

# Water permeability prediction of sponge city pavement materials based on different machine learning algorithms

ZHANG Xiong<sup>1</sup>, LV Xinyan<sup>1</sup>, JI Tao<sup>2</sup>, ZHONG Chen<sup>1,\*</sup>

<sup>1</sup>Key Laboratory of Advanced Civil Engineering Materials of Education Ministry, Tongji University, Shanghai, 201804, China

<sup>2</sup>School of Architecture and Civil Engineering, Nanjing Institute of Technology, Nanjing, 211167, China

**Abstract.** Permeable pavement material is one of the most important supporting materials in the construction of sponge city, and its water permeability is the most important performance index. The water permeability test of permeable pavement materials is a tedious and complicated experimental work. It is of great research significance to predict the water permeability of permeable pavement materials through structural parameters modeling. In this paper, the database is first established by experimental means, and then the prediction models of LASSO (Least absolute shrinkage and selection operator), SVR (Support vector regression) and GBR (Gradient Boosting Regression) machine learning algorithms are established. Through the four factors of particle size, particle size distribution, shape parameters and binder content predict the water permeability of sponge city pavement materials. The results show that different machine learning algorithms have different sensitivity to the distribution of data samples. The fitting effect of GBR model water permeability prediction is better than that of SVR and LASSO models. The test value-predicted value MSE is 0.0051 and  $R^2$  is 0.92, which can effectively predict the water permeability of sponge city pavement materials.

## 1 Introduction

To solve the problem of traditional cities, the concept of "sponge city" came into being in the era of water shortage, water quality pollution, urban waterlogging and other urban water problems in various countries around the world. It has great significance for accelerating urban modernization and ecological civilization construction. The pavement materials used in the construction of sponge city can be well used in the design of permeable pavement and the ecological purification system due to their good permeability and biocompatibility, and thus become the main supporting materials in the construction of sponge city<sup>[1-3]</sup>.

Water permeability is one of the most important properties of sponge city pavement materials. In the United States, Europe, Canada and other places, the permeability coefficient is used to reflect the water permeability of pavement materials. For the on-site permeability testing of the project, the penetration ring is mostly used, but the national standards have not been unified<sup>[4]</sup>. The results of the water permeability test using the different permeable ring show a large difference. At the same time, this detection method has high requirements on pavement resettlement points and test environment; For the laboratory permeability testing, two types of constant-head permeameters and variable-head permeameters are mainly used in domestic and overseas<sup>[5]</sup>. The permeability coefficient is converted through the permeation amount per unit time and the

length, width and height of the test piece to evaluate the water permeability performance, but there are no relevant regulations on the influence factors such as the test piece and the penetration device. Regardless of the method used to study the water permeability of pavement materials, obtaining a large amount of data through experiments requires a lot of manpower, material resources, and time, and is susceptible to external conditions, resulting in lower experimental efficiency and greater errors.

At present, most of the domestic and foreign sponge city materials permeation performance research models are physical models: Xinxin Li<sup>[6]</sup> et al. idealized permeable concrete into a three-phase composite material composed of coarse aggregate, mortar and ITZ, and used the zero thickness interface element to model the interface transition zone structure. The results show that based on the characteristic assumption of ITZ, this method can be used to predict the water permeability of concrete. The numerical results are in good agreement with the experimental results, and the model simulation results are stable; Masad<sup>[7]</sup> et al. used the finite difference method and the finite element method to analyze the water infiltration form inside the pavement structure, and validated the numerical scheme by simulating the fluid flow in the ideal microstructure in parallel cracks and stacked cubes. The results show that the numerical results are in good agreement with the closed-form solution and can be used to simulate the fluid flow in the microstructure of actual porous media,

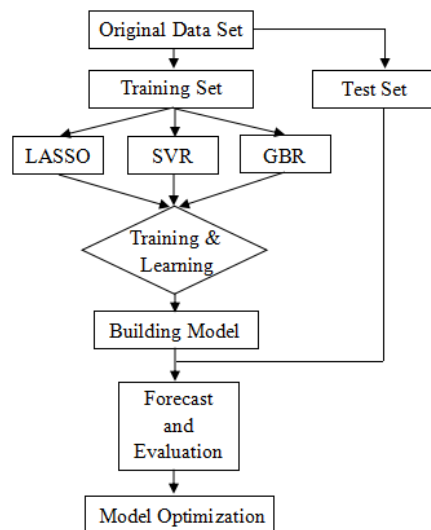
\* Corresponding author: zhongchen0227@163.com

but the permeability is greatly affected by the number of iterations and the resolution of the 3D image; Kuang<sup>[8]</sup> et al. used the finite element model to explore the relationship between the pore distribution characteristics and the penetration rate of the pavement material. The hydraulic conductivity(k) was measured by collecting test samples of permeable pavement on site, and an improved KCM model was proposed based on the pore structure model including the KCM model and used to predict the k value of the permeable pavement structural parameters. Combining the prediction result k with the continuous rainfall simulation model used in history can be used to provide a nomogram of the inspection of permeable pavement as a low-impact development (LID) infrastructure component. Although these physical models have achieved remarkable results in terms of theoretical thinking, mechanism of action, model construction, etc., there are too many presupposed assumptions and there are still some gaps between their presuppositions and the actual experimental conditions.

The machine learning method is to learn statistical rules in a large number of training samples by using the algorithm model, so as to make predictions for unknown events. With the rise of artificial intelligence in recent years, machine learning methods are gradually combined with the field of materials science and gradually promoted and applied. Use different machine learning methods for the large number of experimental data sets currently available to predict material performance parameters and provide theoretical guidance for material experiments and applications<sup>[9]</sup>. Tibshirani R(1996)<sup>[10]</sup> has proposed a new variable selection method called LASSO inspired by the "Bridge Regression" proposed by Frank and the "Nonnegative Garrote" proposed by Bireman. Its advantage lies in its fast computing power; SVR was invented by Boser, Guyon and Vapnik. It first appeared in the paper of the Conference on Computational Learning Theory in 1992<sup>[11]</sup>. It can convert non-linear problems into linear problems, thereby solving the problem of high-dimensional model construction; The GBR algorithm is based on the boosting algorithm framework<sup>[12]</sup>. It has performed well in large-scale data competitions in recent years (such as Tianchi, Kaggle, etc.) and has received great attention. Its advantage lies in its low deviation. In view of the many factors affecting the performance of sponge city pavement materials and their non-linear laws, the model based on the three machine learning algorithms of LASSO, SVR, and GBR was selected for the prediction study of the performance of sponge city pavement materials. The training samples they require are few and have high accuracy.

The machine learning algorithm is used to predict the permeation performance of sponge cities, as shown in Fig. 1. The original data set is divided into a training set and a test set. Different machine algorithm models are trained and learned through the training set, and then the trained machine algorithm model is used to predict the test set to obtain a regression fit. result. In this paper, three different machine learning algorithms, LASSO, SVR, and GBR, are used to establish a mathematical regression model based on the machine learning

algorithm through experimental data. Through the four performance parameters of particle size, particle size distribution, shape parameter and binder content, water permeability prediction is performed, and the prediction accuracy of the model is compared and evaluated.



**Fig. 1** Workflow of predictive model based on machine learning algorithm

## 2 Experimental methods and principle

### 2.1 Raw materials

The sponge city pavement material studied in this work is resin-based composite permeable material. The raw materials are mainly aggregate and binder. In order to make the performance indexes of the resin-based composite permeable material comparable, the same binder is used uniformly.

**Aggregate:** Different types of aggregates are selected nationwide. The basic physical parameters are shown in Table 1. Image analysis method is used to obtain the particle group parameters, as shown in Fig. 2.

The particle size of the aggregate is taken as the minimum ellipse short axis of the aggregate, as shown in Fig. 3.

Roundness is expressed as the ratio of the area of a circle of equal perimeter to the projected area of the particle projection, which characterizes how similar the particle projection is to the circle. The closer the value is to 1, the rounder the particles. Its definition is:

$$Roundness = \frac{Perimeter^2}{4 \times \pi \times Area} \quad (1)$$

**Binder:** The binder used in the test is a two-component binder, where component A is E44 thermosetting epoxy resin with an epoxy value of 0.41-0.47 and the liquid is white and transparent; component B is a polyamine type 593 alicyclic Family curing agent, the liquid is light yellow transparent. The curing of the binder needs to be carried out under dry conditions. The strength of one day after curing can reach 85% of the maximum strength. It has the advantages of rapid

prototyping test blocks for testing. Its basic properties are shown in Table 2.

**Table 1** Twelve kinds of aggregate basic performance parameters

No.	Bulk density	True density	Average grain diameter	Roundness
A	1.55	2.65	0.49	1.05
B	1.41	2.66	0.10	1.14
C	1.48	2.65	0.38	1.06
D	1.47	2.60	0.50	1.15
E	1.48	2.63	0.46	1.12
F	1.38	2.63	0.88	1.22
G	1.33	2.65	1.08	1.18
H	1.39	2.67	1.20	1.19
I	1.22	2.66	0.82	1.11
J	1.43	2.65	0.70	1.19
K	1.41	2.66	0.23	1.13
L	1.54	2.62	0.60	1.2

**Table 2** Properties of AB type epoxy resin

Viscosity (20°C) / (MPa·s)	Compressive strength /MPa	Tensile Strength /MPa	Bond strength /MPa	Elongation /%
60-100	≥35	>15	>3.0	8

### 2.2.1 Preparation

For the preparation and testing of resin-based composite permeable pavement material test pieces, refer to "Standard\_JG\_T376-2012 Sand-based Permeable Brick". Weigh a certain amount of sand according to the recipe and pour it into the mixing pot, add a certain amount of epoxy resin, curing agent and admixture, and then put it in a planetary mortar mixer to mix evenly to obtain a uniformly mixed epoxy resin mortar mixture. The frame after removing the middle ribbed plate with cement mortar triple mold is a plate-shaped mold for making road material test pieces. The epoxy resin mixture is evenly laid in it, scraped flat with a scraper, covered with a 160mm×132mm×10mm steel plate and 160mm×80mm×34mm steel pad is placed in the press to pressurize to 2.5MPa, and the static pressure is formed for 90s. The molded epoxy resin mortar mixture can be removed after 24 hours of natural curing to obtain a resin-based composite permeable pavement material sample, as shown in Fig. 4(a). In this work, the self-made water permeability tester (10cm×10cm×20cm) is used to test the water permeability of quantitative water (500ml) as shown in Fig. 4(b).

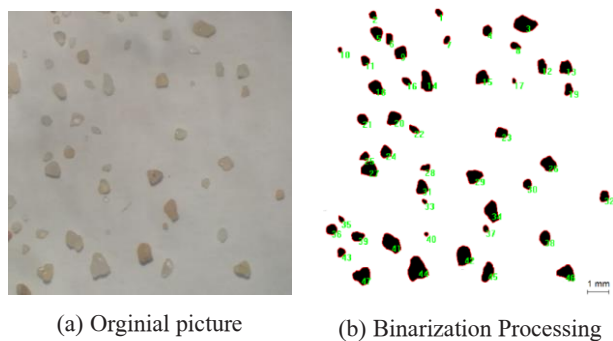
### 2.2.2 Test

This work uses a self-made water permeability tester to test the water permeability. The device is composed of transparent plexiglass cuboid (10cm × 10cm × 20cm) opened at the upper and lower ends, and the test material samples are bonded at the lower end, as shown in Fig. 4(c). In this experiment, the vertical seepage velocity of quantitative water is used to express the permeability of the sample. Quantitative water (500ml) is used, that is, the fixed height of the permeable material specimen is 5cm at the top. Record the penetration time of quantitative water from injection to the end of penetration, recorded as  $t_1$ , and fresh water is injected again to repeat the measurement and record the penetration time  $t_2$ . Water penetration rate  $v$  is calculated as follows:

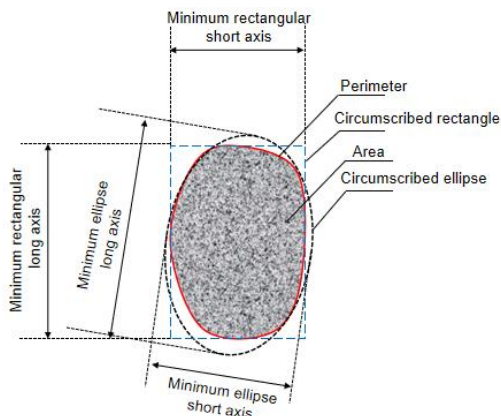
$$v = \frac{V}{S \cdot t} \tag{2}$$

$$\bar{t} = \frac{t_1 + t_2}{2} \tag{3}$$

In the formula:  
 $v$ — the water permeability rate of the test block, accurate to  $1.00 \times 10^{-2}$ cm/s;  
 $t$ — average time of water penetration, s;  
 $t_1$ — first water penetration time, s;  
 $t_2$ — second water penetration time, s;  
 $S$ — cross-sectional area,  $\text{cm}^2$ , where  $S$  is the cross-sectional area of the plexiglass rectangular cylinder, i.e.,  $100 \text{ cm}^2$ .



**Fig. 2** Image analysis to obtain aggregate particle group parameters



**Fig. 3** Aggregate particle size

## 2.2 Sample preparation

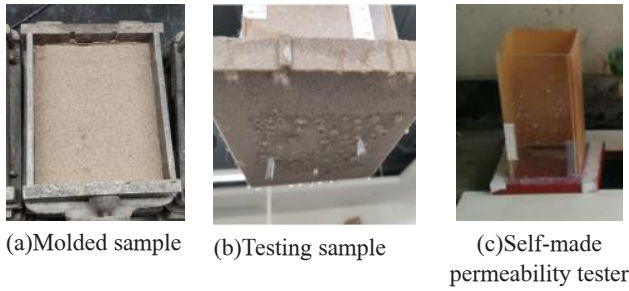


Fig. 4 Experimental sample preparation and testing

## 2.3 Principles and methods

### 2.3.1 LASSO<sup>[10,13]</sup>

LASSO was first proposed by Robert Tibshirani in 1996 and is a data dimensionality reduction method. The principle is to obtain a more refined model by constructing a penalty function, and automatically compress the independent variables that have little or no influence on the explanatory variables.

The LASSO parameter estimation is defined as follows:

$$\hat{\beta}(\text{lasso}) = \arg_{\beta} \min \left\| y - \sum_{j=1}^p x_j \beta_j \right\|^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (4)$$

In the formula:  $x_j$  and  $y$  represent explanatory variables and response variables respectively;  $\beta_j$  represents the regression coefficient;  $\lambda \in [0, \infty)$  is the harmonic parameter. The first part of the function can be regarded as an excellent expression of model fitting, and the second part is expressed as a penalty. Among them, the degree of LASSO regression complexity adjustment is controlled by the parameter  $\lambda$ : the larger the  $\lambda$ , the greater the amount of shrinkage, the greater the penalty for the linear model with more variables, and the fewer variables selected by the final model.

### 2.3.2 SVR<sup>[11,14]</sup>

SVR is a variant model of Support Vector Machine (SVM) proposed by Vapnik et al. The goal of SVR is to find a function that can well approximate the training instance by minimizing the prediction error, and when the error is minimized, maximize the flatness of the function to reduce the risk of fitting.

Suppose the sample pair of the training set in the sample space is  $\{(x_i, y_i)\}$ ,  $i = 1, 2, \dots, l$ ,  $l$  is the sample pair of the training set,  $x_i = [x_i^1, x_i^2, \dots, x_i^n]^T$ ,  $y_i \in \mathbb{R}$  are the corresponding output values. Now let  $f(x) = \omega^T \varphi(x) + a$ , where  $\omega$  is the weight vector and  $a$  is the threshold. SVR achieves the goal of regression by introducing the  $\varepsilon$  insensitive loss function. If the difference  $f(x_i) - y_i$  between the predicted value  $f(x_i)$  and the sample value  $y_i$  is less than the given  $\varepsilon$ , it is considered to be lossless. Then the SVR model is expressed as:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (5)$$

$C$  represents the penalty parameter,  $C > 0$ , when the value of  $C$  is large, the penalty for error classification increases, and when the value of  $C$  is small, the penalty for error classification decreases. The larger the  $C$  value, the better the fitting effect.  $\xi_i$  and  $\xi_i^*$  represent relaxation variables. For ease of processing, we refer to  $\alpha_i$  and  $\alpha_i^*$  in the Lagrange function and the kernel function  $K(x_i, x_j)$  to transform the original problem into a dual problem, and finally get the regression function:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x_j) + a \quad (6)$$

### 2.3.3 GBR<sup>[12,15]</sup>

GBR is an algorithm model that learns from its errors. It is essentially brainstorming, integrating a bunch of poor learning algorithms to learn. The core idea of the GBR model is to generate a CART tree based on the negative gradient direction fitting of the loss function. The GBR model uses the CART tree as a single model, can process multiple types of data, has low learning error.

Suppose the data set is  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ , the loss function is  $L(x_N, y_N)$ , and the number of leaf nodes in each regression tree is  $J$ , Divide its input space into  $J$  disjoint regions  $R_{1m}, R_{2m}, \dots, R_{jm}$ , and estimate a constant value  $k_{jm}$  for each region, the regression tree  $g_m(x)$  formula is:

$$g_m(x) = \sum_{j=1}^J (k_{jm} U), \quad x \in R_{jm} \quad (7)$$

$$U(x \in R_{jm}) = \begin{cases} 1, & x \in R_{jm} \\ 0, & \text{others} \end{cases} \quad (8)$$

Initial model:

$$p_0(x) = \arg \min \sum_{i=1}^n L(y_i, \rho) \quad (9)$$

The regression tree continuously iterates, producing residuals, for  $m = 1$  to  $M$ ,  $m$  represents the  $m$ -th tree, and the residuals generated at each step generate the regression tree again, where the gradient descent step size is:

$$\rho_m = \arg \min \sum_{i=1}^n L(y_i, p_{m-1}(x_i) + \rho g_m(x_i)) \quad (10)$$

After each step, the model can be updated,  $lr$  represents the learning efficiency:

Final output model  $p_m(x)$ :

$$p_m(x) = p_{m-1}(x) + lr * \rho_m g_m(x) \quad (11)$$

The prediction accuracy is mainly affected by the number of regression trees ( $M$ ), the depth of the tree ( $J$ ), and the learning efficiency ( $lr$ ), where the number of regression trees represents the number of basic learners, and the depth of the regression tree represents the number of nodes generated by the tree. The learning efficiency is set to prevent the model from overfitting to reduce the impact of each basic model on the final result.

## 2.4 Performance evaluation

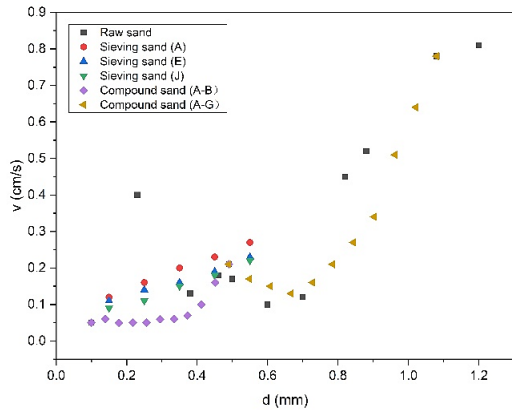


MSE (Mean Square Error) is a measure that reflects the degree of difference between the predicted value and the true value. And  $R^2$  is an indicator that measures the overall fit of the regression equation. The formula is as follows:

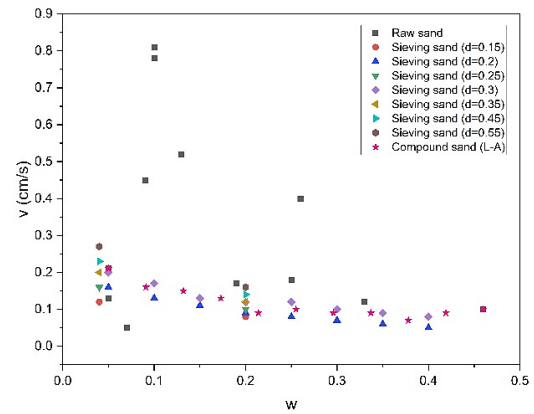
$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (13)$$

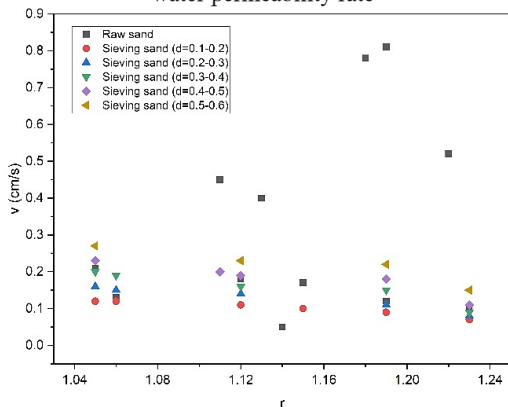
### 3 Experimental results and discussions



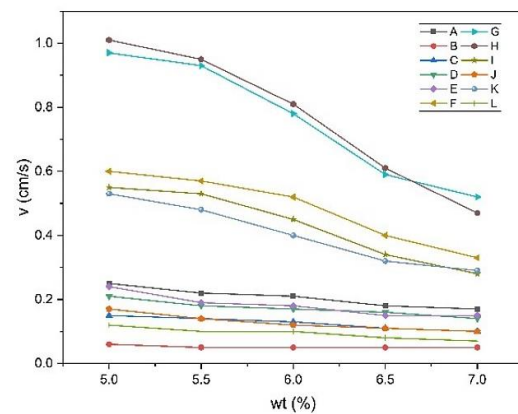
(a) Relationship between aggregate size and water permeability rate



(b) Relationship between aggregate standard deviation and water permeability rate



(c) Relationship between roundness of aggregate and water permeability rate



(d) Relationship between binder materials wt% and water permeability rate

**Fig. 5** Relationship between characteristic parameters and water penetration rate

Fig. 5(a) shows the relationship between water permeability rate and particle size when A-J 12 kinds of sand are used as aggregate substrates. It can be seen that the particle size of the raw sand increases and the water penetration rate shows an upward trend as a whole, but it does not show a linear characteristic. This is because the aggregate has batch differences, including gradation differences and granular differences. The experiment of sieving sands of A, E and J was selected, and the results showed that the water permeability of the three sands of A, E and J increased as the particle size increased. Mix B sand and G sand separately with A sand to thoroughly explore the effect of aggregate particle size on water permeability. Compared with the results of screening

In the formula:  $y_i$  represent true value;  $\hat{y}_i$  represent predicted value;  $\bar{y}$  represent average value;  $n$  represent the number of data samples.

sand A, with the addition of coarse particles (G sand), the preferred particle size increases, the water permeability rate is increasing; when fine particles (B sand) are added, the particle size is preferably reduced, and water permeability rate is reduced. The above results indicate that the aggregate particle size increases, the close packing between the particles decreases, the voids increase, and the water permeability rate also increases.

The effect of particle size distribution on the permeability of sponge city pavement materials is shown in Fig. 5(b). From the relationship between the standard deviation of twelve kinds of raw sand and its water permeability rate, when the aggregate is single graded, i.e. the standard deviation is close to 0, the water

permeability rate is higher, but due to the influence of particle size, the experimental results still show a large difference. Using sieved sand for further experiments, when the aggregate particle size is the same, the water permeability rate decreases with the increase of the standard deviation. Reducing the distribution range of the aggregate particle size is beneficial to increase the material permeability rate. When the single gradation A sand is mixed into the continuous gradation L sand, the finer A sand is filled into the void of L sand when the amount of incorporation is small, so the stacking structure is more compact and the water permeability rate is reduced to the minimum; When a large amount of A sand incorporation becomes the dominant aggregate, the standard deviation decreases, the particle size distribution becomes narrow, the porosity increases, so the water permeability rate increases.

The roundness is the most sensitive parameter among the shape parameters, which has a great influence on the material accumulation structure and the shape of the water flow channel. Fig. 5(c) shows the relationship between roundness and water permeability rate when 12 kinds of sand are used as raw sand and sieved sand aggregate substrate. It can be seen from the figure that if the impact of aggregate particle size is not avoided, the roundness of the raw sand and the data of the water permeability rate are scattered and disorderly: the permeability rate of sand F with poor roundness ( $r=1.22$ ) is  $0.81\text{cm/s}$ , while that of another sand L with poor roundness ( $r=1.23$ ) is only  $0.10\text{cm/s}$ , indicating that the impact of the particle size on the water penetration rate at this time is much greater than the roundness. The results show that the roundness increases and the permeability rate of pavement materials decreases. The worse the roundness is, i.e., the larger the R value is, at this time, the aggregate particles are irregular, the accumulation gap is smaller when they are stacked, and the water permeability rate is lower.

Fig. 5(d) shows the change of the water permeability rate of pavement materials with the amount of binder materials when A-L 12 kinds of sand are used as aggregate substrates. It is not difficult to see that with the increase of binder mixing amount, the permeable rate of pavement material decreases in different degrees. Among them, H sand used as aggregates had the largest decrease, the dosage increased from 5% to 7%, and the penetration rate decreased by 53.31%. The more the amount of binder material is added, the thicker the slurry layer coated on the surface of the aggregate is, the accumulation gap between the particles is reduced, and the water flow channel is also reduced, which ultimately leads to a decrease in the water permeability rate.

In summary, the particle size, particle size distribution, shape parameters, and the amount of binder material all have a great influence on the permeability of sponge city pavement materials. The water permeability rate increases with different amplitudes as the particle size increases, the standard deviation decreases, the roundness decreases, and the binder content decreases.

## 4 Establishment and verification of prediction model

In order to select a suitable prediction model for the permeability of sponge city pavement materials, combined with the previous experimental data, by writing the Python code of the prediction model, the three prediction models of LASSO, SVR and GBR were established, and the three models were compared and optimized through regression analysis. The prediction model of water permeability of sponge city pavement materials that is the most suitable for the four performance parameters of particle size, particle size distribution, shape parameter and binder content is obtained. The process of building a model for predicting water permeability of sponge city pavement materials with different machine algorithms is shown in Fig. 6: Among them, the four performance parameters of particle size( $d$ ), relative standard deviation( $w$ ), roundness( $r$ ), and binder content( $c$ ) are independent variables, and the water permeability rate( $v$ ) is the dependent variable.

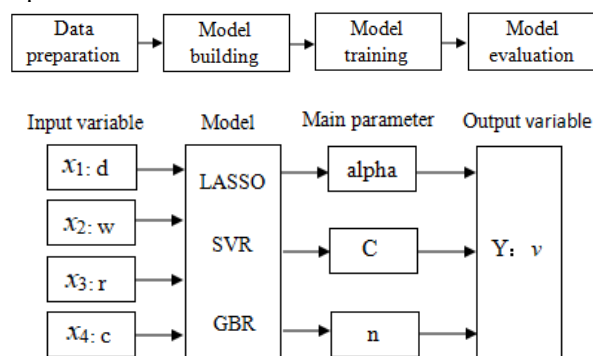
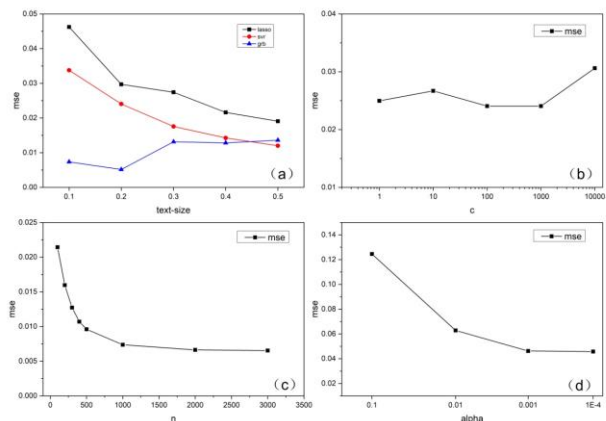


Fig. 6 Workflow chart for model establishment

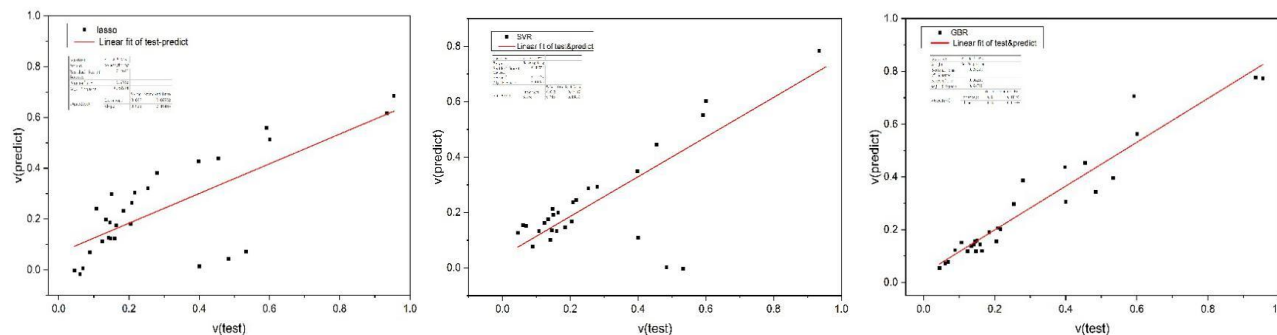
The ratio of the test set to the training set is determined by studying the changes in MSE (mean square error), as shown in Fig. 7. It can be seen from Fig. 7(a) that when the test set proportion is 20%, i.e. the training set proportion is 80%, both the MSE of the three models can be guaranteed to be in a low state, at the same time, it avoids the situation that the training set proportion is too small, resulting in large errors and insufficient accuracy. Debug the model, as shown in Fig. 7(b)(c)(d), when the MSE of the prediction model is stable and the lowest, three model parameters are determined:  $\alpha=0.001$ ,  $C=102$ ,  $n=1000$ , this The MSE values are 0.0297, 0.0240, and 0.0051, respectively.



**Fig. 7** Determination of key parameters of three prediction models

Table 3 lists the comparison results of predicted value and test value under the three models, and Fig. 8 shows the fitting of predicted value and actual value of the three models when 20% data is the test set. It can be seen from the figure that the data points are more

concentrated, and the GBR model has the best fit,  $R^2$  is 0.92; the SVR model and the LASSO model have a poor fit,  $R^2$  is 0.61 and 0.53 respectively. Table 4 lists the variance analysis results of the three models' prediction fits. It can be seen that the differences in the permeation performance results of the prediction models: LASSO>SVR>GBR. Comparing the proportions of qualified data of the three models under different error ranges, the results are shown in Table 5. When the acceptable errors are the same, the number of qualified data in the prediction model is: GBR>SVR>LASSO. According to the analysis of Fig. 8, Table 4 and Table 5, the accuracy of the LASSO model after running under the data set is inferior to the other two models in all aspects. The fitting results between the actual value and the predicted value of the SVR model and the GBR model are: The error range is similar at 0.20 cm/s, and the GBR model is slightly better than the SVR model at 0.01 and 0.03 cm/s, with 39.29% and 53.57%, respectively.



**Fig. 8** Three models predicted value-actual value fitting situation diagram

**Table 3** Comparison of test and predicted values under different models

No.	Test	Predict1	Predict2	Predict3	No.	Test	Predict1	Predict2	Predict3
1	0.454	0.438	0.444	0.453	15	0.254	0.321	0.286	0.298
2	0.151	0.298	0.192	0.159	16	0.069	0.006	0.151	0.078
3	0.398	0.426	0.349	0.437	17	0.280	0.380	0.293	0.386
4	0.935	0.615	0.784	0.777	18	0.124	0.111	0.163	0.118
5	0.208	0.263	0.237	0.206	19	0.148	0.124	0.213	0.118
6	0.534	0.071	(0.004)	0.396	20	0.185	0.233	0.146	0.191
7	0.061	(0.017)	0.153	0.071	21	0.045	(0.002)	0.126	0.054
8	0.601	0.513	0.602	0.562	22	0.089	0.068	0.077	0.122
9	0.484	0.042	0.003	0.343	23	0.134	0.198	0.176	0.138
10	0.592	0.557	0.551	0.706	24	0.164	0.175	0.198	0.119
11	0.217	0.303	0.244	0.201	25	0.205	0.181	0.168	0.155
12	0.107	0.239	0.132	0.151	26	0.142	0.127	0.100	0.145
13	0.146	0.185	0.136	0.156	27	0.401	0.013	0.110	0.304
14	0.954	0.686	0.934	0.773	28	0.160	0.122	0.133	0.144

**Table 4** Analysis of variance of three models

Variance analysis	DF	Sum of square	Mean square	F value	Prob>F
Lasso Model	1	0.56	0.56	31.14	7.33E-6

	Error	26	0.46	0.02	43.69	5.21E-7
	Total	27	1.02			
SVR	Model	1	0.83	0.83	325.51	3.33E-16
	Error	26	0.5	0.02		
	Total	27	1.33			
GBR	Model	1	1.12	1.12	325.51	3.33E-16
	Error	26	0.09	0		
	Total	27	1.21			

**Table 5** Proportion under three model error ranges

Model	LASSO Model		SVR Model		GBR Model	
	Number	Proportion	Number	Proportion	Number	Proportion
≤0.01	1	3.57%	5	17.86%	11	39.29%
≤0.03	10	35.71%	11	39.29%	15	53.57%
≤0.05	15	53.57%	19	67.86%	21	75.00%
≤0.10	21	75.00%	24	85.71%	22	78.57%
≤0.20	25	89.29%	25	89.29%	28	100.00%

Based on the comparison of MSE and  $R^2$ , the highest accuracy is the GBR model, followed by the SVR model, and the LASSO model is not suitable for the selection of sponge city pavement material water permeable model under this data set.

## 5 Conclusion

In this paper, starting from the characteristic performance parameters of sponge city pavement materials, water permeability prediction models based on three machine learning algorithms of LASSO, SVR and GBR is constructed. The experimental data is used to predict the trained prediction model and the regression fitting results are obtained. Based on the comparative analysis of the prediction models based on different machine learning algorithms, we find the prediction model that is most suitable for the water permeability of sponge city pavement materials.

(1) Based on the study of water permeability of sponge city pavement materials, experimental data is obtained through sample design and testing methods. In the study of the effect of material composition and structure on water permeability: The aggregate particle size increases and water permeability rate increases; the standard deviation decreases and the water permeability rate increases; the roundness decreases and the water permeability rate increases; the content of binder materials increases and the water permeability rate decreases.

(2) Establishing prediction models based on different machine learning algorithms through the acquisition of preliminary experimental data. Among them, the GBR model water permeation performance prediction fitting effect is better than the SVR and LASSO models, the test value-predicted value MSE is 0.0051 and  $R^2$  is 0.92. GBR model can effectively predict the water permeability of sponge city pavement materials.

There are many factors influencing the water permeability of sponge city pavement materials. This article is based on the test results and combined with machine learning methods to predict and analyze the

representative four characteristic parameters (particle size, particle size distribution, shape parameter, binder content). More characteristic parameters need to be supplemented in future research, and the proportion and number of data sets will be expanded to make the performance prediction effect more universal, so as to improve the prediction accuracy and wider practicality of the water permeability of sponge city pavement materials.

## References

1. Budge T. Sponge cities and small towns: a new economic partnership // Rogers M F, Jones R, eds. The Changing Nature of Australia's Country Town. Ballarat, Australia: Victorian Universities Regional Research Network Press, 2006.
2. YU Kongjian, LI Dihua, YUAN Hong, et al. "Sponge City": Theory and practice[J]. City Planning Review, 2015, 39(06):26-36.(in Chineses)
3. Jun X, Yongyong Z, Lihua X, et al. Opportunities and challenges of the Sponge City construction related to urban water issues in China[J]. Science China(Earth Sciences), 2017, 60(04):652-658.
4. ZHAO Jin-hui, ZHAO Ya-qian, Xie Xi, et al. Comparison of Field Permeability Test Methods for Permeable Pavement[J]. Chian Water &Wastewater, 2019, 35(05):114-118+124.(in Chinese)
5. Pan Xiyang. Research on Preparation and Performance of Porous Cement Concrete Pavement Material[D]. Chang'an University, 2010.(in Chinese)
6. Li X, Xu Q, Chen S. An experimental and numerical study on water permeability of concrete[J]. Construction & Building Materials, 2016, 105:503-510.
7. Al-Omari A, Masad E. Three dimensional simulation of fluid flow in X-ray CT images of porous media[J]. International Journal for Numerical & Analytical Methods in Geomechanics, 2004, 28(13):1327-1360.



8. X. Kuang, J. Sansalone, G. Ying, et al. Pore-structure models of hydraulic conductivity for permeable pavement[J]. *Journal of Hydrology*, 2011, 399(3):148-157.
9. ZHENG Wei-da, ZHANG Hui-ran, HU Hong-qing, et al. Performance prediction of perovskite materials based on different machine learning algorithms[J]. *The Chinese Journal of Nonferrous Metals*, 2019, 29(04):803-809.(in Chinses)
10. Tibshirani R . Regression Shrinkage and Selection Via the Lasso[J]. *Journal of the Royal Statistical Society*, 1996, 58(1):267-288.
11. B. E. Boser, I. M. Guyon, V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the 5th annual ACM Workshop on computation learning theory*. ACM Press. 1992, 144-152.
12. Freund Y, Schapire R, Abe N. A short introduction to boosting[J]. *Journal-Japanese Society for Artificial Intelligence*, 1999, 14(5): 771–780.
13. Fan Li, Yiming Yang, Eric P. Xing. From LASSO regression to feature vector machine[C]// *Advances in Neural Information Processing Systems 18* [Neural Information Processing Systems, NIPS 2005, December 5-8, 2005, Vancouver, British Columbia, Canada]. DBLP, 2005.
14. Yuan Z, Huang B. Prediction of protein accessible surface areas by support vector regression[J]. *Proteins Structure Function & Bioinformatics*, 2004, 57(3):558-564.
15. GONG Yue, LUO Xiao-Qin, WANG Dian-hai,et al. Urban travel time prediction based on gradient boosting regression tress[J]. *Journal of Zhejiang University(Engineering Science)*, 2018, 52(03):453-460.(in Chinses)