



Modelling of the impact of water quality on the infiltration rate of the soil

Balraj Singh¹ · Parveen Sihag² · Surinder Deswal²

Received: 29 May 2017 / Accepted: 3 January 2019 / Published online: 14 January 2019
© The Author(s) 2019

Abstract

The concept behind of this paper is to check the potential of the three regression-based techniques, i.e. M5P tree, support vector machine (SVM) and Gaussian process (GP), to estimate the infiltration rate of the soil and to compare with two empirical models, i.e. Kostiakov model and multi-linear regression (MLR). Totally, 132 observations were obtained from the laboratory experiments, out of which 92 observations were used for training and residual 40 for testing the models. A double-ring infiltrometer was used for experimentation with different concentrations (1%, 5%, 10% and 15%) of impurities and different types of water quality (ash and organic manure). Cumulative time (T_f), type of impurities (I_t), concentration of impurities (C_i) and moisture content (W_c) were the input variables, whereas infiltration rate was considered as target. For SVM and GP regression, two kernel functions (radial based kernel and Pearson VII kernel function) were used. The results from this investigation suggest that M5P tree technique is more precise as compared to the GP, SVR, MLR approach and Kostiakov model. Among GP, SVR, MLR approach and Kostiakov model, MLR is more accurate for estimating the infiltration rate of the soil. Thus, M5P tree is a technique which could be used for modelling the infiltration rate for the given data set. Sensitivity analyses suggest that the cumulative time (T_f) is the major influencing parameter on which infiltration rate of the soil depends.

Keywords Double-ring infiltrometer · Gaussian process · Support vector regression · M5P tree model

Introduction

The process in which water moves into the soil through the top surface soil is called the infiltration, and the rate by which it enters into the soil is called the infiltration rate (Haghighi et al. 2010). It plays the important role in the hydrologic cycle. There are many factors which influence the infiltration rate, that is, rainfall intensity, suction head, humidity, water content, types of impurities, field density and humidity. It is associated with the surface runoff and

groundwater recharge (Uloma et al. 2014) and also helpful in water supply system, landslides, design of irrigation, flood control system and drainage (Igbadun and Idris 2007). With the help of infiltration rate, we can easily find out sorptivity and unsaturated hydraulic conductivity of the soil (Chow et al. 1988; Scotter et al. 1982). Hydraulic properties of soil are necessary for design of drainage system (Brooks and Corey 1964). At catchment level, infiltration characteristic is one of the dominant factors in determining the flooding condition (Bhave and Sreeja 2013). The soil capacity of infiltration affects the amount of surface flow (Diamond and Shanley 2003). Infiltration rate in soil is inversely proportion to the water-holding capacity of soil (Singh et al. 2014). Physical changes of soil also affect the infiltration rate (Gupta and Gupta 2008; Smith 2006; Micheal 1978).

Water quality of soil is also affected the infiltration rate and ultimately affected the natural and artificial ground water recharge. Generally, there are many impurities present in the earth surface which can easily mix with the water and changes the quality of the water. Many people studied about

✉ Balraj Singh
balrajzinder@gmail.com

Parveen Sihag
parveen12sihag@gmail.com

Surinder Deswal
deswal.leo@gmail.com

¹ Civil Engineering Department, National Institute of Technology, Hamirpur, Hamirpur 177005, India

² Civil Engineering Department, National Institute of Technology, Kurukshetra, Kurukshetra 136118, India

the concept of the water quality and infiltration. Singh et al. (2017) used the two types of impurities (ash and organic manure) in his study with three soft computing techniques (M5P model tree, artificial neural network and random forest) and found that random forest predicts the infiltration rate well as compared to the other methods. Sihag (2018) studied the infiltration rate by mixing different proportions fly ash and rice husk ash in sand with fuzzy logic and artificial neural network and found that artificial neural network outperforms the fuzzy logic. Singh et al. (2017a, 2018) and Sihag and Singh (2018) utilised various infiltration models (empirical model) in his study to calculate the infiltration rate of the soil for the given study area. Tiwari et al. (2017) used the generalised regression neural network, MLR, M5P model tree and SVM to predict the cumulative infiltration of soil and found that SVM works well than the other techniques. Various researchers have been used various soft computing techniques in hydraulics and environmental engineering applications (Sihag et al. 2017b, c, 2018a; Haghiabi et al. 2018; Nain et al. 2018a; Tiwari et al. 2018; Parsaie et al. 2017a, b; Shiri et al. 2016, 2017; Parsaie and Haghiabi 2015, 2017; Parsaie 2016; Azamathulla et al. 2016; Baba et al., 2013). These researchers found that these techniques work exceptionally well. Keeping it in the view, the focus of this investigation is on the prediction of the infiltration rate by using M5P tree, GP, MLR and SVM. Furthermore, the results were also compared with the empirical model (Kostiakov 1932) and sensitivity analysis was performed to find out the most important influencing parameter for predicting the infiltration rate of the soil.

Soft computing techniques

The soft computing technique is one of the most relevant and modern techniques used in the civil engineering problems (Sihag et al. 2018b, c; Nain et al. 2018b; Haghiabi et al. 2017; Kisi et al. 2017; Parsaie et al. 2017c; Kisi et al. 2015; Parsaie and Haghiabi 2014; Shiri and Kisi 2012). In this investigation, GP, SVM and M5P tree models were used. The description of the GP, SVM and M5P tree is given below.

Gaussian process (GP) regression

GP regression relies upon the postulation that nearby observation must share the information mutually and it is an approach for mentioning earlier straight over the function space. The simplification of Gaussian distribution is known as Gaussian regression. The matrix and vector of Gaussian distribution are expressed as covariance and mean in GP regression. Due to having earlier knowledge of function reliance and data, the validation for generalisation is not

essential. The GP regression models are capable of recognising the foresee distribution consequent to the input test data (Rasmussen and Williams 2006).

A GP is the collection of numbers of random variable, and any finite number of them has a collective multivariate Gaussian distribution. Assuming u and v stand for input and output domain accordingly, thereupon \times pairs (g_i, h_i) are drawn freely and equivalently distribution. For regression, it is assumed that $h \subseteq R_e$; then, a GP on p is expressed by the mean function $v_0: u R_e$ and covariance function $\mu: u \times u R_e$. Readers are requested to follow the Kuss (2006a, b) to get the exhaustive details of GP.

Support vector machine (SVM)

This method was first proposed by Vapnik (1998) and based on statistical learning theory. Main principle of SVM is optimal separation of classes. From the separable classes, SVM selects the one which have lowest generalisation error from infinite number of linear classifier or set upper limit to error which is generated by structural risk minimisation. In this way, the maximum margin between two classes can be found from the selected hyperplane and sum of distances of the hyperplane from the nearby point of two classes will set highest margin between two classes. Readers are requested to follow the Smola (1996) to get the exhaustive details of SVM. Cortes and Vapnik (1995) gave the idea of kernel function for nonlinear support vector regression.

M5P tree

M5P tree (Quinlan 1992) is a binary decision tree that uses linear regression function at the leaf (terminal node) which helps in predicting continuous numerical attributes. This method involves two stages for generation of model tree. First stage consists of splitting criteria to generate a decision tree. Splitting criteria for this method are based on treating the standard deviation of class value. Splitting process causes less standard deviation in child node as compared to parent node and thus considered as pure (Quinlan 1992). Out of all possible splits, M5P tree chooses the one that maximises the error reductions. This process of splitting the data may overgrow the tree which may cause over fitting. So, the next stage involves in removing over fitting using pruning method. It trims overgrown trees by substituting the subtrees with linear regression function. In this technique of tree generation, parameter space is split into surfaces and building a linear regression model in each of them.

M5P tree algorithm utilises standard deviation of the class value reaching at terminal node which measures the error value at that node and evaluates the expected reduction in error. Standard reduction is given as

$$\text{SDR} = \text{sd}(N) - \sum \frac{|N_i|}{|N|} \text{sd}(N_i) \quad (1)$$

where N depicts a set of examples that arrive at the node. N_i depicts i th outcome of subset of examples of potential set, and sd is the standard deviation.

Conventional models

In this investigation, two conventional models were used. The description of the conventional models was listed below.

Multi-linear regression (MLR)

The parameters for multi-linear regression analysis include $f(t)$ with T_f , I_t , C_i and W_c ; therefore, a following functional relationship may be initially assumed:

$$f(t) = k T_f^a \cdot I_t^b \cdot C_i^c \cdot W_c^d \quad (2)$$

where k is the proportionality constant.

$$\log f(t) = \log k + a \log T_f + b \log I_t + c \log C_i + d \log W_c \text{ (Taking log)} \quad (3)$$

There are four explanatory variables in the multi-linear equation. Now to develop a multi-linear model, $\log f(t)$ is taken as the output parameter and the four explanatory variables, namely $\log T_f$, $\log I_t$, $\log C_i$ and $\log W_c$, are taken as input parameters. The output of the multi-linear regression provided the values of k , a , b , c and d and, in turn, the developed equation of the form (3). The developed multi-linear regression equation is as follows:

$$f(t) = 104 \left(\frac{I_t^{1.310} \cdot C_i^{0.007}}{T_f^{0.66} \cdot W_c^{0.270}} \right) \quad (4)$$

where I_t is 1 for ash and 2 for organic manure.

Kostiakov model

The details of the Kostiakov model (Kostiakov 1932) are as follows:

$$f(t) = a T_f^{-b} \quad (5)$$

where a and b are constants.

After solving Eq. (5) with the measure infiltration rate with time, it will become in the form of Eq. (6).

$$f(t) = 114.6 T_f^{-0.68} \quad (6)$$

Materials and methodology

In this investigation, two double-ring infiltrometers were used to calculate the infiltration rate of soil. These consist of two rings, i.e. inner ring and outer ring, with diameter 300 mm and 450 mm, respectively, as shown in Fig. 1. The instrument was driven 100 mm into the soil out of 300 mm which is the total depth of the instrument and it was done with the fallen weight type hammer strike uniformly without disturbing the top layer of the soil. Both the rings were filled with equal depth of water and note down the initial depth of water in inner ring because the water from the inner ring went downwards directly not laterally. The moisture content of soil was also calculated before each experiment by using gravitational method.

The experimentations were done in the Hydraulics Laboratory, NIT Kurukshetra, India. The soil of NIT Kurukshetra is loam soil, and elevation from the sea level is 274 m. The climate of the Kurukshetra is cold in winter and dry in the summer except the monsoon season (normal annual rainfall



Fig. 1 Double-ring infiltrometer

582 mm). Infiltration rate was calculated with water quality which was the mixture of the water and different concentrations of impurities, i.e. 1%, 5%, 10% and 15%, and different types of impurities, i.e. ash and organic manure, which are a by-product and generally present in the study area. Two double-ring infiltrometers were driven into the soil paralleled to each other: One is filled with fixed concentration of ash and other with organic manure. Furthermore, infiltration rate was measured up to a fixed time interval which is 180 min because after 180 min, infiltration rate attains the steady infiltration rate (Sihag et al. 2017). The details of the experimental procedure along with the range of the infiltration rate are summarised in Table 1, and plot for the infiltration rate versus time with ash and organic manure is depicted in Fig. 2. As indicated in Table 1 and Fig. 2, infiltration rate is inversely proportional to the time. Initial infiltration rate of the water with impurities ash was higher than water with impurities organic impurities, but final infiltration rate of the soil of water having organic manure was higher than the ash. In case of the organic manure, infiltration rate increases with time when time reached to 90 min; then, it decreased with time (Singh 2015).

Data set

The experiments for the measured infiltration rate were performed in between January 2015 and May 2015 in Hydraulics Laboratory, Civil Engineering Department (NIT Kurukshetra). The geographical co-ordinates of the study area are 29.9490° N and 76.8173° E. The soil present in the campus is poorly permeable which has the low tendency of the infiltration rate. Totally, 132 observations were obtained from the field experiments out of which 92 observations were used for training and residual 40 for testing the models. Cumulative time (min), type of impurities (organic manure/ash),

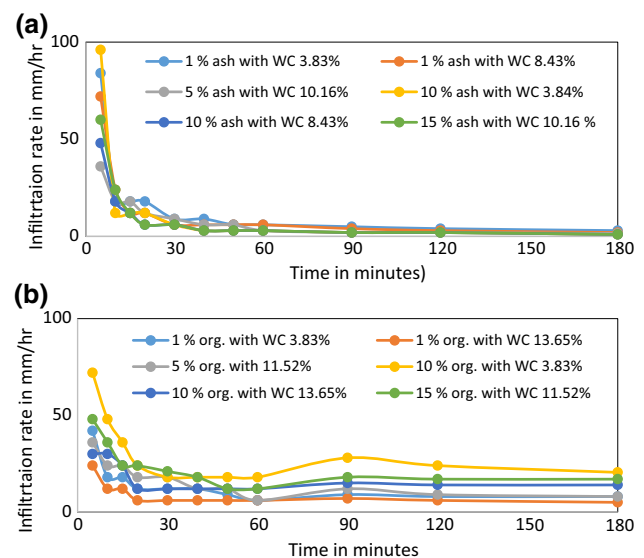


Fig. 2 Result analysis of the infiltration rate with different water qualities: **a** ash and **b** organic manure

concentration of impurities (%) and moisture content (%) were the input variables, whereas infiltration rate (mm/h) was output. The features and correlation matrix of the data set are given in Tables 2 and 3.

Detail of kernel functions

The SVM- and GP-based regression approaches design includes the scheme of kernel function. There are several kernel functions in GP and SVM. In this study, two kernel functions were used with GP and SVM technique.

1. Radial basis kernel (RBF) = $e^{-\gamma|a-b|^2}$

Table 1 Details of the experimental procedure along with the range of infiltration rate

Time (min.)	Type of impurities	Concentration of impurity (%)	Water content (%)	Range of the infiltration rate (mm/h)
5	Organic manure, ash	1, 5, 10, 15	3.83, 8.43, 10.16, 11.51, 13.65	24–84
10	Organic manure, ash	1, 5, 10, 15	3.83, 8.43, 10.16, 11.51, 13.66	12–48
15	Organic manure, ash	1, 5, 10, 15	3.83, 8.43, 10.16, 11.51, 13.67	12–36
20	Organic manure, ash	1, 5, 10, 15	3.83, 8.43, 10.16, 11.51, 13.68	6–24
30	Organic manure, ash	1, 5, 10, 15	3.83, 8.43, 10.16, 11.51, 13.69	6–21
40	Organic manure, ash	1, 5, 10, 15	3.83, 8.43, 10.16, 11.51, 13.70	3–18
50	Organic manure, ash	1, 5, 10, 15	3.83, 8.43, 10.16, 11.51, 13.71	3–18
60	Organic manure, ash	1, 5, 10, 15	3.83, 8.43, 10.16, 11.51, 13.72	3–18
90	Organic manure, ash	1, 5, 10, 15	3.83, 8.43, 10.16, 11.51, 13.73	2–28
120	Organic manure, ash	1, 5, 10, 15	3.83, 8.43, 10.16, 11.51, 13.74	2–24
180	Organic manure, ash	1, 5, 10, 25	3.83, 8.43, 10.16, 11.51, 13.75	2.21

Table 2 Features of the data used

Variables	Minimum	Maximum	Mean	SD	Kurtosis	Skewness
<i>Training data set</i>						
T_f (min.)	5	180	57.6630	53.3414	0.2810	1.1566
C_i (%)	1	15	7.2174	5.2011	− 1.3709	0.1210
W_c (%)	3.83	13.65	8.7021	3.6694	− 1.3878	− 0.2521
$f(t)$ (mm/h)	1	84	15.1793	15.9358	6.1863	2.3142
<i>Testing data set</i>						
T_f (min.)	5	180	53.3750	48.6271	1.3901	1.3840
C_i (%)	1	15	6.5000	5.0637	− 1.3367	0.2401
W_c (%)	3.83	13.65	8.2635	3.8315	− 1.5351	− 0.0002
$f(t)$ (mm/h)	2	96	15.3875	16.7426	13.6287	3.2757
<i>All data set</i>						
T_f (min.)	5	180	56.3636	51.8111	0.4986	1.2096
C_i (%)	1	15	7.0000	5.1512	− 1.3555	0.1572
W_c (%)	3.83	13.65	8.5692	3.7101	− 1.4396	− 0.1741
$f(t)$ (mm/h)	1	96	15.2424	16.1204	8.2289	2.5971

Table 3 Correlation matrix of input data set

Variables	T_f (min.)	C_i (%)	W_c (%)	$f(t)$ (mm/h)
T_f (min.)	1.0000	0.0000	0.0000	− 0.4322
C_i (%)	0.0000	1.0000	0.1796	0.1077
W_c (%)	0.0000	0.1796	1.0000	− 0.1033
$f(t)$ (mm/h)	− 0.4322	0.1077	− 0.1033	1.0000

$$2. \text{ Pearson VII kernel function (PUK)} = \left(1 / \left[1 + \left(2\sqrt{\|a - b\|^2} \sqrt{2^{(1/\omega)} - 1} / \sigma \right)^2 \right]^\omega \right)$$

where γ , σ and ω are kernel parameters. It is well known that GP and SVM estimation performance depends on a good setting of meta-parameters, parameters Gaussian noise, C , γ , σ and ω . The selections of Gaussian noise, C , γ , σ and ω control the prediction (regression) model complexity. In this study, a physical method was used to select primary parameters (i.e. C , γ , σ , ω and Gaussian noise). In order to minimise the RMSE and to maximise the CC, suitable values of various primary parameters are selected. The same kernel-specific parameters were taken for GP regression as well as for SVM. Table 4 enlists all the optimal values of the primary parameters for GP, SVM and M5P tree model.

Statistical performance evaluation criteria

Correlation coefficient (CC) and root-mean-square error (RMSE) values were calculated to investigate the performance of GP, SVM and M5P tree modelling approaches.

Table 4 Primary parameters using GP, SVM and M5P tree

Approaches	Primary parameters
M5P tree	$m = 4$
GP with RBF	Gaussian noise = 0.80, $\gamma = 3.5$
GP with PUK	Gaussian noise = 0.80, $\omega = 0.02$, $\sigma = 0.5$
SVM with RBF	$C = 2$, $\gamma = 3.5$
SVM with PUK	$C = 2$, $\omega = 0.02$, $\sigma = 0.5$

Coefficient of correlation (CC)

The coefficient of correlation (CC) is computed as

$$CC = \frac{m \sum_{i=1}^m o_i t_i - \left(\sum_{i=1}^m o_i \right) \left(\sum_{i=1}^m t_i \right)}{\sqrt{m \left(\sum_{i=1}^m o_i^2 \right) - \left(\sum_{i=1}^m o_i \right)^2} \sqrt{m \left(\sum_{i=1}^m t_i^2 \right) - \left(\sum_{i=1}^m t_i \right)^2}} \quad (7)$$

Root-mean-square error (RMSE)

The root-mean-square error (RMSE) is computed as:

$$RMSE = \sqrt{\frac{1}{m} \left(\sum_{i=1}^m (o_i - t_i) \right)^2} \quad (8)$$

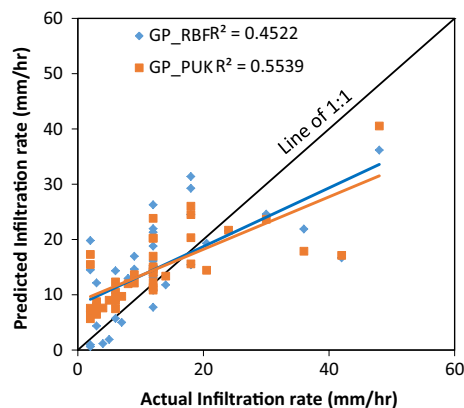
where o_i is the calculated values of infiltration rate, t_i is the estimated values of infiltration rate and m is the number of observations.

Results and discussion

This section of this investigation focuses on predicting performance of the proposed three soft computing techniques, i.e. GP, SVM and M5P tree, and two empirical

Table 5 Results of the different modelling approaches and empirical models for training and testing data set

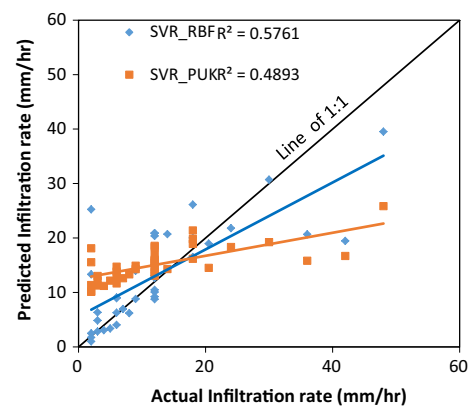
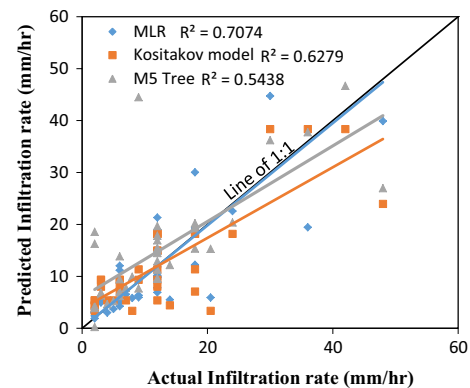
Approaches	Training data set		Testing data set	
	CC	RMSE	CC	RMSE
GP_RBF	0.7753	10.4193	0.4374	14.9329
GP_PUK	0.9937	8.3337	0.4115	15.8927
SVM_RBF	0.8162	9.8525	0.5278	14.1891
SVM_PUK	1.0000	0.0933	0.4289	15.4325
M5P	0.9072	7.7038	0.8490	9.4356
MLR	0.6262	12.6476	0.4405	15.9657
Kostiakov model	0.7112	11.3162	0.4608	15.0521

**Fig. 3** Predicted infiltration rate of soil using GP_RBF and GP_PUK

models, i.e. MLR and Kostiakov model. The ability of these soft computing models is depending upon the primary parameters, and the values of primary parameters are listed in Table 4. In this study, input variables were T_p , I_p , C_i and W_c and output was $f(t)$. The results of these soft computing techniques with empirical model are given in Table 5.

Figure 3 gives the scattered details of actual and predicted values of the infiltration rate of the soil by GP regression with RBF and PUK kernel function. It is clear from Fig. 3 that both the kernels function failed to predict a good result for the infiltration rate of the soil. But in comparison with RBF and PUK kernel of GP regression, RBF kernel function works well with CC and RMSE 0.4374 and 14.9329 (refer Table 5), respectively.

The same data set was also used for SVM-based regression techniques. Figure 4 gives the scattered details of the infiltration rate of the soil by using SVM regression techniques with RBF and PUK kernel function. Same like GP regression techniques, SVM also failed to predict the good quality of result for the infiltration rate of the soil. But the results from the SVM techniques were little bit good from the GP regression

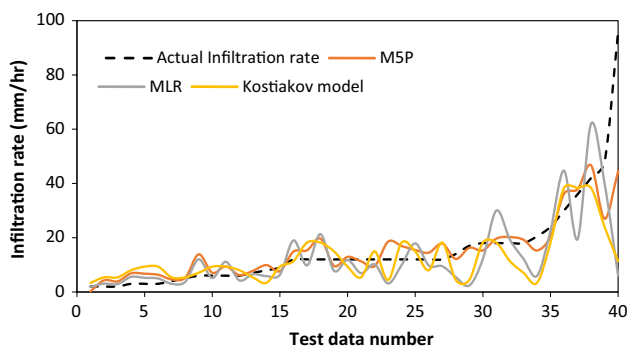
**Fig. 4** Predicted infiltration rate of soil using SVR_RBF and SVR_PUK**Fig. 5** Predicted infiltration rate of soil by using MLR, M5P tree and Kostiakov model

techniques. The values of CC and RMSE for RBF kernel with SVM were 0.5278 and 14.1891, respectively (refer Table 4).

The prediction of infiltration rate of the soil by M5P tree techniques, multi-linear regression and Kostiakov model was also performed by the same data set. Figure 5 gives the scattered details of the infiltration rate of the soil by using MLR, Kostiakov model and M5P tree. It is also clear from Fig. 5 that all the scatters from the M5P tree model are nearby to agreement line than the other two models. Also, the value of CC is much high (0.8490) and RMSE is much less (9.4356) than the other two techniques. In comparison with MLR and Kostiakov model, the prediction of the MLR and Kostiakov model was almost same with CC (0.4405 and 0.4806) and RMSE (15.9657 and 15.0521), respectively.

Table 6 Statistical information of the infiltration rate with different soft computing techniques and empirical models

Approaches	Minimum	Maximum	Mean	SD	Kurtosis	Skewness
<i>Training data set</i>						
Actual	1.00	84.00	15.1793	15.9358	6.1863	2.3142
GP_RBF	0.77	37.17	15.0424	9.4499	−0.7493	0.3505
GP_PUK	7.20	45.74	15.1815	7.6481	4.3109	1.9056
SVM_RBF	0.91	43.24	13.3878	9.8245	0.1587	0.8043
SVM_PUK	1.02	83.94	15.2025	15.8756	6.2540	2.3261
M5P tree	0.50	46.69	14.5036	10.6377	0.9470	1.1090
MLR	1.83	63.04	12.8963	11.4244	4.5405	1.9507
Kostiakov model	3.35	38.36	13.4477	10.3690	0.7651	1.2985
<i>Testing data set</i>						
Actual	2	96	15.3875	16.7426	13.6287	3.2757
GP_RBF	−1.61	36.17	14.7312	8.5849	−0.0222	0.2881
GP_PUK	12.36	20.75	15.1809	1.7746	1.1680	0.8328
SVM_RBF	1.00	39.52	13.3489	8.7057	0.6250	0.7803
SVM_PUK	10.11	25.85	15.1678	3.2476	1.6642	0.9674
M5P tree	0.25	46.69	15.1861	10.6086	2.2856	1.5043
MLR	1.90	62.03	12.5197	12.5803	6.1334	2.3440
Kostiakov model	3.35	38.36	12.3710	9.2193	2.7331	1.6930

**Fig. 6** Variation of infiltration rate using the different regression approaches

Comparison of the results

A comparison of all the techniques and models was done to find out the most efficient technique in prediction of the infiltration rate of the soil. The performance of the M5P tree is good with performance evaluation parameters ($CC = 0.8490$ and $RMSE = 9.4356$ mm/h) than the other model and techniques, while among GP and SVM, SVM with RBF kernel outperforms than other kernel function with values of CC and $RMSE$ 0.5278 and 14.1891 mm/h, respectively. Table 6 gives the statistical information of actual and predicted values of the infiltration rate with different soft computing techniques and empirical models.

Figure 6 provides a plot of MLR, M5P tree and Kostiakov model with actual infiltration rate in increasing order

Table 7 Sensitivity analysis using M5P tree model

Combinations of the variables	Parameter removed	M5P tree model	
		CC	RMSE (mm/h)
T_f, I_t, C_i, W_c		0.9072	7.7038
I_t, C_i, W_c	T_f	0.1971	15.5368
T_f, C_i, W_c	I_t	0.8557	9.0473
T_f, I_t, W_c	C_i	0.8822	8.2711
T_f, I_t, C_i	W_c	0.8418	9.0481

of the values. This figure suggests that the predicted values using M5P model follows same path which is followed by actual values. But when the actual values of infiltration rate were very high, prediction from all the techniques gave the large error because there were large fluctuations in the infiltration rate at starting point. Hence, it is clear from Table 4 and Fig. 6 that M5P tree was the best technique which can be predicted the values of the infiltration rate in the absence of the infiltration data under the same conditions.

Sensitivity analysis (SA)

SA is the test in which we find the most important input parameter or parameters which affect the infiltration rate of the soil most. In this investigation, sensitivity analysis was done by removing the one parameter one by one in each case. M5P tree model was used to carry out the sensitivity analysis by using same primary parameters. Table 7 summarises the

results of the sensitivity analysis. Outcomes from Table 7 suggest that cumulative time is the most important parameter to predict the infiltration rate of the soil for this data set.

Conclusions

Knowledge of infiltration process is essential for agriculture, hydrologic study, watershed management, irrigation system design and drainage design. In this investigation, three soft computing techniques (SVR, GP and M5P tree) and two empirical models (MLR and Kostiakov model) were used to estimate the infiltration rate of the soil with different water qualities. The obtained results concluded that the M5P tree model is the most efficient model to predict the infiltration rate of the soil with different water qualities than the SVR, GP, MLR and Kostiakov model, whereas the results of SVM were more suitable as compared to the GP and MLR and also gave better prediction than Kostiakov model. Thus, M5P tree model was the most suitable model for predicting the infiltration rate of the soil. Finally, SA suggests that cumulative time is an essential parameter which affects the infiltration rate of the soil with different water qualities using M5P model tree for this data set.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Azamathulla HM, Haghiabi AH, Parsaie A (2016) Prediction of side weir discharge coefficient by support vector machine technique. *Water Sci Technol Water Supply* 16(4):1002–1016
- Baba APA, Shiri J, Kisi O, Fard AF, Kim S, Amini R (2013) Estimating daily reference evapotranspiration using available and estimated climatic data by adaptive neuro-fuzzy inference system (ANFIS) and artificial neural network (ANN). *Hydrol Res* 44(1):131–146
- Bhave S, Sreeja P (2013) Influence of initial soil condition on infiltration characteristics determined using a disk infiltrometer. *ISH J Hydraul Eng* 19(3):291–296
- Brooks RH, Corey AT (1964) Hydraulic properties of porous media and their relation to drainage design. *Trans ASAE* 7(1):26–0028
- Chow VT, Maidment DR, Mays LW (1988) *Applied hydrology*. McGraw-Hill, New York
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
- Diamond J, Shanley T (2003) Infiltration rate assessment of some major soils. *Ir Geogr* 36(1):32–46
- Gupta BL, Gupta AMIT (2008) *Water resources systems and management*, 2nd edn. Standard Publishers Distributors, Delhi, pp 510–535
- Haghiabi AH, Parsaie A, Ememgholizadeh S (2017) Prediction of discharge coefficient of triangular labyrinth weirs using Adaptive Neuro Fuzzy inference system. *Alexandria Eng J* 57:1773–1782
- Haghiabi AH, Nasrolahi AH, Parsaie A (2018) Water quality prediction using machine learning methods. *Water Qual Res J* 53(1):3–13
- Haghighi F, Gorji M, Shorafa M, Sarmadian F, Mohammadi MH (2010) Evaluation of some infiltration models and hydraulic parameters. *Span J Agric Res* 8(1):210–217
- Igbadun HE, Idris UD (2007) Performance evaluation of infiltration models in a hydromorphic soil. *Niger J Soil Environ Res* 7(1):53–59
- Kisi O, Shiri J, Karimi S, Shamshirband S, Motamedi S, Petkovic D, Hashim R (2015) A survey of water level fluctuation predicting in Urmia Lake using support vector machine with firefly algorithm. *Appl Math Comput* 270:731–743
- Kisi O, Keshavarzi A, Shiri J, Zounemat-Kermani M, Omran EE (2017) Groundwater quality modeling using neuro-particle swarm optimization and neuro-differential evolution techniques. *Hydrol Res* 48(6):1508–1519
- Kostiakov AN (1932) On the dynamics of the coefficient of water percolation in soils and the necessity of studying it from the dynamic point of view for the purposes of amelioration. *Trans Sixth Comm Int Soc Soil Sci* 1:7–21
- Kuss M (2006) Gaussian process models for robust regression, classification, and reinforcement learning. Doctoral dissertation, Ph.D. thesis, Technische Universität, Darmstadt
- Kuss M (2006) Gaussian process models for robust regression, classification, and reinforcement learning. Doctoral dissertation, Technische Universität
- Micheal AM (1978) *Irrigation, theory and practice*. Vikas Press Private Limited, New Delhi
- Nain SS, Garg D, Kumar S (2018a) Investigation for obtaining the optimal solution for improving the performance of WEDM of super alloy Udimet-L605 using particle swarm optimization. *Eng Sci Technol Int J* 21(2):261–273
- Nain SS, Sihag P, Luthra S (2018b) Performance evaluation of fuzzy-logic and BP-ANN methods for WEDM of aeronautics super alloy. *MethodsX* 5(1):890–908. <https://doi.org/10.1016/j.mex.2018.04.006>
- Parsaie A (2016) Predictive modeling the side weir discharge coefficient using neural network. *Model Earth Syst Environ* 2(2):63
- Parsaie A, Haghiabi A (2014) Predicting the side weir discharge coefficient using the optimized neural network by genetic algorithm. *Sci J Pure Appl Sci* 3(3):103–112
- Parsaie A, Haghiabi A (2015) The effect of predicting discharge coefficient by neural network on increasing the numerical modeling accuracy of flow over side weir. *Water Resour Manag* 29(4):973–985
- Parsaie A, Haghiabi AH (2017) Improving modelling of discharge coefficient of triangular labyrinth lateral weirs using SVM, GMDH and MARS techniques. *Irrig Drain* 66(4):636–654
- Parsaie A, Azamathulla HM, Haghiabi AH (2017a) Prediction of discharge coefficient of cylindrical weir-gate using GMDH-PSO. *ISH J Hydraulic Eng* 24:116–123
- Parsaie A, Najafian S, Omid MH, Yonesi H (2017b) Stage discharge prediction in heterogeneous compound open channel roughness. *ISH J Hydraulic Eng* 23(1):49–56
- Parsaie A, Yonesi H, Najafian S (2017c) Prediction of flow discharge in compound open channels using adaptive neuro fuzzy inference system method. *Flow Meas Instrum* 54:288–297
- Quinlan JR (1992, November) Learning with continuous classes. In: 5th Australian joint conference on artificial intelligence, vol 92. pp 343–348
- Rasmussen CE, Williams CK (2006) *Gaussian processes for machine learning*, vol 1. MIT Press, Cambridge, p 248

- Scotter DR, Clothier BE, Harper ER (1982) Measuring saturated hydraulic conductivity and sorptivity using twin rings. *Soil Res* 20(4):295–304
- Shiri J, Kisi O (2012) Estimation of daily suspended sediment load by using wavelet conjunction models. *ASCE J Hydrol Eng* 17(9):986–1000
- Shiri J, Shamshirband S, Kisi O, Karimi S, Bateni SM, HosseiniNazhad SH, Hashemi A (2016) Prediction of water-level in the Urmia lake using the extreme learning machine approach. *Water Resour Manag* 30:5217–5229
- Shiri J, Keshavarzi A, Kisi O, Karimi S (2017) Using soil easily measured parameters for estimating soil water capacity: soft computing approaches. *Comput Electron Agric* 141:327–339
- Sihag P (2018) Prediction of unsaturated hydraulic conductivity using fuzzy logic and artificial neural network. *Model Earth Syst Environ* 4:189–198
- Sihag P, Singh B (2018) Field evaluation of infiltration models. *Tech-nogenic Ecol Saf* 4(2/2018):3–12
- Sihag P, Tiwari NK, Ranjan S (2017a) Estimation and inter-comparison of infiltration models. *Water Sci* 31(1):34–43
- Sihag P, Tiwari NK, Ranjan S (2017b) Modelling of infiltration of sandy soil using gaussian process regression. *Model Earth Syst Environ* 3(3):1091–1100
- Sihag P, Tiwari NK, Ranjan S (2017c) Prediction of unsaturated hydraulic conductivity using adaptive neuro-fuzzy inference system (ANFIS). *ISH J Hydraul Eng*. <https://doi.org/10.1080/09715010.2017.1381861>
- Sihag P, Jain P, Kumar M (2018a) Modelling of impact of water quality on recharging rate of storm water filter system using various kernel function based regression. *Model Earth Syst Environ* 4:61–68
- Sihag P, Singh B, Vand AS, Mehdipour V (2018b) Modeling the infiltration process with soft computing techniques. *ISH J Hydraul Eng*. <https://doi.org/10.1080/09715010.2018.1464408>
- Sihag P, Tiwari NK, Ranjan S (2018c) Support vector regression-based modeling of cumulative infiltration of sandy soil. *ISH J Hydraul Eng*. <https://doi.org/10.1080/09715010.2018.1439776>
- Singh B (2015) Impact of water quality on infiltration rate of soil. M.Tech. dissertation, National Institute of Technology Kurukshetra
- Singh B, Sihag P, Singh D (2014) Study of infiltration characteristics of locally soils. *J Civ Eng Environ Technol* 1:9–13
- Singh B, Sihag P, Singh K (2017) Modelling of impact of water quality on infiltration rate of soil by random forest regression. *Model Earth Syst Environ* 3(3):999–1004
- Singh B, Sihag P, Singh K (2018) Comparison of infiltration models in NIT Kurukshetra campus. *Appl Water Sci* 8(2):63
- Smith B (2006) *The farming handbook*. University of Natal Press, Pietermaritzburg
- Smola AJ (1996) Regression estimation with support vector learning machines. Doctoral dissertation, Master's thesis, Technische Universität München
- Tiwari NK, Sihag P, Ranjan S (2017) Modeling of infiltration of soil using adaptive neuro-fuzzy inference system (ANFIS). *J Eng Technol Educ* 11(1):13–21
- Tiwari NK, Sihag P, Kumar S, Ranjan S (2018) Prediction of trapping efficiency of vortex tube ejector. *ISH J Hydraul Eng*. <https://doi.org/10.1080/09715010.2018.1441752>
- Uloma AR, Samuel AC, Kingsley IK (2014) Estimation of Kostia-kov's infiltration model parameters of some sandy loam soils of Ikwuano-Umuahia, Nigeria. *Open Trans Geosci* 1(1):34–38
- Vapnik V (1998) *Statistical learning theory*. Wiley, New York

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