Machine Learning-Based Assessment of Liquefaction Susceptibility in Tailings Dams

Maryam Saaly, Klohn Crippen Berger, Canada
Ali Reza Zafarani, Klohn Crippen Berger, Canada
Vicki Nguyen, Klohn Crippen Berger, Canada

Abstract

Soil liquefaction is a failure mechanism that has contributed to catastrophic tailings dam failures in the past. To limit the potential of liquefaction triggering, geotechnical engineers must assess the susceptibility of key soil units to liquefaction. Traditional liquefaction assessment methods, such as empirical correlations, can be time-consuming. Therefore, this study investigates the use of machine learning (ML) techniques to classify liquefaction susceptibility of tailings using only a limited set of raw CPT-based parameters, including cone resistance (qt), sleeve friction (fs), friction ratio (FR), as well as USGS soil classification.

A dataset compiled from previous liquefaction case studies at tailings facilities across the world was used to develop two ML classification models using the XGBoost algorithm. In the first model, liquefiable and non-liquefiable layers were identified based on interpreted CPT data using the Robertson (2021) method. The model was trained on both the raw CPT outputs and the interpreted CPT-based outputs including state parameter and clean sand equivalent normalized cone resistance ($Q_{tn,cs}$). A ML model was developed based on the outputs of the raw CPTs and obtained values of the interpreted CPT. Additionally a second ML model was developed incorporating only the raw CPT output (i.e., qt, fs, and FR) as well as the USGS soil classification of that layer. The objective is to evaluate whether a reliable classification of liquefiable layers is possible using only raw CPT-based field data, to be able to trust the ML model in-future and use the model instead of interpreting the raw CPT data.

Results show that the full-feature model achieves 100% accuracy, while the reduced model achieves 87% accuracy. These findings highlight the potential of ML approaches to support preliminary screening in data-limited environments, offering a conservative yet practical decision-support tool for geotechnical engineers. By leveraging machine learning techniques, the liquefaction assessment process can be streamlined and applied more consistently across large datasets, enabling rapid evaluation of CPT records

where traditional empirical or analytical methods would be time-consuming or may require manual interpretation.

Introduction

Soil liquefaction is a critical phenomenon affecting the stability of tailings dams, especially in seismic zones or under undrained loading conditions. In tailings facilities, the loss of strength due to liquefaction can result in flow failures, uncontrolled release of tailings, and significant environmental and socio-economic consequences (Bhattacharya et. al (2011), Bray et. al. (2014)). Traditional liquefaction evaluation methods, such as the simplified procedure or correlation-based approaches, typically rely on detailed in-situ testing and often assume idealized conditions. While these tools are valuable, they may not be well-suited for highly variable tailings deposits or data-sparse projects.

As tailings dams are increasingly scrutinized for safety and performance, there is a growing need for innovative assessment tools that can work with limited data, adapt to a variety of geological settings, and enhance predictive accuracy. Recent advances in artificial intelligence and data-driven approaches offer an opportunity to augment geotechnical engineering practices through machine learning (ML). In particular, ensemble-based methods like Extreme Gradient Boosting (XGBoost) have shown strong classification performance in geotechnical applications involving nonlinear and multivariate data (Demir et. al. (2022)).

This study explores the application of XGBoost models for predicting the liquefaction potential of tailings using CPT-based and classification inputs. Two distinct models were developed: one incorporating a full set of geotechnical features obtained from the cone penetration tests (CPTs), and another limited to a smaller subset of four parameters. The study compares the performance of these models and discusses their implications for early-stage screening and decision-making under data constraints.

Methodology

Data Source and Compilation

The dataset used in this study was compiled from CPTs, laboratory investigations, and classification data from tailings storage facilities located across the world. In total 16,000 data points were collected from the reviewed CPTs and corresponding laboratory data. Each data point corresponds to a reading interval of the CPT and includes inputs such as cone resistance (qt), sleeve friction (fs), friction ratio (FR), pore pressure (PP), soil behavior type index (Ic), and others. The corresponding liquefaction label ("Y" or "N") was based on the results from a liquefaction assessment that was completed for each CPT in accordance with the method prescribed by Robertson (2021). In addition to the CPT data, if available, grain size distribution tests were collected in the dataset in form of USGS classification of that specific soil layer.

As stated above, the CPT data were processed using Robertson (2021) methodology to identify the potentially liquifiable layers. The key outputs of this CPT interpretation methodology for liquefaction susceptibility are the state parameter (ψ), SBT index (I_c) and clean sand equivalent normalized cone resistance ($Q_{tn,cs}$). The following equations are commonly used to define the aforementioned parameters.

The normalized cone resistance, Q_t is defined as:

$$Q_t = \frac{q_c - \sigma_{v0}}{\sigma'_{v0}} \tag{1}$$

Where:

q_c is the measured cone tip resistance, σ_{v0} is total vertical stress, and σ'_{v0} is the effective vertical stress.

To account for overburden effects, Q_t is further corrected as:

$$Q_{tn} = Q_t \cdot \left(\frac{P_a}{\sigma'_{v0}}\right)^n \tag{2}$$

Where P_a is the atmospheric pressure, and n is a stress normalization exponent.

To account for the influence of fines content on cone resistance, clean sand equivalent normalized cone resistance ($Q_{tn,cs}$) is defined by:

$$Q_{tn,cs} = Q_{tn} \cdot (1 + \frac{FC}{100}) \tag{3}$$

where *FC* is the fines content (%).

The normalized friction ratio, F_r is given by:

$$F_r = \frac{f_s}{q_c - \sigma_{v0}} \times 100 \tag{4}$$

Where f_s is the sleeve friction resistance.

SBT index (I_c) is calculated as:

$$I_C = \sqrt{(3.47 - \log_{10}(Q_t))^2 + (\log_{10}(F_r) + 1.22)^2} \tag{5}$$

The state parameter (ψ) which represents the distance of a soil from its critical state, can be estimated from CPT data as:

$$\psi = a - b \cdot \log_{10}(Q_{tn})$$

Where *a* and *b* are material-based coefficients.

Data Preprocessing

Categorical variables, such as USGS soil classification and tailings type, were transformed into a numerical format suitable for modeling using one-hot encoding. The target variable, liquefaction status, was binarized (1 = liquefied, 0 = non-liquefied).

Two feature sets were prepared:

- Full model: raw and processed CPT input features
- Reduced model: 4 features qt, fs, FR, and USGS classification

Model Training and Evaluation

The dataset was randomly split into two sets for training (80% of the data) and testing (20% of the data). XGBoost classifiers were trained on both feature sets using scikit-learn wrappers. Although XGBoost is a boosting algorithm and does not rely on traditional bagging (i.e., training on bootstrapped samples), it incorporates bagging-like techniques such as row and column subsampling to enhance generalization and reduce overfitting. Model performance was evaluated using accuracy (i.e., the proportion of total predictions that were correct), precision (i.e., the proportion of positive predictions that were actually correct), recall (i.e., the proportion of actual positives that were correctly identified), and confusion matrices (a table showing true vs. predicted classifications).

The modeling workflow was implemented in Python using the following libraries: pandas for data manipulation, scikit-learn for splitting and evaluation, XGBoost for training, and seaborn/matplotlib for visualization.

Results and Discussion

Model Performance Overview

The full-feature XGBoost model, which incorporated raw and interpreted CPT output variables, achieved almost a classification accuracy of 100%, with perfect recall in identifying liquefiable layers. This demonstrates the model's effectiveness in utilizing comprehensive geotechnical data for high-confidence predictions.

In contrast, the reduced-feature model, which used only four accessible parameters (qt, fs, FR, and USGS classification), achieved a lower accuracy of 87%. While the model still maintained a high recall (i.e., successfully identifying the liquefiable layers), it exhibited a greater number of misclassifications of non-liquefiable samples, resulting in a higher false positive rate (i.e., the layer is non-liquefiable but the model identified it as liquefiable). Although conservative misclassifications are safer from a design perspective, the reduction in overall accuracy highlights the trade-off between data simplicity and predictive precision. Importantly, the false negative rate remained low, indicating the model rarely missed identifying actual liquefiable layers, a critical consideration for risk-averse applications. It is believed that with a larger dataset, the model could be trained better and potentially achieve better generalization and improved performance across all metrics, including accuracy (the proportion of total predictions that were correct), precision (the proportion of positive predictions that were actually correct), and recall (the proportion of actual positives that were correctly identified). The confusion matrix for the full-feature model and reduced-feature model are presented in Figure 1 and Figure 2, respectively.

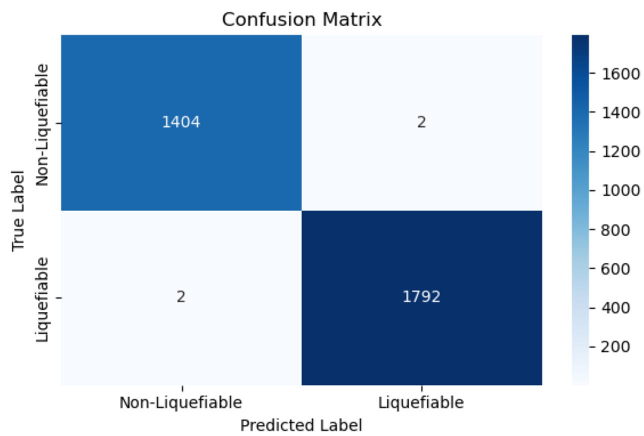

Figure 1: Confusion Matrix for the Model Using all Features

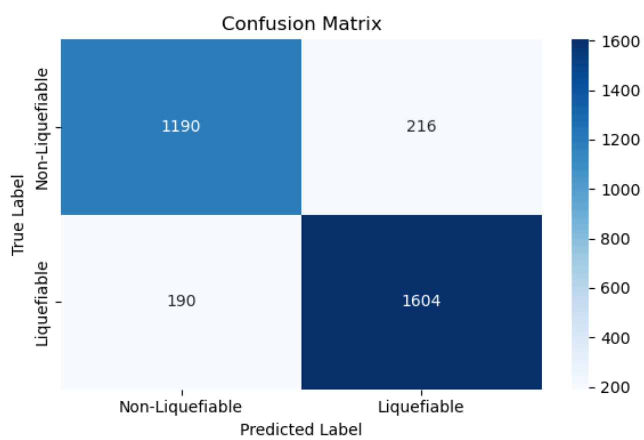

Figure 2: Confusion Matrix for the Model Using Reduced Features

Feature Contribution

Feature importance rankings from the reduced number of features indicated that USGS classification and f_s were important predictors of liquefaction. As expected, the state parameter and Q_{tncs} were the dominant prediction factors for the model that considered all features. The USGS classification also contributed to model decisions, reinforcing its value in cases where lab data are available. However, only limited number of layers had corresponding grain size distribution tests. Therefore, if more laboratory test data were made available in the dataset for which the ML model was trained, more accurate decisions regarding the layer could be made.

Comparison and Implications

The findings show that while reduced models can still support initial screening, their predictive reliability may be compromised in nuanced geological conditions. For high-stakes applications, especially in seismicprone or critical infrastructure sites, the full-feature model offers a more robust basis for decision-making.

Nonetheless, the reduced-feature model's ability to identify all known liquefiable layers without false negatives suggests it could still serve as a conservative, low-data alternative when access to complete geotechnical datasets is constrained.

Table 1: Model Performance Comparison

Model	Accuracy	Precision	False Negatives	False Positives
Raw CPT and USGS class.	87%	88%	190	216
Raw and interpreted CPT and USGS class	100%	100%	2	2

Conclusion

This study demonstrates the effectiveness of machine learning, particularly XGBoost, in assessing liquefaction susceptibility of tailings using CPT-based and classification data. Two models were developed and tested: one incorporating a comprehensive set of geotechnical features (i.e., raw and processed CPTs and USGS classification), and another relying on a reduced set of just four inputs (i.e., raw CPTs and USGS classification). The full-feature model achieved strong classification performance with almost 100% accuracy, while the reduced-feature model achieved 87% accuracy.

Although the reduced-feature model showed a decline in overall accuracy, the false negative predictions are less than 5%. This suggests that the reduced-feature model may serve as an effective preliminary screening tool, helping to identify cases where a more detailed liquefaction assessment is warranted, particularly in early project stages or when rapid decision-making is required.

Incorporating ML-based liquefaction models into engineering practice enables a more data-informed and adaptable approach, especially in regions where traditional methods may be limited by sparse or costly data collection. Future work may focus on expanding the dataset with additional tailings types and incorporating real-time monitoring data to enhance model adaptability and robustness. into the geotechnical workflow, especially during early-stage investigations. The findings emphasize the potential of data-driven methods to complement traditional assessments and guide more efficient resource allocation for further testing and monitoring.

References

Bhattacharya, S., Hyodo, M., Goda, K., Tazoh, T., & Taylor, C. A. (2011). Liquefaction of soil in the Tokyo Bay area from the 2011 Tohoku (Japan) earthquake. Soil Dynamics and Earthquake Engineering, 31(11), 1618-1628.

Demir, S., & Sahin, E. K. (2022). Comparison of tree-based machine learning algorithms for predicting liquefaction potential using canonical correlation forest, rotation forest, and random forest based on CPT data. Soil Dynamics and Earthquake Engineering, 154, 107130.

Bray, J., Cubrinovski, M., Zupan, J., & Taylor, M. (2014). Liquefaction effects on buildings in the central business district of Christchurch. Earthquake Spectra, 30(1), 85-109.

Robertson, P. K. (2021). Evaluation of cyclic liquefaction potential based on the CPT. In Seismic Behaviour of Ground and Geotechnical Structures: Special Volume of TC 4 (pp. 269-277). CRC Press.