American Journal of Civil Engineering
2026, Vol. 14, No. 6, pp. 411–424
https://doi.org/10.11648/j.ajce.20261406.11

Research Article

Settlement Response and Prediction of Soft-Soil Subgrade Widening Based on Abaqus Numerical Simulation and Machine Learning

Zheng Ping[1, 4, 5], Lei Bingbing[1, 4, 5, *], Guo Chongjin[1, 3], Zhang Jiangpeng[1, 4], Chen Feijie[2], Zhao Qifeng[1], Wang Kaiyuan[1, 3]

[1]Xinjiang Transportation Planning Survey and Design Institute Co., Ltd, Urumqi, China

[2]Xinjiang Transportation Investment (Group) Co., Ltd, Urumqi, China

[3]Xinjiang Key Laboratory for Safety and Health of Transportation Infrastructure in Alpine and High-Altitude Mountainous Areas, Xinjiang Transportation Planning, Survey and Design Institute Co., Ltd., Urumqi, China

[4]Xinjiang Engineering Research Center for Disaster Prevention and Control Technology of Mountain Transportation Infrastructure, Xinjiang Transportation Planning, Survey and Design Institute Co., Ltd., Urumqi, China

[5]College of Water Conservancy and Civil Engineering, Xinjiang Agricultural University, Urumqi, China

Abstract

Highway subgrade widening on soft ground is often accompanied by pronounced settlement and differential deformation due to the combined effects of construction conditions, moisture variation, and temperature changes. In this study, a two-dimensional plane-strain finite element model was developed in Abaqus to investigate the settlement response of a widened subgrade under varying compaction, moisture, and temperature conditions. The widening process was simulated through staged construction and incremental loading. The vertical displacement at the edge of the newly widened subgrade, together with the corresponding state variables, was extracted to construct a numerical dataset. Four regression algorithms, namely linear regression, random forest, XGBoost, and LightGBM, were then employed to develop surrogate models for reproducing the finite element settlement response. Model performance was evaluated using 10-fold cross-validation, with the coefficient of determination (R^2) and root mean square error (RMSE) adopted as evaluation metrics. The numerical results showed that surface settlement generally increased from the existing subgrade toward the widened side, with relatively large vertical displacement occurring near the edge of the new subgrade. Temperature effects were mainly concentrated within the subgrade body and the shallow foundation soil. Among the four machine-learning models, the random forest model achieved the best overall predictive performance, with a mean R^2 of 0.9828 and a mean RMSE of 0.0014 m. Furthermore, incorporating additional soil-state variables, including dry density, void ratio, and degree of saturation, improved the ability of the surrogate model to reproduce the numerical simulation results. These findings demonstrate that combining Abaqus-based numerical simulation with machine-learning surrogate modeling can effectively characterize the settlement response and parameter–response relationships of widened subgrades on soft ground within the investigated parameter ranges. The proposed framework can provide a useful reference for settlement-response analysis, identification of influential parameters, and efficient finite-element-assisted evaluation of similar subgrade-widening problems.

*Correspondence: Lei Bingbing (2265346728@qq.com)

Received: 19 October 2026; **Accepted:** 5 November 2026; **Published:** 25 November 2026

Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Keywords

Soft Ground, Subgrade Widening, Settlement Prediction, Abaqus Multiphysics Simulation, Machine Learning, Differential Settlement

1. Introduction

As existing expressways increasingly enter the reconstruction and expansion stage, subgrade widening has become an important engineering measure for increasing traffic capacity and extending the service life of transportation infrastructure. However, the additional fill load applied during widening alters the stress state of the existing subgrade and underlying foundation soil, inducing additional settlement on the widened side and pronounced differential deformation between the existing and newly constructed subgrades [27, 35]. This problem is particularly significant for subgrades constructed over soft ground, where the soil is generally characterized by high compressibility, low permeability, and pronounced consolidation behavior. Consequently, substantial settlement can develop under additional loading, further aggravating differential settlement between the existing and widened sections [31]. Therefore, clarifying the settlement response of highway subgrades during widening over soft ground is essential for understanding the deformation mechanisms of reconstruction and expansion projects and for optimizing engineering design.

Extensive studies have been conducted on the settlement and deformation behavior associated with subgrade widening over soft ground. Chen and Zhang [9] investigated the deformation characteristics of a pile-supported widened embankment over soft ground using centrifuge model tests and examined the effects of the existing embankment and pile foundation on deformation beneath the widened section. Wang and Huang [28] established a two-dimensional finite element model to investigate the stress redistribution, deformation, and differential settlement induced by embankment widening over soft ground. Chen et al. [10] developed a two-dimensional finite element approach incorporating the creep behavior of soft soils to investigate the evolution of consolidation settlement under embankment loading, highlighting the importance of time-dependent effects and soil compressibility in long-term settlement. Weng et al. [30] conducted full-scale model tests to investigate the deformation response of pavement structures subjected to differential settlement of widened embankments, providing experimental evidence for evaluating the effects of differential subgrade deformation on pavement performance.

With advances in environmental geotechnics and unsaturated soil mechanics, it has become increasingly recognized that soil deformation is governed not only by external loading but also by environmental factors such as pore-water conditions, matric suction, and temperature variations. Gens [15] emphasized the complex interactions between soils and their surrounding environment, noting that unsaturated conditions, suction variations, and thermal effects can modify soil structure and mechanical behavior and therefore need to be considered beyond the framework of conventional soil mechanics. Abed and Sołowski [3] developed a thermo-hydro-mechanical (THM) coupling framework for unsaturated soils, derived the corresponding governing equations, and implemented them using the finite element method. For highway widening projects over soft ground in regions experiencing substantial temperature variations, thermal effects may influence settlement by altering soil moisture migration and associated mechanical properties. Accordingly, incorporating the coupled effects of temperature, moisture transport, and stress–deformation behavior into conventional mechanical analyses can provide a more comprehensive understanding of subgrade settlement under complex environmental conditions.

Although THM numerical models are capable of describing the interactions among thermal, hydraulic, and mechanical fields, repeated finite element analyses can become computationally expensive when a large number of parameters and loading scenarios are considered. To improve the efficiency of parametric numerical analyses, finite-element-generated datasets have increasingly been combined with machine-learning methods to establish surrogate relationships between numerical model inputs and engineering responses [17, 18, 29]. Gao et al. [14] generated a large-scale finite element dataset in Abaqus for pile composite foundations subjected to embankment loading and employed a Transformer model to learn the relationships among geological conditions, embankment loading, pile parameters, and settlement. Abbasi et al. [2] generated a parametric finite element dataset using the iSight–Abaqus framework to investigate differential settlement between existing and widened road subgrades and compared the ability of several machine-learning algorithms to reproduce the settlement response. The integration of finite element simulation with machine-learning-based surrogate modeling has therefore emerged as a practical approach for reducing the computational cost associated with repeated simulations in multi-parameter analyses while providing rapid approximations of numerical input–output relationships.

Despite these advances, studies addressing the combined effects of non-freezing temperature, moisture conditions, and compaction state on the settlement response of widened highway subgrades over soft ground remain relatively limited. In

addition, the influence of different types of input variables on the representational capability of machine-learning surrogate models requires further investigation. Therefore, this study focuses on highway subgrade widening over soft ground in the cold and arid region of southern Xinjiang, China. A sequentially coupled thermo-hydro-mechanical finite element model accounting for non-freezing temperature effects is developed, and parametric numerical simulations are performed to generate settlement-response samples. Linear regression, random forest, XGBoost, and LightGBM are subsequently employed to construct surrogate models of the finite element responses. The effects of basic input variables, including water content, degree of compaction, and temperature, are further compared with those of soil-state variables, including dry density, void ratio, and degree of saturation, in terms of their influence on surrogate-model performance. The study focuses on the settlement response and surrogate-model characteristics within the prescribed model assumptions and parameter ranges, with the aim of providing a reference for multi-parameter numerical analysis and rapid estimation of settlement responses.

2. Materials and Methods

2.1. Theoretical Framework

This study considers highway subgrade widening over soft ground in the cold and arid region of southern Xinjiang, China. Because the subgrade is covered by pavement and its behavior is strongly influenced by moisture conditions, the present study focuses on the effects of temperature variations under non-freezing conditions on moisture migration and the mechanical response of the subgrade. Ambient temperature was prescribed as the thermal boundary condition, and the temperature field within the subgrade and underlying foundation was first obtained through heat-transfer analysis and subsequently transferred to the seepage–stress analysis. Accordingly, the thermal effects considered in this study are primarily represented by temperature-induced variations in hydraulic and mechanical properties under non-freezing conditions. The development of a freezing front, ice–water phase transition, frost heave, and thaw settlement are beyond the scope of the present study.

The deformation and stress state of the soil skeleton are governed by the effective stress principle. For unsaturated soils, the pore-water state influences the effective stress and consequently the mechanical response of the soil. A Bishop-type effective stress formulation was adopted, with the degree of saturation used to account for the influence of unsaturation on effective stress [20]:

$$\sigma' = \sigma - S_r u_w I \tag{1}$$

where σ' is the effective stress tensor, σ is the total stress tensor, S_r is the degree of saturation, u_w is the pore-water pressure, and I is the second-order identity tensor. Based on the principle of virtual work, the mechanical equilibrium equation under quasi-static conditions can be expressed as

$$\nabla \cdot \sigma + \rho_m g = 0 \tag{2}$$

where ∇ denotes the divergence operator, ρ_m is the average density of the soil mixture, and g is the gravitational acceleration vector. Together with the constitutive relationships, Eq. (2) describes the equilibrium among external loading, self-weight, and the internal stress state of the soil.

Moisture transport within the soil is governed by the law of mass conservation. Considering changes in pore volume caused by deformation of the soil skeleton as well as variations in fluid density, the fluid continuity equation is written as

$$\frac{\partial (n S_r \rho_w)}{\partial t} + \nabla \cdot (\rho_w v_w) = 0 \tag{3}$$

where n is the porosity, ρ_w is the density of water, t is time, and v_w is the seepage velocity.

To close the governing equations for unsaturated flow, the hydraulic characteristics of the unsaturated soil must be defined, including the nonlinear relationships between degree of saturation and matric suction and between relative permeability and degree of saturation. The van Genuchten model, which is widely used for numerical simulation of unsaturated flow, was adopted to describe these hydraulic characteristics [26]. The van Genuchten parameters were selected on the basis of previously reported experimental studies on unsaturated soils and adjusted according to the soil properties of the study region. The corresponding parameters were obtained from fitting soil–water characteristic curves. The relationships are expressed as

$$S_e = \frac{S_r - S_{\mathrm{res}}}{S_{\mathrm{sat}} - S_{\mathrm{res}}} = [1 + (\alpha \mid u_w \mid)^n]^{-m} \tag{4}$$

And

$$k_r(S_e) = S_e^{1/2} \left[1 - \left(1 - S_e^{1/m}\right)^m\right]^2 \tag{5}$$

where S_e is the effective degree of saturation; S_{sat} and S_{res} are the saturated and residual degrees of saturation, respectively; α, n, and m are fitting parameters, with $m = 1 - 1/n$; and k_r is the relative permeability.

The influence of temperature on the seepage field is primarily introduced through the temperature dependence of the dynamic viscosity of water. An increase in temperature reduces water viscosity and consequently enhances the saturated permeability of the soil, representing a conventional pathway for thermal–hydraulic interaction in THM analyses of unsaturated soils. In the present model, the saturated permeability was treated as a temperature-dependent parameter and combined with the relative permeability to obtain the effective permeability. The temperature–seepage relationship was established

based on the generalized Darcy law as

$$k(T) = k_0 \frac{\mu_0}{\mu(T)} k_r(S_e) \quad (6)$$

where the temperature dependence of water viscosity governs the variation in permeability.

Under the small-strain assumption, thermal expansion of the soil skeleton itself was neglected. The total strain increment was therefore decomposed into elastic and plastic components:

$$d\varepsilon = d\varepsilon^e + d\varepsilon^p \quad (7)$$

During the elastic stage, the elastic modulus was defined as a function of temperature to represent the influence of temperature variations on soil stiffness. Temperature-controlled triaxial experiments have demonstrated that temperature variations can affect the elastic modulus and other mechanical properties of soils [7].

Different constitutive models were adopted according to the mechanical characteristics of the individual soil layers. Owing to the high compressibility of the soft foundation soil, the Modified Cam-Clay (MCC) model was employed to characterize its elastoplastic behavior. The MCC model is a critical-state elastoplastic constitutive model capable of describing the stress–strain response, volumetric deformation, and strength characteristics of clayey soils [16]. In contrast, the compacted fill of the existing subgrade and the underlying stable soil layer exhibit stronger structural characteristics and are subjected to relatively lower stress levels. These materials were therefore represented using the widely adopted Mohr–Coulomb elastoplastic model, providing a balance between numerical applicability and the availability of model parameters.

To account for cumulative settlement induced by long-term cyclic traffic loading, an empirical cumulative plastic-strain model based on the Drucker–Prager criterion was introduced. Such models characterize the accumulation of plastic strain under repeated loading using stress invariants and have been applied to the prediction of permanent deformation in subgrade soils [33]. In the adopted formulation, ε^{acc} denotes the cumulative plastic strain, N is the number of loading cycles, J_2 is the second invariant of deviatoric stress, and I_1 is the first stress invariant. The remaining coefficients are regression parameters, while the stress-state-related variables are associated with the Mohr–Coulomb strength parameters of the material.

Heat transfer is governed by the principle of energy conservation. Assuming local thermal equilibrium among the soil particles, pore water, and pore air, the governing heat-conduction equation is expressed as

$$(\rho C)_{\text{eff}} \frac{\partial T}{\partial t} = \nabla \cdot (\lambda_{\text{eff}} \nabla T) \quad (8)$$

where T is the temperature field, $(\rho C)_{\text{eff}}$ is the equivalent volumetric heat capacity of the soil mixture, and λ_{eff} is the effective thermal conductivity tensor.

2.2. Sequential THM Coupling in Abaqus

Thermo-hydro-mechanical numerical simulations were performed using the Abaqus finite element platform to generate the dataset required for subsequent machine-learning analysis. Based on continuum mechanics and thermodynamics, both the subgrade fill and the underlying soft foundation soil were treated as unsaturated porous media. The numerical framework incorporates the mechanical equilibrium equation, fluid continuity equation, heat-conduction equation, and the corresponding constitutive relationships. Because environmental temperature changes occur over relatively long timescales and soil deformation develops slowly, the heat generated by mechanical deformation was assumed to be negligible compared with heat transfer driven by the external thermal environment. Therefore, the feedback of the mechanical field on the thermal field was neglected. A sequential coupling strategy consisting of one-way thermal transfer and two-way hydro-mechanical coupling was adopted to implement the THM analysis using the built-in analysis procedures and material-definition capabilities of Abaqus/Standard [12, 19, 22]. A time-dependent ambient-temperature boundary condition was applied to the subgrade surface, whereas the lateral and bottom boundaries of the model were assumed to be thermally insulated. After completion of the heat-transfer analysis, the nodal temperature results were imported as spatially and temporally varying predefined field variables into the subsequent soil-consolidation analysis. In this manner, the temperature field was transferred unidirectionally to the hydraulic and mechanical fields.

The effects of temperature on the hydraulic and mechanical behavior were incorporated through temperature-dependent material properties. In the Abaqus material module, the saturated permeability and elastic parameters were specified as functions of temperature, with the corresponding parameter values at different temperatures entered in tabular form in accordance with the adopted temperature–property relationships [11]. This approach enabled the nonlinear interaction among temperature, seepage, stress, and deformation to be represented within the sequential coupling framework.

2.3. Finite Element Model and Model Parameters

A two-dimensional plane-strain assumption was adopted for the numerical model. This simplification is appropriate because the longitudinal dimension of a highway subgrade is substantially greater than its transverse dimensions and has been widely adopted in numerical analyses of highway embankment widening [28]. The computational domain consisted of the asphalt pavement layers, the existing and newly widened subgrade fills, the underlying soft foundation soil, and the deeper stable soil layer.

The total width of the subgrade crest was 26 m, the fill height was 4 m, and the side slope was 1:1.5. Following general boundary-setting principles for geotechnical finite element analyses, the model was extended sufficiently in the horizontal direction to minimize the influence of lateral boundaries on the settlement field. In the vertical direction, the computational domain extended to the underlying stable stratum to reduce the influence of the bottom boundary. At the interface between the existing and widened subgrades, a stepped excavation geometry was adopted in accordance with the engineering design to reproduce the actual contact configuration between the two sections.

The construction process was simulated through staged filling and incremental loading. Soil elements were progressively activated over multiple analysis steps, and self-weight was applied to reproduce the stress evolution associated with layer-by-layer construction [4]. The newly widened subgrade was constructed in four filling layers. Including construction of the base and pavement surface layers, a total of 12 construction stages and 19 numerical analysis steps were considered. Self-weight of the subgrade was included throughout the analysis. According to the corresponding construction conditions, each soil layer was assigned an initial void ratio, initial geostatic stress, initial pore-water pressure, initial temperature, and initial degree of saturation. Horizontal displacements were constrained along the two lateral boundaries, while both horizontal and vertical displacements were constrained at the bottom boundary. Hydraulic and thermal boundary conditions were prescribed according to the environmental conditions represented in the model. To account for stress redistribution caused by the stiffness contrast between the existing and newly widened subgrades, a contact interaction was defined along their interface [1]. The mesh discretization and material assignment of the numerical model are shown in Figure 1.

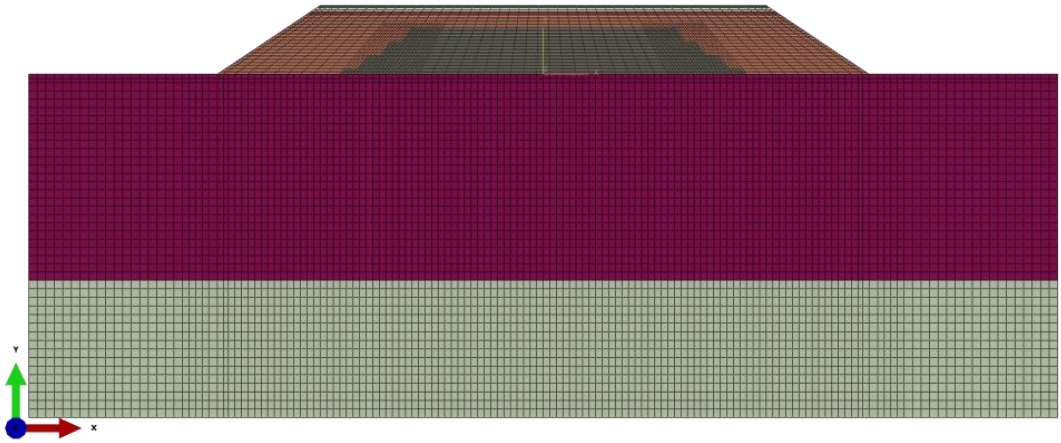


Figure 1. *Mesh discretization and material assignment of the finite element model.*

To represent different soil compaction states, the numerical model characterized the degree of compaction by adjusting physical-state parameters such as the initial void ratio. Considering the relatively dense state of the existing subgrade and foundation soil after long-term consolidation, the initial void ratio of the compacted soft foundation zone and the underlying stable soil layer was set to 0.47. The initial degree of saturation was assumed to vary gradually with depth, increasing from 0.90 near the ground surface to full saturation at the bottom of the model, corresponding to an initial water content of 15%. The loading conditions mainly consisted of the initial self-weight stress field and external loads. First, gravitational loading was applied according to the density of each soil layer to establish the initial geostatic equilibrium. The initial geostatic stress and pore-water pressure were assumed to increase linearly with depth. Subsequently, the element activation technique in Abaqus was adopted to simulate the staged construction process of the newly widened subgrade. Pavement structural loads and traffic loads were then applied, with the external load ranging from 17.6 to 20.58 kN/m^2. The material parameters required for the numerical model included mechanical, hydraulic, and thermophysical properties. The mechanical parameters mainly consisted of elastic modulus, Poisson's ratio, internal friction angle, and constitutive model parameters. The hydraulic parameters included saturated permeability and the van Genuchten soil-water characteristic parameters, while the thermophysical parameters included thermal conductivity and volumetric heat capacity. Temperature-dependent parameters were assigned according to the predefined temperature–property relationships. The material parameters were selected based on previous finite element studies of highway subgrade widening and soft-soil behavior, with appropriate adjustments considering the engineering characteristics of the studied region. The adopted parameters are summarized in Table 1.

Table 1. Material Parameters for Subgrade and Subsoil.

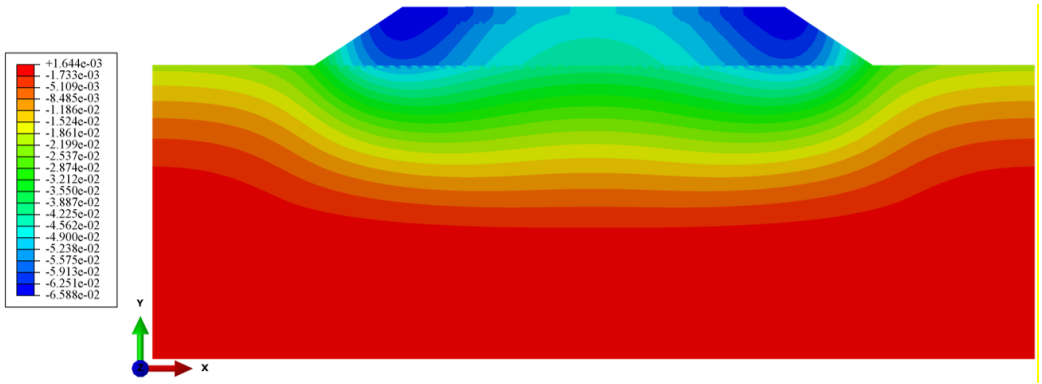
Material	E (MPa)	ν	φ (°)	k (m·d⁻¹)
Upper surface (wearing) course	1400	0.25	-	-
Middle surface course	1200	0.25	-	-
Lower surface course	1000	0.3	-	-
Existing subgrade soil	50	0.32	20	0.00004~0.0002
New subgrade soil	35	0.35	22	0.00004~0.0002
Soft subsoil layer	15	0.3	19	0.00004~0.0002
Stable soil layer	35	0.3	20	0.00006

2.4. Numerical Settlement Response and Dataset Construction

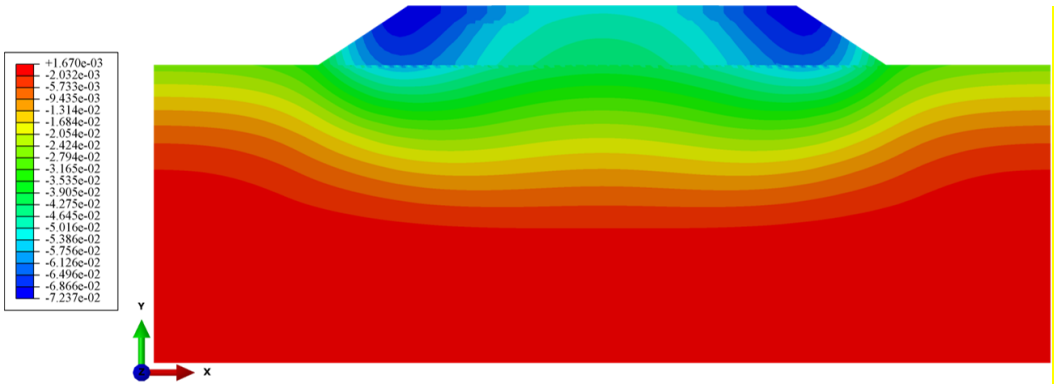
Figure 2 presents the distributions of vertical displacement in the widened subgrade and underlying foundation soil under different temperature conditions. Under the reference condition ($T = 0°C$), the region of relatively large settlement was mainly concentrated within the newly widened subgrade and the underlying soft foundation soil, while the magnitude of vertical displacement gradually decreased with increasing depth and horizontal distance.

Under the warming condition, the overall spatial pattern of settlement remained generally unchanged, whereas the settlement magnitude in the shallow region increased. Within the sequential coupling framework adopted in this study, an increase in temperature reduces the dynamic viscosity of water and consequently enhances soil permeability, thereby accelerating pore-water pressure dissipation and effective stress transfer. Meanwhile, changes in temperature-dependent mechanical parameters further affect the deformation response of the soil. Consequently, a greater degree of settlement developed within the same analysis period under the warming condition.

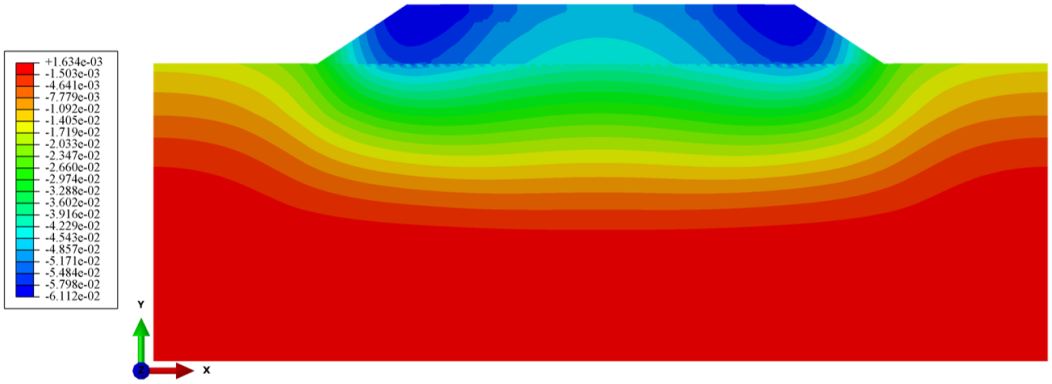
By contrast, under the lower-temperature condition, the reduced seepage capacity slowed pore-water pressure dissipation and consolidation deformation, resulting in a comparatively smaller degree of settlement development. These trends represent the numerical responses obtained under the temperature-dependent hydraulic and mechanical parameters and modeling assumptions adopted in this study.



(a) Simulated vertical settlement of the widened subgrade and foundation soil at $T = 0°C$.



(b) Simulated vertical settlement of the widened subgrade and foundation soil under the warming condition



(c) Simulated vertical settlement of the widened subgrade and foundation soil under the lower-temperature condition

Figure 2. *Simulated vertical settlement of the subgrade and foundation soil after widening under different temperature conditions.*

The surface vertical displacement profile of the widened subgrade is shown in Figure 3. Settlement generally increased from the centerline of the existing subgrade toward the newly widened section, with relatively large vertical displacement occurring at the outer edge of the widened subgrade.

Based on this spatial distribution, the vertical displacement at the edge of the newly widened subgrade was selected as the target response for subsequent machine-learning modeling. Combination scenarios were established by varying the degree of compaction, water content, and temperature, and the corresponding numerical results and soil-state variables were extracted from each simulation to construct a structured dataset. The resulting parameter combinations were used to systematically characterize settlement responses under different input conditions within the prescribed parameter ranges and to provide the data required for subsequent surrogate-model training, validation, and feature-contribution analysis.

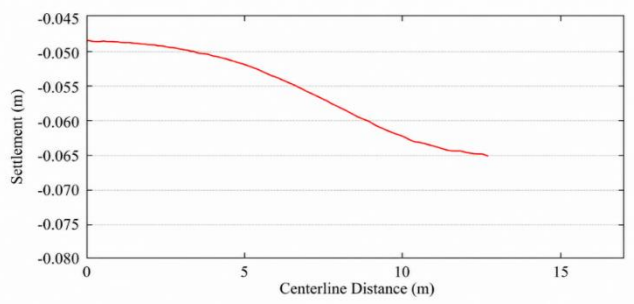


Figure 3. *Surface vertical settlement profile of the widened subgrade at $T = 0°C$ (x-axis: distance from the centerline of the existing subgrade).*

2.5. Machine-learning Models

To compare the capability of different machine-learning methods to reproduce the settlement responses obtained from the numerical simulations, four regression models were considered: linear regression, random forest (RF), XGBoost, and LightGBM. Among these models, random forest was selected as the primary surrogate model for subsequent analyses. Ran-

dom forest, originally proposed by Breiman [6], is an ensemble-learning method that constructs multiple decision trees through bootstrap sampling and random feature selection. The outputs of the individual decision trees are subsequently aggregated to generate the final prediction. This ensemble strategy reduces the sensitivity of the prediction to perturbations in individual training samples and thereby improves model generalization performance. Random forest has been applied to settlement prediction problems involving soft ground and soft-soil embankments. Sun et al. [23] developed a random-forest-based model for soft-ground settlement prediction using engineering data with multiple feature parameters. Chen et al. [8] compared random forest regression, XGBoost, and multilayer perceptron models for settlement prediction of a soft-soil embankment and found that random forest and XGBoost exhibited relatively favorable predictive accuracy and robustness in their case study. The basic architecture of the random forest model is illustrated in Figure 4.

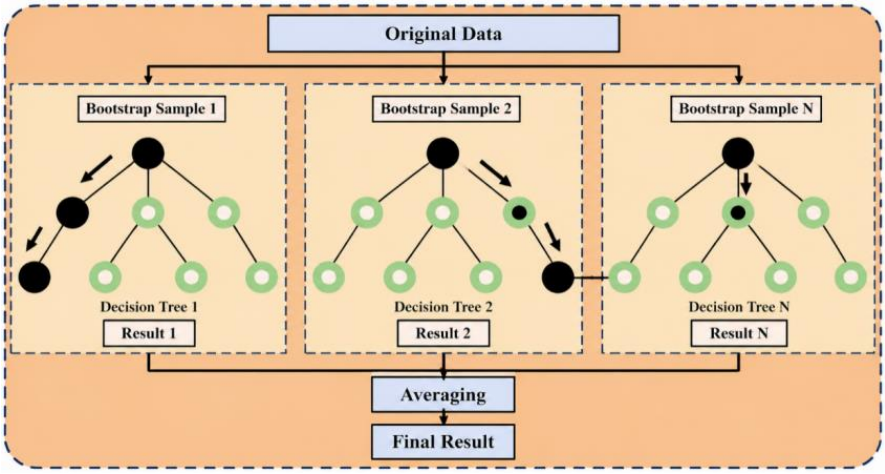


Figure 4. *Basic architecture of the random forest model.*

2.6. Cross-validation and Model Evaluation

To provide a comprehensive assessment of the generalization performance of the machine-learning models, k-fold cross-validation was employed for model validation and evaluation. In this study, $k = 10$was adopted. All samples were randomly divided into 10 approximately equal-sized subsets. During each cross-validation iteration, nine subsets were used for model training, while the remaining subset was used for testing. Model performance on the test subset was evaluated using the coefficient of determination (R^2) and root mean square error (RMSE).

The procedure of the 10-fold cross-validation adopted in this study is illustrated in Figure 5. For each model, the training and testing procedure was repeated 10 times so that each subset served as the test set once. The evaluation metrics obtained from the 10 test folds were then averaged to characterize the overall generalization performance of each model. The average $R^2$and RMSE values were subsequently used to compare the performance of the different regression models.

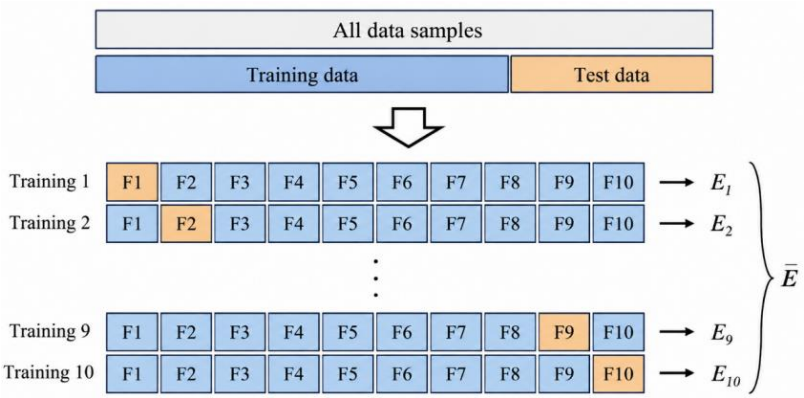
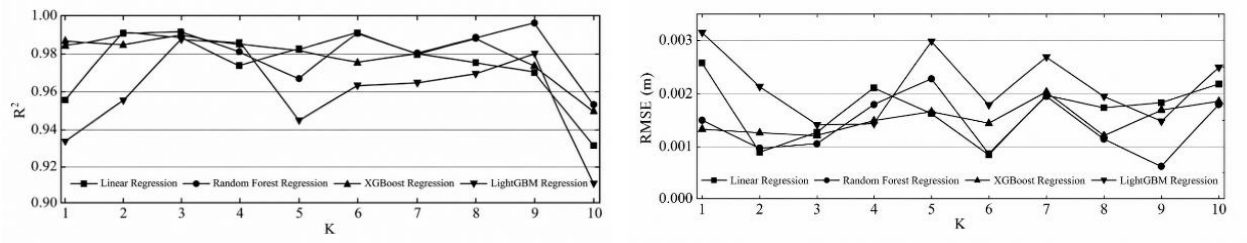


Figure 5. *Schematic representation of the 10-fold cross-validation procedure.*

3. Results and Discussion

3.1. Model Validation and Comparison

The results of the 10-fold cross-validation for the four machine-learning models are presented in Figure 6 and Table 2. In terms of the mean cross-validation metrics, the random forest model achieved the highest mean R^2 and the lowest mean RMSE, with a mean R^2 of 0.9828 and a mean RMSE of 0.0014 m. The XGBoost model yielded a mean R^2 of 0.9800 and a mean RMSE of 0.0015 m, showing an overall performance close to that of random forest. Linear regression achieved a mean R^2 of 0.9744 and a mean RMSE of 0.0017 m, whereas LightGBM yielded a mean R^2 of 0.9596 and a mean RMSE of 0.0022 m. These results indicate that, within the numerical dataset and parameter ranges considered in this study, all four models were capable of reproducing the relationship between the input parameters and the numerically calculated settlement response, with random forest exhibiting the best overall mean evaluation metrics. Because model performance was assessed using a single 10-fold cross-validation procedure, the following comparisons are based primarily on the mean evaluation metrics and do not imply statistical significance.



(a) Coefficient of determination (R^2) (b) Root mean square error (RMSE)

Figure 6. *Results of 10-fold cross-validation for the different machine-learning algorithms.*

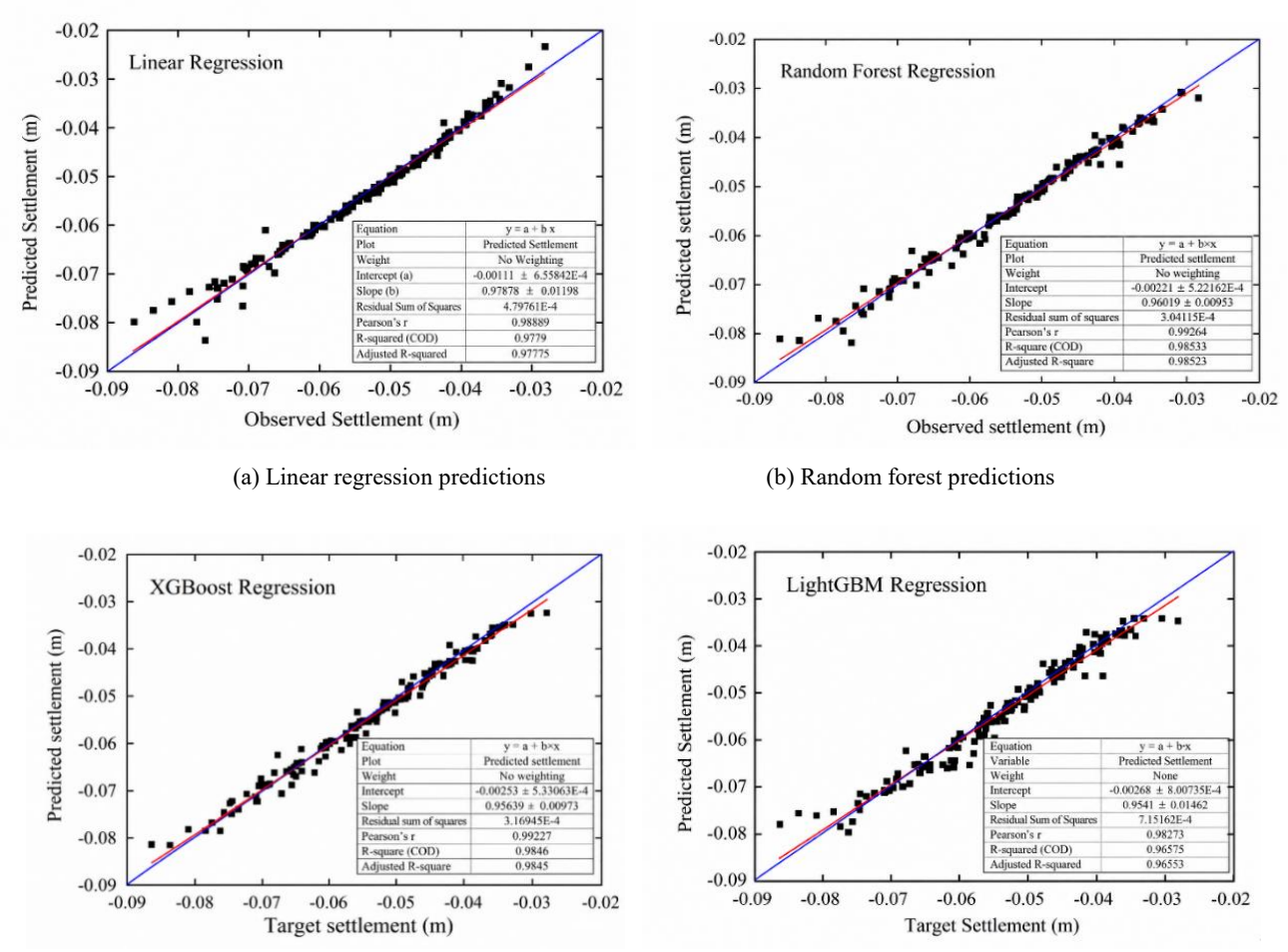
Table 2. *Results of 10-fold cross-validation and mean performance metrics for the different machine-learning algorithms.*

Fold	Linear regression		Random forest		XGBoost		LightGBM	
	R^2	RMSE (m)	R^2	RMSE (m)	R^2	RMSE (m)	R^2	RMSE (m)
1	0.9559	0.0026	0.9850	0.0015	0.9881	0.0013	0.9341	0.0032
2	0.9921	0.0009	0.9907	0.0010	0.9843	0.0013	0.9555	0.0021
3	0.9888	0.0013	0.9924	0.0011	0.9899	0.0012	0.9862	0.0014
4	0.9742	0.0021	0.9813	0.0018	0.9870	0.0015	0.9881	0.0014
5	0.9833	0.0016	0.9673	0.0023	0.9826	0.0017	0.9444	0.0030
6	0.9916	0.0009	0.9911	0.0009	0.9758	0.0014	0.9631	0.0018
7	0.9810	0.0020	0.9812	0.0020	0.9796	0.0020	0.9645	0.0027
8	0.9756	0.0017	0.9894	0.0011	0.9880	0.0012	0.9693	0.0020
9	0.9703	0.0018	0.9964	0.0006	0.9746	0.0017	0.9804	0.0015
10	0.9317	0.0022	0.9536	0.0018	0.9503	0.0019	0.9106	0.0025
Average	0.9744	0.0017	0.9828	0.0014	0.9800	0.0015	0.9596	0.0022

Random forest constructs multiple diverse decision trees through bootstrap sampling and random feature selection and reduces the prediction variance of individual trees through ensemble averaging, thereby generally providing favorable stability and generalization capability [5, 13]. Under the sample size and feature dimensionality considered in this study, the random forest model exhibited a strong ability to reproduce the numerical simulation results.

XGBoost also demonstrated a strong nonlinear fitting capability, and its mean performance was close to that of random

forest, although some variation was observed among the individual cross-validation folds. The performance metrics of linear regression were only moderately lower than those of the two ensemble-learning models. This suggests that, within the parameter ranges considered in this study, the overall relationship between the input variables and settlement at the edge of the widened subgrade exhibited a pronounced approximately linear component, allowing the linear model to explain a substantial proportion of the variation in the numerical results. In contrast, LightGBM exhibited greater variability in predictive performance across several cross-validation folds, with an R^2 of 0.9106 in the tenth fold. This behavior may be associated with the limited sample size and the sensitivity of model-complexity parameters to the partitioning of the training data. In particular, under the leaf-wise tree-growth strategy, an inadequate match between parameters such as the number of leaves and the minimum number of samples per leaf and the size of the dataset may lead to stronger fitting of local sample characteristics and consequently affect model stability across different validation subsets [21, 24, 32]. The relationship between the numerically calculated and model-predicted settlement values is further illustrated in Figure 7. Overall, the predictions from all four models were predominantly distributed close to the 1:1 reference line, indicating that the models were generally capable of reproducing the settlement trends contained in the numerical dataset. A comparison across different settlement ranges further showed that all four models exhibited varying degrees of underestimation when the settlement exceeded 0.07 m, accompanied by an increase in prediction error. The number of samples within this high-settlement range was relatively limited, which may have constrained the ability of the models to adequately learn the characteristics of large-settlement cases. In addition, more complex interactions among the influencing factors may occur within this range, thereby increasing the difficulty of model fitting. The underestimation was relatively more pronounced for linear regression and LightGBM, whereas random forest and XGBoost provided comparatively better fits for the large-settlement samples. Considering both the 10-fold cross-validation results and the predictive performance across different settlement ranges, random forest achieved the most favorable overall mean evaluation metrics within the numerical dataset and parameter ranges considered in this study. It was therefore selected as the candidate surrogate model for the subsequent parameter sensitivity analysis and rapid estimation of numerical settlement responses.



(a) Linear regression predictions

(b) Random forest predictions

(c) XGBoost predictions

(d) LightGBM predictions

Figure 7. *Scatter plots of numerically calculated and predicted settlement values for the four machine-learning models.*

3.2. Feature Importance Analysis

To investigate the relative contribution of different input variables to the vertical displacement response, the feature importance obtained from the random forest model was used to evaluate the influence of each input variable. Feature importance reflects the contribution of individual variables to the model's identification of variations in vertical displacement [25]. For variables with physical or mathematical relationships, the importance values may also reflect information-sharing characteristics among correlated features. Therefore, the feature importance results were interpreted in combination with the physical meanings of the parameters and their relationships.

The feature importance results are presented in Figure 8. The importance of the input variables exhibited clear differences. Among all features, the degree of saturation (S_r) showed the highest importance, followed by the degree of compaction (K). Water content (w) also contributed to the model response, whereas dry density (ρ_d), void ratio (e), and temperature (T) exhibited relatively lower individual importance values. Overall, parameters associated with moisture conditions and soil compaction state constituted the primary information sources for the random forest model in reproducing vertical displacement variations. The degree of saturation represents the extent to which pore space is occupied by water. Its variation reflects both the water-storage condition and pore-structure characteristics of the soil and is closely associated with unsaturated seepage behavior and effective stress response. Therefore, S_r exhibited a relatively high contribution within the surrogate model. The degree of compaction reflects the density state formed during construction and directly influences pore structure, soil stiffness, and compressibility. Water content represents the overall moisture condition of the soil. Previous studies have demonstrated that variations in water conditions and compaction state can significantly affect the permanent deformation characteristics of compacted subgrade soils [34], which is consistent with the feature importance results obtained in this study.

Dry density (ρ_d) and void ratio (e) also describe the soil compaction state; however, both variables are closely related to the degree of compaction. When the maximum dry density ($\rho_{d,\max}$) is determined, the relationship can be expressed as:

$$\rho_d = K\rho_{d,\max} \tag{9}$$

Furthermore, under a given particle density condition, the void ratio can be derived from dry density. Therefore, K, ρ_d, and e describe the soil density state from different perspectives, and part of their information is shared within the machine-learning model. The relatively lower individual importance values of ρ_d and e therefore mainly reflect the distribution of overlapping density-state information among correlated features rather than a lack of physical significance. The temperature variable (T) exhibited the lowest feature importance, indicating that within the non-freezing temperature range considered in this study, its contribution to vertical displacement was lower than that of moisture-related and compaction-related parameters. This result is consistent with the sequential coupling framework adopted in this study, in which temperature mainly influences seepage–deformation behavior by modifying hydraulic parameters and related mechanical properties. Because ice–water phase transition, frost heave, and thaw settlement were not considered, temperature acted as a secondary coupling factor in the investigated conditions.

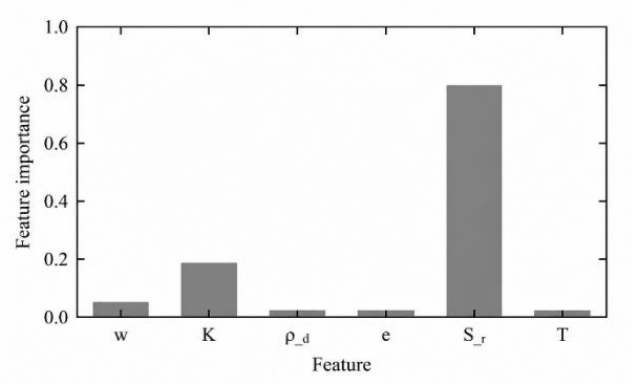


Figure 8. *Feature importance of the input variables.*

3.3. Effect of Input Feature Selection on Surrogate-model Performance

The random forest model described above used six input variables, including water content (w), degree of compaction (K), temperature (T), dry density (ρ_d), void ratio (e), and degree of saturation (S_r). Among these variables, w, K, and T can be regarded as basic parameters representing construction and environmental conditions, whereas ρ_d, e, and S_r further describe the soil density state, pore structure, and water-filling condition, respectively. To investigate the influence of different input-variable representations on surrogate-model performance, a three-feature random forest model was constructed using only w, K, and T as input variables. The model was evaluated using the same 10-fold cross-validation procedure as the six-feature model. The corresponding results are presented in Figure 9. The results show that the three-feature model achieved a mean R^2 of 0.9256 and a mean RMSE of 0.00294 m. Although the model maintained an acceptable fitting performance across different validation folds, some variation remained among different data partitions. Compared with the six-feature random forest model, the mean R^2 decreased from 0.9828 to 0.9256, while the mean RMSE increased from 0.0014 m to 0.00294 m. These results indicate that using only water content, degree of compaction, and temperature enables the model to capture the general variation

trend of vertical displacement within the numerical dataset; however, introducing additional soil-state variables significantly improves the ability of the model to reproduce the finite element responses. The difference in performance between the two input schemes mainly reflects the distinction between basic parameters and soil-state variables in representing the physical condition of the soil. Degree of compaction and water content describe the overall construction compaction level and moisture condition, whereas dry density, void ratio, and degree of saturation provide more direct descriptions of soil density structure, pore characteristics, and pore-water conditions. Although these two groups of parameters exhibit clear physical relationships, explicitly incorporating soil-state variables allows the random forest model to directly utilize information related to density state, pore structure, and water filling conditions during decision-tree partitioning. This enhances the identification of nonlinear relationships between input variables and vertical displacement. The results are consistent with the feature importance analysis in Section 3.2, where the degree of saturation (S_r) exhibited a relatively high contribution in the six-feature model. This indicates that transforming basic construction and moisture parameters into more direct soil-state descriptors can improve the surrogate model's ability to reproduce finite element settlement responses.

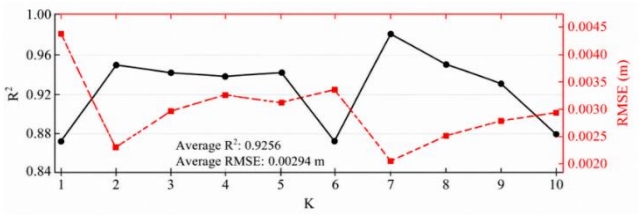


Figure 9. Results of 10-fold cross-validation for the random forest model with limited input features.

Therefore, the three-feature model demonstrates reasonable response characterization capability under simplified input conditions, whereas the six-feature model achieves improved performance by incorporating physically meaningful state variables. Within the numerical dataset and parameter ranges considered in this study, combining basic construction/environmental parameters with soil-state parameters provides a more comprehensive representation of the current subgrade condition and improves the fitting capability of the random forest surrogate model for vertical displacement responses.

These results indicate that, in finite-element-driven surrogate modeling, model performance is influenced not only by the number of input variables but also by the physical representation of the selected features and their relationship with the target response.

4. Conclusions

In this study, a sequentially coupled thermo-hydro-mechanical finite element model considering non-freezing temperature effects was developed for highway subgrade widening over soft ground in the cold and arid region of southern Xinjiang. Based on the parameterized numerical simulations, machine-learning surrogate models were established to investigate settlement responses and the influence of different input features. The main conclusions are summarized as follows:

1) After subgrade widening, relatively large settlements were mainly concentrated within the newly widened subgrade and the underlying soft foundation soil. The surface vertical displacement generally increased from the existing subgrade toward the widened side, with pronounced settlement occurring near the edge of the newly widened subgrade. Under the non-freezing temperature conditions considered in this study, temperature effects on settlement response were weaker than those associated with moisture conditions and compaction state.
2) All four machine-learning models exhibited relatively high fitting capability for reproducing the numerical settlement responses. Among them, the random forest model achieved the best overall performance, with a mean R^2 of 0.9828 and a mean RMSE of 0.0014 m based on 10-fold cross-validation. XGBoost showed comparable performance, while linear regression ranked second, and LightGBM exhibited relatively larger variations among some validation folds.
3) The feature importance analysis indicated that the contribution of the input variables followed the order: degree of saturation (S_r), degree of compaction (K), water content (w), dry density (ρ_d), void ratio (e), and temperature (T). The three-feature random forest model using w, K, and T achieved a mean R^2 of 0.9256 and a mean RMSE of 0.00294 m. After incorporating dry density, void ratio, and degree of saturation, the model performance improved, with the mean R^2 increasing to 0.9828 and the mean RMSE decreasing to 0.0014 m. These results demonstrate that incorporating physically meaningful soil-state variables can enhance the ability of surrogate models to characterize finite element settlement responses.

Abbreviations

LR	Linear Regression
Light GBM	Light Gradient Boosting Machine
MCC	Modified Cam-Clay
R^2	Coefficient of Determination
RF	Random Forest
RMSE	Root Mean Square Error
THM	Thermo-Hydro-Mechanical
XG Boost	Extreme Gradient Boosting

Author Contributions

Zheng Ping: Conceptualization, Funding acquisition, Project administration, Writing – original draft, Resources Lei

Lei Bingbing: Methodology, Supervision, Validation, Writing review & editing

Guo Chongjin: Software, Formal analysis, Data curation

Zhang Jiangpeng: Investigation, Visualization

Chen Feijie: Funding acquisition, Resources

Zhao Qifeng: Data curation, Investigation

Wang Kaiyuan: Software, Visualization

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Abbasi, S. J., Weng, X., and Iqbal, M. J. (2024). Numerical analysis of differential settlement in road due to widening considering different reinforcement techniques. Appl. Sci. 14, 1740. https://doi.org/10.3390/app14051740

[2] Abbasi, S. J., Minqjie, H., Weng, X., et al. (2026). Quantitative prediction of differential settlement based on machine learning techniques. Sci. Rep. https://doi.org/10.1038/s41598-026-60671-y

[3] Abed, A. A., and Sołowski, W. T. (2017). A study on how to couple thermo-hydro-mechanical behaviour of unsaturated soils: physical equations, numerical implementation and examples. Comput. Geotech. 92, 132–155. https://doi.org/10.1016/j.compgeo.2017.07.021

[4] Badarinath, R., and El Naggar, H. (2021). Effects of construction sequence and pace on the stability of high embankments founded on soft marine clay. Transp. Geotech. 30, 100635. https://doi.org/10.1016/j.trgeo.2021.100635

[5] Belgiu, M., and Drăguţ, L. (2016). Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

[6] Breiman, L. (2001). Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

[7] Cekerevac, C., and Laloui, L. (2004). Experimental study of thermal effects on the mechanical behaviour of a clay. Int. J. Numer. Anal. Methods Geomech. 28, 209–228. https://doi.org/10.1002/nag.332

[8] Chen, D., Wu, Y., Liu, J., Wu, H., and Ren, Y. (2024a). Field observation and settlement prediction study of a soft soil embankment under rolling dynamic compaction. Appl. Sci. 14, 6454. https://doi.org/10.3390/app14156454

[9] Chen, T., and Zhang, G. (2024b). Centrifuge modeling of pile-supported embankment on soft soil base for highway widening. Soils Found. 64, 101422. https://doi.org/10.1016/j.sandf.2023.101422

[10] Chen, Z. J., Feng, W. Q., Yin, J. H., and Shi, X. S. (2023). Finite element model and simple method for predicting consolidation displacement of soft soils exhibiting creep underneath embankments in 2-D condition. Acta Geotech. 18, 2513–2528. https://doi.org/10.1007/s11440-022-01741-z

[11] Dassault Systèmes Simulia Corp. (2025). Abaqus 2025 Documentation: Linear Elastic Behavior and Permeability in Abaqus/Standard. Available at: https://help.3ds.com [Accessed October 2026].

[12] Fei, K., and Dai, D. (2018). Experimental and numerical study on the behavior of energy piles subjected to thermal cycles. Adv. Civ. Eng. 2018, 3424528. https://doi.org/10.1155/2018/3424528

[13] Friedrich, S., Groll, A., Ickstadt, K., Kneib, T., Pauly, M., Rahnenführer, J., and Friede, T. (2023). Regularization approaches in clinical biostatistics: a review of methods and their applications. Stat. Methods Med. Res. 32, 425–440. https://doi.org/10.1177/09622802221133557

[14] Gao, S., Chen, C., Jiang, X., Zhu, S., Cai, H., and Li, W. (2024). Transformer-based settlement prediction model of pile composite foundation under embankment loading. Comput. Geotech. 176, 106783. https://doi.org/10.1016/j.compgeo.2024.106783

[15] Gens, A. (2010). Soil–environment interactions in geotechnical engineering. Géotechnique 60, 3–74. https://doi.org/10.1680/geot.9.P.109

[16] Graham, J., Tanaka, N., Crilly, T., and Alfaro, M. (2001). Modified Cam-Clay modelling of temperature effects in clays. Can. Geotech. J. 38, 608–621. https://doi.org/10.1139/t00-125

[17] He, X., Xu, H., Sabetamal, H., and Sheng, D. (2020). Machine learning aided stochastic reliability analysis of spatially variable slopes. Comput. Geotech. 126, 103711. https://doi.org/10.1016/j.compgeo.2020.103711

[18] Liu, C., Wang, Z., Liu, H., Cui, J., Huang, X., Ma, L., and Zheng, S. (2024). Prediction of surface settlement caused by synchronous grouting during shield tunneling in coarse-grained soils: a combined FEM and machine learning approach. Undergr. Space 16, 206–223. https://doi.org/10.1016/j.undsp.2023.10.001

[19] Lu, H., Wang, C., Bouazza, A., Ding, X., Ou, Q., Hu, X., and Kong, G. (2026). Field-validated thermo-mechanical behavior of energy pile groups in high-altitude regions. Geothermics 140, 103725. https://doi.org/10.1016/j.geothermics.2026.103725

[20] Lu, N., Godt, J. W., and Wu, D. T. (2010). A closed-form equation for effective stress in unsaturated soil. Water Resour. Res. 46, W05515. https://doi.org/10.1029/2009WR008646

[21] Park, D. J., Park, M. W., Lee, H., Kim, Y. J., Kim, Y., and Park, Y. H. (2021). Development of machine learning model for diagnostic disease prediction based on laboratory tests. Sci. Rep. 11, 7567. https://doi.org/10.1038/s41598-021-87171-5

[22] Sterpi, D., Coletto, A., and Mauri, L. (2017). Investigation on the behaviour of a thermo-active diaphragm wall by thermo-mechanical analyses. Geomech. Energy Environ. 9, 1–20. https://doi.org/10.1016/j.gete.2016.10.001

[23] Sun, C., Yu, T., Li, M., Wei, H., and Tan, F. (2024). Study on settlement prediction of soft ground considering multiple feature parameters based on ISSA-RF model. Sci. Rep. 14, 5067. https://doi.org/10.1038/s41598-024-55236-w

[24] Sun, X. (2025). Application of an improved LightGBM hybrid integration model combining gradient harmonization and Jacobian regularization for breast cancer diagnosis. Sci. Rep. 15, 2569. https://doi.org/10.1038/s41598-025-86014-x

[25] Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., and Zeileis, A. (2008). Conditional variable importance for random forests. BMC Bioinformatics 9, 307. https://doi.org/10.1186/1471-2105-9-307

[26] Van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. 44, 892–898. https://doi.org/10.2136/sssaj1980.03615995004400050002x

[27] Wang, C., Xie, L., Liu, Z., Wu, M., Zhang, T., Cai, G., and Liu, S. (2024). Study on settlement deformation law of new and old subgrade of expressway reconstruction and expansion based on CPTU. Transp. Geotech. 49, 101392. https://doi.org/10.1016/j.trgeo.2024.101392

[28] Wang, H., and Huang, X. M. (2011). Stress and deformation due to embankment widening with different treatment techniques. J. Cent. South Univ. Technol. 18, 1304–1310. https://doi.org/10.1007/s11771-011-0837-9

[29] Wang, T., Chen, W., Li, T., Connolly, D. P., Luo, Q., Liu, K., and Zhang, W. (2023). Surrogate-assisted uncertainty modeling of embankment settlement. Comput. Geotech. 159, 105498. https://doi.org/10.1016/j.compgeo.2023.105498

[30] Weng, X. L., Zhu, H. H., Chen, J. X., Liang, D. P., Shi, B., and Zhang, C. C. (2015). Experimental investigation of pavement behavior after embankment widening using a fiber optic sensor network. Struct. Health Monit. 14, 46–56. https://doi.org/10.1177/1475921714548935

[31] Wu, J., Xuan, Y., Deng, Y., Li, X., Zha, F., and Zhou, A. (2021). Combined vacuum and surcharge preloading method to improve Lianyungang soft marine clay for embankment widening project: a case. Geotext. Geomembr. 49, 452–465. https://doi.org/10.1016/j.geotexmem.2020.10.013

[32] Wu, T., Zhang, W., Jiao, X., Guo, W., and Hamoud, Y. A. (2020). Comparison of five boosting-based models for estimating daily reference evapotranspiration with limited meteorological variables. PLoS ONE 15, e0235324. https://doi.org/10.1371/journal.pone.0235324

[33] Yesuf, G. Y., and Hoff, I. (2015). Finite element modelling for prediction of permanent strains in fine-grained subgrade soils. Road Mater. Pavement Des. 16, 392–404. https://doi.org/10.1080/14680629.2015.1013053

[34] Zhang, J., Zhang, A., Li, J., and Fan, H. (2024). Enhanced understanding on permanent deformation behaviour of subgrade compacted clay under long-term cyclic loading. Soil Dyn. Earthq. Eng. 187, 108972. https://doi.org/10.1016/j.soildyn.2024.108972

[35] Zheng, G., Wang, J., Zhou, H., Xia, B., and Bui, P. D. (2025). Settlement characteristics and evaluation approach of embankment widening over soft clay. Can. Geotech. J. 62, 1–10. https://doi.org/10.1139/cgj-2024-0413