Water Supply

Water Supply Vol 22 No 3, 2847 doi: 10.2166/ws.2021.430

An expert system for predicting the infiltration characteristics

Balraj Singh[a,*], Isa Ebtehaj[b], Parveen Sihag[c] and Hossein Bonakdari[b]

[a] Department of Civil Engineering, Panipat Institute of Engineering and Technology, Panipat 132102, India
[b] Department of Soils and Agri-Food Engineering, Laval University, Québec, QC, G1V0A6, Canada
[c] Department of Civil Engineering, Chandigarh University, Chandigarh, India
*Corresponding author. E-mail: balrajzinder@gmail.com

IE, 0000-0002-6906-629X; HB, 0000-0001-6169-3654

ABSTRACT

Infiltration plays a fundamental role in streamflow, groundwater recharge, subsurface flow, and surface and subsurface water quality and quantity. This study includes a comparative analysis of the two machine learning techniques, M5P model tree (M5P) and Gene Expression Programming (GEP), in predictions of the infiltration characteristics. The models were trained and tested using the 7 combination (CMB1 – CMB7) of input parameters; moisture content (m), bulk density of soil (D), percentage of silt (SI), sand (SA) and clay (C), and time (t), with output parameters; cumulative infiltration (CI) and infiltration rate (IR). Results suggested that GEP has an edge over M5P to predict the IR and CI with R, RMSE and MAE values 0.9343, 15.9667 mm/hr & 8.7676 mm/hr, and 0.9586, 9.2522 mm and 7.7865 mm for IR and CI, respectively with CMB1. Although the M5P model also gave good results with R, RMSE and MAE values 0.9192, 14.1821 mm/hr, and 19.2497 mm/hr, and 0.8987, 11.2144 mm and 18.4328 mm for IR and CI, respectively, but lower than GEP. Furthermore, single-factor ANOVA and uncertainty analysis were used to show the significance of the predicted results and to find the most efficient soft computing techniques respectively.

Key words: cumulative infiltration, gene expression programming, infiltration rate, M5P model tree

HIGHLIGHTS

- Infiltration characteristics is predicted using two soft computing techniques; M5P and GEP.
- Linear relationship model is generated from M5P and GEP.
- Single-factor ANOVA and uncertainty analysis is done for the best selected models.

ABBREVIATIONS

CI	Cumulative infiltration
IR	Infiltration rate
M5P	M5P model tree
GEP	Gene Expression Programming
M	Moisture content of soil
D	Bulk density of soil
SI	Silt
SA	Sand
C	Clay
t	Time of infiltration
ANN	Artificial neural network
RF	Random forest
GA	Genetic algorithms
ETs	Expression trees
DRI	Double-ring infiltrometer
LLR	Local Linear Regression
DLLR	Dynamic Local Linear Regression
ANOVA	Analysis of variance
MAE	Mean absolute value

RMSE	Root mean square error
R	Coefficient of correlation
UA	Uncertainty analysis
ANFIS	Adaptive network based fuzzy inference system
GMDH	Group method of data handling
MARS	Multivariate adaptive regression splines
SVR	Support Vector Regression
FFA	Firefly algorithm
PSO	Particle swarm optimization
MLR	Multiple Linear Regression
GRG	Generalized Reduced Gradient
MGGP	Multigene Genetic Programming
Z	A set of examples that reaches the nodes
Z_i	The subset of examples that is the ith output of the potential set
SD	Standard deviation
ETs	Expression trees
k_i	Estimated value
j_i	Actual value
p	Number of observations
CFD	Computational fluid dynamics

INTRODUCTION

Water and soil have a vibrant relationship (Patle *et al.* 2018). A good water management system requires an efficient control of the infiltration characteristics of the soil (Singh *et al.* 2018a). The infiltration characteristics consist of two terms, infiltration rate and cumulative infiltration. The cumulative infiltration refers to the total amount of water that infiltrates into the soil and the infiltration rate is the rate by which it infiltrates into the soil (Haghighi *et al.* 2010). Good knowledge of the infiltration characteristics would help in a wide range of problems such as artificial and natural groundwater recharge, flooding, pollution of underground water, the optimum amount of water for irrigation, and runoff water (Dahan *et al.* 2007). It is also the most dominant factor in the accurate prediction of the flooding conditions in any catchment (Bhave & Sreeja 2013). An irrigation scheme should be planned by considering the lateral flow of water, for the efficient and effective utilization of irrigation water (Chowdary *et al.* 2006). Infiltration characteristics also help us in measuring irrigation efficiency, growth in crop yields, and minimizing soil erosion (Adeniji *et al.* 2013). There are many factors such as soil properties and texture, water content, humidity, rainfall intensity, and field density that affect the infiltration characteristics. Soil properties and texture affect the water-holding capacity of the soil (Gupta & Gupta 2008). Sand contains a comparatively larger pore size than clay and thus has a high infiltration rate and very low water-holding capacity (Micheal 1978; Smith 2006). Infiltration characteristics also play a significant role in the prediction of runoff in designing hydraulic structures as well as water resources planning and management (Heinz *et al.* 2007; Souchère *et al.* 2010).

Many researchers have used different types of infiltration models such as the Philip model (Philip 1957), Green and Ampt (Green & Ampt 1911), Holton model (Holtan 1961), Singh-Yu model (Singh & Yu 1990), Kostiakov model (Kostiakov 1932), Huggins–Monke model (Huggins & Monke 1966), Novel model (Sihag *et al.* 2017), Horton model (Horton 1938) and revised modified Kostiakov model (Parhi *et al.* 2007) with a different type of the soil. Mishra *et al.* (2003) compared 14 infiltration models on 243 datasets with different types of soil and found Singh-Yu gave the precise result as well Huggins–Monke, Holton, and Horton models, which also gave good results but comparatively lower than the Singh-Yu model. Mirzaee *et al.* (2014) introduced a new model that is modified Kostiakov which works well with silty clay, clay loam, and loam soil in Iran. Sihag *et al.* (2017) also worked in the infiltration field and introduced a new model (Novel model) in the soil of India (Kurukshetra). Zakwan (2018) used various infiltration models on 16 datasets (worldwide), in which the Holton model was the best suitable model and the Novel model by Sihag *et al.* (2017) worked well in 5 cases.

In the 21st century, machine learning techniques such as GEP, ANN, M5P, and RF, became one of the most dominant and used techniques in the field of water resources, structure, hydrology, and environmental engineering (Mehdipour & Memarianfard 2017; Mehdipour *et al.* 2017; Parsaie & Haghiabi 2017; Angelaki *et al.* 2018; Haghiabi *et al.* 2018; Sihag *et al.* 2018a, 2018b, 2020; Singh *et al.* 2018b, 2019a, 2019b; Vand *et al.* 2018; Mohanty *et al.* 2019; Kumar *et al.* 2020; Malik *et al.* 2020; Pandey *et al.* 2020; Pandhiani *et al.* 2020; Mohammed *et al.* 2021; Rehamnia *et al.* 2021). CFD is also a method that can be used in water resources problems (Al-Obaidi 2019; Al-Obaidi & Mohammed 2019; Al-Obaidi 2021), but machine learning

techniques inside traditional fluid simulations can improve both accuracy and speed (Kochkov *et al.* 2021), Karbasi & Azamathulla (2016) used the GEP technique in the prediction of characteristics of a hydraulic jump over a rough bed. The performance of the GEP is slightly better than other modeling techniques (ANN and SVM). Shabanlou *et al.* (2018) employed GEP to determined scour dimensions around submerged vanes and found that it estimated the scour dimensions precisely. Gholami *et al.* (2018) applied GEP to determine the stable threshold channel slope. Khozani *et al.* (2018) compared the GEP and ANN, in which GEP gave the superior result for predicting the shear stress distribution in circular open channels.

As compared to GEP, the M5P model tree is not very popular among the machine learning techniques but successfully used in civil engineering problems (Bhattacharya & Solomatine 2005; Pal & Deswal 2009; Singh *et al.* 2010; Goyal & Ojha 2011). Ayoubloo *et al.* (2011) compared two machine learning techniques, the M5P model tree and ANN, to calculate local scour downstream of spillways and found M5P predicts the result efficiently. Singh *et al.* (2017) also employed the M5P model tree, ANN, and RF to predict the infiltration rate of the soil. Goyal & Ojha (2011) found M5P model tree gave more accurate results than ANN for the topic of the local scour downstream of a ski jump bucket. The main advantage of the GEP and M5P model tree is that they give linear models to predict the output at any instance. The present investigation is originated by setting the following goals (i) to develop M5P and GEP models for prediction of infiltration characteristics, infiltration rate as well as cumulative infiltration (ii) to compare the results of M5P and GEP models to find the most accurate values of infiltration characteristics (iii) to generate the linear relationship equation for calculating the infiltration characteristics by M5P and GEP. By achieving these objectives, two machine learning-based models were developed and their results were compared by using performance evaluation parameters and novel linear relationship equations were generated to calculate the values of IR and CI at any instance.

MATERIALS AND METHODS

M5P

M5P, initially introduced by Quinlan (1992), is used to develop a decision tree by engaging the linear regression function method at nodes to construct a model that develops a relationship among input and output variables. The splitting approach is applied at each node instead to gain the maximum information to minimize the variation in the intra subset class value down to each branch. The splitting process will be converged when there are diminutive variations among the class values of the instances or only a few instances are left, or when the tree is pruned back. The developed tree indicates a very good structure and estimation precision due to showing more probable linearity at the leaf node (Singh *et al.* 2017). The equation to calculate the standard deviation reduction is as follows:

$$SDR = sd(Z) - \sum \frac{|Z_i|}{|Z|} sd(Z_i) \quad (1)$$

The main advantage of the M5P model is that it can deal with a continuous and categorical variable and is also able to handle the missing values of variables. Figure 1 provided the concept of the M5P in which M_i are the models and n_i are the split nodes of the tree.

GEP

GEP proposed by Ferreira (2002) is a search technique that involves computer programs. It is a developed method with the base of GA and has been widely implemented in current studies. The computer programs of GEP are all encoded in linear chromosomes, which are then articulated or translated into ETs. A concise flowchart of GEP is shown in Figure 2. The first step of this program to solve any problem is to produce the initial population, which happens with arbitrary births of chromosomes and in the latter, the chromosomes convert to ETs that are examined by performance criteria to represent the solubility of produced ETs. If the outcomes convince the performance criteria, population generating stops, and if the results are not satisfactory, the system regenerates with some improvement to make a new generation with improved value and this process occurs until the best results are achieved. These results are generated based on the primary parameters and details of the primary parameters are summarized in Table 1. The advantage of GEP is to compare the chromosomes of a symbolic and linear string of fixed length. For further explanation about GEP, readers are referred to Ferreira (2006) and Ebtehaj *et al.* (2015).

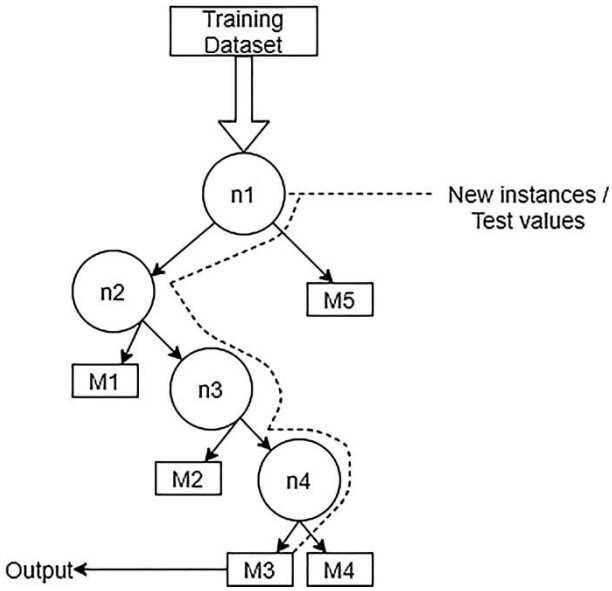

Figure 1 | A M5P model tree.

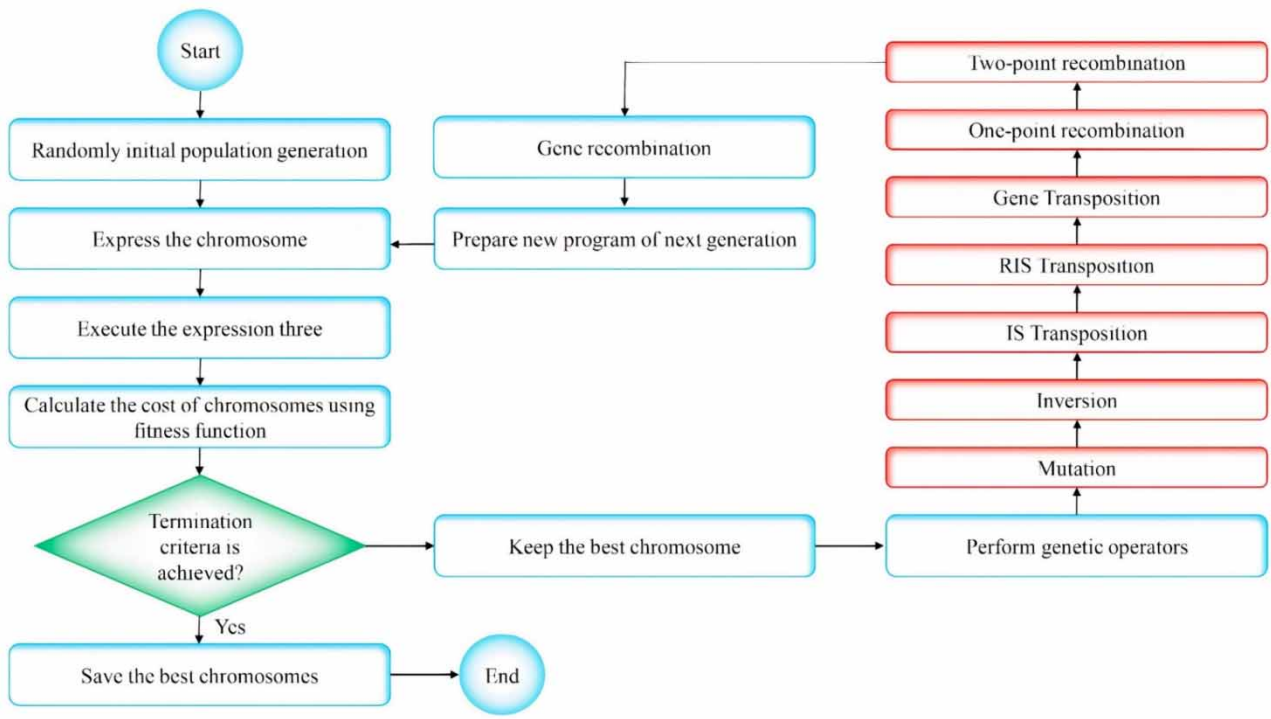

Figure 2 | Description of gene expression programming.

Study area and methodology

The infiltration characteristics data are experimentally determined in four districts; that is, Kurukshetra, Hisar, Kaithal, and Karnal of Haryana state, India. These four districts are situated in the north part of India, which is very far from the sea which results in a very low temperature (1 °C to 7 °C) in the winter and high temperature (40 °C to 45 °C) in the summer. The geographical representation of all the districts along with the location of the experimentation point is shown in Figure 3. There is a total of 17 locations on which experimentation was done (Figure 3). All the locations are located 29.14° to 29.96° north

Table 1 | Details of the primary parameters

Techniques	Parameters	Setting
M5P	M	4
GEP	Population size	60
	No. of chromosomes	50
	No. of generations	50,000
	No. of genes	3
	Linking function	Addition
	Function set	Pow, Exp, Not, +, −, ×
	Inversion rate	0.0450
	Mutation rate	0.0013
	IS transportation rate	0.20
	Gene transportation rate	0.30
	RIS transportation rate	0.20
	One point recombination rate	0.20
	Gene recombination rate	0.40

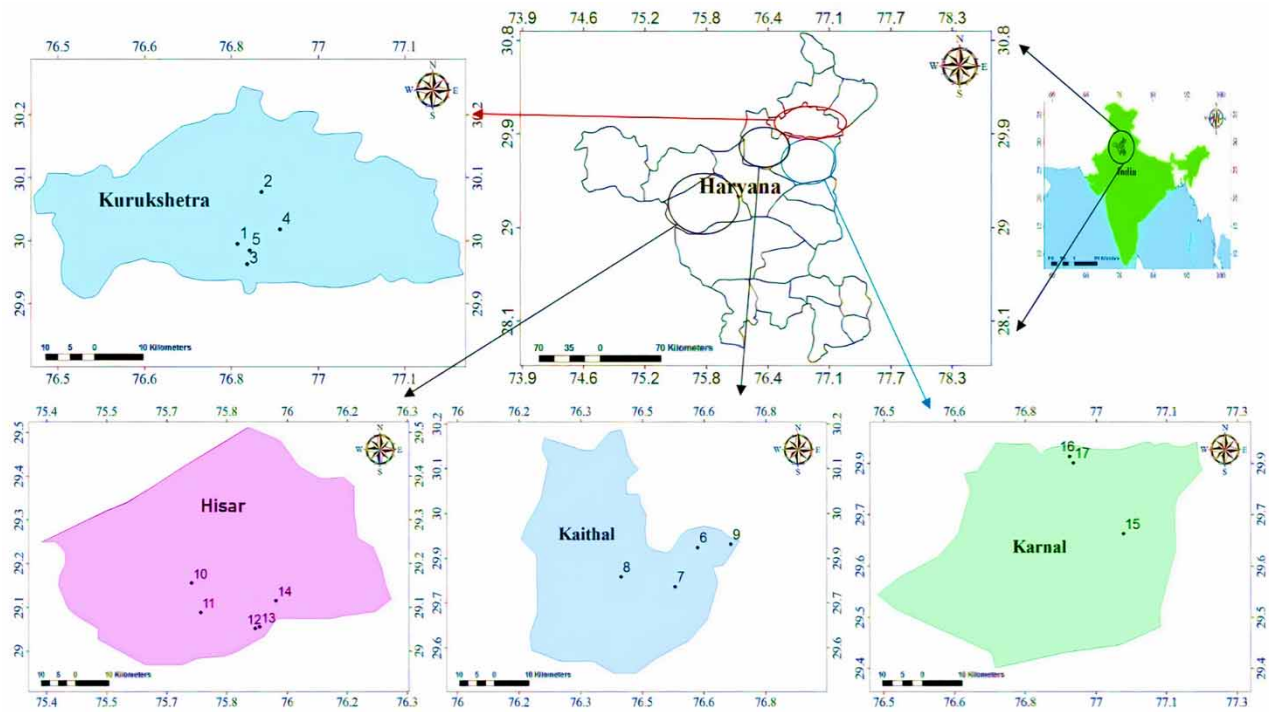

Figure 3 | Geographical representation of all the locations.

latitude and 75.72° to 76.99° east longitude. The rainfall occurs in this part of India due to two factors; monsoon (June to Aug.) and western disturbance (Dec. to Feb.) but approximately 95% of the total rainfall occurs due in the monsoon period. The soil present is loam, sandy loam, clay loam, and sandy in Kurukshetra, Kaithal, Karnal, and Hisar respectively.

The infiltration characteristics consist of the two properties, IR and CI of water. The IR and CI were determined by using DRI (ASTM 2009). The DRI is the standard instrument for measuring the infiltration characteristics of the soil. The DRI is consists of two cylinders, the outer (diameter 600 mm) and the inner cylinder (diameter 300 mm), connected with iron strips as displayed in Figure 4. 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 striking uniformly without disturbing the top layer of the soil. Both the rings were filled with an equal depth of water and note down the initial depth of water in the inner ring because the water from the inner ring went downwards directly not laterally. The total amount of water that infiltrates through the process is

Figure 4 | DRI setup.

called CI and the rate of water or amount of water infiltrate per time is the IR. Simultaneously, the soil samples were also collected to find the properties of the soil such as percentages of SI, SA and C, D, and M. The percentage of SI, SA and C were calculated by the hydrometer test (ASTM 2007) and D and M were calculated by the Proctor test (Connely *et al.* 2008) and oven-dry method (Rowe 2018) respectively. The main reason for calculating these parameters is to know about the soil of the study area. The detailed description of these soil properties is tabulated in Table 2 for the locations of four districts.

RESULTS AND DISCUSSIONS

The structure of results for this investigation is as follows: firstly, analysis of the dataset which includes the descriptive characteristics of the dataset, the making of various combination models, and details of the performance evaluation parameters followed by the prediction of the IR and CI using M5P and GEP and performance comparison among each model. Finally, single-factor ANOVA and uncertainty analysis are done for the best fit models for both of the techniques.

Table 2 | Description of the soil properties

Sr. No.	Locations	SA (%)	C (%)	SI (%)	D (g/cm^3)	M (%)
1	Kurukshetra	47.5	25.2	27.3	1.63	12.74
2		50.71	23.19	26.10	1.66	19.85
3		39.84	52.34	7.82	1.58	19.74
4		42.85	24.00	33.15	1.58	18.39
5		48.74	29.38	21.88	1.66	8.77
6	Kaithal	59.58	30.72	9.70	1.60	14.21
7		26.63	41.82	31.55	1.54	18.55
8		46.7	16.56	36.74	1.64	5.28
9		24.81	39.62	35.57	1.236	19.06
10	Hisar	79.73	6.45	13.825	1.603	8.47
11		84.14	7.83	8.026	1.749	7.47
12		66.63	19.62	13.75	1.703	12.71
13		44.27	23.12	32.61	1.65	11.56
14		45.67	39.16	15.17	1.62	8.33
15	Karnal	29.12	66.14	7.74	1.61	18.60
16		19.73	62.41	17.86	1.61	15.37
17		32.71	54.21	13.08	1.75	7.64

Analysis of dataset

To check the effectiveness of the M5P and GEP model, M in percentage, D in kg/m^3, SI, SA and C in percentage and T in minutes were used as input parameters and on the other hand, output variables are CI and IR. The total dataset was divided into two sections; the training section (2/3 of the total dataset) and the testing section (1/3 of the total dataset). The study was carried out in two sets, the first set was the analysis of the IR and the second set was the analysis of the CI. The IR and CI dataset consists of 185 observations, which tend to result in 125 observations in the training section and 60 observations in the testing section. Details of the characteristics of the training and testing section data are summarized in Table 3.

A total of 7 models (CMB1-CMB7) were created by using the input parameters for IR as well as CI. CMB 1 contained all the parameters, while CMB 2 to CMB 7 was created by removing one input parameter in each combination. These combinations were created to examine the effects of each of the parameters in the prediction of infiltration characteristics. The details of the input model's combination are given in Table 4.

The performance evaluation parameters were used to compare the predicted values of the M5P and GEP. MAE (Malik *et al.* 2021a; Singh *et al.* 2021a, 2021b), RMSE (Aradhana *et al.* 2021; Bhoria *et al.* 2021; Malik *et al.* 2021b), and R (Arora *et al.* 2019; Malik *et al.* 2021c; Sepahvand *et al.* 2021) was used in this study to compare the effectiveness of the modeling techniques. The expressions of these parameters were explained in Equations (2)–(4). The range of MAE and RMSE lies in between 0 and ∞ (Malik *et al.* 2021b). The obtained results of performance evaluation parameters (MAE, RMSE and R) for IR and CI were tabulated in Figure 5 using the M5P and GEP techniques. Secondly, a single factor ANOVA was used in the comparison of the statistically predicted values. This method is the hypothesis method to find out the significant or insignificant differences among two or more modeling techniques. The single factor ANOVA technique gives three values; F-value, *p*-value, and F-critical. If the *p*-value is more than 0.05 and the F value is less than F critical then the result of that particular approach is insignificant and vice versa (Singh *et al.* 2017).

$$\mathrm{MAE} = \frac{1}{p}\sum_{i=1}^{p}|j_i - k_i| \tag{2}$$

Table 3 | Characteristics of the training section data and testing section data

	T	SA	C	SI	D	M	IR (mm/hr)	CI (mm)
Training section								
Mean	52.08	47.0804	32.4470	20.4725	1.6165	13.2013	57.4888	46.9876
Standard Error	4.7594	1.5933	1.5464	0.9199	0.0096	0.4411	5.3054	5.7798
Median	30	45.67	29.38	17.86	1.6220	12.7400	36	25
Mode	20	45.67	39.16	15.17	1.6220	8.3300	24	19
Standard Deviation	53.2120	17.8139	17.2901	10.2847	0.1077	4.9326	59.3162	64.6203
Sample Variance	2831.5258	317.3363	298.9491	105.7770	0.0116	24.3315	3518.4120	4175.7920
Kurtosis	0.7797	−0.2910	−0.7365	−1.4957	6.4497	−1.5157	1.7261	11.9799
Skewness	1.4171	0.6085	0.4158	0.2367	−2.2777	0.0233	1.5093	3.2818
Testing section								
Mean	46.5000	44.5281	33.9819	21.4898	1.6106	13.4749	47.2933	38.5441
Standard Error	6.0800	2.2802	2.2530	1.3172	0.0148	0.6451	5.9284	5.3716
Median	30	44.9700	29.3800	19.8700	1.6220	13.4750	30	25.2500
Mode	30	19.7300	62.4100	17.8600	1.6120	15.3700	18	7
Standard Deviation	47.0961	17.6630	17.4517	10.2035	0.1147	4.9976	45.9212	41.6087
Sample Variance	2218.0510	311.9838	304.5645	104.1120	0.0131	24.9768	2108.7620	1731.2860
Kurtosis	3.2485	−0.0357	−0.8760	−1.5059	5.6825	−1.5102	1.1467	9.9792
Skewness	1.9753	0.6548	0.3552	0.1163	−2.2054	−0.1352	1.3116	2.7150

Table 4 | Details of model input and model output with the model number

Model no.	Model input	Model output
CMB1	M, D, SI, C, SA, and T	IR & CI
CMB 2	D, SI, C, SA, and T	IR & CI
CMB 3	M, D, SI, C, SA, and T	IR & CI
CMB 4	M, D, SI, C, SA, and T	IR & CI
CMB 5	M, D, SI, C, SA, and T	IR & CI
CMB 6	M, D, SI, C, SA, and T	IR & CI
CMB 7	M, D, SI, C, SA, and T	IR & CI

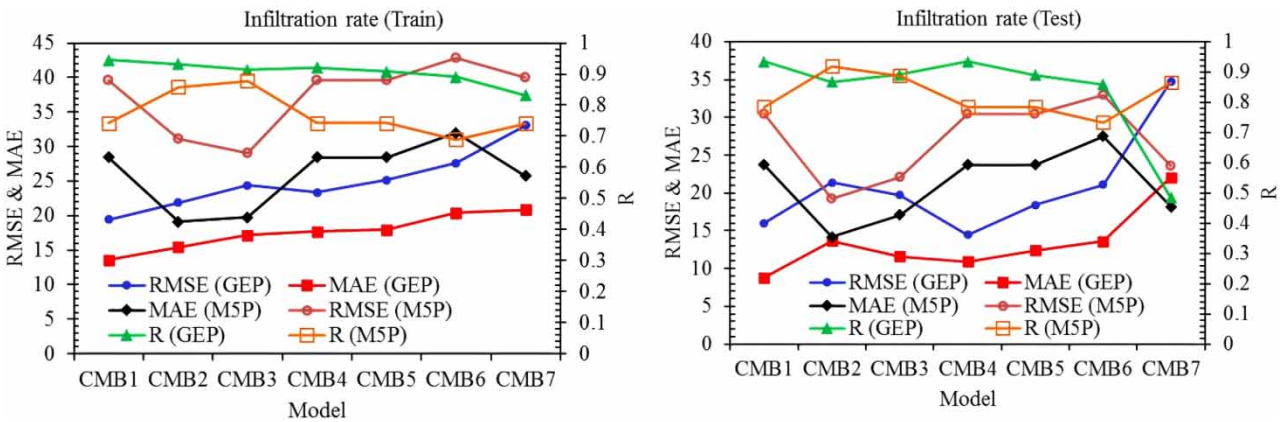

Figure 5 | Performance assessment parameters using M5P and GEP techniques (IR).

$$\mathrm{RMSE} = \sqrt{\frac{1}{p}\left(\sum_{i=1}^{p}(j_i - k_i)\right)^2} \tag{3}$$

$$R = \frac{\sum_{i=1}^{p}(j_i - \bar{j})(k_i - \bar{k})}{\sqrt{\sum_{i=1}^{p}(j_i - \bar{j})^2 \sum_{i=1}^{N}(k_i - \bar{k})^2}} \tag{4}$$

IR

To get the best prediction of the IR, M5P and GEP were used. The details of the performance evaluation parameters are listed in Figure 4. GEP provided good results in the prediction of the IR. An output from Figure 5 suggests that the values of R vary from 0.4847 to 0.9343, RMSE values vary from 14.4233 mm/hr to 34.7474 mm/hr and MAE values vary from 8.7676 mm/hr to 22.0840 mm/hr for the training dataset for all seven models. Among the 7 models, the model with CMB 1 gave the highest accurate result with R, RMSE, and MAE values equal to 0.9343, 15.9667 mm/hr, and 8.7676 mm/hr respectively. CMB 7 was the worst model combination. Taylor graph (Figure 6) results for GEP and M5P show that the best model for these techniques is CMB1 and CMB2, respectively. The performance of the best accurate model combinations was plotted in Figure 7. The figure combined the two outputs; the first is the variation of the predicted values and actual values with the number of the dataset and the second is the scattered plot of the actual values and predicted values of IR. The value of R is 0.9159 for the scatters of the predicted and actual values of the IR with CMB 1. Thus, CMB 1 was the best-fitted combination model to predict the IR with GEP. M5P also provided good results for the IR. The values of R, RMSE, and MAE values vary from 0.7333 to 0.9192, 14.1821 mm/hr to 27.4874 mm/hr, and 19.2475 mm/hr to 32.9872 mm/hr respectively (Figure 5). The model with CMB 2 gave the most precise result as compared to the other model combinations. Due to Figure 7 for the M5P. The value of

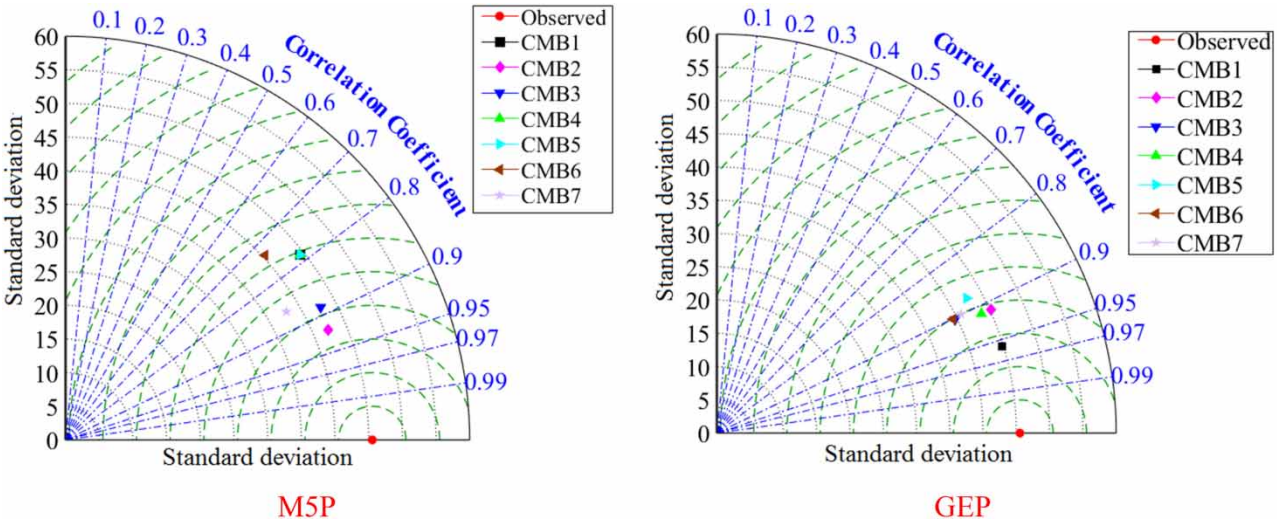

Figure 6 | Taylor diagram for all introduced models by M5P and GEP (IR).

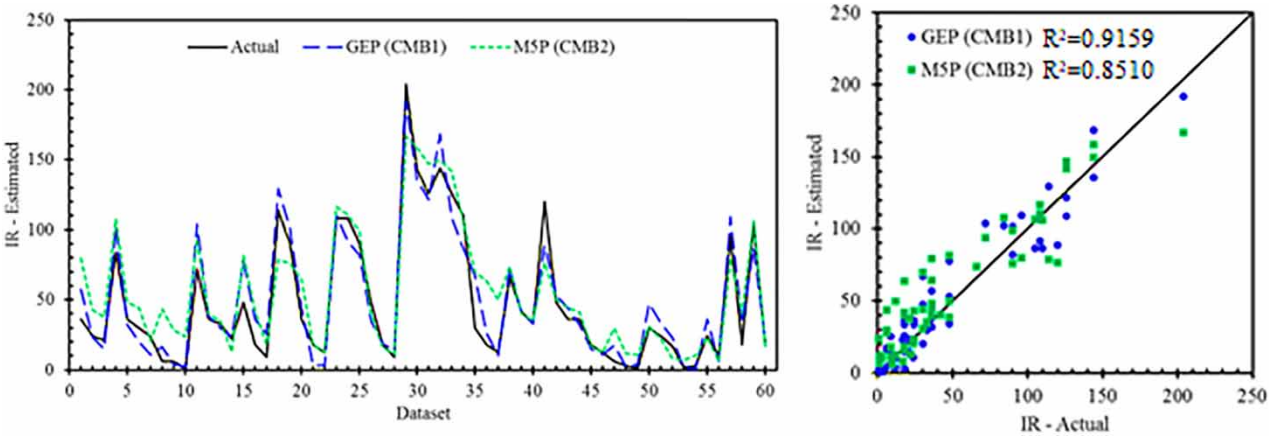

Figure 7 | The performance of the IR with the most accurate model using GEP (CMB 1) and M5P (CMB 2).

R (0.9192) is the highest and values of the RMSE and MAE (14.1851 mm/hr and 19.2475 mm/hr) are lowest for CMB 2 compared to other model combinations. Similarly, the value of R for this case is 0.8510. The major benefit of the M5P and GEP is that they provided a linear relationship between the input and output variables. The details of the linear relation for both of the models are summarized in Table 5. In M5P, if C <= 21.37, linear relation 1 (LM num 1) was followed, if C> 21.37 and T <= 17.5, linear relation 2 (LM num 2) was followed and so on. A total of five linear models were given by the M5P model with model combination CMB 2. For GEP, only one model, M num1, was created which is listed in Table 5. CMB 1 and CMB 2 were the best-fitted combination model in the prediction of IR for GEP and M5P, respectively. The values of R, RMSE, and NAE were 0.9343, 15.9667 mm/hr and 8.7676 mm/hr, and 0.9192, 14.1821 mm/hr, and 19.2497 mm/hr for GEP and M5P, respectively (Figure 5). The output from Figures 4–6 suggests that the GEP technique is more accurate than the M5P technique. The plot of the GEP technique is more symmetrical than M5P with the highest values of R (0.9159), which is much higher than M5P (R = 0.8510) (Figure 7). Thus, the GEP technique is the most precise technique to predict the IR with CMB 1.

CI

In this section, the prediction of the CI with 7 models' combinations was done using GEP and M5P. The values of the performance evaluation parameters i.e. R, RMSE, and MAE were also provided in Figure 8. The values of R, RMSE and MAE varies from 0.6311 to 0.9586, 9.2522 mm to 19.5626 mm and 7.7856 mm to 14.2538 mm and 0.5366 to 0.8987, 11.2144 mm to 32.6036 mm and 18.4328 mm to 54.3111 mm for GEP and M5P, respectively. CMB 1 and CMB 6 were the best-fitted

Table 5 | Linear relation model provided by M5P for IR using CMB 2

M5P	GEP
C <= 21.37: LM1 (31/59.976%) C >21.37: | T <= 17.5: LM2 (26/77.865%) | T > 17.5: | | SA <= 41.345: LM3 (30/12.659%) | | SA >41.345: | | | C <= 30.05: LM4 (27/8.392%) | | | C > 30.05: LM5 (11/9.541%) LM num: 1 IR = −0.3958 * T − 1.1449 * SA − 8.5026 * C + 316.7511 LM num: 2 IR = −4.4779 * T + 1.7928 * SA +39.2196 LM num: 3 IR = −0.1571 * T+ 0.8181 * SA +0.2557 LM num: 4 IR = −0.1716 * T + 1.072 * SA +0.1278 * SI − 6.3953 LM num: 5 IR = −0.1834 * T + 2.0062 * SA − 0.3486 * SI − 30.8452	M num: 1 IR = (SA + (cos((SA-M)).*((M − 3.887726).*D))) + ((((SA + 9.966279)-(M − 9.966279))./T).* 9.280182) + (SI-((log(C)-sin(SA)).*(SA.^(1.0./3.0)))) + (((((0.257995-M).* (−2.68805)) + M).*cos(log(C))) + ((−3.535065)./cos((((SI + D).*SI).*(−3.535065 − 3.556549)))).

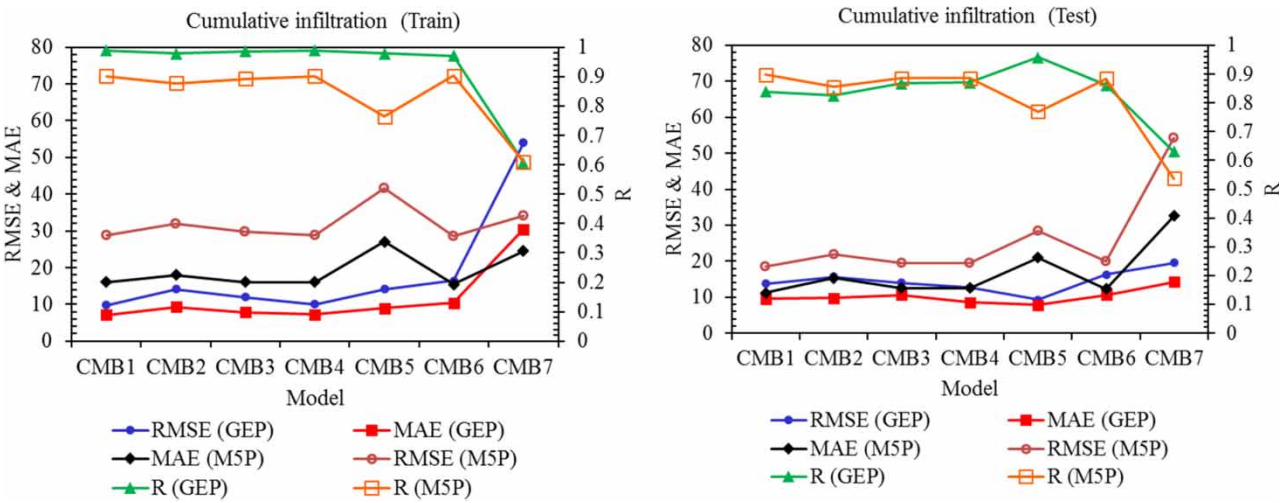

Figure 8 | Performance assessment parameters using M5P and GEP techniques (CI).

models with GEP and M5P by the Taylor diagram and performance graph (Figure 9 and 10). The values of R for these combinations were 0.9586 and 0.8987, which was the highest among all model combinations and the values of RMSE and NSE were 9.2522 mm and 11.2144 mm and 7.7856 mm and 18.4328 mm, which was lowest from CMB 5 with GEP and CMB 1 with M5P. But CMB 7 was the worst model combination with both the modeling techniques. The performance of the GEP and M5P with the best model (CMB 1 and CMB 6 respectively) is provided by Figures 9 and 10. Similarly, the performance combined two outputs which were explained in the IR section. The linear relation model for the CI is listed in Table 6. It is clear from Table 6 that a total of 9 linear models was developed for the CMB 6 using M5P.

Similarly, CMB 1 and CMB 6 were the most accurate combination model to predict the CI with GEP and M5P respectively. Figure 8 suggests the value of R^2 which was higher in the case of the GEP (0.8949) than M5P (0.8349) and all the plots of the GEP were in symmetry. It also suggests that the GEP has a good result of R, RMSE, and MAE (0.9586, 9.2522 mm, and 7.7865 mm) for GEP than M5P (0.8987, 11.2144 mm, and 18.4328 mm). Thus, in CI also, the GEP technique is the most precise technique with CMB 1.

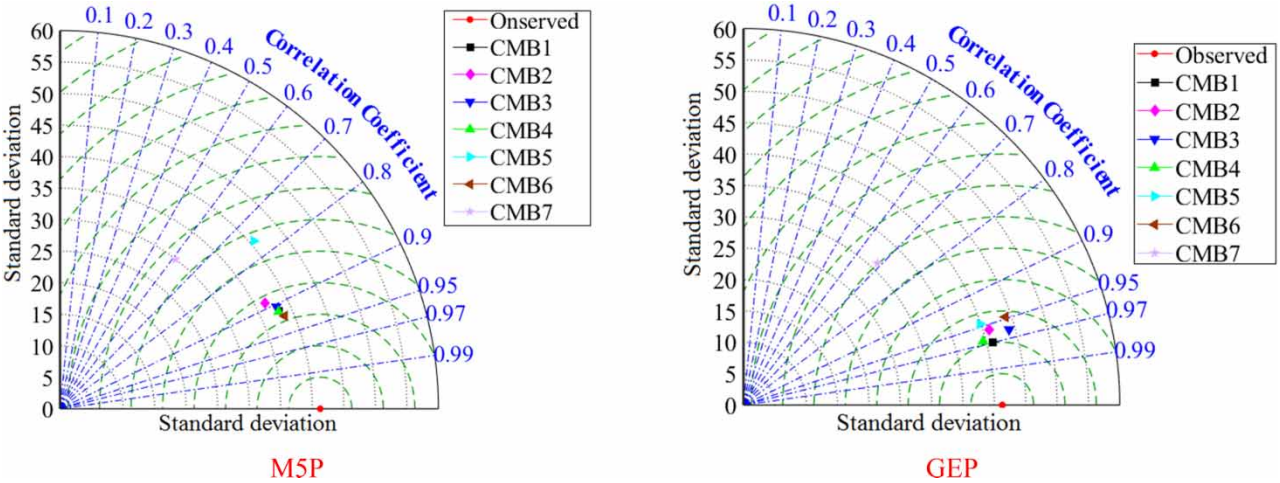

Figure 9 | Taylor diagram for all introduced models by M5P and GEP (CI).

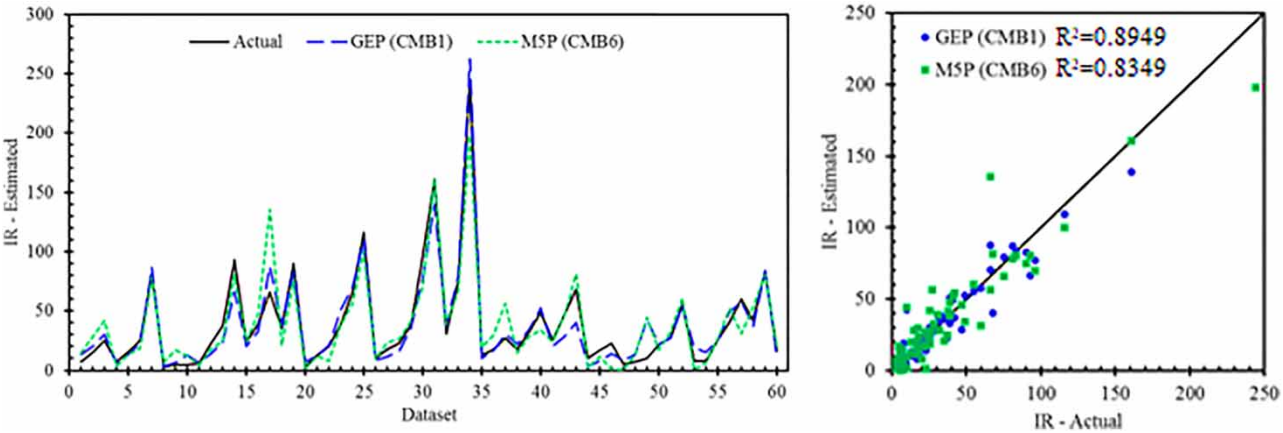

Figure 10 | The performance of the IR with the most accurate model using GEP (CMB 1) and M5P (CMB 6).

A comparison was done with the previous studies to check the potential best machine learning techniques; GEP. The selected previous studies are Vand *et al.* 2018, Sihag *et al.* 2017, 2018b, Singh *et al.* 2019a; Sepahvand *et al.* 2021. Figure 11 shows the result of the comparison, which suggests that the result of the current study is best compared to the previous studies. The order of the results is as follows: Singh *et al.* 2017, 2019a; Sihag *et al.* 2018b, Vand *et al.* 2018; Sepahvand *et al.* 2021; Current study. Thus, the result of the current study outperformed the previous studies.

Additionally, both the machine learning techniques (M5P and GEP) are capable of providing the explicit equation (provided in Tables 5 and 6), which can be useful in calculating the infiltration characteristics. These equations may be used in such regions where the measuring of infiltration characteristics is very difficult. Furthermore, accurate estimation of soil infiltration rate and thereby runoff rate will help to develop proper soil management strategies and conservation measures to minimize the risk of erosion and land degradation.

Single factor ANOVA

The single factor ANOVA for both of the techniques is depicted in Table 8. For an insignificant result, the *P*-value should be more than 0.05 and F-value should be less than F critical. Table 7 suggests that in both of the techniques, the *P*-value is more than 0.05 and F-value is less than F critical for both IR and CI. Thus, the results of both of the predictors are insignificant for IR and CI.

Table 6 | Liner relation model provided by M5P (CMB6) and GEP (CMB1) for CI

M5P	GEP
T <= 55: | M <= 8.62: | | T <= 12.5: LM1 (7/5.005%) | | T > 12.5: | | | SA <= 46.185: LM2 (9/4.207%) | | | SA >46.185: LM3 (13/12.117%) | M > 8.62: | | SA <= 41.345: | | | SA <= 33.235: LM4 (21/3.35%) | | | SA >33.235: LM5 (6/1.851%) | | SA >41.345: LM6 (37/13.068%) T >55: | M <= 14.79: | | C <= 18.09: LM7 (6/59.051%) | | C > 18.09: LM8 (14/72.846%) | M > 14.79: LM9 (12/26.853%) LM num: 1 CI = 0.9269 * T + 0.365 * SA − 0.379 * C + 0.0524 * SI − 13.7013 * D − 1.1599 * M + 42.2674 LM num: 2 CI = 1.0951 * T + 0.5658 * SA − 0.4946 * C − 13.7013 * D- 1.1599 * M + 37.4434 LM num: 3 CI = 1.2975 * T + 0.365 * SA − 0.8058 * C − 13.7013 * D − 1.1599 * M + 47.5746 LM num: 4 CI = 0.4137 * T + 0.1036 * SA − 0.0518 * C − 13.7013 * D + 0.061 * M + 17.4881 LM num: 5 CI = 0.3773 * T − 0.0277 * SA − 0.0518 * C − 13.7013 * D + 0.028 * M + 21.1074 LM num: 6 CI = 0.5896 * T + 0.2873 * SA +0.9715 * C − 13.7013 * D − 1.1148 * M + 10.0137 LM num: 7 CI = 0.8363 * T + 2.5836 * SA − 96.3248 * D − 3.7514 * M + 90.6576 LM num: 8 CI = 0.5766 * T + 2.275 * SA − 236.6134 * D − 3.7514 * M + 351.3984 LM num: 9 CI = 0.4467 * T + 2.0005 * SA − 115.537 * D − 4.4842 * M + 173.5502	M num:1 CI = ((T + M)./((sin(C) + SI).*sin(SA))) + ((((SI./d5)./2.341217) + (4.281494-M)) + ((T.^(1.0./3.0))-(T./SI))) + ((C + T)./(tan((M./ 9.998108)).*(C./ 9.888886))).

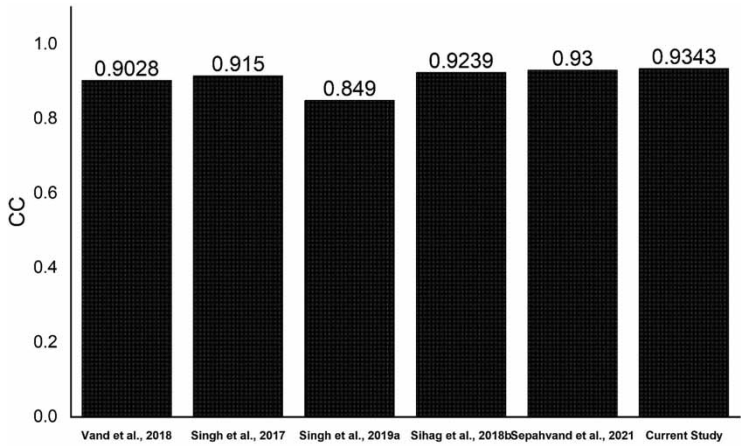

Figure 11 | Comparison of current results with previous studies.

Table 7 | Single-factor ANOVA for IR and CI using GEP and M5P

Infiltration characteristics	Techniques	Model	F- value	*P*-value	F critical	Result
IR	GEP	CMB 1	0.1068	0.7444	3.9215	Insignificant
	M5P	CMB 2	0.0063	0.9367	3.9215	Insignificant
CI	GEP	CMB 5	0.1068	0.7444	3.9215	Insignificant
	M5P	CMB 1	0.0063	0.9367	3.9215	Insignificant

Table 8 | UA for the above-mentioned soft computing techniques

Infiltration characteristics	Soft computing techniques	Mean prediction error $\overline{er}$	SD_{er}	95% prediction error interval
IR	M5P CMB2	−0.0019	0.0153	±0.064
	GEP CMB1	+0.0006	0.0079	±0.048
CI	M5P CMB6	−0.0016	0.00186	±0.068
	GEP CMB1	+0.0009	0.0096	±0.053

UA

The accurate technique, GEP, underwent UA after Single Factor ANOVA. In this investigation, UA gives a nice performance description of GEP models. Several researchers have used UA for data testing (Azimi *et al.* 2018, 2019). Uncertainty is the difference of actual and estimated values $er_i = j_i - k_i$. The mean difference and standard deviation (SD_{er}) of predicted values are calculated with $\overline{er} = \sum_{i=1}^{m} er_i$ and $SD_{er} = \sqrt{\sum_{i=1}^{m} (er_i - \overline{er})^2 / n - 1}$. The positive and negative signs in the prediction error specify the model underestimation and overestimation of estimated values. Parameters $\overline{er}$ and SD_{er} are used to define a certainty band around the prediction error values with the Wilson score method without continuity correction. Table 8 gives the details of the UA. In IR, GEP CMB1 gave the minimum values of the indices used for the UA ($\overline{er} = +0.0006$, $SD_{er} = 0.0079$ and 95% prediction error interval $= \pm 0.0480$) and also for CI, GEP CMB1 gave the minimum values. Hence, CMB1 gave the best results for predicting the infiltration characteristics of soil which were also interpreted by the performance assessment parameters (Figures 4 and 7), Taylors diagram (Figures 5 and 8), and performance graph (Figures 6 and 9).

CONCLUSIONS

Infiltration characteristics measuring is a time-consuming and complicated task for water resource and agriculture researchers. Thus, a trustworthy soft computing technique can be a perfect replacement. For this objective, M5P and GEP were used to model the infiltration characteristics with seven model combinations. The following conclusions are drawn from this investigation:

- GEP is the most efficient technique to predict the infiltration characteristics (both IR and CI) of the soil of four districts of Haryana, India.
- CMB 1, with input combination M, D, SI, C, SA, and T, gave the best values of the performance evaluation parameters (R, RMSE, and MAE = 0.9343, 15.9667, and 8.7676) compared to the other model combination to predict the IR.
- CMB 1 with input combinations M, D, SI, C, SA, and T, was the best model in the prediction of CI with R, RMSE, and MAE values equal to 0.9586, 9.2522, and 7.7865.
- Linear relationships among the input and output variables were given to find the values of infiltration characteristics at any instance by the GEP and M5P soft computing techniques.
- A comparison with past studies also revealed that the GEP model of this study is superior.
- Single-factor ANOVA suggested that both the techniques gave insignificant results in the prediction of the IR and CI with *P*-value more than 0.05 and F-value less than F critical.
- UA also gave results that concluded GEP is the best soft computing technique to predict the infiltration characteristics.

The limitations of the study are that it works on a limited dataset where the types of soil are of the same characteristics. Also, the study is not applicable in the prediction of infiltration characteristics with vegetables as all the experiments were performed in bare land. While the findings based on M5P and GEP are thought to be promising for future studies. Therefore, it would be worthy to the extent the current analysis with advanced machine learning techniques (i.e. GMDH, MARS, and particularly hybrid (conventional + optimization) techniques including ANN + FFA, ANN + PSO, ANFIS + GA, ANFIS + PSO, SVR + GA, SVR + PSO, etc. The use of these soft computing techniques will open up a new chance for the modeling of infiltration characteristics.

FUNDING

No funding source available.

CONFLICT OF INTEREST

Authors have no conflict of interest.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories at https://drive.google.com/drive/folders/1-JD_CNHaGO-pQTXJ2KI4fL9xqGpVqgw9?usp=sharing.

REFERENCES

Adeniji, F. A., Umara, B. G., Dibal, J. M. & Amali, A. A. 2013 Variation of infiltration rates with soil texture. A laboratory study. *International Journal of Engineering and Innovative Technology* **3** (2), 454–459.

Al-Obaidi, A. R. 2019 Monitoring the performance of centrifugal pump under single-phase and cavitation condition: a CFD analysis of the number of impeller blades. *Journal of Applied Fluid Mechanics* **12** (2), 445–459.

Al-Obaidi, A. R. 2021 Numerical investigation on effect of various pump rotational speeds on performance of centrifugal pump based on CFD analysis technique. *International Journal of Modeling, Simulation, and Scientific Computing* **12** (5), 2150045.

Al-Obaidi, A. R. & Mohammed, A. A. 2019 Numerical investigations of transient flow characteristic in axial flow pump and pressure fluctuation analysis based on the CFD technique. *Journal of Engineering Science & Technology Review* **12** (6), 70–79.

Angelaki, A., Singh Nain, S., Singh, V. & Sihag, P. 2018 Estimation of models for cumulative infiltration of soil using machine learning methods. *ISH Journal of Hydraulic Engineering* **27**, 1–8.

Aradhana, A., Singh, B. & Sihag, P. 2021 Predictive models for estimation of labyrinth weir aeration efficiency. *Journal of Achievements in Materials and Manufacturing Engineering* **105** (1), 18–32.

Arora, S., Singh, B. & Bhardwaj, B. 2019 Strength performance of recycled aggregate concretes containing mineral admixtures and their performance prediction through various modeling techniques. *Journal of Building Engineering* 24, 100741.

ASTM, D 2007 Standard test method for particle-size analysis of soils. ASTM International, West Conshohocken, PA.

ASTM International 2009 *Standard Test Method for Infiltration Rate of Soils in the Field Using Double-Ring Infiltrometer*. ASTM International West Conshohocken, PA.

Ayoubloo, M. K., Azamathulla, H. M., Ahmad, Z., Ghani, A. A., Mahjoobi, J. & Rasekh, A. 2011 Prediction of scour depth in downstream of ski-jump spillways using soft computing techniques. *International Journal of Computers and Applications* **33** (1), 92–97.

Azimi, H., Bonakdari, H., Ebtehaj, I. & Michelson, D. G. 2018 A combined adaptive neurofuzzy inference system–Firefly algorithm model for predicting the roller length of a hydraulic jump on a rough channel bed. *Neural Computing & Applicatons* **29** (6), 249–258. doi:10.1007/s00521-016-2560-9.

Azimi, H., Bonakdari, H. & Ebtehaj, I. 2019 Gene expression programming-based approach for predicting the roller length of a hydraulic jump on a rough bed. *ISH Journal of Hydraulic Engineering* **27**, 77–87.

Bhattacharya, B. & Solomatine, D. P. 2005 Neural networks and M5 model trees in modelling water level–discharge relationship. *Neurocomputing* **63**, 381–396.

Bhave, S. & Sreeja, P. 2013 Influence of initial soil condition on infiltration characteristics determined using a disk infiltrometer. *ISH Journal of Hydraulic Engineering* **19** (3), 291–296.

Bhoria, S., Sihag, P., Singh, B., Ebtehaj, I. & Bonakdari, H. 2021 Evaluating Parshall flume aeration with experimental observations and advance soft computing techniques. *Neural Computing and Applications* **24**, 1–15.

Chowdary, V. M., Rao, M. D. & Jaiswal, C. S. 2006 Study of infiltration process under different experimental conditions. *Agricultural Water Management* **83** (1), 69–78.

Connely, J., Jensen, W. & Hamon, P. 2008 *Proctor Compaction Testing. Construction Management Program*. Nebraska Department of Roads, Nebraska.

Dahan, O., Shani, Y., Enzel, Y., Yechieli, Y. & Yakirevich, A. 2007 Direct measurements of floodwater infiltration into shallow alluvial aquifers. *Journal of Hydrology* **344** (3-4), 157–170.

Ebtehaj, I., Bonakdari, H., Zaji, A. H., Azimi, H. & Sharifi, A. 2015 Gene expression programming to predict the discharge coefficient in rectangular side weirs. *Applied Soft Computing* **35**, 618–628.

Ferreira, C. 2002 Gene expression programming in problem solving. In: *Soft Computing and Industry*. Springer, London, pp. 635–653.

Ferreira, C. 2006 *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*. Vol. 21. Springer, London.

Gholami, A., Bonakdari, H., Zeynoddin, M., Ebtehaj, I., Gharabaghi, B. & Khodashenas, S. R. 2018 Reliable method of determining stable threshold channel shape using experimental and gene expression programming techniques. *Neural Computing and Applications* **31** (10), 1–19.

Goyal, M. K. & Ojha, C. S. P. 2011 Estimation of scour downstream of a ski-jump bucket using support vector and M5 model tree. *Water Resources Management* **25** (9), 2177–2195.

Green, W. H. & Ampt, G. A. 1911 Studies on soil phyics. *The Journal of Agricultural Science* **4** (1), 1–24.

Gupta, B. L. & Gupta, A. M. I. T. 2008 *Water Resources Systems and Management*. 2nd edn. Standard Publishers Distributors, Delhi 110006, pp. 510–535.

Haghighi, F., Gorji, M., Shorafa, M., Sarmadian, F. & Mohammadi, M. H. 2010 Evaluation of some infiltration models and hydraulic parameters. *Spanish Journal of Agricultural Research* **8** (1), 210–217.

Haghiabi, A. H., Nasrolahi, A. H. & Parsaie, A. 2018 Water quality prediction using machine learning methods. *Water Quality Research Journal* **53** (1), 3–13.

Heinz, I., Pulido-Velazquez, M., Lund, J. R. & Andreu, J. 2007 Hydro-economic modeling in river basin management: implications and applications for the European water framework directive. *Water Resources Management* **21** (7), 1103–1125.

Holtan, H. N. 1961 *Concept of Infiltration Estimates in Watershed Engineering. ARS41-51*. U.S. Department of Agricultural Service, Washington, DC.

Horton, R. I. 1938 The interpretation and application of runoff plot experiments with reference to soil erosion problems. *Soil Science Society of America Proceedings* **3**, 340–349.

Huggins, L. F. & Monke, E. J. 1966 *The Mathematical Simulation of the Hydrology of Small Watersheds*. Technical Report No. 1, Purdue Water Resources. Lafayette: Research Centre.

Karbasi, M. & Azamathulla, H. M. 2016 GEP to predict characteristics of a hydraulic jump over a rough bed. *KSCE Journal of Civil Engineering* **20** (7), 3006–3011.

Khozani, Z. S., Bonakdari, H. & Ebtehaj, I. 2018 An expert system for predicting shear stress distribution in circular open channels using gene expression programming. *Water Science and Engineering* **11** (2), 167–176.

Kochkov, D., Smith, J. A., Alieva, A., Wang, Q., Brenner, M. P. & Hoyer, S. 2021 Machine learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences* **118** (21), 1–8.

Kostiakov, A. N. 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. *Transactions Sixth Commission International Society of Soil Science* **1**, 7–21.

Kumar, M., Kumari, A., Kushwaha, D. P., Kumar, P., Malik, A., Ali, R. & Kuriqi, A. 2020 Estimation of daily stage–discharge relationship by using data-driven techniques of a perennial river, India. *Sustainability* **12** (19), 7877.

Malik, A., Tikhamarine, Y., Souag-Gamane, D., Kisi, O. & Pham, Q. B. 2020 Support vector regression optimized by meta-heuristic algorithms for daily streamflow prediction. *Stochastic Environmental Research and Risk Assessment* **34** (11), 1755–1773.

Malik, A., Tikhamarine, Y., Souag-Gamane, D., Rai, P., Sammen, S. S. & Kisi, O. 2021a Support vector regression integrated with novel meta-heuristic algorithms for meteorological drought prediction. *Meteorology and Atmospheric Physics* **133** (3), 891–909.

Malik, A., Tikhamarine, Y., Al-Ansari, N., Shahid, S., Sekhon, H. S., Pal, R. K., Rai, P., Pandey, K., Singh, P., Elbeltagi, A. & Sammen, S. S. 2021b Daily pan-evaporation estimation in different agro-climatic zones using novel hybrid support vector regression optimized by Salp swarm algorithm in conjunction with gamma test. *Engineering Applications of Computational Fluid Mechanics* **15** (1), 1075–1094.

Malik, A., Tikhamarine, Y., Sammen, S. S., Abba, S. I. & Shahid, S. 2021c Prediction of meteorological drought by using hybrid support vector regression optimized with HHO versus PSO algorithms. *Environmental Science and Pollution Research* **28** (29), 39139–39158.

Mehdipour, V. & Memarianfard, M. 2017 Application of support vector machine and gene expression programming on tropospheric ozone prognosticating for Tehran metropolitan. *Civil Engineering Journal* **3** (8), 557–567.

Mehdipour, V., Memarianfard, M. & Homayounfar, F. 2017 Application of gene expression programming to water dissolved oxygen concentration prediction. *International Journal of Human Capital in Urban Management* **2** (1), 39–48.

Micheal, A. M. 1978 *Irrigation, Theory and Practice*. Vikas Press Private Limited, New Delhi.

Mirzaee, S., Zolfaghari, A. A., Gorji, M., Dyck, M. & Ghorbani Dashtaki, S. 2014 Evaluation of infiltration models with different numbers of fitting parameters in different soil texture classes. *Archives of Agronomy and Soil Science* **60** (5), 681–693.

Mishra, S. K., Tyagi, J. V. & Singh, V. P. 2003 Comparison of infiltration models. *Hydrological Processes* **17** (13), 2629–2652.

Mohammed, A., Yaseen, Z. M., Heddam, S., Malik, A. & Kisi, O. 2021 Advanced machine learning models development for suspended sediment prediction: comparative analysis study. *Geocarto International* 1–25. DOI: 10.1080/10106049.2021.1933210.

Mohanty, S., Roy, N., Singh, S. P. & Sihag, P. 2019 Estimating the strength of stabilized dispersive soil with cement clinker and fly ash. *Geotechnical and Geological Engineering* **37** (4), 2915–2926.

Pal, M. & Deswal, S. 2009 M5 model tree based modelling of reference evapotranspiration. *Hydrological Processes: An International Journal* **23** (10), 1437–1443.

Pandey, K., Kumar, S., Malik, A. & Kuriqi, A. 2020 Artificial neural network optimized with a genetic algorithm for seasonal groundwater table depth prediction in Uttar Pradesh, India. *Sustainability* **12** (21), 8932.

Pandhiani, S. M., Sihag, P., Shabri, A. B., Singh, B. & Pham, Q. B. 2020 Time-series prediction of streamflows of Malaysian rivers using data-driven techniques. *Journal of Irrigation and Drainage Engineering* **146** (7), 04020013.

Parhi, P. K., Mishra, S. K. & Singh, R. 2007 A modification to Kostiakov and modified Kostiakov infiltration models. *Water Resources Management* **21** (11), 1973–1989.

Parsaie, A. & Haghiabi, A. H. 2017 Mathematical expression of discharge capacity of compound open channels using MARS technique. *Journal of Earth System Science* **126** (2), 20.

Patle, G. T., Sikar, T. T., Rawat, K. S. & Singh, S. K. 2018 Estimation of infiltration rate from soil properties using regression model for cultivated land. *Geology, Ecology, and Landscapes* **3** (1), 1–13.

Philip, J. R. 1957 Theory of infiltration. *Soil Science* **83**, 345–357.

Quinlan, J. R. 1992 Learning with continuous classes. In: *5th Australian Joint Conference on Artificial Intelligence*, Vol. 92, pp. 343–348.

Rehamnia, I., Benlaoukli, B., Jamei, M., Karbasi, M. & Malik, A. 2021 Simulation of seepage flow through embankment dam by using a novel extended Kalman filter based neural network paradigm: case study of Fontaine Gazelles Dam, Algeria. *Measurement* **176**, 109219.

Rowe, R. O. S. I. A. 2018 *Soil Moisture. Biosystems Engineering*. Auburn University, Auburn, Alabama, United States.

Sepahvand, A., Singh, B., Ghobadi, M. & Sihag, P. 2021 Estimation of infiltration rate using data-driven models. *Arabian Journal of Geosciences* **14** (1), 1–11.

Shabanlou, S., Azimi, H., Ebtehaj, I. & Bonakdari, H. 2018 Determining the scour dimensions around submerged vanes in a 180° bend with the gene expression programming technique. *Journal of Marine Science and Application* **17** (2), 233–240.

Sihag, P., Tiwari, N. K. & Ranjan, S. 2017 Estimation and inter-comparison of infiltration models. *Water Science* **31** (1), 34–43.

Sihag, P., Singh, B., Gautam, S. & Debnath, S. 2018a Evaluation of the impact of fly ash on infiltration characteristics using different soft computing techniques. *Applied Water Science* **8** (6), 187.

Sihag, P., Singh, B., Sepah Vand, A. & Mehdipour, V. 2018b Modeling the infiltration process with soft computing techniques. *ISH Journal of Hydraulic Engineering* **26** (2), 1–15.

Sihag, P., Kumar, M. & Singh, B. 2020 Assessment of infiltration models developed using soft computing techniques. *Geology, Ecology, and Landscapes* **5** (4), 1–11.

Singh, V. P. & Yu, F. X. 1990 Derivation of infiltration equation using systems approach. *Journal of Irrigation and Drainage Engineering* **116** (6), 837–858.

Singh, K. K., Pal, M. & Singh, V. P. 2010 Estimation of mean annual flood in Indian catchments using backpropagation neural network and M5 model tree. *Water Resources Management* **24** (10), 2007–2019.

Singh, B., Sihag, P. & Singh, K. 2017 Modelling of impact of water quality on infiltration rate of soil by random forest regression. *Modeling Earth Systems and Environment* **3** (3), 999–1004.

Singh, B., Sihag, P. & Singh, K. 2018a Comparison of infiltration models in NIT Kurukshetra campus. *Applied Water Science* **8** (2), 63.

Singh, B., Sihag, P., Singh, K. & Kumar, S. 2018b Estimation of trapping efficiency of vortex tube silt ejector. *International Journal of River Basin Management* **19** (3), 1–38.

Singh, B., Sihag, P. & Deswal, S. 2019a Modelling of the impact of water quality on the infiltration rate of the soil. *Applied Water Science* **9** (1), 15.

Singh, B., Sihag, P., Tomar, A. & Sehgal, A. 2019b Estimation of compressive strength of high-strength concrete by random forest and M5P model tree approaches. *Journal of Materials and Engineering Structures «JMES»* **6** (4), 583–592.

Singh, A., Singh, B. & Sihag, P. 2021a Experimental investigation and modeling of aeration efficiency at Labyrinth Weirs. *Journal of Soft Computing in Civil Engineering* **5** (3), 15–31.

Singh, B., Sihag, P., Singh, V. P., Sepahvand, A. & Singh, K. 2021b Soft computing techniques-based prediction of water quality index. *Water Supply*.

Smith, B. 2006 *The Farming Handbook*. University of Kwazulu-Natal Press and CTA, 6700 AJ Wageningen, The Netherlands, 37-132.

Souchère, V., Millair, L., Echeverria, J., Bousquet, F., Le Page, C. & Etienne, M. 2010 Coconstructing with stakeholders a role-playing game to initiate collective management of erosive runoff risks at the watershed scale. *Environmental Modelling & Software* **25** (11), 1359–1370.

Vand, A. S., Sihag, P., Singh, B. & Zand, M. 2018 Comparative evaluation of infiltration models. *KSCE Journal of Civil Engineering* **22** (10), 4173–4184.

Zakwan, M. 2018 Comparative analysis of the novel infiltration model with other infiltration models. *Water and Environment Journal* **33** (4), 1–13.

First received 7 July 2021; accepted in revised form 28 November 2021. Available online 10 December 2021