buildings

MDPI

Article

Prediction of the Unconfined Compressive Strength of Salinized Frozen Soil Based on Machine Learning

Huiwei Zhao [1,2] and Hui Bing [1,2,*]

1 State Key Laboratory of Frozen Soil Engineering, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China; zhaohuiwei@nieer.ac.cn
2 University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: binghui@lzb.ac.cn

Abstract: Unconfined compressive strength (UCS) is an important parameter of rock and soil mechanical behavior in foundation engineering design and construction. In this study, salinized frozen soil is selected as the research object, and soil GDS tests, ultrasonic tests, and scanning electron microscopy (SEM) tests are conducted. Based on the classification method of the model parameters, 2 macroscopic parameters, 38 mesoscopic parameters, and 19 microscopic parameters are selected. A machine learning model is used to predict the strength of soil considering the three-level characteristic parameters. Four accuracy evaluation indicators are used to evaluate six machine learning models. The results show that the radial basis function (RBF) has the best UCS predictive performance for both the training and testing stages. In terms of acceptable accuracy and stability loss, through the analysis of the gray correlation and rough set of the three-level parameters, the total amount and proportion of parameters are optimized so that there are 2, 16, and 16 macro, meso, and micro parameters in a sequence, respectively. In the simulation of the aforementioned six machine learning models with the optimized parameters, the RBF still performs optimally. In addition, after parameter optimization, the sensitivity proportion of the third-level parameters is more reasonable. The RBF model with optimized parameters proved to be a more effective method for predicting soil UCS. This study improves the prediction ability of the UCS by classifying and optimizing the model parameters and provides a useful reference for future research on salty soil strength parameters in seasonally frozen regions.

Keywords: machine learning model; sulfuric salinized frozen soil; unconfined compressive strength; macro–meso–micro three-level characteristic parameter

check for updates

Citation: Zhao, H.; Bing, H. Prediction of the Unconfined Compressive Strength of Salinized Frozen Soil Based on Machine Learning. *Buildings* **2024**, *14*, 641. https://doi.org/10.3390/buildings14030641

Academic Editor: Eugeniusz Koda

Received: 9 November 2023
Revised: 16 January 2024
Accepted: 18 January 2024
Published: 29 February 2024

1. Introduction

Cold and severe cold regions account for approximately 75% of the total land area in China [1,2], and over 66% of these regions are salinized [3]. The overlapping part of these two regions is referred to as the seasonally salinized frozen soil area. These areas are widely distributed in western, northern and northeastern China [4]. With the increasing demand for resource development and infrastructure construction in the above-mentioned areas, research on the physical and mechanical properties of salinized frozen soil has begun. Additional requirements have been put forward for the bearing capacity of foundation soil layers and the stability of embankments, foundation pits, and natural soil slopes [5], especially UCS [6,7], such as how to obtain UCS quickly and accurately. At the same time, the artificial freezing method is used for reinforcement in offshore and submarine engineering construction [8]. In the event of a sudden water inrush in a formation rich in high salinity water, liquid nitrogen freezing is used to rescue the emergency, quickly forming a freezing curtain to stop the water [9]. In these engineering practices, UCS is usually used as an index for engineering performance evaluation. Therefore, UCS has extensive demand in engineering practice and is an important parameter to pay attention to in the design and construction of related geotechnical engineering. Studying the UCS of saline soil under freeze–thaw cycles is of great significance to improving the stability of related engineering construction.

The freeze–thaw cycle and salinity characteristics are important factors affecting the UCS of salinized frozen soil. The effect of freeze–thaw cycles on the physical and mechanical properties of saline soil has been reported in a large number of studies; that is, repeated freeze–thaw cycles change the soil structure and cause strength damage [10–12]. Salinity characteristics are mainly expressed in the form of salt type and content [13].

Ultrasonic testing is a test method that indirectly reflects the meso physical and mechanical characteristics of soil through transmission signal changes. The inhomogeneity of the medium inside the soil causes the attenuation of the acoustic energy. Thus, ultrasonic waves can be used to resolve structural defects such as cracks and holes in the soil [14,15], evaluate the mechanical properties and stability [16–19], and reflect the meso properties of soil features. However, in current research, only a small number of ultrasound parameters are usually introduced as the research scope, and there are problems such as the incomplete utilization of parameters and a small overall number of parameters.

The mechanical properties of porous materials largely depend on their micro pore structure [20–22]. Therefore, exploring the correlations between pore structure parameters and macro behavior is of great significance for understanding the macro mechanical properties of porous materials [23]. To date, the rapid development of micro technology has greatly improved the micro study of soil pore characteristics. For example, through scanning electron microscopy technology, the geometric shape and size of soil pores can be directly observed [24–28], and quantitative statistical analysis can be carried out after combining this technique with relevant image analysis software (Particles (Pores) and Cracks Analysis System (PCAS) V2.3; Image-ProPlus 5.1; etc.). However, there are few studies on the quantitative relationships between micro characteristic parameters, macro mechanical indicators, and related data-driven models.

The traditional method used to determine the UCS of salinized frozen soil is the uniaxial compression test. Although the results are accurate, the sample preparation period is long, and operating the equipment is very time-consuming. Therefore, the method is often limited in engineering practice. Machine learning can be used to analyze large amounts of data from various sources to achieve a comprehensive prediction of the output results [29], and compared with experimental methods, it has many advantages, such as high accuracy, high speed, and low cost [30]. Therefore, a large number of machine learning models have been developed and used in the field of geotechnical engineering in the past three decades, including ANNs, SVMs, LSTM, CNNs, and GANs [31]. Table 1 lists the applications of some machine learning models in soil UCS prediction. However, there are few reports on UCS prediction of salinized frozen soil. At the same time, in current research, there are problems such as only using of a single type of parameter, inputting a small number of parameters into a model, and considering more macro factors and less combinations of soil meso and micro parameters.

Therefore, in this study, a machine learning model was used to predict the UCS of salinized frozen soil. Unlike previous machine learning models that only considered macro parameters as input, in this paper, the model parameter classification method is applied, and macro, meso, and micro factors are considered. The three-level characteristic parameters, i.e., the macro, meso, and micro parameters, are obtained through experiments and used as input parameters, and relevant data-driven models are constructed based on six machine learning models to obtain a multiscale comprehensive prediction of soil UCS. Through the analysis of the prediction accuracy, stability, and parameter sensitivity of the optimal model, the prediction performance of the machine learning model for salinized frozen soil UCS driven by the three-level characteristic parameters is explored to provide a new reference for further improving the prediction ability of a model for UCS.

Table 1. Soil UCS prediction using an artificial intelligence model.

Model	Input Parameters	Soil Type	Performance	Reference
BP; PSO-BP	5	Treated fibrous peat soil	BP: R = 0.928; MSE = 2.14 PSO-BP: R = 1.000; MSE = 0.73	Dehghanbanadaki et al. [32]
REG; RDF; BLR	5	Soil improvement such as OPT	REG: R^2 = 0.91; RMSE = 0.31 RDF: R^2 = 0.89; RMSE = 0.34 BLR: R^2 = 0.91; RMSE = 0.31	Eyo and Abbey. [33]
ANN	8	Cohesive soils stabilized with geopolymer	R^2 = 0.9808; RMSE = 0.8808	Ngo et al. [34]
KNN; XGB	4	Marl soil treated by cement and lignosulfonate	KNN: R^2 = 0.811; RMSE = 151.408 XGB: R^2 = 0.954; RMSE = 74.878	Shafiei et al. [35]
RF	7	Geopolymer stabilized clayey soil	R^2 = 0.9757; RMSE = 0.9815	Zeini et al. [36]
GB-ML	8	TCEF-soils	R^2 = 0.900; RMSE = 0.335	Eyo et al. [37]
GB-PSO	12	Australia-EBCA-soils	R^2 = 0.9655; RMSE = 0.1633	Tran. [38]
BAS-BP	5	CPF soils	R = 0.9594; RMSE = 0.1727	Zhang et al. [39]

Note: R: Pearson's correlation coefficient; R^2: determination coefficient; RMSE: root mean square error; MSE: mean square error; BP: back-propagation; REG: multiple linear regression; RDF: random decision forest; BLR: Bayesian linear regressor; XGB: extreme gradient boosting; KNN: k-nearest neighbor; RF: random forest; TCEF Soils: soils treated with calcium-based additives blended with eco-friendly pozzolans; GB-ML: machine learning using the gradient boosting; Australia-EBCA-Soils: earth building sites in Canberra, Australia; GB-PSO: gradient boosting machines-particle swarm optimizer; CPF Soils: cement stabilized soil incorporating solid waste and propylene fiber; BAS-BP: beetle antennae search BP.

The remainder of this article is described below. Section 2, " Experiment ", introduces the test soil samples and basic characteristics, basic test methods, and test results in detail. This creates the premise for the subsequent formulation of the basic hypothesis and basic methods of the methodology. Section 3, "Methodology", first proposes the basic hypothesis, the three-level characteristic interaction hypothesis and the basic method, the new method of parameter expansion classification. Then, based on this, three-level characteristic parameters of macro–meso–micro were constructed, and a model data set was created to provide parameters and data sets for the subsequent methodology to be applied to the model. Section 4, "Methodology applied to models", describes six machine learning models to which the methodology is specifically applied. Four statistical indices are used for model performance evaluation, model analysis process, and hyperparameter optimization. Section 5, "Results and discussion", shows the performance evaluation of six machine learning models for UCS prediction before and after parameter optimization. The section also contains sensitivity analysis of individual parameters of optimal models, sensitivity analysis of the third-level parameters of the model, and model comparison and limitation analysis. Section 6, "Conclusions and summary", gives the main conclusions.

2. Experiment

The overall experimental process is shown in Figure 1. The wave speed, SEM images, and UCS are obtained through experimental methods, and the interaction between them is analyzed.

2.1. Basic Physical Properties of Soil Samples and Sample Preparation

The test soil was obtained from Lanzhou, China. After soil is collected, desalination is carried out first. The specific process refers to the desalination treatment of loess-like saline soil provided by Hui Bing et al. [40]. After desalination, air-dry and pass through a 2 mm sieve for later use. The soil ion content before and after desalination is shown in Table 2, which meets the relevant test requirements [4,41]. The particle size distribution and basic physical properties of the soil samples after desalination are shown in Figure 2 and Table 3, respectively. The test soil sample is silty clay (ASTM D2487-17 (2020) [42]).

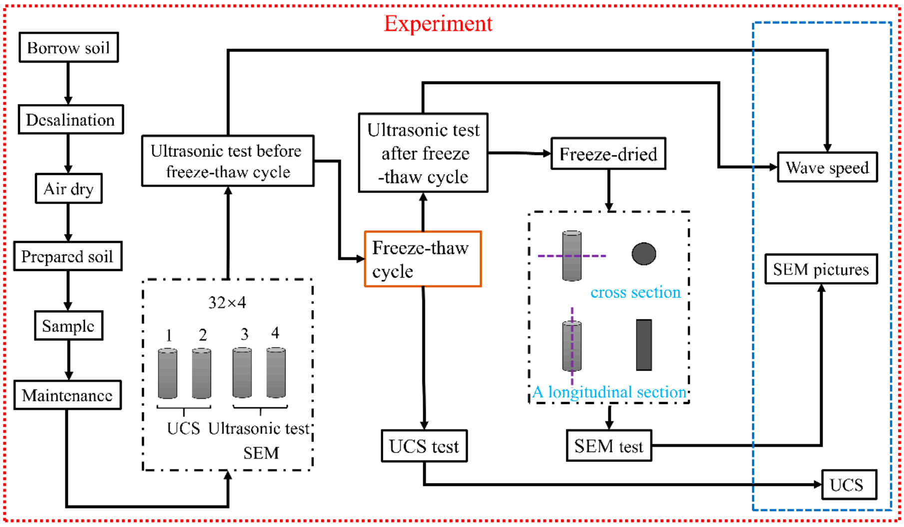

Figure 1. Schematic diagram of the experimental research path.

Table 2. Ion content of the soil samples before and after desalination.

	Cation Content/%				Anion Content/%			Total Salt Content/%
	Na^+	K^+	Mg^{2+}	Ca^{2+}	Cl^-	$SO_4{}^{-2}$	$NO_3{}^-$	
Before	0.2355	0.0036	0.0238	0.1624	0.1144	0.7756	0.0020	1.3173
After	0.0058	0.0013	0.0022	0.0142	0.0021	0.0095	0.0001	0.0352

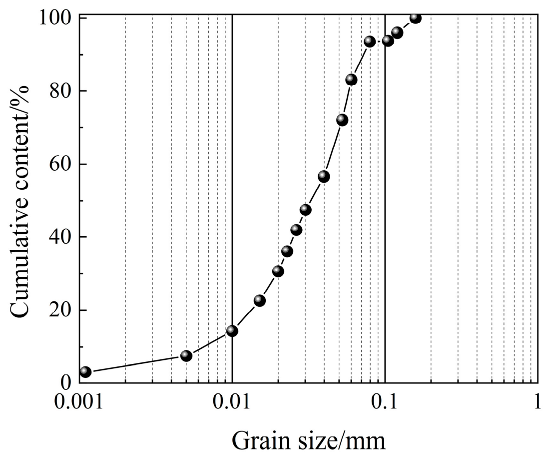

Figure 2. Particle size distribution curve of the soil samples.

Table 3. Basic physical properties of the soil samples.

Physical Index	Plastic Limit w_P/%	Liquid Limit w_L/%	Plasticity Index I_P/%	Maximum Dry Density/$g{\cdot}cm^{-3}$	Optimum Moisture Content/%	Specific Gravity of Soil	BET Specific Surface Area/$m^2{\cdot}g^{-1}$
Value	8.38	33.38	24.99	1.68	12.64	2.71	4.86

The control variables of the experimental design are salt content (S) and the number of freeze–thaw cycles (N). The moisture content of the sample is the optimal moisture content of 13%, and the dry density is the maximum dry density of 1.68 g/cm^3. Add a certain amount of anhydrous sodium sulfate to the deionized water required for the target water content to prepare a salt solution, and stir evenly with dry soil. Seal and let stand for 24 h to distribute water and salt evenly. Use an automatic sample preparation machine to prepare a sample to be tested with a diameter of 39.1 mm and a height of 80.0 mm.

The prepared soil samples were immediately wrapped in plastic wrap to prevent moisture loss. First, we let them stand for 12 h in a foam insulation box, and then put them into copper-like saturator molds to limit the deformation of the samples in any direction. Finally, they were put into a programmable ultra-low temperature testing machine for freeze–thaw cycle testing. Referring to ASTM-D560/D560M (2016) [43], after preliminary testing, a cycle was set to last 8 h, with freezing and thawing each taking 4 h. The freezing temperature is −20 and the melting temperature is 20 °C.

The salt contents of the test settings are 0%, 0.5%, 1%, and 2%, and the numbers of freeze–thaw cycles are 0, 1, 3, 5, 10, 20, 30, and 50 times, respectively, with a total of 32 test types. Each type contains 4 soil samples for parallel testing, totaling 128 soil samples. Among them, a set of UCS parallel tests was completed on two mechanical test soil samples that were in a melted state after undergoing freeze–thaw cycle tests. The other two wave speed test soil samples were subjected to a parallel wave speed test before the start of the freeze–thaw cycle test, which is called the wave speed test before freezing and thawing (compressional wave and shear wave speed test before freezing and thawing). After the target number of freezing and thawing cycles is completed, wait for it to appear in a melted state and then conduct a parallel wave speed test, which is called the wave speed test after freezing and thawing (compressional wave and shear wave speed test after freezing and thawing). Finally, after the wave speed test is completed, it is freeze-dried and subjected to two sets of (cross-section, longitudinal section) SEM parallel tests.

2.2. Test Methods and Procedures

2.2.1. UCS Test

UCS testing is performed at room temperature (approximately 20 °C). Refer to ASTM-D2166, (2016) [44], with the help of the British GDS triaxial testing system. The testing equipment mainly consists of three parts: the pressurization system, the back pressure control system, and the measurement system. The test settings include a strain rate of 1 mm/min, a confining pressure of 0 kPa, a non-consolidated and non-drained method, and data collection once every 3 s. All results are the arithmetic mean of two parallel samples from each group. The vertical load range is (50 kPa–500 kPa). The specific process of UCS test is as follows:

(1) Remove the plastic wrap of the sample, polish the surface to ensure smoothness, and then start the UCS test.
(2) After the test is completed, the stress–strain data corresponding to the .gds format is automatically output. Subsequently, through Origin drawing, the peak stress of the curve is selected as the UCS value of the sample.

2.2.2. Ultrasonic Testing

Ultrasonic testing is performed at room temperature. The RSM-SY5(T) non-metallic acoustic wave detector developed by the Wuhan Institute of Geotechnical Sciences, Chinese Academy of Sciences is used. The instrument mainly consists of pressure-bearing transmitting and receiving transducers, acoustic wave detectors, and wires. The wave speed is measured using the acoustic pulse transmission method, and the compressional and shear wave transducer frequencies are 50 kHz and 200 kHz, respectively. The parameter settings before and after freezing and thawing are the same, the compressional wave setting transmit pulse width is 300, and the gain is 1000 times. The shear wave transmit pulse width is 300 and the gain is 4000 times. All results are the arithmetic mean of two parallel samples from each group. The specific process is as follows:

(1) Add a sample with the upper and lower bottom surfaces brushed with Vaseline as coupling agent. Note that the lower bottom surface of the sample is close to the receiving transducer, and the upper bottom surface is close to the pressure-bearing transmitting transducer. After the sample loading is completed, the wave speed test begins.

(2) Each test item is drawn and recorded three times as waveform data by the acoustic wave detector. After the collection is completed, .SHT and .SHD format files are output. Subsequently, with the help of the oscilloscope software corresponding (RSM acoustic intensity analysis software, V1.1.220310) to the detector and the initial arrival wave method [45,46], the wave propagation time and the sample acoustic wave speed are obtained.

2.2.3. SEM Test

SEM tests were performed at room temperature using Quanta Training-X50 series, an instrument which mainly consists of three parts: electronic optical system, signal detection and amplification system, and vacuum and power supply system. The test setting scan time is 5 us and HV is 20.00 kV. The specific testing process is as follows:

(1) Freeze-dry the sample and obtain a fresh cross-section cube of about 8 mm by cutting near the middle part of the sample along the direction perpendicular to the Z-axis as a cross-section sample (The Z-axis is parallel to the height direction of the sample). Cut along the direction perpendicular to the X and Y axes to obtain a fresh cross-section cube of about 8 mm as a longitudinal cross-section sample. The sample to be tested is then fixed on the metal stage, coated, and sent to the sample chamber.
(2) Start the test program and enter relevant test parameters. Focus at high magnification first, and then look for a suitable area under low magnification conditions. After selecting the test area, the position will no longer move, and 500X, 1000X, and 2000X shooting will be performed in sequence.

Through the above SEM test, SEM photos of the same observation point and different magnifications of the sample to be tested can be obtained. The acquisition of SEM characteristic parameters will use PCAS software (Particles (Pores) and Cracks Analysis System (PCAS) V2.3) [47]. It was jointly developed by Dr. Liu Chun and his team from the School of Earth Science and Engineering of Nanjing University. It is a piece of software that specializes in automatic identification, geometric quantification, and statistical analysis of rock and soil particles, pores, and crack images. The specific process is shown in Figure 3. According to relevant information [23,48,49] and preliminary tests, the relevant threshold parameters are determined as follows: threshold: 75; element radius (pixel) 2.1; minimum area (pixel) 50.

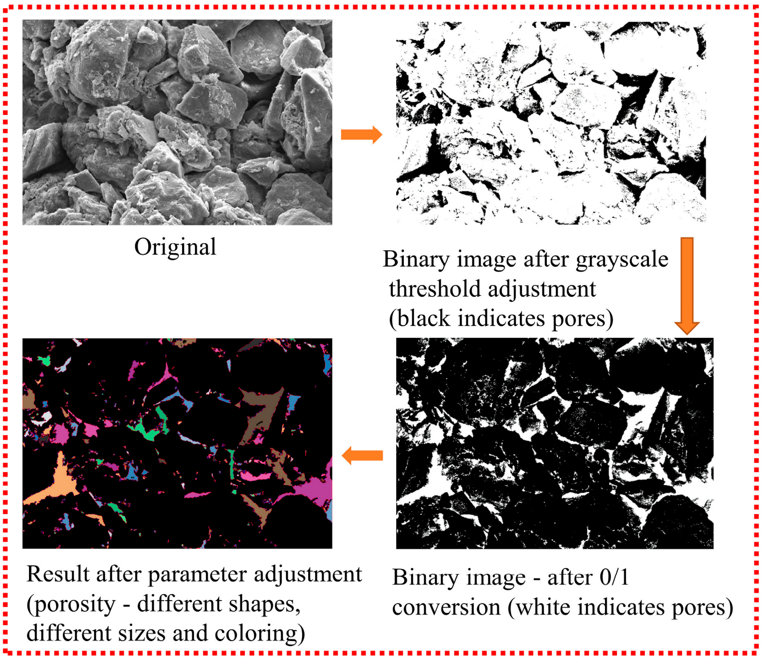

Figure 3. The process of obtaining SEM characteristic parameters in PCAS software.

2.3. Analysis of Test Results

2.3.1. The Impact of Control Factors on UCS

As shown in Figure 4, UCS shows obvious stages of characteristics change as the number of freeze–thaw cycles increases. Taking N = 5 and 30 as the node, it is divided into three stages. According to the changes in each stage, the first, second, and third stages are named adjustment period, dynamic fluctuation period, and stable period.

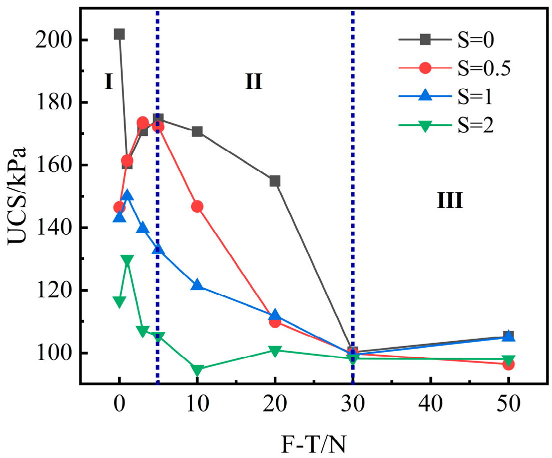

Figure 4. Relationship between controlling factors and soil UCS. Among them, I, II, and III represent the first, second, and third stages respectively; the blue dotted line is the stage dividing line.

As the number of freeze–thaw cycles increases, the overall UCS of the soil decreases. In the first stage, when there is no salt, UCS first decreases and then increases; when there is salt, UCS first increases and then decreases, and the overall fluctuation range is large. As the salt content increases, UCS decreases. In the second stage, UCS decreased significantly as a whole, with the largest fluctuation. In the third stage, the overall UCS remained basically unchanged, with the smallest fluctuation. The overall strength increased and decreased in the first stage, showing that the adjustment and adaptation of the internal soil particles of the soil caused the strength to fluctuate up and down, so it was named the adjustment period. In the second stage, the strength was greatly reduced and the deterioration rate was extremely fast. This corresponds to the abnormal activity or dynamic fluctuation of the internal characteristics of the soil, causing a rapid decline in strength, so it is named the dynamic fluctuation period. In the third stage, the strength remained stable and the deterioration was not obvious, which corresponds to the balance and stability of the internal characteristics of the soil, so it is named the stable period. As the salt content increases, the soil UCS decreases overall. When the salt content is 2%, the overall UCS fluctuation amplitude is small as the freeze–thaw cycle changes, the corresponding adjustment period will be shortened and ends early, the dynamic fluctuation period will be early and extended, and the stable period remains unchanged.

2.3.2. Influence of Control Factors on Wave Speed

As shown in Figure 5, the two parameters of compressional wave velocity after freezing and thawing and shear wave velocity after freezing and thawing also show obvious stages of characteristics change with the increase in the number of freeze–thaw cycles. Taking N = 5 and 30 as the node, it is also divided into three stages, corresponding to the adjustment period, the dynamic fluctuation period, and the stable period. Comparing the compressional wave velocity of the soil before freezing and thawing, as the number of freezing and thawing cycles increases, the overall compressional wave velocity after freezing and thawing decreases significantly in the first stage, and the fluctuations are strong. In the second stage, the overall trend increases and the fluctuation is the strongest. The third stage is slightly smaller overall, with the weakest fluctuations. At the same time, comparing the shear wave velocity of the soil before freezing and thawing, as the number of freeze–thaw cycles increases, the overall shear wave velocity after freezing and thawing increases significantly in the first stage, and the fluctuation is the strongest. Although there is a decline in the second stage, the overall trend is still increasing and the fluctuations are strong. The third stage is slightly smaller overall, with the weakest fluctuations. It shows that the relatively large adjustment changes of the soil particle aggregates and the internal meso structure of the soil in the first stage are not conducive to the propagation of compressional waves but are conducive to the propagation of shear waves. After passing the critical point and reaching the second stage, the most active changes or dynamic fluctuations of the meso structure are conducive to the propagation of both compressional and shear waves. In the third stage, the changes in the meso structure tend to stagnate, resulting in relatively stable changes in compressional and shear wave speeds.

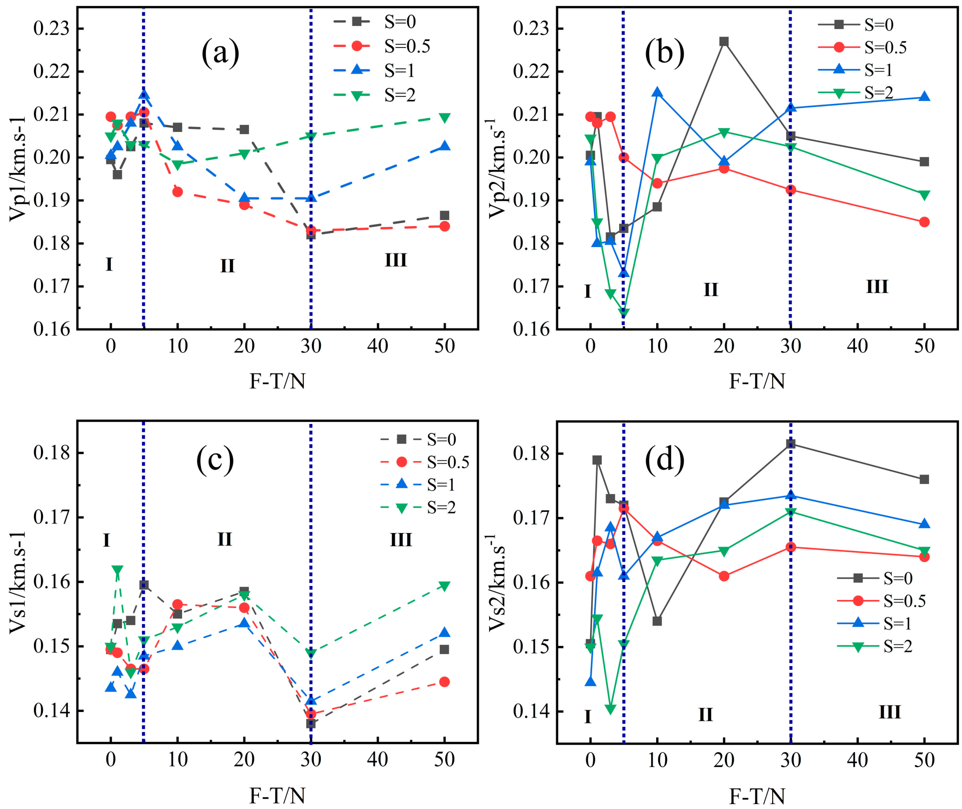

Figure 5. The relationship between controlling factors and soil wave speed: (**a**) V_{p1}-compressional wave speed before freeze–thaw cycle; (**b**) V_{p2}-compressional wave speed after freeze–thaw cycle; (**c**) V_{s1}-shear wave speed before freeze–thaw cycle; (**d**) V_{s2}-shear wave speed after freeze–thaw cycle wave speed. Among them, I, II, and III represent the first, second, and third stages respectively; the blue dotted line is the stage dividing line.

As the salt content increases, the compressional and shear wave speeds after freezing and thawing generally decrease below the initial values. It shows that the compressional and shear wave speeds generally decrease after the addition of salt, and the presence of salt will weaken the propagation of wave speed in the soil.

2.3.3. The Influence of Control Factors on SEM Characteristic Parameters

The choice of magnification is very important when performing quantitative analysis based on SEM images. Even increased magnification may result in a reduced overall perspective on microstructural characterization. However, this article aims to eliminate as much as possible the inaccurate identification of the shape of the soil pore system and errors in parameter statistical analysis caused by insufficient magnification [50]. Therefore, the

SEM image with the maximum magnification, i.e., 2000X, was selected for microstructural parameter analysis [51,52].

PCAS analysis of SEM images obtained from SEM experiments can obtain 19 characteristic parameters, from which the following 4 representative parameters are selected [4,50] to explore the changes of each parameter with controlling factors. The relevant analysis is as follows:

(1) Probability entropy

Probabilistic entropy describes the directional characteristics of pore systems.

$$H = -\sum_{i=1}^{n} P_i log_n P_i \tag{1}$$

where H is the probability entropy; P_i represents the percentage of pores within a specific range, and the value of H is between 0 and 1.

(2) Probability distribution index

The probability distribution index describes the area distribution characteristics of the pore system. Defined by a probability distribution function, it refers to the density of pore area in a specific area.

$$F(S) = dN/(N \cdot dS) \tag{2}$$

where N is the total number of pores and dN is the number of pores within a specific dS.

(3) Fractal dimension

Fractal dimension describes the shape distribution characteristics of the pore system. It refers to the variation pattern of shape complexity with its area.

$$\log(C) = Df/2 \cdot \log(S) + c_1 \tag{3}$$

where c_1 is a constant. Plot $C - S$ on log–log coordinates; $log(C)$–$log(S)$ data will exhibit a simple near-linear form, with the slope of the approximate line being $Df/2$.

(4) Porosity

Porosity reflects the absolute volume proportion of pores and changes in the microstructure of soil particles.

$$n = \frac{S_0}{S_1} \times 100\% \tag{4}$$

where n is the apparent porosity % of the soil and S_0 and S_1 are the areas of pores and particles, respectively, μm^2.

As shown in Figure 6, the four parameters of probability entropy, probability distribution index, fractal dimension, and porosity all show obvious stages of characteristics change with the increase in the number of freeze–thaw cycles. Taking N = 5 and30 as the node, it can still be divided into three stages, which still correspond to the adjustment period, the dynamic fluctuation period, and the stable period.

As shown in Figure 6a, with the increase in the number of freeze–thaw cycles, the probability entropy does not change significantly in the first and third stages, and the fluctuations in the second stage are strong, with an overall slight increase. It shows that the directionality of the pore system in the first and third stages is strong, and the directionality of the pore system in the second stage is weakened. The corresponding directional characteristics of soil particles are strengthened, i.e., part of the surface contact between soil particles is converted into point contact, which is not conducive to the strength properties [53].

As shown in Figure 6b, as the number of freeze–thaw cycles increases, the probability distribution index has no obvious change pattern in the first stage alone, the fluctuation is the strongest in the second stage, and the fluctuation intensity weakens in the third stage, but in the second and third stages, the overall decreasing trend of stages is similar. It shows that in the first order, although the pore system is adjusting, the overall area distribution characteristics are relatively stable. The area distribution characteristics of the pore system in the second and third stages began to continuously weaken, i.e., the number of small-area pores decreased, and some small pores were converted into large pores [23]. In particular, the second stage fluctuates the most, indicating that the area of the pore system has the largest density conversion rate in a specific region, which is the most detrimental to the strength properties. Although the third stage continued the conversion trend of the second stage to a certain extent, the adverse effect on strength was weakened because the conversion rate was significantly reduced.

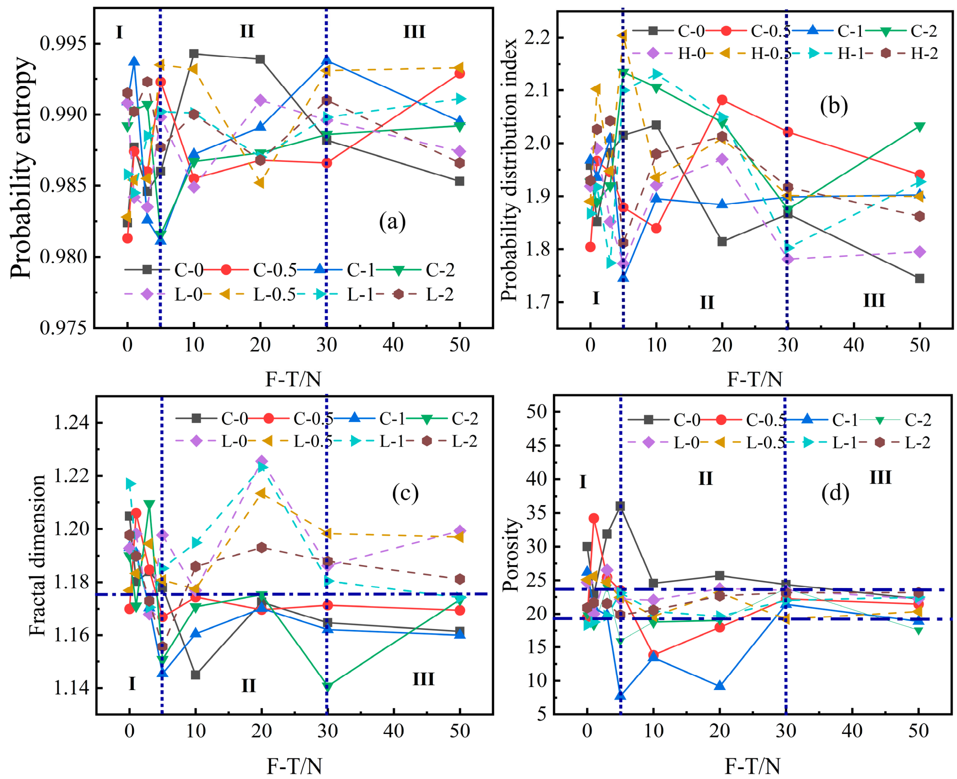

Figure 6. The relationship between controlling factors and soil SEM characteristic parameters: (**a**) probability entropy; (**b**) probability distribution index; (**c**) fractal dimension; (**d**) porosity; C-cross section, L-longitudinal section; 0, 0.5, 1, 2- is the salt content. Among them, I, II, and III represent the first, second, and third stages respectively; the blue dotted line is the stage dividing line.

As shown in Figure 6c, as the number of freeze–thaw cycles increases, the overall fractal dimension decreases significantly in the first stage. Starting from the second stage, the fractal dimensions of the longitudinal and cross sections differentiated, although both showed an overall fluctuation trend of first increasing and then decreasing in the second stage, and the fluctuation amplitude was the largest. However, there is an obvious limit of Df = 1.175. The fractal dimension of the longitudinal section always fluctuates above the limit, and the fractal dimension of the cross section always fluctuates below the limit. In the third stage, the differentiation trend of the fractal dimensions in the longitudinal and cross sections is maintained, but the fluctuation range of the fractal dimensions in the longitudinal and cross sections is the smallest. It shows that the boundary complexity of the pore system in the first stage has a weakening trend, and the corresponding boundary complexity of the soil particles increases as a whole. The soil is in an adjustment and adaptation period; its strength increases and decreases, and it begins to develop in the direction of deterioration. There are obvious differences in the boundary complexity of the pore system on the longitudinal and cross sections in the second stage, even though the overall fluctuation trends are similar. At the same time, the overall boundary complexity of the pore system fluctuates the most, and the corresponding contact mode between soil particles also changes drastically, resulting in a continuous and substantial reduction in strength. However, the boundary complexity of the pore system in the longitudinal section is generally stronger than that in the cross section, or the boundary complexity of the pore system in the longitudinal section is more sensitive to the soil fluctuation trend. In the third stage, the boundary complexity of the longitudinal and cross-sectional pore systems still maintains obvious differentiation, while the overall fluctuation trend is still similar. This shows that the overall fluctuation of the boundary complexity of the pore system has slowed down, resulting in the fluctuation of the intensity also becoming stable simultaneously. However, the situation that the boundary complexity of the pore system in the longitudinal section is greater than that in the cross section still exists.

As shown in Figure 6d, as the number of freeze–thaw cycles increases, the porosity has no obvious change pattern in the first stage. Starting from the second stage, the porosity of the longitudinal and cross sections also shows differentiation, and the overall fluctuation trends are also different. The overall fluctuation range of the longitudinal section is small, essentially between 18.5% and 23.5%, while the overall fluctuation range of the cross section is larger, and the overall range is outside 18.5%–23.5%. In the third stage, the porosity differentiation in the longitudinal and cross sections disappears, and the overall fluctuation trend converges. The fluctuation range is the smallest, essentially between 18.5% and 23.5%. It shows that the absolute volume proportion of pores in the first stage is unstable and the soil microstructure changes strongly. In the second stage, there are obvious differences in the absolute volume proportions of the pore system on the longitudinal and cross sections, and the overall fluctuation trends are different. The overall fluctuations of the pore system are greater in the cross section than in the longitudinal section. Specifically, the overall absolute volume ratio of the pore system in the cross section when containing salt is smaller than that in the longitudinal section. Salt affects the absolute volume proportion of the pore system and changes in soil microstructure, causing strength to differ according to salt content. The greater the salt content, the smaller the fluctuation in the absolute volume proportion of pores, and the smaller the decrease in strength. In the third stage, the difference in absolute volume proportion of the pore system in the longitudinal section and cross section disappears. At the same time, the overall fluctuation trend is similar, the fluctuations are slowing down, and the ability to maintain the current state is strong. The changes in soil microstructure tend to stagnate, resulting in the basic formation of the strength pattern of the soil.

This is different from the freeze–thaw cycle that actively changes the soil pore system directly and affects the SEM characteristic parameters. Salt can only be integrated into the soil system through crystallization and dissolution, and has an impact with the help of freezing and thawing. Therefore, salt undergoes crystallization and dissolution under the action of freeze–thaw cycles, which is generally not conducive to the cementation ecology within the soil and activates the development of the pore system. Specifically, there is no obvious pattern in the influence of probability entropy, probability distribution index, and fractal dimension among SEM characteristic parameters. The second-stage differentiation effect on the porosity in longitudinal and cross sections is more significant and is sensitive to the absolute volume change of the pore system. Overall, it is not conducive to the change of soil microstructure and weakens the fluctuation of strength.

In summary, Table 4 is used to conduct a comparative analysis of changes in soil macroscopic (UCS), mesoscopic (wave speed), and microscopic (SEM characteristic parameters) parameters under the influence of control factors. It was found that the changes in the three-level parameters can be divided into three stages, and the changes and fluctuations in each stage have a good correspondence. That is, the subscripts 1, 2, and 3 of each stage of the three-level parameters have a good correspondence with the subscripts 1, 2, and 3 of the stages. This shows that the three-level characteristics of salinized soil under the action of freeze–thaw cycles are not isolated from each other but have some connection. However, limited by limited experimental data and complex actual changes, it is difficult to conduct in-depth qualitative research. Faced with this difficulty, an attempt was made to propose a hypothesis linking the three-level characteristics of soil. By using new methods to bring hypotheses into machine learning models, we offer a useful attempt to build data-driven mathematical models with the help of machine learning models.

Table 4. Comparison of changes in macro (UCS), meso (wave speed), and micro (SEM characteristic parameters) in three stages.

Stage	Macro	Meso	Micro
Stage 1: Adjustment period	I: A_1; B_1	* I: (A_1); (C_1) * II: (A_1); (B_2)	** I: <A_1>; <B_1> ** II: <A_1>; <B_1> ** III: <A_1>; <C_1> ** IV: <A_1>; <B_1>
Stage 2: Dynamic fluctuation period	I: A_2; B_2	* I: (A_2); (B_2) * II: (A_2); (B_1)	** I: <A_2>; <B_2> ** II: <A_2>; <C_2> ** III: <A_2>; <D_2>; <F_2> ** IV: <A_2>; <E_2>; <F_2>
Stage 3: Stable period	I: A_3; B_3	* I: (A_3); (B_3) * II: (A_3); (B_3)	** I: <A_3>; <B_3> ** II: <A_3>; <B_3> ** III: <A_3>; <C_3>; <D_3> ** IV: <A_3>; <C_3>

Note: I: UCS; A_1: medium fluctuation; B_1: decrease first and then increase or increase first and then decrease, i.e., both increase and decrease; A_2: maximum fluctuation; B_2: overall decrease; A_3: minimum amplitude fluctuation; B_3: the overall situation remains basically unchanged. * I: compressional wave velocity after freeze–thaw cycle; * II: shear wave velocity after freeze–thaw cycle; (A_1): medium amplitude fluctuation; (B_1): overall increase, moderate amplitude; (C_1): overall decrease, moderate amplitude; (A_2): maximum fluctuation; (B_2): the overall increase is the largest; (C_2): overall decrease, with the largest amplitude; (A_3): minimum amplitude fluctuation; (B_3): the overall value remains basically unchanged, with a slight decrease. ** I: probabilistic entropy; ** II: probability distribution index; ** III: fractal dimension; ** IV: porosity; <A_1>: medium fluctuation; <B_1>: there is no obvious development trend overall, and the amplitude is medium; <C_1>: overall decrease, moderate amplitude; <A_2>: maximum fluctuation; <B_2>: the overall increase is the largest; <C_2>: overall decrease, with the largest amplitude; <D_2>: the overall value first increases and then decreases, with the largest amplitude; <E_2>: the overall value first decreases and then increases, with the largest amplitude; <F_2>: there is a differentiation between the longitudinal section and the cross section as a whole; <A_3>: minimum amplitude fluctuation; <B_3>: the overall situation remains basically unchanged, with a slight decrease; <C_3>: the whole remains unchanged; <D_3>: there are continuous differentiation in longitudinal and transverse sections throughout the whole.

3. Methodology

The overall methodological research is shown in Figure 7. Through the application of experimental results, basic hypotheses, basic methods, and basic parameters are constructed to form a database that can be used for subsequent use of machine learning models.

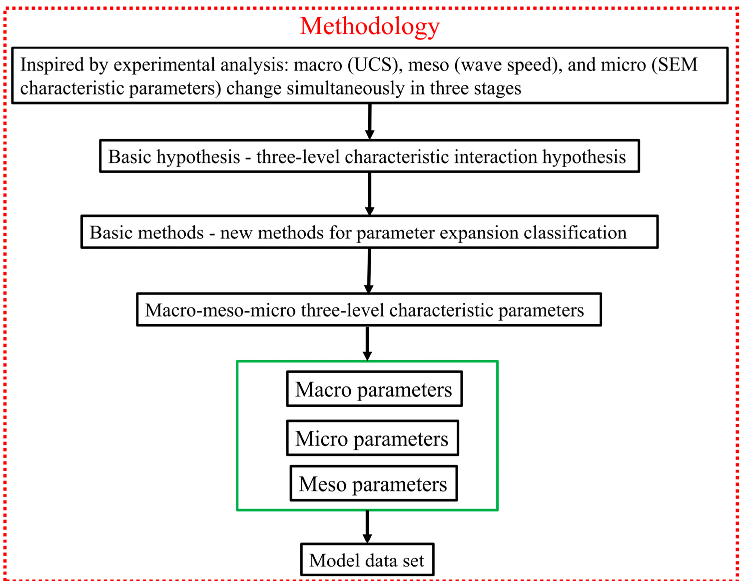

Figure 7. Schematic diagram of methodological research.

3.1. Basic Hypothesis-Three-Level Characteristic Interaction Hypothesis

The three-level characteristic interaction hypothesis is shown in Figure 8. The macro control factors (control variables in this study) affect the macro strength characteristic—UCS changes. However, this is not a direct effect, i.e., path ①, an inter-mediate path. The middle path is divided into two parts: paths ② and ③. Between them, the macro control factors in path ② directly affect the meso properties of the soil first, such as causing crack development and defect derivation at the meso scale (between the centimeter scale of the macro test and the micron scale of the micro analysis), thus affecting the macro UCS of the soil. The macro control factors in path ③ directly affect the micro properties of the soil, such as causing the soil particles on the micron scale inside the soil to become broken and denuded and the pores to expand or shrink, thereby affecting the macro UCS of the soil. It is worth noting that paths ② and ③ are not independent, and the micro properties of the soil directly affect the meso properties of the soil, i.e., path ④. At the same time, the meso characteristics restrict the further development of the micro characteristics to a certain extent, i.e., path ⑤, forming a complex dynamic equilibrium interaction system as a whole.

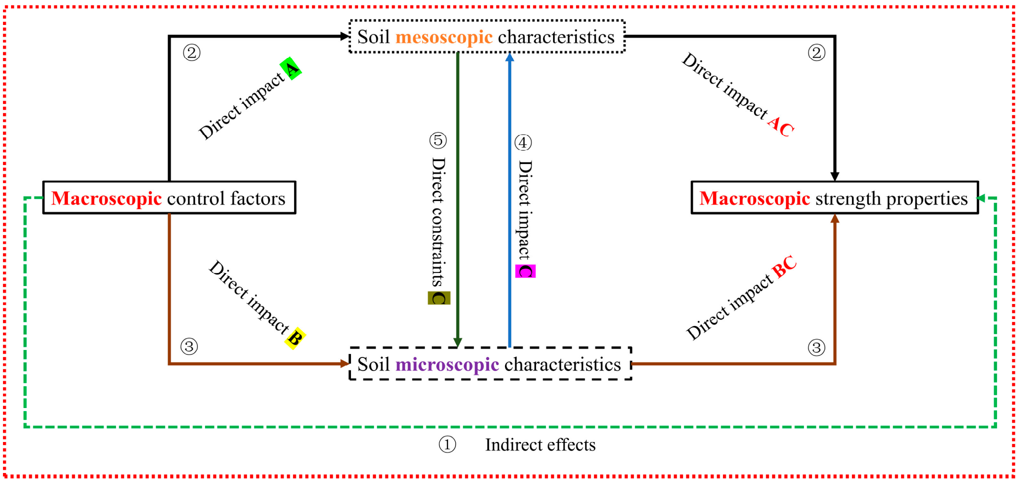

Figure 8. Schematic diagram of three-level characteristic interaction hypothesis logical relationship.

3.2. Basic Methods-New Methods for Parameter Expansion Classification

The core idea of the three-level characteristic interaction hypothesis is that the three-level characteristics of soil are interconnected, and the interactions on the third-level scale have a certain synchronicity. Therefore, it is necessary to expand the types and quantities of parameters and divide the levels of parameters. Based on this, a new method for expansion classification of machine learning model parameters is proposed. Among them, expansion refers to this method's expansion and construction of the number and level of parameters at the input end of the model. We enrich parameters to ensure that the information supplied to the model at the input end contains as much and significantly different information as possible, especially relevant scale differences. Classification refers to this method by clearly distinguishing and positioning parameters. During the model running process, we deliberately pay attention to the performance of parameters of different groups and give the parameters differentiated treatment.

3.3. Macro–Meso–Micro Three-Level Characteristic Parameters

3.3.1. Macro Parameters

The experimental control variables, the content of anhydrous sodium sulfate and the number of freeze–thaw cycles, are the macro parameters. The parameter codes are shown in Table 5, and the number of parameters is two.

Table 5. Macro parameter codes and definitions.

Parameter code	X_1	X_2
Definition	S	N

3.3.2. Meso Parameters

The ultrasonic characteristic parameters are selected as the meso parameters, the parameter codes are shown in Tables 6 and 7, and the number of parameters is 38.

Table 6. Ultrasonic velocity codes and definitions.

Parameter code	X_3	X_4	X_5	X_6
Definition	V_{p2}	V_{s2}	V_{p1}	V_{s1}

Table 7. Codes and definitions of the characteristic parameters derived from the wave velocity.

Parameter Code	**Definition**	**Parameter Code**	**Definition**	**Parameter Code**	**Definition**
X_7	$\left\lvert V_{p1}-V_{p2}\right\rvert$	X_{19}	$\lvert\mu_2\rvert$	X_{31}	$\left\lvert 1-\left(\Delta V_p/V_{p1}\right)^2\right\rvert$
X_8	$\lvert V_{s1}-V_{s2}\rvert$	X_{20}	V_{p1}^2	X_{32}	$\left\lvert 1-(\Delta V_s/V_{s1})^2\right\rvert$
X_9	$\left\lvert \Delta V_p/V_{p1}\right\rvert$	X_{21}	V_{s1}^2	X_{33}	$\lvert E_2/E_1\rvert$
X_{10}	$\lvert\Delta V_s/V_{s1}\rvert$	X_{22}	$\lvert G_1\rvert$	X_{34}	$\lvert G_2/G_1\rvert$
X_{11}	$\lvert V_{p2}/V_{s2}\rvert$	X_{23}	$\lvert\mu_1\rvert$	X_{35}	$\lvert\mu_2/\mu_1\rvert$
X_{12}	$\lvert V_{p2}-V_{s2}\rvert$	X_{24}	$\lvert E_2-E_1\rvert$	X_{36}	$\lvert\Delta E/E_1\rvert$
X_{13}	$\lvert\Delta V_p-\Delta V_s\rvert$	X_{25}	$\lvert G_2-G_1\rvert$	X_{37}	$\lvert\Delta G/G_1\rvert$
X_{14}	V_{p2}^2	X_{26}	$\lvert\mu_2-\mu_1\rvert$	X_{38}	$\lvert\Delta\mu/\mu_1\rvert$
X_{15}	V_{s2}^2	X_{27}	$\lvert 1-E_2/E_1\rvert$	X_{39}	$\lvert E_2\rvert$
X_{16}	ΔV_p^2	X_{28}	$\lvert 1-G_2/G_1\rvert$	X_{40}	$\lvert E_1\rvert$
X_{17}	ΔV_s^2	X_{29}	$\left\lvert 1-\left(V_{p2}/V_{p1}\right)^2\right\rvert$		
X_{18}	$\lvert G_2\rvert$	X_{30}	$\left\lvert 1-(V_{s2}/V_{s1})^2\right\rvert$		

Note: the characteristic parameters derived from the wave velocity are directly constructed through the parameter definition relational formulas on the basis of V_{p1}, V_{p2}, V_{s1}, and V_{s2} obtained from ultrasonic tests, with test data as source data. V_{p2}: compressional wave velocity after freeze–thaw (km/s); V_{s2}: shear wave velocity after freeze–thaw (km/s); V_{p1}: compressional wave velocity before freeze–thaw (km/s); V_{s1}: shear wave velocity before freeze–thaw (km/s); E: elastic modulus (GPa); G: shear modulus (GPa); μ: Poisson's ratio; and $E=\rho V_s^2\frac{\left(3V_p^2-4V_s^2\right)}{\left(V_p^2-2V_s^2\right)}$; $G=\rho V_s^2$; $\mu=\frac{\left(V_p^2-2V_s^2\right)}{2\left(V_p^2-V_s^2\right)}$.

The ultrasonic characteristic parameters are composed of two types of parameters: (1) 4 ultrasonic velocities ($X_3, \ldots, X_6$), as shown in Table 6, and (2) 34 wave velocity-derived characteristic parameters ($X_7, \ldots, X_{40}$), as shown in Table 7.

3.3.3. Micro Parameters

The SEM characteristic parameters as chosen as the micro parameters, the parameter codes are shown in Table 8, and the number of parameters is 19.

3.4. Model Data Set

A comprehensive dataset was created through experiments, and a total of 32 UCS values, 32 macro data values, 32 meso data values, and 192 micro data values were collected for salinized frozen soil. The overall composition is a 192 × 60 machine learning dataset, as shown in Table 9. Table 10 presents the statistical analysis of this dataset.

Table 8. SEM characteristic parameter codes and definitions.

Parameter Code	Definition	Parameter Code	Definition	Parameter Code	Definition
X_{41}	Image area	X_{48}	Average form factor	X_{55}	APDI
X_{42}	Total region area	X_{49}	Maxima length	X_{56}	PPDFD
X_{43}	Region number	X_{50}	Average length	X_{57}	Sorting Coefficient
X_{44}	Region percentage	X_{51}	Maxima width	X_{58}	UC
X_{45}	Maxima region area	X_{52}	Average width	X_{59}	Curvature Coefficient
X_{46}	Average region area	X_{53}	Probability Entropy		
X_{47}	Average perimeter	X_{54}	Fractal dimension		

Note: APDI: area probability distribution index; PPDFD: pore porosity distribution fractal dimension; UC: uniformity coefficient; and "region" specifically refers to soil pores.

Table 9. The composition and distribution of the parameters in the dataset.

Columns / Rows	Macro Parameters	Meso Parameters	Micro Parameters	UCS
1–32	A	B	C_1	D
33–64	A	B	C_2	D
65–96	A	B	C_3	D
97–128	A	B	C_4	D
129–160	A	B	C_5	D
161–192	A	B	C_6	D

Note: the 192 rows of UCS in the dataset were composed of the same 32 UCS values (denoted as D) repeated 6 times. The 192 rows of macro parameters consist of six identical 32 macro data values (denoted as A). The 192 rows of meso parameters consist of six identical 32 meso data values (denoted B). The 192 lines of micro parameters are composed of 192 micro data values (among them, C_1 is 2000X, cross-section SEM characteristic parameter value; C_2 is 2000X, longitudinal section SEM characteristic parameter value; C_3 is 1000X, cross-section SEM characteristic parameter value; C_4 is 1000X, longitudinal section SEM characteristic parameter value; C_5 is 500X, cross section SEM characteristic parameter value; and C_6 is 500X, longitudinal section SEM characteristic parameter value).

Table 10. Basic statistical analysis of the dataset.

Variables	Unite	Symbol	Mean	Max	Min	St.D	Sk	Ku
UCS	kPa	UCS	131.21	201.73	94.73	30.20	0.46	−1.10
S	%	X_1	0.87	2	0	0.73	0.43	−1.16
N	1	X_2	14.87	50	0	16.45	1.08	−0.09
V_{S1}	km/s	X_3	0.15	0.19	0.13	0.01	0.89	−0.05
V_{S2}	km/s	X_4	0.16	0.18	0.14	0	−0.64	−0.04
V_{P1}	km/s	X_5	0.2	0.25	0.18	0.01	0.92	0.31
V_{P2}	km/s	X_6	0.19	0.19	0.16	0.01	−0.35	−0.27
$\left\lvert V_{p1}-V_{p2} \right\rvert$	km/s	X_7	0.01	0.06	0	0.01	0.81	0.05
$\lvert V_{s1} - V_{s2} \rvert$	km/s	X_8	0.01	0.04	0	0	1.08	1.8
$\lvert \Delta V_p / V_{p1} \rvert$	1	X_9	0.08	0.26	0	0.06	0.63	−0.33
$\lvert \Delta V_s / V_{s1} \rvert$	1	X_{10}	0.08	0.34	0	0.07	1.5	3.38
$\lvert V_{p2} / V_{s2} \rvert$	1	X_{11}	1.19	1.37	1.04	0.08	0.24	−0.67
$\lvert V_{p2} - V_{s2} \rvert$	km/s	X_{12}	0.03	0.05	0	0.01	0.04	−0.92
$\lvert \Delta V_p - \Delta V_s \rvert$	km/s	X_{13}	0.01	0.04	0	0.01	0.97	0.59
V_{p2}^2	km^2/s^2	X_{14}	0.03	0.05	0.02	0	−0.17	−0.29
V_{s2}^2	km^2/s^2	X_{15}	0.02	0.03	0.01	0	−0.51	−0.18
ΔV_p^2	km^2/s^2	X_{16}	0	0	0	0	2.28	5.7
ΔV_s^2	km^2/s^2	X_{17}	0	0	0	0	3.28	12.7
$\lvert G_2 \rvert$	GPa	X_{18}	0.05	0.06	0.03	0	−0.51	−0.18
$\lvert \mu_2 \rvert$	1	X_{19}	1.02	4.46	0.05	1	1.77	2.82
V_{p1}^2	km^2/s^2	X_{20}	0.04	0.06	0.03	0	1.09	0.56
V_{s1}^2	km^2/s^2	X_{21}	0.02	0.03	0.01	0	1.03	0.15
$\lvert G_1 \rvert$	GPa	X_{22}	0.04	0.07	0.03	0	1.03	0.15

Table 10. *Cont.*

Variables	Unite	Symbol	Mean	Max	Min	St.D	Sk	Ku
$\lvert\mu_1\rvert$	1	X_{23}	0.19	0.56	0.01	0.13	0.89	0.44
$\lvert E_2-E_1\rvert$	GPa	X_{24}	0.43	2.09	0	0.56	1.82	2.36
$\lvert G_2-G_1\rvert$	GPa	X_{25}	0	0.02	0	0	1	1.48
$\lvert\mu_2-\mu_1\rvert$	1	X_{26}	0.84	4.28	0	0.99	1.83	2.91
$\lvert 1-E_2/E_1\rvert$	1	X_{27}	0.84	1.22	0.02	0.3	−1.16	0.
$\lvert 1-G_2/G_1\rvert$	1	X_{28}	0.18	0.8	0	0.16	1.84	4.8
$\left\lvert 1-\left(V_{p2}/V_{p1}\right)^2\right\rvert$	1	X_{29}	0.15	0.45	0	0.12	0.43	−0.72
$\left\lvert 1-(V_{s2}/V_{s1})^2\right\rvert$	1	X_{30}	0.18	0.8	0	0.16	1.84	4.8
$\left\lvert 1-\left(\Delta V_p/V_{p1}\right)^2\right\rvert$	1	X_{31}	0.98	1	0.93	0.01	−2.02	4.29
$\left\lvert 1-(\Delta V_s/V_{s1})^2\right\rvert$	1	X_{32}	0.98	1	0.88	0.02	−3.84	16.
$\lvert E_2/E_1\rvert$	1	X_{33}	0.29	1.46	0	0.38	1.99	3.02
$\lvert G_2/G_1\rvert$	1	X_{34}	1.11	1.8	0.77	0.21	0.92	1.43
$\lvert\mu_2/\mu_1\rvert$	1	X_{35}	8.59	37.88	0.84	9.19	1.45	1.6
$\lvert\Delta E/E_1\rvert$	1	X_{36}	0.84	1.22	0.02	0.3	−1.16	0.7
$\lvert\Delta G/G_1\rvert$	1	X_{37}	0.18	0.8	0	0.16	1.84	4.8
$\lvert\Delta\mu/\mu_1\rvert$	1	X_{38}	7.73	36.88	0.03	9.38	1.44	1.51
$\lvert E_2\rvert$	GPa	X_{39}	0.09	0.66	0	0.14	2.7	7.16
$\lvert E_1\rvert$	GPa	X_{40}	0.49	2.27	0.03	0.58	1.69	1.77
Image area	pixel	X_{41}	1,569,032	1,574,400	1,545,216	5522.56	−3.34	11.16
Total region area	μm^2	X_{42}	348,390.76	854,526	121,255	68,664.57	3.41	24.31
Region number	1	X_{43}	848.65	1542	159	255.59	0.06	0.07
Region percentage	%	X_{44}	7.67	54.38	0.19	11.36	1.3	1.09
Max region area	μm^2	X_{45}	56,331.4	295,866	9333	44,069.51	2.41	8.29
Average region area	μm^2	X_{46}	453.75	1357.64	227.57	183.62	2.1	5.92
Average perimeter	μm	X_{47}	98.08	146.84	78.67	11.38	1.05	1.58
Average form factor	1	X_{48}	0.38	0.42	0.33	0.01	−0.45	0.14
Max length	μm	X_{49}	498.44	1447.68	200.48	199.56	1.45	3.35
Average length	μm	X_{50}	27.14	35.47	23.21	2.24	0.97	1.07
Max width	μm	X_{51}	281.62	662	99.61	103.78	1.05	1.21
Average width	μm	X_{52}	15.51	19.84	13.46	1.11	0.85	0.95
Probability Entropy	1	X_{53}	0.98	0.99	0.96	0	−1.91	6.2
Fractal dimension	1	X_{54}	1.19	1.26	1.14	0.02	0.26	−0.07
APDI	1	X_{55}	1.98	2.28	1.7	0.11	0.23	−0.49
PPDFD	1	X_{56}	1.97	2.54	1.43	0.2	0.17	−0.1
Sorting Coefficient	1	X_{57}	1.39	4.69	1.05	0.35	5.67	44.18
Uniformity Coefficient	1	X_{58}	1.75	4.54	1.09	0.38	2.36	13.56
Curvature Coefficient	1	X_{59}	1.18	2.26	0.41	0.26	1.86	5.2

Note: St.D-standard deviation; Min-minimum; Max-maximum; Sk-skewness; Ku-kurtosis. Bold lines represent macro parameters, italics represent micro parameters, and the rest are UCS and meso parameters.

A total of 59 characteristic parameters in the dataset are used as input variables to predict UCS using six machine learning models. Figure 9 shows the correlation between the considered characteristic parameters and UCS. Furthermore, to reasonably train and evaluate the predictive performance of each model, the entire dataset was randomly divided into two groups, namely the training set (76%) (147 × 60) and the testing set (24%) (45 × 60).

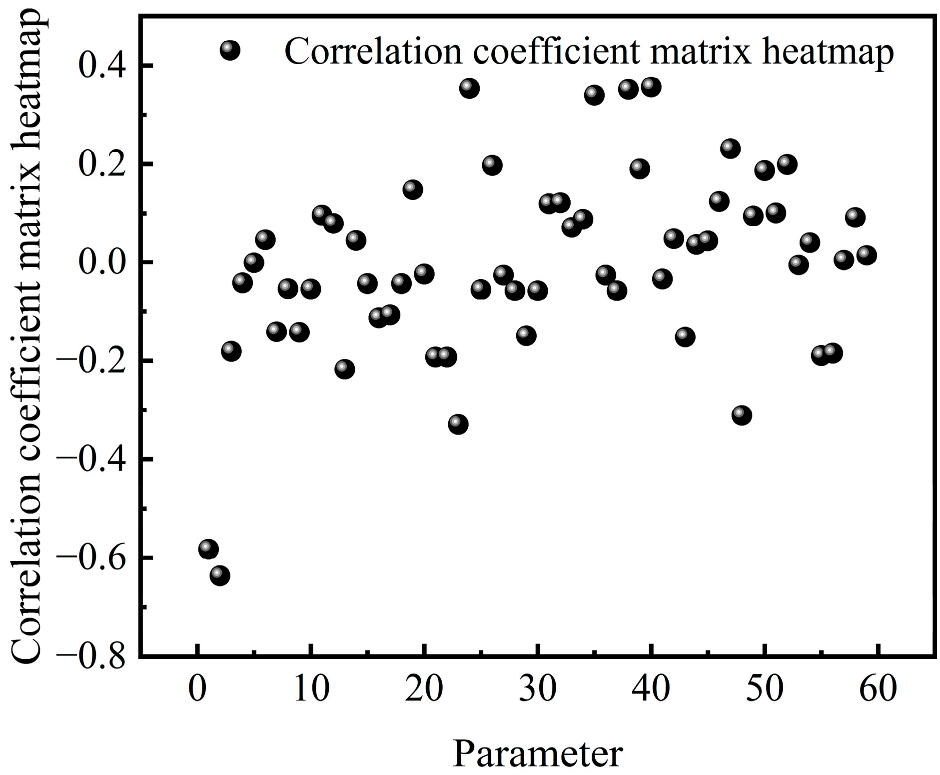

Figure 9. The correlation between the parameters in the dataset and UCS.

4. Methodology Applied to the Model

In order to realize the hypothesis in Section 3.1, the following six representative models are selected from machine learning models widely used in the field of geotechnical engineering. As a platform and tool, we use the method in Section 3.2 to perform predictive analysis of soil UCS. Among them, the three-level characteristic interaction hypothesis was successfully brought into the model through the new model parameter expansion classification method, reflecting the macro–meso–micro three-level response characteristics of the parameters. We verify the validity of the above assumptions and methods based on the relevant characteristics displayed by the model (model accuracy and parameter sensitivity).

4.1. 6 Machine Learning Models

It is intensely important to develop a suitable machine learning model for the accurate prediction of UCS of salinized frozen soil. In this study, six typical machine learning methods are used.

(1) Support vector machine (SVM)

SVM is a machine learning regression method based on statistical theory that has obvious advantages in dealing with linearly separable and linearly inseparable problems. It has the ability to calculate high-dimensional and multi-complexity inputs and has excellent generalizability and generally high prediction accuracy [54,55].

(2) Genetic algorithm optimized BP (GA-BP)

GA-BP is a global heuristic optimized stochastic search BP neural network based on the concept of natural selection and genetics and performs well in solving high-dimensional, nonlinear, and strong noise problems [56,57].

(3) Random forest (RF)

RF is a supervised regression ensemble learning method consisting of a bagging framework and an independent decision tree and has unique advantages in data utilization and performance evaluation mechanisms [58]. An increase in the number of decision trees usually does not lead to overfitting, and it is widely used in solving nonlinear and high-dimensional data problems [59].

(4) Radial basis function (RBF)

The RBF is an artificial neural network model based on the radial basis function and has good performance in terms of function approximation and clustering. It can deal with relatively complex input and output relationships, and the training speed is fast. Therefore, it is widely used in the field of geotechnical engineering [60,61].

(5) Long short-term memory (LSTM)

LSTM is a special recurrent neural network (RNN) [62] for simulating data with long-term dependencies, efficiently maintaining and updating the internal state and preserving long-term step information. It has the advantages of long-term dependent data modeling, noise robustness, and parameter adaptive ability [63,64].

(6) Particle swarm optimization algorithm BP (PSO-BP)

PSO-BP is a BP optimization algorithm that uses individual local information and global information in the group to guide a search and has the advantages of fewer adjustable parameters and strong hyperparameter selection ability [65,66]. It can effectively address nonlinear, nonconvex, and multimodal problems and is widely used to solve various optimization problems [67].

4.2. Evaluation Indicators

The performance of the six models was evaluated using the following four statistical indicators: root mean square error (RMSE), coefficient of determination (R^2), Willmott's index (WI), and variance accounted for (VAF). The R^2, WI, and VAF values of the corresponding optimal model should be higher, and the RMSE value should be lower. The above indicators are defined as follows [68–71]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - y_i\right)^2} \tag{5}$$

$$\mathrm{WI} = 1 - \left[\frac{\sum_{i=1}^{n}\left(Y_i - y_i\right)^2}{\sum_{i=1}^{n}\left(\left|y_i - \overline{Y}\right| + \left|Y_i - \overline{Y}\right|\right)^2}\right] \tag{6}$$

$$\mathrm{R}^2 = 1 - \frac{\left[\sum_{i=1}^{n}\left(Y_i - y_i\right)\right]^2}{\left[\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)\right]^2} \tag{7}$$

$$VAF = \left[1 - \frac{\mathrm{var}(Y_i - y_i)}{\mathrm{var}(Y_i)}\right] \times 100\% \tag{8}$$

where n is the number of samples in the training and testing stages, Y_i and y_i are the actual and predicted UCS values of the i-th sample, respectively, and $\overline{Y}$ and $\overline{y}$ are the mean values of the actual and predicted UCS, respectively.

4.3. Model Analysis and Hyperparameters

For the 59 total parameters and the 34 parameters obtained after parameter optimization, the UCS predictions with the six machine learning models were obtained. The optimal model was selected through the four main model evaluation parameters and four optimal model screening methods in turn. Then, the overall parameter sensitivity and the third-level characteristic parameter sensitivity analyses of the optimal model were carried out, and the prediction effect of the optimal model and the third-level response characteristics were comprehensively evaluated. The specific process is shown in Figure 10.

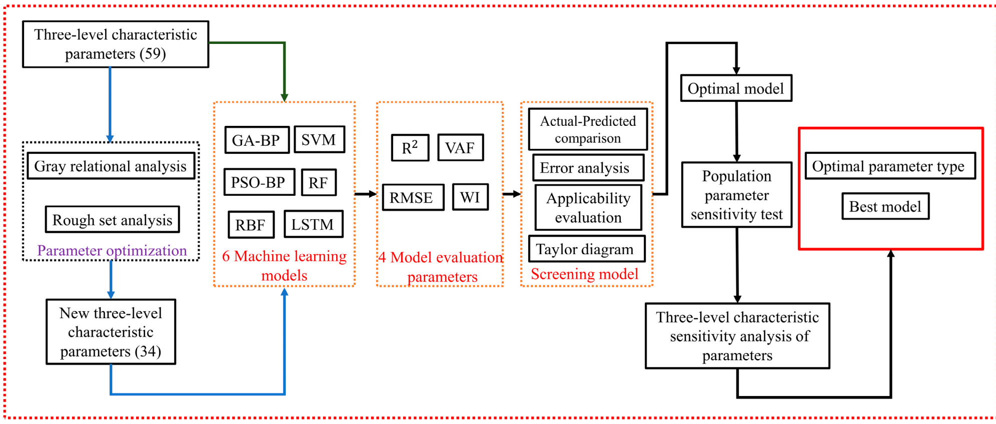

Figure 10. Flow chart of the model analysis.

According to the relevant information referenced in the early stage and the changes in the adaptive characteristics of the construction process model itself, the hyperparameters of the six machine learning models were determined, as shown in Table 11. At this time, the model does not exhibit over-fitting phenomena.

Table 11. Hyperparameters of the six machine learning models.

Model	Hyperparameter
SVM	PF = 4.0; RP = 0.8
GA-BP	NI = 10^5; ET = 10^{-6}; LR = 10^{-2}; NH = 7; GA = 50; PS = 5
RF	NDT =100; MNL = 5
RBF	ESR = 100
LSTM	LL = 4; MNI = 1200; ILA = 10^{-2}; LIDF = 0.5
PSO-BP	NI = 10^5; ET = 10^{-6}; LR = 10^{-2}; NH = 7; LF = 4.494; NPU = 30; PS = 5

Note: PF: penalty factor; RP: radial basis function parameter; NI: number of iterations; ET: error threshold; LR: learning rate; NH: number of hidden layer nodes; GA: genetic algebra; PS: population size; NDT: number of decision trees; MNL: minimum number of leaves; ESR: expansion speed of the radial basis function; LL: LSTM layer; MNI: maximum number of iterations; ILA: initial learning rate; LIDF: learning rate drop factor; LF: learning factor; NPU: number of population updates.

5. Results and Discussion

5.1. Model Prediction of the 59 Total Parameters

The evaluation of each model was carried out using four evaluation indicators, and the performance indicators and related grade scores of each model in the training stage are shown in Table 12. The RBF has the best performance and the highest grade scores in the four performance indicators. The GA-BP has slightly worse performance than then RBF, the SVM and LSTM are close to the middle, and the PSO-BP and RF perform the worst. However, all six models have good UCS prediction performance in the training stage.

Table 12. Fifty-nine parameters corresponding to the prediction performance evaluation of the six models in the training stage.

Model	Performance	and	Rank						Total
	R2	Score	RMSE	Score	WI	Score	VAF (%)	Score	
SVM	0.9989	4	1.0094	4	0.9997	4	99.8886	4	16
GA-BP	0.9994	5	0.7286	5	0.9999	5	99.9438	5	20
RF	0.9864	1	3.5871	1	0.9963	1	98.6484	1	4
RBF	**1**	**6**	$\mathbf{3.79 \times 10^{-7}}$	**6**	**1**	**6**	**100**	**6**	**24**
LSTM	0.9985	3	1.0931	3	0.9996	3	99.8869	3	12
PSO-BP	0.9907	2	2.9826	2	0.9977	2	99.0997	2	8

The bold line represents the optimal model.

The regression relationship between each model's actual and predicted UCS during the training stage is shown in Figure 11. The red boxplots in the figure show the statistical results of the actual and predicted values of UCS, including the median, minimum, maximum, upper quartile, and lower quartile. When the actual and predicted values are exactly equal, the data points are distributed on the black diagonal line (Y = X), while the dashed line indicates that the predicted value is allowed to deviate by 10%. Most of the points in each model are concentrated on the black diagonal line, a few points fall between the diagonal line and the 10% line, and very few points are distributed outside the 10% line. The RBF model not only has the most points on the black diagonal line but also has the highest values of the R^2, WI and VAF and the lowest value of the RMSE. The difference between the predicted value and the actual value in the statistical results of the RF model is the largest (median = 126.49 and 129.85).

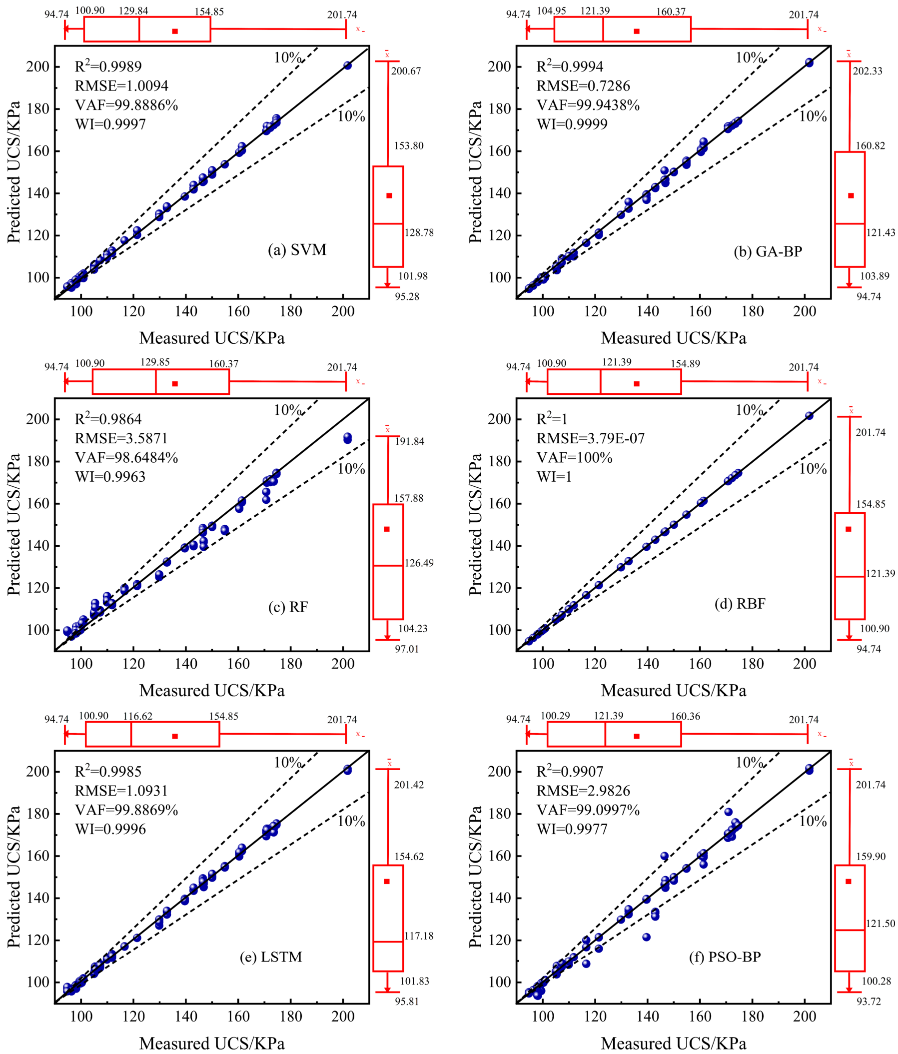

Figure 11. The regression diagram of the six models in the training stage with the 59 total parameters.

Figure 12 shows the error analysis of all models during the training stage, including the maximum and minimum errors and the standard deviation of all errors of the models. The error analysis of each model is significantly different, especially for models with similar scores in terms of the model performance indicators and related grade score tables. All error indicators of the RBF model are significantly lower than those of the other models.

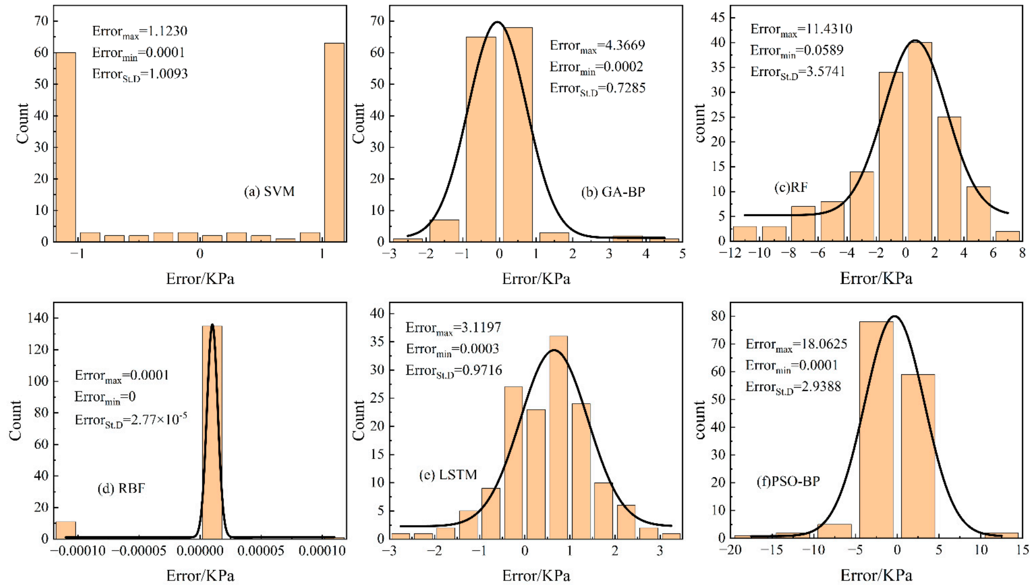

Figure 12. Error maps of the six models in the training stage with the 59 total parameters.

Since the training model with the best performance index with the training set may perform poorly in the testing stage, only the model verified with the testing set is generally officially used as the real model for UCS prediction. Table 13 shows the evaluation indicators and grade scores of the model in the testing stage. Among them, the RBF is still the best model and still obtains the highest grade scores for the four performance indicators. The LSTM performs slightly worse than the RBF. However, at this time, the performance of the GA-BP and RF is in the middle, and the performance of PSO-BP and SVM is the worst. At the same time, the six models still have good UCS prediction performance in the testing stage.

Table 13. The prediction performance evaluation of the six models in the testing stage with the 59 total parameters.

Model	Performance	and	Rank						Total
	R^2	Score	RMSE	Score	WI	Score	VAF (%)	Score	
SVM	0.9218	1	8.4106	1	0.9765	1	92.3190	1	4
GA-BP	0.9788	4	4.0217	4	0.9946	4	97.9077	4	16
RF	0.9738	3	4.5067	3	0.9925	3	97.3905	3	12
RBF	0.9998	6	0.4238	6	0.9999	6	99.9774	6	24
LSTM	0.9946	5	2.4451	5	0.9986	5	99.4615	5	20
PSO-BP	0.9366	2	6.9283	2	0.9847	2	94.1932	2	8

The regression relationship between the actual and predicted UCS of the models in the testing stage is shown in Figure 13. Most of the points in the RBF and LSTM models are concentrated on the black diagonal line, and a few points fall between the diagonal line and the 10% line. Most of the points in the GA-BP, RF, and PSO-BP models fall between the diagonal line and the 10% line, and a few points are distributed outside the 10% line. Nearly half of the points in the SVM fall between the diagonal and the 10% line, and the remaining points are distributed outside the 10% line. At the same time, the RBF model still has the most points on the black diagonal line. In addition, the R^2, WI, and VAF values are the highest, and the RMSE value is the lowest. The difference between the predicted value and the actual value in the statistical results of the SVM model is the largest (median = 130.03 and 121.39).

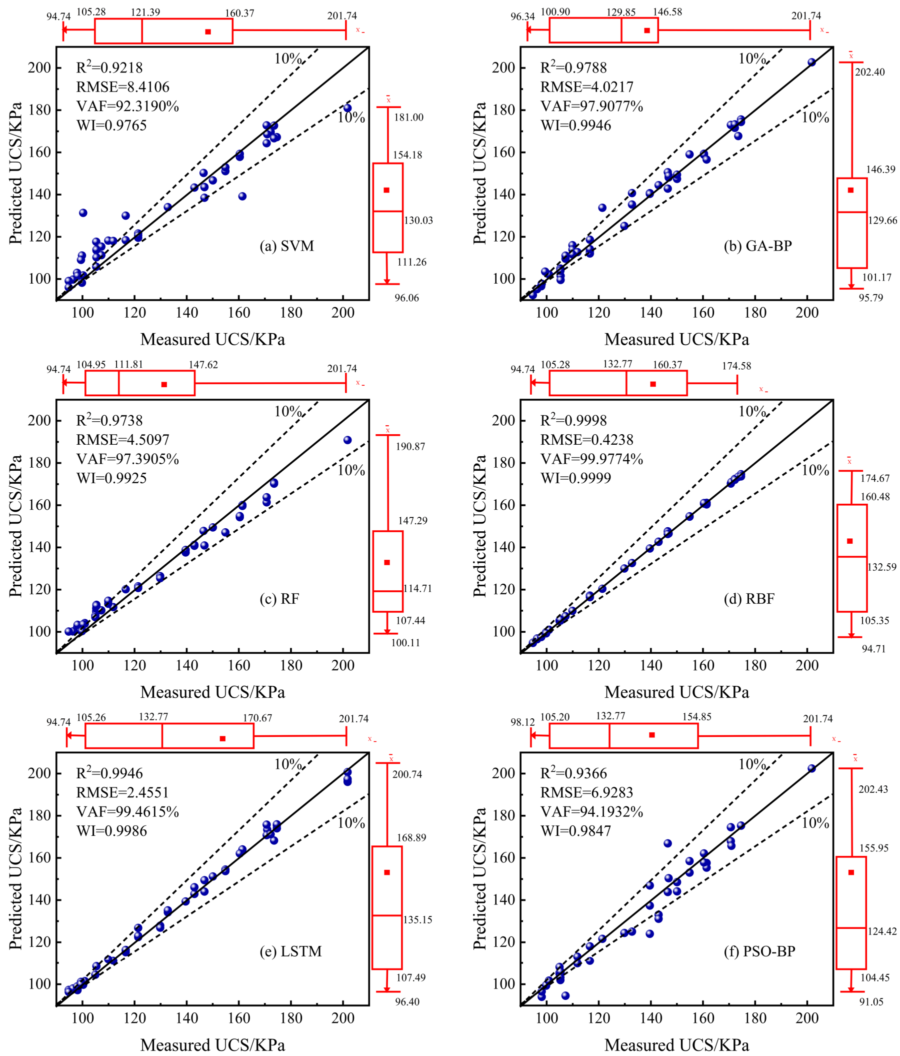

Figure 13. The regression diagram of the six models in the testing stage with the 59 total parameters.

Figure 14 shows the error analysis of all models in the testing stage. It can be observed from the figure that the differences in the error analysis of each model are also obvious, especially for models with similar scores in terms of the model performance indicators and related grade score tables. All error indicators of the RBF model are significantly lower than those of the other models.

5.1.1. 59-Parameter Optimal Model

It is not sufficient to sort the prediction performance of the six machine learning models only through the performance indicators and related grade scores in the model training and testing stages, the regression relationship diagram between the actual value and the predicted value, and the model error diagram. Thus, the Taylor diagram and the model applicability evaluation chart were introduced for the following screening and sorting of the optimal model.

The Taylor diagram is a model verification method that is widely used in the field of machine learning, and it can generally be divided into three parts, namely standard deviation, correlation coefficient, and root mean square error. As shown in Figure 15, the blue line reflects the correlation coefficient, the green line is the root mean square error, and the black line is the standard deviation. The reference point (red solid circle) is set as follows: training (SD: 30.0; RMSE: 0; and R: 1), testing (SD: 15.0; RMSE: 0; and R: 1). The order of all models in the training stage is RBF > GA-BP > SVM > LSTM > PSO-BP > RF, and the order in the testing stage is RBF > LSTM > GA-BP > RF > PSO-BP > SVM. The RBF model is closest to the reference point and performs best. However, the RF model is the farthest from the reference point in the training stage, and the SVM model is farthest from the reference point in the testing stage, and their performances are relatively poor.

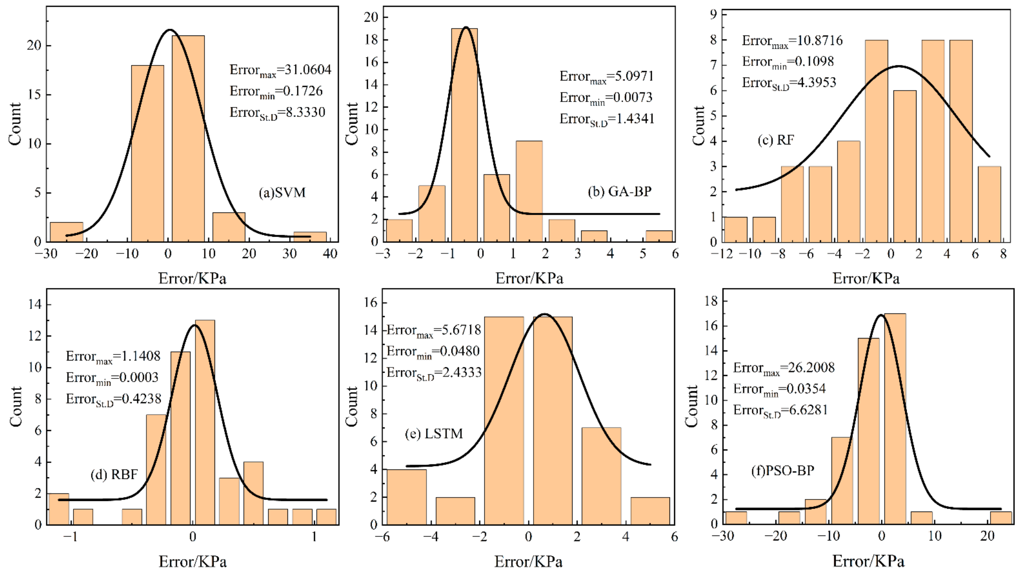

Figure 14. Error maps of the six models in the testing stage with the 59 total parameters.

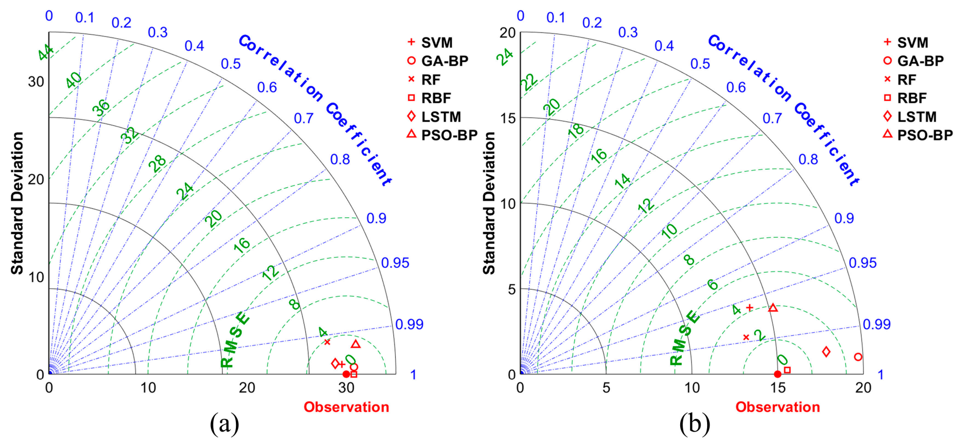

Figure 15. Taylor diagrams of the six machine learning models with the 59 total parameters. (**a**) Training stages; (**b**) testing stages.

The model applicability evaluation is shown in Figure 16. In addition to using the R^2, VAF, and WI in the model evaluation parameters, the mean absolute error (MAE) and mean square error (MSE) [72], which can further reflect the true state of the error, are introduced, and the relevant parameters in the formula refer to the same as above. An excellent model should have a larger R^2, VAF, and WI and a smaller MAE and MSE, and the larger the difference between the two is, the better the model. The model sorting in the training stage is RBF > GA-BP > SVM > LSTM > PSO-BP > RF, and the model sorting in the testing stage is RBF > LSTM > GA-BP > RF > PSO-BP > SVM. Thus, the sorting results are essentially consistent with the above Taylor diagram.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|Y_i - y_i| \quad (9)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - y_i)^2 \quad (10)$$

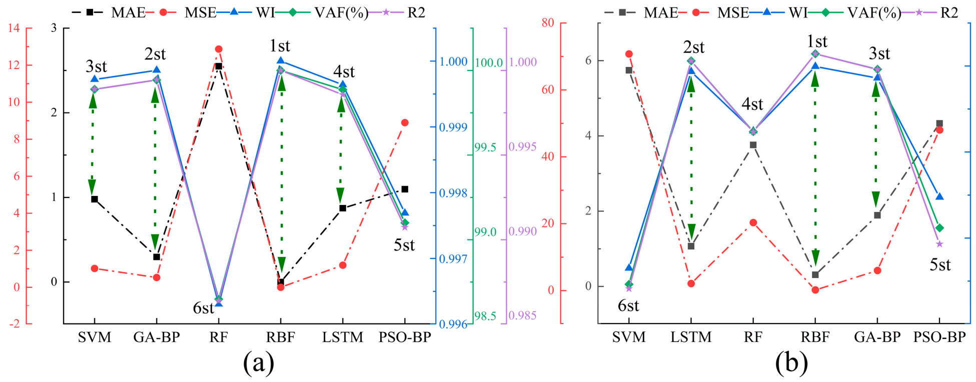

Figure 16. The model applicability evaluation comparison of the six models in terms of the statistical indicators with the 59 total parameters. (**a**) Training stages; (**b**) testing stages.

In summary, the RBF is the optimal model for the 59 parameters. In considering the importance of the testing stage to the true application of the model and that the ranking difference between the models in the Taylor diagram and the model applicability evaluation diagram is small, the comprehensive ranking of machine learning models with the total 59 parameters is RBF > LSTM > GA-BP > RF > PSO-BP > SVM.

5.1.2. Single Parameter Sensitivity Test of 59-Parameter Optimal Model

After obtaining the optimal model, a sensitivity analysis was introduced to determine the importance of each input parameter, and at the same time, the importance of each level of the three-level characteristic parameters was calculated. The ratio R_i of the factor default model testing error $RMSE_i$ to the full factor model testing error $RMSE$ is defined as the degree of influence of the i-th factor default on the output factor UCS, i.e., the sensitivity index [73,74].

$$R_i = \frac{RMSE_i}{RMSE} \tag{11}$$

where the larger R_i is, the more sensitive the factor.

The sensitivity changes of each factor during the training stage are shown in Figure 17a. Among them, the 2 macro parameters have the greatest sensitivity; 11 of the 38 meso parameters have a sensitivity greater than 1, 24 fluctuate between 0.5–1, and 3 are less than 0.5; 7 of the 19 micro parameters have a sensitivity greater than 1, 10 fluctuate between 0.5–1, and 2 are less than 0.5. The testing stage results are shown in Figure 17b. The macro parameters are still the most sensitive; 11 of the meso parameters have a sensitivity greater than 1, and 27 fluctuate between 0.5 and 1; 7 of the micro parameters have a sensitivity greater than 1, and 12 fluctuate between 0.5 and 1.

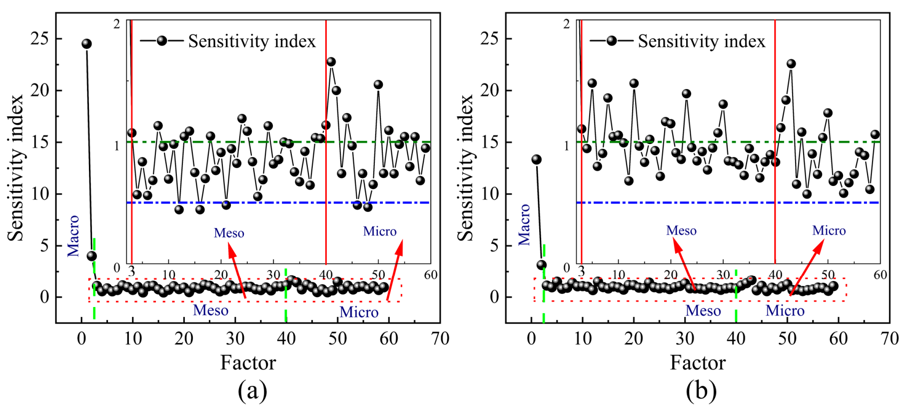

Figure 17. Parameter sensitivity analysis of RBF with the 59 total parameters. (**a**) Training stages; (**b**) testing stages.

As shown in Figure 18, compared with the overall sensitivity of the parameters in the training stage, the overall sensitivity of the macro parameters in the testing stage decreased, while the overall sensitivities of the meso and micro parameters increased slightly. The sensitivity distribution of the third-level parameters is extremely unbalanced. Although the number of meso and micro parameters occupies an absolute majority of the overall parameters, the sensitivity proportion in the model as a whole is extremely low.

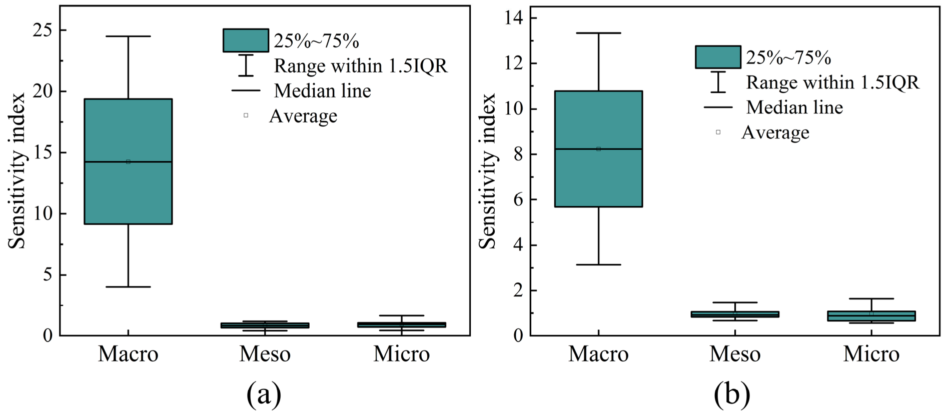

Figure 18. The three-level parameter sensitivity boxplots of the RBF with the 59 total parameters. (**a**) Training stages; (**b**) testing stages.

This may be because macro parameters are controlling factors, and both meso and micro parameters change due to macro parameter changes. At the same time, as controlling factors, the macro parameters have an effect on UCS and play a role as a guide or a catalyst in the model; that is, for the model accuracy and stability, these factors play an important role, while for the sensitivity analysis, they play a minor role. However, the sensitivity gap of the three-level characteristic parameters is too large, which is not conducive to explaining the influence of the parameter classification method of the model on the UCS prediction ability of the model. Therefore, this problem is solved by optimizing the number and proportion of parameters.

5.2. Parameter Optimization

The number and proportion of the 59 input parameters were optimized mainly through the following two methods: (1) gray correlation analysis between parameters and UCS, and (2) rough set analysis between parameters and UCS.

Gray correlation analysis is a mathematical method used to calculate the correlation coefficient of two sequences by studying the geometric proximity between subsequences and parent sequences [75,76]. It is suitable for more accurately locating correlation characteristics in 'poor information' and 'gray relationships', where the sample size is small and the change law is partially known. It has the advantages of requiring a small amount of calculation and not easily contradicting the results of a qualitative analysis. With the 59 input parameters as the subsequences, UCS as the parent sequence, and a resolution coefficient of $\rho = 0.5$, the specific results are shown in Figure 19a. The overall gray correlation coefficient of parameters is larger, and the performance of meso and micro parameters is better.

Rough set theory is a mathematical method used to analyze fuzzy and uncertain knowledge [77,78]. Under the premise of maintaining a certain classification ability, concept classification rules are derived through redundancy elimination without prior information. Using the rough set software ROSE2 V2.2 developed by Poznan University of Technology in Poland, based on the attribute importance reduction algorithm named manual search, the attribute importance between the subsequence and the parent sequence is calculated. The specific calculation results are shown in Figure 19b. The overall rough set attribute importance of macro parameters is relatively high, indicating that there is no redundancy; the importance of meso and micro parameter attributes is clearly graded, indicating that there is obvious redundancy, especially for meso parameters, and there are a number of parameters with an attribute importance of 0.

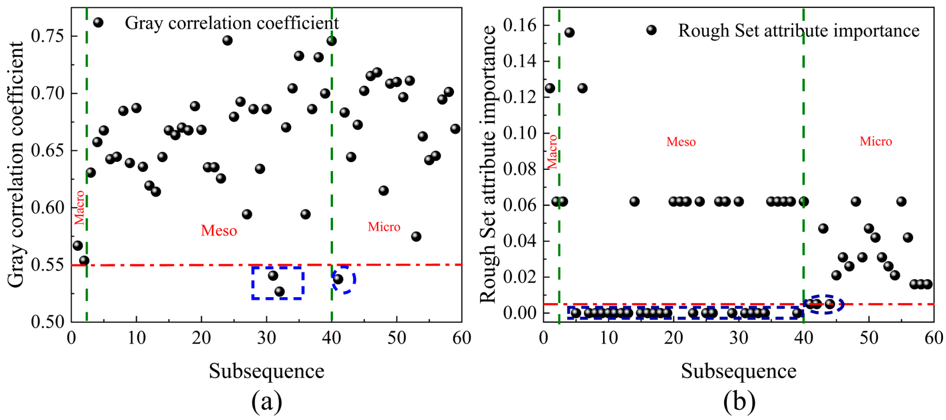

Figure 19. Parameter number and proportion optimization diagram of the 59 parameters. (**a**) Gray correlation analysis; (**b**) rough set analysis.

After comprehensively removing the related parameters of the gray correlation coefficient ≤ 0.5500 and rough set attribute importance ≤ 0.0050, the remaining parameters after elimination are shown in Table 14. At this time, there are 34 parameters in total, among which macro:meso:micro = 1:8:8.

Table 14. Three-level characteristic parameters after parameter number and proportion optimization.

Type	Quantity	Specific
Macro	2	X_1; X_2
Meso	16	X_3; X_4; X_6; X_{14}; X_{20}; X_{21}; X_{22}; X_{24}; X_{27}; X_{28}; X_{30}; X_{35}; X_{36}; X_{37}; X_{38}; X_{40}
Micro	16	X_{43}; X_{45}; X_{46}; X_{47}; X_{48}; X_{49}; X_{50}; X_{51}; X_{52}; X_{53}; X_{54}; X_{55}; X_{56}; X_{57}; X_{58}; X_{59}

5.3. Model Prediction of the 34 Optimized Parameters

To facilitate the comparative study of the performance of each model before and after parameter optimization, the hyperparameter settings in the six models with the 34 parameters after parameter optimization are kept consistent with those of the 59 parameter models before optimization.

We referred to the analysis of the performance indicators and related grade scores of the model in the training and testing stages before parameter optimization, the regression relationship analysis between the actual and predicted UCS, the error analysis of the model, the Taylor diagram of the optimal model selection process, and the model applicability evaluation analysis. After optimization, the evaluation analysis of each model shows that the comprehensive ranking of the six models with the 34 optimized parameters is RBF > LSTM > SVM > GA-BP > RF > PSO-BP. The optimal model is still RBF, and its performance indicators in the training and testing stages are shown in Table 15.

Table 15. The prediction performance evaluation of the RBF with the 34 optimized parameters in the training and testing stages.

Stage	R^2	RMSE	WI	VAF (%)
Training	1	1.37×10^{-5}	1	100
Testing	0.9868	3.7166	0.9967	98.8338

Single Parameter Sensitivity Test of the 34-Parameter Optimal Model

The sensitivity changes of each parameter in the training stage are shown in Figure 20a, among which the sensitivity values of two macro parameters are the largest, 15 of the

16 meso parameters have values greater than 1 and 1 is between 0.5 and 1, and the 16 micro parameters all have values greater than 1. The testing stage results are shown in Figure 20b, in which the macro parameters are still the most sensitive; 12 of the meso parameters have values greater than 1, and 4 fluctuate between 0.5–1; 12 of the micro parameters have values greater than 1, and 4 fluctuate between 0.5–1.

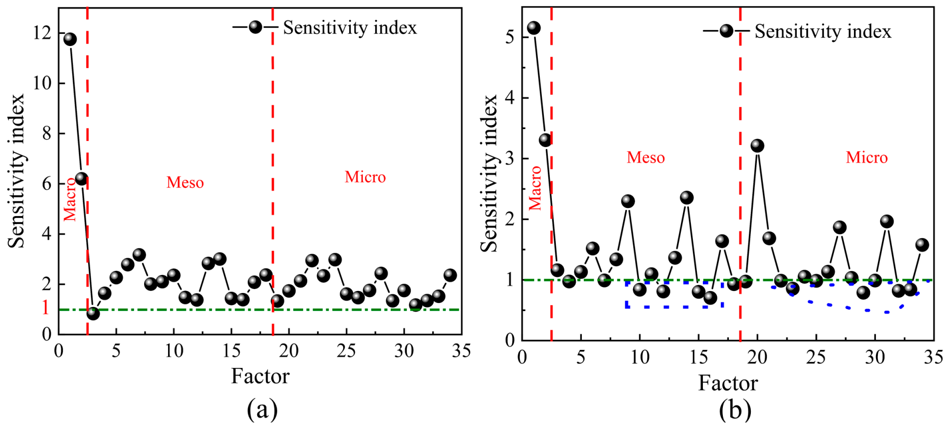

Figure 20. Parameter sensitivity analysis of the RBF model with the 34 optimized parameters. (**a**) Training stages; (**b**) testing stages.

As shown in Figure 21, compared with sensitivity values in the training stage, the sensitivity values of the three-level parameters in the testing stage decreased as a whole, among which the macro parameters decreased the most, while the meso and micro parameters decreased slightly overall. However, the proportion of meso and micro parameters in the overall sensitivity of the third-level parameters increased significantly, and the problem of unbalanced distribution of the sensitivity of the third-level parameters was better resolved.

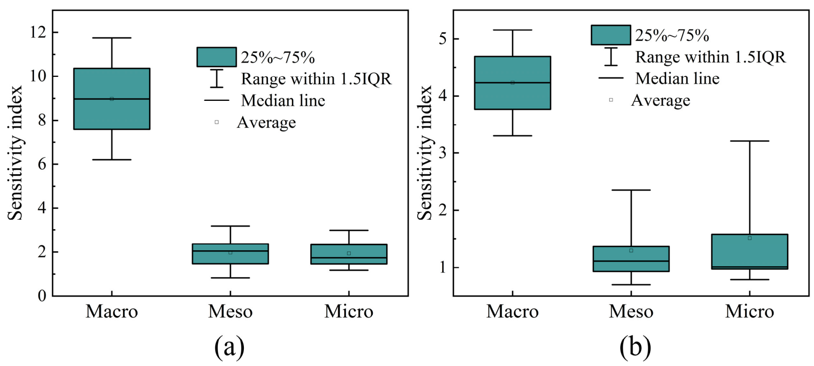

Figure 21. The three-level parameter sensitivity boxplots of the RBF model with the 34 optimized parameters. (**a**) Training stages; (**b**) testing stages.

In summary, under the premise of a limited loss of model accuracy and stability, the optimization of the number and proportion of parameters via gray correlation and rough set has greatly improved the sensitivity proportion of meso and micro parameters in the optimal BRF model, indicating that parameter optimization is conducive to the uniform distribution of the three-level parameter sensitivity.

5.4. Sensitivity Analysis of Three-Level Parameter Sets for 59 and 34 Parameter Models

A three-level parameter set sensitivity analysis was performed. That is, it would be informative to examine the impact of removing certain parameter sets (such as meso or micro parameter sets) on the RBF prediction accuracy of the optimal model. This helps in understanding the relative importance of each parameter scale in the model's predictive power. Specifically, this analysis is achieved by defaulting the three-level parameters of the model by level. The relevant analysis is as follows.

As shown in Figure 22, when the three-level parameters are defaulted by level, the sensitivity of the meso parameters in the training and testing stages of the 59-parameter model is extremely prominent. Among them, in the training and testing stages, respectively, the macro–meso parameter sensitivity difference is 3 and 0 orders of magnitude. The micro–meso parameter sensitivity difference is between 4 and 8 orders of magnitude. Micro–macro parameter sensitivities differ between 1 and 8 orders of magnitude. Similarly, the sensitivity of meso parameters in the 34-parameter model is extremely prominent. In the training and testing stages, respectively, the macro–meso parameter sensitivity difference is 1 and 0 orders of magnitude. The micro–meso parameter sensitivity difference is between 4 and 8 orders of magnitude. Micro–macro parameter sensitivities differ between 3 and 8 orders of magnitude.

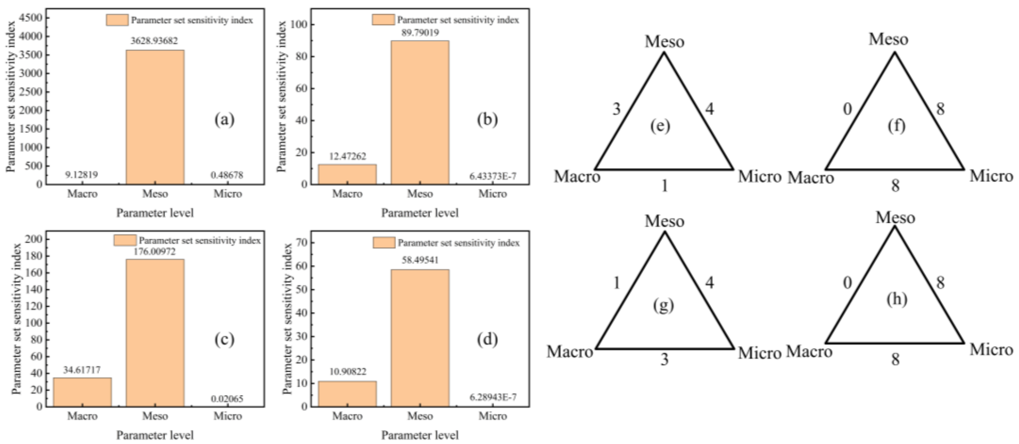

Figure 22. Default sensitivity changes by level of three-level parameters in the 59-parameter model and the 34-parameter model: (**a**,**e**) 59-parameter model training stage; (**b**,**f**) 59-parameter model testing stage; (**c**,**g**) 34-parameter model training stage; (**d**,**h**) 34-parameter model testing stage.

The expected phenomenon of relatively uniform distribution of three-level parameter sensitivity by level did not appear. This shows that after parameter classification, the overall sensitivity differences of parameters at each level are quite different by level. Parameter optimization also does not affect this difference, which is especially noticeable during the testing phase. The importance of meso parameters in the model is of primary importance, i.e., the mesoscale, as the interconnection link between three-level features, plays a decisive role in the influence of the model strength (in Figure 8, the direct effects B and BC that work at the meso scale are stronger than the direct effects A and AC that work at the micro scale). This proves that the interaction between three-level characteristics basically follows the control factors first from the macro scale to the meso scale, and then from the meso scale to the micro scale, thereby affecting the intensity. Alternatively, the strength deterioration occurs sequentially from the micro to the meso, and then from the meso to the macro. It further verifies the accuracy of the basic hypothesis—the three-level characteristic interaction hypothesis—while illustrating the effectiveness of the basic method—the new method of parameter expansion classification.

Of course, the parameter optimization performed in Section 5.2 mentioned above is still necessary because it can improve the sensitivity of any single parameter among the three-level parameters, even though it has a limited effect on level-by-level overall sensitivity optimization of tertiary parameters. At the same time, taking into account the number of parameters and the subsequent optimization work of the overall model, continuous parameter optimization is an inevitable choice.

5.5. Model Comparison and Limitation Analysis

5.5.1. Model Comparison

Table 16 lists the performance indicators of different machine learning methods used in different literatures with different soil UCS prediction accuracy. The current study has the highest R^2 and a relatively small RMSE value. Of course, other studies have also reported similar performance. At the same time, it is important to note that the datasets and machine learning methods used in each study are different, so direct comparison of performance values may not always be appropriate [79]. Nevertheless, it is obvious that the optimal model RBF constructed in this study using the basic hypothesis—the three-level characteristic interaction hypothesis—and the basic method—the new method of parameter expansion classification—provides accurate predictions with the largest R^2 value and relatively low RMSE value.

Table 16. Comparison of prediction performance of soil UCS from different studies.

Model	Soil	Parameters	Parameter Type	Performance	References
Before parameter optimization: SVM; GA-BP; RF; RBF; LSTM; PSO-BP	Saline soil in Lanzhou, China	59	Macro–meso–micro	SVM: R^2 = 0.9218; RMSE = 8.4160; GA-BP: R^2 = 0.9788; RMSE = 4.0217; RF: R^2 = 0.9738; RMSE = 4.5067; RBF: R^2 = 0.9998; RMSE = 0.4238; LSTM: R^2 = 0.9946; RMSE = 2.4451; PSO-BP: R^2 = 0.9366; RMSE = 6.9283;	This article
After parameter optimization: RBF	Saline soil in Lanzhou, China	34	Macro–meso–micro	RBF: R^2 = 0.9868; RMSE = 3.7166;	This article
EPR modelling A; B; C	Adelaide Industrial (AI) sand	1; 4; 4	All macro	EPR-A: R^2 = 0.714; RMSE = 1.461; EPR-B: R^2 = 0.885; RMSE = 0.374; EPR-C: R^2 = 0.939; RMSE = 0.273;	Ahenkorah et al. [80]
OEM; ANNs	Soils from around the world	9	All macro	OEM: R^2 = 0.61; MSE = 370,860; ANNs: R^2 = 0.65; MSE = 457,271;	Taffese and Abegaz. [81]
MGGP; ANNs	Geopolymer-stabilized clayey soil	9	All macro	MGGP: R^2 = 0.942; MSE = 2.366; ANNs: R^2 = 0.964; MSE = 1.500;	Soleimani et al. [82]
MR; ANNs; SVM	The soils selected were from Coimbra area	8	All macro	MR: R^2 = 0.59; RMSE = 0.56; ANNs: R^2 = 0.91; RMSE = 0.26; SVM: R^2 = 0.93; RMSE = 0.23;	Tinoco et al. [83]
ERBF; RBF; POLY	Three different types of clayey soil	7	All macro	ERBF: R = 0.9938; RMSE = 0.2586; RBF: R = 0.9901; RMSE = 0.8679; POLY: R = 0.9737; RMSE = 1.6277;	Mozumder et al. [84]
BP	Sulfate silty sand from the central desert of Iran	4	All macro	BP: R^2 = 0.9917; RMSE = 0.037;	Ghorbani et al. [85]
FN; MARS	Cement-stabilized soil	7	All macro	FN: R = 0.95; RMSE = 0.34; MARS: R = 0.95; RMSE = 0.31;	Suman et al. [86]
SNN-LogS	India soil	5	All macro	SNN-LogS: R = 0.95184; MSE = 0.09021;	Tiwari and Satyam. [87]

Note: R: Pearson's correlation coefficient; R^2: determination coefficient; RMSE: root mean square error; MSE: mean square error; SVM: support vector machines; GA-BP: genetic algorithm optimized BP; RF: random forest; RBF: radial basis kernel function; LSTM: long short-term memory; PSO-BP: particle swarm optimization algorithm BP; EPR: evolutionary polynomial regression; OEM: optimizable ensemble technique; ANNs: artificial neural networks; MGGP: multi-gen genetic programming; MR: multiple regression; POLY: polynomial kernel function; ERBF: exponential radial basis kernel function; BP: back propagation; FN: functional networks; MARS: multivariate adaptive regression splines; SNN-LogS: artificial neural network (ANN) was combined with the cross validation (LOOCV) method as CNN, and logS was the activation function; Soils from around the world: stabilized soils utilizing a diverse set of stabilized soils collected from around the world, the data set includes a variety of soils from 12 nations in Africa, Asia, Europe, North America, and Oceania; The soils selected were from Coimbra area: (located near Coimbra city, Portugal), ranging from cohesive to cohesionless soils, organic to nonorganic soils, presenting different geotechnical properties; India soil: the soil was collected at a depth of 2.5 m at the Indore campus of the Indian institute of technology in Madhya Pradesh, India.

5.5.2. Analysis of Model Advantages and Limitations

This study considers different numbers and proportions of macro–meso–micro three-level characteristic variables before and after parameter optimization to predict the UCS of salinized frozen soil. Its advantages and limitations are as follows:

Advantages:

(1) High model accuracy: the overall accuracy of the machine learning model built based on the three-level characteristic interaction hypothesis and the new method of parameter expansion classification is higher; in particular, the optimal model has the highest accuracy.
(2) The model parameters are highly interpretable: the model constructed using the new method has a basis for the expansion and classification of input parameters, and the boundaries between parameters are clear. This greatly increases the interpretability of parameters and can provide a reference for subsequent model parameter selection.
(3) There is a large space for model optimization: the current model is only a preliminary exploration of a new method for expanding and classifying model parameters, and there is a lot of room for optimization. Among them, there is a lot of room for optimization in terms of compressing the number of model parameters, optimizing parameter proportions, simplifying parameter construction, further improving model accuracy, and increasing practicality.

Limitations:

(1) Insufficient model transferability: there are currently relatively few data on multi-level parameters of salinized frozen soil in other literature, so it is impossible to obtain diversified data from different literature to verify the transferability of the constructed model.
(2) The cost of data acquisition is high: the data collection process in the model requires different experiments, which is relatively cumbersome, and the cost of data set acquisition is high.
(3) The parameters are complex and the model is not practical enough: although the number of model parameters has been reduced after parameter optimization, the number of current model parameters is still too large, and the structure is relatively complex, which will increase the difficulty of practical application, thus leading to the model's insufficient practicality.
(4) There are limitations in the generalizability of the conclusions: the results currently obtained are only applicable to the soil samples used in this study and may not be considered as a general rule for other data sets. It is unclear whether they can be generalized to other soil bodies and other materials.

In future work, the authors will collect a continuously updated and easily accessible database containing a variety of soil types to improve the generalizability of the proposed model. The database will include data samples containing more input variables, such as other macro-control factors and meso and micro parameters obtained by other means. We will collect soil shear strength (c, φ), expand the output from UCS to more strength parameters, or add frost heave, etc., as output parameters. At the same time, we will strengthen the continuous optimization of hyperparameters to improve model accuracy. In addition, we will continue to expand the model types, add predictive performance comparisons with other artificial intelligence models, and focus on detailed discussions of the algorithms used in these models. Finally, the proposed model is promoted to be incorporated into the construction system and the feasibility of applying the current model to the practice of saline frozen soil engineering is explored.

6. Conclusions and Summary

This paper takes salinized frozen soil as the research object. The response of the three-level characteristic parameters of macro–meso–micro was analyzed through experiments. The basic hypothesis—the three-level characteristic interaction hypothesis—was proposed.

And for the needs of application in subsequent machine learning models, a basic method—a new method of parameter expansion classification—is proposed. A model database was constructed through the data obtained from the experiment, and six models including SVM, GA-BP, RF, RBF, LSTM, and PSO-BP were applied. The UCS prediction of the salinized frozen soil based on the machine learning model based on macro–meso–micro three-level characteristic response is realized. We answer the problem of parameter selection at the input end of the machine learning model and the interpretability of model parameters. The main conclusions are as follows:

(1) In the experiment, with the increase in the control factors (number of freeze–thaw cycles, salt content), the macro, meso, and micro parameters all showed obvious stages of characteristics change. Taking the number of freeze–thaw cycles as 5 and 30 as the node, it can be divided into three stages. According to the changes in each stage, the first, second, and third stages are named the adjustment period, the dynamic fluctuation period, and the stable period. The fluctuation characteristics of each stage of the third-level parameters correspond well and show synchronized response characteristics.
(2) The ranking of the six machine learning models in terms of the UCS prediction with 59 parameters is RBF > LSTM > GA-BP > RF > PSO-BP > SVM. The optimal RBF model has the best prediction performance for UCS, with the largest R^2, WI, and VAF values and the smallest RMSE value. However, in view of the large difference in the sensitivity distribution of the third-level parameters, the RBF model needs to be further improved.
(3) The massive number of input parameters and obvious proportional differences in the use of the model parameter expansion classification method are the main reasons for the large sensitivity gap of the third-level parameters of the RBF model with 59 parameters. Through gray correlation and rough set analysis between parameters and UCS, optimizing the total number of parameters and the proportion of three-level parameters can effectively solve this problem.
(4) The ranking of the six machine learning models in terms of the UCS prediction with 34 parameters is RBF > LSTM > SVM > GA-BP > RF > PSO-BP, and the optimal model is still RBF. The accuracy and stability of the RBF model are slightly lower, but the sensitivity distribution of the three-level parameters is more reasonable, which can better reflect the macro–meso–micro three-level characteristic response of parameters and is more effective for UCS prediction.
(5) The actual performance of the 59- and 34-parameter models shows that the comprehensive macro–meso–micro three-level characteristic response of the soil can effectively improve the UCS prediction ability of the model. It is proven that it is necessary to expand and classify the input parameters in the soil UCS machine learning model prediction based on the basic hypothesis and basic methods. This approach not only considers the parameters more comprehensively and makes the logical relationship between parameters clearer and more interpretable but also helps to improve the prediction accuracy and stability of the model.

Author Contributions: Conceptualization, H.Z.; methodology, H.Z.; software, H.Z.; formal analysis, H.B.; investigation, H.Z.; data curation, H.Z.; writing—original draft preparation, H.Z.; writing—review and editing, H.B.; visualization, H.Z.; supervision, H.B.; project administration, H.B.; funding acquisition, H.B. All authors have read and agreed to the published version of the manuscript.

Funding: The work was funded by the Independent Project of the State Key Laboratory of Frozen Soil Engineering (No. SKLFSE-ZT-202211).

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

Acknowledgments: The authors greatly appreciate the constructive comments and language help from Yan Lu from Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences.

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

References

1. Meng, F.D.; Zhai, Y.; Li, Y.B.; Zhao, R.F.; Li, Y.; Gao, H. Research on the effect of pore characteristics on the compressive properties of sandstone after freezing and thawing. *Eng. Geol.* **2021**, *286*, 106088. [CrossRef]
2. *Gb/t 50176-2016*; Code for Thermal Design of Civil Building. China Planning Press: Beijing, China, 2016; pp. 147–152.
3. Lv, Q.F.; Jiang, L.S.; Ma, B.; Zhao, B.H.; Huo, Z.S. A study on the effect of the salt content on the solidification of sulfate saline soil solidified with an alkali-activated geopolymer. *Constr. Build. Mater.* **2018**, *176*, 68–74. [CrossRef]
4. You, Z.; Lai, Y.; Zhang, M.; Liu, E. Quantitative analysis for the effect of microstructure on the mechanical strength of frozen silty clay with different contents of sodium sulfate. *Environ. Earth Sci.* **2017**, *76*, 143. [CrossRef]
5. Li, K.; Li, Q.; Liu, C. Impacts of Water Content and Temperature on the Unconfined Compressive Strength and Pore Characteristics of Frozen Saline Soils. *KSCE J. Civ. Eng.* **2022**, *26*, 1652–1661. [CrossRef]
6. Shen, M.D.; Zhou, Z.W.; Zhang, S.J. Effect of stress path on mechanical behaviours of frozen subgrade soil. *Road Mater. Pavement Des.* **2022**, *23*, 1061–1090. [CrossRef]
7. Wan, X.S.; Lai, Y.M.; Wang, C. Experimental study on the freezing temperatures of saline silty soils. *Permafr. Periglac. Process.* **2015**, *26*, 175–187. [CrossRef]
8. Wallis, S. Freezing under the sea rescues Oslofjord highway tunnel. *Tunnel* **1999**, *8*, 19–26.
9. Wang, H.; Tong, M. Properties and field application of the grouting material for water blocking during thawing of frozen wall of deep sand layer. *Arab. J. Geosci.* **2021**, *14*, 1429. [CrossRef]
10. Zhang, Y.; Yang, Z.H.; Liu, J.K.; Fang, J.H. Impact of cooling on shear strength of high salinity soils. *Cold Reg. Sci. Technol.* **2017**, *141*, 122–130. [CrossRef]
11. Carteret, R.D.; Buzzi, O.; Fityus, S.; Liu, X.F. Effect of naturally occurring salts on tensile and shear strength of sealed granular road pavements. *J. Mater. Civ. Eng.* **2014**, *26*, 04014010. [CrossRef]
12. Xiao, Z.A.; Lai, Y.M.; You, Z.M.; Zhang, M.Y. The phase change process and properties of saline soil during cooling. *Arab. J. Sci. Eng.* **2017**, *42*, 3923–3932. [CrossRef]
13. Lai, Y.M.; Xu, X.T.; Dong, Y.H.; Li, S.Y. Present situation and prospect of mechanical research on frozen soils in China. *Cold Reg. Sci. Technol.* **2013**, *87*, 6–18. [CrossRef]
14. Du, C.C.; Li, D.Q.; Ming, F.; Liu, Y.H.; Shi, X.Y. Wave propagation characteristics in frozen saturated soil. *Sci. Cold Arid Reg.* **2018**, *10*, 95–103.
15. Han, F.H.; Zhang, Z.Q. Properties of 5-year-old concrete containing steel slag powder. *Powder Technol.* **2018**, *334*, 27–35. [CrossRef]
16. Canivell, J.; Martin-del-Rio, J.J.; Alejandre, F.J.; García-Heras, J.; Jimenez-Aguilar, A. Considerations on the physical and mechanical properties of lime-stabilized rammed earth walls and their evaluation by ultrasonic pulse velocity testing. *Constr. Build. Mater.* **2018**, *191*, 826–836. [CrossRef]
17. Martín-del-Rio, J.J.; Canivell, J.; Falcon, R.M. The use of non-destructive testing to evaluate the compressive strength of a lime-stabilised rammed-earth wall: Rebound index and ultrasonic pulse velocity. *Constr. Build. Mater.* **2020**, *242*, 118060. [CrossRef]
18. Wang, J.N.; Huang, L.; Wu, C.Y.; Jiang, T. Mechanical properties and microstructure of saline soil solidified by alkali-activated steel slag. *Ceram.–Silikáty* **2021**, *66*, 339–346. [CrossRef]
19. Choobbasti, A.J.; Samakoosh, M.A.; Kutanaei, S.S. Mechanical properties soil stabilized with nano calcium carbonate and reinforced with carpet waste fibers. *Constr. Build. Mater.* **2019**, *211*, 1094–1104. [CrossRef]
20. Kim, J.; Choi, Y.C.; Choi, S. Fractal characteristics of pore structures in GGBFS-based cement pastes. *Appl. Surf. Sci.* **2018**, *428*, 304–314. [CrossRef]
21. Chen, X.D.; Wu, S.X.; Zhou, J.K. Influence of porosity on compressive and tensile strength of cement mortar. *Constr. Build. Mater.* **2013**, *40*, 869–874. [CrossRef]
22. Atzeni, C.; Pia, G.; Sanna, U. A geometrical fractal model for the porosity and permeability of hydraulic cement pastes. *Constr. Build. Mater.* **2010**, *24*, 1843–1847. [CrossRef]
23. Zhu, Z.D.; Huo, W.W.; Sun, H.; Ma, B.R.; Yang, L. Correlations between unconfined compressive strength, sorptivity and pore structures for geopolymer based on SEM and MIP measurements. *J. Build. Eng.* **2023**, *67*, 106011. [CrossRef]
24. Zhang, X.W.; Kong, L.W.; Yang, A.W.; Sayem, H.M. Thixotropic mechanism of clay: A microstructural investigation. *Soils Found.* **2017**, *57*, 23–35. [CrossRef]
25. Gao, Q.F.; Jrad, M.; Hattab, M.; Fleureau, J.M.; Ameur, L.I. Pore morphology, porosity, and pore size distribution in kaolinitic remolded clays under triaxial loading. *Int. J. Geomech.* **2020**, *20*, 04020057. [CrossRef]
26. Jia, R.; Lei, H.Y.; Li, K. Compressibility and microstructure evolution of different reconstituted clays during 1D compression. *Int. J. Geomech.* **2020**, *20*, 04020181. [CrossRef]

27. Liu, X.Y.; Zhang, X.W.; Kong, L.W.; Wang, G.; Liu, H.H. Formation mechanism of collapsing gully in southern China and the relationship with granite residual soil: A geotechnical perspective. *Catena* **2022**, *210*, 105890. [CrossRef]
28. Zhang, X.W.; Wang, G.; Liu, X.Y.; Xu, Y.Q.; Kong, L.W. Microstructural analysis of pore characteristics of natural structured clay. *Bull. Eng. Geol. Environ.* **2022**, *81*, 473. [CrossRef]
29. Hassanien, A.E.; Darwish, A.; Abdelghafar, S. Machine learning in telemetry data mining of space mission: Basics, challenging and future directions. *Artif. Intell. Rev.* **2020**, *53*, 3201–3230. [CrossRef]
30. Moein, M.M.; Saradar, A.; Rahmati, K.; Mousavinejad, S.H.G.; Bristow, J.; Aramali, V.; Karakouzian, M. Predictive models for concrete properties using machine learning and deep learning approaches: A review. *J. Build. Eng.* **2023**, *63*, 105444. [CrossRef]
31. Baghbani, A.; Choudhury, T.; Costa, S.; Reiner, J. Application of artificial intelligence in geotechnical engineering: A state-of-the-art review. *Earth-Sci. Rev.* **2022**, *228*, 103991. [CrossRef]
32. Dehghanbanadaki, A.; Sotoudeh, M.A.; Golpazir, I.; Keshtkarbanaeemoghadam, A.; Ilbeigi, M. Prediction of geotechnical properties of treated fibrous peat by artificial neural networks. *Bull. Eng. Geol. Environ.* **2019**, *78*, 1345–1358. [CrossRef]
33. Eyo, E.U.; Abbey, S.J. Machine learning regression and classification algorithms utilised for strength prediction of OPC/by-product materials improved soils. *Constr. Build. Mater.* **2021**, *284*, 122817. [CrossRef]
34. Ngo, A.Q.; Nguyen, L.Q.; Tran, V.Q. Developing interpretable machine learning-Shapley additive explanations model for unconfined compressive strength of cohesive soils stabilized with geopolymer. *PLoS ONE* **2023**, *18*, e0286950. [CrossRef] [PubMed]
35. Shafiei, A.; Aminpour, M.; Hasanzadehshooiili, H.; Ghorbani, A.; Nazem, M. Mechanical characterization of marl soil treated by cement and lignosulfonate under freeze-thaw cycles: Experimental studies and machine-learning modeling. *Bull. Eng. Geol. Environ.* **2023**, *82*, 200. [CrossRef]
36. Zeini, H.A.; Al-Jeznawi, D.; Imran, H.; Bernardo, L.F.A.; Al-Khafaji, Z.; Ostrowski, K.A. Random Forest Algorithm for the Strength Prediction of Geopolymer Stabilized Clayey Soil. *Sustainability* **2023**, *15*, 1408. [CrossRef]
37. Eyo, E.U.; Abbey, S.J.; Booth, C.A. Strength predictive modelling of soils treated with calcium-based additives blended with eco-friendly pozzolans-A machine learning approach. *Materials* **2022**, *15*, 4575. [CrossRef] [PubMed]
38. Tran, V.Q. Hybrid gradient boosting with meta-heuristic algorithms prediction of unconfined compressive strength of stabilized soil based on initial soil properties, mix design and effective compaction. *J. Clean. Prod.* **2022**, *355*, 131683. [CrossRef]
39. Zhang, G.B.; Ding, Z.Q.; Wang, Y.F.; Fu, G.H.; Wang, Y.; Xie, C.F.; Zang, Y.; Zhao, X.; Lu, X.Y.; Wang, X.Y. Performance Prediction of Cement Stabilized Soil Incorporating Solid Waste and Propylene Fiber. *Materials* **2022**, *15*, 4250. [CrossRef]
40. Bing, H.; Zhang, Y.; Ma, M. Impact of desalination on physical and mechanical properties of Lanzhou loess. *Eurasian Soil Sci.* **2017**, *50*, 1444–1449. [CrossRef]
41. Lai, Y.; Liao, M.; Hu, K. A constitutive model of frozen saline sandy soil based on energy dissipation theory. *Int. J. Plast.* **2016**, *78*, 84–113. [CrossRef]
42. *ASTM D2487–17*; Standard Practice for Classification of Soils for Engineering Purposes (Unified Soil Classification System). ASTM International: West Conshohocken, PA, USA, 2020.
43. *ASTM-D560/D560M*; Standard Test Methods for Freezing and Thawing Com-Pacted Soil-Cement Mixtures. ASTM International: West Conshohocken, PA, USA, 2016.
44. *ASTM-D2166*; Standard Test Method for Unconfined Compressive Strength of Cohesive Soil. ASTM International: West Conshohocken, PA, USA, 2016.
45. Brignoli, E.G.; Gotti, M.; Stokoe, K.H. Measurement of shear waves in laboratory specimens by means of piezoelectric transducers. *Geotech. Test. J.* **1996**, *19*, 384–397.
46. Leong, E.C.; Cahyadi, J.; Rahardjo, H. Measuring shear and compression wave velocities of soil using bender–extender elements. *Can. Geotech. J.* **2009**, *46*, 792–812. [CrossRef]
47. Liu, C.; Shi, B.; Zhou, J.; Tang, C. Quantification and characterization of microporosity by image processing, geometric measurement and statistical methods: Application on SEM images of clay materials. *Appl. Clay Sci.* **2011**, *54*, 97–106. [CrossRef]
48. Liu, C.; Tang, C.S.; Shi, B.; Suo, W.B. Automatic quantification of crack patterns by image processing. *Comput. Geosci.* **2013**, *57*, 77–80. [CrossRef]
49. Gu, K.; Shi, B.; Liu, C.; Jiang, H.; Li, T.; Wu, J. Investigation of land subsidence with the combination of distributed fiber optic sensing techniques and microstructure analysis of soils. *Eng. Geol.* **2018**, *240*, 34–47. [CrossRef]
50. Tang, C.S.; Lin, L.; Cheng, Q.; Zhu, C.; Wang, D.W.; Lin, Z.Y.; Shi, B. Quantification and characterizing of soil microstructure features by image processing technique. *Comput. Geotech.* **2020**, *128*, 103817. [CrossRef]
51. Liu, X.; Deng, W.; Wang, S.; Liu, B.; Liu, Q. Experimental investigation on microstructure and surface morphology deterioration of limestone exposed on acidic environment. *Constr. Build. Mater.* **2023**, *377*, 131065. [CrossRef]
52. Tang, C.; Shi, B.; Liu, C.; Zhao, L.; Wang, B. Influencing factors of geometrical structure of surface shrinkage cracks in clayey soils. *Eng. Geol.* **2008**, *101*, 204–217. [CrossRef]
53. Jiang, N.; Li, H.; Liu, Y.; Li, H.; Wen, D. Pore microstructure and mechanical behaviour of frozen soils subjected to variable temperature. *Cold Reg. Sci. Technol.* **2023**, *206*, 103740. [CrossRef]
54. Lawal, A.I.; Idris, M.A. An artificial neural network-based mathematical model for the prediction of blast-induced ground vibrations. *Int. J. Environ. Stud.* **2020**, *77*, 318–334. [CrossRef]

55. Somvanshi, M.; Chavan, P.; Tambade, S.; Shinde, S.V. A review of machine learning techniques using decision tree and support vector machine. In Proceedings of the 2016 International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 12–13 August 2016; pp. 1–7. [CrossRef]
56. Liu, Y.Q.; Jiang, C.Q.; Lu, C.P.; Wang, Z.; Che, W.L. Increasing the Accuracy of Soil Nutrient Prediction by Improving Genetic Algorithm Backpropagation Neural Networks. *Symmetry* **2023**, *15*, 151. [CrossRef]
57. Liu, D.W.; Liu, C.; Tang, Y.; Gong, C. A GA-BP neural network regression model for predicting soil moisture in slope ecological protection. *Sustainability* **2022**, *14*, 1386. [CrossRef]
58. Dai, Y.; Khandelwal, M.; Qiu, Y.G.; Zhou, J.; Monjezi, M.; Yang, P.X. A hybrid metaheuristic approach using random forest and particle swarm optimization to study and evaluate backbreak in open-pit blasting. *Neural Comput. Appl.* **2022**, *34*, 6273–6288. [CrossRef]
59. Zhang, P.; Yin, Z.Y.; Jin, Y.F.; Chan, T.H.T. A novel hybrid surrogate intelligent model for creep index prediction based on particle swarm optimization and random forest. *Eng. Geol.* **2020**, *265*, 105328. [CrossRef]
60. Imanian, H.; Shirkhani, H.; Mohammadian, A.; Cobo, J.H.; Payeur, P. Spatial interpolation of soil temperature and water content in the land-water interface using artificial intelligence. *Water* **2023**, *15*, 473. [CrossRef]
61. Huo, D.Y.; Chen, J.S.; Zhang, H.; Shi, Y.R.; Wang, T.Y. Intelligent prediction for digging load of hydraulic excavators based on RBF neural network. *Measurement* **2023**, *206*, 112210. [CrossRef]
62. Bui, D.T.; Prakash, I.A. Sustainable Development of Urban Underground Space Using Deep Learning Method Based on LSTM at Substation Site in Southern Vietnam. In *Lecture Notes in Civil Engineering*; Springer: Berlin/Heidelberg, Germany, 2021; Volume 187, pp. 127–142.
63. Zamani, M.G.; Nikoo, M.R.; Rastad, D.; Nematollahi, B. A comparative study of data-driven models for runoff, sediment, and nitrate forecasting. *J. Environ. Manag.* **2023**, *341*, 118006. [CrossRef] [PubMed]
64. Huang, F.N.; Zhang, Y.K.; Zhang, Y.; Shangguan, W.; Li, Q.L.; Li, L.; Jiang, S.J. Interpreting Conv-LSTM for Spatio-Temporal Soil Moisture Prediction in China. *Agriculture* **2023**, *13*, 971. [CrossRef]
65. Mulumba, D.M.; Liu, J.K.; Hao, J.; Zheng, Y.N.; Liu, H.Q. Application of an Optimized PSO-BP Neural Network to the Assessment and Prediction of Underground Coal Mine Safety Risk Factors. *Appl. Sci.* **2023**, *13*, 5317. [CrossRef]
66. Wang, X.F.; Dong, X.P.; Zhang, Z.S.; Zhang, J.M.; Ma, G.W.; Yang, X. Compaction quality evaluation of subgrade based on soil characteristics assessment using machine learning. *Transp. Geotech.* **2022**, *32*, 100703. [CrossRef]
67. Fallahi, S.; Taghadosi, M. Quantum-behaved particle swarm optimization based on solitons. *Sci. Rep.* **2022**, *12*, 13977. [CrossRef] [PubMed]
68. Benemaran, R.S.; Esmaeili-Falak, M. Predicting the Young's modulus of frozen sand using machine learning approaches: State-of-the-art review. *Geomech. Eng.* **2023**, *34*, 507–527. [CrossRef]
69. Kardani, N.; Aminpour, M.; Raja, M.N.A.; Kumar, G.; Bardhan, A.; Nazem, M. Prediction of the resilient modulus of compacted subgrade soils using ensemble machine learning methods. *Transp. Geotech.* **2022**, *36*, 100827. [CrossRef]
70. Yu, Z.; Shi, X.Z.; Zhou, J.; Gou, Y.G.; Huo, X.F.; Zhang, J.H.; Armaghani, D.J. A new multikernel relevance vector machine based on the HPSOGWO algorithm for predicting and controlling blast-induced ground vibration. *Eng. Comput.* **2022**, *38*, 1905–1920. [CrossRef]
71. Roshan, M.J.; Rashid, A.S.A.; Wahab, N.A.; Tamassoki, S.; Jusoh, S.N.; Hezmi, M.A.; Daud, N.N.N.; Apandi, N.M.; Azmi, M. Improved methods to prevent railway embankment failure and subgrade degradation: A review. *Transp. Geotech.* **2022**, *37*, 100834. [CrossRef]
72. Rebouh, R.; Boukhatem, B.; Ghrici, M.; Tagnit-Hamou, A. A practical hybrid NNGA system for predicting the compressive strength of concrete containing natural pozzolan using an evolutionary structure. *Constr. Build. Mater.* **2017**, *149*, 778–789. [CrossRef]
73. Liong, S.Y.; Lim, W.H.; Paudyal, G.N. River stage forecasting in Bangladesh: Neural network approach. *J. Comput. Civ. Eng.* **2000**, *14*, 1–8. [CrossRef]
74. Suthar, M. Applying several machine learning approaches for prediction of unconfined compressive strength of stabilized pond ashes. *Neural Comput. Appl.* **2020**, *32*, 9019–9028. [CrossRef]
75. Tong, S.C.; Li, G.R.; Li, X.L.; Li, J.F.; Zhai, H.; Zhao, J.Y.; Zhu, H.L.; Liu, Y.B.; Chen, W.T.; Hu, X.S. Soil Water Erosion and Its Hydrodynamic Characteristics in Degraded Bald Patches of Alpine Meadows in the Yellow River Source Area, Western China. *Sustainability* **2023**, *15*, 8165. [CrossRef]
76. Yao, M.; Wang, Q.; Yu, Q.B.; Wu, J.Z.; Li, H.; Dong, J.Q.; Xia, W.T.; Han, Y.; Huang, X.L. Mechanism Study of Differential Permeability Evolution and Microscopic Pore Characteristics of Soft Clay under Saturated Seepage: A Case Study in Chongming East Shoal. *Water* **2023**, *15*, 968. [CrossRef]
77. Wong, M.; Parker, G. Reanalysis and correction of bed-load relation of Meyer-Peter and Müller using their own database. *J. Hydraul. Eng.* **2006**, *132*, 1159–1168. [CrossRef]
78. Ang, J.C.; Tang, J.Y.; Chung, B.Y.H.; Chong, J.W.; Tan, R.R.; Aviso, K.B.; Chemmangattuvalappil, N.G.; Thangalazhy-Gopakumar, S. Development of predictive model for biochar surface properties based on biomass attributes and pyrolysis conditions using rough set machine learning. *Biomass Bioenergy* **2023**, *174*, 106820. [CrossRef]

79. Nofalah, M.H.; Ghadir, P.; Hasanzadehshooiili, H.; Aminpour, M.; Javadi, A.A.; Nazem, M. Effects of binder proportion and curing condition on the mechanical characteristics of volcanic ash-and slag-based geopolymer mortars; machine learning integrated experimental study. *Constr. Build. Mater.* **2023**, *395*, 132330. [CrossRef]
80. Ahenkorah, I.; Rahman, M.M.; Karim, M.R.; Beecham, S. Unconfined compressive strength of MICP and EICP treated sands subjected to cycles of wetting-drying, freezing-thawing and elevated temperature: Experimental and EPR modelling. *J. Rock. Mech. Geotech.* **2023**, *15*, 1226–1247. [CrossRef]
81. Taffese, W.Z.; Abegaz, K.A. Prediction of compaction and strength properties of amended soil using machine learning. *Buildings* **2022**, *12*, 613. [CrossRef]
82. Soleimani, S.; Rajaei, S.; Jiao, P.; Sabz, A.; Soheilinia, S. New prediction models for unconfined compressive strength of geopolymer stabilized soil using multi-gen genetic programming. *Measurement* **2018**, *113*, 99–107. [CrossRef]
83. Tinoco, J.; Alberto, A.; Venda, P.; Correia, A.G.; Lemos, L.A. data-driven approach for q_u prediction of laboratory soil-cement mixtures. *Procedia Eng.* **2016**, *143*, 566–573. [CrossRef]
84. Mozumder, R.A.; Laskar, A.I.; Hussain, M. Empirical approach for strength prediction of geopolymer stabilized clayey soil using support vector machines. *Constr. Build. Mater.* **2017**, *132*, 412–424. [CrossRef]
85. Ghorbani, A.; Hasanzadehshooiili, H. Prediction of UCS and CBR of microsilica-lime stabilized sulfate silty sand using ANN and EPR models; application to the deep soil mixing. *Soils Found.* **2018**, *58*, 34–49. [CrossRef]
86. Suman, S.; Mahamaya, M.; Das, S.K. Prediction of maximum dry density and unconfined compressive strength of cement stabilised soil using artificial intelligence techniques. *Int. J. Geosynth. Ground Eng.* **2016**, *2*, 11. [CrossRef]
87. Tiwari, N.; Satyam, N. Coupling effect of pond ash and polypropylene fiber on strength and durability of expansive soil subgrades: An integrated experimental and machine learning approach. *J. Rock. Mech. Geotech.* **2021**, *13*, 1101–1112. [CrossRef]