International Journal of GEOMATE, June, 2026 Vol.30, Issue 142, pp.118-125
ISSN: 2186-2982 (P), 2186-2990 (O), Japan, DOI: https://doi.org/10.21660/2026.142.g15157
Geotechnique, Construction Materials and Environment

UNDERSTANDING THE MECHANISMS OF LIQUEFACTION-INDUCED SOIL EJECTA USING INTERPRETABLE MACHINE LEARNING

*Emerzon S. Torres [1,2], and Jonathan R. Dungca[2]

[1]School of Innovation and Sustainability, De La Salle Univ., Biñan, Laguna, Philippines
[2]Dept. of Civil Eng., De La Salle Univ., Manila, Philippines

*Corresponding Author, Received: 29 May 2025, Revised: 07 March 2026, Accepted: 10 March 2026

ABSTRACT: Soil ejecta, commonly observed as sand boils, is a prominent manifestation of soil liquefaction during earthquakes, often leading to ground settlement, structural damage, and post-earthquake site instability. Despite their impact, traditional liquefaction models largely focus on triggering mechanisms and rarely address the conditions that govern ejecta formation. This study applies Rough Set Machine Learning (RSML), an interpretable rule-based algorithm, to analyze 96 historical case histories of liquefaction, including 85 cases with observed soil ejecta. Six geotechnical and seismic parameters—moment magnitude, peak ground acceleration, groundwater table depth, average critical layer depth, corrected SPT N-value, and fines content—were used as model inputs. IF–THEN rules were induced and evaluated using support, strength, certainty, and coverage metrics. Sensitivity and interaction analyses revealed the critical influence of corrected penetration resistance, groundwater conditions, and fines content on ejecta occurrence. Notably, soil ejecta was found to occur even in moderately dense or fine-grained soils, challenging conventional assumptions. The parameter interaction map and scenario map further highlighted the combined effects of seismic and subsurface conditions. This interpretable model provides valuable insights into post-liquefaction behavior and supports improved hazard assessment, mitigation planning, and resilience strategies. The findings contribute to the advancement of data-driven approaches in earthquake geotechnical engineering.

Keywords: Sand boils, Interpretable machine learning, Rough set theory, Liquefaction hazard assessment, Artificial intelligence

1. INTRODUCTION

Soil liquefaction remains one of the most critical and complex phenomena in soil dynamics [1]. Its occurrence is governed by dynamic, non-linear interactions among multiple variables, including seismic parameters (magnitude, frequency), soil properties (grain size, density), hydraulic conditions (pore water pressure, drainage), and broader geotechnical factors (soil stratification, confining pressure). A prominent and damaging manifestation of this complex phenomenon is soil ejecta, as shown in Fig. 1. Often observed as sand boils, soil ejecta significantly contributes to ground settlement, structural damage, and post-earthquake site instability [2]. The ejection of subsurface material can compromise foundation support, damage underground utilities, and cause severe differential settlement and structural cracking.

Despite the severe consequences of soil ejecta, current liquefaction assessment models primarily focus on general triggering mechanisms and often overlook the specific factors governing ejecta formation. The fundamental challenge lies in the inability of traditional empirical methods, such as simplified stress-based approaches, to capture complex relationships dictating soil behavior. By treating parameters largely in isolation, these methods oversimplify the intricate, interconnected nature of soil properties and seismic conditions. While some studies have calculated liquefaction indices based on Standard Penetration Tests (SPT) [3] or explored lateral spreading mechanisms [4], a notable gap remains in the analysis of historical liquefaction databases that explicitly account for the presence and characteristics of soil ejecta.

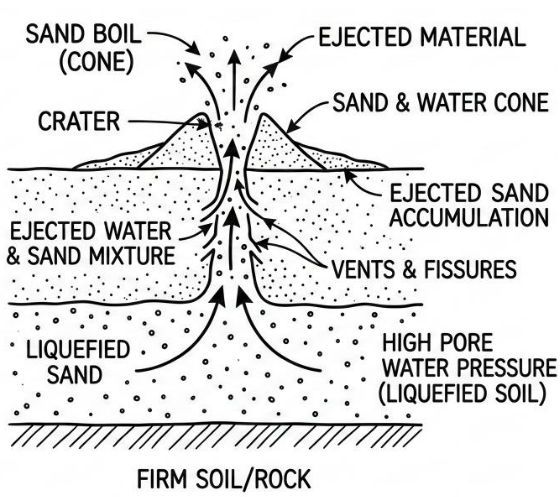

Fig.1 Soil ejecta formation during soil liquefaction.

To overcome the limitations of traditional models, researchers have increasingly turned to historical data, leveraging big data and artificial intelligence to unravel the complexities of seismic soil behavior. However, while advanced machine learning models (like neural networks) offer improved predictive accuracy, they often act as "black boxes" lacking practical interpretability. Interpretable machine learning, particularly Rough Set Machine Learning (RSML), offers a compelling solution. RSML captures non-linear relationships and complex parameter interactions while maintaining strict transparency in its predictions. Unlike traditional stress-based approaches that rely on conservative simplifications, RSML uncovers hidden patterns directly from case histories, generating clear, physically interpretable IF-THEN rules. This bridges the gap between sophisticated computational analysis and practical engineering applications.

Some studies have successfully applied interpretable machine learning in geotechnical engineering applications. Some studies applied RSML in the triggering of liquefaction using historical SPT-based liquefaction data [5-6]. These studies showed that interpretable machine learning models align well with the state-of-practice stress-based model and can significantly predict triggering if soil liquefaction by applying sets of IF-THEN rules.

To address the critical gap in current assessments, this study employs RSML to analyze and extract patterns from historical case histories of liquefaction-induced soil ejecta. By developing an interpretable, rule-based model, this research aims to identify which individual parameters and parameter combinations most significantly influence the occurrence of sand boils. Ultimately, the generated IF-THEN rules will not only predict potential outcomes but also provide deeper insights into the underlying mechanisms of soil ejecta, serving as a robust, practical supplement to traditional liquefaction assessment methods.

2. RESEARCH SIGNIFICANCE

The significance of this study lies in its dual contribution to geotechnical hazard assessment and applied machine learning. By focusing explicitly on soil ejecta, this research addresses a critical blind spot in traditional seismic risk evaluations. Methodologically, the application of Rough Set Machine Learning (RSML) advances the field beyond overly simplified empirical methods and opaque "black-box" artificial intelligence by offering a transparent, rule-based approach. The concrete outputs of this study—specifically the decision rules, sensitivity, and parametric analyses—provide researchers and engineers with valuable insights into the mechanisms underlying ejecta formation. By extracting physically interpretable IF-THEN rules from historical data, the model equips practitioners with a highly accessible tool to evaluate complex, non-linear parameter interactions. Consequently, these findings can directly guide the selection of suitable ground improvement techniques and disaster risk reduction strategies tailored to specific site conditions. Ultimately, this enhanced predictive capability will facilitate more reliable site assessments and the improved protection of critical infrastructure, while the unconventional patterns revealed by the induced rules highlight vital new directions for future geotechnical research.

3. METHODS

The research flow is illustrated in Fig. 2. From the soil liquefaction database [7-8], a total of 96 case histories involving soil ejecta were utilized in this study. Minor data cleansing was performed, followed by discretization to prepare the data for machine learning. The Rough Set Exploration System (RSES 2.2) [9], a rough set-based software, was employed to induce decision rules from the constructed decision table. Rule pruning and shortening were applied to eliminate redundant and weak rules. The resulting rule sets were evaluated to identify the most robust and informative set. The optimal rule set was subsequently subjected to sensitivity and parametric analyses to assess the influence of individual parameters and their interactions.

3.1 Soil Ejecta Database

From the available global database of liquefied sites, 96 events were selected for this study, as summarized in Table 1. Among these, 85 sites exhibited manifestations of soil ejecta, while 11 showed no such occurrence. Six conditional attributes (input parameters) and one decision attribute (output parameter) were collected and used for the simulation. Table 2 shows the sample spreadsheet of raw data. The field names and definitions, along with the discretization schemes and frequency distributions of the variables, are presented in Table 3. The decision attribute represents the presence or absence of observed soil ejecta.

3.2 Rough Set Machine Learning

Rough Set Theory (RST), developed by Pawlak (1982), addresses vagueness and uncertainty in knowledge-based systems [10]. In RST, knowledge is typically represented using a decision table composed of rows and columns, where attributes are categorized into condition and decision attributes. Each row in the decision table represents a rule in the form: IF (conditions) … THEN (decision). An inductive process is applied to extract rules from the data contained in the decision table. RST has been successfully applied in various machine learning tasks, including clustering, feature selection, and rule induction. In this study, rule induction was carried out to generate interpretable IF-THEN statements that define patterns associated with soil ejecta occurrence. The induced rules were evaluated using four key validation statistics: *support*, *strength*, *certainty factor*, and *coverage*. *Support* indicates the number of observations that follow a given rule, while *strength* is the proportion of these supports relative to the total number of observations in the dataset. The *certainty factor* reflects the likelihood that an observation belongs to the decision class when the conditions of a rule are met. Lastly, *coverage* measures the extent to which a particular rule accounts for the instances within the decision class.

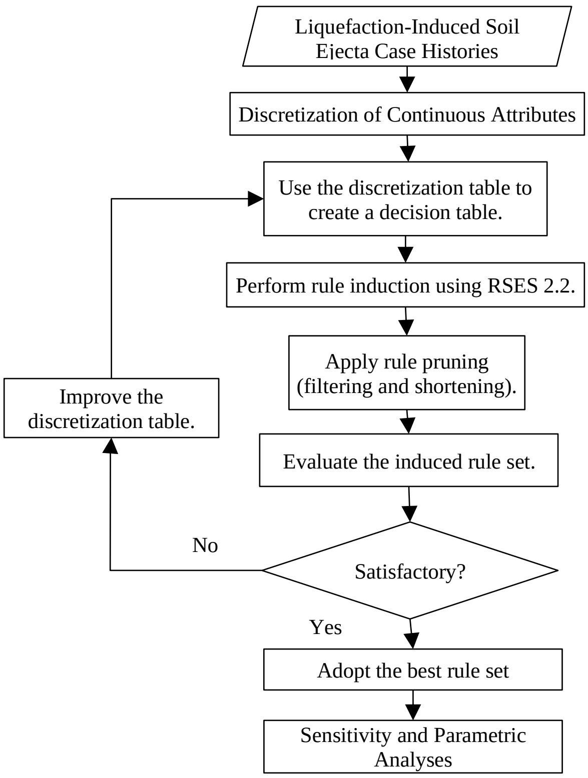

Fig.2 General framework of the study.

Table 4 presents a sample decision table used in this study, with attributes discretized into two to three bins: very high, high, and low. This table was input into the RSES 2.2 to generate decision rules along with their corresponding validation statistics.

3.3 Sensitivity and Parametric Analyses

Ablation-based sensitivity analysis was conducted to assess the importance of each input attribute in the rule-based model. By systematically removing one condition attribute at a time and measuring the change in model accuracy through cross-validation (10 and 50 k-folds), the analysis revealed each attribute's contribution to model performance. This method provided valuable insights into the stability and robustness of the model and the interplay between input parameters.

Table 1 Soil ejecta database used.

Earthquake Event	No.of Sites	Earthquake Event	No. of Sites
1944 Tohnankai	3	1979 Imperial Valley	4
1948 Fukui	2	1981 WestMorland	3
1964 Niigata	4	1983 Nihonkai-Chubu June 21	1
1968 Tokachi-Oki	3	1983 Nihonkai-Chubu May 26	12
1971 San Fernando	2	1987 Superstition Hills	1
1975 Haicheng	3	1989 Loma Prieta	14
1976 Guatemala	2	1990 Luzon	1
1976 Tangshan	1	1993 Kushiro-Oki	2
1977 Argentina	3	1994 Northridge	3
1978 Miyagiken-Oki Feb 20	1	1995 Hyogoken-Nambu (Kobe)	5
1978 Miyagiken-Oki June 12	14	1999 Kocaeli	12
Total no. of sites		96	

Table 2 Sample spreadsheet of raw data for sand boils rule-based model.

ID	EQ Event	M	a_{max}	Avg Depth	Depth of GWT	(N1)60	FC	Sand boil?
1	1944 Tohnankai	8.1	0.2	5.2	2.1	8.2	10	Y
2	1944 Tohnankai	8.1	0.2	4.3	2.4	3.4	30	Y
3	1944 Tohnankai	8.1	0.2	3.7	2.1	1.7	27	Y
4	1948 Fukui	7	0.4	4	1.2	11.8	0	Y
5	1948 Fukui	7	0.35	7.5	3.7	21.1	4	Y
6	1964 Niigata	7.6	0.09	3.3	1	4.7	5	Y
…	…	⋮	⋮	⋮	⋮	⋮	⋮	…
95	1999 Kocaeli	7.5	0.4	9.6	1.7	16.5	11	N
96	1999 Kocaeli	7.5	0.3	5.1	0.8	14.4	19	N

Note: Y=Yes, N=No

Table 3 Field names, attributes, and discretization of variables.

Attributes	Definition	Low		High		Very High	
		Range	Count	Range	Count	Range	Count
M	Moment Magnitude	[5.9, 7.5)	20	[7.5, 8.3]	76	-	-
a_{max} (g)	Maximum Acceleration	[0.09, 0.28)	51	[0.28, 0.693)	43	[0.693, 0.84]	2
Avg depth (m)	Average Depth	[1.8, 4.9)	47	[4.9, 10. 3)	46	[10.3, 11.5]	3
Depth GWT (m)	Depth of Groundwater Table	[0, 1.7)	47	[1.7, 4.275]	38	[4.275, 7.2]	11
$(N_1)_{60}$	Corrected SPT N Value	[1.7, 10.3)	51	[10.3, 24.1)	44	[24.1, 25.9]	1
FC (%)	Fines Content	[0,8)	42	[8, 60.5)	43	[60.5, 92]	11
Sand boil?	No or Yes	No	11	Yes	85	-	-

To enhance interpretability, a parametric analysis in the form of a parameter interaction map was developed using insights from sensitivity analysis and rule-based interactions. Ablation analysis measured the impact of removing individual parameters on model accuracy, classifying their importance as high, moderate, or low based on performance delta from 100-fold cross-validation. Additionally, parameter interactions were quantified by the frequency of co-occurrence within the induced rule set, categorized as strong, moderate, or weak. The interaction map visually represents these relationships, offering a clearer understanding of how parameters influence each other and contribute to soil ejecta susceptibility.

4. RESULTS AND DISCUSSION

This study developed a liquefaction-induced soil ejecta rule-based model using RSML. Its primary objective was to identify and understand the key factors influencing the occurrence of soil ejecta. The model utilizes six conditional attributes to determine a single decision attribute: the presence or absence of soil ejecta.

To improve model performance, an iterative optimization of the discretization process was conducted until the best rule-based model emerged, yielding the most favorable validation statistics. The model is characterized by a highly imbalanced distribution of decision classes, with 85 events showing the presence and only 11 showing the absence of soil ejecta. This imbalance reflects the frequent occurrence of ejecta in liquefaction events and introduces challenges in rule induction. The selected optimal model was selected, achieving 96.4% accuracy and 87.5% coverage, balancing predictive performance with reliable representation of liquefaction-induced soil ejecta. Table 4 presents the optimal rule set for the soil ejecta rule-based model, comprising 24 rules—1 for the absence and 23 for the presence of soil ejecta. This distribution reflects the common occurrence of soil ejecta in liquefaction events. As shown in Fig. 2, several rules (specifically Rules 2–8 and 14) exhibit coverage factors exceeding 20%, indicating strong explanatory power within the dataset.

Table 4 Sample decision table used in this study.

M	a_{max} (g)	Avg depth (m)	Depth GWT (m)	$(N_1)_{60}$	FC	Sand boil?
H	L	L	H	L	L	No
H	L	H	H	H	H	No
H	L	H	L	L	H	No
H	H	H	H	H	L	No
H	H	L	H	VH	L	No
L	VH	H	VH	H	H	No

Note: H = High, L=Low, VH=Very High

Notably, soil ejecta was observed even at lower ranges of moment magnitude and peak ground acceleration, suggesting that ejecta can occur under varied seismic conditions once liquefaction is initiated. This observation aligns with recent findings by [9] on the complex nature of liquefaction-induced ground deformation. Furthermore, the presence of ejecta was associated with corrected SPT N-values ranging from 10.3 to 24.1 and high fines content (60.5–92%).

These findings suggest that excess pore pressure generation may occur even in moderately dense or fine-grained soils—challenging conventional assumptions. Similar conclusions were drawn by [5,10], who emphasized the nuanced behavior of silty soils under liquefaction conditions. These insights highlight the multifaceted nature of soil ejecta formation and underscore the importance of data-driven approaches in advancing our understanding of liquefaction phenomena. Although the 24 rules demonstrate high certainty, the model's generalizability is constrained by strength and coverage factors that generally do not exceed 33%.

Table 5 The chosen best rule set for the sand boils rule-based model.

Rules	M (magnitude)	A_{max} (maximum acceleration)	Avg depth (m) Average Depth	Depth GWT (m) Depth of Groundwater Table	$(N_1)_{60}$ Corrected SPT N Value	FC Fines Content	Soil Ejecta?
1	Low			Very High	High		No
2				Low		Low	Yes
3			Low		High		Yes
4			Low			High	Yes
5		Low		Low			Yes
6					Low	High	Yes
7		Low				Low	Yes
8					High	Low	Yes
9	Low				Low		Yes
10	Low		Low				Yes
11	Low			High			Yes
12				Very High	Low		Yes
13	Low	Low					Yes
14			High		Low		Yes
15	High			Very High			Yes
16	Low			Low			Yes
17	Low					Very High	Yes
18		Low				Very High	Yes
19			Very High				Yes
20		Low		Very High			Yes
21	Low					Low	Yes
22			High			Very High	Yes
23				Very High		Low	Yes
24		High			Low	Low	Yes

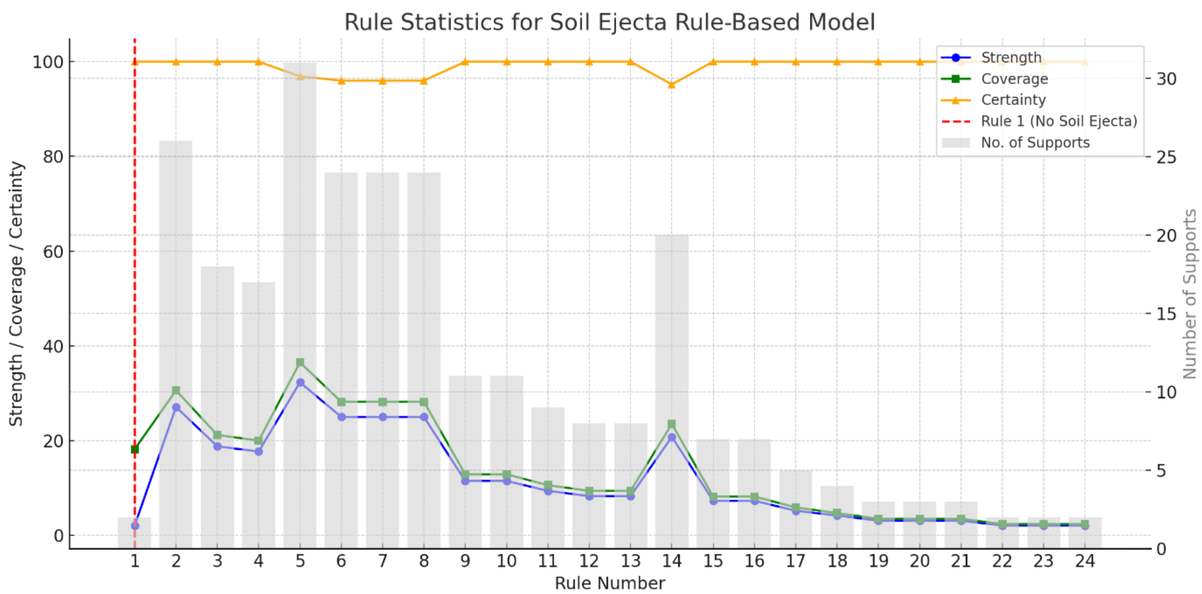

Fig. 3 Validation rule statistics for the generated rules.

This limitation is likely due to the relatively small dataset (96 cases). Despite this, Rule 5 emerged as the most influential, exhibiting the highest coverage. It can be expressed as follows:

"If the maximum acceleration is low and the groundwater table is shallow, then soil ejecta is likely to occur."

This rule indicates that even with peak ground acceleration below 0.28g, the occurrence of soil ejecta remains probable when the groundwater table is shallow (approximately 1.7 meters). This finding highlights the critical role of groundwater conditions in soil ejecta formation, even under moderate seismic loading [2,11].

The sensitivity analysis of the soil ejecta rule-based model demonstrates its strong predictive capability, with baseline accuracies of 77.8% (10-fold cross-validation) and 78.0% (50-fold). These values are within the published efficiency of other state-of-practice liquefaction models [12]. Among the parameters analyzed, corrected penetration resistance emerged as the most critical, as its removal resulted in the largest drop in accuracy (62.2% and 60.0%, respectively), highlighting the importance of soil density and stiffness in resisting liquefaction. Groundwater table depth was the second most influential factor, with its removal decreasing accuracy to 65.6% and 66.0%, affirming that high water tables increase susceptibility to liquefaction by reducing effective stress. The average depth of the critical layer also significantly impacted model performance, aligning with observations that shallow liquefiable layers facilitate the upward movement of fluidized soil. Notably, fines content showed a strong influence, reflecting its role in both liquefaction potential and soil erodibility.

The parameter interaction map highlights interactions spanning various levels of importance and strength. The most significant interaction occurs between moment magnitude and groundwater table depth, emphasizing their joint role in enhancing liquefaction potential and driving soil ejecta formation.

Prolonged shaking enables pore pressure build-up, while shallow groundwater conditions reduce effective stress, promoting upward movement of liquefied soil. These conditions also contribute to uneven ground settlements and extensive ejecta in stratified subsurface profiles. Moderate interactions—such as those between corrected SPT N-values and fines content, and between maximum acceleration and fines content—underscore the combined influence of soil density, composition, and seismic energy. The model's interaction patterns mirror those of the other published RSML-based models [4,11], indicating that the same parameters influencing liquefaction triggering also govern post-liquefaction surface manifestations. This continuity highlights the importance of interpretable models in uncovering the physical dependencies behind complex geotechnical behavior.

Figure 5 scenario map categorizes the probability of liquefaction-induced soil ejecta based on key geotechnical parameters. It identifies three conditions: high ejecta probability, low ejecta probability, and silty soil ejecta, each associated with specific groundwater table depth (GWT), corrected SPT N-value, $(N_1)_{60}$, and soil composition. The high ejecta probability zone is characterized by a shallow GWT, low $(N_1)_{60}$, and shallow soil depth, leading to consistent sand boils. Conversely, the low ejecta probability zone is linked to deep groundwater, high $(N_1)_{60}$, and lower seismic shaking magnitude (M), making sand boils unlikely. Additionally, silty soil ejecta is highlighted as a separate case, where high fines content combined with a shallow GWT and low $(N_1)_{60}$ may still lead to possible sand boils, though with different material behavior. This aligns with recent observations of recent study [13], who documented liquefied soil ejecta in fine-grained soils, including some with plasticity indices higher than 30%.

Table 6 Soil ejecta sensitivity analysis.

Attribute Removed	Total Accuracy (k=10 folds)	Total Accuracy (k=50 folds)
No Attribute Removed	77.8	78.0
Moment Magnitude	70.0	70.0
Maximum Acceleration	75.6	76.0
Average Depth of Critical Layer	70.0	72.0
Depth of Groundwater Table	65.6	66.0
Corrected Penetration Resistance	62.2	60.0
Fines Content	64.9	62.0

5. CONCLUSIONS

This study developed and evaluated a rule-based model for liquefaction-induced soil ejecta using RSML, offering an interpretable, data-driven framework for understanding soil behavior under seismic loading. By analyzing 96 case histories and generating IF-THEN rules, the model identified key parameters influencing soil ejecta, including

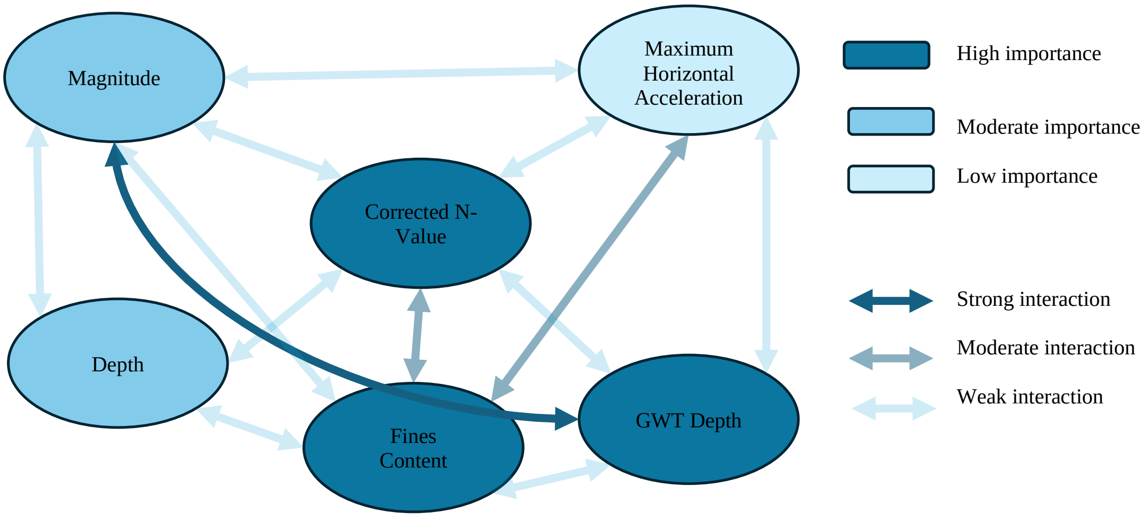

Fig. 4 Parameter interaction map for liquefaction-induced soil ejecta.

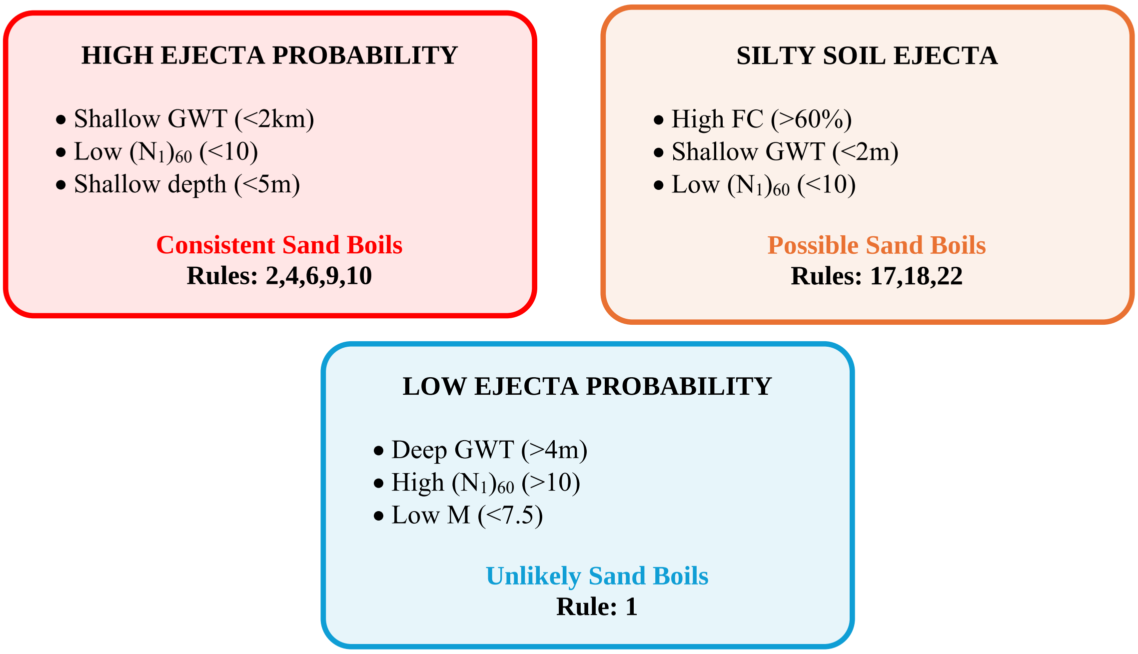

Fig. 5 Scenario map for liquefaction-induced soil ejecta.

corrected SPT N-value, groundwater table depth, fines content, and an average depth of the liquefiable layer.

Sensitivity and interaction analyses revealed that ejecta occurrence is not solely dependent on strong seismic events but also significantly influenced by shallow groundwater conditions and soil composition, especially in silty soils.

The model achieved a high accuracy of 96.4% and a coverage of 87.5%, demonstrating its robustness despite the dataset's class imbalance. Although generalizability is limited by the small sample size, the model's findings align with field observations and existing literature. The interaction map further highlighted complex interdependencies between parameters, emphasizing the need for multi-parameter analysis in liquefaction studies.

Overall, this study contributes to the current understanding of soil ejecta that can be used for hazard assessment, risk reduction, and resilience planning. Future research should expand the dataset and explore hybrid models to further refine predictive capabilities and support infrastructure safety in earthquake-prone regions.

6. ACKNOWLEDGMENTS

The authors would like to express their deepest gratitude to DOST-ERDT for funding this research, and to the DLSU Geotechnical Engineering Division Faculty for their support.

7. REFERENCES

[1] Latifi F.E., and Baba K., Predicting Liquefaction Susceptibility in North-East Morocco: Comparative Analysis of Semi-empirical Methods and UBC3D-PLM Model, Civil Engineering and Architecture, Vol. 12, Issue 3, 2024, pp.1474–1489.
[2] Kramer S.L., Chapter 9, Geotechnical Earthquake Engineering, Prentice Hall, Inc., 1996, pp.348–417.
[3] Kim H.S., Kim M., Baise L.G., and Kim B., Local and Regional Evaluation of Liquefaction Potential Index and Liquefaction Severity Number for Liquefaction-Induced Sand Boils in Pohang, South Korea, Soil Dynamics and Earthquake Engineering, Vol. 141, 2021, pp.1–16.
[4] Torres E.S., and Dungca J.R., An Interpretable Machine Learning Approach in Understanding Lateral Spreading Case Histories, International Journal of GEOMATE, Vol. 26, Issue 116, 2024, pp.110–117.
[5] Torres E., and Dungca J., Prediction of Soil Liquefaction Triggering Using Rule-Based Interpretable Machine Learning, Geosciences, Vol. 14, Issue 6, 2024, 23.
[6] Torres E., and Dungca J., Interpretable AI for Site-Adaptive Soil Liquefaction Assessment, Geosciences, Vol. 16, Issue 1, 2026, 25.
[7] Boulanger R.W., and Idriss I.M., CPT and SPT Based Liquefaction Triggering Procedures, UCD/CGM-14/01, Center for Geotechnical Modeling, University of California, Davis, 2014, pp.1–134.
[8] Cetin K.O., Seed R.B., Kayen R.E., Moss R.E.S., Bilge H.T., Ilgac M., and Chowdhury K., Dataset on SPT-Based Seismic Soil Liquefaction, Data in Brief, Vol. 20, 2018, pp.544–548.
[9] Bazan J.G., and Szczuka M., The Rough Set Exploration System, in Transactions on Rough Sets III, Peters J. F. and Skowron A. (Eds.), Springer, Berlin, Vol. 3400, 2005, pp.37–56.
[10] Pawlak Z., Rough Sets, International Journal of Computer & Information Sciences, Vol. 11, Issue 5, 1982, pp.341–356.
[11] Ntritsos N., and Cubrinovski M., Ground-Motion Effects on Liquefaction Response, Soil Dynamics and Earthquake Engineering, Vol. 177, 2024, pp.1–15.
[12] Maurer B.W., and Sanger M.D., Why “AI” Models for Predicting Soil Liquefaction Have Been Ignored, Plus Some That Shouldn’t Be, Earthquake Spectra, Vol. 39, Issue 3, 2023, pp.1883–1910.
[13] Cetin, K. O., Soylemez, B., Guzel, H., & Cakir, E. Soil liquefaction sites following the February 6, 2023, Kahramanmaraş-Türkiye earthquake sequence, Bulletin of Earthquake Engineering, Vol. 23, Issue 3, 2025, 921-944.