Hindawi
Advances in Civil Engineering
Volume 2021, Article ID 3230343, 9 pages
https://doi.org/10.1155/2021/3230343

Research Article

Liquefaction Evaluation Based on Shear Wave Velocity Using Random Forest

Lu Liu,[1,2] Shushan Zhang,[1,2] Xiaofei Yao,[1,2] Hongmei Gao,[1,2] Zhihua Wang,[1,2] and Zhifu Shen[1,2]

[1]*Urban Underground Space Research Center, Nanjing Tech University, Nanjing 211816, China*
[2]*Institute of Geotechnical Engineering, Nanjing Tech University, Nanjing 211816, China*

Correspondence should be addressed to Hongmei Gao; hongmei54@163.com

Received 27 August 2021; Revised 8 November 2021; Accepted 9 November 2021; Published 8 December 2021

Academic Editor: Xinyu Ye

Copyright © 2021 Lu Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Liquefaction evaluation on the sands induced by earthquake is of significance for engineers in seismic design. In this study, the random forest (RF) method is introduced and adopted to evaluate the seismic liquefaction potential of soils based on the shear wave velocity. The RF model was developed using the Andrus database as a training dataset comprising 225 sets of liquefaction performance and shear wave velocity measurements. Five training parameters are selected for RF model including seismic magnitude (M_w), peak horizontal ground surface acceleration ($a_{\max}$), stress-corrected shear wave velocity of soil (V_{s1}), sandy-layer buried depth (d_s), and a new introduced parameter, stress ratio (k). In addition, the optimal hyperparameters for the random forest model are determined based on the minimum error rate for the out-of-bag dataset (ERR_{OOB}) such as the number of classification trees, maximum depth of trees, and maximum number of features. The established random forest model was validated using the Kayen database as testing dataset and compared with the Chinese code and the Andrus methods. The results indicated that the random forest method established based on the training dataset was credible. The random forest method gave a success rate for liquefied sites and even a total success rate for all cases higher than 80%, which is completely acceptable. By contrast, the Chinese code method and the Andrus methods gave a high success rate for liquefaction but very low for non-liquefaction which led to the increase of engineering cost. The developed RF model can provide references for engineers to evaluate liquefaction potential.

1. Introduction

Soil liquefaction occurs if saturated sand suffers loss of strength and modulus due to the increase of excess pore pressure when subjected to strong earthquake loading. It was not until the 1964 Niigata earthquake in Japan and the Alaska earthquake in the United States that people fully realized the harm of sand liquefaction. These two earthquakes caused the damage of many buildings and the loss of people's lives and property, which highlighted the need to study the liquefaction potential of saturated sand.

Many scholars have presented different methods for liquefaction evaluation based on in-situ tests such as shear wave velocity (V_s) [1–5], standard penetration test (SPT) [6–9], and cone penetration test (CPT) [10–13]. Compared with SPT-based and CPT-based liquefaction evaluation methods, the advantage of V_s-based methods is that they are less sensitive to soil characteristics [4]. In addition, shear wave velocity tests can be easily performed in soils while SPT and CPT tests are limited by the soil type such as gravelly sands. These advantages make the V_s-based method irreplaceable and have a bright development prospect for liquefaction evaluation. At present, the most widely accepted liquefaction evaluation method based on shear wave velocity is proposed by Andrus and Stokoe [1] by establishing the liquefaction assessment curves based on the correlation between *CRR* (liquefaction resistance) and V_s. Kayen et al. [2] further presented updated V_s-based liquefaction assessment curves based on 422 case histories and concluded that the correlation developed between *CRR* and V_s is

insensitive to fines content (FC). Shen et al. [14] established an updated liquefaction assessment curves based on 261 cases histories combining the Andrus database with new collected cases from Canterbury earthquake. However, these V_s-based methods are commonly empirical or semiempirical and are easily limited by local data. In addition, there are several factors influencing the liquefaction evaluation such as sandy-layer buried depth (d_s), ground water table (d_w), and peak horizontal ground surface acceleration ($a_{\max}$). Unfortunately, the existing V_s-based methods cannot clearly explain the relationship between influencing factors and liquefaction potential.

With the continuous development of computer technology, machine learning methods such as adaptive neuro fuzzy inference system (ANFIS), artificial neural network (ANN), and support vector machine (SVM) make it possible to solve the above problems. The advantage of machine learning method is that it is not necessary to consider the relationship between input and output variables and can obtain the accurate prediction by relying on the collected data itself. Xue and Yang [15] adopted the ANFIS model for the assessment of liquefaction potential, which provided more accurate results than traditional empirical methods including seed simplified methods [16] and CPT-based method proposed by Stark and Olson [17]. Hanna et al. [18] established an ANN model for liquefaction evaluation utilizing 12 parameters related to soil and seismic characteristics. Zhao et al. [19] employed the SVM method to assess the soil liquefaction based on SPT and CPT data through the particle swarm optimization (PSO) [20], for searching the kernel functions and training parameters. These methods get a satisfactory result compared with traditional empirical methods, but there still exist some shortcomings. For example, ANN and ANFIS approaches are time consuming in terms of the optimal selection of parameters due to the fact that the number of training parameters is excessive and the model is easy to fall into overfitting. The SVM method is difficult to operate large-scale training data and is sensitive to the choice of parameters and kernel function. At present, there is still no effective method to solve this problem. Moreover, the machine learning models established by single dataset may not perform well in other liquefaction datasets [21]. Kohestani et al. [22] also stated that machine learning methods have a limited domain of applicability and are mostly case dependent. Therefore, it is necessary to improve these existing methods or seek for other more advanced methods for liquefaction evaluation.

The random forest (RF) is an ensemble learning algorithm developed by Breiman [23] based on a combination of a large set of decision trees. The advantage of the random forest is that it is simpler in selection of hyperparameter and can solve the overfitting problem [22, 24]. RF method has been successfully used for solving the geotechnical engineering such as landslide [25], ground surface settlements [24, 26], the prediction of soil shear strength [27], and bearing capacity of foundations [28]. However, few studies have been reported about the RF model applications in the liquefaction evaluation. Kohestani et al. [22] reported the evaluation of liquefaction potential based on CPT data using RF method. Nejad et al. [29] established a RF model for predicting the occurrence or nonoccurrence of liquefaction based on the shear wave velocity data collected by Kayen et al. [2]. However, these liquefaction models using RF cannot be well verified by other datasets since limited dataset was utilized by separating training dataset from testing set randomly.

In this study, the liquefaction evaluation was conducted based on shear wave velocity by using the random forest. A total of 225 cases from Andrus et al. [1] and 336 cases from Kayen et al. [2] with respect to liquefaction and nonliquefaction histories are used as training dataset and testing dataset, respectively. The performance of the model using the RF method for liquefaction evaluation proposed in this study is compared with the Chinese code method [30] and the method proposed by Andrus et al. [1], respectively.

2. Method

2.1. Random Forest. The random forest (RF) is an intelligent recognition method based on statistical learning theory [23]. Many predictors are generated based on the strategy of ensemble learning, which can be applied for solving classification and regression problems by classification tree and regression tree, respectively. The purpose of this study is to evaluate the liquefaction potential of soil. There are two results for liquefaction evaluation including liquefaction or nonliquefaction; thus, classification tree is adopted herein. Figure 1 shows the general architecture of the random forest for classification, where X is the training dataset, n is the number of classification trees (represented by *n_estimator* in machine learning tool, Scikit-learn), and *max_depth* is the maximum depth of the tree which plays an important role in controlling the complexity and size of the tree. The construction process of the random forest is summarized as follows.

(1) New training subsets X_n are generated by randomly drawing samples with replacement from original training dataset (X). Each training subset contains approximately two-thirds of the elements of X (called bootstrap samples) and the remaining elements are called out-of-bag (OOB) samples. OOB samples can be used to evaluate the generalization ability of the random forest model based on the calculation of OOB error rate (ERR_{OOB}) for training subsets, which can be also utilized to determine the optimal hyperparameters of the random forest model [28] such as *n_estimator*, *max_depth*, and *max_features* (the number of training parameters in random subset at each node).

(2) For each subset X_n, the classification tree is growing. At each node of the tree, rather than choosing the best spilt among all training parameters in classification tree, *max_features* training parameters is randomly selected and the best spilt is chosen among them based on calculating the value of Gini index as follows [31, 32]:

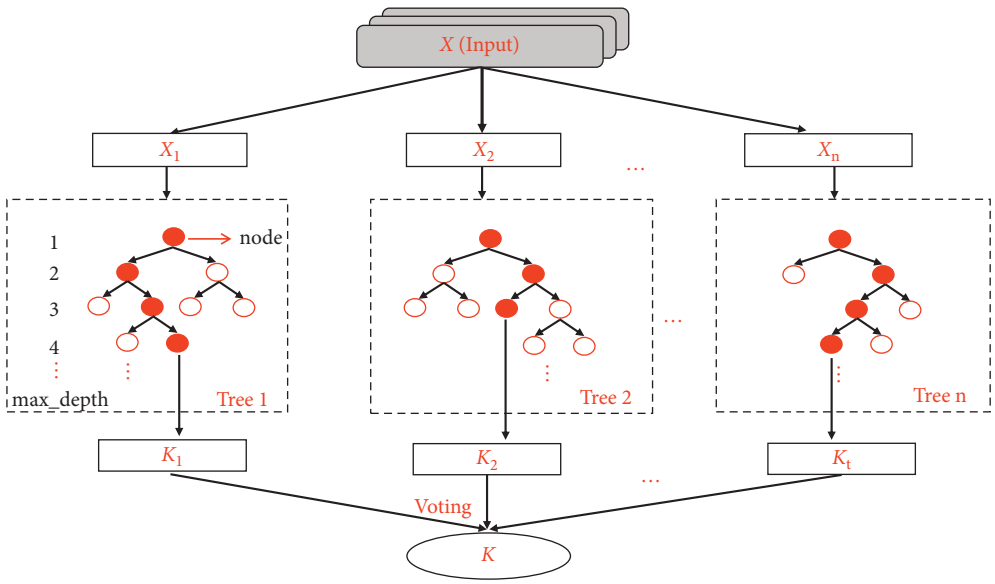

Figure 1: A general architecture of the random forest.

$$\mathrm{Gini}(X_n) = 1 - \sum_{i=1}^{t} P_i^{\,2}, \tag{1}$$

where P_i represents the possibility of class K_t in X_n and t is the number of the classes which is 2 for liquefaction evaluation since the classification result only has two types: liquefaction or nonliquefaction.

(3) The random forest produces classification results K_t for each classification tree. The final classification result (K) is obtained based on the voting results by following the principle that the minority is subordinate to the majority [22].

2.2. Database. The database published by Andrus et al. [1] (hereafter called the Andrus database) was used as training dataset in this study to develop the liquefaction evaluation model using random forest. Numerous scholars have used the Andrus database to develop the new model for liquefaction evaluation and the authority of Andrus database is widely recognized [5, 33]. The fundamental parameters such as M_w (moment magnitude), d_s (sandy-layer buried depth), d_w (depth of ground water table), V_{s1} (overburden pressure-corrected shear wave velocity), and $a_{\max}$ (peak horizontal ground surface acceleration) are provided in Table 1. As shown in Table 1, the Andrus database collected 225 cases of shear wave velocity data from 26 earthquakes and more than 70 sites, including 96 liquefaction and 129 nonliquefaction cases. According to the Chinese seismic intensity table [34], the dataset is divided into seismic intensities VI, VII, VIII, and IX, as shown in Table 2.

The database published by Kayen et al. [2] (hereafter called the Kayen database), given in Table 3, also consists of soil and seismic parameters including M_w, d_s, d_w, V_{s1}, and $a_{\max}$. The Kayen database was used as testing dataset which consists of 415 case studies collected from 256 sites of nine earthquakes mainly distributed in Asia, Greece, the United States, and China. Excluding 79 sets of cases which were duplicated with the Andrus database, a total of 336 cases were used as testing dataset and the classification is shown in Table 4.

2.3. Selection of Training Parameters. The accuracy of the random forest method for liquefaction evaluation is highly related to the selection of influence parameters, which can be divided into three categories including the intensity of ground motion, buried condition of soil layer, and compactness of soil. There are 9 liquefaction influencing parameters in the Andrus and Kayen database. Among these parameters, the seismic magnitude (M_w), peak horizontal ground surface acceleration ($a_{\max}$), and cyclic stress ratio (CSR) reflect the intensity of ground motion; the sandy-layer buried depth (d_s), depth of water table (d_w), total vertical stress of soil at the depth considered (σ_r), effective vertical stress of soil at the same depth ((σ_r'), and shear stress reduction factor (r_d) reflect the buried conditions of soil layer; the stress-corrected shear wave velocity (V_{s1}) characterizes the compactness of soils. Some of these parameters are coupled with each other and have strong correlation which leads to the inaccuracy of the model. Therefore, it is necessary to analyze the correlation of these parameters and select the training parameter that is less relevant to the others to establish the random forest model.

Figure 2 shows the correlation matrix of these parameters. It can be seen that four parameters d_s, d_w, σ_r, and σ_r' that reflect the buried condition of soils have strong correlation, which will reduce the accuracy of the model. Therefore, a new parameter k is introduced herein to reflect the buried condition of soil layer. Yao et al. [35] proposed the stress ratio, k, to establish a logistic regression model for liquefaction evaluation and achieved satisfying evaluating results. The stress ratio, k, is defined as

Table 1: Database published by Andrus et al. [1].

No.	Earthquake	M_w	No. of cases	d_s (m)	d_w (m)	V_{s1} (m/s)	$a_{\max}$ (g)
1	1906 San Francisco	7.7	12	4.2~8	2.4~6.1	131~168	0.32~0.36
2	1957 Daly City	5.3	5	3.5~7.9	2.7~5.9	105~220	0.11
3	1964 Nigita Japan	7.5	4	3.2~6.2	1.2~5	112~162	0.16
4	1975 Haicheng China	7.3	6	3~10.2	0.5~1.5	98~147	0.12
5	1979 Imperial Valley, California	6.5	11	3~4.7	1.5~2.7	90~173	0.12~0.51
6	1980 Mid-Chiba Japan	5.9	2	6.1~14.8	1.3	155~195	0.08
7	1981 Westmorland, California	5.9	11	3~4.7	1.5~2.4	90~173	0.02~0.36
8	1983 Borah Peak, Idaho	6.9	18	1.9~3.7	0.8~3	94~274	0.23~0.46
9	1985 Chiba-Ibaragi-Kenkyo, Japan	6.0	2	6.1~14.8	1.3	155~195	0.05
10	1986 Taiwan (event Lsst2)	5.3	4	5.3~6.1	0.5	127~156	0.05
11	1986 Taiwan (event Lsst3)	5.5	4	5.3~6.1	0.5	127~156	0.02
12	1986 Taiwan (event Lsst4)	6.6	4	5.3~6.1	0.5	127~156	0.22
13	1986 Taiwan (event Lsst6)	5.4	4	5.3~6.1	0.5	127~156	0.04
14	1986 Taiwan (event Lsst7)	6.6	4	5.3~6.1	0.5	127~156	0.18
15	1986 Taiwan (event Lsst8)	6.2	4	5.3~6.1	0.5	127~156	0.04
16	1986 Taiwan (event Lsst12)	6.2	4	5.3~6.1	0.5	127~156	0.18
17	1986 Taiwan (event Lsst13)	6.2	4	5.3~6.1	0.5	127~156	0.08
18	1986 Taiwan (event Lsst16)	7.6	4	5.3~6.1	0.5	127~156	0.14
19	1987 Chiba-Toho-Oki, Japan	6.5	1	9	6.2	150	0.1
20	1987 Elmore Ranch, California	5.9	11	3.4~4.7	1.5~2.7	90~173	0.03~0.24
21	1987 Superstition Hills, California	6.5	11	3.0~4.7	1.5~2.7	90~173	0.18~0.21
22	1989 Loma Prieta, California	7.0	67	2.3~9.9	0.6~6.1	91~209	0.1~0.45
23	1993 Kushiro –Oki Japan	8.3	2	4.2~4.5	0.9~1.9	135~152	0.41
24	1993 Hokkaido-Nansei-Oki, Japan	8.3	4	2.0~7.0	1.0~1.4	74~143	0.15~0.19
25	1994 Northridge, California	6.7	3	4.4~5.6	3.4	129~160	0.51
26	1995 Hyogo-Ken Nanbu, Japan	6.9	19	3.3~11.5	1.5~7	110~214	0.12~0.65

Table 2: Data classification based on the Andrus database [1].

Seismic intensity	L	NL	Total
VI	0	34	34
VII	23	46	69
VIII	35	35	70
IX	38	14	52
Total	96	129	225

Note. L represents liquefied sites; NL represents nonliquefied sites.

$$k = \sigma_r / \sigma_r'. \tag{2}$$

It can be shown from Figure 2 that, compared with d_w, k is less relevant to σ_r and σ_r'. In addition, r_d only depends on the buried depth of sand, d_s. Thus, the stress ratio, k, and sandy-layer buried depth, d_s, are selected as training parameters for the random forest model to reflect the buried condition of soil layer. *CSR* and $a_{\max}$ are two highly correlated factors, while $a_{\max}$ is more easily obtained than *CSR*. Thus, $a_{\max}$ and M_w are selected to reflect the intensity of ground motion. Moreover, the corrected shear wave velocity of soil, V_{s1}, reflects the compactness of soil. Shear wave velocity are of great significance for the establishment of the liquefaction evaluation model. In summary, five factors including d_s, k, $a_{\max}$, M_w, and V_{s1} are used as training parameters in the random forest model. Table 5 shows the statistical information of training parameters used for model training and normal distribution is selected for these parameters.

2.4. Optimizing the Random Forest Hyperparameters. The best prediction accuracy of the random forest model can be obtained by tuning hyperparameters. As presented in Section 2.1, three hyperparameters need to be optimized including *n_estimator*, *max_depth*, and *max_features*. The value of ERR_{OOB} for each combination of hyperparameters was calculated, and we chose the optimal combination for three hyperparameters corresponding to minimum ERR_{OOB} [36]. The *n_estimator* was initially set as 1–100 and *max_depth* was set as 1–20. Usually, *max_features* < M (the number of training parameters), which is 5 in this study as discussed above. The default *max_features* is $[\log_2(M) + 1]$ [36] and then is decreasing and increasing until the minimum ERR_{OOB} is obtained. The optimal hyperparameters were determined as *max_features* = 2, *max_depth* = 7, and *n_estimator* = 13. The same optimal hyperparameters were used in the random forest model to evaluate the liquefaction potential based on the new Kayen database [2]. The predicted result can be produced by using python programming language in Scikit-learn [37].

Table 3: Database published by Kayen et al. [2].

No.	Earthquake	M_w	No. of cases	d_s (m)	d_w (m)	V_{s1} (m/s)	$a_{\max}(g)$
1	1906 San Francisco	7.7	2	4.75~9.85	2.4~6.1	172~172.1	0.32~0.366
2	1948 Fukui	7.1	11	1.1~9.5	0.5~1.3	134~312.5	0.5
3	1964 Niigata Japan	7.5	9	3.6~10.5	1.2~5	122.2~170.6	0.16~0.18
4	1968 Tokachi Oki	7.9	5	2.75~6.5	0.8~1	109.9~148.5	0.15~0.23
5	1973 Miyagi Ken Oki	7.4	11	3~9	0.5~2.7	110.2~171.8	0.2~0.32
6	1975 Haicheng	7.1	6	1.9~10.25	0.5~1.5	101.3~152.1	0.12
7	1976 Tangshan	8.0	24	1.35~8.15	0.7~3.7	140.9~336.6	0.2~0.45
8	1978 Miyagi Ken Oki	6.7	8	4.3~7.5	0.5–2.6	140.9~172.2	0.12~0.14
9	1979 Imperial Valley	6.5	11	3~4.65	1.4~2.7	98.3~203.9	0.12~0.51
10	1980 mid Chiba	5.9	2	6.15~14.8	1.4	152.7~153.8	0.08
11	1981 Westmorland	5.9	11	3.15~4.65	1.4~2.7	98.3~203.9	0.02~0.36
12	1983 Nihonkai Chubu	7.7	8	4~7.075	0.4~1.5	123.4~184.2	0.12~0.28
13	1983 Nihonkai Chubu Aftershock	7	2	4.4~6	0.4~1.0	123.4~139	0.12
14	1983 Bora Peak	6.9	24	2~3.65	0.8~3	114.9~328.6	0.23~0.5
15	1986 Lotung LSST	6~6.2	6	3.5	0.5	159.2	0.04~0.18
16	1986 Chiba Ibaragi Kenkyo	6	2	6.15~14.8	1.4	152.7~153.8	0.06
17	1987 Chiba Toho Oki	6.5	1	9.1	6.2	110.8	0.1
18	1987 Superstition Hills	6.5	11	3.15~4.65	1.4~2.7	99.4~211.7	0.15~0.2
19	1987 Elmore Ranch	5.9	11	3.15~4.65	1.4~2.7	98.3~211.7	0.03~0.15
20	1989 Loma Prieta	7	59	2.15~9.85	0.6~6.4	124~239.6	0.11~0.47
21	1993 Kushiro	7.6	8	2.55~9	1.0~2.0	118.5~197.6	0.4~0.46
22	1993 Hokkaido Nansei Oki	7.7	27	1.45~7.5	0.5~3	81.7~213	0.14~0.50
23	1995 Hyogo Nambu	7	85	2.5~18.5	0.5~7	124.3~298.2	0.12~0.60
24	1999 Chi Chi	7.6	14	2.15~12.7	0.5~3	147.4~217.8	0.25~0.70
25	1999 Druce	7.4	2	7.25~7.5	2.4	137.3~146.7	0.4
26	2000 Tottori Seibu	6.8	3	5.6~6.0	1.0~2	100.1~129.5	0.33~0.38
27	2001 Geiyo Hiroshima	6.8	5	2.5~10.5	1.0~3	145.9~179.3	0.25~0.26
28	2002 Denali Fault	7~7.9	9	1.75~2.35	0.5~1	149.1~362.9	0.36~0.39
29	2003 Sanriku Minami	7	11	3.0~9.0	0.5~2.7	110.2~172.2	0.22~0.28
30	2003 Tokachi Oki	7.8	6	4.5~9.0	1.0~2	109.9~197.6	0.09~0.4
31	2003 Tokachi Oki Aftershock	7.1	1	6.5	1.0	109.9	0.13
32	2007 Niigata Chuetsu Oki	6.6	2	3.5~5.0	2.4	136.7~157.8	0.68
33	2008 Achaia Elia	6.5	2	5.5	2.4	191.0~193.4	0.33
34	2011 Tohoku	9.0	2	4.55~4.6	2.4	139.0~139.5	0.27
35	2011 Tohoku Aftershock	7.4	10	2.9~9.0	2.4	111.1~172.7	0.2~0.76

Table 4: Data classification based on the database published by Kayen et al. [2].

Seismic intensity	L	NL	Total
VI	None	4	4
VII	23	17	40
VIII	80	29	109
IX	145	35	180
X	3	None	3
Total	251	85	336

Note. L represents liquefied sites; NL represents nonliquefied sites.

3. Results and Analysis

The random forest model trained by the Andrus database is compared with the Chinese code method and the Andrus method by using the same database. Accordingly, these two methods will be introduced in brief first. Moreover, the established random forest model is validated by evaluating the actual cases and comparing with these two methods by using the Kayen database which was not used in the training process.

3.1. The Chinese Code Method for Liquefaction Evaluation. The critical shear wave velocity, V_{scri}, is used as an index to evaluate soil liquefaction in the Chinese code [30], which is expressed as

$$V_{\text{scri}} = V_{s0}\sqrt{\left(d_s - 0.0133{d_s}^2\right)}\left[1 - 0.185\frac{d_w}{d_s}\right], \quad (3)$$

where V_{s0} = a reference value of shear wave velocity, which is 65 m/s, 90 m/s, and 130 m/s for seismic intensity VII, VIII, and IX, respectively.

When the stress-corrected shear wave velocity, V_{s1}, is smaller than V_{scri}, the site is evaluated as liquefied; otherwise, it is nonliquefied. It should be noted that the Chinese code method does not evaluate the sites distributed in seismic intensity VI and X. Therefore, the cases in seismic intensity VI and X were excluded from comparisons among different methods.

3.2. The Andrus Method for Liquefaction Evaluation. Andrus et al. proposed a method for liquefaction evaluation [1], called the Andrus method hereafter, which is written as

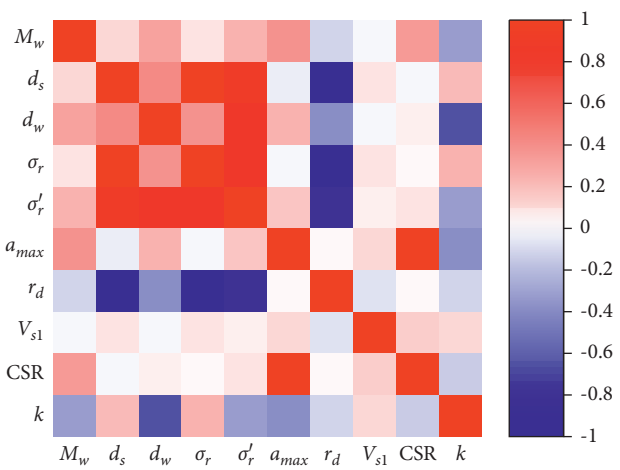

Figure 2: The correlation matrix of training parameters.

Table 5: Statistics of the Andrus dataset for training ($N = 225$) [1].

Training parameters	d_s (m)	k	$a_{\max}$ (g)	M_w	V_{s1} (m/s)
Maximum	15	2.51	0.65	8.3	318
Minimum	1.9	1.07	0.02	5.3	99
Median	5.3	1.41	0.18	6.9	159
Mean	5.46	1.57	0.23	6.72	161.85
Standard deviation	2.22	0.45	0.15	0.65	31.01

$$\mathrm{CSR} = \frac{\tau_{\mathrm{eq}}}{\sigma_r'} = 0.65\frac{a_{\max}}{g}\frac{\sigma_r}{\sigma_r'}r_d,$$
$$\mathrm{CRR} = \left(0.022\left(\frac{V_{s1}}{100}\right)^2 + 2.8\left(\frac{1}{V_{s1}{}^* - V_{s1}} - \frac{1}{V_{s1}{}^*}\right)\right)\mathrm{MSF},$$
$$V_{s1} = V_S\left(\frac{P_a}{\sigma_r'}\right)^{0.25}, \tag{4}$$

where CRR = soil liquefaction resistance, V_s = measured shear wave velocity, P_a is a reference stress of 100 kPa, and V_{s1}^* = limiting upper value of V_{s1}.

Once CSR > CRR, liquefaction is predicted to occur; otherwise, liquefaction does not happen. More details about the method can be referred to reference [1].

3.3. Comparison among Different Methods. It should be noted that 34 sets of cases in seismic intensity VI are removed in order to compare the random forest method with the Chinese code method. Table 6 shows the success rates of liquefaction evaluation based on the Andrus database by different methods. The total success rate is defined as the ratio of site number evaluated successfully to total site number. Take the total success rate of the Chinese code method as an example. As displayed in Table 2, there are 96 liquefied sites and 95 nonliquefied sites in seismic intensity VII, VII, and IX. Of these, 86 liquefied sites and 28 nonliquefied sites are successfully evaluated by the Chinese code method. Thus, the total success rates, S_{total}, is obtained as

Table 6: Success rates given by different methods based on the Andrus database [1].

Evaluation method	L (%)	NL (%)	Total (%)
The Chinese code method	89.6	29.5	59.7
The Andrus method	97.9	34.7	66
The Random forest	100	95.8	96.9

$$S_{\mathrm{total}} = (86 + 28)/(96 + 95) \times 100\% = 59.7\%. \tag{5}$$

For liquefied sites, the success rates using the Chinese code method and the Andrus method are 89.6% and 97.9%, respectively. However, the success rates for nonliquefied sites are only 29.5% and 34.7%, respectively. These two methods with total success rates smaller than 70% tend to be conservative. The random forest method gives a satisfactory total success rate of 96.9%, along with 100% at liquefied sites. To be noted, three methods all give satisfactory success in terms of predicted result of liquefied sites presented in Table 6, which are close to or more than 90%. The random forest method shows a significant difference in the prediction of nonliquefied sites as compared with the other two methods. Many nonliquefied sites are misjudged as liquefied sites by the Chinese code method and the Andrus method while the random forest method correctly classifies most cases of nonliquefied sites. In conclusion, the other two methods are fairly conservative. The random forest method can satisfactorily predict both liquefied sites and nonliquefied sites.

The success rates of liquefaction evaluation obtained by the random forest model based on the Kayen database are depicted in Table 7 by comparing with the Chinese code method and the Andrus method, noting that 7 sets of cases were excluded from the Kayen database. It is concluded that the Chinese code method and the Andrus method give success rate greater than 90% at liquefied sites for most seismic intensities except 78.3% for the Chinese code method at seismic intensity VII. The total success rates for liquefied sites obtained by the random forest method are more than 80%. However, the success rates for nonliquefied sites given by the Chinese code method and the Andrus method are too low. For example, the success rates given by these two methods at seismic intensity VIII and IX are both less than 15%. Therefore, it tends to be conservative and increase the engineering cost if these two methods are used for seismic design. For liquefied sites, the Chinese code method and the Andrus method give the total success rates of more than 95% (regardless of seismic intensity), and the random forest method gives a value of 83.1% which has reached an acceptable level. However, the total success rates for nonliquefied sites given by the Chinese code method and the Andrus method are 23.5% and 15.4%, respectively, while 75.3% is given by the random forest method which is far higher. Moreover, the total success rate given by the random forest method for all cases is 81.2% which is the highest of all methods. It can be inferred that the random forest method is effective and reliable. Moreover, the random forest method developed in this study differs from the conventional studies which divided a database into 70% as the training dataset

Table 7: Success rates given by different methods based on Kayen database [2].

Seismic intensity	L/NL/Total	The Chinese code method	The Andrus method	The random forest
VII	L	78.3	100	78.3
	NL	52.9	52.9	94.1
	Total	67.5	80.0	85.0
VIII	L	97.5	98.8	85.0
	NL	13.8	3.4	62.1
	Total	75.2	73.4	78.9
IX	L	100	94.5	82.8
	NL	17.1	11.4	77.1
	Total	83.9	78.3	81.7
Total	L	97.2	96.4	83.1
	NL	23.5	15.4	75.3
	Total	79.0	76.9	81.2

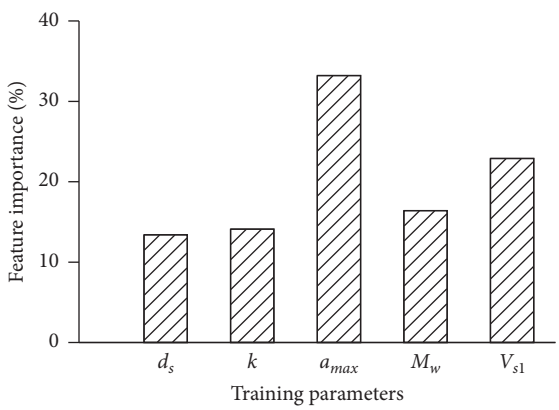

Figure 3: Feature importance of training parameters.

and remained 30% as the testing dataset. Two independent databases including the Andrus database and Kayen database are selected as the training dataset and testing dataset, respectively. The results indicate that the random forest model developed in this study is not only applicable in the range of training data, which limited the generalization [21, 22]. The application of the random forest is expanded in the engineering, which can provide references for similar projects.

The feature importance of the random forest model is used to output the importance of each training parameter, whose sum is equal to 100% [38]. Figure 3 shows the feature importance of each training parameter. It is worth noting that $a_{\max}$ and V_{s1} are the two most important parameters for predicting soil liquefaction, which is in line with the conclusion given by Nejad et al. [29] and Yang et al. [39]. The other three parameters also show 15% feature importance for liquefaction evaluation, which indirectly proves that it is reasonable to choose these five parameters as training parameters and the liquefaction evaluation results are also credible.

Even though the random forest method has been successfully applied in soil liquefaction evaluation, there still exist some limitations that need to be mentioned for further research. The accuracy of the random forest model may be further improved by tuning other hyperparameters. On the other hand, even though the contribution of each training parameter to soil liquefaction was discussed through the correlation analysis; however, similar with other machine learning methods, the feature importance of training parameters for the random forest is sensitive to the selection of training dataset. Thus, more datasets need to be collected in order to further confirm the relationship among training parameters.

4. Conclusions

The random forest method is utilized to evaluate the soil liquefaction. The selection of training parameters is discussed by adopting the Andrus database as the training dataset. Three hyperparameters in the random forest were optimized by correlation analysis and the liquefaction evaluation based on the random forest was established and examined by using the Kayen database as testing dataset. The success rates under different seismic intensities were obtained and compared with the Chinese code method and the Andrus method. The main conclusions can be drawn as follows.

(1) Based on the Andrus database, four parameters including the sandy-layer buried depth (d_s), peak horizontal ground surface acceleration ($a_{\max}$), seismic magnitude (M_w), stress-corrected shear wave velocity (V_{s1}), and an introduced parameter-stress ratio (k) were used as training parameters by correlation analysis for the random forest model.

(2) The optimal hyperparameters were determined by the minimum OOB error rate including the number of classification trees, maximum depth of classification trees, and maximum number of features corresponding to 13, 7, and 2, respectively.

(3) The success rates for liquefaction evaluation of three methods were compared based on the Andrus database, which indicated that the random forest model developed based on the training dataset was credible.

(4) Based on the Kayen database which was not used in the training dataset, the success rates of liquefaction evaluation were compared among three methods including the Chinese code method, the Andrus method, and the random forest method established

in this study. The Chinese code method and the Andrus methods gave a high success rate exceeding 90% for liquefied sites. The random forest method gave a success rate for liquefied sites and even a total success rate for all cases higher than 80%, which was completely acceptable. However, the success rates for nonliquefied sites given by the Chinese code method and the Andrus method were too low leading to the increase of engineering cost which might not be accepted.

Data Availability

The Andrus database used to establish the random forest model for liquefaction evaluation were taken from Andrus et al.'s study [1] (https://www.nist.gov/publications/draft-guidelines-evaluating-liquefaction-resistance-using-shear-wave-velocity) and the Kayen database collected from Kayen et al. [2] is available online in the ASCE Library (https://doi.org/10.1061/(ASCE)GT.1943-5606.0000743).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to express their thanks for the financial support to this study from the Project of the National Natural Science Foundation of China (grant no. 52178336), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (grant nos. 19KJB560015 and 18KJA560002), and Middle-Aged and Young Science Leaders of Qinglan Project of Universities in Jiangsu Province, China (grant nos. QL20200203 and QL20210210).

References

[1] R. D. The Andrus, K. H. Stokoe, R. M. Chung, and C. H. Juang, "Draft guidelines for evaluating liquefaction resistance using shear wave velocity measurements and simplified Procedures," *NIST Interagency/Internal Report (NISTIR)*, vol. 6277, 1999.

[2] R. Kayen, R. E. S. Moss, E. M. Thompson, and R. B. Seed, "Shear-wave velocity-based probabilistic and deterministic assessment of seismic soil liquefaction potential," *Journal of Geotechnical and Geoenvironmental Enigeering*, vol. 140, no. 4, pp. 407–419, 2014.

[3] G. X. Chen, Q. Wu, K. Zhao, Z. F. Shen, and J. Yang, "A binary packing material-based procedure for evaluating Soil liquefaction triggering during earthquakes," *Journal of Geotechnical and Geoenvironmental Enigeering*, vol. 146, no. 6, Article ID 04020040, 2020.

[4] P. K. Robertson, "Comparing CPT and vs liquefaction triggering methods," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 141, no. 9, Article ID 04015037, 2015.

[5] A. Johari, A. R. Khodaparast, and A. A. Javadi, "An analytical approach to probabilistic modeling of liquefaction based on shear wave velocity," *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, vol. 43, no. 1, pp. S263–S275, 2019.

[6] H. B. Seed, I. M. Idriss, and I. Arango, "Evaluation of liquefaction potential using field performance data," *Journal of Geotechnical Engineering*, vol. 109, no. 3, pp. 458–482, 1983.

[7] I. M. Idriss and R. W. Boulanger, "Semi-empirical procedures for evaluating liquefaction potential during earthquakes," *Soil Dynamics and Earthquake Engineering*, vol. 26, no. 2–4, pp. 115–130, 2006.

[8] M. Z. Rahman and S. Siddiqua, "Evaluation of liquefaction-resistance of soils using standard penetration test, cone penetration test, and shear-wave velocity data for Dhaka, Chittagong, and Sylhet cities in Bangladesh," *Environmental Earth Sciences*, vol. 76, no. 5, p. 207, 2017.

[9] A. Johari and A. R. Khodaparast, "Modelling of probability liquefaction based on standard penetration tests using the jointly distributed random variables method," *Engineering Geology*, vol. 158, no. 158, pp. 1–14, 2013.

[10] C. H. Juang, H. Yuan, D.-H. Lee, and P.-S. Lin, "Simplified cone penetration test-based method for evaluating liquefaction resistance of soils," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 129, no. 1, pp. 66–80, 2003.

[11] T. L. Youd and I. M. Idriss, "Liquefaction resistance of soils: summary report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of liquefaction resistance of soils," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 127, no. 4, pp. 297–313, 2001.

[12] R. W. Boulanger and I. M. Idriss, "CPT-based liquefaction triggering procedure," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 142, no. 2, Article ID 04015065, 2016.

[13] A. Johari and A. R. Khodaparast, "Analytical reliability assessment of liquefaction potential based on cone penetration test results," *Scientia Iranica*, vol. 21, no. 5, pp. 1549–1565, 2014.

[14] M. Shen, Q. Chen, J. Zhang, W. Gong, and C. Hsein Juang, "Predicting liquefaction probability based on shear wave velocity: an update," *Bulletin of Engineering Geology and the Environment*, vol. 75, no. 3, pp. 1199–1214, 2016.

[15] X. Xue and X. Yang, "Application of the adaptive neuro-fuzzy inference system for prediction of soil liquefaction," *Natural Hazards*, vol. 67, no. 2, pp. 901–917, 2013.

[16] H. B. Seed and I. M. Idriss, "Simplified procedure for evaluating soil liquefaction potential," *Journal of the Soil Mechanics and Foundations Division*, vol. 97, no. 9, pp. 1249–1273, 1971.

[17] T. D. Stark and S. M. Olson, "Liquefaction resistance using CPT and field case histories," *Journal of Geotechnical Engineering*, vol. 121, no. 12, pp. 856–869, 1995.

[18] A. M. Hanna, D. Ural, and G. Saygili, "Evaluation of liquefaction potential of soil deposits using artificial neural networks," *Engineering Computations*, vol. 24, no. 1, pp. 5–16, 2007.

[19] H. B. Zhao, Z. L. Ru, and S. Yin, "Updated support vector machine for seismic liquefaction evaluation based on the penetration tests," *Marine Georesources & Geotechnology*, vol. 25, no. 3-4, pp. 209–220, 2007.

[20] S. K. Mishra, S. Pattnaik, and D. Patnaik, "Estimating plans along with cost in multiple query processing environments by applying particle swarm optimization technique," *American Journal of Information Systems*, vol. 2, no. 3, pp. 52–55, 2014.

[21] J. Zhang and Y. Wang, "An ensemble method to improve prediction of earthquake-induced soil liquefaction: a multi-

dataset study," *Neural Computing & Applications*, vol. 33, no. 5, pp. 1533–1546, 2021.

[22] V. R. Kohestani, M. Hassanlourad, and A. Ardakani, "Evaluation of liquefaction potential based on CPT data using random forest," *Natural Hazards*, vol. 79, no. 2, pp. 1079–1089, 2015.

[23] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 10, pp. 5–32, 2001.

[24] V. R. Kohestani, M. R. Bazargan-lari, and J. Asgari-Marnani, "Prediction of maximum surface settlement caused by earth pressure balance shield tunneling using random forest," *Journal of AI and Data Mining*, vol. 5, no. 1, pp. 127–135, 2017.

[25] V. H. Dang, N. D. Hoang, L. M. D. Nguyen, D. T. Bui, and P. Samui, "A novel GIS-based random forest machine algorithm for the spatial prediction of shallow landslide susceptibility," *Forests*, vol. 11, no. 118, pp. 1–20, 2020.

[26] J. Zhou, X. Z. Shi, K. Du, X. Y. Qiu, X. B. Li, and H. S. Mitri, "Feasibility of random-forest approach for prediction of ground settlements induced by the construction of a shield-driven tunnel," *International Journal of Geomechanics*, vol. 17, no. 6, Article ID 04016129, 2017.

[27] B. T. Pham, C. C. Qi, L. S. Ho, and T. N. Thoi, "A novel hybrid soft computing model using random forest and particle swarm optimization for estimation of undrained shear strength of soil," *Sustainability*, vol. 12, 2218 pages, 2020.

[28] V. R. Kohestani, M. Vosoghi, M. Hassanlourad, and M. Fallahnia, "Bearing capacity of shallow foundations on cohesionless soils: a random forest based approach," *Civil Engineering Infrastructures Journal*, vol. 50, no. 1, pp. 35–49, 2017.

[29] A. S. Nejad, E. Guler, and M. Ozturan, "Evaluation of liquefaction potential using random forest method and shear wave velocity results," in *Proceedings of the 2018 International Conference on Applied Mathematics and Computational Science (ICAMCS. NET)*, IEEE, Budapest, Hungary, pp. 23–233, Oct 2018.

[30] GB 50021–2009, *Code for Investigation of Geotechnical Engineering*, China Architecture and Building Press, Beijing, 2009.

[31] J. Zhang, G. Ma, Y. Huang, J. Aslani, and B. Nener, "Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression," *Construction and Building Materials*, vol. 210, pp. 713–719, 2019.

[32] J. Zhang, D. Li, and Y. H. Wang, "Predicting uniaxial compressive strength of oil palm shell concrete using a hybrid artificial intelligence model," *Journal of Building Engineering*, vol. 30, no. 7, Article ID 101282, 2020.

[33] R. Sun and X. M. Yuan, "Depth-consistent Vs-based approach for soil liquefaction evaluation," *Chinese Journal of Geotechnical Engineering*, vol. 41, no. 3, pp. 439–447, 2018.

[34] GB 18306–2015, *Seismic Ground Motion Parameters Zonation Map of China*, China Standard Press, Beijing, China, 2015.

[35] X. F. Yao, L. Liu, Z. X. Wang, Z. F. Shen, and H. M. Gao, "A V_S-Based logistic regression method for liquefaction evaluation," *Advances in Civil Engineering*, vol. 2021, Article ID 5535387, 11 pages, 2021.

[36] A. Verikas, A. Gelzinis, and M. Bacauskiene, "Mining data with random forests: a survey and results of new tests," *Pattern Recognition*, vol. 44, no. 2, pp. 330–349, 2011.

[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, and B. Thirion, "Scikit-learn: machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[38] W. G. Zhang, C. Z. Wu, Y. Q. Li, L. Wang, and P. Samui, "Assessment of pile drivability using random forest regression and multivariate adaptive regression splines," *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, vol. 10, pp. 1–14, 2019.

[39] Y. Yang, R. Sun, and X. M. Yuan, "Liquefaction probability formula of shear wave velocity based on conventional parameters of soil layer," *Rock and Soil Mechanics*, vol. 40, no. 304, pp. 292–301, 2019.