Structural Durability & Health Monitoring

Tech Science Press

https://doi.org/10.32604/sdhm.2026.075421

ARTICLE

Prediction of Asphalt Pavement Rutting Depth Based on Multi-Model Fusion of Stacking Algorithm

Chenhui Peng[1], Jinbiao Tang[1] and Derun Zhang[1,2,*]

[1]School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan, China
[2]National Center of Technology Innovation for Digital Construction, Huazhong University of Science and Technology, Wuhan, China

*Corresponding Author: Derun Zhang. Email: derunzhang@hust.edu.cn

Received: 31 October 2025; Accepted: 19 January 2026; Published: 18 May 2026

ABSTRACT: Rutting is a serious issue in asphalt pavement, which may reduce the pavement driving quality and safety. Accurately predicting rutting depth is a crucial task in pavement engineering, providing crucial decision support for asphalt pavement design and maintenance. However, accurate prediction of pavement rutting still remains a significant challenge for pavement engineers. This research first selects the loading number, temperature, dynamic modulus, asphalt layer thickness, and base layer type and thickness as candidate features. Data preprocessing, including outlier handling and feature selection, is then performed. Finally, based on the stacking algorithm, a multi-model fusion approach for predicting rutting depth in asphalt pavements is proposed, using ridge regression (RidgeR), K-nearest neighbor (KNN), multilayer perceptron (MLP), and random forest (RF) models as base models, and support vector machine (SVM) as a meta-model. The model is optimized using a Bayesian optimization model. Results demonstrate the feasibility of using correlation analysis for feature selection. Seven features, including axle weight, upper layer temperature, and middle layer modulus, were selected as predictive features. While all the basic models achieved good prediction accuracy, the stacking ensemble model exhibited lower variance and bias, demonstrating superior generalization capability. The asphalt pavement rutting depth prediction method based on the stacking algorithm multi-model fusion proposed in this research can accurately predict the rutting depth.

KEYWORDS: Rutting prediction; asphalt pavement; stacking algorithm; Bayesian optimization model

1 Introduction

Currently, asphalt pavement is widely used on most major traffic arteries in China. However, rutting is a major problem affecting asphalt pavement, severely impacting driving comfort. It also poses a significant challenge to traffic safety, especially in extreme weather conditions such as rain and snow. Rutting can cause vehicles to skid, leading to unpredictable traffic accidents [1–3]. In addition, rutting accelerates the occurrence and development of other pavement defects, significantly and adversely affecting the service life of asphalt pavements [4–7]. Therefore, accurately predicting rut depth is a crucial task in current pavement engineering. This work provides key decision-making information for the design and maintenance of asphalt pavements [8,9]. However, the causes of rutting deformation in asphalt pavements are extremely complex, which involve numerous factors, including material properties, environmental conditions, design and construction, and traffic loadings. Therefore, its prediction and mechanism research have always been important issues for road design engineers worldwide.

Prediction methods of asphalt pavement rutting can be classified into two categories: empirical testing method and machine learning prediction method. In the empirical testing method, the empirical mechanical models are often obtained through extensive experiments and elastic/viscoelastic layered system theory. Kong et al. [10] modified the MEPDG model using real data from RIOHTrack by introducing α, effectively reducing the impact of the accumulated number of load repetitions on rutting prediction. Hasan et al. [11] constructed a logistic regression model that effectively captured the correlation between the physical and chemical properties of the mixture and its rutting performance. Wang et al. [12] proposed the variable-order fractional Burgers rutting model based on the mechanical deformation characteristics of the mixture during rutting development. However, although the mechanical empirical simulation method has excellent prediction performance, its practicality is limited. With the continuous development of computer technology, machine learning methods are becoming more and more popular [13,14]. Zhang et al. [15] proposed a rutting prediction model based on multivariate transfer entropy and graph neural network. This model can analyze the significant causal relationship between feature variables and rutting, and shows excellent accuracy and robustness in the prediction task. Usanga et al. [16] trained a Gaussian process regression model and conducted sensitivity analysis on different rutting performance evaluation indicators. The study clarified the variables that have a key impact on rutting performance prediction. Gardezi et al. [17] used the gene expression programming method to predict the rutting depth of asphalt pavement. The experimental results show that the model has excellent prediction ability. Alnaqbi et al. [18] collected rutting data from the Long-Term Pavement Performance (LTPP) database. Multiple machine learning models showed excellent prediction performance on this dataset.

Despite significant progress in rutting prediction research in recent years, its performance and practical application are still severely constrained by two major problems. On the one hand, there is a lack of efficient and universal data processing, feature selection, and modeling frameworks [4]; on the other hand, effective and reliable asphalt pavement data are also seriously insufficient [19]. Therefore, there is still much work to be done in rutting prediction research.

To predict asphalt pavement rutting depth, this paper proposes a multi-model fusion method based on the stacking algorithm. This method combines six models: ridge regression, K-nearest neighbor, support vector machine, multilayer perceptron, and random forest. Factors such as load time, temperature, dynamic modulus, asphalt layer thickness, and base layer type and thickness are incorporated into the model reference features. By stacking and fusing the prediction results of each model, this method effectively integrates the advantages of each model. The overall prediction accuracy is significantly improved on the dataset.

2 Data

2.1 Data Source

The data for this study is sourced from RIOHTrack [20]. RIOHTrack is the first full-scale pavement test loop in my country's road engineering field. Located in Beijing, it was completed in October 2015 and officially put into operation in December 2016. The road has 19 asphalt pavement structure types, which can be divided into 7 structural types, as shown in Table 1.

Table 1: Table of test section pavement structure classification.

Type	Structural Layer Number
12 cm asphalt layer Semi-rigid base pavement	STR1~STR3
Rigid base asphalt pavement	STR4, STR5
18 cm asphalt layer Semi-rigid base pavement	STR6~STR9
Inverted asphalt pavement,	STR10, STR12
24–28 cm asphalt layer Semi-rigid base pavement	STR11, STR13, STR14
36 cm asphalt layer Semi-rigid base pavement	STR15~STR17
Full-thickness asphalt pavement	STR18, STR19

Note: The asphalt layer thickness, base type and thickness of various structures are different. In subsequent modeling, it is necessary to select certain features for different structures to distinguish them.

2.2 Data Acquisition and Analysis

To accurately predict pavement rutting depth, it's necessary to screen the factors that influence rutting depth and select appropriate data features to represent these factors. The axle load times is clearly a key factor affecting rutting depth, so it's included as one of the features. In addition, this paper also selected factors such as temperature, dynamic modulus, asphalt thickness, base layer type, and base layer thickness as candidate features. Obviously, rutting depth is the target of prediction.

(1) Rutting Depth

When extracting rutting depth, in order to distinguish the characteristics of different pavement structures, this paper eliminated the rutting depth near the junction of the two pavement structures. A box plot was used to detect and process abnormal data, accurately depicting the discrete distribution characteristics of the data. Adequate data cleaning is conducive to improving prediction accuracy. Fig. 1 shows the variation of rut depth with axle load times for 19 different road surface structures.

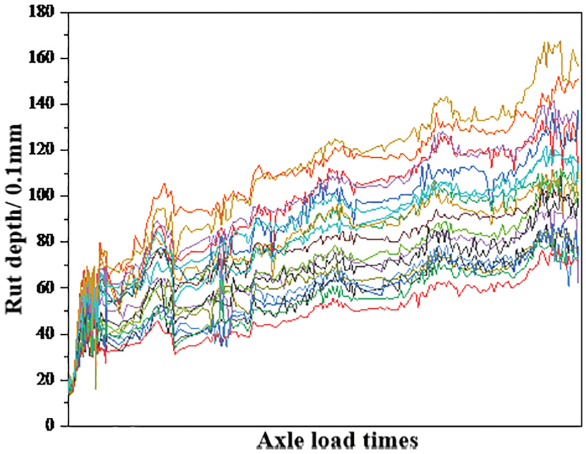


Figure 1: Graph of rut depth as a function of axle load times.

(2) Temperature

Temperature is an important factor affecting the rutting depth of asphalt pavement. When the temperature is high, asphalt pavement defects are more likely to form and intensify. This paper selects four temperature data, namely the temperature of the top three asphalt layers and the ambient temperature. If there are fewer than three asphalt layers, the temperature at the bottom of the second layer is used to fill the missing value of the third layer. Based on previous research experience, the temperature values are all the average temperature values of the 15 days before the rutting measurement time. If there are missing values in these 15 days, the average temperature values of the 5 days before and after the same day in the previous year are used to fill the missing values. On the one hand, temperature is a feature that needs to be used when predicting rutting depth; on the other hand, temperature is an indispensable parameter when calculating dynamic modulus.

(3) Dynamic modulus

In the prediction and analysis of asphalt pavement rutting, dynamic modulus is often indispensable. In addition to sorting out the existing dynamic modulus data of asphalt mixtures, this paper also searches for missing data and supplements them. The dynamic moduli of three asphalt mixtures, "Recycled-AC25", "SBS-SAC12" and "SBS-PAC13" are mainly supplemented [21–23]. The data are shown in Table 2. The Sigmoid model is further used to fit the dynamic modulus master curve, as shown in Fig. 2a, and the dynamic modulus of the layer at the corresponding time is calculated using the real-time temperature, as shown in Fig. 2b.

Table 2: Dynamic modulus of supplementary asphalt mixture.

Structure	Alpha	Delta	Labda	Beta	Gama	C1	C2
Recycled-AC25	1.808860	4.474533	0.409568	1.802823	0.671768	15.862650	140.563130
SBS-SAC12	1.818105	4.532275	0.482371	1.360056	0.562702	7.637120	60.619647
SBS-PAC13	0.000136	4.498614	0.846938	1.787140	0.289065	45.767951	308.305957

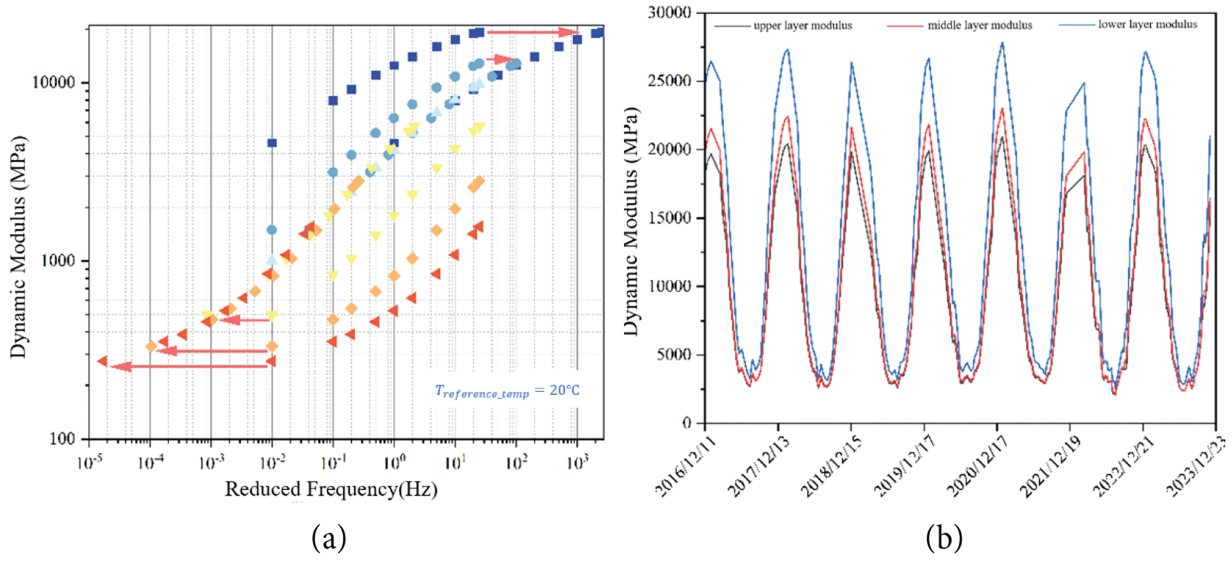


Figure 2: (**a**) Dynamic modulus master curve (taking SBS-AC13 as an example). (**b**) Dynamic modulus variation diagram of typical 4, 6, and 8 cm asphalt layer structures (taking STR7 as an example).

(4) Asphalt layer thickness

Semi-rigid base pavement structures are divided into four types based on the asphalt layer thickness. It can be seen that the asphalt layer thickness has a significant impact on pavement performance, and its influence on rutting depth is self-evident, typical road surface structures of RIOHTrack are shown in Fig. 3 below. To facilitate the unification of feature quantities, this paper selects the thickness of the top three asphalt layers as a surrogate feature. For structures with less than three asphalt layers, the thickness of the last asphalt layer is set to 0. However, the top asphalt layer thickness of 19 structures is the same and can be directly discarded.

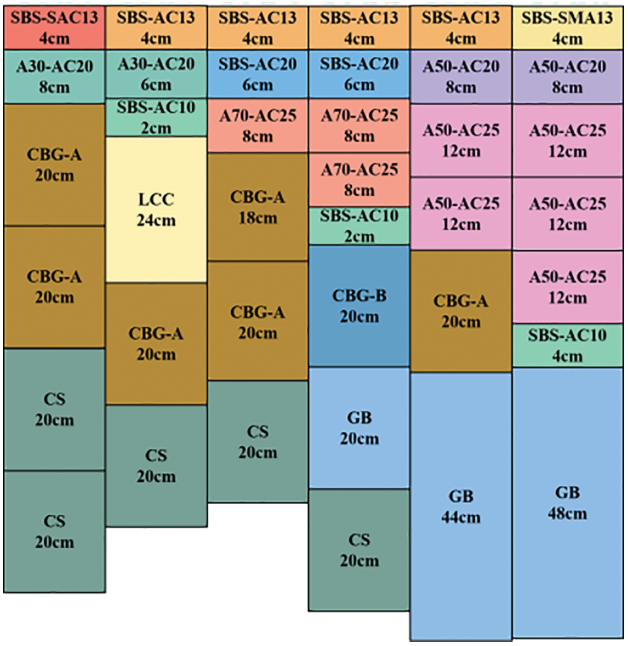


Figure 3: Typical road surface structure diagram of RIOHTrack.

(5) Base type and thickness

Base type and thickness are also important characteristics for distinguishing pavement structures. To facilitate model identification, this article uses a numbering system to distinguish base types: "CBG-A" is numbered "1", "LCC" is numbered "2", "CC" is numbered "3", "CBG-B" is numbered "4", and "GB" is numbered "5". The base thickness is represented by the total base thickness.

The distribution table of all candidate feature data is shown in Table 3.

Table 3: Distribution of candidate feature data.

Feature Name		Min	Max	Average	SD
Road structure	First layer modulus, MPa	1784.623	31,079.478	8195.629	5726.750
	Second layer modulus, MPa	2047.191	34,168.198	13,061.079	8399.001
	Third layer modulus, MPa	2252.764	32,966.033	14,024.697	8347.043

(Continued)

Table 3 (continued)

	Feature Name		Min	Max	Average	SD
Input		Thickness of the second asphalt layer, cm	6.000	10.000	7.578	1.042
		Thickness of the third asphalt layer, cm	0.000	40.000	12.947	11.955
		Total thickness of base layer, cm	20.000	80.000	56.526	12.086
		First layer base type	—	—	—	—
	Temperature	Bottom temperature of the first asphalt layer, °C	−5.760	39.818	21.502	12.051
		Bottom temperature of the second asphalt layer, °C	−5.381	39.799	21.385	11.780
		Bottom temperature of the third asphalt layer, °C	−5.026	38.478	21.322	11.461
		Ambient temperature, °C	−5.027	30.239	17.594	9.744
	Load	Axle load times, 10,000 times	2.840	9930.931	4632.050	3174.393
Output	Service performance	Rutting depth, 0.1 mm	7.510	167.720	74.181	28.957

2.3 Data Processing

(1) Outlier Processing

Experiments often have certain errors. For machine learning tasks, the raw data obtained will inevitably contain some data that is far away from the majority of samples. These data are usually called outliers. The essence of a machine learning model is to establish a mapping relationship from input to output. Abnormal data usually deviates far from the majority of samples. When the model learns this deviation, it may cause the model to deviate and reduce the accuracy of the model; if the model over-adapts to this deviation, it may also cause the model to overfit and reduce its generalization ability [24]. Therefore, before inputting data into the machine learning model, it is necessary to remove abnormal data in the dataset.

This paper uses a boxplot to handle data outliers. This method arranges the features from largest to smallest and calculates the lower quartile Q1, median Q2, and upper quartile Q3 of each feature based on the data's location distribution and dispersion. It is assumed that the reasonable range of values for each set of data should meet the requirements of Eq. (1). Data outside this range are considered outliers and are removed. This outlier handling does not include structural layer thickness, as it is a fixed value in each sample.

$$Q_1 - k\,(Q_3 - Q_1) \le x \le Q_1 + k\,(Q_3 - Q_1) \tag{1}$$

where k is the anomaly threshold, which determines the strength of anomaly detection. k is usually set to 1.5 or 3. k = 1.5 means strong recognition, eliminating even minor anomalies; k = 3 means weak recognition, retaining most of the data.

The original dataset contained 3933 samples, and the sample size remained at 3933 after outlier removal. Thanks to rigorous raw data collection methods and meticulous preprocessing techniques, no obvious

outliers were detected in the rut data used in this study. The data distribution is reasonable and meets the requirements for subsequent machine learning model construction.

(2) Feature selection

Feature selection is very important for machine learning. Obviously, not all features will have a significant impact on rutting depth. Introducing too many features will increase the dimension of the data, which will not only significantly increase the training cost of the model, but may also cause the model to overfit [25]. Therefore, the feature selection method will be used to select features that have a significant impact on rutting depth. By detecting and analyzing the features, removing redundant features, and obtaining a smaller feature subset, on the one hand, the model training cost can be reduced, and on the other hand, the model prediction accuracy can be effectively improved.

This paper uses correlation analysis for feature selection and uses the Pearson correlation coefficient to calculate the correlation between candidate features. The calculation formula is shown in Eq. (2). A correlation heat map is plotted, as shown in Fig. 4. For features with high correlation, one is selected as the predictive feature. After calculation and analysis, a total of seven features were ultimately determined as predictive features: axle load times, second layer modulus, thickness of the second asphalt layer, thickness of the third asphalt layer, total thickness of base layer, first layer base type, and bottom temperature of the first asphalt layer.

$$\gamma = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2} \cdot \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}} \tag{2}$$

where n is the sample size, x_i and y_i are the i-th observation values of the sample, and $\overline{x}$ and $\overline{y}$ are the sample means.

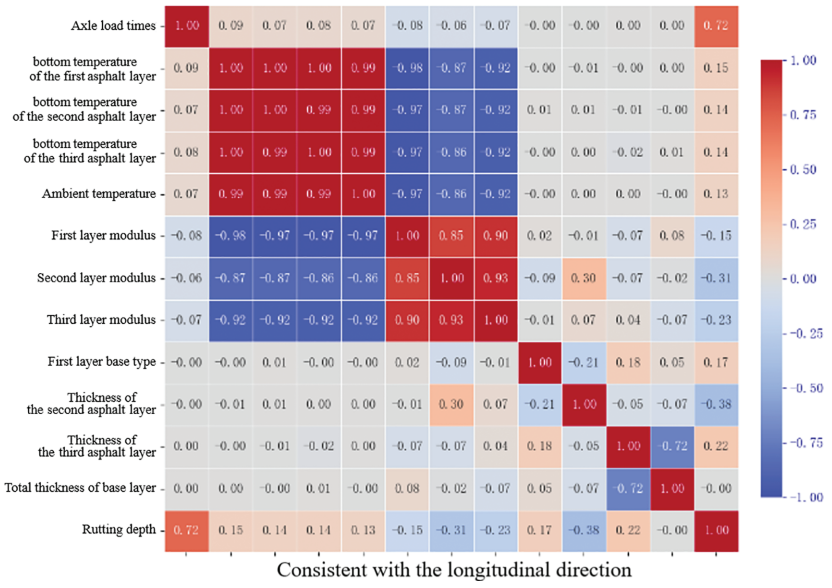


Figure 4: Heatmap of correlation between candidate features.

3 Methods

3.1 Base Model

(1) Ridge Regression Model (RidgeR)

RidgeR is a linear regression technique that incorporates L2 regularization into the least squares method. It mitigates overfitting by penalizing the magnitude of regression coefficients, which effectively reduces the influence of less relevant features while maintaining a linear input–output assumption. Its coefficient estimation is given in Eq. (3).

$$\hat{\beta}^{ridge} = \left(X^T + \lambda I\right)^{-1} X^T y \tag{3}$$

where $\hat{\beta}^{ridge}$ is the coefficient estimate of the RidgeR model, X is the matrix of independent variables, y is the matrix of dependent variables, λ is the ridge parameter used to control the degree of regularization, and I is the identity matrix.

(2) k-nearest neighbor algorithm (KNN)

KNN is a supervised, instance-based learning algorithm that does not require an explicit training phase. For regression tasks, it identifies the k closest samples in the feature space using a distance metric and outputs the average of their target values. The choice of k is crucial to balance sensitivity to noise and preservation of local patterns [26].

(3) Support Vector Machine Model (SVM)

SVM is a supervised model based on structural risk minimization, renowned for its powerful generalization ability. It employs an ε-insensitive loss function and kernel methods to map data into a higher-dimensional space [27]. Through dual optimization, it identifies support vectors that define the regression function, effectively capturing complex, nonlinear relationships.

(4) Multilayer Perceptron Model (MLP)

A multilayer perceptron model consists of an input layer, one or more hidden layers, and an output layer, which has been widely used in civil engineering [28]. Its topology is shown in Fig. 5. By applying nonlinear activation functions in the hidden layers, it can approximate complex nonlinear mappings between inputs and outputs. The model is essentially a stack of fully connected layers trained via backpropagation.

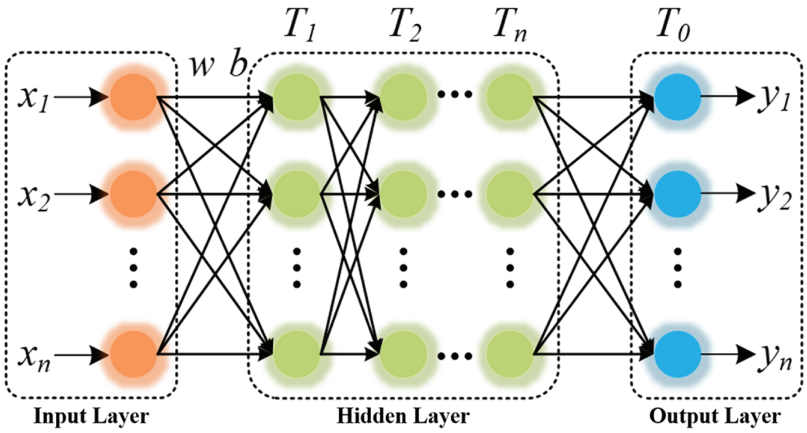


Figure 5: Multilayer perceptron model topology.

(5) Random Forest Model (RF)

Random forest is an ensemble learning method based on a bagging strategy, consisting of multiple decision trees. The general structure of the random forest is shown in Fig. 6. Each tree is trained on a bootstrap sample of the data, and a random subset of features is considered for splitting at each node. The final prediction is the average of all tree outputs, which enhances generalization and stability by reducing model variance. Hyperparameter tuning is essential to control overfitting.

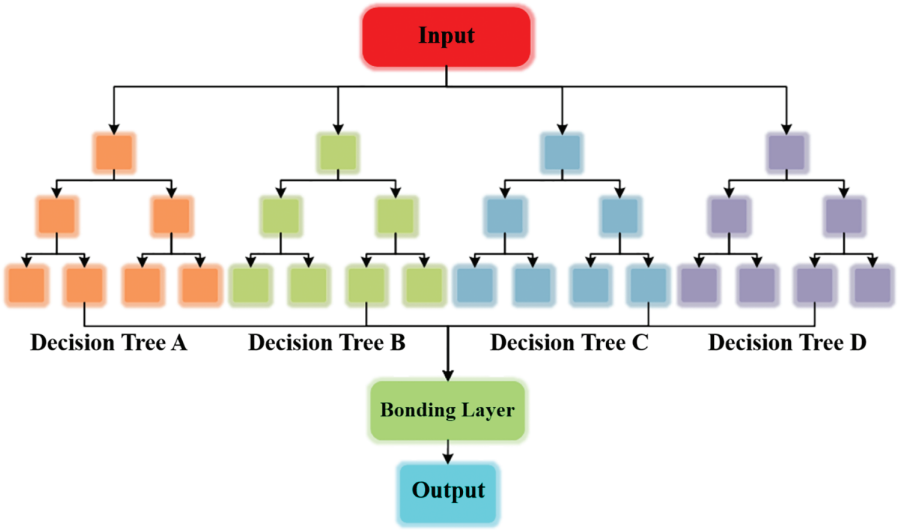


Figure 6: Schematic diagram of the random forest model process.

3.2 Stacking Algorithm

The core differences of the current mainstream ensemble learning methods are concentrated in three categories: bagging, boosting, and stacking [29]. Bagging is based on "parallel independence" and draws conclusions by training multiple base learners separately and averaging/weighting the results; boosting relies on "sequential correction" to improve accuracy. Subsequent models will fit the deviation of the previous model. Theoretically, the accuracy can reach 100% when the number of models is large enough; stacking is a fusion of the two, and its core is the "multi-layer structure" (as shown in Fig. 7). The first layer uses different types of base learners to process the original data in parallel, and the second layer uses a low-complexity meta-model to integrate the output of the base learners, which can avoid overfitting and explore deep mapping relationships [30]. However, it should be noted that too many base model layers will significantly increase training costs and the risk of overfitting.

The basic algorithm models described above serve as the foundation for the stacking algorithm, with the SVM model serving as the meta-model. This study explored linear regression, decision tree, and SVM models. Linear regression failed to improve model performance, while decision tree models led to overfitting in the stacked model. Ultimately, SVM was chosen as the meta-model because the stacked model effectively reduced bias and dispersion. These basic models differ significantly from one another, enabling them to extract input-output mappings from different perspectives. Furthermore, model stacking and integration can effectively reduce the bias and variance of individual models, ensuring model accuracy and improving generalization.

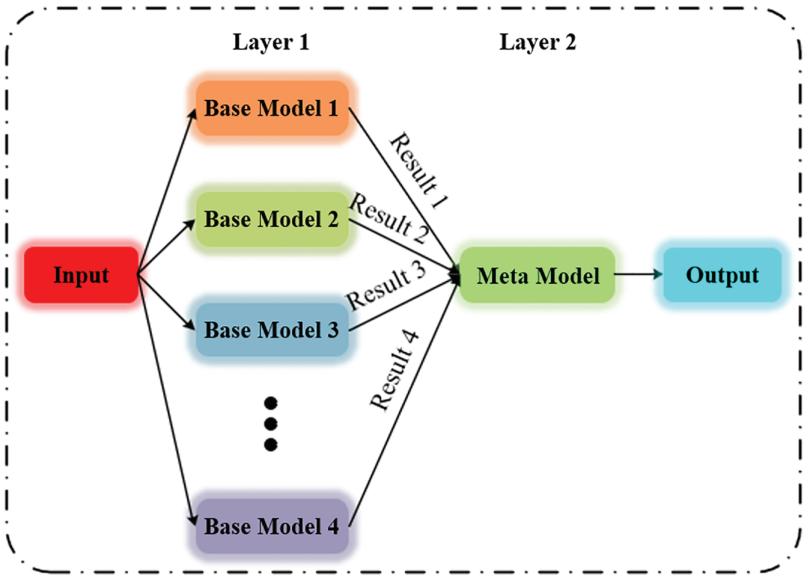


Figure 7: Schematic diagram of the stacking model based on the stacking algorithm.

3.3 Bayesian Optimization

As the name suggests, the Bayesian optimization model is based on Bayes' theorem. The algorithm calculates posterior probabilities based on prior knowledge and observed data. Leveraging this knowledge and observed data, it can more effectively search the parameter space, improving search efficiency. Fig. 8 illustrates how the Bayesian optimization model optimizes machine learning model parameters.

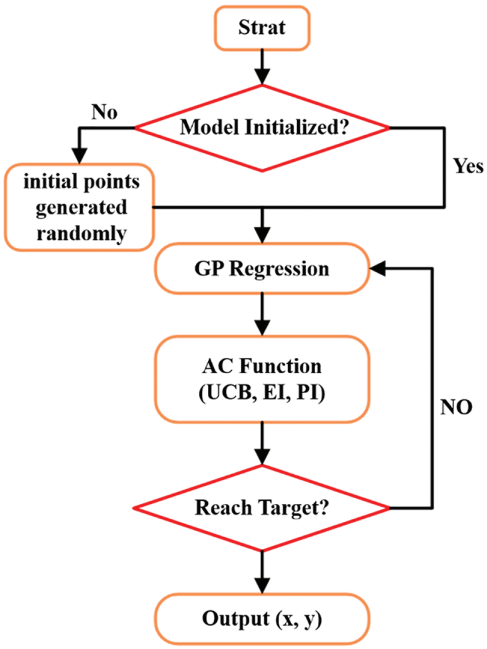


Figure 8: Flowchart of parameter optimization of Bayesian optimization model.

The predictive power of a machine learning model directly depends on appropriate hyperparameters [31], which need to be manually set before training the machine learning model. To obtain a forward

design model with better performance, this paper will use the Bayesian optimization algorithm. This algorithm typically uses Gaussian process regression to establish the objective function and updates the best-performing model hyperparameters based on probability. Compared with traditional random search and grid search methods, this algorithm can find appropriate hyperparameters more efficiently.

Before performing hyperparameter optimization, the original data is divided into a training set and a test set based on a stratified sampling strategy, with the training set accounting for 80% and the test set accounting for 20%. At the same time, all data are normalized based on the data distribution of the training set. The normalization formula is shown in Eq. (4).

$$z_i = \frac{x_i - \mu_i}{\sigma_i} \tag{4}$$

where: x_i is the original value of a certain feature; μ_i is the mean of the corresponding variable; σ_i is the standard deviation of the corresponding variable; z_i is the standardized value of the corresponding variable.

4 Results and Discussion

4.1 Hardware and Hyperparameter Tuning

In terms of hardware, an Intel(R) Core (TM) i7-9750H central processing unit (CPU) and an NVIDIA GeForce GTX 1650 graphics processing unit (GPU) were used. In terms of software, the operating system was Windows 10, Python 3.10 was adopted as the programming language, PyTorch 2.8.0 served as the deep learning library, and CUDA 12.6 was utilized to accelerate and optimize the computation, training and inference processes of the neural network.

The hyperparameter optimization results are shown in Table 4 below. The hyperparameters of the meta model are also presented in the table.

Table 4: Hyperparameter spaces and their corresponding values for each model.

Machine Learning Models	Hyperparameters	Space of Hyperparameters	Values
RidgeR	Alpha	0.000001–10	6.773
KNN	N_neighbors	3–20	6
SVM	Kernel	'linear', 'poly', 'rbf', 'sigmoid'	'rbf'
	C	0.1–1000	857.977
	Epsilon	0.01–1	0.39698
MLP	Hidden layer sizes	10, 15, 20, 25, 30, (10, 10), (15, 15), (20, 20), (25, 25), (30, 30)	(20, 20)
	Alpha	0.0001–0.1	0.01760
	Solver	'adam', 'sgd', 'lbfgs'	'lbfgs'
RF	N_estimators	10–1000	860
	Max_depth	None, 5–21	9
Stacking meta model	Kernel	'linear', 'poly', 'rbf', 'sigmoid'	'linear'
	C	0.1–1000	240.816
	Epsilon	0.01–1	0.80171

4.2 Evaluation Metrics

To evaluate the performance of the prediction model, this paper uses Mean Absolute Error (MAE), Mean Square Error (MSE), and R^2 as model evaluation indicators. MAE reflects the average deviation between the predicted value and the true value. The smaller its value, the higher the prediction accuracy of the model. MSE amplifies the impact of outliers by squaring the error. The smaller its value, the better the model training effect. R^2 is used to evaluate the goodness of fit of the regression model. The closer its value is to 1, the better the model fit. The calculation formulas for each evaluation indicator are shown in Eqs. (5)–(7).

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i| \tag{5}$$

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \overline{y}_i)^2} \tag{7}$$

where m is the number of samples; y_i is the true value of the i-th sample; $\hat{y}_i$ is the predicted value of the i-th sample.

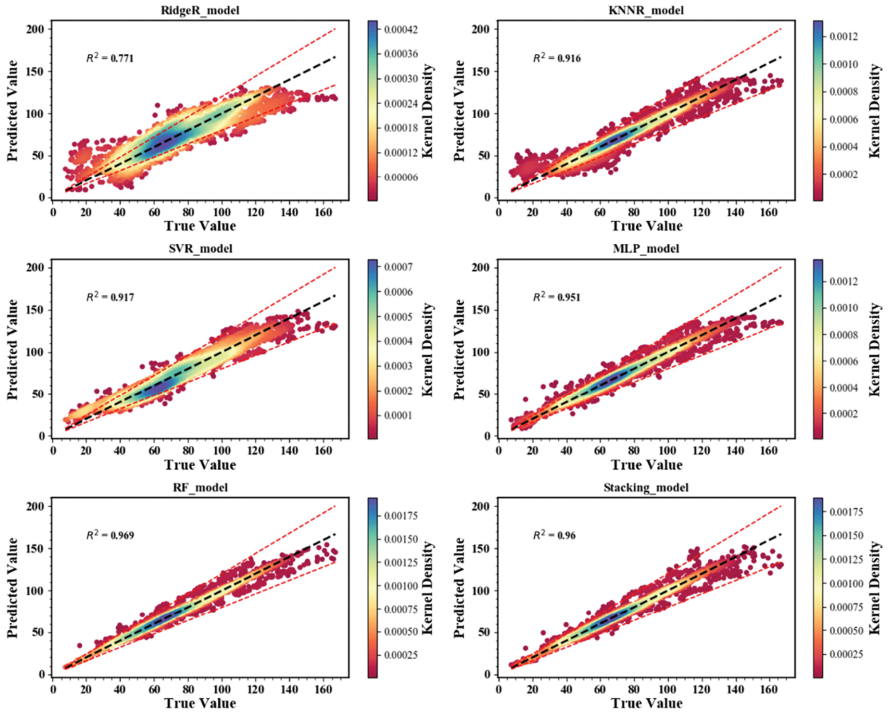
4.3 Performance Comparison

The prediction performance evaluation indicators of the five basic models and the stacking fusion model are shown in Table 5. As can be seen from the table, the overall performance of the RidgeR model is significantly inferior to that of the other five models. The performance of the KNN, SVM, MLP, and RF models on the training set and the test set all show obvious differences, and the generalization performance is insufficient. At the same time, it can be found that the RF model has the highest prediction performance on the training set, with an R^2 of 96.9%, which is due to the fusion of multiple decision trees in the random forest. Although the performance of the stacking model on the training set is slightly worse than that of the RF model, it performs best on the test set, and the model performance does not show a significant decline. This shows that compared with the basic model, the model stacking method can effectively reduce the prediction bias and discreteness of the model, and improve the generalization ability and prediction accuracy of the model.

Table 5: Model evaluation indicators.

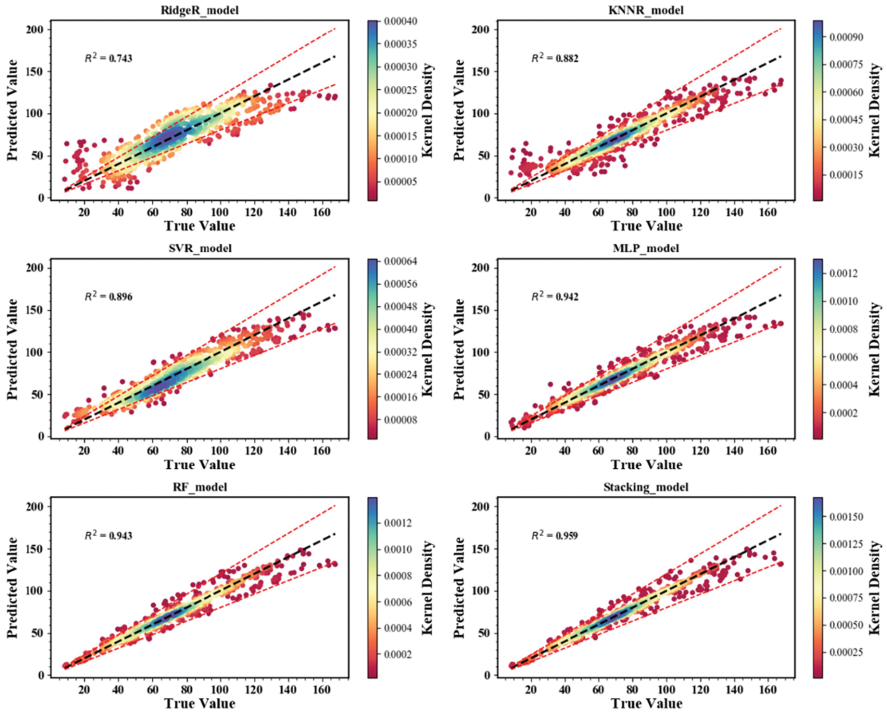
Model	Evaluation Indicators					
	Training Set			Test Set		
	MAE	MSE	R^2	MAE	MSE	R^2
RidgeR	0.372	0.229	0.771	0.397	0.258	0.743
KNN	0.192	0.084	0.916	0.227	0.118	0.882
SVM	0.234	0.083	0.917	0.255	0.104	0.896
MLP	0.149	0.049	0.951	0.163	0.059	0.942
RF	0.110	0.031	0.969	0.152	0.057	0.943
Stacking	0.120	0.040	0.960	0.125	0.041	0.959

Furthermore, Fig. 9 shows the kernel density estimation scatter plots of the training set and the test set. In the figure, the black dotted line represents the baseline, the red dotted line represents the 20% error line, and the color change of the scatter plot represents the distribution density of the points. As can be seen from the figure, the scatter plot is generally distributed near the baseline. In the RidgeR and KNN models, when the rutting depth is too small or too large, especially in the early stage of rutting, there will be a significant deviation between the predicted value and the true value. For the SVM, MLP, RF and stacking models, the scatter plots are generally distributed within the 20% error line, and there is a high-density distribution area near the baseline. These models all show excellent prediction performance. The stacking fusion model also shows a higher scatter kernel density in the test set, thus achieving the best performance, which is very suitable for predicting the rutting depth of asphalt pavement.



(a) training set

Figure 9: (Continued)



(b) test set

Figure 9: Kernel density estimation scatter plot.

5 Conclusions

(1) Based on the extracted layered temperature results, it can be found that although there are certain differences in the temperature of different layers within a day, the temperature effects of different layers are basically the same within a loading cycle.

(2) Correlation analysis results show that the temperature changes at different layers are highly correlated with the modulus changes, and the change trends are basically consistent. Among them, the Axle load times have a high correlation with the rutting depth, which plays a key role in rutting depth prediction.

(3) The results of multiple machine learning models show that the use of the stacking fusion model can effectively improve the accuracy and generalization ability of rutting depth prediction, meeting the actual engineering needs.

The model proposed in this study still has certain limitations. In terms of data processing, the handling of temperature and modulus is still in its preliminary stages. The proposed model fails to fully capture the spatial three-dimensional structure of the road surface. The model's adaptability and accuracy under complex conditions cannot be guaranteed. Regarding model training, this paper's model fails to incorporate uncertainty assessment. These two limitations collectively restrict the model's application in practical engineering scenarios.

Acknowledgement: The authors also acknowledge the Research Institute of Highway Ministry of Transport for providing the long-term monitoring data of RIOHTrack.

Funding Statement: The authors acknowledge the financial support of the Natural Science Foundation of Hubei Province of China (grant number: 2025AFD739).

Author Contributions: Chenhui Peng: writing—original draft, methodology, funding acquisition, project administration, validation, writing—review & editing. Jinbiao Tang: writing—original draft, methodology, investigation, data curation. Derun Zhang: methodology, data curation, writing—review & editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data and materials in the current study are available from the corresponding author on a reasonable request.

Ethics Approval: Not applicable.

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

References

1. Kou B, Cao J, Huang W, Ma T, Shi Z. Rutting prediction model of asphalt pavement based on RIOHTrack full-scale ring road. Measurement. 2025;242:115915. doi:10.1016/j.measurement.2024.115915.
2. Radhakrishnan V, Chowdari GS, Reddy KS, Chattaraj R. Evaluation of wheel tracking and field rutting susceptibility of dense bituminous mixes. Road Mater Pavement Des. 2019;20(1):90–109. doi:10.1080/14680629.2017.1374998.
3. Luo W, Qin Y, Zhang D, Li L. Measurement of pavement rutting trajectories on two-lane highway using the 3D line scanning laser system. Int J Pavement Eng. 2023;24(2):2149753. doi:10.1080/10298436.2022.2149753.
4. Singh AK, Sahoo JP. Rutting prediction models for flexible pavement structures: a review of historical and recent developments. J Traffic Transp Eng Engl Ed. 2021;8(3):315–38. doi:10.1016/j.jtte.2021.04.003.
5. Zhao X, Shen A, Ma B. Temperature adaptability of asphalt pavement to high temperatures and significant temperature differences. Adv Mater Sci Eng. 2018;2018(1):9436321. doi:10.1155/2018/9436321.
6. Dai X, Jia Y, Wang S, Gao Y. Evaluation of the rutting performance of the field specimen using the hamburg wheel-tracking test and dynamic modulus test. Adv Civ Eng. 2020;2020(1):9525179. doi:10.1155/2020/9525179.
7. Chopra T, Parida M, Kwatra N, Chopra P. Development of pavement distress deterioration prediction models for urban road network using genetic programming. Adv Civ Eng. 2018;2018:1253108. doi:10.1155/2018/1253108.
8. Liu G, Chen L, Qian Z, Zhang Y, Ren H. Rutting prediction models for asphalt pavements with different base types based on RIOHTrack full-scale track. Constr Build Mater. 2021;305:124793. doi:10.1016/j.conbuildmat.2021.124793.
9. Chen S, Cao J, Wan Y, Shi X, Huang W. Enhancing rutting depth prediction in asphalt pavements: a synergistic approach of extreme gradient boosting and snake optimization. Constr Build Mater. 2024;421:135726. doi:10.1016/j.conbuildmat.2024.135726.
10. Kong W, Huang W, Guo W, Wei Y. Modification of MEPDG rutting model based on RIOHTrack data. Int J Pavement Eng. 2023;24:2201500. doi:10.1080/10298436.2023.2201500.
11. Hasan MM, Hasan MA, Tarefder RA. New model for predicting rutting performance of superpave asphalt mixes. Transp Res Rec J Transp Res Board. 2022;2676(8):174–85. doi:10.1177/03611981221083298.
12. Wang Y, Yan J, Huang W, Rutkowski L, Cao J. Variable-order fractional derivative rutting depth prediction of asphalt pavement based on the RIOHTrack full-scale track. Sci China Inf Sci. 2023;66(5):152205. doi:10.1007/s11432-022-3647-7.
13. Cheng C, Wang L, Zhou X, Wang X. Predicting rutting development using machine learning methods based on RIOCHTrack data. Appl Sci. 2024;14(8):3177. doi:10.3390/app14083177.
14. Majidifard H, Jahangiri B, Rath P, Alavi AH, Buttlar WG. A deep learning approach to predict Hamburg rutting curve. Road Mater Pavement Des. 2021;22(9):2159–80. doi:10.1080/14680629.2021.1886160.
15. Zhang J, Cao J, Huang W, Shi X, Zhou X. Rutting prediction and analysis of influence factors based on multivariate transfer entropy and graph neural networks. Neural Netw. 2023;157:26–38. doi:10.1016/j.neunet.2022.08.030.
16. Usanga IN, Ikeagwuani CC, Etim RK, Attah IC. Predictive modeling of modified asphalt mixture rutting potentials: machine learning approach. Iran J Sci Technol Trans Civ Eng. 2023;47(6):4087–101. doi:10.1007/s40996-023-01192-w.

17. Gardezi H, Ikrama M, Usama M, Iqbal M, Jalal FE, Hussain A, et al. Predictive modeling of rutting depth in modified asphalt mixes using gene-expression programming (GEP): a sustainable use of RAP fly ash, and plastic waste. Constr Build Mater. 2024;443:137809. doi:10.1016/j.conbuildmat.2024.137809.
18. Alnaqbi AJ, Zeiada W, Al-Khateeb GG, Hamad K, Barakat S. Creating rutting prediction models through machine learning techniques utilizing the long-term pavement performance database. Sustainability. 2023;15(18):13653. doi:10.3390/su151813653.
19. Zhang J, Cao J, Huang W, Shi X, Ji X, Zhou X. A hybrid framework for asphalt pavement rutting prediction modeling and influencing factors analysis based on multilevel wavelet decomposition and transfer entropy. Appl Math Model. 2023;121:714–30. doi:10.1016/j.apm.2023.05.024.
20. Li S, Fan M, Xu L, Tian W, Yu H, Xu K. Rutting performance of semi-rigid base pavement in RIOHTrack and laboratory evaluation. Front Mater. 2021;7:590604. doi:10.3389/fmats.2020.590604.
21. Jiang K, Dong M, Feng Z, Zhang H, Zhang H, Liu J. Viscoelastic characterization of multi-fragmented asphalt concrete based on hirsch model. Highway. 2025;70(5):43–9. (In Chinese).
22. Peng C. The study of materials and properties of porous asphalt concrete at plateau region [master's thesis]. Xi'an, China: Chang'an University; 2024. (In Chinese). doi:10.26976/d.cnki.gchau.2023.002193.
23. Yin L. Study on road performance and mechanical properties of the warm recycled asphalt mixture with high volume RAP [master's thesis]. Jinan, China: Shandong Jianzhu University; 2024. (In Chinese). doi:10.27273/d.cnki.gsajc.2023.000989.
24. Liu J, Liu F, Zheng C, Zhou D, Wang L. Optimizing asphalt mix design through predicting effective asphalt content and absorbed asphalt content using machine learning. Constr Build Mater. 2022;325:126607. doi:10.1016/j.conbuildmat.2022.126607.
25. Jukte NR, Swamy AK. Deep learning based methodological approach for prediction of dynamic modulus and phase angle of asphalt concrete. Eng Appl Artif Intell. 2025;145:110269. doi:10.1016/j.engappai.2025.110269.
26. Zhang D, Peng C, Ma J, Wang W, Borg RP, Lewis O. Reverse design for mixture proportion of asphalt concrete in high-temperature regions based on the materials informatics. Intell Transp Infrastruct. 2025;4:liaf020. doi:10.1093/iti/liaf020.
27. Zhang W, Lee D, Lee J, Lee C. Residual strength of concrete subjected to fatigue based on machine learning technique. Struct Concr. 2022;23(4):2274–87. doi:10.1002/suco.202100082.
28. Xu P, Zeng Z, Miao Y, Zhang D, Fu C. Field aging characterization of asphalt pavement based on the artificial neural networks and gray relational analysis. J Mater Civ Eng. 2023;35(7):04023188. doi:10.1061/jmcee7.mteng-15004.
29. Sandeep MS, Tiprak K, Kaewunruen S, Pheinsusom P, Pansuk W. Shear strength prediction of reinforced concrete beams using machine learning. Structures. 2023;47:1196–211. doi:10.1016/j.istruc.2022.11.140.
30. Guan Y, Zhang B, Li Z, Zhang D. Enhanced flow number prediction of asphalt mixtures using stacking ensemble-based machine learning model and grey relational analysis. Constr Build Mater. 2025;463:140001. doi:10.1016/j.conbuildmat.2025.140001.
31. Le TH, Nguyen HL, Pham BT, Nguyen MH, Pham CT, Nguyen NL, et al. Artificial intelligence-based model for the prediction of dynamic modulus of stone mastic asphalt. Appl Sci. 2020;10(15):5242. doi:10.3390/app10155242.