (hour of a day and day of a week) to simulate the pattern of water consumption and organic pollutant discharge and (ii) incorporation of time-series data learning method by adding COD values up to n lags of the current time as inputs to map not only the nonlinearity of the influent character but also its time variability.

The current inlet COD concentration at any time depends on the previous concentration values. Therefore, the machine must have an awareness of the past. However, which past values are critical and how long they remain important depend on model training, and thus, the system must be flexible in that regard. The standard KELM model does not map the temporal dependence of variables. Hence, the network is not proper for modelling time-series data. Since no information of past values is stored, the model fails to simulate time-series data accurately. Therefore, a function for informing input and output variables (window of data) needs to be constructed to model time-dependent variables for predicting current value (t) from the n lag values (Najafzadeh and Zeinolabedini 2019). Such historical values of the variables are known as memory values (Verma et al. 2013). As the memory captures the past time-series information, the model can learn the dynamic behaviour (Wei and Kusiak 2014) and nonlinearity in time series.

The available wastewater parameters used in this study include flow rate, NH_4 , pH, EC and temperature. In the modified model, the current values except COD denoting as

$x_{i(t)}$ and the COD values for the past n days along with the other parameters denoting $y_{i(t)}$ as inputs are presented in Eqs. (37) and (38). Thus, the time and historical COD values are externally embedded as short-term memory in the function presented in Eq. (39). The networks exposed to deep learning can be enhanced by increasing the numbers of input variables and strengthening the prediction accuracy because enhanced supervised learning can be achieved through the construction of background information.

$$\text{Input data sets} = \begin{pmatrix} \text{Flow rate}_t \\ \text{Tem}_t \\ \text{pH}_t \\ \text{NH}_4_t \\ \text{EC}_t \end{pmatrix}, \begin{pmatrix} \text{COD}_{t-1} \\ \text{Flow rate}_{t-1} \\ \text{Tem}_{t-1} \\ \text{pH}_{t-1} \\ \text{NH}_4_{t-1} \\ \text{EC}_{t-1} \end{pmatrix}, \dots, \begin{pmatrix} \text{COD}_{t-n} \\ \text{Flow rate}_{t-n} \\ \text{Tem}_{t-n} \\ \text{pH}_{t-n} \\ \text{NH}_4_{t-n} \\ \text{EC}_{t-n} \end{pmatrix} \quad (37)$$

$$\text{Input data sets} = x_i(t), y_i(t-1), \dots, y_i(t-n) \quad (38)$$

$$\text{COD}_{\text{in}}(t) = f(x_i(t), y_i(t-1), \dots, y_i(t-n)) \quad (39)$$

where $x_{i(t)}$ is the set of input parameters at time (t) and $y_{i(t)}$ is the set of input parameters at time t along with the COD concentration at one step back from t .

As the meaningful time lags between time-series variables are unknown, the model should cover the maximum time length to reduce the risk of missing important past data points (Kim et al. 2018). Therefore, the process of history extension continues until the maximum prediction accuracy is achieved. The best model topology is the one having the best fit based on the evaluation metrics. As the database should be divided into learning and test sets, they should be independent to provide an unbiased estimate of error, usually ensured by randomized splitting. Nevertheless, the time correlation between input datasets would be lost in that case. To cope with this issue, datasets were chronologically arranged and then divided into training and test sets. In this study, seven models were trained with input datasets, including $y_{i(t-1)}$ in scenario III, and deep learning models with different input datasets, including $y_{i(t-1)}$ and $y_{i(t-2)}$, in scenario IV.

This study compared the performance of the basic KELM model and hybrid KELM models and six advanced algorithms for four different input combinations/scenarios. The available wastewater parameters ($x_{i(t)}$) were considered input data for the scenario I. The hour of the day and day of the

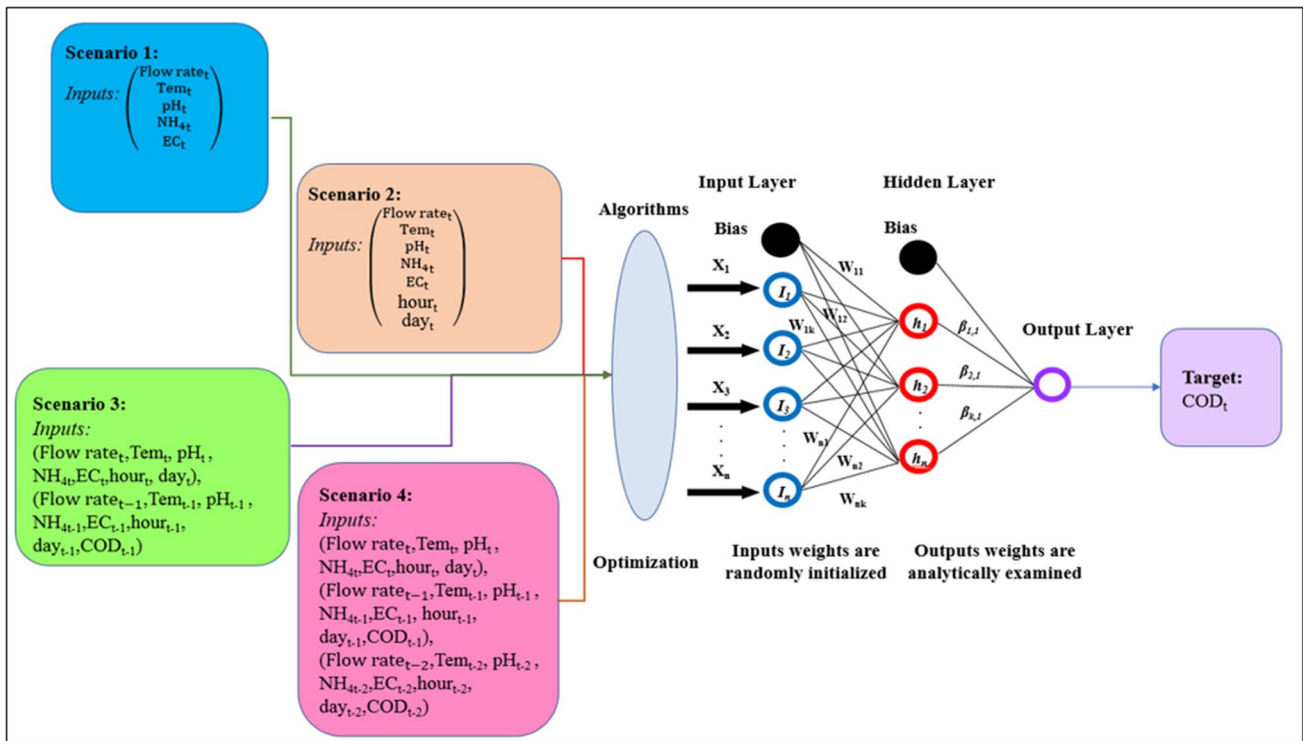


Fig. 5 The predictive model structure and the modelling scenarios employed in this study

Table 1 Characteristics of inlet wastewater in Mashhad WWTP No. 4 and comparison of wastewater quality variations in weekdays and workdays (\pm indicates standard deviation)

Parameter and unit	Raw wastewater			Workdays			Weekends		
	Ave	Max	Min	Ave	Max	Min	Ave	Max	Min
COD (mg l ⁻¹)	787 ± 87	985	554	783 ± 84	982	563	798 ± 94	985	554
EC (µS cm ⁻¹)	1839 ± 119	2359	1449	1839 ± 120	2359	1449	1838 ± 115	2170	1630
pH	8.4 ± 0.1	8.8	7.9	8.4 ± 0.1	8.8	8.1	8.4 ± 0.2	8.8	7.9
NH ₄ (mg l ⁻¹)	58 ± 14	89	29	58 ± 13	89	31	57 ± 15	89	29
Temperature (°C)	27.4 ± 0.5	28.3	25.6	27.4 ± 0.4	28.2	25.9	27.4 ± 0.5	28.3	25.6
Flow rate (m ³ h ⁻¹)	3304 ± 444	4286	1915	3296 ± 445	4061	1915	3326 ± 443	4286	2048

week were added to $x_{i(t)}$ as the supplementary input data in scenario II. Scenarios III and IV are the models that include the time of data recorded and the time-series data learning method, respectively. In scenario III, the models are trained

by input datasets of $x_i(t)$, $y_i(t - 1)$ and time of data (hours and days). The input datasets were $x_{i(t)}$, $y_{i(t - 1)}$ and $y_{i(t - 2)}$ in scenario IV. Figure 5 shows the predictive models' structures and training for different scenarios.

Table 2 The correlation values between the input parameters and the predicted COD

	COD (mg l ⁻¹)	EC (µs)	NH ₄ (mg l ⁻¹)	pH	Temperature	Flow rate (m ³ h ⁻¹)
COD (mg l ⁻¹)	1.0000					
EC (µs)	0.1850	1.0000				
NH ₄ (mg l ⁻¹)	-0.1455	0.2134	1.0000			
pH	-0.2824	0.3268	0.2807	1.0000		
Temperature (°C)	0.0275	-0.1816	-0.1320	-0.4559	1.0000	
Flow rate (m ³ h ⁻¹)	0.4489	0.2382	-0.0962	-0.5109	0.3542	1.0000

Table 3 The statistical performance of the hybrid models for COD prediction in training and testing phases (scenario I)

Models	RMSE	MAE	MAPE	NSE	WI	R^2
Training						
KELM-SSA	0.107	0.086	0.207	0.721	0.907	0.727
KELM-FA	0.116	0.091	0.381	0.676	0.888	0.681
KELM-PSO	0.143	0.116	0.335	0.499	0.788	0.512
KELM-GA	0.121	0.093	0.215	0.634	0.871	0.636
KELM-GWO	0.119	0.095	0.345	0.649	0.874	0.656
KELM-SCA	0.129	0.099	0.228	0.584	0.845	0.587
KELM	0.136	0.109	0.345	0.539	0.817	0.546
Testing						
KELM-SSA	0.130	0.105	0.428	0.567	0.848	0.569
KELM-FA	0.144	0.118	0.269	0.471	0.811	0.479
KELM-PSO	0.157	0.126	0.298	0.406	0.734	0.410
KELM-GA	0.151	0.119	0.466	0.482	0.817	0.487
KELM-GWO	0.146	0.112	0.271	0.496	0.808	0.497
KELM-SCA	0.151	0.116	0.452	0.496	0.809	0.496
KELM	0.170	0.132	0.304	0.331	0.698	0.336

Data collection and pre-processing

Mean hourly records of wastewater parameters measured by sensors over a season from July to September 2018 were collected and pre-processed. The database includes 1644 values with a total number of data points of 9858. Table 1 shows the descriptive statistics of inlet variables on workdays and weekends separately. Table 2 reports the correlation to show the association between the predictors and the COD. The table shows that the flow rate has a substantial influence on the COD amount.

Before building the modelling networks, the data points were normalized between 0 and 1. Datasets were then chronologically arranged and divided into training and test sets. The database was similarly partitioned into two subsets for all paradigms and scenarios. Seventy-five percent of the first part of the dataset was used to train the model, and the remaining 25% was used to test the models' precision. The models were built using MATLAB 9.2 mathematical software.

Model accuracy criteria

Different statistical metrics were used to evaluate the performance of the models developed in this study, including root mean square error (RMSE), mean absolute errors (MAE), mean absolute percentage error (MAPE), the Nash–Sutcliffe efficiency (NSE), Willmott Index of agreement (WI) and coefficient of determination (r^2)

(Sharafati et al. 2018). The metrics can be calculated using Eqs. (40) to (45) (Yaseen 2021).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_{\text{obs},i} - Y_{\text{pred},i})^2} \quad (40)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |Y_{\text{obs},i} - Y_{\text{pred},i}| \quad (41)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_{\text{obs},i} - Y_{\text{pred},i}}{Y_{\text{obs},i}} \right| \quad (42)$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (Y_{\text{obs},i} - Y_{\text{pred},i})^2}{\sum_{i=1}^n (Y_{\text{obs},i} - \bar{Y}_{\text{obs}})^2} \quad (43)$$

$$\text{WI} = 1 - \frac{\sum_{i=1}^n (Y_{\text{obs},i} - Y_{\text{pred},i})^2}{\sum_{i=1}^n (|Y_{\text{pred},i} - \bar{Y}_{\text{obs}}| + |Y_{\text{obs},i} - \bar{Y}_{\text{obs}}|)^2} \quad (44)$$

$$r^2 = 1 - \frac{\sum_{i=1}^n (Y_{\text{obs},i} - Y_{\text{pred},i})^2}{\sum_{i=1}^n (Y_{\text{obs},i} - \bar{Y}_{\text{obs}})^2} \quad (45)$$

where n is the number of the observed data and $Y_{\text{obs},i}$, $Y_{\text{pred},i}$ and \bar{Y}_{obs} are the observed, predicted and mean observed values, respectively.

Fig. 6 The scatterplots of the observed and predicted inlet COD concentrations by different KELM models in testing phase (scenario I)

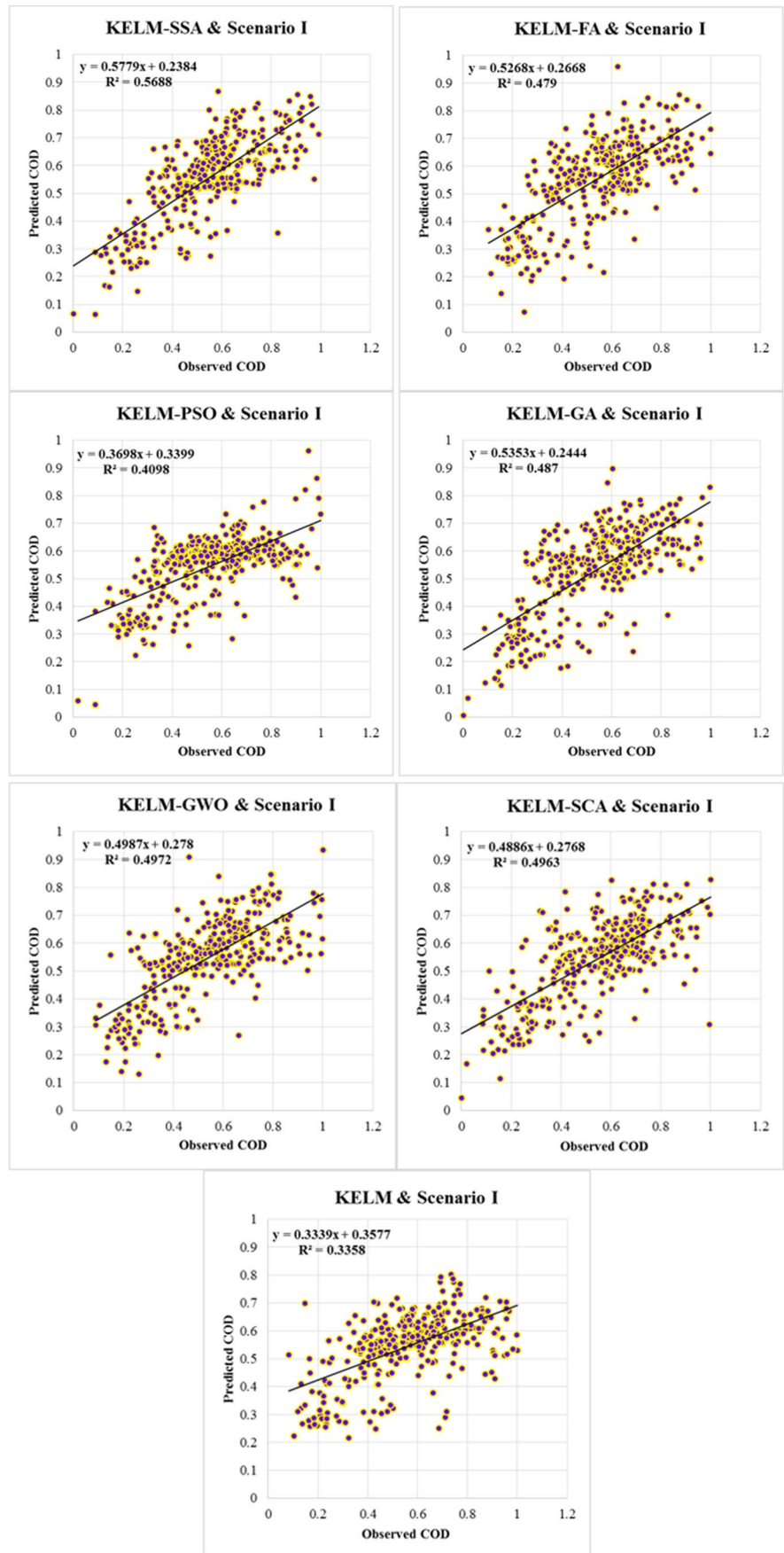


Table 4 The statistical performance of the hybrid models for COD prediction in training and testing phases (scenario II)

Models	RMSE	MAE	MAPE	NSE	WI	R^2
Training						
KELM-SSA	0.071	0.053	0.573	0.876	0.965	0.877
KELM-FA	0.078	0.056	0.615	0.847	0.956	0.849
KELM-PSO	0.143	0.117	0.286	0.495	0.781	0.512
KELM-GA	0.116	0.080	0.234	0.671	0.903	0.679
KELM-GWO	0.097	0.070	0.248	0.768	0.931	0.769
KELM-SCA	0.128	0.105	0.306	0.597	0.838	0.620
KELM	0.149	0.122	0.348	0.429	0.730	0.453
Testing						
KELM-SSA	0.120	0.092	0.218	0.664	0.903	0.677
KELM-FA	0.120	0.092	0.282	0.666	0.904	0.678
KELM-PSO	0.145	0.119	0.748	0.492	0.788	0.503
KELM-GA	0.129	0.091	0.207	0.582	0.877	0.600
KELM-GWO	0.115	0.089	0.226	0.668	0.897	0.670
KELM-SCA	0.135	0.111	0.273	0.550	0.815	0.571
KELM	0.169	0.137	0.381	0.415	0.704	0.453

Modelling results and analysis

As this study focuses on predicting real-time COD concentrations, the performance of different models was compared based on their relative efficiency. Obtained results for different input scenarios are presented in the following subsections.

The performance of basic and hybrid models (scenario I)

The basic and hybrid models' performance for scenario I during the training and testing phases is presented in Table 3. The results show the better capability of KELM-SSA in predicting inlet COD concentrations in the current time (with RMSE = 0.129 and NSE = 0.566) compared to other models. The performance of KELM-FA, KELM-GA, KELM-GWO and KELM-SCA networks was found relatively similar. The KELM-PSO and KELM showed the lowest performance with RMSE = 0.157 and 0.169 and NSE = 0.405 and 0.330, respectively.

To evaluate the performance of the paradigms and the influence of training in different scenarios, the r^2 values of the predicted and measured data were compared using scatterplots for the testing phase. The results presented in Fig. 6 reveal a better correlation ($r^2 = 0.5688$) for hybrid KELM-SSA than for the other models. The lowest coefficient of 0.3358 was obtained using the basic KELM network. The results indicate that the performance of models is unsatisfactory and needs to be improved.

The performance of models trained by input variables and time of records (scenario II)

Table 4 shows the effectiveness of adding the time of data (hour and day) as inputs. Following the addition of hours and weekdays, the networks showed different behaviour. The results showed significant improvement in the accuracy of all the models. The KELM-SSA model was found more reliable with minimum prediction errors. Though improvement in prediction accuracy was noticed for all the models, the ranking of the models based on the prediction accuracy was the same as that obtained for scenario I.

The scatterplots showing the performance of the models for scenario II are presented in Fig. 7. By adding time records, the highest correlation coefficients were obtained for KELM-FA ($r^2 = 0.6780$) and KELM-SSA ($r^2 = 0.6768$) for scenario II. However, the highest enhancement was noticed for KELM-FA, KELM and KELM-GWO by 41.55%, 35% and 34.7%, respectively.

The performance of models trained by the time of records and time-series data learning (scenario III)

Table 5 shows a significant increase in the performance of all models by incorporating a time-series pattern learning technique. Comparison of results obtained for scenarios II and III revealed that a higher improvement is possible using time-series learning than consumer patterns as input. The KELM-SSA model was still found to provide the highest prediction accuracy for scenario III. However, the highest enhancement in accuracy was obtained for KELM-PSO.

Fig. 7 The scatterplots of the observed and predicted inlet COD concentrations by different KELM models in testing phase (scenario II)

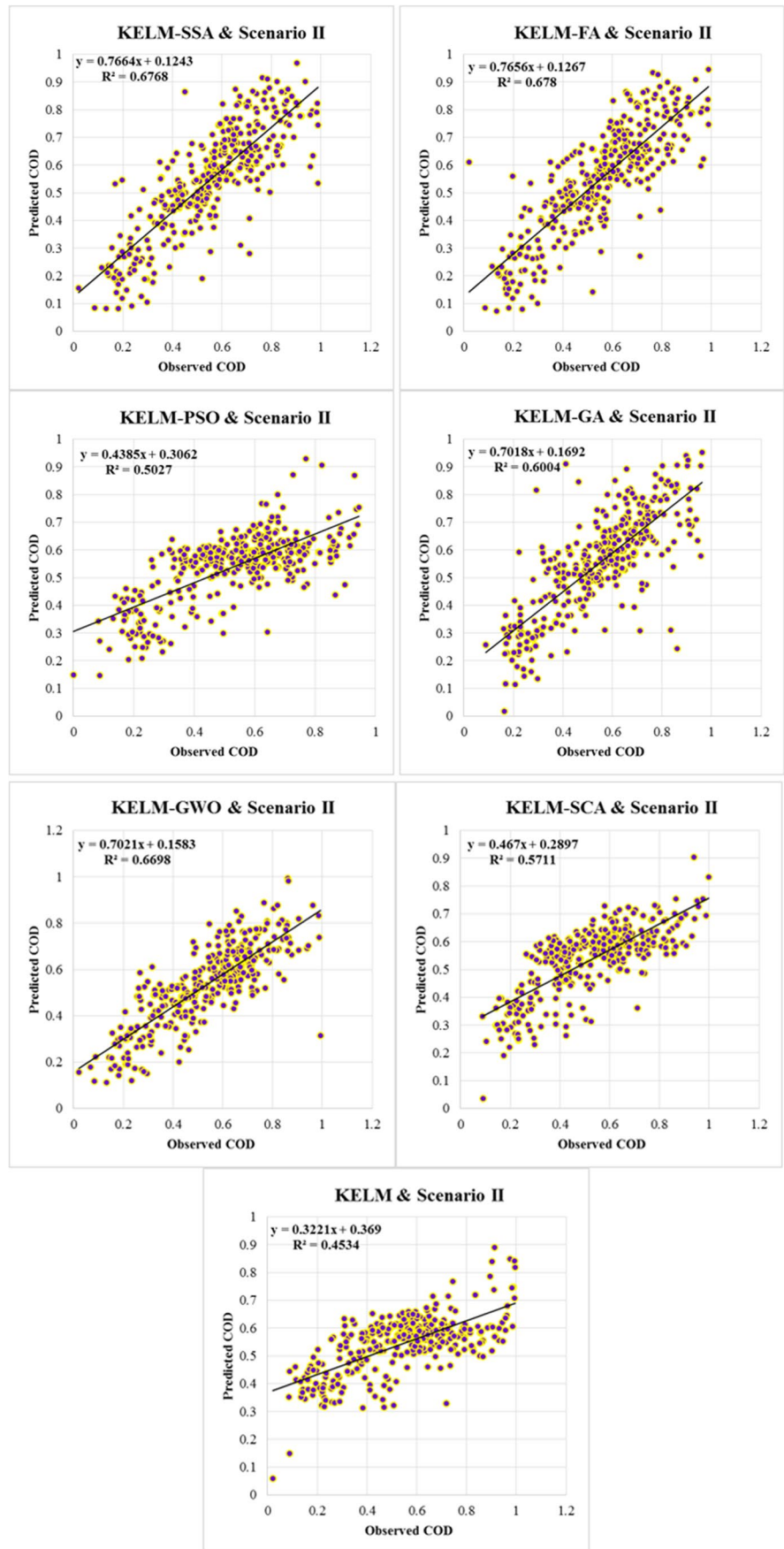


Table 5 The statistical performance of the hybrid models for COD prediction in training and testing phases (scenario III)

Models	RMSE	MAE	MAPE	NSE	WI	R^2
Training						
KELM-SSA	0.067	0.049	0.107	0.890	0.969	0.892
KELM-FA	0.073	0.052	0.300	0.873	0.964	0.874
KELM-PSO	0.073	0.052	0.309	0.871	0.964	0.871
KELM-GA	0.076	0.054	0.174	0.856	0.958	0.858
KELM-GWO	0.078	0.055	0.313	0.851	0.957	0.852
KELM-SCA	0.070	0.048	0.995	0.880	0.966	0.880
KELM	0.106	0.087	0.222	0.720	0.888	0.794
Testing						
KELM-SSA	0.071	0.051	0.117	0.876	0.966	0.876
KELM-FA	0.091	0.064	0.139	0.781	0.936	0.782
KELM-PSO	0.076	0.055	0.114	0.846	0.956	0.847
KELM-GA	0.091	0.063	0.140	0.803	0.944	0.803
KELM-GWO	0.090	0.060	0.157	0.797	0.940	0.798
KELM-SCA	0.083	0.059	0.134	0.831	0.954	0.832
KELM	0.116	0.093	0.353	0.687	0.871	0.758

Scatterplots showing the model performance for scenario II (Fig. 8) support the results obtained using the statistical metrics. The performance of KELM, KELM-PSO and KELM-SCA showed the highest increase in correlation (125.8%, 106.6% and 67.7%, respectively) compared to that obtained for scenario I. KELM-SSA and KELM were still the highest ($r^2=0.876$) and lowest ($r^2=0.7584$) performing models in predicting real-time inlet COD concentration, respectively. The largest improvement in the rank was noticed for KELM-PSO, which attained 2nd rank compared to 6th rank for scenario I.

The performance of models trained by the time of records and time-series data learning (scenario IV)

The statistical performance of the models for scenario IV is presented in Table 6. The incremental trend of model precision remained relatively constant for scenario IV. The results indicate that adding historical data up to two lags is optimum to obtain the best accuracy. However, the progress in the performance is still obtainable for the other models. Like the previous scenarios, the KELM-SSA model showed the best performance with RMSE = 0.072 and NSE = 0.870 (testing phase), while the KELM model still provided the poorest results.

Scatterplots of the models for scenario IV are presented in Fig. 9. The scatterplots show an inconsistency in results obtained using statistical metrics. The correlation coefficient of KELM-SSA dropped slightly ($r^2=0.8717$). The KELM still provides the least correlation but a slight improvement in accuracy (up to 5.3%) for this scenario.

The results revealed that the training paradigms considering time lag values as input and time-series learning up to a maximum of two lags can improve model prediction accuracy.

Comparison of scenario performance

Taylor diagrams were prepared to compare the prediction ability of the models for each scenario. The diagram can compare models' precision based on multiple statistical indices (Abba et al. 2020). The Taylor diagram for each scenario is given in Fig. 10. The distance of a model on the diagram from the observation (actual) exhibits their efficiency. The model which lies nearer to the actual point gives the best performance. The RMSE error is presented as proportional to the distance from the actual point, while the standard deviation (SD) of the predicted COD values is proportional to the radial distance from the origin of the diagram. In general, if the SD of the predicted values is higher than the SD of the observed data, the model is considered to overestimate the observation and vice versa. Figure 10 a shows that KELM-SSA and KELM-FA models could simulate the amplitude of the variations (i.e. the standard deviation) much better than the other models, leading to a smaller RMS error. The KELM-GA, KELM-SCA and KELM-GWO showed nearly the same correlation with the observed data. In contrast, the poorest performing models were KELM and KELM-PSO, which were far away from the SD line of observed data and showed the lowest correlation coefficient. Figure 10 b vividly demonstrates the increase in the accuracy of all models in scenario II. However, the models showed different

Fig. 8 The scatterplots of the observed and predicted inlet COD concentrations by different KELM models in testing phase (scenario III)

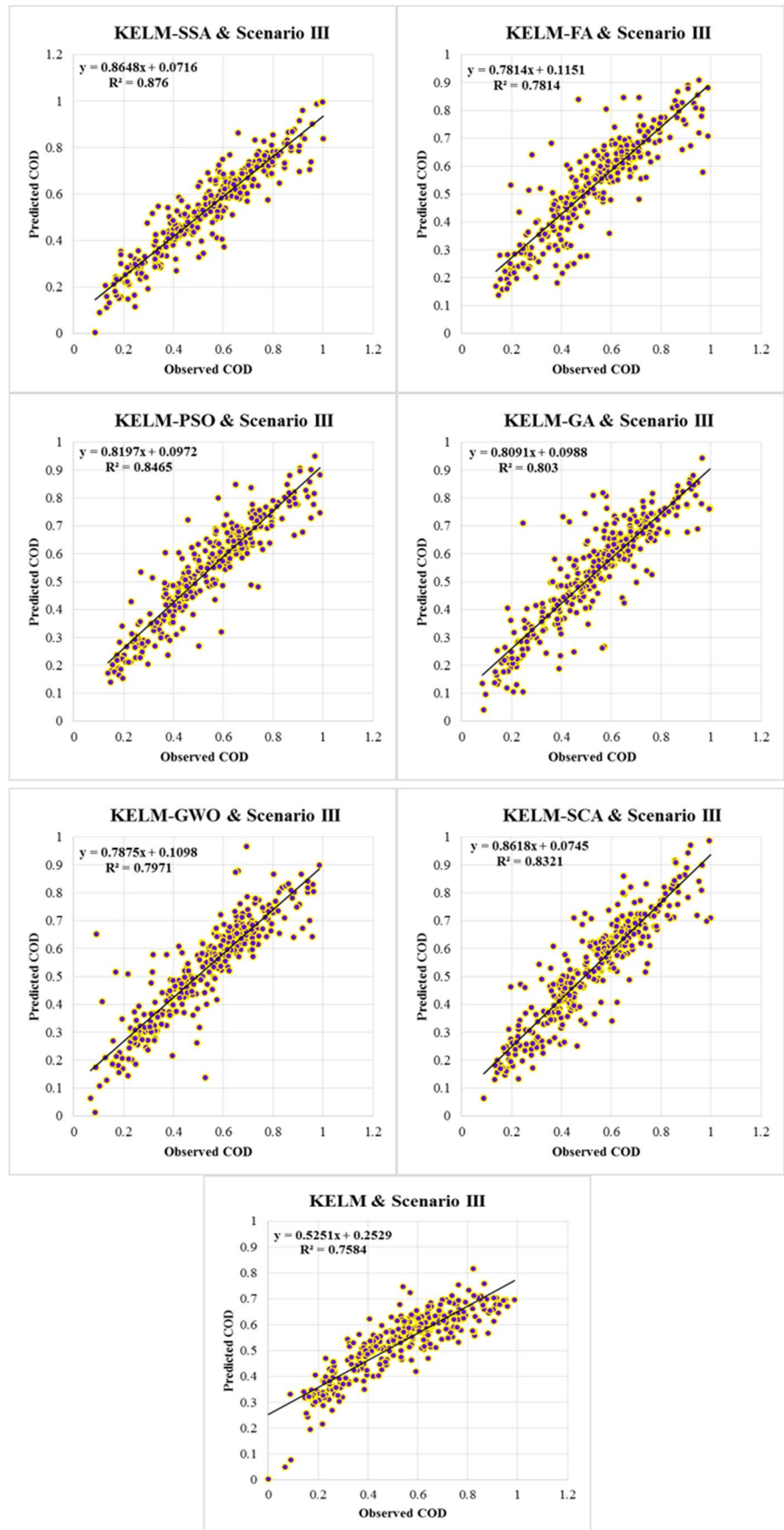


Table 6 The statistical performance of the hybrid models for COD prediction in training and testing phases (scenario IV)

Models	RMSE	MAE	MAPE	NSE	WI	R^2
Training						
KELM-SSA	0.058	0.044	0.101	0.918	0.977	0.920
KELM-FA	0.073	0.055	0.119	0.870	0.962	0.872
KELM-PSO	0.070	0.052	0.114	0.878	0.987	0.881
KELM-GA	0.066	0.049	0.108	0.893	0.970	0.895
KELM-GWO	0.071	0.052	0.115	0.877	0.964	0.880
KELM-SCA	0.063	0.045	0.103	0.901	0.972	0.902
KELM	0.095	0.075	0.513	0.781	0.926	0.801
Testing						
KELM-SSA	0.073	0.054	0.118	0.871	0.963	0.872
KELM-FA	0.085	0.063	0.132	0.812	0.945	0.813
KELM-PSO	0.081	0.058	0.131	0.847	0.955	0.850
KELM-GA	0.078	0.055	0.129	0.847	0.956	0.848
KELM-GWO	0.090	0.065	0.139	0.802	0.941	0.804
KELM-SCA	0.083	0.057	0.136	0.840	0.956	0.841
KELM	0.089	0.070	0.160	0.787	0.930	0.799

behaviours with the addition of new inputs. The accuracy enhancement of KELM-PSO, KELM-SCA and KELM models was lower than that of KELM-SSA, KELM-FA and KELM-GWO models. However, hybrid models were still superior to the basic KELM model, which shows the influence of optimization algorithms on the accuracy of ELM models. The performance of the KELM-SSA model was still closest to the actual point, with a correlation coefficient of nearly 0.7.

Figure 10 c shows the response of the models to the addition of time records and data of wastewater quality up to an hour antecedent value as inputs. A promising improvement was noticed for all the models. Unlike scenarios I and II, more similarity among the models except KELM was seen in correlation coefficient and SD. The KELM-PSO and KELM-SCA showed the poorest performances for scenario I, whereas the models rose to the second and third ranks in scenario III, respectively. This enhancement indicates a significant influence of time-series data learning on model performance. The KELM-SSA network as the best model for predicting real-time COD concentrations appeared to be much closer to the actual point for scenario III than scenarios I and II.

The performance of the model for scenario IV is presented in Fig. 10d. The figure shows a significant enhancement in the accuracy of the models by adding two lag values coupled with the time of records as inputs. In this scenario, the similarity of the model's response rose substantially. The accuracy enhancement of models was considerably effective, particularly for KELM and KELM-PSO networks. The KELM-SSA was still the

closest to the actual point, with a standard deviation very similar to the observed one. As the precision improvement in RMSE and correlation coefficient were negligible for the best network (KELM-SSA), historical data and time-series learning as inputs were not further considered. Therefore, the hybrid KELM-SSA model was considered the best model for predicting real-time influent COD. It can be remarked that time-series learning and adding the time of records as inputs led to a significant increase in the precision of KELM models. However, the SD of KELM-SSA simulated COD for all scenarios was less than the observed SD. Therefore, it can be inferred that the paradigms still do not simulate expected fluctuations, particularly above the average values.

Since Mashhad WWTP is known as a large-scale sewage treatment plant, energy-saving and cost-effective management is vital in the operation of the plant. Excess energy in aeration tanks is sometimes consumed in WWTP because the operators believe more oxygen is required for effluent treatment. The over-aeration is not cost-effective and environmentally friendly. The excess aeration leads to sludge settlement problems in the secondary clarifiers, which cause the sludge washout from the system. As a result, accurate prediction of real-time wastewater qualities, particularly COD value, can help to reduce extra costs in operation through automation. The results in this study show that the incorporation of advanced algorithms, time-series learning and considering water consumption patterns can enhance real-time COD prediction for providing optimum oxygen amounts linked directly to the PLC and control software.

Fig. 9 The scatterplots of the observed and predicted inlet COD concentrations by different KELM models in testing phase (scenario IV)

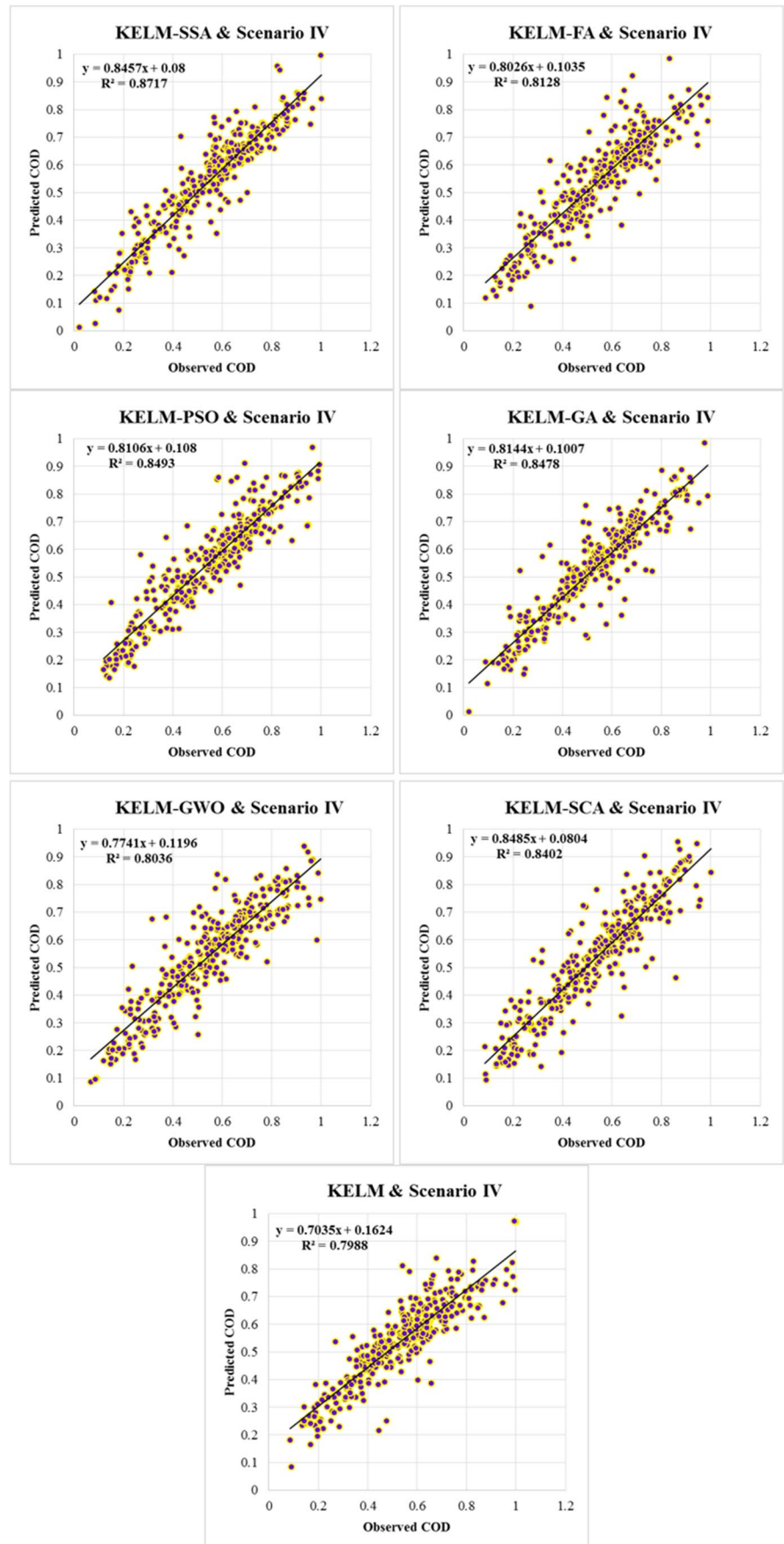
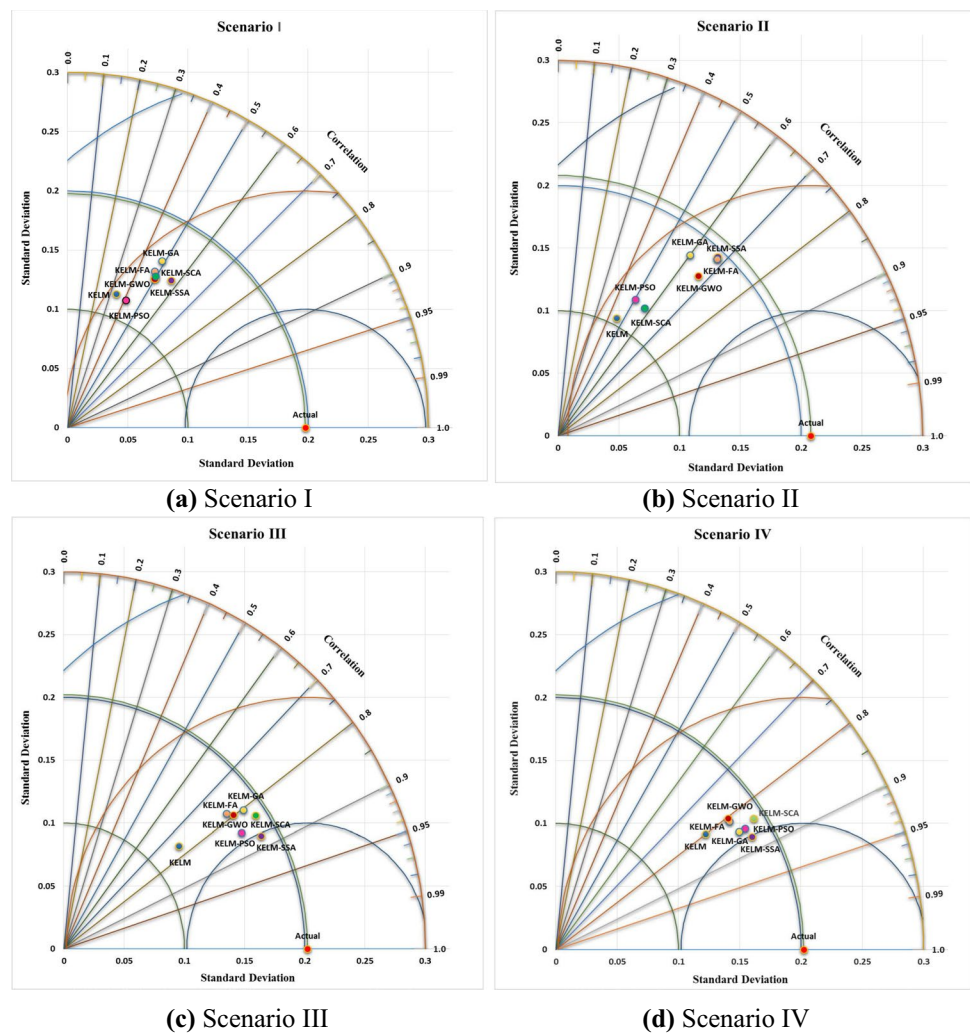


Fig. 10 Taylor diagram for displaying the correlation between actual and predicted inlet COD values by basic KELM model and optimized KELM models through intelligent algorithms: **a** scenario I, **b** scenario II, **c** scenario III and **d** scenario IV



Conclusion

The present study compared the performance of different hybrid machine learning models (i.e. KELM-SSA, KELM-FA, KELM-PSO, KELM-GA, KELM-GWO, KELM-SCA and standalone KELM) in predicting COD at the inlet of a WWTP. Several wastewater parameters, consumer behaviour and COD's time lag values were used to construct different modelling scenarios. The finding of the study can be summarized as follows:

- The hybrid machine learning models are robust and reliable methods for estimating COD at the inlet of a WWTP.
- KELM-SSA showed the best prediction capacity, which is apparently due to hybridization with the SSA optimization algorithm.
- The COD of the studied WWTP showed a high fluctuation, emphasizing the necessity for developing advanced machine learning models for COD prediction.

- Incorporating time lag values of COD as input is the most effective measure for improving hybrid KELM model performance.

By linking an accurate COD prediction model to an automation system, it is possible to manage the aeration supply precisely and prevent the sludge settling problems and malfunctions caused by the over-aeration in the WWTP.

Author contribution Javad Alavi: conceptualization, methodology and writing.

Ahmed A. Ewees: formal analysis, modelling, validation, evaluation and writing.

Sepideh Ansari: investigation, discussion, visualization and writing.

Shamsuddin Shahid: revision, editing, investigation, validation and discussion.

Zaher Mundher Yaseen: supervision, revision, editing, discussion and writing.

Data availability Data can be provided upon request from the corresponding author.

Declarations

Ethics approval The research was conducted and the manuscript was prepared following the ethical guidelines advised by the environmental science and pollution research journal.

Consent to participate Not applicable.

Consent for publication The research is scientifically consenting to be published.

Conflict of interest The authors declare no competing interests.

References

- Abba SI, Hadi SJ, Sammen SS et al (2020) Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination. *J Hydrol* 587:124974. <https://doi.org/10.1016/j.jhydrol.2020.124974>
- Abouzari M, Pahlavani P, Izaditame F, Bigdeli B (2021) Estimating the chemical oxygen demand of petrochemical wastewater treatment plants using linear and nonlinear statistical models—a case study. *Chemosphere* 270:129465
- Abyaneh HZ (2014) Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. *J Environ Heal Sci Eng* 12:40
- Al-Doury MMI, Alwan MH (2019) Phenol removal from synthetic wastewater using batch adsorption scheme. *Tikrit J Eng Sci* 26:31–36. <https://doi.org/10.25130/tjes.26.3.04>
- Al-Obaidi MA, Li JP, Kara-Zaitri C, Mujtaba IM (2017) Optimisation of reverse osmosis based wastewater treatment system for the removal of chlorophenol using genetic algorithms. *Chem Eng J* 316:91–100. <https://doi.org/10.1016/j.cej.2016.12.096>
- Al-Sulttani AO, Al-Mukhtar M, Roomi AB, et al (2021) Proposition of new ensemble data-intelligence models for surface water quality prediction. *IEEE Access*
- Al doury M, Al samerrai H (2019) Treatment of Al Doura oil refinery wastewater turbidity using magnetic flocculation. *Tikrit J Eng Sci* 26:1–8. <https://doi.org/10.25130/tjes.26.1.01>
- Alaba PA, Popoola SI, Olatomiwa L et al (2019) Towards a more efficient and cost-sensitive extreme learning machine: a state-of-the-art review of recent trend. *Neurocomputing* 350:70–90. <https://doi.org/10.1016/j.neucom.2019.03.086>
- Aras E, Toğan V, Berkun M (2007) River water quality management model using genetic algorithm. *Environ Fluid Mech* 7:439–450. <https://doi.org/10.1007/s10652-007-9037-4>
- Bernardelli A, Marsili-Libelli S, Manzini A et al (2020) Real-time model predictive control of a wastewater treatment plant based on machine learning. *Water Sci Technol* 81:2391–2400
- Bolyard SC, Motlagh AM, Lozinski D, Reinhart DR (2019) Impact of organic matter from leachate discharged to wastewater treatment plants on effluent quality and UV disinfection. *Waste Manag.* <https://doi.org/10.1016/j.wasman.2019.03.036>
- Chen X, Yin G, Zhao N, et al (2020) Simultaneous determination of nitrate, chemical oxygen demand and turbidity in water based on UV–Vis absorption spectrometry combined with interval analysis. *Spectrochim Acta Part A Mol Biomol Spectrosc* 118827
- Dairi A, Cheng T, Harrou F, et al (2019) Deep learning approach for sustainable WWTP operation: a case study on data-driven influent conditions monitoring. *Sustain Cities Soc* 50. <https://doi.org/10.1016/j.scs.2019.101670>
- Grady Jr CPL, Daigger GT, Love NG, Filipe CDM (2011) *Biological wastewater treatment*. CRC press
- Hai T, Sharafati A, Mohammed A et al (2020) Global solar radiation estimation and climatic variability analysis using extreme learning machine based predictive model. *IEEE Access* 8:12026–12042. <https://doi.org/10.1109/ACCESS.2020.2965303>
- He L, Tan T, Gao Z, Fan L (2019) The shock effect of inorganic suspended solids in surface runoff on wastewater treatment plant performance. *Int J Environ Res Public Health*. <https://doi.org/10.3390/ijerph16030453>
- He YL, Geng ZQ, Zhu QX (2015) Data driven soft sensor development for complex chemical processes using extreme learning machine. *Chem Eng Res Des* 102:1–11. <https://doi.org/10.1016/j.cherd.2015.06.009>
- Huang G-B, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern B Cybern* 42:513–529. <https://doi.org/10.1109/TSMCB.2011.2168604>
- Huang G, Zhu Q, Siew C (2004) Extreme learning machine: a new learning scheme of feedforward neural networks. In: *IEEE International Joint Conference on Neural Networks* 985–990
- Ibrahim A (2019) Effect of the horizontal perforated plates on the turbidity removal efficiency in water treatment plant of Tikrit University. *Tikrit J Eng Sci* 26:38–42. <https://doi.org/10.25130/tjes.26.4.06>
- Kim K, Kim D-K, Noh J, Kim M (2018) Stable forecasting of environmental time series via long short term memory recurrent neural network. *IEEE Access* 6:75216–75228
- Li Z, Huang S, Chen J (2019) A novel method for total chlorine detection using machine learning with electrode arrays. *RSC Adv* 9:34196–34206. <https://doi.org/10.1039/c9ra06609h>
- Lin M, Zhang C, Su C (2016) Prediction of effluent from WWTPS using differential evolutionary extreme learning machines. In: *2016 35th Chinese Control Conference (CCC)*. IEEE 2034–2038
- Liu H, Zhang Y, Zhang H (2020) Prediction of effluent quality in paper-making wastewater treatment processes using dynamic kernel-based extreme learning machine. *Process Biochem*
- Man Y, Hu Y, Ren J (2019) Forecasting COD load in municipal sewage based on ARMA and VAR algorithms. *Resour Conserv Recycl* 144:56–64. <https://doi.org/10.1016/j.resconrec.2019.01.030>
- Mirjalili S, Gandomi AH, Mirjalili SZ et al (2017) Salp swarm algorithm: a bio-inspired optimizer for engineering design problems. *Adv Eng Softw* 0:1–29. <https://doi.org/10.1016/j.advengsoft.2017.07.002>
- Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. *Adv Eng Softw*. <https://doi.org/10.1016/j.advengsoft.2013.12.007>
- Mirjalili SM, Mirjalili SZ, Saremi S, Mirjalili S (2020) Sine cosine algorithm: theory, literature review, and application in designing bend photonic crystal waveguides. In: *Nature-Inspired Optimizers*. Springer 201–217
- Naganna S, Deka P, Ghorbani M et al (2019) Dew point temperature estimation: application of artificial intelligence model integrated with nature-inspired optimization algorithms. *Water*. <https://doi.org/10.3390/w11040742>
- Najafzadeh M, Zeinolabedini M (2019) Prognostication of waste water treatment plant performance using efficient soft computing models: an environmental evaluation. *Measurement* 138:690–701
- Revilla M, Galán B, Viguri JR (2016) An integrated mathematical model for chemical oxygen demand (COD) removal in moving bed biofilm reactors (MBBR) including predation and hydrolysis. *Water Res.* <https://doi.org/10.1016/j.watres.2016.04.003>
- Salih SQ, Alakili I, Beyaztas U, et al (2020) Prediction of dissolved oxygen, biochemical oxygen demand, and chemical oxygen demand using hydrometeorological variables: case study of Selangor River, Malaysia. *Environ Dev Sustain* 1–20

- Sanikhani H, Deo RC, Yaseen ZM et al (2018) Non-tuned data intelligent model for soil temperature estimation: a new approach. *Geoderma* 330:52–64. <https://doi.org/10.1016/j.geoderma.2018.05.030>
- Sharafati A, Asadollah SBHS, Neshat A (2020) A new artificial intelligence strategy for predicting the groundwater level over the Rafsanjan aquifer in Iran. *J Hydrol*. <https://doi.org/10.1016/j.jhydrol.2020.125468>
- Sharafati A, Yasa R, Azamathulla HM (2018) Assessment of stochastic approaches in prediction of wave-induced pipeline scour depth. *J Pipeline Syst Eng Pract* 9. [https://doi.org/10.1061/\(ASCE\)PS.1949-1204.0000347](https://doi.org/10.1061/(ASCE)PS.1949-1204.0000347)
- Singh P, Kansal A (2018) Energy and GHG accounting for wastewater infrastructure. *Resour Conserv Recycl* 128:499–507. <https://doi.org/10.1016/j.resconrec.2016.07.014>
- Srinu Naik S, Pydi Setty Y (2014) Optimization of parameters using response surface methodology and genetic algorithm for biological denitrification of wastewater. *Int J Environ Sci Technol* 11:823–830. <https://doi.org/10.1007/s13762-013-0266-4>
- Tchobanoglous G, Burton FL, Stensel D (2003) *Wastewater engineering: treatment and reuse*, 4th edn. McGraw-Hill Companies Inc, New York
- Tiyasha T, Tung TM, Bhagat SK et al (2021) Functionalization of remote sensing and on-site data for simulating surface water dissolved oxygen: development of hybrid tree-based artificial intelligence models. *Mar Pollut Bull* 170:112639
- Tung TM, Yaseen ZM (2020) A survey on river water quality modelling using artificial intelligence models: 2000–2020. *J Hydrol* 585:124670
- Verma A, Wei X, Kusiak A (2013) Predicting the total suspended solids in wastewater: a data-mining approach. *Eng Appl Artif Intell* 26:1366–1372. <https://doi.org/10.1016/j.engappai.2012.08.015>
- Wang Z, Man Y (2021) Artificial intelligence algorithm application in wastewater treatment plants: case study for COD load prediction. In: *Applications of Artificial Intelligence in Process Systems Engineering*. Elsevier 143–164
- Wang Z, Man Y, Hu Y et al (2019) A deep learning based dynamic COD prediction model for urban sewage. *Environ Sci Water Res Technol* 5:2210–2218. <https://doi.org/10.1039/c9ew00505f>
- Wei C, Wu H, Kong Q et al (2019) Residual chemical oxygen demand (COD) fractionation in bio-treated coking wastewater integrating solution property characterization. *J Environ Manage*. <https://doi.org/10.1016/j.jenvman.2019.06.001>
- Wei X, Kusiak A (2014) Short-term prediction of influent flow in wastewater treatment plant. *Stoch Environ Res Risk Assess* 29:241–249. <https://doi.org/10.1007/s00477-014-0889-0>
- Yang XS (2010) Firefly algorithm, leavy flights and global optimization. *Res Dev Intell Syst* 135–146. <https://doi.org/10.1007/978-1-84882-983-1>
- Yaseen ZM (2021) An insight into machine learning models era in simulating soil, water bodies and adsorption heavy metals: review, challenges and solutions. *Chemosphere* 130126
- Yaseen ZM, Ali M, Sharafati A, et al (2021) Forecasting standardized precipitation index using data intelligence models: regional investigation of Bangladesh. *Sci Rep* 11. <https://doi.org/10.1038/s41598-021-82977-9>
- Yu T, Bai Y (2019) Comparative study of optimization intelligent models in wastewater quality prediction. *Proc - 2018 Int Conf Sensing Diagnostics, Progn Control SDPC 2018*:221–225. <https://doi.org/10.1109/SDPC.2018.8664791>
- Yu Q, Liu R, Chen J, Chen L (2019) Electrical conductivity in rural domestic sewage: an indication for comprehensive concentrations of influent pollutants and the effectiveness of treatment facilities. *Int Biodeterior Biodegradation* 143:104719
- Zhang S, Tan W, Wang Q, Wang N (2019) A new method of online extreme learning machine based on hybrid kernel function. *Neural Comput Appl* 31:4629–4638. <https://doi.org/10.1007/s00521-018-3629-4>
- Zhu JJ, Kang L, Anderson PR (2018) Predicting influent biochemical oxygen demand: balancing energy demand and risk management. *Water Res* 128:304–313. <https://doi.org/10.1016/j.watres.2017.10.053>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.