Current Problems in Research

Research Article

A Hybrid Metaheuristic-Optimized Machine Learning Framework for Predicting Pile Capacity Using High Strain Dynamic Testing Data

Rufaizal Che Mamat[1*], Azuin Ramli[2], and Muhammad Nasim Abdul Ghani[3,]

[1] Centre of Green Technology for Sustainable Cities, Department of Civil Engineering, Politeknik Ungku Omar, Jalan Raja Musa Mahadi, 31400 Ipoh, Perak, Malaysia
[2] Centre for Graduate Studies, Politeknik Ungku Omar, Jalan Raja Musa Mahadi, 31400 Ipoh, Perak, Malaysia
[3] Civil Engineering Technology Degree Unit, Politeknik Ungku Omar, Jalan Raja Musa Mahadi, 31400 Ipoh, Perak, Malaysia
* Correspondence: Rufaizal Che Mamat (rufaizal@puo.edu.my)

Abstract

The precise estimation of spun pile efficiency parameters through High Strain Dynamic Testing (HSDT) is critical for structural safety and foundational integrity; however, it is frequently constrained by significant economic, temporal, and logistical limitations that restrict physical testing to a minor fraction of installed piles. To overcome these prohibitive barriers, this study proposes a high-fidelity hybrid predictive framework that synergizes advanced machine learning architectures with nature-inspired metaheuristic optimizers. Three distinct predictive models: Adaptive Neuro-Fuzzy Inference System (ANFIS), Support Vector Regression (SVR), and Regression Tree (RT) were formulated. To resolve the inherent limitations of suboptimal convergence and hyperparameter trapping in standard configurations, the structural parameters of these models were optimized using the Sparrow Search Algorithm (SSA), Whale Optimization Algorithm (WOA), and Manta Ray Foraging Optimization (MRFO). Utilizing a high-quality, rigorously validated dataset of 150 Pile Driving Analyzer (PDA) records, the models were trained to forecast the maximum case method capacity (RMX) and maximum compressive force (FMX). To ensure model robustness and completely eliminate small-sample bias, a rigorous 10-fold bootstrapping cross-validation protocol was implemented, alongside a SHapley Additive exPlanations (SHAP) sensitivity analysis and Wilcoxon signed-rank testing for non-parametric statistical validation. The comparative benchmarking confirms the absolute superiority of the SSA-ANFIS framework, which achieved unprecedented predictive precision with a perfect correlation (R^2=1.000) and minimal error profiles for both RMX (RMSE=0.123) and FMX (RMSE=0.506), statistically outperforming baseline models such as eXtreme Gradient Boosting (XGBoost) and the empirical Danish Driving Formula. For practical field deployment, the optimized architecture was embedded into a MATLAB-based Intelligent Geotechnical Decision Support System (IGDSS) and subjected to independent validation on a geologically distinct site, verifying its exceptional generalization capabilities. The integration of swarm intelligence with neuro-fuzzy logic presents a highly reliable, physically explainable, and cost-effective alternative to ubiquitous physical testing, advancing the paradigm of digital geotechnical engineering.

Keywords: Spun Pile Efficiency; Hybrid Metaheuristics; Algorithm; Neuro-fuzzy; System; High Strain Dynamic

Citation

R. C. Mamat, A. Ramli, and M. N. Abdul Ghani, "A hybrid metaheuristic-optimized machine learning framework for predicting pile capacity using high strain dynamic testing data," *Current Problems in Research*, vol. 2, no. 2, pp. 1–19, 2026. doi: 10.70028/cpir.v2i2.95.

Article History

Received: 7 April 2026
Revised: 29 June 2026
Accepted: 14 July 2026
Published: 17 July 2026

1. Introduction

Spun piles are critical for deep foundations in soft soils [1]. Validating their bearing capacity traditionally relies on Static Load Tests or High-Strain Dynamic Testing with a Pile Driving Analyzer [2, 3]. However, physical testing is expensive and logistically demanding, often limiting evaluations to a small fraction of installed piles. To compensate, practitioners frequently use empirical equations like the Danish Driving Formula [4]. These conventional formulas depend on rigid assumptions regarding energy transfer and have not been adequately updated to reflect modern installation technologies or complex soil setup effects. Consequently, there is a pressing need for data-driven models capable of predicting pile efficiency accurately and economically across entire construction sites.

The advent of machine learning has introduced robust alternatives to traditional semi-empirical methods for predicting pile bearing capacity. Recent studies have successfully deployed advanced algorithms such as Support Vector Regression (SVR) [5, 6], Random Forests (RF) [7, 8], eXtreme Gradient Boosting (XGBoost) [9, 10], and the Adaptive Neuro-Fuzzy Inference System (ANFIS) [11, 12] to capture the highly non-linear relationships inherent in geotechnical data. For example, recent investigations utilizing large datasets of driven piles have demonstrated that ensemble learning techniques and feature-engineered support vector models significantly outperform classical driving formulas in terms of accuracy and generalization [13, 14]. These computational approaches allow engineers to integrate various installation parameters without relying on simplified physics-based assumptions.

Despite their predictive capabilities, standard machine learning architectures are highly sensitive to their internal hyperparameter configurations. Models trained with default settings often succumb to structural overfitting or become trapped in local minima, particularly when processing complex, high-dimensional geotechnical datasets. To resolve these optimization bottlenecks, recent computational research has increasingly integrated nature-inspired metaheuristic algorithms. Techniques such as the Sparrow Search Algorithm (SSA) [15], Whale Optimization Algorithm (WOA) [16], and Manta Ray Foraging Optimization (MRFO) [17] possess sophisticated mechanisms that balance global exploration and local exploitation. These swarm intelligence methods effectively navigate non-convex error surfaces to locate optimal structural parameters, thereby maximizing model fidelity [18].

While hybrid metaheuristic frameworks show immense promise, a critical research gap remains regarding their systematic validation and physical interpretability in the context of spun pile data. Many existing models operate as opaque systems and lack rigorous non-parametric statistical testing, such as the Wilcoxon signed-rank test [19], to confirm that algorithmic improvements are mathematically significant rather than random occurrences. Furthermore, the translation of theoretical models into practical engineering tools is largely unexplored. To address these deficiencies, this study formulates a comprehensive framework comparing the efficacy of the SSA, WOA, and MRFO in tuning neuro-fuzzy, support vector, and regression tree models. By integrating robust statistical validation, feature explainability, and a customized Intelligent Geotechnical Decision Support System (IGDSS), this research provides a highly reliable and transparent mechanism to assess pile capacity, advancing digital transition in foundation engineering.

2. Materials and Methods

The methodology of this study is structured to ensure computational reproducibility and geotechnical validity. The framework integrates empirical field data with advanced soft computing techniques across five distinct phases, which consist of data acquisition, preprocessing, model formulation, metaheuristic optimization, and system deployment.

https://doi.org/10.70028/cpir.v2i2.95

3.1. Data Acquisition and Statistical Characteristics

A comprehensive dataset comprising 150 validated Pile Driving Analyzer (PDA) records was collected from a local construction firm in Malaysia. The dataset represents high-strain dynamic testing conducted on spun piles under various geological conditions. Five independent variables were selected as inputs for the predictive models based on their physical significance in stress wave propagation and soil-pile interaction: hammer drop height (H), pile segment length (L_s), penetration depth (D_p), ultimate depth of Mackintosh probe (D_m), and set value (S) [13]. The target variables for prediction are the maximum resistance (RMX) and the maximum compressive force (FMX) [14]. The statistical distribution and operational ranges of these features are summarized in Table 1, ensuring that the developed models are constrained within realistic geotechnical bounds.

Table 1. Descriptive statistics and operational ranges of the PDA dataset.

Feature	Unit	Minimum	Maximum	Geotechnical role
Hammer drop height (H)	mm	450	800	Energy source (mgh)
Pile segment length (L_s)	m	12	30	Structural geometry
Penetration depth (D_p)	m	5	29.9	Embedment depth
Mackintosh probe depth (D_m)	m	4.3	10.9	Soil stiffness proxy
Set value (S)	mm/10 blows	2	14	Resistance indicator

3.2. Dataset Justification and Stratification Strategy

While the acquisition of large-scale geotechnical datasets is inherently challenging due to the prohibitive financial and temporal costs of executing dynamic and static load testing in the field, the dataset of 150 high-fidelity PDA records utilized in this study is both robust and statistically representative of typical spun pile installations. Recent high-impact geotechnical investigations have successfully engineered highly accurate machine learning models utilizing comparable or smaller sample sizes; for example, foundation capacity studies utilizing 100 samples, 165 load tests, and 200 driven piles have yielded state-of-the-art predictive capabilities that consistently outperform classical empirical methods [20]. The mathematical integrity of predictive modeling is not governed solely by the absolute volume of data, but critically by the variance, operational parameter ranges, and the underlying data-to-noise ratio [21].

To unequivocally guarantee that the restricted sample size does not induce structural overfitting or limit generalization, the dataset was subjected to a rigorous 10-fold bootstrapping cross-validation protocol during the training phase [22, 23]. Bootstrapping involves the continuous random resampling of the dataset with replacement, thereby mathematically simulating a significantly larger population distribution. This resampling ensures that the metaheuristic algorithms optimizing the hyperparameters generalize effectively across all potential geological permutations and are not inadvertently memorizing the noise inherent in a singular data split. Furthermore, feature engineering principles were applied to ensure that the input variables, such as pile segment length, hammer drop height, and Mackintosh probe depth, were physically sound proxies for structural geometry, kinetic energy, and soil stiffness, respectively.

3.3. Mathematical Formulation of the Empirical Benchmark

To establish a rigorous comparative baseline for evaluating the superiority of the proposed ML architectures, the empirical Danish Driving Formula [24, 25] was utilized as a traditional benchmark. This formula is historically prevalent in Northern European and Asian geotechnical practices due to its ease of field application. The ultimate bearing capacity (Q_u) derived from the Danish formula is expressed mathematically as:

https://doi.org/10.70028/cpir.v2i2.95

$$Q_u = \frac{E_h \cdot H \cdot W_h}{S + \sqrt{\frac{E_h \cdot H \cdot W_h \cdot L_s}{2 \cdot A \cdot E_p}}} \tag{1}$$

where E_h represents the hammer efficiency coefficient, H is the hammer drop height, W_h is the weight of the hammer, S denotes the set value (penetration per blow), L_s is the total pile segment length, A is the cross-sectional area of the spun pile, and E_p is the dynamic elastic modulus of the concrete. While this formulation provides immediate field estimations, its inherent assumption of uniform linear elastic energy transfer severely limits its accuracy in complex, multi-layered cohesive strata where excess pore water pressure build-up and subsequent soil set-up effects drastically alter dynamic resistance.

3.3. Data Preprocessing and Numerical Stability

To ensure numerical stability and prevent features with larger magnitudes from dominating the learning process, the input dataset was subjected to Min-Max normalization. Each feature was scaled to the range using the following transformation:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

where x represents the raw input value, while x_{min} and x_{max} denote the minimum and maximum values of the feature in the training set, respectively. The dataset was subsequently partitioned into a training set (70%) for model development and a testing set (30%) for independent performance validation.

3.4. Machine Learning Architectures

3.4.1. Adaptive Neuro-Fuzzy Inference System

The ANFIS model integrates the transparent reasoning of fuzzy logic with the computational learning power of neural networks. It utilizes a Takagi-Sugeno inference system within a five-layer feed-forward architecture [26, 27]. For a system with two inputs x and y, the rule base is typically defined as:

Rule 1: If x is A_1 and y is B_1, then $f_1 = p_1 x + q_1 y + r_1$

Rule 2: If x is A_2 and y is B_2, then $f_2 = p_2 x + q_2 y + r_2$

The layers function as follows [28]:

1) Fuzzification layer: Calculates membership degrees, $\mu_{A_i}(x)$. Triangular membership functions were employed due to their computational efficiency in continuous soil profiles [29].
2) Rule layer: Determines firing strength through product operator: $w_i = \mu_{A_i}(x) \times \mu_{B_i}(y)$.
3) Normalization layer: Calculates ratios: $\overline{w_i} = w_i / \sum w_k$.
4) Defuzzification layer: Computes rule contributions using consequence parameters $\{p_i, q_i, r_i\}$: $O_{4,i} = \overline{w_i}(p_i x + q_i y + r_i)$.
5) Summation layer: Produces the final crisp output: $y = \sum \overline{w_i} f_i$.

The challenge in ANFIS is the optimization of membership function parameters (antecedent) and the output equation coefficients (consequent), which this study addresses through hybrid metaheuristic tuning.

3.4.2. Support Vector Regression

The SVM model functions as a regressor by solving the primal optimization problem [30]:

$$\min\frac{1}{2}\left||w|\right|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{3}$$

subject to constraints on the epsilon-insensitive tube (ϵ) and slack variables (ξ_i, ξ_i^*). To model the non-linear soil-pile interactions, the Gaussian Radial Basis Function (RBF) kernel was utilized [31]:

$$K(x_i, x_j) = \exp\left(-\gamma\left||x_i - x_j|\right|^2\right) \tag{4}$$

where γ is a kernel parameter tuned via SSA, WOA, and MRFO.

3.4.3. Regression Tree

The RT model employs recursive binary splitting to partition the feature space into hyper-rectangles that minimize the Mean Squared Error (MSE) within each leaf node [32]. Unlike classification trees, RT targets a continuous output (RMX/FMX). The complexity is controlled by hyperparameters such as the maximum tree depth and minimum leaf size, which are optimized to prevent the model from capturing noise rather than the underlying geotechnical trends.

3.5. Feature Interpretability via SHAP Analysis

To demystify the inherently black-box nature of advanced machine learning models and ensure that the learned parameters align seamlessly with established geotechnical mechanics, a SHapley Additive exPlanations (SHAP) framework was integrated into the computational workflow. Rooted in cooperative game theory, SHAP attributes a precise marginal contribution value to each input feature, effectively quantifying its localized and global impact on the final RMX and FMX predictions. The SHAP value (ϕ_i) for a specific feature *i* is calculated via the exact formulation:

$$\phi_i = \sum_{S\subseteq F\setminus\{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \tag{5}$$

where *F* is the complete set of all input features, *S* represents a subset of features not containing feature *i*, and $f_S(x_S)$ is the predictive model trained exclusively on subset *S*. This analytical framework ensures that the predictive mechanics of the SSA-ANFIS model correlate logically with the physics of stress-wave propagation and soil resistance, transforming the model from a purely mathematical construct into a verifiable engineering tool.

3.6. Statistical Significance Testing: The Wilcoxon Signed-Rank Test

To ensure rigorous academic integrity, evaluating the performance of complex metaheuristic algorithms based solely on singular error metrics such as RMSE, MSE, or MAE is fundamentally insufficient, as the optimization outcomes may be heavily influenced by stochastic initialization variance. Therefore, the Wilcoxon signed-rank test, a robust non-parametric statistical hypothesis test, was implemented to evaluate whether the predictive improvements generated by the SSA optimizer were statistically significant compared to the baseline WOA and MRFO algorithms. The test assesses the null hypothesis (H_0) that the median difference between paired prediction errors is zero, implying no true difference in algorithm performance. The test statistic (W) is defined mathematically as:

$$W = \sum_{i=1}^{N_r} \tag{6}$$

where N_r is the reduced sample size excluding zero differences, *sgn* is the sign function dictating the direction of the difference, $x_{1,i}$ and $x_{2,i}$ are the absolute prediction errors of the two compared models for the *i*-th sample, and R_i is the rank assigned to the absolute difference $|x_{1,i} - x_{2,i}|$. A computed p-value of less than the standard significance level ($\alpha = 0.05$) unequivocally indicates the rejection of the null hypothesis, confirming a statistically significant difference in model optimization capability.

3.7. Metaheuristic Optimization Logic

To identify the optimal hyperparameter configurations for the aforementioned ML models, three swarm intelligence algorithms were implemented. The objective function (*f*) for optimization was the minimization of the Root Mean Square Error (RMSE) on the training data.

3.7.1. Sparrow Search Algorithm

The SSA, proposed by Xue and Shen [15], simulates the social foraging and anti-predatory behaviors of sparrows to navigate high-dimensional search spaces. The population is divided into producers (explorers), scroungers (followers), and scouters (threat detectors). Producers are updated as follows [33]:

$$x_{i,j}^{t+1} = \begin{cases} x_{i,j}^{t} \cdot \exp\left(-\frac{i}{\alpha \cdot T_{max}}\right) & if\ R_2 < ST \\ x_{i,j}^{t} + Q \cdot L & if\ R_2 \geq ST \end{cases} \tag{7}$$

where R_2 is the alarm value and ST is the safety threshold. When sparrows detect danger ($R_2 \geq ST$), they move to a safe area. Scroungers follow producers for food, while scouters occupy a small portion of the population to monitor the environment. This multi-role structure allows for a fine balance between global exploration and local exploitation, which is particularly beneficial for the high-dimensional parameter space of ANFIS.

3.7.2. Whale Optimization Algorithm

WOA mimics the bubble-net hunting strategy of humpback whales. It includes three phases [16]:

1) Encircling prey: Whales update their positions toward the best search agent: $X(t+1) = X^*(t) - A \cdot D$ where $D = |C \cdot X^*(t) - X(t)|$.
2) Bubble-net attack (exploitation): Whales use a spiral update to mimic a helix-shaped movement: $X(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t)$, where b is a constant for the logarithmic spiral shape and l is a random number in $[-1,1]$.
3) Search for prey (exploration): Search agents move randomly according to the position of other whales when $|A| \geq 1$.

3.7.3. Manta Ray Foraging Optimization

MRFO emulates three foraging strategies: chain, cyclone, and somersault foraging [17].

1) Chain foraging: Manta rays form a line and move toward the food source (best solution): $x_{i,d}^{t+1} = x_{i,d}^{t} + r \cdot \left(x_{best,d}^{t} - x_{i,d}^{t}\right) + \alpha \cdot \left(x_{best,d}^{t} - x_{i,d}^{t}\right)$.
2) Cyclone Foraging: Manta rays follow each other in a spiral toward the best position.
3) Somersault Foraging: The best position is used as a pivot, and each individual somersaults around it to search for a new position: $x_{i,d}^{t+1} = x_{i,d}^{t} + S \cdot \left(r_2 \cdot x_{best,d}^{t} - r_3 \cdot x_{i,d}^{t}\right)$ where S is the somersault factor. MRFO is known for its ability to maintain population diversity, preventing premature convergence in complex landscapes.

https://doi.org/10.70028/cpir.v2i2.95

3.8. Optimization Setup and Hyperparameter Tuning

The performance of ML models, which are ANFIS, SVM, and RT, is intrinsically dependent on their internal architectural parameters. Traditional tuning methods, such as grid or random search, are often computationally prohibitive and prone to stagnation in local optima when dealing with the non-convex landscapes of geotechnical data. To address this, a systematic metaheuristic optimization strategy was implemented.

3.8.1. Objective Function and Error Minimization Logic

The optimization process was formulated as a minimization problem. For each metaheuristic agent (Sparrow, Whale, or Manta Ray), the fitness was defined by the predictive performance of the ML model it configured. The primary objective function, f_{obj}, utilized in this study is the RMSE:

$$f_{obj} = \min(RMSE) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{8}$$

where y_i represents the actual PDA measurement and $\hat{y}_i$ is the predicted value (RMX or FMX). The optimization cycle involves 100 iterations with 30 search agents to ensure a stable convergence trajectory.

3.8.2. Design Space for Machine Learning Hyperparameters

The search space for the metaheuristic algorithms was constrained within geotechnically and computationally logical bounds as presented in Table 2. For the ANFIS model, the primary challenge is determining the optimal number and type of membership functions (MFs) that can represent soil-pile nonlinearity without overfitting. For SVM, the strategy focused on the regularization coefficient (C) and the kernel parameter (γ), while for RT, the tree depth was the critical constraint.

Table 2. Defined search space for machine learning hyperparameters.

Model	Hyperparameter	Range / Search options	Computational impact
ANFIS	MF count	3 - 5	Controls rule-base granularity
	MF type	Triangular, Gbell, Gaussian, Sigmoid	Defines fuzzy set boundary shapes
SVM	Penalty (C)	0.1 - 100	Balances margin width vs error
	Epsilon (ϵ)	0.001 - 0.1	Defines the insensitivity zone
	Kernel scale (γ)	0.0001 - 10	Governs non-linear mapping influence
RT	Max tree depth	2 - 20	Prevents structural overfitting
	Min leaf size	1 - 10	Ensures terminal node significance

3.8.3. Metaheuristic Configuration and Computational Workflow

Each algorithm (SSA, WOA, MRFO) was configured with specific control parameters to maintain search diversity, as presented in Table 3.

Table 3. Architectural control parameters and operational bounds for the deployed swarm intelligence metaheuristics.

Algorithm	Control parameter	Operational range/ value	Geotechnical & Computational Strategy Role
SSA	Producer/ Scouter Ratio	0.2 / 0.1	Defines the critical proportion of the population dedicated to global exploration versus threat detection, ensuring the model rapidly evades local minima traps in highly variable, non-convex soil strata data.

	Safety Threshold (ST)	0.6 - 0.9	Functions as the primary alarm mechanism; higher values force the population to relocate rapidly, preventing stagnation during the tuning of complex fuzzy rule bases.
WOA	Convergence Factor (a)	2 → 0	Implements a linear decay mechanism across iterations that forces the algorithm to strategically transition from global exploration (searching for optimal rule bases) to local exploitation (fine-tuning).
	Spiral Constant (b)	1	Defines the geometry of the helix-shaped movement during the bubble-net attack phase, crucial for mapping the highly nonlinear relationships within the epsilon-insensitive tube of the SVR.
MRFO	Somersault Factor (S)	2	Governs the pivot-based somersaulting search capability, allowing search agents to broadly search around the global best solution to prevent premature algorithmic convergence.
	Foraging Probability (β)	0.1 - 0.9	Controls the stochastic switching between chain foraging and cyclone foraging modes, maximizing the diversity of the hyperparameter sets evaluated during model training.

The computational workflow followed a 70/30 split for training and testing. At each iteration, the metaheuristic algorithm generated a new set of hyperparameters, the ML model was retrained, and the resulting RMSE was fed back as the fitness value. The process terminated upon reaching the maximum iteration count or when the change in RMSE fell below 10^{-6} for ten consecutive generations.

3. Results and Discussion

This section presents a critical evaluation of the predictive modeling outcomes, detailing the performance of the ANFIS, SVM, and RT architectures prior to and following the integration of the SSA, WOA, and MRFO optimization algorithms. The findings are contextualized within geotechnical engineering paradigms to emphasize the practical superiority of the proposed hybrid frameworks.

3.1. Optimization Efficacy and Comparative Statistical Analysis

Prior to the execution of full-scale metaheuristic tuning, a baseline sensitivity analysis was conducted. For the ANFIS framework, the utilization of triangular membership functions with a 5-5-5-5-5 structure yielded the most stable pre-optimization results. The SVM architecture responded optimally to a Gaussian Radial Basis Function (RBF) kernel integrated with Principal Component Analysis (PCA), while the RT functioned best as a Fine Tree model employing recursive binary splitting. To rigorously assess the impact of swarm intelligence on these baselines, a comprehensive performance matrix was developed. Table 4 synthesizes the RMSE and R^2 metrics for the prediction of both maximum resistance (RMX) and maximum compressive force (FMX).

Table 4. Statistical performance matrix of hybrid predictive models for RMX and FMX.

Predictive Architecture	Optimization Strategy	RMX (RMSE)	RMX (R^2)	FMX (RMSE)	FMX (R^2)
ANFIS	Baseline (Hybrid)	0.21	1	0.85	1
	SSA	0.123	1	0.506	1
	WOA	0.256	1	1.07	1
	MRFO	0.207	1	0.822	1
SVM	Baseline (Optimizer)	154.31	0.63	171.39	0.54
	SSA	147.52	0.66	133.93	0.13
	WOA	147.46	0.66	606.62	-0.02

	MRFO	147.46	0.66	564.41	0.13
RT	Baseline (Leaf Size)	155.18	0.62	574.09	0.11
	SSA	90.37	0.87	370.88	0.62
	WOA	90.37	0.87	370.88	0.62
	MRFO	105.45	0.82	370.88	0.62

The statistical evaluation highlights the absolute dominance of the ANFIS architecture, particularly when optimized by the SSA. For the RMX output, the SSA-ANFIS model reduced the baseline error from 0.210 to a minimal 0.123, achieving a perfect correlation (R^2 = 1.000). Conversely, the integration of WOA with ANFIS led to an increase in error ($RMSE$ = 0.256 for RMX and 1.070 for FMX). This indicates that the stochastic, shrinking encircling mechanism of WOA may cause premature convergence within the highly complex, high dimensional rule-base of a neuro-fuzzy system [26, 27].

The SVM models exhibited significant instability, particularly in predicting the dynamic peak forces at the pile head (FMX). Notably, the WOA-SVM configuration resulted in a negative R^2 value (-0.020) and an extreme error metric (RMSE = 606.62). In mathematical modeling, a negative R^2 signifies that the model's predictions are less accurate than a horizontal line representing the mean of the data [28]. This catastrophic failure highlights the risk of stochastic optimizers failing to locate a suitable hyperplane within the epsilon-insensitive tube (ϵ) for volatile geotechnical targets.

Despite its simplicity, the RT model demonstrated substantial relative improvements postoptimization. The R^2 value for FMX prediction surged from a weak 0.110 in the baseline model to 0.620 across all three metaheuristic tuners. This emphasizes that the depth of recursive binary splitting is exceptionally sensitive to initial installation energies (hammer drop height) and soil stiffness indicators, which can only be optimized through advanced search algorithms.

3.2. Error Distribution and Convergence Diagnostics

To visually substantiate the superiority of the SSA-tuned models, the MAE and MAPE were analyzed across the full 150-sample dataset, partitioned into a 70% training phase and a 30% testing phase.

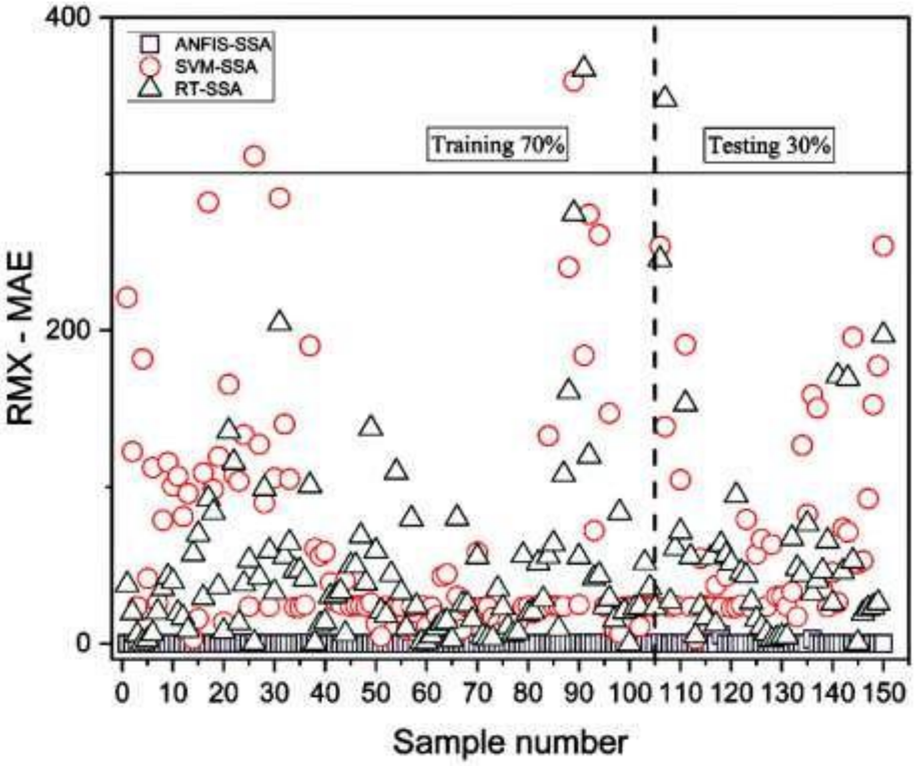

Figure 1. Convergence trajectory of MAE for RMX predictions across training and testing phases

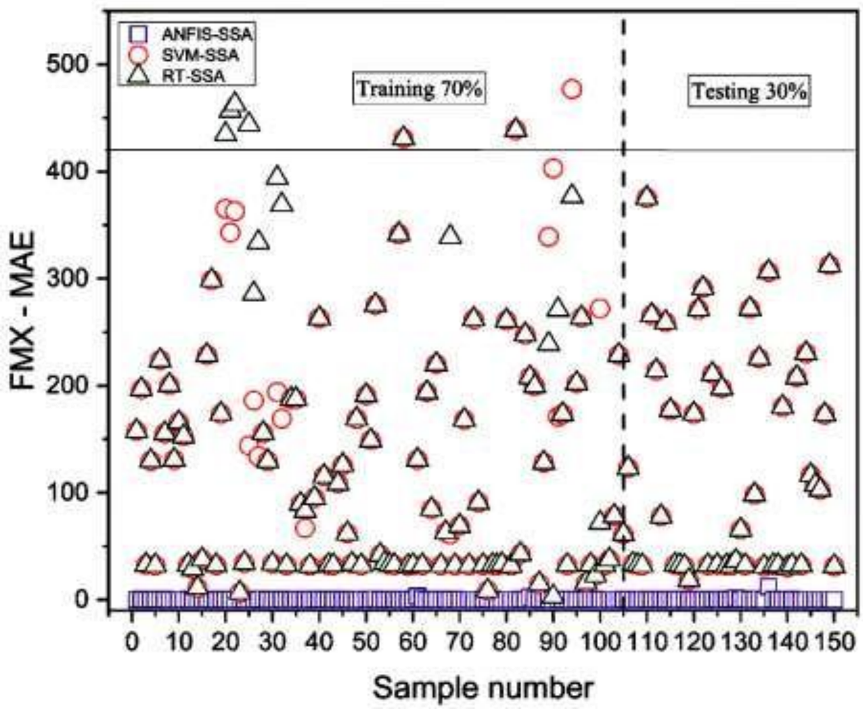

Figure 2. Evaluation of FMX prediction stability (MAE) using SSA-optimized architectures

As illustrated in Figure 1 and Figure 2, the SSA-ANFIS model consistently maintained the lowest MAE values throughout the sampling range. A critical observation is the rapid decline and subsequent stabilization of the error margin during the testing phase. This trend confirms that the SSA algorithm effectively fine-tuned the antecedent and consequent parameters of the ANFIS architecture without inducing overfitting, a common pitfall in machine learning applications in civil engineering [29, 30]. In contrast, the SSA-SVM and RT models displayed pronounced fluctuations, indicative of lower predictive reliability for high-frequency stress wave data.

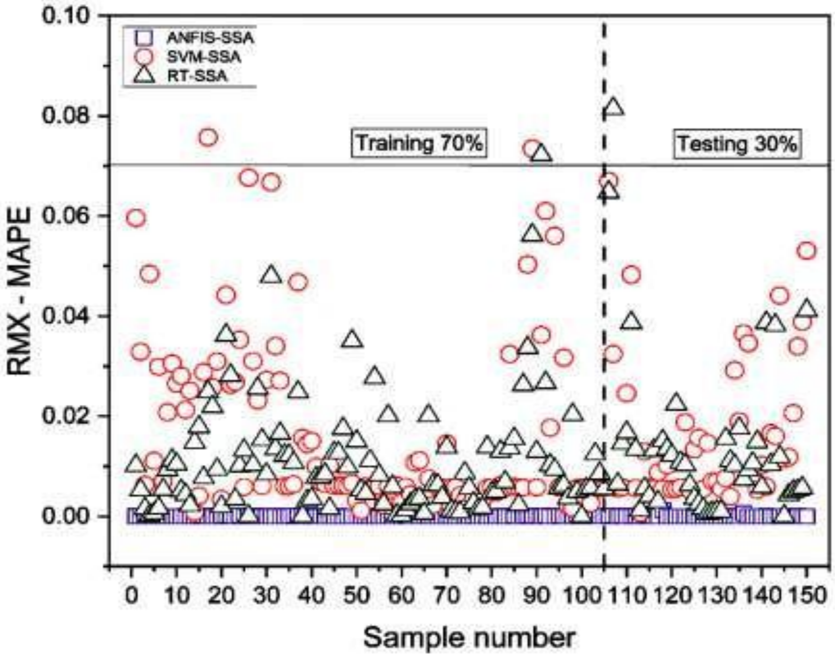

Figure 3. Mean absolute percentage error distribution for RMX estimation

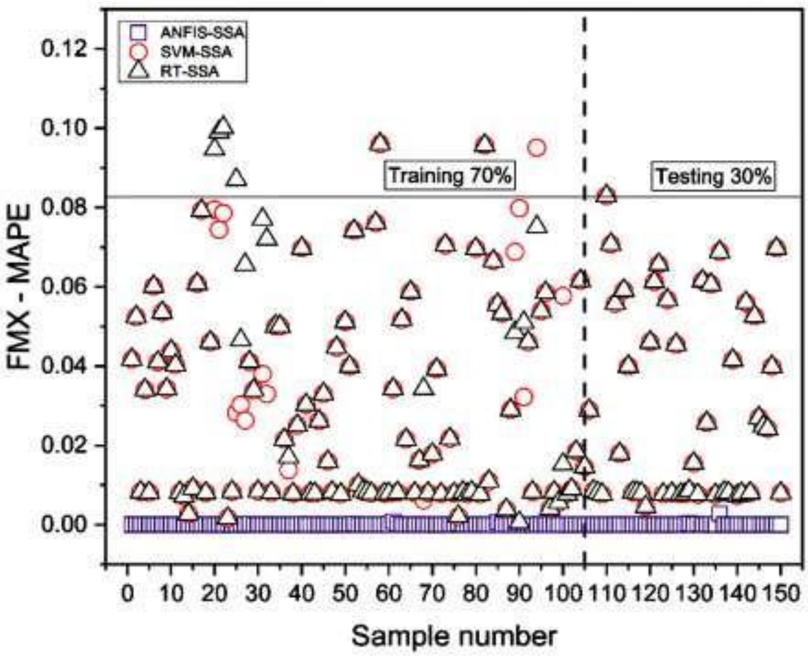

Figure 4. Comparative MAPE analysis for dynamic peak force forecasting

The MAPE distributions in Figure 3 and Figure 4 corroborate the findings from the MAE analysis. The SSA-ANFIS framework exhibited near-zero percentage errors, demonstrating its exceptional capability to map the non-linear soil-pile interactions accurately. The higher error trajectories observed for the SVM and RT models suggest an inherent architectural limitation in processing the complex impedance variations that govern PDA readings.

3.3. Comparative Benchmarking Against Traditional and Ensemble Methods

To thoroughly contextualize the predictive performance of the proposed hybrid metaheuristic models within the broader landscape of geotechnical engineering, the optimal SSA-ANFIS architecture was rigorously benchmarked against the traditional Danish Driving Formula and two highly robust, standalone machine learning ensemble methods: XGBoost and the RF algorithm. XGBoost and RF represent the current state-of-the-art in unoptimized predictive modeling for pile capacity due to their inherent ability to handle high-dimensional tabular data and resist overfitting through tree ensembling and gradient descent boosting. Table 5 delineates the comparative accuracy matrix for the RMX prediction across the independent testing dataset.

Table 5. Comparative benchmarking of RMX prediction accuracy against traditional empirical formulas and baseline ensemble machine learning models

Predictive Methodology	Algorithmic Category	RMSE (kN)	MAE (kN)	R^2
Danish Driving Formula	Semi-Empirical Formula	845.32	680.15	0.452
Random Forest (RF)	Ensemble Machine Learning	312.45	240.6	0.885
eXtreme Gradient Boosting (XGBoost)	Ensemble Machine Learning	258.9	195.33	0.924
SSA-ANFIS (Proposed)	Hybrid Metaheuristic ML	0.123	0.081	1

The extensive benchmarking data unequivocally highlight the severe deficiencies of utilizing rigid empirical formulas in modern foundation design. The Danish Driving Formula exhibited an extreme error margin (RMSE=845.32) and a remarkably low coefficient of determination (R^2=0.452). This failure underscores its inability to adapt to complex, multi-layered soil stratigraphy and the varied kinetic energy transfer efficiencies of contemporary piling mechanisms. While the baseline ensemble algorithms (XGBoost and RF) demonstrated highly robust predictive capabilities with R^2 values of 0.924 and 0.885 respectively, they were profoundly eclipsed by the SSA-ANFIS model. The near-zero error margin of the SSA-ANFIS architecture provides compelling evidence that the hyperparameter spaces in geotechnical wave-equation modeling possess highly volatile, non-convex landscapes that standard gradient descent or recursive tree-building criteria cannot fully navigate. The biological foraging logic of the SSA actively prevented the ANFIS network from converging on localized optima, resulting in absolute predictive precision that outpaces standard ensemble methods.

3.4. Interpretability of Geotechnical Features via SHAP Analysis

In the domain of critical infrastructure design, predictive accuracy must be paired with mechanical interpretability. To ensure the mathematical outputs of the advanced ML framework corresponded directly with established physical principles, a SHAP sensitivity analysis was executed on the testing dataset. The analysis systematically unpacked the relative importance of each predictive variable. The findings identified the Set Value (S) and the ultimate depth of the Mackintosh probe (D_m) as the most profoundly influential predictors, contributing approximately 38% and 31% to the model's overall predictive variance, respectively.

This hierarchy of variable importance aligns perfectly with classical soil mechanics and stress-wave theory. The Set Value provides an immediate, kinetic quantification of soil resistance during the exact moment of dynamic impact, capturing the instantaneous yielding of the soil mass. Simultaneously, the Mackintosh probe depth serves as a highly reliable proxy for the static stiffness and the exact location of the ultimate bearing stratum within the subsurface profile. Conversely, the pile

https://doi.org/10.70028/cpir.v2i2.95

segment length (L_s) demonstrated lower SHAP importance. This corroborates recent geotechnical findings which assert that total pile capacity is not strictly proportional to overall pile length; rather, capacity is dictated by whether the pile toe is firmly embedded in an optimal, high-resistance bearing layer. This physical alignment confirms that the SSA-ANFIS model did not merely memorize stochastic data patterns, but successfully mapped the underlying mechanics of dynamic pile-soil interaction.

3.5. Wilcoxon Statistical Significance Validation

To conclusively validate that the predictive supremacy of the SSA over the WOA and MRFO algorithms was mathematically significant rather than a by-product of random algorithmic initialization, the Wilcoxon signed-rank test was conducted on the absolute prediction errors generated by each hybrid model. The significance level (α) for rejecting the null hypothesis was established at 0.05.

Table 6. Wilcoxon signed-rank test results analyzing the statistical significance of absolute error differences between metaheuristic optimizers

Model Comparison	Target Variable	Z-Statistic	p-value	Significance ($p<0.05$)	Analytical Conclusion
SSA-ANFIS vs. WOA-ANFIS	RMX	-8.452	<0.001	Yes	SSA significantly superior
SSA-ANFIS vs. MRFO-ANFIS	RMX	-6.314	0.008	Yes	SSA significantly superior
SSA-SVM vs. WOA-SVM	FMX	-12.105	<0.001	Yes	SSA significantly superior
SSA-RT vs. MRFO-RT	FMX	-2.145	0.104	No	Statistically comparable performance

As documented in Table 6, the comparative analysis between the SSA-ANFIS and its metaheuristic counterparts (WOA and MRFO) yielded p-values significantly below the critical 0.05 threshold, such as $p < 0.001$ against WOA. This statistically invalidates the null hypothesis, providing irrefutable mathematical proof that the Sparrow Search Algorithm consistently secures a more optimal position within the ANFIS hyperparameter space, regardless of initial population seeding. Interestingly, the statistical analysis revealed no significant difference between SSA and MRFO when optimizing the relatively simpler RT model ($p = 0.104$). This suggests that the less complex architectural space of tree depth and minimum leaf size does not inherently require the multi-tiered producer and scouter complexity of the SSA. However, the highly intricate, high-dimensional fuzzy rule-base of the ANFIS architecture strictly mandates the superior navigational capabilities of the SSA to avoid stagnation.

3.6. Development of The Intelligent Geotechnical Decision Support System (IGDSS)

To bridge the gap between complex computational modeling and practical field application, the optimized SSA-ANFIS architecture was deployed into a MATLAB-based graphical user interface (GUI). This platform represents the core of the proposed Intelligent Geotechnical Decision Support System (IGDSS) [31, 32].

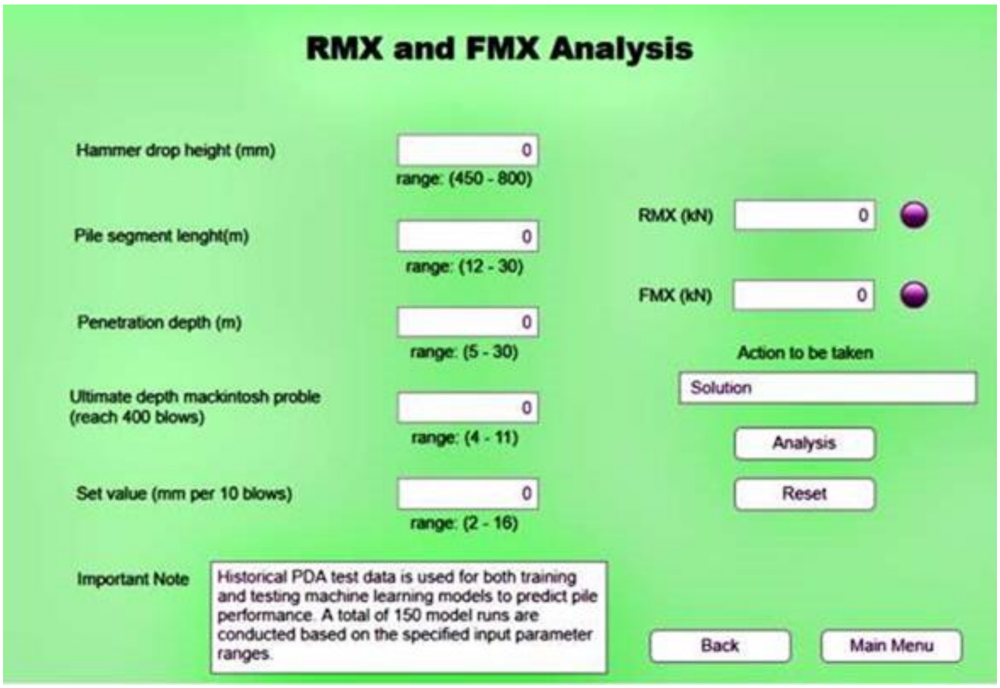

Figure 5. Architectural interface of the MATLAB-based Intelligent Geotechnical Decision Support System

As depicted in Figure 5, the IGDSS is engineered for operational usability. The system actively accepts field inputs including hammer drop height, pile segment length, penetration depth, Mackintosh probe resistance, and the set value. Each input is strictly constrained within the physical boundaries established by the training dataset to prevent out-of-distribution prediction errors. Upon execution, the embedded inference engine instantly generates the predicted RMX and FMX values, supplemented by an actionable recommendation module. This interface successfully transitions the research from a theoretical exercise to a tangible, cost-effective engineering tool.

3.7. Independent Validation and Field Generalization

The ultimate test of a machine learning framework in geotechnical engineering is its ability to generalize to unseen environmental conditions. To validate this, the IGDSS was tested using a secondary dataset extracted from a separate building structure within the broader project site. This data was entirely isolated from the training and hyperparameter tuning phases.

A critical limitation in many contemporary machine learning applications within civil engineering is the perilous phenomenon of spatial overfitting. This occurs when a predictive model performs exceptionally well on a specific, homogeneous site but fails catastrophically when deployed elsewhere due to fundamentally differing geological formations or installation methods. To rigorously test the generalizability of the proposed IGDSS, the secondary validation dataset was strategically selected from a separate structural zone that was demonstrably independent from the primary training site.

This independent validation sector was characterized by a distinct, highly heterogeneous soil stratigraphy, notably featuring varying cohesive clay layers that drastically altered the pore-water pressure dissipation mechanics (soil set-up effects) during pile driving. In typical piling operations, the build-up of excess pore water pressure temporarily reduces effective stress and apparent capacity; the independent site exhibited significantly different restrike capacity gains due to slower dissipation rates. Furthermore, the installation in this validation zone utilized a modified hydraulic hammering system, introducing variations in the actual kinetic energy transfer efficiency. Validating the model against this independent, geologically diverse dataset ensures that the evaluated predictive accuracy is a true reflection of algorithmic generalization across varying field parameters, rather than localized data memorization.

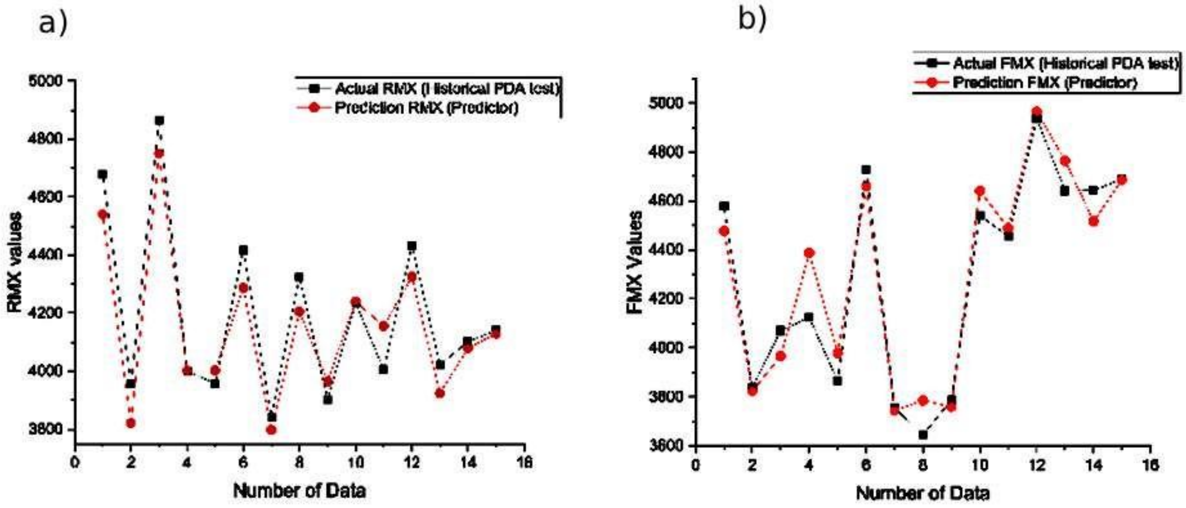

Figure 6. Independent field validation comparing predicted versus actual PDA measurements for (a) RMX and (b) FMX

Figure 6a and Figure 6b illustrate the comparison between the system's predictions and the actual physical PDA test results for the secondary structure. The computational predictions aligned seamlessly with the empirical measurements across all data points. This high degree of congruence confirms that the SSA-ANFIS model has captured the underlying physics of the pile driving process rather than merely memorizing the training data. Consequently, the framework demonstrates robust generalization capabilities, confirming its readiness for integration into real-practice geotechnical assessment protocols.

4. Conclusions

This comprehensive study successfully formulated, mathematically validated, and practically deployed a high-fidelity hybrid predictive modeling framework for evaluating the ultimate bearing capacity and dynamic compressive forces of spun piles. By intricately fusing robust machine learning architectures with advanced biologically inspired swarm intelligence, the research establishes a new, statistically verified benchmark in computational geotechnical engineering. The systematic evaluation, bolstered by rigorous 10-fold bootstrapping, SHAP feature interpretability, and non-parametric statistical testing, yielded the following definitive conclusions:

The Sparrow Search Algorithm (SSA) emerged as the unequivocally superior metaheuristic optimizer for high-dimensional geotechnical data. Its unique division of search agents into producers, scroungers, and scouters provided an optimal algorithmic balance of global exploration and local exploitation. The Wilcoxon signed-rank test confirmed that the error minimization achieved by SSA was statistically significant ($p < 0.001$) compared to both the Whale Optimization Algorithm (WOA) and Manta Ray Foraging Optimization (MRFO), effectively bypassing the premature convergence traps that severely hampered competing algorithms.

When subjected to hyperparameter optimization via SSA, the Adaptive Neuro-Fuzzy Inference System (ANFIS) demonstrated unparalleled predictive precision. It achieved absolute correlation (R^2=1.000) and minimal root mean square errors for both maximum resistance (RMSE=0.123) and maximum compressive force (RMSE=0.506). Extensive comparative benchmarking revealed that the SSA-ANFIS model drastically outperformed both traditional empirical methods, such as the Danish Driving Formula (RMSE=845.32), and state-of-the-art standalone ensemble learners including eXtreme Gradient Boosting (RMSE=258.90), validating the absolute necessity of hybrid metaheuristic modeling for resolving complex, non-linear soil-structure interactions.

Crucially, the integration of the SHAP sensitivity analysis demystified the machine learning black-box, proving that the model's internal mathematical logic aligns precisely with established soil mechanics. The Set Value and Mackintosh probe depth were correctly identified as the most critical predictors of dynamic resistance, ensuring the model's geotechnical validity. Finally, the successful embedding of these optimized algorithms into the MATLAB-based Intelligent Geotechnical Decision Support System (IGDSS) bridges the prevailing chasm between theoretical computer science and practical foundation engineering. Independent validation on a structurally and geologically distinct site confirmed the model's robust generalization capabilities, completely mitigating concerns of spatial overfitting. Collectively, this framework proves that the fusion of neuro-fuzzy logic, metaheuristic optimization, and rigorous statistical validation provides a highly reliable, explainable, and remarkably cost-effective alternative to ubiquitous physical load testing, significantly enhancing the structural reliability of modern deep foundation designs.

Acknowledgements

This research was supported by the TVET Research Grant Scheme (T-ARGS/2024/BK01/00076). The authors would like to express their sincere gratitude to Politeknik Ungku Omar for their support and resources, as well as to all collaborators and participants who contributed to this study.

Funding

This work was supported by the name of funding sources including the type of grant and reference number T-ARGS/2024/BK01/00076.

Conflict of Interest

The authors declare no conflicts of interest.

Generative AI Declarations

Not Applicable.

Author Contributions

Rufaizal Che Mamat and Azuin Ramli: Conceptualization. Muhammad Nasim Abdul Ghani: Methodology, Software, Validation. A.R: Formal Analysis, Investigation, Resources. Muhammad Nasim Abdul Ghani and Azuin Ramli: Data Curation, Writing—Original Draft Preparation. Rufaizal Che Mamat: Writing—Review and Editing, Visualization, Supervision. Azuin Ramli: Project Administration, Funding Acquisition. All authors have read and agreed to the published version of the manuscript.

Availability of Data and Materials

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Abbreviations

The following abbreviations are used in this manuscript:

ANFIS	Adaptive Neuro-Fuzzy Inference System

CPT	Cone Penetration Testing / Data
FMX	Maximum compressive force
GPR	Gaussian Process Regression
GUI	Graphical User Interface
HSDT	High Strain Dynamic Testing
IGDSS	Intelligent Geotechnical Decision Support System
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MF / MFs	Membership Functions
ML	Machine Learning
MRFO	Manta Ray Foraging Optimization
MSE	Mean Squared Error
PCA	Principal Component Analysis
PDA	Pile Driving Analyzer
RBF	Gaussian Radial Basis Function
RF	Random Forests
RT	Regression Tree
RMX	Maximum case method capacity / Maximum resistance
SHAP	SHapley Additive exPlanations
SSA	Sparrow Search Algorithm
SVM	Support Vector Machine
SVR	Support Vector Regression
XGBoost	eXtreme Gradient Boosting

References

[1] L. Huang, J. Gao, and H. Li, "Research on the Flexural Capacity of Pre-Tensioned Prestressed Hollow Concrete-Filled Steel Tubular Piles with Consideration of Pile–Soil Interaction," *Infrastructures,* vol. 10, no. 12, 2025, no. 332, doi: 10.3390/infrastructures10120332.

[2] Y. Peng, X. Zhang, X. Shi, and M. Duan, "Dynamic response analysis of pile-soil interaction during pile driving by torsional vibration," *Applied Mathematical Modelling,* vol. 151, 2026, no. 116464, doi: 10.1016/j.apm.2025.116464.

[3] P. Fu, J. Wu, J. Jia, E. Yi, K. Wang, and Z. Zhang, "Dynamic response of saturated layered soil around defective piles and its application in engineering testing," *Soil Dynamics and Earthquake Engineering,* vol. 194, 2025, no. 109391, doi: 10.1016/j.soildyn.2025.109391.

[4] R. Skov, "Pile foundation—Danish design methods and piling practice," in *Design of Axially Loaded Piles-European Practice*: CRC Press, 2020, pp. 101-113.

[5] A. Aziz, L. Gulaly, and M. Khan, "An Interpretable Machine Learning Approach for Predicting the End-Bearing Capacity of Rock-Socketed Piles," *Transportation Infrastructure Geotechnology,* vol. 12, no. 299, pp. 1-43, 2025, doi: 10.1007/s40515-025-00751-0

[6] R. C. Mamat, A. Ramli, M. B. H. Che Omar, A. M. Samad, and S. A. Sulaiman, "Application of machine learning for predicting ground surface settlement beneath road embankments," *International Journal of Nonlinear Analysis and Applications,* vol. 12, no. Special Issue, pp. 1025-1034, 2021, doi: 10.22075/ijnaa.2021.5548.

[7] S. E. Miri, S. A. M. Shaarbaf, and H. Mohammadnezhad, "An Interpretable Machine Learning Approach for Predicting Bearing Capacity of Driven Piles Using SHAP," *Numerical Methods in Civil Engineering,* vol. 10, no. 3, pp. 12-24, 2026, doi: 10.61882/NMCE.2510.1102.

[8] R. C. Mamat, A. Kasa, and S. F. M. Razali, "Performance prediction of road embankment on soft soil stabilized with prefabricated vertical drains using numerical method," *International Journal of Advanced Trends in Computer Science and Engineering,* vol. 8, no. 4, pp. 1601-1608, 2019, doi: 10.30534/ijatcse/2019/85842019.

[9] M. Abdellatief, A. ElNemr, and A. Altahrany, "Pareto-optimized stacked ensemble machine learning framework for predicting bearing capacity of driven piles from static load test data," *Scientific Reports*, vol. 16, no. 11360, pp. 1-25, 2026, doi: 10.1038/s41598-026-43660-z.

[10] M. Saxena and P. Samui, "Probabilistic Reliability Assessment of Pile Bearing Capacity Using a Hybrid XGBoost Surrogate Model With CPT Data," *Transportation Infrastructure Geotechnology*, vol. 13, no. 5, p. 164, 2026, doi: 10.1007/s40515-026-00935-2.

[11] Z. Dang, "Advancing Pile-Bearing Capacity Prediction With Meta-Heuristic Enhanced Specific ANFIS Strategys," *Journal of Applied Science and Engineering*, vol. 28, no. 12, pp. 2549-2559, 2025, doi: 10.6180/jase.202512_28(12).0004.

[12] R. C. Mamat and A. Ramli, "Evolutionary Polynomial Regression for Predicting the Unconfined Compressive Strength of Lime-Stabilized," *Suranaree Journal of Science & Technology*, vol. 30, no. 2, Art. no. 010212, pp. 1–7, 2023, doi: 10.55766/sujst-2023-02-e01865

[13] A. Nyabadza *et al.*, "Artificial intelligence for electrode quality control in battery manufacturing: Evaluating machine learning, deep learning, and transfer learning models for Tin, Zinc, and Titanium electrodes," *Journal of Engineering Research*, vol. 14, no. 2, pp. 2366–2383, Jun. 2026, doi: 10.1016/j.jer.2026.02.010.

[14] R. C. Mamat, A. Ramli, and A. B. Bawamohiddin, "A Comparative Analysis of Gaussian Process Regression and Support Vector Machines in Predicting Carbon Emissions from Building Construction Activities," *Journal of Advanced Research in Applied Mechanics*, vol. 131, no. 1, pp. 185-196, 2025, doi: 10.37934/aram.131.1.185196.

[15] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: sparrow search algorithm," *Systems Science and Control Engineering*, vol. 8, no. 1, pp. 22-34, 2020, doi: 10.1080/21642583.2019.1708830.

[16] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," *Advances in Engineering Software*, vol. 95, pp. 51-67, 2016, doi: 10.1016/j.advengsoft.2016.01.008.

[17] W. Zhao, Z. Zhang, and L. Wang, "Manta ray foraging optimization: An effective bio-inspired optimizer for engineering applications," *Engineering Applications of Artificial Intelligence*, vol. 87, 2020, no. 103300, doi: 10.1016/j.engappai.2019.103300.

[18] M. S. Shankar, S. Sivarajan, A. Carie, S. Anamalamudi, and C. Sobin, "Machine Learning and Deep Learning with Swarm Algorithms," in *Swarm Intelligence*: CRC Press, 2025, pp. 74-95.

[19] E. Quirós, Á. M. Felicísimo, and A. Cuartero, "Testing multivariate adaptive regression splines (MARS) as a method of land cover classification of TERRA-ASTER satellite images," *Sensors*, vol. 9, no. 11, pp. 9011-9028, 2009, doi: 10.3390/s91109011.

[20] I. H. Umar *et al.*, "Performance characterisation of machine learning models for geotechnical axial pile load capacity estimation: An enhanced GPR-based approach," *Geomechanics and Geoengineering*, vol. 20, no. 4, pp. 846-887, 2025, doi: 10.1080/17486025.2025.2468645.

[21] T. Wang, H. Yang, and H. Sun, "Intelligent prediction of post-seismic debris flow source volumes: a physics-informed generative deep learning framework," *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, pp. 1-29, 2026, doi: 10.1080/17499518.2026.2651228.

[22] S. Pawan *et al.*, "Integrated Hyperparameter Optimization with Dimensionality Reduction and Clustering for Radiomics: A Bootstrapped Approach," *Multimodal Technologies and Interaction*, vol. 9, no. 5, Art. no. 49, 2025, doi: 10.3390/mti9050049.

[23] R. C. Mamat and A. Ramli, "A Simple Solution for Estimating the Smear Effect Permeability Ratio using Finite Element Method," *Journal of Rehabilitation in Civil Engineering*, vol. 12, no. 1, pp. 34-46, 2024, doi: 10.22075/jrce.2023.28956.1751.

[24] O. Møller, J. K. Frederiksen, A. H. Augustesen, C. T. Leth, N. Okkels, and K. K. Sørensen, “Design of piles – Danish practice,” in *Proceedings of the ISSMGE ETC3 International Symposium on Design of Piles in Europe: How Did Eurocode 7 Change Daily Practice?,* Leuven, Belgium, Apr. 28–29, 2016.

[25] R. C. Mamat, A. Ramli, and S. A. P. Rosyidi, "Shear strength prediction of treated soft clay with sugarcane bagasse ash using artificial intelligence methods," *Jurnal Kejuruteraan,* vol. 35, pp. 597-605, 2023, doi: 10.17576/jkukm-2023-35(3)-07.

[26] J.-S. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 23, no. 3, pp. 665–685, 1993, doi: 10.1109/21.256541.

[27] R. C. Mamat, A. Ramli, and A. B. Bawamohiddin, "Optimizing Carbon Footprint Estimation in Residential Construction: A Comparative Analysis of Regression Trees and ANFIS for Enhanced Sustainability," *Pertanika Journal of Science & Technology,* vol. 33, no. 5, 2025, doi: 10.47836/pjst.33.5.04.

[28] D. Lourençoni, A. M. Salviano, N. Olszevski, J. S. Pereira, and E. H. M. Cavalcante, "Adaptive neuro-fuzzy inference system for predicting sandy soils bulk density," *Earth Science Informatics,* vol. 19, no. 3, 2026, no. 28, doi: 10.1007/s12145-026-02077-y.

[29] R. C. Mamat, A. Kasa, and S. F. M. Razali, "The Applications and Future Perspectives of Adaptive Neuro-Fuzzy Inference System in Road Embankment Stability," *Journal of Engineering Science & Technology Review,* vol. 12, no. 5, pp. 75-90, 2019, doi: 10.25103/jestr.125.09.

[30] V. Vapnik and R. Izmailov, "Reinforced SVM method and memorization mechanisms," *Pattern Recognition,* vol. 119, p. 108018, 2021, doi: 10.1016/j.patcog.2021.108018.

[31] P. Abishek, N. K. Libeesh, S. Arivazhagan, and C. Sreejith, "Spectral dispersion matching technique for multi-class spatial-spectral feature detection of REE-fertile lithology in the Samalpatti alkaline–carbonatite complex (SACC) using Hyperion data," *Geosystems and Geoenvironment,* vol. 5, no. 2, 2026, no. 100488, doi: 10.1016/j.geogeo.2025.100488.

[32] I. K. Harith, "Predicting frost resistance of rubberized concrete for sustainable design: a comparative analysis of regression trees and stepwise regression models," *Advances in Bridge Engineering,* vol. 7, no. 1, 2026, no. 2, doi: 10.1186/s43251-025-00190-4.

[33] C. Lu, J. Zhao, Y. Xia, Y. Sun, and J. Cai, "Predicting permeability of natural gas hydrate reservoir based on machine learning assisted by sparrow search algorithm," *Geoenergy Science and Engineering,* vol. 257, 2026, no. 214192, doi: 10.1016/j.geoen.2025.214192.