



Article

Machine Learning Predictive Analysis of Liquefaction Resistance for Sandy Soils Enhanced by Chemical Injection

Yuxin Cong¹, Toshiyuki Motohashi², Koki Nakao¹ and Shinya Inazumi^{3,*}

¹ Graduate School of Engineering and Science, Shibaura Institute of Technology, Tokyo 135-8548, Japan; na23107@shibaura-it.ac.jp (Y.C.); na21105@shibaura-it.ac.jp (K.N.)

² Osaka Bousui Construction Co., Ltd., Osaka 543-0016, Japan; motohashi@obcc.co.jp

³ College of Engineering, Shibaura Institute of Technology, Tokyo 135-8548, Japan

* Correspondence: inazumi@shibaura-it.ac.jp; Tel.: +81-358598360

Abstract: The objective of this study was to investigate the liquefaction resistance of chemically improved sandy soils in a straightforward and accurate manner. Using only the existing experimental databases and artificial intelligence, the goal was to predict the experimental results as supporting information before performing the physical experiments. Emphasis was placed on the significance of data from 20 loading cycles of cyclic undrained triaxial tests to determine the liquefaction resistance and the contribution of each explanatory variable. Different combinations of explanatory variables were considered. Regarding the predictive model, it was observed that a case with the liquefaction resistance ratio as the dependent variable and other parameters as explanatory variables yielded favorable results. In terms of exploring combinations of explanatory variables, it was found advantageous to include all the variables, as doing so consistently resulted in a high coefficient of determination. The inclusion of the liquefaction resistance ratio in the training data was found to improve the predictive accuracy. In addition, the results obtained when using a linear model for the prediction suggested the potential to accurately predict the liquefaction resistance using historical data.

Keywords: artificial intelligence; chemical injection; cyclic undrained triaxial test; liquefaction; machine learning; sandy soil



Citation: Cong, Y.; Motohashi, T.; Nakao, K.; Inazumi, S. Machine Learning Predictive Analysis of Liquefaction Resistance for Sandy Soils Enhanced by Chemical Injection. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 402–419. <https://doi.org/10.3390/make6010020>

Academic Editor: Francesco Flammini

Received: 18 January 2024

Revised: 8 February 2024

Accepted: 12 February 2024

Published: 14 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The significant structural damage often caused by the settlement or tilting of structures, due to the liquefaction of saturated sandy soils during large earthquakes, has long been a major concern in the field of geotechnical engineering, as shown in Figure 1. This phenomenon, which can have serious consequences, was particularly documented in seminal studies [1–4]. The sudden instability of the ground during such events can lead to the catastrophic destruction of buildings and infrastructures, resulting in significant economic losses as well as the tragic loss of human life. This critical issue was further highlighted in [5,6]. These concerns have led to a significant increase in the study and development of activities aimed at improving liquefaction resistance and developing other mitigation methods. This focus was particularly highlighted by the groundbreaking work of [7,8], which contributed to a better understanding of these challenges.

In response to these critical challenges, the chemical injection method has emerged as a prominent and innovative solution for mitigating subsurface liquefaction risks [9–13]. This technique involves injecting chemical agents into sandy soils to increase their stability and cohesion. However, the effective implementation of this method and the accurate execution of designs depend heavily on the availability of precise and reliable data on the liquefaction resistance of the targeted chemically treated sandy soils [14–18]. Traditionally, liquefaction resistance has been assessed using cyclic undrained triaxial tests, which are fundamental to

building comprehensive databases reflecting a variety of test conditions and results [19–24]. Despite their critical importance in understanding soil behavior, these tests are often time-consuming, expensive, and labor-intensive. In addition, they are limited by the diverse nature of the properties of sandy soils found in different geographic regions, requiring a large number of experiments for a thorough and comprehensive data collection [25–27].

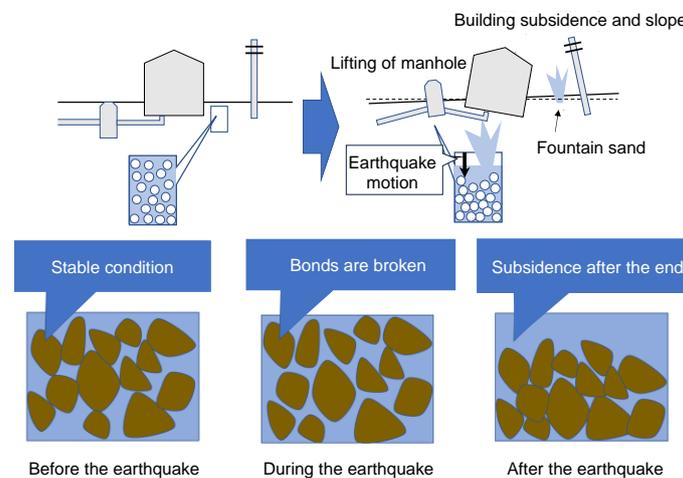


Figure 1. General principle of liquefaction phenomenon.

To address these challenges and limitations, the present study introduces a novel and state-of-the-art approach that employs machine learning and ensemble learning techniques [28–30]. The authors of this study propose a predictive model for evaluating the liquefaction resistance of sandy soils treated with solution-type chemical agents. This model is a synergistic combination of existing experimental data and advanced algorithms with artificial intelligence (AI) [31–34]. This innovative approach makes it possible to predict the liquefaction resistance of sandy soils prior to testing and to develop efficient strategies. The method is not only efficient, but also cost-effective, providing significant advances in the formulation of liquefaction mitigation strategies and enhancing risk assessment capabilities in geotechnical engineering [35,36]. The methodology of this study begins with the meticulous collection and analysis of data from cyclic undrained triaxial tests. The data form the basis of the machine learning database used in the study. Employing ensemble learning techniques, the authors successfully integrate the results of different prediction models to produce more accurate and reliable predictions. The primary goal of this study is to comprehensively assess the risk of sandy soil liquefaction and to provide reliable guidance for the design and implementation of chemical injection methods. It is expected that the development of this innovative non-experimental prediction method will contribute significantly to the sustainable development and advancement of geotechnical engineering practices. This approach not only minimizes the environmental impact, but also significantly reduces the time and costs associated with traditional soil testing methods. By introducing AI as a tool to assist in the execution of experiments traditionally used to predict liquefaction resistance in geotechnical engineering [37,38], this research aims to revolutionize the field and improve the safety and stability of sandy soils in earthquake-prone areas.

The application of AI in this context is particularly noteworthy, as it represents a paradigm shift in how geotechnical engineering challenges are addressed. By harnessing the power of machine learning, the study bypasses the limitations of traditional experimental methods. The AI-driven model is able to synthesize large amounts of experimental data, learn from different soil conditions, and adapt to different chemical treatments. This leads to a more holistic understanding of soil behavior under seismic activity, providing engineers with a powerful tool for predicting soil response in real-world scenarios. Of particular importance in this study is the use of ensemble learning techniques. Ensemble

learning involves combining multiple machine learning models to improve prediction accuracy [39], thereby reducing the likelihood of erroneous predictions that could lead to unsafe engineering practices. This approach ensures that the predictive model is not based on a single dataset or algorithm, but is a robust composite of multiple predictive insights, resulting in a more reliable and trustworthy predictive model. In addition, the study's approach to integrating AI with traditional geotechnical engineering practices is an exemplary model of interdisciplinary innovation. By bridging the gap between advanced computational techniques and practical engineering applications, the study sets a precedent for future studies in the field. It demonstrates the potential of AI to improve the accuracy and efficiency of engineering solutions, thereby contributing to the development of safer and more resilient infrastructures. The study not only addresses the immediate challenge of predicting and mitigating the liquefaction of sandy soils, but also opens new avenues for study and innovation in geotechnical engineering. By harnessing the power of AI and machine learning, it presents a forward-looking approach that could revolutionize the field, leading to more sustainable, efficient, and safer engineering practices [40]. This study not only contributes to the academic body of knowledge, but also has practical implications for the construction industry, urban planning, and disaster risk management, especially in earthquake-prone regions.

2. General Evaluation and Countermeasures to Liquefaction of Sandy Soils

2.1. Chemical Injection Method

The liquefaction of sandy soils during seismic events is a major challenge in geotechnical engineering [41,42]. It poses a risk to the stability and integrity of structures built on such soils. In response, chemical injection has emerged as a promising technique for mitigating liquefaction in sandy soils [9–13]. Liquefaction occurs when saturated sandy soils lose their strength and stiffness in response to an applied stress, such as an earthquake, resulting in fluid-like behavior [43]. Chemical injection, also known as soil grouting, involves injecting chemical solutions into soils to improve their physical and mechanical properties, thereby increasing their resistance to liquefaction.

The chemical injection process typically involves the use of materials such as silicates, polyurethanes, or acrylamides. When injected into the soil, these chemicals react with the soil particles or with each other to form a solidified matrix that binds the soil particles together, increasing their density and shear strength. One common approach is to use sodium silicate, a water-soluble silicate that reacts with calcium chloride to form a gel-like substance. This substance fills the voids between the soil grains, reducing porosity and increasing soil cohesion. Another approach is to use organic polymers that solidify when injected, creating a network of polymer chains that bind the soil particles together. Chemical injection has several advantages. It is a relatively quick process compared to other soil stabilization methods and can be applied to specific areas without the need for extensive excavation or the disruption of existing structures. In addition, the method can be tailored to different soil types and conditions [13].

Despite its advantages, the chemical injection method faces several challenges and limitations. The use of chemicals raises environmental concerns. Some chemicals used in the process can be harmful to the environment, especially if they leach into groundwater. Selecting environmentally friendly chemicals that do not compromise soil stability is critical. In addition, the long-term effectiveness of the treatment is uncertain. Over time, the injected chemicals may degrade or the bond between soil particles may weaken, reducing the effectiveness of the treatment. Achieving uniform distribution of the chemical solution throughout the soil is challenging. Inhomogeneous treatment can result in uneven soil properties that may not effectively mitigate liquefaction hazards. In addition, the process can be costly, especially for large-scale applications. The cost of chemicals and the need for specialized equipment and personnel can be significant. Finally, not all sandy soils are suitable for chemical injection. The method is less effective in soils with high organic content, or those that are too coarse or too fine. Effective monitoring and quality control

are essential to ensure successful treatment. This includes monitoring the distribution of chemicals, the reaction process, and the final soil properties.

Recent advances in chemical injection technology have focused on improving the environmental sustainability of the process. Researchers are exploring the use of biodegradable and non-toxic chemicals that minimize environmental impact while maintaining soil stability [44–46]. In addition, new techniques are being developed to inject chemicals more uniformly and efficiently using advanced monitoring systems and precision application equipment [47]. These systems allow for real-time adjustments during the injection process, improving the consistency and effectiveness of the treatment. Studies are also being conducted to determine the long-term performance of chemically stabilized soils under various environmental conditions. This research is critical to understanding the durability of chemical injections and developing maintenance strategies to ensure the continued effectiveness of the treatment over time [48]. The economic aspects of chemical injection are also being addressed, with cost-benefit analyses being conducted to compare the method with alternative soil stabilization techniques [49]. These analyses consider not only the direct costs of the chemicals and the application process, but also the potential savings from reduced damage during seismic events. Chemical injection for liquefaction mitigation in sandy soils offers a viable solution for improving soil stability in seismic areas [44]. However, it is imperative to address the environmental, technical, and economic challenges associated with this method. Future studies should focus on developing environmentally friendly chemicals, improving application techniques for uniform soil treatment, and evaluating the long-term performance of treated soils. With advances in technology and a better understanding of soil behavior, chemical injection has the potential to become a more effective and sustainable option for liquefaction mitigation in sandy soils.

2.2. Cyclic Undrained Triaxial Test

In the specialized field of geotechnical engineering, the cyclic undrained triaxial test stands out as a fundamental technique for evaluating the ability of chemically treated sandy soils to resist liquefaction, a phenomenon that can severely compromise the structural integrity of buildings and other infrastructure during an earthquake. This test meticulously replicates the complex stress conditions that soils experience during seismic activity, mimicking the rapid loading and unloading patterns typical of such events. As a result, it provides invaluable data on the dynamic behavior of soils under these extreme conditions. Specifically, the cyclic undrained triaxial test measures soil strength and deformation characteristics without allowing water to drain from the soil sample, which closely mimics the rapid loading conditions during an earthquake. This aspect is critical to understanding how chemically stabilized soils behave when subjected to seismic forces, providing insight into their structural stability and the effectiveness of chemical treatments in enhancing their resistance to liquefaction. The knowledge gained from this testing is critical in informing engineers and researchers about the limitations and capabilities of treated soils to withstand seismic forces, thus playing an important role in the design and implementation of safer, more resilient construction projects in earthquake-prone areas. The studies [19–24] provide comprehensive insight into the procedure, application, and significance of the cyclic undrained triaxial test in the broader context of improving soil stability and safety in the face of natural disasters.

Liquefaction is the phenomenon in which saturated sandy soils significantly lose their strength and stiffness in response to an applied load, such as seismic shaking, causing them to behave like a liquid. The cyclic undrained triaxial test is a laboratory test designed to evaluate the resistance of soils to liquefaction, which is particularly important for soils that have been treated with chemical agents for stabilization. Figure 2 shows the typical appearance of the cyclic undrained triaxial test. The test involves the cyclic loading of a cylindrical soil sample in a triaxial chamber. The soil sample is first saturated and then subjected to axial cyclic loading at a controlled frequency and amplitude. The test is undrained, meaning that no water can enter or leave the soil sample during the test. This

condition simulates the rapid loading that occurs during earthquakes. Parameters, such as axial stress, axial strain, pore water pressure, and volume change, are recorded. The number of cycles the soil can withstand before failure (defined by a certain level of strain or pore pressure) is used to evaluate its resistance to liquefaction.

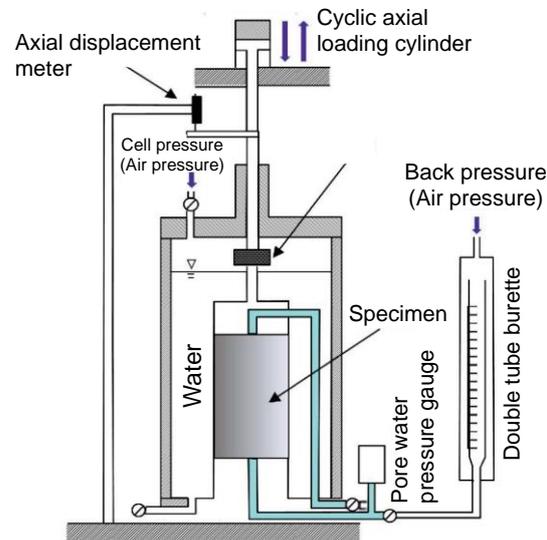


Figure 2. Typical appearance of cyclic undrained triaxial test.

The cyclic undrained triaxial test is widely recognized for its ability to replicate the stress conditions experienced by soils during earthquakes. It provides valuable data on the behavior of chemically treated soils, including the stiffness, strength, and pore pressure response, which are critical for evaluating the liquefaction potential [19–24].

Despite its advantages, the cyclic undrained triaxial test faces several challenges. Obtaining and preparing undisturbed soil samples for testing is challenging. Sample disturbance can significantly affect the test results, making it difficult to accurately represent the in situ soil conditions. Each test is conducted on a small-scale soil sample, which may not accurately represent the behavior of the soil mass in the field due to scale effects. For chemically treated soils, it is difficult to ensure the uniform distribution of the chemical agent throughout the sample. Inconsistent treatment can lead to variable results that do not accurately reflect the true behavior of the treated soil. The test is complex and requires sophisticated equipment and skilled personnel, making it expensive and time-consuming. The accurate measurement of the pore water pressure during the test is critical, but can be challenging, especially in sands with low permeability. It is difficult to ensure the repeatability and reliability of the test results due to the inherent variability of the soil properties and the sensitivity of the test to experimental conditions.

The cyclic undrained triaxial test is an essential tool for evaluating the liquefaction resistance of chemically treated sandy soils. However, overcoming the challenges associated with sample preparation, scale effects, chemical interactions, test complexity, and measurement accuracy is critical to obtaining reliable results. Future advances in test procedures, equipment, and analytical methods are needed to overcome these challenges. By improving the cyclic undrained triaxial test, it can continue to be a valuable method for evaluating the effectiveness of chemical treatments for mitigating the liquefaction hazards of sandy soils.

2.3. Liquefaction Resistance Ratio

The concept of the liquefaction resistance ratio, often derived from the cyclic undrained triaxial test, is a critical parameter in geotechnical engineering, particularly in assessing the stability of soils under seismic conditions. This ratio is a measure of a soil's ability to resist liquefaction, a phenomenon in which saturated soil loses much of its strength and

stiffness in response to an applied stress, such as an earthquake, causing it to behave like a fluid. In the context of the cyclic undrained triaxial test, the liquefaction resistance ratio is defined as the ratio of the cyclic stress required to cause liquefaction in a soil sample to the maximum cyclic stress experienced by the soil during an earthquake [14–18]. Liquefaction in this test is typically identified by a specific criterion, such as reaching a predetermined level of axial strain or a significant increase in pore water pressure, indicating a loss of soil strength. The cyclic undrained triaxial test involves subjecting a cylindrical soil sample, saturated and confined in a triaxial chamber, to controlled cyclic axial loading. The loading simulates the stress conditions that the soil would experience during seismic events. The cyclic stress required to induce liquefaction is determined by gradually increasing the stress amplitude of the loading cycles until the soil sample reaches the failure criterion. The liquefaction resistance ratio is an index that indicates the resistance of the sandy soil to liquefaction. Specifically, it refers to the cyclic stress amplitude ratio when the axial strain amplitude reaches 5% or the excess pore water pressure ratio reaches 95% and the number of cyclic loads is 20 [13]. This index is calculated from the liquefaction intensity curve [13,45] obtained as a result of the cyclic undrained triaxial test, as shown in Figure 3. The cyclic undrained triaxial test simulates liquefaction phenomena under compacted and undrained conditions in a testing machine. The collected undisturbed specimen is compacted under the original effective confining pressure, subjected to cyclic shear stress equivalent to the stress during an earthquake, and tested. During the test, experiments are performed at multiple cyclic stress levels and the number of cyclic loads at which both axial strain amplitudes reach 5% is determined. A liquefaction intensity curve is constructed from these data.

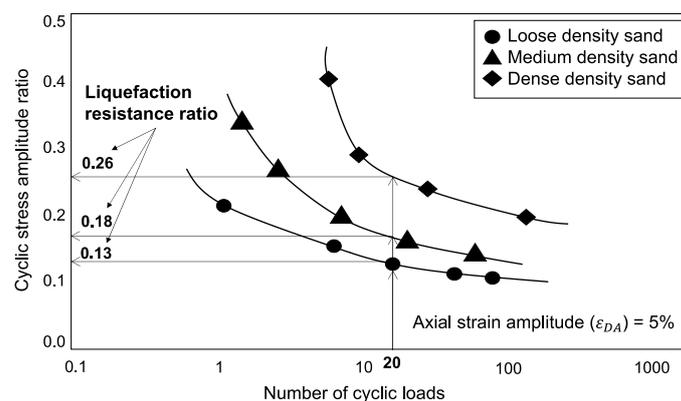


Figure 3. Liquefaction intensity curves obtained from cyclic undrained triaxial tests.

3. Machine Learning Predictive Analysis

3.1. Ensemble Learning

Ensemble learning methods use multiple machine learning algorithms to produce weakly predictive results based on features extracted through a variety of projections on the data and fuse the results with various voting mechanisms to achieve a better performance than that obtained by any constituent algorithm alone [46]. Neural networks have attracted attention in the field of machine learning due to their high expressiveness in modeling non-linear data. On the other hand, gradient boosting decision trees excel in terms of interpretability and accuracy. It is expected that the combination of these two methods will improve the predictive accuracy of the model.

A neural network is an interconnected collection of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron. The processing capability of the network is stored in the inter-unit connection strengths or weights, which are obtained through a process of adaptation to, or learning from, a set of training patterns [47–50]. A model with many hidden layers is called deep learning. Multiple inputs and outputs are possible, and neural networks enable prediction, judgment, and classification. As shown in

Figure 4, data are input into the input layer, the features are input with indicators of the data, and the final results are calculated by inputting neurons into the output layer.

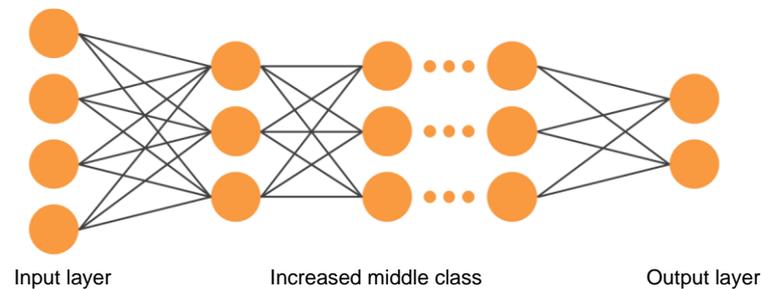


Figure 4. Schematic prediction flow of neural network.

The gradient boosting decision tree is an algorithm that learns multiple decision trees sequentially, using the residuals from the previous decision tree in the learning process of the next decision tree. This method also uses a gradient descent to minimize the errors in the predicted values [51–53].

The ensemble model proposed in this study combines two models: a neural network and a gradient boosting decision tree. As shown in Figure 5, it determines the weighted average of the predictions from these models to generate the final prediction. Both models are known for their high predictive performance in terms of tabular training data, and it is expected that the combination of these models, through ensemble modeling, will further improve in accuracy.

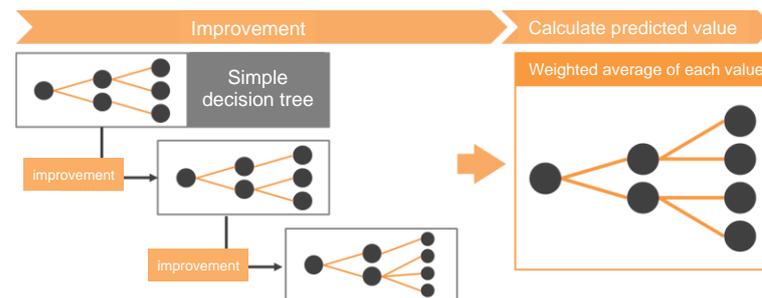


Figure 5. Schematic prediction flow of proposed ensemble model.

The decision to use both a gradient boosting decision tree as well as a neural network, especially within an ensemble learning framework, was driven by several key factors, including the following:

- (1) **Interpretability and transparency:** Decision trees provide a clear and interpretable structure, making it easy to understand how predictions are made. This is particularly important in this study because it involves complex geotechnical data, where providing clear insight into how the model reaches its conclusions is critical to gaining acceptance and trust from the engineering community.
- (2) **Handling of non-linear relationships:** The nature of the dataset in this study, which includes various soil parameters and their interactions, exhibits non-linear patterns. Gradient boosting decision trees are adept at handling such non-linear relationships, making them a suitable choice for predictive analysis in this study.
- (3) **Flexibility with different types of data:** The dataset in this study includes a mix of numeric and categorical variables (such as soil type, chemical composition, etc.). Decision trees can handle this variety without extensive preprocessing, simplifying the modeling process.
- (4) **Robustness against outliers and missing values:** Decision trees are less sensitive to outliers and can handle missing data efficiently, which is a significant advantage given the variability and occasional gaps in geotechnical data.

- (5) Effective with ensemble methods: Ensemble learning techniques, which combine multiple machine learning models to improve prediction accuracy, were used. Decision trees integrate well with such ensemble methods (e.g., gradient boosting decision tree) and often result in models that are more accurate and robust than those based on a single algorithm.

The gradient boosting decision tree approach is chosen for its transparency, robustness, and effectiveness in dealing with the specific characteristics of dataset in this study. This approach will allow this study to develop a predictive model that is not only accurate, but also interpretable and reliable for assessing the liquefaction resistance of chemically treated sandy soils.

Before constructing the ensemble model, a method is used to optimize the generalization ability of each model. Specifically, the training data are divided into several subsets, and a technique called random search is used for cross-validation to optimize the hyperparameters.

3.2. Preparation of Dataset

Datasets play a crucial role in the implementation of machine learning. They are divided into two main categories: training data and test data. Training data are used for model training, which is the necessary basis for acquiring the generalization ability.

3.2.1. Details of Training Data

In this study, data from cyclic undrained triaxial tests on chemically improved sandy soils, conducted to determine the liquefaction resistance, were used. These data include specimen conditions, test conditions, and test results. Specifically, the variable elements shown in Table 1 were extracted from previous test records to form the training data. A total of 272 specimens from 68 cases of cyclic undrained triaxial tests were used. One case corresponds to one site. In order to obtain the liquefaction resistance as shown in Figure 3, at least four specimens must be used for the cyclic undrained trial tests conducted for each case (each site). All 272 specimens were chemically improved sandy soils with 6%, 9%, and 12% silica concentrations, with four specimens being collected from each of the 68 sites.

Table 1. Variable elements related to cyclic undrained triaxial test employed in training data.

Category	Variable Elements
Condition parameters for specimens of chemically improved sandy soils	Dry density (g/cm ³)
	Fine particle content (%)
	Effective confining pressure (kN/m ²)
	Unconfined compressive strength (kN/m ²)
	Silica gel concentration of injected chemical solution (%)
Results obtained by cyclic undrained triaxial test	Increase in silica content (mg/g)
	Number of cycles to reach 5% strain in both amplitudes
	Number of cycles to reach 95% excess pore pressure ratio
	Cyclic stress amplitude ratio
	Liquefaction resistance ratio *

*: Refers to the cyclic amplitude stress ratio when the axial strain amplitude reaches 5% or the excess pore water pressure ratio reaches 95% and the number of cyclic loads is 20.

3.2.2. Details of Test Data

The test data in this study are based on the above training data. However, the test data exclude the target variables of the training data and mainly consist of explanatory variables from the training data.

3.3. Distinguishing Explanatory and Target Variables

In this study, the training data are used to train an ensemble model and, after the learning process, predictions are made by inputting test data. During this process, the predicted values are compared with the target variables of the training data to validate the predictive performance of the machine learning model. In this study, the distinction between the explanatory variables and the target variables is made for four cases, namely Case-1, Case-2, Case-3, and Case-4, as shown in Table 2. An example of the training data used in this study, i.e., data for 2 of the 272 specimens, are presented in Table 3. The target variable for Case-3 and Case-4 is the same. The difference is that Case-3 makes predictions without liquefaction resistance, while Case-4 makes them with liquefaction resistance.

Table 2. Explanatory and target variables employed for each prediction case.

Case	Explanatory Variables	Target Variables
Case-1	Variable elements shown in Table 1 excluding the liquefaction resistance ratio and the target variable	Number of cycles to reach 5% strain in both amplitudes
Case-2	Variable elements shown in Table 1 excluding the liquefaction resistance ratio and the target variable	Number of cycles to reach 95% excess pore pressure ratio
Case-3	Variable elements shown in Table 1 excluding the liquefaction resistance ratio and the target variable	Cyclic stress amplitude ratio
Case-4	Variable elements shown in Table 1 excluding the target variable	Cyclic stress amplitude ratio

Table 3. Example of employed training data (data of 2 out of 272 specimens were extracted).

Variable	Variable Elements	Data for 2 of 272 Specimens	
Explanatory variables	Dry density (g/cm ³)	1.684	1.484
	Effective confining pressure (kN/m ²)	90	165
	Fine particle content (%)	14.8	11.4
	Unconfined compressive strength (kN/m ²)	539	483
	Silica gel concentration of injected chemical solution (%)	12	12
	Increase in silica content (mg/g)	11.62	7.79
	Number of cycles to reach 5% strain in both amplitudes	18	6.5
	Number of cycles to reach 95% excess pore pressure ratio	37	38.4
Target variable	Repetitive stress amplitude ratio		

3.4. Evaluation of Prediction Accuracy

The coefficient of determination (R^2) quantifies the proportion of variance explained by a statistical model and is an important summary statistic of biological interest [54]. It is also widely used in machine learning. This metric quantitatively indicates how well the predicted values of the target variables, generated by a machine learning model using test data, match the actual values of the target variables in the training data. When the predictions of a machine learning model are perfectly accurate, R^2 is equal to 1, while it approaches 0 when the predictions are completely unrelated to the actual values.

The formula for the coefficient of determination (R^2) is defined as Equation (1):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

where n is the total amount of data, y_i is the i -th actual value, \hat{y}_i is the i -th predicted value, and \bar{y} is the average of all the actual values.

The coefficient of determination (R^2) provides a numerical measure of how well the predictions from a machine learning model match the actual values in the training data, ranging from 0 (no correlation) to 1 (perfect correlation).

The coefficient of determination (R^2) is traditionally used to assess the correlation between variables, particularly in linear regression models. However, this study chooses R^2 as the primary metric for several reasons specific to the context and objectives of this study:

- (1) Interpretability in the context of geotechnical engineering: In geotechnical engineering, particularly in studies involving practitioners and engineers, the interpretability of the model output is critical. R^2 , as a widely recognized and understood metric, provides a clear and direct measure of how well the model's predictions match the actual observed data.
- (2) Quantifying the explanation of variance: The primary goal of this study was to develop a model that could accurately predict liquefaction resistance based on various input characteristics. R^2 effectively quantifies the proportion of variance in the dependent variable that can be predicted from the independent variables, which directly aligns with the goal of this study.
- (3) Model evaluation in machine learning: In the context of machine learning, particularly ensemble methods, R^2 is still a relevant metric for evaluating predictive models. It provides a concise summary of model performance, especially when dealing with continuous outcome variables, as in this study.

4. Results and Discussion

4.1. Selecting Target Variables

Looking closely at Case-1, shown in Figure 6a, the y -axis represents the actual values of the target variable in the training data, while the x -axis represents the values predicted by the ensemble model. In an ideal scenario, the yellow points would align closely along the upward sloping red line, indicating a high degree of accuracy with minimal error between the actual and predicted values. However, in Case-1, the coefficient of determination (R^2) is significantly low at -0.0790 . This low accuracy is primarily due to the distribution characteristics of the target variable, "number of cycles to reach 5% strain in both amplitudes", within the training data. Although the range in training data for this goal variable is from 0 to 450, the occurrence of values above 200 is extremely rare. This disproportionate distribution is likely to skew the model's predictive ability, negatively affecting its accuracy.

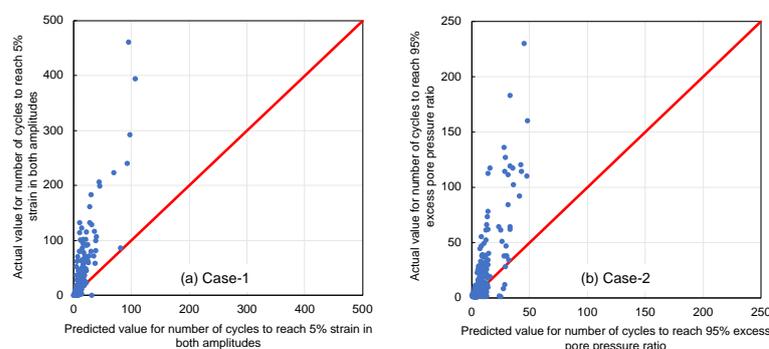


Figure 6. Degrees of deviation between predicted and measured values in (a) Case-1 and (b) Case-2.

In contrast, when examining Case-2, shown in Figure 6b, the alignment of many yellow points is seen to be closer to the red line, implying an improved prediction accuracy

compared to Case-1. However, the coefficient of determination (R^2) remains relatively low at 0.0296. This suggests that, similar to Case-1, the intrinsic characteristics of the target variable (number of cycles to reach 95% excess pore pressure ratio) influence the prediction results. It implies that, while the alignment of the data points may visually suggest accuracy, the underlying distribution and nature of the target variable play more significant roles in determining the actual predictive accuracy.

The target variable is changed to “repetitive stress amplitude ratio” in Case-3 and Case-4. This change results in a remarkable agreement between the experimental results and the predicted values. The results of Case-3 and Case-4 are provided in Figure 7, showing a clear and consistent alignment of the yellow points along the red line, visually confirming the high accuracy of the model.

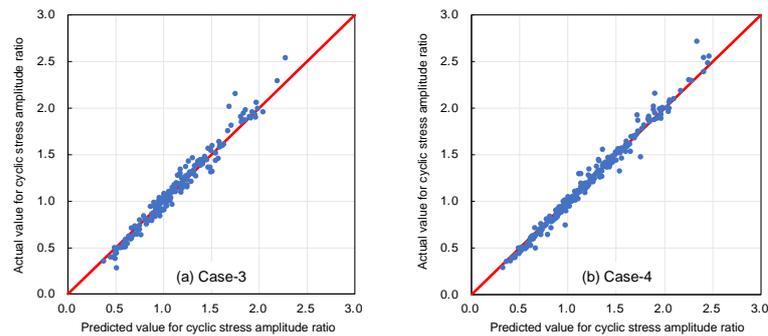


Figure 7. Degrees of deviation between predicted and measured values in (a) Case-3 and (b) Case-4.

These observations underscore a crucial aspect: the predictive accuracy of the ensemble model developed in this study is strongly influenced by the selection of both explanatory and target variables. The distributional characteristics and inherent nature of these variables are key determinants of the model’s effectiveness.

Furthermore, this analysis highlights the importance of considering skewness and outliers in the training data. The presence of outliers or a skewed distribution can lead to model overfitting or underperformance, as seen in Case-1 and Case-2. In addition, the study highlights the potential limitations of relying solely on graphical representations to assess accuracy. While the visual alignment of data points provides an intuitive understanding of model performance, it does not always capture the nuances of the predictive accuracy, especially in cases of non-uniform data distributions. In addition, the results suggest that preprocessing techniques, such as the normalization or transformation of target variables, could potentially improve model performance. These techniques could help mitigate the problems posed by skewed distributions or outliers, thereby improving the model’s ability to generalize and predict more accurately.

This detailed analysis confirms that the careful selection and preprocessing of explanatory and target variables are critical to improving the predictive accuracy of ensemble models. This insight is invaluable for future studies and applications, emphasizing the need for a thorough understanding of data characteristics and the application of appropriate statistical methods in predictive modeling. The findings from these cases provide a foundation for developing more robust and accurate predictive models in various fields, especially when data distributions are complex or skewed.

4.2. Selecting Explanatory Variables

As shown in Table 4, there are eight types of explanatory variables in Case-3 and Case-4, each of which was predicted nine times. The first time, all eight explanatory variables were used simultaneously, namely Case-3 and Case-4, and their results are shown in Figure 7. The second time, the remaining explanatory variables were used, except dry density, namely Case-3(a) and Case-4(a), and their results are shown in Figure 8a. The third time, the remaining explanatory variables were used, except effective confining pressure, namely Case-3(b) and Case-4(b), and their results are shown in Figure 8b. The

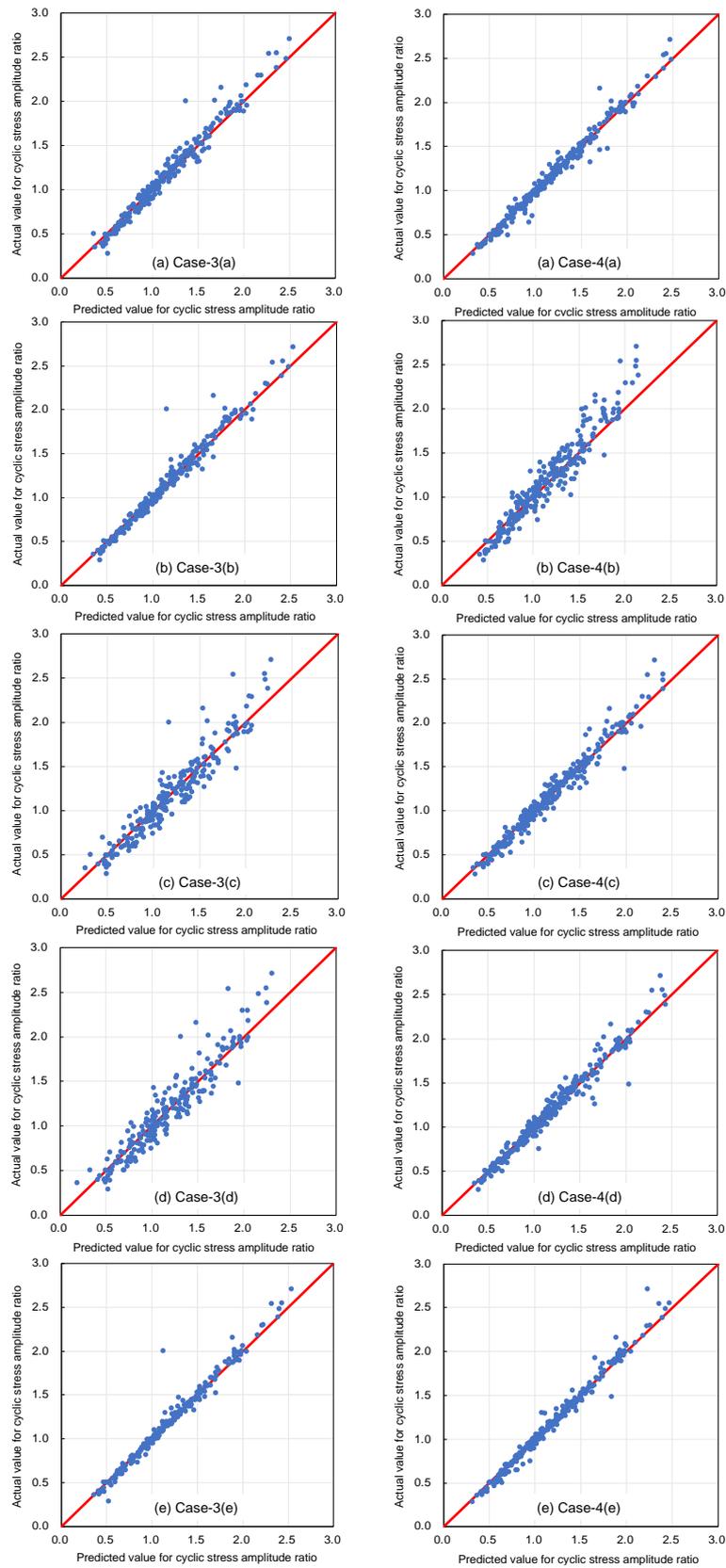


Figure 8. Cont.

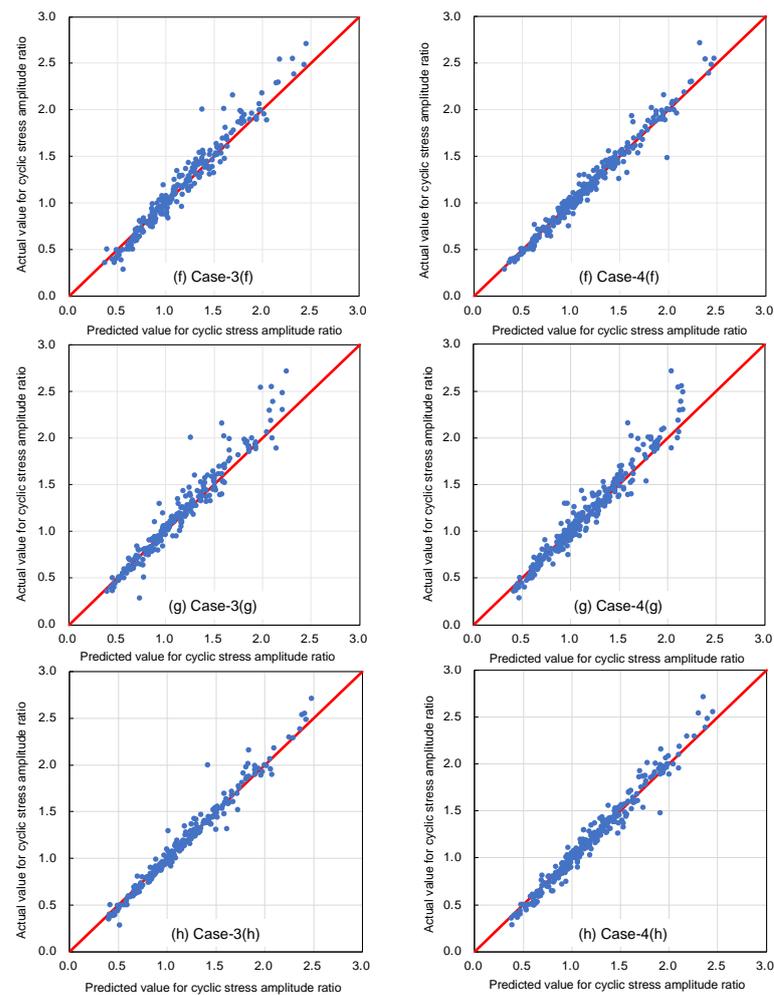


Figure 8. Degrees of deviation between predicted and measured values in (a–h) Cases-3(a) to (h) and (a–h) Cases-4(a) to (h).

However, it is important to note that uniaxial compressive strength is a critical parameter that reflects the strength characteristics of the local sandy soil. Given its importance, its exclusion raises questions about the potential impact on the predictive accuracy when applied to test results from other sites with different soil characteristics. Therefore, while the current study suggests its lesser significance in the context of the dataset used, further investigation is needed to validate this finding across different soil types and conditions. In addition, when all the explanatory variables were used in the prediction model, Case-4 showed a coefficient of determination (R^2) of 0.85. This high value indicates a remarkable level of accuracy and reliability in the predictions. It suggests that integrating ensemble learning methods into the analysis can significantly improve the model's ability to predict the liquefaction resistance in cyclic undrained triaxial tests with high accuracy.

The analysis in Case-4 underscores the importance of selecting appropriate explanatory variables and the potential impact of each on the accuracy of the predictive model. It also highlights the value of ensemble learning methods in improving predictive capabilities, especially in complex geotechnical scenarios such as liquefaction resistance prediction. The results of this study provide a solid foundation for future studies and practical applications in the field of geotechnical engineering.

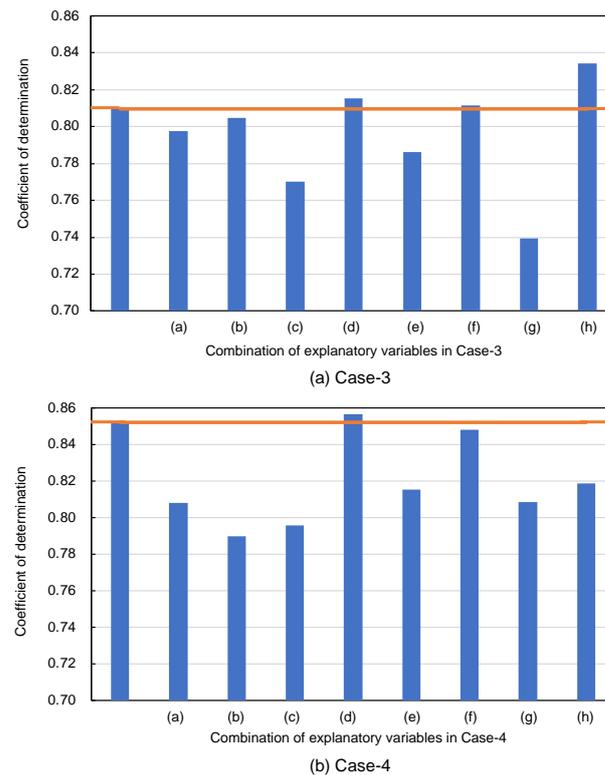


Figure 9. Comparisons of accuracy for predictions in Case-3 (a), Cases-3(a) to (h), Case-4 (b) and Cases-4(a) to (h).

5. Conclusions

In this study, the primary objective was to evaluate the liquefaction resistance of solution-type chemical sandy soil amendments using a novel approach. By utilizing existing experimental databases and artificial intelligence (AI), we sought to achieve accurate predictions without the need to conduct physical experiments. This methodology focused on analyzing data from 20 loading cycles of cyclic undrained triaxial tests and evaluating the impact of various explanatory variables, leading to an investigation of the optimal combinations of these variables for making predictions.

The results of this study are summarized as follows:

- (1) For the development of a predictive model, it is highly recommended to designate the liquefaction resistance ratio as a dependent variable and the other parameters as explanatory variables. This approach allows for a more focused analysis and provides more reliable predictions of the soil behavior under liquefaction conditions.
- (2) The exploration of combinations of explanatory variables revealed that using all available variables tends to produce a more stable coefficient of determination (R^2). This stability is critical to the reliability of the model, especially in applications where precision is paramount.
- (3) Including the liquefaction resistance ratio in the training dataset significantly increases the predictive accuracy of the model. This finding underscores the importance of this particular variable in understanding and predicting the behavior of chemically enhanced sandy soils under stress.
- (4) The results of using AI for making predictions highlight the potential of accurately predicting liquefaction resistance using historical data. This approach not only saves time and resources, but also opens new avenues for studies in soil mechanics and geotechnical engineering.
- (5) In addition, this study aimed to validate the effectiveness of the solution-type chemical improvement of sandy soils against liquefaction through an AI-based analysis of existing data from cyclic undrained triaxial tests. The results of this study confirmed

that high-precision predictions are achievable using the explanatory variables listed in Table 1. In particular, excluding uniaxial compressive strength as an explanatory variable resulted in the highest accuracy, followed closely by scenarios using all explanatory variables. This suggests a nuanced relationship between the variables and their predictive power that warrants further investigation.

Looking ahead, several challenges and opportunities emerge. A key area for future study is to expand the training dataset to include test results from multiple sites. This would improve the generalizability and accuracy of the model and provide a more comprehensive understanding of soil behavior under different geological conditions. In addition, the role of uniaxial compressive strength as an explanatory variable merits further investigation. Its inclusion or exclusion from the model has significant implications for predictive accuracy, suggesting a complex interplay with other variables.

Another future direction is to explore more advanced AI techniques and algorithms. The potential of machine learning and deep learning for improving the predictive models for soil liquefaction resistance is vast and largely untapped. These advanced methods could uncover deeper insights into soil behavior and provide more robust predictive tools for geotechnical engineers. However, the authors believe that even the predictions of advanced AI techniques and algorithms will only serve to interpolate the predictions based on actual experimental works.

In conclusion, this study represents a significant step forward in the application of AI for predicting soil liquefaction resistance. It not only demonstrates the feasibility of using AI for such predictions, but also sets the stage for more sophisticated analyses and applications in the field of geotechnical engineering. The integration of AI with traditional soil mechanics offers a promising avenue for future studies, with the potential to revolutionize the way in which soil improvement and liquefaction resistance analyses are approached.

Author Contributions: Conceptualization, S.I. and T.M.; methodology, S.I.; software, Y.C.; validation, K.N. and S.I.; formal analysis, Y.C. and K.N.; investigation, T.M.; resources, T.M.; data curation, Y.C. and K.N.; writing—original draft preparation, Y.C.; writing—review and editing, S.I.; visualization, S.I.; supervision, S.I.; project administration, S.I.; funding acquisition, T.M. and S.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, Shinya Inazumi, upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kuribayashi, E.; Tatsuoka, F. Brief review of liquefaction during earthquakes in Japan. *Soils Found.* **1975**, *15*, 81–92. [[CrossRef](#)] [[PubMed](#)]
2. Huang, Y.; Yu, M. Review of soil liquefaction characteristics during major earthquakes of the twenty-first century. *Nat. Hazards* **2013**, *65*, 2375–2384. [[CrossRef](#)]
3. Hasheminezhad, A.; Bahadori, H. Three dimensional finite difference simulation of liquefaction phenomenon. *Int. J. Geotech. Eng.* **2019**, *15*, 245–251. [[CrossRef](#)]
4. Nakao, K.; Inazumi, S.; Takahashi, T.; Nontananandh, S. Numerical simulation of the liquefaction phenomenon by MPSM-DEM coupled CAES. *Sustainability* **2022**, *14*, 7517. [[CrossRef](#)]
5. Lo, R.C.; Wang, Y. Lessons learned from recent earthquakes-geoscience and geotechnical perspectives. In *Advances in Geotechnical Earthquake Engineering—Soil Liquefaction and Seismic Safety of Dams and Monuments*; IntechOpen: Rijeka, Croatia, 2012; pp. 1–42. [[CrossRef](#)]
6. Hazout, L.; Zitouni, Z.E.A.; Belkhatir, M.; Schanz, T. Evaluation of static liquefaction characteristics of saturated loose sand through the mean grain size and extreme grain sizes. *Geotech. Geol. Eng.* **2017**, *35*, 2079–2105. [[CrossRef](#)]
7. Bao, X.; Ye, B.; Ye, G.; Zhang, F. Co-seismic and post-seismic behavior of a wall type breakwater on a natural ground composed of liquefiable layer. *Nat. Hazards* **2016**, *83*, 1799–1819. [[CrossRef](#)]
8. Bao, X.; Jin, Z.; Cui, H.; Chen, X.; Xie, X. Soil liquefaction mitigation in geotechnical engineering: An overview of recently developed methods. *Soil Dyn. Earthq. Eng.* **2019**, *120*, 273–291. [[CrossRef](#)]

9. Gallagher, P.M.; Pamuk, A.; Abdoun, T. Stabilization of liquefiable soils using colloidal silica grout. *J. Mater. Civ. Eng.* **2007**, *19*, 33–40. [[CrossRef](#)]
10. Sayehvand, S.; Kalantari, B. Use of grouting method to improve soil stability against liquefaction—A review. *Electron. J. Geotech. Eng.* **2012**, *17*, 1559–1566.
11. Verma, H.; Ray, A.; Rai, R.; Gupta, T.; Mehta, N. Ground improvement using chemical methods: A review. *Heliyon* **2021**, *7*, e07678. [[CrossRef](#)]
12. Yoon, J.C.; Su, S.W.; Kim, J.M. Method for prevention of liquefaction caused by earthquakes using grouting applicable to existing structures. *Appl. Sci.* **2023**, *13*, 1871. [[CrossRef](#)]
13. Motohashi, T.; Sasahara, S.; Inazumi, S. Strength assessment of water-glass sand mixtures. *Gels* **2023**, *9*, 850. [[CrossRef](#)]
14. Yoshimi, Y.; Tanaka, K.; Tokimatsu, K. Liquefaction resistance of a partially saturated sand. *Soils Found.* **1989**, *29*, 157–162. [[CrossRef](#)] [[PubMed](#)]
15. Mele, L.; Flora, A. On the prediction of liquefaction resistance of unsaturated sands. *Soil Dyn. Earthq. Eng.* **2019**, *125*, 105689. [[CrossRef](#)]
16. Park, S.S.; Nong, Z.Z.; Woo, S.W. Liquefaction resistance of Pohang sand. In *Earthquake Geotechnical Engineering for Protection and Development of Environment and Constructions*; CRC Press: London, UK, 2019; pp. 4359–4365.
17. Khashila, M.; Hussien, M.N.; Karray, M.; Chekired, M. Liquefaction resistance from cyclic simple and triaxial shearing: A comparative study. *Acta Geotech.* **2021**, *16*, 1735–1753. [[CrossRef](#)]
18. Ni, X.Q.; Zhang, Z.; Ye, B.; Zhang, S. Unique relation between pore water pressure generated at the first loading cycle and liquefaction resistance. *Eng. Geol.* **2022**, *296*, 106476. [[CrossRef](#)]
19. Toyota, H.; Takada, S. Variation of liquefaction strength induced by monotonic and cyclic loading histories. *J. Geotech. Geoenviron. Eng.* **2017**, *143*, 04016120. [[CrossRef](#)]
20. Zhang, J.; Yang, Z.; Yang, Q.; Yang, G.; Li, G.; Liu, J. Liquefaction behavior of fiber-reinforced sand based on cyclic triaxial tests. *Geosynth. Int.* **2021**, *28*, 316–326. [[CrossRef](#)]
21. Amer, M.I.; Kovacs, W.D.; Aggour, M.S. Cyclic simple shear size effects. *J. Geotech. Eng.* **1987**, *113*, 7, 693–707. [[CrossRef](#)]
22. Nong, Z.Z.; Park, S.S.; Lee, D.E. Comparison of sand liquefaction in cyclic triaxial and simple shear tests. *Soils Found.* **2021**, *61*, 1071–1085. [[CrossRef](#)]
23. Nong, Z.Z.; Park, S.S.; Jeong, S.W.; Lee, D.E. Effect of cyclic loading frequency on liquefaction prediction of sand. *Appl. Sci.* **2021**, *10*, 4502. [[CrossRef](#)]
24. Arpit, J.; Mittal, S.; Shukla, S.K. Liquefaction proneness of stratified sand-silt layers based on cyclic triaxial tests. *J. Rock Mech. Geotech. Eng.* **2023**, *15*, 1826–1845. [[CrossRef](#)]
25. Hyodo, M.; Yasuhara, K.; Hirao, K. Prediction of clay behaviour in undrained and partially drained cyclic triaxial tests. *Soils Found.* **1992**, *32*, 117–127. [[CrossRef](#)]
26. Gu, C.; Wang, J.; Cai, Y.; Yang, Z.; Gao, Y. Undrained cyclic triaxial behavior of saturated clays under variable confining pressure. *Soil Dyn. Earthq. Eng.* **2012**, *40*, 118–128. [[CrossRef](#)]
27. Ghayoomi, M.; Suprunenko, G.; Mirshekari, M. Cyclic triaxial test to measure strain-dependent shear modulus of unsaturated sand. *Int. J. Geomech.* **2017**, *17*, 04017043. [[CrossRef](#)]
28. Dong, X.; Yu, Z.; Cao, W.; Shi, Y.; Ma, Q. A survey on ensemble learning. *Front. Comput. Sci.* **2020**, *14*, 241–258. [[CrossRef](#)]
29. Chongzhi, W.; Lin, W.; Zhang, W. Assessment of undrained shear strength using ensemble learning based on Bayesian hyperparameter optimization. In *Modeling in Geotechnical Engineering*; Academic Press: Cambridge, MA, USA, 2021; pp. 309–326. [[CrossRef](#)]
30. Wang, Z.Z.; Hu, Y.; Guo, X.; He, X.; Kek, H.Y.; Ku, T.; Goh, S.H.; Leung, C.F. Predicting geological interfaces using stacking ensemble learning with multi-scale features. *Can. Geotech. J.* **2023**, *60*, 7. [[CrossRef](#)]
31. Shahin, M.A. Artificial intelligence in geotechnical engineering: Applications, modeling aspects, and future directions. In *Metaheuristics in Water, Geotechnical and Transport Engineering*; Elsevier: Amsterdam, The Netherlands, 2013; pp. 169–204.
32. Jong, S.C.; Ong, D.E.L.; Oh, E. State-of-the-art review of geotechnical-driven artificial intelligence techniques in underground soil-structure interaction. *Tunn. Undergr. Space Technol.* **2021**, *113*, 103946. [[CrossRef](#)]
33. Baghbani, A.; Choudhury, T.; Costa, S.; Reiner, J. Application of artificial intelligence in geotechnical engineering: A state-of-the-art review. *Earth-Sci. Rev.* **2022**, *228*, 103991. [[CrossRef](#)]
34. Sharma, S.; Ahmed, S.; Naseem, M.; Alnumay, W.S.; Singh, S.; Cho, G.H. A survey on applications of artificial intelligence for pre-parametric project cost and soil shear-strength estimation in construction and geotechnical engineering. *Sensors* **2021**, *21*, 463. [[CrossRef](#)]
35. Huang, Y.; Wen, Z. Recent developments of soil improvement methods for seismic liquefaction mitigation. *Nat. Hazards* **2015**, *76*, 1927–1938. [[CrossRef](#)]
36. Cardellicchio, A.; Ruggieri, S.; Nettis, A.; Renò, V.; Uva, G. Physical interpretation of machine learning-based recognition of defects for the risk management of existing bridge heritage. *Eng. Fail. Anal.* **2023**, *149*, 107237. [[CrossRef](#)]
37. Mele, L.; Tian, J.T.; Lirer, S.; Flora, A.; Koseki, J. Liquefaction resistance of unsaturated sands: Experimental evidence and theoretical interpretation. *Géotechnique* **2019**, *69*, 541–553. [[CrossRef](#)]
38. Fahim, A.K.F.; Rahman, M.Z.; Hossain, M.S.; Kamal, A.M. Liquefaction resistance evaluation of soils using artificial neural network for Dhaka City, Bangladesh. *Nat. Hazards* **2022**, *113*, 933–963. [[CrossRef](#)]

39. Kotsiantis, S.B.; Zaharakis, I.D.; Pintelas, P.E. Machine learning: A review of classification and combining techniques. *Artif. Intell. Rev.* **2006**, *26*, 159–190. [[CrossRef](#)]
40. Pólvara, A.; Nascimento, S.; Lourenço, J.S.; Scapolo, F. Blockchain for industrial transformations: A forward-looking approach with multi-stakeholder engagement for policy advice. *Technol. Forecast. Soc. Change* **2020**, *157*, 120091. [[CrossRef](#)]
41. Tuttle, M.P.; Hartleb, R.; Wolf, L.; Mayne, P.W. Paleoliquefaction studies and the evaluation of seismic hazard. *Geosciences* **2019**, *9*, 311. [[CrossRef](#)]
42. Dhakal, R.; Cubrinovski, M.; Bray, J.D. Geotechnical characterization and liquefaction evaluation of gravelly reclamations and hydraulic fills (Port of Wellington, New Zealand). *Soils Found.* **2020**, *60*, 1507–1531. [[CrossRef](#)]
43. Yang, M.; Taiebat, M.; Radjaï, F. Liquefaction of granular materials in constant-volume cyclic shearing: Transition between solid-like and fluid-like states. *Comput. Geotech.* **2022**, *148*, 104800. [[CrossRef](#)]
44. Mitchell, J.K. Mitigation of liquefaction potential of silty sands. In *Research to Practice in Geotechnical Engineering*; American Society of Civil Engineers: Reston, VA, USA, 2008; pp. 433–451. [[CrossRef](#)]
45. Kramer, S.L.; Mitchell, R.A. Ground motion intensity measures for liquefaction hazard evaluation. *Earthq. Spectra* **2006**, *22*, 413–438. [[CrossRef](#)]
46. Zhou, Z.H. *Ensemble Methods: Foundations and Algorithms*; Chapman & Hall/CRC Machine Learning & Pattern Recognition; CRC Press: Boca Raton, FL, USA, 2012.
47. Gurney, K. *An Introduction to Neural Networks*, 1st ed.; CRC Press: Boca Raton, FL, USA, 1997.
48. Zhao, Z.; Chow, T.L.; Rees, H.W.; Yang, Q.; Xing, Z.; Meng, F.R. Predict soil texture distributions using an artificial neural network model. *Comput. Electron. Agric.* **2009**, *65*, 36–48. [[CrossRef](#)]
49. Zhong, L.; Guo, X.; Xu, Z.; Ding, M. Soil properties: Their prediction and feature extraction from the LUCAS spectral library using deep convolutional neural networks. *Geoderma* **2021**, *402*, 115366. [[CrossRef](#)]
50. Pham, B.T.; Nguyen, M.D.; Ly, H.B.; Pham, T.A.; Hoang, V.; Le, V.H.; Le, T.T.; Nguyen, H.Q.; Bui, G.L. Development of artificial neural networks for prediction of compression coefficient of soft soil. In Proceedings of the 5th International Conference on Geotechnics, Civil Engineering Works and Structures, Virtual, 18–20 October 2020; pp. 1167–1172.
51. Chakraborty, D.; Elhegazy, H.; Elzarka, H.; Gutierrez, L. A novel construction cost prediction model using hybrid natural and light gradient boosting. *Adv. Eng. Inform.* **2020**, *46*, 101201. [[CrossRef](#)]
52. Lee, S.; Vo, T.P.; Thai, H.T.; Lee, J.; Patel, V. Strength prediction of concrete-filled steel tubular columns using Categorical Gradient Boosting algorithm. *Eng. Struct.* **2021**, *238*, 112109. [[CrossRef](#)]
53. Guo, R.; Fu, D.; Sollazzo, G. An ensemble learning model for asphalt pavement performance prediction based on gradient boosting decision tree. *Int. J. Pavement Eng.* **2022**, *23*, 3633–3646. [[CrossRef](#)]
54. Nakagawa, S.; Johnson, P.C.D.; Schielzeth, H. The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J. R. Soc. Interface* **2017**, *14*, 1–11. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.