ORIGINAL ARTICLE

Feasibility of a novel predictive model based on multilayer perceptron optimized with Harris hawk optimization for estimating of the longitudinal dispersion coefficient in rivers

Sima Ohadi[1] · Seyed Arman Hashemi Monfared[1] (ORCID) · Mehdi Azhdary Moghaddam[1] · Mohammad Givehchi[1]

Abstract

Protecting water resources from pollution is one of the most important challenges facing water management researchers. The governing equation for river pollution is mostly the advection–dispersion equation, with considering the longitudinal dispersion coefficient as its most important effective parameter. The purpose of this paper is to develop a new framework for accurate prediction of the longitudinal dispersion coefficient of rivers based on artificial intelligence (AI) methods. To do this, we used a combination of multilayer perceptron (MLP), one of the most robust neural networks, and a novel metaheuristic algorithm, namely Harris hawk optimization (HHO). Besides, two optimized MLP models with particle swarm optimization (PSO) and imperialist competitive algorithm (ICA) were utilized to demonstrate the accuracy of the proposed model. To evaluate the developed models, 164 series of data collected from previous studies, including hydraulic and geometric parameters of rivers, were used. The indicated results proved the efficiency of the HHO to improve the optimum auto-selection of the AI models. Thus, the recorded results show very high accuracy of the newly developed model, MLP-HHO compared to others. Furthermore, to increase the prediction accuracy, a K-means clustering technique is coupled with MLP-HHO model during dividing the data to train and test categories. The proposed hybrid K-means-MLP-HHO model with coefficient of determination (R^2) and root mean square error (RMSE), of 0.97 and 30.94 m^2/s, respectively, significantly outperformed all existing and AI-based models. Furthermore, the sensitivity analysis showed that the flow width is the most influential factor in predicting the longitudinal dispersion coefficient.

Keywords Longitudinal dispersion coefficient · Hybrid artificial models · Multilayer perceptron · Harris hawk optimization

1 Introduction

Due to the current situation of world water resources and the increasing problem of surface and groundwater pollution, water quality management is prioritized by many managers and researchers in this field. A source of contamination released into the water is rapidly diffused into it and transported downstream along with the water flow due to molecular motion, turbulence and uneven velocity at the cross-sectional area of the flow [1].

The governing equation for pollutant transmission is the advection–diffusion equation, expressed as Eq. (1) [2].

$$\left(\frac{\partial C}{\partial t}\right) + U\left(\frac{\partial C}{\partial x}\right) = K\left(\frac{\partial^2 C}{\partial x^2}\right) \tag{1}$$

where C is the cross-sectional average concentration $\left(\frac{kg}{m^3}\right)$, U is the cross-sectional average velocity $\left(\frac{m}{s}\right)$, x is direction of the mean flow, t is time (s) and K is representing the longitudinal dispersion coefficient in rivers $\left(\frac{m^2}{s}\right)$.

It is a partial differential equation that can be solved using numerical methods and defining the appropriate boundary and initial conditions [3, 4]. According to this equation, the longitudinal dispersion coefficient is considered the most important parameter that the accurate estimation of which is critical in hydraulic engineering [5]. This coefficient can be measured directly by sampling the

✉ Seyed Arman Hashemi Monfared
  hashemi@eng.usb.ac.ir

[1] Department of Civil Engineering, Faculty of Engineering, University of Sistan and Baluchestan, Zahedan, Iran

upstream and downstream sections of the river. This is difficult due to the complex geometry of the natural channel bed and the various hydraulic prevailing conditions. Thus, it is usually calculated using empirical and theoretical relations [6].

The longitudinal dispersion coefficient was firstly introduced by Taylor as a measure of the longitudinal dispersion process by the advection–diffusion equation [7]. Then, Elder began to extend the Taylor method. Assuming that the velocity profile in the vertical direction is logarithmic, he proposed Eq. (2) to estimate the longitudinal dispersion coefficient for wide channels [8].

$$\frac{K}{HU_*} = 5.93 \tag{2}$$

where H and U_* are the depth of flow and the bed shear velocity, respectively.

Although the Elder's equation was considered a great success in predicting the longitudinal dispersion coefficient, ignoring changes in cross-sectional velocity profiles showed a large difference between the observed values and the predicted values [1, 9]. According to Fisher, the transverse velocity profile was more important than the vertical velocity profile for predicting the longitudinal dispersion coefficient in the wide channel. Accordingly, he proposed a triangular integral [Eq. (3)] to estimate the longitudinal dispersion coefficient in wide channels [10].

$$K = -\frac{1}{A} \int_0^W hu' \int_0^y \frac{1}{\varepsilon_y h} \int_0^y hu' \mathrm{d}y \, \mathrm{d}y \, \mathrm{d}y \tag{3}$$

in which A and W represent the cross-sectional area and the channel width, respectively. h is the local flow depth, u' is the difference between the local depth-averaged velocity and the cross-sectional average velocity and ε_y is the transverse turbulent diffusion coefficient.

Many attempts have been made so far, such as tracking measurements, empirical models and theoretical relationships, to estimate the longitudinal dispersion coefficient [5].

Kashefipour and Falconer proposed a new equation for predicting the longitudinal dispersion coefficient for natural channels using dimensional analysis and regression of hydraulic flow parameters. A comparison of statistical parameters showed that the method was more accurate than other equations for calculating the longitudinal dispersion coefficient. By linearly combining of their proposed equation with the equation proposed by Soe and Cheong, they also proposed another equation to estimate the longitudinal dispersion coefficient. The efficiency of this equation was proved by analyzing it using statistical methods [11].

Continuing efforts to find a more accurate way to calculate the dispersion coefficient, optimization algorithms have become more popular in recent decades due to the high speed of soft computing methods. Sahay and Dutta used the genetic algorithm (GA) and Etemad-Shahidi and Taghipour used the M5 tree model, a flexible computational method, to estimate the longitudinal dispersion coefficient of natural rivers [9, 12]. Li et al. used the differential evolution (DE) algorithm to estimate the longitudinal dispersion coefficient of rivers by minimizing the sum-square error. This study used observational data from 29 rivers in the USA to evaluate the performance of the proposed algorithm [13]. Sattar and Gharabaghi used the gene expression programming (GEP) model to develop an empirical relationship between longitudinal dispersion coefficient and control variables including Froude number, dimensional ratio and bed roughness. For this purpose, they used 150 series of natural flow data including geometric and hydraulic parameters of the flow. Comparing the proposed model with other methods in terms of uncertainty indicated the higher accuracy of the GEP model [14]. Alizadeh et al. used the Bayesian network method to provide a model for predicting the longitudinal dispersion coefficient of natural rivers. To increase the accuracy of the proposed model, they used cluster analysis for data processing couples with Bayesian network [6]. Kargar et al. used the M5 model tree (M5P) model to predict longitudinal dispersion coefficient as a function of flow depth, channel width, average velocity and shear velocity [15].

Consequently, many different kinds of techniques have been developed for estimating the longitudinal dispersion coefficient in rivers. They consist of soft computing techniques (e.g., GA, DE, PSO, etc.), artificial intelligence methods (e.g., ANN, GEP, Model Tree, etc.) and empirical (e.g., mathematical, statistical, etc.). A brief description of previous models/equations is presented in Table 1.

According to the investigations, there are various experimental studies on empirical equations for estimating the longitudinal dispersion coefficient in rivers; however, they are limited with the laboratory conditions (e.g., bed configuration, channel width and flow properties) and they have not reliable results for all types of channels [1, 16, 17]. In addition, the used AI-based models, in previous studies, shown some limitations like slow convergence, needs to find the best tuning parameters of method and other drawbacks that need user interaction. Therefore, it is necessary to develop a model with higher accuracy to estimate the longitudinal dispersion coefficient [11]. Among AI-based predictive models, ANN has high teaching and learning power. It has been used in recent years, especially to predict complex issues [18–24]. However, it is challenging for users to choose the suitable ANN parameters, such as size of input layers, learning curve and

Table 1 Empirical and artificial intelligence-based equations for prediction of longitudinal dispersion coefficient

Author(s) (year)	Category	Formula
Seo and Cheong [63]	Empirical	$\frac{K}{HU_*} = 5.915 \left(\frac{U}{U_*}\right)^{1.428} \left(\frac{W}{H}\right)^{0.62}$
Deng et al. [64]	Empirical	$\frac{K}{HU_*} = \frac{0.15}{8\varepsilon_t} \left(\frac{U}{U_*}\right)^2 \left(\frac{W}{H}\right)^{5/3}$
Zeng and Huai [65]	Empirical	$\frac{K}{HU_*} = 5.4 \left(\frac{W}{H}\right)^{0.7} \left(\frac{U}{U_*}\right)^{1.13}$
Disley et al. [57]	Empirical	$\frac{K}{HU_*} = 3.563 \left(\frac{U}{\sqrt{gH}}\right)^{-0.4117} \left(\frac{W}{H}\right)^{0.6776} \left(\frac{U}{U_*}\right)^{1.0132}$
Etemad-Shahidi and Taghipour [12]	Artificial intelligence	$\frac{K}{HU_*} = \begin{cases} 15.49 \left(\frac{W}{H}\right)^{0.78} \left(\frac{U}{U_*}\right)^{0.11} & if \ \frac{W}{H} \leq 30.6 \\ 14.12 \left(\frac{W}{H}\right)^{0.61} \left(\frac{U}{U_*}\right)^{0.85} & if \ \frac{W}{H} > 30.6 \end{cases}$
Alizadeh et al. [6]	Soft computing	$\frac{K}{HU_*} = \begin{cases} 5.319 \left(\frac{W}{H}\right)^{1.206} \left(\frac{U}{U_*}\right)^{0.075} & if \ \frac{W}{H} \leq 28 \\ 9.931 \left(\frac{W}{H}\right)^{0.187} \left(\frac{U}{U_*}\right)^{1.802} & if \ \frac{W}{H} > 28 \end{cases}$
Kargar et al. [15]	Artificial intelligence	$\frac{K}{HU_*} = \begin{cases} 1.6896 \left(\frac{W}{H}\right) + 20.0124 \left(\frac{U}{U_*}\right) + 393.3346 & if \ \frac{W}{H} \leq 47.23 \\ 2.8759 \left(\frac{W}{H}\right) + 181.7915 \left(\frac{U}{U_*}\right) + 339.5557 & if \ \frac{W}{H} > 47.23 \end{cases}$
Memarzadeh et al. [16]	Artificial intelligence	$\frac{K}{HU_*} = \begin{cases} 0.2694 \left(\frac{W}{H}\right)^{2.2456} & if \ \frac{W}{H} \leq 27 \\ \left(0.35 + 8.7 \left(\frac{H}{W}\right)\right) \times \left(6.4 + 8 \left(\frac{W}{H}\right)\right) \times \left(\frac{U}{U_*}\right)^{0.5} & if \ \frac{W}{H} > 27 \end{cases}$

the number of hidden layers. These parameters are generally determined based on the user experience using the MATLAB toolbox with a limited number of learning algorithms, leading to non-optimized ANN.

Nowadays, well-known metaheuristic algorithms such as PSO, GA and ABC are exploited as tools to find optimal ANN parameters in many cases, such as predicting the axial compression capacity of columns [25–27]; modeling evaporation [28–30]; and prediction of river flow [31–33].

Metaheuristic optimization techniques are considered for their simplicity, flexibility, derivation-free mechanism and overcoming local optimization problem. But according to the NFL[1] theory, none of the metaheuristic methods are appropriate for solving all optimization problems [34, 35]. In other words, a particular metaheuristic may show incredible results to solve a series of specific problems, but this algorithm may have poor performance for a number of other issues [36]. Therefore, the original contribution of this work is investigating the capability of a new hybrid expert system (MLP combined with a recent metaheuristic algorithm, namely HHO) to predict the longitudinal dispersion coefficient in rivers so as to provide a robust model

for solving such a complex environmental engineering problem. In order to evaluate the feasibility of the proposed HHO-based training algorithm in improving prediction performance of MLP, this training tool was compared with the other two powerful known metaheuristic optimizers, ICA and PSO by integrating into MLP as an efficient training tool for estimating the longitudinal dispersion coefficient in rivers. A review of studies on the methods of determining the longitudinal dispersion coefficient of rivers indicates that hybrid HHO-MLP has not been used for this purpose up to now. Current study proposes three hybrid models: HHO-MLP, PSO-MLP and GA-MLP, for streamflow forecasting problems. For this purpose, 164 observed data set was gathered. In the first stage, the longitudinal dispersion coefficient under influence hydraulic parameters including depth, width, velocity and shear velocity was predicted by each of these developed models with the aim of minimizing the mean square error. Then, the observed data are compared with the results of the above models and the most powerful algorithm for optimizing MLP has been identified using the statistical indexes.

The rest of the paper is organized as follows: in Sect. 2 describes the proposed hybrid models. The details of

---

[1] No Free Lunch.

databases and the employed statistical indicators are presented in Sects. 3 and 4, respectively. The results of the numerical experiment and their discussions are presented in Sect. 5 while some conclusion remarks are given in Sect. 6.

2 Materials and methods

2.1 Concept and theory background

According to previous studies, some of the important parameters affecting the longitudinal dispersion coefficient of rivers are:

(1) The geometric properties of the flow, including the width of the flow path (W), the bed shape factor (S_f), the slope of the channel floor (S), the sinuosity (σ), the coefficient of the roughness of the channel (k_s) and the depth of flow (H);

(2) Hydraulic properties of flow, including the average flow velocity (U) and shear velocity (U_*); and

(3) The properties of the fluid, including the density of the fluid (ρ) and the kinematic viscosity of the flow (μ).

which can be expressed as Eq. (4) using dimensional analysis [37]:

$$\frac{K}{U_* H} = f'\left(\rho \frac{HU}{\mu}, \frac{W}{H}, \frac{U}{U_*}, \sigma, \frac{U}{\sqrt{gH}}, \frac{k_s}{H}, S_f, S \right) \quad (4)$$

where $\frac{K}{U_* H}$ is the dimensionless dispersion coefficient parameter, $\rho \frac{HU}{\mu}$ is the Reynolds number, $\frac{W}{H}$ is the width-to-depth ratio and $\frac{U}{U_*}$ is the flow resistance term. Here, the direct influence of parameters such as the slope of the channel, the shape of the bed and the sinusoidal value of the river path, which are not easily measurable and their effect can be seen in the term of resistance, is omitted. The effect of the Reynolds number can be also ignored due to the usually turbulent flow of rivers. Equation (4) can finally be written as Eq. (5) [6]:

$$\frac{K}{U_* H} = f''\left(\frac{W}{H}, \frac{U}{U_*} \right) \quad (5)$$

where the longitudinal dispersion coefficient (K) depends on the following four parameters: U average flow velocity, H flow depth, W width of flow and U_* bed shear velocity.

2.2 Artificial neural network (ANN)

ANN is an idea for processing information like the brain, inspired by the biological nervous system. This system comprises many processing elements called neurons, which work in unison to solve a problem [38]. ANN is considered a tool for estimating the approximate performance, applicable in places where there is a nonlinear and complex relationship between input(s) or predictor(s) and output(s) [39, 40].

ANNs need to be trained with certain learning algorithms to approximate the goal and achieve the desired results [38, 41]. Numerous network learning algorithms have been developed by examining simple, nonlinear mathematical solutions based on human biological neurons. Learning algorithms are a series of processes, which help to adjust the weights of the network. During the training process, a network can learn tasks and similarities without a predefined program. The learning system usually lacks any prior information about the problem such as various input parameters set in the input data layers [42, 43]. The most common learning algorithm is the backpropagation (BP) algorithm. It is developed based on the downward slope optimization method, which can minimize the network error between the desired values and the estimated values [41, 44].

One of the most common neural networks is the multi-layer perceptron. It is also known as backpropagation because its training algorithm is usually backpropagation. The perceptron neural network method was introduced by Rosenblatt [45]. The input layer only acts as an intermediary with the outside world. Each node sends an unchanged input to the next layer of neurons. Therefore, the number of input layer nodes is equal to the number of independent variables. The number of neurons in the output layer indicates the number of dependent variables. Also, the number of nodes in each layer will depend on the complexity of the problem.

2.3 Metaheuristic algorithm

Metaheuristic algorithms are a flexible method for solving complex optimization problems based on swarm intelligence, inspired by biological and physical systems in nature. These methods use a special mechanism that many particles are scattered in the problem space and simultaneously search for optimal solutions. These methods are called nature-inspired, population-based or metaheuristic algorithms. Considering various constraints, metaheuristic algorithms also achieve a set of non-dominant solutions as the best solution of the solution set, or hopefully have a short distance from the original solution set [46, 47].

2.3.1 Particle swarm optimization algorithm (PSO)

The particle swarm optimization (PSO) algorithm was first introduced by Eberhart, inspired by the social behavior of birds [48]. This algorithm considers each solution as a bird

in a bird group, called a particle. In such a structure, birds have individual intelligence and group behavior, which leads them to the goal.

First, the process begins with a bunch of randomly selected particles, each as a solution to the problem. Then, a bunch of particles move and the search space is searched to find the best solution.

In this algorithm, each particle has an adaptive velocity, representing the vector of the particle's move in the search space and memory, which means it remembers its best position up to that point. Particles are measured in terms of value based on the criterion function of the optimization

$$v_i^{k+1} = w \times v_i^k + c_1 \times r_1 \times \left(P_{best_i} - x_i^k\right) + c_2 \times r_2 \\ \times \left(g_{best_i} - x_i^k\right) \tag{6}$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \tag{7}$$

where v_i^k and x_i^k are the velocities and positions of the particle i in the kth iteration, respectively. c_1 and c_2 are acceleration constants, r_1 and r_2 are also random numbers obtained from a uniform distribution between 0 and 1. w is the inertia weight factor. Algorithm 1 presents the PSO pseudocode.

Algorithm 1 Pseudo-code of PSO algorithm.

Start

Initialize the position and velocity randomly for each position

   For *each particle*

      Calculate fitness value

      If the fitness value is better than the best fitness value (P_{best}) in history

      Set current value as the new P_{best}

   End

Choose the particle with the best fitness value of all the particles as the g_{best}

   For *each particle*

      Calculate particle velocity according Eq. 6

      Update particle position according Eq. 7

   End

   Wille maximum iterations or minimum error criteria is not attained

   End

---

problem [49]. For each particle, the best solution and position are stored in the variables P_{best} and g_{best}, respectively [48]. Then, in each iteration, the particles accelerate toward their P_{best} and g_{best} based on the velocity vector v_i. As a result, in each iteration, a new position is found in the search space for each particle by adding their displacement vector. Finally, g_{best} is considered the optimal solution to the problem. The velocity (v_i^{k+1}) and new position (x_i^{k+1}) of the particles are updated in each iteration by the following equations [50]:

2.3.2 Imperialist competitive algorithm (ICA)

The imperialist competitive algorithm (ICA) is inspired from a human social phenomenon. Like other evolutionary algorithms, this algorithm starts with an initial population. Each member of this population is referred to as a "country." Each country has a set of characteristics, which determine its location in the search space. Among these points, the points with the lowest cost, according to the optimization function, are considered imperialist and the

rest as colonies [51]. The normalized cost for each colonizer is initially calculated as follows:

$$C_n = c_n - \max_i\{c_i\} \tag{8}$$

where c_n is the cost of the nth imperialist and C_n is the normalized cost for this imperialist. The normalized power for each imperialist (P_n) is calculated by Eq. (9). On this basis, the colonial countries are divided among these imperialists:

$$P_n = \left| \frac{C_n}{\sum_{i=1}^{N_{imp}} C_i} \right| \tag{9}$$

The initial number of colonies of an imperialist will be equal to:

$$N \cdot C_{.n} = round\{P_n, N_{col}\} \tag{10}$$

where $N \cdot C_{.n}$ is the initial number of colonies of nth empire and N_{col} is the total number of colonies in the population of the primary countries.

Following the division of all the colonies between the emperors and the formation of the first empires, the colonies began to move toward the emperors. This movement models the policy of absorption. Figure 1 shows the movement of a colony toward the emperor.

As shown in Fig. 1, d is the direct distance between the imperialist and the colonies. A colony moves toward the imperialist, not directly and linearly, but at an angle θ to the size of x. Then the amount of deflection angle and the size of move are determined uniformly in randomly determined intervals.

$$X \sim U(0 \cdot \beta \times d) \tag{11}$$
$$\theta \sim U(-\gamma \cdot \gamma) \tag{12}$$

where β and γ are arbitrary numbers, which determine the clone search space around the empire.

As colonial countries move toward imperialist countries, some of them may achieve better colonial status. In this case, the colonial country and the imperialist country
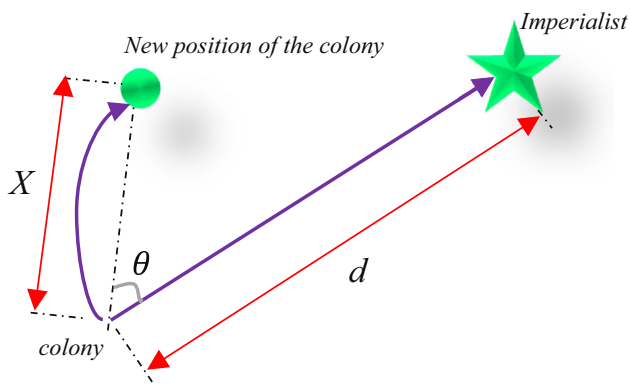

Fig. 1 Moving colonies toward their relevant imperialist

change their place together. This is followed by colonial competition. At this stage, the weakest colony is selected from the weakest empire and given to a stronger empire. The stronger an empire, the more likely it is to be chosen. Finally, when an empire loses all its colonies, it is excluded from the list of empires and granted to other empires as a colony during a colonial rivalry. The evolutionary process takes place in a loop and continues until a stop criterion is met.

2.3.3 Harris hawk optimization (HHO)

Harris hawk optimization (HHO) is a population-based optimization algorithm proposed by Heidari et al. [52]. Inspired by Harris hawks' behavior in nature, this algorithm is based on cooperation between falcons to hunt prey [53, 54]. It uses a solid strategy: a group of Harris hawks launches a joint attack from different directions to catch the prey by surprise. These hawks cooperate in the process of the attack. Meanwhile, the Harris hawks' leader attacks the target prey, chases it and suddenly disappears. The next Harris hawk continues the pursuit action. At this point, the Harris hawk utilizes different attack strategies for different prey escape modes. In this way, the prey is exhausted and eventually hunted. In the HHO algorithm, these steps are mathematically modeled in three basic steps: exploration, transition and exploitation (Fig. 2) [52–55].

The exploration phase involves waiting, searching and discovering the proposed prey. In other words, this step determines the position of the hawk before finding the prey [55].

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1|X_{rand}(t) - 2r_2X(t)| & q \geq 0.5 \\ (X_{rabbit}(t) - X_m(t)) - r_3(LB + r_4(UB - LB)) & q < 0.5 \end{cases} \tag{13}$$

where $X(t)$ and $X(t+1)$ represent the position of the hawks in the t and $t+1$ iteration, respectively. $X_{rabbit}(t)$ indicates the position of the prey. r_1, r_2, r_3, r_4 and q are random numbers inside (0, 1), which are updated during each iteration. LB and UB represent the lower and upper bound of the decision variables, respectively. X_{rand} is the position of a randomly selected hawk. Concerning $X_i(t)$ as the position of each hawk in the iteration of t and N as the total number of hawks, X_m, representing the average position of the hawks, can be calculated using the following equation [52]:

$$X_m(t) = \frac{1}{N}\sum_{i=1}^{N} X_i(t) \tag{14}$$

The intermediate phase is defined as transition from exploration to exploitation which expressing the function of wasting energy prey during the escape. Due to the

reduced energy of the prey during the escape, the prey energy is modeled using the following equation:

$$E = 2E_0\left(1 - \frac{t}{T}\right) \tag{15}$$

where E represents the energy of escape of the prey and T represents the maximum number of iterations. E_0, varying randomly in the range $(-1, 1)$ per repetition, indicates the initial state of prey energy. If $|E| \geq 1$, the HHO algorithm executes the exploration phase; However, if $|E| < 1$, exploitation occurs.

In the third phase of the HHO algorithm, entitled exploitation, four surprise attack models are proposed based on prey escape behavior and hawk pursuit strategies. In this phase, r is considered the prey chance parameter. $r < 0.5$ means a successful escape of prey and contrariwise. In this phase, based on the escape energy E, soft besiege $(|E| \geq 0.5))$ or hard besiege $((|E| < 0.5))$ is performed by Harris hawk (Fig. 1) [52, 56]. Algorithm 2 presents the HHO pseudocode.

Algorithm 2 Pseudo-code of HHO algorithm.

Inputs: *The population size N and maximum number of iterations T*

Outputs: *The location of rabbit and its fitness value*

Initialize the random population X_i(i = 1, 2, . . . , N)

while *(stopping condition is not met)* **do**

    Calculate the fitness values of hawks

    Set X_{rabbit} as the location of rabbit (best location)

    for *(each hawk (X_i))* **do**

        Update the initial energy $E_0$and jump strength J   ▷

$E_0 = 2rand() - 1, J = 2(1 - rand())$

        Update the E

        if *($|E| \geq 1$)* **then**   ▷ *Exploration phase*

            Update the location vector

        if *($|E| < 1$)* **then**   ▷ *Exploitation phase*

            if *(r $\geq$ 0.5 and $|E| \geq$ 0.5)* **then**   ▷ *Soft besiege*

                Update the location vector

            else if *(r $\geq$ 0.5 and $|E| <$ 0.5)* **then**   ▷ *Hard besiege*

                Update the location vector

            else if *(r $<$ 0.5 and $|E| \geq$ 0.5)* **then**   ▷ *Soft besiege*

with progressive rapid dives

                Update the location vector

            else if *(r $<$ 0.5 and $|E| <$ 0.5)* **then**   ▷ *Hard besiege*

with progressive rapid dives

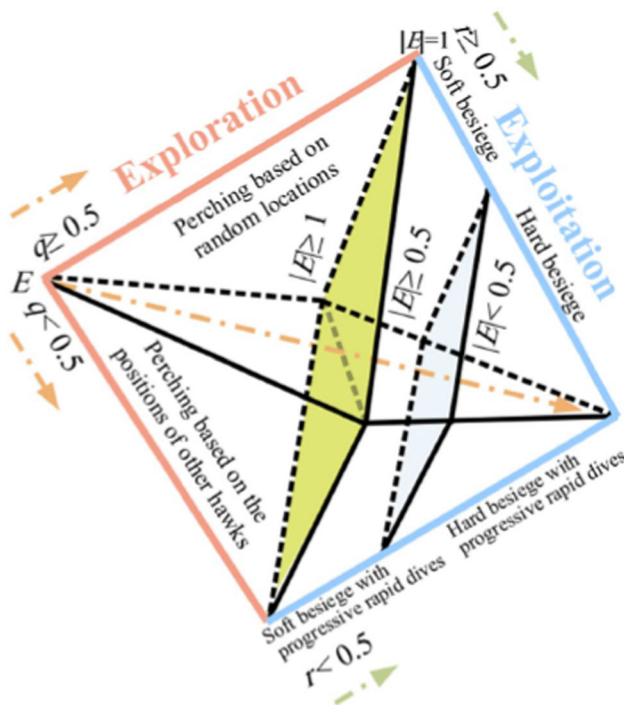   *Update the location vector*

Return *X_{rabbit}*

Fig. 2 Different phases of HHO [52]

As explained, the three metaheuristic algorithms, PSO, ICA and HHO are used in this work to improve the prediction performance of the MLP. The metaheuristic optimizers need to be integrated with an MLP to optimize its weights and biases. In brief, the solution of the metaheuristic learning tools is a matrix of optimized computational parameters to redesign a robust MLP using them. The algorithm of the proposed hybrid MLP model described by the following steps:

Step 1: Initialize the problem of training MLP as an optimization approach.
Step 2: Initialize the main parameters of metaheuristic algorithms (e.g., the maximum iteration, upper and lower bounds of variables and the number of search agents).
Step 3: Generate the random solutions (e.g., decision variables) and calculate the value of objective function for each solution.
Step 4: Use the metaheuristic algorithms to find the optimal values of decision variables.
Step 5: Repeat Step 4 until the finish criteria of algorithms is satisfied. The best values of weights and biases are obtained.

Figure 3 shows the flowchart of the proposed hybrid MLP methods.

3 Data and statistical analysis

Hydraulic and geometric flow data are essential for accurate estimation of the longitudinal dispersion coefficient. Thus, in this study a range of data that gathered from the previous studies where the required parameter values were provided are exploited to investigate the abilities of the proposed hybrid AI models for robust and efficient prediction of the longitudinal dispersion coefficient. This information includes 164 series of data related to the flow characteristics, such as the width (W), depth (H), velocity (U), shear velocity (U_*) and the measured longitudinal dispersion coefficient (K) [6].

Table 2 report the statistical properties of the utilized database in terms of the minimum, maximum, mean and standard deviation values. Moreover, the coefficient of correlation between the input variables and the longitudinal dispersion coefficient is reported in Table 2. Besides, Fig. 4 shows the frequency histogram of the input and output variables. Based on the table and figures illustrated results, the utilized data have different type and range, which complicates the problem of estimating the longitudinal dispersion coefficient of rivers.

The statistical analysis results showed a significant correlation between the longitudinal dispersion coefficient and the width-to-depth ratio (W/H) [57]. This ratio is used in most developed experimental relationships of estimating the longitudinal dispersion coefficient to improve the accuracy of the predictive models. Therefore, this paper also exploited this ratio for categorizing different flows to increase the accuracy of the proposed predictive AI model [6]. Accordingly, the 164 data sets were divided into two groups as $\frac{W}{H} \leq 28$ and $\frac{W}{H} > 28$.

4 Statistical evaluation indicators

To compare the results efficiency of the proposed predictive neural network model integrated with the indicated metaheuristic algorithms above as a novel framework for accurate modeling of the longitudinal dispersion coefficient, several statistical criteria and parameters are employed. The utilized statistical indicators include the
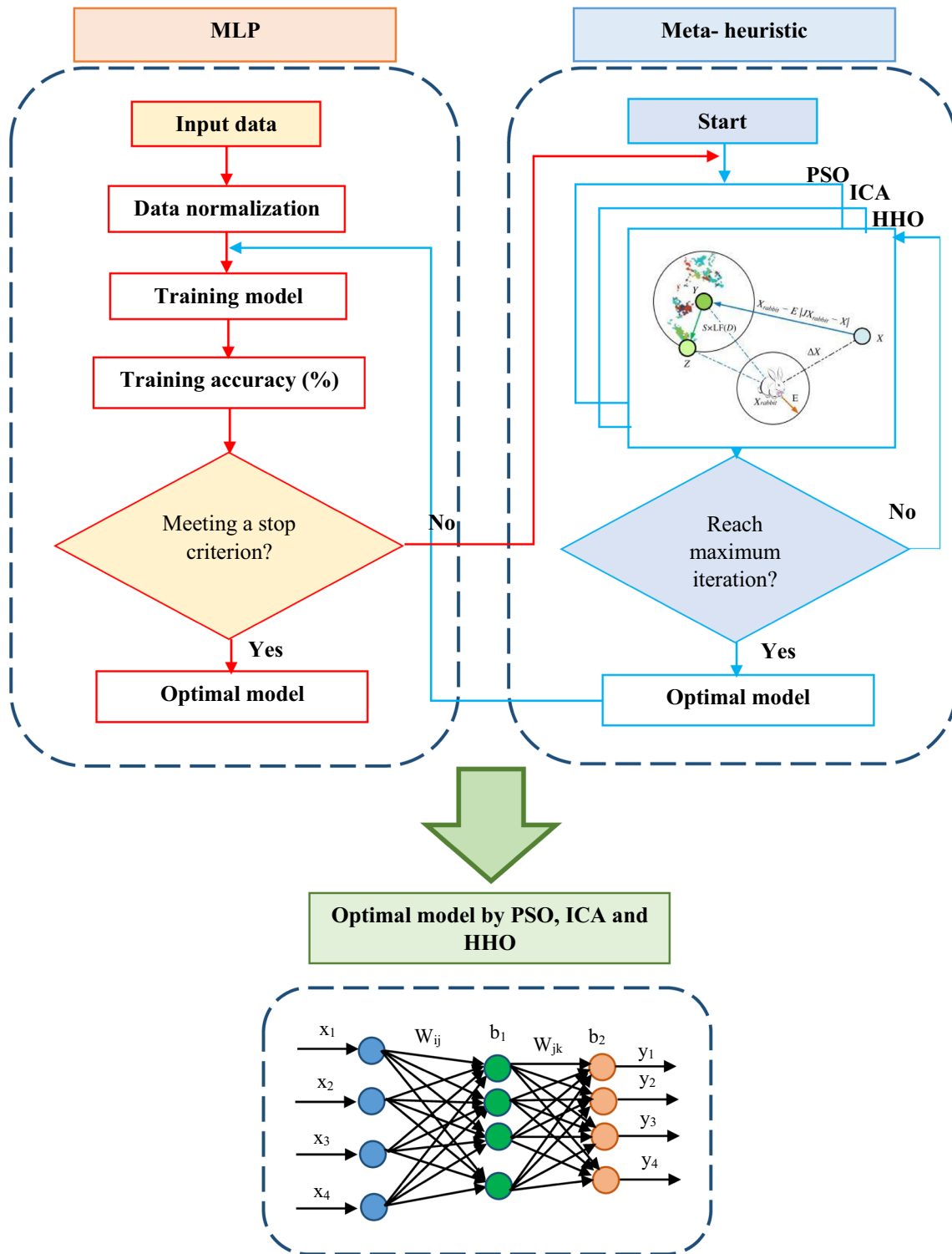
Fig. 3 Flowchart of the proposed hybrid models based on the MLP and metaheuristic algorithms

mean bias error (*MBE*), the root mean square error (*RMSE*), the normalized root mean square error (NRMSE), the mean absolute error (*MAE*), coefficient of determination (R^2), t-statistics and uncertainty at 95% (U_{95}), which the equation of each index is presented as follows [58, 59]:

$$MBE = \frac{1}{n}\sum_{i=1}^{n}\left(K^i_{(observed)} - K^i_{(predicted)}\right) \qquad (16)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(K^i_{(observed)} - K^i_{(predicted)}\right)^2} \qquad (17)$$

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(K^i_{(observed)} - K^i_{(predicted)}\right)^2}}{\overline{K}_{(observed)}} \qquad (18)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|K^i_{(observed)} - K^i_{(predicted)}\right| \qquad (19)$$

$$t-statistics = \frac{1}{n}\sqrt{\frac{(n-1)MBE^2}{RMSE^2 - MBE^2}} \qquad (20)$$

$$U_{95} = 1.96\left(SD^2 + RMSE\right)^{1/2} \qquad (21)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(K^i_{(observed)} - K^i_{(predicted)}\right)^2}{\sum_{i=1}^{n}\left(\overline{K_{(observed)}} - K^i_{(predicted)}\right)^2} \qquad (22)$$

In the above-mentioned formulas, $K^i_{(observed)}$, $K^i_{(predicted)}$ and $\overline{K_{(observed)}}$ are the observed value, the predicted and the observed mean values of the longitudinal dispersion coefficient (K), respectively. It should be mentioned that the model with the lowest values of the statistical indicators as MBE, RMSE, NRMSE, MAE, t-statistics and U_{95} is

considered to be the best performing model for the prediction of the longitudinal dispersion coefficient among the others.

5 Results and discussion

5.1 Statistical evaluation of the models' performance

In order to achieve the best performance of the meta-heuristic algorithms in each optimization problem, it is necessary to determine a set of setting parameters using the trial and error method to minimize or maximize the desired objective function. The optimal setting parameters values of PSO algorithm and MLP are tabulated in Table 3. It is noticeable that only PSO has setting parameters among the used algorithms.

As mentioned before, the 164 data sets presented in Sect. 3 were divided into two groups as $\frac{W}{H} \leq 28$ and $\frac{W}{H} > 28$. Then the developed MLP models based on the metaheuristic algorithms were implemented separately for two groups of data, training data (80%) and test data (20%).

Tables 4 and 5 show the performance of the developed predictive AI models, including MLP, MLP-ICA, MLP-PSO and MLP-HHO, using the statistical parameters of MBE, RMSE, NRMSE, R^2, MAE, t-statistic and U_{95}.

Investigating the obtained results using the statistical parameters presented in Tables 4 and 5, it is noticeable that the closer the value of a statistical parameter to 0, the more the predicted results of the longitudinal dispersion coefficient is accurate. As a result, the algorithm has shown better performance in predicting data.

Therefore, according to Tables 4 and 5, it can be said that all the three MLP models optimized using ICA, HHO and PSO have better performance than the MLP standalone model and predict the dispersion coefficient more accurately. The values of the statistical parameters for the proposed models are lower than those given by the MLP.

All the proposed models showed an acceptable performance in calculating the longitudinal dispersion coefficient. However, the new MLP-HHO hybrid AI model yield the best results based on most of the statistical indicators used in the prediction process, including the training, testing and overall phases. Such that, the yielded RMSE was 28.5178, 1.5809 and 23.3026, in the training, testing and overall stages, respectively, for the first group of data ($\frac{W}{H} \leq 28$). In addition, the RMSE was 66.4141, 48.6601 and
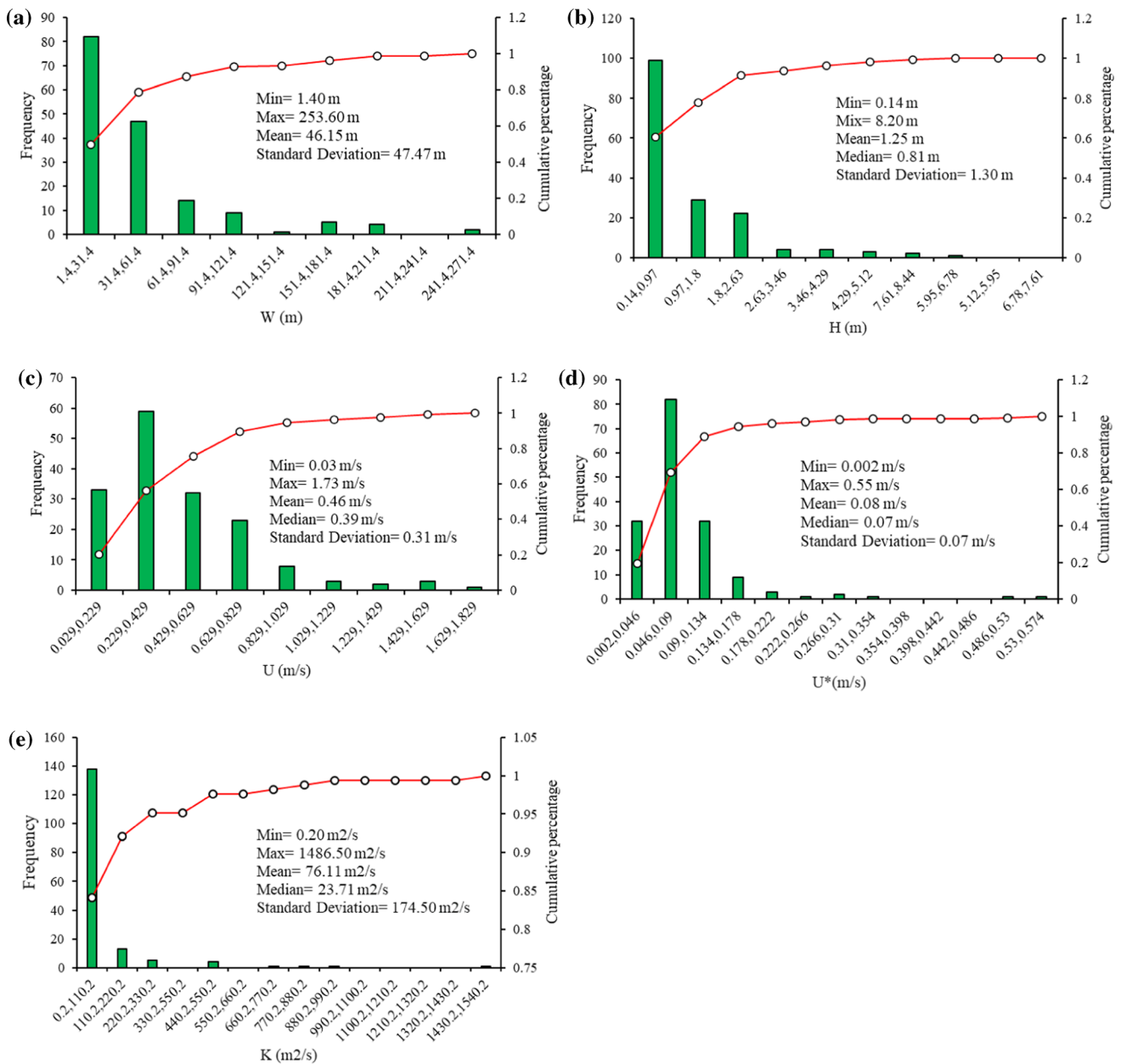
Table 2 Summary of the geometric, hydraulic and dispersion coefficient data sets employed in this study

Variables	Min	Mean	Max	SD
W	1.41	46.14	253.6	47.47
H	0.14	1.25	8.2	1.30
U	0.03	0.465	1.73	0.314
U_*	0.002	0.086	0.553	0.069
$\frac{W}{H}$	2.20	46.52	403.75	45.39
$\frac{U}{U_*}$	0.77	6.68	20.25	4.63
$K\left(\frac{m^2}{s}\right)$	0.20	76.97	1480.50	175.40
$\frac{K}{U_*H}$	3.08	1127.25	37,140.74	3509.64

Fig. 4 Frequency histograms of the input/output variables for the longitudinal dispersion coefficient prediction, **a** channel width, W (m), **b** flow depth, H (m), **c** cross-sectional average velocity, U (m/s), **d** shear velocity, U_* (m/s) and **f** experimental results of dispersion coefficient, K (m²/s)

61.0201, in the training, testing and overall phases, respectively, for the second group data of ($\frac{W}{H} > 28$). In the other side, the lowest performance among the proposed models according to the indicators metrics was found to be obtained by the MLP in which the recorded values of RMSE were 85.3782, 4.4405 and 69.7582 in the training, testing and overall phases for data of $\frac{W}{H} \leq 28$, respectively. Also, RMSE values of the MLP for the group data of $\frac{W}{H} > 28$ were 104.2541, 143.9363 and 119.0904 in the

training, testing and overall phases, respectively. Thus, the correlation of MLP Showed low abilities for accurate modeling of the longitudinal dispersion coefficient among the proposed models.

In addition to the investigated statistical parameters, the correlation coefficient (R^2) was used to evaluate the accuracy of the longitudinal dispersion coefficient predicted using the mentioned AI models, according to Eq. (22). Figures 5 and 6 show the scatter plots of all three

Table 3 Sitting parameters of the models

Method	Parameter	Value
PSO	Inertia weight factor	0.3
	C_1	1
	C_2	1
MLP	Number of neurons	15
	Number of layers	3
	Activation function	Sigmoid (hyperbolic tangent)
	Range partitions (weights and biases)	$[-3, +3]$

Table 4 Average values of statistical indicators for the proposed hybrid models based on the MLP approach for the longitudinal dispersion coefficient prediction in train, test and overall phases $\left(\frac{W}{H} \le 28\right)$

Models	Performance indicators					
	MBE	RMSE	NRMSE	MAE	t-Statistics	U_{95}
Train						
MLP	− 8.054	85.3782	1.258	50.2183	0.015168	167.5766
MLP-ICA	**− 1.7578**	67.6144	0.99626	47.4571	**0.0041628**	133.4561
MLP-PSO	− 10.8734	62.8191	0.9256	43.7102	0.028132	122.258
MLP-HHO	− 5.949	**28.5178**	**0.42019**	**16.9048**	0.034143	**55.6582**
Test						
MLP	− 1.1539	4.4405	0.47997	1.8924	0.06009	9.3644
MLP-ICA	− 1.3417	4.4898	0.4853	3.091	0.069925	9.3687
MLP-PSO	− 1.8627	3.7773	0.40829	2.3995	0.12658	7.4828
MLP-HHO	**0.23647**	**1.5809**	**0.17088**	**1.0403**	0.033781	**3.9318**
Overall						
MLP	− 5.754	69.7582	1.4434	34.1097	0.010866	137.2399
MLP-ICA	**− 1.6191**	55.2677	1.1436	32.6684	**0.0038477**	109.2542
MLP-PSO	− 7.8698	51.338	1.0622	29.94	0.020366	100.4199
MLP-HHO	3.9318	**23.3026**	**0.48216**	**11.6166**	0.022211	**46.0163**

Bold values indicate the best outcome results among the others

Table 5 Average values of statistical indicators for the hybrid proposed models based on the MLP approach for the longitudinal dispersion coefficient prediction in train, test and overall phases $\left(\frac{W}{H} > 28\right)$

Models	Performance indicators					
	MBE	RMSE	NRMSE	MAE	t Statistics	U_{95}
Train						
MLP	**− 0.23029**	104.2541	1.1696	31.2848	**0.0002603**	205.3152
MLP-ICA	1.0972	88.4535	0.9923	44.5657	0.0014618	174.3329
MLP-PSO	− 1.2436	85.7734	0.96223	46.6699	0.0017087	169.0755
MLP-HHO	4.3828	**66.4141**	**0.74505**	**29.1567**	0.0077934	**130.8664**
Test						
MLP	− 17.0563	143.9363	1.5247	56.7687	0.019612	281.1126
MLP-ICA	− 30.7388	134.9982	1.43	36.781	0.038428	258.6505
MLP-PSO	− 18.9331	94.1055	0.99682	51.8578	0.033753	181.6729
MLP-HHO	**− 5.1711**	**48.6601**	**0.51544**	**27.4568**	**0.017563**	**95.8142**
Overall						
MLP	− 5.8914	119.0904	1.31	39.8588	0.0047659	234.1105
MLP-ICA	− 9.614	106.4108	1.1705	41.9465	0.0087291	208.6939
MLP-PSO	− 7.1952	88.6642	0.97528	48.4154	0.0078343	174.1891
MLP-HHO	**1.1684**	**61.0201**	**0.6712**	**28.5848**	**0.0018428**	**120.5538**

Bold values indicate the best outcome results among the others
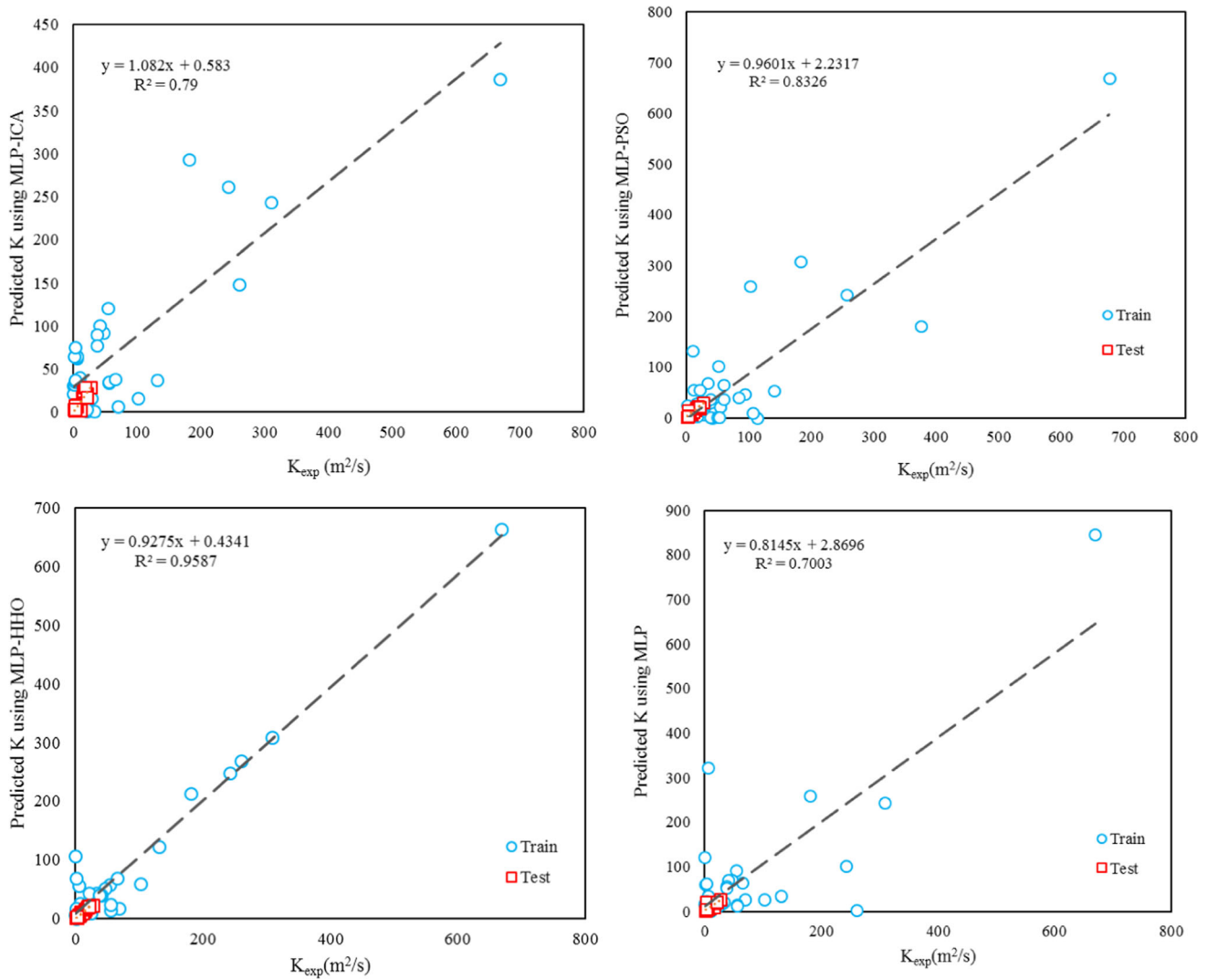
Fig. 5 Scatterplots of the proposed hybrid MLP models in the train and test phases $\left(\frac{W}{H} \leq 28\right)$

developed hybrid and the MLP models for both groups, $\frac{W}{H} \leq 28$ and $\frac{W}{H} > 28$. It is noted that the R^2 indicates the uncertainty of the proposed model. The closer it is to 1, the greater the correlation between the predicted data and the observational data.

In each figure, the linear equation between the predicted and the measured values is shown as Eq. (23).

$$y = aX + b \tag{23}$$

where y and X represent the predicted and the actual longitudinal dispersion coefficient values, respectively; a and b also represent the linear ratios and bias, respectively. As shown from the figures, the MLP-HHO hybrid model obtained the highest correlation coefficient for both data groups with value equal to $R^2 = 0.9587$ for data of $\frac{W}{H} \leq 28$

and $R^2 = 0.9761$ for data of $\frac{W}{H} > 28$ in testing phase. It is worth noting that the R^2 of all optimized MLP hybrid models is higher than the MLP model, that indicating the superiority of the proposed models than the MLP. So that, R^2 value of MLP-HHO, MLP-PSO and MLP-ICA shows 36, 18 and 12% increase than MLP for data group of $\frac{W}{H} \leq 28$, respectively. Also, for data of $\frac{W}{H} > 28$, this value for MLP-HHO, MLP-PSO and MLP-ICA is 41, 26 and 14% higher than MLP, respectively.

Moreover, the $K_{(predicted)}/K_{(observed)}$ ratio in Figs. 7 and 8 has been used to illustrate the uncertainty of the proposed hybrid MLP models results using the train and test phases. According to the figures, the ratio results for both train and test phases are in lower rang in the case of MLP-HHO. The $K_{(predicted)}/K_{(observed)}$ for overall data of MLP-HHO
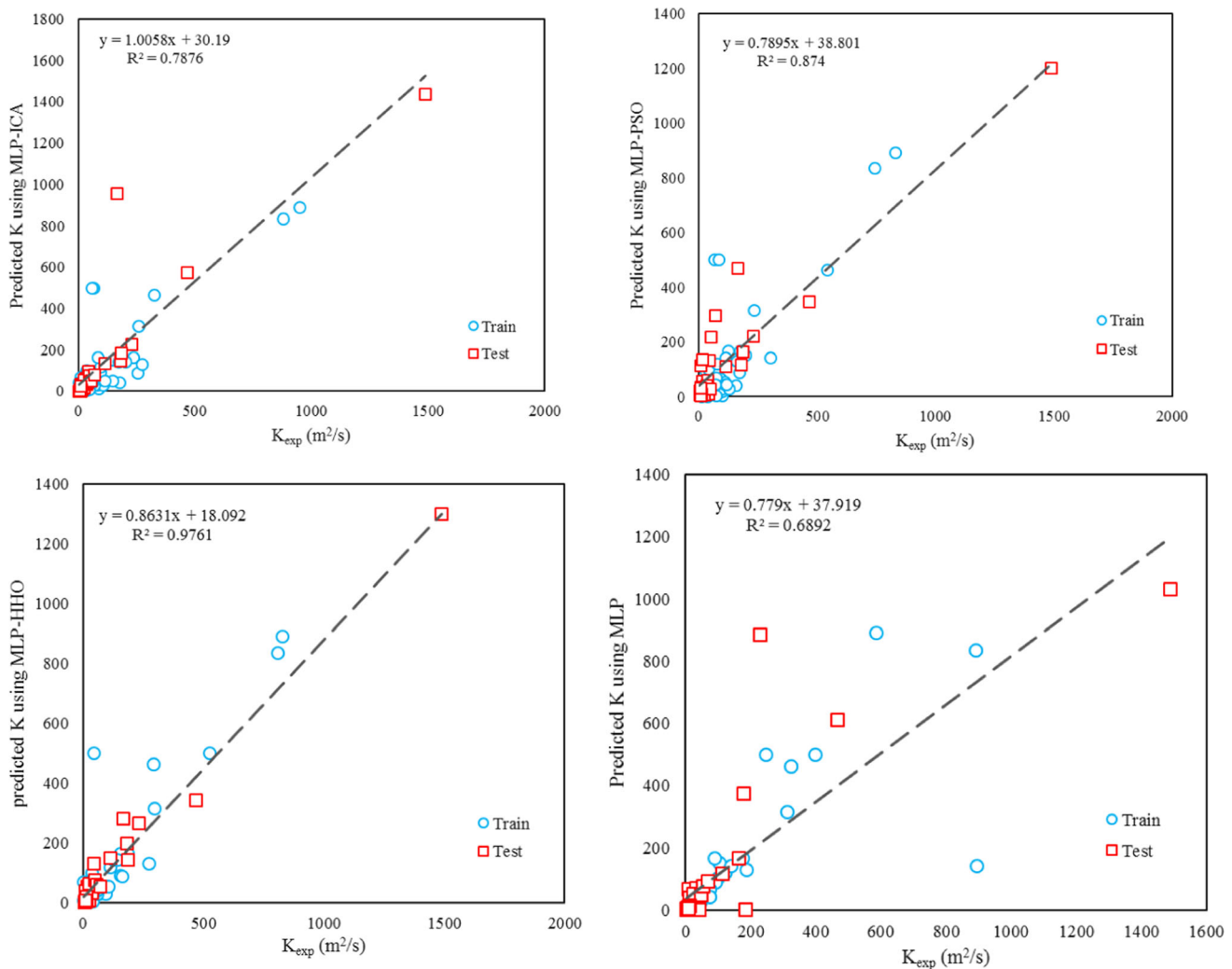
Fig. 6 Scatterplots of the proposed hybrid MLP models in the train and test phases $\left(\frac{W}{H} > 28\right)$

included in the range of [0.246, 138.284] and [0.024, 6.879] for $\frac{W}{H} \leq 28$ and $\frac{W}{H} > 28$ respectively, that have the lowest prediction uncertainties compared to the other investigated models.

5.2 Taylor diagram

In the following, the Taylor diagram (TD) is plotted to show the accuracy of the developed models. In fact, the TD express the performance of models in summarized form. Thus, the relationship between predicted and observed values is displayed by correlation coefficient, standard deviations and the root mean square difference (RMSD) [60, 61].

In the TD, which is plotted as a semicircle or quadrilateral, the values of the correlation coefficient in the

radius, the values of standard deviations in the form of concentric circles with the circle center [0, 0] and the RMSD values in the form of concentric circles with the reference point as the center are shown [60]. In this diagram, the evaluation is done in a way in which the values of the standard deviations, correlation coefficient and RMSD deviation from the models are specified on the diagram as points [62]. Then the point that has a closer distance to the reference point (observed data) represented by the black dot has more accurate prediction credibility [61].

As shown in Figs. 9 and 10, the developed MLP-HHO model is more accurate than the MLP-ICA, MLP-PSO and MLP in predicting the longitudinal dispersion coefficient for both group of data including $\frac{W}{H} \leq 28$ and $\frac{W}{H} > 28$. The MLP-HHO has the closest distance to the reference point
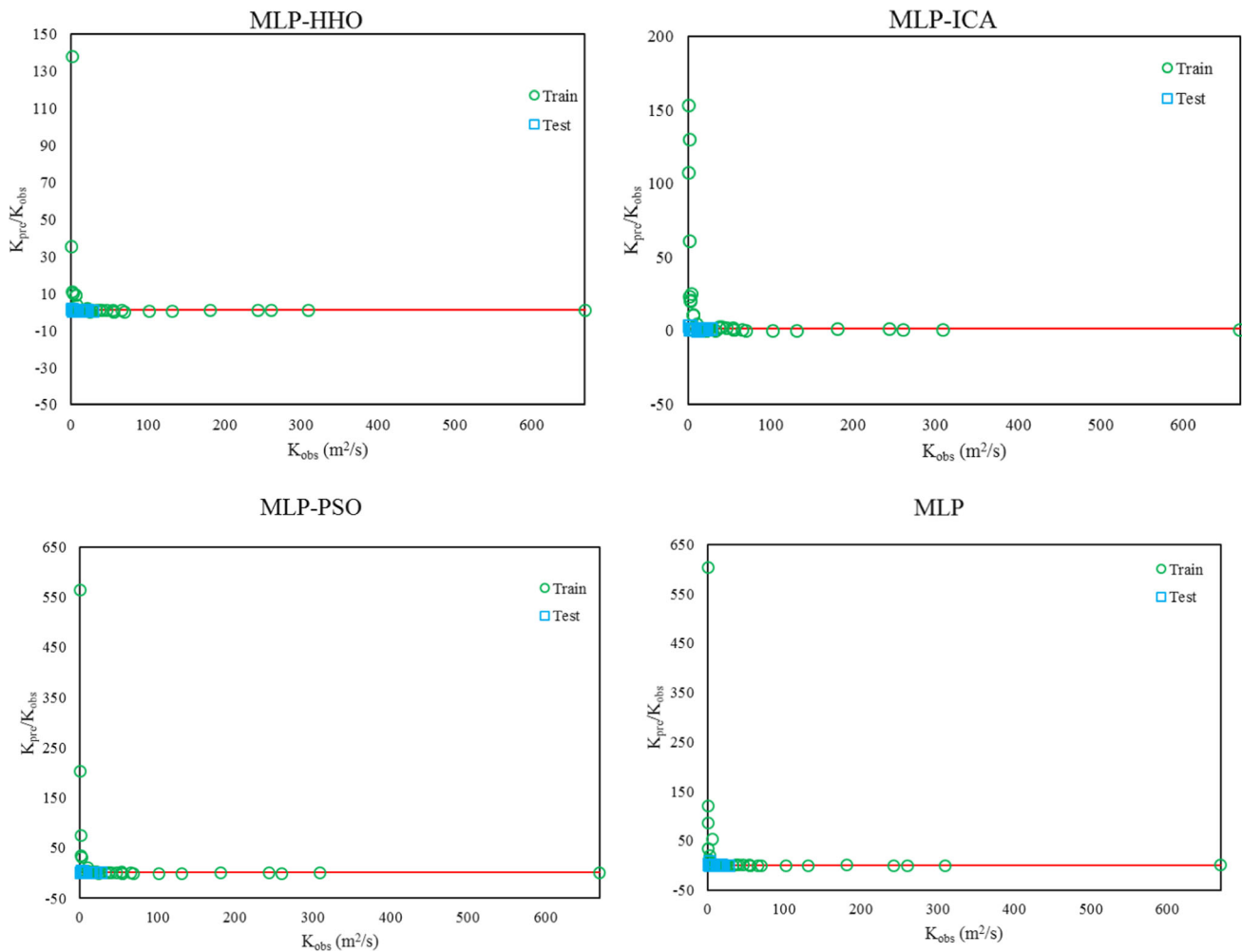
Fig. 7 Comparison of the experimental to predicted longitudinal dispersion coefficient using the proposed hybrid MLP models in the train and test phases $\left(\frac{W}{H} \leq 28\right)$

that refers to observed data, followed by the MLP-PSO model that indicated average results with a close correlation coefficient value to the observed data by 0.9. However, both MLP-ICA and MLP models show low results compared to the other models, which indicate that these models have the weak abilities for modeling the longitudinal dispersion coefficient in rivers.

5.3 Prediction performance validation

To confirm the superiority of the proposed hybrid models, the results of MLP-based models were also compared with the artificial intelligence-based models and empirical equations from previous works. In this regard, the statistical metrics (MBA, RMSE, NRMSE, MAE, t-statistics and U95) were applied for modeling the longitudinal dispersion coefficient for all data sets (see Table 6). According to the

results, it is clear that the MLP-HHO is superior to the other existing equations with respect to all performance indicators. Afterward, the proposed MLP-ICA model (RMSE = 77.750, MAE = 41.99 and R^2 = 0.80) and the proposed soft computing-based model by Alizadeh et al. [6] (RMSE = 89.825, MAE = 45.159 and R^2 = 0.73) ranked second and third, respectively. Moreover, the empirical-based equation by Seo and Cheong [63] (RMSE = 85.390, MAE = 40.804 and R^2 = 0.76) was able to provide a good prediction of the longitudinal dispersion coefficient.

Regarding Fig. 11, it is clear that the MLP-HHO model has highest accuracy in comparison with other models (selected empirical and soft computing-based equations), which indicates a good correlation between simulated and measured values.
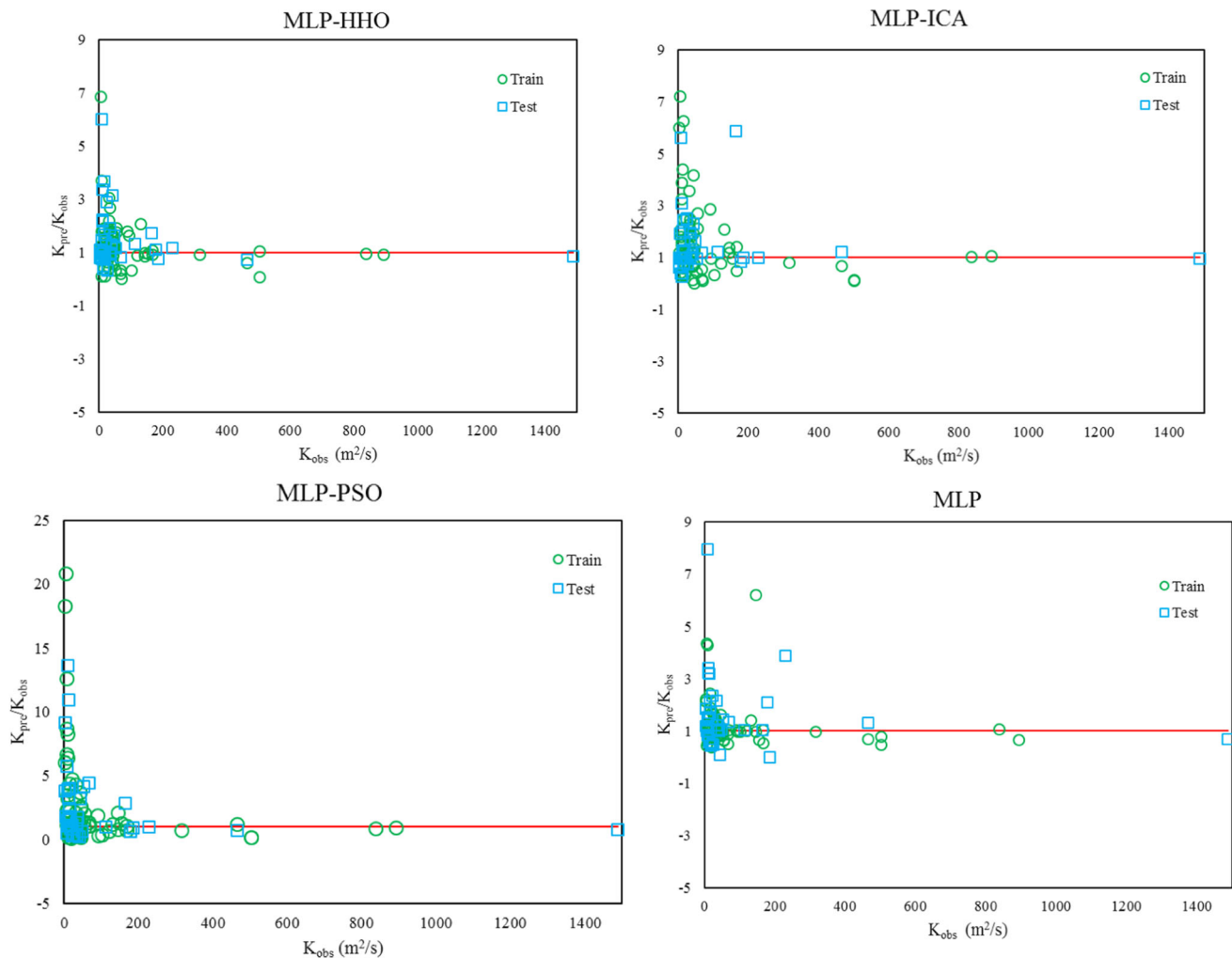
Fig. 8 Comparison of the experimental to predicted longitudinal dispersion coefficient using the proposed hybrid MLP models in the train and test phases $\left(\frac{W}{H} > 28\right)$

Figure 12 compares the observed and predicted values of the longitudinal dispersion coefficient using the three MLP models based on HHO, ICA and PSO, as well as MLP and the other existing equations. Each model's performance can be understood conceptually by displaying the measured values of the data versus the predicted values for each model. As shown in Fig. 12, the developed hybrid MLP models based on the three metaheuristic algorithms have higher accuracy than the MLP model. Also, among the three MLP models based on metaheuristic algorithms, the MLP-HHO hybrid model can accurately identify the prediction pattern and provide accurate longitudinal dispersion coefficients.

Another statistical indicator used to illustrate the performance of the developed MPL models is the discrepancy ratio (DR), which can be calculated as Eq. (24):

$$DR = log\frac{K_{predi}}{K_{obsi}} \tag{24}$$

where K_{predi} and K_{obsi} are ith value of the predicted and observed longitudinal dispersion coefficient obtained from the considered models respectively.

DR indicates the correlation between the longitudinal dispersion coefficients measured and the estimated values by each model, in such a way that the predicted values are the same as the measured values $DR = 0$. If the predicted values are higher than the measured values, $DR > 0$, also $DR < 0$ when the predicted models were underestimate than the measured values. The closer the DR is to 0, the closer the predicted dispersion coefficients are to the measured values. In other words, the model has shown better performance in estimating this parameter.

Fig. 9 Plot of results in Taylor diagram for $\left(\frac{W}{H} \leq 28\right)$
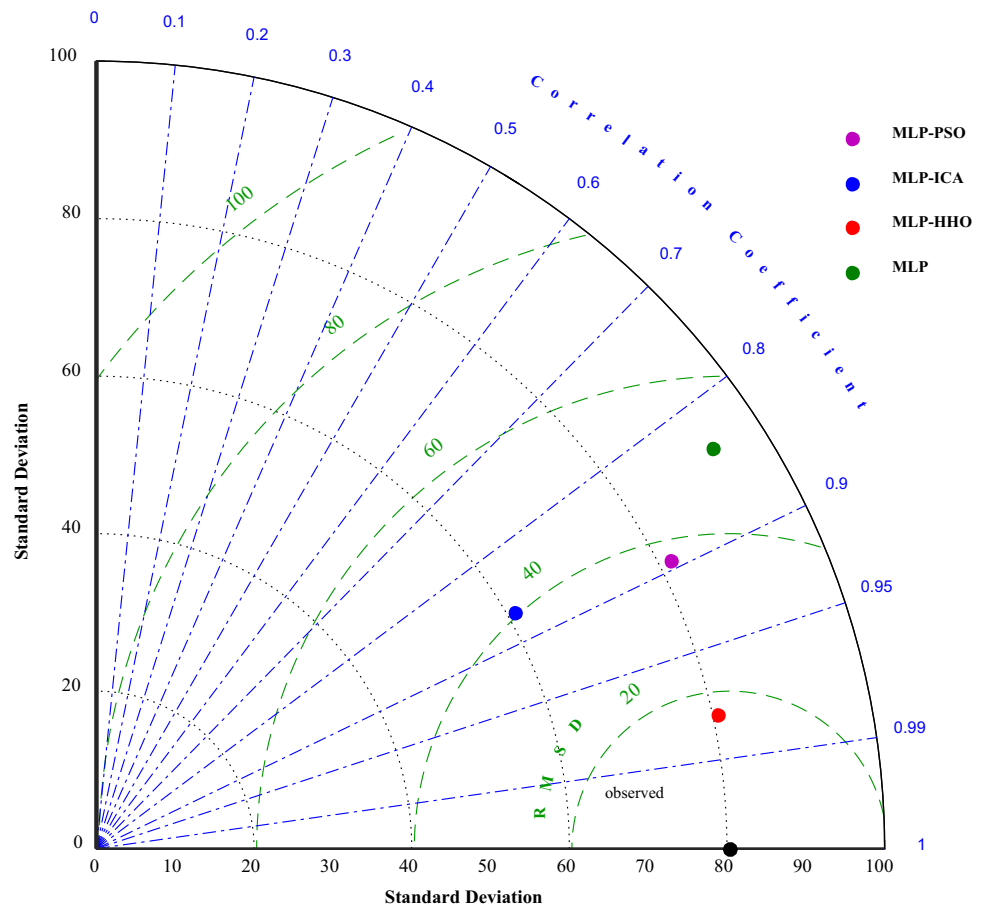
According to Table 7, DRs are closer to 0 using the MLP-HHO model than the other MLP-based models, empirical equations and AI models, indicating a more successful performance of the MLP-HHO despite the acceptable solution provided by all the hybrid MLP models in estimating the longitudinal dispersion coefficient.

5.4 Boosting the proposed MLP-HHO model

Training and testing data are typically determined randomly in prediction models. As mentioned earlier, this study has also utilized a random approach to select data in both training and testing phases for building the proposed model. The use of this approach is efficient and effective while limited data is available. However, when dealing with a large amount of data, the use of this method may decrease the model accuracy during some testing phases due to the random selection of data, which may lead to training the model based on some certain data range, and in the testing phase, the model encounters data with a different range from the training data, thereby reducing the

model accuracy significantly. Since the proposed MLP-HHO model provided the highest performance compared with the other models in the previous steps, we proposed a more robust model improvement for forecasting the longitudinal dispersion coefficient by adding a preprocess step. For this purpose, we have employed the K-means model, which is arranged as unsupervised machine learning method, to find the data average (the cluster center) based on the W/H ratio. Given the data of dispersion, we initially considered the number of cluster centers equal to 5 and divided all the data based on the cluster centers. Therefore, in the training phase, the model was trained based on an equal number of data from each group, so that, the robustness of the model could be maintained against data dispersion during testing phase. Table 8 shows the cluster center as well as the number of data placed in each cluster according to the W/H ratio.

After developing the combined model, we tested it by all the data including 164 data series addressed in Sect. 3. Table 9 shows the values of the statistical parameters calculated for the data simulated by the K-means-MLP-HHO

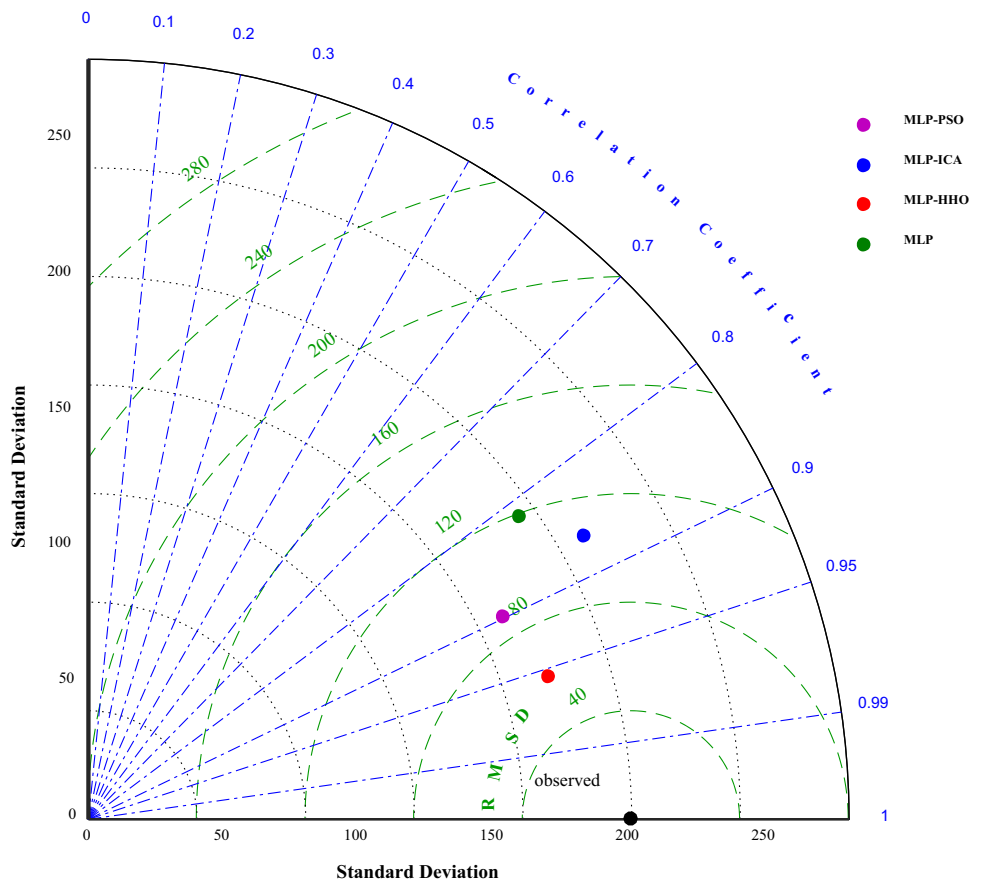**Fig. 10** Plot of results in Taylor diagram for $\left(\frac{W}{H} > 28\right)$

Table 6 Comparison between the existing and proposed models

Models	Performance indicators					
	MBE	RMSE	NRMSE	MAE	t-Statistics	U_{95}
Deng et al. [64]	− 19.145	94.448	1.241	51.818	0.016	182.274
Seo and Cheong [63]	7.308	85.390	1.122	40.804	0.007	167.730
Zeng and Huai [65]	15.670	101.786	1.337	43.827	0.012	198.112
Disley et al. [57]	14.712	117.194	1.540	47.262	0.010	228.869
Etemad-Shahidi and Thaghipour [12]	23.329	118.067	1.551	44.853	0.016	227.846
Alizadeh et al. [6]	− 1.846	89.825	1.180	45.159	0.002	176.997
Memarzadeh et al. [16]	36.272	145.610	1.913	51.147	0.020	277.4102
Kargar et al. [15]	− 25.008	135.228	1.777	62.027	0.015	261.470
MLP-ICA	− 7.430	77.750	1.022	41.994	0.007	152.674
MLP-PSO	− 6.835	91.920	1.208	38.722	0.006	180.645
MLP-HHO	**− 0.589**	**51.167**	**0.672**	**22.687**	**0.001**	**101.256**
MLP	− 5.844	104.616	1.375	37.861	0.004	205.707

Bold values indicate the best outcome results among the others

model. It is clear that, the use of K-means to cluster the data before modeling has improved the accuracy of the MLP-HHO model during the training and testing phases compared to the random data selection strategy. For example, the RMSE value has decreased from 57.22 to 34.12 m²/s in the training phase and from 24.09 to 17.77

m²/s in the testing phase and overall from 51.16 to 30.99 m²/s. According to the parameter U95, it is clear that the clustering-based model provide the favorable predictions. Also, the lowest uncertainty in prediction of the longitudinal dispersion coefficient (35.687) is observed in the
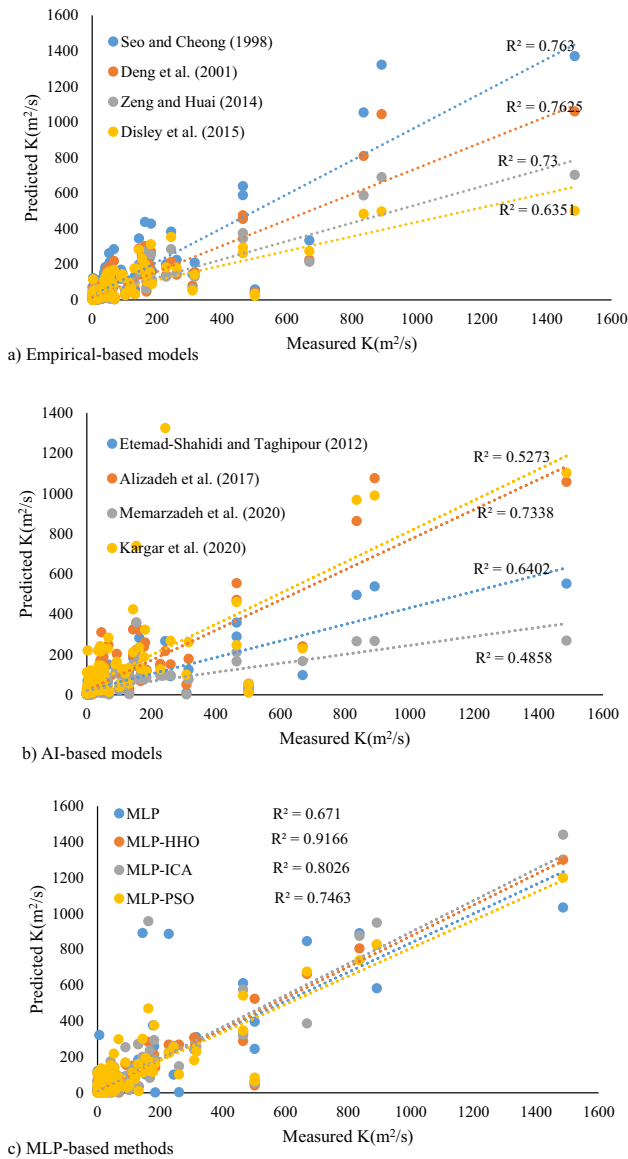
Fig. 11 R^2 of different models for the all data set; **a** empirical-based models; **b** AI-based models; **c** MLP-based models

results obtained by the clustering-based model in testing phase.

Moreover, as Fig. 13 shows, the use of k-means in combination with the proposed model improves the robustness in the testing phase as the coefficient of determination (R^2) of the predicted and real data in both, training and testing phases is equal.

5.5 Sensitivity analysis

Finally, a sensitivity analysis was performed to evaluate the relative importance of the input parameters (W, U, H, U_*)

on the longitudinal dispersion coefficient using the MLP-HHO model. For this purpose, Eq. (25) was utilized

$$r(I_j, O) = \frac{\sum_{i=1}^{n}(I_{j,i} - \overline{I}_j)(O_i - \overline{O})}{\sqrt{\sum_{i=1}^{n}(I_{j,i} - \overline{I}_j)^2 \sum_{i=1}^{n}(O_i - \overline{O})^2}} \quad (25)$$

where I and $\overline{I}$ are the input parameter, related to the longitudinal dispersion coefficient and its average value, respectively. O and $\overline{O}$ denote the predicted longitudinal dispersion coefficient output and its average value using K-means-MLP-HHO model, respectively.

The highest value for r by the input I_j indicates the greater impact on the predicted output. Figure 14 shows the sensitivity analysis of the input data on the longitudinal dispersion coefficient by the K-means-MLP-HHO model. As indicated, the width parameter W ($r = 0.573$), the average flow velocity U ($r = 0.553$) and the flow depth H ($r = 0.409$) by the highest values of the relevancy factor are recognized as the most important factor on the longitudinal dispersion coefficient estimation in rivers.

6 Conclusion

In this study, a new hybrid model using the artificial intelligence (AI) methods is developed to estimate the rivers' longitudinal dispersion coefficient accurately.

The proposed model consists of a powerful multilayer perceptron neural network and a new metaheuristic optimization algorithm called HHO, inspired by the Harris hawk behavior. Two other well-known algorithms, ICA and PSO, were used to train the MLP in order to illustrate the high accuracy of the developed MLP-HHO. For this purpose, 164 series of measured data gathered from previous studies in two groups based on width-to-depth ratio, $\frac{W}{H} \leq 28$ and $\frac{W}{H} > 28$, were evaluated. The investigations on the models' efficiency using the statistical indicators show that the combination of the metaheuristic and MLP algorithms performs better than MLP standalone in solving the complex problem of longitudinal dispersion coefficient in rivers.

According to the results, the MLP-HHO model, with $R^2 = 0.9587$ for $\frac{W}{H} \leq 28$ data and $R^2 = 0.9761$ for $\frac{W}{H} > 28$ data in the testing phase, is the best proposed AI model for predicting the longitudinal dispersion coefficient in rivers than other models. Finally, to increase the robustness of MLP-HHO model, a classification technique, namely K-means, is coupled with it. Thus, a certain percentage of data was selected for better model learning that includes all data ranges. The performance of K-means-MLP-HHO was
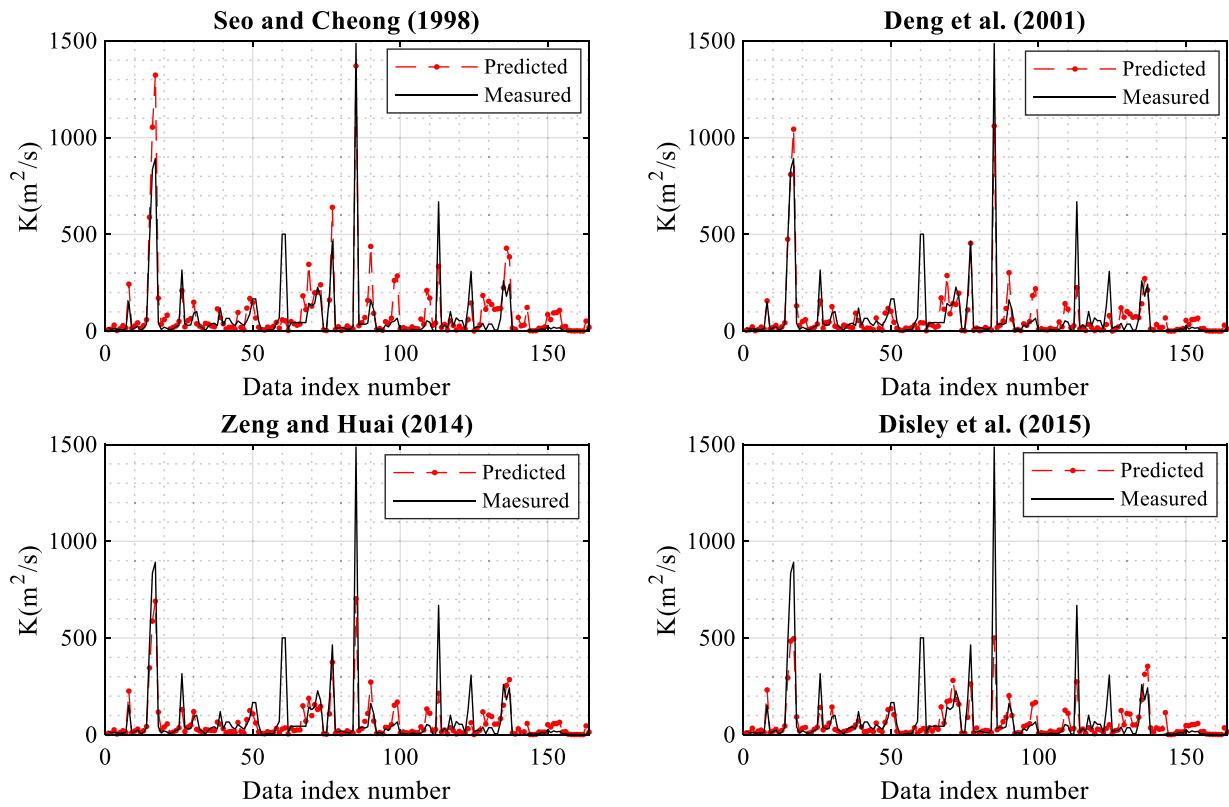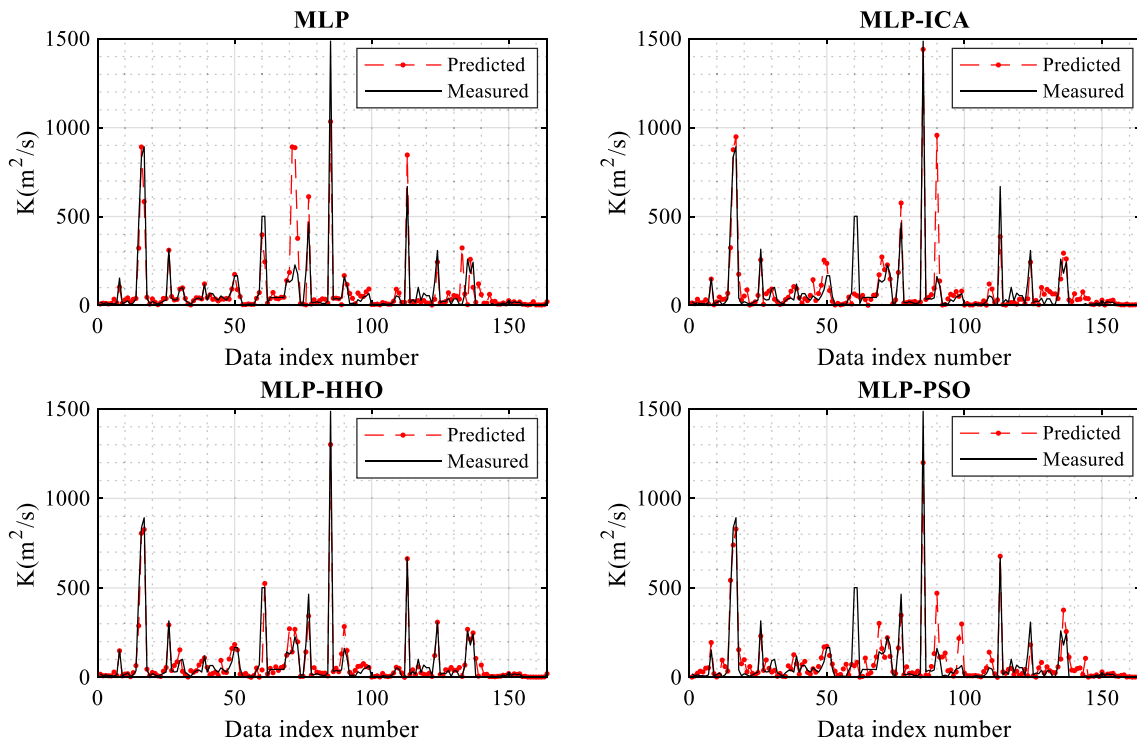
a) Empirical-based model
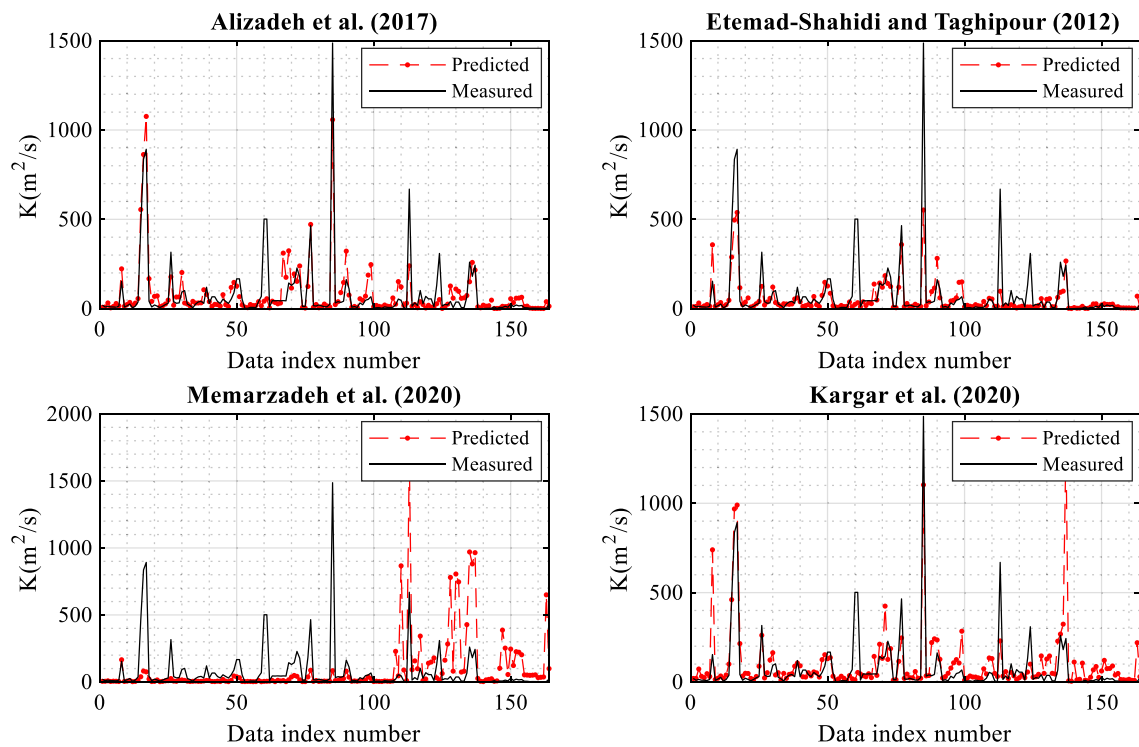
b) AI-based model

Fig. 12 Time series plots for the all data set; **a** empirical-based models; **b** AI-based models; **c** MLP-based models

c) MLP-based model

Fig. 12 continued

Table 7 DR values of the hybrid models developed to predict the longitudinal dispersion coefficient

Model	DR
Deng et al. [64]	[− 1.032, 2.153]
Seo and Cheong [63]	[− 1.287, 1.838]
Zeng and Huai [65]	[− 1.227, 1.795]
Disley et al. [57]	[− 1.343, 2.261]
Etemad-Shahidi and Thaghipour [12]	[− 1.497, 1.424]
Alizadeh et al. [6]	[− 1.240, 1.791]
Memarzadeh et al. [16]	[− 2.068, 2.351]
Kargar et al. [15]	[− 1.633, 2.348]
MLP	[− 1.826, 2.781]
MLP-ICA	[− 1.186, 1.701]
MLP-PSO	[− 1.148, 1.891]
MLP-HHO	[− 0.6091, 1.614]

Table 8 Centroids of data sets in K-means clustering technique

	Centroids	Number of samples
Class 1	19.63	65
Class 2	96.60	24
Class 3	46.95	63
Class 4	403.75	6
Class 5	313.08	6

compared with MLP-HHO model. The boosting K-means-MLP-HHO model with $R^2 = 0.97$, was the best performing model for estimating the longitudinal dispersion coefficient in rivers. Furthermore, the sensitivity analysis related to the input parameters using the relevancy error factor shows that the flow width (W) and the average flow velocity (U) have the highest influence on the longitudinal dispersion coefficient. As the relevancy error factor was 0.573 and 0.553, respectively, for these parameters. Furthermore, the shear velocity (U_*) has the least influence on the behavior of longitudinal dispersion coefficient in rivers.

Moreover, the prediction results may be affected by the uncertainty of the input data, it is suggested to develop the

Table 9 Comparison between the clustering and random-based selection MLP-HHO model for train, test and overall data

Models	Performance indicators					U_{95}
	MBE	RMSE	NRMSE	MAE	t-Statistics	
Train						
Random-based model (MLP-HHO)	− 2.19	57.223	0.640	26.522	0.003	113.064
Clustering-based model (*K*-means-MLP-HHO)	− 3.261	34.125	0.437	20.889	0.008	67.555
Test						
Random-based model (MLP-HHO)	9.215	24.098	0.641	10.797	0.064	44.690
Clustering-based model (*K*-means-MLP-HHO)	− 1.459	17.773	0.2524	11.128	0.013	35.687
Overall						
Random-based model (MLP-HHO)	− 0.589	51.167	0.672	22.687	0.001	101.256
Clustering-based model (*K*-means-MLP-HHO)	− 2.82	30.944	0.406	18.508	0.007	61.374
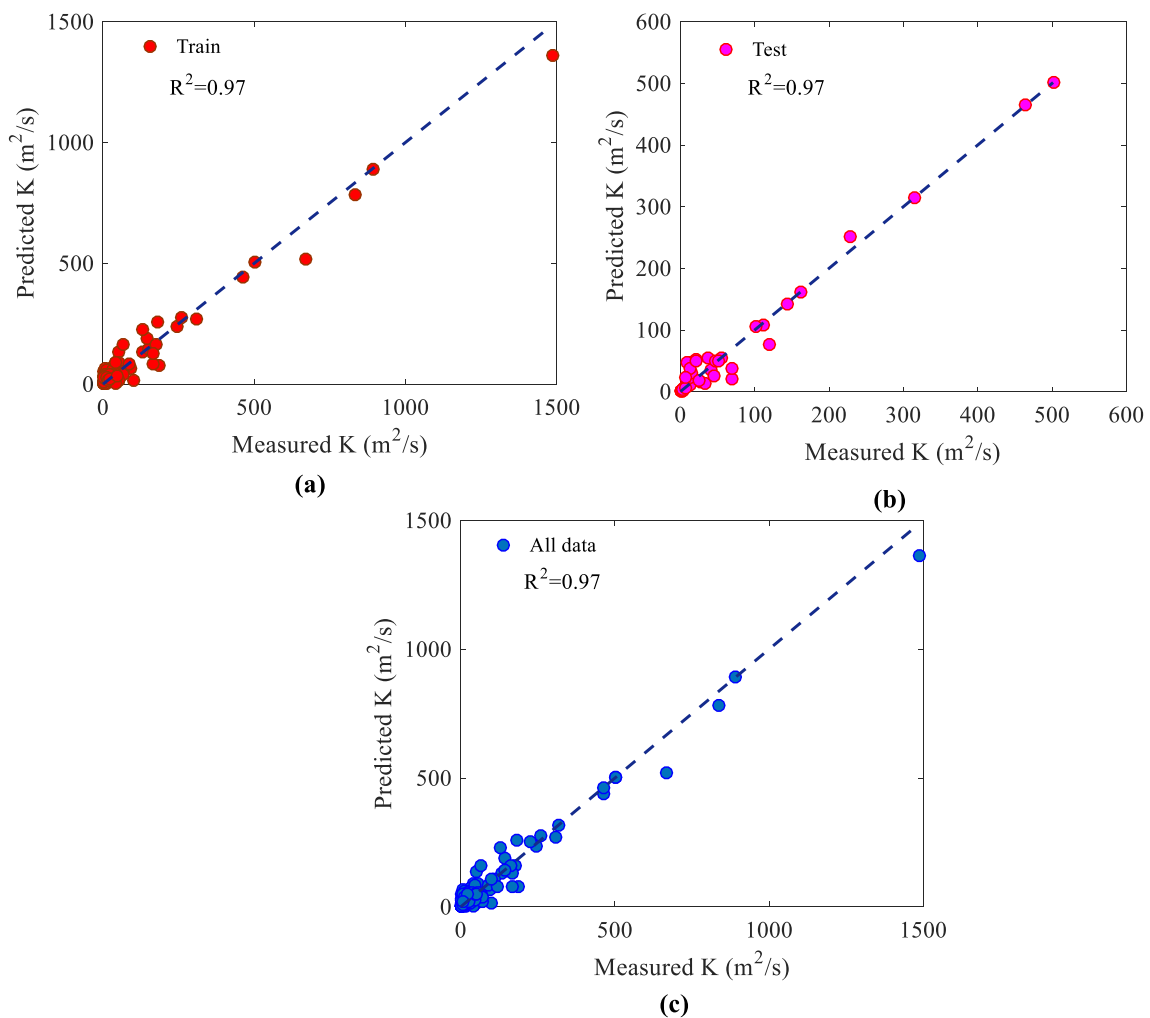
Fig. 13 Cross plots of the implemented *K*-means-MLP-HHO
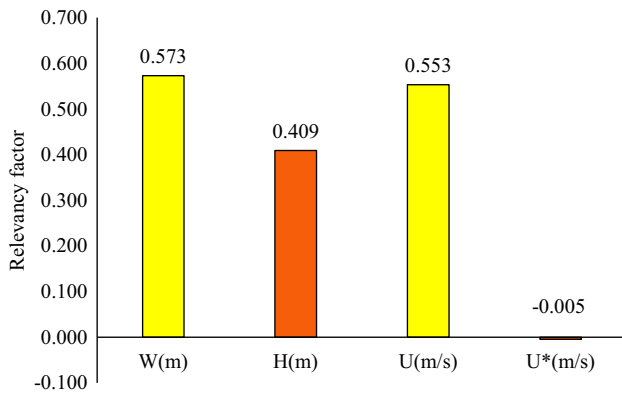
Fig. 14 Sensitivity analysis of the input parameters using *K*-means-MLP-HHO (all data)

probabilistic models in future research and consider the uncertainty of the input data to obtain more reliable results.

Declarations

Conflict of 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.

References

1. Alizadeh MJ, Shahheydari H, Kavianpour MR et al (2017) Prediction of longitudinal dispersion coefficient in natural rivers using a cluster-based Bayesian network. Environ Earth Sci 76:1–11. https://doi.org/10.1007/s12665-016-6379-6

2. Hashemi Monfared SA, Mirbagheri SA, Sadrnejad SA (2014) A three-dimensional, integrated seasonal separate advection-diffusion model (ISSADM) to predict water quality patterns in the Chahnimeh reservoir. Environ Model Assess 19:71–83. https://doi.org/10.1007/s10666-013-9376-0

3. Baghbanpour S, Kashefipour SM (2012) Numerical modeling of suspended sediment transport in rivers (case study: Karkheh River). J Sci Technol Agric Nat Resour 16:45–58

4. Boddula S, Eldho TI (2017) A moving least squares based meshless local petrov-galerkin method for the simulation of contaminant transport in porous media. Eng Anal Bound Elem 78:8–19. https://doi.org/10.1016/j.enganabound.2017.02.003

5. Noori R, Ghiasi B, Sheikhian H, Adamowski JF (2017) Estimation of the dispersion coefficient in natural rivers using a granular computing model. J Hydraul Eng 143:04017001. https://doi.org/10.1061/(asce)hy.1943-7900.0001276

6. Alizadeh MJ, Ahmadyar D, Afghantoloee A (2017) Improvement on the existing equations for predicting longitudinal dispersion coefficient. Water Resour Manag 31:1777–1794. https://doi.org/10.1007/s11269-017-1611-z

7. Taylor GI (1954) The dispersion of matter in turbulent flow through a pipe. Proc R Soc Lond Ser A Math Phys Sci 223:446–468. https://doi.org/10.1098/rspa.1954.0130

8. Elder JW (1959) The dispersion of marked fluid in turbulent shear flow. J Fluid Mech 5:544–560. https://doi.org/10.1017/S0022112059000374

9. Sahay RR, Dutta S (2009) Prediction of longitudinal dispersion coefficients in natural rivers using genetic algorithm. Hydrol Res 40:544–552. https://doi.org/10.2166/nh.2009.014

10. Fisher H (1968) Dispersion predictions in natural streams. J Sanit Eng Div 94:927–944

11. Kashefipour S, Falconer RA (2002) Longitudinal dispersion coefficients in natural channels. Water Res 36:1596–1608. https://doi.org/10.1016/S0043-1354(01)00351-7

12. Etemad-Shahidi A, Taghipour M (2012) Predicting longitudinal dispersion coefficient in natural streams using M5′ model tree. J Hydraul Eng 138:542–554. https://doi.org/10.1061/(asce)hy.1943-7900.0000550

13. Li X, Liu H, Yin M (2013) Differential evolution for prediction of longitudinal dispersion coefficients in natural streams. Water Resour Manag 27:5245–5260. https://doi.org/10.1007/s11269-013-0465-2

14. Sattar AMA, Gharabaghi B (2015) Gene expression models for prediction of longitudinal dispersion coefficient in streams. J Hydrol 524:587–596. https://doi.org/10.1016/j.jhydrol.2015.03.016

15. Kargar K, Samadianfard S, Parsa J et al (2020) Estimating longitudinal dispersion coefficient in natural streams using empirical models and machine learning algorithms. Eng Appl Comput Fluid Mech 14:311–322. https://doi.org/10.1080/19942060.2020.1712260

16. Memarzadeh R, Ghayoumi Zadeh H, Dehghani M et al (2020) A novel equation for longitudinal dispersion coefficient prediction based on the hybrid of SSMD and whale optimization algorithm. Sci Total Environ 716:137007. https://doi.org/10.1016/j.scitotenv.2020.137007

17. Dehghani M, Zargar M, Riahi-Madvar H, Memarzadeh R (2020) A novel approach for longitudinal dispersion coefficient estimation via tri-variate archimedean copulas. J Hydrol 584:124662

18. Jafari-Asl J, Ben Seghier MEA, Ohadi S, van Gelder P (2021) Efficient method using Whale Optimization Algorithm for reliability-based design optimization of labyrinth spillway. Appl Soft Comput 101:107036. https://doi.org/10.1016/j.asoc.2020.107036

19. Julie MD, Kannan B (2012) Attribute reduction and missing value imputing with ANN: prediction of learning disabilities. Neural Comput Appl 21:1757–1763. https://doi.org/10.1007/s00521-011-0619-1

20. Wróbel J, Kulawik A (2019) Calculations of the heat source parameters on the basis of temperature fields with the use of ANN. Neural Comput Appl 31:7583–7593. https://doi.org/10.1007/s00521-018-3594-y

21. Shukla V, Bandyopadhyay M, Pandya V et al (2020) Artificial neural network based predictive negative hydrogen ion helicon plasma source for fusion grade large sized ion source. Eng Comput. https://doi.org/10.1007/s00366-020-01060-5

22. Wang L, Von Laszewski G, Huang F et al (2011) Task scheduling with ANN-based temperature prediction in a data center: a simulation-based study. Eng Comput 27:381–391. https://doi.org/10.1007/s00366-011-0211-4

23. Jalal M, Grasley Z, Gurganus C, Bullard JW (2020) A new nonlinear formulation-based prediction approach using artificial neural network (ANN) model for rubberized cement composite. Eng Comput. https://doi.org/10.1007/s00366-020-01054-3

24. Koopialipoor M, Fahimifar A, Ghaleini EN et al (2020) Development of a new hybrid ANN for solving a geotechnical problem related to tunnel boring machine performance. Eng Comput 36:345–357. https://doi.org/10.1007/s00366-019-00701-8

25. Mai SH, Ben Seghier MEA, Nguyen PL et al (2020) A hybrid model for predicting the axial compression capacity of square

concrete-filled steel tubular columns. Eng Comput. https://doi.org/10.1007/s00366-020-01104-w

26. Luat NV, Shin J, Lee K (2020) Hybrid BART-based models optimized by nature-inspired metaheuristics to predict ultimate axial capacity of CCFST columns. Eng Comput. https://doi.org/10.1007/s00366-020-01115-7

27. Le LM, Ly HB, Pham BT et al (2019) Hybrid artificial intelligence approaches for predicting buckling damage of steel columns under axial compression. Materials (Basel) 12:1670. https://doi.org/10.3390/ma12101670

28. Zounemat-Kermani M, Kisi O, Piri J, Mahdavi-Meymand A (2019) Assessment of artificial intelligence-based models and metaheuristic algorithms in modeling evaporation. J Hydrol Eng 24:04019033. https://doi.org/10.1061/(asce)he.1943-5584.0001835

29. Jahandideh-Tehrani M, Jenkins G, Helfer F (2020) A comparison of particle swarm optimization and genetic algorithm for daily rainfall-runoff modelling: a case study for Southeast Queensland, Australia. Optim Eng. https://doi.org/10.1007/s11081-020-09538-3

30. Ghorbani MA, Kazempour R, Chau KW et al (2018) Forecasting pan evaporation with an integrated artificial neural network quantum-behaved particle swarm optimization model: a case study in Talesh, Northern Iran. Eng Appl Comput Fluid Mech 12:724–737. https://doi.org/10.1080/19942060.2018.1517052

31. Azad A, Farzin S, Kashi H et al (2018) Prediction of river flow using hybrid neuro-fuzzy models. Arab J Geosci 11:718. https://doi.org/10.1007/s12517-018-4079-0

32. Najafzadeh M, Tafarojnoruz A (2016) Evaluation of neuro-fuzzy GMDH-based particle swarm optimization to predict longitudinal dispersion coefficient in rivers. Environ Earth Sci 75:157. https://doi.org/10.1007/s12665-015-4877-6

33. Gholami A, Bonakdari H, Ebtehaj I et al (2018) Uncertainty analysis of intelligent model of hybrid genetic algorithm and particle swarm optimization with ANFIS to predict threshold bank profile shape based on digital laser approach sensing. Measurement 121:294–303. https://doi.org/10.1016/j.measurement.2018.02.070

34. Jafari-Asl J, Azizyan G, Monfared SAH et al (2021) An enhanced binary dragonfly algorithm based on a V-shaped transfer function for optimization of pump scheduling program in water supply systems (case study of Iran). Eng Fail Anal 123:105323. https://doi.org/10.1016/j.engfailanal.2021.105323

35. Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. IEEE Trans Evol Comput 1:67–82. https://doi.org/10.1109/4235.585893

36. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007

37. Te CV (1959) Open-channel hydraulics. McGraw-Hill, New York

38. Khandelwal M, Marto A, Fatemi SA et al (2018) Implementing an ANN model optimized by genetic algorithm for estimating cohesion of limestone samples. Eng Comput 34:307–317. https://doi.org/10.1007/s00366-017-0541-y

39. Gao W, Guirao JLG, Basavanagoud B, Wu J (2018) Partial multi-dividing ontology learning algorithm. Inf Sci (NY) 467:35–58. https://doi.org/10.1016/j.ins.2018.07.049

40. Armaghani DJ, Momeni E, Abad SVANK, Khandelwal M (2015) Feasibility of ANFIS model for prediction of ground vibrations resulting from quarry blasting. Environ Earth Sci 74:2845–2860. https://doi.org/10.1007/s12665-015-4305-y

41. Liu L, Moayedi H, Rashid ASA et al (2020) Optimizing an ANN model with genetic algorithm (GA) predicting load-settlement behaviours of eco-friendly raft-pile foundation (ERP) system.

Eng Comput 36:421–433. https://doi.org/10.1007/s00366-019-00767-4

42. Gao W, Raftari M, Rashid ASA et al (2020) A predictive model based on an optimized ANN combined with ICA for predicting the stability of slopes. Eng Comput 36:325–344. https://doi.org/10.1007/s00366-019-00702-7

43. Moayedi H, Mehrabi M, Mosallanezhad M et al (2019) Modification of landslide susceptibility mapping using optimized PSO-ANN technique. Eng Comput 35:967–984. https://doi.org/10.1007/s00366-018-0644-0

44. Dreyfus G (2005) Neural networks: methodology and applications. Springer Science & Business Media

45. Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 65:386–408. https://doi.org/10.1037/h0042519

46. Tukey JW (1962) The future of data analysis. Ann Math Stat 33:1–67

47. Montalvo I, Izquierdo J, Pérez R, Tung MM (2008) Particle swarm optimization applied to the design of water supply systems. Comput Math Appl 56:769–776. https://doi.org/10.1016/j.camwa.2008.02.006

48. Eberhart R, Kennedy J (1995) New optimizer using particle swarm theory. In: Proceedings of the 6th international symposium on micro machine and human science, pp 39–43

49. Jafari-Asl J, Sami Kashkooli B, Bahrami M (2020) Using particle swarm optimization algorithm to optimally locating and controlling of pressure reducing valves for leakage minimization in water distribution systems. Sustain Water Resour Manag 6:1–11. https://doi.org/10.1007/s40899-020-00426-3

50. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of ICNN'95—international conference on neural networks, pp 1942–1948

51. Atashpaz Gargari E, Hashemzadeh F, Rajabioun R, Lucas C (2008) Colonial competitive algorithm: a novel approach for PID controller design in MIMO distillation column process. Int J Intell Comput Cybern. https://doi.org/10.1108/17563780810893446

52. Heidari AA, Mirjalili S, Faris H et al (2019) Harris hawks optimization: algorithm and applications. Futur Gener Comput Syst 97:849–872. https://doi.org/10.1016/j.future.2019.02.028

53. Abbasi A, Firouzi B, Sendur P (2019) On the application of Harris hawks optimization (HHO) algorithm to the design of microchannel heat sinks. Eng Comput. https://doi.org/10.1007/s00366-019-00892-0

54. Moayedi H, Abdullahi MM, Nguyen H, Rashid ASA (2019) Comparison of dragonfly algorithm and Harris hawks optimization evolutionary data mining techniques for the assessment of bearing capacity of footings over two-layer foundation soils. Eng Comput. https://doi.org/10.1007/s00366-019-00834-w

55. Zhong C, Wang M, Dang C et al (2020) First-order reliability method based on Harris Hawks Optimization for high-dimensional reliability analysis. Struct Multidiscipl Optim 62:1951–1968. https://doi.org/10.1007/s00158-020-02587-3

56. Moayedi H, Osouli A, Nguyen H, Rashid ASA (2019) A novel Harris hawks' optimization and k-fold cross-validation predicting slope stability. Eng Comput. https://doi.org/10.1007/s00366-019-00828-8

57. Disley T, Gharabaghi B, Mahboubi AA, Mcbean EA (2015) Predictive equation for longitudinal dispersion coefficient. Hydrol Process 29:161–172. https://doi.org/10.1002/hyp.10139

58. Ben Seghier MEA, Corriea JAFO, Jafari-Asl J et al (2021) On the modeling of the annual corrosion rate in main cables of suspension bridges using combined soft computing model and a novel nature-inspired algorithm. Neural Comput Appl 33:15969–15985. https://doi.org/10.1007/s00521-021-06199-w

59. Seghier MEAB, Höche D, Zheludkevich M (2022) Prediction of the internal corrosion rate for oil and gas pipeline: Implementation of ensemble learning techniques. J Nat Gas Sci Eng 99:104425

60. Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res Atmos 106:7183–7192. https://doi.org/10.1029/2000JD900719

61. Granata F, Gargano R, de Marinis G (2020) Artificial intelligence based approaches to evaluate actual evapotranspiration in wetlands. Sci Total Environ 703:135653. https://doi.org/10.1016/j.scitotenv.2019.135653

62. Simão ML, Videiro PM, Silva PBA et al (2020) Application of Taylor diagram in the evaluation of joint environmental distributions' performances. Mar Syst Ocean Technol 15:151–159. https://doi.org/10.1007/s40868-020-00081-5

63. Seo IW, Cheong TS (1998) Predicting longitudinal dispersion coefficient in natural streams. J Hydraul Eng 124:25–32. https://doi.org/10.1061/(ASCE)0733-9429(1998)124:1(25)

64. Deng ZQ, Singh VP, Bengtsson L (2001) Longitudinal dispersion coefficient in straight rivers. J Hydraul Eng 127(11):919–927

65. Zeng Y, Huai W (2014) Estimation of longitudinal dispersion coefficient in rivers. J Hydro Environ Res 8(1):2–8