

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Evaluation of Liquefaction-Induced Settlement Using Random Forest and REP Tree Models: Taking Pohang Earthquake as a Case of Illustration

*Mahmood Ahmad, Xiaowei Tang and Feezan Ahmad*

## Abstract

A liquefaction-induced settlement assessment is considered one of the major challenges in geotechnical earthquake engineering. This paper presents random forest (RF) and reduced error pruning tree (REP Tree) models for predicting settlement caused by liquefaction. Standard penetration test (SPT) data were obtained for five separate borehole sites near the Pohang Earthquake epicenter. The data used in this study comprise of four features, namely depth, unit weight, corrected SPT blow count and cyclic stress ratio. The available data is divided into two parts: training set (80%) and test set (20%). The output of the RF and REP Tree models is evaluated using statistical parameters including coefficient of correlation ( $r$ ), mean absolute error (MAE), and root mean squared error (RMSE). The applications for the aforementioned approach for predicting the liquefaction-induced settlement are compared and discussed. The analysis of statistical metrics for the evaluating liquefaction-induced settlement dataset demonstrates that the RF achieved comparatively better and reliable results.

**Keywords:** liquefaction, random forest, REP tree, settlement

## 1. Introduction

The evaluation of liquefaction-induced settlements has become an extremely significant issue about the foundations of different buildings, nuclear power plants, and earth dams on sandy soil deposits. Saturated sand deposits when are endured during an earthquake, pore water pressures are known to develop contributing to liquefaction or loss of shear strength. The pore water pressure then begin to dissipate primarily towards the ground surface, followed by a change in the volume of soil deposits which is manifested on the ground surface as settlements. Settlements caused by liquefaction are conventionally predicted using analytical or numerical methods.

Tokimatsu and Seed [1] developed a technique for predicting ground post liquefaction settlements based on volumetric strain, SPT N-value and cyclic stress ratio (CSR) relationships in the case of completely liquefied saturated sands transformed from an experimental relationship between relative sand density, volumetric strain, and maximum shear strain. Ishihara and Yoshimine [2] used an alternative approach to estimate ground settlements based on the safety factor, by means of the maximum shear strain which is an essential factor affecting the post-liquefaction volumetric strain. The liquefaction-induced settlement during the earthquake can be identified if the safety factor and relative density are established. Furthermore, the simplified method was constructed only by a relation between relative density, the factor of safety against liquefaction (FS) and volumetric strain ( $\epsilon_v$ ) to quantify the settlement of a site where the safety factor of safety against liquefaction was obtained. By combining earthquake intensity and SPT N-value with empirical equations to cause measurement error and lead to significant prediction error [3].

Analytical method used to assess liquefaction-induced settlements is based on the effective stress analysis of dynamic response which accounts for the generation and dissipation of excess pore water pressures. When used to evaluate post-liquidation settlements in saturated sand deposits, the volume compressibility coefficient of the sand is required which is very difficult to determine for the liquefied sand layer [4]. Shamoto et al., [4] suggested a simplified approach for estimating liquefaction-induced settlements of saturated sand deposits, based on the experimental evidence that there is an almost linear relationship between the function of the void ratio and the logarithm of the maximum shear strain induced during cyclic loading.

In numerical analysis, earthquake-induced liquefaction in the free-field may be interpreted as a 1D phenomenon occurring along a vertical soil column in which seismic-induced cyclic shear and compressive forces increase the pore pressure and hence cause a reduction in the transient soil strength and stiffness. Reconsolidation arises in the soil after liquefaction due to the dissipation of the excess pore pressure ( $\Delta u$ ) by means of water flow, resulting in the vertical settlement of the ground surface [5].

Park et al. [6] established a simple and sustainable method for predicting liquefaction-induced settlement using ANN. Tang et al. [3] found that the ANN and Bayesian Belief Networks (BBN) predictive outcomes are better than the Ishihara and Yoshimine simplified approach.

Pohang earthquake ( $M_w = 5.4$ ) that hit the Heunghae Basin around Pohang city had a liquefaction-induced damages—settlement and lateral displacement. In this study liquefaction-induced settlement is considered as a case of illustration. Several efforts have been made since the event to evaluate the post-earthquake damages [7–11]. Nevertheless, the liquefaction-induced settlement has received little attention. Settlement caused by liquefaction is commonly calculated by taking into account various factors and following several sophisticated analytical and numerical procedures. Nevertheless, in most cases it may not be possible to acquire such parameters in the field, as some of the required data may not be obtainable. The main purpose of this study is to evaluate liquefaction-induced settlement based on the database of field observations. To achieve this purpose, the random forest and REP tree techniques are used to develop two new models for evaluation of liquefaction-induced settlement. Although these techniques have been successfully applied in many domains, the application in geotechnical earthquake engineering is limited based on the literature surveys.

The remainder of this chapter is organized as follows: Section 2 briefly provides the description of data acquisition for liquefaction-induced settlement calculation.

Section 3 presents the methodology used to evaluate settlement caused due to earthquakes; an overview of the random forest and Rep tree techniques. Section 4 presents the development of the liquefaction-induced settlement models. Detailed results of the proposed models are discussed by performance evaluation measures are presented in Section 5, followed by conclusions in Section 6.

2. Data acquisition

In this study, Park et al. [6] collected database from the Integrated DB Centre of National Geotechnical Information, Korea [12] and the UBCSAND constitutive effective stress model [13] was used to develop predictive models. SPT data were obtained for five different borehole sites near the epicenter of the earthquake at Pohang. The input parameters for the RF and REP Tree models are depth (m), unit weight (kN/m<sup>3</sup>), corrected SPT blow count ( $N_{1(60)}$ ) and cyclic stress ratio (CSR) and the output is the observed settlement (mm). For details about the database, readers can refer to Park et al. [6]. The summary of the data base comprised 100 data points (20 data for each borehole) along with the corresponding settlement values is shown in **Table 1**.

Borehole	Depth (m)	Unit Weight (kN/m <sup>3</sup> )	$N_{1(60)}$	CSR	Settlement (mm)
BH-A-1	1	20	11	0.33	0.5
	2	20	11	0.31	0.5
	3	20	14	0.29	0.8
	4	20	16	0.28	1.4
	5	20	5	0.27	3.3
	6	20	10	0.26	3.4
	7	20	5	0.27	3.4
	8	20	6	0.29	2.5
	9	20	9	0.3	1.6
	10	20	9	0.31	1
	11	18	25	0.31	0.4
	12	18	25	0.31	0.3
	13	18	25	0.32	0.2
	14	18	25	0.32	0.3
	15	18	25	0.32	0.3
	16	18	25	0.32	0.3
	17	18	25	0.32	0.3
	18	18	25	0.32	0.2
	19	18	25	0.31	0.3
	20	18	25	0.31	0.3
BH-A-2	1	20	15	0.35	0.4
	2	20	17	0.32	0.8
	3	20	17	0.3	1.6

Borehole	Depth (m)	Unit Weight (kN/m <sup>3</sup> )	N <sub>1(60)</sub>	CSR	Settlement (mm)
	4	20	7	0.29	3.1
	5	20	6	0.27	2.8
	6	21	13	0.31	1.4
	7	21	18	0.34	0.8
	8	21	13	0.36	0.9
	9	21	11	0.37	1.1
	10	21	13	0.36	0.8
	11	16	2	0.36	0
	12	16	1	0.37	0
	13	16	1	0.38	0
	14	16	1	0.39	0
	15	16	1	0.39	0
	16	16	1	0.39	0
	17	16	1	0.39	0
	18	16	1	0.39	0
	19	16	1	0.38	0
	20	16	1	0.37	0
BH-A-3	1	18	6	0.24	0.6
	2	18	8	0.28	1.4
	3	18	10	0.3	2
	4	18	10	0.29	2.3
	5	18	11	0.28	2
	6	18	10	0.3	1.8
	7	18	11	0.32	1.4
	8	18	11	0.33	1.3
	9	18	12	0.34	1.2
	10	18	13	0.34	1
	11	21	25	0.34	0.7
	12	21	25	0.33	0.6
	13	21	25	0.33	0.6
	14	21	25	0.33	0.5
	15	21	25	0.32	0.5
	16	21	25	0.32	0.4
	17	21	25	0.31	0.5
	18	21	25	0.31	0.4
	19	21	25	0.3	0.4
	20	21	25	0.3	0.5
BH-A-4	1	20	5	0.23	1.1
	2	20	7	0.27	1.9
	3	20	18	0.27	1.6

Borehole	Depth (m)	Unit Weight (kN/m <sup>3</sup> )	N <sub>1(60)</sub>	CSR	Settlement (mm)
	4	20	9	0.27	2.8
	5	20	6	0.26	2.8
	6	20	11	0.31	1.6
	7	20	9	0.34	1.4
	8	21	25	0.36	0.6
	9	21	25	0.38	0.6
	10	21	25	0.38	0.6
	11	21	25	0.38	0.6
	12	21	25	0.37	0.5
	13	16	7	0.22	0
	14	16	1	0.21	0
	15	16	0	0.21	0
	16	16	2	0.22	0
	17	16	3	0.22	0
	18	16	3	0.22	0
	19	16	3	0.22	0
	20	16	3	0.22	0
BH-A-5	1	20	11	0.32	0.5
	2	20	10	0.31	1.6
	3	20	9	0.29	2.6
	4	20	11	0.3	1.9
	5	20	11	0.32	1.5
	6	20	10	0.33	1.4
	7	20	15	0.34	0.9
	8	20	15	0.35	0.8
	9	21	25	0.34	0.6
	10	21	25	0.34	0.6
	11	21	25	0.34	0.6
	12	21	25	0.33	0.6
	13	21	25	0.33	0.7
	14	21	25	0.33	0.7
	15	18	15	0.33	1.1
	16	18	11	0.33	1.4
	17	18	12	0.32	1
	18	16	14	0.32	0.1
	19	16	10	0.32	0
	20	16	7	0.32	0
Note: Borehole (BH-A-5) data comprised of 20 data points is used as testing dataset in this study.					

**Table 1.**  
Summary of liquefaction-induced settlement database.

### 3. Methodology

#### 3.1 Random forest

Random Forest (RF) is an ensemble machine learning technique driven by the development of a large number of decision trees that is produced by Leo Breiman [14]. Unlike DT, which uses all the features to construct a tree-like classification graph, RF uses an “efficient bagging” learning algorithm which integrates random selection of features with bagging. If one or a few features are very good predictors for target performance, it will pick this subset of features to construct a tree-like graph. This type of sample is known as the Bootstrap Sample. Using bagging techniques, these models are fitted with the above bootstrap samples, and then combined by voting. RF improves reliability and precision, reduces uncertainty and helps avoid overfitting.

Bootstrap aggregation or bagging is used to determine an appropriate number of trees with the size and nature of the training set. The RF prediction can be expressed as: by averaging the predictions from the individual regression trees;

An optimal number of trees are calculated by bootstrap aggregation or bagging with the size and nature of the training set. By averaging the predictions from the individual regression trees; The RF prediction can be expressed as:

$$\hat{g}(x) = \frac{1}{N} \sum_{n=1}^N g_n(x) \quad (1)$$

where  $\hat{g}(x)$  represents the RF prediction from the total of  $N$  trees, and  $g_n(x)$  denotes the prediction of each individual tree with the input  $x$ . In addition, an approximation of the uncertainty of the prediction can be made as the standard deviation of the predictions from all the trees, which can be expressed as:

$$\sigma = \sqrt{\frac{\sum_{n=1}^N (g_n(x) - \hat{g}(x))^2}{N - 1}} \quad (2)$$

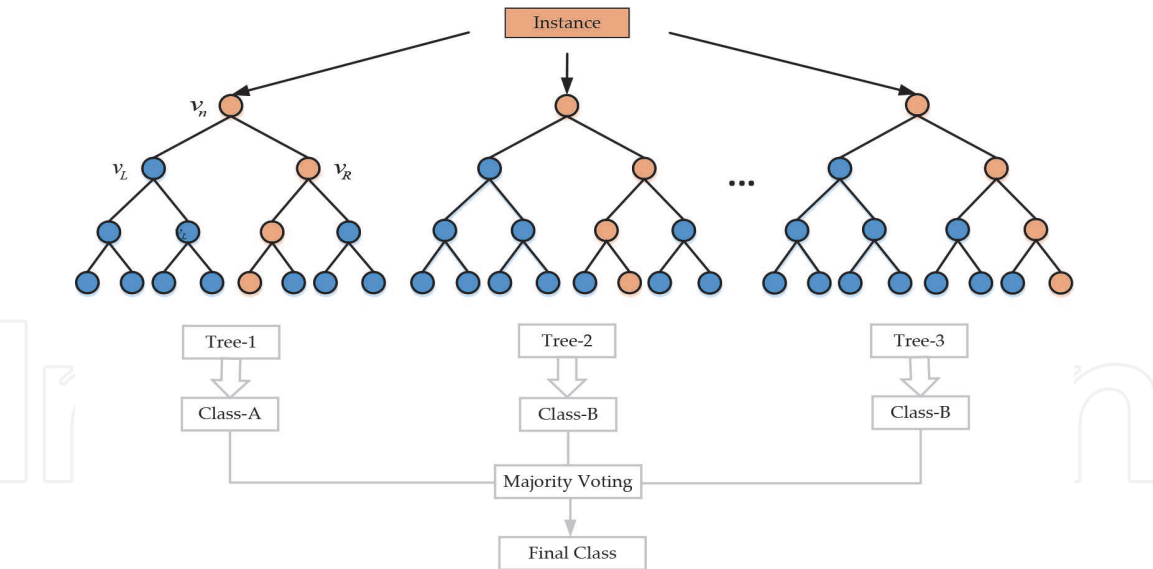
**Figure 1** demonstrates the method of classifying RF with the  $N$  trees. Starting from the root node ( $\nu_n$ ), after comparison with certain parameters or threshold values, samples are moved to the right node ( $\nu_R$ ) or the left node ( $\nu_L$ ). Repeat this partition until a terminal node is reached and get a classification tag (in this case, classes A or B). For classification task, the ensemble prediction is achieved by majority voting rule as a combination of the results of the individual trees [15].

#### 3.2 REP tree

The reduced error pruning tree (REP Tree) is an ensemble model of decision tree (DT) and The REP Tree (Reduced Error Pruning Tree) is an ensemble model of decision tree (DT) and reduced error pruning (REP) algorithms, equally good for classification and regression problems [16]. The REP Tree algorithm generates a decision regression tree by dividing and pruning the regression tree based on the importance of the highest knowledge benefit ratio ( $IGR$ ) [17]; The  $IGR$  values were determined via Eq. (3) based on the entropy ( $E$ ) function.

$$IGR(x, S) = \frac{E(S) - \sum_{i=1}^n \frac{E(S_i)|S_i|}{|S|}}{-\sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}} \quad (3)$$





**Figure 1.**  
Schematic representation of a RF classifier with N trees.

The IGR considers all the predictors of liquefaction-induced settlement with subset  $S_i$  from the training dataset ( $S$ ):  $i = 1, 2, \dots, n$  successive pruning steps. Since complex decision trees can result in a model being overfitted and less interpretable, REP helps to reduce complexity by removing the DT structure's leaves and branches [16, 18–20].

## 4. Liquefaction-induced settlement model development

### 4.1 Preparing training and testing datasets

The manner in which data are divided into training and test data sets in data mining procedures has a substantial effect on the results [21–23]. The statistical parameters for the input variables include the minimum, maximum, mean and standard deviation of the training and test datasets, as shown in **Table 2**. Data set splitting was done to assess the generalization efficiency and predictive ability of the developed models. The related performance of the training and testing datasets suggests that the developed models can be applied to the trained ranges. In the testing the ranges of input and output parameters often occur in the training datasets as shown in **Table 2**. The training and testing datasets' statistical consistency enhances the performance of the developed models and thus helps to properly assess them.

To ensure comparability, the RF and REP Tree models are proposed using the same training and test datasets. Using these models, liquefaction-induced settlements are predicted, and an analysis of the detailed performance of these models will find the optimum model afterwards. If the performance of this model on the training and test datasets is adequate then it can be adopted for development.

### 4.2 Evaluation measures

In this study, three evaluation measures, mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient ( $r$ ) are used to evaluate and compare the performance of the models. The MAE, RMSE and  $r$  are three useful statistical measures which provide some useful insights into the prediction model, of which the MAE is an average of the sum of the differences between the values predicted by a model and the actual values, the RMSE is a standard deviation of the



Dataset	Statistical parameter	Depth (m)	Unit Weight (kN/m <sup>3</sup> )	N <sub>1(60)</sub>	CSR	Settlement (mm)
Training	Minimum	1	16	0	0.21	0
	Maximum	20	21	25	0.39	3.4
	Mean	10.50	18.85	13.14	0.31	0.89
	Standard deviation	5.80	1.89	9.14	0.05	0.92
Testing	Minimum	1	16	7	0.29	0
	Maximum	20	21	25	0.35	2.6
	Mean	10.5	19.4	15.55	0.3255	0.93
	Standard deviation	5.92	1.76	6.66	0.01	0.65

**Table 2.**  
*Statistical parameters of the training and testing datasets.*

differences, and the correlation coefficient ( $r$ ) is a statistical measure representing the percentage of the variance for a model a dependent variable that’s described by an independent variable, and their expressions are as follows [24]:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i| \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2} \tag{5}$$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \tag{6}$$

where  $y_i$  and  $x_i$  are the observed and predicted value of  $i^{th}$  sample of the data respectively,  $\bar{x}$  and  $\bar{y}$  are the mean values of the observed and predicted values respectively, and  $n$  is the total number of samples. MAE can be given as a more natural and unambiguous index compared with RMSE to quantify errors between the estimated and actual observed values [25, 26]. RMSE was used as a standard statistical metric to assess output of a model [27]. The larger correlation coefficient ( $r$ ) and lower mean absolute error (MAE) values, and the root mean squared error (RMSE) present a higher accuracy of predicted results.

5. Results and discussion

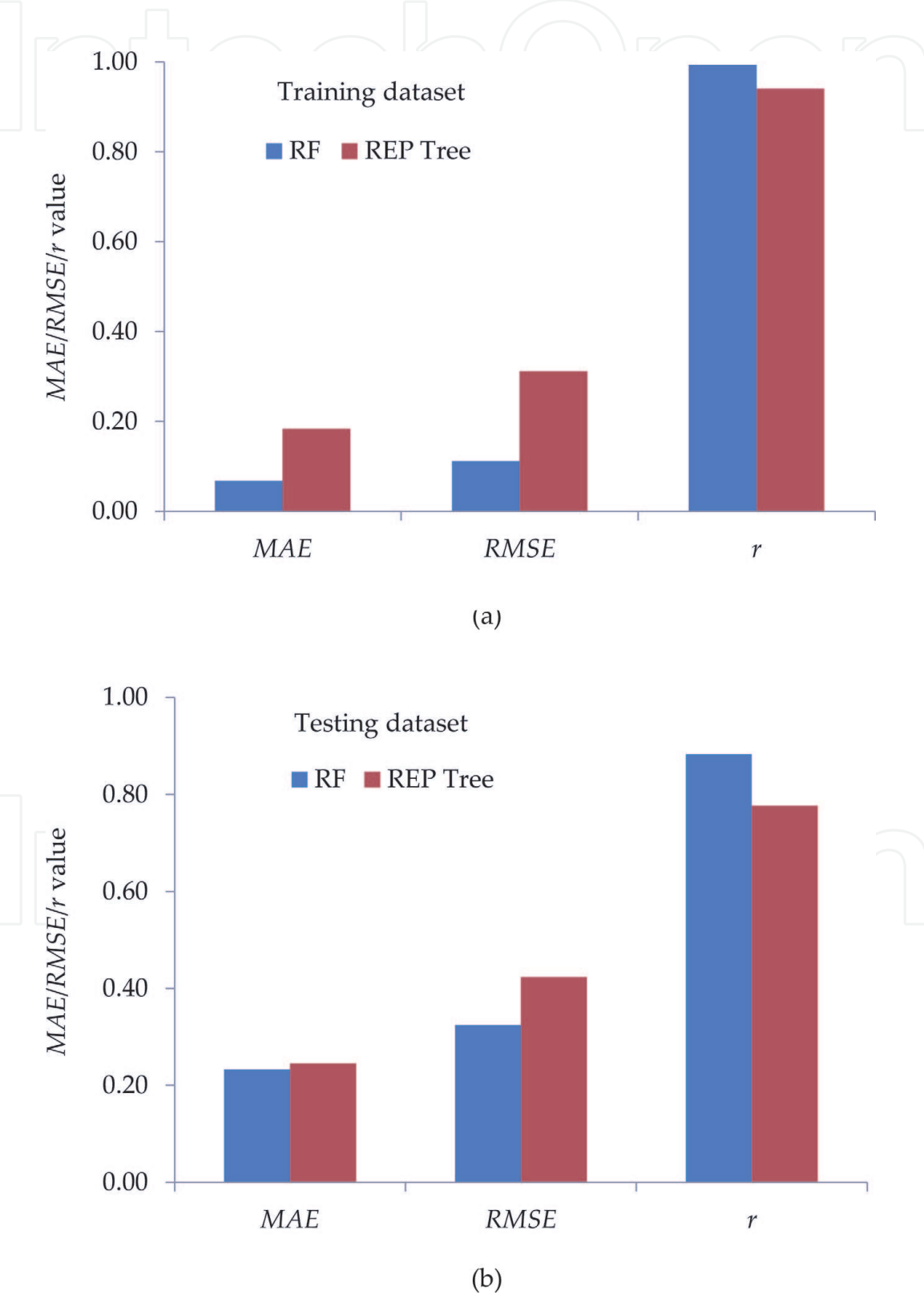
Theoretically, a specific model can be obtained when the model parameters are correctly selected and updated. The optimum values are obtained by trial and error

Algorithm	Parameters
RF	Minimum total weight of instances in a leaf: 1; minimum portion of the variance of all the data to be present in a node to be split in regression tress: 0.001; random number seed used to pick attributes: 1; K value: 0
REP Tree	Maximum tree depth: -1; minimum total instance weight in the leaf: 2; minimum likelihood of variance: 0.001; fold number: 3; seed number: 1

**Table 3.**  
*Model optimum modeling parameters.*

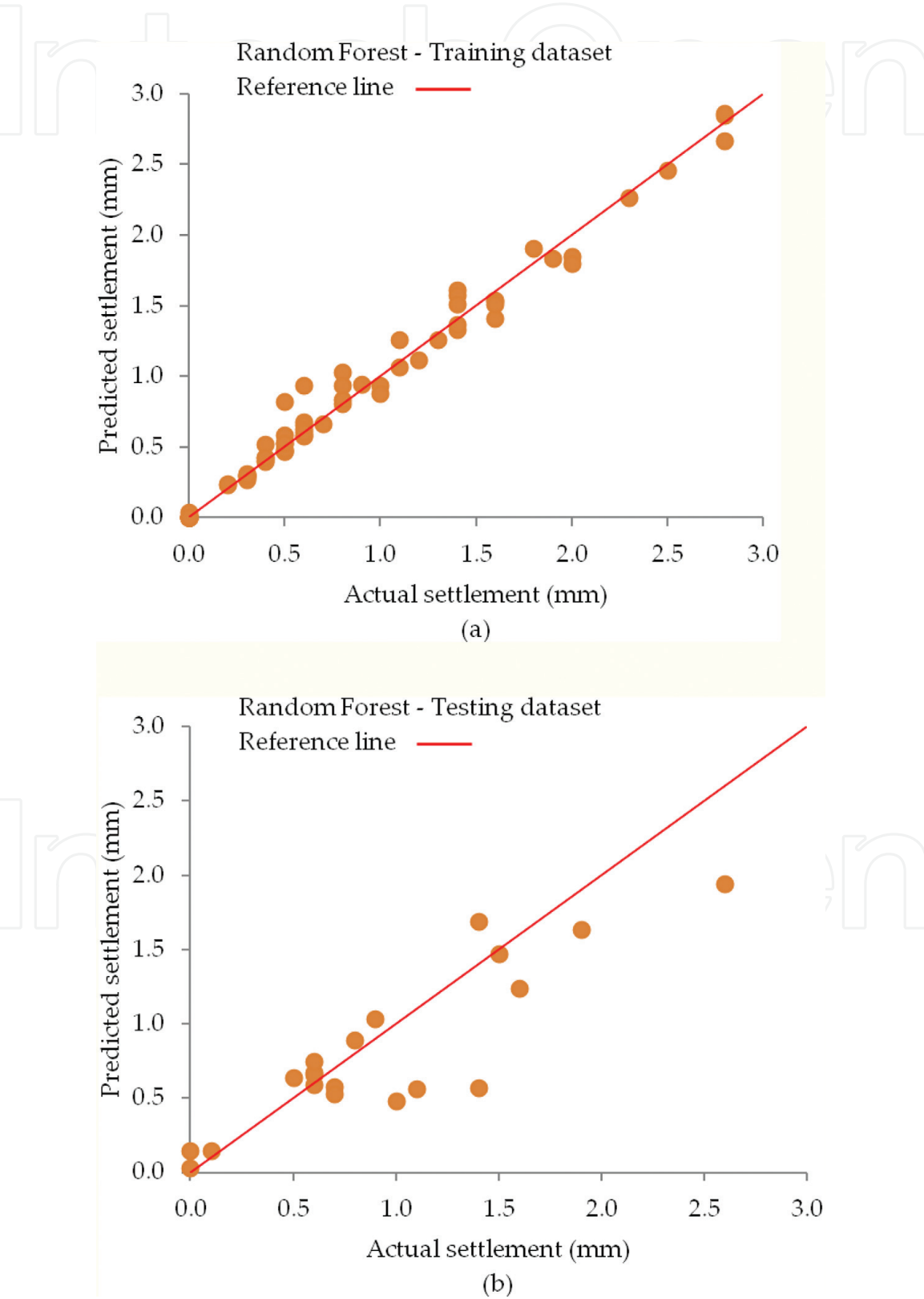
using parameter setting. The optimum value for each machine learning parameter is illustrated in **Table 3**. In the proposed RF and REP Tree models the most significant parameters are the number of seeds and the minimum total weight of instances in a leaf during the modeling process.

The RF and REP Tree predictive results were obtained from the datasets for training and testing datasets. The MAE, RMSE and correlation coefficient ( $r$ ) were subsequently determined on the basis of the Eqs. (4)–(6) shown in **Figure 2** that depicts RF and REP Tree models performance, respectively. For the RF model the



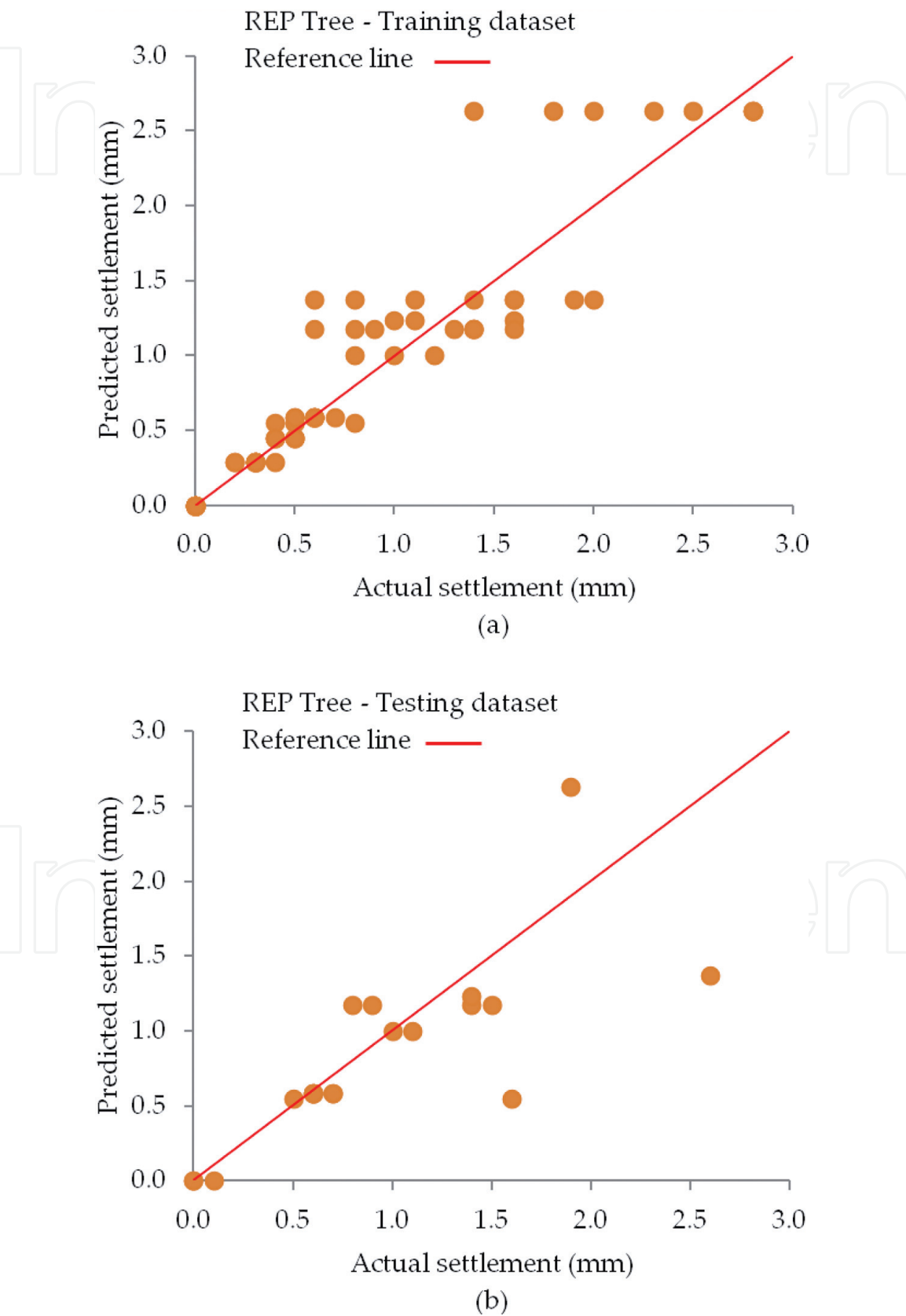
**Figure 2.**  
Comparison of MAE, RMSE, and  $r$  values from the RF and REP tree models.

training data prediction is higher than the test dataset prediction. The  $r$  values for the training data and testing data are found 0.9935 and 0.8833, respectively. For the REP Tree model, the training data  $r$  value ( $= 0.9405$ ) indicates marginally better results than that for the testing data ( $= 0.777$ ). It is obvious to judge that the performance of RF model in training and testing datasets is higher than that of REP Tree model. **Figure 2** presents bar graphs comparing the mean absolute error (MAE), the root mean squared error (RMSE), and the correlation coefficient ( $r$ )



**Figure 3.**  
*Training and testing of the RF model.*

for both models' training and test datasets. The MAE calculates the variance in the error term by term and reduces the significance of large errors; the *RMSE* value is more concentrated on large errors than on small ones. The RF model has lower *MAE* and *RMSE* values while higher *r* value, showing that in both training and testing datasets, the RF model provides adequate prediction of liquefaction-induced



**Figure 4.**  
*Training and testing of the REP tree model.*

settlement. Additionally, the results of training and testing were shown in **Figures 3 and 4**, showing the projected settlements are plotted with the actual data. One can see that settlements were predicted more accurately by the RF model than by the REP Tree model. While the REP Tree model few settlements cases are relatively under predicted as compared to the RF model.

## **6. Conclusions**

This paper explores the potential of RF and REP Tree models for predicting liquefaction-induced settlement using field data. The models were trained and tested based on the Pohang city liquefaction-induced settlement database. Both models assess liquefaction-induced settlement with substantial contributing factors such as depth, unit weight, corrected SPT blow count and cyclic stress ratio. The performance of the models presented is measured using statistical parameters such as the correlation coefficient ( $r$ ), MAE, and RMSE. The RF model indicates a better performance with respect to the training and testing datasets. From this analysis it can be inferred that the RF model works well in predicting liquefaction-induced settlement as opposed to the REP Tree model. Since, artificial intelligence-based approaches are data-dependent and their output can vary depending on the dataset, the quality and number of training datasets and the size of the experiments. Finally, it is obvious that the proposed models are open to develop and accumulation of more data will provide much better evaluation of liquefaction-induced settlements.

## **Acknowledgements**

The work presented in this paper was part of the research sponsored by the Key Program of National Natural Science Foundation of China under Grant No. 51639002 and National Key Research and Development Plan of China under Grant No. 2018YFC1505300-5.3.

## **Conflict of interest**

The authors declare no conflict of interest.

IntechOpen

## Author details

Mahmood Ahmad<sup>1,2</sup>, Xiaowei Tang<sup>1\*</sup> and Feezan Ahmad<sup>1</sup>

1 State Key Laboratory of Coastal and Offshore Engineering, Dalian University of Technology, Dalian, China

2 Department of Civil Engineering, University of Engineering and Technology Peshawar (Bannu Campus), Bannu, Pakistan

\*Address all correspondence to: [tangxw@dlut.edu.cn](mailto:tangxw@dlut.edu.cn)

## IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 



## References

- [1] Tokimatsu, K.; Seed, H.B. Evaluation of settlements in sands due to earthquake shaking. *Journal of geotechnical engineering* **1987**, *113*, 861–878.
- [2] Ishihara, K.; Yoshimine, M. Evaluation of settlements in sand deposits following liquefaction during earthquakes. *Soils and foundations* **1992**, *32*, 173–188.
- [3] Tang, X.-W.; Bai, X.; Hu, J.-L.; Qiu, J.-N. Assessment of liquefaction-induced hazards using Bayesian networks based on standard penetration test data. *Natural Hazards and Earth System Sciences* **2018**, *18*, 1451–1468.
- [4] Shamoto, Y.; Sato, M.; ZHANG, J. Simplified estimation of earthquake-induced settlements in saturated sand deposits. *Soils and Foundations* **1996**, *36*, 39–50.
- [5] Da Fonseca, A.V.; Millen, M.; Gómez-Martínez, F.; Romão, X.; Quintero, J. DELIVERABLE D3. 1 State of the art review of numerical modelling strategies to simulate liquefaction-induced structural damage and of uncertain/random factors on the behaviour of liquefiable soils. **2017**.
- [6] Park, S.-S.; Ogunjinmi, P.D.; Woo, S.-W.; Lee, D.-E. A Simple and Sustainable Prediction Method of Liquefaction-Induced Settlement at Pohang Using an Artificial Neural Network. *Sustainability* **2020**, *12*, 4001.
- [7] Kang, S.; Kim, B.; Bae, S.; Lee, H.; Kim, M. Earthquake-induced ground deformations in the low-seismicity region: A case of the 2017 M5. 4 Pohang, South Korea, earthquake. *Earthquake Spectra* **2019**, *35*, 1235–1260.
- [8] Choi, J.H.; Ko, K.; Gihm, Y.S.; Cho, C.S.; Lee, H.; Song, S.G.; Bang, E.S.; Lee, H.J.; Bae, H.K.; Kim, S.W. Surface Deformations and Rupture Processes Associated with the 2017 Mw 5.4 Pohang, Korea, Earthquake Surface Deformations and Rupture Processes Associated with the 2017 Mw 5.4 Pohang, Korea, Earthquake. *Bulletin of the Seismological Society of America* **2019**, *109*, 756–769.
- [9] Naik, S.P.; Kim, Y.-S.; Kim, T.; Su-Ho, J. Geological and structural control on localized ground effects within the Heunghae Basin during the Pohang Earthquake (MW 5.4, 15th November 2017), South Korea. *Geosciences* **2019**, *9*, 173.
- [10] Gihm, Y.S.; Kim, S.W.; Ko, K.; Choi, J.-H.; Bae, H.; Hong, P.S.; Lee, Y.; Lee, H.; Jin, K.; Choi, S. Paleoseismological implications of liquefaction-induced structures caused by the 2017 Pohang earthquake. *Geosci. J* **2018**, *22*, 871–880.
- [11] Kim, H.-S.; Sun, C.-G.; Cho, H.-I. Geospatial assessment of the post-earthquake hazard of the 2017 Pohang earthquake considering seismic site effects. *ISPRS International Journal of Geo-Information* **2018**, *7*, 375.
- [12] Integrated DB Center of National Geotechnical Information, SPT Database. Available online: <http://www.geoinfo.or.kr> (accessed on August 20, 2020).
- [13] Park, S.-S. Liquefaction evaluation of reclaimed sites using an effective stress analysis and an equivalent linear analysis. *Journal of the Korean Society of Civil Engineers* **2008**, *28*, 83–94.
- [14] Breiman, L. *Random Forests-Random Features* (# 567); Technical report, Dept, of Statistics, Univ. of California, Berkeley: 1999.
- [15] Criminisi, A.; Shotton, J.; Konukoglu, E. Decision forests: A unified framework for classification,

regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends® in Computer Graphics and Vision* **2012**, 7, 81–227.

[16] Quinlan, J.R. Simplifying decision trees. *International journal of man-machine studies* **1987**, 27, 221–234.

[17] Khosravi, K.; Pham, B.T.; Chapi, K.; Shirzadi, A.; Shahabi, H.; Revhaug, I.; Prakash, I.; Bui, D.T. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Sci. Total Environ.* **2018**, 627, 744–755.

[18] Pham, B.T.; Prakash, I.; Singh, S.K.; Shirzadi, A.; Shahabi, H.; Bui, D.T. Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* **2019**, 175, 203–218.

[19] Mohamed, W.N.H.W.; Salleh, M.N. M.; Omar, A.H. A comparative study of reduced error pruning method in decision tree algorithms. In Proceedings of 2012 IEEE International conference on control system, computing and engineering; pp. 392–397.

[20] Galathiya, A.; Ganatra, A.; Bhensdadia, C. Improved decision tree induction algorithm with feature selection, cross validation, model complexity and reduced error pruning. *International Journal of Computer Science and Information Technologies* **2012**, 3, 3427–3431.

[21] Ahmad, M.; Tang, X.-W.; Qiu, J.-N.; Ahmad, F. Evaluating Seismic Soil Liquefaction Potential Using Bayesian Belief Network and C4. 5 Decision Tree Approaches. *Applied Sciences* **2019**, 9, 4226.

[22] Javadi, A.A.; Rezaia, M.; Nezhad, M.M. Evaluation of liquefaction induced lateral displacements using genetic

programming. *Computers and Geotechnics* **2006**, 33, 222–233.

[23] Rezaia, M.; Javadi, A.A. A new genetic programming model for predicting settlement of shallow foundations. *Canadian Geotechnical Journal* **2007**, 44, 1462–1473.

[24] Barnston, A.G. Correspondence among the correlation, RMSE, and Heidke forecast verification measures; refinement of the Heidke score. *Weather and Forecasting* **1992**, 7, 699–709.

[25] Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geoscientific model development* **2014**, 7, 1247–1250.

[26] Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, 30, 79–82.

[27] Veerasamy, R.; Rajak, H.; Jain, A.; Sivadasan, S.; Varghese, C.P.; Agrawal, R.K. Validation of QSAR models-strategies and importance. *Int. J. Drug Des. Discov* **2011**, 3, 511–519.