Assessment of infiltration models developed using soft computing techniques
Parveen Sihag (a), Munish Kumar (b) and Balraj Singh (c)

(a) Department of Civil Engineering, National Institute of Technology, Kurukshetra, India; (b) Department of Civil Engineering, Panipat Institute of Engineering & Technology, Samalkha, India

ABSTRACT

In this study, predicting ability of support vector machines (SVM), Gaussian process (GP), artificial neural network (ANN), and Random forests (RF) based regression approaches was tested on the infiltration data of soil samples having different compositions of sand, silt, clay, and fly ash. In addition to this, their performances were compared with the Kostiakov model (KM) and Philip’s model (PM). Dataset containing a total of 392 observations was collected from the experimental measurements of soil infiltration rate on different soil samples. Out of the total dataset, 272 recordings were randomly selected for training and the residual 120 observations were selected for validation of the developed models. Standard statistical parameters were used to measure the predicting ability of various developed models. The result suggests that the best performance could be achieved by Polynomial kernel function-based GP regression (GP_Poly) with coefficient of correlation values as 0.9824, 0.9863, Bias values as 0.0006, −2.3542, root-mean-square error values as 47.7336, 40.3026, and Nash Sutcliffe model efficiency values as 0.9655, 0.9727 using training and testing dataset, respectively. Furthermore, time is found as the most influencing input variable for predicting the infiltration rate when GP_Poly-based model is used to predict the infiltration rate.

1. Introduction

Infiltration plays a very important part in hydrology related to above and under the surface of the earth as well as irrigation; it has earned an enormous consideration from hydrologists. A noticeable number of infiltration models have been developed for the estimation of infiltration rate and they can be further classified into subcategories (Mishra, Tyagi, & Singh, 2003): (i) physical model, (ii) semi-empirical model, and (iii) empirical model. Physical and semi-empirical models depend on the derived laws and the equations. Smith (1972), Green and Ampt (1911), Horton (1938), Holtan (1961) are their examples. Empirical model is dependent on the field and laboratory experimental data. Modified Kostiakov (Smith, 1972), Soil conservation service (1972) are the examples of empirical models. Clearly, there are numerous infiltration models but still, their fitness is not clear for the actual world conditions.

Various researchers and scientists have done a lot of research to compare the infiltration models. Skaggs, Huggins, Monke, and Foster (1969) have compared Green-Ampt, Philips, Horton, and Holton models for dissimilar soils with different surface conditions and moisture contents. The Horton and Holton models evaluated an accurate steady-state infiltration rate. Whisler and Bower (1970) analysed the Philip’s model, Green-Ampt model, and numerical model for estimating the infiltration rate of different soil profiles. Rawls, Brakensiek, and Miller (1983) focused on the progress of methods to estimate infiltration model parameters. GENSTAT which is a statistical tool was used by Ogbe, Jayeoba, and Ode (2011) to fit four infiltration models. Sihag, Tiwari & Ranjan (2017a) proposed a Novel Model for the estimation of infiltration rate through soil. This study includes two most popular infiltration models: Kostiakov Model (KM) and Philip Model (PM).

1.2. KM

Kostiakov model (KM) proposed by Kostiakov (1932) is as follows:

\[ f(t) = xt^{-\gamma} \]  

(1)

where \( f(t) \) is the infiltration rate (mm/hr) at time \( t \), \( t \) is time of infiltration (sec), \( x \) and \( y \) are constants in this equation.

1.3. PM

Philip (1957) introduced a physical model by resolving Richards equation (non-linear partial differential equation, Richards, 1931). The PM is indicated as (Philips, 1957):

CONTACT Parveen Sihag (parveen12sihag@gmail.com) Department of Civil Engineering, National Institute of Technology, Kurukshetra 136119, India

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group on behalf of the International Water, Air & Soil Conservation Society(INWASCON). This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
\[ f(t) = \frac{1}{2} St^{-0.5} + A \]  

where \( f(t) \) is the infiltration rate (mm/hr), \( t \) is time (sec), \( S \) is the Sorptivity (mm/hr\(^{0.5}\)), and \( A \) is the rate factor (mm/hr).

Modelling techniques are trending and popular nowadays and showed successful application in the field of infiltration in order to simulate the infiltration rate of soil with basic soil variables (Kumar & Sihag, 2019; Sihag et al., 2018a, 2018b; Singh, Sihag, & Singh, 2017). Within few decades, artificial intelligence approaches like GP, SVM, artificial neural network, M5 tree, fuzzy logic, ANFIS, and random forest regression have been successfully implemented in water resources engineering problems (Ghorbani, Khatibi, Goel, FazeliFard, & Azani, 2016; Kumar, Tiwari, & Ranjan, 2018; Sihag, Tiwari, & Ranjan, 2017b) and observed to be working admirably.

In this paper, modelling methods, viz., GP, SVM, ANN, and RF-based regression are investigated and the abilities of these strategies in modelling the infiltration rate of water through sandy soil are explored. Further, the performances of these modelling methods are compared with KM and PM infiltration models.

2. Soft computing techniques

2.1. Support vector machines (SVM)

This method is a regression and classification approach which originates from statically learning theory (Cortes & Vapnik, 1995). The SVMs classification techniques depend on the standard of ideal division of classes. In the event that the classes are divisible, this strategy chooses, from amongst the endless number of linear classifiers, the one with minimal generalization error. Along these lines, the chosen hyperplane will be one that leaves the most extreme edge between the two classes, where edge is characterized as the total of the separations of the hyperplane from the nearest purpose of the two classes. It very well may be accomplished by anticipating the first arrangement of factors into a higher dimensional element space and figuring a straight characterization issue in the element space (Aghelpour, Mohammadi, & Biazar, 2019; Smola, 1996; Vapnik, 1998).

2.2. Gaussian processes regression (GP)

The Gaussian (GP) models depend on the presumption that nearby observations ought to pass on data about one another. They indicate an earlier specifically over function space. Therefore, the GP is a natural generalization distribution whose covariance is a matrix and mean is a vector. The Gaussian method is based upon the function, whereas distribution relies upon the vector. Due to earlier information pertaining to the function, the validation is not necessary for speculation and Gaussian process regression model can comprehend the prescient distribution related to test input Rasmussen (2006).

A Gaussian procedure is characterized as an accumulation of arbitrary factors, any limited number which has a joint multivariate Gaussian distribution. The \( n \) number of pairs \((x_i, y_i)\) has been made through the \((\chi \times y)\) which indicates the input and output data domain, correspondingly. It is assumed that \( y \subseteq R \), accordingly the GP on \( \chi \) is uttered by mean function \( \mu: \chi \rightarrow \mathbb{R} \) and covariance function \( \kappa : \chi \times \chi \rightarrow \mathbb{R} \).

2.3. Artificial neural networks (ANN)

The artificial neural system (ANN) is widely drawn in for numerical prescence and grouping (Aggarwal, Goel, & Singh, 2012; Jahani & Mohammadi, 2019; Kia et al., 2012; Moazenzadeh & Mohammadi, 2019). It is manufactured with quantities of handling components and includes three essential layers, for example, the information layer, hidden layer, and output layer correspondingly. The channel in the midst of the layer is used to make the weight relationship in the midst of the hubs. Each node is similar to biological neuron and performs mostly two tasks. It has done the total of the information values and weight related with each interaction. Further, this summation is yielded over activation function to make the result. By giving the weight, the system creates a result which is existed close to the watched target result. The complete detail about ANN is given by Haykin (1999). In this current investigation, one hidden layer is used for the model development.

2.4. Random forest regression (RF)

Random Forest is one of the latest techniques which is used for the classification and regression-based analyses. In this technique, the number of trees with different verities was used for forecasting or estimation. Tree predictors were used numerical values as randomly to class labels in random forest classifier (Breiman, 1999). Random forest regression used in this study contains an assembly of input parameters or arbitrarily chosen parameters at each node to grow a tree. RF technique requires only two user-defined parameters such as the number of parameters used at each node and the amount of trees (Breiman, 1999).

3. Methodology and dataset

An experimental investigation was performed using mini-disk infiltrometer (Decagon Devices Inc., 2006) in the laboratory at the National Institute of
TABLE 1. Characteristics of the material selected in this study.

| Properties | Specific gravity | D_95 (mm) | C_s | Colour |
|------------|-----------------|-----------|-----|--------|
| Sand       | 2.48            | 0.438     | 3.129 | White  |
| Rice husk ash | 1.89            | 0.19      | 3.2  | Black  |
| Fly ash    | 2.07            | 0.18      | 2.7333 | Gray   |

Technology, Kurukshetra, India. The materials chosen for this study were sand, rice husk ash, and fly ash. All the observations had been taken on pre-assumed mixing proportion and other initial conditions such as moisture content and bulk density. The characteristics of the material are indicated in Table 1.

Total 392 observations were collected from the experiments, in which 272 observations, randomly taken from the total data, were used for preparing the models and rest 120 were used to validate/test the developed model performance. Time (T), sand content (S), rice husk ash (R), fly ash (Fa), suction head (s), bulk density (B), and moisture content (w) were chosen as input variables in this study, whereas infiltration rate was taken as the output variable. Table 2 records the statistical features of training and testing dataset that were taken in this investigation.

### Table 2. Characteristics of training and testing dataset selected.

| Parameter | Unit     | Min^m | Max^m | Mean   | St. Dev. | Min^m | Max^m | Mean   | St. Dev. |
|-----------|----------|-------|-------|--------|----------|-------|-------|--------|----------|
| T         | sec      | 6.75  | 6056.8| 341.26 | 511.93   | 22.93 | 4731.2| 372.16 | 504.33   |
| S         | %        | 50    | 90    | 61.65  | 14.50    | 50    | 90    | 61.58  | 14.84    |
| R         | %        | 5     | 45    | 21.01  | 13.23    | 5     | 45    | 21.58  | 13.63    |
| Fa        | cm       | 0.5   | 6     | 1.215  | 0.989    | 0.5   | 6     | 1.19   | 1.01     |
| s         | gm/cc    | 0.84  | 1.73  | 1.18   | 0.23     | 0.84  | 1.73  | 1.17   | 0.24     |
| B         | mm/hr    | 10.25 | 1341.36| 433.67 | 256.28   | 11.54 | 1039.27| 402.96 | 244.84   |

5. Performance criteria

Correlation of coefficient (C.C), Bias, root-mean-square error (RMSE), and Nash–Sutcliffe model efficiency coefficient (E) are the performance evaluation criteria, which are implemented in the current study to evaluate the fitting ability of the models.

#### 5.1. Coefficient of correlation

C.C is used for evaluating the performance of any model using numeric values. The C.C is given as:

\[
C.C = \frac{n \sum a^2 - (\sum a)^2 (\sum b)^2}{\sqrt{n(\sum a^2)^2 - (\sum a)^4} \sqrt{n(\sum b^2)^2 - (\sum b)^4}}
\]

where

- \(a\) = actual data values.
- \(b\) = Estimated (computed) data values.

The range of correlation coefficient is: -1 to +1. Zero value indicates that there is no relation between the actual and predicted data.

#### 5.2. Bias

The bias is the average difference among actual and predicted values. Its value is defined by Equation (4):

\[
Bias = \frac{\sum a \cdot b}{n}
\]

#### 5.3. Root-mean-square error (RMSE)

Root-mean-square error is generally selected to measure numeric assessment. RMSE is calculated as:

\[
RMSE = \sqrt{\frac{1}{n} \sum (a - b)^2}
\]

### Table 3. Optimal value of primary parameters.

| Approach | Kernel function | User-defined parameters |
|----------|-----------------|-------------------------|
| GP       | RBF             | Noise = 0.001, \( y = 0.005 \) |
| SVM      | RBF             | C = 8, \( y = 0.005 \) |
| ANN      | -               | H = 12, L = 0.2, iterations = 1500, M = 0.1 |
| RF       | -               | k = 0.1, m = 1, l = 1000 |

4. Detail of kernel functions and user-defined parameters of soft computing techniques

In this investigation, the two most popular kernel functions, i.e., polynomial \(K(x, x') = ((x \cdot x')^d)\) and radial basis kernel function \(K(x, x') = e^{-\|x-x'|^2}\) were implemented in GP and SVM, where \(d\) and \(y\) are the variables of polynomial and radial basis kernel function correspondingly.

In ANN, hidden layer neurons (H), learning rate (L), iterations, and momentum (M) are adjusted during model learning and testing, while in RF regression, setting of two user-defined parameters: no. of trees (I) and features allowed at each node (m) are required during training and testing.

Several manual adjustments were tried by tuning the user-defined parameters of soft computing techniques aiming to obtain the minimal error between the predicted and actual output. The optimal values of primary parameters are listed in Table 3 for SVM, GP, ANN, and RF-based models.
5.4. Nash–Sutcliffe efficiency coefficient (E)

It is implemented to examine the predicting power of the models. It is expressed as (Nash and Sutcliffe, 1970):

\[ E = 1 - \frac{\sum_{i=1}^{n} (a - b)^2}{\sum_{i=1}^{n} (a - \bar{a})^2} \]

(6)

When the value of \( E \geq 90\% \), it means that the performance of the model is a satisfactory performance, if value \( E \) lies between 80% and 90%, it means that the performance is fairly good, and a value \( \leq 80\% \) shows that the performance of the model is an unsatisfactory performance.

The combined use of \( C.C \), Bias, RMSE, and \( E \) provides a sufficient evaluation of every model’s performance and favours a judgment of the precision of the six modelling approaches implemented in the current study.

6. Results

The coefficients of Kostiakov (empirical) model and Philip’s model (physical) were derived using least-squares methods.

Kostiakov model:

\[ f(t) = 3810t^{-0.46} \]

(7)

Philip’s model:

\[ f(t) = \frac{1}{2} 5145.87t^{-0.5} + 190.79 \]

(8)

The performance of Kostiakov and Philip’s models is compiled in Table 4. It indicates that the performances of Kostiakov model and Philip’s model are unsatisfactory. Nash–Sutcliffe model efficiency coefficient was achieved 0.3417 for Kostiakov model and 0.3401 for Philip’s model. As shown in Table 5, single factor ANOVA results indicate that \( F \)-values (4.397386) were greater than \( f \)-critical (3.880827), and \( P \)-values (0.03705) were less than 0.05, which suggests that the variation among predicted values by Kostiakov model and actual values is significant. Figure 1 indicates the flowchart of the methodology. Figure 2 illustrates the agreement diagram among actual and predicted outcomes of the infiltration rate for both models.

Figure 3 displays the agreement diagram between the actual and predicted data of infiltration rate of sand, obtained by using GP regression-based Radial basis kernel and Polynomial kernel functions with the testing dataset. Table 4 suggests that the GP regression-based Polynomial kernel function shows superiority over Radial basis kernel function. GP regression-based Polynomial kernel function acquired higher values of \( C.C \) (0.9863), as well as \( E \) (0.9727), while lower values of Bias (−2.3524) and RMSE (40.3026). Nash–Sutcliffe model efficiency coefficient suggests that RBF kernel has a fairly good performance, but Polynomial kernel-based GP regression has a very good satisfactory performance. From Table 5, single factor ANOVA results suggest that the variation among predicted values by GP_RBF and GP_Poly and the actual values is not considerable.

Figure 4 displays the agreement of actual and predicted value of infiltration rate of sand obtained by using SVM regression-based Radial basis kernel and Polynomial kernel functions with the test data. SVM-based Polynomial kernel functions show closer agreement to the line of perfect prediction. Table 4 suggests that the SVM regression-based Polynomial kernel function performs better than Radial basis kernel function. SVM regression-based Polynomial kernel function acquired higher values of \( C.C \) (0.9629), \( E \) (0.9229) and lower value of Bias (4.5753) and RMSE (67.7070). Nash–Sutcliffe model efficiency coefficient suggests that RBF kernel has a fairly good performance, but Polynomial kernel-based SVM regression has a very good satisfactory performance. According to Table 5, single factor ANOVA suggests that the variation among predicted values by SVM_RBF and SVM_Poly and the actual values is not considerable.

Figures 5 and 6 display the agreement of actual and predicted value of infiltration rate of sand observed by using ANN and Random Forest models with the testing dataset. The infiltration data predicted by the RF model (Figure 6) lie closer to the agreement line when compared with ANN (Figure 5). Table 4 suggests that the Random Forest model works better than ANN and the higher value of the Nash–Sutcliffe model efficiency coefficient (0.9565) gained by RF suggests its satisfactory performance over ANN. Table 5 shows that single factor ANOVA suggests that the difference in

### Table 4. Performance assessing parameters using training and testing dataset.

| Approaches       | Training dataset |                | Test dataset |                |
|------------------|------------------|----------------|--------------|----------------|
|                  | \( C.C \) | Bias | RMSE | \( E \) | \( C.C \) | Bias | RMSE | \( E \) |
| Kostiakov model  | 0.7096 | 6.3307 | 208.4940 | 0.3402 | 0.6324 | 57.0095 | 197.8236 | 0.3417 |
| Philip’s model   | 0.7013 | 0.0012 | 182.0260 | 0.4971 | 0.6231 | 20.6110 | 198.0730 | 0.3401 |
| G P RBF          | 0.9088 | 0.0005 | 106.7934 | 0.8275 | 0.9398 | −13.3398 | 84.5843 | 0.8798 |
| G P Poly         | 0.9624 | 0.0006 | 47.7336 | 0.9655 | 0.9863 | −2.3542 | 40.3026 | 0.9727 |
| SVM, RBF         | 0.7731 | 28.0965 | 164.7065 | 0.5897 | 0.8397 | −3.5768 | 122.5228 | 0.7046 |
| SVM, Poly        | 0.9247 | 17.9179 | 100.1534 | 0.8483 | 0.9629 | 4.5753 | 67.7070 | 0.9229 |
| ANN              | 0.9523 | −45.1273 | 90.1993 | 0.8769 | 0.9667 | −47.8140 | 78.7458 | 0.8957 |
| Random Forest    | 0.9958 | 0.0439 | 25.1966 | 0.9904 | 0.9798 | −0.5078 | 50.8485 | 0.9565 |
predicted values by ANN and Random Forest and the actual values is insignificant.

Table 5. Outcomes of ANOVA single factor test.

| Approaches     | F       | P-value  | F critical |
|----------------|---------|----------|------------|
| Kostiakov model| 4.397386| 0.03705  | 3.880827   |
| Philip’s model | 0.722505| 0.596145 | 3.880827   |
| G_P_RBF       | 0.193155| 0.660703 | 3.880827   |
| G_P_Poly      | 0.005641| 0.940191 | 3.880827   |
| SVM_RBF       | 0.018806| 0.891039 | 3.880827   |
| SVM_Poly      | 0.023136| 0.879232 | 3.880827   |
| ANN           | 2.322978| 0.128803 | 3.880827   |
| Random Forest | 0.000279| 0.986682 | 3.880827   |

Figure 7 displays the agreement of the actual and the predicted values of infiltration rate of sand obtained by Kostiakov, Philip, GP, SVM, ANN, and Random Forest regression approaches with the testing dataset. Table 4 suggests that GP-based Polynomial kernel function works better than other approaches for this dataset. Figure 7 indicates that Polynomial kernel-based (GP and SVM), ANN, and Random Forest approaches work better than Kostiakov and Philip’s models. Nash–Sutcliffe model efficiency coefficient suggests that GP, SVM, ANN, and Random
Forest have satisfactory performances. Single factor ANOVA also suggests that there is not so much variation in actual and predicted values using these approaches.

Figure 8 indicates the residuals plot for the testing period using six modeling approaches. This figure suggests that GP, SVM, ANN, and Random Forest approaches have less residual than Kostiakov and Philip’s models.

7. Discussion
Infiltration prediction models are considered important for the management of stormwater and groundwater recharge and are utilized to predict the nature of subsurface recharge as well as surface runoff. The soil infiltration characteristics largely dependent on the properties of soil and the availability of moisture...
into the soil. Table 6 contains the modeling studies addressing the performance of various soft computing models in simulating the infiltration through soil. A lot of research has been conducted in the field as well as in the laboratory pertaining to the estimation of infiltration characteristics, viz., hydraulic conductivity, cumulative infiltration and infiltration rate of soil, recharging rate, and permeability of soil. Most of the modelling studies recognized the application of ANN on the laboratory as well as field data and substantial number of researchers found appreciable results with this technique (Anari et al., 2011;
Esmaeinejad et al., 2015; Schaap & Leij, 1998; Sedaghat et al., 2016; Sihag, 2018, 2018; Sy, 2006). Observing some recent field-based studies, illustrated in Table 6, RF regression is recognized as superior modeling method in predicting the infiltration rate and hydraulic conductivity of soil in the region of Kurukshetra, India (Kumar & Sihag, 2019; Sihag et al., 2019; Singh et al., 2017). Regarding the infiltration characteristics, the learning suitability of SVM and GP regression methods cannot be ignored as some of the studies suggest better accuracy in measuring the infiltration properties in contrast to some of the other popular soft computing approaches (Das et al., 2011; Elbisy, 2015; Sihag, Tiwari & Ranjan, 2017b; Sihag et al., 2018a, 2018b; Vand et al., 2018; Sihag et al., 2018d; Singh et al., 2019a). Reviewing these data-mining-based studies; ANN, SVM, GP, and RF regression techniques come out to be strong modelling tools in determining the infiltration characteristics of soil in the field as well as in the laboratory (Table 6), so the authors acknowledged the combined utility of these modelling tools in an attempt to compare the prediction performance in simulating the infiltration rate through mixed soil of variable material of different characteristics (sand, rice husk ash, and fly ash) with basic soil properties. The infiltration estimation comparison yielded the highest estimation accuracy with a polynomial kernel-based GP regression followed by RF regression. Results of GP and RF regression methods are way

### Table 6. Comparison of soft computing-based models (past studies).

| Se. No. | Author | Applied modelling approaches | Predicted variable | Best performing model |
|---------|--------|------------------------------|--------------------|----------------------|
| 1       | Schaap & Leij (1998) | ANN | Saturated hydraulic conductivity (field) | ANN |
| 2       | Sy (2006) | ANN, PM, and GAM | Infiltration rate (field) | ANN |
| 3       | Erzin, Gumaste, Gupta, & Singh (2009) | ANN, MRA | Hydraulic conductivity (laboratory) | ANN |
| 4       | Anari, Darani, & Nafarzadegan (2011) | ANN, ANFIS, LLR, and DLLR | Infiltration rate (field) | ANN |
| 5       | Das, Samui, & Sabat (2011) | ANN and SVM | Field hydraulic conductivity (field) | SVM |
| 6       | Yilmaz, Marschalko, Bednark, Kaynar, & Fotova (2012) | ANN and ANFIS | Permeability of soil (laboratory) | ANFIS |
| 7       | Arshad, Sayyad, Mosaddeghi, & Gharabaghizadeh (2013) | ANFIS, MLR, RF-BFNN, and MLPPNN | Unsaturated hydraulic conductivity (field) | ANFIS and RBFNN |
| 8       | Al-Sulaiman (2015) | ANFIS | Unsaturated hydraulic conductivity (field) | Gaussian membership based ANFIS |
| 9       | Elbisy (2015) | SVM (radial, Linear, sigmoid) | Hydraulic conductivity (field) | Radial basis SVM |
| 10      | Esmaeinejad, Ramezanpour, Seyedmohammadi, & Shahanpour (2015) | ANN, MLR, and Rosetta model | Pedotransfers functions (field) | ANN |
| 11      | Al-Sulaiman & Aboukarima (2016) | ANN and interactive application using C# | Unsaturated hydraulic conductivity (field) | Interactive application using C# |
| 12      | Sedaghat, Bayat, & Sinegan (2016) | ANN | Unsaturated hydraulic conductivity (laboratory) | ANN |
| 13      | Rahmati (2017) | GMDH, MLR, and ANN | Pedotransfers functions (field) | GMDH |
| 14      | Singh et al. (2017) | RF, ANN, and MSP | Unsaturated hydraulic conductivity (laboratory) | RF |
| 15      | Sihag, Tiwari, & Ranjan (2019) | ANFIS, ANN, and MLR | Unsaturated hydraulic conductivity (laboratory) | MLR |
| 16      | Sihag et al. (2017a) | SVR, GP, and MLR | Cumulative infiltration (laboratory) | GP |
| 17      | Sihag et al. (2018c) | SVR and GP | Recharging rate of stormwater filter system (laboratory) | GP |
| 18      | Sihag et al. (2018b) | SVR, MLR, MSP, and GRNN | Cumulative infiltration (laboratory) | SVR |
| 19      | Sihag et al. (2018a) | ANN, GP, GRNN, and GEP | Infiltration rate (field) | ANN |
| 20      | Sihag (2018) | Fuzzy logic and ANN | Unsaturated hydraulic conductivity (laboratory) | ANN |
| 21      | Vand, Sihag, Singh, & Zand (2018) | GP, SVR, and MLR | Infiltration rate (field) | SVR |
| 22      | Angelaki, Singh Nain, Singh, & Sihag (2018) | SVM, ANN, and ANFIS | Cumulative infiltration (laboratory) | ANFIS |
| 23      | Sihag, Singh, Gautam, & Debnath (2018d) | ANN, SVM, GP, RF and MSP | Cumulative infiltration (laboratory) | SVM |
| 24      | Sihag, Tiwari, & Ranjan (2019a) | MSP, Bagged MSP, RF, and Bagged RF | Cumulative infiltration (laboratory) | Bagged MSP |
| 25      | Singh, Sihag, Pandhiani, Debnath, & Gautam (2019a) | SVM, GP, RF, and MLR | Permeability of the soil (laboratory) | SVM |
| 26      | Sihag, Karimi, & Angelaki (2019b) | RF, MSP, and MLR | Unsaturated hydraulic conductivity (field) | RF |
| 27      | Gautam, Sihag, Tiwari, & Ranjan (2019) | ANFIS | Cumulative infiltration of the soil (laboratory) | ANFIS |
| 28      | Singh, Sihag, & Deswal (2019b) | SVM, GP, MSP, and MLR | Infiltration rate (laboratory) | MSP |
| 29      | Sihag, Esmaeilebiki, Singh, Eftehah, & Bonakdari (2019c) | ANFIS, ANFIS-FDA, and ANFIS-PSO | Unsaturated hydraulic conductivity (field) | ANFIS-FDA and ANFIS-PSO |
| 30      | Kumar & Sihag (2019) | RF and ANFIS | Infiltration rate (field) | RF |
| 31      | Patle, Sikar, Rawat, & Singh (2019) | MLR | Infiltration rate (field) | MLR |
better than the other applied modelling methods, as well as other two popular conventional models (Kostiakov and Philip). The performance of Polynomial kernel is found best with both SVM and GP regression, as compared to RBF kernel. So this study and some past studies (Kumar & Sihag, 2019; Sihag et al., 2019; Singh et al., 2017) affirm that RF model is a reasonably good predictor of infiltration characteristics in the field as well as in laboratory and can be successfully employed in both cases.

### 8. Sensitivity study

Sensitivity study was carried out to find the most influencing input variable for the infiltration rate of soil. For this investigation, the best performing model was selected (GP_Poly). The number of training datasets was prepared by eliminating a single input variable at a time and outcomes were listed in the form of CC and RMSE for the test dataset. The higher variation in the value of RMSE observed from Table 7 concludes that time is the most influencing variable for estimating the infiltration rate of soil.

### 9. Conclusion

This paper inspects the performance comparison of Kostiakov model, Philip’s model, GP, SVM, ANN, and RF approaches in approximating the infiltration rate of water through soil. Results of performance assessing parameters suggest that Polynomial kernel-based GP approach works superior to conventional models, SVM, ANN, and RF models with coefficient of correlation values as 0.9863, Bias values as –2.3542, root-mean-square error values as 40.3026 and Nash–Sutcliffe model efficiency values as 0.9727 using testing dataset. One of the most imperative conclusions is that Polynomial kernel Function works better than Radial basis kernel function with both GP and SVM approaches. RF regression is found as second best modelling alternative after GP_Poly regression. All the artificial intelligence techniques work superior to conventional models. ANNOVA single factor outcomes suggest that the variation in the predicted values by artificial intelligence techniques is not considerable. Results of sensitivity study conclude that the most influencing input variable is time for estimating the infiltration rate of soil for this dataset.

### Compliance with Ethical Standards

Conflict of Interest: no conflict of interest: Parveen Sihag, Munish Kumar, and Balraj Singh declared that there is no conflict of interest.

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

### Disclosure statement

No potential conflict of interest was reported by the authors.

### ORCID

Parveen Sihag [http://orcid.org/0000-0002-7761-0603](http://orcid.org/0000-0002-7761-0603)
Munish Kumar [http://orcid.org/0000-0002-2995-0216](http://orcid.org/0000-0002-2995-0216)
Balraj Singh [http://orcid.org/0000-0002-0381-4363](http://orcid.org/0000-0002-0381-4363)

### References

Aggarwal, S. K., Goel, A., & Singh, V. P. (2012). Stage and discharge forecasting by SVM and ANN techniques. *Water Resources Management*, 26(13), 3705–3724.

Aghelpour, P., Mohammadi, B., & Biazar, S. M. (2019). Long-term monthly average temperature forecasting in some climate types of Iran, using the models SARIMA, SVR, and SVR-FA. *Theoretical and Applied Climatology*, 138, 1471–1480.

Al-Sulaiman, M. A. (2015). Applying of an adaptive neuro fuzzy inference system for prediction of unsaturated soil hydraulic conductivity. *Biosciences, Biotechnology Research Asia*, 12(3), 2261–2272.

Al-Sulaiman, M. A., & Aboukarima, A. M. (2016). Distribution of natural radionuclides in the surface soil in some areas of agriculture and grazing located in west of Riyadh, Saudi Arabia. *Journal of Applied Life Sciences International*, 7(2), 1.

Anari, P. L., Darani, H. S., & Nafarzadegan, A. R. (2011). Application of ANN and ANFIS models for estimating total infiltration rate in an arid rangeland ecosystem. *Research Journal of Environmental Sciences*, 5(3), 236.

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*, 1–8. doi.org/10.1080/09715010.2018.1531274

Arshad, R. R., Sayyad, G., Mosaddeghi, M., & Gharabaghi, B. (2013). Predicting saturated hydraulic conductivity by artificial intelligence and regression models. *ISRN Soil Science*, 2013. doi.org/10.1155/2013/308159

Breiman, L. (1999). Random forests - Random Features. *Technical Report 567*, Statistics Department, University of California, Berkeley.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.

Das, S. K., Samui, P., & Sabat, A. K. (2011). Prediction of field hydraulic conductivity of clay liners using an
artificial neural network and support vector machine. *International Journal of Geomechanics*, 12(5), 606–611.

Devices, D. (2006). *Mini disk infiltrometer user's manual*. Pullman, WA: Decagon Devices.

Elbisy, M. S. (2015). Support Vector Machine and regression analysis to predict the field hydraulic conductivity of sandy soil. *KSCE Journal of Civil Engineering*, 19(7), 2307–2316.

Erzin, Y., Gumaste, S. D., Gupta, A. K., & Singh, D. N. (2009). Artificial neural network (ANN) models for determining hydraulic conductivity of compacted fine-grained soils. *Canadian Geotechnical Journal*, 46(8), 955–968.

Esmaeilejad, L., Ramezanpour, H., Seyedmohammadi, J., & Shabanpour, M. (2015). Selection of a suitable model for the prediction of soil water content in north of Iran. *Spanish Journal of Agricultural Research*, 13 (1), 1202.

Gautam, S., Sihag, P., Tiwari, N. K., & Ranjan, S. (2019). Neuro-Fuzzy approach for predicting the Infiltration of soil. In Agnihotri A., Reddy K., Bansal A. (eds). *Environmental Geotechnology* (pp. 221–228). Singapore: Springer.

Ghorbani, M. A., Khatibi, R., Goel, A., FazeliFard, M. H., & Azani, A. (2016). Modeling river discharge time series using support vector machine and artificial neural networks. *Environmental Earth Sciences*, 75(8), 1–13.

Green, W. H., & Ampt, G. (1911). Studies on soil physics, 1. The flow of air and water through soils. *The Journal of Agricultural Science*, 4, 1–24.

Haykin, S. (1999). *Neural Networks: A comparative Foundation*. Macmillan Publishing, Prentice Hall New York, USA.

Holman, H. N. (1961). Concept for infiltration estimates in watershed engineering. URS41–41.US. Washington, D.C.: US Department of Agriculture Service.

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

Jahani, B., & Mohammadi, B. (2019). A comparison between the application of empirical and ANN methods for estimation of daily global solar radiation in Iran. *Theoretical and Applied Climatology*, 137(1–2), 1257–1269.

Kia, M. B., Pirasteh, S., Pradhan, B., Mahmud, A. R., Sulaiman, W. N. A., & Moradi, A. (2012). An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. *Environmental Earth Sciences*, 67(1), 251–264.

Kostiakov, A. N. (1932). On the dynamics of the coefficients of water percolation in soils. *In Sixth Commission, International Society of Soil Science, Part A*, 1.15–21.

Kumar, M., & Sihag, P. (2019). Assessment of infiltration rate of soil using empirical and machine learning-based models. *Irrigation and Drainage*, 68, 588–601.

Kumar, M., Tiwari, N. K., & Ranjan, S. (2018). Prediction of oxygen mass transfer of plunging hollow jets using regression models. *ISH Journal of Hydraulic Engineering*, 1–8. doi:10.1080/09715010.2018.1435311

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

Moazenzadeh, R., & Mohammadi, B. (2019). Assessment of bio-inspired metaheuristic optimisation algorithms for estimating soil temperature. *Geoderma*, 353, 152–171.

Nash, J. E., & Sutcliffe, J. V. (1970). River forecasting using conceptual models: part 1-a discussion of principles. *Journal Of Hydrology*, 10(3), 280–290. doi:10.1016/0022-1694(70)90255-6

Ogbe, V. B., Javeoba, O. J., & Ode, S. O. (2011). Comparison of four soil infiltration models on a sandy soil in Lafia, Southern Guinea Savanna Zone of Nigeria. *Production Agriculture and Technology*, 7(2), 116–126.

Patle, G. T., Sikar, T. T., Rawat, K. S., & Singh, S. K. (2019). 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: 1. The infiltration equation and its solution. *Soil Science*, 83, 435–448.

Rahmati, M. (2017). Reliable and accurate point-based prediction of cumulative infiltration using soil readily available characteristics: A comparison between GMDH, ANN, and MLR. *Journal of Hydrology*, 551, 81–91.

Rasmussen, C. E. (2006). *Gaussian processes for machine learning*. Cambridge: Journal of Machine Learning Research, MIT Press.

Rawls, W. J., Brakensiek, D. L., & Miller, N. (1983). Green-Ampt infiltration parameters from soils data. *Journal of Hydraulic Engineering, 109*(1), 62–70.

Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Physics*, 1(5), 318–333.

Schaap, M. G., & Leij, F. J. (1998). Using neural networks to predict soil water retention and soil hydraulic conductivity. *Soil and Tillage Research*, 47(1–2), 37–42.

Sedaghat, A., Bayat, H., & Sinegani, A. S. (2016). Estimation of soil saturated hydraulic conductivity by artificial neural networks ensemble in smectitic soils. *Eurasian Soil Science*, 49(3), 347–357.

Sihag, P. (2018). Prediction of unsaturated hydraulic conductivity using fuzzy logic and artificial neural network. *Modeling Earth Systems and Environment*, 4(1), 189–198.

Sihag, P., Esmaeilebeiki, F., Singh, B., Ebtehaj, I., & Bonakdari, H. (2019c). Modeling unsaturated hydraulic conductivity by hybrid soft computing techniques. *Soft Computing*, 23, 12897–12910.

Sihag, P., Jain, P., & Kumar M. (2018c). Modelling of impact of water quality on recharging rate of storm water filter system using various kernel function based regression. *Modeling Earth Systems and Environment*, 4(1), 61–68. doi:10.1007/s40808-017-0410-0

Sihag, P., Karimi, S. M., & Angelaki, A. (2019b). Random forest, M5P and regression analysis to estimate the field unsaturated hydraulic conductivity. *Applied Water Science*, 9(5), 129.

Sihag, P., Singh, B., Gautam, S., & Debnath, S. (2018d). 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., Vand, A. S., & Mehdipour, V. (2018a). Modeling the infiltration process with soft computing techniques. *ISH Journal of Hydraulic Engineering*, 1–15. doi:10.1080/09715010.2018.1464408

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

Sihag, P., Tiwari, N. K., & Ranjan, S. (2017a). Modelling of infiltration of sandy soil using gaussian process
regression. *Modeling Earth Systems and Environment*, 3(3), 1091–1100.

Sihag, P., Tiwari, N. K., & Ranjan, S. (2017b). Modelling of infiltration of sandy soil using gaussian process regression. *Modeling Earth Systems and Environment*, 3(3), 1091–1100.

Sihag, P., Tiwari, N. K., & Ranjan, S. (2018b). Support vector regression-based modeling of cumulative infiltration of sandy soil. *ISH Journal of Hydraulic Engineering*, 26(1), 44-50.

Sihag, P., Tiwari, N. K., & Ranjan, S. (2019). Prediction of unsaturated hydraulic conductivity using adaptive neuro-fuzzy inference system (ANFIS). *ISH Journal of Hydraulic Engineering*, 25(2), 132–142.

Sihag, P., Tiwari, N. K., & Ranjan, S. (2019a). Prediction of cumulative infiltration of sandy soil using random forest approach. *Journal of Applied Water Engineering and Research*, 7(2), 118–142.

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

Singh, B., Sihag, P., Pandhiani, S. M., Debnath, S., & Gautam, S. (2019a). Estimation of permeability of soil using easy measured soil parameters: Assessing the artificial intelligence-based models. *ISH Journal of Hydraulic Engineering*, 1–11, doi.org/10.1080/09715010.2019.1574615.

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.

Skaggs, R. W., Huggins, L. E., Monke, E. J., & Foster, G. R. (1969). Experimental evaluation of infiltration equations. *Transactions of the ASAE*, 12(6), 822–828.

Smith, R. E. (1972). The infiltration envelope: Results from a theoretical infiltrometer. *Journal of Hydrology*, 17, 1–21.

Smola, A. J. (1996). *Regression estimation with support vector learning machines* (Doctoral dissertation, Master’s thesis). Technische Universität München.

Soil Conservation Service (SCS). (1972). *National engineering handbook, Section 4: Hydrology*. Washington DC: Department of Agriculture.

Sy, N. L. (2006). Modelling the infiltration process with a multi-layer perceptron artificial neural network. *Hydrological Sciences Journal*, 51(1), 3–20.

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

Vapnik, V. N. (1998). *Statistical learning theory* (Vol. 3). New York: Wiley.

Whisler, F. D., & Bower, H. (1970). Comparison of methods for calculating vertical drainage and infiltration for soil. *Journal of Hydrology*, 10, 12–19.

Yılmaz, I., Marschalko, M., Bednarik, M., Kaynar, O., & Fojtova, L. (2012). Neural computing models for prediction of permeability coefficient of coarse-grained soils. *Neural Computing and Applications*, 21(3), 957–968.