applied sciences

MDPI

Article

A Novel Approach to Swell Mitigation: Machine-Learning-Powered Optimal Unit Weight and Stress Prediction in Expansive Soils

Ammar Alnmr [1,*], Richard Ray [1] and Mounzer Omran Alzawi [2]

[1] Department of Structural and Geotechnical Engineering, Széchenyi István University, 9026 Gyor, Hungary; ray@sze.hu
[2] Department of Geotechnical Engineering, Tishreen University, Latakia P.O. Box 2237, Syria; momran1@scs-net.org
* Correspondence: alnmr.ammar@hallgato.sze.hu

Abstract: Expansive soils pose significant challenges to structural integrity, primarily due to volumetric changes that can lead to detrimental consequences and substantial economic losses. This study delves into the intricate dynamics of expansive soils through loaded swelling pressure experiments conducted under diverse conditions, encompassing variations in the sand content, initial dry unit weight, and initial degree of saturation. The findings underscore the pronounced influence of these factors on soil swelling. To address these challenges, a novel method leveraging machine learning prediction models is introduced, offering an efficient and cost-effective framework to mitigate potential hazards associated with expansive soils. Employing advanced algorithms such as decision tree regression (DTR), random forest regression (RFR), gradient boosting regression (GBR), extreme gradient boosting (XGBoost), support vector regression (SVR), and artificial neural networks (ANN) in the Python software 3.11 environment, this study aims to predict the optimal applied stress and dry unit weight required for soil swelling mitigation. Results reveal that XGBoost and ANN stand out for their precision and superior metrics. While both performed well, ANN demonstrated exceptional consistency across training and testing phases, making it the preferred choice. In the tested dataset, ANN achieved the highest R-squared values (0.9917 and 0.9954), lowest RMSE (7.92 and 0.086), and lowest MAE (5.872 and 0.0488) for predicting optimal applied stress and dry unit weight, respectively.

Keywords: clay; sand (additives); swelling pressure; loaded swelling pressure; partial saturation; machine learning

Citation: Alnmr, A.; Ray, R.; Alzawi, M.O. A Novel Approach to Swell Mitigation: Machine-Learning-Powered Optimal Unit Weight and Stress Prediction in Expansive Soils. *Appl. Sci.* **2024**, *14*, 1411. https://doi.org/10.3390/app14041411

Academic Editors: Edyta Plebankiewicz, Jiwei Zhang, Hui Cao and Song Zhang

Received: 9 January 2024
Revised: 27 January 2024
Accepted: 6 February 2024
Published: 8 February 2024

1. Introduction

Expansive soils present a significant challenge for geotechnical engineering, as they are widespread in nature and their significant deformations are associated with changes in suction and the degree of saturation [1–6]. In the unsaturated zone above the phreatic groundwater level, the soil moisture content varies significantly over seasons, necessitating a comprehensive understanding of expansive soil behavior and precise specifications for optimal characterization to mitigate risks [7,8]. When facilities and roads are constructed on expansive soils, they can cause significant damage and high costs. For example, in the USA, losses resulting from cracks in buildings and roads built on expansive soils were estimated at about USD 10 billion in 1985 AD, with half of that amount spent on repairing roads [9]. Similar losses have been reported in [10–12]. According to Nelson and Miller [13], the financial loss resulting from the devastating effects of expansive soils would be greater than the loss caused by earthquakes or floods. This highlights the importance of the continued deepening of research conducted on these types of soils to limit their damage. The objective of this study is to comprehensively examine the influence of the initial degree of saturation, initial dry unit weight, and the percentage of added sand on the swelling

characteristics of expansive soils. This investigation entails experimental analysis through loaded swelling pressure tests and employs diverse machine learning algorithmic models implemented within the Python programming environment for predictive purposes. This paper introduces a new method that relies on ensuring that the weight of the structure is aligned with the properties of the expansive soil, such as dry unit weight, degree of saturation, and liquid limit. This approach aims to reduce the negative effects of expansive soils on structures and offers an alternative to traditional soil improvement methods.

Sand is widely known as a granular material with a high bearing capacity that can alter the properties of cohesive soil, including plasticity, compactness, and resistance [14–18]. When mixed with cohesive soil in various ratios, sand replaces fine soil particles with coarser ones, creating a better gradient of soil grain that increases cohesion and friction and improves swelling properties. This research seeks to explore the swelling behavior of partially saturated expansive soils in response to variations in their sand content. Additionally, it endeavors to predict the optimal dry unit weight and applied stress parameters that result in minimizing swell amplitude, thereby contributing to the mitigation of associated risks. Despite previous studies that have explored the use of inert materials in expansive soils [17,19–22] and their contributions to improving expansive soils and reducing hazards, many questions and issues on the behavior of expansive soils remain unresolved. Within this framework, our objective is to assess the influence of varying proportions of sand under differing saturation levels and dry unit weights on the physical and swelling attributes of expansive soils. This study endeavors to optimize the applied stress to attain minimal swelling, thus contributing to a more comprehensive understanding of expansive soil behavior.

The detrimental effects of expansive soils have been widely documented, and the poor evaluation of these soils can result in significant damage and financial losses [1–4,9–13,23]. Correctly classifying the degree of shrinkage and swelling of these soils is crucial for successful foundation treatment and ensuring the stability of structures built on them. However, expansive soils in different regions exhibit varying physical and engineering characteristics due to differences in their composition and environment [23]. Assessing the degree of shrinkage and expansion in expansive soils is a complex problem, as influencing factors behave with characteristics of fuzziness and uncertainty that are difficult to interpret using traditional methods. Consequently, it is necessary to develop suitable methods for analysing the classification of swell and shrinkage in expansive soils. To address this challenge, we propose a novel approach that uses machine learning to predict the optimal dry unit weight and applied stress required to mitigate swelling in expansive soils.

To assess the degree of risk associated with expansive soils, it is important to identify the main cause of hazards, which is volumetric changes [24]. Therefore, an effective solution must be developed to reduce these changes and limit the associated risks. This can be achieved by following these steps [24]:

(1) Collecting information on the damage that has occurred in the study area, the risk reduction methods employed, and their effectiveness.
(2) Conducting a comprehensive analysis of the soil and its properties (in situ and laboratory testing) by a geotechnical engineering expert.
(3) Classifying the degree of severity of the expansive soil based on the findings from steps 1 and 2.
(4) Proposing an appropriate risk mitigation strategy.

In the present study, the primary focus is on steps 3 and 4, which are considered crucial as they involve the selection of an appropriate method to mitigate the risks associated with expansive soils. The first two steps (i.e., collection of information related to damages and a thorough study of soil properties) are also essential, particularly step 2, which requires the expertise of a geotechnical specialist. Therefore, it is imperative to choose an experienced specialist with in-depth knowledge of expansive soils.

The classification process in this study relies on predicting the optimal dry unit weight, a value that corresponds to achieving the minimal swell amplitude. Subsequent to this

predictive determination, a comparative assessment in relation to the site dry unit weight is executed. In cases where the predicted dry unit weight manifests as inferior to the site's empirically observed dry unit weight, the classification of the expansive soil is assigned a precarious characterization. In contrast, when the predicted value of the dry unit weight exceeds the site's dry unit weight, the expansive soil is classified as stable, implying the absence of swelling-induced structural damage. It is worth noting that the magnitude of perceived risk is determined by the quantitative difference between the predicted and actual dry unit weight. It is important to note, however, that the precise quantification of this risk gradient is not the major focus of the current investigation.

Based on the literature, various methods have been used to address the issues related to expansive soils, such as controlling the level of compactness, chemical improvement using lime or cement, and pre-moistening of expansive soil. However, these methods are often expensive and time-consuming to implement, thus emphasizing the need for an alternative approach.

Henceforth, this study aims to propose a method that would effectively address the issue of expansive soil by reducing its volumetric changes and limiting its risks. Specifically, the focus of this paper is on step 4, which involves recommending a suitable approach to mitigate the identified risks. It is worth noting that step 2 plays a critical role in this process, and hence, a geotechnical specialist with extensive experience in expansive soils must be selected for this task. This study proposes a stress-based method that leverages the characteristics of the expansive soil. The proposed approach is primarily based on the machine learning model, which is developed through experimental work, as will be discussed later. The steps are outlined as follows:

- Place a high permeability soil layer beneath the entire building with sufficient thickness to ensure uniform moisture conditions under the structure [25].
- Use a machine learning model to determine the optimal dry unit weight for expansive soil as a fill material, or the optimal stress to achieve the lowest volumetric change.
- Design a suitable foundation that can deliver the required stress based on the building's specifications.
- Isolate any external factors that could influence soil moisture content, such as nearby trees, to prevent tree roots from absorbing moisture from beneath the structure [24].

The scientific literature has documented various experimental methods for characterizing the phenomenon of soil swelling, with some of the most commonly used methods including those outlined in references [26–29]. In order to gain a deeper understanding of the behavior of expansive soils, other researchers have sought to develop new models, as seen in references [30–32]. These models can be used to estimate the behavior of expansive soils based on simple and easily performed experiments, such as those that utilize Atterberg's limits and grain-size analysis. However, the development or modification of such models requires prior knowledge of the relationships between data that depend on a multitude of external factors related to soil composition, which are difficult to define using only traditional statistical methods due to their interdependence.

Given the complexity of expansive soils, numerous researchers have strived to integrate machine learning (ML) with geotechnical reliability analysis, with the objective of enhancing computational accuracy and efficiency. This pursuit has yielded a variety of successful applications [33–37]. In the realm of geotechnical reliability analysis, the fundamental objective of machine learning (ML) is to reconstruct intricate high-dimensional implicit performance functions by leveraging insights from meticulously curated datasets. These datasets predominantly encompass diverse input stochastic variables such as liquid limit, plastic index, unit weight, and degree of saturation. Correspondingly, the datasets include relevant quantities of interest, such as swell amplitude or swell pressure. Utilizing this methodology, the application of the ML analysis model offers an efficient means of accurately predicting outcomes in the geotechnical field. This accuracy is confirmed through rigorous training and thorough validation, demonstrating the model's ability to meet intended performance benchmarks with precision and computational efficiency.

Geotechnical analysis has effectively utilized a wide range of machine learning (ML) algorithms. These algorithms include artificial neural networks (ANN) [38–40], support vector machines (SVM) [41], relevant vector machines (RVM) [42], gradient boosting regression (GBR) [43], decision tree regression (DTR) [43], K nearest neighbor regression (KNR) [43], random forest regression (RFR) [43], particle swarm optimization (PSO) [44], extreme learning machines (ELM) [45], multivariate adaptive regression splines (MARS) [45,46], and extreme gradient boosting (XGBoost) [37,47].

Machine learning (ML) has been extensively used by researchers to model the behavior of and solve problems associated with expansive soil. For instance, Najjar et al. [48] and Najjar and Basheer [49] used ANNs to model swelling, and they concluded that the ANN technique was superior to multivariate regression analysis. ANNs were also employed by Doris et al. [50] to predict the vertical surface movement of soil due to shrinkage and swelling. In addition, ANNs were utilized by Ashayeri et al. [51] and Aissa Mamoune [52] to estimate the amplitude and swelling pressure of unsaturated clay. Similarly, Banu Ikizler et al. [53], Erzin and Güneş [54], and Merouane and Aissa Mamoune [38] estimated the swelling pressure of expansive soils using ANNs. Furthermore, ANNs were used by Dutta et al. [39] to predict the free swell index of the expansive soil. These studies have demonstrated the effectiveness of the artificial neural network technique. Ikeagwuani [55] employed three distinct machine learning models, namely multivariate adaptive regression splines (MARS), random forest, and gradient boosting machine models, for the purpose of predicting the California bearing ratio (CBR) of expansive soil subgrade amended with constituents such as sawdust ash, ordinary Portland cement, and quarry dust. The outcomes of the study demonstrated that the random forest model exhibited enhanced predictive efficacy as compared to the MARS and gradient boosting machine models. Eyo et al. [56] employed a diverse set of machine learning algorithms to assess and forecast the behavior of soils characterized by varied plastic properties when subjected to expansive behavior under inundation conditions. The results highlighted that the support vector machine (SVM) achieved higher accuracy compared to other methods such as the artificial neural network (ANN), logistic regressor (LR), and random decision forest (RDF).

Geotechnical engineers have shown a growing interest in machine learning due to its remarkable ability to model nonlinear problems with multiple variables. Machine learning algorithms can establish non-linear relationships between variables and provide reasonably accurate predictions [57]. In fact, several geotechnical studies have shown that the models developed by machine learning are much more effective than those developed by multiple linear regression (MLR) [53,57,58].

In contrast to many previous studies and models that often overlooked the interactive effects of parameters, our research uniquely addresses the combined influence of the sand content, initial degree of saturation, and dry unit weight on soil swelling. While existing studies have made valuable contributions by focusing on individual factors such as the initial degree of saturation and dry unit weight, they have not adequately considered two critical aspects: the impact of sand content on soil swelling characteristics and the determination of optimal applied stress and dry unit weight to reduce volume changes in expansive soil. This study stands out for its significant importance in predicting the optimal applied stress and dry unit weight for mitigating soil volumetric changes, leveraging machine learning as a highly effective tool for this purpose. The study introduces a novel approach in the application of established machine learning algorithms. Although the algorithms themselves are not novel, their application to predict optimal applied stress and dry unit weight for expansive soils, while considering the combined influence of multiple parameters, represents a unique and valuable contribution to the field, marking a notable advancement in the understanding and application of these predictive models. To achieve the objectives of this study, the definitions of 'optimal unit weight' and 'optimal applied stress' are paramount, precisely delineated through thorough laboratory investigations employing loaded swelling pressure tests. The primary goal is to identify the specific dry unit weight and applied stress levels, referred to as the 'optimal unit weight' and 'optimal

applied stress,' at which swelling amplitudes in expansive soils become negligible. This is explicitly defined in the experimental work of the study, specifying that the condition for negligible swelling amplitudes is when they reach zero.

2. Materials and Methods

2.1. Materials and Classification Tests

To achieve the research objective, expansive clay soil was excavated from Demsarkho-Lattakia at a depth of 3 m from the ground level, surpassing the backfill soils to access the targeted soil layer. Fine-grained marine sand was used for mixing with the expansive clay, which was washed to eliminate fine particles and obtain clean sand suitable for experiments. Mixtures were prepared by adding different percentages of sand (10%, 20%, 30%, 40%, and 50%) to the expansive soils based on the dry weight. To determine the grain size distribution, experiments were conducted using the dry sieving method for fine sand and hydrometer sedimentation for clay soil according to ASTM standards [59,60]. Figure 1 illustrates the granular gradient curves of the mixed soil, based on the percentage of sand added to it. Specific weight tests were carried out in accordance with ASTM D854-98 [61], revealing a specific gravity of 2.7 for the expansive clay and 2.65 for the sand. The chemical composition of clay has been listed in Table 1.

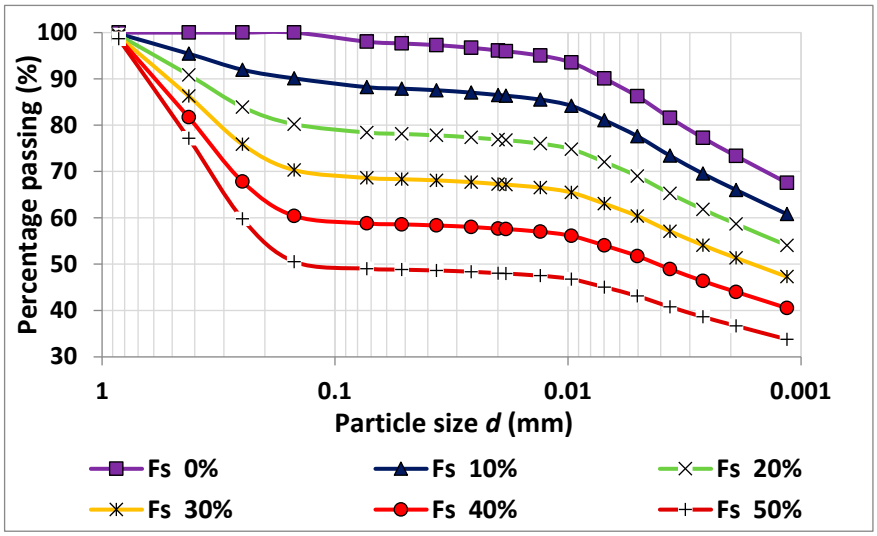

Figure 1. The granular gradient curves of the tested mixtures.

Table 1. Chemical properties of clay.

Chemical Composition	%
Alumina (Al_2O_3)	11.5
Ferric (Fe_2O_3)	5.5
Calcium (CaO)	12
Magnesium (MgO)	2.4
Silica (SiO_2)	48.8
Sodium (Na_2O)	1.2
Potassium (K_2O)	0.36
Loss of ignition (LoI)	17.24

The Atterberg Limits experiments were performed in accordance with ASTM international standard D854-14 [62] on mixtures with varying percentages of sand, ranging from 0 to 50%. The results of the experiments, depicted in Figure 2, reveal the variation in the Liquid Limit (LL), Plastic Limit (PL), and Plastic Index (PI) with respect to the quantity of sand added. These findings are in line with previous studies investigating the effect of adding sand to the expansive soil for the purpose of soil improvement [14,22,63].

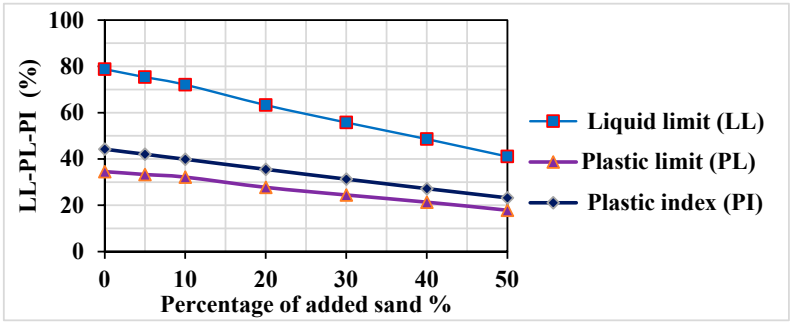

Figure 2. The relationship curves between the liquid limit, plastic limit, and plastic index with the percentage of added sand.

The clear clayey soil used in this study, without sand, was classified as A-7-5 according to the AASHTO classification system [64] and as CH according to the Unified Soil Classification System [65]. The degree of volume change was determined to be high to very high according to [66–68]. The swelling pressure of the utilized expansive soil, prepared under standard proctor parameters of γ_d = 13.95 kN/m^3 and SR = 91.2%, revealed a substantial swelling pressure of almost 200 kPa while the swelling pressure of the utilized expansive soil, prepared at γ_d = 15.3 kN/m^3 and SR = 75%, was observed to be almost 520 kPa. The percentage of clay particles (with a diameter smaller than 0.02 mm) was found to be 73.9%. The free swell [69] for clear clayey soil was determined to be 127%.

Standard proctor experiments were carried out to determine the maximum dry unit weight and optimal moisture for each mixing percentage, as illustrated in Figures 3 and 4. The results indicate a continuous increase in the maximum dry unit weight, which is consistent with the findings of previous studies [63,70]. This can be attributed to the reduction in soil pore volume and suction stress resulting from the replacement of a soft component, which can hold a large amount of water, with a coarse component that has a low water-holding capacity. This is further supported by the optimal moisture values of the mixtures, which decrease as the percentage of added sand increases, as shown in Figure 4. For more details on the Proctor, Atterberg, and consolidation tests, see Alnmr and Ray [71].

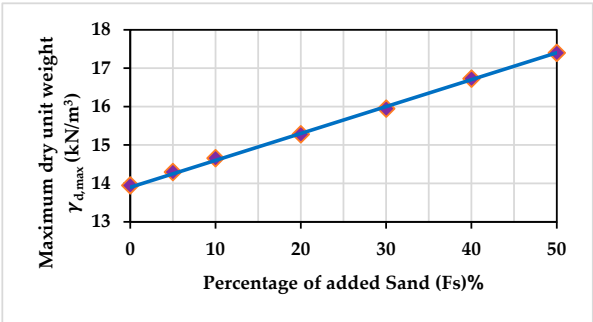

Figure 3. Relationship between maximum dry unit weight and percentage of added sand.

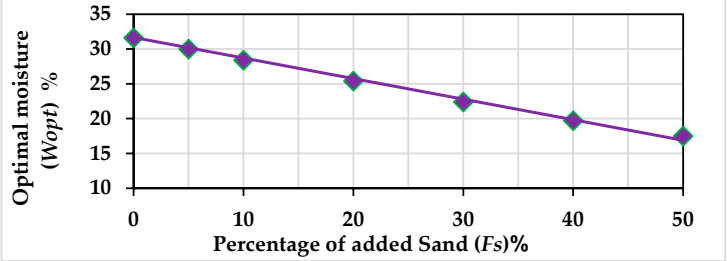

Figure 4. Relationship between optimal moisture and percentage of added sand.

2.2. Samples Preparation

The sample preparation involved dry mixing with the required sand percentage to achieve homogeneity, followed by the addition of the appropriate amount of water to reach the desired degree of saturation and thorough mixing, as shown in Figure 5a. The samples were then placed in plastic bags to retain moisture, as depicted in Figure 5b, and left in an insulated container for a day to allow for uniform moisture distribution and suction. Finally, the samples were formed to the desired unit weight using a hydraulic piston, as illustrated in Figure 5c–e. As visually demonstrated in Figure 5c, the soil sample was positioned within the confines of the ring. Subsequently, static pressure was meticulously applied through the utilization of a hydraulic piston. This pressure application was executed with precision to attain the targeted dry unit weight while concurrently ensuring a level and uniform surface, as depicted in Figure 5d,e.

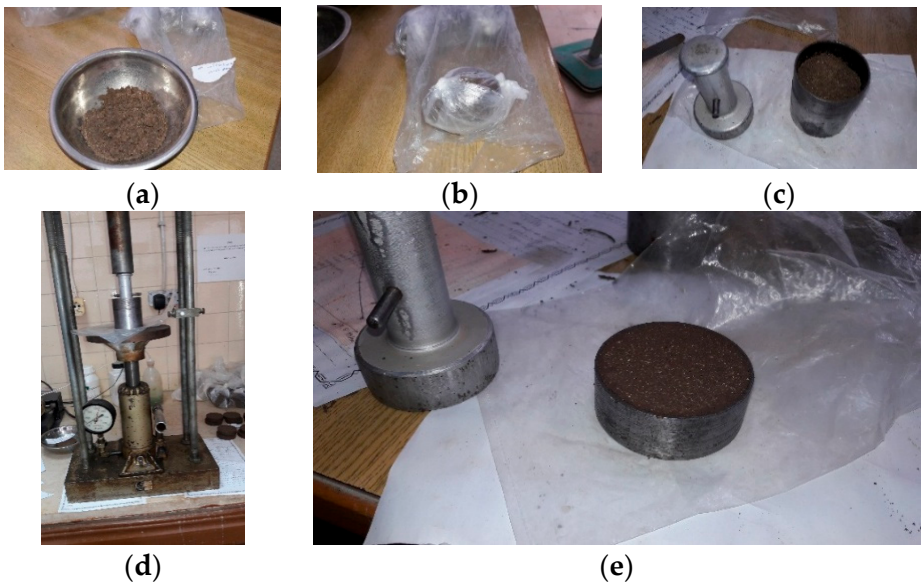

Figure 5. Images of the sample formation process step-by-step: (**a**) mixing sand with expansive soil, (**b**) placing the sample in plastic bags for moisture retention, (**c**) positioning it within the ring, (**d**) applying meticulous static pressure through a hydraulic piston, and (**e**) presenting the final prepared specimen.

2.3. Loaded Swelling Pressure Tests

The loaded swelling pressure tests were performed on samples with different densities and degrees of saturation for each sand mixing percentage. In this method, a procedure was employed involving the loading of three or more specimens at various pressures. This process enabled the specimens to absorb water, leading to either swelling or compression until equilibrium positions were achieved. These positions aligned linearly on the swell versus logarithmic pressure plot as shown in Figure 6. While this methodology mandates the use of a minimum of three identical specimens, it provides the advantage of comparably reduced time demands. In the context of the loaded swelling pressure tests, the soil specimens were meticulously placed within a cylindrical metal ring, characterized by dimensions measuring 71.4 mm in diameter and a height of 20 mm. To mitigate any potential effects of friction between the soil and the inner surface of the ring, a lubricant was applied to the lateral sides of the ring.

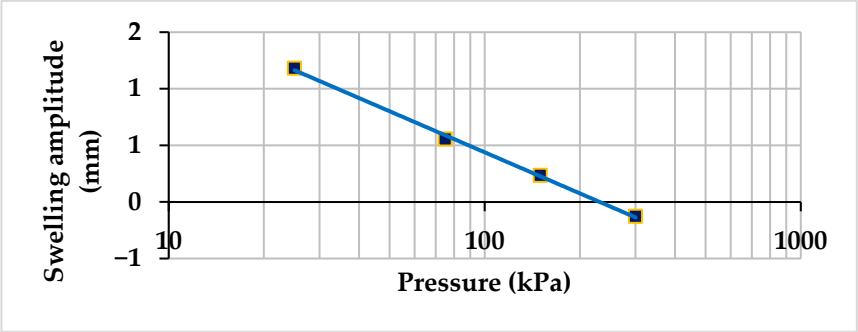

Figure 6. Relationship between swelling amplitude and applied stress in loaded swelling pressure test for initial condition (Fs = 10%, γ_d = 14.66 kN/m^3, SR = 75%).

After meticulous sample preparation, as illustrated in Figure 5, filter papers were inserted both above and below the soil sample to prevent the ingress of fine soil particles into the porous discs' interstices. Subsequently, a loading cap was installed atop the upper porous disc, and the entirety of this composite assembly was carefully positioned within the oedometer cell. Finally, the load transfer frame was installed onto the loading cap, as exemplified in Figure 7. The specimens were subjected to varying stress levels (25, 75, 150, 300 kPa). Afterwards, the specimens underwent immersion in distilled water, inducing swelling or settlement contingent upon the initial conditions and applied stress. Following the attainment of equilibrium in deformation, the swelling amplitude was meticulously recorded for each applied stress.

1. **Device frame**
2. **Oedometer cell**
3. **Displacement measurement's watch**
4. **Load transfer frame**
5. **Load balance piece**
6. **Magnification arm**
7. **Magnification's arm support screw**
8. **Weight carrier**

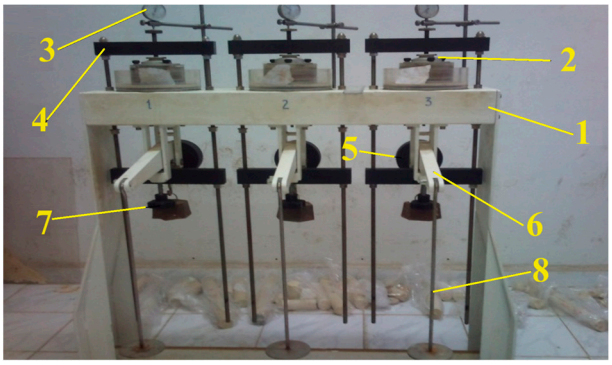

Figure 7. Consolidation test device (oedometer).

The findings presented in Figure 6, specific to initial conditions (Fs = 10%, γ_d = 14.66 kN/m^3, SR = 75%), revealed a distinctive swelling pressure of 232 kPa. This observation signifies that, under these specified initial conditions, the optimal dry unit weight of 14.66 kPa should align with an applied stress of 232 kPa to achieve a swelling amplitude of 0. Correspondingly, for γ_d = 14.66 kN/m^3, the optimal applied stress for these specific initial conditions was determined to be 232 kPa. These outcomes contribute significantly to our understanding of the complex interplay between initial conditions, optimal applied stress, and optimal dry unit weight in expansive soils.

Based on the results of the experiments, a dataset of 657 unconfined compression tests was created for use in developed machine learning models. The input data include: liquid limit (LL), clay content (Fc), silt content (Fsi), sand content (Fs), specific degree of saturation (SR), maximum dry unit weight (γ_{dmax}) of proctor test, applied stress (σ), and dry unit weight (γ_d), This dataset is used to predict the optimal applied stress (σ_{min}) and optimal dry unit weight (γ_{max}). The key steps of the applied methodology are depicted in Figure 8. Figure 9 displays the distribution histograms and density plots for the dataset.

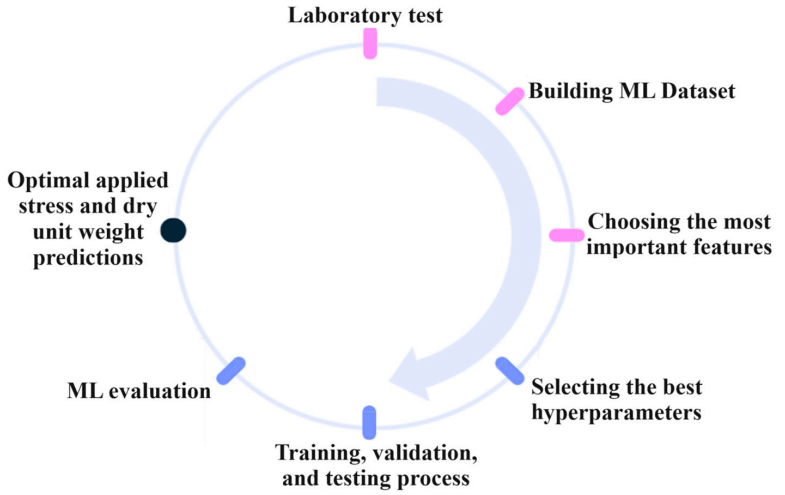

Figure 8. The key steps of the applied methodology.

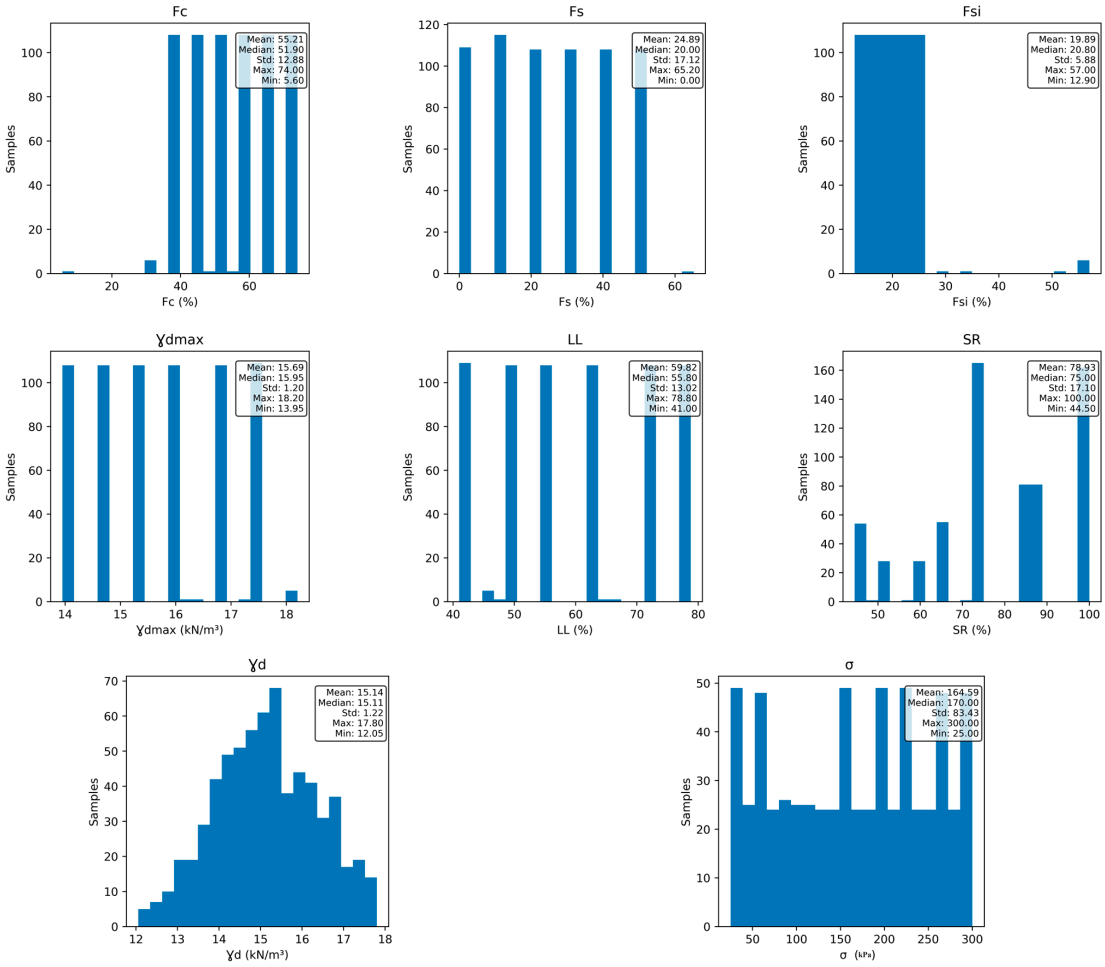

Figure 9. Histograms of distribution and density plots for the variables studied in this study.

2.4. Selection of Variables

Correlation analysis is a useful tool for analyzing correlations between different inputs inside datasets. However, because datasets vary in nature, the assumptions underpinning each correlation approach typically differ, limiting their application and robustness.

The final selection criteria are removing variables that are difficult to collect and choosing data that are easily accessible. The Pearson, Kendall, and Spearman correlation

coefficients are computed among various parameters in this study to investigate the correlation between input and output parameters, as shown in Figure 10. The correlation between factors is represented by values of 0, 1, and −1, with 1 representing a strong positive connection and −1 representing a significant negative correlation. When the three approaches (Pearson, Kendall, and Spearman) were compared across different datasets and inputs, the variable, γ_d consistently had the highest correlation with applied stress (σ). Whatever the makeup of the dataset or the underlying assumptions of each correlation approach, this consistent pattern indicates a solid and robust link between γ_d and σ. As a result, these two parameters, as well as LL, γ_{dmax}, and SR, were chosen as inputs due to their importance in soil categorization and ease of assessment via simple experiments. Table 2 displays the statistical properties of the variables chosen. Table 3 shows the inputs and output boundaries for the developed models.

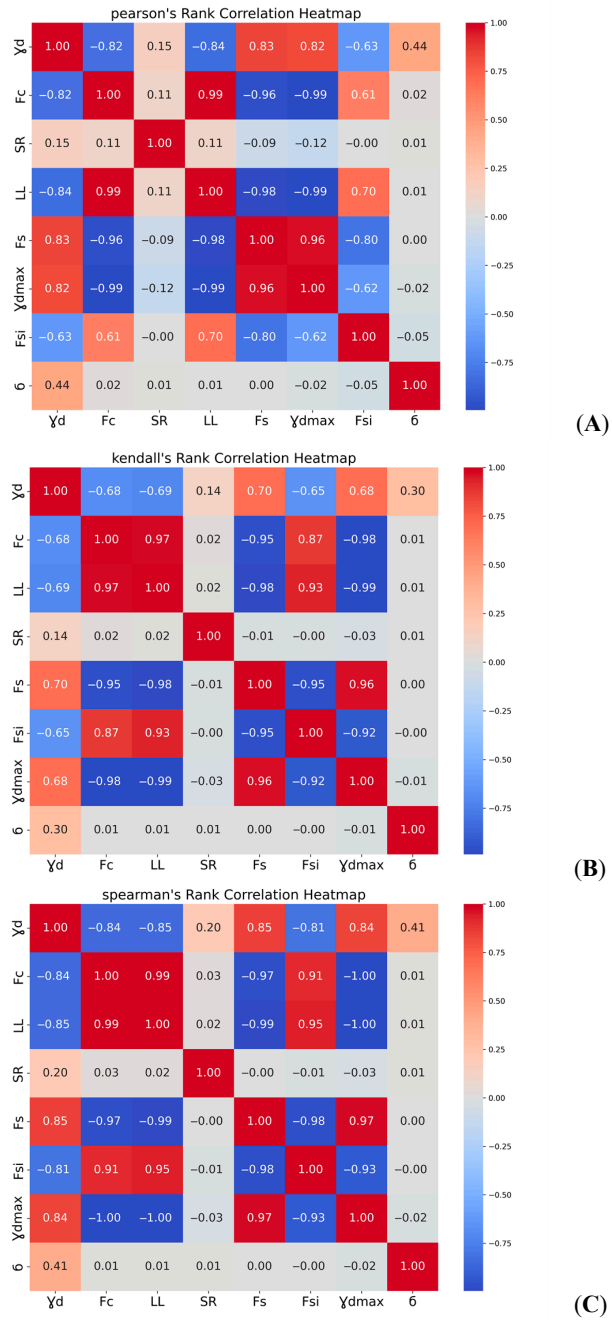

Figure 10. Correlation coefficients used to examine the influence of various input parameters on 'qu' and 'Et'; (**A**) Pearson, (**B**) Kendall, (**C**) Spearman.

Table 2. Summary of the statistical characteristics of the variables.

	LL (%)	SR (%)	γ_d (kN/m^3)	γ_{dmax} (kN/m^3)	Applied Stress [σ] (kN/m^2)
mean	59.82	78.9	15.14	15.69	164.59
std	13.02	17.1	1.22	1.20	83.43
min	41.00	44.5	12.05	13.95	25.00
25%	48.60	65.0	14.26	14.66	87.50
50%	55.80	75.0	15.11	15.95	170.00
75%	72.10	88.0	16.05	16.73	240.00
max	78.80	100.0	17.80	18.20	300.00

Table 3. Boundaries, inputs, and outputs used for the developed models.

Models to Predict Dry Unit Weight (γ_{max})			Models to Predict Applied Stress (σ_{min})	
Parameters	**Minimum**	**Maximum**	**Minimum**	**Maximum**
	Input parameters		Input parameters	
LL (%)	41	79	41	79
SR (%)	45	100	45	100
σ (kN/m^2)	25	300		
γ_{dry} (kN/m^3)			12	17.8
γ_{dmax}	13.95	18.2	13.95	18.2
	Output parameter		Output parameter	
γ_{max} (kN/m^3)	12	17.8		
σ_{min} (kN/m^2)			25	300

2.5. Machine Learning Algorithms Used in the Study

In this section, a comprehensive comparison is conducted among the algorithms employed in this study, including decision tree regression (DTR), random forest regression (RFR), gradient boosting regression (GBR), support vector regression (SVR), and artificial neural networks (ANN). These algorithms have been chosen for their common use in previous studies.

2.5.1. Decision Tree Regression

Decision tree regression, or DTR, represents a hierarchical data structure characterized by a dynamic arrangement of branches and nodes. Within this structure, nodes exhibit lines going outwards, and some nodes, termed 'leaves', lack such extensions. The mechanism involves the division of data points intended for regression or categorization into two or more distinct categories using specific internal nodes. In the training phase, input variable values undergo comparisons against designated functions. The algorithm systematically endeavors to construct optimal decision trees by iteratively minimizing the fitness function. At various junctures within each context of independent variables, the dataset undergoes division. During this process, the algorithm computes prediction errors, representing disparities between projected and actual values, guided by the fitness function. The determination of the optimal split point hinges on identifying the variable that yields the smallest fitness function value, a process that entails evaluating split point errors across each variable. In the context of decision tree regression, several key hyperparameters play pivotal roles in shaping the model's performance. One such parameter is the 'max_depth,' which dictates the maximum depth of the decision tree. A deeper tree can capture intricate relationships in the training data but risks overfitting. On the other hand, 'min_samples_split' determines the minimum number of samples required to split an internal node. It influences the threshold for further division, thereby affecting the model's

generalization capability. Additionally, the 'min_samples_leaf' parameter establishes the minimum number of samples necessary to constitute a leaf node [72].

2.5.2. Random Forest Regression

Random forest regression (RFR) stands out as a powerful ensemble learning algorithm widely recognized for its efficacy in both academic and practical settings, particularly in tasks involving classification and regression. The strength of RFR lies in its innovative use of bootstrap aggregation, which facilitates the creation of an ensemble comprising diverse, randomly constructed, and unpruned decision trees. The ensemble is meticulously crafted by systematically altering a group of decision trees, and a pivotal element in this process is the introduction of random feature selection. This strategic approach ensures a rich diversity of decision trees, making the judicious selection of different attributes imperative. Key hyperparameters govern the behavior of the RFR model. The max_depth parameter controls the maximum depth of individual decision trees, influencing the model's capacity to capture intricate relationships. The parameter min_samples_leaf determines the minimum number of samples required in a leaf node, guarding against nodes with too few samples. Simultaneously, min_samples_split sets the minimum number of samples needed to split an internal node, contributing to model generalization. The critical n_estimators parameter specifies the number of trees in the ensemble, influencing overall model performance. The careful tuning of these hyperparameters is essential to strike a balance, avoiding overfitting and ensuring the generalization capability of the random forest model. The RFR's ability to aggregate votes from diverse decision trees makes it a robust tool for diverse applications, with detailed explanations available in prior studies [55,73,74].

2.5.3. Gradient Boosting Regression

Gradient boosting regression (GBR) stands as a formidable ensemble-trained supervised machine learning model renowned for its predictive capabilities. This method operates by amalgamating numerous simple models into a singular composite model, a technique recognized as boosting. Boosting is often characterized as an additive model due to its incremental addition of basic models while maintaining the model's trees unchanged. The amalgamation of more fundamental models consistently enhances predictive accuracy. Central to gradient boosting is the reduction in losses through gradient descent, an iterative optimization procedure of the first order. In the realm of gradient boosting, decision trees function as weak learners, employing a squared error loss function. GBR orchestrates the training of a weak model to map features to the anticipated residuals of that weak model. These residuals are seamlessly integrated into the input of the current model, steering it toward the desired outcome. Through recurrent iterations of this process, the overall predictive accuracy of the GBR model steadily advances. The effectiveness of gradient boosting regression (GBR) is profoundly influenced by key hyperparameters that govern its behavior. The subsample parameter determines the fraction of samples used for fitting the individual base learners, playing a role in mitigating overfitting. The parameter n_estimators specifies the number of boosting stages or trees, directly impacting the complexity and performance of the model. The min_samples_split hyperparameter dictates the minimum number of samples required to split an internal node, influencing the model's tendency to partition the data. Simultaneously, min_samples_leaf sets the minimum number of samples needed in a terminal or leaf node, contributing to the prevention of overly specific nodes. The max_depth parameter controls the maximum depth of the individual decision trees, managing their capacity to capture complex relationships. Lastly, the learning_rate hyperparameter governs the contribution of each tree to the model, regulating the impact of new trees on the overall ensemble [75,76].

2.5.4. Extreme Gradient Boosting

Extreme gradient boosting (XGBoost), a prominent implementation of gradient boosting machines (GBM), stands out as a high-performing tool in supervised learning, versatile

for both regression and classification challenges. Its accelerated execution speed, particularly in out-of-core computations, makes it a preferred choice among data scientists. Powered by advanced mechanisms such as tree pruning, regularization, and gradient boosting, XGBoost synergistically enhances predictive capabilities, demonstrating adaptability in handling missing data instances and providing regularization strategies. To unlock its full potential, meticulous hyperparameter tuning is paramount. The learning rate and maximum depth, among other key hyperparameters, play crucial roles in optimizing efficiency. The 'colsample_bytree' parameter controls the subsample ratio of columns when constructing each tree, influencing model diversity. The 'subsample' parameter determines the fraction of training samples used for tree fitting, contributing to the algorithm's robustness. 'N_estimators' specifies the number of boosting rounds, impacting the overall model complexity. Together, these hyperparameters contribute to the fine-tuning process, ensuring XGBoost operates at its peak performance for various predictive tasks [77].

2.5.5. Support Vector Regression

Support vector machine (SVM) has found extensive application in diverse predictive scenarios, encompassing both classification and regression tasks. In the realm of regression, it takes on the name of support vector regression (SVR) [78]. In regression, the model undergoes a transformation of inputs into a higher-dimensional space, constituting the foundation of SVR's training process grounded in the principles of structural risk minimization (SRM) [79]. A pivotal role in this intricate transformation is assigned to a kernel function, responsible for mapping inputs to the higher-dimensional space. The key hyperparameters critical to SVR's performance are C, epsilon, and kernel. The parameter C influences the trade-off between a smooth decision boundary and accurate fitting of the training data. Epsilon defines the margin of tolerance, determining the sensitivity of the model to errors. The choice of kernel, whether linear, polynomial, or radial basis function (RBF), significantly affects the model's ability to capture complex relationships within the data. Each hyperparameter plays a crucial role in shaping the SVR model's accuracy and generalization capability, demanding careful consideration during the model tuning process.

2.5.6. Artificial Neural Networks

Artificial neural networks (ANN) constitute a computational framework comprised of artificial neurons designed to mimic information transfer processes akin to the human brain, enabling knowledge acquisition. They create sophisticated input–output models capable of discerning intricate relationships within multidimensional data, finding applications across various engineering disciplines [80,81]. ANN displays diverse network configurations, encompassing single and multiple layers. For instance, the feedforward back propagation neural network (FFBP) features an input layer for data reception, an intermediate hidden layer, and culminates in an output layer providing outcome-specific insights in response to input stimuli. Key hyperparameters governing the performance of ANN models include learning_rate, batch size, hidden layers function, and linkage between the hidden layer and the ultimate output layer function. The learning_rate determines the step size during optimization, influencing the convergence speed and model stability. Batch size regulates the number of samples processed before updating model parameters, impacting both computational efficiency and model generalization. The choice of the hidden layers function dictates the activation function applied to neurons within the hidden layers, shaping the network's ability to capture complex patterns. The linkage function between the hidden layer and the ultimate output layer is pivotal in defining how information flows through the network, determining the model's capacity to map inputs to accurate outputs. Figure 11 depicts an ANN model with four neurons in the input layer, two hidden layers with five neurons in each, and an output layer.

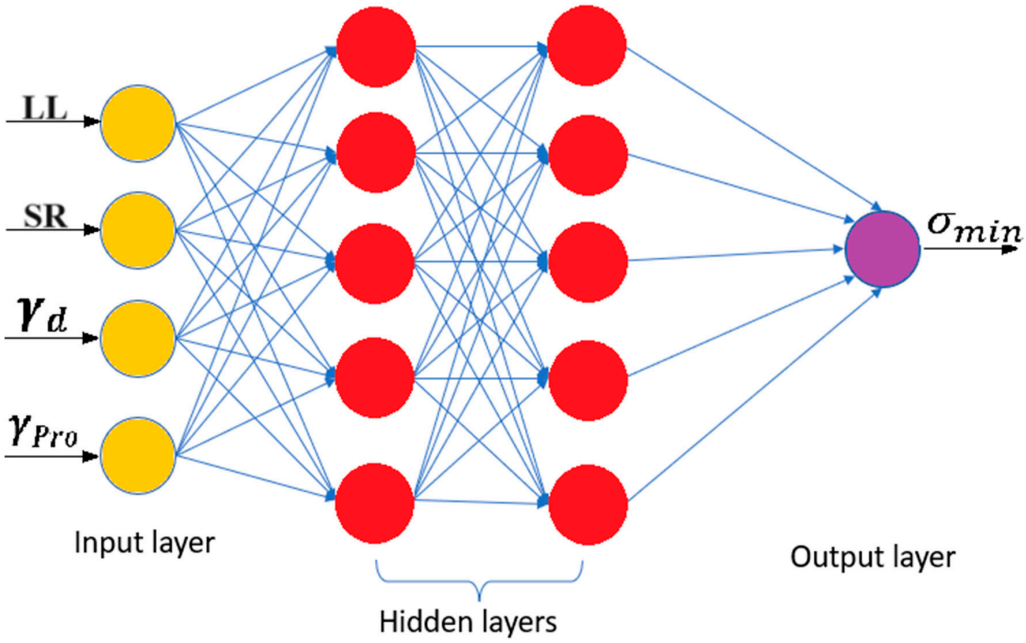

Figure 11. An example of a visualized neural network plot with two hidden layers of five neurons each and four features.

In the realm of machine learning, hyperparameters play a pivotal role in shaping the behavior and performance of models. These parameters serve as configuration settings, influencing how a model learns patterns from data. The careful adjustment of hyperparameters is crucial to achieving optimal model accuracy and overall effectiveness in solving specific tasks. A thoughtful tuning of hyperparameters is, therefore, essential to harness the full potential of machine learning algorithms, enabling them to adapt and perform well across diverse datasets.

2.6. Building the Dataset and Evaluation of the Models

In this study, the previously mentioned algorithms outlined in Section 2.2 are utilized to construct developed models to predict the applied stress and dry unit required to achieve the lowest swell amplitude in expansive soil. To formulate these models, various geotechnical identification parameters of expansive soil will be introduced as inputs to the machine learning models. These parameters include the liquid limit (*LL*), initial degree of saturation (SR), maximum dry unit weight (γ_{dmax}) as determined by the standard Proctor test, and the dry unit weight (γ_d) to predict the optimal applied stress (σ_{min}), while the parameters including liquid limit (*LL*), initial degree of saturation (SR), maximum dry unit weight (γ_{dmax}), and the applied stress (σ) will be used to predict the optimal dry unit weight (γ_{max}) as shown in Table 3. The output parameter will be the applied stress and dry unit weight required to achieve the lowest level of swell amplitude (almost zero) in expansive soil.

In evaluating the effectiveness of the developed models, the initial step involves subjecting them to scrutiny using the training data. Subsequently, to assess the models' generalization capabilities, the method of 'Cross-Validation' is employed. This involves testing the models with datasets distinct from those used during the training phase. In this research, a total of 657 datasets were used in developing each predictive model, specifically designed to estimate the applied stress needed to achieve minimal volumetric changes. Among these datasets, 70% were allocated for the training phase, while the remaining 30% were carefully reserved for the crucial purposes of validation and testing, employing the Cross-Validation methodology.

The dataset used in this study comprises 657 systematically curated entries, enhancing the precision of our machine learning models. Specifically, 648 entries were generated

from experimental findings, systematically covering various parameters such as sand percentages, initial degrees of saturation, and dry unit weights. For each sand percentage, 16 samples, corresponding to unique liquid limits, were generated. Within each degree of saturation, four initial dry unit weights were considered, resulting in 16 samples for each percentage. Each sample underwent replication four times and testing with applied stress values (25–300 kPa) to identify dry unit weight and stress levels nullifying swell amplitude. To address the significant impact of applied stress on predictive accuracy, we employed a condensed approach for each sand percentage. This involved deriving trendline equations capturing the relationship between applied stress and the corresponding dry unit weight, with the goal of nullifying swell amplitude for each saturation degree. These equations enabled the projection of dry unit weights across a range of stress values (25–300 kPa) in systematic 10 kPa intervals. The dataset was thus expanded to encompass 27 consolidated stress levels, each multiplied by four initial saturation degrees and further multiplied by six distinct sand percentages, resulting in the generation of 648 datasets. Additionally, nine data points from previous research by Zou [82] and Rosenbalm [83] were included, contributing to the compilation of 657 systematically curated datasets.

During the training, validation, and testing process of the models, the performance was evaluated by calculating the root mean square error (RMSE) (Equation (1)), mean absolute error (MAE) (Equation (2)), and the coefficient of determination (R^2) (Equation (3)). The RMSE is the root of average squared difference between the predicted outputs and the actual targets, where a lower RMSE value indicates better performance. On the other hand, the correlation coefficient (R), which is the square root of the coefficient of determination, measures the degree of correspondence between the predicted outputs and the actual targets. A value of R close to 1 indicates a more accurate estimation.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i' - Y_i\right)^2}{n}} \tag{1}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\left(Y_i' - Y_i\right)\right| \tag{2}$$

$$R^2 = \left(\frac{\sum_{j=1}^{N}\left(Y - \overline{Y}\right)\left(Y_j - \overline{Y_j}\right)}{\sqrt{\sum_{j=1}^{N}\left(Y - \overline{Y}\right)^2}\sqrt{\sum_{j=1}^{N}\left(Y_j - \overline{Y_j}\right)^2}}\right)^2 \tag{3}$$

where:

Y, Y_j—refers to the observed values and expected values by machine learning models.
$\overline{Y}$, $\overline{Yj}$—are the average of Y, Y_j, respectively.
N—represents the number of data.

2.7. Hyperparameters Optimization

This study adopts a systematic approach to hyperparameter tuning, focusing on optimizing machine learning models exclusively through the GridSearchCV technique. GridSearchCV is carefully applied to XGBoost, GBR, SVR, DTR, ANN, and RFR. This method entails an iterative exploration of a predetermined grid of hyperparameter values, conducting cross-validation for each combination to identify the optimal set that maximizes model performance. Given the smaller and more manageable hyperparameter spaces of these models, a comprehensive search proves both practical and advantageous [74,84,85]. The study acknowledges that both the GBR and ANN models feature larger and more complex hyperparameter spaces, necessitating more time for optimization. However, for the sake of consistency, the study opts to utilize GridSearchCV for GBR as well. This strategic decision ensures a thorough and systematic exploration of the hyperparameter space, aligning with the study's methodology and eliminating the need for Randomized Search CV [85].

The TensorFlow-based artificial neural network (ANN) assumes a pivotal role in capturing intricate patterns within the dataset. The ANN architecture, comprised of interconnected nodes organized into layers, is employed for regression tasks, mapping input features to continuous output values. Hyperparameter optimization for TensorFlow-based ANNs involves a comprehensive exploration of the parameter space, encompassing network architecture parameters (e.g., layer count, neuron count per layer, and activation functions) as well as training process parameters (e.g., learning rate, batch size, and the number of training epochs). This systematic process aims to identify the hyperparameter combination that maximizes predictive accuracy, considering various configurations and leveraging TensorFlow's capabilities for efficient training and evaluation.

Table 4 shows the optimal configurations for each model concerning γ_{max} and σ_{min}, respectively.

Table 4. The optimal hyperparameters for models.

Model	Hyperparameter	Values Range	Optimal Values	
			σ_{min}	γ_{max}
DTR	max_depth	5 to 15	14	13
	min_samples_leaf	1 to 5	1	1
	min_samples_split	1 to 5	3	1
RFR	max_depth	5 to 15	14	10
	min_samples_leaf	1 to 5	1	1
	min_samples_split	1 to 5	1	1
	n_estimators	50 to 150	110	130
GBR	subsample	0.5 to 1	0.8	0.9
	n_estimators	50 to 250	250	200
	min_samples_split	2 to 20	17	10
	min_samples_leaf	1 to 20	3	2
	max_depth	3 to 15	10	5
	learning_rate	0.01 to 0.4	0.1	0.2
XGBoost	colsample_bytree	0.6 to 1	1	0.8
	learning_rate	0.01 to 0.4	0.1	0.2
	max_depth	1 to 15	7	6
	n_estimators	50 to 250	200	200
	subsample	0.5 to 1	0.8	0.6
SVR	C	0.1 to 10,000	1000	10
	epsilon	0.001 to 100	10	0.01
	kernel	linear, poly, rbf	rbf	rbf
ANN	number of neurons in layer1	2 to 100	8	8
	number of neurons in layer2	2 to 100	56	56
	learning_rate	0.01 to 0.4	0.05	0.025
	batch size	10 to 50	16	12
	hidden layers function	ReLU, tanh, linear, Sigmoid	ReLU	tanh
	linkage between the hidden layer and the ultimate output layer function	ReLU, tanh, linear, Sigmoid	linear	ReLU

3. Results and Discussion

3.1. Experimental Work Results and Discussion

The relationship between the swelling amplitude and the dry unit weight of the soil samples was examined for each percentage of added sand, with an initial degree of saturation of 75%, as shown in Figure 12. The results indicate that the swell amplitude increases as the dry unit weight of the soil samples increases, for all percentages of added sand. It should be noted that the swelling values decrease with an increase in the applied stress to the soil. Similar observations were made for the other initial degree of saturation.

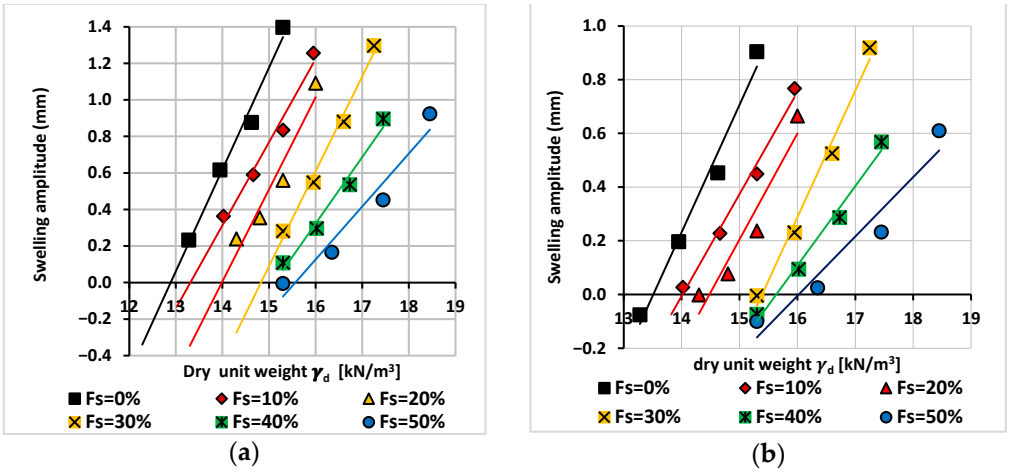

Figure 12. Swelling amplitude vs. dry unit weight for different percentages of sand at 75% degree of saturation; (**a**) applied stress = 75kPa, (**b**) applied stress = 150kPa.

For each applied stress considered as σ_{min}, the dry unit weight (γ_{max}) at which volumetric changes become insignificantly small (indicated by a swelling amplitude of 0) was determined. This determination is illustrated in Figure 13, which presents the correlation between dry unit weight and the percentage of added sand at various applied stress levels that lead to a swelling amplitude of 0, while maintaining a degree of saturation at 75%. An analysis of Figure 13 reveals that, under the same applied stress resulting in minimal volumetric changes (swelling amplitude = 0), the necessary dry unit weight increases proportionally with the increase in sand content up to approximately 30%. Beyond this point, the rate of increase becomes less prominent. Furthermore, it is noteworthy that as the dry unit weight increases, the required applied stress to achieve zero swell amplitude also increases. Similar results hold true for other degrees of saturation as well. Consequently, structures with varying weights require a careful consideration of foundation dimensions. This ensures that the applied stress aligns with the initial dry density and degree of saturation, ultimately minimizing foundation heave.

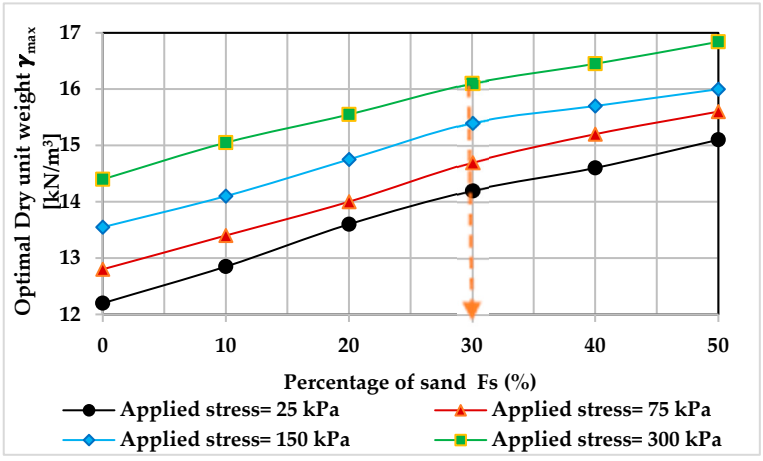

Figure 13. The dry unit weight corresponding to swelling amplitude = 0 vs. sand percentage for 75% degree of saturation.

In view of the complexity of the swelling phenomenon and the significant variability in the results obtained through statistical methods, machine learning is proposed to assist in predicting the applied stress required to minimize volumetric changes. The inputs to the model will comprise data obtained from simple experiments that are not overly

time-consuming or resource-intensive, while the outputs will be used to compare and validate the model's predictions against experimental data.

3.2. Machine Learning Results and Discussion

The Results and Discussion section initiates with a comparative analysis aimed at evaluating the strengths and weaknesses of each model in relation to others, offering nuanced insights. Subsequently, a detailed examination of each model's training and test datasets is conducted separately, ensuring a comprehensive understanding of their individual performance.

3.2.1. Comparison of Machine Learning Models

This section provides a concise comparative analysis of various machine learning models used in this study, with a specific focus on their performance across training validation and testing datasets. The evaluation includes both quantitative metrics, as presented in Table 5, and visual representations depicted in Figures 14 and 15. Additionally, Figure 16A–F and Figure 17A–F provide graphical illustrations that collectively enhance the understanding of the relative predictive capabilities of these models. Based on the metrics presented in Table 5 and supported by Figures 14 and 15, ANN and XGBoost had superior performance measures. However, ANN performed better than XGBoost in predicting σ_{min}. ANN and XGBoost achieved the highest R-squared values and the lowest RMSE and MAE, establishing themselves as the top performers in predicting applied stress and dry unit weight. GBR followed closely in second place, showcasing strong predictive capabilities, while SVR and RFR ranked third in terms of predictive accuracy. However, DTR demonstrated a less favorable performance, indicated by its performance metrics and instances of overestimation for specific data points. It is important to note that while SVR achieved metrics similar to GBR, its lower ranking was due to the presence of mispredicted points that deviated significantly from the best fit trend line (ideal line) as shown in Figures 16 and 17. As a result, in predicting σ_{min} and γ_{max} for the expansive soil under consideration, ANN and XGBoost outperformed GBR, RFR, SVR, and DTR.

Table 5. The metrics for different algorithms used in the study.

Algorithm	Metrics	σ_{min}				γ_{max}			
		Training	Validation	Testing	r_d (%)	Training	Validation	Testing	r_d (%)
DTR	R^2	0.995	0.9337	0.9372	5.8	1	0.9931	0.9845	1.6
	RMSE	5.79	22.63	21.79	276.3	0	0.10344	0.158	-
	MAE	3.65	17.414	16.414	349.7	0	0.0816	0.0965	-
RFR	R^2	0.9946	0.9698	0.9636	3.1	0.999	0.9937	0.9924	0.7
	RMSE	5.975	15.273	16.6	177.8	0.0374	0.0985	0.111	196.8
	MAE	4.728	12.344	12.954	174.0	0.0244	0.0687	0.0696	185.2
GBR	R^2	0.9997	0.9839	0.9789	2.1	0.999	0.9954	0.9954	0.4
	RMSE	1.385	11.16	12.64	812.6	0.01	0.085	0.087	770.0
	MAE	1.017	8.332	9.172	801.9	0.0081	0.052	0.0463	471.6
XGBoost	R^2	0.9999	0.9841	0.9856	1.4	0.9999	0.9985	0.9976	0.2
	RMSE	0.932	11.1	10.43	1019.1	0.01	0.048	0.062	520.0
	MAE	0.7035	8.19	7.92	1025.8	0.0068	0.035	0.039	473.5
SVR	R^2	0.9554	0.971	0.972	1.7	0.9967	0.9954	0.9936	0.3
	RMSE	17.26	15.07	14.60	15.4	0.069	0.0836	0.102	47.8
	MAE	10.79	10.622	11.37	5.4	0.0237	0.0444	0.0488	105.9
ANN	R^2	0.9946	0.994	0.9917	0.3	0.9963	0.9977	0.9954	0.1
	RMSE	6.01	6.82	7.92	31.8	0.073	0.06	0.086	17.8
	MAE	4.625	5.085	5.872	27.0	0.051	0.048	0.061	19.6

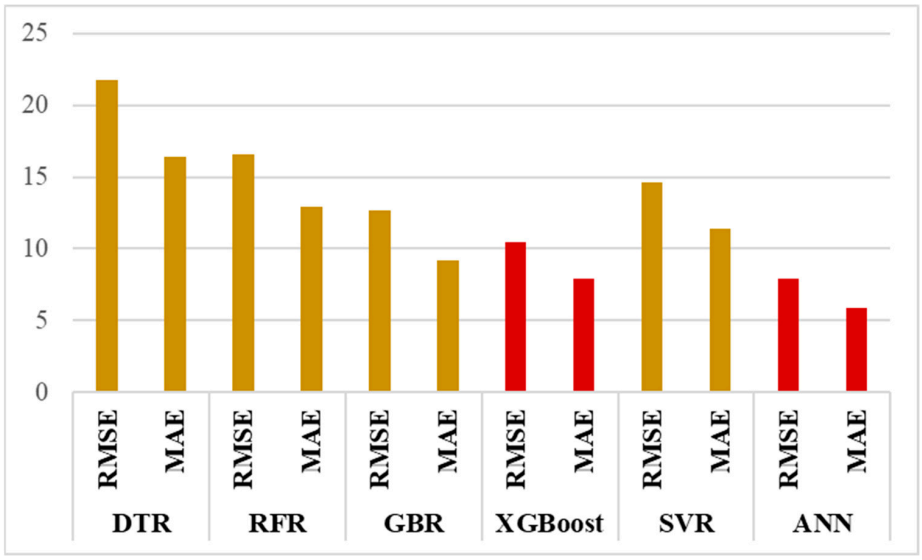

Figure 14. Comparison of RMSE and MAE values in testing sets across various models of σ_{min} predictions.

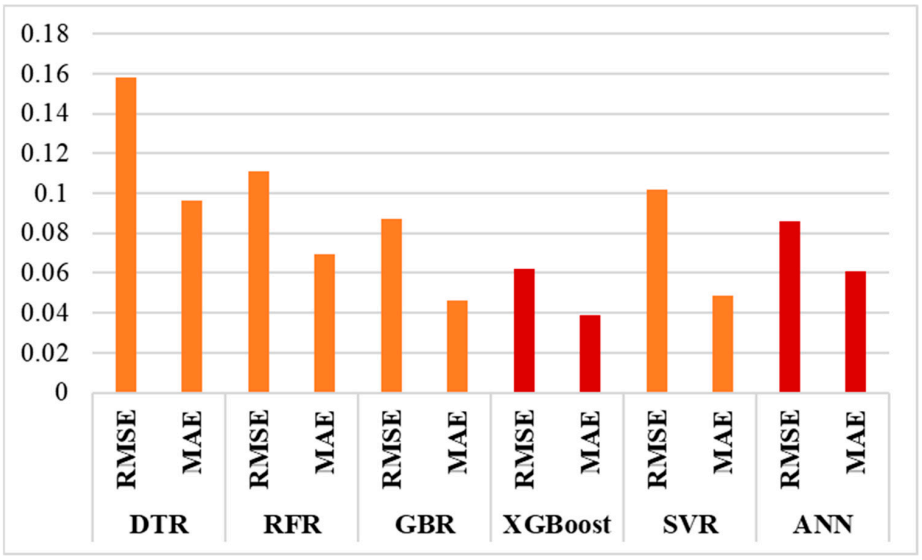

Figure 15. Comparison of RMSE and MAE values in testing sets across various models of γ_{max} predictions.

It should be highlighted that the generalization capacity of an ML model is measured by how well it predicts the testing (unseen) dataset. Metric scores in the testing stage unsurprisingly show a decrease in performance for all models, albeit somewhat, due to the unknown data, when compared to the training stage. To quantify this performance gap, Equation (4) [86] defines the degradation rate, or r_d:

$$r_d = \left| \frac{m_{train} - m_{test}}{m_{train}} \right| \times 100\% \quad (4)$$

where m_{train} and m_{test} are the values of a specific measure throughout the testing and training stages.

All performance measurements reveal that among all models, the ANN produced the least degree of performance degradation. The SVR comes next, although its metrics were not the best. The performance of the other algorithms fell dramatically, especially when it came to the RMSE and MAE for DTR. The ANN produced the most consistent results across both performance phases (training and testing), while the XGBoost and ANN produced the most accurate predictions and highest metrics as shown in Figures 14 and 15.

So, ANN is recommended as the best model as it achieved a high performance with the least degree of performance degradation.

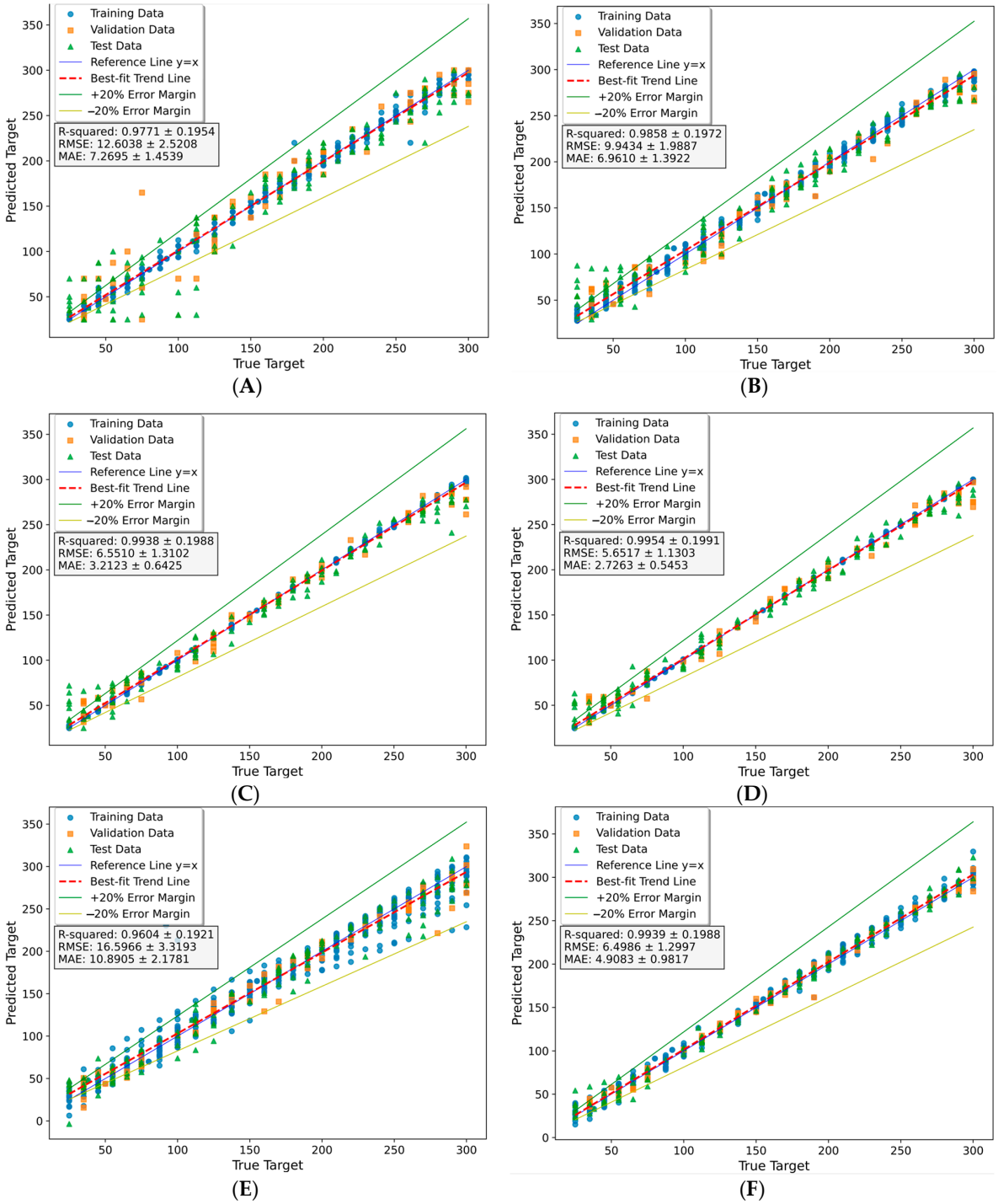

Figure 16. Comparison of actual and predicted targets for σ_{min} models; (**A**) DR; (**B**) RFR; (**C**) GBR; (**D**) XGBoost; (**E**) SVR; (**F**) ANN.

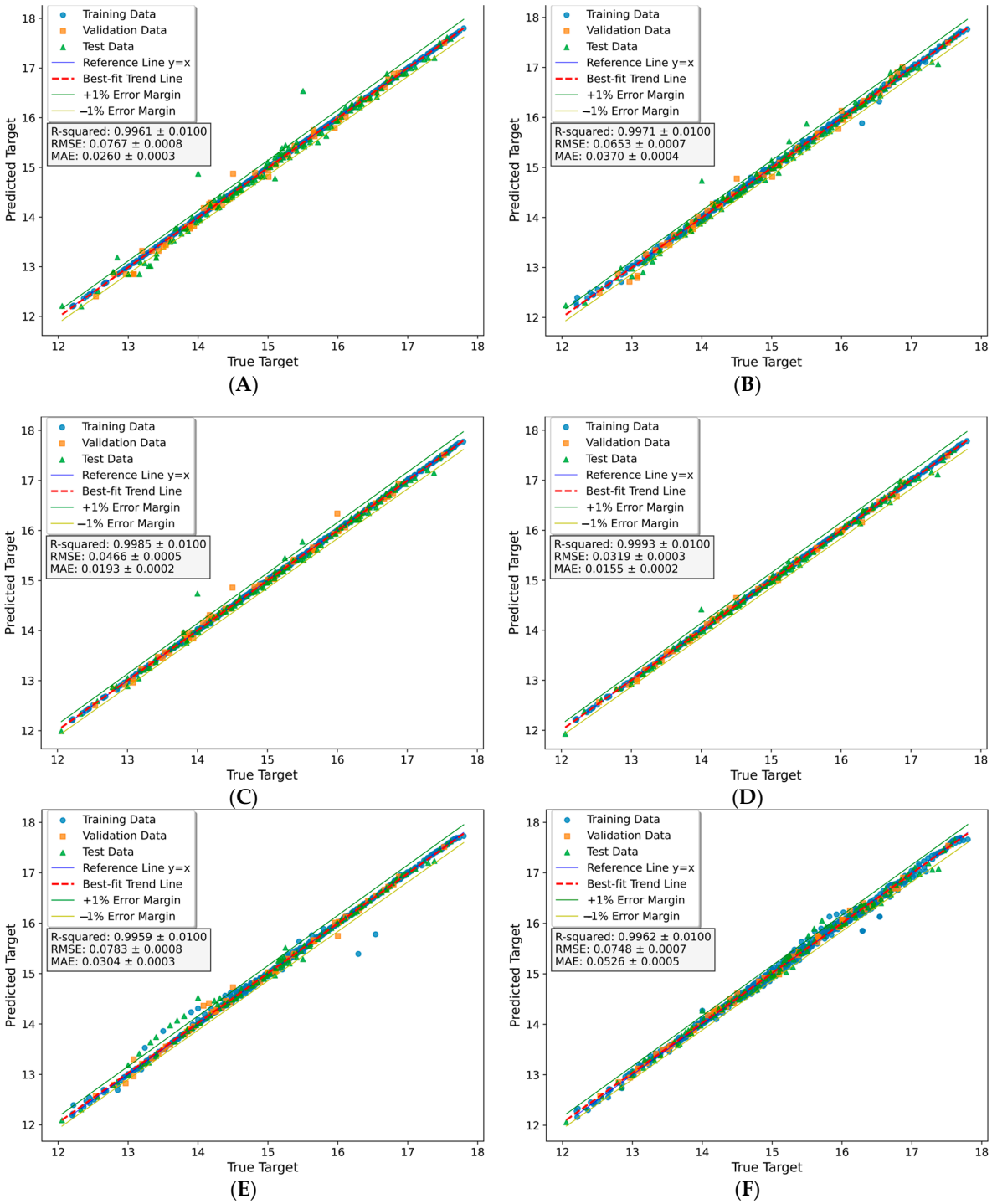

Figure 17. Comparison of actual and predicted targets for γ_{max} models; (**A**) DR; (**B**) RFR; (**C**) GBR; (**D**) XGBoost; (**E**) SVR; (**F**) ANN.

The visualization in Figures 16 and 17 provides a thorough analysis of the models' predictive capabilities of σ_{min} and γ_{max}, respectively. Figures 16 and 17 depict the comparison between actual and predicted values for training, validation, and test sets. In the training set, most of the models demonstrated excellent accuracy, achieving an R-squared value of almost 1 and a low RMSE and MAE. This indicates that it effectively captured the variability in the training data and made minimal prediction errors. However, it is worth noting that some data points notably deviated from the best fit trend line (deal line), indicating limitations of the SVR and DTR models in specific situations.

ANN and XGBoost Models clearly aligned closely with the ideal line, demonstrating their accurate predictions across a range of stress levels. GBR models followed this trend with minimal deviations from the reference line, although there were some exceptions. RFR also showed good alignment overall, with occasional deviations. In contrast, DTR and SVR exhibited more noticeable differences from the ideal line, indicating variations between actual and predicted.

Furthermore, determining an acceptable error margin is critical in determining the training dataset's adequacy. Figures 16 and 17 show a 20% error margin for σ_{min} models and a 1% error margin for γ_{max} models on each cross plot to demonstrate this. The prediction of γ_{max} requires increased precision, with the smallest possible margin of error due to the significant influence even minor changes can have on its behavior, as shown in Figure 17, where most projected values fall within the 1% error margin boundaries. Both the σ_{min} and γ_{max} models performed well, with most predictions falling within the error margin lines during the training, validation, and testing phases. This demonstrates that the training dataset was enough to support the σ_{min} and γ_{max} predictive models using the ANN and XGBoost algorithm proposed in this study.

3.2.2. Models Results

DTR Model Results

Figure 18 provides a comprehensive analysis of the DTR model's predictive performance for σ_{min} and γ_{max} across both training and test datasets. The training set exhibits exceptional accuracy, as indicated by a coefficient of determination (R-squared) of 0.995 and 1 for σ_{min} and γ_{max}, respectively, along with low RMSE and MAE values (5.79 and 3.65 for σ_{min}, 0 for γ_{max}). This suggests robust capturing of training data variance and minimal prediction errors. However, in the test set, performance diminished, with R-squared values of 0.9372 and 0.9845 for σ_{min} and γ_{max}, respectively, and higher RMSE and MAE values (21.79 and 16.414 for σ_{min}, 0.158 and 0.0965 for γ_{max}). This confirms a notable degradation in predictions for unseen data. The combined assessment of metrics and visualization highlights the DTR model's limited ability to align predictions with actual values, particularly evident in the deviation of certain unseen data points from the best-fit trend line in specific contexts.

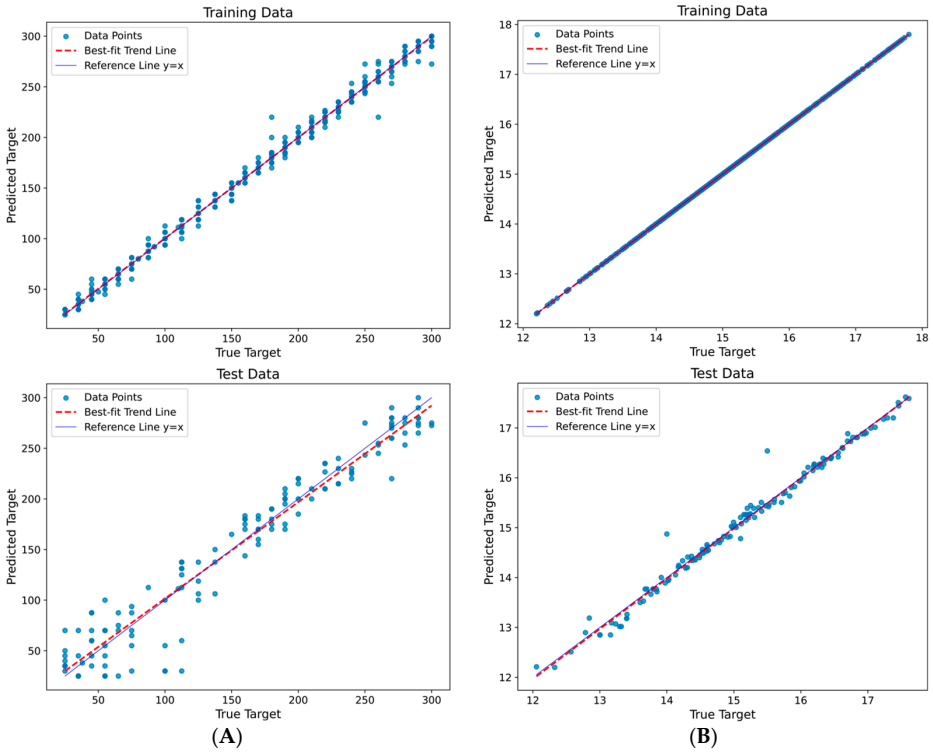

Figure 18. Comparing the actual and predicted values for the training and test datasets in the DTR model for (**A**) σ_{min} and (**B**) γ_{max}.

RFR Model Result

Figure 19 presents a visualization of the RFR model's predictive performance for σ_{min} and γ_{max} across both training and test datasets. The training set showcased exceptional accuracy, with a coefficient of determination (R-squared) of 0.9946 and 0.999 for σ_{min} and γ_{max}, respectively. Additionally, the model exhibited low RMSE and MAE values (5.975 and 4.728 for σ_{min}, 0.0374 and 0.244 for γ_{max}), indicating robust capturing of training data variance and minimal prediction errors, akin to DTR but with a slight advantage favoring DTR.

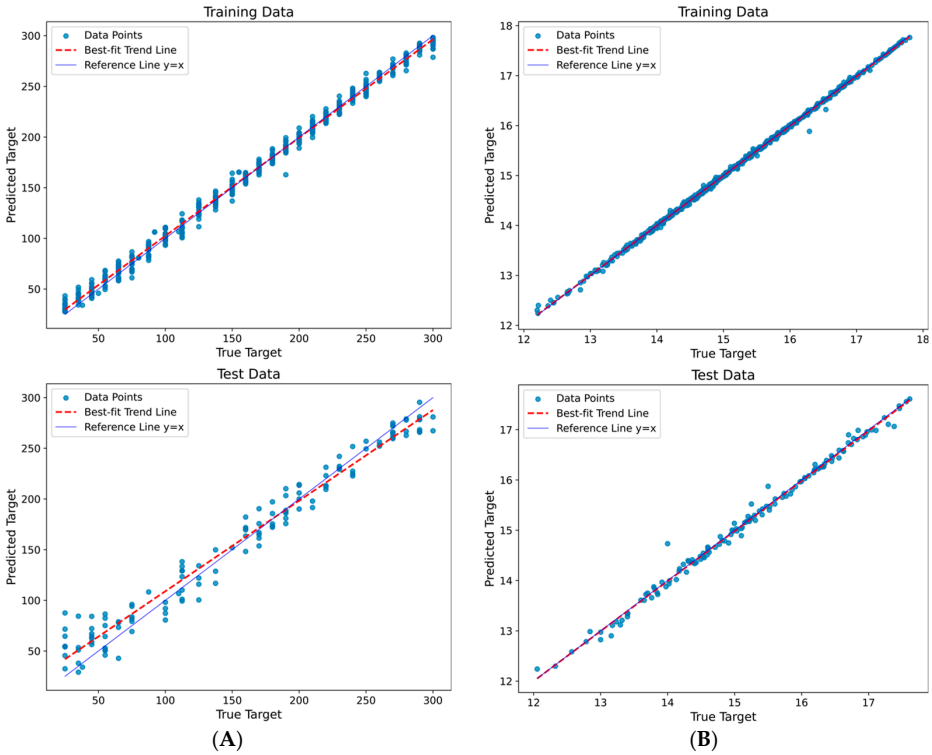

Figure 19. Comparing the actual and predicted values for the training and test datasets in the RFR model for (**A**) σ_{min} and (**B**) γ_{max}.

In the test set, RFR outperformed DTR, with R-squared values of 0.9636 and 0.9924 for σ_{min} and γ_{max}, respectively, along with lower RMSE and MAE values (16.6 and 12.954 for σ_{min}, 0.111 and 0.0696 for γ_{max}). While there was still a degradation in predictions for unseen data, it is notably better than DTR. The combined assessment of metrics and visualization indicates that the RFR model had a strong ability to align predictions with actual values.

Clear evidence suggests that the RFR model surpassed the DTR model, showcasing enhanced predictive capabilities. However, it is essential to note that although the number of data points deviating from the best-fit trend line had significantly been reduced compared to the DTR model, some instances still exist, particularly for lower stress values. These deviations indicate ongoing challenges in specific situations, emphasizing the need to explore other models to address these discrepancies.

GBR Model Results

Figure 20 provides a visualization of the *GBR* model's predictive abilities for σ_{min} and γ_{max} across both training and test datasets. In the training set, the model demonstrated exceptional accuracy with a coefficient of determination (R-squared) of 0.9997 and 0.999 for σ_{min} and γ_{max}, respectively. The low RMSE and MAE values (1.385 and 1.017 for σ_{min}, 0.01 and 0.0081 for γ_{max}) underscored the robust capturing of training data variance and minimal prediction errors, although slightly better than RFR. However, in the test set, GBR

outperformed RFR with R-squared values of 0.9789 and 0.9954 for σ_{min} and γ_{max}, respectively. The RMSE and MAE values (12.64 and 9.172 for σ_{min}, 0.087 and 0.0463 for γ_{max}) indicate a degradation in predictions on unseen data compared to the training set. While GBR's metrics were higher than RFR, its overall performance was considerably better. The reason lies in GBR achieving superior metrics for the training set compared to the RFR model.

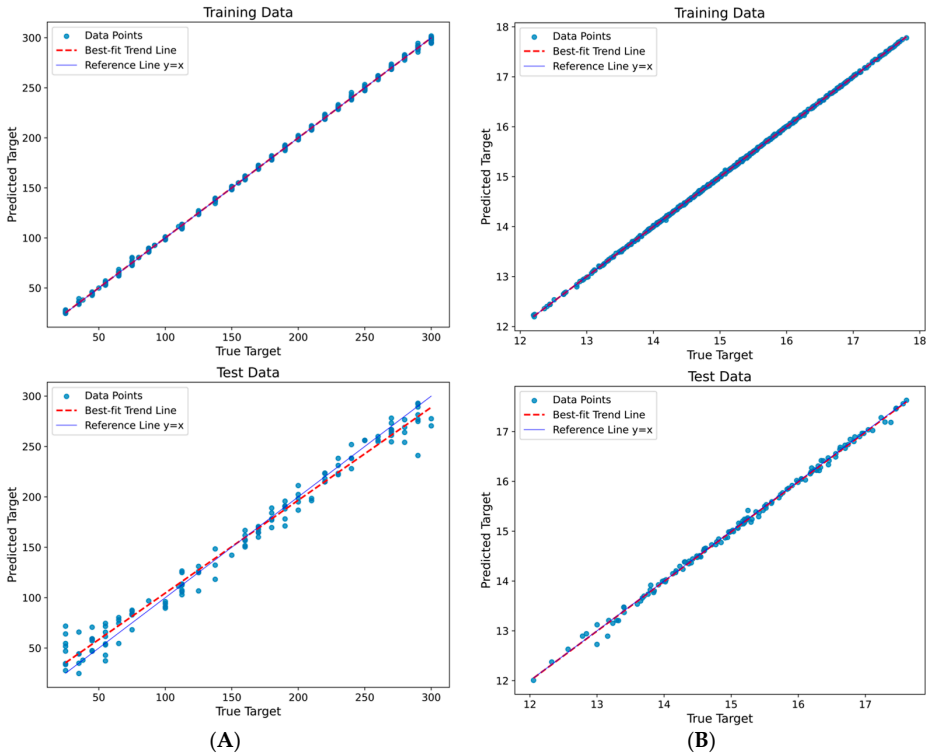

Figure 20. Comparing the actual and predicted values for the training and test datasets in the GBR model for (**A**) σ_{min} and (**B**) γ_{max}.

Metrics and visualization collectively demonstrate the RFR model's ability to match predictions with actual values. Despite a slight increase in prediction errors compared to the training set, the model showcased a notable capacity to generalize predictions on unseen data.

In summary, GBR surpassed RFR in performance, showcasing enhanced predictive capabilities. However, deviations from the ideal line persisted, particularly for low stress values. GBR is highly recommended for dry unit weight predictions due to minimal deviations from the ideal line.

XGBoost Model Results

The analysis presented in Figure 21 delves into the *XGBoost* model's predictive prowess concerning σ_{min} and γ_{max} across both training and test datasets. Notably, the training set showcased exceptional accuracy, boasting a coefficient of determination (R-squared) of 0.9999 for both σ_{min} and γ_{max}. Additionally, the model exhibited low RMSE and MAE values (0.932 and 0.7035 for σ_{min}, 0.01 and 0.0068 for γ_{max}), signifying its adeptness in capturing training data variance with minimal prediction errors. While comparable to GBR, the *XGBoost* model slightly outperformed GBR in the training set.

In the test set, XGBoost outshone GBR, evident in the R-squared values of 0.9856 and 0.9976 for σ_{min} and γ_{max}, respectively. Correspondingly, the RMSE and MAE values (10.43 and 7.92 for σ_{min}, 0.062 and 0.039 for γ_{max}) reinforced XGBoost's proficiency in delivering precise predictions on independent data. Although there was a marginal increase in prediction errors on unseen data compared to RFR, XGBoost's metrics and performance

surpassed GBR. The key differentiator lies in XGBoost achieving higher metrics for the training set than the GBR model.

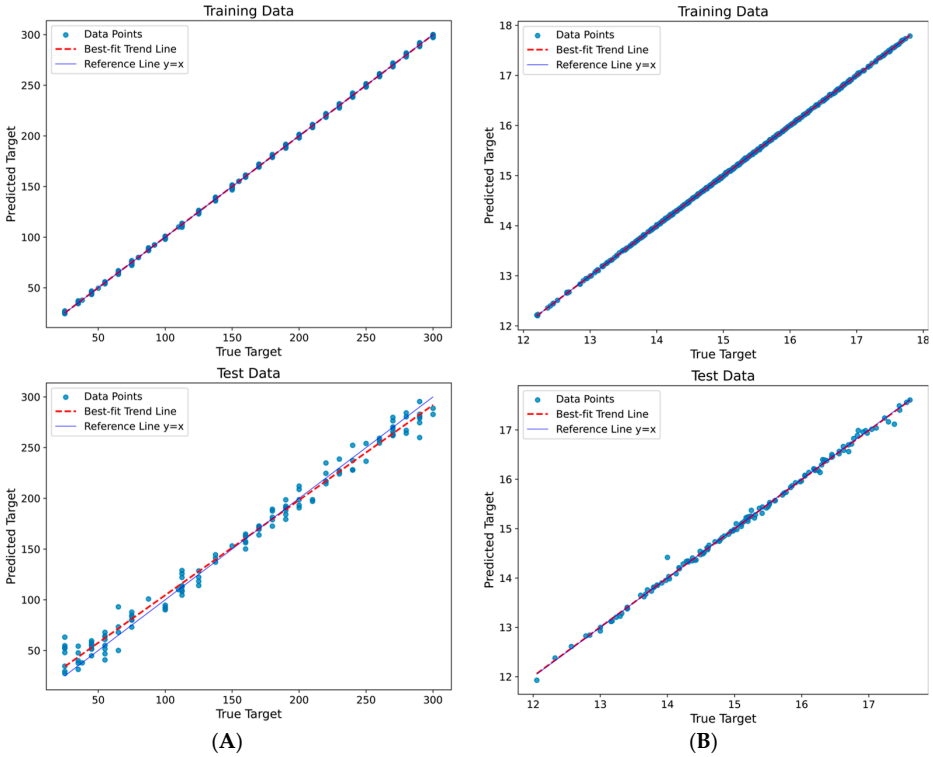

Figure 21. Comparing the actual and predicted values for the training and test datasets in the XGBoost model for (**A**) σ_{min} and (**B**) γ_{max}.

In summary, XGBoost unequivocally outperformed the DTR, RFR, and GBR models in terms of predictive accuracy and closely aligned with the ideal line. Despite minor deviations from the ideal line among certain data points, these discrepancies remained within acceptable bounds. Crucially, these instances hold significance, particularly in predictions at low stress levels, where the predicted stress occasionally exceeds the actual value. This serves as a safety measure, preventing swelling and thereby enhancing the model's overall predictive utility and practical applicability.

SVR Model Results

Figure 22 offers an analysis of the SVR model's predictive performance for σ_{min} and γ_{max} across both training and test datasets. The training set exhibited a coefficient of determination (R-squared) of 0.9554 for σ_{min} and 0.9967 for γ_{max}, along with an RMSE of 17.26 for σ_{min} and 0.069 for γ_{max}, and an MAE of 10.79 for σ_{min} and 0.0237 for γ_{max}. Despite being less accurate than previous models in the training set, SVR demonstrated acceptable predictive capabilities. In the test set, SVR performed slightly better than DTR and nearly matched RFR's performance, evident in R-squared values of 0.972 for σ_{min} and 0.9936 for γ_{max}. The RMSE and MAE values (14.6 and 11.37 for σ_{min}, 0.102 and 0.0488 for γ_{max}) confirmed SVR's effectiveness in providing reasonably accurate predictions on independent data. However, SVR exhibited some discrepancies from the ideal line, highlighting its limitations compared to RFR, GBR, and XGBR models. While SVR can deliver accurate predictions, there were instances where data points deviated significantly from the ideal line, akin to the limitations observed in the DTR model.

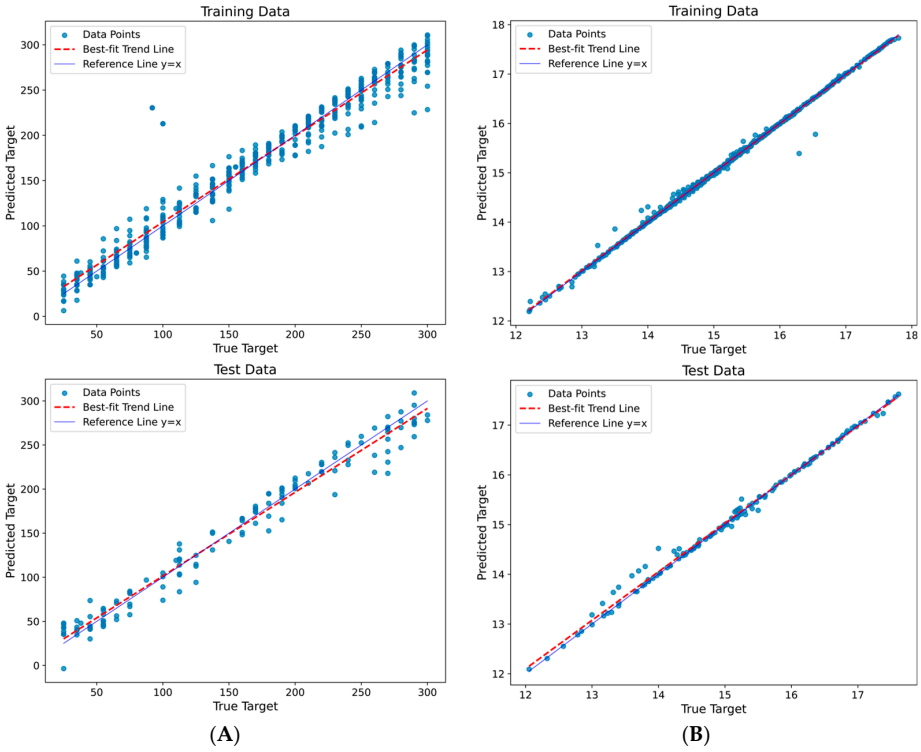

Figure 22. Comparing the actual and predicted values for the training and test datasets in the SVR model for (**A**) σ_{min} and (**B**) γ_{max}.

In direct comparison to RFR, GBR, and XGBR models, SVR fell short in its predictive ability. Despite providing reasonably accurate predictions, the noticeable deviations from the ideal line suggest caution in using SVR for stress and dry unit weight predictions. Given its relative inferiority, it is advisable to prioritize RFR, GBR, and XGBR models for more accurate and reliable predictions in applied stress and dry unit weight scenarios.

ANN Model Results

Figure 23 provides an examination of the ANN model's predictive capacity for σ_{min} and γ_{max} across both training and test datasets. The training set exhibited a coefficient of determination (R-squared) of 0.9946 for σ_{min} and 0.9963 for γ_{max}, along with an RMSE of 6.01 for σ_{min} and 0.073 for γ_{max}, and a low MAE of 4.625 for σ_{min} and 0.051 for γ_{max}. Although displaying slightly less accuracy than its predecessors in the training set, the ANN model demonstrated notable predictive capabilities. In the test set, the performance of the ANN model stands out, ranking among the top models and closely approaching the performance of XGBoost and GBR models. The ANN model achieved R-squared values of 0.9917 for σ_{min} and 0.9954 for γ_{max}, with an RMSE of 7.92 for σ_{min} and 0.086 for γ_{max}, along with an MAE of 5.872 for σ_{min} and 0.061 for γ_{max}. These results underscore the proficiency of the ANN model in delivering reliable predictions for unseen (test) data instances. A comparative analysis with previous models revealed that the ANN model holds a favorable position, offering near-perfect accuracy and minimal degradation, as evidenced in Table 5. This high level of performance positions the ANN Model as a valuable tool for predicting applied stress and dry unit weight.

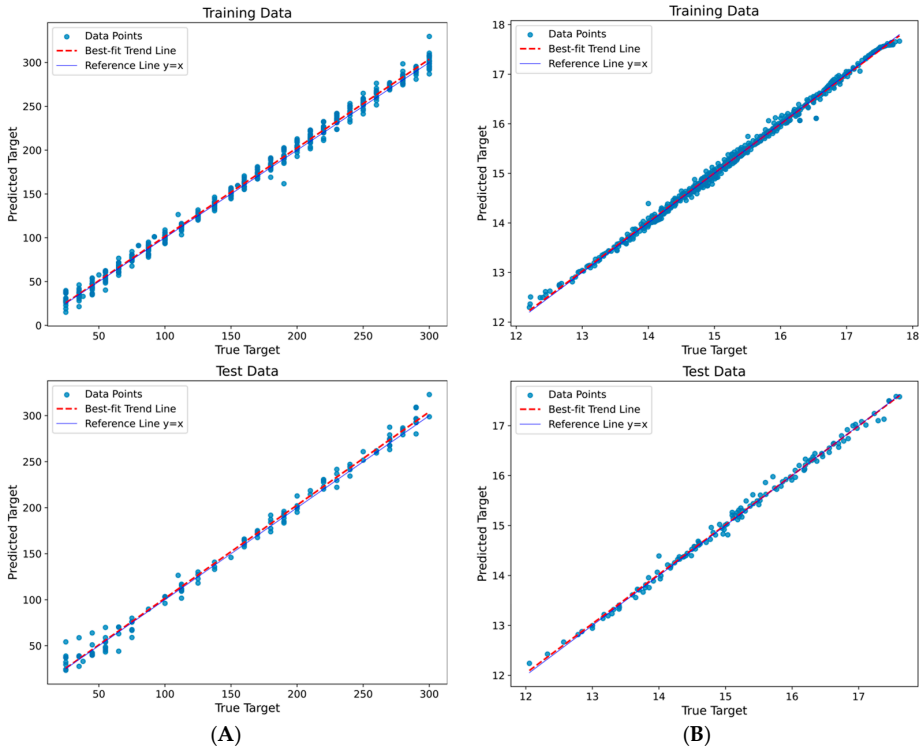

Figure 23. Comparing the actual and predicted values for the training and test datasets in the ANN model for (**A**) σ_{min} and (**B**) γ_{max}.

3.2.3. Streamlined Interface for ANN Model Predictions

The introduction of a user-friendly interface complements this approach by offering a simplified, one-click prediction process. This user-friendly interface ensures that users, regardless of their technical proficiency, can effortlessly harness the predictive capabilities of the ANN model. For example, to utilize the ANN Optimal Dry Unit Weight Model, one can open the 'ANN Gama model.ipynb' Python file using the Jupyter Notebook app (see Supplementary Materials). With a simple press of Shift + Enter, the interface swiftly materializes, providing a user-friendly experience as shown in Figure 24. This same process seamlessly applies to the ANN Optimal Applied Stress model, ensuring a consistent and accessible user interaction.

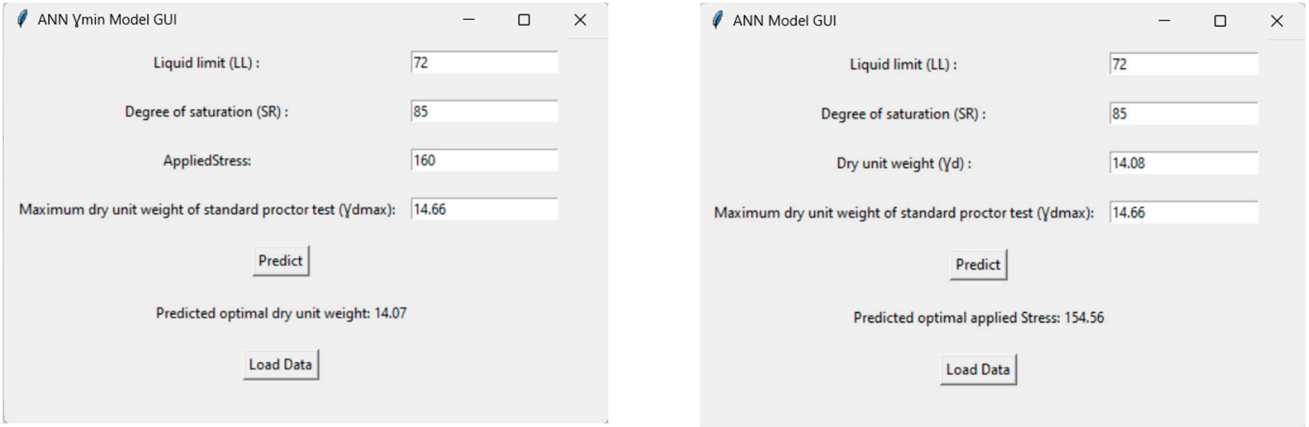

Figure 24. User-friendly interface.

4. Conclusions

This study delved into the complex behaviors of expansive soils, emphasizing key factors such as initial dry unit weight, initial degree of saturation, sand content, and applied stress in influencing soil swelling amplitude. A cost-effective methodology was proposed, leveraging stress application based on soil engineering properties, demonstrating a substantial reduction in swelling amplitude and offering an effective approach to mitigate associated risks.

The introduction of an innovative machine learning-based methodology has enabled the prediction of optimal dry unit weight and stress levels for controlling soil swelling. Among the various models, XGBoost and ANN have emerged as frontrunners, showcasing exceptional performance with the highest R-squared values: 0.9856 and 0.9917 for σ_{min} predictions, and 0.9976 and 0.9954 for γ_{max} predictions. These models have also demonstrated the lowest RMSE (10.43 and 7.92 for σ_{min}, 0.062 and 0.086 for γ_{max}) and the lowest MAE (7.92 and 5.872 for σ_{min}, 0.039 and 0.0488 for γ_{max}), respectively.

Importantly, the ANN model exhibits the least degree of performance degradation, underlining its robustness in providing reliable predictions. This proposed methodology holds significant potential to advance geotechnical engineering practices, empowering informed decision-making in construction projects involving expansive soils and minimizing potential damage.

This study's limitations arise from depending on a specific dataset tailored to a particular range for each parameter, as indicated in Table 3. Additionally, Atterberg limits should be positioned slightly above and near line A on the Casagrande chart. This commitment to achieving accurate predictions may, however, pose constraints on the generalizability of the findings. Future research in this field should focus on expanding the dataset and refining risk mitigation strategies related to expansive soils, thereby contributing to the continual advancement of geotechnical engineering practices. These enhancements aim to further solidify the applicability and effectiveness of the proposed approach in real-world scenarios. In conclusion, while the current study provides valuable insights, addressing these limitations through broader datasets and refined strategies will be crucial for advancing the reliability and robustness of predictive models in geotechnical engineering.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app14041411/s1.

Author Contributions: Conceptualization, A.A., M.O.A. and R.R.; investigation, A.A. and M.O.A.; experimental work, AA.; writing—original draft preparation, A.A.; modeling, A.A. and R.R.; writing—review and editing, R.R., M.O.A., and A.A.; supervision, R.R. and M.O.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. Funding for open access granted by Szechenyi István University (SZE).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Stoll, S.C.; Henning, S.R.; Bagley, A.D.; Wieghaus, K.T. Foundation Damage Assessments and Structural Repairs. In *Forensic Engineering*; American Society of Civil Engineers: Denver, CO, USA, 2022; pp. 166–174. ISBN 9780784484548.
2. Fredlund, D.G.; Rahardjo, H. *Soil Mechanics for Unsaturated Soils*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1993; ISBN 9780470172759.
3. Mokhtari, M.; Dehghani, M. Swell-Shrink Behavior of Expansive Soils, Damage and Control. *Electron. J. Geotech. Eng.* **2012**, *17*, 2673–2682.

4. Federal Highway Administration. Chapter 4: Soil and Rock Behavior. In *A Quarter Century of Geotechnical Research*; FHWA-RD-98-139; Federal Highway Administration: Washinton, DC, USA, 1999.
5. Sawangsuriya, A.; Jotisankasa, A.; Anuvechsirikiat, S. Classification of Shrinkage and Swelling Potential of a Subgrade Soil in Central Thailand. In *Unsaturated Soils: Research and Applications*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 325–331.
6. Márta, F. Development of the Classification of High Swelling Clay Content Soils of Hungary Based on Diagnostic Approach. Ph.D. Thesis, Szent István University, Gödöllő, Hungary, 2012.
7. Teodosio, B.; Kristombu Baduge, K.S.; Mendis, P. A Review and Comparison of Design Methods for Raft Substructures on Expansive Soils. *J. Build. Eng.* **2021**, *41*, 102737. [CrossRef]
8. Ijaz, N.; Ye, W.; ur Rehman, Z.; Dai, F.; Ijaz, Z. Numerical Study on Stability of Lignosulphonate-Based Stabilized Surficial Layer of Unsaturated Expansive Soil Slope Considering Hydro-Mechanical Effect. *Transp. Geotech.* **2022**, *32*, 100697. [CrossRef]
9. Steinberg, M.L. *Controlling Expansive Soil Destructiveness by Deep Vertical Geomembranes on Four Highways*; Transportation Research Board: Washinton, DC, USA, 1985; ISBN 0309039231.
10. Goodarzi, A.R.; Akbari, H.R.; Salimi, M. Enhanced Stabilization of Highly Expansive Clays by Mixing Cement and Silica Fume. *Appl. Clay Sci.* **2016**, *132–133*, 675–684. [CrossRef]
11. Kolay, P.K.; Ramesh, K.C. Reduction of Expansive Index, Swelling and Compression Behavior of Kaolinite and Bentonite Clay with Sand and Class C Fly Ash. *Geotech. Geol. Eng.* **2016**, *34*, 87–101. [CrossRef]
12. Salimi, M.; Ilkhani, M.; Vakili, A.H. Stabilization Treatment of Na-Montmorillonite with Binary Mixtures of Lime and Steelmaking Slag. *Int. J. Geotech. Eng.* **2020**, *14*, 295–301. [CrossRef]
13. Nelson, J.; Miller, D.J. *Expansive Soils: Problems and Practice in Foundation and Pavement Engineering*; John Wiley & Sons: Hoboken, NJ, USA, 1997.
14. Roy, T.K. Influence of Sand on Strength Characteristics of Cohesive Soil for Using as Subgrade of Road. *Procedia-Soc. Behav. Sci.* **2013**, *104*, 218–224. [CrossRef]
15. Mohamed, K.; Abdelkrim, M.; Lakhdar, M. Problematic Soil Mechanics in the Algerian Arid and Semi-Arid Regions: Case of M'sila Expansive Clays. *J. Appl. Eng. Sci. Technol.* **2015**, *1*, 37–41.
16. Al Rawi, O.S.; Assaf, M.N.; Hussein, N.M. Effect of Sand Additives on the Engineering Properties of Fine Grained Soils. *ARPN J. Eng. Appl. Sci.* **2018**, *13*, 3197–3206.
17. Phanikumar, B.R.; Dembla, S.; Yatindra, A. Swelling Behaviour of an Expansive Clay Blended with Fine Sand and Fly Ash. *Geotech. Geol. Eng.* **2021**, *39*, 583–591. [CrossRef]
18. Alnmr, A.; Ray, R.P. Review of the Effect of Sand on the Behavior of Expansive Clayey Soils. *Acta Tech. Jaurinensis* **2021**, *14*, 521–552. [CrossRef]
19. Lamara, M.; Gueddouda, M.K.; Benabed, B. Stabilisation Physico-Chimique Des Sols Gonflants (Sable de Dune + Sel). *Rev. Française Géotechn.* **2006**, *115*, 25–35. [CrossRef]
20. Prasad, C.R.V.; Sharma, R.K. Influence of Sand and Fly Ash on Clayey Soil Stabilization. *IOSR J. Mech. Civ. Eng.* **2014**, *334*, 36–40.
21. Nagaraj, H.B. Influence of Gradation and Proportion of Sand on Stress–Strain Behavior of Clay–Sand Mixtures. *Int. J. Geo-Eng.* **2016**, *7*, 19. [CrossRef]
22. Srikanth, V.; Mishra, A.K. Atterberg Limits of Sand-Bentonite Mixes and the Influence of Sand Composition. In *Geotechnical Characterisation and Geoenvironmental Engineering*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 139–145. [CrossRef]
23. Dasog, G.S.; Mermut, A.R. Expansive Soils and Clays. *Encycl. Earth Sci. Ser.* **2013**, 297–300. [CrossRef]
24. Jones, L.D.; Jefferson, I. Expansive Soils. In *ICE Manual of Geotechnical Engineering. Volume 1: Geotechnical Engineering Principles, Problematic Soils and Site Investigation*; Burland, J., Ed.; ICE Manual of Geotechnical Engineering: London, UK, 2012.
25. Alnmr, A.; Ray, R. Numerical Simulation of Replacement Method to Improve Unsaturated Expansive Soil. *Pollack Period.* **2023**, *18*, 41–47. [CrossRef]
26. Dawson, R.F.; Altmeyer, W.T.; Barber, E.S.; DuBose, L.A. Discussion of "Engineering Properties of Expansive Clays". *Trans. Am. Soc. Civ. Eng.* **1956**, *121*, 664–675. [CrossRef]
27. Seed, H.B.; Woodward, R.J.; Lundgren, R. Prediction of Swelling Potential for Compacted Clays. *Trans. Am. Soc. Civ. Eng.* **1963**, *128*, 1443–1477. [CrossRef]
28. Ranganatham, B.V.; Satyanarayana, B. A Rational Method of Predicting Swelling Potential for Compacted Expansive Clays. In Proceedings of the 6th International Conference on Soil Mechanics and Foundation Engineering, Montréal, QC, Canada, 8–15 September 1965; pp. 92–96.
29. Snethen, D.R. Evaluation of Expedient Methods for Identification and Classification of Potentially Expansive Soils. In Proceedings of the Fifth International Conference on Expansive Soils 1984, Adelaide, Australia, 21–23 May 1984; pp. 22–26.
30. Al-Shayea, N.A. The Combined Effect of Clay and Moisture Content on the Behavior of Remolded Unsaturated Soils. *Eng. Geol.* **2001**, *62*, 319–342. [CrossRef]
31. Yilmaz, I. Indirect Estimation of the Swelling Percent and a New Classification of Soils Depending on Liquid Limit and Cation Exchange Capacity. *Eng. Geol.* **2006**, *85*, 295–301. [CrossRef]
32. Çimen, Ö.; Keskin, S.N.; Yıldırım, H. Prediction of Swelling Potential and Pressure in Compacted Clay. *Arab. J. Sci. Eng.* **2012**, *37*, 1535–1546. [CrossRef]
33. Ling, Q.; Zhang, Q.; Wei, Y.; Kong, L.; Zhu, L. Slope Reliability Evaluation Based on Multi-Objective Grey Wolf Optimization-Multi-Kernel-Based Extreme Learning Machine Agent Model. *Bull. Eng. Geol. Environ.* **2021**, *80*, 2011–2024. [CrossRef]

34. Liu, L.; Zhang, S.; Cheng, Y.M.; Liang, L. Advanced Reliability Analysis of Slopes in Spatially Variable Soils Using Multivariate Adaptive Regression Splines. *Geosci. Front.* **2019**, *10*, 671–682. [CrossRef]
35. Wang, H.; Zhang, L.; Yin, K.; Luo, H.; Li, J. Landslide Identification Using Machine Learning. *Geosci. Front.* **2021**, *12*, 351–364. [CrossRef]
36. Ray, R.; Kumar, D.; Samui, P.; Roy, L.B.; Goh, A.T.C.; Zhang, W. Application of Soft Computing Techniques for Shallow Foundation Reliability in Geotechnical Engineering. *Geosci. Front.* **2021**, *12*, 375–383. [CrossRef]
37. Wang, L.; Wu, C.; Tang, L.; Zhang, W.; Lacasse, S.; Liu, H.; Gao, L. Efficient Reliability Analysis of Earth Dam Slope Stability Using Extreme Gradient Boosting Method. *Acta Geotech.* **2020**, *15*, 3135–3150. [CrossRef]
38. Merouane, F.Z.; Mamoune, S.M.A. Prediction of Swelling Parameters of Two Clayey Soils from Algeria Using Artificial Neural Networks. *Math. Model. Civ. Eng.* **2018**, *14*, 11–26. [CrossRef]
39. Dutta, R.K.; Singh, A.; Gnanandarao, T. Prediction of Free Swell Index for the Expansive Soil Using Artificial Neural Networks. *J. Soft Comput. Civ. Eng.* **2019**, *3*, 47–62. [CrossRef]
40. Cho, S.E. Probabilistic Stability Analyses of Slopes Using the ANN-Based Response Surface. *Comput. Geotech.* **2009**, *36*, 787–797. [CrossRef]
41. Li, S.; Zhao, H.B.; Ru, Z. Slope Reliability Analysis by Updated Support Vector Machine and Monte Carlo Simulation. *Nat. Hazards* **2013**, *65*, 707–722. [CrossRef]
42. Li, T.Z.; Pan, Q.; Dias, D. Active Learning Relevant Vector Machine for Reliability Analysis. *Appl. Math. Model.* **2021**, *89*, 381–399. [CrossRef]
43. Kardani, N.; Aminpour, M.; Nouman Amjad Raja, M.; Kumar, G.; Bardhan, A.; Nazem, M. Prediction of the Resilient Modulus of Compacted Subgrade Soils Using Ensemble Machine Learning Methods. *Transp. Geotech.* **2022**, *36*, 100827. [CrossRef]
44. Yi, P.; Wei, K.; Kong, X.; Zhu, Z. Cumulative PSO-Kriging Model for Slope Reliability Analysis. *Probabilistic Eng. Mech.* **2015**, *39*, 39–45. [CrossRef]
45. Kumar, M.; Samui, P. Reliability Analysis of Pile Foundation Using ELM and MARS. *Geotech. Geol. Eng.* **2019**, *37*, 3447–3457. [CrossRef]
46. Shen, H.; Li, J.; Wang, S.; Xie, Z. Prediction of Load-Displacement Performance of Grouted Anchors in Weathered Granites Using FastICA-MARS as a Novel Model. *Geosci. Front.* **2021**, *12*, 415–423. [CrossRef]
47. Zhang, W.; Wu, C.; Tang, L.; Gu, X.; Wang, L. Efficient Time-Variant Reliability Analysis of Bazimen Landslide in the Three Gorges Reservoir Area Using XGBoost and LightGBM Algorithms. *Gondwana Res.* **2022**, *123*, 41–53. [CrossRef]
48. Najjar, Y.M.; Basheer, I.A.; Mcreynolds, R. Neural Modeling of Kansas Soil Swelling. *Transp. Res. Rec. J. Transp. Res. Board.* **1996**, *1526*, 14–19. [CrossRef]
49. Najjar, Y.M.; Basheer, I.A. *Modeling of Soil Swelling via Regression and Neural Network Approaches*; Kansas Department of Transportation: Topeka, KS, USA, 1998.
50. Doris, J.J.; Rizzo, D.M.; Dewoolkar, M.M. Forecasting Vertical Ground Surface Movement from Shrinking/Swelling Soils with Artificial Neural Networks. *Int. J. Numer. Anal. Methods Geomech.* **2008**, *32*, 1229–1245. [CrossRef]
51. Ashayeri, I.; Yasrebi, S. Free-Swell and Swelling Pressure of Unsaturated Compacted Clays; Experiments and Neural Networks Modeling. *Geotech. Geol. Eng.* **2009**, *27*, 137–153. [CrossRef]
52. Mamoune, S.M.A. Characterization and Modelling of the Clays of Tlemcen Using Neural Networks. Ph.D. Thesis, University Abou Bakr Belkaid, Tlemcen, Algeria, 2009.
53. Ikizler, S.B.; Aytekin, M.; Vekli, M.; Kocabaş, F. Prediction of Swelling Pressures of Expansive Soils Using Artificial Neural Networks. *Adv. Eng. Softw.* **2010**, *41*, 647–655. [CrossRef]
54. Erzin, Y.; Güneş, N. The Prediction of Swell Percent and Swell Pressure by Using Neural Networks. *Math. Comput. Appl.* **2011**, *16*, 425–436. [CrossRef]
55. Ikeagwuani, C.C. Estimation of Modified Expansive Soil CBR with Multivariate Adaptive Regression Splines, Random Forest and Gradient Boosting Machine. *Innov. Infrastruct. Solut.* **2021**, *6*, 199. [CrossRef]
56. Eyo, E.U.; Abbey, S.J.; Lawrence, T.T.; Tetteh, F.K. Improved Prediction of Clay Soil Expansion Using Machine Learning Algorithms and Meta-Heuristic Dichotomous Ensemble Classifiers. *Geosci. Front.* **2022**, *13*, 101296. [CrossRef]
57. Amanabadi, S.; Vazirinia, M.; Vereecken, H.; Vakilian, K.A.; Mohammadi, M.H. Comparative Study of Statistical, Numerical and Machine Learning-Based Pedotransfer Functions of Water Retention Curve with Particle Size Distribution Data. *Eurasian Soil. Sci.* **2019**, *52*, 1555–1571. [CrossRef]
58. Bachir, R.; Mohammed, A.M.S.; Habib, T. Using Artificial Neural Networks Approach to Estimate Compressive Strength for Rubberized Concrete. *Period. Polytech. Civ. Eng.* **2018**, *62*, 858–865. [CrossRef]
59. *ASTM D6913/D6913M-17*; Standard Test Method for Particle-Size Analysis of Soils. ASTM International: Conshohocken, PA, USA, 2017. [CrossRef]
60. *ASTM D7928-17*; Standard Test Method for Particle-Size Distribution (Gradation) of Fine-Grained Soils Using the Sedimentation (Hydrometer) Analysis. ASTM International: Conshohocken, PA, USA, 2017. [CrossRef]
61. *D854-14*; Standard Test Methods for Specifc Gravity of Soil Solids by Water Pycnometer. ASTM International: Conshohocken, PA, USA, 2014. [CrossRef]
62. *ASTM D4318-17e1*; Standard Test Methods for Liquid Limit, Plastic Limit, and Plasticity Index of Soils. ASTM International: Conshohocken, PA, USA, 2017. [CrossRef]

63. Atemimi, Y.K. Effect of the Grain Size of Sand on Expansive Soil. In *Proceedings of the Key Engineering Materials*; Trans Tech Publications Ltd: Wollerau, Switzerland, 2020; Volume 857, pp. 367–373.
64. AASHTO. AASHTO Standard Method of Test for The Classification of Soils and SoilAggregate Mixtures for Highway Construction Purposes, Test Designation M145-91. In *Standard Specifications for Transportation Materials and Methods of Sampling and Testing*; AASHTO: Washington, DC, USA, 2002.
65. *ASTM D2487-17e1*; Standard Practice for Classification of Soils for Engineering Purposes (Unified Soil Classification System). ASTM International: Conshohocken, PA, USA, 2017. [CrossRef]
66. Raman, V. Identification of Expansive Soils from the Plasticity Index and the Shrinkage Index Data. *Indian. Eng. Calcutta* **1967**, *11*, 17–22.
67. Sowers, G.F.; Kennedy, C.M. High Volume Change Clays of the South-Eastern Coastal Plain. In Proceedings of the Third Panamerican Conference on Soil Mechanics and Foundation Engineering, Caracas, Venezuela, 1967; pp. 99–120.
68. Dakshanamurthy, V.; Raman, V. A Simple Method of Identifying an Expansive Soil. *Soils Found.* **1973**, *13*, 97–104. [CrossRef]
69. Prakash, K.; Sridharan, A. Free Swell Ratio and Clay Mineralogy of Fine-Grained Soils. *Geotech. Test. J.* **2004**, *27*, 220–225. [CrossRef]
70. Mallikarjuna Rao, K.; Subba Rao, G.V.R. Influence of Coarse Fraction on Characteristics of Expansive Soil–Sand Mixtures. *Int. J. Geosynth. Ground Eng.* **2018**, *4*, 19. [CrossRef]
71. Alnmr, A.; Ray, R. Investigating the Impact of Varying Sand Content on the Physical Characteristics of Expansive Clay Soils from Syria. *Geotech. Geol. Eng.* **2023**, 1–17. [CrossRef]
72. Tso, G.K.F.; Yau, K.K.W. Predicting Electricity Energy Consumption: A Comparison of Regression Analysis, Decision Tree and Neural Networks. *Energy* **2007**, *32*, 1761–1768. [CrossRef]
73. Rodriguez-Galiano, V.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M. Machine Learning Predictive Models for Mineral Prospectivity: An Evaluation of Neural Networks, Random Forest, Regression Trees and Support Vector Machines. *Ore Geol. Rev.* **2015**, *71*, 804–818. [CrossRef]
74. Zeini, H.A.; Al-Jeznawi, D.; Imran, H.; Bernardo, L.F.A.; Al-Khafaji, Z.; Ostrowski, K.A.; Kazmi, S.; Zeini, H.A.; Al-Jeznawi, D.; Imran, H.; et al. Random Forest Algorithm for the Strength Prediction of Geopolymer Stabilized Clayey Soil. *Sustainability* **2023**, *15*, 1408. [CrossRef]
75. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
76. Bardhan, A.; Kardani, N.; Alzo'ubi, A.K.; Roy, B.; Samui, P.; Gandomi, A.H. Novel Integration of Extreme Learning Machine and Improved Harris Hawks Optimization with Particle Swarm Optimization-Based Mutation for Predicting Soil Consolidation Parameter. *J. Rock. Mech. Geotech. Eng.* **2022**, *14*, 1588–1608. [CrossRef]
77. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
78. Lai, V.; Ahmed, A.N.; Malek, M.A.; Afan, H.A.; Ibrahim, R.K.; El-Shafie, A.; El-Shafie, A. Modeling the Nonlinearity of Sea Level Oscillations in the Malaysian Coastal Areas Using Machine Learning Algorithms. *Sustainability* **2019**, *11*, 4643. [CrossRef]
79. Wu, J.; Liu, H.; Wei, G.; Song, T.; Zhang, C.; Zhou, H. Flash Flood Forecasting Using Support Vector Regression Model in a Small Mountainous Catchment. *Water* **2019**, *11*, 1327. [CrossRef]
80. Ibrahem Ahmed Osman, A.; Najah Ahmed, A.; Chow, M.F.; Feng Huang, Y.; El-Shafie, A. Extreme Gradient Boosting (Xgboost) Model to Predict the Groundwater Levels in Selangor Malaysia. *Ain Shams Eng. J.* **2021**, *12*, 1545–1556. [CrossRef]
81. Ali, I.; Alharbi, O.M.L.; Alothman, Z.A.; Badjah, A.Y.; Alwarthan, A.; Basheer, A.A. Artificial Neural Network Modelling of Amido Black Dye Sorption on Iron Composite Nano Material: Kinetics and Thermodynamics Studies. *J. Mol. Liq.* **2018**, *250*, 1–8. [CrossRef]
82. Zou, W.L.; Han, Z.; Ye, J. bing Influence of External Stress and Initial Density on the Volumetric Behavior of an Expansive Clay during Wetting. *Environ. Earth Sci.* **2020**, *79*, 211. [CrossRef]
83. Rosenbalm, D.C.; Zapata, C.E.; Houston, S.L.; Kavazanjian, E.; Witczak, M.W. Volume Change Behavior of Expansive Soils Due to Wetting and Drying Cycles. Ph.D. Thesis, Arizona State University, Tempe, AZ, USA, 2013.
84. Lizama, E.; Morales, B.; Somos-Valenzuela, M.; Chen, N.; Liu, M. Understanding Landslide Susceptibility in Northern Chilean Patagonia: A Basin-Scale Study Using Machine Learning and Field Data. *Remote Sens.* **2022**, *14*, 907. [CrossRef]
85. Scikit-Learn Developers Scikit-Learn. Machine Learning in Python. Available online: https://scikit-learn.org/stable/modules/grid_search.html#randomized-parameter-search (accessed on 21 January 2024).
86. Chen, Y.; Xu, Y.; Jamhiri, B.; Wang, L.; Li, T. Predicting Uniaxial Tensile Strength of Expansive Soil with Ensemble Learning Methods. *Comput. Geotech.* **2022**, *150*, 104904. [CrossRef]