





OPEN

Investigating a hybrid extreme learning machine coupled with Dingo Optimization Algorithm for modeling liquefaction triggering in sand-silt mixtures

Mohammed Majeed Hameed^{1,2}, Adil Masood³, Aman Srivastava⁴, Norinah Abd Rahman^{5,6}, Siti Fatin Mohd Razali^{5,6}, Ali Salem^{7,8} & Ahmed Elbeltagi⁹

Liquefaction is a devastating consequence of earthquakes that occurs in loose, saturated soil deposits, resulting in catastrophic ground failure. Accurate prediction of such geotechnical parameter is crucial for mitigating hazards, assessing risks, and advancing geotechnical engineering. This study introduces a novel predictive model that combines Extreme Learning Machine (ELM) with Dingo Optimization Algorithm (DOA) to estimate strain energy-based liquefaction resistance. The hybrid model (ELM-DOA) is compared with the classical ELM, Adaptive Neuro-Fuzzy Inference System with Fuzzy C-Means (ANFIS-FCM model), and Sub-clustering (ANFIS-Sub model). Also, two data pre-processing scenarios are employed, namely traditional linear and non-linear normalization. The results demonstrate that non-linear normalization significantly enhances the prediction performance of all models by approximately 25% compared to linear normalization. Furthermore, the ELM-DOA model achieves the most accurate predictions, exhibiting the lowest root mean square error (484.286 J/m³), mean absolute percentage error (24.900%), mean absolute error (404.416 J/m³), and the highest correlation of determination (0.935). Additionally, a Graphical User Interface (GUI) has been developed, specifically tailored for the ELM-DOA model, to assist engineers and researchers in maximizing the utilization of this predictive model. The GUI provides a user-friendly platform for easy input of data and accessing the model's predictions, enhancing its practical applicability. Overall, the results strongly support the proposed hybrid model with GUI serving as an effective tool for assessing soil liquefaction resistance in geotechnical engineering, aiding in predicting and mitigating liquefaction hazards.

Keywords Liquefaction, Earthquake, Dingo Optimization Algorithm, Non-linear normalization

The concept of liquefaction in soil science revolves around the phenomenon wherein soil undergoes a loss of strength and stiffness, adopting liquid-like properties. This usually occurs during an earthquake or other seismic activities when a sudden increase in pore water pressure within the soil causes it to lose its ability to support dynamic loading. When liquefaction happens, the ground essentially turns into a fluid-like state, leading to the sinking, tilting, or sliding of structures built on the affected soil. Soil liquefaction increases the intensity of the damage caused by earthquakes. Assessment of the various literature surveys on liquefaction-induced

¹Department of Civil Engineering, Al-Maarif University College, Ramadi, Iraq. ²Department of Computer Science, Al-Maarif University College, Ramadi, Iraq. ³Department of Natural and Applied Sciences, TERI School of Advanced Studies, New Delhi, India. ⁴Department of Civil Engineering, Indian Institute of Technology (IIT) Kharagpur, Kharagpur, West Bengal 721302, India. ⁵Department of Civil Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia. ⁶Smart and Sustainable Township Research Centre (SUTRA), Universiti Kebangsaan Malaysia (UKM), 43600 Bangi, Selangor, Malaysia. ⁷Civil Engineering Department, Faculty of Engineering, Minia University, Minia 61111, Egypt. ⁸Structural Diagnostics and Analysis Research Group, Faculty of Engineering and Information Technology, University of Pécs, Pécs, Hungary. ⁹Agricultural Engineering Department, Faculty of Agriculture, Mansoura University, Mansoura 35516, Egypt. ✉email: mohmmag1@gmail.com; salem.ali@mik.pt.e.hu

devastation reveals five significant occurrences, namely the earthquakes in Niigata in 1964, Alaska in 1964, Kobe in 1995, Christchurch in 2011, and Sulawesi in 2018^{1–5}. Under the initial stress conditions, seismic waves generated during an earthquake cause the ground to shake, subjecting the soil to dynamic loading in the form of cyclic shear stress. This cyclic loading rearranges soil particles, increasing pore water pressure within the soil due to the inability of water to escape quickly. As pore water pressure rises, effective stress that holds the soil particles together, gradually decreases until it reaches zero or becomes negative, causing the soil to lose its cohesion and behave like a fluid. Structures atop liquefied soil may sink or tilt, resulting in significant damage. After the seismic event ceases, excess pore water pressure gradually dissipates, but the damage remains. The occurrence of liquefaction is contingent upon several factors within the soil, including its composition, density, saturation level and intensity of seismic shaking^{6,7}.

To mitigate the potential loss of life and property, it is crucial to conduct a liquefaction hazard analysis at sites vulnerable to this phenomenon as a fundamental step in seismic risk management⁸. Numerous studies have been conducted to improve the precision of liquefaction prediction methods; discussed here. The stress-based approach is a widely used method for assessing liquefaction potential. It determines the soil's likelihood of liquefying by examining its stress state during seismic loading. It estimates the cyclic shear stress imposed on the soil and compares it with the soil's resistance, often considering factors such as shear stress history and soil properties^{9–11}. The energy-based approach focuses on the energy imparted by seismic waves and compares it with the energy dissipated by the soil during cyclic loading. This method considers factors such as stress–strain hysteresis loops to assess liquefaction susceptibility^{12–14}. There are techniques like the empirical methods, which rely on observed correlations between liquefaction and factors such as standard penetration test (SPT) N-values or shear wave velocity^{15–17}. Empirical methods often offer practical solutions based on historical data but may lack precision in diverse geological settings.

Numerical modeling techniques, such as finite element analysis and discrete element modeling, enable detailed simulations of soil behavior, accounting for complex factors like soil heterogeneity and three-dimensional stress states^{18,19}. However, the phenomenon of liquefaction is inherently nonlinear, and the existing empirical and statistical models found in the literature have limitations in accurately predicting and achieving desirable outcomes. Additionally, certain approaches, such as model-based finite element methods, often rely on numerous assumptions and extensive data requirements. To enhance prediction accuracy and flexibility, Machine Learning (ML) models can offer a viable alternative to the existing techniques. This is primarily attributed to the capability of ML models to identify intricate relationships between input features and output^{20,21}. Also, ML models can effectively capture complex patterns and dependencies, allowing them to adapt and learn from the data, ultimately leading to improved prediction accuracy and a more flexible approach in various domains^{22–24}.

In the field of liquefaction prediction, various ML techniques have recently seen a surge in adoption for assessing liquefaction potential²⁵. Some researchers have developed support vector regression (SVR) and extreme gradient boosting (XGBoost) based models to predict liquefaction using various soil and seismic^{26,27}. These studies have demonstrated promising results in terms of predictive accuracy. Moreover, the use of artificial neural network (ANN) prediction-based models, and adaptive neuro-fuzzy inference system (ANFIS) models have been implemented to forecast soil liquefaction, resulting in improved prediction accuracy^{28–30}. Other researchers have investigated the potential of extreme learning machines (ELM), a modified version of ANN, for predicting soil liquefaction³¹. Their findings indicated that ELM exhibited a higher prediction capacity, achieving an accuracy of 87.5. Given the strong predictive performance of classical ML models, the challenge lies in finding the optimal hyperparameters that significantly impact model predictions. To address this, geotechnical researchers have turned to various metaheuristic algorithms, including genetic algorithm (GA), grey wolf optimization (GWO), and differential evolution (DE)^{32,33}. By incorporating these algorithms into hybrid models, researchers have observed a substantial improvement in prediction accuracy³⁴.

In recent years, the integration of machine learning techniques into geotechnical engineering has shown promising potential for addressing complex soil mechanics phenomena. A recent study has explored the application of ANN based models for estimating geotechnical properties of lime-stabilised clayey soils, demonstrating the versatility of machine learning approaches in geotechnical engineering-based problems³⁵. Other researchers have investigated machine learning methods for predicting the maximum dry density of soil samples³⁶. It was observed that the constructed models exhibited significant generalization capability when predicting correlations among soil properties. Besides, other scholars have employed various state-of-the-art modeling techniques for estimating the ground liquefaction potential, highlighting the utility of ML approaches in framing soil liquefaction decisions under given conditions³⁷. Moreover, some investigations have explored the role of tunnel boring machine (TBM) tail-grouting on surface settlement in coarse-grained soils, emphasizing the importance of numerical modeling and field testing in geotechnical engineering practices³⁸. In addition, several researchers demonstrated the effectiveness of supervised learning methods in modeling unconfined compressive strength prediction of stabilized soils, showcasing the potential of ML techniques in simulating geotechnical properties under real conditions³⁹. Moreover, other studies have predicted the potential for earthquake-induced liquefaction in fine-grained soil with minimal uncertainty and human intervention, showcasing the effectiveness of machine learning methods in evaluating liquefaction prediction performance for practical earthquake engineering applications^{40,41}. These studies collectively illustrate the wide-ranging applicability of ML methods in geotechnical engineering domain, motivating further exploration of ML-based approaches for tackling soil liquefaction prediction challenges.

According to the available research⁴², there is a recognized shortage of studies in the prediction of soil liquefaction potential within the field of geological/geotechnical engineering⁴³. Consequently, in the present study, liquefaction potential was assessed using a ML-based approach which integrated the Extreme Learning Machine (ELM) with a novel meta-heuristic algorithm referred to as the Dingo Optimization Algorithm (DOA).

The DOA played a crucial role in optimizing the proposed hybrid model (ELM-DOA), resulting in improved performance and efficacy⁴⁴.

The study further explores the application of non-linear normalization technique and compares the accuracy of ML models using this technique with the commonly used linear normalization approach. Furthermore, the prediction accuracy of the ELM-DOA model has been compared with classic ELM, ANFIS, and other ML models. The research also aims to investigate the use of non-linear normalization in ML model development and evaluate its predictive accuracy compared to the widely used linear normalization method. It seeks to highlight the potential advantages and effectiveness of non-linear normalization techniques in improving ML model performance. Finally, the study introduces a graphical user interface (GUI) tailored to aid engineers and researchers in effectively utilizing the developed model.

Methods and materials

Data collection

The energy dissipation theories were first introduced by⁴⁵ who provided an understanding of the pore water pressure generation inside the soil skeleton. The dissipated energy (W) required to trigger liquefaction in a soil volume can be obtained from undrained cyclic triaxial tests, which are highly dependent on effective confining pressure. In terms of the silt-sand mixture, several parameters taken into consideration for predicting soil composition are relative density (D_r), uniformity coefficient (C_u), mean grain size (D_{50}), initial mean effective confining pressure (σ_c), and percentage of fine content (FC). The database adopted in this study is based on laboratory tests conducted during previous studies^{46–49}. Including these parameters is essential for predicting soil composition and liquefaction resistance in sand-silt mixtures. Recent research provides insights into identifying and benchmarking significant factors related to soil liquefaction, further supporting the discussion on the selection of input parameters⁵⁰. For instance, relative density (D_r) reflects the compactness of soil particles, which directly influences soil strength and susceptibility to liquefaction. Similarly, parameters such as uniformity coefficient (C_u) and mean grain size (D_{50}) offer insights into soil gradation and particle size distribution, affecting soil stability and deformation characteristics during cyclic loading. Initial mean effective confining pressure (σ_c) plays a crucial role in confining soil particles and controlling pore water pressure generation, while the percentage of fine content (FC) influences soil cohesion and internal friction angle, impacting soil liquefaction potential.

The dataset used in this study comprises 144 samples (see Tables S-1 and S-2), which were utilized to train the energy-based liquefaction resistance of sand-silt mixtures. Out of these samples, 100 were allocated for training the models, while the remaining samples were used for testing purposes. In this study, the data was divided into training and testing sets randomly, following a widely used approach in previous works^{51–55} to reduce bias between the two sets. To ensure repeatability in the data division process, a specific random seed value was employed and recorded. This allows for the exact replication of the division, ensuring consistent results. After this step, the training and testing data were stored and used for training the ML models. It is worth noting that the overall data, including the training and testing sets, are provided in the ESM Appendix. Importantly, there is no repeated data in the dataset, as each observation is uniquely assigned to either the training or testing set. This ensures the integrity and validity of the data used for model training and evaluation.

The statistics in Table 1 include the maximum (X_{max}), average (X_{av}), minimum (X_{min}), standard deviation ($X_{st.d}$), skewness (S_{sks}), and median (X_{med}) values for each parameter. Table 1 shows that the statistical values

Category	Statistics	Sigma c (Kp)	Dr (%)	FC (%)	Cu	D50 (mm)	W (J/m ³)
Training samples	X_{max}	400	93	60	28.120	0.260	12,230
	X_{av}	103.470	62.807	14.710	8.101	0.165	2372.130
	X_{min}	14.130	9.380	0	1.520	0.029	385
	$X_{st.d}$	83.872	15.641	17.544	9.655	0.056	2156.920
	S_{sks}	2.531	-0.657	0.997	1.349	-0.403	2.285
	X_{med}	100	64.595	5	2.270	0.15	1705.450
Testing Samples	X_{max}	400	88.020	60	28.120	0.26	15,000
	X_{av}	92.13	60.505	13.627	6.277	0.155	2075.800
	X_{min}	28.400	25	0	1.520	0.029	398.200
	$X_{st.d}$	75.647	13.078	17.239	8.293	0.049	2545.110
	S_{sks}	3.129	-0.572	0.989	2.004	0.067	3.452
	X_{med}	82.740	62	5	2.270	0.15	1317
Overall samples	X_{max}	400	93	60	28.120	0.26	15,000
	X_{av}	100.060	62.114	14.384	7.552	0.162	2283.02
	X_{min}	14.130	9.380	0	1.520	0.029	385
	$X_{st.d}$	81.389	14.909	17.4	9.275	0.054	2275.520
	S_{sks}	2.6837	-0.609	0.996	1.515	-0.268	2.744
	X_{med}	100	62.700	5	2.27	0.15	1609

Table 1. Statistical analysis of the experimental data.

(mean, maximum, median, st. dev, etc.) of the predictors (e.g., D50, Cu, Dr, FC, and Sigma c) in the training dataset are very similar to those in the testing dataset. Regarding liquefaction resistance (W), the average value in the training dataset is slightly higher than in the testing dataset by approximately 12%. However, the minimum value in the training dataset is lower than the corresponding value in the testing dataset by approximately 3.430%. These findings indicate a minor variation in the specific variable but do not suggest any significant bias in the overall data division procedure. Furthermore, the dataset has an overall maximum capacity energy of $15,000 \text{ J/m}^3$, an average capacity energy of 2283.021 J/m^3 , and a minimum capacity energy of 385 J/m^3 . Overall, such statistics offer a synopsis of the dataset, providing an understanding of the range, distribution, and central tendency of various parameters pertaining to the energy-based liquefaction resistance of sand-silt mixtures.

Extreme learning machine

The Extreme Learning Machine (ELM) stands out in machine learning for its simplicity and efficiency. Positioned within the category of feedforward neural networks, ELM distinguishes itself through a single-layer architecture, a departure from the conventional multi-layer structures found in traditional neural networks⁵⁶. Notably, the connections between input and hidden layer neurons are randomly initialized^{20,57}, contributing to the algorithm's adaptive nature. The main structure of the ELM can be illustrated in Fig. 1⁵⁸.

ELM's learning process is marked by its remarkable speed, which makes it particularly adept at handling large datasets. The weights calculation, crucial for the model's performance, is achieved through a single learning iteration, eliminating the need for iterative optimization⁵⁹. Despite its streamlined approach, ELM exhibits robust generalization capabilities, allowing it to navigate complex patterns within data effectively^{58,60,61}. Equations (1–3) provide a breakdown of the equations associated with the ELM. Each neuron's output in the hidden layer is determined by the weighted sum of the input data ($w_i x$) plus a bias term (b_i), passed through an activation function g , as shown in Eq. (1). Here, H_i represents the output of the hidden neuron i ; g is the activation function, typically a sigmoid, Gaussian, or ReLU function; w_i denotes the input weights connected to neuron i ; x is the input data vector; b_i represents the bias term for neuron i .

$$H_i = g(w_i x + b_i) \quad (1)$$

The output layer weights are calculated by performing a pseudo-inverse operation (H^+) on the hidden layer output matrix (H) multiplied by the target output matrix (T), as shown in Eq. (2)⁶². Here, Y represents the predicted output matrix; H is the hidden layer output matrix obtained from the input data; $beta$ denotes the output layer weights. The final predicted output (Y) is obtained by multiplying the hidden layer output matrix (H) with the calculated output layer weights ($beta$), as shown in Eq. (3)⁶³.

$$beta = H^+ \times T \quad (2)$$

$$Y = H \times beta \quad (3)$$

These equations form the core of an ELM's functioning, where the network learns the mapping from input data to output by initializing input weights randomly and calculating output weights analytically without iterative optimization⁶⁴. This approach results in fast learning and prediction capabilities in ELM. The suitability of ELM

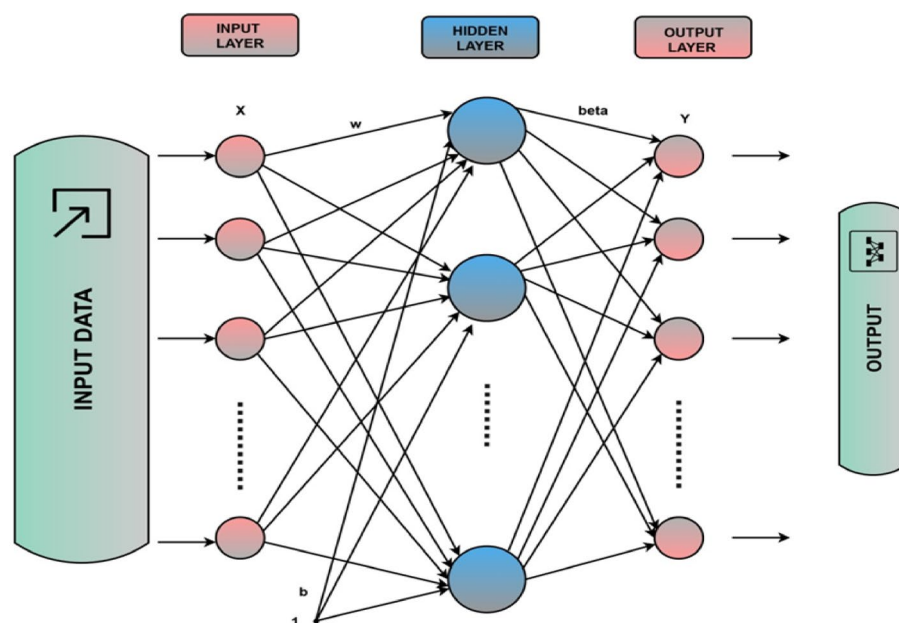


Figure 1. The basic structure of the ELM model.

for liquefaction resistance prediction in sand-silt mixtures becomes evident in its ability to capture intricate relationships within the dataset. This adaptability extends to various data types, including geological data, aligning well with the complexity inherent in liquefaction resistance prediction. In the context of the hybrid model, ELM serves as the foundational machine learning algorithm, and its integration with the Dingo Optimization Algorithm (DOA) elevates its predictive accuracy. With a proven track record in various fields, ELM demonstrates versatility and reliability, although it is essential to note its limitations in interpretability, particularly when compared to more complex models^{31,56,60,65,66}. In summary, the integration of ELM in the hybrid model underscores its efficiency, rapid learning capabilities, and adaptability to the intricacies of liquefaction resistance prediction.

Dingo Optimization Algorithm

The Dingo Optimization Algorithm (DOA) introduces a novel meta-heuristic strategy inspired by the cooperative hunting behavior of dingoes⁴⁴. As a meta-heuristic algorithm, DOA emulates the collaborative hunting approach employed by dingoes in the wild. Its core characteristics involve leveraging dingo-inspired search principles incorporating concepts such as scent marking, exploration, and cooperation into its search strategy⁶⁷. Operating on a population-based approach, the algorithm maintains a collection of solutions, each representing a potential optimization solution. DOA's evolutionary components emulate the cooperative behavior of dingoes through scent marking, serving as an analogy for information sharing among individuals in the population. The algorithm adapts and evolves over iterations, mirroring the dynamic nature of dingo behavior during hunting⁶⁷⁻⁶⁹. DOA is an enhancement strategy integrated with the hybrid model's Extreme Learning Machine (ELM). This integration creates a synergistic relationship, introducing adaptability to the learning process of ELM, ultimately contributing to improved convergence and effectiveness in predicting liquefaction resistance.

DOA's principles are more conceptual and inspired by dingo behavior rather than having explicit mathematical equations. The algorithm operates based on the coordination, movement, scent marking, and cooperative strategies observed in dingoes during hunting. While it lacks precise mathematical formulations, DOA utilizes these principles in a population-based approach to iteratively improve solutions to optimize an objective function⁷⁰. Each "iteration" in DOA involves updates to solution positions influenced by concepts derived from dingo behavior, contributing to an evolutionary optimization process. The conceptual representation of the iterative updates in DOA is provided below:

(A) *Iterative update process in DOA (conceptual representation)*

- Movement and coordination: In each iteration, dingoes (representing solutions) coordinate their movements based on scent marking and interactions within the population.
- Solution position: Analogous to the dingo movement, each solution's position is updated within the search space to explore and exploit potential areas that optimize the objective function.

(B) *Scent marking and cooperation*

- Scent marking analogy: Similar to dingo communication through scent marking, solutions share information within the population, influencing each other's movement and decision-making.
- Cooperative strategies: Solutions interact cooperatively by sharing information or influencing each other's positions in the search space, akin to dingoes' cooperative hunting behavior.

(C) *Population-based approach*

- Population initialization: The algorithm starts with an initial population of solutions representing potential solutions to the optimization problem.
- Evolutionary process: Over iterations, solutions evolve and adapt within the search space, driven by the principles inspired by dingo behavior, aiming to optimize the objective function.

In summary, the DOA methodology operates based on principles inspired by dingo behavior, integrating coordination, movement, scent marking, and cooperative strategies observed in dingoes during hunting. Each iteration involves updating solution positions within the search space, influenced by interactions and information sharing akin to dingo behavior⁷¹. These principles guide the evolutionary optimization process, aiming to iteratively improve solutions toward optimizing the objective function without specific mathematical formulations. The algorithm's effectiveness lies in its conceptual implementation inspired by natural behaviors rather than explicit mathematical equations.

Beyond its application in the hybrid model, DOA exhibits versatility and robustness in tackling a broad spectrum of optimization problems⁷². Drawing inspiration from nature, the algorithm efficiently explores solution spaces, balancing exploration and exploitation. However, DOA's effectiveness may vary depending on the nature of the optimization problem, emphasizing its problem-specific performance. Sensitivity to parameter choices necessitates careful configuration, and the algorithm's relatively recent emergence in the field of meta-heuristics positions it as an area of emerging research with applications across diverse domains⁶⁷⁻⁶⁹. In summary, the integration of DOA with ELM showcases an evolutionary approach that enhances the overall capabilities of the liquefaction resistance prediction model.

Adaptive Neuro-Fuzzy Inference System (ANFIS)

The Adaptive Neuro-Fuzzy Inference System (ANFIS) is a hybrid computational model that combines the adaptability of neural networks with the interpretability of fuzzy logic. It is designed to model complex relationships and patterns within data effectively. ANFIS integrates fuzzy logic systems with neural networks, utilizing a learning algorithm to adjust its parameters and adapt to the underlying data distribution. This adaptability makes ANFIS well-suited for modeling nonlinear systems, pattern recognition, and solving complex decision-making problems. The synergy between fuzzy logic and neural networks in ANFIS provides a robust framework for capturing and interpreting intricate relationships in diverse datasets^{73–76}. In ANFIS, membership functions determine how well input data belongs to different fuzzy sets. These memberships are combined to calculate the firing strengths of individual rules, which are then normalized to provide the contribution of each rule to the output, as shown in Eqs. (4–6). The membership grade $[\mu_i(x)]$ for input data x in fuzzy set i is calculated using Gaussian or other membership functions, as shown in Eq. (4)⁷⁷. Here, c_i represents the center of the membership function for the fuzzy set i ; σ_i is the width parameter; m is a fuzziness exponent affecting the shape of the membership function.

$$\mu_i(x) = \frac{1}{1 + \left(\frac{x-c_i}{\sigma_i}\right)^{2m}} \quad (4)$$

The firing strength $[w_i(x)]$ of rule i is determined by the product of individual input memberships, as shown in Eq. (5). n represents the number of input variables⁷⁸. $\mu_{ij}(x)$ denotes the membership grade of input x in the j th fuzzy set of the i th rule.

$$w_i(x) = \prod_{j=1}^n \mu_{ij}(x) \quad (5)$$

The normalized firing strength $[w_i^*(x)]$ is calculated to determine the contribution of rule i to the overall output, as shown in Eq. (6). Here, N denotes the total number of rules.

$$w_i^*(x) = \frac{w_i(x)}{\sum_{i=1}^N w_i(x)} \quad (6)$$

ANFIS with fuzzy C-means (ANFIS-FCM)

The ANFIS-FCM hybrid model represents a sophisticated approach to liquefaction resistance prediction in sand-silt mixtures by synergizing the ANFIS with Fuzzy C-Means (FCM) clustering. ANFIS is acknowledged for its adaptability, employing a learning mechanism to adjust fuzzy inference system parameters dynamically⁷⁹. This adaptability is enhanced by the integration of fuzzy logic, providing crucial interpretability that aids in understanding the rationale behind predictions^{58,59,80}. On the other hand, FCM contributes a robust clustering strategy, dividing the dataset into fuzzy partitions based on similarity and utilizing membership functions to assign degrees of belongingness to each cluster. This data segmentation capability enhances the model's capacity to capture nuanced relationships within the sand-silt mixture data^{81–83}. More specifically, FCM aims to cluster data points by iteratively updating the membership matrix based on the distances between data points and cluster centroids. This process continues until the membership matrix stabilizes, assigning data points to clusters, as shown in Eq. (7)^{58,59,80}. Here, u_{ij} represents the membership of data point i in cluster j ; c is the number of clusters; x_i denotes the i -th data point; v_j is the centroid of cluster j ; m is a fuzziness parameter (typically set between 1 and 2).

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|}\right)^{\frac{2}{m-1}}} \quad (7)$$

Integrating ANFIS and FCM in the hybrid model forms a powerful enhancement strategy. By combining adaptive neuro-fuzzy inference with data-driven clustering, ANFIS-FCM offers a comprehensive modeling ability that captures intricate relationships, providing a nuanced understanding of liquefaction resistance prediction. In the broader context, ANFIS-FCM contributes to the evolutionary approach of the hybrid model, showcasing its role in enhancing adaptability and efficiency in liquefaction resistance prediction^{81–83}. The hybrid model, with its advanced optimization and predictive capabilities, offers improved accuracy and insights into the liquefaction resistance of sand-silt mixtures⁸⁴.

ANFIS with sub-clustering (ANFIS-sub)

The ANFIS with sub-clustering (ANFIS-sub) hybrid model represents an innovative approach to enhance the accuracy of liquefaction resistance prediction in sand-silt mixtures. Rooted in the ANFIS, the model capitalizes on its renowned adaptability, merging the strengths of fuzzy logic and neural networks to intricately model relationships within the dataset. ANFIS dynamically adjusts its fuzzy inference system parameters through a learning mechanism, fostering adaptability to the underlying patterns in the sand-silt mixture data. Integrating fuzzy logic enhances interpretability, providing valuable insights into the rationale behind predictions^{60–62}. ANFIS-Sub introduces a sub-clustering technique to refine the clustering process within the dataset further. This refinement allows for a more granular analysis by breaking down clusters into subgroups, capturing finer distinctions within the sand-silt mixture data. The hybrid model strategically integrates ANFIS with sub-clustering techniques, creating a synergy that leverages adaptive neuro-fuzzy inference and fine-grained sub-clustering.

This integration enhances the model's ability to represent complex patterns in liquefaction resistance prediction, contributing to a more accurate and nuanced understanding of the task^{85–87}. To emphasize further, Sub-clustering techniques aim to further refine clusters created by the ANFIS membership functions. This refinement involves updating cluster centroids and memberships within clusters to capture finer distinctions within the data. In a Sub-clustering process, the centroids of clusters may be updated iteratively. For example, in the case of K-Means sub-clustering, the centroid v_j of cluster j can be updated using Eq. (8)^{60–62}. Here, v_j represents the updated centroid of cluster j ; N denotes the total number of data points. u_{ij} represents the membership of data point i in cluster j ; x_i is the i -th data point. The membership grades (u_{ij}) within clusters are recalculated based on the distances between data points and cluster centroids using Eq. (7).

$$v_j = \frac{\sum_{i=1}^N u_{ij} \times x_i}{\sum_{i=1}^N u_{ij}} \quad (8)$$

ANFIS-Sub's accuracy is systematically compared with classic ELM and other ANFIS-based models in the evaluation phase. The comparison serves as a robust benchmark for predictive performance, employing rigorous performance metrics such as accuracy, precision, and recall to comprehensively assess and compare each model's accuracy. In the broader evolutionary approach of the hybrid model, ANFIS-Sub plays a crucial role, contributing to enhanced adaptability and efficiency in liquefaction resistance prediction. In summary, the ANFIS with Sub-clustering (ANFIS-Sub) hybrid model not only advances the accuracy of prediction but also aligns with the overarching evolutionary approach, showcasing its pivotal role in enhancing the scientific rigor of liquefaction resistance prediction methodologies^{85–87}.

Model development

This study investigates the effectiveness of a hybrid model, named ELM-DOA, which combines the ELM algorithm with the DOA algorithm, for predicting the required strain energy to initiate liquefaction in sand-silt mixtures. The DOA algorithm is employed to optimize the parameters of ELM, such as weight and bias values, by minimizing the objective function, which is the root mean square error (RMSE). The DOA algorithm in the ELM-DOA model employs a nature-inspired approach to optimization by mimicking the hunting behavior of dingoes. It iteratively explores the search space by updating the weight and bias values of the ELM model based on fitness evaluation. During each iteration, the algorithm strikes a balance between exploration and exploitation⁴⁴. Exploration allows the algorithm to search a wide range of parameter values, enabling it to discover potentially better solutions in different regions of the search space. Exploitation focuses on refining the parameter values around promising regions that have shown good fitness, aiming to exploit the local information and further improve the model's performance. The DOA algorithm continues this iterative process until a termination criterion is met. The termination criterion could be a predefined number of iterations or a threshold for improvement in fitness. By reaching convergence or meeting the termination criterion, the algorithm ensures that the ELM-DOA model is optimized to the best extent possible within the given search space. By combining exploration and exploitation strategies, the DOA algorithm effectively optimizes the ELM model by finding the optimal weight and bias values that minimize the objective function (RMSE). This optimization process enhances the performance of the ELM-DOA model in predicting the required strain energy for liquefaction in sand-silt mixtures.

To assess the accuracy and reliability of the hybrid model, it is compared with well-known benchmark models including ANFIS-FCM, classical ELM, and ANFIS-Sub models. The hyperparameters of these models are determined through a trial-and-error approach aimed at minimizing the objective function (RMSE). Once the optimal condition is reached, the training process is halted, and the results are saved for future comparison. The development process of the prediction models is illustrated in Fig. 2. It should be noted that the data used in this study is randomly divided into two phases: a training phase, which constitutes 70% of the total collected samples, and a testing phase, which utilizes the remaining 30% of the samples. The applied random division (70/30) is commonly employed in regression-related machine learning studies, as observed in previous works^{88–91}. The 70% training data allocation allows for sufficient data to train the model, while the 30% testing data provides ample data for evaluating the model's generalization capabilities. This division strategy achieves a balanced distribution of data between training and testing, facilitating accurate model training and reliable assessment of its performance.

In order to ensure fair feature comparison, two normalization procedures have been employed in this paper. The first procedure involves classic normalization, where features and labels are linearly scaled to fit within the range of zero to one using the following equation (Eq. 9):

$$Xn_i = \frac{X_i - \alpha}{\beta - \alpha} \quad (9)$$

where, Xn_i is the normalized vector of i th the sample X is the vector that is needed to normalize while, the coefficients β , and α are maximum and minimum values which computed from training data set.

In this study, a random data division procedure was used to split the data into training and testing phases, which is widely employed to minimize bias in previous researches. Avoiding bias between training and testing data is crucial for a machine learning model to possess strong generalization capacity. Comparing the strain energy-based liquefaction resistance between the training and testing data, as shown in Fig. 3, it was observed that the normalized liquefaction resistance values less than 0.1 accounted for approximately 52% in the training set, similar to around 56.8% in the testing set, indicating similar characteristics for lower values in both datasets. Moving to the Less than 0.6 (<0.6) interval, both the training and testing sets covered a higher percentage of the total values, with the training set representing 96% and the testing set representing 97.67% of the total values, indicating a more significant representation of hazard and non-hazard events within this interval. Furthermore,

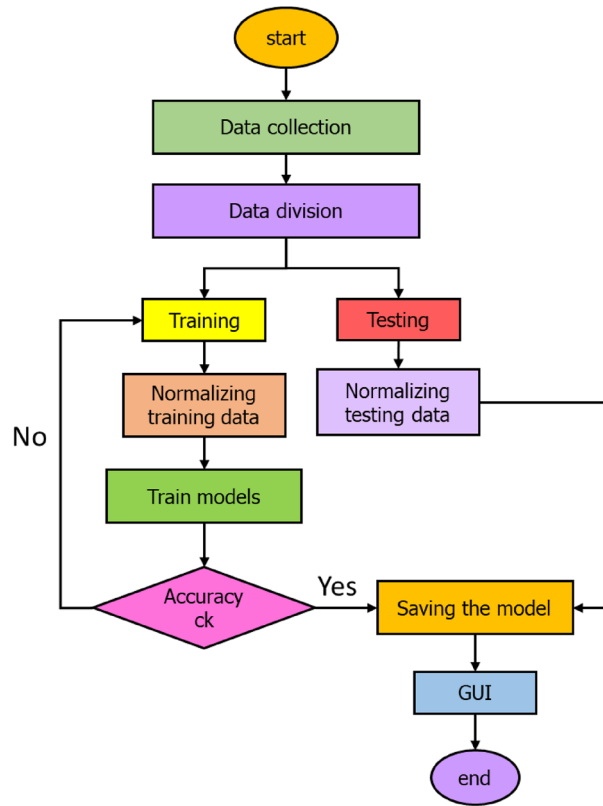


Figure 2. The primary process of model development.

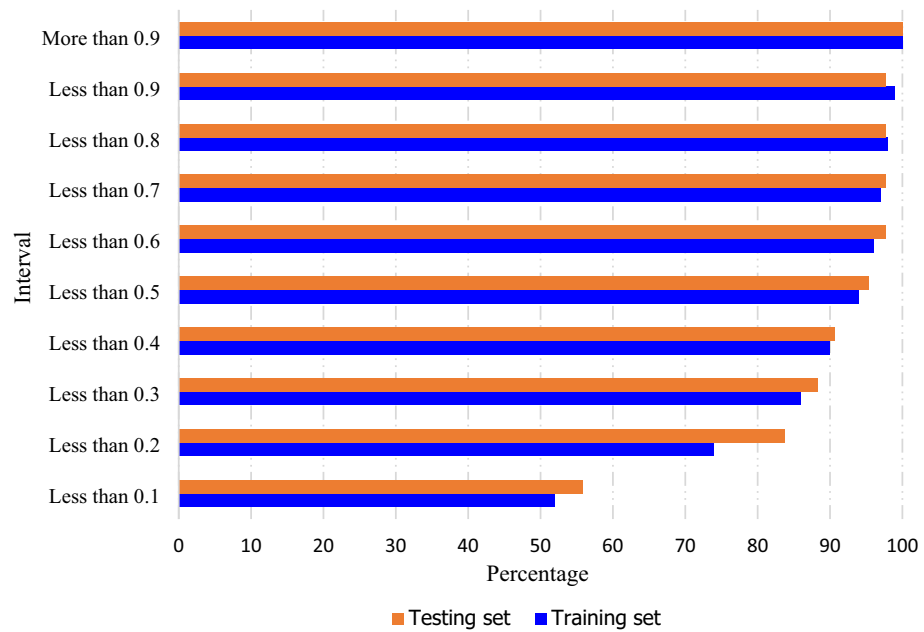


Figure 3. Distribution of hazard and non-hazard events across liquefaction resistance intervals in training and testing sets.

in the More than 0.9 (>0.9) interval, both the training and testing sets encompassed 100% of the total values, providing a complete representation of extreme liquefaction resistance values in both sets. These comparisons highlight the efforts to ensure a fair distribution of hazard and non-hazard events across different intervals as a percentage of the total values in both the training and testing sets. Consequently, the applied approach effectively mitigates zero sampling bias and enhances the reliability of the model's predictions.

The second scenario involves non-linear normalization, which includes applying a logarithm base 10 to the data before using the aforementioned equation. Subsequently, the data is linearly scaled between zero and one using the same formula. Then, a de-normalizing procedure is proposed to return data to its normal scales. Then, a de-normalization procedure is employed to restore the data to its original scales⁹². Lastly, (GUI) has been developed, specifically tailored to the best prediction model, to aid engineers and researchers in effectively utilizing the best predictive model of this study.

Statistical metrics

Seven statistical performance indices have been applied in this study to assess the accuracy of the prediction models. These indices serve as metrics to quantify the prediction errors and assess the match between predicted values and measured values⁹³. The metrics used include (*RMSE*), mean absolute error (*MAE*), maximum absolute percentage relative error (*erMax*), mean absolute percentage error (*MAPE*), and uncertainty at 95% (*U₉₅*). Additionally, other metrics such as the correlation coefficient (*R*), Willmot index (*WI*), and Nash coefficient (*NSE*) have been utilized to gauge the accuracy of the predictions and determine how well they align with the measured values. The mathematical expressions of the applied metrics are provided below (Eqs. 10–17)^{94–96}:

$$MAPE\% = \frac{100}{n} \sum_{i=1}^n \left| \frac{X_{obs_i} - X_{pred_i}}{X_{obs_i}} \right| \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_{obs_i} - X_{pred_i})^2} \quad (11)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |X_{obs_i} - X_{pred_i}| \quad (12)$$

$$R = \frac{\sum_{i=1}^n [(X_{obs_i} - \bar{X}_{obs})(X_{pred_i} - \bar{X}_{pred})]}{\sqrt{\sum_{i=1}^n (X_{obs_i} - \bar{X}_{obs})^2 \sum_{i=1}^n (X_{pred_i} - \bar{X}_{pred})^2}} \quad (13)$$

$$WI = 1 - \frac{\sum_{i=1}^n (X_{obs_i} - X_{pred_i})^2}{\sum_{i=1}^n (|X_{pred_i} - \bar{X}_{obs}| + |X_{obs_i} - \bar{X}_{obs}|)^2} \quad (14)$$

$$NSE = 1 - \frac{\sum_{i=1}^n |X_{obs_i} - X_{pred_i}|}{\sum_{i=1}^n |X_{obs_i} - \bar{X}_{obs}|} \quad (15)$$

$$erMax = \max \left(\left| \frac{X_{obs_i} - X_{pred_i}}{X_{obs_i}} \right| \right) \quad (16)$$

$$U_{95} = 1.96(RMSE^2 + SD^2)^{0.5} \quad (17)$$

where, the n , $X_{obs_i} - X_{pred_i}$, are representing the total number of samples, measured value of i th sample, and predicted value, respectively. Also, the terms \bar{X}_{obs} , and \bar{X}_{pred} are the mean values of observed and predicted values. Finally, the *SD* is the standard deviation of the forecasted errors.

Modeling result

Prediction results: both scenarios

The estimation of liquefaction resistance, a key parameter indicating soil's capacity to withstand temporary liquid-like behavior during cyclic loading, was performed using four distinct machine learning techniques: ELM, ELM-DOA, ANFIS-Sub, and ANFIS-FCM. In order to maximize the effectiveness of our modeling process, the performances of our models were rigorously evaluated across distinct phases, encompassing both training and testing stages, under two varied scenarios of data preprocessing (Tables 2 and 3). Furthermore, it is important to note that the hyperparameters for all models are provided in Table S-3 in the Supplementary File. Notably, in the first scenario, linear data normalization was employed, whereas in the second scenario, non-linear data normalization was utilized. The utilization of metrics such as *MAE*, *RMSE*, *MAPE*, *R*, *erMax*, *NSE*, and *WI* facilitate a thorough examination of the model's effectiveness. It is noteworthy that higher *R*, *NSE*, and *WI* values signify a stronger correlation between the predicted and actual values, indicating the model's robust performance. Additionally, lower *RMSE*, *MAE*, *MAPE*, and *erMax* values further reinforce the accuracy of the model, as they indicate minimal discrepancies between predicted and observed values. The results obtained from both scenarios reveal that the ANFIS-based models suffer from overfitting. This is evident in the lower prediction accuracy during the testing phase and higher accuracy during the training phase, as compared to other models. Also, the prediction accuracy of the models significantly improved when using nonlinear normalization (scenario 2)

Data mode	Models	Statistical parameters						
		MAE (J/m ³)	RMSE (J/m ³)	MAPE%	R	erMax	NSE	WI
Training	ELM-DOA	623.713	952.048	30.75	0.896	1.291	0.582	0.943
	ELM	800.894	1182.843	42.00	0.834	2.202	0.463	0.902
	ANFIS-Sub	9.668	55.580	0.410	1.000	0.129	0.994	1.000
	ANFIS-FCM	464.033	737.395	22.76	0.939	0.884	0.689	0.968
Testing	ELM-DOA	505.821	680.956	39.950	0.969	2.251	0.665	0.982
	Classic ELM	730.990	969.364	56.770	0.929	2.489	0.516	0.957
	ANFIS-Sub	757.163	1371.102	43.920	0.866	1.964	0.499	0.886
	ANFIS-FCM	574.232	879.408	39.880	0.937	2.289	0.620	0.966

Table 2. The performance of the proposed models during the training phase: Classic normalization (scenario-1).

Data mode	Models	Statistical parameters						
		MAE (J/m ³)	RMSE (J/m ³)	MAPE%	R	erMax	NSE	WI
Training	ELM-DOA	468.965	834.157	17.600	0.923	0.961	0.686	0.956
	ELM	475.853	846.841	17.600	0.920	1.015	0.681	0.954
	ANFIS-Sub	158.861	316.836	5.700	0.989	0.681	0.894	0.994
	ANFIS-FCM	353.617	687.021	14.200	0.949	0.591	0.763	0.971
Testing	ELM-DOA	335.162	484.286	24.900	0.983	1.021	0.778	0.991
	Classic ELM	404.416	556.129	28.200	0.979	1.100	0.732	0.988
	ANFIS-Sub	679.824	1014.447	51.500	0.916	3.079	0.550	0.954
	ANFIS-FCM	512.849	857.677	27.700	0.958	1.072	0.660	0.963

Table 3. The performance of the proposed models during the training phase: Log normalization (scenario-2).

compared to classical linear normalization, as stated in scenario one. The findings presented in Table 2 reveal that all the ML models showed good predictive performance, particularly the ELM-DOA model, which exhibited slightly better results compared to all other models under scenario-1 during the training and testing stages with $MAE = 623.713 \text{ J/m}^3$ and 505.821 , $RMSE = 952.048 \text{ J/m}^3$ and 680.956 J/m^3 , $MAPE = 30.750$ and 39.950 , $R = 0.896$ and 0.969 , $erMax = 1.291$ and 2.251 , $NSE = 0.582$ and 0.665 and $WI = 0.943$ and 0.982 , respectively. For further assessment, additional analyses were conducted using different proportions of the training set and testing set, including a 50:50 split and a 60:30 split, in addition to the standard 70:30 ratio. The quantitative results are summarized in Fig. 4. The findings consistently indicate that the 70:30 training/testing ratio yielded the best outcomes, leading to improved performance across all models. This ratio demonstrated good accuracy in predicting liquefaction resistance. In contrast, the other ratios resulted in less accurate predictions. These findings suggest that a higher proportion of training data (70%) in comparison to testing data (30%) is preferable for achieving optimal performance in the prediction of liquefaction resistance.

In scenario-2, the ELM-DOA model demonstrated superior performance with higher NSE (0.686 and 0.778), R (0.923 and 0.983), and WI (0.956 and 0.991), and lower MAE (468.965 J/m^3 and 335.162 J/m^3), $RMSE$ (834.157 J/m^3 and 484.286 J/m^3), $MAPE$ (0.176 and 0.249) and $erMax$ (0.961 and 1.021) for both training and testing phases (Table 3). The reported results highlight the remarkable accuracy achieved by the proposed models in both the training and testing phases. This serves as strong evidence supporting the validity of the optimization process employed. During the learning phase, the models effectively minimize the objective function ($RMSE$), indicating their ability to capture the underlying patterns and relationships within the data. Consequently, during the testing phase, the models showcase a high level of accuracy in their predictions, further affirming their reliability and effectiveness.

Figure 5 displays the performance plot of all the models during the testing phase for the second scenario. It is observed from the figure that the prediction pattern of all the models exhibits commendable conformity with the actual data pattern. Notably, the ELM-DOA model's predictions display superior agreement and align more closely with the observed patterns compared to the other models. This suggests that the ELM-DOA model performs exceptionally well in capturing and predicting the underlying patterns during the testing phase, highlighting its efficacy in this context. The graph underscores the reliability of this hybrid ELM-DOA model in accurately reflecting the observed data trends.

The scatter details of the observed and the predicted liquefaction resistance values for all the developed models during the testing phase are shown in Fig. 6. Notably, all models exhibit a good fit with the observed data. It is worth highlighting that the ELM-DOA model, characterized by the highest R^2 value (0.935), displays a particularly denser scattering pattern around the isline regression line in comparison with the other ML models. This phenomenon underscores the robust performance of the ELM-DOA model, as the tight clustering signifies a high level of accuracy and reliability in predicting liquefaction resistance values.

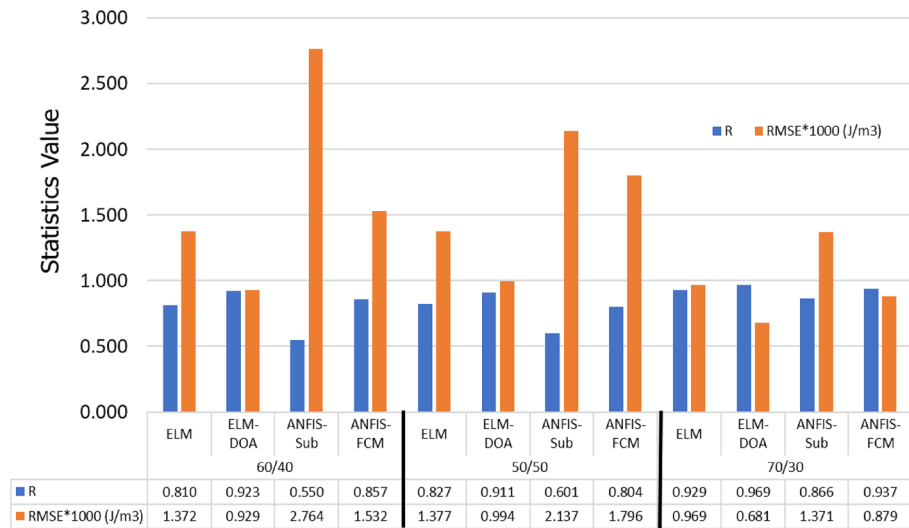


Figure 4. Comparing model performance with varying training/testing ratios for liquefaction resistance prediction.

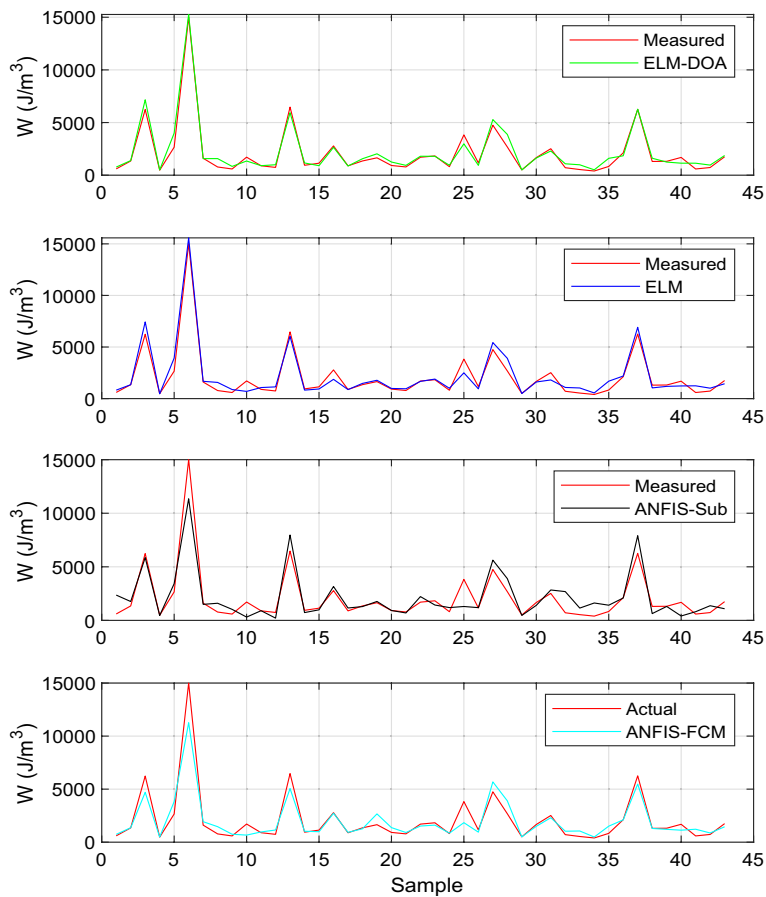


Figure 5. Variation of Predicted values of liquefaction resistance using ELM-DOA, ELM, ANFIS-sub, and ANFIS-FCM in comparison to actual liquefaction resistance values.

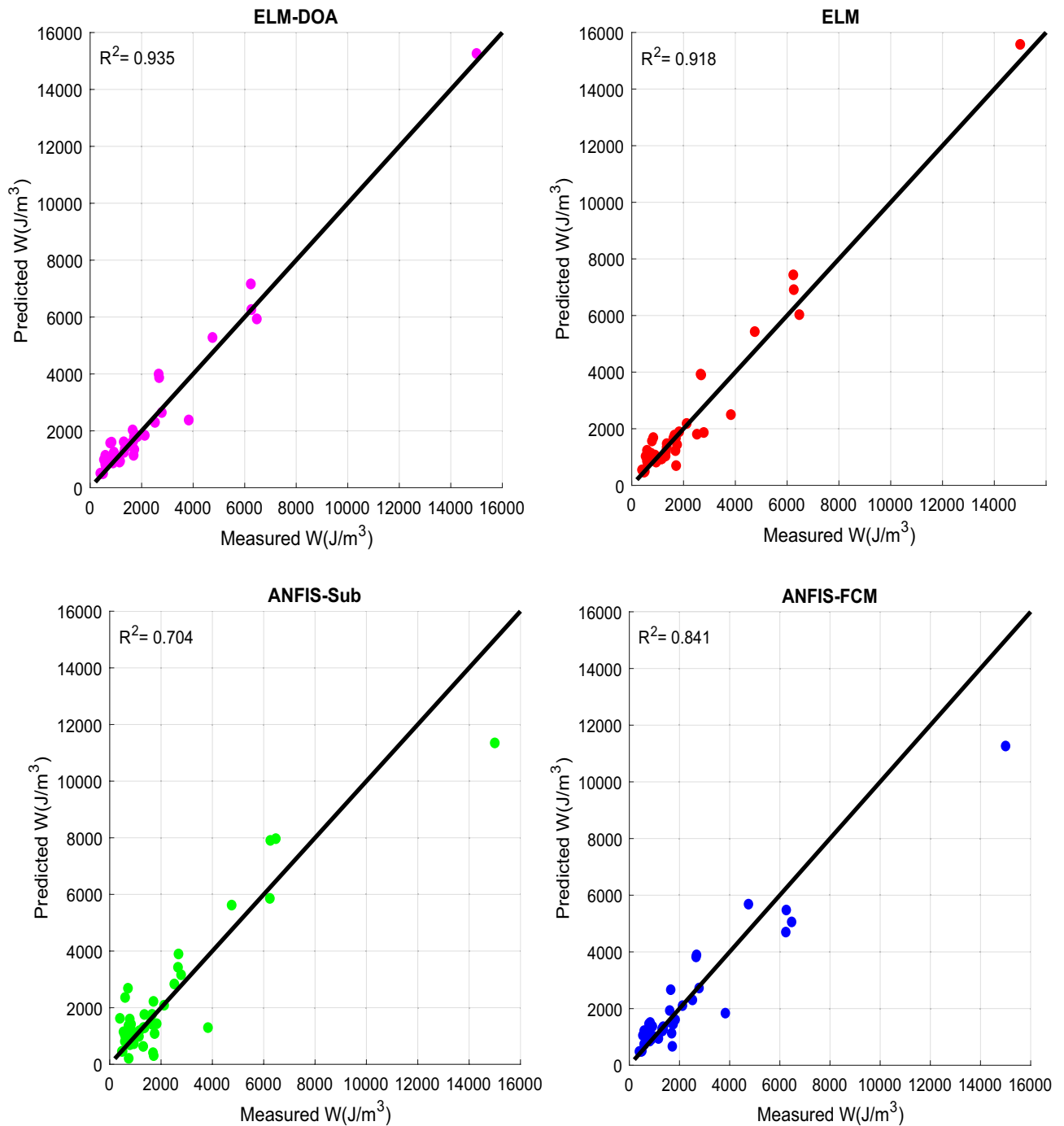


Figure 6. Scatter plots of observed and predicted liquefaction resistance values for all the considered models for testing dataset.

The Violin cum box plot is a unique graphical representation that seamlessly merges the violin plot and box plot, featuring the advantages of both. This hybrid plot displays multiple layers and includes a mean marker, akin to the traditional violin and box plots (see Fig. 7a)⁹². Moreover, Fig. 7b displays the violin and box plots for various ML models (ELM-DOA, ELM, ANFIS-sub and ANFIS-FCM) used to estimate liquefaction resistance values for the testing datasets. It was observed that the violin representing the observed values closely mirrors that of the ELM-DOA-based violin in comparison to other models for the testing data. The ELM-DOA box exhibits trends identical to those of the observed value's box, characterized by an equivalent mean value (represented by an asterisk in Fig. 7). Furthermore, the outer layer of the violin cum box plot for both the actual and ELM-DOA model demonstrates a notable degree of symmetry when compared to other machine learning models in estimating liquefaction resistance values. It is important to note that, the ELM-DOA model demonstrates more efficient prediction of extreme values compared to other models. It can be concluded that the ELM-DOA model performs the best, followed by the ELM model, while the other two ANFIS models exhibit poor performance.

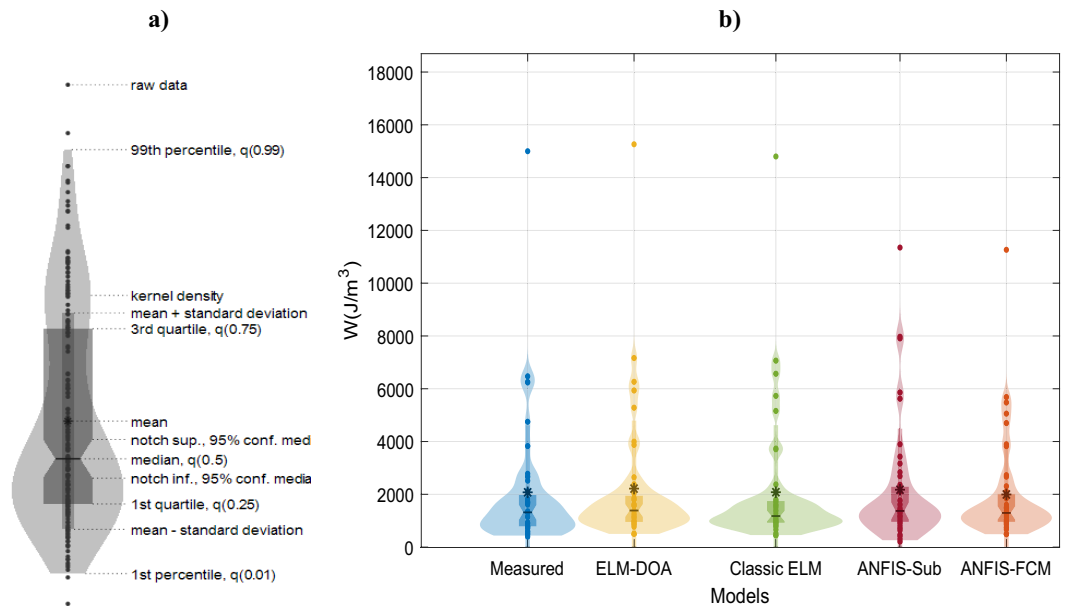


Figure 7. Violin cum box plot for applied ML techniques. (a) a visual representation highlighting the essential characteristics of the violin plot⁹⁷, and (b) a comparison between the observed values and the corresponding predictions generated by ML models.

The uncertainty analysis is performed to assess the level of confidence in a model’s predictions. This analysis helps in assessing the reliability and robustness of the proposed models by quantifying the potential range of true values for the predicted outcomes. The U_{95} interval represents the range where it is expected to find the true value for approximately 95% of similar experiments. Figure 8 demonstrates the results of the uncertainty analysis attained for the applied models i.e., ELM, ELM-DOA, ANFIS-sub and ANFIS-FCM. The results presented in the bar chart (Fig. 8) reveal that ELM-DOA demonstrates the lowest uncertainty, boasting a normalized U_{95} value of 0.0878, outperforming ELM ($U_{95} = 0.099$), ANFIS-sub ($U_{95} = 0.262$) and ANFIS-FCM ($U_{95} = 0.154$). These findings underscore the superior stability and consistency of ELM-DOA, indicating its reduced sensitivity compared to other machine learning models concerning input data variations.

The Taylor diagram visually compares model performances using three key statistical metrics—correlation coefficient (R), standard deviation (SD), and root-mean-square error ($RMSE$)—in a 2-D plot, assessing their simultaneous variations and highlighting distance-based distinctions between the observed and predicted data. The depicted liquefaction resistance data is represented as a point on the x-axis with $R = 1$ and $SD = 2500$ (Fig. 6). Based on Fig. 9, it is evident that all four models demonstrated strong prediction performance during the testing stage, exhibiting a correlation coefficient exceeding 0.90 and SD under 2700 (with all model predictions falling

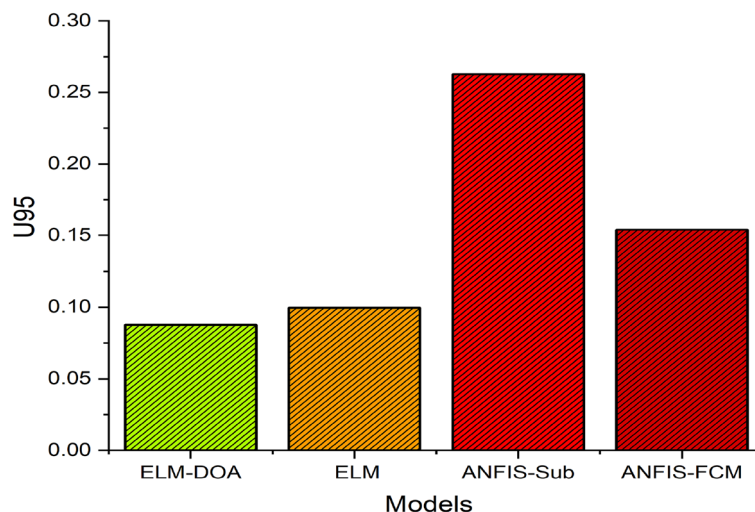


Figure 8. U_{95} values for the applied ML models.

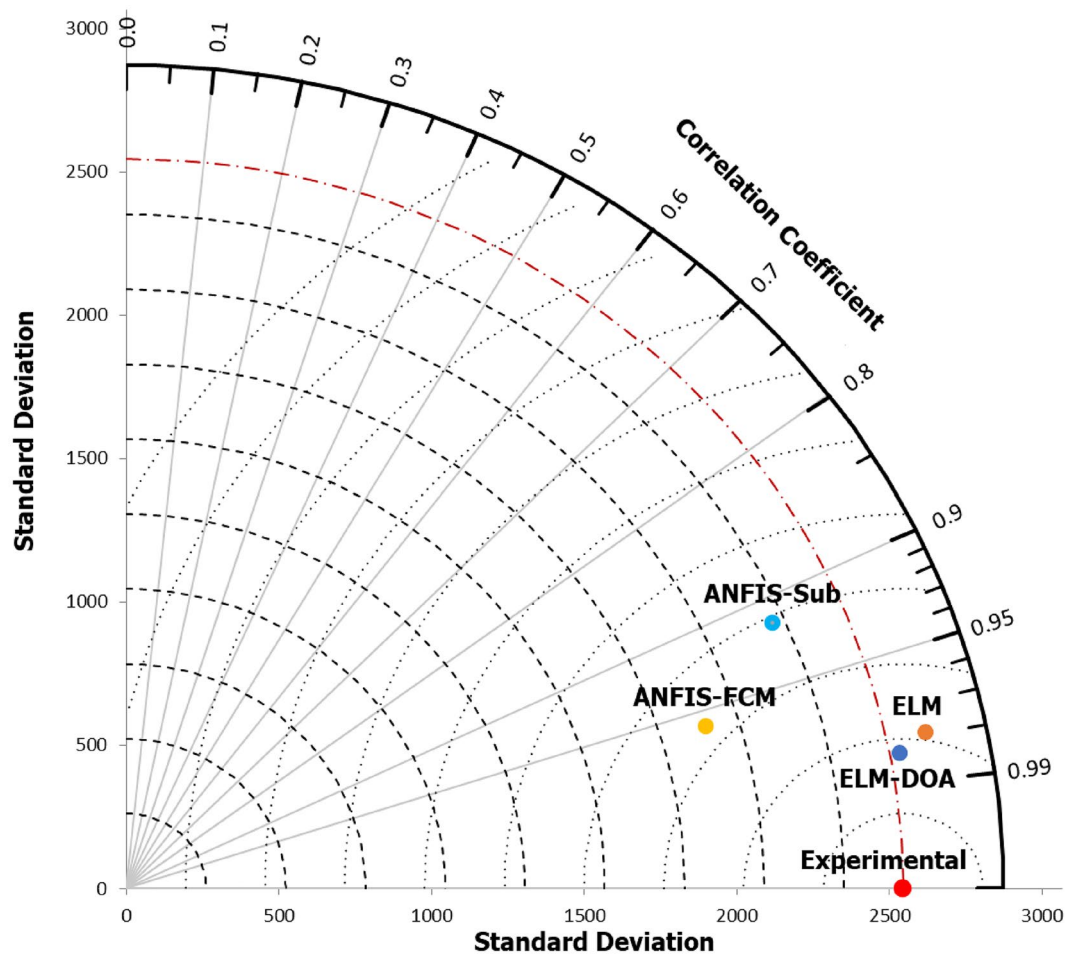


Figure 9. Comparative analysis of prediction models and measured liquefaction resistance using a Taylor Diagram.

within the SD curve). Notably, the ELM-DOA model (represented by the dark blue point), positioned closest to the observed data point, exhibited exceptional accuracy and outperformed other ML models in predicting liquefaction resistance values. Overall, the developed models exhibited exceptional performance throughout their training and testing phases, consistently surpassing benchmarks across various statistical metrics and graphical analyses. Their ability to accurately predict outcomes aligned closely with actual values, underscoring their robust generalization capacity. Notably, these models showcased a remarkable consistency in their predictions, illustrating their reliability across diverse datasets.

Validation of ELM-DOA with previous works

Hybrid ML models have been widely employed to estimate the probability of liquefaction on a regional or global scale. However, this study breaks new ground by employing a novel predictive ensemble model to enhance liquefaction potential in sand-silt mixtures. The proposed ELM-DOA method yielded enhanced liquefaction resistance estimates compared to other methods in the literature. Its results demonstrated statistical performance on par with or superior to alternative approaches, highlighting its effectiveness in accurately predicting liquefaction resistance values. For example, some researchers developed numerous hybrid ML models for the prediction of liquefaction probability in soils⁹⁸. The performance, accuracy, and reliability of the models were evaluated using different assessment and evaluation tools. The hybrid model with an $R^2=0.713$ presented the highest accuracy in terms of prediction of the liquefaction probability in soils. Moreover, other investigation reported the development of two hybrid models based on Support Vector Machines, Radial Basis Function and Grey wolf optimization algorithm³². The models were developed on the inputs of earthquake magnitude, water table, total vertical stress and other important parameters. The results of the study suggested that the hybrid models with higher R^2 values (0.757 and 0.692) outperformed the other models in terms of prediction accuracy. Furthermore, other study applied advanced hybrid ML models for forecasting soil liquefaction potential for railway embankments built on fine soil deposits⁹⁹. The moisture content, wet density, dry density, liquid limit, plastic limit, and plasticity index were incorporated as inputs for the model development. Later it was concluded that the developed hybrid model (ANFIS-FF) showed good prediction accuracy in terms of R^2 (0.900) for liquefaction prediction. Overall, the proposed hybrid model (ELM-DOA) in this study demonstrates higher

accuracy in predicting liquefaction resistance, with an R^2 value exceeding 0.93. This indicates the superiority of the models developed in previous work.

Graphical User Interface (GUI)

As mentioned in the preceding section, the suggested ELM-DOA model exhibited significant promise in estimating the strain energy-based liquefaction resistance. Given the nature of hybrid machine learning models, the integration into daily engineering practice necessitates a bridge between complex algorithms and user-friendly applications. To fulfill this requirement, we developed a MATLAB-based graphical user interface (GUI) for the ELM-DOA model, which is demonstrated in Fig. 10. This GUI is not just a testament to the model's expansive practical application but also a crucial tool to facilitate its use among engineers who may not have specialized expertise in hybrid ML models. Figure 7 shows a graphical illustration that delineates the extensive scope of applicability for the proposed ELM-DOA model. Through the effective utilization of the GUI software, we successfully conducted precise predictions for liquefaction resistance ($W = 2646.13$) at $\sigma = 82.74$, $Dr\% = 72$, $Fc\% = 5$, $Cu = 1.88$ and $D50 = 0.148$. The designed feature substantially augments the practical and academic value of the research, as it provides a standardized platform for performance assessment and comparison, streamlining future advances in the field.

Discussion

The study revealed that integrating DOA and ELM in a hybrid model yielded markedly superior performance in predicting liquefaction resistance compared to individual ML models such as ELM. The ELM-DOA model serves to augment the predictive precision of the conventional ELM by meticulously ascertaining the most advantageous weights and bias values. Through comprehensive evaluation based on statistical criteria and graphical assessments, the hybrid model (ELM-DOA) not only surpasses the performance of the traditional ELM but also outshines other widely recognized hybrid models, including ANFIS-sub, and ANFIS-FCM. This aligns with similar findings in scientific literature, highlighting the efficacy of hybrid models in enhancing predictive accuracy for simulating liquefaction resistance values¹⁰⁰. The findings presented in the study contribute to the growing body of evidence supporting the advantages of hybrid models in addressing complex geotechnical engineering challenges such as stress-strain modeling of soils, analysis of piles bearing capacity, assessment of

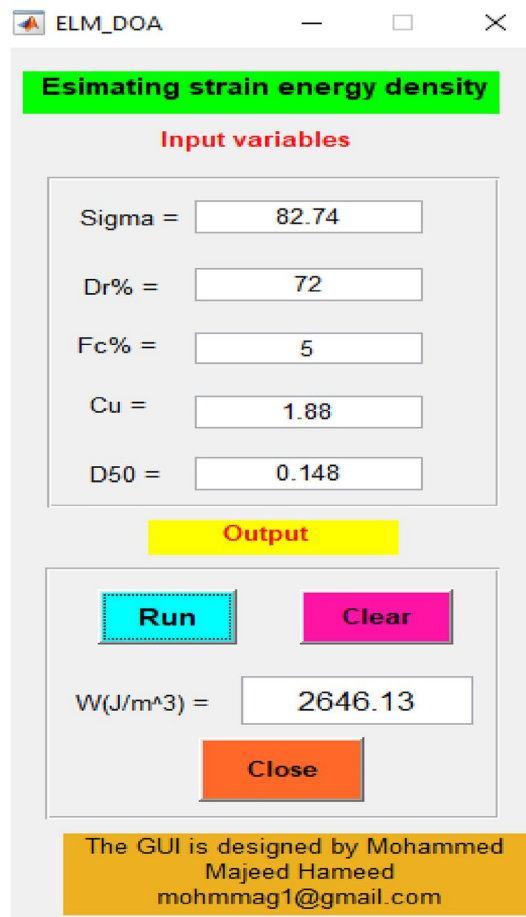


Figure 10. GUI of the ELM-DOA model based on liquefaction resistance value prediction.

settlement in shallow foundations, investigation into earthquake-induced liquefaction, and scrutiny of seismic lateral spreading phenomena^{33,101}.

The Dingo Optimization Algorithm (DOA), a novel bio-inspired approach, addresses optimization problems by emulating the social behavior of Australian dingoes. Drawing inspiration from dingo hunting techniques involving persecution, grouping, and scavenging, DOA enhances overall efficiency through three search strategies and four formulated rules. This combination achieves a delicate equilibrium between exploitation and exploration across the search space. The proposed model outperforms other individual models developed in the present study for predicting liquefaction resistance values in a sand-silt mixture, effectively. Its unique blend of dingo-inspired strategies positions it as a promising solution for such prediction challenges, showcasing its superiority in comparison to other comparable models.

Accurate prediction of soil liquefaction induced by seismic loads is crucial for geotechnical engineers due to its potential for uncontrolled lateral spreading, posing significant threats to regional civil engineering projects. Incidents of liquefaction can compromise the stability of structures, heightening the risk to infrastructure in the area. While conventional in situ tests like the Standard Penetration Test and Cone Penetration Test have been conducted, their susceptibility to uncertainties has prompted a shift towards more accurate methodologies. Consequently, geotechnical engineers are increasingly turning to advanced techniques, including machine learning (ML). ML's precision in analyzing complex data enables a more nuanced evaluation of soil behavior during seismic events, offering a refined understanding of liquefaction risks. This transition to ML applications underscores a progressive approach, empowering engineers to enhance assessments, optimize designs, and implement targeted mitigation strategies in seismic-prone regions. The hybrid ML-model (ELM-DOA) developed in our study has shown excellent performance, as evidenced by higher values *NSE* (0.778), *R* (0.983) and *WI* (0.991), and lower values of *MAE* (335.162), *RMSE* (484.286), *MAPE* (0.249) and *erMax* (1.021).

The results indicate that only the ELM-DOA and ELM models demonstrate good generalization capacity, whereas the ANFIS models encounter difficulties in producing accurate estimates during the testing phase. Additionally, the study findings reveal that the generalization capacity of all models is significantly enhanced when non-linear normalization is applied. This is attributed to the fact that non-linear normalization reduces the data variance, thereby facilitating effective model training and generating more accurate estimates. To illustrate this point, considering the ELM-DOA model as the best one, the *MAE* during the testing phase using linear normalization was 505.821 J/m³. However, after applying nonlinear normalization, the *MAE* improved to 335.162 J/m³, indicating an increase in prediction accuracy of approximately 33.74%. The other ML models achieved varied improvements, with the ELM model experiencing a 44.68% improvement as the *MAE* reduced from 730.990 to 404.416 J/m³, while the ANFIS-Sub model showed a 10.21% improvement, and the ANFIS-FCM model demonstrated a 10.69% improvement. Overall, applying nonlinear normalization improved the overall accuracy of all models by approximately 25%. This improvement was observed across all the applied models, indicating the effectiveness of nonlinear normalization in enhancing the prediction performance of these models. In addition to evaluating our models against internal benchmarks, a comparative analysis with respect to previous iterations developed in earlier studies was conducted. This comprehensive examination offered valuable insights into the evolution and refinement of the modeling techniques, ultimately affirming the success of the applied optimization process.

The results highlight that the ELM-DOA model represents a robust, accurate, and intelligent methodology for predicting soil liquefaction. The model showcases its effectiveness not only in geotechnical engineering but also holds the potential to serve as a valuable tool in various other engineering disciplines. The integration of advanced optimization techniques, such as Dingo Optimizer, with ELM, enhances the predictive capabilities, making it reliable in capturing the intricate patterns associated with soil liquefaction dynamics.

Conclusion

In this study, the predictive capabilities of the suggested hybrid model (ELM-DOA) were evaluated for predicting strain energy requirements necessary to initiate liquefaction in sand-silt mixtures. To comprehensively evaluate its effectiveness, the performance of ELM-DOA was compared with the standalone ELM model and adaptive neuro-fuzzy inference with fuzzy c-means (ANFIS-FCM), and sub-clustering (ANFIS-Sub) models, respectively. Also, two data preprocessing scenarios are utilized in this work. The first scenario is traditional linear normalization, and the second scenario is non-linear normalization. The findings indicate that non-linear normalization significantly improves the prediction performance of all models by approximately 25% compared to linear normalization. The results demonstrated that the ELM-DOA model outperformed the other models in accurately predicting the liquefaction resistance values. Moreover, the ELM-DOA model, created by integrating the novel metaheuristic algorithm (DO) with ELM, exhibited superior performance compared to standalone ELM, and other models. It achieved a remarkable minimum forecasting error of *RMSE* = 505.821 and 335.162 during both scenarios 1 and 2, respectively. The ELM-DOA model, developed using a novel metaheuristic algorithm (DO) and ELM, performed significantly better than the standalone ELM model and other hybrid models. The ELM-DOA model depicting minimum forecasting errors of *RMSE* = 505.821 and 335.162, was 30% and 17% more accurate than the standalone ELM model during scenarios 1 and 2, respectively. Similarly, the ELM-DOA model showcased improvements of 33% and 50.60% in forecasting accuracies during scenarios 1 and 2 compared to the ANFIS-Sub model. Finally, in the case of ANFIS-FCM, the ELM-DOA presented an improved accuracy of 11.91% and 34.64% during both scenarios. Overall, this study suggests that the ELM-DOA hybrid model is a promising approach for accurately predicting liquefaction values, which is crucial for making informed decision-making in geotechnical engineering. Also, it may facilitate the development of resilient and sustainable solutions in the face of potential challenges posed by seismic activities and other geological factors.

Limitations and recommendations

Soil liquefaction prediction research encounters several limitations. Firstly, the reliance on limited data for training and testing models may not fully capture the complexities and variability of real-world scenarios. However, acquiring high-quality, comprehensive real-life datasets is challenging, as it requires extensive fieldwork, laboratory testing, and data collection. This limitation restricts the availability of diverse and representative datasets for model development and validation. Lastly, complex hybrid models, such as ELM-DOA, face challenges in terms of interpretability, making it difficult to understand the underlying mechanisms driving their predictions. Thus, Enhancing the model's interpretability is crucial for establishing trust and confidence among stakeholders, including geotechnical engineers, researchers, and decision-makers. consequently, the recommendations of this study are:

1. Foster collaborative data sharing initiatives within the geotechnical engineering community to facilitate the creation of comprehensive and diverse datasets for model development and validation.
2. Explore interpretable ML techniques or model-agnostic interpretability methods to shed light on the decision-making process of complex hybrid models.
3. Conduct further studies to evaluate the robustness and generalization capabilities of ML models across different geological and environmental conditions.
4. Continuously refine and improve optimization algorithms, such as the Dingo Optimization Algorithm, to enhance the performance and efficiency of hybrid ML models for soil liquefaction prediction.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request. Additionally, all the experimental samples utilized in this study can be found in the Supplementary File, namely Tables S-2 and S-3.

Received: 1 February 2024; Accepted: 30 April 2024

Published online: 11 May 2024

References

1. Ishihara, K. & Koga, Y. Case studies of liquefaction in the 1964 Niigata Earthquake. *Soils Found.* **21**, 35–52 (1981).
2. Youd, T. L. Ground failure investigations following the 1964 Alaska Earthquake. in *Proceedings of the 10th National Conference in Earthquake Engineering, Earthquake Engineering Research Institute, Anchorage, AK* (2014).
3. Toda, S., Hataya, R., Abe, S. & Miyakoshi, K. The 1995 Kobe earthquake and problems of evaluation of active faults in Japan. *Eng. Geol.* **43**, 151–167 (1996).
4. Giona Bucci, M. *et al.* Associations between sediment architecture and liquefaction susceptibility in fluvial settings: The 2010–2011 Canterbury Earthquake Sequence, New Zealand. *Eng. Geol.* **237**, 181–197 (2018).
5. Sassa, S. & Takagawa, T. Liquefied gravity flow-induced tsunami: First evidence and comparison from the 2018 Indonesia Sulawesi earthquake and tsunami disasters. *Landslides* **16**, 195–200 (2019).
6. Su, D., Ming, H. Y. & Li, X. S. Effect of shaking strength on the seismic response of liquefiable level ground. *Eng. Geol.* **166**, 262–271 (2013).
7. Wang, Y., Cao, T., Gao, Y. & Shao, J. Experimental study on liquefaction characteristics of saturated Yellow River silt under cycles loading. *Soil Dynam. Earthq. Eng.* **163**, 107457 (2022).
8. Geyin, M., Maurer, B. W. & van Ballegooy, S. Lifecycle Liquefaction Hazard Assessment and Mitigation. in *Geo-Congress 2020* 312–320 (American Society of Civil Engineers Reston, VA, 2020).
9. Kim, S. & Park, K. Proposal of liquefaction potential assessment procedure using real earthquake loading. *KSCE J. Civ. Eng.* **12**, 15–24 (2008).
10. ElGhoraiby, M. A., Park, H. & Manzari, M. T. Stress-strain behavior and liquefaction strength characteristics of Ottawa F65 sand. *Soil Dynam. Earthq. Eng.* **138**, 106292 (2020).
11. Silver, M. L. & Park, T. K. Liquefaction potential evaluated from cyclic strain-controlled properties tests on sands. *Soils Foundations* **16**, 51–65 (1976).
12. Kokusho, T. Energy-based liquefaction evaluation for induced strain and surface settlement—Evaluation steps and case studies. *Soil Dynam. Earthq. Eng.* **143**, 106552 (2021).
13. Chen, Y.-R., Chen, J.-W., Hsieh, S.-C. & Chang, Y.-T. Evaluation of soil liquefaction potential based on the nonlinear energy dissipation principles. *J. Earthq. Eng.* **17**, 54–72 (2013).
14. Jain, A., Mittal, S. & Shukla, S. K. Energy-based approach to study liquefaction triggering in homogeneous and stratified soils under consolidated undrained cyclic loading. *Eng. Geol.* **321**, 107151 (2023).
15. Tokimatsu, K. & Yoshimi, Y. Empirical correlation of soil liquefaction based on SPT N-value and fines content. *Soils Foundations* **23**, 56–74 (1983).
16. Cetin, K. O. *et al.* The use of the SPT-based seismic soil liquefaction triggering evaluation methodology in engineering hazard assessments. *MethodsX* **5**, 1556–1575 (2018).
17. Daag, A. S., Halasan, O. P. C., Magnaye, A. A. T., Grutas, R. N. & Solidum, R. U. Empirical correlation between standard penetration resistance (SPT-N) and shear wave velocity (Vs) for soils in Metro Manila, Philippines. *Appl. Sci.* <https://doi.org/10.3390/app12168067> (2022).
18. Karamitros, D. K., Bouckovalas, G. D., Chaloulos, Y. K. & Andrianopoulos, K. I. Numerical analysis of liquefaction-induced bearing capacity degradation of shallow foundations on a two-layered soil profile. *Soil Dynam. Earthq. Eng.* **44**, 90–101 (2013).
19. Kusakabe, R., Ichimura, T., Fujita, K., Hori, M. & Wijerathne, L. A finite element analysis method for simulating seismic soil liquefaction based on a large-scale 3D soil structure model. *Soil Dynam. Earthq. Eng.* **123**, 64–74 (2019).
20. Hameed, M. M., AlOmar, M. K., Al-Saadi, A. A. & AlSaadi, M. A. Inflow forecasting using regularized extreme learning machine: Haditha reservoir chosen as case study. *Stoch. Environ. Res. Risk Assess.* **36**, 4201–4221. <https://doi.org/10.1007/s00477-022-02254-7> (2022).
21. Alomar, M. K. *et al.* Data-driven models for atmospheric air temperature forecasting at a continental climate region. *PLoS One* **17**, e0277079 (2022).

22. Shi, M.-L., Lv, L. & Xu, L. A multi-fidelity surrogate model based on extreme support vector regression: Fusing different fidelity data for engineering design. *Eng. Comput. (Swansea)* **40**, 473–493 (2023).
23. Long, X., Mao, M., Su, T., Su, Y. & Tian, M. Machine learning method to predict dynamic compressive response of concrete-like material at high strain rates. *Defence Technol.* **23**, 100–111 (2023).
24. Rai, P., Pei, H., Meng, F. & Ahmad, M. Utilization of marble powder and magnesium phosphate cement for improving the engineering characteristics of soil. *Int. J. Geosynth. Ground Eng.* **6**, 31 (2020).
25. Ahmad, M., Tang, X.-W., Qiu, J.-N. & Ahmad, F. Evaluating seismic soil liquefaction potential using Bayesian belief network and C45 decision tree approaches. *Appl. Sci.* <https://doi.org/10.3390/app9204226> (2019).
26. Samui, P., Kim, D. & Sitharam, T. G. Support vector machine for evaluating seismic-liquefaction potential using shear wave velocity. *J. Appl. Geophys.* **73**, 8–15 (2011).
27. Jas, K. & Dodagoudar, G. R. Explainable machine learning model for liquefaction potential assessment of soils using XGBoost-SHAP. *Soil Dynam. Earthq. Eng.* **165**, 107662 (2023).
28. Kumar, D. R., Samui, P. & Burman, A. Prediction of probability of liquefaction using soft computing techniques. *J. Inst. Eng. (India) Series A.* **103**, 1195–1208 (2022).
29. Egbueri, J. C., Igwe, O., Omeka, M. E. & Agbasi, J. C. Development of MLR and variedly optimized ANN models for forecasting the detachability and liquefaction potential index of erodible soils. *Geosyst. Geoenviron.* **2**, 100104 (2023).
30. Jangir, H. K. & Satavalekar, R. Evaluating Adaptive Neuro-Fuzzy Inference System (ANFIS) to assess liquefaction potential and settlements using CPT test data. *J. Soft Comput. Civ. Eng.* **6**, 119–139 (2022).
31. Zhang, Y., Qiu, J., Zhang, Y. & Wei, Y. The adoption of ELM to the prediction of soil liquefaction based on CPT. *Nat. Hazards* **107**, 539–549 (2021).
32. Cai, M. *et al.* Integrating the LSSVM and RBFNN models with three optimization algorithms to predict the soil liquefaction potential. *Eng. Comput.* **38**, 3611–3623 (2022).
33. Zhou, J., Huang, S., Wang, M. & Qiu, Y. Performance evaluation of hybrid GA-SVM and GWO-SVM models to predict earthquake-induced liquefaction potential of soil: a multi-dataset investigation. *Eng. Comput.* **38**, 4197–4215 (2022).
34. Zhang, J. & Wang, Y. An ensemble method to improve prediction of earthquake-induced soil liquefaction: A multi-dataset study. *Neural Comput. Appl.* **33**, 1533–1546 (2021).
35. Taleb Bahmed, I. *et al.* Prediction of geotechnical properties of clayey soils stabilised with lime using artificial neural networks (ANNs). *Int. J. Geotech. Eng.* **13**, 191–203 (2019).
36. Zhang, P., Yin, Z.-Y. & Jin, Y.-F. Machine learning-based modelling of soil properties for geotechnical design: Review, tool development and comparison. *Arch. Comput. Methods Eng.* **29**, 1229–1245 (2022).
37. Ozsagir, M., Erden, C., Bol, E., Sert, S. & Özocak, A. Machine learning approaches for prediction of fine-grained soils liquefaction. *Comput. Geotech.* **152**, 105014 (2022).
38. Liu, C. *et al.* The role of TBM asymmetric tail-grouting on surface settlement in coarse-grained soils of urban area: Field tests and FEA modelling. *Tunnel. Underground Space Technol.* **111**, 103857 (2021).
39. Taffese, W. Z. & Abegaz, K. A. Prediction of compaction and strength properties of amended soil using machine learning. *Buildings.* <https://doi.org/10.3390/buildings12050613> (2022).
40. Ghani, S., Kumari, S. & Ahmad, S. Prediction of the seismic effect on liquefaction behavior of fine-grained soils using artificial intelligence-based hybridized modeling. *Arab. J. Sci. Eng.* **47**, 5411–5441 (2022).
41. Kumar, D. R., Samui, P. & Burman, A. Prediction of probability of liquefaction using soft computing techniques. *J. Inst. Eng. India Series A.* **103**, 1195–1208 (2022).
42. Zhou, J., Huang, S., Zhou, T., Armaghani, D. J. & Qiu, Y. Employing a genetic algorithm and grey wolf optimizer for optimizing RF models to evaluate soil liquefaction potential. *Artif. Intell. Rev.* **55**, 5673–5705 (2022).
43. Ahmad, M., Tang, X. & Ahmad, F. Evaluation of liquefaction-induced settlement using random forest and REP tree models: taking pohang earthquake as a case of illustration. in *Natural Hazards-Impacts, Adjustments and Resilience* (IntechOpen, 2020).
44. Bairwa, A. K., Joshi, S. & Singh, D. Dingo optimizer: A nature-inspired metaheuristic approach for engineering problems. *Math. Probl. Eng.* **2021**, 2571863 (2021).
45. Berrill, J. B. & Davis, R. O. Energy dissipation and seismic liquefaction of sands: Revised model. *Soils Foundations* **25**, 106–118 (1985).
46. Baziar, M. H., Jafarian, Y., Shahnazari, H., Movahed, V. & Amin Tutunchian, M. Prediction of strain energy-based liquefaction resistance of sand-silt mixtures: An evolutionary approach. *Comput. Geosci.* **37**, 1883–1893 (2011).
47. Tao, M. *Case History Verification of the Energy Method to Determine the Liquefaction Potential of Soil Deposits.* (Case Western Reserve University, 2003).
48. Rokoff, M. D. The influence of grain-size characteristics in determining the liquefaction potential of a soil deposit by the energy method. (1999).
49. Kanagalingam, T. *Liquefaction Resistance of Granular Mixes Based on Contact Density and Energy Considerations.* (State University of New York at Buffalo, 2006).
50. Ahmad, M., Tang, X.-W., Qiu, J.-N. & Ahmad, F. Interpretive structural modeling and MICMAC analysis for identifying and benchmarking significant factors of seismic soil liquefaction. *Appl. Sci.* <https://doi.org/10.3390/app9020233> (2019).
51. Heddam, S. *et al.* Chapter 1—Predicting dissolved oxygen concentration in river using new advanced machines learning: Long-short term memory (LSTM) deep learning. in (ed. Pourghasemi, H. R. B. T.-C. in E. and E. S.) 1–20 (Elsevier, 2022). <https://doi.org/10.1016/B978-0-323-89861-4.00031-2>.
52. Parveen, N., Zaidi, S. & Danish, M. Development of SVR-based model and comparative analysis with MLR and ANN models for predicting the sorption capacity of Cr(VI). *Process Safety Environ. Protect.* **107**, 428–437 (2017).
53. Were, K., Bui, D. T., Dick, Ø. B. & Singh, B. R. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecol. Indic.* **52**, 394–403 (2015).
54. Kaingo, J., Tumbo, S. D., Kihupi, N. I. & Mbilinyi, B. P. Prediction of soil moisture-holding capacity with support vector machines in dry subhumid tropics. *Appl. Environ. Soil Sci.* **2018**, 9263296 (2018).
55. Tabarsa, A., Latifi, N., Osouli, A. & Bagheri, Y. Unconfined compressive strength prediction of soils stabilized using artificial neural networks and support vector machines. *Front. Struct. Civ. Eng.* **15**, 520–536 (2021).
56. Huang, G. B., Zhu, Q. Y. & Siew, C. K. Extreme learning machine: Theory and applications. *Neurocomputing* **70**, 489–501 (2006).
57. Masood, A., Niazkar, M., Zakwan, M. & Piraei, R. A machine learning-based framework for water quality index estimation in the Southern Bug River. *Water.* <https://doi.org/10.3390/w15203543> (2023).
58. Masood, A. *et al.* Improving PM2.5 prediction in New Delhi using a hybrid extreme learning machine coupled with snake optimization algorithm. *Sci. Rep.* **13**, 1–17 (2023).
59. Zhang, J., Li, Y., Xiao, W. & Zhang, Z. Non-iterative and fast deep learning: multilayer extreme learning machines. *J. Franklin. Inst.* **357**, 8925–8955 (2020).
60. Ding, S., Xu, X. & Nie, R. Extreme learning machine and its applications. *Neural Comput. Appl.* **25**, 549–556 (2014).
61. Wang, J., Lu, S., Wang, S.-H. & Zhang, Y.-D. A review on extreme learning machine. *Multimed. Tools Appl.* **81**, 41611–41660 (2022).

62. Kang, M., Chen, H. & Dong, J. Adaptive visual servoing with an uncalibrated camera using extreme learning machine and Q-learning. *Neurocomputing* **402**, 384–394 (2020).
63. Hameed, M. M., Mohd Razali, S. F., Wan Mohtar, W. H. M., Ahmad Alsaydalani, M. O. & Yaseen, Z. M. Deep learning versus hybrid regularized extreme learning machine for multi-month drought forecasting: A comparative study and trend analysis in tropical region. *Heliyon* **10**, e22942 (2024).
64. Hameed, M. M., Razali, S. F. M., Mohtar, W. H. M. W., Rahman, N. A. & Yaseen, Z. M. Machine learning models development for accurate multi-months ahead drought forecasting: Case study of the Great Lakes, North America. *PLoS One* **18**, e0290891 (2023).
65. Ghani, S., Kumari, S. & Bardhan, A. A novel liquefaction study for fine-grained soil using PCA-based hybrid soft computing models. *Sādhanā* **46**, 113 (2021).
66. Wang, J., Lu, S., Wang, S. H. & Zhang, Y. D. A review on extreme learning machine. *Multimed. Tools Appl.* **81**, 41611–41660 (2021).
67. Almazán-Covarrubias, J. H., Peraza-Vázquez, H., Peña-Delgado, A. F. & García-Vite, P. M. An improved Dingo optimization algorithm applied to SHE-PWM modulation strategy. *Appl. Sci.* **12**, 992 (2022).
68. Peraza-Vázquez, H. *et al.* A bio-inspired method for engineering design optimization inspired by dingoes hunting strategies. *Math. Probl. Eng.* **2021**, 9107547 (2021).
69. Ramya, K. & Ayothi, S. Hybrid dingo and whale optimization algorithm-based optimal load balancing for cloud computing environment. *Trans. Emerg. Telecommun. Technol.* **34**, e4760 (2023).
70. Nayak, S. R., Khadanga, R. K., Arya, Y., Panda, S. & Sahu, P. R. Influence of ultra-capacitor on AGC of five-area hybrid power system with multi-type generations utilizing sine cosine adopted dingo optimization algorithm. *Electr. Power Syst. Res.* **223**, 109513 (2023).
71. Cai, W. & Duan, F. Task scheduling for federated learning in edge cloud computing environments by using adaptive-greedy dingo optimization algorithm and Binary Salp Swarm Algorithm. *Future Internet*. <https://doi.org/10.3390/fi15110357> (2023).
72. Muazu, A. A., Hashim, A. S. & Sarlan, A. Review of nature inspired metaheuristic algorithm selection for combinatorial t-way testing. *IEEE Access* **10**, 27404–27431 (2022).
73. Zaghoul, M. S., Hamza, R. A., Iorhemen, O. T. & Tay, J. H. Comparison of adaptive neuro-fuzzy inference systems (ANFIS) and support vector regression (SVR) for data-driven modelling of aerobic granular sludge reactors. *J. Environ. Chem. Eng.* **8**, 103742 (2020).
74. Ghani, S. & Kumari, S. Plasticity-based liquefaction prediction using support vector machine and adaptive neuro-fuzzy inference system. *Lecture Notes Civ. Eng.* **300**, 515–527 (2023).
75. Deif, M., Hammam, R. & Solyman, A. Adaptive neuro-fuzzy inference system (ANFIS) for rapid diagnosis of COVID-19 cases based on routine blood tests. *Int. J. Intel. Eng. Syst.* **14**, 178–189 (2021).
76. Tulla, P. S. *et al.* Daily suspended sediment yield estimation using soft-computing algorithms for hilly watersheds in a data-scarce situation: a case study of Bino watershed, Uttarakhand. *Theor. Appl. Climatol.* **155**, 4023–4047. <https://doi.org/10.1007/s00704-024-04862-5> (2024).
77. Ehteram, M. *et al.* Performance improvement for infiltration rate prediction using hybridized Adaptive Neuro-Fuzzy Inferences System (ANFIS) with optimization algorithms. *Ain Shams Eng. J.* **12**, 1665–1676 (2021).
78. Babanezhad, M., Masoumian, A., Nakhjiri, A. T., Marjani, A. & Shirazian, S. Influence of number of membership functions on prediction of membrane systems using adaptive network based fuzzy inference system (ANFIS). *Sci. Rep.* **10**, 16110 (2020).
79. Kanagaraj, N. An adaptive neuro-fuzzy inference system to improve fractional order controller performance. *Intell. Autom. Soft Comput.* **35** (2023).
80. Adnan, R. M. *et al.* Enhancing accuracy of extreme learning machine in predicting river flow using improved reptile search algorithm. *Stochastic Environ. Res. Risk Assessment* **37**, 3063–3083 (2023).
81. Adeleke, O., Akinlabi, S. A., Jen, T. C. & Dunmade, I. Prediction of municipal solid waste generation: An investigation of the effect of clustering techniques and parameters on ANFIS model performance. *Environ. Technol.* **43**, 1634–1647 (2022).
82. Hussain, W., Merigó, J. M., Raza, M. R. & Gao, H. A new QoS prediction model using hybrid IOWA-ANFIS with fuzzy C-means, subtractive clustering and grid partitioning. *Inf. Sci. (N Y)* **584**, 280–300 (2022).
83. Jafari, M. M., Ojaghlo, H., Zare, M. & Schumann, G. J. P. Application of a novel hybrid wavelet-ANFIS/fuzzy C-means clustering model to predict groundwater fluctuations. *Atmosphere*. **12**, 9 (2020).
84. Chen, W., Chen, X., Peng, J., Panahi, M. & Lee, S. Landslide susceptibility modeling based on ANFIS with teaching-learning-based optimization and Satin bowerbird optimizer. *Geosci. Front.* **12**, 93–107 (2021).
85. Yilmaz, S., İlhan, R. & Feyzullahoğlu, E. Estimation of adhesive wear behavior of the glass fiber reinforced polyester composite materials using ANFIS model. *J. Elastomers Plastics* **54**, 86–110 (2022).
86. Pramod, C. P. & Pillai, G. N. K-Means clustering based Extreme Learning ANFIS with improved interpretability for regression problems. *Knowl. Based Syst.* **215**, 106750 (2021).
87. Kumar, R., Sahu, M. & Mohdiwale, S. Two class motor imagery classification based on ANFIS. *Lecture Notes Electr. Eng.* **601**, 703–711 (2020).
88. Pham, B. T., Son, L. H., Hoang, T.-A., Nguyen, D.-M. & Tien Bui, D. Prediction of shear strength of soft soil using machine learning methods. *Catena (Amst)*. **166**, 181–191 (2018).
89. Pham, B. T. *et al.* A novel artificial intelligence approach based on Multi-layer Perceptron Neural Network and Biogeography-based Optimization for predicting coefficient of consolidation of soil. *Catena (Amst)* **173**, 302–311 (2019).
90. Tunçay, T., Alaboz, P., Dengiz, O. & Başkan, O. Application of regression kriging and machine learning methods to estimate soil moisture constants in a semi-arid terrestrial area. *Comput. Electron. Agric.* **212**, 108118 (2023).
91. Iqbal, M., Onyelowe, K. C. & Jalal, F. E. Smart computing models of California bearing ratio, unconfined compressive strength, and resistance value of activated ash-modified soft clay soil with adaptive neuro-fuzzy inference system and ensemble random forest regression techniques. *Multisc. Multidiscip. Model. Exp. Design* **4**, 207–225 (2021).
92. Hameed, M. M., Mohd Razali, S. F., Wan Mohtar, W. H. M. & Yaseen, Z. M. Improving multi-month hydrological drought forecasting in a tropical region using hybridized extreme learning machine model with Beluga Whale Optimization algorithm. *Stochastic Environ. Res. Risk Assessment*. **37**, 4963–4989 (2023).
93. Masood, A. & Ahmad, K. Prediction of PM_{2.5} concentrations using soft computing techniques for the megacity Delhi, India. *Stochastic Environ. Res. Risk Assessment* **37**, 625–638 (2023).
94. Hameed, M. M., Khaleel, F., AlOmar, M. K., Mohd Razali, S. F. & Alsaadi, M. A. Optimising the selection of input variables to increase the predicting accuracy of shear strength for deep beams. *Complexity* **2022**, (2022).
95. Hameed, M. M., Abed, M. A., Al-Ansari, N. & Alomar, M. K. Predicting compressive strength of concrete containing industrial waste materials: Novel and hybrid machine learning model. *Adv. Civ. Eng.* **2022**, 5586737 (2022).
96. Mamata, R., Ramlia, A., *et al.* (2022). Slope stability prediction of road embankment using artificial neural network combined with genetic algorithm. *journalarticle.ukm.myRC Mamata, A Ramlia, MRM Yazidb, A Kasab, SFM Razalib, MN BastamcJurnal Kejuruteraan, 2022•journalarticle.ukm.my*.
97. al_goodplot—boxplot and violin plot—File Exchange - MATLAB Central. https://www.mathworks.com/matlabcentral/fileexchange/91790-al_goodplot-boxplot-violin-plot.

98. Kumar, D. R., Samui, P. & Burman, A. Prediction of probability of liquefaction using hybrid ANN with optimization techniques. *Arab. J. Geosci.* **15**, 1–21 (2022).
99. Ghani, S. & Kumari, S. Prediction of soil liquefaction for railway embankment resting on fine soil deposits using enhanced machine learning techniques. *J. Earth Syst. Sci.* **132**, 145 (2023).
100. Kumar, D. R., Samui, P. & Burman, A. Prediction of probability of liquefaction using hybrid ANN with optimization techniques. *Arab. J. Geosci.* **15**, (2022).
101. Mohammed, M., Sharafati, A., Al-Ansari, N. & Yaseen, Z. M. Shallow foundation settlement quantification: Application of hybridized adaptive neuro-fuzzy inference system model. *Adv. Civ. Eng.* **2020**, (2020).

Acknowledgements

The authors would like to thank Al-Maarif University College for supporting this research.

Author contributions

Conceptualization, Methodology: Mohammed Majeed Hameed, Adil Masood, Aman Srivastava, Norinah Abd Rahman, Siti Fatin Mohd Razalid, Ali Salem, Ahmed Elbeltagi, and Ali Salem; Formal analysis: Mohammed Majeed Hameed, Adil Masood and Aman Srivastava, Software: Mohammed Majeed Hameed and Siti Fatin Mohd Razalid; Visualization: Mohammed Majeed Hameed and Adil Masood; Writing- Original draft preparation: Mohammed Majeed Hameed; Writing—review and editing: Ali Salem, Aman Srivastava, and Ahmed Elbeltagi. All authors reviewed the manuscript.

Funding

Open access funding provided by University of Pécs.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-61059-6>.

Correspondence and requests for materials should be addressed to M.M.H. or A.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024