Geomechanics and Engineering, *Vol. 40, No. 3 (2025) 151-163*
https://doi.org/10.12989/gae.2025.40.3.151

Data-driven modeling for interfacial behaviors between frozen soil and existing structures for applications of artificial ground freezing

Sangyeong Park[1,2a], Chaemin Hwang[1b], Byeonghyun Hwang[1c] and Hangseok Choi*[1]

[1]*Departmentof Civil, Environmental and Architectural Engineering, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Korea*
[2]*Department of Petroleum Engineering, Texas A&M University, 400 Bizzell Street, College Station, Texas, 77840, USA*

(*Received April 8, 2024, Revised January 13, 2025, Accepted January 20, 2025*)

Abstract. When the artificial ground freezing technique is applied near existing underground structures, adfreezing behavior, characterized by ice bonding between the frozen soil and the existing structures, becomes a critical factor in assessing the stability of these structures. In this study, punch shear test data were employed to evaluate adfreezing behaviors at the frozen soil-structure interface under zero confinement conditions, representing critical states. Since machine learning (ML) algorithms have offered a powerful data-driven predictive modeling in geotechnical engineering, this study discussed the application of ML approaches to broaden the feasibility of the punch shear test for assessing the adfreezing behavior. Four ML algorithms, i.e., support vector regression (SVR), feedforward neural network (FNN), random forest (RF), and extreme gradient boosting (XGB), were adopted to develop predictive models based on the punch shear test results. To ensure optimal model performance, Bayesian optimization and five-fold cross-validation methods were employed to effectively train the ML models and identify the best hyperparameter combinations for each model. The predictive performance of these models was compared using three regression metrics: root-mean- square error (RMSE), mean absolute error (MAE), and determination coefficient (R^2). The models were ranked based on their performance as follows: XGB > RF > FNN > SVR. Among them, the XGB model demonstrated the highest accuracy, with an RMSE of 0.0037, an MAE of 0.0015, and an R^2 of 0.9999. The reliability and interpretability of the XGB model were further enhanced through post-hoc analysis estimating the prediction interval and SHAP values.

Keywords: adfreezing; artificial ground freezing; interfacial behavior; machine learning; post-hoc analysis; punch shear test

1. Introduction

Adfreezing behavior refers to the interaction between frozen soil and structures, characterized by ice bonding on the structure's surface. Artificial ground freezing (AGF) is a widely utilized technique in construction projects, serving as an auxiliary method to achieve waterproofing and enhance ground strength (Andersland 2003, Schmall *et al.* 2006, Son *et al.* 2021, Tan *et al.* 2021, Zhou *et al.* 2022). AGF has gained recognition as a reliable and sustainable technique due to its unique advantages, including applicability to almost all types of geological strata, the uniformity of the frozen soil body, strong cementation with structures, and minimal environmental impact. Among the many design considerations for AGF, adfreezing behavior stands out as a critical factor that can significantly influence structural stability. This aspect becomes particularly important in restoration projects that employ AGF, as outlined in references (Quanbin *et al.* 2018, Wang *et al.* 2018, Alzoubi *et al.* 2020, Wang *et al.* 2020, Tan *et al.* 2023), where the frozen soil body provides support to existing underground structures and withstands earth pressure. Therefore, it is essential to adequately evaluate adfreezing behavior to ensure the sustainable application of AGF.

To assess the adfreezing behavior, a direct shear test, involving the application of normal stress, is conventionally conducted (Wen *et al.* 2016, Sun *et al.* 2021, Pan *et al.* 2022, Zhang *et al.* 2022, Fuping *et al.* 2023). The adfreezing characteristics are influenced by various factors, including soil properties, initial water content, structure roughness, freezing temperature, shear rate, and normal stress (Jin *et al.* 2020, Tang *et al.* 2020, Wang *et al.* 2020). Based on prior information obtained from direct shear tests, several predictive models have been developed to simulate the adfreezing stress–displacement relationship for arbitrary conditions without experiments. For example, He *et al.* (2021) proposed an analytical model based on the disturbed state concept. Xiong *et al.* (2021) constructed a simple nonlinear model using a combination of power and exponential functions. Chen *et al.* (2022) reproduced this relationship using a backpropagation neural network (NN) and bidirectional long short-term memory (LSTM) model. However, it is important to note that the direct shear test cannot directly assess the adfreezing behavior under zero normal stress conditions. In such cases, it indirectly

*Corresponding author, Professor
E-mail: hchoi2@korea.ac.kr
[a]Ph.D.
[b]Ph.D. Candidate
[c]Ph.D. Candidate

ISSN: 2005-307X (Print), 2092-6219 (Online)

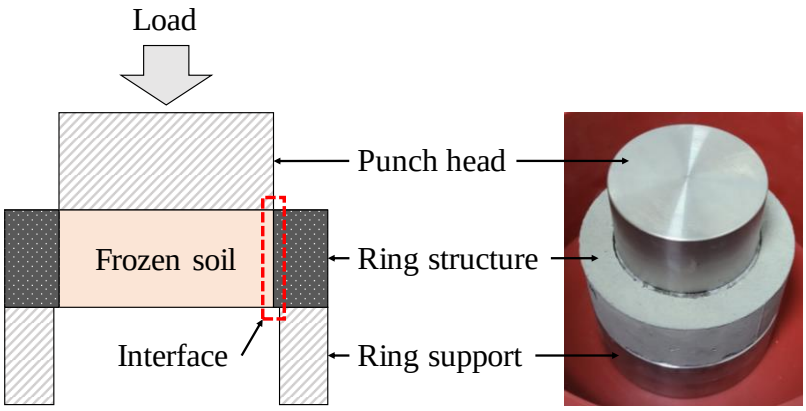


Fig. 1 Configuration of the punch shear test

estimates the unconfined adfreezing behavior by incorporating additional assumptions, such as the Mohr-Coulomb criterion (Ladanyi 1995).

In a previous study (Park *et al.* 2022), the punch shear test was validated as a reliable technique for assessing the adfreezing behavior of frozen soil–structure interfaces under unconfined conditions in the critical state. Comparable to the direct shear test, the punch shear test requires the development of a predictive model to improve the feasibility of the test. However, due to a limited number of experimental cases and ongoing research into the punch shear test, there is insufficient prior information available regarding the adfreezing behavior through this test. Consequently, the application of conventional predictive modeling approaches, such as empirical or analytical methods, is not feasible. As a result, a data-driven approach emerges as the most suitable methodology for the punch shear test.

Machine learning (ML), a robust data-driven methodology, serves as an exceptional and dependable alternative to existing indirect approaches in geotechnical engineering (Bello *et al*. 2015, Fang *et al*. 2018, Lawal and Idris 2020, Pham *et al.* 2022). When equipped with sufficient and high-quality data, ML-based methods can handle complex and nonlinear systems, establishing models without the need for assumptions or prior information (Pham *et al*. 2021, Baghbani *et al*. 2022, Kim *et al.* 2022). Furthermore, ML-based approaches often unveil useful hidden relationships between features, thereby enhancing an understanding of these characteristics beyond existing knowledge.

In the realm of ML models, there often exists a trade-off relationship between accuracy and interpretability, where more intricate models generally achieve higher levels of accuracy. Even when input features from a database are meticulously selected and preprocessed, considering prior information, it can be challenging to explain how the output values are derived from these inputs. This lack of clarity can impede the reliability and interpretability of ML models. In response to this challenge, post-hoc techniques have been developed to explain the predictions made by complex models, thereby enhancing the dependability of ML models (Barredo *et al.* 2020, Linardatos *et al*. 2021). These post-hoc techniques serve to improve the interpretability of black-box models, providing insights into the relationships between input features and predictions, while also assessing the confidence level of these models.

This study aims to address the challenge of predicting adfreezing behaviors under critical conditions using data-driven ML models. Moreover, the post-hoc techniques, including prediction interval estimation and SHAP value analysis, were employed to enhance the understanding of model predictions, thereby improving the reliability and transparency of the ML model.

2. Experimental data

2.1 Punch shear tests

The database used in this study was obtained through a series of the punch shear tests (Park *et al.* 2022), utilizing a setup that included a punch head, frozen soil-ring structure, and ring support (as illustrated in Fig. 1). The rings were made of concrete and stainless steel, and three soil specimens (soil 1, soil 2, and soil 3) were filled into the rings to create the soil-ring assemblage. The concrete consisted of ordinary Portland cement and water, mixed at a ratio of 100 kg to 14 liters. Table 1 provides an overview of the physical properties of these soil specimens. The initial water content for soil 1 and soil 2 was set to be 10, 13, 16, and 19%, while for soil 3, it was 54, 59, 64, and 67%. Following the completion of specimen preparation, the soil-ring assemblages were frozen at temperatures of –5, –10, –15, and –20°C for more than 24 h.

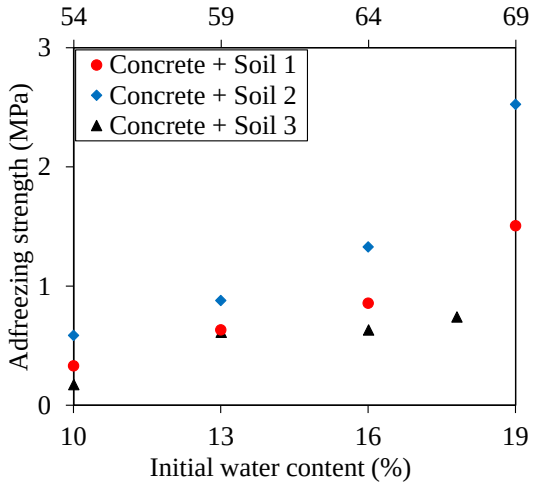
During testing, the load (L) applied to the punch head was transferred to the frozen soil-ring interface, causing shear stress on it. The relationship between the displacement of the punch head and the adfreezing stress (σ_{adf}) calculated by Eq. (1) was determined to exhibit adfreezing behavior.

$$\sigma_{adf} = \frac{L}{\pi DH} \tag{1}$$

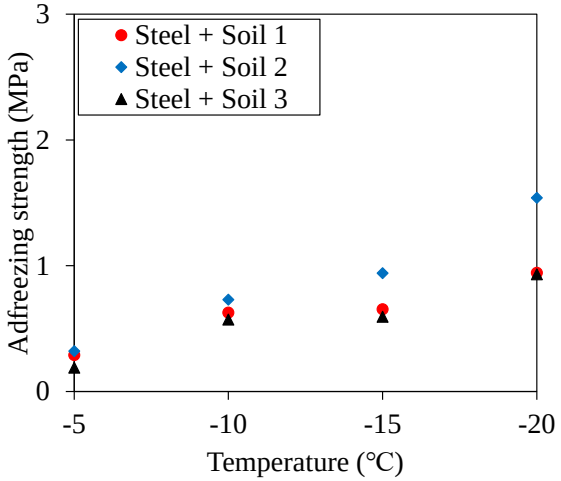
where D is the inner diameter of the ring structure (the same as the diameter of the frozen soil specimen), and H is the height of the ring structure.

Table 1 Physical properties of soil specimens

	Specific gravity (Gs)	Mass median diameter (D_{50}, mm)	Sand fraction (F_{sand})	Silt fraction (F_{silt})	Clay fraction (F_{clay})	Classification (USCS)
Soil 1	2.65	1.0219	100	0	0	SP
Soil 2	2.66	0.2181	98	0	0	SP
Soil 3	2.72	0.0086	19	51	21	CH



(a) Adfreezing strength at -15°C



(b) Adfreezing strength at a constant initial water content

Fig. 2 Experimental results for different ring assemblages

The adfreezing strength was determined as the maximum adfreezing stress observed at the point of slip failure along the soil-ring interface and summarized in Fig. 2. In the case of the concrete ring assemblage, as depicted in Fig. 2(a), the adfreezing behaviors were assessed at a `freezing temperature of –15°C, with variations in the initial water content. On the other hand, under the steel ring condition, as shown in Fig. 2(b), the adfreezing behaviors were evaluated with a constant initial water content (specifically, 16% for soil 1 and soil 2, and 64% for soil 3) while varying the freezing temperature. It was observed that as the initial water content increased at a constant freezing temperature, ice cementation also increased due to an increase in ice content, subsequently leading to increasing the adfreezing strength. Moreover, the adfreezing strength exhibited an upward trend with decreasing temperature, attributed to lower temperatures reducing the unfrozen water content and enhancing the ice strength. For the coarse-grained soils (soil 1 and soil 2), as mass median diameter (D_{50}) decreased, the normalized roughness increased, thereby increasing interfacial resistance. Finally, the concrete ring assemblage exhibited a higher adfreezing strength than the steel ring assemblage because the concrete ring had a rougher surface than the steel ring, leading to greater friction.

2.2 Data analysis

The experimental data consisted of 24 cases, encompassing approximately 42,000 data points that included various parameters such as displacement, adfreezing stress, mass median diameter (D_{50}), freezing temperature, initial water content, unfrozen water content, and ring material (concrete or steel). The calculation of unfrozen water content followed the method proposed by Aukenthaler (2016), which empirically estimated the water content using an exponential function based on the variables of temperature and soil texture, including the fractions of sand (F_{sand}), silt (F_{silt}), and clay (F_{clay}). Notably, higher temperatures and increased fine content were associated with higher levels of unfrozen water content, as shown in Fig. 3. Table 2 summarizes the statistical characteristics of the database, excluding the structural material, which is considered a categorical variable.

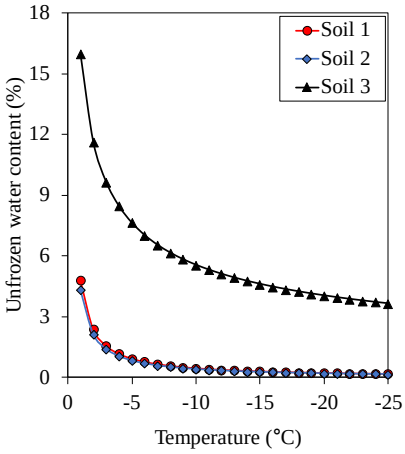


Fig. 3 Unfrozen water content with temperature calculated by the semi-empirical method

Table 2 Statistical descriptions for the database

Feature	Unit	Mean	Std.	Min	Q1	Median	Q3	Max
Displacement	mm	1.99	1.93	0.00	0.47	1.18	3.13	7.65
Adfreezing stress	MPa	0.52	0.45	0.00	0.19	0.45	0.63	2.52
D_{50}	mm	0.27	0.37	0.01	0.01	0.22	0.22	1.02
Temperature	°C	−14.88	2.36	−20.00	−15.00	−15.00	−15.00	−5.00
Initial water content	%	37.14	23.80	10.00	16.00	19.00	64.00	67.00
Unfrozen water content	%	2.12	1.96	0.20	0.27	0.41	4.09	6.90

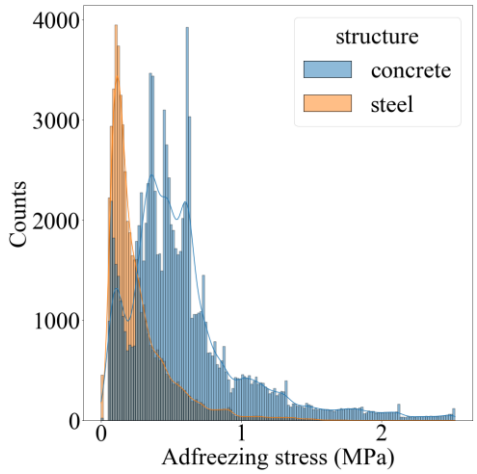


Fig. 4 Histogram according to the ring material

It's worth noting that the distribution of data points according to the ring material (concrete or steel) was imbalanced. As shown in Fig. 4, the distribution of adfreezing stress across the dataset revealed that data from the concrete ring accounted for 80.94% of the total, while data from the steel ring made up the remaining 19.06%. In addition, the adfreezing stress distribution for the steel ring exhibited a narrower range compared to that of the concrete ring, consistent with the experimental results.

3. Development of predictive models

3.1 Machine learning algorithms

This study evaluated the predictive performance of four ML algorithms: two individual models and two ensemble models consisting of multiple decision trees (DTs). The individual models included support vector regressor (SVR) and feedforward neural network (FNN) having one hidden layer, while the ensemble models consisted of random forest (RF) and extreme gradient boosting (XGB).

3.1.1 Support vector regressor (SVR)

SVR, which is used to address regression problems, employs a support vector machine algorithm. SVR with ε-insensitive defines an ε-tube with a radius of ε from the SVR model and quantifies the difference between the actual value and the ε-tube as the error; errors smaller than ε between the predicted and actual values are disregarded.

The objective functions for SVR with ε-insensitive are formulated in Eqs. (2) and (3), where ω represents the weight vector, C is the regularization parameter, n is the number of training datasets, ξ is the slack variable indicating the penalty for the out-of-margin samples, K denotes the kernel function, y stands for the actual value, and ε is the width of the margins. In addition, SVR, which is based on a linear model, solves nonlinear problems by mapping the input space into a higher dimension using a kernel function. In this study, the radial basis function (RBF) described in Eq. (4) was employed as the kernel function, where γ denotes the kernel parameter, and $\|x - x_i\|^2$ represents the squared Euclidean distance between the two feature vectors.

$$\text{Minimize:} \quad \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}|\xi_i| \tag{2}$$

$$\text{Subject to:} \quad |(\omega^T K_{RBF}(x, x^*) + b) - y_i| \le \varepsilon + |\xi_i| \tag{3}$$

$$K_{RBF}(x, x_i) = exp(-\gamma\|x - x_i\|^2) \tag{4}$$

3.1.2 Feedforward neural network (FNN)

An FNN comprises one input layer, one or more hidden layers, and one output layer, with each layer consisting of multiple interconnected nodes. In this study, a single hidden-layer structure was adopted. While the input and output layers matched the number of nodes, it was necessary to determine the optimal number of nodes in the hidden layers. The nodes within the hidden layers utilize a nonlinear activation function for the weighted sum of their inputs, along with a bias term as expressed in Eq. (5), where H represents the output value from one node, f_{act} is the activation function, n_{in} is the number of input features, ω is the weight vector, x is the input value, and b is the bias. The weight and bias values of the network were tuned using a backpropagation algorithm, incorporating a learning rate based on the output layer errors. The objective function of FNN is presented in Eq. (6), where the mean-squared error function is combined with the L2 regularization term to mitigate overfitting. In this equation, α denotes the regularization parameter.

$$H = f_{act}\left(\sum_{i=1}^{n_{in}} \omega_i x_i + b\right) \tag{5}$$

$$\text{Minimize:} \quad \frac{1}{n}\left(\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \alpha\sum_{i=1}^{n}\omega_i^2\right) \tag{6}$$

3.1.3 Random forest (RF)

RF employs an ensemble learning technique that utilizes multiple DTs through the bagging method. An ensemble of DTs, known as the RF, is generated and trained using bootstrapped samples from the training data and random feature subsets. When the RF calculates the output value, each tree independently generates a prediction, and the final prediction is determined by aggregating the individual tree predictions in parallel, as expressed in Eq. (7). In this study, the loss function for model training was the mean-squared error (MSE), and the hyperparameters of the RF were associated with the control of the DTs.

$$y = \frac{1}{N}\sum_{i=1}^{N} DT_i(x_i) \tag{7}$$

where N is the number of trees, DT_i is the ith decision tree, and x_i is the ith bootstrapped input.

3.1.4 Extreme gradient boosting (XGB)

XGB is another ensemble learning technique that utilizes multiple DTs. However, it differs from RF in that it uses a boosting method instead of bagging. The boosting technique is a sequential learning method that calculates the residual of the predecessor (i.e., the difference between the predicted and actual values), multiplies the residual value by the learning rate, and incorporates this weighted residual when learning the next predictor. The gradient boosting model (GBM) is a DT-based model that employs the boosting technique. XGB involves an additional regularization term in the GBM to address the issue of overfitting. The objective function of XGB combines the loss function, $\sum_{i=1}^{n} l(y_i, \hat{y}_i)$, with the regularization function, $\Omega(f)$, as represented by Eq. (8). The regularization function is defined by Eq. (9), where T is the number of leaves, ω is the vector of scores on the leaves of a tree, and γ and λ are the penalty parameters. In this study, the loss function used for model training was the mean-squared error (MSE).

$$\text{Minimize:} \quad \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \Omega(f) \tag{8}$$

$$\Omega(f) = \gamma T + \frac{\lambda}{2}\sum_{j=1}^{T} \omega_j^2 \tag{9}$$

3.2 Model implementation

Fig. 5 illustrates a flowchart outlining the model development procedure. The four ML models predicted the adfreezing stress using the six additional features in the database.

3.2.1 Data preprocessing

The ring material, which is a categorical variable, was binarized using the one-hot encoding method and converted into a numerical variable format suitable for ML algorithms. In order to mitigate the influence of varying scales among input features, the input dataset was standardized using Eq. (10). Addressing data imbalance associated with the ring material, the entire database was partitioned into a 70% training dataset and a 30% test dataset using a stratified random sampling technique with respect to the ring material. Consequently, the training and test datasets were established to maintain the same ratio of the concrete and steel rings as that in the original database, as shown in Table 3.

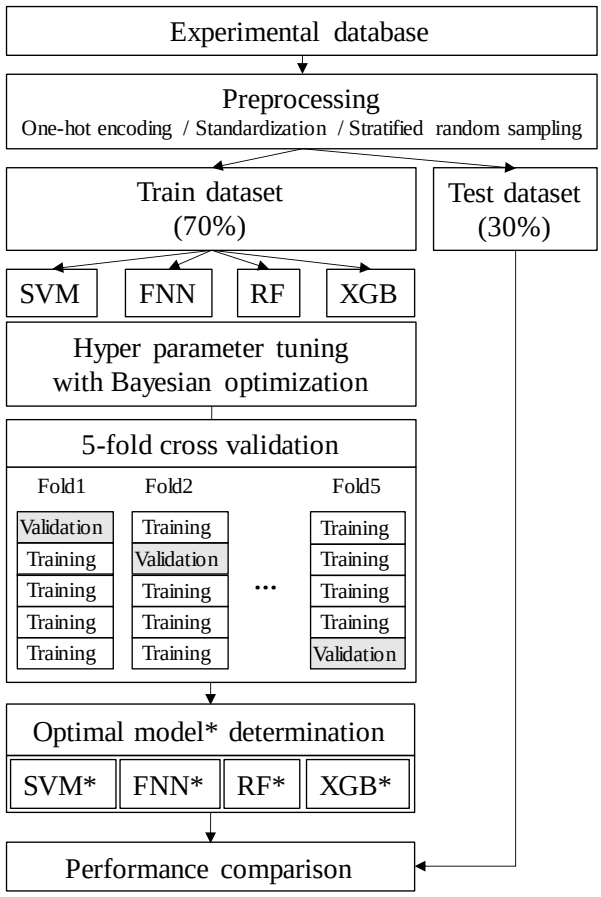


Fig. 5 Flowchart of model development procedure

Table 3 Stratified random sampling results

Unit: %	Concrete ring structure	Steel ring structure
Entire database	80.9379	19.0624
Training dataset	80.9386	19.0614
Test dataset	80.9354	19.0646

$$Z = \frac{X - \bar{X}}{s} \tag{10}$$

where Z is the scaled input dataset, X is the raw input dataset, $\bar{X}$ is the mean of the input dataset, and s is the standard deviation of the input dataset.

3.2.2 Model development

The optimal hyperparameters for the four ML models were determined through a Bayesian search along with the five-fold cross-validation method. Previous studies have demonstrated the effectiveness of grid and randomized searches in tuning hyperparameters in various machine-learning models (Li *et al.* 2022, Odebiri *et al.* 2022, Tang and Na 2021, Tarek *et al.* 2023). These approaches have

Table 4 Hyperparameters of the ML models

Model	Hyperparameters	Search space; Type; Distribution	Optimal value
SVR	γ	$[10^{-5}, 10^{4}]$; Real; Log-uniform	0.41182
	C	$[10^{-5}, 10^{4}]$; Real; Log-uniform	0.00879
	ε	$[10^{-5}, 10^{4}]$; Real; Log-uniform	0.00117
FNN	Hidden layer sizes	[1, 500]; Integer; Uniform	401
	Activation function	[identity, logistic, tanh, ReLU]; Categorical	ReLU
	α	$[10^{-5}, 1]$; Real; Log-uniform	0.00182
	Learning rate	$[10^{-5}, 10^{-2}]$; Real; Log-uniform	1.86078×10-5
RF	Number of DTs	[1, 500]; Integer; Uniform	420
	Maximum depth of DT	[1, 30]; Integer; Uniform	23
	Minimum number of samples required to split a node	[2, 30]; Integer; Uniform	21
	Minimum number of samples required in a leaf	[1, 30]; Integer; Uniform	7
XGB	Number of DTs	[1, 500]; Integer; Uniform	396
	Maximum depth of DT	[1, 30]; Integer; Uniform	12
	Maximum number of leaves	[1, 30]; Integer; Uniform	30
	Learning rate	$[10^{-4}, 1]$; Real; Log-uniform	0.05365
	L1 regularization parameter	$[10^{-2}, 10]$; Real; Uniform	0.01
	L2 regularization parameter	$[10^{-2}, 10]$; Real; Uniform	10.0

been widely adopted to tune model hyperparameters for improving model performance. However, dealing with an increased number of hyperparameters and larger search spaces, tuning the hyperparameters can become computationally intensive. To address this issue, this study utilized Bayesian reasoning to optimize the hyperparameters, thereby reducing computational costs.

Bayesian search is an iterative strategy that combines a probabilistic surrogate model and an acquisition function to optimize the performance of an ML model. The surrogate model, which is typically constructed using Gaussian processes, provides a probabilistic representation of the objective function with the given hyperparameters. Subsequently, the acquisition function guides the search process by selecting the most promising set of hyperparameters for assessment, based on the surrogate model's estimates. As new evaluations are performed, the surrogate model continuously updates, thus enhancing its accuracy. Through this iterative process, Bayesian optimization can determine the optimal combination of hyperparameters, resulting in the best-performing ML model. Comprehensive details on Bayesian optimization can be found in Snoek *et al*. (2012) and Frazier (2018).

Five-fold cross-validation was applied to enhance the reliability of results (Stone 1974). In the training process, the training dataset was randomly partitioned into five subsets to implement the five-fold cross-validation. In each fold, four of these subsets were used for training, and the remaining subset was reserved for model validation. The set of hyperparameters that yielded the highest accuracy in the five-fold cross-validation was determined to be the optimal hyperparameter combination.

Table 4 lists the hyperparameters of each ML model, the respective search spaces for these hyperparameters, and the optimal values determined by a combination of Bayesian optimization and five-fold cross-validation.

3.2.3 Performance evaluation

The predictive performance of the ML models was assessed using three regression metrics, namely root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2), represented by Eqs. (11)-(13), respectively. RMSE and MAE quantify the degree of similarity between predicted and observed adfreezing stress, with smaller values indicating greater accuracy. Meanwhile, R^2 assesses the linearity between predicted and experimental data, with values ranging from 0 to 1. An R^2 value approaching 1 signifies a strong correlation, indicating superior predictive performance of the ML models.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i^{pred} - y_i^{obs}\right)^2} \tag{11}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i^{pre} - y_i^{obs}\right| \tag{12}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_i^{pre} - y_i^{obs}\right)^2}{\sum_{i=1}^{N}\left(y_i^{pre} - \bar{y}_i^{obs}\right)^2} \tag{13}$$

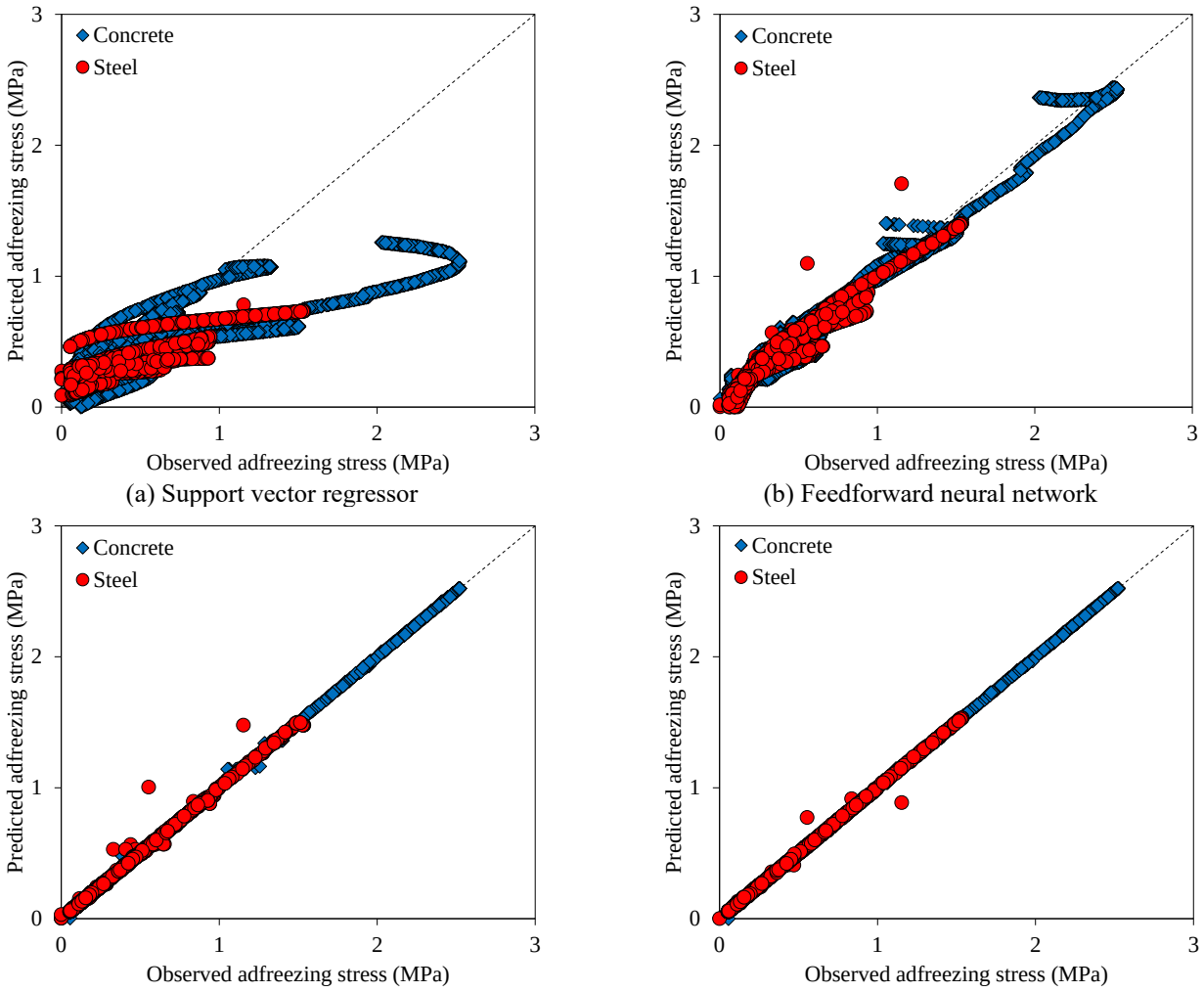
where N is the number of datasets, y_i^{pred} is the predicted value, y_i^{obs} is the observed value, and $\bar{y}_i^{obs}$ is the mean of the observed values.

4. Results and comparisons

This section compares the overall predictive performance of the optimal models, generated by each algorithm using a test dataset. Two post-hoc analyses, one focusing on prediction uncertainty and the other on prediction interpretation, were performed specifically on the model that exhibited the best performance. Finally, a

Table 5 Comparison of the model performance by regression metrics

Regression metrics	Criteria	Material	SVR	FNN	RF	XGB
RMSE	RMSE → 0	Concrete	0.3122	0.0863	0.0031	0.0019
		Steel	0.2182	0.0772	0.0145	0.0079
		Total	0.2965	0.0847	0.0069	0.0037
MAE	MAE → 0	Concrete	0.1508	0.0697	0.0013	0.0012
		Steel	0.1627	0.0606	0.0044	0.0024
		Total	0.1531	0.0680	0.0019	0.0015
R^2	$R^2 \rightarrow 1$	Concrete	0.6897	0.9688	1.0000	1.0000
		Steel	0.6035	0.9377	0.9978	0.9993
		Total	0.5911	0.9667	0.9998	0.9999



(a) Support vector regressor (b) Feedforward neural network

(c) Random forest (d) Extreme gradient boosting

Fig. 6 Correlations between observed and predicted adfreezing strengths

parametric study was conducted to analyze the prediction outcomes of the model under untrained conditions.

4.1 Performance comparison

Fig. 6 shows the observed adfreezing stress plotted against the predicted adfreezing stress for each model employing the optimal set of hyperparameters using the test dataset, categorized by ring materials (concrete and steel). In the case of individual models (SVR and FNN), the continuous evolution of adfreezing stress was overall underestimated compared to the actual values. This underestimation might be attributed to the relatively simplistic model structures, which struggled to capture the nonlinearity of the adfreezing behavior. In FNN, accuracy can be improved by enhancing model complexity, such as adopting multiple hidden layers. However, ensemble models (RF and XGB), with their higher complexity,

accurately predicted adfreezing stress, and non-continuous and scattered outliers were primarily observed in case of the steel ring.

The model performance was quantitatively evaluated using the three regression metrics, as listed in Table 5. The ensemble models outperformed the individual models, with XGB exhibiting the best performance. As the machine learning model allows more sophisticated hyperparameter tuning with an increase in the training data volume, the developed models generally showed improved performance in case of the concrete ring, which represented approximately 81% of the data, compared to the steel ring (approximately 19% of the data). It's noteworthy that the non-continuous outliers in Fig. 6 were mainly observed in the case of the steel ring. However, the RMSE values for SVR and FNN, as well as the MAE value for FNN, were higher (indicating poor performance) for the concrete ring compared to the steel ring. These results can be attributed to the overall adfreezing stress underestimation due to model complexity.

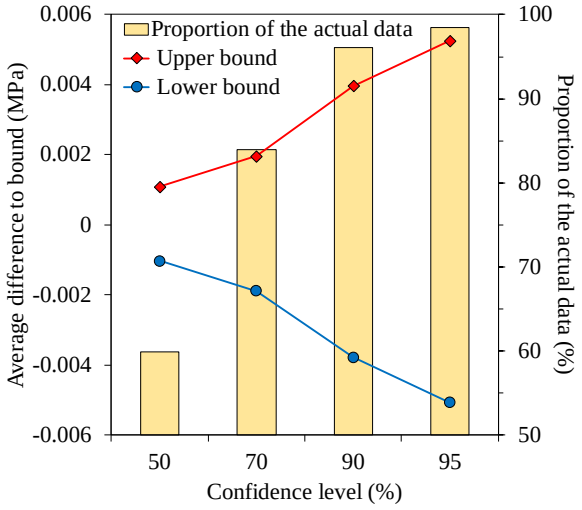


Fig. 7 Analysis results of the conformal prediction method with XGB

4.2 Prediction uncertainty

Uncertainty in the prediction of ML models pertains to the inherent limitations associated with ML approaches in predicting outcomes, stemming from factors such as incomplete data, model complexity, and inherent randomness within the underlying processes. Therefore, assessing prediction uncertainty becomes essential to evaluate the reliability and robustness of ML models.

The conformal prediction method primarily serves to quantify uncertainty in a single-output model (Shafer and Vovk 2008, Barber *et al*. 2021). In the conformal prediction method, a model $\mu(x)$: $\mathbb{R}^d \rightarrow \mathbb{R}$, trained with independent and identically distributed training data $(X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R}$ $(i = 1, \ldots, n)$, predicts $\mu(X_{n+1})$ based on a given new input vector X_{n+1} and provides a prediction interval around $\mu(X_{n+1})$ that is likely to contain the actual value of Y_{n+1}. This prediction interval, denoted as $\hat{C}_{n,\epsilon}(X_{n+1})$, carries a probability P, as defined in Eq. (14), at a given error probability ϵ. The confidence level can be expressed as $1 - \epsilon$, and following the law of large numbers, a 95% confidence level is typically deemed valid.

$$P\{Y_{n+1} \in \hat{C}_{n,\epsilon}(X_{n+1})\} \geq 1 - \epsilon \tag{14}$$

In this study, MAPIE—an open-source Python library (Taquet et al., 2022)—was employed to determine the prediction interval and provide its upper and lower bounds using a test dataset. The analysis was conducted by varying the confidence levels to 50%, 70%, 90%, and 95%. For each confidence level, the distances between the predictions and the bounds were calculated, and the amount of actual data included in the prediction interval (between the lower and upper bounds) was counted.

Fig. 7 shows the average difference between the predictions and the bounds, as well as the proportion of actual data situated within the prediction interval according to the confidence level. As the confidence level increased, the average distance to the bounds also increased. Consequently, the proportion of actual data encompassed within the prediction interval also increased. At a 95% confidence level, despite the average distances to the bounds remaining within 0.006, a remarkable 98.42% of the actual data was included in the prediction interval, signifying the reliability of the developed XGB model.

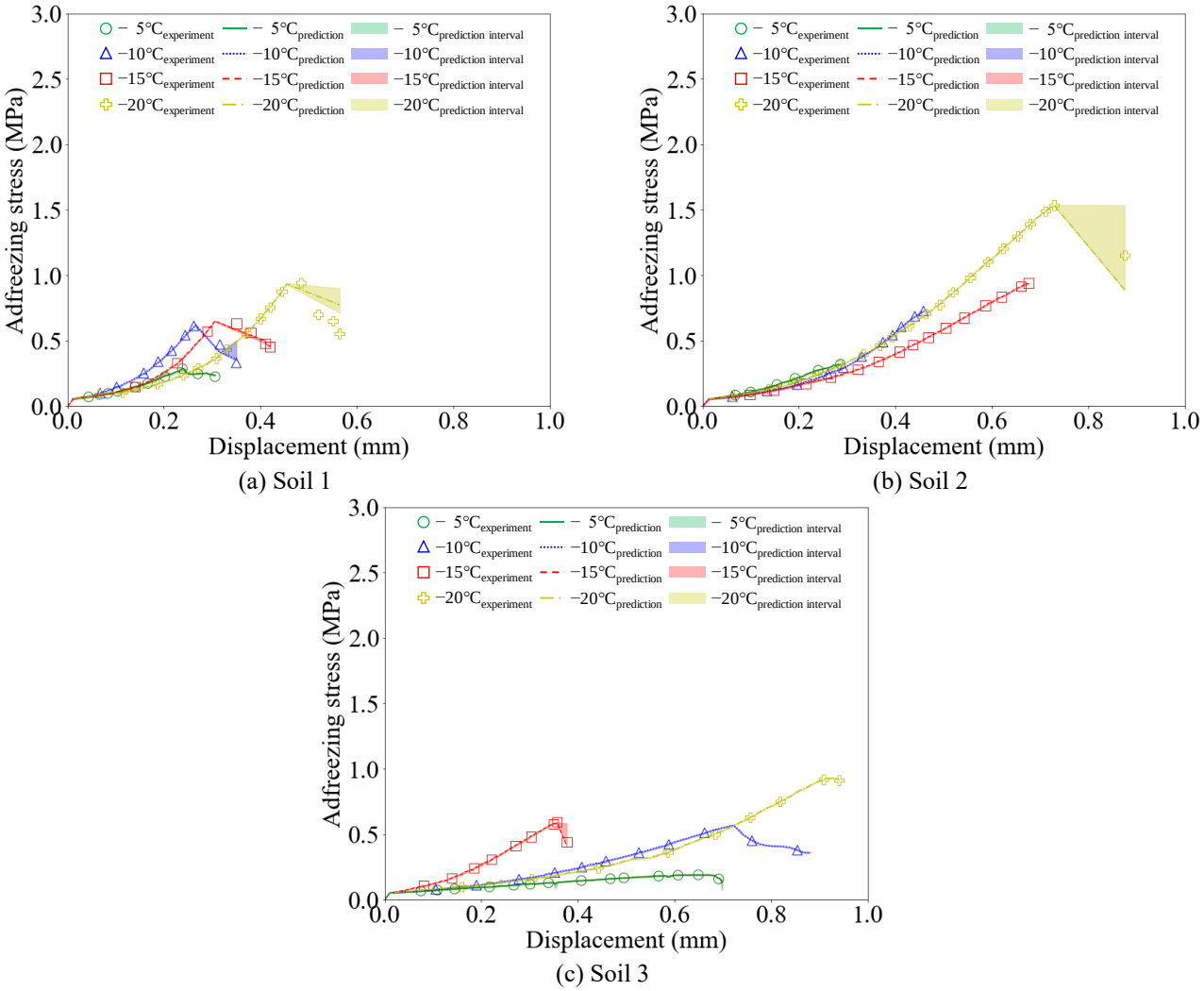
Fig. 8 displays the predicted results at a 95% confidence level for the steel ring case (note that the prediction intervals for the concrete ring case were exceptionally narrow). In Fig. 8, the symbols denote the actual data points, the solid lines represent the XGB predictions, and the shades indicate the prediction intervals. While the XGB model generally predicted the adfreezing behavior with narrow prediction intervals, relatively broader prediction intervals were evident toward the end, where slip failure occurred after reaching the maximum adfreezing stress. The model's uncertainty increased in this region due to the limited data available after the test ceased upon slip failure, leading to the expansion of the prediction interval. No doubt that the prediction interval would have narrowed if there had been sufficient data available after slip failure.

4.3 Feature influence analysis

To analyze the effect of experimental factors on adfreezing behavior prediction, this study adopted the SHapley Additive exPlanations (SHAP) approach (Lundberg and Lee 2017), which is based on the concept of Shapley values (Shapley 1953) from cooperative game theory. SHAP provides importance scores for each feature and calculates the impact of each predictor on the prediction. The fundamental idea behind SHAP is to use a linear explanation model for predicting the original trained model with an additive feature attribution method. By denoting the original trained model as $f(x)$, the linear explanation model $g(z')$ can be described by Eq. (15).

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z_i' \tag{15}$$

where z' is a simplified local input, ϕ_0 is the baseline of the model, M is the number of input features, and ϕ_i is the SHAP value corresponding to the i-th feature.



(a) Soil 1 (b) Soil 2

(c) Soil 3

Fig. 8 Adfreezing stress–displacement curves with a steel ring structure

In this study, the TreeSHAP algorithm was employed to reduce computational costs, which is a variant of SHAP proposed by Lundberg *et al.* (2018) specially designed for tree-based ML models. The SHAP values for the i-th feature are computed as follows

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|M|-|S|-1)!}{|M|!} [f_x(S \cup \{i\}) - f_x(S)] \quad (16)$$

$$f_x(S) = E[f(x)|x_s] \quad (17)$$

where N is the set of all input features, S is a subset of the features, and $E[f(x)|x_s]$ is the expected value of the original model with the features in set S.

In Fig. 9, the left chart displays the average of the absolute SHAP values for global feature importance, while the right side presents the SHAP values according to the feature values to estimate the effect of each predictor on the output. The analysis indicates that the top three features (unfrozen water content, displacement, and initial water content) have a significant effect on the output, compared to the bottom four features (concrete, D_{50}, temperature, and steel), as shown in Fig. 9. The feature of the steel ring did not affect the output.

In contrast, the right side of Fig. 9 illustrates that, for the unfrozen-water content and concrete, the SHAP values decreased with higher feature values, indicating a negative correlation with adfreezing stress prediction. Conversely, for D_{50}, the SHAP values increased with higher feature values, suggesting a positive correlation. For the displacement and initial water content, high feature values were located in the middle of each SHAP value distribution. The temperature did not exhibit a clear correlation with the predicted output. Although the experimental results showed that each feature affected adfreezing behavior (referring to Section 2.1), the SHAP analysis results may not fully align with the actual mechanism driving the influence of several factors. This discrepancy arises because the SHAP results for each feature are calculated solely based on the given data values.

4.4 Generalization performance

A parametric study was conducted under unseen hypothetical conditions to assess the generalization performance of the XGB model. Fig. 10 shows the adfreezing strengths obtained from the predicted adfreezing

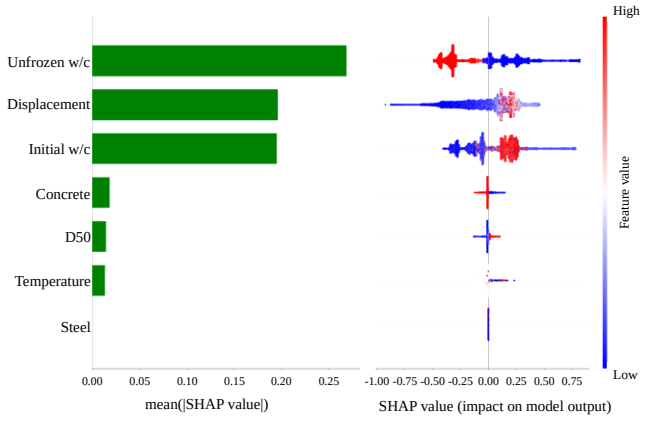


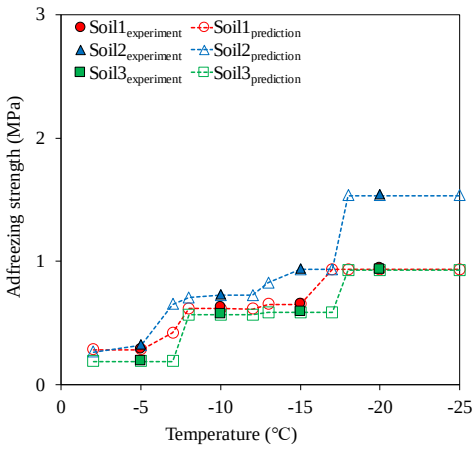
Fig. 9 TreeSHAP analysis results

stress–displacement curves under various conditions, alongside the experimental results (from Fig. 2). In the case of the steel ring, the adfreezing strengths were predicted for different freezing temperatures at a constant initial water content (i.e., 16% for soil 1 and soil 2 and 64% for soil 3). In the case of the concrete ring, the adfreezing strengths were predicted for varying initial water contents at a constant freezing temperature of –15°C.

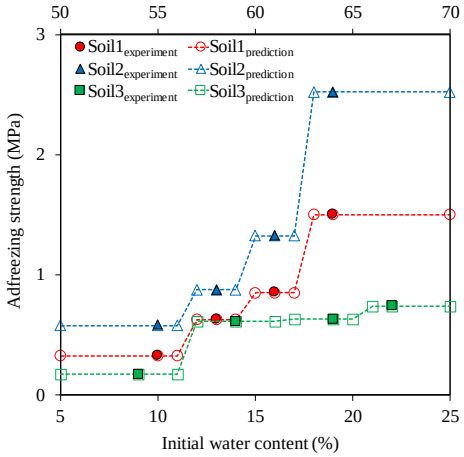
The steel ring showed that the XGB model provided reasonable predictions for the adfreezing strengths between within the temperature range of –5 and –15°C, as illustrated in Fig. 10(a), suggesting its potential for interpolation. However, at temperatures below –15°C, the XGB model failed to deliver satisfactory interpolation performance due to the exponential changes in the unfrozen water content, which had the most significant impact on the model's predictions. These changes were related to the semi-empirical method used to calculate unfrozen water content in this study. As shown in Fig. 3, the unfrozen water content exhibited significant variations corresponding to temperature changes within the higher temperature range, which the XGB model effectively captured. However, in the low temperature range, the unfrozen water content varied slightly with temperature changes, resulting in reduced model performance. It's worth noting that extrapolation beyond the temperature range covered by the trained data (i.e., temperatures above –5°C and below –25°C) was also unfeasible.

The generalization performance of the XGB model was less effective for the concrete ring when compared to the steel ring. As illustrated in Fig. 10(b), when utilizing the unseen initial water content as the input parameter, the XGB model struggled to generate acceptable interpolated and extrapolated values. This suggests that while the initial water content was the feature with the third-highest influence on the prediction, its influence was insufficient to enable effective interpolation or extrapolation.

If the XGB model was trained using more experimental datasets conducted under various conditions, the influence of the input features on the adfreezing stress prediction would differ. This could potentially enhance the model's ability to perform both interpolation and extrapolation.



(a) Steel ring structure



(b) Concrete ring structure

Fig. 10 Experimented and predicted adfreezing strengths

5. Discussion

While the XGB model developed in this study exhibited excellent performance, it's important to acknowledge that machine learning approaches—data-driven methods that rely solely on training data without incorporating prior scientific knowledge—come with inherent limitations. These limitations include the potential for producing incorrect or misleading predictions if the training data is biased, sparse, or outdated. This study assessed how the performance of the XGB model was affected by the imbalance in data quantity between the concrete ring and the steel ring. Notably, prediction uncertainty significantly increased in regions with sparse data, particularly near the slip failure. The SHAP analysis results may not be consistent with prior information until the XGB model is further trained with experimental data obtained under a wider array of conditions, ideally leading to larger and more comprehensive databases. Furthermore, while the XGB model was capable of providing reasonable predictions within the range of the trained data (permitting limited interpolation), it produced less reliable values outside this established range.

Although a post-hoc analysis provides valuable insights into prediction results, it can only offer a retrospective view

and cannot guarantee that a model will generalize well with new data. In this study, the prediction interval played a crucial role in enhancing prediction robustness by providing a reliable prediction range. The SHAP analysis helped understand why the XGB model made specific predictions and quantified the influence of various factors. However, when prediction uncertainty becomes exceedingly high for certain reasons, the prediction interval may not provide useful information. In addition, it's essential to acknowledge that SHAP values are calculated based on the provided dataset and may not fully account for prior information.

Recent research has explored a new paradigm that combines machine learning (ML) approaches with prior scientific knowledge, aiming to overcome the inherent limitations of typical ML methods (Raissi *et al*. 2019, Rai and Sahu 2020, Zhang *et al*. 2022). These hybrid ML models, enriched with prior information, exhibit enhanced robustness to changes in data distribution and offer more interpretable results. Moreover, the use of prior knowledge can alleviate the need for extensive training datasets, making ML approaches applicable to smaller and imbalanced datasets. Therefore, there is a growing need for further research to actively implement hybrid ML approaches, improving both model performance and interpretability.

6. Conclusions

This study investigated the use of machine learning (ML) models to enhance the applicability of the punch shear test in evaluating adfreezing behavior at frozen soil-structure interfaces under zero confinement conditions. Using a dataset of 24 experimental cases, predictive models were developed and validated with various ML algorithms, followed by post-hoc analyses to improve their reliability and interpretability. The key findings and conclusions of this study are summarized as follows:

- Optimal hyperparameter sets were effectively determined through Bayesian optimization and five-fold cross-validation approach. Model performance was assessed using three regression metrics: RMSE, MAE, and R2. The models' prediction performance ranked in the following order: XGB > RF > FNN > SVR. Notably, the XGB model outperformed the others, achieving an RMSE of 0.0037, MAE of 0.0015, and R^2 of 0.9999.
- The predictive performance of the ML model was affected by both the complexity of the model structure and the quantity of training data. Ensemble models (RF and XGB), known for their high structural complexity, demonstrated superior prediction performance to individual models (SVR and FNN), which shows a trade-off relationship between prediction accuracy and model complexity. Additionally, the ensemble models displayed greater prediction accuracy when applied to the concrete ring, which had a larger dataset, as opposed to the steel ring case.
- The reliability and interpretability of the XGB model were enhanced by quantifying prediction uncertainty and assessing the influence of input features through post-hoc techniques such as prediction intervals and SHAP value analysis. However, these post-hoc techniques provide a retrospective perspective on prediction results and do not guarantee strong generalization performance when applied to new data.
- It is necessary to discuss the inherent limitations of the ML models and a potential solution such as hybrid ML models. The hybrid model combines prior information with the ML approach, offering a means to address the shortcomings of data-driven modeling.

Acknowledgments

This research was supported by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure and Transport (No. RS-2024-00410248).

References

Alzoubi, M.A., Xu, M., Hassani, F.P., Poncet, S. and Sasmito, A.P. (2020), "Artificial ground freezing: A review of thermal and hydraulic aspects", *Tunn. Undergr. Sp. Tech.*, **104**, 103534. https://doi.org/10.1016/j.tust.2020.103534.

Andersland, O.B. and Ladanyi, B. (2003), *Frozen Ground Engineering*, John Wiley & Sons.

Aukenthaler, M. (2016), "The frozen & unfrozen Barcelona basic model", Ph.D. Dissertation, Delft University of Technology.

Baghbani, A., Choudhury, T., Costa, S. and Reiner, J. (2022), "Application of artificial intelligence in geotechnical engineering: A state-of-the-art review", *Earth-Science Rev.,* **228**, 103991. https://doi.org/10.1016/j.earscirev.2022.103991.

Barber, R.F., Candès, E.J., Ramdas, A. and Tibshirani, R.J. (2021), "Predictive inference with the jackknife+", *Ann. Stat.,* **49**, 486-507. https://doi.org/10.1214/20-AOS1965.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F. (2020), "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI", *Inf. Fusion.*, **58**, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012.

Bello, O., Holzmann, J., Yaqoob, T. and Teodoriu, C. (2015), "Application of artificial intelligence methods in drilling system design and operations: A review of the state of the art", *J. Artif. Intell. Soft Comput. Res*., **5**, 121-139. https://doi.org/10.1515/jaiscr-2015-0024

Chen, W., Luo, Q., Liu, J., Wang, T. and Wang, L. (2022), "Modeling of frozen soil-structure interface shear behavior by supervised deep learning", *Cold Reg. Sci. Technol*., **200**, 103589. https://doi.org/10.1016/j.coldregions.2022.103589.

Fang, J., Feng, Z., Cao, S.J. and Deng, Y. (2018), "The impact of ventilation parameters on thermal comfort and energy-efficient control of the ground-source heat pump system", *Energ. Buildings*, **179**, 324-332. https://doi.org/10.1016/j.enbuild.2018.09.024.

Frazier, P.I. (2018), "A tutorial on Bayesian optimization", *arXiv Prepr.,* arXiv1807.02811.

Fuping, Z., Jianguo, S., Wenting, G. and Pengfei, H. (2023), "Experimental study on freezing strength of soil-concrete lining interface in cold regions", *Geofluids*, https://doi.org/10.1155/2023/8910226.

He, P.F., Mu, Y.H., MA, W., Huang, Y.T. and Dong, J.H. (2021), "Testing and modeling of frozen clay–concrete interface behavior based on large-scale shear tests", *Adv. Clim. Chang. Res.*, **12**, 83-94. https://doi.org/10.1016/j.accre.2020.09.010.

Jin, H., Lee, J., Zhuang, L. and Byu, B.H. (2020), "Laboratory investigation of unconfined compression behavior of ice and frozen soil mixtures", *Geomech. Eng.*, **22**(3), 219-226. https://doi.org/10.12989/gae.2020.22.3.219.

Kim, D., Kwon, K., Pham, K., Oh, J.Y. and Choi, H. (2022) "Surface settlement prediction for urban tunneling using machine learning algorithms with Bayesian optimization", *Automat. Constr.*, **140**, 104331. https://doi.org/10.1016/j.autcon.2022.104331.

Ladanyi, B. (1995), "Frozen soil — structure interfaces", *Stud. Appl. Mech.*, **42**, 3-33. https://doi.org/10.1016/S0922-5382(06)80004-8.

Lawal, A.I. and Idris, M.A. (2020), "An artificial neural network-based mathematical model for the prediction of blast-induced ground vibrations" *Int. J. Environ. Stud.*, **77**, 318-334. https://doi.org/10.1080/00207233.2019.1662186.

Li, K.Q., Liu, Y. and Kang, Q. (2022), "Estimating the thermal conductivity of soils using six machine learning algorithms", *Int. Commun. Heat Mass Transf.*, **136**, 106139. https://doi.org/10.1016/j.icheatmasstransfer.2022.106139.

Linardatos, P., Papastefanopoulos, V. and Kotsiantis, S. (2021), "Explainable ai: A review of machine learning interpretability methods", *Entropy.*, **23**, 1-45. https://doi.org/10.3390/e23010018.

Lundberg, S.M. and Lee, S.I. (2017), "A unified approach to interpreting model predictions", *Adv. Neural Inf. Process. Syst.,* **30**.

Lundberg, S.M., Erion, G.G. and Lee, S.I. (2018), "Consistent individualized feature attribution for tree ensembles", *arXiv preprint*, arXiv:1802.03888.

Odebiri, O., Mutanga, O. and Odindi, J. (2022), "Deep learning-based national scale soil organic carbon mapping with Sentinel-3 data", *Geoderma*, **411**, 115695. https://doi.org/10.1016/j.geoderma.2022.115695.

Pan, R., Yang, P. and Yang, Z. (2022), "Experimental study on the shear behavior of frozen cemented sand-structure interface", *Cold Reg. Sci. Technol.*, **197**, 103516. https://doi.org/10.1016/j.coldregions.2022.103516.

Park, S., Hwang, C., Choi, H., Son, Y. and Ko, T.Y. (2022), "Experimental study for application of the punch shear test to estimate adfreezing strength of frozen soil-structure interface", *Geomech. Eng.*, **29**(3), 281-290. https://doi.org/10.12989/gae.2022.29.3.281.

Pham, K., Jung, S., Park, S., Kim, D. and Choi, H. (2022), "Bayesian neural network for estimating stress-strain behaviors of frozen sand", *KSCE J. Civ. Eng.*, **26**, 933-941. https://doi.org/10.1007/s12205-021-0432-z.

Pham, K., Park, S., Choi, H. and Won, J. (2021) "Data-driven framework for predicting ground temperature during ground freezing of a silty deposit", *Geomech. Eng.,* **26**(3), 235-251. https://doi.org/10.12989/gae.2021.26.3.235.

Quanbin, S., Ping, Y. and Guoliang, W. (2018), "Experimental research on adfreezing strengths at the interface between frozen fine sand and structures", *Sci. Iran.*, **25**(2), 663-674. https://doi.org/10.24200/SCI.2017.20005.

Rai, R., Sahu and C.K. (2020), "Driven by data or derived through physics? A review of hybrid physics guided Machine learning techniques with Cyber-Physical System (CPS) focus", *IEEE Access*, **8**, 71050-71073. https://doi.org/10.1109/ACCESS.2020.2987324.

Raissi, M., Perdikaris, P. and Karniadakis, G.E. (2019), "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations", *J. Comput. Phys.*, **378**, 686-707. https://doi.org/10.1016/j.jcp.2018.10.045.

Schmall, P.C. and Braun, B. (2006), "Ground freezing—a viable and versatile construction technique", *Current Practices in Cold Regions Eng.*, 1-11. https://doi.org/10.1061/40836(210)29.

Shafer, G. and Vovk, V. (2008), "A tutorial on conformal prediction", *J. Mach. Learn. Res.,* **9**, 371–421.

Shapley, L. (1953), "A value for n-person games", Princet. Univ. Press. https://doi.org/10.1097/01.ede.0000417297.03956.ad.

Snoek, J., Larochelle, H. and Adams, R.P. (2012), "Practical bayesian optimization of machine learning algorithms", *Adv. Neural Inf. Process. Syst.*, **25**.

Son, Y., Ko, T.Y., Lee, D., Won, J., Lee, I.M. and Choi, H. (2021), "Applicability of liquid air as novel cryogenic refrigerant for subsea tunnelling construction", *Geomech. Eng.,* **27**(2), 179-187. https://doi.org/10.12989/gae.2021.27.2.179.

Stone, M. (1974), "Cross-validatory choice and assessment of statistical predictions", *J. R. Stat. Soc. Ser. B.*, **36**(2), 111-133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x.

Sun, T., Gao, X., Liao, Y. and Feng, W. (2021), "Experimental study on adfreezing strength at the interface between silt and concrete", *Cold Reg. Sci. Technol.,* **190**, 103346. https://doi.org/10.1016/j.coldregions.2021.103346.

Tan, Y., Lu, Y. and Wang, D. (2021), "Catastrophic failure of Shanghai metro line 4 in July, 2003: Occurrence, emergency response, and disaster relief", *J. Perform. Constr. Fac.*, **35**(1), 04020125. https://doi.org/10.1061/(ASCE)CF.1943-5509.0001539.

Tan, Y., Lu, Y. and Wang, D. (2023), "Catastrophic Failure of Shanghai Metro Line 4 in July 2003: Postaccident Rehabilitation", *J. Perform. Constr. Fac.*, **37**(2), 04023006. https://doi.org/10.1061/JPCFEV.CFENG-4135.

Tang, L., Du, Y., Liu, L., Jin, L., Yang, L. and Li, G. (2020), "Effect mechanism of unfrozen water on the frozen soil-structure interface during the freezing-thawing process", *Geomech. Eng.*, **22**(3), 245-254. https://doi.org/10.12989/gae.2020.22.3.245

Tang, L. and Na, S.H. (2021), "Comparison of machine learning methods for ground settlement prediction with different tunneling datasets", *J. Rock Mech. Geotech. Eng.*, **13**, 1274-1289. https://doi.org/10.1016/j.jrmge.2021.08.006.

Taquet, V., Blot, V., Morzadec, T., Lacombe, L. and Brunel, N. (2022), "MAPIE: an open-source library for distribution-free uncer-tainty quantification", *arXiv Prepr.,* arXiv2207.12274.

Tarek, Z., Elshewey, A.M., Shohieb, S.M., Elhady, A.M., El-Attar, N.E., Elseuofi, S. and Shams, M.Y. (2023), "Soil erosion status prediction using a novel random forest model optimized by random search method" *Sustain.*, **15**. https://doi.org/10.3390/su15097114.

Wang, D., Wang, T., Xu, D. and Zhou, G. (2020), "Estimation of spatial autocorrelation variations of uncertain geotechnical properties for the frozen ground", *Geomech. Eng.*, **22**(4), 339-348. https://doi.org/10.12989/gae.2020.22.4.339.

Wang, S., Wang, Q., Qi, J. and Liu, F. (2018), "Experimental study on freezing point of saline soft clay after freeze-thaw cycling", *Geomech. Eng.*, **15**(4), 997-1004. https://doi.org/10.12989/gae.2018.15.4.997.

Wang, T., Zhou, G., Wang. J. and Wang, D. (2020), "Impact of spatial variability of geotechnical properties on uncertain settlement of frozen soil foundation around an oil pipeline", *Geomech. Eng.*, **20**(1), 19-28. https://doi.org/10.12989/gae.2020.20.1.019.

Wen, Z., Yu, Q., Ma, W., Dong, S., Wang, D., Niu, F. and Zhang, M. (2016), "Experimental investigation on the effect of fiberglass reinforced plastic cover on adfreeze bond strength", *Cold Reg. Sci. Technol.,* **131**, 108-115. https://doi.org/10.1016/j.coldregions.2016.07.009

Xiong, M., He, P., Mu, Y. and Na, X. (2021), "Modeling of concrete-frozen soil interface from direct shear test results", *Adv. Civ. Eng.*, https://doi.org/10.1155/2021/7260598.

Zhang, P., Yin, Z.Y. and Sheil, B. (2022), "A physics-informed data-driven approach for consolidation analysis", *Geotechnique*, 1-12. https://doi.org/10.1680/jgeot.22.00046.

Zhang, Q., Zhang, J., Zhang, T., Liu, S., Li, Y. and Xu, G. (2022), "Experimental analysis and mechanical model of interaction between warm frozen silt and pre-bored grouted planted nodular pile", *Cold Reg. Sci. Technol.*, **202**, 103632. https://doi.org/10.1016/j.coldregions.2022.103632.

Zhou, X.M., Jiang, G., Li, F., Gao, W., Han, Y., Wu, T. and Ma, W. (2022), "Comprehensive review of artificial ground freezing applications to urban tunnel and underground space engineering in China in the last 20 years", *J. Cold Reg. Eng.*, **36**(3), 04022002. https://doi.org/10.1061/(ASCE)CR.1943-5495.0000273.

GC