J. Indian Water Resour. Soc.,
Vol. 38, No. 3, July, 2018

PREDICTION OF PERMEABILITY OF SOIL USING RANDOM TREE

Parveen Sihag[1], Siraj Muhammed Pandhiani[2], Varun Singh[3], Sourav Debnath[4]

ABSTRACT

The permeability of the soil is an essential parameter for hydrological modeling and groundwater recharge related studies. In this research, the permeability of soil was predicated using Random tree (RT) approach. To compare the performance of RT with other types of soft computing techniques, adaptive neuro-fuzzy inference systems (ANFIS) and artificial neural network (ANN) prepared as well. The dataset was collected by falling head permeability test for different percentage of sand and fly ash mixture. Total 95 data collectedthrough this experiment out of which 66 data used for training, and 29 data used for testing. There are five input variable Percentage of sand (S), the percentage of fly ash (Fa), specific gravity (G), time (T) and Head (H). The output variable was permeability of soil (k). Results of RT, ANN and ANFIS showed that all these models have a high ability for modelling and predicting permeability of soil; however, RT is a bit more precise. The sensitivity analysis of RT showed that the percentage of sand (S) and time (T) are the most effective parameters in the discharge coefficient.

Keywords: *Permeability of soil; falling head permeability test; Random tree; adaptive neuro-fuzzy inference systems; artificial neural network*

INTRODUCTION

Soils are permeable materials due to presence of voids in it, which allow the flow of water from higher energy to lower energy. The permeability of the soil is one of the critical engineering properties in various soil related problems like as yield of wells, cutoff wall, seepage through the dam, diaphragm wall, seepage through and below the earth structures, etc. Exact knowledge of permeability of the soil is essential for seepage under and through dams and other hydraulic structure related problems. The permeability of the soil is influenced by various factors like the texture of the soil, voids ratio, type of soil, the density of soil, impurities in soil and degree of saturation, etc. Direct measurement of permeability of the soil is very laborious, time-consuming and challenging task for engineers in the field. The variation of coefficient of permeability is up to 10 orders of magnitude from coarse to very fine-grained soils (Mitchell,1993). The earlier study on the coefficient of permeability denotes that the coefficient of permeability is exceptionally variable for similar soil (Nagy et al., 2013). The permeability of the soil is measured using laboratory constant and falling head methods, which are simple to perform. However, it is very tedious and laborious process to execute in undisturbed samples from sandy soil deposits. On the other hand, indirect methods were also established for prediction of permeability of soil using efficiently computable properties. Several analytical models are suggested in the literature to find out the permeability of soil (Hazan, 1892; Kozeny, 1927; Kenney et al., 1984; Rosas et al., 2014). Theoretical approaches proposed by (Terzaghi and Peck, 1964; Alyamani and Sen, 1993; Li and Zand, 2011; Ranaivomanana et al., 2017) are quite complicated. A few researchers have created models utilizing prescient techniques for assessing the permeability of soil from effortlessly quantifiable properties of soils. The forecasting methods for the estimation of the permeability of soil are developed by various methods such as ANN, support vector machine (SVM). These techniques have suggested for estimation of permeability of soil as encouraging results achieved by various past studies (Gupta and Chitra, 2015; Erzin et al., 2009; Dolzyk and Chmielewska, 2014; Yan et al., 2017).Traditional modeling techniques are based on empirical relationships developed from the experimental data.

During last few years, researchers have investigated the potential of artificial intelligence techniques such as artificial neural network, Support vector machine, Gaussian process, Fuzzy Logic, M5P model tree, Generalized neural network, ANFIS, Random forest, etc. (Sihag 2018; Sihag et al., 2018a; Sihag et al., 2018b) in various problems in the field of civil engineering. Best of author's knowledge no one use Random forest technique for the estimation of permeability of soil. The objective of the paper is to examine the capability of ANN, Random tree and ANFIS for estimation of the permeability of sandy soil mixed with fly ash.

MATERIAL AND METHODOLOGY

Locally available sand was collected from the campus of National Institute of Technology Kurukshetra, Kurukshetra, Haryana, India, and fly ash was collected from Panipat thermal power plant, Panipat, Haryana, India. To check the effect of fly ash on the permeability of sand various amount of fly ash was added at different replacement ratios. Total 21 samples were prepared for testing. Each soil sample has undergone different laboratory test.

I) Specific Gravity test was done according to IS: 2720 (part 3)-1980.

II) Falling head permeability test was done according to IS: 2720 (part 17)-1986

III) Relative density test was conducted according to IS: 2720 (part14)-1983.

I) Specific Gravity test: The objective of this test is to find out the specific gravity of the sample. Specific gravity is the ratio between weights of soil in the air to the weight of the equal amount of water at 27 degrees Celsius. By this specific

Email (corresponding Author): parveen12sihag@gmail.com

1 *National Institute of Technology, Kurukshetra, India*

2 *University College Jubail. Saudi Arabia*

3 *National Institute of Technology, Kurukshetra, India*

4 *National Institute of Technology, Kurukshetra, India*

Manuscript No. : 1488

gravity, we can also find the density of soil. The pycnometer is used to find out the specific gravity. This test is conducted according to IS: 2720 (part 3)-1980 guideline.

II) Falling Head Permeability test: Falling head permeability test was conducted according to

IS: 2720 (part 17)-1986.The falling head permeability test (Fig 1) permits the water flowing through a small soil sample attached to a graduated standpipe providing a measurement of the water head along with the volume of water passing through the sample. We can find the permeability of the sample by the following equation

$$k = \frac{aL}{At} \ln \left(\frac{h1}{h2}\right) \quad (1)$$

Where:

k= Coefficient of permeability in cm/s

a= Area of falling head tube in cm^2

A= Area of the specimen in cm^2

L= Length of the specimen in cm

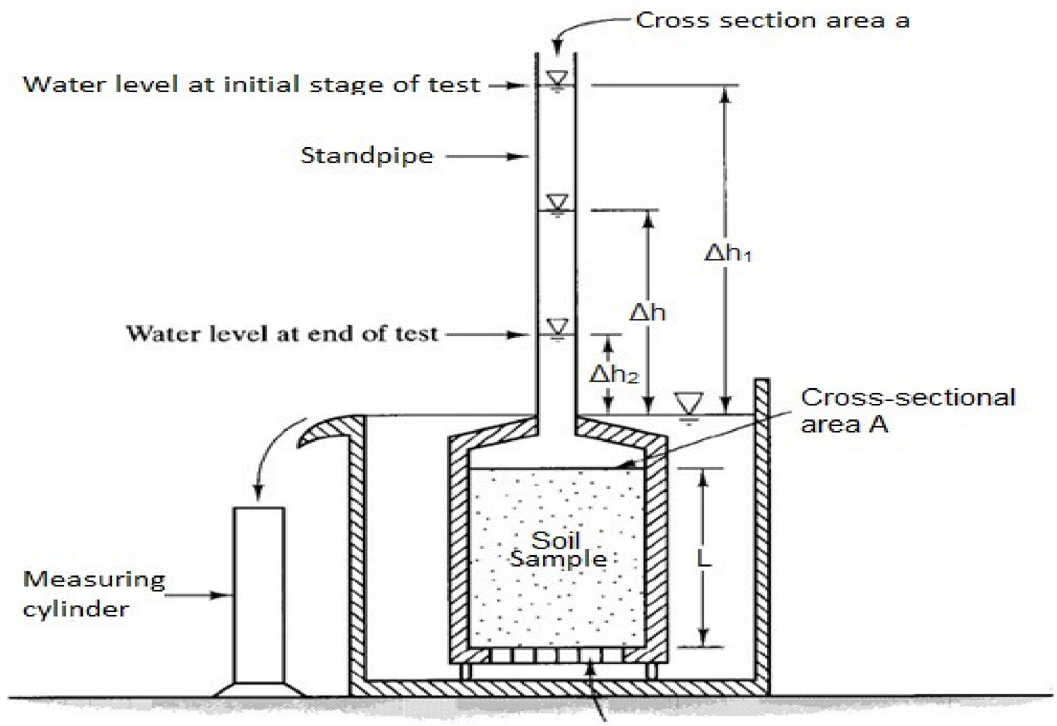


Figure 1. Falling Head Apparatus

Property of the Material:

The property of Sand and fly ash are given in Table 1 and Table 2 respectively.

Table 1:Property of Sand

Sr. No.	Property	Value
1	Effective size (D_{10})	0.190 mm
2	D_{60}	0.450 mm
3	D_{30}	0.300 mm
4	The coefficient of Uniformity (C_u)	2.368
5	The coefficient of curvature (C_c)	1.05
6	Type of soil	SP
7	Specific gravity	2.65

Table 2. Property of Fly ash

Sr No.	Property	Value
1	Specific Gravity	2.1
2	C_u	2.73
3	C_c	0.9056
4	Color	Gray

ANFIS:

ANFIS is one of the most critical soft computing methods. It is a combination of Sugeno fuzzy inference model with ANN,so it gave the more precise results as compared to the ANN and Sugeno fuzzy inference (Sihag et al., 2017). ANFIS approach based on adaptive and non-adaptive nodes in different layers (Singh et al., 2018). Figure 2 shows the structure of the ANFIS model (first-order Sugeno fuzzy model) having 2 inputs (a & b), 4 rules and one target (c). The first order Sugeno type is implemented to develop two if-then rules as follows:

- **Rule 1** If a is X1 and b is Y1, then $c1 = m_1\,a + n_1\,b + p_1$,
- **Rule 2** If a is X2 and b is Y2, then $c2 = m_2\,a + n_2\,b + p_2$,

Figure 2 has five layers; every layer executes a unique role explained below:

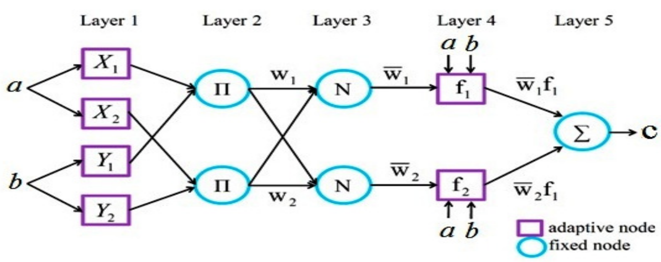


Figure 2: Structural plan of first-order Sugeno fuzzy model.

First layer

The first layer is used to select the input parameters. In the ANFIS approach, the input parameters were introduced as adaptive nodes. The membership functions for the adaptive nodes defined as follow,

$$M_{1i} = \mu_{Xi}(a), \qquad i=1,2 \tag{2}$$

$$M_{2i} = \mu_{Yi}(b), \qquad i=1,2 \tag{3}$$

Where,

M_{1i} & M_{2i} is the fuzzy membership grades

X_i & Y_i is the fuzzy sets

μ_{Xi} & μ_{Yi} is the fuzzy MFs,

The range of fuzzy MFs 0 to 1 as follows:

$$\mu_{Xi}(a) = \frac{1}{1+\left|\frac{a-t_i}{r_i}\right|^{s_i}} \tag{4}$$

Where,

r_i, s_i and t_i are the nonlinear parameters in the form of membership functions on linguistic labels.

Second layer

Fixed number of nodes contained by the second layer. The multiplication function is defined as follow,

$$wki = \mu_{Xi}(a) \times \mu_{Yi}(b), i=1,2 \tag{5}$$

Third layer

The third layer also contains fixed nodes. The third layer contains summation function is as follow,

$$\overline{wk_i} = \frac{wk_i}{wk_1 + wk_2}, \quad i=1,2 \tag{6}$$

Fourth layer

The fourth layer consists of adaptive nodes and is known as the functional layer. The functional layer is defined by,

$$\overline{wk_i}f_i = wk_i(m_i a + n_i b + p_i), \quad i=1,2 \tag{7}$$

m_i, n_i and p_i are the characteristics of the first-order polynomial.

Target layer

The target layer consists of only one node. This layer calculates the sum of all the input parameters calculated at various stages. The target layer is defined by,

$$c = \sum wk_i f_i = \frac{\sum_i wk_i f_i}{\sum_i wk_i} \qquad i=1,2 \tag{8}$$

Sugeno type ANFIS model is used in this paper to model the punching load of slabs. Four types of Membership functions (MFs) such as trapezoidal, triangular, generalized bell-shaped, Gaussian functions are used in this paper.

Artificial neural networks(ANN):

The artificial neural network (ANN) is an artificial intelligence based approach generally used for the exact forecast of civil engineering problems (Nain et al., 2018; Sihag et al., 2018c). ANN is a parallel knowledge processing system containing a set of neurons arranged in layers. It contains an input layer, hidden layers and at last target layer. The target layer is the ultimate processing part. The neurons are linked by weight in every layer of the neurons in a successive layer during the learning process. For further information, readers may refer to Haykin (2010). In the current study, one hidden layer is used in the ANN model.

Random Tree (RT):

The random tree is a tree that is produced by a stochastic method. With k arbitrary features at every node, a Random tree is a tree grown at arbitrary from a set of possible trees. In the randomtree, each tree has an equal possibility of being sampled. The distribution of trees is uniform. Random trees can be developed economically, and the mixture of big sets of arbitrary trees generally leads to precise models. Random tree models have been extensively developed in the field of Machine Learning in the last few years.

DATA SET:

The data set selected for the RT, ANN and ANFIS models were generated from the experiments conducted in soil mechanics laboratory of N.I.T Kurukshetra, as summarized in Table 3. The objective of the model to estimate the k (coefficient of permeability) of the sand fly ash mixture depends upon the degree of the train data set. Out of 95 data as many as (66) datasets used as training data and remaining, 29 were used for testing data. Input variables include the percentage of fly ash (Fa), the percentage of sand (S), specific gravity (G), time (T) and head (H) whereas the coefficient of permeability (k) is considered as output. The range of different input parameters applied in RT, ANN, and ANFIS techniques are listed in Table 3.

Table 3: Features of the dataset

Parameters	**Units**	**Training dataset**					
		Minimum	**Maximum**	**Mean**	**St Dev.**	**Kurtosis**	**Skewness**
S	%	5	95	48.9394	27.9493	-1.2434	0.0399
Fa	%	5	95	51.0606	27.9493	-1.2434	-0.0399
G	-	2.1275	2.6225	2.3692	0.1537	-1.2434	0.0399
T	min.	5	150	23.7121	29.2116	11.2086	3.3058
H	cm	104.1	172.2	138.7167	17.3601	-0.8027	-0.3425
k	m/day	0.0106	0.7430	0.1951	0.1094	8.6869	1.6554
		Testing dataset					
S	%	5	95	52.4138	26.8800	-1.0885	-0.0793
Fa	%	5	95	47.5862	26.8800	-1.0885	0.0793
G	-	2.1275	2.6225	2.3883	0.1478	-1.0885	-0.0793
T	min.	5	120	21.0345	24.6178	10.9885	3.2604
H	cm	104	163.4	140.1690	16.4472	-0.7402	-0.3101
k	m/day	0.0118	0.2981	0.1864	0.0805	-0.0610	-0.8020
		Total dataset					
S	%	5	95	50.0000	27.5314	-1.2068	0.0000
Fa	%	5	95	50.0000	27.5314	-1.2068	0.0000
G	-	2.1275	2.6225	2.3750	0.1514	-1.2068	0.0000
T	min.	5	150	22.8947	27.7870	11.1005	3.2939
H	Cm	104	172.2	139.1600	17.0125	-0.7914	-0.3355
k	m/day	0.0106	0.7430	0.1925	0.1011	8.3311	1.3519

RESULT AND DISCUSSION

Results of ANFIS:

The Sugeno fuzzy rule-based ANFIS models have trial and error process. For this purpose, the collected data set, listed in Table 3, was implemented and arbitrarily separated into two parts of training and testing. Designing of the ANFIS model includes defining the number of hidden layer(s), neurons, number and shape of membership functions. In the study, the number of membership functions was included one by one to every input, and target parameters and then the ANFIS model was prepared and test. The training data set was about 70% of the total collected data, and the rest (30%) was used for testing. Results of the ANFIS model to estimate the permeability of soil are shown in Figure 3 and Table 4. As shown in Table 4, Gaussian Function(GAUSSMF) based ANFIS model works better than Triangular MFs based ANFIS model with CC=0.9034 and RMSE = 0.0382 cm.

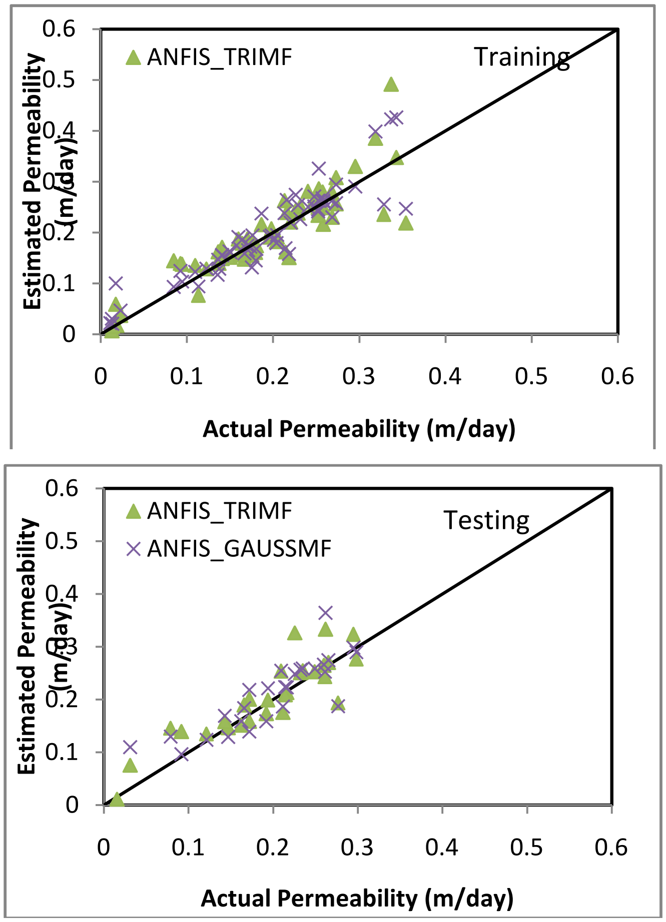


Figure 3: Performance of ANFIS models

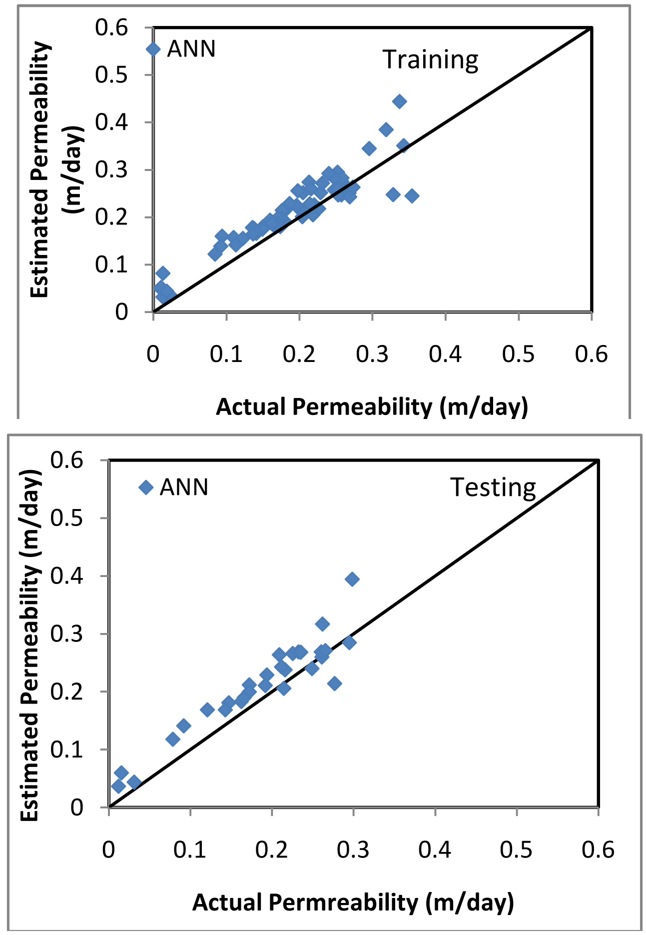


Figure 4: Performance of ANN models

Table 4: Performance of soft computing models

Approaches	Training		Testing	
	C.C.	R.M.S.E.	C.C.	R.M.S.E.
ANN	0.9008	0.0521	0.9083	0.0376
Random Tree	0.9999	0.0015	0.9125	0.0356
ANFIS_TRIMF	0.8705	0.0534	0.9025	0.0374
ANFIS_GAUSSMF	0.8718	0.0532	0.9035	0.0382

Table 5: Primary parameters

Approaches	Optimum Parameters
ANN	Hidden Layer=7, L=0.4, m=0.2
Random Tree	k=10, m=1

Results of ANN:

Developing the ANN model is also a trial and error process, it includes a number of neurons in hidden layer, number of hidden layers, momentum, learning rate, iteration, etc. primary parameters of ANN are listed in Table 5. Out of the data set, 70% was used for training, and the rest (30%) was considered for testing the model. In this study ANN model include single hidden layer with 3 neurons, momentum = 0.2, learning rate = 0.4 and iteration = 2000. The performance of the ANN model is shown in Figure 4 for training and testing stages. As shown in Table 4, the ANN model was obtained C.C. = 0.9083 and RMSE=0.0376 for testing stage. Generally examining Table 4 and Figure 4 implies that the accuracy of the ANN model was more suitable than ANFIS based models for estimation of Permeability of soil.

Results of Random Tree:

Random tree is a controlled Classifier; it is a collective learning algorithm that generates lots of individual learners. It engages a bagging idea to construct a random set of data for constructing a decision tree. In the standard tree, every node is split using the best split among all variables. In a random forest, every node is split using the best among the subset of predictors randomly chosen at that node. The algorithm can deal with both classification and regression problems. Of the total collected data, 70% was used for training, and the rest (30%) was used for testing. Detail of the primary parameters of RT is listed in Table 5. Results of Random tree-based models are shown in Figure 6 and Table 5. The RT-based model has suitable for estimation of the performance of permeability of soil with CC = 0.9124 and RMSE = 0.0355 cm.

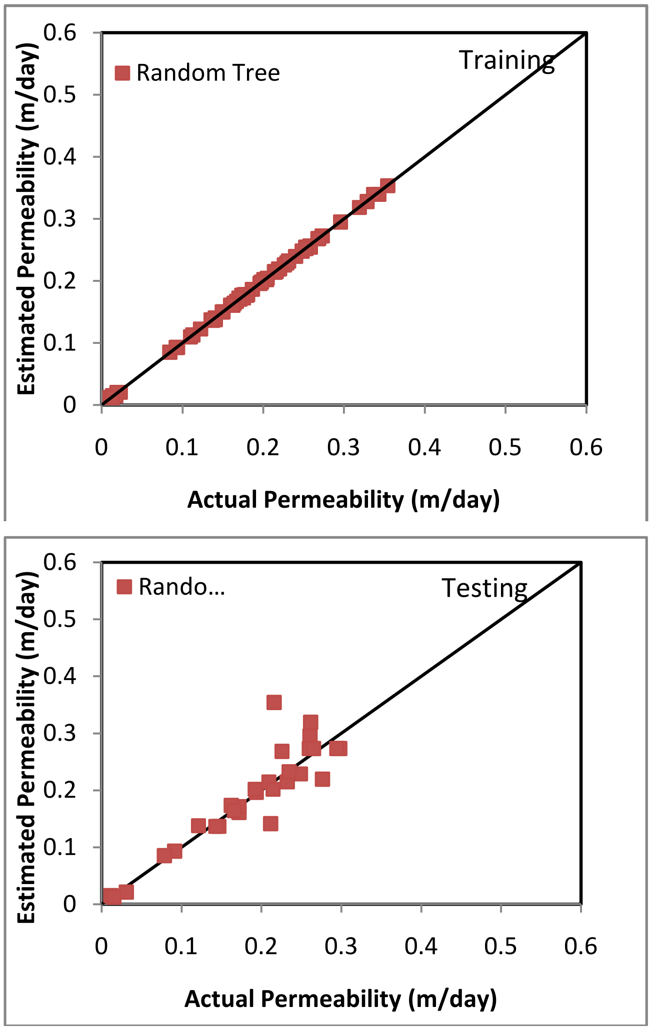
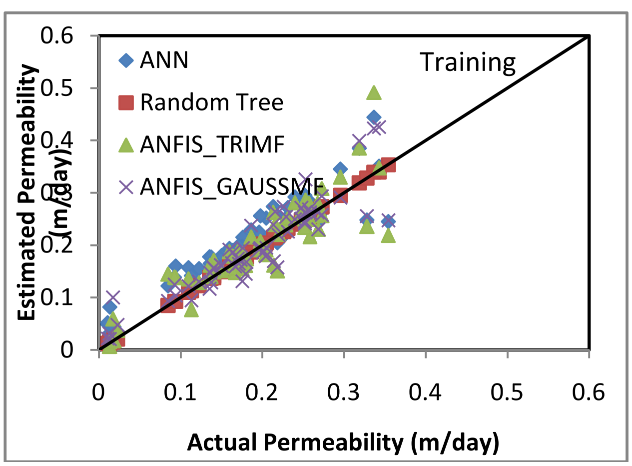


Figure 5: Performance of the RT-based model

Comparison of models:

Figure 6 shows the agreement plot of actual vs. predicted values using ANFIS, ANN, and RT-based models for training and testing stages. By and large examining, Figure 6 and Table 5 the RT-based model is more suitable than ANFIS and ANN-based models for predicting the permeability of the soil. The lower value of RMSE suggests that the RT model is better than other models.



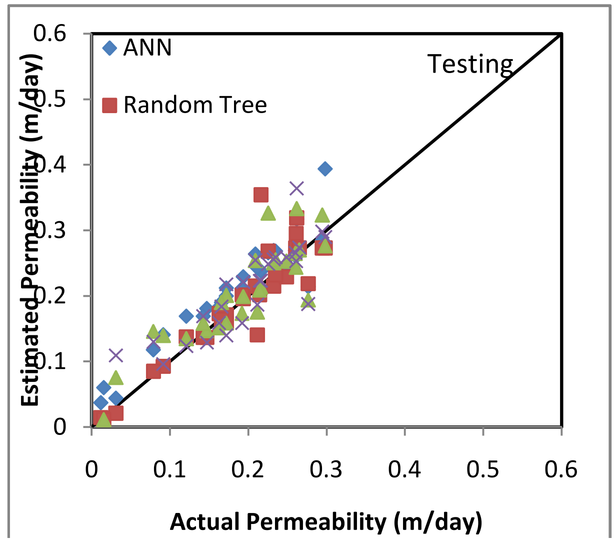


Figure 6: Performance of ANFIS, ANN,andRT-based models

Sensitivity study:

A sensitivity study was carried out to determine the mainly significant input parameter in the estimation of permeability of the soil. For this, the RT model was selected which is outperforming the above-discussed models. Several sets of training data were generated by eliminating the single input parameter at a time and outcomes were recorded in terms of CC and RMSE with testing data set. Results from Table 5 suggest that the percentage of sand and time have an important role in estimating the permeability of soil in comparison to other input parameters.

Table 6: Sensitivity analysis using Random Tree

Input combination	**Removed parameter**	**Random Tree**	
		Coefficient of correlation	**Root mean square error (m/day)**
S, Fa, G, T, H		0.9121	0.0357
Fa, G, T, H	S	0.8534	0.0464
S, G, T, H	Fa	0.9121	0.0357
S, Fa, T, H	G	0.9121	0.0357
S, Fa, G, H	T	0.9367	0.0314
S, Fa, G, T	H	0.9361	0.0315

CONCLUSION

Estimation of the permeability of the soil is essential for the irrigation system, agriculture, and groundwater recharging related studies. In this study, the permeability of the soil was estimated using ANFIS, ANN and Random tree (RT). Outcomes of the RT suggested that percentage of sand and time are the most critical parameter on the permeability of the soil. Results of the present study suggest that the performance of the RT-based model is better than ANN and ANFIS based models. Based on the obtained results, the RT model has a suitable capability to estimate the permeability of the soil. The ANN also provides better performance than ANFIS based models.

REFERENCES

1. Mitchell,1993. Fundamentals of Soil Behavior, second ed., John Wiley and Sons Inc., New York, 456 pages.
2. Nagy L, Tabacks A, Huszak T, Mahler A, Varga G., 2013. In: Proceedings of the 18th International Conference on SoilMechanics and Geotechnical Engineering, Paris, pp399–402
3. Hazan, A, 1892. Some physical properties of sands and gravels: Mass. State Board of Health, Ann. Report, pp 539-556.
4. Kozeny J., 1927. Uberkapillareleitung des wassersimboden: Sitzungsber. Acad. Wiss. Wien, Vol. 136, pp 271-306.
5. Kenney TC, Lau D, OfoegbuGI, 1984. Permeability of compacted granular materials: Canadian Geotechnical Journal, Vol. 21, pp 726-729.
6. Rosas J, Lopez O, Missimer TM, Coulibaly KM, Dehwah AH, Sesler K, Lujan LR, Mantilla D, 2014. Determination of hydraulic conductivity from grain-size distribution for different depositional environments. *Groundwater*, *52*(3), pp399-413.
7. Terzaghi K, Peck RB, 1964. Soil Mechanics in Engineering Practice: John Wiley and Son, New York.
8. Alyamani MS, Sen Z., 1993. Determination of hydraulic conductivity from grain size distribution curves: Ground Water, Vol. 31, pp 551-555.
9. Li X, ZangY., 2011 Study on permeability test and curve prediction of expansive soil [C]// The 2nd International Conference on Mechanic Automation and Control Engineering, MACE 2011-Proceedings. IEEE Computer Society, 8(7), pp 6900–6903.
10. Ranaivomanana H, Razakamanantsoa A, Amiri O, 2017. Permeability Prediction of Soils Including Degree of Compaction and Microstructure, In:Int. J. Geomech, ASCE, 04016107, pp 1-11.
11. Gupta M, Chitra R., 2015. Artificial Neural Networks for assessing permeability characteristics of soilsInternational Journal of Engineering Sciences & Research Technology, pp 338-346.
12. ErzinY,Gumaste SD, Gupta AK, Singh DN, 2009.Artificial neural network (ANN) models for determining hydraulic conductivity of compacted fine-grained soils. Canadian Geotechnical Journal, 46(8), pp955-68.
13. Dolzyk K, Chmielewska I, 2014. Predicting the coefficient of permeability of non-plastic soils, In: Soil Mechanics and Foundation Engineering, 51(5), pp213-218.
14. Yan WM, Chiu CF, Yuen KV, 2017. Prediction and modeling of permeability function and its application to the evaluation of breakthrough suction of a two-layer capillary barrier. Canadian Geotechnical Journal, 54(6), pp778-88.
15. Sihag P, 2018. Prediction of unsaturated hydraulic conductivity using fuzzy logic and artificial neural network. *Modeling Earth Systems and Environment*, pp1-10.
16. Sihag P, Jain P, Kumar M, 2018a. Modeling of impact of water quality on recharging rate of stormwater filter system using various kernel function based regression *Modelling Earth Systems and Environment*, pp1-8.
17. Sihag P, Tiwari NK and Ranjan S, 2018b. Support vector regression-based modeling of cumulative infiltration of sandy soil. *ISH Journal of Hydraulic Engineering*, pp1-7.
18. IS 2720 (Part 3), 1980. Methods of test for soils: determination of specific gravity. Section 1: Fine-grained soils.
19. IS: 2720 (Part 17), 1986. Methods of test for soils, laboratory determination of permeability.
20. Sihag P, Tiwari NK, Ranjan S, 2017. Prediction of unsaturated hydraulic conductivity using adaptive neuro-fuzzy inference system (ANFIS). *ISH Journal of Hydraulic Engineering*, pp1-11.
21. Singh, B., Sihag, P., Singh, K. and Kumar, S., 2018. Estimation of trapping efficiency of vortex tube silt ejector. *International Journal of River Basin Management*, (just-accepted), pp.1-38.
22. Nain, S.S., Sihag, P. and Luthra, S., 2018. Performance Evaluation of Fuzzy-Logic and BP-ANN Methods for WEDM of Aeronautics Super Alloy. *Methods X*.
23. Sihag, P., Singh, B., Sepah Vand, A. and Mehdipour, V., 2018c. Modeling the infiltration process with soft computing techniques.*ISH Journal of Hydraulic Engineering*, pp.1-15.
24. Haykin, S., 2010. Neural networks: a comprehensive foundation, 1999. *Mc Millan, New Jersey*.