Predicting the reduction of embankment pressure on the surface of the soft ground reinforced by sand drain with random forest regression

Predicting the reduction of embankment pressure on the surface of the soft ground reinforced by sand drain with random forest regression

Tiep Duc Pham, Nang Duc Bui, Tuan Tien Nguyen and Hieu Chi Phan

Le Quy Don Technical University, Hanoi, Vietnam.

E-mail: phamductiep@lqdtu.edu.vn

Abstract. The consolidation acceleration of embankment with sand drain has been studied in many researchs and standards. However, the pressure distribution on the surface of soft ground with the appearance of sand drain has not focused. Intuitively, because of the much higher stiffness of sand drain compared to this of soft soil, the stress concentration at the top of sand drain occurred along with the pressure decrease on the surface of soft ground. In this paper, the Finite Element Analysis (FEA) is implemented to obtain a labeled database with inputs are sand drain and soft soil moduluses, diameter of sand drains and distance between them. The predicted variable is the ratio of pressure on the surface of soft ground with and without sand drain (K) obtained based on simulation with Plaxis. Consequently, the developed database used as the input of a machine learning model, the Random Forest Regression (RFR). To the end, observations from FEA reinforced the initial intuition of this phenomenon and a predicting model for K also proposed with Random Forest Regression.

1. Introduction

The use of sand drain to accelarate the consolidation process has been studied and widely accepted in many researches (e.g. [1, 2, 3] .etc) and codes (e.g. [4, 5].etc). Lekha et al. [1] proposed a generalized governing equation for sand drain consolidation with time-dependent loading. Wan-Huan and Shuai [2] applied differential quadrature method for unsaturated consolidation. Koy et al. in [3] used numerical analysis for clay consolidation including the heat injection and sand drain. Another concern on sand drain is the increase of the stabitility of the overall structure ([6, 7]). Ramkrishnan et al. [6] simulated the behaviour of the seepage while Mahmood et al. [7] stressed on stability of dam foundation with sand drain.

To the authors' best knowledge, the change of pressure on the surface of soft soil with the appearance of sand drain are not discussed or, at least, not fully developed in literature. According to [2] and [3], sand drain improves the drainage efficiency without compacting the soft soil. Practically, the sand drain already has the density to some level due to the previous construction process. This density is nessessary to maintain the shape stability and drainage ability of the sand drain. The void ratio of sand drain can be found by Eq.1 [8] with an assumption that sand drain diameter is constant:

$$\varepsilon_{tk} = \frac{\varepsilon_0 - (k_{yc} - 1)}{k_{yc}} \quad (1)$$

where: ε_{tk} is sand void ratio after construction process; ε_o is initial sand void ratio, obtained by experiment; k_{yc} is the required ratio of the volume of sand V_0 and the design volume of the sand drain

V_{tk}. k_{yc} commly ranges within (1.2÷1.3) and appear in Eq.1a:

$$V_0 = k_{yc} \times V_{tk} \tag{1a}$$

Considering to the fact that the medium sand to coarse sand are used in practice, ε_o are choosen within [0.7÷1.0] in this paper. Subtituting this range to Eq.1, ε_{tk} ranges within [0.31÷0.67] or sand in the sand drain is classified as medium to very dense [9]. Arcording to [9], the corresponding sand modulus (E_c), angle of internal friction (φ_c), and cohesion (c_c) are [30÷50] MPa, [35÷40] degree and [1÷3] kPa, respectively.

Comparing to the sand modulus, the soft soil modulus ($E_{đy}$) is significantly lower at less than 5 MPa [10]. This difference is the cause of the stress concentration at the top of the sand drains and the reduction of pressure on the surface of soft soil layer. However, current standards only considered this effect on sand pile [11] (Fig.1) and there is an ignorance of this effect on sand drain (e.g. [4, 5]) leading to the over estimation of calculated results versus the monitoring results [12].

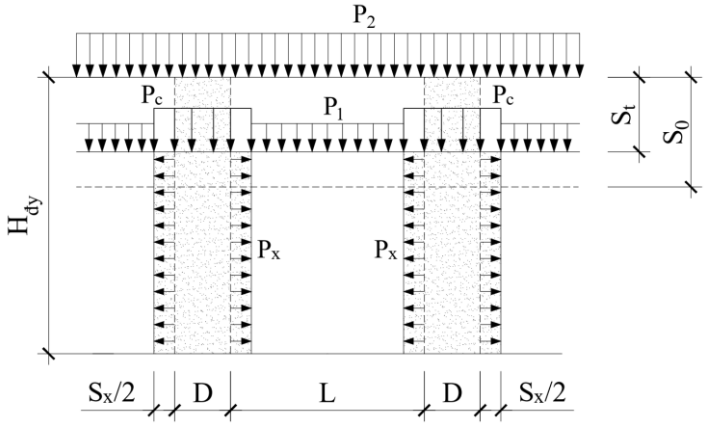


Figure 1. The scheme for calculating the soft ground with sand piles ($H_{đy}$ - thickness of soft soil, L - distance between sand piles, D - diameter of sand pile, S_0 - vertical settlement of soft ground without sand pile, S_t - vertical settlement of soft ground with sand pile, $S_x/2$ - horizontal displacement of sand pile, P_1 - pressure on the surface of the soft ground, P_c - pressure on the top of sand pile, P_2 - pressure of embankment [11]).

Onsite monitoring results on settlement of soft soil with the sand drain is consistently lower than the settlement of soft soil without the sand drain at around 5 to 15% [12]. Evgeniev I.E [12] proposed Eq.2 as an approximation of settlement of soil with sand drain:

$$S = \left(\frac{e_0 - e_p}{1 + e_0} - \frac{d_c^2}{L^2} \right).h \tag{2}$$

where: e_0 is the initial void ratio; e_p is the void ratio under the external load; d_c is the diameter of sand drain; L is the distance among sand drains; h is the thickness of soft soil with sand drain.

Dobrov E.M [13] has been analytically and numerically proofed the existance of the “hanging effect” insteads of the well-known arching effect in the soft soil domain. The embankment presssure on soft soil depends on distance among sand drains, diameter of them, moduluses of soft soil and material made of sand drain. These are the importance variables focused in this paper as the input features to predict the ratio of embankment pressure on the surface of soft ground with and without sand drain (designated as K). The numerical analyses implemented by Finite Element method with Plaxis 2D software.

Regression modelling the K metric from the database of such Finite Element Analysis (FEA) then implement as a supervised learning task in machine learning. Along with Artificial Neutral Network (ANN) and Support Vector Machine (SVM), Random Forest is a powerful machine learning tool [14]. This ensemble learning method developed by Ho in [15, 16] and then commonly used for geotechnical problems for predicting concerned variables. For instance, Kohestani et al. in [17] implement Random

Forest Regression (RFR) to predict liquefaction potential of soil under an earthquake based on field records with the comparison to ANN and SVM models. In a further study of Kohestani, RFR is the best model for predicting maximum surface settlement of soil due to earthquake [3]. The maximum surface settlement caused by tunnel construction are the predicted variable of RFR in [18]. Unxial compressive strength and Young modulus of rock have been successfully predicted in [19]. In this study, a RFR model is developed to predict the K metric based on the data obtained from Finite Element Analysis (FEA).

2. Methodology

2.1 Numerical method for stress analysis the embankment on soft soil reinforced by sand drain
The FEA with Plaxis 2D is implemented in this study to model the embankment on soft soil reinforced by sand drain using the triangular plane element type. Fig.2 provides the diagram of the problem with three types of soils (i.e. the soft soil, embankment soil and the sand drain) along with their dimensions. The Mohr - Coulomb model is used for the nonlinear behaviour of the materials incorperating with the Newton Raphson method to solve the nonlinear equations.

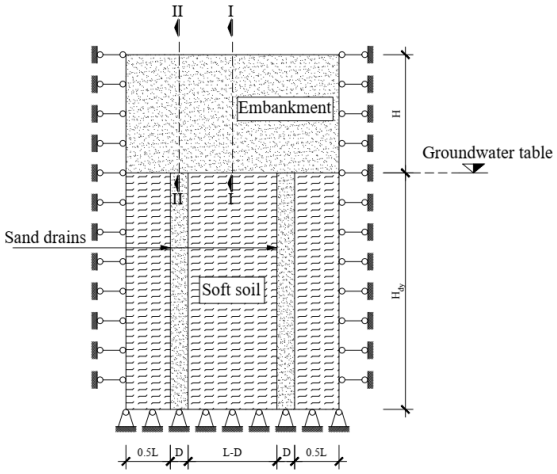


Figure 2. Diagram of the vertical sand drains

Dimensions of sand drain includes the height (H_{dy}, equals to the thickness of the soft soil), the diameter (D) and distance between sand drains (L). The deformation modulus (E_{dy}), angle of internal frictions (φ_{dy}), cohesions (C_{dy}) of soft soil and the deformation modulus of sand drain (E_c) are also the critical input variables. The E_c corresponding to the angle of internal frictions (φ_c) and cohesions (C_c) [5]. The embankment soil has the thickness H and self weight γ. The ground water table is constant at the surface of the soft ground. The critical output of FEA is the vertical pressure on the surface of the soft ground with sand drain p_1. Designating p_2 as the vertical pressure on the surface of the soft ground without sand drain ($p_2 = \gamma . H$), the K ratio of p_1/p_2 can be written as:

$$K = f(E_c, c_c, \varphi_c, E_{dy}, c_{dy}, \varphi_{dy}, D, L, H, \gamma) \quad (3)$$

To develop a simplified model, variable selection is conducted to reduce the features space. The strong correlation of E_c φ_c, c_c for sand drain and E_{dy} φ_{dy}, c_{dy} for soft soil leads to the represent of E_c and E_{dy} only is sufficient for model developing. Also, because of the the appearance of H and γ in both p_1 and p_2 (i.e. the normalization characteristic of K), these variables can be removed in the predicting model. Consequently, the f function in Eq.3 can be condensed as:

$$K = f(E_c, D, L, E_{dy}) \quad (4)$$

Table 1 provides the database composed by all input variables and the p_1 obtained by 144 runs on Plaxis. The predicted variable (K) then calculated by p_1/p_2.

Table 1. Database obtained from FEA with Plaxis

#	E_c	φ_c	c_c	D	L	E_{dy}	φ_{dy}	c_{dy}	H	γ	p_1	$p_2=\gamma.H$	$(p_1-p_2)/p_1$	K
	kPa	Degree	kPa	m	m	kPa	Degree	kPa	m	kN/m^3	kPa	kPa	%	
1	30000	35.0	1.0	0.35	1.55	2000	8.30	8.00	3.00	20.00	-56.81	-60.00	5.32	0.9468
2	35000	36.5	1.5	0.35	1.55	2000	8.30	8.00	3.00	20.00	-56.50	-60.00	5.83	0.9417
3	40000	38.0	2.0	0.35	1.55	2000	8.30	8.00	3.00	20.00	-56.17	-60.00	6.38	0.9362
4	45000	39.0	2.5	0.35	1.55	2000	8.30	8.00	3.00	20.00	-55.94	-60.00	6.77	0.9323
5	50000	40.0	3.0	0.35	1.55	2000	8.30	8.00	3.00	20.00	-55.71	-60.00	7.15	0.9285
6	30000	35.0	1.0	0.35	1.55	4100	14.00	20.00	3.00	20.00	-53.22	-60.00	11.30	0.8870
7	35000	36.5	1.5	0.35	1.55	4100	14.00	20.00	3.00	20.00	-51.98	-60.00	13.37	0.8663
8	40000	38.0	2.0	0.35	1.55	4100	14.00	20.00	3.00	20.00	-50.81	-60.00	15.32	0.8468
9	45000	39.0	2.5	0.35	1.55	4100	14.00	20.00	3.00	20.00	-49.73	-60.00	17.12	0.8288
10	50000	40.0	3.0	0.35	1.55	4100	14.00	20.00	3.00	20.00	-48.92	-60.00	18.47	0.8153
11	40000	38.0	2.0	0.35	1.40	4100	14.00	20.00	3.00	20.00	-49.12	-60.00	18.13	0.8187
12	40000	38.0	2.0	0.35	1.75	4100	14.00	20.00	3.00	20.00	-52.82	-60.00	11.97	0.8803
13	40000	38.0	2.0	0.35	2.45	4100	14.00	20.00	3.00	20.00	-56.12	-60.00	6.47	0.9353
14	40000	38.0	2.0	0.35	3.15	4100	14.00	20.00	3.00	20.00	-57.74	-60.00	3.77	0.9623
15	40000	38.0	2.0	0.35	3.85	4100	14.00	20.00	3.00	20.00	-58.59	-60.00	2.35	0.9765
16	40000	38.0	2.0	0.35	4.55	4100	14.00	20.00	3.00	20.00	-59.08	-60.00	1.53	0.9847
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
134	30000	35.0	1.0	0.45	3.60	4100	14.00	20.00	3.00	20.00	-58.09	-60.00	3.19	0.9682
135	35000	36.5	1.5	0.45	3.60	4100	14.00	20.00	3.00	20.00	-57.60	-60.00	4.00	0.9600
136	40000	38.0	2.0	0.45	3.60	4100	14.00	20.00	3.00	20.00	-57.12	-60.00	4.79	0.9521
137	45000	39.0	2.5	0.45	3.60	4100	14.00	20.00	3.00	20.00	-56.66	-60.00	5.57	0.9443
138	50000	40.0	3.0	0.45	3.60	4100	14.00	20.00	3.00	20.00	-56.20	-60.00	6.34	0.9366
139	35000	36.5	1.5	0.35	1.55	4100	14.00	20.00	3.00	20.00	-51.88	-60.00	13.53	0.8647
140	35000	36.5	1.5	0.35	1.55	4100	14.00	20.00	3.50	20.00	-60.61	-70.00	13.41	0.8659
141	35000	36.5	1.5	0.35	1.55	4100	14.00	20.00	4.00	20.00	-69.42	-80.00	13.22	0.8678
142	35000	36.5	1.5	0.35	1.55	4100	14.00	20.00	4.50	20.00	-79.07	-90.00	12.14	0.8786
143	35000	36.5	1.5	0.35	1.55	4100	14.00	20.00	5.00	20.00	-88.75	-100.00	11.25	0.8875
144	35000	36.5	1.5	0.35	1.55	4100	14.00	20.00	5.50	20.00	-98.47	-110.00	10.48	0.8952

2.2. The principal of Random Forest Regression

In Random Forest method, each "tree" of the forest refers to the Decision Tree, which splitting the input space of a single feature k after each node of the tree with a threshold t_k into the left and the right subset. A decision tree uses the Classification And Regression Tree algorithm [20] for training with the objective or the empirical risk function in the following form for regression problem

$$\min\left(J\left(k,t_k\right)=\frac{m_{left}}{m}MSE_{left}+\frac{m_{right}}{m}MSE_{right} \right) \tag{5}$$

where: $MSE_{left/right}$ is the Mean Squared Error of the left/right subset, formulation for MSE provided in Table 4; $m_{left/right}$ is the sample of the left/right subset.

Fig.2 provides an illustration of a decision tree. Starting with the root node which has no children with the depth is zero, the children nodes has the depth is d+1 with d is the depth of their parent. The nodes without any children are defined as the leaf nodes. Decision tree in Fig.2 has the maximum depth d_{max} = 3 with 8 leaf nodes.

In Random Forest algorithm, each tree in the forest is a Decision Tree trained by a subset of database randomly chosen from the database. In the database with few features as in this paper (i.e. 4 features), sampling features [16] are not conducted. The random choice is the so-called bagging or bootstrap aggregating if sampling with replacement is conducted. Without the replacement, the sampling is designated as pasting. Results from all Decision Trees are used to aggregate the final prediction in the voting process. The mechanism of Random Forest algorithm is provided in Fig.3.

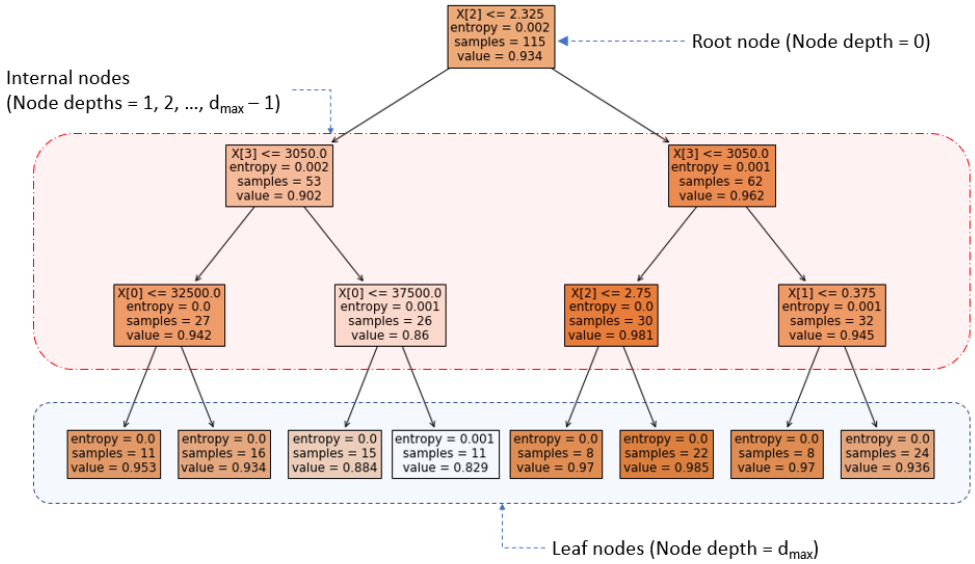


Figure 3. Example of a Decision Tree

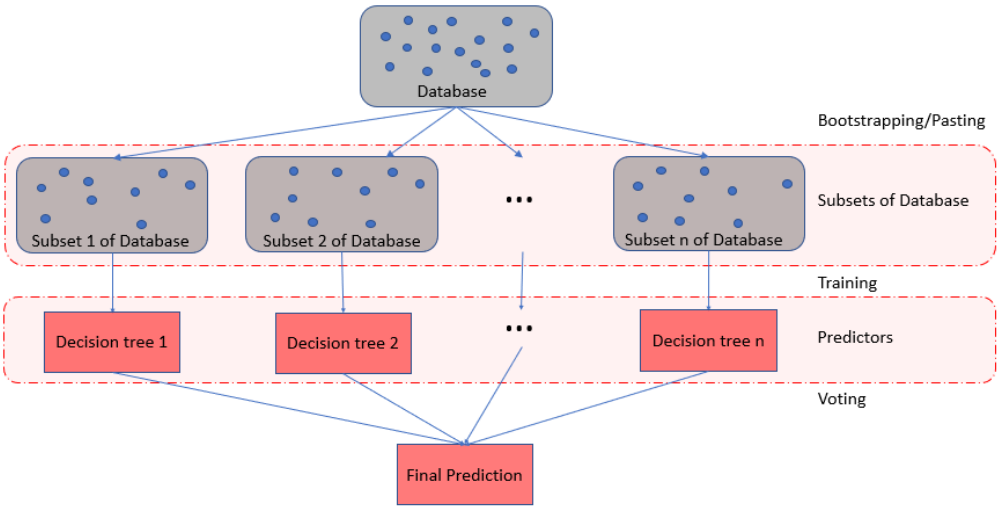


Figure 4. Mechanism of predicting process in Random Forest Regression

The Random Forest Regression is implemented with Python, a programing language. In this process, the overall database is divided into train set and test set with the ratio of 0.8/0.2, respectively. The train set are used for training process in Fig.3 and the test set used for validating the obtained model. The training process includes the grid search, a process of scanning for the appropriate model by gradually changing hyperparameters. Bootstrap or pasting, number of decision trees, minimum samples split, minimum samples leaf and max depth are hyperparameters considered in the fine-tuning process. Explaination of these hyperparameters provided in Fig.3. Feature importance is calculated based on the average depth of node of each feature because the feature closer to the root node tend to be more importance.

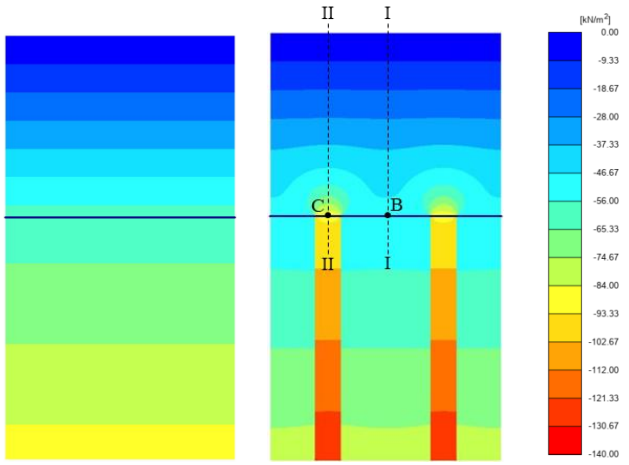
3. Results and Discussion

3.1. Results from FEA

An arbitrary FE simulation is choosen in the database for observation on the stress concentration phenomenon with the D=0.35m, L=1.55m and the material properties provided in Table 2. Fig.4 illustrates the contours of vertical effective stress under the embankment with and without sand drain.

Table 2. Properties of soils in the observation (Fig.5 and 6)

Type of soil	γ (kN/m^3)	φ (Degree)	c (kPa)	E (kPa)
Embankment	20	40	3	50000
Sand (in sand drain)	19	38	2	40000
Soft soil	17	14	20	4100



(a) (b)

Figure 5. The contours map of effective stress σ_z'(kN/m^2) (a) without sand drain; (b) with sand drain

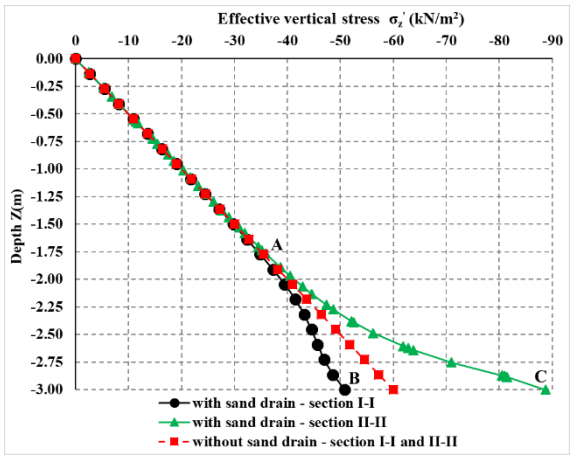


Figure 6. The distribution of vertical effective stress in section I-I and II-II

Fig.5 illustrates the contours map of stress distribution in soil. The distribution of the vertical effective stress along the cross section I-I and II-II (Fig.1 and 5) due to the depth provided in Fig.6. Base on the data from FEA (Table 1), the contour plot of K ratio with the variation of L/D and $E_c/E_đ$ is established in Fig.7.

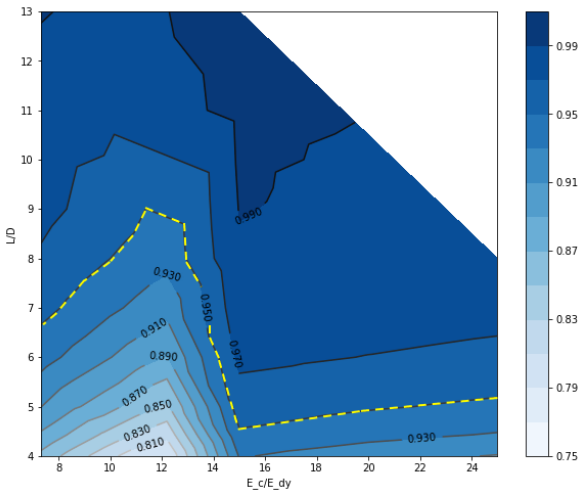


Figure 7. Contours of K ratio vesus L/D and $E_c/E_đ$

3.2. Results of RFR

The set of candidate hyperparameters for the model provided in Table 3 for the grid search. The grid search is conducted twice, the first search scanning for loose grid of hyperparameters. Once the first search finished, the second search is implemented with finer interval of hyperparameters distributed

around the best hyperparameters found in the first search. The "best" set of parameters found based on its lowest MSE on the training set. Table 4 provides the MSE of the final model along with 2 other evaluation metrics, the Mean Absolute Error, MAE and the Coefficient of determination, R^2. Fig.8 illustrates the plot of simulated K versus predicted K. Fig.9 provides the results of the feature importance analysis.

Table 3. Grid search for hyperparameter tuning

Hyperparameter	Explaination	First grid search	"Best" hyperparameter*	Second grid search	"Best" hyperparameter**
Bootstrap	Whether bootstrap samples used	True, False	True	True	True
n_estimators	Number of decision trees	100, 1000, 10000	100	50, 100, 200, 300, 400	100
min_samples_split	The minimum number of samples required to split an internal node	2, 5, 10	2	1, 2, 3, 4	1
min_samples_leaf	The minimum number of samples required to be at a leaf node	1, 2, 4	2	2	2
max_depth	maximum depth of the tree (d_{max})	10, 20, 30, 40, 50, None	40	30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50	36

* After First grid search
** After Second grid search

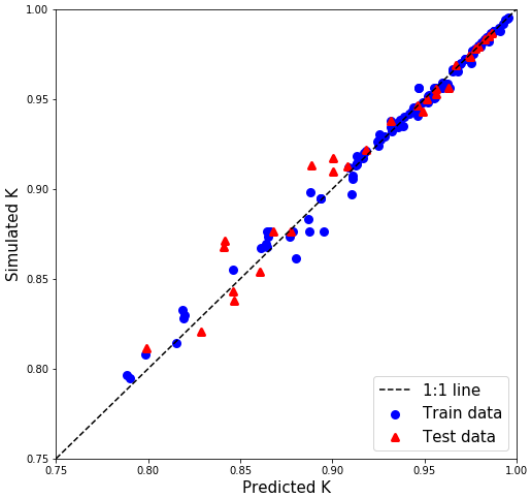


Figure 8. Scatter plot of simulated K versus predicted K.

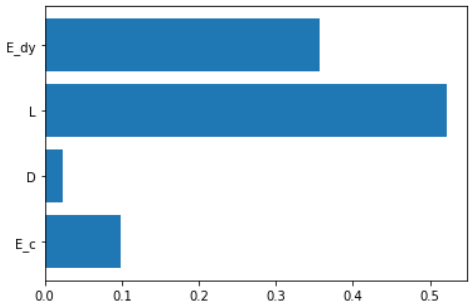


Figure 9. Feature importance of input variables

Table 4. Evaluation metrics for chosen model

Metric	Equation	Result on train set	Result on test set
Mean squared error	$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2$	0.00003	0.00012
Mean absolute error	$MAE = \frac{1}{n}\times\sum_{i=1}^{n}\lvert y_i - f_i\rvert$	0.00309	0.00711
Coefficient of determination	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$	0.98983	0.96156

Table 5. Input variables and K from test set (FEA) versus predicted K from RFR

#	E_c	D	L	E_{dy}	K_test (simulated with FEA)	K_pred (predicted with RFR)
1	40000	0.350	1.550	4100	0.8468	0.8317
2	35000	0.400	1.600	4100	0.8411	0.8699
3	50000	0.400	2.000	4100	0.8459	0.8460
4	30000	0.400	2.400	2000	0.9749	0.9735
5	40000	0.400	3.200	4100	0.9516	0.9518
6	30000	0.450	3.150	4100	0.9565	0.9563
7	45000	0.450	3.150	2000	0.9777	0.9783
8	35000	0.400	2.000	2000	0.9563	0.9538
9	35000	0.450	3.150	4100	0.9462	0.9462
10	30000	0.400	1.600	2000	0.9317	0.9361
11	45000	0.450	2.700	4100	0.9006	0.9173
12	40000	0.450	2.250	4100	0.8772	0.8707
13	50000	0.400	3.200	2000	0.9830	0.9835
14	50000	0.450	1.800	2000	0.9085	0.9123
15	45000	0.450	2.700	2000	0.9673	0.9687
16	40000	0.400	2.800	2000	0.9791	0.9794
17	45000	0.400	2.000	2000	0.9492	0.9453
18	45000	0.350	1.550	4100	0.8288	0.8208
19	50000	0.450	2.700	4100	0.8887	0.9107
20	30000	0.450	2.700	2000	0.9745	0.9736
21	30000	0.350	1.550	2000	0.9630	0.9561
22	45000	0.450	1.800	4100	0.7991	0.8181
23	35000	0.400	3.200	2000	0.9864	0.9866
24	45000	0.400	2.000	4100	0.8608	0.8608
25	40000	0.450	1.800	2000	0.9185	0.9212
26	35000	0.450	1.800	4100	0.8416	0.8736
27	40000	0.450	3.600	2000	0.9851	0.9852
28	35000	0.350	1.550	4100	0.8678	0.8764
29	45000	0.400	2.400	4100	0.9002	0.9116

3.3. Discussion

As can be seen in Fig.5, results from FEA matched with the intuition of the stress decrease in soft ground and increase in sand drain. This conclusion is further supported by Fig.6. From point 0, the lines presented for relationship between vertical stress versus depth of section I-I and II-II overlapped with this of the embankment self weight (i.e. $\gamma \times H$). The seperation occurred at point A (z=-1.75m, σ'_z=-35 kN/m2) with the appearance of the stress concentration around the head of the sand drain due to the difference in moduluses of sand and soft soil. The stress versus depth relationship turned into nonlinear with a decrease in section I-I and an increase in II-II. To the end (z=-3m), the stress at section II-II is almost double the stress at section I-I (i.e. 89 kN/m^2 versus 51 kN/m^2). The vertical pressure at the surface of the soft soil reduced from 60 kN/m^2 to 51 kN/m^2 compared to the hydrostatic pressure or K=0.8235.

Statistical analysis on database in Table 1 shows a relatively wide range of K from 0.793 to 0.995. This ratio has a strong positive correlation with L/D as can be seen in Fig.7. The dashed contour presented for the boundary of K larger than 0.95, which can be used as the suggestion of considering to the effect of pressure reduction on soft ground. This threshold can be subjectively choosen for practical purposes.

For the developed RFR model, it is reasonable for the error of the model (i.e. MSE and MAE) on the train set relatively low compared to the predicted variable of K which ranges within 0.793 to

0.995. The model also has the low errors on the test set implies the regulation or the successfully avoidance of overfitting of the chosen model. With such low errors, the model reasonably has high R^2 which closes to the maximum of 1. The scatter plot in Fig.8 illustrates the success of the model with the narrow scattering of the datapoints around the 1:1 line. Table 5 provides a detailed comparison of predicted K from the RFR versus the K from simulation with FEA on the test set. It is noted that the test set is not contributed to the training process or the developed model has not "seen" these data before.

Fig.9 provides the results of the feature importance analysis with L is ranked first as the most effect variable to the K with the importance level of 0.5129. It is reasonably to rank the Modulus of soft soil E_{dy} to the second important variable and followed by the Modulus of sand drain (E_c). Diameter of the sand drain has the least effect to K with the importance level of 0.1025.

4. Conclusions

The paper focuses on the pressure reduction on the surface of the soft ground with sand drain. FEA and RFR are incorporated to develop the regression model for predicting the ratio of pressure on the soft soil surface with and without sand drain (K). An observation on a random run is provided with discussion in detail to reinforce the initial intuition. Simulation with FEA provides a sufficient database for predicting K with R^2 on the test set of database up to 0.9616. Important factors from RFR reveals the critical role of distance between sand drains and the least impact of diameter of sand drain to K.

Inputs of the machine learning model (i.e. E_c, E_{dy}, L, D) can be easily collected and made the model applicable. Prediciting the decrease factor of pressure on the surface of soft ground with and without sand drain is useful for the practical consolidation problem. The study proposes to consider this reduction along with the conventional approach of focused on consolidation merely. Results can be the preliminary work for implementing the guidance of finding reduction factor K in standards or for practical design.

References

[1] Lekha K, Krishnaswamy N and Basak P 1998 Consolidation of clay by sand drain under time-dependent loading *J. Geot. and Geoenv. Eng., ASCE* **124**(1) 91-94
[2] Wan-Huan Z and Shuai T 2012 Unsaturated consolidation in a sand drain foundation by differential quadrature method *Procedia Earth and Planetary Science* **5**(2012) 52-57
[3] Koy C and Yune C Y 2017 Numerical Analysis on Consolidation of Soft Clay by Sand Drain with Heat Injection *Journal of the Korean Geotechnical Society*, **33**(11) 45-57
[4] Ministry of Transport of Vietnam 2000 *22TCN262:2000 Survey and design procedures for motorway sub-grade on soft* (Hanoi: Transport Publishing House) (in Vietnamese)
[5] MOST of Vietnam 2017 *TCVN 11713:2017 Soft ground improvement by sand drains - Construction and acceptance* (Hanoi: Construction Publishing House) (in Vietnamese)
[6] Ramkrishnan R, Karthik V, Unnithan M S, Kiran Balaji R, Athul Vinu M and Venugopalan A 2017 Stabilization of seepage induced soil mass movements using sand drains *Geotech. Eng. J. SEAGS & AGSSEA* **48**(4) 129-137
[7] Mahmood M S, Akhtarpour A and Alali A A A 2019 Mechanical Behavior of Dam Foundation with Vertical Sand Drain, Case Study: Sombar Dam *Journal of Engineering and Technological Sciences* **51**(3) 380-391
[8] Pham T D, Cao H V and Tran CV 2019 Observation of embankment pressure decrease for sand pile *Journal of Science and Technology for Civil Engineering* **1** 63-69 (in Vietnamese)
[9] MOST of Vietnam 2012 *TCVN 9362:2012 Specifications for design of foundation for buildings and structures* (Hanoi: Construction Publishing House) (in Vietnamese)
[10] Kézdi Á and Rétháti L 1974 *Handbook of soil mechanics* vol 1 (Amsterdam: Elsevier)

[11] Mintransstroya SSSR 1989 *Guide to SNiP 2.05.02-85 Guide to the design of the subgrade of roads on soft soils* Put into effect by an order of the Ministry of transport of Russia No OS-1067-R from 03.12.2003 (Moscow: Publishing house 'Construction') (in Russian)
[12] Evgeniev I E 1968 *Construction of highways through swamps* (Moscow: Publishing house 'Transport') pp 166-167 (in Russian)
[13] Dobrov E M, Bao Nguyen Nhu 2017 Accounting the effect of hovering of bulk soil while applying the flexible pile caps on soft road foundations *Journal of science and technology in the road industry* **1** 21-23 (in Russian)
[14] Géron A 2019 *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (California: O'Reilly Media)
[15] Ho T K 1995 Random decision forests *Pro. of 3rd Int. Conf. on document analysis and recognition (Montreal)* (Montreal: IEEE) p 278
[16] Ho T.K 1998 The random subspace method for constructing decision forests *IEEE Trans. Pattern Anal. Mach. Intell* **20**(8) 832-844
[17] Kohestani V, Hassanlourad M and Ardakani A 2015 Evaluation of liquefaction potential based on CPT data using random forest *Natural Hazards* **79**(2) 1079-1089.
[18] Zhou J, Shi X, Du K, Qiu X, Li X, and Mitri H S 2017 Feasibility of random-forest approach for prediction of ground settlements induced by the construction of a shield-driven tunnel *International Journal of Geomechanics* **17**(6) 04016129
[19] Matin S, Farahzadi L, Makaremi S, Chelgani S C and Sattari G 2018 Variable selection and prediction of uniaxial compressive strength and modulus of elasticity by random forest *Applied Soft Computing* vol 70 (Amsterdam: Elsevier) pp 980-987
[20] Breiman L, Friedman J, Stone C J and Olshen R A 1984 Classification and regression trees *Classification and regression trees* (Monterey: Brooks/Cole Publishing)