



New hybrid nature-based algorithm to integration support vector machine for prediction of soil cation exchange capacity

Samad Emamgholizadeh¹ · Babak Mohammadi²

Accepted: 28 July 2021 / Published online: 6 August 2021
© The Author(s) 2021. This article is published with open access at Springerlink.com

Abstract

Soil cation exchange capacity (CEC) strongly influences the chemical, physical, and biological properties of soil. As the direct measurement of the CEC is difficult, costly, and time-consuming, the indirect estimation of CEC from chemical and physical parameters has been considered as an alternative method by researchers. Accordingly, in this study, a new hybrid model using a support vector machine (SVM), coupling with particle swarm optimization (PSO), and integrated invasive weed optimization (IWO) algorithm is developed for estimating the soil CEC. The physical and chemical data (i.e., clay, organic matter (OM), and pH) from two field sites of Taybad and Semnan in Iran were used for validating the new proposed approach. The ability of the proposed model (SVM-PSOIWO) was compared with the individual model (SVM) and the hybrid model (SVM-PSO). The results of the SVM-PSOIWO model were also compared with those of existing studies. Different performance evaluation criteria such as RMSE, R^2 , MAE, RRMSE, and MAPE, Box plots, and scatter diagrams were used to test the ability of the proposed models for estimation of the CEC values. The results showed that the SVM-PSOIWO model with the RMSE (R^2) of $0.229 \text{ Cmol} + \text{kg}^{-1}$ (0.924) was better than those of the SVM and SVM-PSO models with the RMSE (R^2) of $0.335 \text{ Cmol} + \text{kg}^{-1}$ (0.843) and $0.279 \text{ Cmol} + \text{kg}^{-1}$ (0.888), respectively. Furthermore, the ability of the SVM-PSOIWO model compared with existing studies, which used the genetic expression programming, artificial neural network, and multivariate adaptive regression splines models. The results indicated that the SVM-PSOIWO model estimates the CEC more accurately than existing studies.

Keywords Soil cation exchange capacity · Soil physics · Support vector machine · Particle swarm optimization · Invasive weed optimization algorithm

1 Introduction

The soil cation exchange capacity (CEC) is the total number of exchangeable cations that held in the soil by electrostatic forces at a specified pH in the unit weight (Amini et al. 2005; Velde and Bauer 2014). It is commonly referred to the number of negative charges in soil. CEC is one of the important chemical properties of soil, which shows the ability of soil to maintain positively charged

ions, and also it is a good index to indicate the quality, fertility, and efficiency of soil (Arias et al. 2005; Khaledian et al. 2017). Even though it is possible to measure the CEC directly, the acquisition process is difficult and expensive, especially in Iran due to more significant amounts of lime and gypsums (Amini et al. 2005; Carpena et al. 1972; McBratney et al. 2002). Hence, several methods have been developed to estimate the CEC from soil properties, which can be easily measured. In general, there are two main groups of literature for estimating the CEC. The first group of studies focuses on developing regression-based empirical models called pedotransfer functions (PTFs). This set of methods tried to establish empirical relation between the CEC and physical and chemical properties of soil, such as soil pH, soil texture, and organic matter, which can be easily measured (Bell and Van Keulen 1995; Drake and Motto 1982; Fooladmand 2008; Ghorbani et al. 2015;

✉ Babak Mohammadi
babak.mohammadi@nateko.lu.se

¹ Department of Water and Environmental Engineering,
Faculty of Civil Engineering, Shahrood University of
Technology, Shahrood, Iran

² Department of Physical Geography and Ecosystem Science,
Lund University, Sölvegatan 12, SE-223 62 Lund, Sweden

Krogh et al. 2000; Manrique et al. 1991). In recent years, the second group of studies involving CEC estimates is related to machine learning methods such as support vector machines (SVM), artificial neural networks (ANN), adaptive neuro-fuzzy inference system (ANFIS), genetic expression programming (GEP), and others. In recent years, several studies, such as Emamgholizadeh (2012); Parhizkar et al., (2015); Emamgholizadeh et al., (2017); Parsaie et al., (2018a, b); Maroufpoor et al., (2018); Emamgholizadeh et al. (2018); Emamgholizadeh, and Karimi (2019); Bazoobandi et al., (2019); Parsaie et al., (2018a,b), have reported successful applications of these intelligent models to estimate parameters in soil science, water engineering, and civil engineering, for modeling soil CEC in a nonlinear framework and create relationships between inputs (physicochemical properties of soil data) and output (CEC) (da Silva et al. 2018; Emamgholizadeh et al. 2015; Ghorbani et al. 2015; Jafarzadeh et al. 2016; Kashi et al. 2014; Keshavarzi and Sarmadian 2010; Keshavarzi et al. 2017; Liao et al. 2014). One of the benefits of using artificial intelligence models over pedotransfer functions (PTFs)-based models lies in not depending on specific functions with unusual patterns but just in the training process. Furthermore, the accuracy of these models to retrieve the CEC was better than regression-based PTFs models particularly, when the relationship between input and output data is unknown, and also there is a nonlinear and complex relationship between them (Emamgholizadeh et al. 2015).

Although most of the previous studies indicate the superiority of the data-driven models in comparison with the regression-based PTFs models, it is possible to reduce the learning error and increase the performance of these models by coupling them with meta-heuristic optimization algorithms such as particle swarm optimization (PSO), invasive weed optimization (IWO), genetic algorithm (GA), and cultural algorithm (CA). The literature of past studies indicates that the integrated forecasting methods outperform the individual predictions (Da and Xiurun 2005; Mohammadi and Mehdizadeh 2020; Holland 1975; Kennedy and Eberhart 1995; Mehrabian and Lucas 2006; Meshram et al. 2019; Ndiritu and Daniell 2001; Reynolds 1994; Tien Bui et al. 2018; Mohammadi et al. 2020a). Therefore, in the current study, two meta-heuristic optimization algorithms, namely the PSO, and the IWO were used to predict CEC. The IWO is a nature-inspired meta-heuristic optimization technique, which proposed by Mehrabian and Lucas (Mehrabian and Lucas 2006), inspired by the dynamic growth of the weed's colony and can be used for continuous function optimization. Also, the PSO algorithm was introduced by Kennedy and Eberhart (1995). This technique is a population-based search

algorithm inspired on the social behavior of birds within a flock.

In recent years, the PSO and IWO algorithms are successfully used for improving the prediction ability in soil science. For example, Moazenzadeh and Mohammadi (2019) utilized a hybrid of bio-inspired meta-heuristic optimization algorithms and SVM model to assess the soil temperature. Their findings indicated that the proposed hybrid algorithm was a powerful computational tool for estimating soil temperature compared to the SVM model. Xue et al. (2014) applied hybrid SVM-PSO to predict slope stability of soil, and they stated that using the PSO algorithm enhanced the forecasting accuracy of the SVM model, and the PSO-SVM can be used as a powerful model to estimate the slope stability. Rui et al. (2019) used the PSO-SVM for estimation of the total organic carbon (TOC) content from DT (acoustic log), DEN (bulk density), GR (natural gamma-ray), SP (natural potential), and some array resistivity logged similar M2R1 to M2RX. They found that this model can be used as an efficient and reliable method to estimate TOC content. Also, other researchers such as Tang et al. (2019), Wang et al. (2013) Du et al. (2017) used the PSO-SVM in their studies. Moazenzadeh et al. (2018) applied hybrid support vector machines (SVM) with meta-heuristic optimization algorithms to estimate evaporation values. The results showed that the hybrid model produced a better-estimated result compared to the SVM model alone. In another study, Ghorbani et al. (2017) estimated the field capacity and wilting point of the soil combining the SVM and firefly algorithm models, and they had shown that the hybrid model performs well compared to the SVM method alone. Additionally, Mohammadi et al. (2020b) used a hybrid of grey wolf optimizer and SVR method for modeling lake water level, and they stated that the hybrid model outperforms compared to the SVR model.

Several studies have recently used a hybrid invasive weed optimization algorithm (IWO) and support vector machines (SVM) to find complex relations between inputs and outputs in numerous engineering problems. For example, Huang et al. (2013) applied the invasive weed optimization algorithm (IWO) for optimization of the parameters of support vector machines (SVM). Goli et al. (2018) compared the ability of the invasive weed optimization (IWO) algorithm with three meta-heuristic algorithms, including particle swarm optimization (PSO), genetic algorithm (GA), and cultural algorithm (CA), to improve the artificial intelligence models, namely support vector regression (SVR), multilayer perceptron (MLP), and adaptive-neural-based fuzzy inference system (ANFIS) to predict the demand of dairy products (DDP). Their results showed that using the hybrid IWO and ANFIS model produced better estimation compared to the other hybrid

models. These studies confess that when meta-heuristic optimization algorithms are used to learn the target function in the intelligence models such as SVM, the new hybrid model can be better learned and therefore perform better than the SVM model alone. Due to the aforementioned advantages of the hybrid models, in the current study, the SVM model was coupled with IWOA and PSO methods for estimating the CEC.

There is no work to date, of which we are aware, that has no other study has used SVM-PSOIWO to estimate CEC. The advantage of the proposed new hybrid model (SVM-PSOIWO) is (1) optimization of the objective function to estimate the CEC simultaneously with two methods of meta-heuristic optimization algorithms (IWOA and PSO), (2) local and global search to find the optimal solutions near the definitive answers, so the new proposed model will never drop in the local optimal because all the possible answers at the same time are analyzed, (3) the new proposed model is not sensitive to outlier data and also to the existence of noise in data, and it can consider the extreme values, (4) the optimization of the target function simultaneously by two meta-heuristic optimization algorithms makes it possible to minimize the response of the target function of the SVM model to the extent possible; as a result, the objective function is optimized, and finally, the CEC values are estimated with the highest accuracy and the least error.

The main goal of this study is to examine the ability of the SVM-PSOIWO method to estimate CEC. As a second perspective, a comparison was done between the proposed method and existing methods.

2 Materials and methods

2.1 Case study and data description

The soil physical and chemical data obtained from two field sites, namely Taybad (latitude, 34.6983° N to 34.7000° N; longitude, 60.7667° E to 60.7817° E) and Semnan (latitude, 35.5667° N to 35.5816° N and longitude, 53.4667° E to 53.4817° E), were used in this study (Emamgolizadeh et al. 2015). The area of each site was approximately 400 hectares. Two hundred and fifty soil samples were taken from each site from the top 30 cm of the soil profile. Soil samples randomly and with appropriate distribution were taken from 500 locations in two study sites. The distance between soil sampling points was between 100 and 400 m.

Cation exchange capacity (CEC) determined by Bower's method (Sparks et al. 1996). Also, soil texture, OM percentage, and pH were measured by hydrometer technique (Gee and Bauder 1986), the Walkley–Black

approach (Walkley and Black 1934; Nelson and Sommers, 1982), and pH-meter, respectively. According to the USDA soil classification criteria, the study area has two types of Entisols and Aridisols (USDA, soil taxonomy 2010). Table 1 indicates a summary of the statistical characteristics of the data. To examine the ability of models to predict the CEC, the whole data sets consisted of 500 experimental data points of organic matter (OM), pH, clay, and soil cation exchange capacity (CEC) split into two categories of training and testing based on simple random sampling approach. Overall, 80% of data ($N = 400$ data points) considered for training, and 20% of the remaining data ($M = 100$ data points) for testing (see Fig. 1).

The previous results showed that CEC depends on many factors such as soil texture, organic matter, soil humus content, and soil pH. Among these parameters, OM, clay, and pH are more important than other parameters (Bell and Van Keulen 1995; Brady and Weil 2016; Emamgolizadeh et al. 2015; Fooladmand 2008). Bell and Van Keulen (1995), and Krogh et al. (2000) showed that there is a positive correlation between CEC and soil pH. As the soil pH is increasing, the amount of hydrogen held by organic colloids and silicate clays (kaolinite) is ionized, and replaced; therefore, the number of negative charges on the colloids increases and as a result the CEC value increases (Pratt 1961; Sparks 2012). Soil organic matter (OM) is another important parameter of soil that has a significant contribution to the CEC of the soil due to its high surface area and high electrical charge (negative charges). Studies showed that near the soil surface where the organic matter content is higher, the CEC content increases and, conversely, at lower soil depths, it decreases (Oorts et al. 2003; Parfitt et al. 1995; Sparks 2003). Similar to the OM, and pH, there are several reports on the impact of clay content on the CEC of soils. Clay can absorb and retain cations due to a large number of negative charges on their surface, thereby increasing the amount of the CEC (Amini et al. 2005; Emamgolizadeh et al. 2015; Seybold et al. 2005).

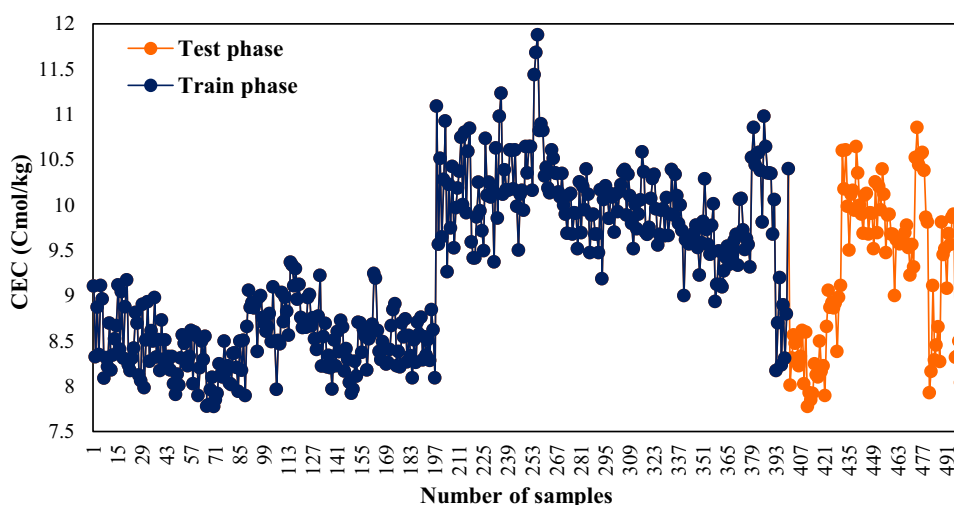
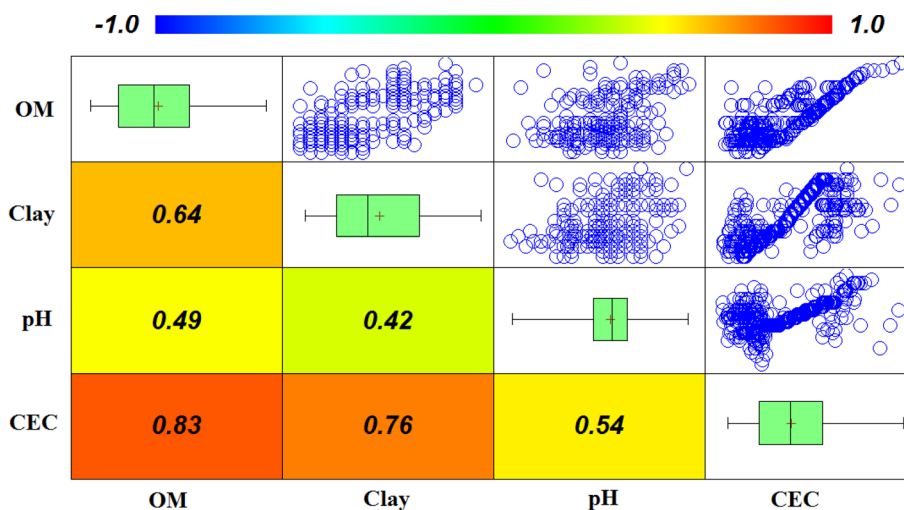
A correlation analysis was done to survey the relationship between soil CEC and clay, OM, and pH (see Fig. 2). For this purpose, the Pearson product-moment correlation performed to find the strength of the linear relationship between variables. Figure 2 shows that there is a high correlation between CEC and OM with $R = 0.83$. Also, the correlation between CEC and clay and pH is 0.76 and 0.54, respectively.

2.2 Support vector machine (SVM)

The SVM method is a supervised learning, and for the first time it was introduced by Vladimir Vapnik in 1995. The support vector machine is an efficient learning system based on bounded optimization theory that utilizes the

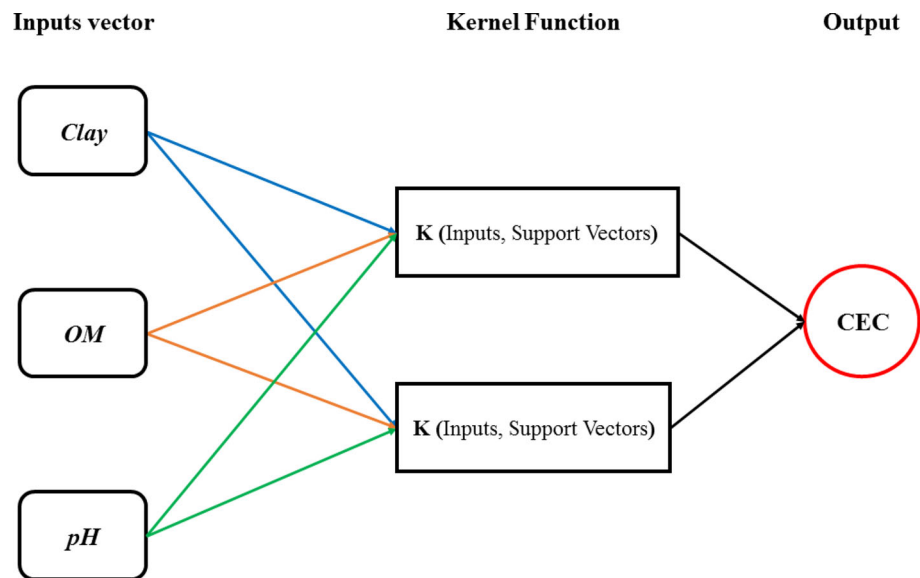
Table 1 Statistical characteristics of data (Emamgholizadeh et al. 2015)

Statistical parameters	CEC ($\text{Cmol}^+ \text{kg}^{-1}$)	Clay (%)	OM (%)	pH
Mean	9.31	17.27	2.18	9.27
Minimum	7.89	10.00	1.20	7.77
Maximum	11.88	27.00	3.72	11.87
Standard Error	0.04	0.20	0.03	0.04
Median	9.23	16.00	2.10	7.90
Mode	9.68	20.00	1.50	7.90
Standard deviation	0.87	4.53	0.56	0.36
Sample variance	0.76	20.52	0.32	0.12
Kurtosis	-1.01	-1.20	-0.89	0.32
Skewness	0.21	0.21	0.31	-0.39

Fig. 1 Sample series plot of CEC measurement data**Fig. 2** Pearson product-moment correlations for all data sets

principle of structural error minimization induction and results in an optimal solution (Tang et al. 2019). To categorize vectors that are not linearly separable, a kernel function such as degreed polynomial, radial basis, or hyperbolic tangent is used to map the observed multidimensional vectors to a space with higher dimensions.

Recently, some researchers suggested radial basic functions (RBF) as a powerful tool for considering as a kernel function in soil and water studies (Moazenazadeh et al. 2018; Mohammadi et al. 2021), and the RBF kernel function parameters were optimized through the trial and

Fig. 3 Schematic structure of the SVM model

error method. Figure 3 shows a schematic structure of the SVM model.

2.3 Particle swarm optimization (PSO)

The PSO meta-heuristic algorithm was first proposed by Kennedy and Eberhart (1995) for optimization of the complicated process. The PSO algorithm is inspired by the collective performance of animal groups such as birds and fishes (Assareh et al. 2010). In this algorithm, a bunch of creatures, which are called particles, spread in the search area. Every single particle approximates its situation relative to the target position. They adjust their position and the velocity based on the current situation and the best position they were already in, and the situation of the best particles in the bunch:

$$V_{id}^t = wV_{id}^{t-1} + C_1r_1(P_{id}^t - x_{id}^t) + C_2r_2(P_{gd}^t - x_{gd}^t), d = 1, 2, \dots, D \quad (1)$$

where x_{id}^t indicates the location of the particle $id = 1, \dots, D$ in iteration t , V_{id}^t is velocity of particle $id = 1, \dots, D$ in iteration t , P_{id}^t is the best location of the particle $id = 1, \dots, D$ in iteration t , P_{gd}^t is the global best position of the article $gd = 1, \dots, D$ in iteration t , w expresses the inertia weight, C_1 expresses the cognition learning factor, C_2 expresses the social learning factor, and r_1 and r_2 denote the random values in $[0, 1]$.

The basic steps for implementing the algorithm are as follows: step (1) generating the initial swarm and assessing it, step (2) evaluation of the fitness of every single particle within the bunch, step (3) update velocity of every single particle according to Eq. 1 and update the position by

$x_{id}^{t+1} = x_{id}^t + v_{id}^t$, step (4) each particle moves to the next position based on the $x_{id}^{t+1} = x_{id}^t + v_{id}^t$, step (5) the algorithm will stop when the termination criterion is satisfied or returned to the step 2.

2.4 Invasive weed optimization (IWO)

The invasive weed optimization (IWO) was introduced by Mehrabian and Lucas (2006). It is an intelligent and evolutionary algorithm for solving optimization problems. In this algorithm, the meta-heuristic procedure is inspired by the dynamic growth performance of the weeds colony in nature (Safari et al. 2020). Also, this iterating algorithm is useful for continuous functions works in five steps consist of initialization, reproduction, spatial dispersal, competitive exclusion, and termination condition (Fig. 4).

Each step in the IWO algorithm is summarized below:

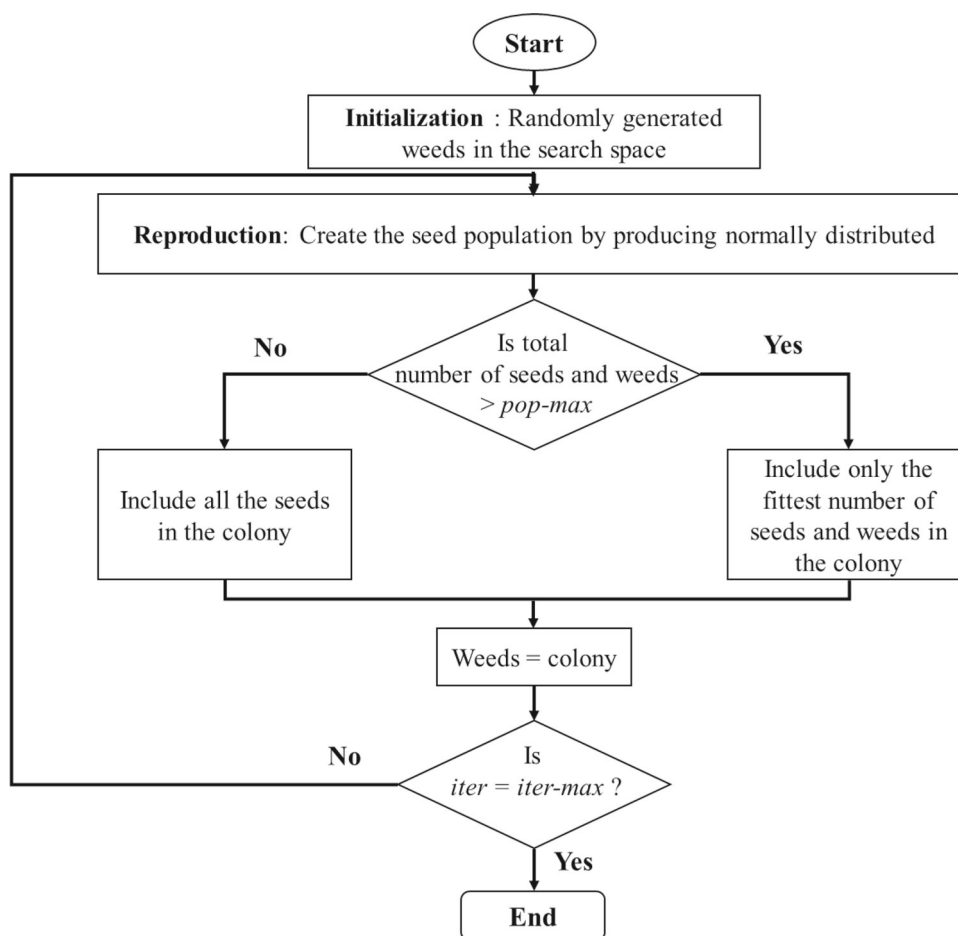
Step 1- Initialization: In the first stage, the initial population of weeds, $X = \{x_1, x_2, \dots, x_{PSO}\}$, is generated in the search space, PSO is the size of the initial population of weeds. Each weed, $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ is an n -dimensional real-valued vector, and each dimension x_{ik} of x_i generated as follows:

$$x_{ik} = lb_k + r \times (up_k - lb_k), i = 1, 2, \dots, PSO, k = 1, 2, \dots, n \quad (2)$$

where r is a uniform random number between 0 and 1. lb_k and up_k denote the lower and upper bounds for the k dimension, respectively.

Step 2- Reproduction: Each weed produces seeds based on its fitness. In fact, the number of seeds (S_i) produced by a weed (x_i) is determined by the fitness of the plant (Eq. 3). A weed that has higher fitness has a greater chance of reproduction.

Fig. 4 Sketch procedure of the basic IWO algorithm



$$s_i = \text{floor} \left(s_{\min} + \frac{s_{\max} - s_{\min}}{f_{\max} - f_{\min}} \times (f(x_i) - f_{\min}) \right) \quad (3)$$

where S_i is the number of seeds generated by weed x_i , $f(x_i)$ stands for the fitness of x_i , so $f_{\min} = \min_{x_i \in X} f(x_i)$ and $f_{\max} = \max_{x_i \in X} f(x_i)$. Floor is a function which rounds the element to the nearest integer towards minus infinity. s_{\min} and s_{\max} define the number of seeds generated by the worst and the best weeds in the population, respectively. The generated seeds in this step have a normal distribution with a mean equal to zero but the variance is different.

Step 3- Spatial Dispersal: In this step, the randomness and adaptation are done in the IWO algorithm. In order to group fitter plants and eliminate inappropriate ones, the nonlinearity at each iteration must be decreased. To achieve this, in each generation over time, the standard deviation (σ) of the normal distribution is reduced from specific initial value (σ_0) to final value (σ_f) according to Eq. 4:

$$\sigma_{\text{iter}} = \frac{(\text{iter}_{\max} - \text{iter})^a}{(\text{iter}_{\max})^a} \times (\sigma_0 - \sigma_f) + \sigma_f \quad (4)$$

σ_{iter} represents the standard deviation at the current iteration; iter , and iter_{\max} define the maximum number of iterations, and α , which is generally set to 3, is a nonlinear modulation index.

Step 4- Competitive Exclusion: In this step, all the weeds in the initialization step and the seeds produced in the reproduction step joint together to form the next generation population. Because the number of weeds does not exceed a given maximum allowable population in a colony, PS_{\max} , the mechanism of the competitive elimination is used to the members of the population, and weeds with lower fitness will be eliminated.

Step 5- Termination Condition: In this step, steps 2 to 4 are repeated until a given termination condition has occurred. Termination condition could be the maximum number of iteration or the maximum elapsed CPU time.

2.5 Hybrid models (SVM-PSO and SVM-PSOIWO)

SVM model does not require complicated calculations, but it needs to adjust network weights and coordinate neurons when performing local convergence and optimization in the network. One of the novelties of this study is to apply

the new hybrid SVM-PSOIWO method in comparison with basic SVM and SVM-PSO to obtain a rapid and efficient method for predicting the CEC in the study area. For optimizing the train performance, the PSO algorithm was then integrated with the ordinary SVM model to construct a single-phase SVM-PSO model. And the PSO aimed to determine the optimized values of the ordinary SVM model parameters (i.e., weights and biases) at the model's training section. Then, two-phase hybrid model (SVM-PSOIWO) was also constructed to further improve the SVM-PSO model for acquiring the best synaptic weights and biases within the two-phase hybrid model's hidden layers. SVM-PSOIWO stops when a mathematical fit between support vector machine weights and the IWO is created, or the maximum number of iterations occurs. It is an estimator hybrid procedure that utilizes both support vector machine capabilities and optimization algorithm capabilities. Some research has shown that such a hybrid technique can predict more successful results (Ghorbani et al. 2017; Moazenzadeh and Mohammadi 2019; Mohammadi and Mehdizadeh 2020). The flowchart of the SVM-PSOIWO is shown in Fig. 5.

2.6 Model performance criteria

The estimated soil CEC values were compared with observed values using five different performance evaluation criteria: The root mean square error (RMSE), the coefficient of determination (R^2), the mean absolute error (MAE), the relative root mean square error (RRMSE), and

Table 2 Mathematical expressions of statistical metrics

Evaluation criteria	Equation
RMSE	$RMSE = \left\{ \left[\sum_1^n (O_i - P_i)^2 \right] / n \right\}^{1/2}$
R^2	$R^2 = \left(\frac{1}{n} \times \frac{\sum (O_i - \bar{O})(P_i - \bar{P})}{(\sigma_o)(\sigma_p)} \right)^2$
MAE	$MAE = \frac{1}{n} \sum_1^n O_i - P_i $
RRMSE	$RRMSE = \frac{\left\{ \left[\sum_1^n (O_i - P_i)^2 \right] / n \right\}^{1/2}}{1/n \sum_1^n P_i} \times 100$
MAPE	$MAPE = \frac{1}{n} \sum_{i=1}^n \left \frac{P_i - O_i}{O_i} \right \times 100$

mean absolute percentage error (MAPE). Table 2 shows mathematical expressions of these performance evaluation criteria,

where O_i is the observed CEC values, P_i is the predicted CEC values, n is the number of CEC data, and the bar denotes the mean of the variable.

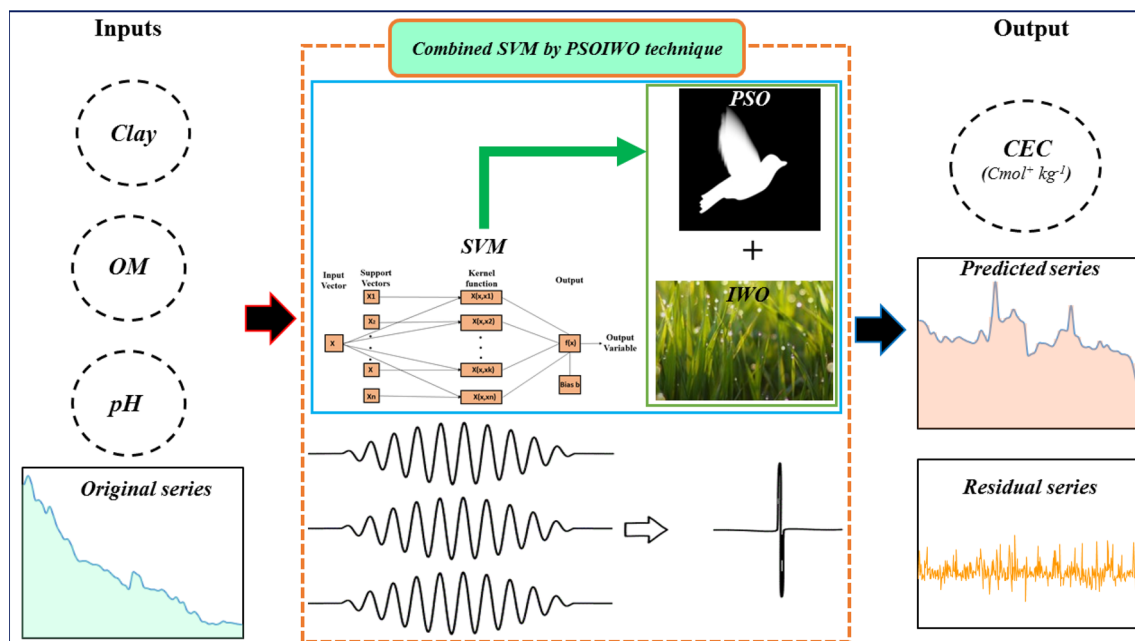


Fig. 5 The schematic diagram includes input and output along with details of the SVM-PSOIWO hybrid method

3 Results and discussion

3.1 CEC estimates from the SVM, SVM-PSO, and SVM-PSOIWO models

In this study, the SVM, SVM-PSO, and SVM-PSOIWO methods were employed to predict the soil CEC value. For this purpose, 80% of the data (400 data points) was used for training predictor models. In addition, 20% of the data was employed in the testing stage. The proper selection of inputs data for models, i.e., SVM, SVM-PSO, and SVM-PSOIWO, has an important role in the accurate estimation of the CEC. For this purpose, based on the Pearson correlation analysis three variables, namely OM, clay, and pH, were selected among different measured physical and chemical parameters as input data to the models. Three scenarios for the input configurations were defined (see Table 3). These scenarios were selected based on the highest correlation of input parameters with the CEC parameter. The first scenario includes OM which has the highest correlation with the CEC, and in the second scenario the clay parameter, which after OM has the highest correlation with the CEC, added into the first scenario, and finally, the third scenario includes OM, clay, and pH.

After designing the different scenarios highlighted in Table 3, the input configurations were introduced to the mentioned models for implementation of them. Tables 4 and 5 show the RMSE, MAE, MAPE, RRMSE, and R^2 of CEC estimates from the SVM, SVM-PSO, and SVM-PSOIWO methods for training and testing stages, respectively. As can be seen in Tables 4 and 5, indicated with the first input configuration (i.e., OM), the RMSE (R^2), of CEC estimates from SVM, SVM-PSO and SVM-PSOIWO are $0.419 \text{ Cmol} + \text{kg}^{-1}$ (0.772), $0.334 \text{ Cmol} + \text{kg}^{-1}$ (0.846), and $0.298 \text{ Cmol} + \text{kg}^{-1}$ (0.888), respectively, for training, and $0.429 \text{ Cmol} + \text{kg}^{-1}$ (0.740), $0.367 \text{ Cmol} + \text{kg}^{-1}$ (0.807), and $0.316 \text{ Cmol} + \text{kg}^{-1}$ (0.857), for testing.

For the second input configuration, by adding the clay to the second input configuration (i.e., OM, and clay), the accuracy of CEC estimates increased. The RMSE varies from a minimum of $0.243 \text{ Cmol} + \text{kg}^{-1}$ to a maximum of $0.408 \text{ Cmol} + \text{kg}^{-1}$ for training and testing stages. Comparing the results of the first and second input configurations indicates that the average RMSE of CEC estimates decreased by 15.17% and 8.82% for training and testing

stages, respectively. Finally, for the third input configuration (i.e., OM, clay, and pH), using these configurations of input data, the average RMSE decreased by 31.95% and 19.78% compared to the first and second input configurations for training and by 24.19% and 17.76% for testing stages, respectively.

Compared to the SVM model, using the SVM-PSO model to estimate the CEC value the accuracy of the model increased, and the average RMSE, MAE, MAPE, and RRMSE decreased by 16.72%, 20.59%, 20.67%, and 16.60% in the testing stage. Similarly, a comparison of the performance of the SVM, SVM-PSO, and SVM-PSOIWO models implied that the SVM-PSOIWO estimation was much more accurate than both SVM and SVM-PSO methods. Also, the findings in Tables 4 and 5 illustrate that the RMSE of SVM-PSOIWO decreased by approximately 35.81% and 19.49% compared to the SVM and SVM-PSO models for training and by 31.64% and 17.92% for testing, respectively. Overall, the results of this study imply that the SVM-PSOIWO model has been able to estimate the CEC values with low error and it suggests the success of the support vector machine (SVM) coupling with particle swarm optimization (PSO) and integrated invasive weed optimization algorithm.

In order to indicate the performance of the SVM, SVM-PSO, and SVM-PSOIWO models, the scatter plot and residual (error plot) of observed and predicted CEC values from the best input configuration (i.e., OM, clay, and pH), are drawn in Fig. 6 for testing phase. Based on this figure, the agreement between the measured and predicted CEC was very good for the SVM-PSOIWO model in training stage ($R^2 = 0.953$, $\text{RMSE} = 0.190 \text{ Cmol} + \text{kg}^{-1}$, $\text{MAE} = 0.132 \text{ Cmol} + \text{kg}^{-1}$), and in testing stage ($R^2 = 0.924$, $\text{RMSE} = 0.229 \text{ Cmol} + \text{kg}^{-1}$, $\text{MAE} = 0.152 \text{ Cmol} + \text{kg}^{-1}$).

Mentioning the optimized parameters of the used models in the hydrological modeling process is very important because it can help researchers measure their new models with optimized parameters (Mohammadi 2019). Concerning this issue, Table 6 lists the optimized parameters and structure of the models used in this study.

Also, in order to compare the SVM, SVM-PSO, and SVM-PSOIWO models to predict the CEC value, the box plot was used. Figure 7 shows the results of models for three scenarios and three models versus the measured data in the testing stage. In the box plot, the statistical characteristics of the measured and predicted soil CEC values are compared. In this figure, the green color represents 25% of the data (first quartile), which is less than the average of the data, and the orange color represents 75% of the data (quartile third). As can be seen, among all used models and scenarios, the SVM-PSOIWOS3 model has the most similar statistical characteristics to the observed values, which

Table 3 The scenarios of the input combinations of models

No	Inputs	Models
1	OM	SVM1 SVM-PSO1 SVM-PSOIWO1
2	OM, Clay	SVM2 SVM-PSO2 SVM-PSOIWO2
3	OM, Clay, pH	SVM3 SVM-PSO3 SVM-PSOIWO3

Table 4 Result of models related to the training phase

Models name	Training phase				
	RMSE (Cmol + kg ⁻¹)	MAE (Cmol + kg ⁻¹)	MAPE (%)	R ²	RRMSE (%)
SVM1	0.419	0.309	3.44	0.772	4.53
SVM2	0.359	0.257	2.855	0.832	3.889
SVM3	0.296	0.202	2.271	0.885	3.209
SVM-PSO1	0.344	0.26	2.881	0.846	3.724
SVM-PSO2	0.298	0.215	2.409	0.888	3.225
SVM-PSO3	0.236	0.151	1.703	0.927	2.553
SVM-PSOIWO1	0.298	0.215	2.409	0.888	3.225
SVM-PSOIWO2	0.243	0.167	1.871	0.923	2.626
SVM-PSOIWO3	0.190	0.132	1.478	0.953	2.058

Table 5 Result of models related to the testing phase

Models name	Testing phase				
	RMSE (Cmol + kg ⁻¹)	MAE (Cmol + kg ⁻¹)	MAPE (%)	R ²	RRMSE (%)
SVM1	0.429	0.311	3.481	0.74	4.632
SVM2	0.408	0.291	3.243	0.765	4.407
SVM3	0.335	0.238	2.67	0.843	3.614
SVM-PSO1	0.367	0.272	2.993	0.807	3.964
SVM-PSO2	0.334	0.242	2.72	0.857	3.605
SVM-PSO3	0.279	0.189	2.118	0.888	3.014
SVM-PSOIWO1	0.316	0.218	2.447	0.857	3.409
SVM-PSOIWO2	0.283	0.203	2.277	0.886	3.058
SVM-PSOIWO3	0.229	0.152	1.722	0.924	2.479

means that the third scenario (i.e., OM, Clay, and pH) are more adequate inputs for modeling CEC. On the other hand, this suggests that the new proposed model, which was first used in this study to predict the CEC, is a successful model and can estimate the CEC values with the least error compared to previous popular models.

In Fig. 8, the comparison between methods (SVM, SVM-PSO, and SVM-PSOIWO) was investigated according to the RRMSE index in the training and testing stages. As expected, in all scenarios and for all used methods, the accuracy of methods in the training stage was better than in the testing stage. Also, based on this index, the use of the third combination of data (i.e., OM, clay, and pH) is the best and most effective combination of input data to estimate the CEC. As shown in Fig. (8), the new model, the SVM-PSOIWO, has been able to reduce the value of the RRMSE index by almost half as compared to the SVM model, which represents a good and positive feature of the newly proposed method. In both the training and testing stages, the results of the new SVM-PSOIWO model are much more satisfactory than SVM and SVM-PSO methods,

so that the SVM-PSOIWO model could significantly reduce the error rate in the CEC estimate.

3.2 Comparing SVM-PSOIWO with previous studies

As shown in Sect. 3.1, the comparison results of CEC estimates for the three methods and SVM-PSOIWO revealed that the best results were achieved when the SVM-PSOIWO model was used with the third combination of data (i.e. OM, clay, and pH). This finding is consistent throughout the study of Emamgholizadeh et al. (2015). To further evaluate the capability of the SVM-PSOIWO method in estimating the CEC parameter, the result of the SVM-PSOIWO model was compared with those of previous studies that used the ANN, GEP, MARS, and MLR models to estimate CEC. The statistical indices of all models in the testing stage are given in Table 7. As shown, compared to other models, the CEC estimates from the SVM-PSOIWO model with R² and RMSE 0.229 Cmol + kg⁻¹, and 0.924 provide accurate results and reduce the RMSE by 9.1%, 28.0%, 38.1%, and 43.9%

Fig. 6 Scatter plot and residual (error plot) of measurement and predict of CEC by SVM3, SVM-PSO3, and SVM-PSOIWO3 models for the testing phase

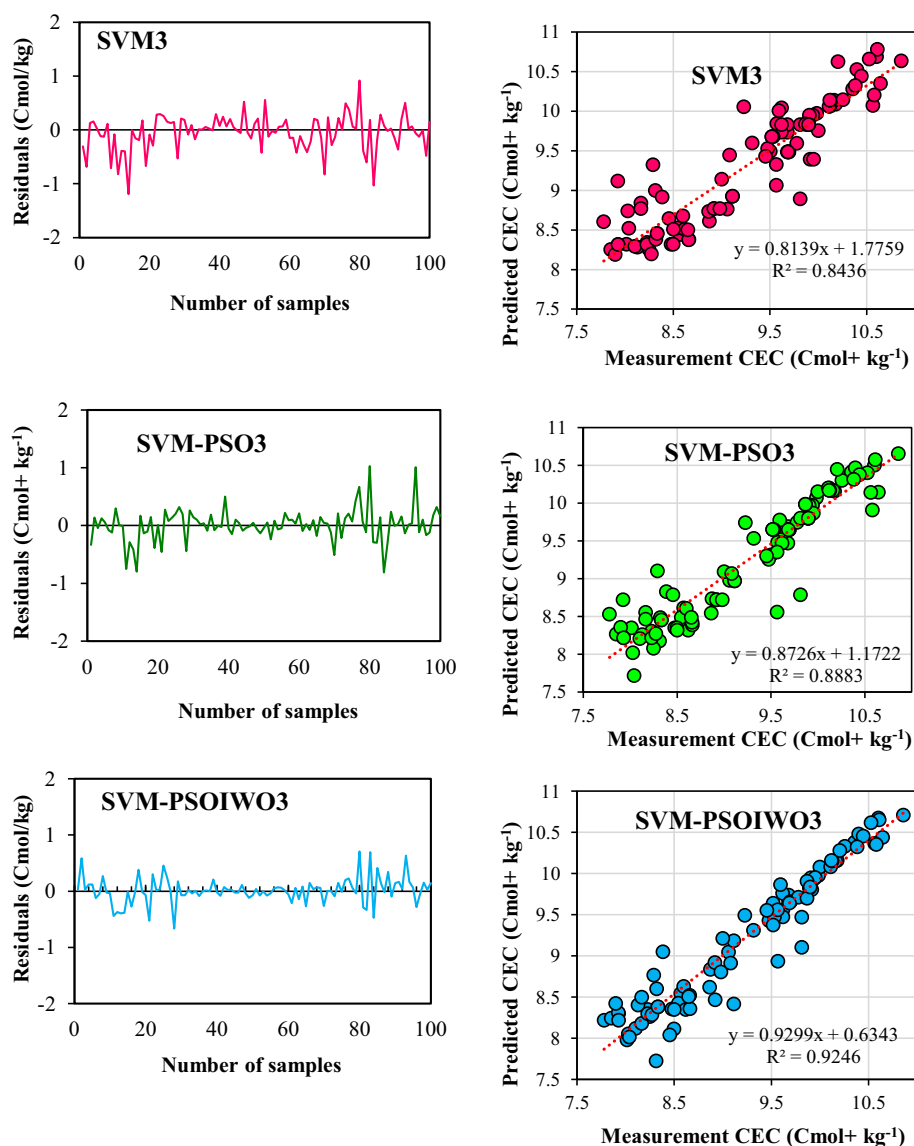


Table 6 Parameters setting for models used

SVM structure		Particle swarm optimization		Invasive weed optimization	
Kernel function	RBF kernel	Maximum number of iterations	500	Maximum number of iterations	500
Bias	0.435	Swarm size	15	Number of an initial population	15
C	1.33–1.85	c1, c2	0.05, 0.05	Maximum number of plant population	25
E	0.19–0.40	r1, r2	0.45, 0.55	Minimum number of seeds	1
Γ	0.42–0.77	w	0.6	Maximum number of seeds	15

compared to ANN, GEP, MARS, and MLR, respectively. Overall, the estimating performance of the SVM model improves when the coupling of this model with particle swarm optimization (PSO) and integrated invasive weed optimization algorithm is used instead of using the conventional SVM model. Also, the results of this study

suggest that SVM-PSOIWO is a viable alternative procedure for the commonly used models such as ANN, GEP, MARS, and MLR models to retrieve CEC.

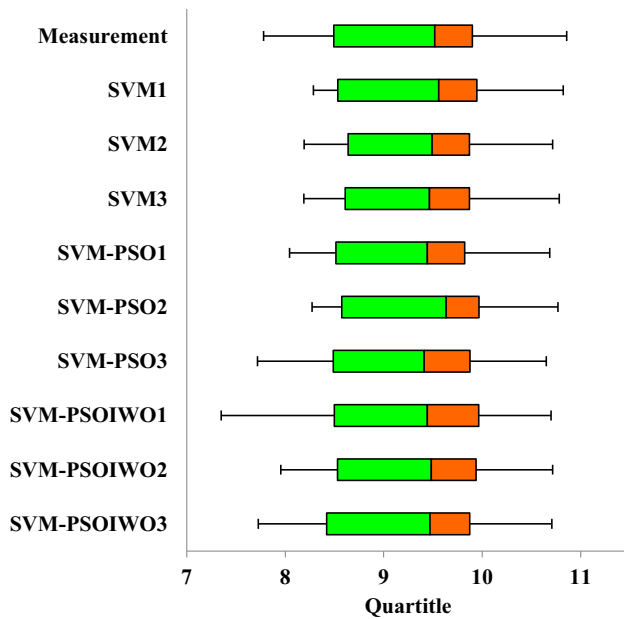


Fig. 7 Box plot of the measured and predicted CEC for the testing phase

4 Conclusion

Soil cation exchange capacity (CEC) is an important parameter in agriculture and soil science. In this research, the SVM-PSOIWO is proposed as a new method for estimating CEC. Accordingly, the physical and chemical data (i.e., clay, OM, and pH) from two field sites of Taybad and Semnan in Iran were used to estimate CEC. For this purpose, three configurations of input data (i.e., clay, OM, and pH) were used to train and test models. It was found that the performance of the three used methods of SVM, SVM-PSO, and SVM-PSOIWO is promising for estimating the CEC as a function of physical and chemical data as input parameters. However, the SVM-PSOIWO performed better than the individual model (SVM) and the hybrid model (SVM-PSO). Moreover, the experiments demonstrated that combinations of clay, OM, and pH are the most effective input parameters for an accurate estimation of the CEC values instead of one and two input combinations of data. In another word, the performance of the models to retrieve the CEC was greatly improved by increasing the number of inputs data. Since the measurements of Clay, OM, and pH are easy and have low cost, therefore, the proposed new hybrid model can be employed to estimate CEC with

Fig. 8 Three-dimensional bar graph of performance models for prediction CEC in both the train and test sections by the RRMSE index

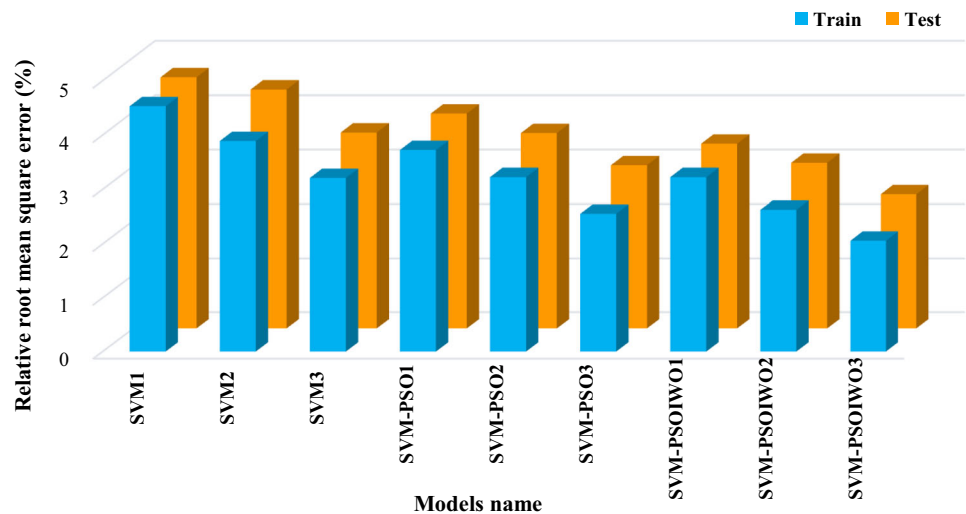


Table 7 Comparing the performance of different studies

Models	MAE (Cmol + kg ⁻¹)	RMSE (Cmol + kg ⁻¹)	R ²
The present study (SVM3)	0.238	0.335	0.843
The present study (SVM-PSO3)	0.189	0.279	0.888
The present study (SVM-PSOIWO3)	0.152	0.229	0.924
ANN (Emamgholizadeh et al. 2015)	0.162	0.252	0.892
MARS (Emamgholizadeh et al. 2015)	0.216	0.318	0.864
GEP (Emamgholizadeh et al. 2015)	0.277	0.37	0.807
MLR (Emamgholizadeh et al. 2015)	0.303	0.408	0.768

acceptable accuracy. The estimated CEC from the SVM, SVM-PSO, and SVM-PSOIWO was also compared with those of existing studies such as ANN, GEP, MARS, and MLR. It was found that the SVM-PSOIWO models estimate the CEC more accurately than those studies. In general, the results of this study showed that the improvement in SVM-PSO provided by the IWO algorithm could be used as a predictive tool along with high-performance optimization to estimate the CEC parameter and other soil and water parameters. The high precision of the proposed method (PSOIWO) can be related to its capability to find the best outcome in the search space, so that this hybrid algorithm simultaneously searches the optimal answer in the local and global search space. Another advantage of this algorithm is that when it finds the optimal solution, all of the other optimal answers analyzed in the neighborhood of the optimal solution, which prevents being trapped in a local optimum. Although the result of the SVM-PSOIWO model demonstrates improvements in the prediction of CEC compared to other artificial intelligence (AI) models, however, same to other AI models, the shortcoming of the proposed model is that it acts like a black box and it must be considered by researchers in their studies.

Funding Open access funding provided by Lund University.

Declarations

Conflict of interest There is no conflict of interest for this research.

Ethical approval This article does not contain any studies with animals performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Amini M, Abbaspour KC, Khademi H, Fathianpour N, Afyuni M, Schulin R (2005) Neural network models to predict cation exchange capacity in arid regions of Iran. *Eur J Soil Sci* 56(4):551–559
- Arias M, Pérez-Novo C, Osorio F, López E, Soto B (2005) Adsorption and desorption of copper and zinc in the surface layer of acid soils. *J Colloid Interface Sci* 288(1):21–29
- Assareh E, Behrang M, Assari M, Ghanbarzadeh A (2010) Application of PSO (particle swarm optimization) and GA (genetic algorithm) techniques on demand estimation of oil in Iran. *Energy* 35(12):5223–5229
- Bazoobandi A, Emamgholizadeh S, Ghorbani H (2019) Estimating the amount of cadmium and lead in the polluted soil using artificial intelligence models. *Eur J Environ Civ Eng*. <https://doi.org/10.1080/19648189.2019.1686429>
- Bell M, Van Keulen H (1995) Soil pedotransfer functions for four Mexican soils. *Soil Sci Soc Am J* 59(3):865–871
- Brady NC, Weil RR (2016) The nature and properties of soils, 15th edn. Pearson Education, London, p 1104
- Carpena O, Lax A, Vahtras K (1972) Determination of exchangeable cations in calcareous soils. *Soil Sci* 113(3):194–199
- da Silva ML, Martins JL, Ramos MM, Bijani R (2018) Estimation of clay minerals from an empirical model for Cation Exchange Capacity: an example in Namorado oilfield, Campos Basin, Brazil. *Appl Clay Sci* 158:195–203
- Da Y, Xiurun G (2005) An improved PSO-based ANN with simulated annealing technique. *Neurocomputing* 63:527–533
- Drake EH, Motto H (1982) An analysis of the effect of clay and organic matter content on the cation exchange capacity of New Jersey soils. *Soil Sci* 133(5):281–288
- Du J, Liu Y, Yu Y, Yan W (2017) A prediction of precipitation data based on support vector machine and particle swarm optimization (PSO-SVM) algorithms. *Algorithms* 10(2):57
- Emamgholizadeh S (2012) Neural network modeling of scour cone geometry around outlet in the pressure flushing. *Glob Nest J* 14:540–549
- Emamgholizadeh S, Bateni S, Shahsavani D, Ashrafi T, Ghorbani H (2015) Estimation of soil cation exchange capacity using genetic expression programming (GEP) and multivariate adaptive regression splines (MARS). *J Hydrol* 529:1590–1600
- Emamgholizadeh S, Shahsavani S, Eslami MA (2017) Comparison of artificial neural networks, geographically weighted regression and Cokriging methods for predicting the spatial distribution of soil macronutrients (N, P, and K). *Chin Geogra Sci* 27(5):747–759
- Emamgholizadeh S, Esmaeilbeiki F, Babak M, Zarehaghi D, Maroufpoor E, Rezaei H (2018) Estimation of the organic carbon content by the pattern recognition method. *Commun Soil Sci Plant Anal* 49(17):2143–2154
- Emamgholizadeh S, Karimi R (2019) A comparison of artificial intelligence models for the estimation of daily suspended sediment load: a case study on the Telar and Kasilian rivers in Iran. *Water Supply* 19(1):165–178
- Fooladmand HR (2008) Estimating cation exchange capacity using soil textural data and soil organic matter content: a case study for the south of Iran. *Arch Agron Soil Sci* 54(4):381–386
- Gee GW, Bauder JW (1986) Particle-size analysis. In: Klute A (ed) *Methods of soil analysis: part 1*. Agronomy handbook No 9. American Society of Agronomy and Soil Science Society of America, Madison, WI, pp 383–411
- Ghorbani H, Kashi H, Hafezi Moghadas N, Emamgholizadeh S (2015) Estimation of soil cation exchange capacity using multiple regression, artificial neural networks, and adaptive neuro-fuzzy inference system models in Golestan Province, Iran. *Commun Soil Sci Plant Anal* 46(6):763–780
- Ghorbani MA, Shamsirband S, Haghi DZ, Azani A, Bonakdari H, Ebtehaj I (2017) Application of firefly algorithm-based support vector machines for prediction of field capacity and permanent wilting point. *Soil Tillage Res* 172:32–38
- Goli A, Khademi Zareh H, Tavakkoli-Moghaddam R, Sadeghieh A (2018) A comprehensive model of demand prediction based on hybrid artificial intelligence and metaheuristic algorithms: a case study in dairy industry. *J Ind Syst Eng* 11(4):190–203

- Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor, MI, USA
- Huang H, Ding S, Zhu H, Xu X (2013) Invasive weed optimization algorithm for optimizing the parameters of mixed kernel twin support vector machines. *J Comput* 8(8):2077–2084
- Jafarzadeh AA, Pal M, Servati M, FazeliFard MH, Ghorbani MA (2016) Comparative analysis of support vector machine and artificial neural network models for soil cation exchange capacity prediction. *Int J Environ Sci Technol* 13(1):87–96
- Kashi H, Emamgholizadeh S, Ghorbani H (2014) Estimation of soil infiltration and cation exchange capacity based on multiple regression, ANN (RBF, MLP), and ANFIS models. *Commun Soil Sci Plant Anal* 45(9):1195–1213
- Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of the IEEE international conference on neural networks. Citeseer 1942–1948
- Keshavarzi A, Sarmadian F (2010) Comparison of artificial neural network and multivariate regression methods in prediction of soil cation exchange capacity (Case study: Ziara region). *Desert* 15(2):167–174
- Keshavarzi A, Sarmadian F, Shiri J, Iqbal M, Tirado-Corbalá R, Omran ESE (2017) Application of ANFIS-based subtractive clustering algorithm in soil cation exchange capacity estimation using soil and remotely sensed data. *Measurement* 95:173–180
- Khaledian Y, Brevik EC, Pereira P, Cerdà A, Fattah MA, Tazikheh H (2017) Modeling soil cation exchange capacity in multiple countries. *CATENA* 158:194–200
- Krogh L, Breuning-Madsen H, Greve MH (2000) Cation-exchange capacity pedotransfer functions for Danish soils. *Acta Agricult Scand Sect B-Plant Soil Sci* 50(1):1–12
- Liao K, Xu S, Wu J, Zhu Q, An L (2014) Using support vector machines to predict cation exchange capacity of different soil horizons in Qingdao City, China. *J Plant Nutr Soil Sci* 177(5):775–782
- Manrique LA, Jones CA, Dyke PT (1991) Predicting cation-exchange capacity from soil physical and chemical properties. *Soil Sci Soc Am J* 55(3):787–794
- Maroufpoor E, Sanikhani H, Emamgholizadeh S, Kişi Ö (2018) Estimation of Wind drift and evaporation losses from sprinkler irrigation systems by different data-driven methods. *Irrig Drain* 67(2):222–232. <https://doi.org/10.1002/ird.2182>
- McBratney AB, Minasny B, Cattle SR, Vervoort RW (2002) From pedotransfer functions to soil inference systems. *Geoderma* 109(1–2):41–73. [https://doi.org/10.1016/S0016-7061\(02\)00139-8](https://doi.org/10.1016/S0016-7061(02)00139-8)
- Mehrabian AR, Lucas C (2006) A novel numerical optimization algorithm inspired from weed colonization. *Eco Inform* 1(4):355–366
- Meshram SG, Ghorbani MA, Deo RC, Kashani MH, Meshram C, Karimi V (2019) New approach for sediment yield forecasting with a two-phase feedforward neuron network-particle swarm optimization model integrated with the gravitational search algorithm. *Water Resour Manag* 33(7):2335–2356
- Moazenzadeh R, Mohammadi B (2019) Assessment of bio-inspired metaheuristic optimisation algorithms for estimating soil temperature. *Geoderma* 353:152–171
- Moazenzadeh R, Mohammadi B, Shamshirband S, Chau K-w (2018) Coupling a firefly algorithm with support vector regression to predict evaporation in northern Iran. *Eng Appl Comput Fluid Mech* 12(1):584–597
- Mohammadi B (2019) Predicting total phosphorus levels as indicators for shallow lake management. *Ecological Indicators* 107: 105664.
- Mohammadi B, Ahmadi F, Mehdizadeh S, Guan Y, Pham QB, Linh NTT, Tri DQ (2020a) Developing novel robust models to improve the accuracy of daily streamflow modeling. *Water Resour Manage* 34(10):3387–3409
- Mohammadi B, Guan Y, Aghelpour P, Emamgholizadeh S, Pillco Zolá R, Zhang D (2020b) Simulation of titicaca lake water level fluctuations using hybrid machine learning technique integrated with grey wolf optimizer algorithm. *Water* 12(11):3015
- Mohammadi B, Guan Y, Moazenzadeh R, Safari MJS (2021) Implementation of hybrid particle swarm optimization-differential evolution algorithms coupled with multi-layer perceptron for suspended sediment load estimation. *CATENA* 198:105024
- Mohammadi B, Mehdizadeh S (2020) Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm. *Agricult Water Manag* 237:106145
- Nelson DW, Sommers LE (1982) Total carbon, organic carbon, and organic matter. In: Page AL, Miller RH, Keeney DR (eds) *Methods of soil analysis. Part II, 2 edn.* American Society of Agronomy, Madison, WI, USA, pp 539–580
- Ndiritu J, Daniell T (2001) An improved genetic algorithm for rainfall-runoff model calibration and function optimization. *Math Comput Modell* 33(6–7):695–706
- Oorts K, Vanlauwe B, Merckx R (2003) Cation exchange capacities of soil organic matter fractions in a Ferric Lixisol with different organic matter inputs. *Agricult Ecosyst Environ* 100(2–3):161–171
- Parfitt R, Giltrap D, Whitton J (1995) Contribution of organic matter and clay minerals to the cation exchange capacity of soils. *Commun Soil Sci Plant Anal* 26(9–10):1343–1355
- Parhizkar S, Ajdari K, Kazemi GA, Emamgholizadeh S (2015) Predicting water level drawdown and assessment of land subsidence in Damghan aquifer by combining GMS and GEP models. *Geopersia* 5(1):63–80
- Parsaie A, Emamgholizadeh S, Azamathulla HM, Haghiabi AH (2018a) ANFIS-based PCA to predict the longitudinal dispersion coefficient in rivers. *Int J Hydrol Sci Technol* 8(4):410–424
- Parsaie A, Emamgholizadeh S, Haghiabi AH, Moradinejad AJWS, Supply TW (2018b) Invest Trap Effic Retent Dams 18(2):450–459
- Pratt P (1961) Effect of pH on the cation-exchange capacity of surface soils. *Soil Sci Soc Am J* 25(2):96–98
- Reynolds RG (1994) An introduction to cultural algorithms. In: Proceedings of the third annual conference on evolutionary programming. World Scientific 131–139.
- Rui J, Zhang H, Zhang D, Han F, Guo Q (2019) Total organic carbon content prediction based on support-vector-regression machine with particle swarm optimization. *J Pet Sci Eng* 180:699–706. <https://doi.org/10.1016/j.petrol.2019.06.014>
- Safari MJS, Mohammadi B, Kargar K (2020) Invasive weed optimization-based adaptive neuro-fuzzy inference system hybrid model for sediment transport with a bed deposit. *J Clean Prod* 276:124267
- Seybold C, Grossman R, Reinsch T (2005) Predicting cation exchange capacity for soil survey using linear models. *Soil Sci Soc Am J* 69(3):856–863
- Sparks DL (2003) Environmental soil chemistry. Academic Press, New York
- Sparks DL (2012) Advances in agronomy, vol 118. Academic Press
- Sparks DL, Page AL, Helmke PA, Leoppert RH, Soltanpour PN, Tabatabai MA, Johnston GT, Summer ME (1996) Methods of soil analysis. Soil Science Society of America and American Society of Agronomy, Madison, Wisconsin
- Tang X, Hong H, Shu Y, Tang H, Li J, Liu W (2019) Urban waterlogging susceptibility assessment based on a PSO-SVM method using a novel repeatedly random sampling idea to select negative samples. *J Hydrol* 576:583–595

- Tien Bui D, Khosravi K, Li S, Shahabi H, Panahi M, Singh VP, Chapi K, Shirzadi A, Panahi S, Chen W, Bin Ahmad B (2018) New hybrids of ANFIS with several optimization algorithms for flood susceptibility modeling. *Water* 10(9):1210
- Velde BD, Bauer A (2014) *Minor element geochemistry at the earth's surface: movement of chemical elements*. Springer, Andreas Bauer. Velde
- Walkley A, Black IA (1934) An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Sci* 37:29–38
- Wang W, Xu DM, Chau K-w, Chen S (2013) Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD. *J Hydroinf* 15(4):1377–1390
- Xue X, Yang X, Chen X (2014) Application of a support vector machine for prediction of slope stability. *Sci China Technol Sci* 57(12):2379–2386

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