

Article

Hybrid Stacking Ensemble Learning for Direct Prediction and Geotechnical Interpretation of Liquefaction Factor of Safety in the Port Sudan Coastal Plain

Ahmed A. H. S. Hussain ¹, Dafalla Wadi ², Wahib Yahya ³, Husam Eldin Ahmed ⁴, Jingang Lü ⁵,
Mohammed Albashir ⁶ and Wenbing Wu ^{1,*}

- ¹ Faculty of Engineering, MOST International Science and Technology Cooperation Base for Geological Engineering, China University of Geosciences, Wuhan 430074, China; ahmedaltaib292@gmail.com
- ² Department of Hydrogeology and Engineering Geology, College of Petroleum Geology and Minerals, University of Bahri, Khartoum 11111, Sudan; wadiadam@cug.edu.cn
- ³ Hubei Provincial Key Laboratory of Oil and Gas Exploration and Development Theory and Technology, China University of Geosciences, Wuhan 430074, China; wahibalwrafy2016@gmail.com
- ⁴ Engineering Tests Lab, Sea Ports Corporation, Port Sudan 33311, Sudan; husameldin55@hotmail.com
- ⁵ Wuhan Municipal Engineering Design & Research Institute Co., Ltd., Wuhan 430014, China; jglv1978@163.com
- ⁶ Department of Petroleum Engineering, Faculty of Earth Resources, China University of Geosciences, Wuhan 430074, China; mohbashax@gmail.com
- * Correspondence: zjuwwb1126@163.com

Featured Application

This work develops a leakage-aware ensemble-learning framework for the direct prediction and geotechnical interpretation of the liquefaction factor of safety in Port Sudan coastal sediments, enabling rapid hazard screening and infrastructure risk mitigation in coastal sedimentary environments.

Abstract

Soil liquefaction is a major seismic hazard that threatens coastal infrastructure. Yet, accurate prediction of the liquefaction factor of safety (FS) remains challenging because of the complex nonlinear interactions among geotechnical and stress-state variables. This study proposes a novel leakage-aware Hybrid Stacking Ensemble (HSE) framework for the direct prediction and geotechnical interpretation of FS in the Port Sudan coastal plain. The proposed model combines Random Forest, Extra Trees, and Gradient Boosting Regressors through a RidgeCV meta-learner while restricting the predictor set to independent spatial, geotechnical, and stress-state variables to prevent data leakage. A database comprising 534 Standard Penetration Test (SPT)-based observations from 61 boreholes was used to evaluate the proposed model against K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), Random Forest, and XGBoost. The Hybrid Stacking Ensemble achieved the highest predictive performance, with a test RMSE of 0.058, an MAE of 0.032, and an R^2 of 0.985, demonstrating improved accuracy, robustness, and generalization over the benchmark models. SHAP-based interpretation identified $N_{l(60)}$ as the dominant predictor of FS, followed by relative density, depth, and stress-related variables, confirming that the proposed framework learned physically meaningful relationships consistent with established liquefaction mechanics. The contribution of this work lies in the development of a Hybrid Stacking Ensemble framework for the direct prediction of the liquefaction factor of safety (FS), integrated with a leakage-free modeling workflow and explainable machine learning techniques, rather than in the development of a new machine learning algorithm.

Academic Editors: Konstantinos G. Megalooikonomou and Leonidas Alexandros S. Kouris

Received: 10 June 2026

Revised: 6 July 2026

Accepted: 7 July 2026

Published: 8 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: soil liquefaction; factor of safety; Hybrid Stacking Ensemble; machine learning; SHAP; Standard Penetration Test; Port Sudan

1. Introduction

Soil liquefaction is a serious seismic hazard in coastal sedimentary deposits, as demonstrated in this study using the Port Sudan Plain as a case study [1–3]. Soil liquefaction increases the risk of damage to vital infrastructure, making it important to improve the accuracy of coastal soil liquefaction prediction using the new capabilities of modern modeling methods and tools [4–6]. Historical earthquakes, including Niigata (1964), Loma Prieta (1989), and Tohoku (2011), have shown that liquefaction-induced damage can be comparable to direct seismic shaking effects.

Liquefaction resistance is commonly evaluated using empirical or semi-empirical stress-based procedures, in which the cyclic stress ratio (CSR) is compared with the cyclic resistance ratio (CRR) to estimate FS [4,7]. Although these procedures are widely used in practice, they depend on correction factors, magnitude-scaling relationships, and case-history calibrations that simplify the coupled effects of density, stress history, grain characteristics, and seismic demand [8,9]. These simplifications can introduce epistemic uncertainty, especially in heterogeneous coastal deposits.

The results of this study fill a knowledge gap regarding the use of machine learning as a tool to study complex geotechnical soil behavior by directly modeling multivariate nonlinear relationships from data, without a predetermined functional form. Importantly, the data-driven models proposed in the manuscript do not replace, but rather complement, traditional procedures, improving the consistency of predictions and filling gaps in geotechnical knowledge [10–12].

Despite numerous publications, the problem of predicting soil liquefaction remains relevant, as existing models in this field lack sufficient predictive accuracy for modern engineering practice. Many existing studies focus on binary liquefaction/non-liquefaction classification, which does not provide the continuous FS needed for engineering design [13–15]. Others emphasize algorithm ranking but do not test a mechanistic hypothesis about how resistance, density, depth, and stress state jointly control FS. Moreover, global feature-importance scores are often reported as interpretability, although they do not explain why a specific borehole interval receives a low- or high-FS prediction [16,17]. Data leakage is also a concern when intermediate liquefaction-response variables are used as inputs [18]. These limitations are especially important in heterogeneous coastal deposits along the western Red Sea margin, where shallow groundwater and variable marine-alluvial sediments create strong spatial and stratigraphic variability.

This study addresses these gaps by developing a leakage-aware, SHAP-interpreted ensemble learning framework for the direct prediction of liquefaction FS on the Port Sudan coastal plain. The framework uses only independent spatial, geotechnical, and stress-state predictors so that the algorithms learn relationships associated with soil resistance, density, and confinement rather than reproducing variables already used to compute FS. Interpretability is treated as a decision-explanation problem: SHAP is used to decompose global and interval-scale predictions into positive and negative feature contributions, thereby explaining how the model arrives at a prediction and why particular layers are assigned low or high FS [17,19].

The Port Sudan coastal plain is an appropriate case study because it combines young coastal sediments, shallow groundwater, and moderate seismicity along the western Red Sea rift [20–23]. Loose granular soils and critical infrastructure, including port facilities and pipelines, increase the need for a reliable, site-specific liquefaction assessment [21,24–27].

Within this context, the central hypothesis of this study is that a new leakage-aware Hybrid Stacking Ensemble model, developed by combining Random Forest, Extra Trees, and Gradient Boosting Regressors through a RidgeCV meta-learner, can better capture the nonlinear interactions among soil resistance, density, depth, and stress-state variables governing the liquefaction factor of safety (FS) than individual machine learning models. It is further hypothesized that these relationships can be interpreted in a geotechnically meaningful manner using SHAP-based explanations.

To test these hypotheses, the study pursues four objectives: (i) develop a leakage-aware framework for direct FS prediction using independent geotechnical and stress-state variables; (ii) propose and evaluate the new Hybrid Stacking Ensemble model and compare its performance with KNN, ANN, Random Forest, and XGBoost; (iii) assess model robustness through testing, cross-validation, learning curves, and residual diagnostics; and (iv) interpret the learned prediction mechanisms using SHAP-based global and local analyses.

By integrating the proposed Hybrid Stacking Ensemble with leakage-aware preprocessing and SHAP-based interpretation, this study provides a reproducible, physically interpretable, and highly accurate framework for preliminary liquefaction-hazard screening in coastal sediments.

2. Study Area: Geological and Geotechnical Setting

2.1. Regional Geology and Subsurface Conditions

Port Sudan lies on a low-relief coastal plain bordered by the Red Sea to the east and Precambrian basement highlands to the west [21,28]. Its geomorphology reflects the combined effects of an ephemeral wadi sediment supply, coastal marine processes, and arid-climate conditions [20,28–30]. Figure 1 shows the geological setting of the study area.

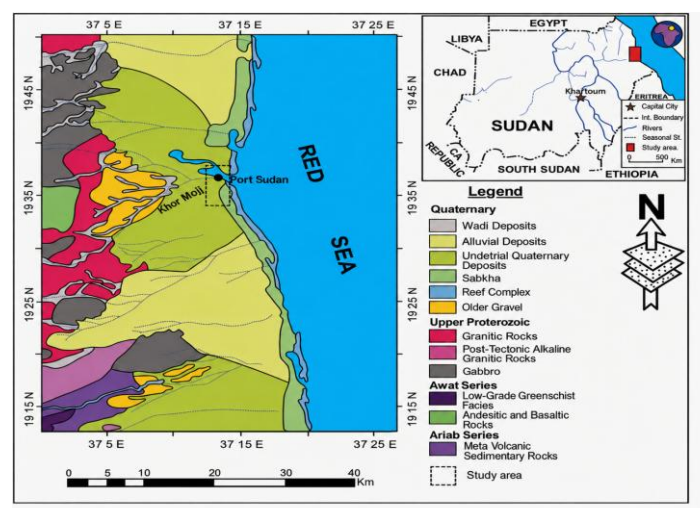


Figure 1. Location map and tectonic framework of Port Sudan (modified after [28]).

Rift-related subsidence has promoted the accumulation of thick Quaternary deposits, including alluvial-fan sands derived from the basement uplands, nearshore marine sands, and interbedded lagoonal silts and sandy silts [21,25,31]. These units are generally young, weakly cemented, and variably sorted. Grain-size data indicate predominantly fine- to medium-grained sands with localized silt lenses, resulting in marked lateral and vertical heterogeneity [21,32].

Geotechnical investigations indicate a typical sequence of loose- to medium-dense sandy fill or aeolian sand overlying saturated fine- to medium-grained sand with variable silt content, intermittent silty sand or sandy silt lenses, and denser sand or weathered coral

fragments at greater depths [21,28]. Representative SPT and lithological profiles for boreholes BH12 and BH14 are shown in Figure 2. The profiles indicate increasing SPT-N values with depth, particularly within the denser sand and weathered limestone layers, whereas lower values occur in softer clayey or silty intervals.

These geological and geotechnical characteristics create strong spatial and vertical heterogeneity in soil properties, providing an appropriate test environment for evaluating the predictive capability and generalization performance of the proposed machine learning framework.

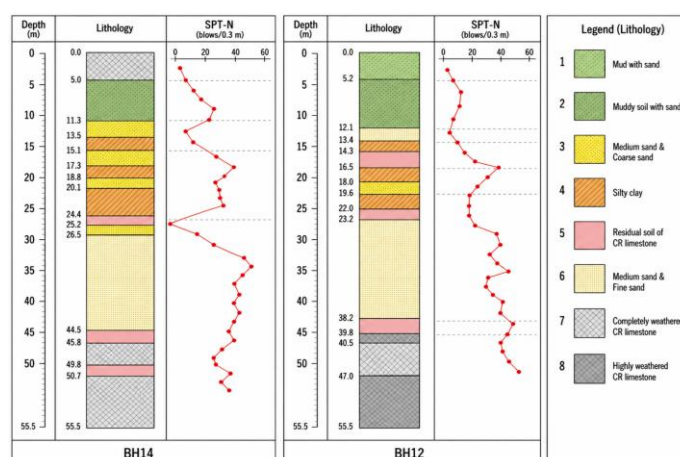


Figure 2. Stratigraphic column and representative borehole profiles (BH12 and BH14).

2.2. Hydrogeological Conditions

Groundwater in the Port Sudan coastal plain is generally shallow due to its proximity to the Red Sea and has limited surface drainage. Measured groundwater levels commonly range from 1.0 to 3.0 m below the ground surface, with variations associated with seasonality and tidal influence [21,26]. These conditions reduce effective overburden stress, maintain saturation in near-surface granular layers, and limit pore-pressure dissipation, thereby increasing liquefaction susceptibility [33,34]. Saline groundwater may also affect soil fabric and interparticle friction, although its influence on cyclic resistance is expected to be secondary to the effects of density and confining stress [34,35].

Although measured groundwater depths typically range from 1.0 to 3.0 m, seasonal recharge, evaporation, and tidal fluctuations may locally modify saturation conditions. Such variations could influence pore-water pressure and, in turn, FS values, particularly in shallow coastal deposits. While these temporal effects are not explicitly modeled in the present dataset, they represent an important source of uncertainty and should be considered in future time-dependent or probabilistic analyses.

3. Methodology

Figure 3 summarizes the machine learning workflow adopted in this study. The procedure includes data preparation, leakage-aware feature selection, training-test partitioning, model training, performance evaluation, model optimization, and model interpretation. The modeling strategy was designed to evaluate whether advanced ensemble learning algorithms can accurately capture the complex nonlinear relationships governing the liquefaction factor of safety (FS) as a function of independent geotechnical, spatial, and stress-state variables. KNN and ANN were employed as representative baseline models, whereas Random Forest, XGBoost, and the proposed Hybrid Stacking Ensemble were evaluated as advanced ensemble learning approaches. The Hybrid Stacking Ensemble combines Random Forest, Extra Trees, and Gradient Boosting Regressors as base learners,

while RidgeCV serves as the meta-learner that integrates the predictions of the individual models to produce the final FS estimate. The optimized models were subsequently evaluated using independent testing data and interpreted using feature-importance analysis and SHAP-based explanations to identify the geotechnical factors controlling low and high liquefaction FS values.

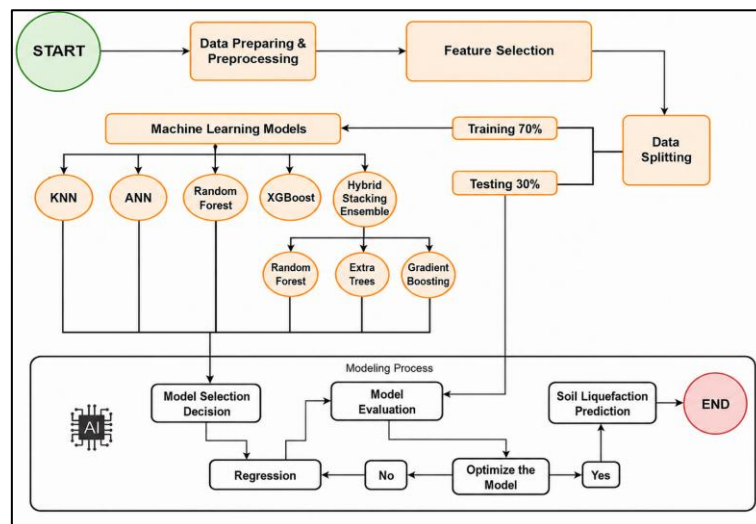


Figure 3. Workflow adopted for model development, evaluation, and prediction.

3.1. Dataset and Preprocessing

A structured geotechnical database was compiled from site-investigation reports for the Port Sudan coastal plain. The database includes borehole logs, SPT records, ground-water measurements, laboratory soil classifications, and seismic-loading parameters. The spatial distribution of the investigation sites is shown in Figure 4. The compiled database contains information from 61 boreholes and 534 SPT intervals covering depths between 0.1 and 80 m, with an average groundwater depth of approximately 1.0 m. Each SPT interval was treated as an independent observation when all variables required for FS calculation were available.

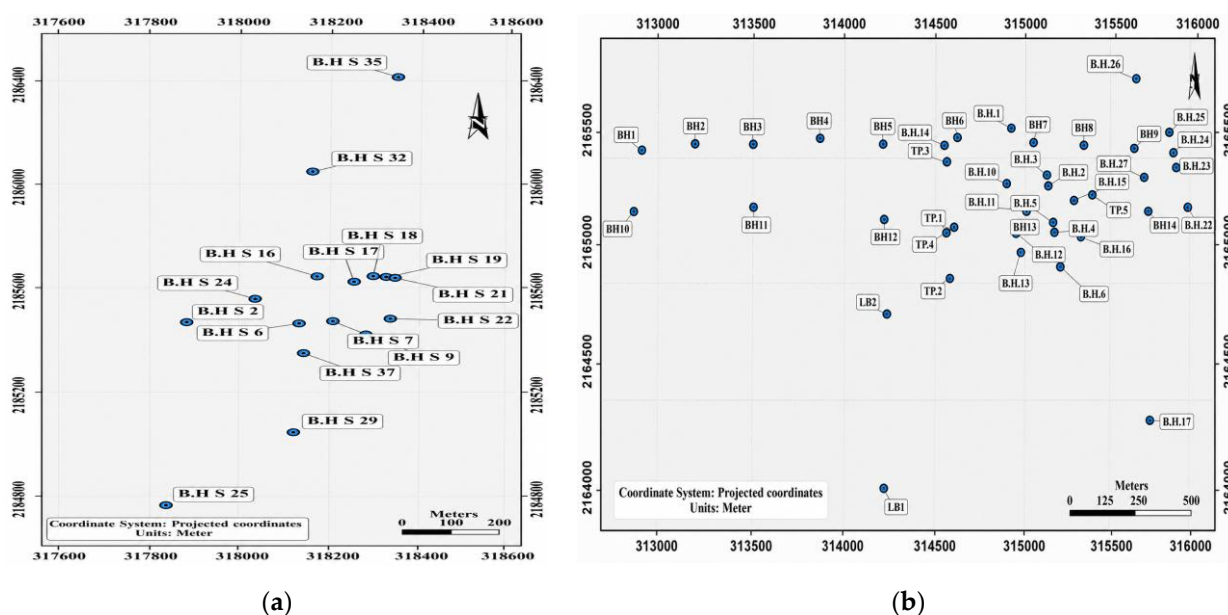


Figure 4. (a) Measurement site in the northern part of the study area. (b) Measurement site in the southern part of the study area.

The exclusion of cohesive soils (CL, CH, and ML) accounts for approximately 4.1% of the original dataset, reflecting their limited relevance to liquefaction assessment under the adopted framework. While this filtering improves consistency in modeling granular soil behavior, it may introduce a selection bias toward sandy deposits. Therefore, the final dataset is representative of liquefiable granular soils rather than the full lithological spectrum of the study area.

3.2. Summary Statistics

Table 1 presents the descriptive statistics of the final predictor variables and the target variable used for machine learning model development. The final dataset comprises 534 observations and covers depths from 0.5 to 80.0 m, with a mean depth of 25.3 m, indicating broad subsurface coverage across the investigated sites.

Measured SPT-N values range from 0 to 58 blows per 30 cm, with a mean of 25.0. The clean-sand-corrected penetration resistance, $N1(60)_{cs}$, ranges from 0 to 60, with a mean value of 17.1. Relative density (D_r) varies from 0 to 100%, with a mean of 50.3%, demonstrating that the dataset includes soil conditions ranging from very loose to dense.

The stress-related variables also exhibit considerable variability. Unit weight ranges from 17.5 to 19.0 kN/m³, whereas total vertical stress varies from 8.8 to 1425.0 kPa, primarily reflecting changes in depth and overburden pressure. The stress-reduction coefficient (rd) ranges from 0.4 to 1.0, while the overburden correction factor (K_σ) varies between 0.7 and 1.0.

The target variable, the liquefaction factor of safety (FS), ranges from 0.1 to 2.0, with a mean of 0.9 and a standard deviation of 0.5, indicating that the dataset contains both potentially liquefiable and relatively stable soil intervals. To ensure a leakage-aware modeling framework, variables directly related to liquefaction response or consequence, including CSR, CRR, liquefaction-induced settlement, lateral displacement, and the Liquefaction Potential Index (LPI), were intentionally excluded from the predictor set. Consequently, the machine learning models learned the relationship between independent geotechnical, spatial, and stress-state variables and the liquefaction factor of safety, thereby reducing circularity and avoiding artificially optimistic predictive performance.

Table 1. Summary statistics of the input variables and target variable used in the machine learning models.

Variable	<i>n</i>	Min	Max	Mean	Std. Dev.
Depth (m)	534	0.5	80.0	25.3	16.3
X (m)	534	312,829	319,809	315,046.85	1596.36
Y (m)	534	2,164,109	2,165,795.85	2,170,856.09	26,681.35
N-SPT (blows per 30 cm)	534	0.0	58.0	25.0	17.6
Unit Weight (kN/m ³)	534	17.5	19.0	17.7	0.3
Total Vertical Stress (kPa)	534	8.8	1425.0	446.6	288.7
$N1(60)_{cs}$	534	0.0	60.0	17.1	9.4
rd	534	0.4	1.0	0.6	0.2
K_σ	534	0.7	1.0	0.9	0.1
D_r (%)	534	0.0	100.0	50.3	22.7
FS	534	0.1	2.0	0.9	0.5

3.3. Target Variable and Predictors

The target variable is the liquefaction factor of safety (FS), defined using the conventional stress-based liquefaction framework as

$$FS = CRR/CSR \quad (1)$$

where

$$CSR = 0.65 \cdot \frac{\sigma_v}{\sigma'_v} \cdot \frac{a_{max}}{g} \cdot r_d \quad (2)$$

$$CRR = f(N_{1(60)cs}, \sigma'_v, FC, MSF) \quad (3)$$

where CRR is the cyclic resistance ratio, and CSR is the cyclic stress ratio. CRR was estimated from corrected SPT-based empirical relationships, whereas CSR was derived from seismic demand and in situ stress conditions following the established liquefaction assessment procedure [7]. Although FS was calculated from CRR and CSR , these two parameters were intentionally excluded from the predictor set to avoid circularity and data leakage. Likewise, post-assessment response variables, including $CRR_{7.5}$, CSR , CSR^* , liquefaction evaluation percentage, vertical settlement, cyclic shear strain, lateral displacement, and the Liquefaction Potential Index (LPI), were excluded because they are derived from or directly related to the target variable [21,33].

The final predictor set comprises spatial, geotechnical, and stress-state variables, including X coordinate, Y coordinate, depth (z), measured SPT blow count ($N_{1(60)}$), clean-sand-corrected SPT resistance ($N_{1(60)cs}$), unit weight, total vertical stress, relative density (Dr), stress-reduction coefficient (r_d), and overburden correction factor ($K\sigma$). These variables provide independent information on location, soil resistance, density, and stress conditions without directly including computed cyclic resistance, cyclic stress, or liquefaction consequence indicators. All predictors are continuous variables; therefore, categorical encoding was not required.

To assess potential multicollinearity among predictors, variance inflation factors (VIFs) were conceptually considered. Strong relationships are expected between depth, total stress, r_d , and $K\sigma$ because these variables share a physical basis. However, tree-based ensemble models are relatively insensitive to multicollinearity with respect to prediction accuracy. The interpretation of correlated variables is therefore primarily supported using SHAP analysis, which quantifies feature contributions while accounting for interaction effects.

Missing predictor values were treated using median imputation within the modeling pipeline. The proportion of missing values for individual features was generally low (<2%), and median imputation was selected for its robustness to outliers. Although alternative imputation methods (e.g., K-Nearest Neighbors or model-based imputation) may provide improved accuracy, the selected approach ensures stability and avoids introducing additional model complexity to a relatively small dataset. The standardization procedure is expressed as

$$x_{scaled} = \frac{x - \mu}{\sigma} \quad (4)$$

where x is the original predictor value, μ is the mean of the training data, and σ is the standard deviation of the training data. Imputation and scaling were fitted only on the training data within the modeling pipeline to prevent preprocessing-related data leakage [11].

The dataset was divided into training and testing subsets using a 70:30 split. To preserve the proportions of liquefiable and non-liquefiable cases, the partitioning was stratified using the engineering threshold $FS < 1$ and $FS \geq 1$. Model robustness and generalization capability were further assessed using shuffled 5-fold cross-validation. All procedures were implemented in Python 3.11 using Scikit-learn and XGBoost libraries within a leakage-aware pipeline. The proposed Hybrid Stacking Ensemble, comprising Random Forest, Extra Trees, and Gradient Boosting Regressors as base learners and RidgeCV as the meta-learner, was trained using out-of-fold predictions to ensure unbiased learning and minimize overfitting while maintaining reproducibility and reducing data leakage risk [18].

3.4. Exploratory Analysis and Feature Screening

Figure 5 shows the correlation matrix for the final predictors and FS. FS exhibits strong positive correlations with soil resistance and density-related variables, particularly N1(60)cs ($r = 0.94$), SPT-N ($r = 0.86$), and relative density (D_r) ($r = 0.85$), indicating that denser soil layers with higher penetration resistance generally possess greater resistance to liquefaction. Moderate positive correlations are observed between FS and both depth and total vertical stress ($r \approx 0.38$), reflecting the beneficial effects of increasing confinement on liquefaction resistance. Conversely, rd and $K\sigma$ exhibit moderate negative correlations with FS ($r \approx -0.41$), consistent with their roles as depth-dependent stress-reduction and overburden correction factors within the conventional liquefaction evaluation framework.

Strong correlations among several predictor variables, including depth and total vertical stress, rd and $K\sigma$, and N1(60)cs and D_r , are expected because these variables are physically related through well-established geotechnical principles rather than data redundancy. These relationships, therefore, represent meaningful engineering dependencies rather than sources of information leakage.

The very strong correlation between FS and N1(60)cs indicates that a large portion of the variability in FS is explained by penetration resistance. However, ensemble models remain necessary to capture interaction effects among resistance, density, stress, and depth, which single-variable or linear regression models cannot fully represent.

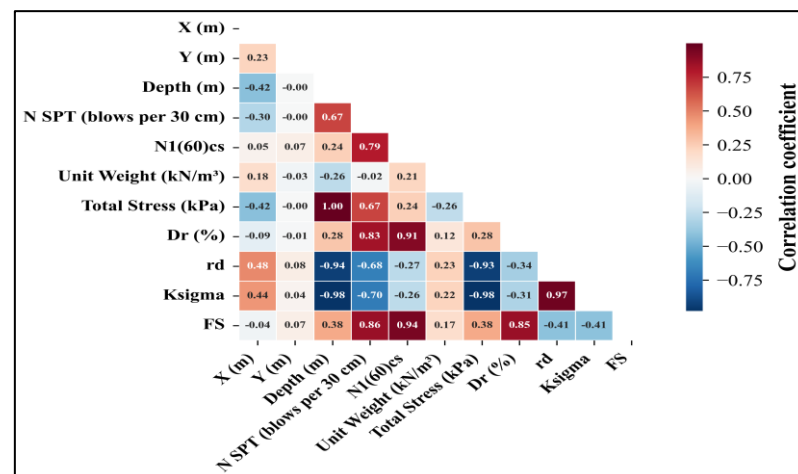


Figure 5. Correlation matrix of predictor variables.

3.5. Machine Learning Algorithms

The five algorithms were not selected solely for benchmarking; rather, they reflect different assumptions about the relationship between geotechnical parameters and the liquefaction factor of safety (FS). KNN examines whether local similarity within the feature space is sufficient for FS estimation. ANN evaluates the capability of a flexible non-linear parametric mapping to capture complex input–output relationships. Random Forest assesses the effectiveness of variance-reduced bagging of decision trees, while XGBoost investigates the performance of regularized sequential boosting. The proposed Hybrid Stacking Ensemble (HSE) model integrates multiple ensemble learning algorithms within a meta-learning framework, testing whether combining diverse predictive perspectives can improve accuracy and robustness beyond what is achievable by individual models. This modeling framework enables a comprehensive evaluation of which machine learning paradigm is most suitable for the nonlinear, interaction-driven, and data-limited characteristics of SPT-based liquefaction assessment.

- K-Nearest Neighbors (KNN)

KNN is a non-parametric, instance-based algorithm used for both regression and classification. In this study, KNN estimates FS from observations with similar geotechnical characteristics. Its performance depends on the number of neighbors and the distance metric: small values of k may overfit local patterns, whereas large values may smooth important variability [13,36,37]. The regression estimate is expressed as

$$\widehat{FS}(x) = \frac{1}{k} \sum_{i \in \mathcal{N}_k(x)} F S_i \quad (5)$$

where $\mathcal{N}_k(x)$ denotes the K-Nearest Neighbors with respect to Euclidean distance.

- Hybrid Stacking Ensemble (HSE)

A Hybrid Stacking Ensemble (HSE) regression model was developed to capture the complex nonlinear relationships between geotechnical variables and the liquefaction factor of safety. Unlike single-model approaches, stacking combines the strengths of multiple learning algorithms through a hierarchical architecture. In this study, Random Forest (RF), Extra Trees (ET), and Gradient Boosting Regressors (GBRs) were employed as base learners, while RidgeCV served as the meta-learner responsible for generating the final prediction.

The base learners independently learn different aspects of the input–output relationships. Random Forest reduces prediction variance through bootstrap aggregation, Extra Trees introduces additional randomness to improve model diversity, and Gradient Boosting Regressors sequentially correct prediction errors to reduce bias. Their predictions are subsequently combined by the RidgeCV meta-model, which determines the optimal weighted contribution of each learner through regularized regression and internal cross-validation. This architecture improves predictive robustness, reduces sensitivity to noise, and enhances generalization performance by leveraging complementary learning mechanisms.

- Artificial Neural Network (ANN)

ANNs are computational models inspired by interconnected biological neurons and are well-suited for approximating nonlinear relationships. The multilayer perceptron used here consists of an input layer, hidden layers, and an output layer. Model weights were adjusted through backpropagation to minimize the difference between observed and predicted FS values [13,38–40]. The general mathematical formulation of the feedforward MLP is given by

$$\widehat{FS} = f(W_2 \cdot f(W_1 X + b_1) + b_2) \quad (6)$$

where W_1 and W_2 are the weight matrices of the hidden and output layers, respectively, b_1 and b_2 are the corresponding bias vectors, X is the input feature vector, and f denotes the activation function. ReLU activation was applied in the hidden layers.

- Random Forest (RF)

Random Forest is an ensemble learning method that combines multiple decision trees trained on bootstrap samples and random feature selection. For regression, predictions are obtained by averaging the outputs of individual trees, which reduces variance and limits overfitting relative to a single decision tree. Random Forest is robust for nonlinear problems, but it can require greater computational effort during prediction [41]. The ensemble prediction is

$$\widehat{FS} = \frac{1}{T} \sum_{t=1}^T f_t(X) \quad (7)$$

where each tree is trained on bootstrapped samples with random feature selection.

- XGBoost

XGBoost is a gradient-boosting ensemble algorithm that builds trees sequentially to minimize a regularized objective function. It is suitable for FS prediction because it can model nonlinear interactions among geotechnical variables, handle missing values, and control overfitting through regularization. However, its performance depends on careful hyperparameter selection [42,43]. The objective function consists of a loss term and a tree-complexity penalty:

$$\mathcal{L} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (8)$$

where the first term represents the loss function, and Ω denotes the regularization term controlling tree complexity, which is given by

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (9)$$

The regularization term penalizes overly complex trees and reduces overfitting.

Table 2 presents the hyperparameters adopted for the five machine learning models developed in this study. For the KNN model, the number of neighbors and the weighting scheme were specified to control the influence of neighboring observations during prediction. The ANN model was configured by defining the hidden-layer architecture, activation function, optimization algorithm, regularization strength, and learning rate. The proposed Hybrid Stacking Ensemble model was constructed using Random Forest, Extra Trees, and Gradient Boosting Regressors as base learners, with RidgeCV for the individual models. For the standalone Random Forest model, the number of trees and the maximum tree depth were specified. In contrast, the XGBoost model was configured by selecting the number of estimators, learning rate, maximum depth, and objective function. These hyperparameter settings were chosen to achieve an appropriate balance between predictive accuracy, model robustness, generalization capability, and computational efficiency.

Table 2. Hyperparameters used for training the five machine learning models.

Model	Hyperparameter	Value
KNN	N neighbors	8
KNN	Weights	Uniform
ANN	Hidden-layer size	(100,100)
ANN	Activation function	ReLU
ANN	Solver	Adam
ANN	Alpha	0.0001
ANN	Learning rate	0.001
Hybrid Stacking Ensemble	Base learners	Random Forest, Extra Trees, Gradient Boosting
Hybrid Stacking Ensemble	Meta-learner	RidgeCV
Hybrid Stacking Ensemble	Cross-validation folds (RidgeCV)	Automatic internal cross-validation
Random Forest	Number of estimators	300
Random Forest	Maximum depth	10
XGBoost	Number of estimators	300
XGBoost	Learning rate	0.05
XGBoost	Maximum depth	3
XGBoost	Objective function	Squared-error regression

3.6. Evaluation Metrics

Model performance was evaluated to test three linked requirements: predictive accuracy, generalization stability, and engineering reliability near the FS = 1 threshold. RMSE and MAE quantify error magnitude, and R^2 measures explained variance. These metrics were interpreted alongside learning curves, cross-validation, and residual diagnostics; therefore, model selection was based not only on numerical accuracy but also on robustness and the absence of a systematic unsafe bias.

Sensitivity to training set size was evaluated indirectly using learning curves, which show consistent improvement as data volume increases. While the high performance suggests a strong predictive capability, the possibility of mild overfitting cannot be entirely ruled out, given the dataset size.

- Root Mean Square Error (RMSE)

RMSE measures the standard deviation of prediction errors and is expressed in the same units as FS. Because errors are squared before averaging, RMSE penalizes large deviations more strongly than small deviations. Lower RMSE values indicate better predictive performance:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (10)$$

where y_i is the observed value, $\hat{y}_i$ is the predicted value, and n is the number of observations.

- Mean Absolute Error (MAE)

MAE provides the average absolute difference between observed and predicted FS values. It is less sensitive to outliers than RMSE and provides a direct measure of typical prediction error, which is useful for evaluating performance near engineering thresholds [44].

$$MAE = \left(\frac{1}{n}\right) \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

- Coefficient of Determination (R^2)

R^2 quantifies the proportion of variance in FS explained by the model. Values close to 1 indicate strong agreement between observed and predicted values, whereas values near 0 indicate performance comparable to predicting the sample mean.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (12)$$

where $\bar{y}$ is the mean of the observed values.

- Model Interpretability

SHAP (SHapley Additive exPlanations) was applied as the main explanation method because a global feature-importance ranking alone does not explain how an individual prediction is formed. SHAP decomposes each model prediction into a baseline model output and additive feature contributions; positive SHAP values increase the predicted FS, whereas negative values reduce it. This provides both global behavior and local explanations for decisions at individual borehole intervals [45,46]. The SHAP value for a feature is:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f(S \cup \{j\}) - f(S)] \quad (13)$$

This value represents the marginal contribution of each feature to the model prediction across all possible feature coalitions.

In this study, global explanation was based on mean absolute SHAP values and beeswarm/dependence plots. In contrast, the local explanation was based on the sign and magnitude of feature contributions for individual low- and high-FS intervals. Conventional tree feature importance was retained only as a preliminary screening tool and was not treated as sufficient evidence of explainable AI [47].

4. Results and Discussion

4.1. Model Performance Comparison

Model performance analysis in this section is interpreted in relation to the scientific question of whether nonlinear interactions among resistance, density, depth, and stress-state variables govern FS. Therefore, the comparison among algorithms is used to test model adequacy for the geotechnical mechanism, not as a standalone ranking exercise.

The final modeling dataset contains 534 observations from 61 boreholes in the Port Sudan coastal plain. Computed FS values range from 0.14 to 1.98, with a mean of 0.90, a standard deviation of 0.49, a median of 0.91, and an interquartile range of 0.86. Approximately 55.1% of the observations have $FS < 1.0$, indicating potentially liquefiable conditions under the adopted seismic scenario. The FS distribution is nearly symmetric to slightly right-skewed, indicating a balanced representation of liquefiable and relatively stable layers. Depth-stratified analysis indicates that the lowest FS values occur mainly within the upper 15–20 m. The upper 15 m has an average FS of approximately 0.63, and 75.4% of these intervals have $FS < 1.0$. Nearly 65% of the lowest-FS quartile occurs within the upper 15 m, confirming that liquefaction susceptibility is concentrated in shallow saturated layers with low penetration resistance and limited confinement.

The scatter plots compare test MAE, R^2 , and RMSE for the five ML models and two baseline regression models based solely on the corrected SPT resistance, $N1(60)_{cs}$. The ranking of the machine learning models remains highly consistent across all three performance metrics. The Hybrid Stacking Ensemble gives the lowest RMSE (0.058), the highest R^2 (0.985), and shares the lowest MAE (0.032) with XGBoost. XGBoost performs similarly well, with $RMSE = 0.061$, $R^2 = 0.984$, and $MAE = 0.032$, followed by Random Forest ($MAE = 0.037$, $R^2 = 0.977$, $RMSE = 0.072$). ANN shows moderate performance, whereas KNN produces slightly larger errors.

The two baseline regression models demonstrated substantially lower predictive capability than all machine learning models. Although $N1(60)_{cs}$ is recognized as the most influential predictor of liquefaction resistance, relying on this parameter alone was insufficient to capture the complex nonlinear interactions among geotechnical and stress-state variables governing the liquefaction factor of safety. These findings demonstrate the advantage of multivariable ensemble learning, with the Hybrid Stacking Ensemble providing the most accurate and robust prediction of FS.

Figure 6 compares observed and predicted FS values for the five ML models. The dashed diagonal line represents perfect prediction. All models reproduce the general upward trend, but the degree of agreement varies. The Hybrid Stacking Ensemble shows the closest correspondence with the observations, with a test R^2 of 0.985 and predictions concentrated near the 1:1 line. XGBoost also performs strongly, with $R^2 = 0.983$ and limited scatter. Random Forest achieves excellent predictive accuracy, with $R^2 = 0.977$ and predictions closely following the ideal line. ANN and KNN provide acceptable but more dispersed predictions, particularly at higher FS values. Overall, the ensemble models, especially the Hybrid Stacking Ensemble, XGBoost, and Random Forest, best capture the nonlinear relationships among soil resistance, stress-state variables, and FS.

Table 3 summarizes the training and testing performance of the five machine learning models and compares test MAE, R^2 , and RMSE for the five ML models and two baseline regression models based solely on the corrected SPT resistance, $N1(60)_{cs}$. The ranking of

the machine learning models remains highly consistent across all three performance metrics. The Hybrid Stacking Ensemble gives the lowest RMSE (0.058), the highest R^2 (0.985), and shares the lowest MAE (0.032) with XGBoost. XGBoost performs similarly well, with RMSE = 0.061, R^2 = 0.984, and MAE = 0.032, followed by Random Forest (MAE = 0.037, R^2 = 0.977, RMSE = 0.072). ANN shows moderate performance, whereas KNN produces slightly larger errors. Overall, the Hybrid Stacking Ensemble demonstrates the strongest predictive capability among the evaluated models.

Table 3. Training and testing performance metrics for all machine learning models used to predict FS.

Model	Hybrid Stacking Ensemble	XGBoost	Random Forest	ANN	KNN	Linear Regression N1(60)cs Only	Polynomial Regression N1(60)cs Only
Train RMSE	0.019	0.018	0.030	0.045	0.120	0.149	0.140
Train MAE	0.014	0.013	0.017	0.027	0.075	0.111	0.110
Train R^2	0.998	0.999	0.996	0.992	0.941	0.908	0.918
Test RMSE	0.058	0.061	0.072	0.125	0.133	0.194	0.269
Test MAE	0.032	0.032	0.037	0.057	0.089	0.115	0.119
Test R^2	0.985	0.984	0.977	0.931	0.923	0.834	0.682

The two baseline regression models demonstrated substantially lower predictive capability than all machine learning models. Although corrected SPT resistance is recognized as the most influential predictor of liquefaction resistance, these results indicate that using this parameter alone cannot adequately characterize the complex interactions among geotechnical and seismic variables governing liquefaction behavior.

The consistency among MAE, RMSE, and R^2 confirms that the Hybrid Stacking Ensemble is the most reliable model for FS prediction, followed closely by XGBoost and Random Forest. The results also indicate that ensemble learning models are better suited to this liquefaction prediction problem than ANN and KNN, owing to their superior ability to capture the dataset's complex nonlinear relationships.

Figure 6 supports the numerical results presented in Table 3. Hybrid Stacking Ensemble, XGBoost, and Random Forest predictions cluster tightly around the 1:1 line, whereas ANN and KNN show greater dispersion. This ranking supports the working hypothesis that direct FS prediction in Port Sudan requires models capable of representing nonlinear interactions and conditional effects, particularly among N1(60)cs, Dr, depth, and stress-state variables. Consequently, the Hybrid Stacking Ensemble and XGBoost were selected for geotechnical interpretation using SHAP.

Figures 7–9 compare test MAE, R^2 , and RMSE for the five ML models and two baseline regression models based solely on the corrected SPT resistance, N1(60)cs. The ranking of the machine learning models remains highly consistent across all three performance metrics. The Hybrid Stacking Ensemble gives the lowest RMSE (0.058), the highest R^2 (0.985), and shares the lowest MAE (0.032) with XGBoost. XGBoost performs similarly well, with RMSE = 0.061, R^2 = 0.984, and MAE = 0.032, followed by Random Forest (MAE = 0.037, R^2 = 0.977, RMSE = 0.072). ANN shows moderate performance, whereas KNN produces slightly larger errors. Overall, the Hybrid Stacking Ensemble demonstrates the strongest predictive capability among the evaluated models.

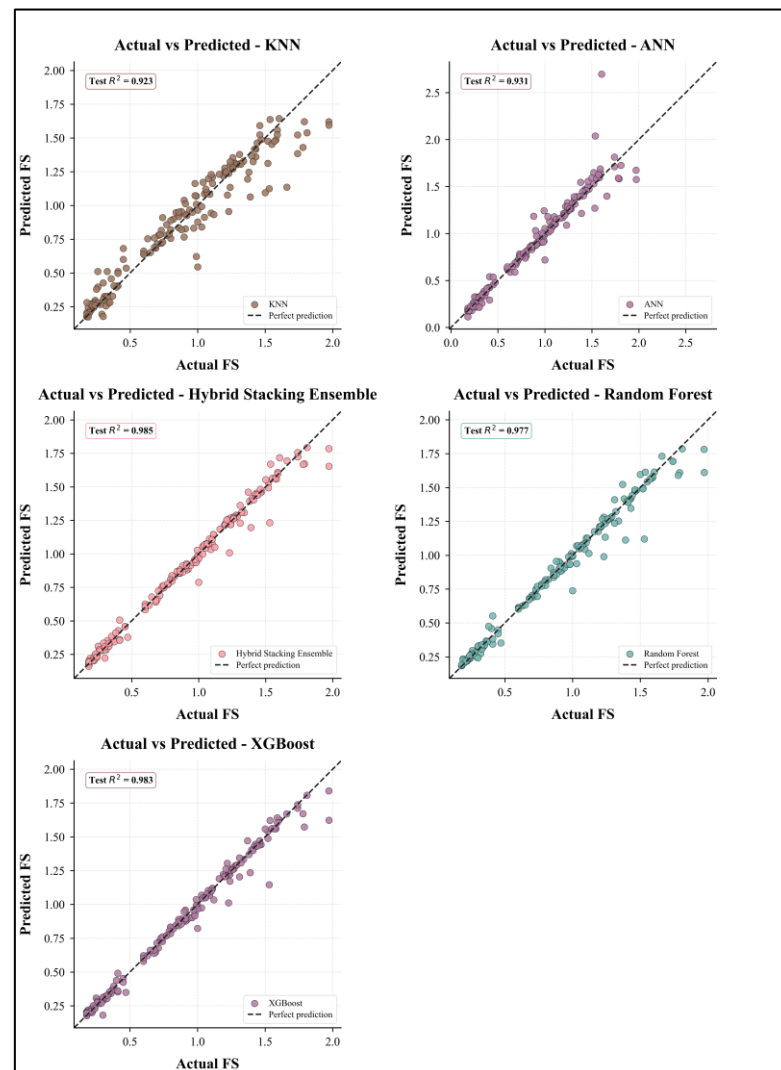


Figure 6. Actual versus predicted liquefaction factor of safety (FS) for the five machine learning models on the testing dataset.

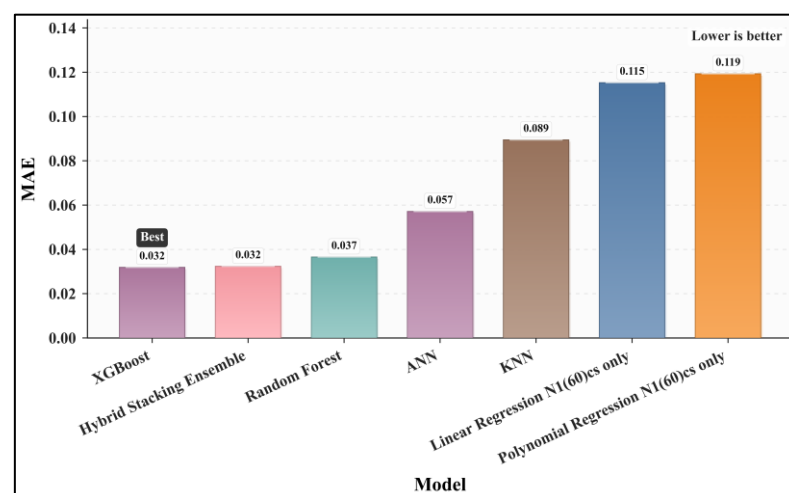


Figure 7. Comparison of test MAE values for the five machine learning models, Linear Regression N1(60)cs only, and Polynomial N1(60)cs only used to predict FS.

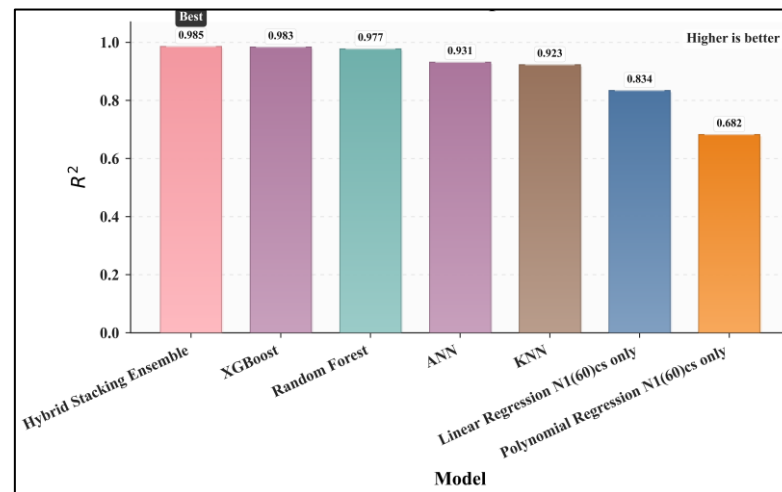


Figure 8. Comparison of test R^2 values for the five machine learning models, Linear Regression N1(60)cs only and Polynomial N1(60)cs only, showing the explained variance of predicted FS.

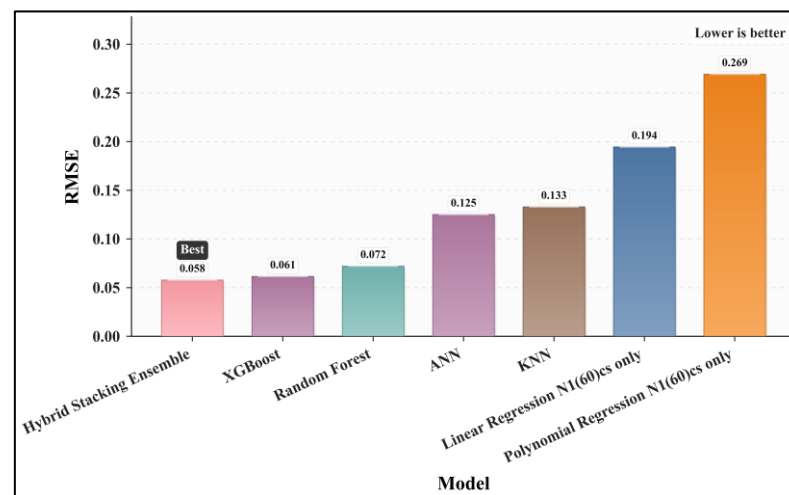


Figure 9. Comparison of test RMSE values for the five machine learning models, Linear Regression N1(60)cs only, and Polynomial N1(60)cs only used to evaluate prediction error magnitude.

4.2. Cross-Validation and Model Robustness

Figure 10 shows the RMSE distribution obtained from 5-fold cross-validation for all evaluated models. The Hybrid Stacking Ensemble consistently achieves low cross-validation RMSE, demonstrating strong predictive stability across different data partitions. XGBoost also shows low and stable RMSE values, followed closely by Random Forest. KNN has higher RMSE values, while ANN displays the largest variability and a high-error outlier, indicating greater sensitivity to data partitioning. The baseline linear regression and polynomial regression models, based solely on N1(60)cs, produce substantially higher RMSE values, highlighting the limitations of single-variable regression for liquefaction prediction. The cross-validation results, therefore, confirm that the Hybrid Stacking Ensemble provides the most stable and reliable performance among all evaluated models.

Figures 11 and 12 present learning curves based on R^2 and RMSE. In general, model performance improves as the training set size increases, indicating that additional data enhances stability and generalization.

As shown in Figure 11, R^2 increases with training set size for most models. The Hybrid Stacking Ensemble maintains the highest and most stable validation performance, with training and validation curves converging rapidly as sample size increases. XGBoost and Random Forest also exhibit strong and stable validation performance. KNN improves

gradually, whereas ANN shows a larger gap between training and validation curves, suggesting reduced generalization capability.

Figure 12 shows a decreasing RMSE with increasing training set size. The Hybrid Stacking Ensemble achieves the lowest validation RMSE across most of the learning process, closely followed by XGBoost and Random Forest, confirming their superior predictive accuracy. KNN improves with additional data, while ANN retains comparatively higher validation errors. These learning curves support the testing results and confirm the robustness of the ensemble models.

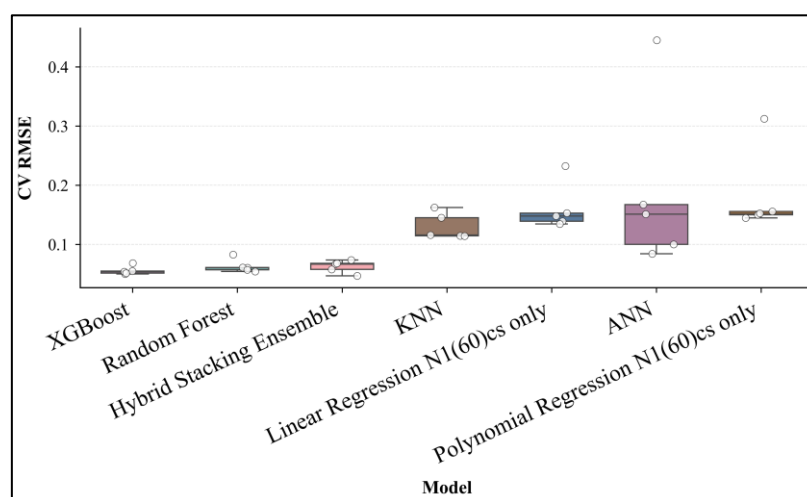


Figure 10. Five-fold cross-validation RMSE distribution for the five machine learning models, Linear Regression N1(60)cs only and Polynomial N1(60)cs only, illustrating model robustness and generalization stability.

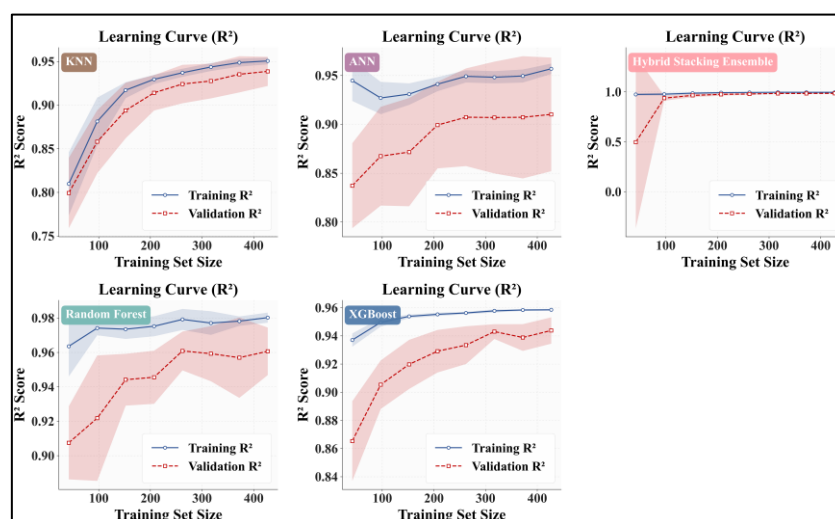


Figure 11. Learning curves based on R^2 for the five machine learning models at different training set sizes. The shaded regions represent ± 1 standard deviation across the cross-validation folds, indicating the variability of model performance at each training set size.

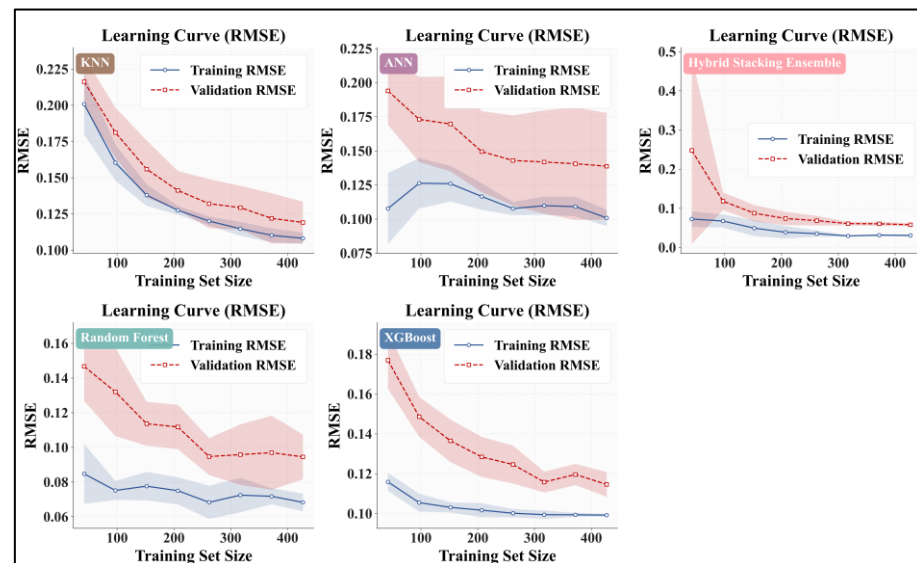


Figure 12. Learning curves based on RMSE for the five machine learning models at different training set sizes. The shaded regions represent ± 1 standard deviation across the cross-validation folds, indicating the variability of model performance at each training set size.

4.3. Residual Error Analysis

Figure 13 presents residuals for the testing dataset. Residuals from the Hybrid Stacking Ensemble, XGBoost, and Random Forest are centered close to zero, indicating limited bias and strong predictive performance. The Hybrid Stacking Ensemble exhibits the most compact residual distribution, corresponding to the smallest prediction errors, while XGBoost and Random Forest also show relatively narrow residual spreads.

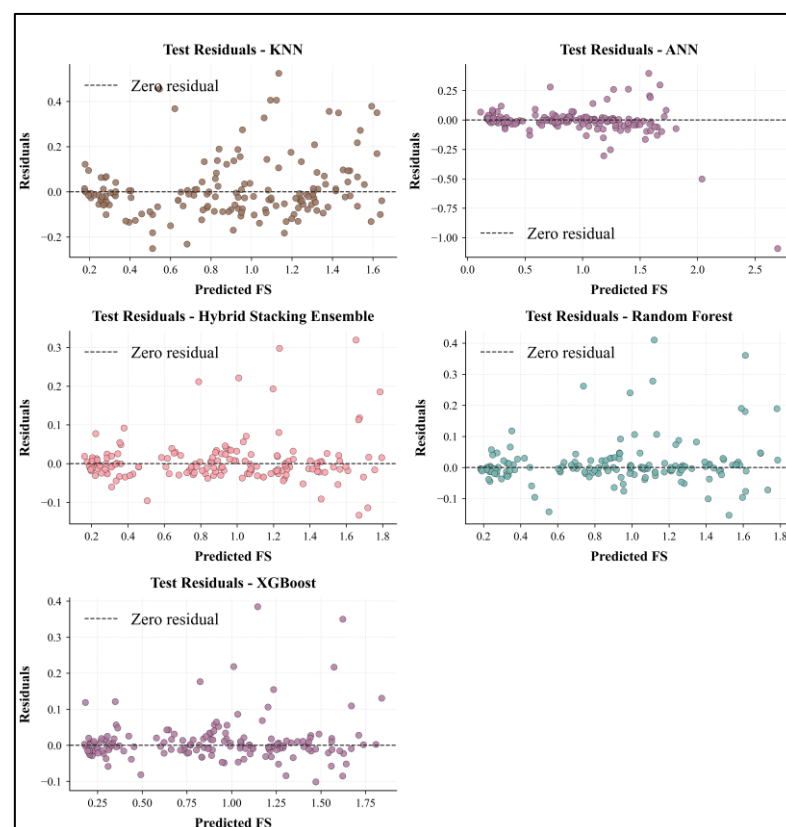


Figure 13. Residual distribution of testing predictions for the five machine learning models, showing prediction errors relative to the zero-residual line.

ANN and KNN display wider residual distributions, indicating lower prediction stability. The ANN model includes several outliers at high predicted FS values, while KNN shows greater scatter across the medium-to-high FS range. The residual analysis, therefore, confirms the performance ranking obtained from the testing results, with the Hybrid Stacking Ensemble providing the most accurate predictions, followed closely by XGBoost and Random Forest.

4.4. Comparison with Previous Studies

In comparison to published materials, this study, based on applied soil liquefaction research, examines the effectiveness of an ensemble learning model with data leakage for direct safety prediction and proposes geotechnically interpretable model-based decision-making approaches. Traditional CSR-CRR approaches estimate FS indirectly through empirical relationships and correction factors [7,48]. These factors include overburden corrections, magnitude-scaling factors, stress-reduction coefficients, and fines-content adjustments, which are derived from simplified assumptions and case-history calibration [1,7]. Although effective in engineering practice, these formulations may not fully represent the nonlinear coupling among density, stress conditions, and cyclic loading.

By contrast, the ML models developed here predict FS directly from independent geotechnical and stress-state variables, without using intermediate parameters such as CSR and CRR. This represents a fundamental methodological difference and reduces the risk of target leakage. Rather than decomposing liquefaction assessment into sequential empirical components, the ML framework learns the combined influence of the input variables. The high accuracy of XGBoost and Random Forest (R^2 up to 0.983 and RMSE as low as 0.061) indicates that these models can reproduce FS variability across heterogeneous subsurface conditions [41].

Conventional procedures also propagate uncertainty through multiple empirical correction factors. In complex depositional settings, such as the Port Sudan coastal plain, spatial heterogeneity, variable fines content, and stratigraphic interlayering may amplify these uncertainties [1,49]. The proposed ML framework reduces reliance on intermediate empirical parameters by mapping independent inputs directly to FS. This approach is especially useful where variable relationships are nonlinear and site-specific.

Interpretability is another important distinction. Conventional methods are transparent because their equations are explicit, although their interactions are restricted to predefined correction factors. ML models are often criticized as black boxes; therefore, this study does not equate global importance with explainability. Instead, SHAP is used to decompose each ensemble prediction into feature-level contributions, allowing the analysis to identify both the dominant global controls and the reasons for low-FS predictions in specific soil intervals [17,19].

To evaluate the performance and scientific contribution of the proposed leakage-aware framework, the results were compared with representative machine learning studies reported in the liquefaction literature. Table 4 summarizes the characteristics and predictive performance of selected previous investigations and the present study.

Table 4. Comparison of representative machine learning studies for liquefaction prediction and the proposed leakage-aware framework.

Study	Objective	Approach	Best Model	Performance
Alobaidi et al. (2019) [15]	Prediction of liquefaction occurrence	Ensemble learning classification framework using historical liquefaction case histories	Ensemble models	Classification-based prediction with improved accuracy over conventional approaches

Kurnaz et al. (2023) [50]	Direct prediction of liquefaction FS	1249 SPT-based liquefaction analyses; ANN, DT, RF, and SVR models	ANN	$R^2 = 0.99$
Ghani et al. (2025) [51]	Liquefaction assessment using ML techniques	Comparative evaluation of multiple AI and ML methods	RF/XGBoost	Demonstrated strong predictive capability for liquefaction assessment
Present Study	Direct prediction of liquefaction FS with a leakage-aware workflow	534 observations from 61 boreholes in Port Sudan; physically screened predictors; SHAP interpretation	Hybrid Stacking Ensemble	$R^2 = 0.985$, RMSE = 0.058, MAE = 0.032

Recent studies have demonstrated that ensemble learning algorithms, particularly XGBoost and Random Forest, provide excellent predictive performance for liquefaction assessment because of their ability to capture complex nonlinear interactions among geotechnical variables. For example, hyper parameterized machine-learning algorithms for direct FS prediction were employed by [50]. Ghani reported the strong performance of Random Forest and XGBoost for liquefaction assessment [51], while Jas and Dodagoudar successfully combined XGBoost with SHAP to improve model interpretability [19]. Consistent with these findings, the present study confirms the effectiveness of tree-based ensemble learning. Furthermore, the proposed Hybrid Stacking Ensemble achieved slightly improved predictive performance by integrating complementary ensemble learners through a RidgeCV meta-learner while maintaining a physically meaningful interpretation using SHAP. These results demonstrate that combining multiple ensemble models can further enhance robustness and predictive accuracy for direct liquefaction factor-of-safety prediction.

4.5. Engineering Interpretation of Model Results

The model results indicate that soil resistance, density, depth, and stress-state conditions mainly control FS in the Port Sudan coastal deposits. The superior performance of XGBoost and Random Forest suggests that the relationship between these variables and FS is strongly nonlinear.

The behavior learned by the ensemble models is geotechnically consistent. Higher SPT-N, N₁(60)_{cs}, and relative density generally indicate denser sand with greater cyclic resistance and higher FS. Conversely, loose saturated sands with low penetration resistance are more susceptible to excess pore-pressure generation and, therefore, have lower FS. Depth and total vertical stress also influence FS through confinement, although this effect is modified by rd and $K\sigma$, which account for depth-dependent seismic-stress reduction and overburden correction. These patterns show that the best-performing models learned physically meaningful relationships and can be used as reliable decision-support tools for liquefaction assessment in Port Sudan.

4.6. Tree-Based Feature Importance as Preliminary Screening

Tree-based feature importance was first used as a preliminary global screening step for the XGBoost, Random Forest, and Hybrid Stacking Ensemble (Figures 14–16). These importance scores summarize how often and how effectively variables contribute to the prediction process, but they do not by themselves explain why a specific prediction is high or low. The results consistently identify N₁(60)_{cs} as the strongest global predictor of FS across all three models. In the Hybrid Stacking Ensemble, N₁(60)_{cs} has the highest importance value (0.232), followed by total stress (0.139) and depth (0.138). In XGBoost, N₁(60)_{cs} is also the most influential variable (0.485), followed by relative density ($Dr = 0.416$). Similarly, Random Forest identifies N₁(60)_{cs} as the dominant predictor, with an importance value of 0.846, followed by Dr (0.100).

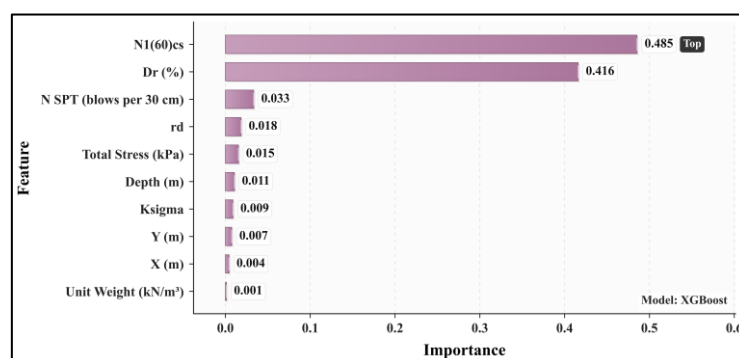


Figure 14. Feature-importance ranking of input variables in the XGBoost model for predicting FS.

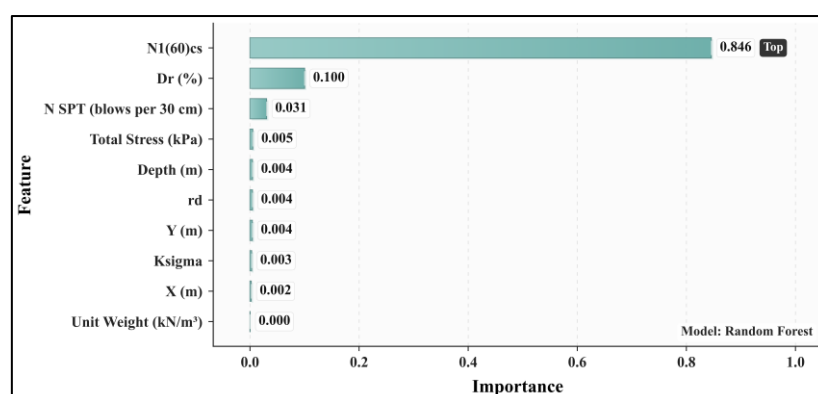


Figure 15. Feature-importance ranking of input variables in the Random Forest model for predicting FS.

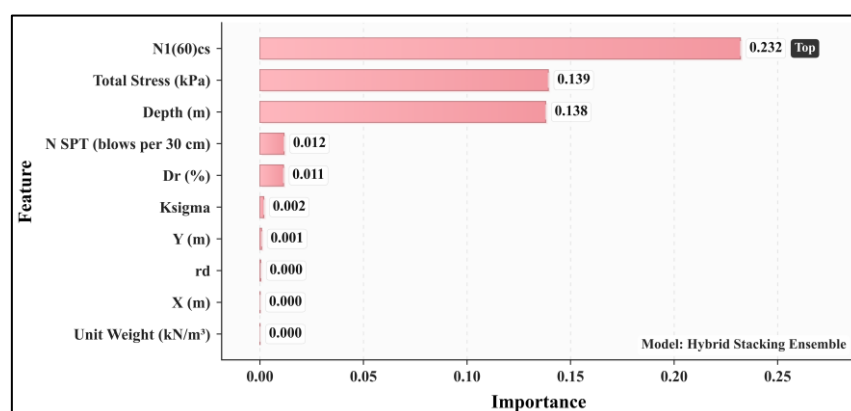


Figure 16. Feature-importance ranking of input variables in the Hybrid Stacking Ensemble model for predicting FS.

This preliminary ranking is geotechnically reasonable because N1(60)cs represents clean-sand-corrected SPT resistance, which is closely related to soil density and cyclic resistance. The Hybrid Stacking Ensemble additionally assigns relatively greater importance to total stress and depth, reflecting their contribution when predictions from multiple base learners are combined. However, these feature-importance values are interpreted only as global screening evidence. The decision-level explanation is provided in the following SHAP analysis, which indicates the direction, magnitude, and case-specific contribution of each variable to the predicted FS.

4.7. SHAP-Based Global and Local Model Explanation

SHAP analysis was used to explain the predictions of the three best-performing ensemble models: the Hybrid Stacking Ensemble, XGBoost, and Random Forest. Unlike tree-

based importance scores, SHAP quantifies how each variable changes each predicted FS relative to the model baseline. We emphasize that SHAP is employed strictly for decision explanation and hypothesis generation, not for causal inference; the observed feature contributions are interpreted within the context of established liquefaction mechanics rather than as standalone evidence of physical causality. The global mean absolute SHAP plots show that $N1(60)_{cs}$ is the dominant control on predicted FS in all three models. Mean absolute SHAP values are approximately 0.26 for the Hybrid Stacking Ensemble, 0.35 for XGBoost, and 0.33 for Random Forest, confirming that clean-sand-corrected SPT resistance provides the largest average contribution to model decisions (Figure 17).

The SHAP beeswarm plots clarify the direction of the model response. Low $N1(60)_{cs}$ values produce negative SHAP contributions and reduce predicted FS, whereas high $N1(60)_{cs}$ values produce positive contributions and increase predicted FS. This response agrees with liquefaction mechanics because loose sands with low corrected penetration resistance are more susceptible to pore-pressure buildup, whereas denser sands have greater cyclic resistance (Figure 18).

SPT-N, depth, and relative density have secondary effects on model output. Higher SPT-N values generally increase predicted FS, while depth contributes through changes in confinement and stress state. Spatial coordinates and unit weight have very small effects, indicating that the predictions are controlled mainly by soil-resistance parameters rather than spatial location alone (Figure 19).

The SHAP dependence plots for $N1(60)_{cs}$ show a nonlinear positive relationship in all three models. As $N1(60)_{cs}$ increases, its SHAP contribution shifts from negative to positive, indicating that higher corrected SPT resistance increases the predicted safety against liquefaction. Likewise, the dependence plots for depth reveal a nonlinear influence on predicted FS, reflecting the combined effects of confinement and stress-state conditions. Overall, the SHAP results confirm that the Hybrid Stacking Ensemble, XGBoost, and Random Forest learned physically meaningful relationships that are consistent with geotechnical principles and support their use for liquefaction assessment in the Port Sudan coastal plain [52,53].

At the local prediction level, the SHAP framework explains the reasons why a given soil interval is predicted to have either low or high FS. Low-FS predictions are driven by negative contributions from low $N1(60)_{cs}$, low SPT-N, low relative density, and limited confinement, which dominate any positive contributions from depth-related stress-state variables. In contrast, high-FS predictions are linked to positive contributions from higher corrected penetration resistance and denser soil conditions. Thus, SHAP explains not only how the models utilize each input variable but also why certain soil layers are assigned low- or high-FS estimates. This provides a more informative and engineer-relevant interpretation than global feature-importance rankings alone.

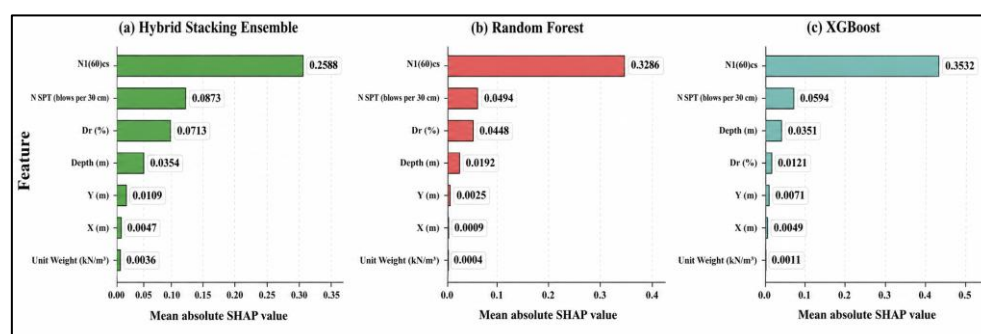


Figure 17. SHAP feature importance for the Hybrid Stacking Ensemble, Random Forest, and XGBoost models, showing the dominant contribution of $N1(60)_{cs}$ to predicted FS.

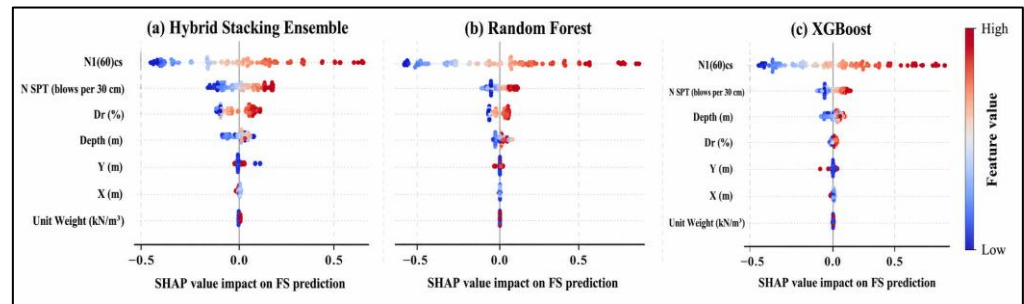


Figure 18. SHAP beeswarm plots for the Hybrid Stacking Ensemble, Random Forest, and XGBoost models, illustrating the direction and magnitude of effects on the predicted FS.

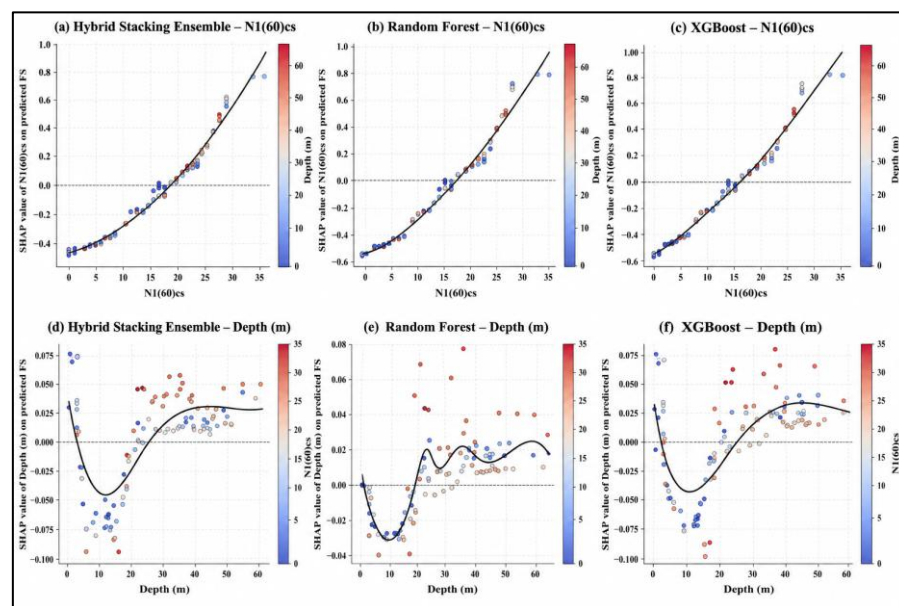


Figure 19. SHAP dependence plots for N1(60)cs and depth, showing the nonlinear influence of corrected SPT resistance and the secondary effect of depth on predicted FS.

4.8. Liquefaction Susceptibility Implication for Port Sudan

The results have direct implications for liquefaction susceptibility assessment in the Port Sudan coastal plain. Low predicted FS values are primarily associated with loose-to-medium-dense saturated sand layers, particularly where groundwater is shallow, and the penetration resistance is low. These conditions are common in coastal and reclaimed deposits, making near-surface layers vulnerable during seismic loading. The strong influence of N1(60)cs, SPT-N, relative density, and depth, identified by the Hybrid Stacking Ensemble, XGBoost, and Random Forest, confirms that soil resistance and density are the main controls on liquefaction safety. Layers with low corrected SPT resistance and low relative density are expected to have lower FS, whereas denser layers at greater depth generally show higher resistance. This interpretation is consistent with the geological setting of Port Sudan, where young marine and alluvial sediments include silty sand and loose sandy deposits.

From an engineering perspective, the Hybrid Stacking Ensemble, XGBoost, and Random Forest can support preliminary liquefaction-hazard screening by identifying zones with low predicted FS. The superior predictive performance of the Hybrid Stacking Ensemble makes it particularly suitable for rapid decision support in port facilities, fuel storage areas, pipelines, transportation corridors, and shallow foundations, where liquefaction-induced settlement or lateral spreading may cause substantial damage. The predictions should nevertheless be used as decision-support information rather than as a substitute for

detailed geotechnical investigations. Zones with low predicted FS should be prioritized for additional site investigation, ground improvement, or liquefaction-mitigation measures.

Several limitations should be acknowledged. The dataset is region-specific; CPT-based parameters were not included, and a single design earthquake scenario was considered. Future work should incorporate multi-scenario seismic loading, probabilistic FS frameworks, CPT and shear-wave velocity data, and numerical cyclic constitutive modeling [54–57].

External validation using independent datasets from different geological regions and depositional environments would provide stronger evidence of the model's transferability and generalization. However, such datasets were not available within the scope of the present study and therefore remain an important direction for future research. Although the proposed Hybrid Stacking Ensemble demonstrated excellent predictive performance for the Port Sudan coastal sediments, future research should validate the model using independent datasets from different geographical regions and geological conditions to further assess its generalization capability and practical applicability. This limitation is common to all data-driven geotechnical models and underscores the intended use of the proposed framework as a preliminary hazard-screening tool rather than a definitive substitute for detailed site investigations.

In addition, recent generalizations of SHAP, such as ShapleyFlow, could be explored in future work to model spatial or stratigraphic flow dependencies in liquefaction predictions, potentially providing richer decision explanations beyond the static feature contributions examined here.

5. Conclusions

This study developed a novel leakage-aware Hybrid Stacking Ensemble (HSE) framework for the direct prediction and geotechnical interpretation of the liquefaction factor of safety (FS) in the Port Sudan coastal plain. The proposed HSE model combines Random Forest, Extra Trees, and Gradient Boosting Regressors as complementary base learners with RidgeCV as a meta-learner, enabling more effective learning of the complex nonlinear relationships among soil resistance, density, depth, and stress-state variables. By excluding intermediate liquefaction-response variables from the predictor set, the proposed framework minimizes data leakage and improves the reliability of model evaluation.

Among the five evaluated machine learning models, the proposed Hybrid Stacking Ensemble achieved the best predictive performance, with a test RMSE of 0.058, MAE of 0.032, and R^2 of 0.985, outperforming XGBoost, Random Forest, ANN, and KNN. The consistency of the testing, cross-validation, learning curve, and residual analyses demonstrates that the HSE model provides accurate, robust, and well-generalized predictions without evidence of significant overfitting.

An important contribution of this study is the integration of high predictive accuracy with transparent model interpretation. Tree-based feature importance was employed as an initial screening tool, while SHAP analysis provided both global and local explanations of model decisions. The interpretation consistently identified $N_{1(60)}$ cs as the dominant controlling factor, followed by relative density, depth, and stress-related variables. These findings agree with established liquefaction mechanics and demonstrate that the proposed HSE model learned physically meaningful geotechnical relationships rather than purely statistical correlations.

From an engineering perspective, the proposed framework provides an efficient decision-support tool for preliminary liquefaction-hazard screening in coastal environments. The ability to directly estimate FS from independent geotechnical variables enables rapid identification of potentially liquefiable zones, supporting infrastructure planning,

site investigation prioritization, and ground-improvement decisions in data-limited regions such as the Port Sudan coastal plain.

Overall, this study demonstrates that combining a new Hybrid Stacking Ensemble architecture, leakage-aware preprocessing, and SHAP-based interpretation provides a reproducible, physically interpretable, and highly accurate framework for direct liquefaction factor-of-safety prediction. Although the developed model is not intended to replace detailed geotechnical investigations, it represents a valuable complement to conventional assessment procedures and advances the application of explainable ensemble learning in geotechnical engineering.

Future work should validate the proposed framework using independent datasets from different geological settings, incorporate CPT and shear-wave velocity measurements, evaluate multiple seismic-loading scenarios, and investigate domain-adaptation strategies to further improve model transferability and practical applicability.

Author Contributions: Conceptualization, A.A.H.S.H.; methodology, A.A.H.S.H. and D.W.; software, A.A.H.S.H. and W.Y.; validation, W.W.; formal analysis, A.A.H.S.H. and W.Y.; investigation and data curation, D.W. and H.E.A.; visualization, A.A.H.S.H.; writing—original draft preparation, A.A.H.S.H.; writing—review and editing, W.W., D.W., W.Y., and M.A.; supervision and funding acquisition, W.W. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant numbers 52538008, 52578446, and 52578436.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset supporting the findings of this study is available from the corresponding author upon reasonable request, subject to institutional and regulatory constraints.

Acknowledgments: The authors gratefully acknowledge the Sea Ports Corporation, Port Sudan, Sudan, for its logistical support and assistance during field investigations and data collection. The authors also thank the China University of Geosciences (Wuhan) for providing the academic and technical environment that supported this research. We extend our thanks for the support provided through the Chinese Government Scholarship Council (CSC) during this study.

Conflicts of Interest: Author Husam Eldin Ahmed was employed by the company Sea Ports Corporation. Author Jingang Lü was employed by the company Wuhan Municipal Engineering Design & Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
CPT	Cone Penetration Test
CRR	Cyclic Resistance Ratio
CRR7.5	Cyclic Resistance Ratio normalized to Mw 7.5
CSR	Cyclic Stress Ratio
Dr	Relative Density
FS	Factor of Safety against Liquefaction
HSE	Hybrid Stacking Ensemble
KNN	K-Nearest Neighbors
$K\sigma$	Overburden Correction Factor
LPI	Liquefaction Potential Index

MAE	Mean Absolute Error
ML	Machine Learning
N-SPT	Standard Penetration Test Blow Count
N1(60)	Energy-corrected SPT resistance
N1(60) _{cs}	Clean-sand-corrected SPT resistance
R ²	Coefficient of Determination
RBF	Radial Basis Function
RF	Random Forest
RMSE	Root Mean Square Error
RSS	Residual Sum of Squares
SHAP	SHapley Additive exPlanations
SPT	Standard Penetration Test
SVM	Support Vector Machine
XGBoost	Extreme Gradient Boosting

References

1. Cetin, K.O.; Seed, R.B.; Der Kiureghian, A.; Tokimatsu, K.; Harder, L.F., Jr.; Kayen, R.E.; Moss, R.E.S. Standard penetration test-based probabilistic and deterministic assessment of seismic soil liquefaction potential. *J. Geotech. Geoenviron. Eng.* **2004**, *130*, 1314–1340.
2. Zhang, Y.P.; Zhang, F.Y.; El Naggar, M.H.; Wu, W.B. Seismic response of pile group embedded in unsaturated soil considering the coupling of kinematic and inertia pile-pile interactions. *Comput. Geotech.* **2025**, *178*, 106937.
3. Wu, W.B.; Di, T.Y.; Yang, X.Y.; El Naggar, M.H.; Zhang, Y.P. Revised pile-pile mutual interaction factors for seismic analysis of end-bearing pile groups embedded in saturated stratum. *Soil Dyn. Earthq. Eng.* **2024**, *182*, 108745.
4. Kavazanjian, E.; Andrade, J.E.; Arulmoli, K.; Atwater, B.F.; Christian, J.T.; Green, R.A.; Kramer, S.L.; Mejia, L.; Mitchell, J.K.; Rathje, E.M.; et al. *State of the Art and Practice in the Assessment of Earthquake-Induced Soil Liquefaction and Its Consequences*; National Academies Press: Washington, DC, USA, 2022.
5. Wang, R.; Fu, P.; Zhang, J.M.; Dafalias, Y.F. DEM study of fabric features governing undrained post-liquefaction shear deformation of sand. *Acta Geotech.* **2016**, *11*, 1321–1337.
6. Huang, Y.; Yu, M. Review of soil liquefaction characteristics during major earthquakes of the twenty-first century. *Nat. Hazards* **2013**, *65*, 2375–2384.
7. Idriss, I.M.; Boulanger, R.W. Semi-empirical procedures for evaluating liquefaction potential during earthquakes. *Soil Dyn. Earthq. Eng.* **2006**, *26*, 115–130.
8. Wijewickreme, D.; Soysa, A.; Verma, P. Response of natural fine-grained soils for seismic design practice: A collection of research findings from British Columbia, Canada. *Soil Dyn. Earthq. Eng.* **2019**, *124*, 280–296.
9. Paolucci, R.; Aimar, M.; Ciancimino, A.; Dotti, M.; Foti, S.; Lanzano, G.; Mattevi, P.; Pacor, F.; Vanini, M. Checking the site categorization criteria and amplification factors of the 2021 draft of Eurocode 8 Part 1–1. *Bull. Earthq. Eng.* **2021**, *19*, 4199–4234.
10. Lu, D.C.; Wang, G.S.; Du, X.L.; Wang, Y. A nonlinear dynamic uniaxial strength criterion that considers the ultimate dynamic strength of concrete. *Int. J. Impact Eng.* **2017**, *103*, 124–137.
11. Viggiani, G. *Geotechnical Aspects of Underground Construction in Soft Ground*, 1st ed.; CRC Press: London, UK, 2012.
12. Zhou, J.; Li, E.; Wang, M.; Chen, X.; Shi, X.; Jiang, L. Feasibility of stochastic gradient boosting approach for evaluating seismic liquefaction potential based on SPT and CPT case histories. *J. Perform. Constr. Facil.* **2019**, *33*, 04019024.
13. Onyelowe, K.C.; Kamchoom, V.; Gnananandarao, T.; Arunachalam, K.P. Developing data driven framework to model earthquake induced liquefaction potential of granular terrain by machine learning classification models. *Sci. Rep.* **2025**, *15*, 21509.
14. Kumar, K.; Samui, P.; Choudhary, S.S. State parameter based liquefaction probability evaluation. *Int. J. Geosynth. Ground Eng.* **2023**, *9*, 76.
15. Alobaidi, M.H.; Meguid, M.A.; Chebana, F. Predicting seismic-induced liquefaction through ensemble learning frameworks. *Sci. Rep.* **2019**, *9*, 11786.
16. Ma, C.; Lu, D.C.; Du, X.L.; Qi, C.Z.; Zhang, X.Y. Structural components functionalities and failure mechanism of rectangular underground structures during earthquakes. *Soil Dyn. Earthq. Eng.* **2019**, *119*, 265–280.
17. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67.
18. Yagis, E.; Ozcan, B.; Genc, B. Effect of data leakage in brain MRI classification using 2D convolutional neural networks. *Sci. Rep.* **2021**, *11*, 22452.

19. Jas, K.; Dodagoudar, G.R. Explainable machine learning model for liquefaction potential assessment of soils using XGBoost-SHAP. *Soil Dyn. Earthq. Eng.* **2023**, *165*, 107662.
20. Mohammed, M.A.A.; Daoud, A.M.A.; Kazem, M.M.; Mohammed, S.H.; Szabó, N.P.; Szűcs, P. Geophysical and multi-criteria decision methods for delineating groundwater potential in coastal terrains: A study from Port Sudan. *Sci. Rep.* **2026**, *16*, 5218.
21. Wadi, D.; Fuad, A.; Wu, W.B. Prediction of cone penetration test data from standard penetration test attributes using support vector regression neural networks. *AlQalam J. Med. Appl. Sci.* **2025**, *8*, 896–906.
22. Ghebreab, W. Tectonics of the Red Sea region reassessed. *Earth-Sci. Rev.* **1998**, *45*, 1.
23. Bosworth, W.; Huchon, P.; McClay, K. The red sea and Gulf of Aden basins. *J. Afr. Earth Sci.* **2005**, *43*, 334–378.
24. Di Martino, A.; Sgattoni, G.; Di Paola, G.; Berti, M.; Amorosi, A. Reconstructing late quaternary Paleovalley systems of Italy through mHVSr: A tool for seismic hazard assessment in modern coastal lowlands. *Earth Space Sci.* **2023**, *10*, e2023EA003112.
25. Samui, P.; Sitharam, T.G. Machine learning modelling for predicting soil liquefaction susceptibility. *Nat. Hazards Earth Syst. Sci.* **2011**, *11*, 1–9.
26. Hamid, O.A. Hydro-environmental impact of Khor Arbaat Dam, Eastern Sudan. *Int. J. Water Resour. Environ. Eng.* **2022**, *14*, 1–11.
27. Lu, D.C.; Meng, F.P.; Zhou, X.; Zhuo, Y.H.; Gao, Z.W.; Du, X.L. A dynamic elastoplastic model of concrete based on a modeling method with environmental factors as constitutive variables. *J. Eng. Mech.* **2023**, *149*, 04023102.
28. Wadi, D.; Wu, W.; Fu, A.; Ahmed, H.E.; Abdalla, R. Correlation between SPT and CPT for sandy soils in port Sudan city using the linear regression model. *Afr. J. Eng. Technol.* **2022**. <https://doi.org/10.47959/ajet.2021.1.1.7>.
29. Babikir, I.A.A.; Nagm, E.; Bamousa, A.O.; Barazi, N. Facies analysis of the quaternary fluvial system of Khor Eit, Red Sea coast, Sudan. *J. Afr. Earth Sci.* **2021**, *184*, 104370.
30. Bird, E.C.F. Sudan. In *Encyclopedia of the World's Coastal Landforms*; Bird, E.C.F., Ed.; Springer: Dordrecht, The Netherlands, 2010; pp. 1013–1014.
31. Ali, S.H. Sedimentary facies, depositional environments and conceptual outcrop analogue (Dam Formation, early Miocene) Eastern Arabian Platform, Saudi Arabia: A new high-resolution approach. *J. Pet. Explor. Prod. Technol.* **2021**, *11*, 2497–2518.
32. Calliari, L.J. Fine grain sediment transport and deposition in the Patos Lagoon-Cassino beach sedimentary system. *Cont. Shelf Res.* **2009**, *29*, 515–529.
33. Wadi, D.; Wu, W.; Malik, I.; Ahmed, H.A.; Makki, A. Assessment of liquefaction potential of soil based on standard penetration test for the upper Benue region. *Environ. Earth Sci.* **2021**, *80*, 254.
34. Dağdeviren, U.; Demir, A.S.; Erden, C.; Kökçam, A.H.; Kurnaz, T.F. Explainable AI using ensemble machine learning with integrated SHapley additive explanations (SHAP)-Borda approach for estimation of the safety factor against soil liquefaction. *Environ. Earth Sci.* **2025**, *84*, 507.
35. Ying, Z.; Cui, Y.J.; Benahmed, N.; Duc, M. Salinity effect on the compaction behaviour, matric suction, stiffness and microstructure of a silty soil. *J. Rock Mech. Geotech. Eng.* **2021**, *13*, 855–863.
36. Kramer, O. K-Nearest Neighbors. In *Dimensionality Reduction with Unsupervised Neighbors*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 13–23.
37. Liu, L.; Zhao, G.; Liang, W. Slope stability prediction using k-NN-based optimum-path forest approach. *Mathematics* **2023**, *11*, 3071.
38. Guardiani, C.; Soranzo, E.; Wu, W. Time-dependent reliability analysis of unsaturated slopes under rapid drawdown with intelligent surrogate models. *Acta Geotech.* **2022**, *17*, 1071–1096.
39. Awad, M.; Khanna, R. Support vector regression. In *Efficient Learning Machines*; Apress: Berkeley, CA, USA, 2015; pp. 67–80.
40. Katsuumi, A.; Cong, Y.; Inazumi, S. AI-Driven prediction and mapping of soil liquefaction risks for enhancing earthquake resilience in smart cities. *Smart Cities* **2024**, *7*, 1836–1856.
41. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
42. Ji, W.; Liao, Y.; Zhang, L. Integrated generative adversarial network and XGBoost for anomaly processing of massive data flow in dispatch automation systems. *Intell. Autom. Soft Comput.* **2023**, *37*, 2825–2848.
43. Megahed, K.; Mahmoud, N.S.; Abd-Rabou, S.E.M. Application of machine learning models in the capacity prediction of RCFST columns. *Sci. Rep.* **2023**, *13*, 20878.
44. Jierula, A.; Wang, S.; Oh, T.M.; Wang, P. Study on accuracy metrics for evaluating the predictions of damage locations in deep piles using artificial neural networks. *Appl. Sci.* **2021**, *11*, 2314.
45. Shah, J.; Alam, M.; Javed, F.; Umar, M.; Ahmad, F. Comparative performance evaluation of machine learning models for predicting the ultimate bearing capacity of granular shallow foundations. *Sci. Rep.* **2025**, *15*, 36525.

46. Waris, K.A.; ur Rahaman, M.A.; Basha, B.M. Detection and prediction of slope stability in unsaturated finite slopes using interpretable machine learning. *Model. Earth Syst. Environ.* **2025**, *11*, 273.
47. Marín-Moreno, H.; Willis, J.; Zhang, Y.; Gourvenec, S. Interpretable XGBoost-based predictions of shear wave velocity from CPTu data. *Mar. Geophys. Res.* **2026**, *47*, 5.
48. Youd, B.T.L.; Idriss, I.M. 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*, 297–313.
49. Boulanger, R.W.; Idriss, I.M. Probabilistic standard penetration test–based liquefaction–triggering procedure. *J. Geotech. Geoenviron. Eng.* **2012**, *138*, 1185–1195.
50. Kurnaz, T.F.; Erden, C.; Kökçam, A.H.; Dağdeviren, U.; Demir, A.S. A hyper parameterized artificial neural network approach for prediction of the factor of safety against liquefaction. *Eng. Geol.* **2023**, *319*, 107109.
51. Ghani, S.; Thapa, I.; Kumari, S.; Correia, A.G.; Asteris, P.G. Revealing the nature of soil liquefaction using machine learning. *Earth Sci. Inform.* **2025**, *18*, 2.
52. Li, K.; Liu, J.; Mu, C.; Zhang, Y.; Ge, L. Effect of fines content on liquefaction resistance of soil using laboratory test and SPT. *Sci. Rep.* **2026**, *16*, 2135.
53. Karimai, K.; Liu, W.; Maruyama, Y. Prediction and factor analysis of liquefaction ground subsidence based on machine-learning techniques. *Appl. Sci.* **2024**, *14*, 2713.
54. Abdalzaher, M.S.; Moustafa, S.S.R.; Yassien, M. Development of smoothed seismicity models for seismic hazard assessment in the Red Sea region. *Nat. Hazards* **2024**, *120*, 12515–12544.
55. Pal, M. Support vector machines-based modelling of seismic liquefaction potential. *Int. J. Numer. Anal. Methods Geomech.* **2006**, *30*, 983–996.
56. Kumar, D.R.; Wipulanusat, W. Advancements in predicting soil liquefaction susceptibility: A comprehensive analysis of ensemble and deep learning approaches. *Sci. Rep.* **2025**, *15*, 26453.
57. Hsiao, C.H.; Kumar, K.; Rathje, E.M. Explainable AI models for predicting liquefaction-induced lateral spreading. *Front. Built Environ.* **2024**, *10*, 1387953.

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