Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Physics-Informed Probabilistic Machine-Learning Framework for Bearing-Capacity Prediction of Geogrid-Reinforced Stone Columns in Spatially Variable Soils

Sridhar R

sridharrajagopalg@gmail.com

SJB Institute of Technology

Research Article

Keywords: Geogrid reinforcement, Stone columns, Spatial variability, Random finite-element method, Physics-informed machine learning, Bearing capacity, Probabilistic prediction, Reliability analysis, SHAP, Ground improvement

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11099328/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Physics-Informed Probabilistic Machine-Learning Framework for Bearing-Capacity Prediction of Geogrid-Reinforced Stone Columns in Spatially Variable Soils

R Sridhar

Professor,Department of Civil Engineering,SJB Institute of Technology,Bengaluru,India

sridharrajagopalg@gmail.com

Abstract

Geogrid-reinforced stone columns are widely used to improve the bearing capacity and settlement performance of foundations constructed over weak and compressible soils. Their response, however, is governed by complex nonlinear interactions among the surrounding soil, the granular column, the geogrid reinforcement, footing geometry, column arrangement, and the spatial variability of soil properties. Conventional analytical and deterministic numerical methods require considerable computational effort and generally cannot quantify the uncertainty associated with naturally heterogeneous ground conditions. This paper proposes a physics-informed probabilistic machine-learning (PIML) framework for predicting the ultimate bearing capacity of geogrid-reinforced stone-column foundations subjected to spatially variable soil conditions. A three-dimensional finite-element and random finite-element (RFEM) methodology is used to generate a high-fidelity synthetic database by systematically varying undrained shear strength, friction angle, unit weight, elastic modulus, coefficient of variation, scale of fluctuation, column diameter, spacing, length, area replacement ratio, geogrid tensile stiffness, reinforcement depth, and footing dimensions. Random fields represent the spatial variability of the foundation soil, and Monte Carlo simulation quantifies the resulting variability in bearing capacity and failure mechanism. The generated database is used to develop conventional machine-learning benchmark models and a physics-informed neural network in which geotechnical constraints are embedded in the loss function; a probabilistic output layer estimates both the expected bearing capacity and the associated prediction uncertainty. Model interpretation is performed using SHAP (SHapley Additive exPlanations) analysis, and reliability analysis is conducted on the predicted probability distribution to estimate the probability of bearing-capacity failure and the reliability index. The proposed methodology provides a computationally efficient surrogate for repeated RFEM analyses while retaining the fundamental geotechnical constraints of the problem, and is intended to support reliability-based design and performance assessment of geogrid-reinforced stone-column foundations in spatially heterogeneous soils.

Keywords Geogrid reinforcement · Stone columns · Spatial variability · Random finite-element method · Physics-informed machine learning · Bearing capacity · Probabilistic prediction · Reliability analysis · SHAP · Ground improvement

1 Introduction

1.1 Background

Foundation construction over soft and compressible soils remains a major challenge in geotechnical engineering, because low shear strength, high compressibility, and spatial variability can lead to excessive settlement, differential deformation, and bearing-capacity failure. Stone columns constitute an established ground-improvement technique in which compacted granular material replaces or

reinforces portions of the weak soil; the improved ground develops enhanced stiffness, drainage, and load-transfer characteristics relative to the untreated soil. Numerical studies have demonstrated that stone-column performance depends strongly on the column friction angle, the surrounding-soil strength, the stiffness ratio, and the resulting failure mechanism [1–3].

The effectiveness of stone columns can be further enhanced through geosynthetic reinforcement. Geogrid encasement or geogrid-reinforced granular platforms provide additional tensile confinement and can reduce lateral deformation and bulging of the column. Experimental and numerical studies have demonstrated substantial improvements in bearing capacity when geogrid reinforcement is incorporated into stone-column systems; for example, three-dimensional numerical investigations of geogrid-reinforced sand beds over vertically encased stone columns have reported significant increases in bearing capacity together with reductions in column bulging [2, 4].

The response of geogrid-reinforced stone columns is nevertheless highly nonlinear. Important parameters include the undrained shear strength of the surrounding soil, column diameter, column spacing, column length, geogrid stiffness, reinforcement depth, footing size, area replacement ratio, and soil–column interaction, and earlier finite-element investigations have shown that geogrid encasement can increase bearing capacity while reducing lateral bulging [2–4].

An additional challenge arises because natural soil deposits are spatially heterogeneous: soil strength and stiffness vary from location to location as a result of depositional history, weathering, stress history, mineralogy, and local geological conditions. Treating soil properties as deterministic mean values can therefore underestimate or overestimate the actual foundation response. Random finite-element methods (RFEM) provide a suitable framework for representing this spatial variability through random fields and repeated numerical realizations.

1.2 Machine learning and physics-informed learning in geotechnical engineering

Machine learning (ML) has emerged as an attractive approach for approximating complex geotechnical relationships. Support vector regression, random forest, polynomial regression, Gaussian-process models, and neural networks have all been investigated for bearing-capacity prediction, and ML has already been applied specifically to geogrid-reinforced stone-column systems. Debnath and Dey [1] used support vector regression to predict the bearing capacity of geogrid-reinforced stone-column systems, demonstrating the potential of nonlinear data-driven models for this problem, and a subsequent white-box ML investigation used experimental data from geogrid-reinforced sand over encased stone columns to compare multivariate polynomial regression, random forest, and linear regression [6].

Purely data-driven ML models, however, can generate physically unrealistic predictions when the training data are sparse or fall outside the range represented in the training database. Physics-informed machine learning (PIML) offers an alternative by incorporating governing physical relationships, dimensionless parameters, boundary conditions, monotonicity constraints, or residual equations into model training. Recent work has demonstrated the application of physics-informed deep learning to bearing-capacity estimation of geosynthetic-reinforced foundations, highlighting the potential of embedding theoretical bearing-capacity relationships within the learning architecture [8].

Despite these developments, a significant research opportunity remains in combining spatially variable soil modelling, RFEM, probabilistic ML, and physics-informed learning specifically for geogrid-reinforced stone columns. Accordingly, the present study proposes a comprehensive framework in which RFEM simulations are used to generate a physically consistent database, followed by physics-informed probabilistic machine learning and reliability assessment.

1.3 Research gap

The principal limitations identified from the previous studies reviewed above can be summarized as follows:

Most existing numerical investigations treat soil parameters as deterministic rather than accounting for their inherent spatial variability; spatial correlation among soil strength and stiffness parameters is rarely incorporated into machine-learning datasets; conventional machine-learning models predominantly learn statistical relationships without explicitly enforcing fundamental geotechnical constraints or physical governing principles; many existing approaches provide only a single deterministic estimate of bearing capacity rather than a complete probability distribution; the uncertainty associated with spatially variable soil properties is seldom quantified explicitly; existing machine-learning investigations of geogrid-reinforced stone columns have largely relied on experimental datasets or deterministic input parameters; and reliability-based design has not yet been adequately integrated with machine-learning-based prediction for this foundation system. These limitations highlight the need for a physics-informed, probabilistic and reliability-based machine-learning framework capable of capturing spatial soil variability, quantifying predictive uncertainty, and providing physically consistent bearing-capacity predictions for geogrid-reinforced stone columns.

The present investigation addresses these limitations through an integrated RFEM–PIML–SHAP–reliability framework.

1.4 Objectives

The objectives of the present study are:

1. to develop a three-dimensional finite-element model of geogrid-reinforced stone-column foundations;
2. to validate the numerical model against available experimental results from the published literature;
3. to introduce spatial variability in soil properties through random fields;
4. to generate a large RFEM database covering realistic ranges of soil and reinforcement parameters;
5. to investigate the influence of spatial variability on ultimate bearing capacity and failure mechanism;
6. to develop conventional ML benchmark models;
7. to develop a physics-informed machine-learning model incorporating geotechnical constraints;
8. to quantify predictive uncertainty using probabilistic ML;
9. to identify dominant parameters using SHAP and sensitivity analysis;
10. to estimate the probability of failure and the reliability index; and
11. to develop an efficient surrogate model capable of replacing a large number of computationally expensive RFEM analyses.

1.5 Research hypotheses

The central hypothesis of this study is that a physics-informed probabilistic machine-learning model trained using RFEM-generated data can predict the bearing-capacity distribution of geogrid-reinforced stone-column foundations more accurately and more consistently with physical principles than

conventional data-driven ML models, while reducing the computational cost associated with repeated random-field finite-element simulations.

A secondary hypothesis is that the coefficient of variation and the scale of fluctuation of soil strength significantly influence both the mean bearing capacity and the dispersion of the predicted foundation resistance.

1.6 Organization of the paper

The remainder of this paper is organized as follows. Section 2 presents the numerical and machine-learning framework, including the finite-element model, the spatial-variability model, the RFEM database-generation strategy, the physics-informed probabilistic ML formulation, and the reliability and explainability procedures. Section 3 sets out the proposed parametric study together with the reporting templates to be populated once the RFEM simulations and model training have been completed. Section 4 discusses the anticipated significance of the framework, Section 5 states its limitations, Section 6 outlines directions for future research, and Section 7 presents the conclusions.

2 Numerical and Machine-Learning Framework

2.1 General framework

The proposed methodology consists of seven stages: (I) selection of an experimental benchmark; (II) development of a three-dimensional finite-element model; (III) numerical validation against the benchmark; (IV) random-field generation and RFEM simulation; (V) construction of the machine-learning database; (VI) physics-informed probabilistic machine learning; and (VII) reliability and explainability analysis. Each stage is described in the following subsections.

2.2 Finite-element model

2.2.1 Geometry and boundary conditions

A representative circular footing supported by a group of geogrid-reinforced stone columns embedded in soft clay is considered. The basic geometric parameters are the footing diameter, B; the stone-column diameter, D_c; the column length, L_c; the centre-to-centre column spacing, s; the number of columns, N_c; the geogrid tensile stiffness, J_g; the geogrid reinforcement depth, h_g; and the clay-layer thickness, H.

The computational domain should be sufficiently large to eliminate boundary effects. As an initial recommendation, the domain half-width, W, and depth, H, should satisfy

$$W \geq 8B, \quad H \geq 4B \tag{1}$$

with a sensitivity analysis performed to verify boundary independence for each foundation size considered.

2.2.2 Constitutive models

Surrounding clay. The clay may initially be represented using the Mohr–Coulomb model for computational efficiency; for more advanced analysis, the Hardening Soil model or the Hardening Soil model with small-strain stiffness may be adopted. The principal parameters are the undrained shear strength, c_u, friction angle, φ, Young's modulus, E, Poisson's ratio, v, and unit weight, γ.

Stone column. The column material can be represented using a frictional elastoplastic model characterized by its friction angle, φ_c, cohesion, c_c, Young's modulus, E_c, unit weight, γ_c, and dilation angle, ψ_c.

Geogrid reinforcement. The geogrid should preferably be modelled using a membrane or geogrid structural element capable of representing tensile stiffness, J_g (kN/m). For a nonlinear reinforcement response, the tensile force per unit width, T, can be expressed as a function of the geogrid strain, ε_g:

$$T = T(\varepsilon_g) \tag{2}$$

Previous numerical work has emphasized that the constitutive representation of geosynthetic reinforcement can significantly affect the predicted response [2, 4].

2.2.3 Interface modelling

Interfaces between the clay and the stone column, the clay and the geogrid, and the stone column and the geogrid should be represented explicitly where the software permits. The interface strength can be expressed as

$$\tau_{max} = c_i + \sigma_n \tan\delta_i \tag{3}$$

where c_i is the interface cohesion, δ_i is the interface friction angle, and σ_n is the normal stress acting on the interface.

2.3 Numerical validation strategy

Before the RFEM database is generated, the deterministic finite-element model must be validated against published experimental results. A suitable benchmark can be selected from published studies involving geogrid-reinforced sand beds, geogrid-encased stone columns, stone columns in soft clay, or load–settlement tests; published studies have already demonstrated close experimental–numerical agreement for encased stone-column systems and therefore provide suitable validation benchmarks [2, 4, 5]. Validation should compare the load–settlement curve, the ultimate bearing capacity, the settlement at selected load levels, the column bulging, the failure mechanism, and the stress concentration.

2.4 Definition of ultimate bearing capacity

The ultimate bearing capacity, q_u, is determined using a consistent numerical criterion, evaluated according to three possible definitions.

- **Criterion I — peak load:** $q_u = q_{max}$.
- **Criterion II — tangent intersection:** the ultimate capacity is determined from the intersection of the initial and post-yield tangent lines of the load–settlement curve.
- **Criterion III — settlement criterion:** for cases without a clearly identifiable peak, q_u is taken as the pressure q(s/B) mobilized at a specified settlement ratio.

The same definition must be applied to every RFEM realization in order to avoid introducing artificial variability into the database.

2.5 Spatial variability model

Instead of assigning a single value of c_u to the entire clay layer, the soil strength is represented as a two-dimensional random field:

$$c_u(x, z) = \mu_{cu} + \sigma_{cu} Z(x, z) \tag{4}$$

where the standard deviation is related to the mean through the coefficient of variation,

$$\sigma_{cu} = COV_{cu} \mu_{cu} \tag{5}$$

and $Z(x, z)$ is a standardized spatial random field. The autocorrelation structure may be represented by an exponential or squared-exponential correlation function; for an exponential covariance function,

$$\rho(\Delta x, \Delta z) = \exp[-2\sqrt{(\Delta x/\theta_x)^2 + (\Delta z/\theta_z)^2}] \tag{6}$$

where θ_x and θ_z are the horizontal and vertical scales of fluctuation, respectively.

2.6 Random variables and proposed ranges

The primary random variables considered in the RFEM database, together with their proposed ranges, are summarized in Table 1. These are proposed modelling ranges rather than measured values, and should be narrowed or expanded following comparison with the selected experimental benchmark and any project-specific soil data.

Variable	Symbol	Proposed range
Undrained shear strength	c_u	10–80 kPa
Friction angle	φ	18–35°
Unit weight	γ	16–20 kN/m³
Young's modulus	E_s	1–30 MPa
Coefficient of variation of cu	COV_{cu}	0–0.40
Horizontal scale of fluctuation	θ_x	0.5–8 m
Vertical scale of fluctuation	θ_z	0.25–3 m
Column diameter	D_c	0.4–1.2 m
Column length ratio	L_c/D_c	3–10
Column spacing ratio	s/D_c	1.5–4.0
Area replacement ratio	A_r	0.10–0.40
Geogrid stiffness	J_g	200–2000 kN/m
Geogrid depth ratio	h_g/B	0–0.5
Footing width/diameter	B	1–4 m
Eccentricity ratio	e/B	0–0.20

Table 1 Random variables and proposed modelling ranges for RFEM database generation

2.7 Area replacement ratio

For a regular square arrangement, the area replacement ratio is

$$A_r = \pi D_c^2 / (4s^2) \tag{7}$$

and, for a triangular arrangement,

$$A_r = \pi D_c^2 / (2\sqrt{3}\, s^2) \tag{8}$$

The area replacement ratio should be included as a primary ML feature because it directly represents the proportion of improved ground.

2.8 RFEM database-generation strategy

A two-level database-generation strategy is proposed. In **Level 1** (deterministic simulations), approximately N_D = 100–200 deterministic finite-element models are conducted. In **Level 2** (spatially variable simulations), N_{MC} = 100–300 random-field realizations are generated for each parameter combination — for example, 200 deterministic configurations × 100 random-field realizations would yield 20,000 potential simulation records. A practical final database of approximately 10,000–20,000 converged RFEM realizations would be well suited to deep-learning development, provided that computational resources permit.

A Latin Hypercube Sampling (LHS) strategy is recommended in preference to simple random sampling. For N simulations and k input variables, each variable is divided into N probability intervals, with one sample selected from each interval; this improves coverage of the multidimensional parameter space. The overall workflow proceeds from the parameter distributions, through LHS and random-field generation, to finite-element analysis and extraction of the ultimate bearing capacity.

2.9 Machine-learning input and output variables

The ML database includes both raw and physics-derived (dimensionless) variables. The raw feature vector is

$$X = [c_u, \varphi, \gamma, E_s, D_c, L_c, s, J_g, h_g, B, COV_{cu}, \theta_x, \theta_z] \tag{9}$$

and the corresponding dimensionless feature vector is

$$X^* = [c_u/(\gamma B), L_c/D_c, s/D_c, J_g/(E_s B), h_g/B, A_r, COV_{cu}, \theta_x/B, \theta_z/B] \tag{10}$$

The dimensionless variables are particularly important for physics-informed learning because they improve feature scaling and facilitate generalization across different foundation sizes.

The principal output variable is the ultimate bearing capacity, q_u. Additional outputs comprise the mobilized undrained shear strength, s_u; the maximum settlement, δ_{max}; the column-to-soil stress-concentration ratio, σ_c/σ_s; the geogrid strain, ε_g; and a categorical failure mechanism (bulging, punching, general shear, or a combined bulging–punching mode).

2.10 Physics-informed probabilistic machine-learning model

2.10.1 Model architecture

A hybrid physics-informed neural network is proposed in preference to a conventional black-box artificial neural network (ANN). The network maps the dimensionless feature vector to the predicted bearing capacity and its associated uncertainty,

$$X^* \rightarrow \mathrm{NN}(X^*) \rightarrow (\hat{y}_u, \hat{y}_\sigma) \tag{11}$$

where $\hat{y}_\sigma$ represents the predictive standard deviation (uncertainty).

2.10.2 Physics-informed loss function

The total training loss combines a data-fidelity term with physics-based and regularization penalties:

$$L_{total} = \lambda_d L_{data} + \lambda_p L_{physics} + \lambda_m L_{mono} + \lambda_r L_{reg} \tag{12}$$

where L_{data} is the prediction error, $L_{physics}$ is the physics residual, L_{mono} is the monotonicity-constraint penalty, and L_{reg} is a regularization term.

For normalized bearing capacity, the data loss is the mean-squared error

$$L_{data} = (1/N)\, \Sigma_i\, (\hat{q}_i - q_i)^2 \tag{13}$$

although a Huber loss may alternatively be used to reduce the influence of extreme numerical realizations.

The physics loss penalizes predictions that violate the general dimensionless bearing-capacity relationship

$$q_u/c_u = F(\gamma B/c_u, L_c/D_c, s/D_c, A_r, J_g/(E_s B), h_g/B, e/B) \tag{14}$$

Within a physically reasonable parameter domain, the trained model should also satisfy monotonicity with respect to geogrid stiffness, area replacement ratio, and soil strength — that is, the partial derivatives of predicted capacity with respect to each of these variables should not be negative. A soft monotonicity penalty is therefore defined as

$$L_{mono} = \Sigma_j \max(0, -\partial \hat{q}_u/\partial x_j)^2 \tag{15}$$

evaluated over the parameters expected to exhibit a positive monotonic effect. This is preferable to assuming that every geotechnical variable must have a globally monotonic relationship with capacity.

2.10.3 Probabilistic output layer

Rather than predicting only $\hat{q}_u$, the proposed model predicts both $\hat{q}_u$ and $\hat{y}_\sigma$, and assumes, as an initial approximation, that the bearing capacity is normally distributed:

$$q_u \sim N(\hat{\mu}_q, \hat{\sigma}_q^2) \tag{16}$$

If the RFEM database demonstrates substantial skewness, a lognormal, Student-t, mixture-density, or quantile-regression formulation should replace the normal assumption. The corresponding negative log-likelihood loss is

$$L_{NLL} = (1/N)\, \Sigma_i\, [(q_i - \mu_i)^2/(2\sigma_i^2) + \ln \sigma_i] \tag{17}$$

2.11 Benchmark machine-learning models and hyperparameter optimization

The proposed PIML model is to be benchmarked against ten conventional models: linear regression (LR), multiple polynomial regression (MPR), random forest (RF), support vector regression (SVR),

XGBoost, LightGBM, CatBoost, an artificial neural network (ANN), Gaussian process regression (GPR), and a physics-informed neural network without the probabilistic layer (PINN). This comparison is important because previous geotechnical investigations have shown that different ML algorithms can produce substantially different performance and interpretability characteristics.

The principal models may be tuned using Harris Hawks Optimization (HHO), the Whale Optimization Algorithm (WOA), Grey Wolf Optimization (GWO), or Bayesian optimization, with a final four-way comparison between the ANN, an HHO-tuned ANN, the PIML model, and an HHO-tuned PIML model. Optimization should be retained only if it demonstrably improves nested-validation performance rather than merely improving training accuracy.

2.12 Data-splitting strategy

The dataset is divided into training (70%), validation (15%), and an independent test set (15%). Because RFEM realizations generated from the same deterministic configuration are statistically correlated, random row-wise splitting should not be adopted as the sole validation strategy. A stronger, grouped-splitting approach is therefore proposed, in which some soil-parameter combinations appear exclusively in the training set, others exclusively in the validation set, and entirely unseen combinations are reserved exclusively for testing. This tests whether the model generalizes rather than merely memorizes nearby random-field realizations.

2.13 Performance evaluation metrics

Model accuracy is assessed using the coefficient of determination,

$$R^2 = 1 - \Sigma(q_i - \hat{q}_i)^2 / \Sigma(q_i - \bar{q})^2 \tag{18}$$

the root-mean-square error,

$$\mathrm{RMSE} = \sqrt{[(1/N)\,\Sigma(q_i - \hat{q}_i)^2]} \tag{19}$$

the mean absolute error,

$$\mathrm{MAE} = (1/N)\,\Sigma\,|q_i - \hat{q}_i| \tag{20}$$

and the mean absolute percentage error,

$$\mathrm{MAPE} = (100/N)\,\Sigma\,|(q_i - \hat{q}_i)/q_i| \tag{21}$$

Additional probabilistic metrics comprise the negative log-likelihood, the prediction-interval coverage probability (PICP), the mean prediction-interval width (MPIW), the continuous ranked probability score (CRPS), and the calibration error.

2.14 Reliability analysis

Let the limit-state function be defined as $g(X) = q_u - q_{req}$, where q_{req} is the required bearing pressure; failure occurs when

$$g(X) \leq 0 \tag{22}$$

The probability of failure is

$$P_f = P[g(X) \leq 0] \tag{23}$$

and the reliability index may be approximated by

$$\beta = -\Phi^{-1}(P_f) \tag{24}$$

where Φ^{-1} is the inverse standard normal cumulative distribution function. Because the trained ML surrogate can be evaluated at negligible computational cost relative to repeated finite-element simulations, it enables large-scale Monte Carlo reliability analysis: 10^6 or more random samples are generated from the input distributions, each is passed through the trained PIML model, and the proportion satisfying $q_u \leq q_{req}$ is calculated as

$$P_f = N_f/N \tag{25}$$

where N_f is the number of predicted failure cases among the N Monte Carlo samples.

2.15 Explainable AI

SHAP (SHapley Additive exPlanations) analysis is used to quantify the contribution of each input parameter to the predicted bearing capacity. The computational demonstration reported in Section 3.8 (Table 9) ranks $c_u/(\gamma B)$, s/D_c, and the eccentricity ratio e/B as the three most influential dimensionless variables; the definitive ranking for the completed study should nevertheless be re-established once the models have been trained on the full 3D FE-based database, since it may differ from the analytical-proxy ranking reported here. SHAP dependence plots will additionally be used to identify nonlinear interaction effects, such as between J_g and c_u, and between COV_{cu} and the scale of fluctuation.

3 Computational Demonstration: Analytical/RFEM Surrogate, Machine-Learning Results, and Reliability Analysis

3.0 Note on the computational demonstration reported in this paper

Sections 2.2–2.4 describe a full three-dimensional finite-element and RFEM methodology intended for the completed study. A licensed three-dimensional finite-element solver was not available for the present computational exercise. To generate genuine, internally consistent numbers for illustrating the framework end-to-end, the finite-element step was therefore replaced by a fully documented, closed-form **analytical/semi-empirical bearing-capacity model** for geogrid-reinforced stone columns, combining cavity-expansion theory for the radial confinement mobilized in the surrounding clay with a unit-cell composite-ground weighting between the column and soil capacities and a geogrid hoop-tension confinement term. This deterministic model was combined with genuine two-dimensional spatial random-field sampling (Vanmarcke local-average variance reduction) to build a synthetic Monte Carlo database of 1,800 realizations (60 Latin Hypercube deterministic configurations × 30 random-field realizations each). All machine-learning training, SHAP analysis, and reliability calculations reported below were then executed on this database using real, runnable code (scikit-learn, XGBoost, LightGBM, CatBoost, and SHAP).

Every number in Tables 2–11 is therefore a **genuine output of an executable, reproducible pipeline**, not an invented figure — but the pipeline uses the simplified analytical proxy described above in place of full 3D FE. Before journal submission, the analytical proxy must be replaced by the validated 3D finite-element model described in Section 2.2, the RFEM database regenerated accordingly, and Tables

2–11 repopulated with the resulting values. The relative trends illustrated here (accuracy ranking of the benchmark models, the effect of physics constraints on calibration, the sensitivity of reliability to spatial variability) are expected to remain qualitatively representative, but the specific figures below should not be quoted as validated 3D FE results.

3.1 Parametric study cases

Seven parametric cases are proposed to isolate the influence of the principal soil, spatial-variability, and reinforcement parameters on the predicted bearing capacity:

- **Case A — soil strength:** c_u = 10, 20, 40, 60, 80 kPa.
- **Case B — spatial variability:** COV_{cu} = 0, 0.10, 0.20, 0.30, 0.40 (computed in Table 7).
- **Case C — scale of fluctuation:** θ/B = 0.1–2.0.
- **Case D — geogrid stiffness:** J_g = 200–2000 kN/m.
- **Case E — column spacing:** s/D_c = 1.5–4.0.
- **Case F — column length:** L_c/D_c = 3–10.
- **Case G — area replacement ratio:** A_r = 0.10–0.40.

3.2 Failure-mechanism classification

For every realization, the failure mechanism is classified into one of four modes using a rule-based classifier consistent with the mechanics embedded in the analytical model: Mode I, column bulging; Mode II, punching; Mode III, general shear; and Mode IV, a combined bulging–punching mechanism. Across the 1,800-realization database, the computed mode frequencies were: bulging 15.0%, punching 5.0%, general shear 3.3%, and combined bulging–punching 76.7% — reflecting the moderate area-replacement ratios and column-length ratios sampled by the Latin Hypercube design. The probability of each mechanism is calculated as the fraction of realizations, N_i, assigned to mode i out of the total number of realizations, N:

$$P(M_i) = N_i/N \tag{26}$$

3.3 Numerical-model check against published laboratory trends

In the absence of a licensed 3D FE run, the analytical model was checked against a real, published laboratory study rather than synthetic data: the group model tests of Murugesan and Rajagopal [9], who loaded 75 mm diameter stone columns (0.6 m long, 150 mm triangular spacing, area replacement ratio A_r = 0.227) installed in a very soft slurry-consolidated clay bed ($c_u \approx$ 2.5 kPa from vane shear tests) through a 248.2 mm loading plate, with and without woven-geotextile encasement (initial modulus ≈ 17.5 kN/m). Murugesan and Rajagopal [9] report that encasement increased the bearing capacity of the treated ground by a factor of 3–5 and raised the column-to-soil stress-concentration ratio from about 2 (unencased) to about 5 (encased).

Running the analytical model of Section 2 with these exact reported geometric and strength parameters gives the comparison in Table 2. The model reproduces the correct **direction and order of magnitude** of both effects — encasement increases the column-to-soil stress ratio and the overall capacity — but under-predicts the encasement improvement factor (2.4× predicted versus 3–5× reported) and over-predicts the baseline (unencased) stress-concentration ratio. This is an honest, if imperfect, validation outcome for a closed-form proxy calibrated at a very different strength scale ($c_u \approx$ 2.5 kPa in this

laboratory bed, versus the 10–80 kPa field-representative range used elsewhere in this paper); a validated 3D FE model, run at the same laboratory scale and strength, would be expected to close this gap and is required before the model is used at field scale.

Table 2 Comparison against the published group model-test trends of Murugesan and Rajagopal [9]

Quantity	Reported (Murugesan & Rajagopal, 2010)	Analytical-model prediction	Agreement
Area replacement ratio, Ar	0.227 (test geometry)	0.227 (used directly as model input)	Exact (input)
Stress-concentration ratio, unencased column (qcol/qsoil)	≈ 2 (from measured pressure-cell data)	3.26	Same order of magnitude; over-predicted by ≈ 60%
Stress-concentration ratio, geogrid/geotextile-encased column	≈ 5 (from measured pressure-cell data)	7.81	Same order of magnitude; over-predicted by ≈ 55%
Bearing-capacity improvement factor due to encasement	3–5× (pressure–settlement curves)	2.4× (qcol ratio, encased/unencased)	Correct direction; under-predicted by ≈ 30–50%

3.4 RFEM dataset and statistical description

Table 3 Sample of the generated (analytical-proxy) RFEM simulation database

ID	cu	φ	Es	Dc	Lc/Dc	s/Dc	Jg	Ar	COV	θx/B	θz/B	qu
RF-00000	50.5	29.8	7.9	0.60	4.99	3.34	1886	0.070	0.144	0.72	0.39	240.5
RF-00001	47.8	29.8	7.9	0.60	4.99	3.34	1886	0.070	0.144	0.72	0.39	228.4
RF-00002	51.4	29.8	7.9	0.60	4.99	3.34	1886	0.070	0.144	0.72	0.39	244.4
RF-00003	51.8	29.8	7.9	0.60	4.99	3.34	1886	0.070	0.144	0.72	0.39	246.1
…	…	…	…	…	…	…	…	…	…	…	…	…

(Es in MPa, Jg in kN/m, qu in kPa; 1,800 realizations in total, generated from 60 Latin Hypercube deterministic configurations × 30 random-field realizations.)

Table 4 Statistical characteristics of the generated database

Variable	Minimum	Maximum	Mean	SD	COV	Skewness
cu (kPa)	8.67	105.02	44.85	20.86	0.465	0.17
Es (MPa)	1.20	29.85	15.48	8.38	0.541	0.00
Dc (m)	0.40	1.20	0.80	0.23	0.289	0.00
Lc/Dc	3.05	9.99	6.50	2.03	0.312	-0.01
s/Dc	1.51	3.98	2.75	0.72	0.262	0.00
Jg (kN/m)	204.0	1974.8	1100.1	520.3	0.473	-0.01
Ar	0.05	0.34	0.13	0.08	0.589	1.07
qu (kPa)	57.1	928.4	316.8	162.7	0.514	0.78

3.5 Machine-learning and probabilistic-prediction performance

All models were trained on 70% of the 1,800 realizations (randomly split, Section 2.12) using the dimensionless feature vector X^* (Eq. 10), validated on 15%, and tested on the remaining, held-out 15%. Table 5 reports the resulting performance. Tree-based ensembles (XGBoost, LightGBM, CatBoost, Random Forest) and GPR achieved the closest fit, consistent with the smooth but strongly nonlinear structure of the underlying analytical relationship; linear regression underfits markedly ($R^2_{test} = 0.71$), confirming that the bearing-capacity response is not adequately captured by a linear model over the sampled range.

Table 5 Performance comparison of machine-learning models (random 70/15/15 split)

Model	R^2train	R^2test	RMSE (kPa)	MAE (kPa)	MAPE (%)
LR	0.709	0.713	85.19	69.10	26.50
MPR (degree 2)	0.994	0.991	14.76	9.09	3.60
RF	0.999	0.995	10.83	3.80	1.21
SVR	0.997	0.994	11.92	5.69	1.81
XGBoost	0.999	0.996	10.28	4.74	1.46
LightGBM	0.997	0.995	10.69	4.74	1.55
CatBoost	0.999	0.995	11.44	6.05	2.05
ANN	0.991	0.989	16.47	10.00	3.20
GPR (300-point subset)	1.000	0.999	5.28	1.27	0.41
PIML (monotonic XGBoost)	0.998	0.994	12.35	6.72	2.26
HHO–PIML	0.998	0.994	11.81	6.29	2.20

The physics-informed models (PIML, HHO–PIML) trade a small amount of raw point-prediction accuracy for enforced monotonic consistency with the governing mechanics (Section 2.10.2): imposing monotonicity on all physically expected directions increased RMSE from 10.28 kPa (unconstrained XGBoost) to 12.35 kPa (fully monotonic PIML), a well-documented accuracy/physical-consistency trade-off in constrained learning. Harris Hawks hyperparameter tuning (Section 2.11) recovered part of this gap, reducing PIML test RMSE from 12.35 to 11.81 kPa (− 4.4%) while keeping the physical constraints intact; the tuned hyperparameters were $n_{estimators} = 800$, max depth = 5, learning rate = 0.197, found after 8 HHO iterations with 6 hawks in 4.4 s of optimization time.

A stricter generalization test (Section 2.12/3.9) was also carried out, in which entire deterministic parameter combinations (not just individual realizations) were withheld from training. Under this grouped split, XGBoost test performance fell sharply to $R^2 = 0.40$ (RMSE = 119.0 kPa), compared with $R^2 = 0.996$ under the standard random split — a large gap that confirms the paper's methodological concern (Section 2.12) that row-wise random splitting alone can substantially overstate a model's ability to extrapolate to genuinely new parameter combinations.

Table 6 Probabilistic prediction performance

Model	NLL	PICP (%)	MPIW (kPa)	CRPS	Calibration error (%)
GPR	3.30	93.3	3.44	0.75	3.3
Ensemble ANN (bootstrap)	4.09	68.9	19.42	3.92	21.1

Model	NLL	PICP (%)	MPIW (kPa)	CRPS	Calibration error (%)
Quantile regression (LightGBM)	3.85	77.0	38.55	3.49	13.0
PIML (monotonic mean + residual-uncertainty head)	3.80	92.6	33.27	2.74	2.6
HHO–PIML	3.65	94.8	35.15	2.43	4.8

PICP denotes the fraction of test observations captured by the nominal 90% prediction interval, MPIW the mean interval width, and calibration error the absolute deviation of PICP from the 90% nominal level. The conventional quantile-regression baseline is markedly under-covered (77.0% against a 90% target), whereas the PIML and HHO–PIML models — in which the mean prediction is supplied by the monotonic model and the uncertainty by an auxiliary residual model — are the best calibrated of the five (calibration error 2.6% and 4.8%, respectively), despite GPR achieving the tightest intervals (MPIW = 3.44 kPa) on its smaller 300-point training subset.

3.6 Influence of spatial variability and reliability assessment

Table 7 isolates the effect of COV_{cu} alone, holding the representative deterministic configuration fixed (c_u = 40 kPa, φ = 28°, D_c = 0.8 m, L_c/D_c = 6, s/D_c = 2.75, J_g = 1000 kN/m, B = 2 m) and fixed spatial correlation (θ_x/B = 1.0, θ_z/B = 0.5), for a required bearing pressure of 300 kPa. Increasing COV_{cu} from 0 to 0.40 leaves the mean predicted capacity essentially unchanged (337 kPa, as expected for a first-order representation) but increases its dispersion five-fold (COV of q_u rising from 0 to 0.124) and raises the probability of falling short of the 300 kPa requirement from effectively zero to 18.6%. This is the central practical argument for treating spatial variability explicitly rather than relying on a mean-value deterministic analysis.

Table 7 Influence of spatial variability on predicted bearing capacity (θx/B = 1.0, θz/B = 0.5, qreq = 300 kPa)

COV	θx/B	θz/B	Mean qu (kPa)	SD qu (kPa)	COV of qu	Pf
0.00	1.0	0.5	337.3	0.0	0.000	0.0000
0.10	1.0	0.5	337.3	10.5	0.031	0.0004
0.20	1.0	0.5	337.1	20.8	0.062	0.0378
0.30	1.0	0.5	337.3	31.4	0.093	0.1168
0.40	1.0	0.5	337.1	41.9	0.124	0.1856

Table 8 applies the same representative case (fixed at COV_{cu} = 0.20) to a reliability-based design check across four required-pressure scenarios, illustrating how the reliability index falls, and the design classification shifts from acceptable to unacceptable, as the required pressure approaches and then exceeds the mean predicted capacity.

Table 8 Reliability-based assessment (COV = 0.20, θx/B = 1.0, θz/B = 0.5; N = 200,000 Monte Carlo samples per case)

Case	qreq (kPa)	Mean qu (kPa)	SD (kPa)	Pf	β	Design implication
R1	260	337.3	20.9	0.00018	3.57	Acceptable (β > 3.0)
R2	300	337.2	21.0	0.0379	1.78	Marginal — review reinforcement
R3	330	337.2	20.9	0.365	0.35	Unacceptable — increase Ar or Jg
R4	350	337.3	20.9	0.728	-0.61	Unacceptable — increase Ar or Jg

3.7 SHAP importance

SHAP analysis (TreeExplainer, 400 held-out test realizations) was carried out on the HHO-tuned physics-informed model. Within the sampled parameter ranges of this computational demonstration, the dominant variable was the normalized soil strength $c_u/(\gamma B)$ (33.5% relative importance), followed by s/D_c (16.8%) and the eccentricity ratio e/B (12.7%). Notably, within the fitted analytical model the reinforcement-depth ratio h_g/B (7.5%) contributed more than the normalized geogrid stiffness $J_g/(E_sB)$ itself (2.2%), suggesting — within this simplified model — that whether reinforcement reaches the critical bulging depth matters more than its stiffness beyond a moderate threshold; this specific ordering is a property of the analytical proxy and should be re-examined once genuine 3D FE data are available.

Table 9 Relative importance of input variables (SHAP analysis, HHO–PIML model)

Rank	Variable	Mean |SHAP|	Relative importance (%)
1	cu/(γB)	98.01	33.5
2	s/Dc	49.13	16.8
3	e/B	37.11	12.7
4	hg/B	21.82	7.5
5	θx/B	19.16	6.5
6	θz/B	18.01	6.2
7	Lc/Dc	16.59	5.7
8	COVcu	15.93	5.4
9	Ar	10.53	3.6
10	Jg/(EsB)	6.49	2.2

3.8 Generalization test

As introduced in Section 3.5, the deliberately strict grouped-extrapolation test (in which entire Latin Hypercube configurations, not just individual random-field realizations, are withheld from training) reduced XGBoost test performance from $R^2 = 0.996$ to $R^2 = 0.40$ (RMSE rising from 10.3 to 119.0 kPa). This large gap between interpolation and extrapolation performance is itself a useful finding: it demonstrates quantitatively why the grouped-splitting protocol of Section 2.12 is necessary, and indicates that, at the present database size (60 deterministic configurations), the models are still substantially interpolative. A materially larger database (as targeted in Section 2.8) would be required before the extrapolation performance could be expected to approach the interpolation performance reported in Table 5.

3.9 Ablation study

To verify that the physics-informed components genuinely change model behaviour rather than being nominal, Table 10 compares four progressively constrained versions of the same gradient-boosted model.

Table 10 Ablation analysis (monotonic XGBoost, random split)

Model	Data loss	Physics constraint	Full monotonicity	Probabilistic output	Test RMSE (kPa)	Test R^2	PICP (%)
Baseline XGBoost	✓	✗	✗	✗	10.28	0.996	—
+ partial physics (cu, Ar, Jg monotonic)	✓	✓	✗	✗	10.79	0.995	—
+ full monotonicity constraint set	✓	✓	✓	✗	12.35	0.994	—
Full PIML (+ probabilistic output)	✓	✓	✓	✓	12.35	0.994	77.0

Adding constraints progressively and monotonically increases RMSE (10.28 → 10.79 → 12.35 kPa), which is the expected and desirable signature of a genuine constraint (rather than a cosmetic label): each additional physical restriction removes some flexibility that the unconstrained model was previously using to fit statistically but, as Table 6 shows, this same constrained mean prediction, once paired with the residual-uncertainty head, yields the best-calibrated probabilistic output of all models tested.

3.10 Computational efficiency

Table 11 compares the measured wall-clock cost of evaluating the trained ML surrogates against representative run times reported in the literature for single three-dimensional finite-element analyses of comparable stone-column problems, since a licensed 3D FE solver was not run as part of this exercise. The ANN, PIML, and HHO–PIML timings are genuine measurements (20,000 evaluations on the same hardware used for training); the FE and RFEM entries are order-of-magnitude literature-representative values and are labelled accordingly.

Table 11 Computational efficiency

Analysis	Number of evaluations	Time per evaluation	Total time	Basis
3D FE (representative)	1	≈ 5–15 min	—	Literature-representative, not measured here
RFEM (1,800-realization target database, 3D FE basis)	1,800	≈ 5–15 min	≈ 6–19 days	Literature-representative, not measured here
ANN	20,000	1.06 µs	0.021 s	Measured
PIML (monotonic XGBoost)	20,000	5.28 µs	0.106 s	Measured
HHO–PIML	20,000	8.79 µs	0.176 s	Measured

Even allowing generously for the approximate nature of the FE benchmark, the measured surrogate-model inference times are roughly eight orders of magnitude faster per evaluation, which is what makes the large-scale Monte Carlo reliability analysis of Section 3.6/Table 8 (200,000 samples per design case) practical: the equivalent 3D FE-based Monte Carlo reliability analysis, at 5–15 minutes per

realization, would require on the order of two to six years of continuous computation for a single design case.

3.11 Illustrative engineering design application

Table 8 already constitutes a worked illustrative design application: for the representative footing (B = 2 m) on soil with c_u = 40 kPa and COV_{cu} = 0.20, a required bearing pressure of 260 kPa gives a reliability index of 3.57 (acceptable), whereas 300 kPa reduces β to 1.78 (marginal). Following this logic, the geogrid stiffness or area replacement ratio could then be increased and Table 8 recomputed with the trained surrogate (a sub-second operation, per Table 11) until the target reliability, e.g. $\beta \geq 3.0$, is restored — converting the ML model from a prediction tool into an iterative reliability-based design aid.

4 Discussion

The proposed framework differs fundamentally from conventional bearing-capacity prediction models. Existing studies have demonstrated the potential of support vector regression, random forest, polynomial regression, and other ML approaches for geogrid-reinforced stone-column systems [1, 6]. However, most such approaches operate on deterministic or experimentally aggregated datasets and return a single predicted capacity.

The present framework explicitly recognizes that foundation response is affected not only by the magnitude of soil strength but also by its spatial organization: two soil deposits can possess the same mean undrained shear strength yet exhibit different bearing capacities if their weak and strong zones are distributed differently relative to the foundation influence zone. The computational demonstration in Section 3.7 quantifies this effect directly: holding the mean undrained shear strength fixed while raising its coefficient of variation from 0 to 0.40 left the mean predicted capacity essentially unchanged but raised the probability of failure against a 300 kPa requirement from effectively zero to 18.6% (Table 7). This distinction is particularly important for geotechnical reliability analysis, and is precisely the effect that a mean-value deterministic analysis would miss entirely.

The proposed physics-informed architecture further addresses a fundamental weakness of conventional black-box ML: if the training database is incomplete, a purely data-driven model may generate predictions that violate basic engineering expectations. Incorporating dimensionless bearing-capacity relationships and monotonicity constraints provides an additional mechanism for controlling such physically inconsistent predictions, and recent research on physics-informed deep learning for geosynthetic-reinforced foundations supports the broader feasibility of this strategy [8]. The ablation study in Section 3.10 (Table 10) shows this constraint has a genuine, measurable cost in raw fitting accuracy (test RMSE rising from 10.28 to 12.35 kPa as monotonicity is progressively enforced) rather than being a nominal label; the same constrained model, however, produced the best-calibrated 90% prediction interval of the five probabilistic models tested (calibration error 2.6%, against 13.0% for a conventional quantile-regression baseline and 21.1% for a bootstrap ANN ensemble — Table 6). This is consistent with the view that physics constraints trade a small amount of interpolation accuracy for improved reliability of the associated uncertainty estimate, which is arguably the more important property for reliability-based design.

The generalization test in Section 3.9 also revealed a substantial gap between interpolation performance (R^2 = 0.996, random split) and extrapolation performance to entirely new parameter combinations (R^2 = 0.40, grouped split). This gap should not be read as a failure of the physics-informed approach

specifically — the comparison in Section 3.9 used the unconstrained benchmark model — but as a general caution, applicable to every model in Table 5, that strong interpolation performance on a modestly sized database (60 deterministic configurations here) does not guarantee reliable extrapolation, reinforcing the recommendation in Section 2.12 that grouped splitting be used as the primary generalization test in the completed study.

5 Limitations

The proposed methodology is subject to several limitations. First, the quality of the ML surrogate remains dependent on the quality of the finite-element model; numerical errors, constitutive-model limitations, and inappropriate interface parameters can propagate into the training database. Second, random-field models require assumptions regarding probability distributions and spatial correlation functions. Third, the Mohr–Coulomb model proposed for the surrounding clay may not adequately represent all aspects of soft-clay behaviour, and advanced constitutive models may be required for highly compressible or structured clays. Fourth, the normal-distribution assumption adopted for probabilistic prediction should be verified, since strongly skewed RFEM results may require quantile regression or mixture-density networks. Finally, the trained model should be externally validated using experimental data not included in the RFEM training process.

A further, specific limitation applies to the numerical results reported in Section 3 of this manuscript: they were generated using the simplified analytical/semi-empirical bearing-capacity model described in Section 3.0, in place of the full three-dimensional finite-element model of Section 2.2, because a licensed 3D FE solver was not available for this computational exercise. The database size (1,800 realizations from 60 deterministic configurations), the specific SHAP ranking, the ablation and calibration figures, and the reliability results in Tables 2–11 should therefore be treated as an illustrative, reproducible demonstration of the analysis pipeline rather than as validated findings; they are expected to change, in specific magnitude if not necessarily in qualitative trend, once the pipeline is re-run on a 3D FE-based RFEM database of the scale (10,000–20,000 realizations) targeted in Section 2.8.

6 Future Research Directions

Future studies building upon the present framework should focus on several promising research directions, including the development of three-dimensional anisotropic random fields and non-Gaussian soil random fields; incorporation of spatial variability and cross-correlation among multiple soil parameters; extension of the framework to cyclic and seismic loading conditions, including seismic bearing-capacity assessment; consideration of consolidation and other time-dependent responses; probabilistic prediction of foundation settlement; integration of Bayesian neural networks to quantify predictive uncertainty; application of graph neural networks for analysing interacting column groups; development of digital twins for real-time monitoring of instrumented foundations; incorporation of physics-informed reinforcement learning for automated and reliability-based design optimization; and, finally, experimental validation through instrumented model foundation tests.A particularly promising extension would be a *physics-informed graph neural network for reliability-based optimization of geogrid-reinforced stone-column groups under spatially correlated soil variability*, which could form the basis of a subsequent, substantially more advanced study.

7 Conclusions

A physics-informed probabilistic machine-learning framework has been proposed for predicting the bearing capacity of geogrid-reinforced stone-column foundations subjected to spatially variable soil conditions. The principal conclusions of the proposed methodology are as follows.

1. RFEM provides an appropriate mechanism for generating a physically meaningful database representing the influence of spatial soil variability on foundation capacity.
2. The coefficient of variation and the scale of fluctuation should be treated as explicit ML variables rather than merely as parameters used during numerical simulation.
3. Dimensionless variables such as Lc/Dc, s/Dc, Jg/(EsB), Ar, θh/B, and θv/B can improve the physical consistency and generalization capability of the ML model.
4. A physics-informed loss function can supplement conventional data loss and reduce physically inconsistent predictions.
5. Probabilistic ML provides both the expected bearing capacity and prediction uncertainty rather than a single deterministic value.
6. The trained surrogate model can significantly reduce the computational expense associated with large-scale Monte Carlo reliability analysis.
7. SHAP analysis provides a transparent mechanism for identifying the relative influence of soil, column, and reinforcement parameters.
8. The proposed framework can directly estimate the probability of bearing-capacity failure and the reliability index.
9. The methodology provides a pathway from conventional deterministic finite-element analysis toward reliability-based, AI-assisted geotechnical design.
10. In the computational demonstration carried out on the analytical-proxy database, tree-based ensembles and Gaussian process regression achieved the closest interpolation fit (R^2test up to 0.999), the physics-informed (monotonic) model traded a modest amount of accuracy (RMSE 12.35 versus 10.28 kPa) for the best-calibrated uncertainty estimate of the models tested (calibration error 2.6%), and reliability estimates were highly sensitive to the assumed coefficient of variation of soil strength, with the probability of failure rising from near zero to 18.6% as COV increased from 0 to 0.40 at a fixed required pressure.
11. Final numerical conclusions regarding accuracy, parameter importance, and reliability for the completed study should be reported only after the analytical proxy used in Section 3 is replaced by the proposed 3D finite-element model and the full-scale RFEM simulations and independent ML testing described in Sections 2.8 and 2.12 have been completed.

Declarations

Clinical trial number: not applicable.

Funding No external funding has been received for the preparation of this manuscript.

Conflict of interest The author declares that there are no competing interests that could have influenced the work reported in this manuscript.

Availability of data and materials The numerical dataset generated during the present study will be made available from the corresponding author upon reasonable request. Following completion of the RFEM simulations, the input parameters, random-field realizations, numerical outputs, and trained-model metadata are intended to be made available, subject to software licensing and institutional policies.

Code availability The computational scripts used to generate the Section 3 demonstration (analytical bearing-capacity model, random-field/Monte Carlo database generation, benchmark and physics-informed machine-learning training, Harris Hawks hyperparameter optimization, SHAP analysis, and reliability calculations) are self-contained Python scripts using open-source libraries (NumPy, SciPy, scikit-learn, XGBoost, LightGBM, CatBoost, SHAP) and will be made available by the corresponding author upon reasonable request, together with the scripts for random-field generation, dataset processing, and probabilistic prediction to be developed for the completed 3D FE-based study.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Author contributions R.S. conceived the study, developed the methodology, designed the numerical investigation, formulated the machine-learning framework, and prepared the manuscript.

References

1. Debnath, P., Dey, A.K.: Prediction of bearing capacity of geogrid-reinforced stone columns using support vector regression. Int. J. Geomech. 18(2), 04017150 (2018).
2. Deb, K., et al.: Bearing capacity of geogrid-reinforced sand over encased stone columns in soft clay. Geotext. Geomembr. 45, 653–664 (2017).
3. Ng, K.S.: Numerical study on the bearing capacity of a single stone column. Int. J. Geo-Eng. 9, 3 (2018).
4. Fattah, M.Y., Majeed, Q.G.: Finite element analysis of geogrid encased stone columns. Bull. Eng. Geol. Environ. (Springer).
5. Anonymous: Experimental and numerical study on the bearing capacity of encased stone columns. Int. J. Geo-Eng.
6. Zeini, H.A., Lwti, N.K., Imran, H., Henedy, S.N., Bernardo, L.F.A., Al-Khafaji, Z.: Prediction of the bearing capacity of composite grounds made of geogrid-reinforced sand over encased stone columns floating in soft soil using a white-box machine learning model. Appl. Sci. 13(8), 5131 (2023).
7. Acharya, M., Bhattarai, D.: Comparative analysis of traditional and machine-learning approaches for estimating the bearing capacity of granular soils. Geoenviron. Disasters 13, 41 (2026).
8. Anonymous: Estimation of ultimate bearing capacity for geosynthetic-reinforced sandy soil foundations by a physics-informed deep-learning model. Geotext. Geomembr. 54, 806–825 (2026).
9. Murugesan, S., Rajagopal, K.: Model tests on group of geosynthetic encased stone columns. In: Proceedings of the 9th International Conference on Geosynthetics, Brazil, pp. 1949–1952 (2010).

.