

ARTICLE

AI-Driven Cone Penetration Test-Based Soil Liquefaction Assessment with Incomplete Field Data

Cheng-Yu Ku^{1,2}, Ting-Yuan Wu¹ and Chih-Yu Liu^{1,2,*}

¹Department of Harbor and River Engineering, National Taiwan Ocean University, Keelung, Taiwan

²Center of Excellence for Ocean Engineering, National Taiwan Ocean University, Keelung, Taiwan

*Corresponding Author: Chih-Yu Liu. Email: 20452003@email.ntou.edu.tw

Received: 22 June 2026; Accepted: 15 September 2026; Published: 21 October 2026

ABSTRACT: Assessment of soil liquefaction susceptibility is essential for the seismic design of geotechnical systems. Conventional cone penetration test (CPT)-based procedures generally require complete input data, which may not always be available in engineering practice. To address this limitation, a hybrid artificial intelligence (AI) framework combining synthetic knowledge and field observations was developed for liquefaction assessment under incomplete-data conditions. A synthetic database of 279,936 records was generated using parameter discretization and probability-based sampling to represent diverse geotechnical and seismic conditions. Four machine-learning algorithms, including artificial neural network (ANN), random forest (RF), support vector regression (SVR), and decision tree (DT), were evaluated, and the ANN achieved the best performance, with a coefficient of determination (R^2) of 0.99. A separate classification model was trained using 227 documented field cases labeled as liquefaction or nonliquefaction. Variable importance was evaluated using out-of-bag (OOB) analysis and SHapley Additive exPlanations (SHAP). Based on the resulting importance rankings, 25,920 additional input combinations were generated using importance-guided statistical sampling, with finer discretization for more influential variables and coarser discretization for less influential variables. These records were combined with the original synthetic dataset to form an integrated database of 305,856 records for model retraining. Among the four integrated models, the ANN again showed the best overall performance and was selected as the final model, achieving R^2 of 0.96, root mean square error (RMSE) of 8.65×10^{-3} , mean absolute error (MAE) of 7.92×10^{-3} , Nash–Sutcliffe efficiency (NSE) of 0.96, performance index (PI) of 1.93, RMSE–standard deviation ratio (RSR) of 0.15, and Willmott index (WI) of 0.95. The final model was further evaluated using an independent dataset of 87 field cases, including 55 liquefaction and 32 nonliquefaction observations, and correctly classified all cases in this specific dataset. Variable-importance analysis indicated that earthquake magnitude and peak ground acceleration had the greatest influence on model predictions, while Monte Carlo analysis showed stable predictive behavior with limited variation. The proposed framework integrates established CPT-based liquefaction methodology with field-informed machine learning and provides an efficient approach for liquefaction assessment when complete site information is unavailable.

KEYWORDS: Soil liquefaction; cone penetration test; artificial intelligence; machine learning; geotechnical earthquake engineering

1 Introduction

Evaluating soil liquefaction potential is a key task in geotechnical and geological engineering, as it supports the analysis of soil response and stability during seismic events [1–3]. In earthquake prone regions, loose to medium dense saturated sands and silty sands may experience a rapid buildup of excess pore water

pressure during cyclic shaking, leading to a significant reduction in effective stress and, ultimately, a loss of shear strength and bearing capacity [4]. The mechanical response of soils under such conditions is highly nonlinear and influenced by factors such as soil fabric, stress history, groundwater conditions, and loading characteristics [5,6]. As a result, the prediction of liquefaction behavior cannot rely solely on theoretical formulations, highlighting the importance of empirical and semi empirical evaluation methods that translate field observations into practical engineering indices [7–9].

Since the 1970s, a variety of soil liquefaction assessment procedures have been developed, primarily based on *in situ* testing methods such as the cone penetration test (CPT) and standard penetration test (SPT) [10]. Among these, the soil liquefaction evaluation method proposed by Robertson and Wride has gained widespread acceptance due to its clear conceptual basis, practical applicability, and direct capability to estimate the factor of safety (*FS*) against liquefaction [11,12]. This approach integrates CPT based indicators, such as tip resistance and sleeve friction characteristics, with earthquake induced cyclic loading parameters and soil resistance measures to evaluate the susceptibility of soil deposits to liquefaction [13].

Since its introduction, the Robertson and Wride method has undergone continuous refinement and validation through case histories and field observations, and it has been extensively applied in site characterization, liquefaction hazard assessment, and seismic design practice [14]. One of its key advantages lies in its ability to condense complex soil behavior into a continuous safety factor, allowing engineers to evaluate liquefaction susceptibility using a consistent and transparent criterion [15,16]. However, the traditional application of this method relies heavily on comprehensive and high quality CPT data. In practical engineering projects, field investigations are often constrained by time, cost, and site accessibility, leading to incomplete datasets, missing parameters, or increased subjectivity in data interpretation [17]. Under such conditions, the reliability of liquefaction safety factor estimates may be reduced, thereby increasing uncertainty in engineering decision making [18].

Although the Robertson and Wride procedure provides a well established and reliable framework for liquefaction assessment, its practical application depends on the availability of a complete set of input variables and requires multiple sequential calculations. In practical engineering projects, incomplete CPT records and missing site information are frequently encountered, limiting the direct application of the conventional procedure. The objective of the present study is therefore not to replace the Robertson and Wride methodology, but to improve its applicability by integrating machine learning with both synthetic engineering knowledge and observed field behavior.

Emerging artificial intelligence (AI) and machine learning (ML) based methodologies have become valuable tools for supporting conventional liquefaction analyses. Their ability to manage complex datasets, model nonlinear system behavior, and provide reliable prediction results has contributed to their growing adoption in geotechnical applications [19]. Existing research has shown that machine learning approaches can minimize subjective judgment, promote greater analytical consistency, and improve prediction accuracy in geotechnical engineering applications. The integration of conventional liquefaction procedures with field observation based learning may further improve model applicability when available field information is limited or incomplete. Through data driven learning, ML algorithms can estimate unavailable information and facilitate more reliable prediction of liquefaction safety behavior [20].

Recent studies have further demonstrated the potential of advanced machine learning techniques for liquefaction related prediction problems. For example, Ziaie et al. (2026) [21] developed a hybrid Gradient Boosted Regression Tree (GBRT) model optimized using the Earthworm Optimization Algorithm (EOA) for predicting liquefaction induced lateral displacements. Their results showed that integrating machine learning algorithms with metaheuristic optimization can effectively capture the nonlinear relationships among geotechnical and seismic parameters and improve predictive performance. However, their study

focused on predicting post liquefaction lateral displacement, whereas the present study aims to perform CPT based liquefaction classification under incomplete field data conditions by integrating synthetic engineering knowledge with documented field observations [21].

Motivated by these considerations, this study aims to develop and evaluate a hybrid ML framework that combines the Robertson and Wride methodology with observed field liquefaction behavior. A synthetic database containing 279,936 records is first used to develop a model representing the Robertson and Wride based liquefaction response. A separate classification model is then trained using 227 documented field cases labeled directly according to observed liquefaction and nonliquefaction behavior. Variable importance derived from out-of-bag (OOB) and SHapley Additive exPlanations (SHAP) analyses of the synthetic database is subsequently used to guide the generation of an additional 25,920 statistically distributed input combinations, with finer discretization assigned to more influential variables and coarser discretization to less influential variables. The field trained model is used to generate the corresponding outputs for these 25,920 records, whereas the synthetic trained model provides the outputs for the original 279,936 records. The two model derived datasets are then merged to form an integrated database of 305,856 records, which is used to retrain the ANN, RF, SVR, and DT models. The four algorithms are subsequently retrained using the integrated database. ANN, which was identified as the principal model based on the comparative evaluation in Section 4.2, retained the strongest overall performance after integrated training and was adopted as the final prediction model. This procedure enables conventional engineering knowledge and field informed behavior to be incorporated within a single final prediction model.

2 Robertson and Wride Soil Liquefaction Evaluation Method

The Robertson and Wride soil liquefaction evaluation method is an empirical, CPT based framework widely used to assess liquefaction potential and quantify the factor of safety against liquefaction under seismic loading. Developed from extensive field case histories and CPT databases, the method systematically translates *in situ* soil response into a reproducible engineering index that reflects the combined effects of soil state, effective stress conditions, groundwater level, and earthquake loading characteristics. Owing to its transparent formulation, practical applicability, and strong empirical basis, the method has been widely adopted in geotechnical earthquake engineering practice. Compared with traditional SPT based approaches, the CPT based framework offers improved resolution of soil behavior, reduced operator dependency, and continuous profiling of liquefaction susceptibility with depth, making it a standard tool for site characterization and seismic design. Fig. 1 presents the overall calculation workflow of the Robertson and Wride method, linking CPT measurements and seismic loading parameters to a quantitative liquefaction safety factor. The following subsections describe the definitions and engineering significance of the input parameters used in this framework.

The Robertson and Wride soil liquefaction evaluation method assesses liquefaction resistance through a series of sequential calculations that integrate seismic demand and soil resistance. The procedure begins with the estimation of the cyclic stress ratio (CSR), which represents earthquake induced shear stress demand:

$$CSR = 0.65 \frac{A_{\max}}{g} \frac{\sigma_v}{\sigma'_v} r_d, r_d = \frac{(1 - 0.4113z^{0.5} + 0.04052z + 0.001753z^{1.5})}{(1 - 0.4177z^{0.5} + 0.05729z - 0.006205z^{1.5} + 0.00121z^2)}, \quad (1)$$

where $A_{\max}$ is the peak horizontal ground acceleration, g is gravitational acceleration, σ_v and σ'_v are the total and effective stresses, respectively, and r_d is the depth dependent stress reduction factor. The stress reduction factor is computed as a function of depth to account for soil deformability and wave propagation effects. Soil behavior is subsequently characterized using CPT derived parameters, including cone tip resistance q_c and sleeve friction f_s . These parameters are used to compute the soil behavior type index I_c , defined as:

$$I_c = \left[\left(3.47 - \log \left(\frac{q_c - \sigma_v}{\sigma_v'} \right) \right)^2 + \left(\log \left(\frac{f_s}{q_c - \sigma_v} \times 100\% \right) + 1.22 \right)^2 \right]^{0.5} \quad (2)$$

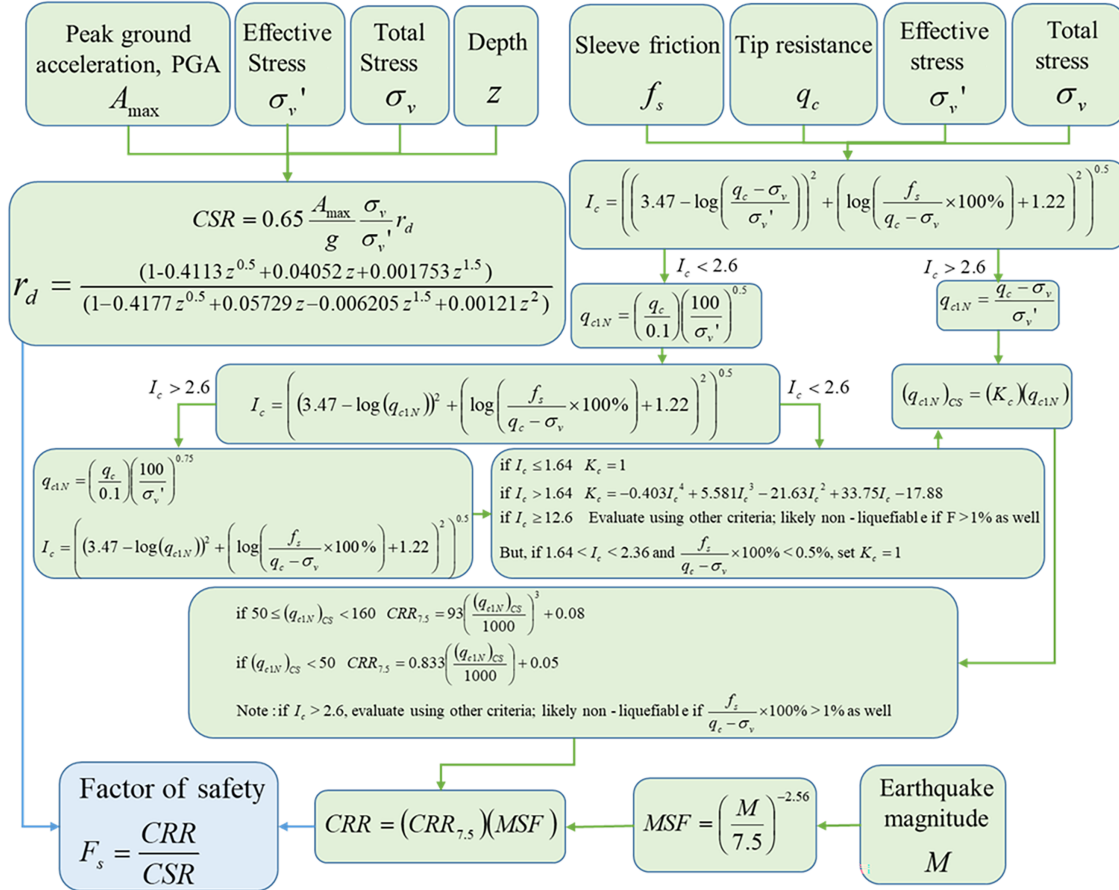


Figure 1: Flowchart of the Robertson and Wride soil liquefaction evaluation method [7].

The value of I_c is applied to classify whether the response is sand like or clay like and to determine the appropriate stress normalization scheme for cone resistance. For sand like soils ($I_c < 2.6$), the normalized cone resistance is calculated as:

$$q_{c1N} = \left(\frac{q_c}{0.1} \right) \left(\frac{100}{\sigma_v'} \right)^{0.5} \quad (3)$$

The soil behavior index is then recalculated using the normalized resistance to ensure consistency between soil classification and stress normalization. If the recalculated I_c exceeds 2.6, an alternative normalization with an exponent of 0.75 is adopted, following the established Robertson and Wride procedure. To incorporate soil behavior effects in evaluating liquefaction resistance, the stress normalized cone resistance is adjusted using the clean sand correction factor K_c , producing an equivalent clean sand value:

$$(q_{c1N})_{CS} = K_c q_{c1N} \quad (4)$$

The correction factor K_c is defined based on the soil behavior type index I_c as follows:

$$\begin{aligned} &\text{if } I_c \leq 1.64 \quad K_c = 1 \\ &\text{if } I_c \geq 1.64 \quad K_c = -0.403I_c^4 + 5.581I_c^3 - 21.63I_c^2 + 33.75I_c - 17.88 \\ &\text{if } I_c \geq 2.6 \quad \text{Evaluate using other criteria; likely non-liquefiable if } F > 1\% \text{ as well,} \\ &\text{but, if } 1.64 < I_c < 2.36 \text{ and } \frac{f_s}{q_c - \sigma_v} \times 100\% < 0.5\%, \text{ set } K_c = 1 \end{aligned} \quad (5)$$

The cyclic resistance ratio for a reference earthquake magnitude of 7.5, cyclic resistance ratio $(CRR)_{7.5}$, is then estimated empirically from $(q_{c1N})_{cs}$:

$$CRR_{7.5} = \begin{cases} 0.93 \left(\frac{(q_{c1N})_{cs}}{1000} \right)^3 + 0.08, & 50 \leq (q_{c1N})_{cs} < 160 \\ 0.833 \left(\frac{(q_{c1N})_{cs}}{1000} \right) + 0.05, & (q_{c1N})_{cs} < 50 \end{cases} \quad (6)$$

For earthquakes with magnitudes other than 7.5, the cyclic resistance ratio is adjusted using a magnitude scaling factor (MSF):

$$MSF = \left(\frac{M}{7.5} \right)^{-2.56}, \quad (7)$$

$$CRR = CRR_{7.5} \times MSF \quad (8)$$

Finally, the factor of safety against liquefaction is computed as:

$$F_s = \frac{CRR}{CSR} \quad (9)$$

Depth is a fundamental parameter in soil liquefaction evaluation, as it controls both the *in situ* stress state and the distribution of earthquake induced shear stress. With increasing depth, overburden pressure raises total and effective stress, generally enhancing liquefaction resistance, while seismic shear stress attenuates due to soil deformability through the stress reduction factor r_d . Consequently, depth directly influences the computation of cyclic stress ratio (CSR) and the normalization of CPT resistance in the Robertson and Wride framework [22].

Total stress σ_v represents the weight of the overlying soil column and defines the overall confining stress acting on the soil. It appears explicitly in the CSR formulation and governs the magnitude of cyclic shear stress induced during seismic loading. Accurate estimation of σ_v is therefore essential for reliable liquefaction demand assessment [23].

Effective stress σ'_v , obtained by subtracting pore water pressure from total stress, governs soil strength and deformation behavior. Liquefaction is fundamentally associated with reductions in σ'_v caused by excess pore pressure generation during cyclic loading. In the Robertson and Wride method, σ'_v is a key variable for both CSR calculation and stress normalization of cone resistance, with lower effective stress generally indicating higher liquefaction susceptibility [24].

Cone tip resistance q_c , obtained from CPT, provides an *in situ* indication of soil stiffness and strength. Higher q_c values typically reflect denser, stronger soils with greater resistance to liquefaction, whereas lower values are associated with loose, contractive soils. To remove stress level effects, q_c is normalized to obtain the stress normalized cone resistance q_{c1N} , which serves as the primary indicator of cyclic resistance in the Robertson and Wride method [25].

Sleeve friction f_s reflects shear resistance mobilized along the CPT sleeve and provides insight into soil texture and fines content. Combined with tip resistance, it is used to compute the friction ratio and the soil behavior type index I_c , which distinguishes sand like from clay like behavior and determines the applicability of CPT based liquefaction correlations. Higher friction ratios are commonly associated with cohesive or silty soils that are less prone to liquefaction [26].

Peak ground acceleration A_{max} represents the intensity of seismic shaking and directly controls the level of cyclic shear stress imposed on the soil. It appears explicitly in the CSR equation, with higher peak ground acceleration (PGA) values corresponding to increased seismic demand and greater liquefaction potential for a given soil condition [13].

Earthquake magnitude M governs the duration and cumulative effects of cyclic loading. Larger magnitude earthquakes induce a greater number of loading cycles, promoting pore pressure accumulation. This effect is accounted for through the MSF , which adjusts cyclic resistance from a reference magnitude of 7.5 to the target event magnitude, ensuring consistent consideration of shaking duration in liquefaction evaluation [27].

The factor of safety against liquefaction F_s is the final output of the Robertson and Wride method and provides a quantitative measure of liquefaction susceptibility. Values less than unity indicate a high likelihood of liquefaction, whereas values greater than unity suggest stable behavior under the specified seismic conditions. By integrating soil properties, stress conditions, and earthquake characteristics, F_s offers a practical and transparent basis for liquefaction hazard assessment and engineering decision making [28].

2.1 Synthetic Dataset Development

To represent the variability associated with the principal variables used in the Robertson and Wride liquefaction assessment procedure, a synthetic database was generated incorporating depth, CPT derived indicators, stress related parameters, and seismic loading characteristics. The dataset was designed to encompass diverse geotechnical and seismic scenarios relevant to liquefaction evaluation. Given the limited availability of field datasets containing both reliable CPT measurements and corresponding seismic records across different geological environments, the use of synthetic data offers an effective means of investigating variable dependencies and assessing predictive model behavior in a controlled framework. To maintain engineering realism and physical plausibility, the parameter ranges were established according to the governing equations of the Robertson and Wride methodology, supplemented by findings reported in previous studies and documented liquefaction case histories. The selected parameter bounds were intended to reflect representative liquefaction susceptible conditions, and the corresponding ranges are presented in Table 1.

Table 1: Summary of the input and output variables used in this study.

No.	Parameter		Unit	Range	Source
1	Input	Field data	Depth (z)	m	0–20
2			Total stress (σ_v)	kPa	0–760
3			Effective stress (σ'_v)	kPa	0–480
4			Tip resistance (q_c)	kPa	0–50,000
5			Sleeve friction (f_s)	kPa	0–1000
6		Seismic loading data	Peak ground acceleration (A_{max})	g	0–0.4
7			Earthquake magnitude (M)	NA	5–9
8	Output	NA	Factor of safety (F_s)	NA	0–10

A synthetic database was generated through random sampling procedures based on predefined probability distributions to represent diverse geotechnical and seismic scenarios relevant to the Robertson and Wride liquefaction assessment methodology. In addition to CPT based measurements, parameters associated with stress conditions and earthquake loading were incorporated to establish a comprehensive input set for estimating the liquefaction factor of safety. Two complementary sampling schemes were adopted for dataset generation, namely uniform interval sampling and probability based stochastic sampling. Both schemes utilized the variables listed in Table 1, with parameter limits and discretization intervals selected to adequately represent the diversity and natural variation of liquefaction related geotechnical conditions.

2.2 Uniform Interval Based Parameter Sampling

Under the uniform discretization scheme, CPT related data were generated by allocating parameter values at regular intervals within physically reasonable bounds defined from previous liquefaction studies and engineering practice. The selected variables included depth (z), total stress (σ_v), effective stress (σ'_v), cone tip resistance (q_c), sleeve friction (f_s), peak ground acceleration (A_{max}), and earthquake magnitude (M). The parameter limits adopted in this study are presented in Table 1. The selected ranges were intended to encompass a broad spectrum of geotechnical and seismic environments. Cone tip resistance values covered conditions from weak fine grained deposits to dense granular soils, whereas sleeve friction values were chosen to reflect variations in soil composition and mechanical behavior. Likewise, total and effective stress levels were defined to represent realistic subsurface stress states encountered at different depths. Seismic loading conditions were represented through varying levels of peak ground acceleration and earthquake magnitude, thereby accounting for differences in shaking intensity and duration. For each parameter combination, the cyclic stress ratio (CSR), cyclic resistance ratio (CRR), and corresponding liquefaction factor of safety (F_s) were computed following the Robertson and Wride evaluation procedure. The resulting sampling framework generated a large and systematically distributed dataset encompassing both liquefaction susceptible and nonliquefaction conditions. This database provides a physically representative domain of input and output variables suitable for the development, training, and validation of machine learning models for liquefaction assessment.

2.3 Statistically Distributed Parameter Sampling

A probabilistic sampling strategy was employed to generate a database encompassing the input variables required for liquefaction evaluation using the Robertson and Wride procedure. The generated dataset was intended to reflect realistic variations in CPT measurements, *in situ* stress conditions, and earthquake

loading characteristics associated with liquefaction susceptible soils. Uniform probability distributions were assigned to depth, total and effective overburden stress, peak ground acceleration, and earthquake magnitude to ensure adequate representation of feasible geotechnical and seismic conditions. Conversely, cone tip resistance and sleeve friction were sampled from lognormal distributions to reproduce the right skewed distributions commonly reported in CPT observations, where moderate values dominate and extreme values occur less frequently. The liquefaction factor of safety was subsequently determined for each parameter combination using the Robertson and Wride methodology. Because both cyclic stress demand and cyclic resistance are governed by complex interactions among multiple variables, the computed safety factors exhibited substantial variability, ranging from highly susceptible to stable soil conditions. The resulting database contained 279,936 records, providing comprehensive coverage of the parameter domain for machine learning model development and performance evaluation. Fig. 2 presents the frequency distributions of the input variables and calculated safety factors. Variables generated from uniform distributions exhibited relatively even coverage across their prescribed domains, whereas cone tip resistance and sleeve friction showed positively skewed distributions consistent with field observations. The distribution of F_s was notably nonuniform, with values concentrated near the threshold of liquefaction and progressively fewer observations at larger safety factors. This pattern reflects the interacting effects of soil characteristics, stress states, and seismic loading conditions. Overall, the generated database successfully captures the variability and nonlinear behavior observed in documented liquefaction case records, thereby providing a physically realistic and statistically representative basis for subsequent machine learning based liquefaction analyses.

The statistical characteristics of the generated variables are further summarized in Fig. 2. Depth is approximately uniformly distributed from 0 to 20 m, with approximately 13,330 to 13,331 records at each sampling level. Its mean and median are approximately 10 m, with a standard deviation of 6.33 m. Total stress is similarly distributed from 0 to 760 kPa, with a mean and median of approximately 380 kPa and a standard deviation of 240.45 kPa. Effective stress ranges from 0 to 480 kPa, with a mean and median of approximately 240 kPa and a standard deviation of approximately 145.33 kPa. These three variables exhibit nearly uniform distributions and therefore do not show distinct modes. In contrast, cone tip resistance shows a nonuniform distribution, with 24,546 records between 0 and 10,000 kPa, 110,045 records between 12,500 and 25,000 kPa, 91,233 records between 27,500 and 37,500 kPa, and 54,112 records between 40,000 and 50,000 kPa. Its mean is approximately 27,372 kPa, median is 26,152 kPa, standard deviation is 5367.48 kPa, and modal value is approximately 25,000 kPa. Sleeve friction follows a similar distribution, with 24,546 records between 0 and 200 kPa, 110,045 records between 250 and 500 kPa, 91,233 records between 550 and 750 kPa, and 54,112 records between 800 and 1000 kPa. Its mean is approximately 547 kPa, median is 516 kPa, standard deviation is 107.27 kPa, and modal value is approximately 500 kPa. Peak ground acceleration ranges from 0 to 0.4 g, with approximately 55,987 to 55,988 records at each level, a mean and median of approximately 0.20 g, and a standard deviation of 0.13 g. Earthquake magnitude ranges from 5 to 9, with a mean and median of approximately 7 and a standard deviation of 1.29. Factor of safety ranges from 0 to 10, with a mean and median of approximately 5 and a standard deviation of 3.16. Peak ground acceleration, earthquake magnitude, and factor of safety are approximately uniformly distributed and therefore do not exhibit meaningful distinct modes.

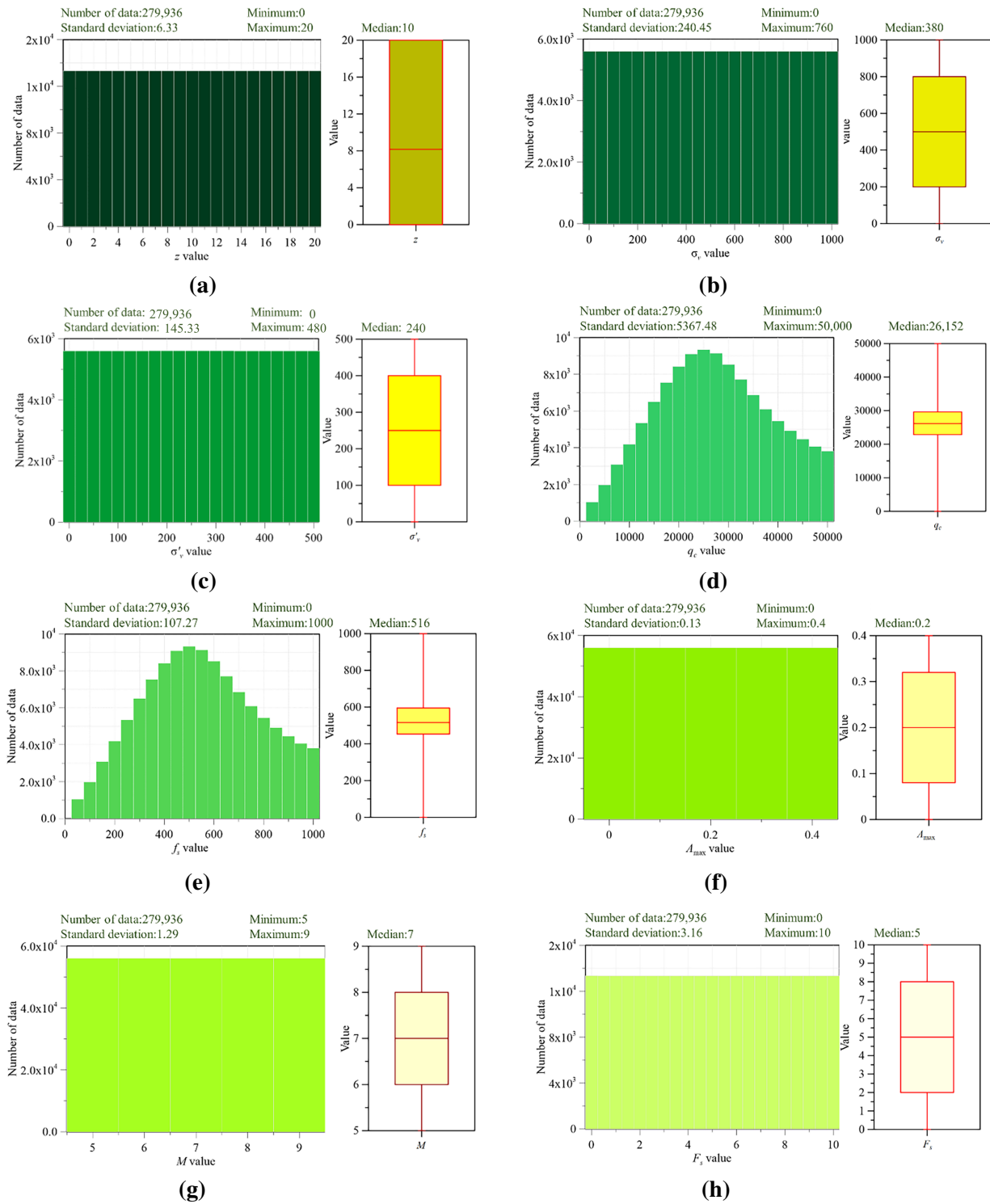


Figure 2: Probability distributions of CPT related input variables adopted in the Robertson and Wride liquefaction assessment framework (a) Depth (z); (b) Total stress (σ_v); (c) Effective stress (σ'_v); (d) Tip resistance (q_c); (e) Sleeve friction (f_s); (f) Peak ground acceleration (A_{max}); (g) Earthquake magnitude (M); and (h) Factor of safety (F_s).

3 AI-Enhanced Robertson and Wride Soil Liquefaction Evaluation Method

To improve the estimation of liquefaction safety within the Robertson and Wride assessment framework, four machine learning algorithms were investigated in this study. Machine learning techniques offer a flexible data driven methodology capable of representing the intricate nonlinear interactions between geotechnical characteristics and seismic loading factors that influence liquefaction response. The predictive models considered include artificial neural networks (ANN), random forests (RF), support vector regression (SVR), and decision trees (DT). The synthetic database was first used to develop models representing the Robertson and Wride based liquefaction response, while field observations were separately used to establish a classification model based on actual liquefaction behavior. Model development involved hyperparameter optimization, training and testing procedures, comparative evaluation of predictive capability, and investigation of the relative contributions of individual input variables. The out-of-bag (OOB) and SHapley Additive exPlanations (SHAP) importance results were subsequently used to guide the statistical generation of 25,920 additional input combinations. Outputs for these additional records were obtained from the field trained model and were combined with the 279,936 records generated using the synthetic trained model. The resulting integrated database of 305,856 records was then used to retrain the ANN, RF, SVR, and DT models, after which their predictive performances were compared to identify the final model. The overall research flowchart is shown in Fig. 3.

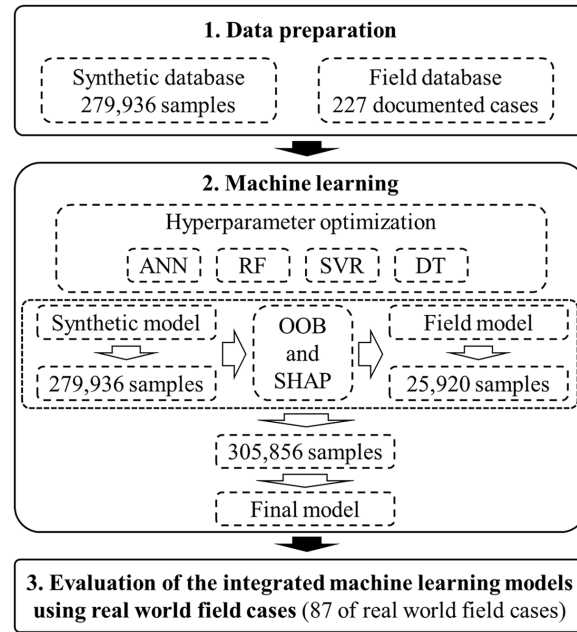


Figure 3: Research flowchart.

3.1 AI-Enhanced Models

Machine learning approaches have gained increasing attention in liquefaction assessment because of their capability to represent intricate dependencies among geotechnical and seismic variables. Within the Robertson and Wride framework, these approaches can be employed to estimate liquefaction safety more objectively while improving the accuracy of factor of safety predictions. The selected algorithms possess distinct advantages in identifying nonlinear patterns, processing multidimensional datasets, and maintaining reliable predictive capability under varying conditions. Model development was carried out using the synthetic database introduced in the preceding section, which encompasses diverse combinations of soil

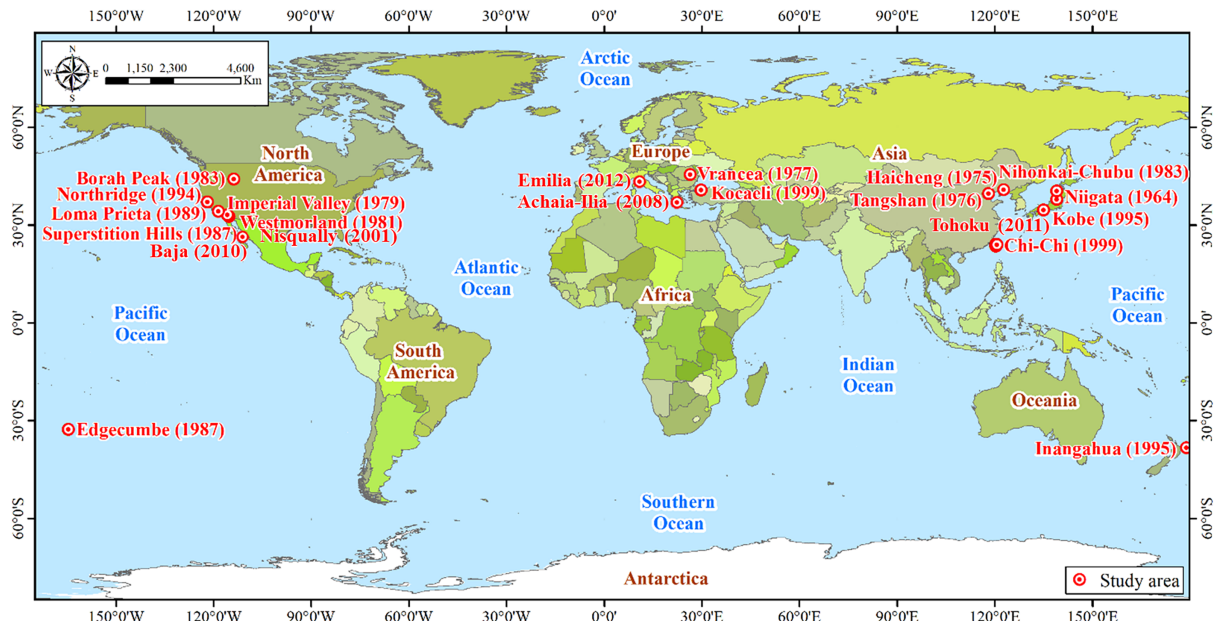
properties, stress environments, and earthquake loading characteristics. A separate classification model was also developed using documented field observations of liquefaction and nonliquefaction behavior. Rather than directly combining the predictions of the synthetic and field models at the decision level, the two trained models were subsequently used to generate complementary model based training records for construction of the final integrated ANN model. To improve model reliability and reduce the risk of overfitting, parameter optimization, validation procedures, and feature interpretation techniques were systematically incorporated into the modeling workflow.

To assess the effectiveness of the proposed AI based framework under real world conditions, 227 documented field case records were collected from previously published earthquake case studies. The dataset includes both liquefaction and nonliquefaction observations and was compiled from CPT based liquefaction studies by Juang et al. (2002), Ku et al. (2004), Moss et al. (2006), Rasanen et al. (2023), and Geyin and Maurer (2021) [31–35], which provide sufficient subsurface and seismic information for field model development. The observed field responses were directly assigned as binary target labels, with 0 representing liquefaction and 1 representing nonliquefaction. A separate field model was then trained using these records with the model configuration selected from the preceding optimization and comparison procedures. The synthetic and field trained models were subsequently used as two complementary sources for constructing the final training database. First, the OOB and SHAP results obtained from the 279,936 record synthetic database were used to determine the relative discretization density of the input variables. Variables exhibiting greater importance were represented using finer discretization, whereas variables with lower importance were represented using coarser discretization. Based on the statistical distributions adopted in this study, this importance guided sampling procedure generated 25,920 additional input combinations.

Table 2 presents the 227 field CPT records used to develop the field based classification model. The geographical distribution of the actual soil samples is shown in Fig. 4. Unlike the synthetic model, which was trained to predict the factor of safety calculated using the Robertson and Wride procedure, the field model was trained directly using observed liquefaction and nonliquefaction outcomes. The trained field model was then used to generate the corresponding output values for these 25,920 combinations, thereby transferring information learned from the 227 documented field cases to a broader input domain. Before construction of the integrated database, the outputs of the two trained models were expressed using the same binary target definition. For the synthetic trained model, the predicted factor of safety was converted to a binary class, with 0 assigned when $F_s < 1$ and 1 assigned when $F_s \geq 1$. The field trained model used the same target convention, with 0 representing liquefaction and 1 representing nonliquefaction. This common output definition enabled the records generated from the two models to be combined within a consistent target space. In parallel, the trained synthetic model was used to generate the output values corresponding to the original 279,936 synthetic input combinations. The 25,920 field informed records and 279,936 synthetic model derived records were then combined to form an integrated database containing 305,856 records. This integrated database was subsequently used to retrain the ANN, RF, SVR, and DT models using their selected configurations. The performances of the four integrated models were then evaluated for comparison. ANN, which had already been identified as the principal model based on the comparative results presented in Section 4.2, continued to exhibit the best overall predictive performance and was therefore retained as the final model. Therefore, the final ANN incorporates information from both the Robertson and Wride based synthetic component and the field observation based component through data level integration and retraining rather than through direct decision level fusion.

Table 2: Summary of 227 documented field CPT cases used for field model development.

Year	Earthquake	Liquefaction	Nonliquefaction	Total
1964	Niigata	1	1	2
1968	Inangahua	1	0	1
1975	Haicheng	1	0	1
1976	Tangshan	5	0	5
1977	Vrancea	0	1	1
1979	Imperial Valley	1	1	2
1981	Westmorland	3	1	4
1983	Borah Peak	4	0	4
1983	Nihonkai-Chubu	3	0	3
1987	Edgumbe	3	3	6
1987	Superstition Hills	1	0	1
1989	Loma Prieta	33	4	37
1994	Northridge	4	0	4
1995	Kobe	11	0	11
1999	Chi-Chi	36	11	47
1999	Kocaeli	16	0	16
2001	Nisqually	9	15	24
2008	Achaia-Ilia	1	1	2
2010	Baja	1	2	3
2011	Tohoku	4	3	7
2012	Emilia	18	28	46

**Figure 4:** Geographical distribution of actual soil samples.

To maintain a clear separation between model development and final field evaluation, an additional set of 87 documented field cases, consisting of 55 liquefaction and 32 nonliquefaction observations, was reserved exclusively for the independent evaluation presented in [Section 5.1](#). These 87 cases were not included

among the 227 field records used to develop the field based classification model. The two datasets are therefore mutually exclusive at the individual case record level, with no overlapping records. Furthermore, the observed liquefaction and nonliquefaction labels of the 87 evaluation cases were not used during development of the field model or retraining of the integrated models and were used only after model development had been completed for final performance evaluation. Although cases associated with the 1999 Chi-Chi earthquake are present in both datasets and some cases were compiled from the same published sources, no individual CPT case record was included in both datasets.

Artificial neural networks (ANNs) were selected as the principal machine learning approach because of their ability to represent complex relationships between geotechnical characteristics and seismic loading conditions that influence liquefaction response. The model architecture consisted of multiple hidden layers and nonlinear activation functions, while training settings were systematically adjusted to achieve reliable convergence and predictive stability. To improve model generalization and limit excessive fitting to the training data, early stopping and validation based monitoring were incorporated during model development. The network architecture and training parameters were systematically optimized to achieve stable predictive performance. Due to their strong capacity for learning intricate patterns from large datasets, ANNs are well suited for estimating liquefaction safety with a high degree of accuracy [36]. Fig. 5 presents the modeling framework for the synthetic database, model training, validation, and performance assessment.

The ANN model was developed in MATLAB, and a data partitioning strategy combined with an early stopping mechanism was adopted to improve model robustness and limit overfitting. The available synthetic dataset was randomly allocated into training (70%), validation (15%), and testing (15%) subsets. Model parameters were estimated using the training subset, whereas the validation subset was used to monitor predictive performance throughout the learning process. During training, model convergence was evaluated by simultaneously tracking errors from both the training and validation datasets. Training was terminated once validation performance began to deteriorate despite continued improvement in the training error, thereby preventing excessive adaptation to the training samples and promoting better predictive performance on previously unseen data. Model performance was further evaluated using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). These metrics were calculated for the training, validation, and testing subsets to examine prediction accuracy, model stability, and overall generalization capability. The predicted factor of safety was subsequently converted into a binary class, with values below 1 assigned to liquefaction and values equal to or greater than 1 assigned to nonliquefaction. Machine-learning and neurocomputing approaches have also been successfully applied to other nonlinear soil and foundation engineering problems.

Random forest (RF) was adopted as a tree based ensemble learning technique that aggregates predictions from multiple decision trees to enhance model stability and predictive reliability. The method is particularly applicable to CPT related liquefaction datasets because it can effectively capture complex interactions among input variables while accommodating heterogeneous data structures with limited data preparation requirements. In addition, RF provides an intrinsic mechanism for evaluating variable contributions, allowing the relative significance of key factors, including effective stress, cone penetration resistance, peak ground acceleration, and earthquake magnitude, to be quantified with respect to liquefaction safety estimation.

To evaluate model robustness and identify potential overfitting, the RF model was developed using a holdout validation strategy. The available dataset was randomly separated into calibration and evaluation subsets, comprising approximately 70% and 30% of the samples, respectively. Model fitting was performed using the calibration subset, whereas predictive capability was assessed using the independent evaluation subset. A substantial discrepancy between the two sets of performance metrics, such as markedly higher R^2 values or lower prediction errors for the calibration data, was interpreted as evidence of excessive model

fitting. Furthermore, RF incorporates bootstrap aggregation as its core ensemble mechanism. This procedure generates numerous decision trees from randomly sampled subsets of the data and combines their outputs to produce the final prediction. By averaging the results obtained from multiple trees, the model reduces prediction variance and improves resistance to overfitting compared with an individual decision tree.

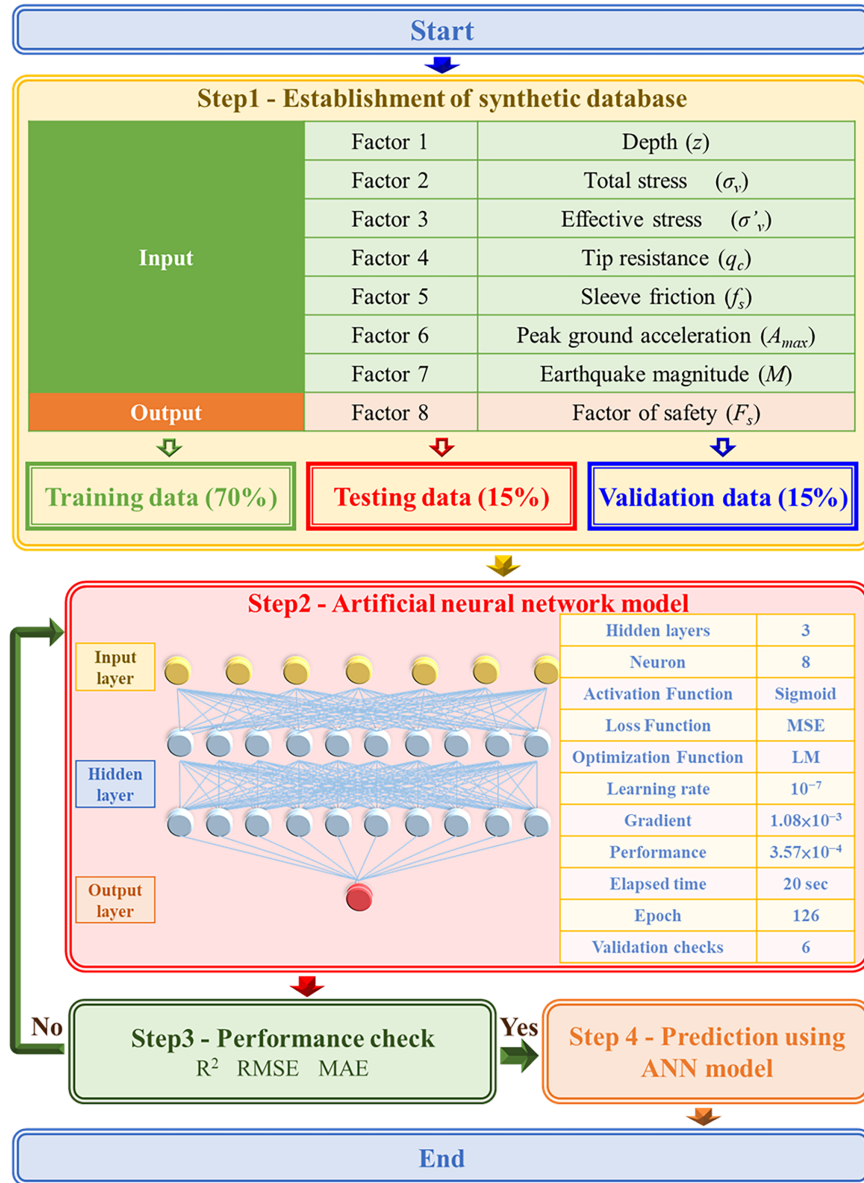


Figure 5: Flowchart of Robertson and Wride soil liquefaction evaluation method using the ANN model.

Support vector regression (SVR) was adopted to model the complex behavior of liquefaction response using a radial basis function kernel. By mapping the original input variables into a transformed feature domain, the algorithm can effectively represent nonlinear relationships among geotechnical and seismic parameters. Model calibration was performed through systematic parameter tuning combined with validation procedures to identify an appropriate set of hyperparameters. These settings contributed to improved predictive reliability and reduced sensitivity to noise and anomalous observations. Although SVR does not

inherently quantify the contribution of individual input variables, its consistent predictive capability and strong generalization characteristics make it a valuable approach for liquefaction safety assessment.

To improve model robustness and reduce overfitting, the SVR model was developed using a holdout validation strategy, with 70% of the data used for model calibration and 30% reserved for evaluation. Hyperparameters were optimized to balance predictive accuracy and model complexity, while the inherent regularization mechanism of SVR enhanced generalization performance. Model effectiveness was assessed using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE), and discrepancies between calibration and evaluation results were used to identify potential overfitting.

Decision tree (DT) models were employed due to their high interpretability. By recursively partitioning geotechnical and seismic variables, DTs can identify hierarchical relationships between input parameters and liquefaction safety predictions. Although generally less robust than ensemble methods, they provide valuable insight into model behavior and the factors influencing liquefaction assessment.

To improve model generalization, DT growth was controlled through several structural constraints, including splitting and node size criteria. Model accuracy was evaluated using R^2 , RMSE, and MAE metrics derived from both calibration and testing datasets. Differences in predictive performance between the two datasets were used to identify potential overfitting and optimize model configuration.

3.2 Performance Assessment

Model performance in this study is evaluated using multiple statistical indicators, including the mean absolute error (MAE), coefficient of determination (R^2), root mean square error (RMSE), Nash Sutcliffe efficiency (NSE), performance index (PI), RMSE to observation standard deviation ratio (RSR), and Willmott index of agreement (WI). Together, these metrics offer complementary insights into prediction accuracy and are employed to facilitate comparison among the developed ML models. The formal definitions of all seven evaluation criteria considered in this work are presented below.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (12)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (13)$$

$$PI = R^2 + \frac{VAF}{100} - RMSE, \quad VAF = \left[1 - \frac{\sum_{i=1}^n [(y_i - \hat{y}_i) - (\bar{y} - \hat{\bar{y}})]^2}{SD_{obs}^2} \right] \times 100\%, \quad (14)$$

$$RSR = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (15)$$

$$WI = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (|y_i - \bar{y}| + |\hat{y}_i - \bar{y}|)^2}, \quad (16)$$

where y denotes the observed value, $\hat{y}_i$ represents the corresponding model estimate, $\bar{y}$ is the average of the predicted values, n denotes the number of samples included in the dataset, VAF denotes the variance accounted for, and SD_{obs} denotes the standard deviation of observed values.

During the subsequent validation phase, these metrics are employed to evaluate prediction accuracy and to characterize the overall performance of the models. All performance metrics were calculated using the normalized target and predicted values.

For the binary classification assessment of the independent field cases presented in [Section 5.1](#), additional performance measures were calculated from the confusion matrix. Although liquefaction and nonliquefaction were numerically coded as 0 and 1, respectively, during model development, liquefaction was designated as the positive class for calculation of the classification metrics. True positive (TP) denotes an observed liquefaction case correctly classified as liquefaction, true negative (TN) denotes an observed nonliquefaction case correctly classified as nonliquefaction, false positive (FP) denotes an observed nonliquefaction case incorrectly classified as liquefaction, and false negative (FN) denotes an observed liquefaction case incorrectly classified as nonliquefaction. The classification metrics were calculated as follows:

$$\text{Sensitivity (Recall)} = TP / (TP + FN), \quad (17)$$

$$\text{Specificity} = TN / (TN + FP), \quad (18)$$

$$\text{Precision} = TP / (TP + FP), \quad (19)$$

$$F1 - \text{score} = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall}), \quad (20)$$

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN). \quad (21)$$

4 Validation of AI-Enhanced Models

Model assessment and integration consisted of parameter optimization, preliminary model performance evaluation, variable importance analysis, parameter sampling evaluation, importance guided field informed data generation, integrated model retraining, final ANN evaluation, and testing under complete and incomplete field data conditions. Together, these investigations were used to examine the predictive accuracy, reliability, and transferability of the ANN, RF, SVR, and DT models for liquefaction safety estimation within the Robertson and Wride framework.

4.1 Hyperparameter Optimization

Parameter optimization was performed using the synthetic database for the ANN, RF, SVR, and DT algorithms to identify the most suitable model configurations for factor of safety prediction within the Robertson and Wride framework. Different optimization strategies were adopted according to the characteristics of each machine learning algorithm. Selected structural parameters were evaluated through comparative analyses, while additional hyperparameters were determined using Bayesian optimization where appropriate. The selected parameter configurations are presented in [Fig. 6](#), and the associated performance results are listed in the corresponding tables. The resulting parameter settings were subsequently fixed and applied to the corresponding models trained using the 227 field records to maintain consistent model complexity and reduce the risk of overfitting associated with the limited field dataset. Accordingly, differences between the synthetic and field models primarily reflect differences in training information rather than model configuration.

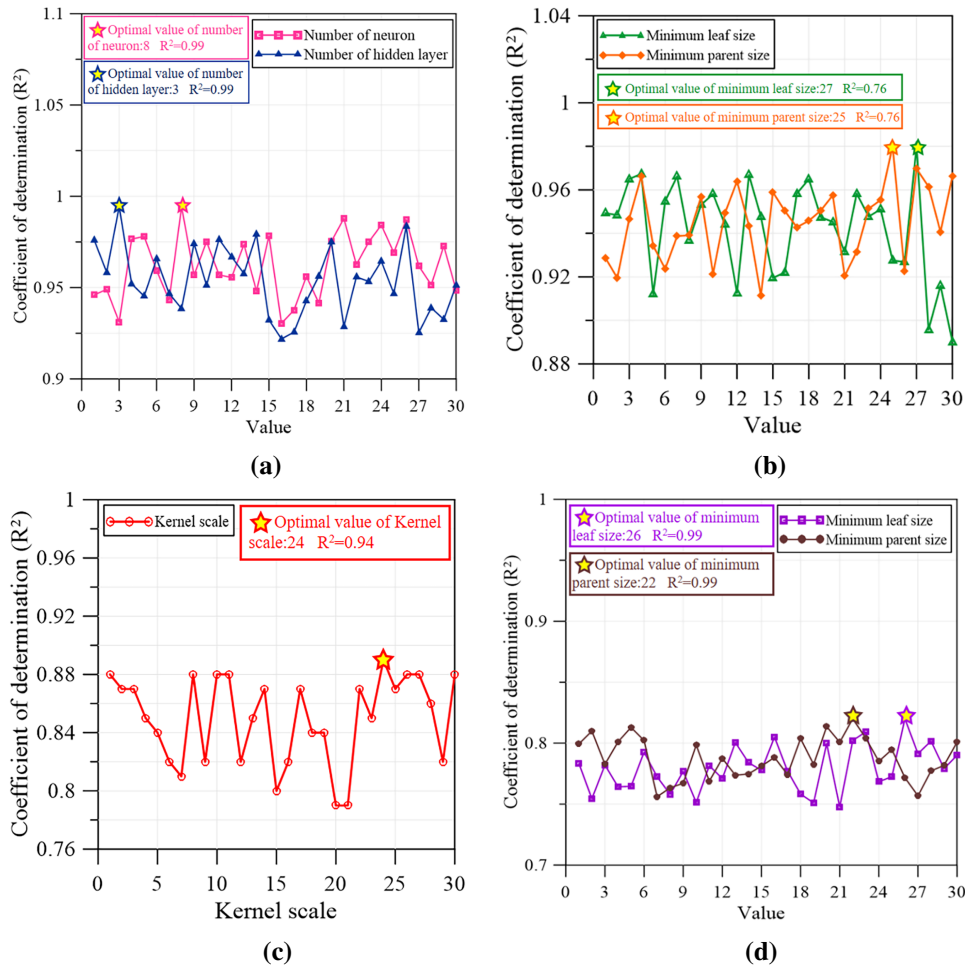


Figure 6: Hyperparameter optimization results for machine learning models: (a) ANN; (b) RF; (c) SVR; and (d) DT.

For ANN, several network configurations were evaluated by adjusting the number of hidden layers and the number of neurons in each hidden layer, as presented in Fig. 6. The highest prediction accuracy was achieved using a structure consisting of three hidden layers with eight neurons per layer, resulting in an R^2 value of 0.99. In addition to the network architecture, the effects of different activation functions were evaluated, and the corresponding performance comparison is presented in Table 3. The influence of different loss functions on ANN performance is summarized in Table 4. The performance of different optimization functions is compared in Table 5, while the effects of different data partition ratios are presented in Table 6. These analyses were used together to determine the optimized ANN configuration adopted for subsequent model development.

Table 3: Performance comparison of ANN models using different activation functions.

Activation Function	R^2 for the Training Dataset	R^2 for the Testing Dataset	R^2 for the Validation Dataset
Sigmoid	0.99	0.99	0.99
Tanh	0.96	0.95	0.95
ReLU	0.91	0.90	0.91

Table 4: Performance comparison of ANN models using different loss functions.

Loss Function	R ² for the Training Dataset	R ² for the Testing Dataset	R ² for the Validation Dataset
MSE	0.99	0.99	0.99
MAE	0.97	0.97	0.96
Huber loss	0.92	0.91	0.91

Table 5: Performance comparison of ANN models using different optimization functions.

Optimization Function	R ² for the Training Dataset	R ² for the Testing Dataset	R ² for the Validation Dataset
Levenberg Marquardt	0.99	0.99	0.99
Gradient Descent	0.95	0.95	0.93
RMSprop	0.89	0.88	0.88

Table 6: Comparison of ANN performance using different data partition ratios.

Data Partition Ratios	R ² for the Training Dataset	R ² for the Testing Dataset	R ² for the Validation Dataset
Training (60%) Testing (20%) Validation (20%)	0.95	0.94	0.94
Training (70%) Testing (15%) Validation (15%)	0.99	0.99	0.99
Training (80%) Testing (10%) Validation (10%)	0.91	0.90	0.91

For RF, a combination of comparative parameter analysis and Bayesian optimization was adopted. The minimum leaf size and minimum parent node size were systematically evaluated through the comparative analysis presented in Fig. 6. The selected values were 27 and 25, respectively. The remaining RF hyperparameters, including the number of trees, maximum tree depth or maximum number of splits, and number of predictor variables randomly selected at each split, were subsequently optimized using Bayesian optimization. This combined procedure enabled both the structural characteristics and additional model complexity parameters of RF to be considered during model calibration. The optimized RF model achieved an R² value of 0.98.

For SVR, the kernel scale was first investigated through the comparative analysis presented in Fig. 6, and the optimal kernel scale was identified as 24. A radial basis function kernel was adopted in this study. The remaining principal SVR hyperparameters, including the box constraint and epsilon insensitive margin, were optimized using Bayesian optimization. Therefore, the SVR calibration procedure considered multiple hyperparameters rather than relying solely on kernel scale optimization. The optimized SVR model achieved an R² value of 0.89.

For the DT model, parameter calibration focused on regulating the sizes of parent and terminal nodes. The best configuration produced an R² value of 0.82 with leaf and parent sizes of 26 and 22, respectively.

While DT offers a clear and interpretable modeling structure, its predictive performance was comparatively lower than that of ANN and RF.

Overall, the hyperparameter tuning procedures demonstrate that the comparison among the four machine learning models was performed using systematically calibrated configurations rather than default parameter settings. ANN hyperparameters were evaluated as shown in Fig. 6 and Tables 3–6, whereas RF and SVR combined comparative parameter analysis with Bayesian optimization for their remaining key hyperparameters.

4.2 Overall Performance of the ML Models

The performance of the four machine learning algorithms was evaluated separately using the synthetic and field datasets based on seven evaluation criteria: the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), Nash Sutcliffe efficiency (NSE), Performance Index (PI), RMSE to observation standard deviation ratio (RSR), and Willmott index of agreement (WI). These indicators were adopted to examine prediction quality and model consistency for liquefaction assessment. As summarized in Table 7, the synthetic data models were evaluated based on their ability to reproduce the factor of safety calculated using the Robertson and Wride procedure. ANN achieved the highest predictive performance, with an R^2 value of 0.99, RMSE of 2.26×10^{-3} , MAE of 1.06×10^{-3} , NSE of 0.99, PI of 1.99, RSR of 0.10 and WI of 0.99. The superior performance of ANN suggests a strong capability for representing the complex interactions among CPT measurements, stress variables, and seismic loading conditions. RF also produced highly accurate predictions, yielding an R^2 value of 0.98, RMSE of 4.15×10^{-3} , MAE of 3.72×10^{-3} , NSE of 0.98, PI of 1.98, RSR of 0.15 and WI of 0.97. The ensemble nature of RF contributed to stable and reliable performance. In contrast, SVR achieved moderate prediction accuracy ($R^2 = 0.89$, RMSE = 7.41×10^{-3} , MAE = 7.39×10^{-3} , NSE = 0.88, PI = 1.82, RSR = 0.23, WI = 0.88), whereas DT produced the least accurate results ($R^2 = 0.82$, RMSE = 9.58×10^{-3} , MAE = 8.97×10^{-3} , NSE = 0.81, PI = 1.79, RSR = 0.29, WI = 0.82). Overall, ANN and RF consistently outperformed SVR and DT, indicating that neural network and ensemble learning approaches are more effective for liquefaction safety prediction using CPT and seismic data.

Table 7: Performance evaluation of machine learning models using the synthetic testing dataset.

	ANN	RF	SVR	DT
R^2	0.99	0.98	0.89	0.82
RMSE	2.26×10^{-3}	4.15×10^{-3}	7.41×10^{-3}	9.58×10^{-3}
MAE	1.06×10^{-3}	3.72×10^{-3}	7.39×10^{-3}	8.97×10^{-3}
NSE	0.99	0.98	0.88	0.81
PI	1.99	1.98	1.82	1.79
RSR	0.10	0.15	0.23	0.29
WI	0.99	0.97	0.88	0.82

Fig. 7 illustrates the comparison between predicted and reference F_s values under optimized model conditions. The ANN and RF models show predictions that closely align with the reference values, confirming their superior predictive capability for liquefaction safety assessment. In contrast, larger deviations are observed for the SVR and DT models, reflecting their comparatively lower accuracy. Overall, these results demonstrate that ANN provides the most accurate and stable predictions of the factor of safety against liquefaction within the Robertson and Wride framework, followed closely by RF, while SVR and DT exhibit limited predictive capability for this application.

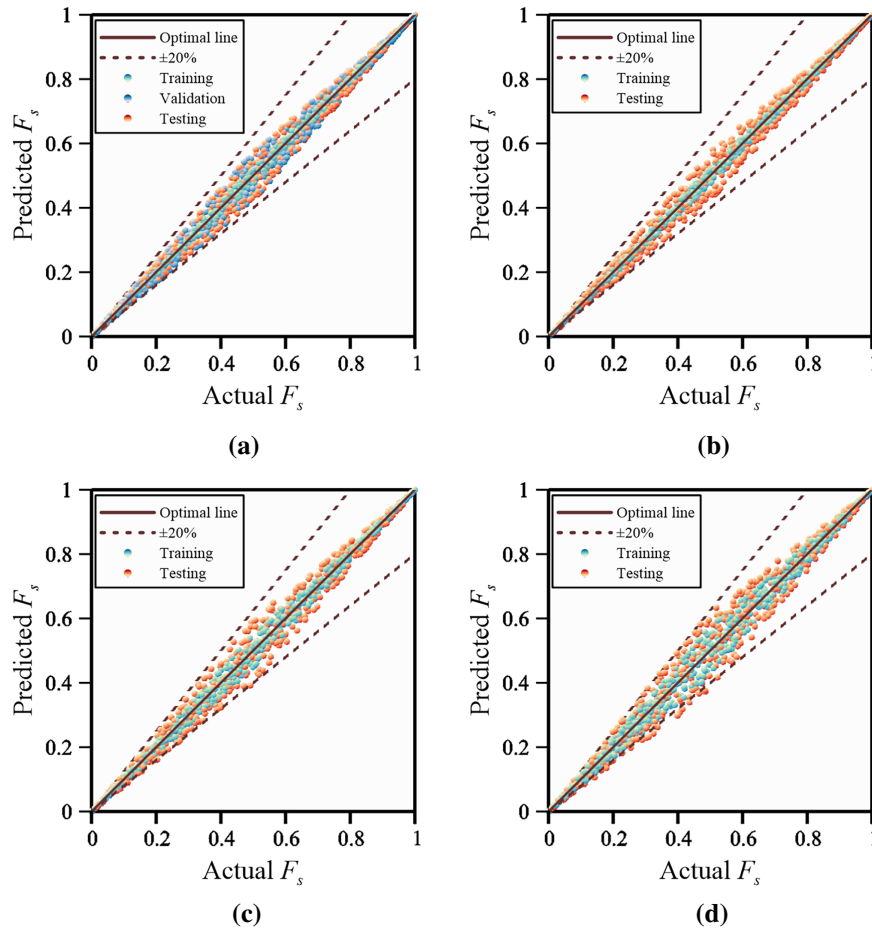


Figure 7: Comparative performance evaluation of optimized machine learning models: (a) Optimized ANN model configuration; (b) Optimized RF model configuration; (c) Optimized SVR model configuration; and (d) Optimized DT model configuration.

The same model configurations were subsequently applied to the field data models trained using observed liquefaction and nonliquefaction cases. As summarized in Table 8, ANN again achieved the highest performance, with an R^2 value of 0.91, RMSE of 7.65×10^{-2} , MAE of 6.89×10^{-2} , NSE of 0.91, PI of 1.85, RSR of 0.19 and WI of 0.91. RF showed comparable performance with an R^2 value of 0.90, while SVR and DT yielded lower R^2 values of 0.78 and 0.77, respectively. These results indicate that ANN and RF maintained relatively strong predictive performance when trained using actual field observations. The consistent ranking of the models for both datasets supports the selection of ANN as the principal architecture for constructing the subsequent integrated prediction model.

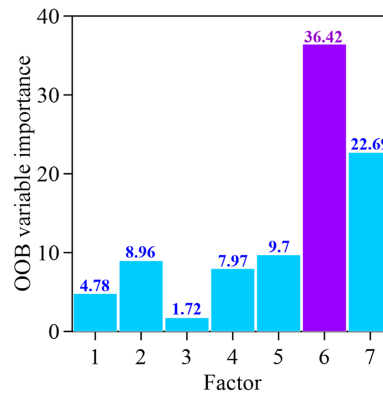
Table 8: Performance evaluation of machine learning models using the field testing dataset.

	ANN	RF	SVR	DT
R ²	0.91	0.90	0.78	0.77
RMSE	7.65×10^{-2}	8.12×10^{-2}	9.89×10^{-2}	9.92×10^{-2}
MAE	6.89×10^{-2}	8.87×10^{-2}	9.96×10^{-2}	9.97×10^{-2}
NSE	0.91	0.90	0.77	0.75
PI	1.85	1.84	1.71	1.69
RSR	0.19	0.22	0.34	0.35
WI	0.91	0.90	0.76	0.74

To further examine potential overfitting, the testing to training RMSE ratio was calculated for each model. For the synthetic dataset, the ratios were 1.02, 1.03, 1.02, and 1.01 for ANN, RF, SVR, and DT, respectively, while the corresponding ratios for the field dataset were 1.01, 1.02, 1.02, and 1.03. Since all ratios were close to 1, the training and testing errors were comparable, indicating no apparent overfitting and satisfactory model generalization.

4.3 Parameter Importance

The contribution of individual variables to liquefaction safety prediction was quantified through RF based OOB analysis. The analysis was conducted using the synthetic database to quantify the relative influence of each input parameter on predictions within the Robertson and Wride framework. The corresponding results are presented in Fig. 8. The OOB importance values for z , σ_v , σ'_v , q_c , f_s , A_{max} , and M are 4.78, 8.96, 1.72, 7.97, 9.70, 36.42, and 22.69, respectively. Among these parameters, peak ground acceleration is the most influential factor, followed by earthquake magnitude, highlighting the importance of seismic loading characteristics in the synthetic model predictions. CPT derived parameters, particularly sleeve friction and cone tip resistance, show moderate to high importance, indicating that soil behavior characteristics significantly influence liquefaction resistance. In contrast, effective stress displays relatively low importance, suggesting that its effect is largely incorporated indirectly through stress normalized resistance parameters. The OOB analysis suggests that both seismic loading conditions and key CPT related parameters play dominant roles in determining liquefaction safety predictions within the Robertson and Wride methodology. Overall, the OOB results indicate that seismic loading conditions and CPT related parameters are the primary contributors to predictions generated within the Robertson and Wride based synthetic model.

**Figure 8:** OOB based importance of input variables.

The contribution of individual variables to liquefaction safety prediction was further quantified using Shapley Additive Explanations (SHAP) analysis based on the optimized preliminary ANN model trained using the 279,936 record synthetic database. This analysis was conducted before construction of the integrated database and was used to evaluate the relative contribution of each input parameter to the preliminary ANN predictions within the Robertson and Wride framework. The corresponding results are presented in Fig. 9. The mean absolute SHAP values for z , σ_v , σ'_v , q_c , f_s , A_{max} , and M are 0.016, 0.029, 0.006, 0.026, 0.032, 0.118, and 0.074, respectively. Among these parameters, peak ground acceleration exhibits the highest mean absolute SHAP value, followed by earthquake magnitude, indicating that seismic loading characteristics exert the strongest influence on the ANN predictions. CPT related parameters, particularly sleeve friction and cone tip resistance, also show notable contributions, indicating their importance in representing subsurface resistance characteristics associated with liquefaction behavior. Total stress exhibits a moderate contribution, whereas effective stress shows the lowest mean absolute SHAP value among the considered variables. The SHAP analysis therefore indicates that seismic loading conditions and key CPT related parameters are the principal contributors to the ANN predictions. Overall, the variable ranking obtained from the SHAP analysis is generally consistent with that obtained from the RF based OOB analysis, providing additional support for the interpretation of variable importance within the proposed machine learning framework. More importantly, the OOB and SHAP importance results were further incorporated into the subsequent data generation procedure. Variables identified as having greater influence were discretized using finer intervals, whereas variables with lower importance were represented using coarser intervals. Combined with the statistical distributions adopted for the input parameters, this importance guided sampling strategy generated 25,920 additional input combinations for field informed model output generation.

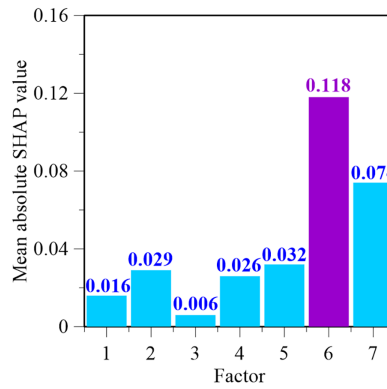


Figure 9: SHAP based importance of input variables.

4.4 Influence of Parameter Sampling on Prediction Accuracy

A set of synthetic databases with different levels of parameter discretization was developed to examine how dataset size influences prediction performance. The resulting databases contained between 2187 and 4,782,969 samples. As shown in Table 9, model accuracy increased with the number of available samples. The smallest dataset produced an R^2 value of 0.74, whereas expanding the database to 78,125 samples improved R^2 to 0.91. A further increase to 279,936 samples yielded an R^2 value of 0.99, accompanied by substantial reductions in RMSE, MAE, and RSR and corresponding increases in NSE, PI, and WI. Beyond this point, only minor performance gains were observed, suggesting that the model had approached convergence. Fig. 10 compares prediction results obtained from uniform interval sampling and probability based sampling schemes. For smaller datasets, probability based sampling generally produced higher prediction accuracy. However, when the database size exceeded approximately 279,936 samples, both sampling approaches

achieved similarly high performance, with R^2 values approaching 0.99. These findings indicate that adequate sample coverage is critical for representing the complex relationships embedded in the Robertson and Wride framework. Increasing the size of the synthetic database substantially improves prediction accuracy and model stability, although the benefits diminish once a sufficiently large dataset has been established. Based on this convergence behavior, the 279,936 record database was retained as the synthetic component for subsequent variable importance analysis and integrated model construction.

Table 9: Prediction accuracy under different parameter sampling resolutions.

Number of Samples	Error Metrics						
	R^2	RMSE	MAE	NSE	PI	RSR	WI
2187	0.74	9.58×10^{-3}	9.39×10^{-3}	0.73	1.68	0.29	0.73
16,384	0.85	8.21×10^{-3}	7.17×10^{-3}	0.84	1.76	0.21	0.85
78,125	0.91	6.54×10^{-3}	4.62×10^{-3}	0.91	1.84	0.15	0.91
279,936	0.99	2.26×10^{-3}	1.06×10^{-3}	0.99	1.99	0.10	0.99
823,543	0.99	2.23×10^{-3}	1.02×10^{-3}	0.99	1.99	0.10	0.99
2,097,152	0.99	2.17×10^{-3}	9.98×10^{-4}	0.99	1.99	0.09	0.99
4,782,969	0.99	2.15×10^{-3}	9.96×10^{-4}	0.99	1.99	0.09	0.99

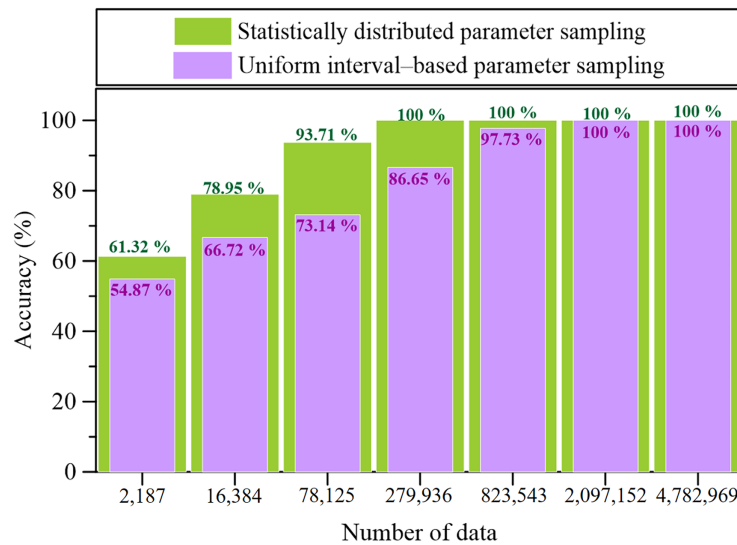


Figure 10: Effect of parameter discretization level on prediction accuracy.

4.5 Construction and Performance Evaluation of the Integrated Models

Following the preliminary model evaluation and variable importance analyses, an integrated training database was constructed to incorporate information from both the synthetic and field based models. The OOB and SHAP importance results were used to control the discretization density of the input variables during additional statistical sampling. More influential variables were divided into finer intervals, whereas less influential variables were represented using coarser intervals. This procedure produced 25,920 additional input combinations following the statistical distributions adopted in this study.

The trained field model was used to generate the output values corresponding to these 25,920 input combinations. In parallel, the trained synthetic model was used to generate the output values for the original

279,936 synthetic input combinations. The resulting 25,920 field informed records were then combined with the 279,936 synthetic model derived records, producing an integrated database containing 305,856 samples.

The integrated database was subsequently used to retrain the ANN, RF, SVR, and DT models using their previously selected model configurations. This additional training stage allowed all four algorithms to learn from the same integrated database containing both synthetic model derived and field informed records. Among the four retrained models, ANN continued to exhibit the best overall predictive performance, confirming the suitability of the ANN architecture selected in [Section 4.2](#). ANN was therefore retained as the final integrated model. The final ANN achieved an R^2 value of 0.96, RMSE of 8.65×10^{-3} , MAE of 7.92×10^{-3} , NSE of 0.96, PI of 1.93, RSR of 0.15, and WI of 0.95. The testing to training RMSE ratio of the final ANN was 1.02, indicating comparable errors between the training and testing datasets and no apparent evidence of overfitting.

5 Evaluation of the Integrated Machine Learning Models Using Real World Field Cases

5.1 Evaluation of Integrated Models Using Real Field Cases

To further evaluate model performance under real earthquake conditions, CPT records from 87 documented field cases, consisting of 55 liquefaction and 32 nonliquefaction observations and associated primarily with the 1999 Chi-Chi earthquake in Taiwan, were compiled from the studies of Juang et al. (2002), Ku et al. (2004), and Moss et al. (2006) [31–33]. These 87 cases constituted an evaluation only dataset and were mutually exclusive at the individual case record level from the 227 field cases used for field model development. Thus, none of the 87 cases or their observed liquefaction/nonliquefaction labels was used to train the field model or to retrain the final integrated models. For the classification assessment, liquefaction was treated as the positive class. Among the 87 independent cases, the final integrated ANN correctly classified all 55 observed liquefaction cases and all 32 observed nonliquefaction cases. The corresponding confusion matrix components were therefore TP = 55, TN = 32, FP = 0, and FN = 0. Consequently, the classification accuracy, sensitivity (recall), specificity, precision, and F1-score were all 1.00 (100%). The confusion matrix and corresponding classification metrics are summarized in [Table 10](#). The observed 100% classification accuracy represents complete agreement between the ANN predictions and the documented observations for this specific 87 case independent evaluation dataset. Because the evaluation dataset is limited in size and is primarily associated with the 1999 Chi-Chi earthquake, this result should not be interpreted as evidence of universal predictive performance across all earthquakes, geological environments, or liquefaction settings.

5.2 Predictions with Missing Input Parameters

In engineering practice, incomplete CPT records frequently lead to the absence of one or more variables required for liquefaction assessment. To investigate the resilience of the final integrated ANN model to incomplete field information, its predictive performance was examined under scenarios in which individual input variables were intentionally omitted. The unavailable variables were substituted with randomly sampled values to simulate the uncertainty associated with field measurements. The resulting changes in liquefaction safety predictions are presented in [Table 11](#). The 100% accuracy reported for the complete input condition refers specifically to the same 87 case independent evaluation dataset described in [Section 5.1](#) and was used only as the baseline for evaluating the effects of missing input variables.

Table 10: Confusion matrix and classification performance of the final ANN for the 87 case independent field evaluation dataset.

Observed Class	Predicted Liquefaction	Predicted Nonliquefaction
Liquefaction	55 (TP)	0 (FN)
Nonliquefaction	0 (FP)	32 (TN)
Performance metric		Value
TP		55
TN		32
FP		0
FN		0
Accuracy		100%
Sensitivity		100%
Specificity		100%
Precision		100%
Recall		100%
F1-score		100%

Table 11: Classification accuracy on the 87 case independent field evaluation dataset under different missing input conditions.

Case	Description	Accuracy (%)
1	All parameters included	100
2	Missing z	92.7
3	Missing σ_v	81.7
4	Missing σ'_v	96.8
5	Missing q_c	85.4
6	Missing f_s	80.2

As reported in [Table 11](#), the model attained 100% accuracy when all input variables were available. This result was used as the benchmark against which the effects of missing data were evaluated. When individual parameters are removed, prediction accuracy generally decreases, with the degree of reduction varying depending on the missing variable. Among the considered cases, missing sleeve friction and cone tip resistance lead to moderate reductions in accuracy to 80.2% and 85.4%, respectively, reflecting the significance of CPT derived soil resistance parameters in characterizing liquefaction susceptibility. The exclusion of effective stress results in a relatively small reduction in accuracy to 96.8%, suggesting that its influence is largely embedded within stress normalized resistance parameters. When total stress is removed, prediction accuracy decreases to 81.7%, indicating its role in defining both seismic demand and stress normalization. In contrast, missing depth produces only a minor impact, with accuracy remaining at 92.7%, implying that depth effects are indirectly captured through stress related variables.

The complete input condition, which achieved 100% agreement with the documented field observations, was adopted as the reference case for evaluating the effects of missing parameters. The results indicate that the proposed machine learning framework maintains satisfactory predictive performance even when CPT information is incomplete. Additional analyses were conducted to examine model response to missing input information. Unavailable variables were represented by randomly sampled values, and repeated simulations

were performed using the trained model. Owing to the stochastic nature of the sampling process, reported accuracies correspond to the mean values obtained from multiple realizations. The required simulation count was identified through convergence assessment. Fig. 11 shows that the estimated accuracy approached a stable value as the number of Monte Carlo iterations increased. Beyond approximately 30 repetitions, only minor changes were observed, indicating that the accuracy curve had approached a plateau and that a stable estimate had been reached. To complement the averaged accuracy, the dispersion across repeated predictions was also computed using Eq. (22). This standard deviation summarizes run to run fluctuations and serves as a quantitative indicator of prediction uncertainty.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}, \quad (22)$$

where σ represents the standard deviation of classification accuracy across the Monte Carlo realizations, n denotes the total number of Monte Carlo realizations, x_i corresponds to the predicted classification from the i th realization, and $\bar{x}$ is the average value of all simulated predictions. The variation of σ as a function of the number of model evaluations is presented in Fig. 12. With additional Monte Carlo realizations, the spread of the outcomes narrows, and the standard deviation levels off once the run count reaches roughly thirty. The emergence of this plateau suggests that the predicted liquefaction classification has attained a repeatable estimate. This convergence supports the stability of the uncertainty quantification procedure and confirms that the adopted averaging approach yields reliable results within the Robertson and Wride CPT based liquefaction evaluation framework.

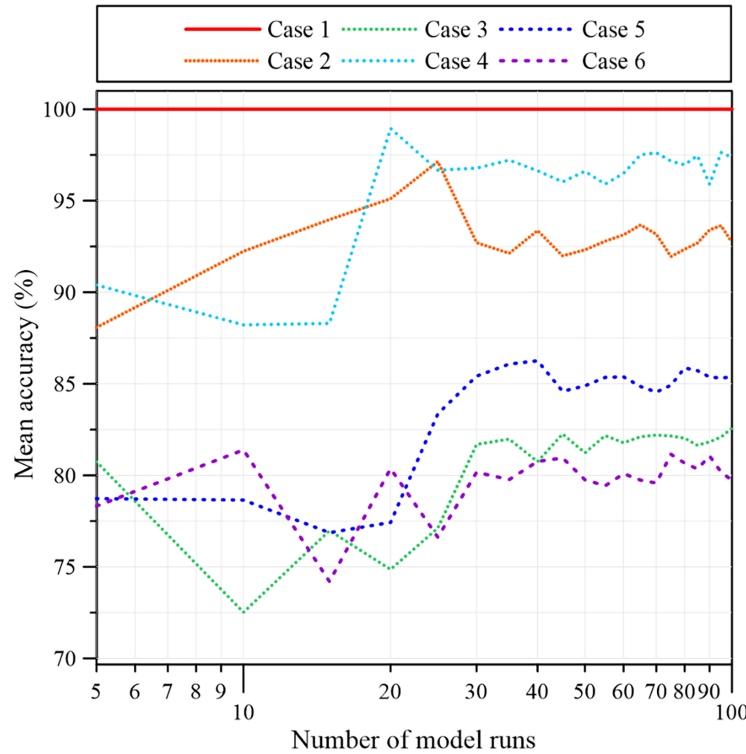


Figure 11: Convergence of mean accuracy for 100 model runs.

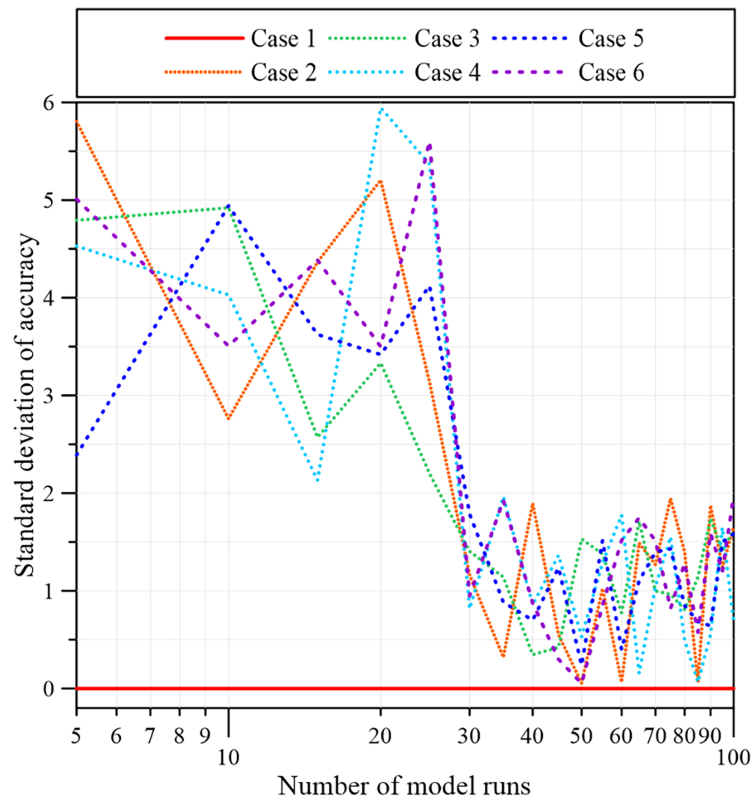


Figure 12: Convergence of standard deviation of accuracy for 100 model runs.

Table 12 compares the classification accuracy of the four integrated machine learning models using the same 87 case independent evaluation dataset. ANN achieved the highest agreement with the documented observations, correctly classifying all 87 cases, whereas RF, SVR, and DT achieved accuracies of 98.4%, 92.3%, and 78.2%, respectively. These results indicate that ANN exhibited the strongest classification performance among the four investigated models for this specific evaluation dataset. However, this comparison is dataset specific and should not be interpreted as establishing a universal performance ranking across all liquefaction conditions.

Table 12: Prediction accuracy of the four integrated machine learning models using 87 independent field cases.

Description	Accuracy (%)
ANN	100
RF	98.4
SVR	92.3
DT	78.2

5.3 Prediction Uncertainty Using Monte Carlo Simulation

Prediction uncertainty was quantified through Monte Carlo analysis using the 87 field CPT records. Under incomplete input conditions, unavailable parameters were repeatedly sampled and the resulting classifications were compared with the documented field observations. The prediction results exhibited limited variation across repeated simulations, indicating stable model behavior under input uncertainty.

The relatively narrow variation further supports the robustness of the proposed framework for liquefaction classification when selected field parameters are unavailable.

6 Conclusions

This study developed a hybrid machine learning framework for liquefaction assessment by integrating information from a synthetic model based on the Robertson and Wride procedure and a field model trained using observed liquefaction and nonliquefaction behavior. Rather than directly combining the predictions of the two models, their outputs were used to construct an integrated training database, which was subsequently used to retrain ANN, RF, SVR, and DT. ANN, previously identified as the principal model in [Section 4.2](#), retained the strongest overall performance and was adopted as the final model. The main conclusions, advantages, and limitations of the present study are summarized as follows:

- (1) Four machine learning algorithms, including ANN, RF, SVR, and DT, were investigated for liquefaction safety estimation. Among the evaluated models, ANN produced the highest prediction accuracy, followed by RF, demonstrating the strongest overall predictive performance among the investigated algorithms. Variable importance analyses based on ANN SHAP values and RF OOB statistics consistently revealed that seismic characteristics, particularly peak ground acceleration and earthquake magnitude, exert the greatest influence on the model predictions. CPT related variables, including cone penetration resistance and sleeve friction, also contributed substantially to model predictions and served as important indicators of subsurface conditions. The resulting importance ranking was further used to guide the discretization density adopted for generation of the field informed input combinations.
- (2) Model accuracy improved with increasing dataset density. Performance approached convergence when the synthetic database reached approximately 279,936 samples, at which point the preliminary synthetic ANN achieved an R^2 value of 0.99. The optimized model configurations were subsequently applied to the field data models to maintain consistent model complexity. These results indicate that sufficient data density is important for achieving stable model performance, while the transfer of the optimized configurations provides a consistent basis for comparison between the synthetic and field models.
- (3) Based on the OOB and SHAP importance results, 25,920 additional statistically distributed input combinations were generated using importance guided discretization. The field trained model was used to generate the corresponding outputs for these records, while the synthetic trained model generated outputs for the original 279,936 records. The two datasets were combined to form an integrated database containing 305,856 records. ANN, RF, SVR, and DT were subsequently retrained using this integrated database. Among the four integrated models, ANN continued to exhibit the best overall performance, confirming the model selection established in [Section 4.2](#). ANN was therefore retained as the final model. The final ANN achieved an R^2 of 0.96, RMSE of 8.65×10^{-3} , MAE of 7.92×10^{-3} , NSE of 0.96, PI of 1.93, RSR of 0.15, and WI of 0.95.
- (4) For the specific independent evaluation dataset consisting of 87 documented field cases, including 55 liquefaction and 32 nonliquefaction observations, the final ANN correctly classified all cases, corresponding to TP = 55, TN = 32, FP = 0, and FN = 0. Accordingly, the accuracy, sensitivity, specificity, precision, recall, and F1-score were all 100%. These 87 evaluation cases were mutually exclusive at the individual case record level from the 227 field cases used for field model development. Nevertheless, the result should be interpreted as dataset specific evidence because the independent evaluation set was limited to 87 cases and was primarily associated with the 1999 Chi-Chi earthquake. It therefore does

not establish universal predictive performance across all liquefaction settings. Monte Carlo analysis further indicated stable prediction behavior under the investigated incomplete input conditions.

A principal advantage of the present framework is that it combines the systematic coverage provided by a large synthetic database with direct information from observed field behavior. Unlike a model developed solely from synthetic calculations, the proposed approach incorporates a separately trained field model to generate field informed training information over an expanded input domain. The field informed records are combined with records generated by the synthetic model, and the resulting integrated database is used to retrain ANN, RF, SVR, and DT. ANN, which had been identified as the principal architecture in [Section 4.2](#), retained the strongest performance after integrated training and was adopted as the final model. This structure enables conventional liquefaction knowledge and information learned from observed field behavior to contribute simultaneously to the final prediction model. In addition, the ability to accommodate incomplete input conditions increases the practical applicability of the framework when complete CPT or stress information cannot be obtained.

Several limitations should also be recognized. First, the field model was developed using 227 documented cases, and its generalization capability may therefore depend on the representativeness of the available field database. Second, the 25,920 field informed records were generated from the outputs of the field trained model rather than from additional direct field observations. Consequently, uncertainty or bias present in the field model may be propagated into the integrated training database. Third, although the 87 field cases used for final evaluation were independent at the individual case record level from the 227 cases used for field model development and included both liquefaction and nonliquefaction observations, they were primarily associated with the 1999 Chi-Chi earthquake. Therefore, the 100% classification accuracy obtained for this evaluation dataset should be interpreted as dataset specific performance rather than evidence of universal generalization across different earthquakes, geological settings, and site conditions. Additional independent datasets covering a broader range of seismic events and geotechnical environments will be required to further assess model generalizability. Fourth, the synthetic database was generated according to the Robertson and Wride procedure and therefore inherits the assumptions and limitations of that conventional method. Finally, the missing input analysis was based on random substitution and Monte Carlo simulation, which may not fully reproduce all dependencies among geotechnical variables encountered under actual site conditions. Future work will develop a user-friendly interface.

Acknowledgement: None.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Cheng-Yu Ku provided conceptualization. Ting-Yuan Wu led data curation and was responsible for validation and visualization. Chih-Yu Liu contributed to drafting and revising the manuscript. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The datasets used in this study are available from the corresponding author upon reasonable request.

Ethics Approval: Not applicable.

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

References

1. Seed HB, Idriss IM. Simplified procedure for evaluating soil liquefaction potential. *J Soil Mech and Found Div.* 1971;97(9):1249–73. doi:10.1061/jsfeaq.0001662.
2. Youd TL, Idriss IM. Liquefaction resistance of soils: summary report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of liquefaction resistance of soils. *J Geotech Geoenviron Eng.* 2001;127(4):297–313. doi:10.1061/(asce)1090-0241(2001)127:4(297).
3. Roy B, Chakraborty P. Liquefaction susceptibility assessment using shear wave velocity: a review with a case study. *Discover Geosci.* 2025;3(1):175. doi:10.1007/s44288-025-00288-2.
4. Seed HB, Lee KL. Liquefaction of saturated sands during cyclic loading. *J Soil Mech and Found Div.* 1966;92(6):105–34. doi:10.1061/jsfeaq.0000913.
5. Schofield AN, Wroth P. *Critical state soil mechanics* (Vol. 310). London, UK: McGraw-Hill; 1968.
6. Yang Q, Zhang JM, Wang R. Fabric quantities governing liquefaction resistance and post-liquefaction shear deformation of granular material. *Comput Geotech.* 2025;185:107344. doi:10.1016/j.compgeo.2025.107344.
7. Robertson PK, Wride CF. Evaluating cyclic liquefaction potential using the cone penetration test. *Can Geotech J.* 1998;35(3):442–59. doi:10.1139/t98-017.
8. Idriss IM, Boulanger RW. Semi-empirical procedures for evaluating liquefaction potential during earthquakes. *Soil Dyn Earthq Eng.* 2006;26(2–4):115–30. doi:10.1016/j.soildyn.2004.11.023.
9. Chiaradonna A, Monaco P. DMT-based liquefaction triggering procedure accounting for the fines content effect. *Soil Dyn Earthq Eng.* 2025;199:109668. doi:10.1016/j.soildyn.2025.109668.
10. Javanbakht A, Molnar S, Sadrekarimi A, Adhikari SR. Performance-based liquefaction analysis and probabilistic liquefaction hazard mapping using CPT data within the Fraser River delta, Canada. *Soil Dyn Earthq Eng.* 2025;189:109101. doi:10.1016/j.soildyn.2024.109101.
11. Saye SR, Olson SM, Franke KW. Common-origin approach to assess level-ground liquefaction susceptibility and triggering in CPT-compatible soils using ΔQ . *J Geotech Geoenviron Eng.* 2021;147(7):04021046. doi:10.1061/(asce)gt.1943-5606.0002515.
12. Yang H, Liu Z. A fused sampling method integrating geotechnical and geophysical data for assessing three-dimensional soil liquefaction-induced damage capacity. *Comput Geotech.* 2024;166:106024. doi:10.1016/j.compgeo.2023.106024.
13. Macedo J, Vergaray L. State-based assessment of cyclic liquefaction manifestation. *Soil Dyn Earthq Eng.* 2024;179:108520. doi:10.1016/j.soildyn.2024.108520.
14. Rasanen RA. Forecasting liquefaction hazard from Seattle fault zone earthquakes: impacts on a major metropolitan area. *Nat Hazards.* 2025;121(16):19117–49. doi:10.1007/s11069-025-07563-y.
15. Hwang JH, Khoshnevisan S, Juang CH, Lu CC. Soil liquefaction potential evaluation—an update of the HBF method focusing on research and practice in Taiwan. *Eng Geol.* 2021;280:105926. doi:10.1016/j.enggeo.2020.105926.
16. Sianko I, Ozdemir Z, Hajirasouliha I, Pilakoutas K. A probabilistic liquefaction hazard analysis: case studies from the Marmara region. *Geotech Geol Eng.* 2025;43(2):103. doi:10.1007/s10706-024-03042-6.
17. Wang Y, Tian HM. Digital geotechnics: from data-driven site characterisation towards digital transformation and intelligence in geotechnical engineering. *Georisk Assess Manage Risk Eng Syst Geohazards.* 2024;18(1):8–32. doi:10.1080/17499518.2023.2278136.
18. Guan Z, Wang Y, Phoon KK. Data-driven recovery of incomplete geotechnical dataset using low-rank matrix completion. *Can Geotech J.* 2025;62:1–20. doi:10.1139/cgj-2024-0781.
19. Liu CY, Ku CY, Chiu YJ, Wu TY. Evaluation of liquefaction potential in central Taiwan using random forest method. *Sci Rep.* 2024;14:27517. doi:10.1038/s41598-024-79127-2.
20. Liu CY, Ku CY, Wu TY, Chiu YJ, Chang CW. Liquefaction susceptibility mapping using artificial neural network for offshore wind farms in Taiwan. *Eng Geol.* 2025;351:108013. doi:10.1016/j.enggeo.2025.108013.
21. Ziaie A, Mehdizadeh B, Safi Jahanshahi F, Ahmadi N, Ghanizadeh AR. Prediction of liquefaction-induced lateral displacements using hybrid GBRT and EOA. *J Soft Comput Civ Eng.* 2026;10(1):e225827. doi:10.22115/SCCE.2025.2061.

22. Soleimani Fard H, Goudarzy M. Influence of surcharge on cone penetration test results and the inspection of various approaches for capturing its effect: a case study. *Int J Geo Eng.* 2021;12(1):17. doi:10.1186/s40703-021-00146-3.
23. Sukkarak R, Tanapalungkorn W, Likitlersuang S, Ueda K. Liquefaction analysis of sandy soil during strong earthquake in Northern Thailand. *Soils Found.* 2021;61(5):1302–18. doi:10.1016/j.sandf.2021.07.003.
24. Wang MX, Leung YF, Lo MK. Incorporating region-variability of model bias into liquefaction-triggering procedures for sandy and gravelly soils through BUS-powered hierarchical Bayesian updating. *Can Geotech J.* 2025;62:1–24. doi:10.1139/cgj-2024-0059.
25. Hutabarat D, Bray JD. Estimating the severity of liquefaction ejecta using the cone penetration test. *J Geotech Geoenviron Eng.* 2022;148(3):04021195. doi:10.1061/(asce)gt.1943-5606.0002744.
26. Konkol J, Gebremeskel BM, Stutz HH. Evaluation of sleeve friction measurement in soft soils using rate interface tests under different boundary conditions. *Can Geotech J.* 2025;62:1–17. doi:10.1139/cgj-2024-0401.
27. Tsai CC, Li PC. Quantifying near-fault motion effects on soil liquefaction through effective stress site response analysis. *Soil Dyn Earthq Eng.* 2024;183:108779. doi:10.1016/j.soildyn.2024.108779.
28. Upadhyaya S, Maurer BW, Green RA, Rodriguez-Marek A. Selecting the optimal factor of safety or probability of liquefaction triggering for engineering projects based on misprediction costs. *J Geotech Geoenviron Eng.* 2021;147(6):04021026. doi:10.1061/(asce)gt.1943-5606.0002511.
29. Ku CY, Wu TY, Liu CY, Hsu WY. AI-enhanced soil classification with incomplete CPT data for offshore wind farm. *Sci Rep.* 2026;16:10589. doi:10.1038/s41598-026-46356-6.
30. Yang Z, Liu X, Guo L, Cui Y, Su X, Ling X. Soil classification and site variability analysis based on CPT—a case study in the Yellow River subaquatic delta, China. *J Mar Sci Eng.* 2021;9(4):431. doi:10.3390/jmse9040431.
31. Juang CH, Yuan H, Lee DH, Ku CS. Assessing CPT-based methods for liquefaction evaluation with emphasis on the cases from the Chi-Chi, Taiwan, earthquake. *Soil Dyn Earthq Eng.* 2002;22(3):241–58. doi:10.1016/s0267-7261(02)00013-1.
32. Ku CS, Lee DH, Wu JH. Evaluation of soil liquefaction in the Chi-Chi, Taiwan earthquake using CPT. *Soil Dyn Earthq Eng.* 2004;24(9–10):659–73. doi:10.1016/j.soildyn.2004.06.009.
33. Moss RE, Seed RB, Kayen RE, Stewart JP, Der Kiureghian A, Cetin KO. CPT-based probabilistic and deterministic assessment of *in situ* seismic soil liquefaction potential. *J Geotech Geoenviron Eng.* 2006;132(8):1032–51. doi:10.1061/(asce)1090-0241(2006)132:8(1032).
34. Rasanen RA, Geyin M, Maurer BW. Select liquefaction case histories from the 2001 Nisqually, Washington, earthquake: a digital data set and assessment of model performance. *Earthq Spectra.* 2023;39(3):1534–57. doi:10.1177/87552930231174244.
35. Geyin M, Maurer B. CPT-based liquefaction case histories from global earthquakes: a digital dataset (version 1) [Dataset]. Austin, TX, USA: DesignSafe-CI; 2021.
36. Şehmusoğlu EH, Kurnaz TE, Erden C. Estimation of soil liquefaction using artificial intelligence techniques: an extended comparison between machine and deep learning approaches. *Environ Earth Sci.* 2025;84(5):130. doi:10.1007/s12665-025-12116-4.