

ORIGINAL ARTICLE

Open Access



Impact of waste foundry sand on drainage behavior of sandy soil: an experimental and machine learning study

Ankit Kumar^{1*} and Aditya Parihar¹

Abstract

The study of drainage behavior is essential for using waste material in geotechnical applications. In this study, sandy soil was replaced with waste foundry sand (WFS) at an incremental interval of 20% by weight. Permeability (k) for each mix was acquired at three relative densities (R_D), i.e., 65%, 75% and 85%, by using the constant head method. Then the results were further processed with machine learning (ML) models to validate the experimental data. The experimental study demonstrated that k would decrease with the increase in relative density and WFS content. A rise in R_D from 65% to 85% resulted in a substantial reduction of up to 140% in the value of k . Moreover, the complete replacement of sand with WFS reduced the value of k by 36%, 51% and 57% for R_D of 65%, 75% and 85%, respectively. The total dataset of 90 observations was divided at a ratio of 63/13/15 into training/validation/testing datasets for ML-AI modeling. Input variables include percentage of sand (BS), replacement with WFS, total head (H), time interval (t) and outflow (Q); and k is the output variable. The methods of artificial neural network (ANN), random forest (RF), decision tree (DT) and multi-linear regression (MLR) are used for k prediction. It is found that the random forest approach performed outstandingly in these methods, with an R^2 value of 0.9955. The performance of all the proposed methods was compared and verified with Taylor's diagram. Sensitivity analysis showed that Q and R_D were the most influential parameters for predicting k values.

Keywords Permeability, Waste foundry sand, Artificial intelligence, Random forest, Artificial neural network, Decision tree

1 Introduction

As a property of soils, permeability measures the speed of water percolation. Low permeability (k -value) means the possible generation of excess pore water pressure and secondary consolidation (Lin et al., 2018; Zhang et al., 2021). The measurements of permeability are essential for geotechnical projects related to groundwater tables or precipitation water. The k -value of soil depends upon many factors, including: grain size properties, void ratio,

fine content, over-consolidation ratio, drainage type, and density of soil or impurities, if any (Cashman & Preene, 2020; Smith, 2014).

The k -value can be identified in direct and indirect methods. For direct methods, tests are carried out on soil samples. In contrast, for indirect methods, the k -value is calculated from empirical formulas based on grain size properties and void ratio (Nagy et al., 2013; Osterhout, 1922). Some direct and indirect methods are listed in Table 1. It is not convenient to perform permeability tests on every soil sample. So, modeling is often used to provide a rough estimate of the actual k -value. The k -value exhibits up to 10 orders of magnitude, ranging from a coarse feature to a very fine feature.

*Correspondence:

Ankit Kumar
akumar_phd19@thapar.edu

¹ Department of Civil Engineering, Thapar Institute of Engineering and Technology, Patiala 147004, India

Table 1 Direct and indirect methods for measuring the *k*-value

| | | |
|------------------|---------------------------------------|--|
| Direct methods | Constant head method (Darcy, 1856) | $k = \frac{Q \times L}{h \times A}$ |
| | Failing head method (Darcy, 1856) | $k = \frac{a \times L}{A \times t} \times \log_e \frac{h_1}{h_2}$ |
| Indirect methods | Hazen equation (Hazen, 1911) | $k = c \times d_{10}^2$ |
| | Slichter equation (Slichter, 1899) | $k = \frac{771 \times d^2}{c}$ |
| | Taylor equation (Taylor, 1948) | $k = d_e^2 \times \frac{\gamma}{\mu} \times \frac{e^3}{1+e} \times c$ |
| | Terzaghi equation (Terzaghi, 1925) | $k = \frac{c}{\mu} \times \left[\frac{n-0.13}{\sqrt[3]{1-n}} \right]^2 \times d_{10}^2$ |
| | Kozenl-Carman equation (Kozeny, 1927) | $k = \frac{1}{k_0 S_g^2} \times \frac{\gamma}{\mu} \times \frac{e^3}{1+e}$ |

Proposed earlier based on some assumptions, indirect methods are suitable for different conventional soils. However, for soils incorporating waste materials, the values of constants would change significantly. So, if various waste materials or composites are present, new equations must be developed based on experimental data. This gap can be bridged by developing certain models with artificial intelligence techniques (Baghbani et al., 2022; Shahin, 2013).

Sand is a kind of natural materials used in the majority of construction projects. In India, the expected demand for sand is 700 MT (in 2017), with an annual growth rate of 6%–7%. Mining has led to a 90% decline in sediment levels in key Asian rivers, putting local communities into risks of flooding, land loss, contaminated drinking water, and crop devastation (Ministry of Mines, 2018). India produces more than 3 million tons of WFS annually. The scarcity of natural sand drives practitioners to replace them with waste and by-products from industries. The replacement materials should meet the strength and other requirements specified in construction laws and codes. One such material which has been used in construction over recent decades is WFS. Many researchers have reported the viability of WFS in various geotechnical applications (Gedik et al., 2018; Heidemann et al., 2021; Javed, 1995; Siddique & Singh, 2011; Sinha et al., 2020; Tittarelli, 2018; Winkler & Bolshakov, 2000), hydraulic or fluid barriers (Abichou et al., 1998, 2000, 2002), retaining-wall backfills (Lee et al., 2001), highway sub-bases (Guney et al., 2006; Javed & Lovell, 1994; Mast & Fox, 1998; Partridge et al., 1999) and ground improvement (Vipulanandan et al., 2000). For all such applications, knowledge on the drainage behaviour of WFS or WFS-incorporating sand is necessary. FHWA (2004) has reported a range of permeability 10^{-3} – 10^{-6} cm/s.

This study proposes a model to measure the permeability of soils incorporating a certain kind of industrial

materials. Sandy soil has been mixed with waste foundry sand at different ratios to cover the range of replacement, and the hydraulic behaviour at different densities has been observed. Few existing research is available that considers the relative density as a governing parameter in indirect methods. The main objective of this study is to explore the drainage behaviour of WFS-incorporated sand in retaining-wall backfills and earthen dam applications.

Sustainable geomaterials are rapidly replacing conventional materials. Due to their accuracy, soft computing methodologies have gained pervasive traction over the preceding two to three decades. Many models have been trained for evaluating material properties based on known input parameters (Dalkilic et al., 2023; Khatti & Grover, 2023a, 2023b, 2023c, 2023d, 2023e; Kumar et al., 2023; Länsivaara et al., 2023; Rabbani et al., 2023). Machine learning modeling can be broadly classified as an approximation, classification and forecasting (Sarker, 2021). This study will adopt artificial neural networks, multi-linear regression, decision trees, and random forest techniques. Several past studies on modelling of *k*-value by using AI techniques are listed in Table 2. In most cases, the input variable is the grain size properties of soil. The performance of models is dependent upon available data sets, correlation between variables, and standard deviation in values.

2 Materials used

2.1 Sand

Sandy soil samples acquired in Punjab, India is used in this study. The grain size properties are presented in Fig. 1 and Table 3. The type of gradation highly impacts the permeability of soil. The specific gravity of soil is determined as per ASTM D854 (2002), while GSD parameters are as per ASTM D422-63 (2007). The soil is classified as poorly graded sand.

Table 2 AI methods used in the past studies on *k*-value

| AI methods | Input variables | Model with highest performance | References |
|----------------------------|---------------------------------------|--------------------------------|---------------------------|
| KNN, SVM, GB, LightGBM, RF | $PI, w_l, w_p, CC, \rho, wc, e$ | GB | Tran (2022) |
| ANN, RF, SVM | $w_l, w_p, CC, \rho, wc, e$ | RF | Pham et al. (2021) |
| ANN, ANFIS | D_{10}, D_{30}, D_{60} | ANFIS | Yilmaz et al. (2012) |
| GPR (Poly, PUK, RBF) | $w_l, w_p, CC, \rho, wc, e$ | PUK | Ahmad et al. (2022) |
| TLBO-ANN | w_l, w_p, CC, G, wc, e | TLBO-ANN | Bui et al. (2022) |
| ANN, MEP, GEP | D, ρ_d | GEP | Rehman et al. (2022) |
| GP, RF, MLR, SVM | S, Fa, G, T, H | SVM | Singh et al. (2021) |
| CANFIS, MLP, SVM, DT, RF | OC, BD, PD | RF | Singh et al. (2020) |
| CART, GMDH | $w_l, w_p, CC, \rho, wc, e$ | GMDH | Torabi et al. (2022a) |
| RF-GA | UCS, Temp, Gp, σ' | RF-GA | Wang et al. (2020) |
| GPR, DT, RVM, ANN | FC, SC, LL, SG, PI, MDD, OMC | GPR | Khatti and Grover (2021a) |
| MLR | $e, D_{10}, D_{30}, D_{50}, C_u, C_c$ | MLR | Feng et al. (2023) |
| ANFIS, DNN, MR | e, γ, d | DNN | Kim and Song (2023) |
| ANN, MLR | FC, SC | MLR | Khatti and Grover (2021b) |
| SSA, EGME, BET, KC | e_{eff}, S_{eff} | KC | Chen et al. (2022) |
| GEP, ANN | $CC, \omega, LL, PL, \gamma, e$ | GEP | Torabi et al. (2022b) |

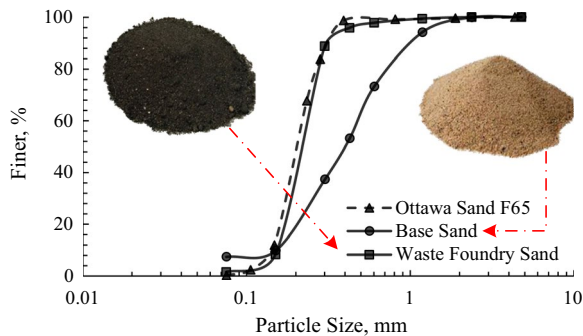


Fig. 1 Grain sizes' distribution curve of the materials used

2.2 Waste foundry sand

In this study, WFS is acquired from an iron foundry located in Ludhiana, India. The grain sizes are presented in Fig. 1. The WFS grain sizes' distribution curve is found similar to that of Ottawa sand at the F65 grade (Carey et al., 2020). The grains of WFS are smaller than those of sand, so the specific gravity is found to be smaller. The index and engineering properties for this type of WFS are reported by Kumar and Parihar (2022).

WFS is classified as poorly graded sand as per the USCS system (ASTM D2487, 2006; Casagrande, 1948).

3 Research methodology

This study is divided into four phases: data generation, data preparation, modeling, and getting conclusions based on the best-performed model (Fig. 2). The data in this study are generated from a series of laboratory scale tests. The experimental output data are further investigated for outliers and multicollinearity. The phases involved will be discussed in detail in the following sections.

3.1 Lab experiments

To cover a wide range of replacements with WFS and to explore the relative density, 18 different compositions are considered (Table 4). At least five distinct readings of permeability for each composition are measured. The time during which a particular amount of water is drained from the samples is also considered as one of the governing parameters.

Permeability is affected by the relative density of soil, as liquid would take more time to flow through denser

Table 3 Grain size properties of the materials used

| Material | Effective size, D_{10} (mm) | Mean size, D_{50} (mm) | C_u | C_c | Classification | G |
|----------|-------------------------------|--------------------------|-------|-------|----------------|------|
| Sand | 0.160 | 0.42 | 3.21 | 2.82 | SP | 2.69 |
| WFS | 0.175 | 0.20 | 1.53 | 0.84 | SP | 2.43 |

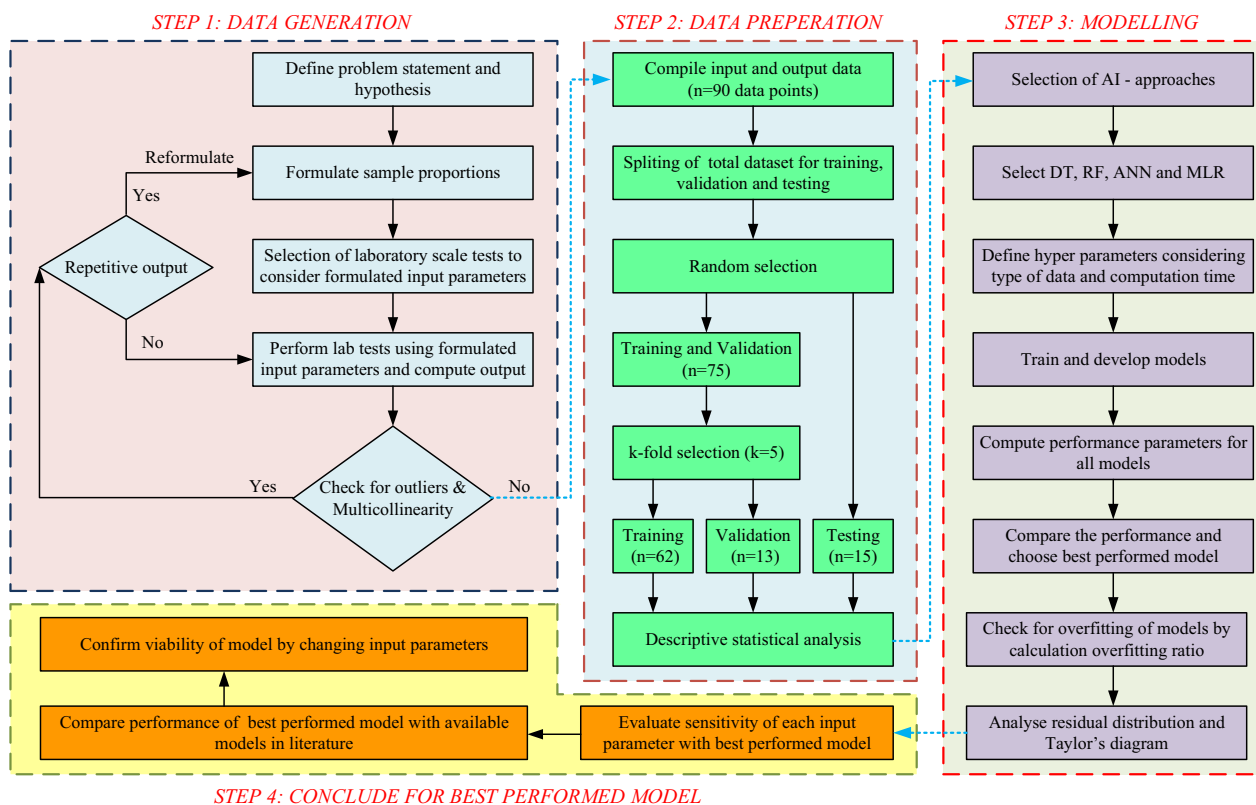


Fig. 2 Flowchart of the research methodology

Table 4 Proportions of different composite soils

| S. No | Base sand (%) | WFS (%) | i_D (%) |
|-------|---------------|---------|-----------|
| 1 | 100 | 0 | 65 |
| 2 | 80 | 20 | 65 |
| 3 | 60 | 40 | 65 |
| 4 | 40 | 60 | 65 |
| 5 | 20 | 80 | 65 |
| 6 | 0 | 100 | 65 |
| 7 | 100 | 0 | 75 |
| 8 | 80 | 20 | 75 |
| 9 | 60 | 40 | 75 |
| 10 | 40 | 60 | 75 |
| 11 | 20 | 80 | 75 |
| 12 | 0 | 100 | 75 |
| 13 | 100 | 0 | 85 |
| 14 | 80 | 20 | 85 |
| 15 | 60 | 40 | 85 |
| 16 | 40 | 60 | 85 |
| 17 | 20 | 80 | 85 |
| 18 | 0 | 100 | 85 |

media. Test for R_D is a preliminary step for sample preparation in the permeability experiment.

3.1.1 Relative density

The relative density of soil composites indicates the compactness of cohesionless materials. The test in this study is conducted as per ASTM D4254 (2000). To measure the minimum density (ρ_{min}), the soil is poured down by free, falling from 2–3 cm height in a relative density mold of 3000 cc, and the mass of soil in the mold is noted down. The maximum density (ρ_{max}) is determined by vibrating the filled mold at a frequency of 3,600 vibrations/min under a 115 kg surcharge for 8 min. After the vibration and removal of the load, the settlement of the loading plate is measured, and the reduced volume is thereby determined. Since the mass remains constant, a reduction in volume in the latter case results in increased density.

The density for different relative densities is calculated by using Eqs. 2 and 3. The values of densities for different relative densities with variation in WFS replacement level are shown in Fig. 8.

$$\rho_{\Delta} = \rho_{max} - \rho_{min} \tag{1}$$

$$\rho_d = \rho_{max} - (R_D * \rho_{\Delta}) \tag{2}$$

3.1.2 Permeability

The permeability of soil indicates the degree of easiness of water to flow through a porous medium. Permeability doesn't depend upon the density and viscosity of the flow-in materials, like hydraulic conductivity. This study measures permeability for all considered cases (Table 4) as per ASTM D2434 (2019). As the soil composites are granular, the constant head method is performed on all compositions. The total head is maintained at 122.5 cm. A Permeameter with a height of 12.73 cm, a diameter of 10 cm, two operating valves, and an air vent is used (Fig. 3). De-aired water is allowed to flow through the sample with two-way drainage. At least five readings of quantity outflow per unit of time for each sample are noted. An equation given by Darcy law (Eq. 3) is adopted for calculating the *k*-value in m/day.

$$k = \frac{Q \times L}{h \times A} \tag{3}$$

where *Q* is the amount of drained water per unit time; *h* is water head; *A* is the cross-sectional area of sample; and *L* is the length of the sample.

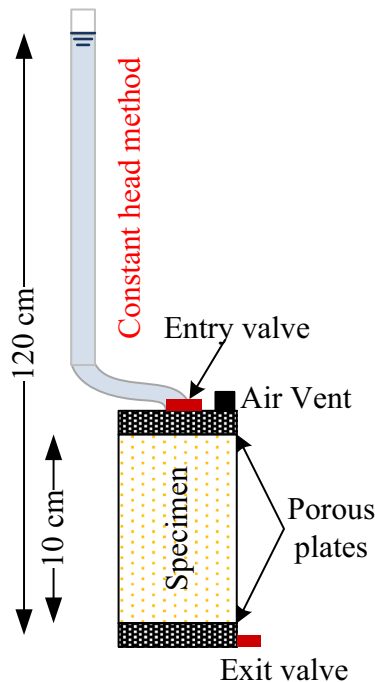


Fig. 3 Setup for permeability test

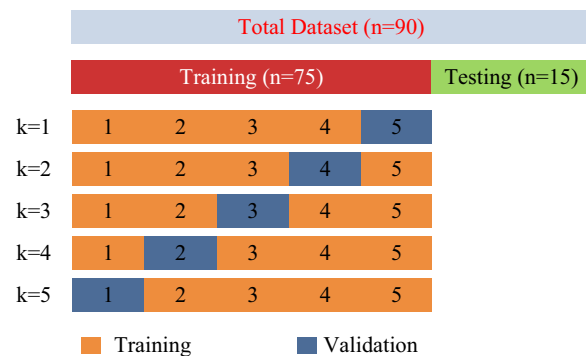


Fig. 4 Splitting of the total data set

3.2 AI-approaches

R programming offers a rich ecosystem of packages, and is specifically designed for machine learning, so it is a versatile tool for data scientists and researchers. Its packages, such as "caret", "mlr" and "tidymodels", provide a wide range of tools and functions to streamline the entire machine-learning workflow. These packages offer well-documented and efficient solutions, ranging from data preprocessing, feature engineering and model selection to training, validation and evaluation. This study uses Rstudio (V 1.4.1564) and R programming platform (V 4.3.1) to access soft computing techniques.

For training the model, data are initially split into three sections: Training data, Validation data and Test data. At the beginning, classifier training is done by using a training data set, followed by using the validation data set to tune the parameters, so as to estimate the skill of the machine learning model on unseen data. In the final stage, the performance of the classifier is tested by using a test data set. According to a widely used thumb rule, the number of data points should be at least ten times the input parameters (Alwosheel et al., 2018; Haykin, 2009). The minimum number of data points required in this study is 70. To avoid overfitting of the model, *k*-fold cross-validation is considered, with the value of *k* as a five-seed value of 42 (Fushiki, 2011). The dataset for training and validation combined should be 85% of the total dataset, with the nearest multiple of 5. The total 90 data points in this study are divided in the ratio of 62/13/15 for training/validation/testing, respectively (Fig. 4). The *k* here represents the fold for cross-validation; it should not be mistaken with *k*, which represents the permeability.

3.2.1 Artificial neural networks (ANN)

This computing technique's working principle is inspired by the biological neural network of human brains. This method was first proposed by McCulloch and Pitts

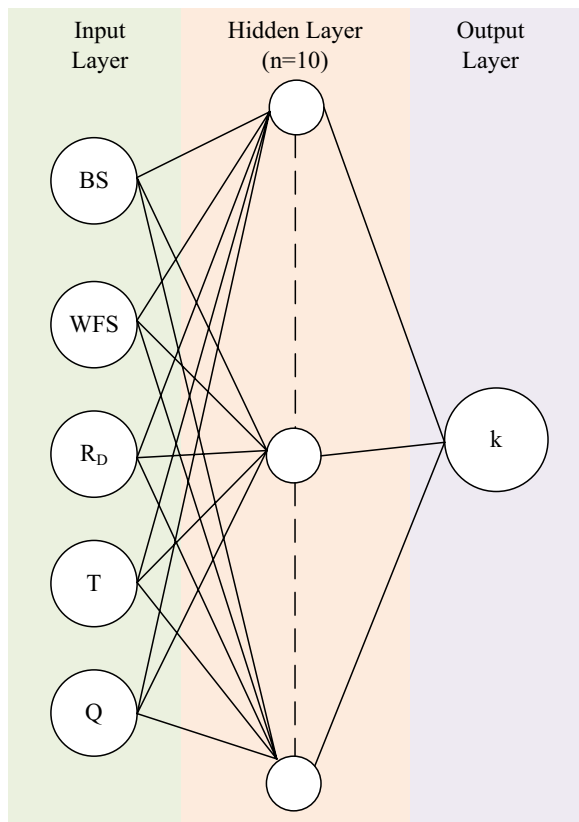


Fig. 5 Architecture of the proposed neural network model

(1943). A group of simulated neurons make an artificial neural network. Every neuron functions as a node linked to other nodes via connections that resemble biological axon-synapse-dendrite connections. A weight is assigned to each link to indicate how strongly one node will affect others (Winston, 1992). Because they can reproduce and model non-linear processes, artificial neural networks have been applied in many disciplines. In civil engineering, they are widely used for soft computing (Lazarovska et al., 2014; Xu et al., 2022; Yang et al., 2021).

Table 5 Hyperparameters for ANN model

| Hyperparameter | Approximate value |
|------------------------------------|-------------------|
| Number of hidden layers | 1 |
| Number of neurons per hidden layer | 10 |
| Activation function | Relu |
| Learning rate | 0.001 |
| Batch size | 64 |
| Number of epochs | 50 |
| Optimizer | Adam |
| Weight initialization | Xavier/glorot |
| Dropout rate | 0.3 |
| Regularization strength | 0.01 |

The input-hidden-output layer schematic is presented in Fig. 5. The hit and trial method is used to select the number of hidden layers, which is found to be optimum at 10. The hyperparameters are optimized for computation time and respective error (Table 5). The activation function chosen for the neurons in the hidden layers is Rectified Linear Unit (ReLU), which helps introduce non-linearity into the model.

3.2.2 Multilinear regression (MLR)

This approach reveals linear relationships between independent (y) and dependent (x) variables. Since multiple-regression takes into account many explanatory variables, it extends the ordinary least-squares regression. The generalized relation is given in Eq. 3, where 'a' represents the intercept and ϵ represents the error. The coefficient b_n is determined by minimizing the sum of the square of residuals after the model is evaluated with statistical parameters.

$$y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n + \epsilon \tag{4}$$

$$k = a(BS)^{b_1} (WFS)^{b_2} (RD)^{b_3} (Q)^{b_4} (T)^{b_5} \tag{5}$$

3.2.3 Decision tree model (DT)

Decision tree is a non-parametric supervised learning method, which can be used for classification and regression. It is aimed at discovering simplistic decision rules derived from data features, so as to build a model that can predict the value of a variable (Fig. 6). This technique is widely used in civil engineering fields, where decisions are often made based on variables' upper or lower limits. For example, if the permeability value is less than 10^{-6} cm/s, the soil will be classified as clay (Desai & Joshi, 2010; Singh et al., 2020). Table 6 outlines the hyperparameters and their respective approximate values used for tuning a decision tree classifier, where the criterion is Gini impurity.

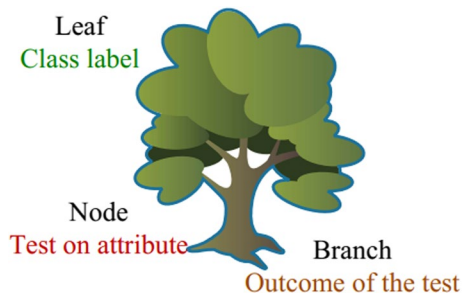


Fig. 6 Parts of the decision tree

Table 6 Hyperparameters for DT model

| Hyperparameter | Approximate value |
|-----------------------|-------------------|
| Criterion | Gini impurity |
| Max depth | None |
| Min samples split | 5 |
| Min samples leaf | 3 |
| Max leaf nodes | None |
| Min impurity decrease | 0 |
| Splitter | Best |
| Class weights | None |
| Random state | 42 |

Table 7 Hyperparameters for RF model

| Hyperparameter | Approximate value |
|-------------------|-------------------|
| Number of trees | 50 |
| Max depth | None |
| Min samples split | 2 |
| Min samples leaf | 1 |
| Max features | Auto |
| Bootstrap | True |
| Criterion | Gini |
| Random state | 42 |
| Class weights | None |
| N_jobs | -1 |

3.2.4 Random forest (RF)

In 1995, the first random decision forest method was developed by Ho (1995). In the fitting process, errors are computed, and the importance of variables is measured. The relevance of variables in a regression or classification task can be ranked by using random forests. Variables that create high values for this score are given higher weightage than those that produce low values. This method solves the problem of overfitting, since the output is based on majority voting or averaging. This technique is widely used in geotechnical engineering to calculate engineering and index properties (Dutta et al., 2019; Rauter & Tschuchnigg, 2021).

In this method, the number of trees for the prediction of the *k*-value is optimized by using an error rate curve, as shown in Fig. 7. More trees than the optimum value may increase the calculation time of the model; meanwhile, a less value can predict erroneous values. The error rate is found to vary insignificantly (can be considered constant) after 50 trees. Fundamental settings or hyperparameters that influence the behavior and performance of the model are tabulated in Table 7.

3.2.5 Limitations of AI models

1. ANN: ANN requires a relatively large amount of data for training, so it may be computationally intensive. It is also often considered a kind of "black-box" models, which makes it a challenging task to interpret their decision-making process. Selecting the exemplary architecture and hyperparameters can be a trial-and-error process, and it is sensitive to initial conditions.
2. MLR: MLR assumes a linear relationship between independent and dependent variables. It might not capture complex non-linear relationships in data. Additionally, it is sensitive to multicollinearity among the predictor variables, which can lead to unstable coefficient estimates.
3. DT: Decision tree assumes that data are non-linearly separable, and it can lead to overfitting, especially when the tree depth is not adequately controlled. It may not perform well on highly imbalanced datasets; and biased trees might be created if one class dominates others.

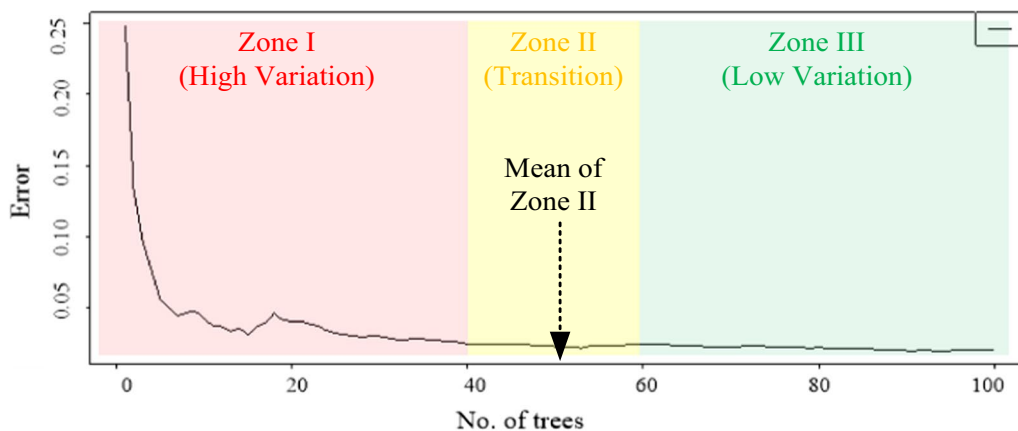


Fig. 7 Error rate progressively for number of trees

4. RF: Random forest is less interpretable than individual decision trees and can be computationally expensive for large datasets. It may not perform optimally when there is a high degree of multicollinearity in the features; and it may struggle with extrapolation as relying on the range of values seen in the training data.

4 Results and discussion

4.1 Experimental results

The variation in density is plotted against WFS content (Fig. 8). It can be seen that the density reduces by 25% with increased WFS content, because WFS exhibits lower dry density values. The experimental results of the

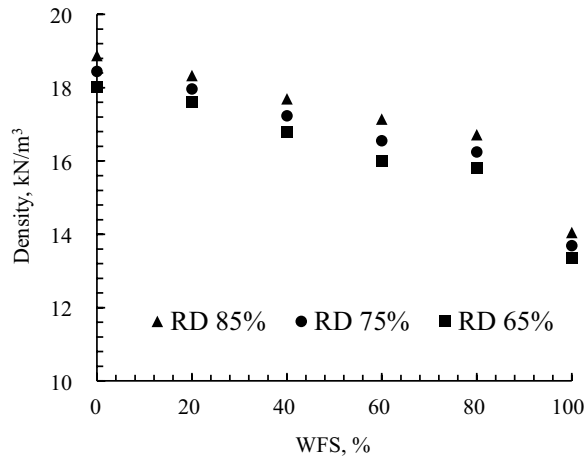


Fig. 8 Variation of density with WFS content

k value for all cases are plotted in Fig. 9. As seen in the surface heat map, the permeability is decreasing as the relative density and WFS content increase. And the permeability also decreases with an increase in the replacement level of WFS. Fully replacement of sand with WFS reduces the k value by 36%, 51% and 57% for R_D values of 65%, 75% and 85%, respectively.

4.2 Statistical features of data sets

The descriptive statistical summary for the training, validation, testing and total data sets is given in Table 8. The summary features all necessary statistical parameters: count, lower and upper bound, mean, standard deviation, kurtosis and skewness. Standard deviation is the maximum in the validation dataset for all parameters. Kurtosis and skewness are reported for all datasets, purposed to measure the symmetries about the center point and the distribution pattern of data. The Pearson's coefficient between two input parameters shows their relationship, and the histogram represents the distribution of data values (Fig. 10).

4.3 Performance of AI models

4.3.1 Results of ANN

The hit and trial method was used to select the number of hidden layers, which is found to be optimum at 10. The 5–10–1 network with 76 weights resulted in SSE value of 0.0782 with a skip-layer connection. More than half of the predicted points are on the negative side of the 1:1 line. Figure 11 shows the output of the ANN model. Error lines of $\pm 20\%$ indicate that the maximum data points

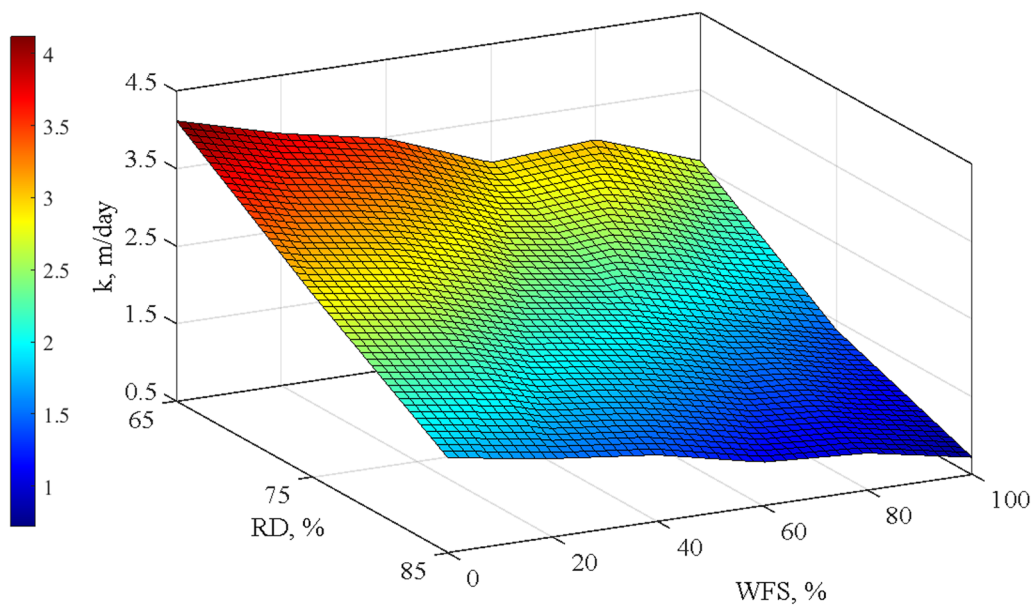


Fig. 9 Variation of permeability with R_D and WFS

Table 8 Statistical features of different data sets

| Variable | Units | Count | Minimum | Maximum | Mean | Standard Deviation | Kurtosis | Skewness |
|--------------------|-------|-------|---------|---------|---------|--------------------|----------|----------|
| Training Dataset | | | | | | | | |
| BS | % | 62 | 0 | 100 | 48.064 | 33.867 | -1.2619 | 0.0884 |
| WFS | % | 62 | 0 | 100 | 51.935 | 33.867 | -1.2619 | -0.0884 |
| R_D | % | 62 | 65 | 85 | 75.483 | 8.381 | -1.5767 | -0.0929 |
| T | sec | 62 | 30 | 150 | 92.419 | 43.218 | -1.3543 | -0.1114 |
| Q | gm | 62 | 17 | 109 | 55.580 | 25.178 | -1.0317 | 0.17817 |
| k | m/day | 62 | 0.6638 | 4.256 | 2.171 | 0.9839 | -1.0385 | 0.17528 |
| Validation Dataset | | | | | | | | |
| BS | % | 13 | 0 | 100 | 63.0769 | 36.3741 | -0.7616 | -0.6609 |
| WFS | % | 13 | 0 | 100 | 36.9231 | 36.3741 | -0.7616 | 0.66086 |
| R_D | % | 13 | 65 | 85 | 75.7692 | 8.62316 | -1.6804 | -0.1636 |
| T | sec | 13 | 30 | 150 | 92.3077 | 43.2346 | -1.3689 | 0.43572 |
| Q | gm | 13 | 29 | 102 | 59 | 26.8359 | -1.2159 | 0.75133 |
| k | m/day | 13 | 1.13236 | 3.982 | 2.30378 | 1.04786 | -1.2159 | 0.75133 |
| Testing Dataset | | | | | | | | |
| BS | % | 15 | 0 | 100 | 46.67 | 34.365 | -1.0909 | 0.08304 |
| WFS | % | 15 | 0 | 100 | 53.33 | 34.365 | -1.0909 | -0.083 |
| R_D | % | 15 | 65 | 85 | 72.33 | 7.03732 | -0.6691 | 0.43303 |
| T | sec | 15 | 30 | 150 | 78 | 40.5674 | -1.2147 | 0.056 |
| Q | gm | 15 | 25 | 110 | 62 | 24.3663 | -0.4054 | 0.42475 |
| k | m/day | 15 | 0.9761 | 4.295 | 2.42 | 0.95143 | -0.4054 | 0.42475 |
| Total Dataset | | | | | | | | |
| BS | % | 90 | 0 | 100 | 50 | 34.3479 | -1.2722 | -5E-18 |
| WFS | % | 90 | 0 | 100 | 50 | 34.3479 | -1.2722 | 5.3E-18 |
| R_D | % | 90 | 65 | 85 | 75 | 8.21071 | -1.517 | 3.1E-17 |
| T | sec | 90 | 30 | 150 | 90 | 42.6641 | -1.3055 | 8.9E-18 |
| Q | gm | 90 | 17 | 110 | 57.14 | 25.1233 | -0.9451 | 0.27479 |
| k | m/day | 90 | 0.6638 | 4.295 | 2.232 | 0.98149 | -0.9499 | 0.2725 |

exist within that limit (Fig. 11a). The maximum error value is found to be -0.55 m/day, which is unacceptable (Fig. 11c). Consequently, this model is inadequate for determining the permeability of sand and WFS mixture.

4.3.2 Results of MLR

The results of MLR model are presented in Fig. 12. As can be seen, all data points get fitted between ± 15% error lines (Fig. 12a). The plot for actual and predicted values along data points is presented in Fig. 12b. The maximum value of the errors is 0.4 m/day (Fig. 12c). MLR assumes that the amount of errors in the residuals is similar at each point of the linear model. This scenario is known as homoscedasticity. This assumption is attributed to the low degree of reliability.

$$k = 0.5051 \frac{(BS)^{0.001} (Q)^{0.03724}}{(RD)^{0.492} (WFS)^{0.0591} (T)^{0.003}}$$

4.3.3 Results of DT

The decision tree analysis predicts the k -values based on Q values. The decision-making process is showcased in Fig. 13. As seen, in a particular box, the k -value is given with the no. of observations (n) and the percentage of observations considered in that condition (%). Condition is written beneath the boxes; all boxes are filled with contrast counter colors (the larger the values, the darker the colors).

Figure 14a presents the scatter plot of actual and predicted k -values acquired from the decision tree model. It is shown that this model performs poorly in predicting the k -value of the sand-WFS mixture; so it is not recommended. A particular predicted value covers a wide range of actual values. The relative error value is 0.6 m/day, the highest among all models (Fig. 14c).

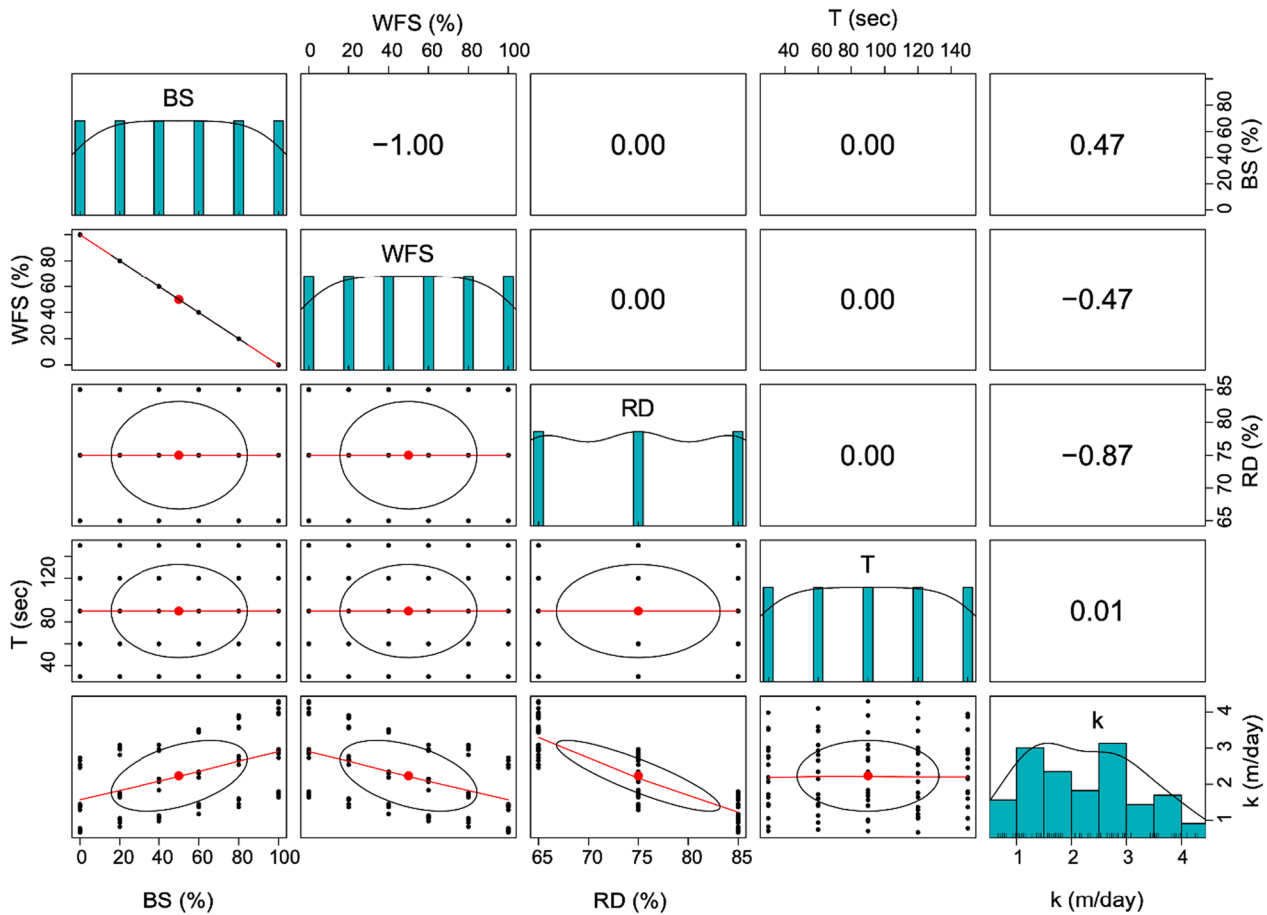


Fig. 10 Distribution of input parameters with Pearson's coefficient values

4.3.4 Results of RF

This approach performs better than other techniques. Results from the RF model are presented in Fig. 15. The error lines to fit the data in the scatter plot are drawn at 15% on the positive side and at 10% on the negative side. For actual *k*-values less than 3.5 m/day, the data points are around the 1:1 line, and the values greater than 3.5 m/day lean towards the negative error line (Fig. 15a). The maximum value of error is found to be 0.45 m/day (Fig. 15c).

4.4 Comparison of performance of different models

4.4.1 Performance parameters

The effectiveness of the proposed models is evaluated by the following performance parameters: coefficient of determination (R^2), mean squared error (MSE), root mean square error (RMSE), performance index (PI),

index of scatter (IOS), index of agreement (IOA), variance accounted for (VAF), and a20 index (Table 9). The mathematical expressions, ideal values and significance of each parameter are also listed in Table 9. Notation y represents actual data, \bar{y} represents the mean of actual data, \hat{y} represents the predicted data, n is the number of data points, and 20 m represents the number of data points, which are in the range of $\pm 20\%$ of the actual data.

The values of these parameters for the training, validation and testing datasets are presented in Tables 10, 11 and 12, respectively. The values demonstrate the performance of different models. The values of R^2 , MSE and RMSE for the testing dataset are less than those for the training and validation datasets.

The values of R^2 for the training, validation and testing datasets are 0.96500, 0.96614 and 0.9126, respectively. The R^2 value for the ANN model is the least of all the proposed models for the testing dataset. For

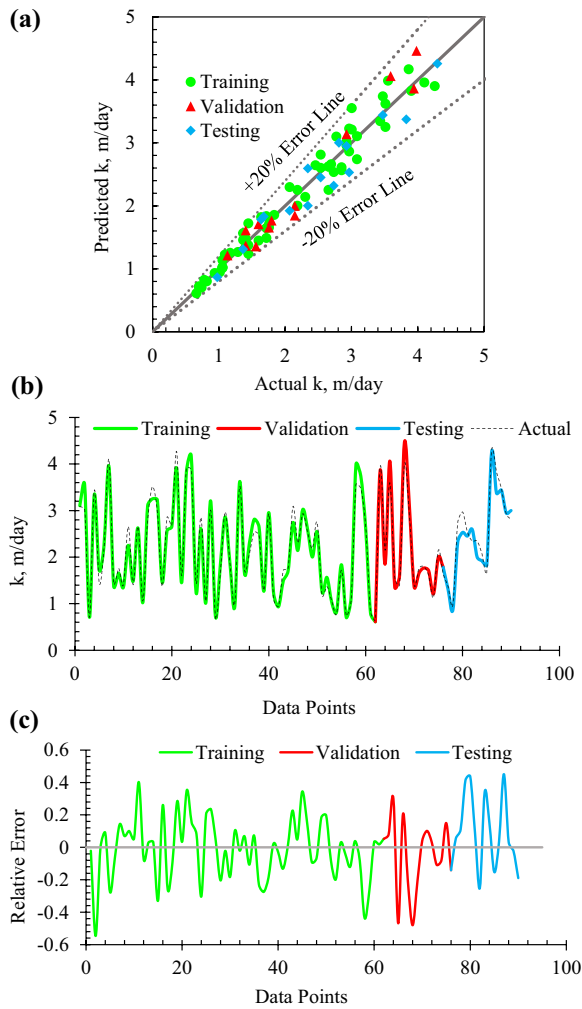


Fig. 11 Results of artificial neural network: **a** performance of the model; **b** actual and predicted k -values; and **c** distribution of errors

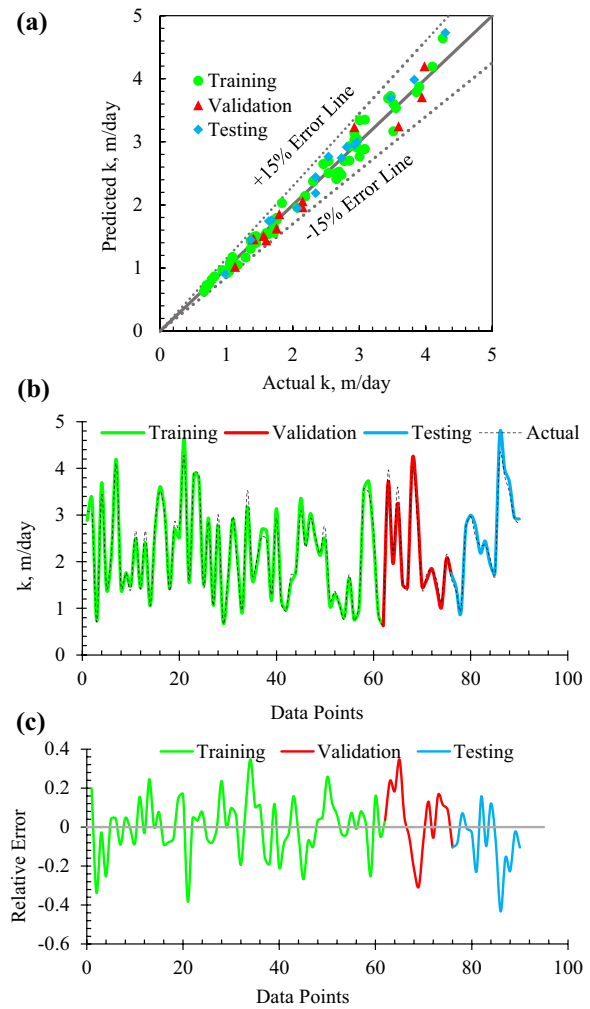


Fig. 12 Results of multi-linear regression model: **a** performance of the model; **b** actual and predicted k -values; and **c** distribution of errors

the training data, the values of R^2 , MSE and RMSE are 0.96106, 0.04289 and 0.2071, respectively.

The performance parameters for RF are optimum for all datasets in a minimum error. The value of R^2 for the training, validation and testing datasets are 0.99314, 0.99374 and 0.9579, respectively. Contrast to the random forest, the multi-linear regression performs well, as the values of R^2 for the training, validation and testing datasets are 0.98066, 0.96854 and 0.9265, respectively. The values of R^2 for the training, validation and testing datasets are 0.96106, 0.95567 and 0.9338, respectively, which shows a poor correlation between the actual and predicted values.

4.4.2 Check for overfitting

Overfitting is a common challenge in machine learning. It means that a model learns the training data to a so excessive degree that it even captures the noises and random fluctuations, instead of only the genuine underlying patterns. As a result, an overfitted model performs well on the training data, but poorly on unseen or new data, leading to bad generalization. Understanding the issue of overfitting is crucial for building up accurate and reliable models on real-world tasks. In this study, the overfitting ratio is computed in Eq. 5. OFR confirms that the RF model is ideally fit.

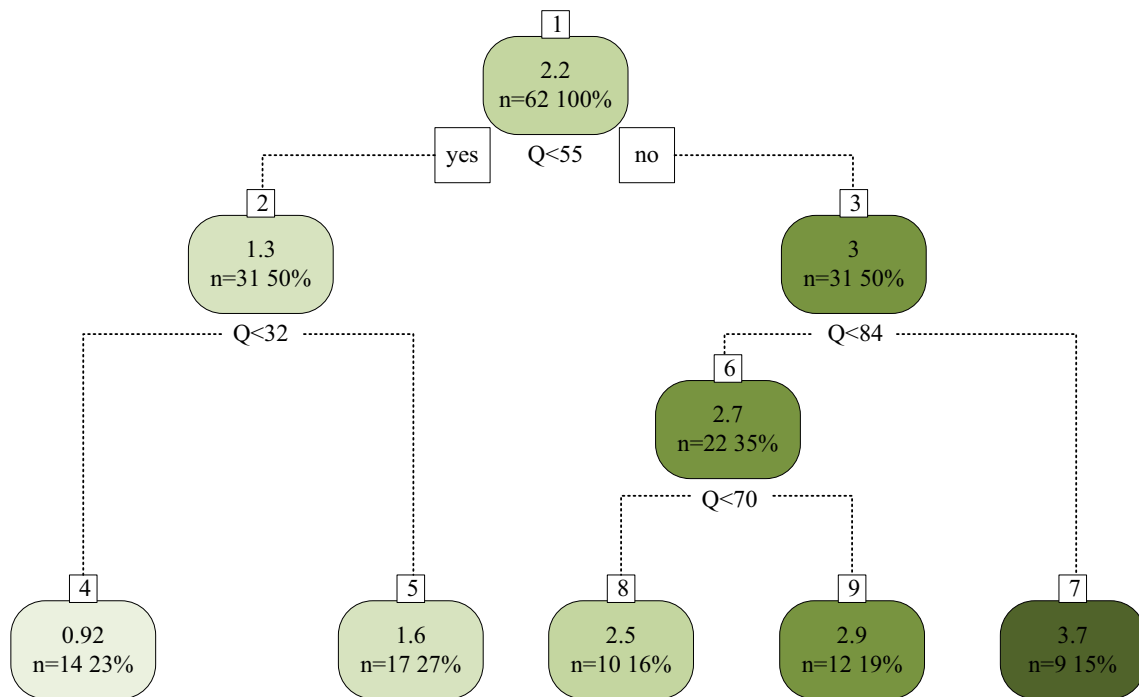
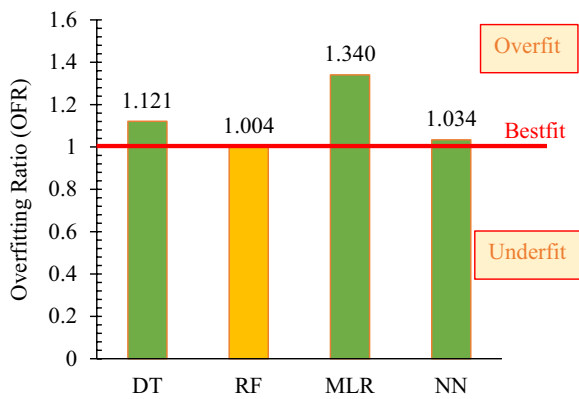


Fig. 13 Decision tree based on Q value

$$OFR = \frac{RMSE_{validation}}{RMSE_{training}}$$



4.4.3 Taylor's diagram

Taylor's diagram quantifies the degree of correspondence between the predicted and actual values. Figure 16 depicts a Taylor diagram, which graphically illustrates the following metrics for all the proposed methods: the value of the Pearson correlation coefficient, the root-mean-square error, and the standard deviation. As can be seen,

(6) the mark of RF is much closer to the actual value point than other marks are.

4.4.4 Distribution of residuals

The error distribution highlights the instances where a model consistently underperforms or overperforms along the data points. The error distribution comparison of all proposed AI approaches (Fig. 17) indicates that the random forest is the best-fit approach for the prediction of the *k*-value of sand-WFS mixtures. The box plots depict the lower and upper values of residuals with outlier points. Investigating these outliers can provide insights into unique scenarios or data points that require special attention. The performance of the models can be compared based on the distance of the median from the origin line. The results are consistent with the performance parameters. Patterns and trends identified from the models' error distribution indicate the absence of systematic errors. These patterns can help increase the robustness of the performance of the models. In addition, the error distribution results show that the models are well-calibrated, particularly in specific prediction ranges.

4.5 Sensitivity analysis

Sensitivity analysis is carried out to determine the most influential input parameters in predicting *k*-values. As RF is an outperforming approach, sensitivity analysis is done

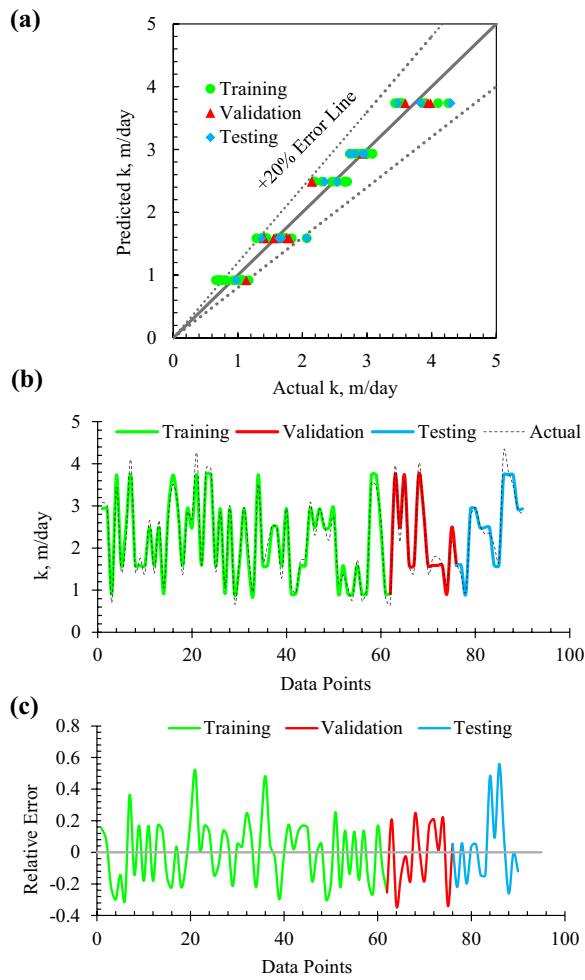


Fig. 14 Results of decision tree model: **a** performance of the model; **b** actual and predicted *k*-values; and **c** distribution of errors

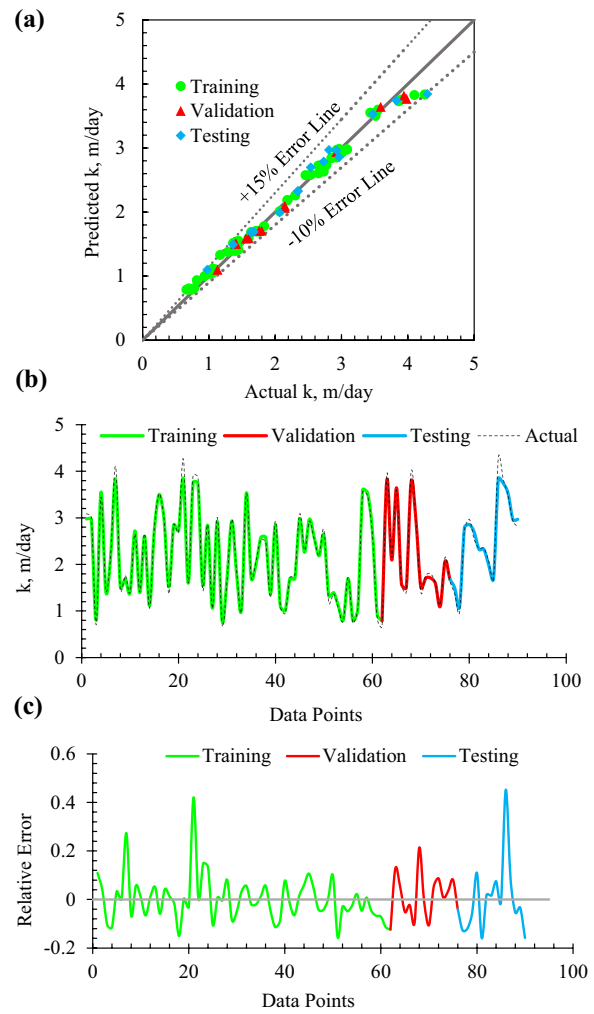


Fig. 15 Results of random forest model: **a** performance of the model; **b** actual and predicted *k*-values; and **c** distribution of errors

Table 9 Insights into the considered performance parameters

| Parameter | Notation | Mathematical expression | Ideal value | Significance |
|------------------------------|----------|--|-------------|---|
| Coefficient of determination | R^2 | $\frac{\sum_{i=1}^n (y-\bar{y})^2 - \sum_{i=1}^n (y-\hat{y})^2}{\sum_{i=1}^n (y-\bar{y})^2}$ | 1 | Measure of the model's goodness of fit |
| Mean Squared Error | MSE | $\frac{1}{n} \sum (y - \hat{y})^2$ | 0 | Quantifies the average squared error between predicted and actual values |
| Root Mean Squared Error | RMSE | $\sqrt{\frac{1}{n} \sum (y - \hat{y})^2}$ | 0 | Provides the same measure as MSE but in the original units of the variable |
| Performance Index | PI | $1 - \frac{\sum_{i=1}^n (y-\hat{y})^2}{\sum_{i=1}^n (y-\bar{y})^2}$ | 1 | Measures how well a model's predictions match the actual values |
| Index of Scatter | IOS | $\frac{\sqrt{\frac{1}{n} \sum (y-\hat{y})^2}}{\bar{y}}$ | 0 | Assessment of scatteredness of data points around a 1:1 line |
| Index of Agreement | IOA | $1 - \frac{\sum_{i=1}^n (\hat{y}-\alpha)}{2 * \sum_{i=1}^n (\alpha-\bar{y})}$ | 1 | Measures agreement between predicted and actual values |
| Variance Accounted For | VAF | $\left[1 - \frac{\text{var}(y-\hat{y})}{\text{var}(y)} \right] * 100$ | 100% | Quantifies the proportion of the total variance in the observed data that is explained by the model's predictions |
| a20 index | a20 | $\frac{20m}{n}$ | 100% | Percentage of predicted values within $\pm 20\%$ of the actual values |

Table 10 Performance parameters for the training dataset

| | R^2 | MSE | RMSE | PI | IOS | IOA | VAF | a20 |
|-----|--------|--------|--------|--------|--------|--------|-------|--------|
| ANN | 0.9650 | 0.0386 | 0.1963 | 0.9636 | 0.0962 | 0.9611 | 99.33 | 88.71 |
| MLR | 0.9807 | 0.0232 | 0.1524 | 0.9791 | 0.0912 | 0.8947 | 98.18 | 100.00 |
| DT | 0.9611 | 0.0429 | 0.2071 | 0.9611 | 0.0708 | 0.9311 | 98.81 | 100.00 |
| RF | 0.9931 | 0.0076 | 0.0869 | 0.9909 | 0.0404 | 0.9520 | 99.17 | 100.00 |

Table 11 Performance parameters for the validation dataset

| | R^2 | MSE | RMSE | PI | IOS | IOA | VAF | a20 |
|-----|--------|--------|--------|--------|--------|--------|-------|--------|
| ANN | 0.9661 | 0.0412 | 0.2030 | 0.9401 | 0.1027 | 0.9555 | 99.31 | 100.00 |
| MLR | 0.9685 | 0.0418 | 0.2040 | 0.9651 | 0.0897 | 0.8520 | 97.71 | 100.00 |
| DT | 0.9557 | 0.0539 | 0.2320 | 0.9555 | 0.0904 | 0.9119 | 98.64 | 100.00 |
| RF | 0.9937 | 0.0076 | 0.0870 | 0.9912 | 0.0386 | 0.9467 | 99.18 | 100.00 |

Table 12 Performance parameters for the testing dataset

| | R^2 | MSE | RMSE | PI | IOS | IOA | VAF | a20 |
|-----|--------|--------|--------|--------|--------|--------|-------|--------|
| ANN | 0.9126 | 0.0633 | 0.2517 | 0.9258 | 0.1132 | 0.9331 | 99.27 | 93.33 |
| MLR | 0.9265 | 0.0144 | 0.1201 | 0.9656 | 0.0994 | 0.7972 | 97.80 | 100.00 |
| DT | 0.9338 | 0.9791 | 0.9867 | 0.9331 | 0.0474 | 0.8714 | 98.60 | 100.00 |
| RF | 0.9679 | 0.0207 | 0.144 | 0.9728 | 0.0569 | 0.9059 | 98.98 | 100.00 |

based on the RF method, and the results are presented in Table 13. This analysis is made by reducting one parameter, and the value of R^2 is noted. The parameters that cause a reduction in R^2 are influential. It is shown that R_D and Q are the most influential parameters, which highly impact the k -values of the composite soil.

$$\hat{s}_i = \hat{s}_a - \hat{s}_r \quad (7)$$

4.6 Comparison with the existing literature

The best architecture model in this study is the random forest for predicting the soil permeability. The proposed model in this study is compared with the models available from the existing literature (Table 14). The R^2 for the actual and predicted datasets of the available models is relatively lower than that of the model proposed in this study.

5 Conclusions

This study explores the drainage behavior of WFS-incorporated sand. The experimental research has been extended to include AI modeling. The following major conclusions are drawn from this study.

- The permeability tends to decrease as the relative density of the soil increases. A notable reduction in the k -value, up to 140%, can be observed when the relative density is increased from 65% to 85%. Similarly, an increase in the replacement level of WFS is associated with the decrease in the permeability. When sand is completely replaced with WFS, there are reductions of 36%, 51%, and 57% in the k -values for the relative density of 65%, 75%, and 85%, respectively.
- The R^2 value and other performance parameters indicate that the relationship between the actual and predicted values is most pronounced in the random forest method. The order of the performance of all the proposed models can be presented as RF > MLR > ANN > DT.
- Taylor's diagram is used to verify the outcomes of all the considered AI approaches, and it proves the good performance of RF, as its mark is nearer to the actual value. The overfitting ratio for RF is close to 1, indicating a strong level of fitness of the model.
- Sensitivity analysis demonstrates that Q and R_D are the most influential parameters for predicting k -values.

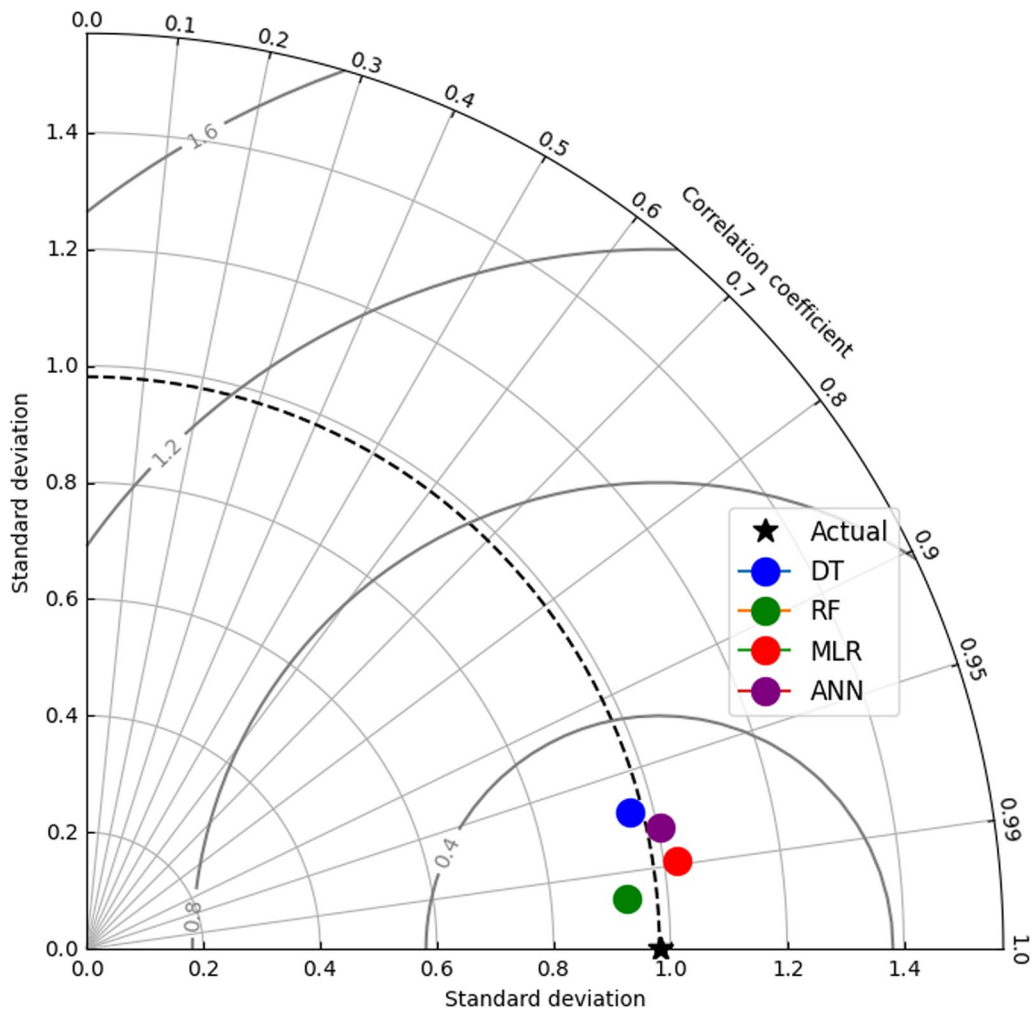


Fig. 16 Taylor diagram for comparison of all proposed models

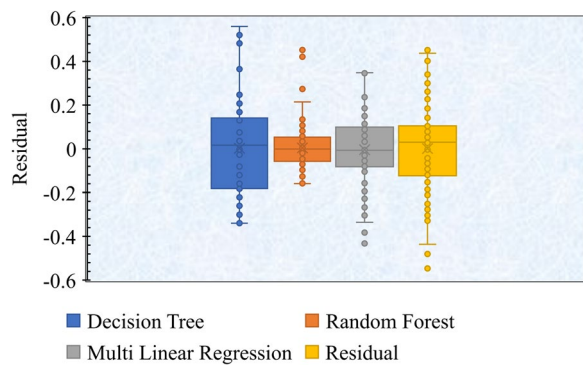


Fig. 17 Distribution of residuals for the proposed models

Table 13 Sensitivity analysis based on the random forest method

| Input parameters | Input parameter removed | R^2 (For total dataset) | Sensitivity | Rank |
|-----------------------|-------------------------|---------------------------|-------------|------|
| BS, WFS, R_D , T, Q | – | 0.9955 | – | |
| WFS, R_D , T, Q | BS | 0.9912 | 0.0043 | 4 |
| BS, R_D , T, Q | WFS | 0.9907 | 0.0048 | 3 |
| BS, WFS, T, Q | R_D | 0.9837 | 0.0118 | 2 |
| BS, WFS, R_D , Q | T | 0.9924 | 0.0031 | 5 |
| BS, WFS, R_D , T | Q | 0.9787 | 0.0168 | 1 |

Table 14 Comparison of the best-performing model with the existing literature

| Input variables | Data size | Best performed model | R ² | References |
|---|-----------|-----------------------------|----------------|---------------------------|
| Sand, silt, clay, BP, PD, TOC | 172 | RF | 0.9022 | Singh et al., (2020) |
| Dielectric constant | 50 | NN | 0.9894 | Uthayakumar et al. (2022) |
| D ₅ , D ₁₀ , D ₃₀ , n | 50 | PSO | 0.92 | Izadi et al. (2022) |
| D ₁₀ , D ₃₀ , D ₆₀ , D ₈₅ , D ₁₀₀ , w, P | 30 | ANN | 0.98 | Erzin et al. (2009) |
| S, Fa, G, T, H | 95 | SVM with Pearson VII kernel | 0.9419 | Singh et al., (2021) |
| BS, WFS, R _D , T, Q | 90 | RF | 0.9955 | This study |

Abbreviations

| | |
|-----------------|--|
| KNN | k-nearest neighbors |
| SVM | support vector machine |
| LightGBM | light gradient boosting machine |
| RF | random forest |
| GB | gradient boosting |
| ANN | artificial neural network |
| MLP | multiple layer perceptron |
| RBF | radial basis function |
| ANFIS | adaptive neuro-fuzzy inference system |
| GPR | gaussian process regression |
| Poly | polynomial |
| PUK | pearson universal kernel |
| TLBO | teaching learning-based optimization |
| MEP | multi-expression-programming |
| GEP | genetic-expression-programming |
| GP | gaussian process |
| MLR | multi-linear-regression |
| CANFIS | co-active neuro-fuzzy inference-system |
| MLP | multilayer perceptron |
| DT | decision tree |
| CART | classification and regression trees |
| GMDH | group method of data handling |
| GA | genetic algorithm |
| k | permeability |
| PI | plasticity index |
| w _l | liquid limit |
| w _p | plastic limit |
| CC | clay content |
| ρ | density |
| wc | water content |
| e | void ratio |
| D ₁₀ | diameter at 10% finer |
| D ₃₀ | diameter at 30% finer |
| D ₆₀ | diameter at 60% finer |
| G | specific gravity |
| ρ _d | dry density |
| D | grain size |
| S | % of sand |
| Fa | % of fly ash |
| T | time |
| H | head |
| OC | organic content |
| BD | bulk density |
| PD | particle density |
| UCS | uniaxial compressive strength |
| Gp | gas pressure |
| Temp | temperature |
| σ' | effective stress |
| Q | rate of flow |
| A | c/s area of sample |
| L | length of flow within soil sample |
| h | total hydraulic head |
| a | c/s area of stand pipe |
| μ | coefficient of viscosity |
| γ | unit weight of water |

| | |
|---------------------------------|--------------------------------|
| h ₁ , h ₂ | hydraulic heads |
| t | time interval |
| d _e | effective grain size |
| c | constant related to e |
| n | porosity |
| S _s | specific surface |
| BS | percentage of sand |
| WFS | fraction of waste foundry sand |

Author contributions

Conceptualization, methodology, software, data curation, writing—original draft preparation, visualization, investigation, formal analysis (AK): Supervision, writing—reviewing and editing (AP).

Funding

Not applicable.

Data availability

Data will be made available on reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 6 June 2023 Revised: 19 November 2023 Accepted: 29 November 2023
Published online: 02 January 2024

References

Abichou, T., Benson, C. H., Edil, T. B., & Freber, B. W. (1998). Using waste foundry sand for hydraulic barriers. *Proceedings of Recycled Materials in Geotechnical Applications*, 86–99.

Abichou, T., Benson, C. H., & Edil, T. B. (2000). Foundry green sands as hydraulic barriers: Laboratory study. *Journal of Geotechnical and Geoenvironmental Engineering*, 126(12), 1174–1183. [https://doi.org/10.1061/\(ASCE\)1090-0241\(2000\)126:12\(1174\)](https://doi.org/10.1061/(ASCE)1090-0241(2000)126:12(1174))

Abichou, T., Benson, C. H., & Edil, T. B. (2002). Foundry green sands as hydraulic barriers: Field study. *Journal of Geotechnical and Geoenvironmental Engineering*, 128(3), 206–215.

Ahmad, M., Keawsawasvong, S., Bin Ibrahim, M. R., Waseem, M., Kashyadeh, K. R., & Sabri, M. M. S. (2022). Novel approach to predicting soil permeability coefficient using Gaussian process regression. *Sustainability*, 14(14), 8781. <https://doi.org/10.3390/su14148781>

Alwosheel, A., van Cranenburgh, S., & Chorus, C. G. (2018). Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of Choice Modelling*, 28, 167–182. <https://doi.org/10.1016/j.jocm.2018.07.002>

ASTM D4254. (2000). *Standard test methods for minimum index density and unit weight of soils and calculation of relative density*. West Conshohocken: ASTM International.

- ASTM D854. (2002). *Standard test methods for specific gravity of soil solids by water pycnometer*. West Conshohocken: ASTM International.
- ASTM D2487. (2006). *Classification and identification of soils for general engineering purposes*. West Conshohocken: ASTM International.
- ASTM D422-63. (2007). *Standard test method for particle-size analysis of soils*. West Conshohocken: ASTM International.
- ASTM D2434. (2019). *Standard test method for permeability of granular soils (constant head)*. West Conshohocken: ASTM International.
- Baghbani, A., Choudhury, T., Costa, S., & Reiner, J. (2022). Application of artificial intelligence in geotechnical engineering: A state-of-the-art review. *Earth-Science Reviews*, 228, 103991. <https://doi.org/10.1016/j.earscrv.2022.103991>
- Bui, Q.-A.T., Al-Ansari, N., van Le, H., Prakash, I., & Pham, B. T. (2022). Hybrid model: teaching learning-based optimization of artificial neural network (TLBO-ANN) for the prediction of soil permeability coefficient. *Mathematical Problems in Engineering*, 2022, 1–9. <https://doi.org/10.1155/2022/8938836>
- Carey, T. J., Stone, N., & Kutler, B. L. (2020). Grain size analysis and maximum and minimum dry density testing of Ottawa F-65 Sand for LEAP-UCD-2017. In *Model Tests and Numerical Simulations of Liquefaction and Lateral Spreading* (pp. 31–44). Springer International Publishing. https://doi.org/10.1007/978-3-030-22818-7_2
- Casagrande, A. (1948). Classification and identification of soils. *Transactions of the American Society of Civil Engineers*, 113, 901–930.
- Cashman, P. M., & Preeene, M. (2020). Permeability of soils and rocks. In *Groundwater Lowering in Construction* (pp. 73–92). CRC Press. <https://doi.org/10.1201/9781003050025-5>
- Chen, J., Tong, H., Yuan, J., Fang, Y., & Gu, R. (2022). Permeability prediction model modified on Kozeny-Carman for building foundation of clay soil. *Buildings*, 12(11), 1798. <https://doi.org/10.3390/buildings12111798>
- Dalkilic, H. Y., Kumar, D., Samui, P., Dixon, B., Yesilyurt, S. N., & Katipoğlu, O. M. (2023). Application of deep learning approaches to predict monthly stream flows. *Environmental Monitoring and Assessment*, 195(6), 705. <https://doi.org/10.1007/s10661-023-11331-5>
- Darcy, H. (1856). *Les Fontaines Publiques de la Ville de Dijon*.
- Desai, V. S., & Joshi, S. (2010). *Application of decision tree technique to analyze construction project data* (pp. 304–313). https://doi.org/10.1007/978-3-642-12035-0_30
- Dutta, R. K., Gnananandarao, T., & Sharma, A. (2019). Application of random forest regression in the prediction of ultimate bearing capacity of strip footing resting on dense sand overlying loose sand deposit. *Journal of Soft Computing in Civil Engineering*, 3(4), 28–40. <https://doi.org/10.22115/scce.2019.137910.1080>
- 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. <https://doi.org/10.1139/T09-035>
- Feng, S., Barreto, D., Imre, E., Ibraim, E., & Vardanega, P. J. (2023). Use of hydraulic radius to estimate the permeability of coarse-grained materials using a new geodatabase. *Transportation Geotechnics*, 41, 101026. <https://doi.org/10.1016/j.jtrgeo.2023.101026>
- FHWA. (2004). *Foundry sands facts for civil engineers*. Federal Highway Administration.
- Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. *Statistics and Computing*, 21(2), 137–146. <https://doi.org/10.1007/s11222-009-9153-8>
- Gedik, A., Lav, A. H., & Lav, M. A. (2018). Investigation of alternative ways for recycling waste foundry sand: An extensive review to present benefits. *Canadian Journal of Civil Engineering*, 45(6), 423–434. <https://doi.org/10.1139/cjce-2017-0183>
- Guney, Y., Aydilek, A. H., & Demirkan, M. M. (2006). Geoenvironmental behavior of foundry sand amended mixtures for highway subbases. *Waste Management*, 26(9), 932–945. <https://doi.org/10.1016/j.wasman.2005.06.007>
- Haykin, S. (2009). *Neural networks and learning machines (Third)*. Pearson.
- Hazen, A. (1911). Discussion: Dams on sand foundations. *Transactions, American Society of Civil Engineers*, 73(11).
- Heidemann, M., Nierwinski, H. P., Hastenpflug, D., Barra, B. S., & Perez, Y. G. (2021). Geotechnical behavior of a compacted waste foundry sand. *Construction and Building Materials*, 277, 122267. <https://doi.org/10.1016/j.conbuildmat.2021.122267>
- Ho, T. K. (1995). Random Decision Forests. *3rd International Conference on Document Analysis and Recognition, Montreal*, 278–282.
- Izadi, H., Roostaei, M., Hosseini, S. A., Soroush, M., Mahmoudi, M., Devere-Bennett, N., Leung, J. Y., & Fattahpour, V. (2022). A hybrid GBPSO algorithm for permeability estimation using particle size distribution and porosity. *Journal of Petroleum Science and Engineering*, 217, 110944. <https://doi.org/10.1016/j.petrol.2022.110944>
- Javed, S. (1995). Uses of waste foundry sands in civil engineering. *Transportation Research R*, 1486, 109–113.
- Javed, S., & Lovell, C. W. (1994). Use of waste foundry sand in highway construction. *Joint Highway Research Project Report, Indiana Department of Transportation*, 1–304.
- Khatti, J., & Grover, K. S. (2021a). Computation of permeability of soil using artificial intelligence approaches. *International Journal of Engineering and Advanced Technology*, 11(1), 257–266. <https://doi.org/10.35940/ijeat.A3220.1011121>
- Khatti, J., & Grover, K. S. (2021b). Determination of permeability of soil for Indian soil classification system using artificial neural network technique. *Invertis Journal of Science & Technology*, 14(2), 49–57. <https://doi.org/10.5958/2454-762X.2021.00005.6>
- Khatti, J., & Grover, K. S. (2023a). Assessment of fine-grained soil compaction parameters using advanced soft computing techniques. *Arabian Journal of Geosciences*, 16(3), 208. <https://doi.org/10.1007/s12517-023-11268-6>
- Khatti, J., & Grover, K. S. (2023b). CBR Prediction of pavement materials in unsoaked condition using LSSVM, LSTM-RNN, and ANN approaches. *International Journal of Pavement Research and Technology*. <https://doi.org/10.1007/s42947-022-00268-6>
- Khatti, J., & Grover, K. S. (2023c). Prediction of compaction parameters for fine-grained soil: Critical comparison of the deep learning and standalone models. *Journal of Rock Mechanics and Geotechnical Engineering*. <https://doi.org/10.1016/j.jrmge.2022.12.034>
- Khatti, J., & Grover, K. S. (2023d). Prediction of UCS of fine-grained soil based on machine learning part 1: Multivariable regression analysis, gaussian process regression, and gene expression programming. *Multiscale and Multidisciplinary Modeling, Experiments and Design*, 6(2), 199–222. <https://doi.org/10.1007/s41939-022-00137-6>
- Khatti, J., & Grover, K. S. (2023e). Prediction of UCS of fine-grained soil based on machine learning part 2: Comparison between hybrid relevance vector machine and Gaussian process regression. *Multiscale and Multidisciplinary Modeling, Experiments and Design*. <https://doi.org/10.1007/s41939-023-00191-8>
- Kim, M. H., & Song, C. M. (2023). Prediction of the soil permeability coefficient of reservoirs using a deep neural network based on a dendrite concept. *Processes*, 11(3), 661. <https://doi.org/10.3390/pr11030661>
- Kozeny, J. (1927). Über Kapillare Leitung der Wasser in Boden. *Royal Academy of Science, Vienna Proc. Class I*, 136, 271–306.
- Kumar, A., & Parihar, A. (2022). State-of-the-art review on sustainability in geotechnical applications of waste foundry sand. *Indian Geotechnical Journal*, 52(2), 416–436. <https://doi.org/10.1007/s40098-021-00580-1>
- Kumar, D. R., Samui, P., Wipulanusat, W., Keawsawasvong, S., Sangjinda, K., & Jitchaijaroen, W. (2023). Soft-computing techniques for predicting seismic bearing capacity of strip footings in slopes. *Buildings*, 13(6), 1371. <https://doi.org/10.3390/buildings13061371>
- Länsivaara, T. T., Farhadi, M. S., & Samui, P. (2023). Performance of traditional and machine learning-based transformation models for undrained shear strength. *Arabian Journal of Geosciences*, 16(3), 183. <https://doi.org/10.1007/s12517-022-11173-4>
- Lazarevska, M., Kneević, M., Cvetkovska, M., & Trombeva-Gavriloska, A. (2014). Application of artificial neural networks in civil engineering. *Tehnicki Vjesnik-Technical Gazette*, 21, 1353–1359.
- Lee, K., Cho, J., Salgado, R., & Lee, I. (2001). Retaining wall model test with waste foundry sand mixture backfill. *Geotechnical Testing Journal*, 24(4), 401–408. <https://doi.org/10.1520/GTJ11137J>
- Lin, D., Wu, H., & Hu, L. (2018). Excess Pore Pressure During One-Dimensional Self-weight Consolidation. In *Proceedings of GeoShanghai 2018 International Conference: Fundamentals of Soil Behaviours* (pp. 407–416). Springer Singapore. https://doi.org/10.1007/978-981-13-0125-4_45
- Mast, D. G., & Fox, P. J. (1998). FHWA/IN/JTRP-98/18, Geotechnical performance of a highway embankment constructed using waste foundry sand. *Joint Transportation Research Program, Indiana Department of Transportation*.

- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133. <https://doi.org/10.1007/BF02478259>
- Ministry of Mines, I. (2018). *Sand Mining Framework*.
- Nagy, L., akács, A. T., Huszák, T., Mahler, A., & Varga, G. (2013). Comparison of permeability testing methods. *International Conference on Soil Mechanics and Geotechnical Engineering*.
- Osterhout, W. J. (1922). Direct and indirect determinations of permeability. *Journal of General Physiology*, 4(3), 275–283. <https://doi.org/10.1085/jgp.4.3.275>
- Partridge, B. K., Fox, P. J., Alleman, J. E., & Mast, D. G. (1999). Field demonstration of highway embankment construction using waste foundry sand. *Transportation Research Record*, 1670, 98–105. <https://doi.org/10.3141/1670-13>
- Pham, B. T., Nguyen, M. D., Al-Ansari, N., Tran, Q. A., Ho, L. S., van Le, H., & Prakash, I. (2021). A comparative study of soft computing models for prediction of permeability coefficient of soil. *Mathematical Problems in Engineering*, 2021, 1–11. <https://doi.org/10.1155/2021/7631493>
- Rabbani, A., Samui, P., & Kumari, S. (2023). Implementing ensemble learning models for the prediction of shear strength of soil. *Asian Journal of Civil Engineering*. <https://doi.org/10.1007/s42107-023-00629-x>
- Rauter, S., & Tschuchnigg, F. (2021). CPT data interpretation employing different machine learning techniques. *Geosciences*, 11(7), 265. <https://doi.org/10.3390/geosciences11070265>
- Rehman, Z., Khalid, U., Ijaz, N., Mujtaba, H., Haider, A., Farooq, K., & Ijaz, Z. (2022). Machine learning-based intelligent modeling of hydraulic conductivity of sandy soils considering a wide range of grain sizes. *Engineering Geology*, 311, 106899. <https://doi.org/10.1016/j.enggeo.2022.106899>
- Sarker, I. H. (2021). Machine learning: algorithms, real-world applications and research directions. *SN Computer Science*, 2(3), 160. <https://doi.org/10.1007/s42979-021-00592-x>
- Shahin, M. A. (2013). Artificial Intelligence in Geotechnical Engineering. In *Metaheuristics in Water, Geotechnical and Transport Engineering* (pp. 169–204). Elsevier. <https://doi.org/10.1016/B978-0-12-398296-4.00008-8>
- Siddique, R., & Singh, G. (2011). Utilization of waste foundry sand (WFS) in concrete manufacturing. *Resources, Conservation and Recycling*, 55(11), 885–892. <https://doi.org/10.1016/j.resconrec.2011.05.001>
- Singh, B., Sihag, P., Pandhiani, S. M., Debnath, S., & Gautam, S. (2021). Estimation of permeability of soil using easy measured soil parameters: Assessing the artificial intelligence-based models. *ISH Journal of Hydraulic Engineering*, 27(sup1), 38–48. <https://doi.org/10.1080/09715010.2019.1574615>
- Singh, V. K., Kumar, D., Kashyap, P. S., Singh, P. K., Kumar, A., & Singh, S. K. (2020). Modelling of soil permeability using different data driven algorithms based on physical properties of soil. *Journal of Hydrology*, 580, 124223. <https://doi.org/10.1016/j.jhydrol.2019.124223>
- Sinha, A. K., Vinoth, M., Shankar, S. R. (2020). Characterization of Foundry Sand Waste Material for Road Construction. *New Building Material & Construction World*. <https://www.nbmcw.com/article-report/infrastructure-construction/roads-and-pavements/characterisation-of-foundry-sand-waste-material-for-road-construction.html>
- Slichter, C. S. (1899). *Heoretical investigation of the motion of ground waters*.
- Smith, I. (2014). *Smith's elements of soil mechanics* (9th ed.). Wiley-Blackwell.
- Taylor, D. W. (1948). Fundamentals of soil mechanics. *Soil Science*, 66(2), 161. <https://doi.org/10.1097/00010694-194808000-00008>
- Terzaghi, K. (1925). Determination of permeability of clays. *Engineering News-Record*, 95(21), 832–836.
- Tittarelli, F. (2018). Waste foundry sand. In *Waste and Supplementary Cementitious Materials in Concrete* (pp. 121–147). Elsevier. <https://doi.org/10.1016/B978-0-08-102156-9.00004-3>
- Torabi, M., Sarkardeh, H., & Mirhosseini, S. M. (2022a). Estimating the permeability coefficient of soil using CART and GMDH approaches. *Water Supply*, 22(8), 6756–6764. <https://doi.org/10.2166/ws.2022.248>
- Torabi, M., Sarkardeh, H., & Mirhosseini, S. M. (2022b). Prediction of soil permeability coefficient using GEP approach. *Numerical Methods in Civil Engineering*. <https://doi.org/10.52547/nmce.2022.414>
- Tran, V. Q. (2022). Predicting and investigating the permeability coefficient of soil with aided single machine learning algorithm. *Complexity*, 2022, 1–18. <https://doi.org/10.1155/2022/8089428>
- Uthayakumar, A., Mohan, M. P., Khoo, E. H., Jimeno, J., Siyal, M. Y., & Karim, M. F. (2022). Machine learning models for enhanced estimation of soil moisture using wideband radar sensor. *Sensors*, 22(15), 5810. <https://doi.org/10.3390/s22155810>
- Vipulanandan, C., Weng, Y., & Zhang, C. (2000). Designing flowable grout mixes using foundry sand, clay and fly ash. *Proceedings of Sessions of Geo-Denver Advances in Grouting and Ground Modification*, 215–233. [https://doi.org/10.1061/40516\(292\)15](https://doi.org/10.1061/40516(292)15)
- Wang, J., Yan, W., Zhijun, W., Wang, Y., Lv, J., & Zhou, A. (2020). Prediction of permeability using random forest and genetic algorithm model. *Computer Modeling in Engineering & Sciences*, 125(3), 1135–1157. <https://doi.org/10.32604/cmescs.2020.014313>
- Winkler, E. S., & Bolshakov, A. (2000). *Characterization of foundry sand waste*. Winston, P. H. (1992). *Artificial intelligence* (3rd ed.). Pearson.
- Xu, H., Chang, R., Pan, M., Li, H., Liu, S., Webber, R. J., Zuo, J., & Dong, N. (2022). Application of artificial neural networks in construction management: A scientometric review. *Buildings*, 12(7), 952. <https://doi.org/10.3390/buildings12070952>
- Yang, X., Guan, J., Ding, L., You, Z., Lee, V. C. S., Mohd Hasan, M. R., & Cheng, X. (2021). Research and applications of artificial neural network in pavement engineering: A state-of-the-art review. *Journal of Traffic and Transportation Engineering (english Edition)*, 8(6), 1000–1021. <https://doi.org/10.1016/j.jtte.2021.03.005>
- Yilmaz, 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(5), 957–968. <https://doi.org/10.1007/s00521-011-0535-4>
- Zhang, L., Dang, F., Gao, J., & Ding, J. (2021). Measurement and investigation on 1-D consolidation permeability of saturated clay considering consolidation stress ratio and stress history. *Geofluids*, 2021, 1–21. <https://doi.org/10.1155/2021/6616331>

Publisher's Note

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

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)