

Article

Not peer-reviewed version

Interpretable Machine Learning for Predicting Splitting Strength of Asphalt Concrete: Insights from SHAP Analysis

[Jianglei Xing](#), [Xiao Tan](#)^{*}, [Yihao Li](#), [Dongzhao Jin](#)^{*}, [Pengwei Guo](#), Yuhuan Wang, [Huiya Niu](#)

Posted Date: 30 March 2026

doi: 10.20944/preprints202603.2259.v1

Keywords: asphalt concrete; splitting strength; TabPFN; explainable artificial intelligence; SHAP interpretation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Interpretable Machine Learning for Predicting Splitting Strength of Asphalt Concrete: Insights from SHAP Analysis

Jianglei Xing ¹, Xiao Tan ^{1,2,*}, Yihao Li ¹, Dongzhao Jin ^{3,*}, Pengwei Guo ⁴, Yuhuan Wang ⁵ and Huiya Niu ⁶

¹ College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing, Jiangsu, 210024, China

² State Key Laboratory of Water Disaster Prevention, Nanjing, Jiangsu, 210024, China

³ Department of Civil, Environmental, and Geospatial Engineering, Michigan Technological University, Houghton, Michigan, 49931, United States

⁴ Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, 2628 CN, Netherlands

⁵ Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, Boulder, Colorado, 80309, United States

⁶ Shanghai Research Institute of Building Science Co., Ltd., Shanghai, 201108, China

* Correspondence: xiaotan@hhu.edu.cn (X.T.); dongj@mtu.edu (D.J.)

Abstract

This paper presents an interpretable machine-learning framework for predicting the splitting strength (ST) of asphalt concrete and supporting data-driven mixture design. A database consisting of 296 samples was established, and 14 input variables related to asphalt properties, aggregate gradation, and fiber characteristics were selected for modeling. Six machine-learning models, namely TabPFN, ANN, SVR, RF, XGBoost, and LightGBM, were developed and compared. Hyperparameter optimization was performed for five models using NSGA-II, while TabPFN was directly applied with its default configuration. The results show that all six models achieved satisfactory predictive capability, whereas TabPFN delivered the best overall performance on the testing set, with the lowest RMSE of 0.28, MAE of 0.21, MAPE of 18.01%, MAD of 0.14, the highest R² of 0.88, and the highest composite score of 0.91. SHAP analysis further revealed that nine dominant variables accounted for 92.0% of the total average contribution, among which Ag9.5, FT, Ag4.75, AC, and Du were the most influential. In addition, favorable parameter ranges for improving ST were quantified, such as Ag9.5 < 66.8%, Ag4.75 < 45.0%, AC < 5.4 wt.%, AV < 3.6%, and Du > 134.7 cm. Finally, a GUI platform integrating prediction and SHAP-based explanation was developed to improve the accessibility and practical applicability of the proposed framework.

Keywords: asphalt concrete; splitting strength; TabPFN; explainable artificial intelligence; SHAP interpretation

1. Introduction

Asphalt concrete has been extensively applied in transportation and hydraulic engineering, including road pavements [1], airport runways [2], parking areas [3], and embankment dams [4]. This wide application is mainly attributed to its favorable waterproofing ability, convenient maintenance, and economic efficiency [5–7]. Nevertheless, under practical service conditions, conventional asphalt concrete remains vulnerable to several problems, such as low-temperature brittleness, high-temperature deformation, and gradual deterioration caused by moisture, repeated loading, aging, and temperature variation [8–10]. These factors can accelerate cracking and adversely affect the durability and serviceability of pavement structures. Among the commonly used performance

indices, splitting strength (ST) is of particular importance because it can directly characterize the tensile resistance and crack susceptibility of asphalt mixtures. Therefore, establishing reliable approaches for ST prediction and identifying the major factors governing its variation are of clear significance for the evaluation and optimization of asphalt concrete.

Conventional experimental evaluation of asphalt concrete performance is generally associated with high cost and low efficiency, since changes in binder content, aggregate properties, or gradation often require repeated and time-consuming laboratory testing [11–16]. To reduce this burden, researchers have developed analytical and empirical approaches to estimate mechanical properties [17]. However, although these methods may achieve satisfactory fitting accuracy in specific cases, their applicability is often constrained by simplified theoretical assumptions and limited adaptability to complex variations in mixture design. Mechanistic-based prediction frameworks have also been employed to assess long-term pavement performance, especially in relation to fatigue and rutting behavior [18]. Nevertheless, the dependence of such approaches on fixed material parameters and predefined structural assumptions reduces their generalization capability when dealing with heterogeneous asphalt mixtures [19].

Machine learning (ML) has recently become an important analytical approach in concrete-materials research [20,21], particularly for identifying hidden patterns and forecasting material properties from complex datasets. Because ML methods are well suited to large, heterogeneous data and can represent nonlinear correlations among multiple material parameters, they have been increasingly adopted for investigating the mechanical performance of engineering materials [22]. In asphalt concrete, models such as artificial neural networks (ANN), random forests (RF), k-nearest neighbors (KNN), and light gradient boosting machines (LightGBM) [21–24], have already been introduced for the prediction of major mechanical properties, and the reported results indicate satisfactory accuracy. For example, RF achieved an R^2 of 0.83 in predicting Marshall stability [23], while KNN and LightGBM yielded even higher predictive performance, reaching an R^2 of 0.90 [24]. These results indicate that ML-based approaches can provide more flexible and accurate predictions than traditional empirical models, particularly when dealing with diverse mixture compositions and complex variable interactions.

Despite the progress achieved so far, the broader use of machine learning in asphalt concrete research is still hindered by several challenges: the machine-learning models adopted in previous studies are mostly conventional, whereas newer and potentially more powerful models have rarely been explored in asphalt concrete research; (2) existing studies have mainly emphasized prediction accuracy, whereas integrated frameworks that combine model comparison, comprehensive evaluation, and interpretability analysis remain limited, thereby restricting a transparent understanding of how input variables influence splitting strength [25]; and (3) publicly accessible platforms for asphalt concrete splitting strength prediction are still scarce, limiting the practical usability of these methods for engineers and researchers.

To address these issues, an interpretable machine-learning framework was established in this study for predicting the ST of asphalt concrete. A literature-derived dataset incorporating asphalt-related properties, aggregate gradation parameters, and mixture-design variables was used to benchmark six ML models, after which the optimal model was selected through a comprehensive performance assessment. To provide insight into the prediction process, SHapley Additive exPlanations (SHAP) was employed to determine the most influential variables and to characterize how they contributed to ST variation. In addition, a graphical user interface (GUI) was built to combine prediction with interpretation, thereby facilitating practical use of the proposed method.

The novelties of this study can be summarized in three aspects: (1) Tabular Prior-data Fitted Network (TabPFN) was introduced as a state-of-art machine-learning paradigm for ST prediction of asphalt concrete, and its ability to deliver excellent performance without dataset-specific hyperparameter tuning was verified, highlighting its efficiency and applicability for small-to-medium-sized tabular datasets; (2) an interpretable machine-learning framework was established for asphalt concrete ST prediction by combining data preprocessing, multi-model comparison,

composite-indicator-based evaluation, and SHAP analysis, which not only improved predictive reliability but also revealed the relative importance of input variables and their influence patterns on ST; and (3) an intuitive graphical user interface (GUI) platform was developed to integrate prediction and interpretation into an interactive tool for asphalt concrete evaluation and design.

2. Methodology

Figure 1 illustrates the workflow of the proposed interpretable machine-learning framework, consisting of four steps:

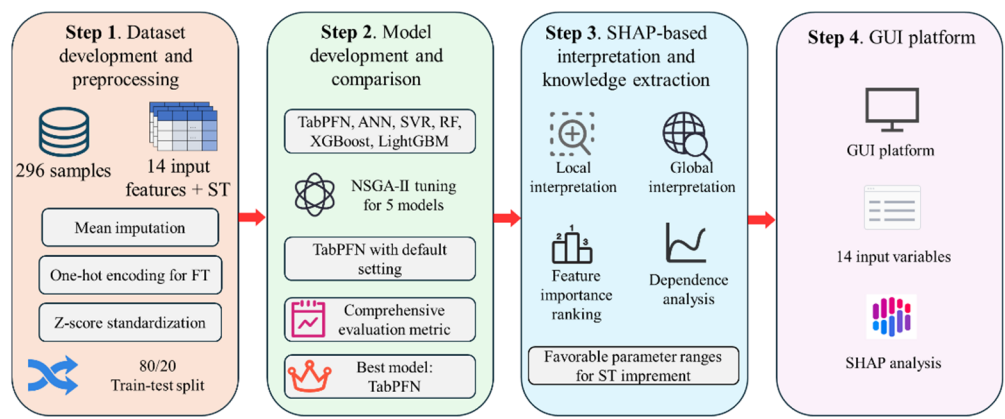


Figure 1. Interpretable machine-learning framework for asphalt concrete splitting strength prediction, SHAP-based interpretation, and GUI application.

(1) Dataset development and preprocessing, where 296 samples with 14 input and ST as the output were prepared using mean imputation, one-hot encoding, Z-score standardization, and an 80:20 train-test split. (2) Model development and comparison, where six models were developed, with NSGA-II used for hyperparameters optimization, and TabPFN identified as the best performer. (3) SHAP-based interpretation for feature importance and parameter analysis. (4) A GUI platform integrating trained model and SHAP results for interactive prediction and explanation.

2.1. Database Development

2.1.1. Database Construction and Description

A database containing 296 asphalt concrete samples was established from relevant studies published between 2008 and 2025 for ST prediction [26–56]. Taking into account the material properties of asphalt, aggregate gradation, and fibers, fourteen input variables were selected and classified into three groups to investigate their effects on splitting strength [57]: asphalt-related features, including asphalt content (AC), penetration (Pe), softening point (SP) and ductility (Du); aggregate-related features, including the passing percentages of 2.36 mm, 4.75 mm, and 9.5 mm aggregates (Ag2.36, Ag4.75, and Ag9.5), air voids (AV), voids in mineral aggregate (VMA), and voids filled with asphalt (VFA); and fiber-related features, namely fiber content (FC), fiber type (FT), tensile strength (TS) and fiber length (FL). When the samples are classified according to fiber type, the database includes basalt fiber, glass fiber, polyester fiber, steel fiber, and no fiber, and their distribution proportions are shown in Table 1.

Table 1. Overview of fiber types and their amounts.

Fiber types	Sample size
Basalt fiber	17
Glass fiber	14
Polyester fiber	20

Steel fiber	4
No fiber	241

2.1.2. Data Analysis

The missing values in the database were filled using mean imputation. Following imputation, the statistical profiles of all variables are summarized in **Table 2**, covering the minimum, maximum, quartile values, mean, and standard deviation for each variable.

Table 2. Distribution of the fourteen input variables and the output variable.

Variable	Unit	Min	Q1	Q2	Q3	Max	Mean	STD
<i>Pe</i>	0.1mm	47	63	71.2	85.9	93	71.65	14.05
<i>Du</i>	cm	98	100	101	150	200	125.65	31.21
<i>SP</i>	°C	44.1	47.2	50	57	73	53	7.87
<i>AC</i>	% by mass	3	4.6	4.9	6.5	8	5.35	1.17
<i>Ag2.36</i>	%	13.9	26.58	32.92	40.15	56	33.54	10.9
<i>Ag4.75</i>	%	23.9	37.9	50.77	58.89	71	47.66	13.53
<i>Ag9.5</i>	%	53	62.76	76.16	81.2	86	72.88	10.3
<i>AV</i>	%	2.54	4.01	4.34	4.95	8	4.41	0.98
<i>VMA</i>	%	12.1	14.94	15.36	16.2	65.6	17.08	8.65
<i>VFA</i>	%	17.11	69.03	72.95	82.59	83.41	72.91	11.81
<i>FC</i>	%	0	0	0	0	3	0.09	0.36
<i>FT</i>	/	/	/	/	/	/	/	/
<i>TS</i>	MPa	0	0	0	0	3250	237.17	735.13
<i>FL</i>	mm	0	0	0	0	12	1.23	2.94
<i>ST</i>	MPa	0.13	0.71	1.1	1.48	5.15	1.33	0.91

The probability density characteristics of the fourteen input variables and the output variable are presented in **Fig. A1**. For each feature (such as *Pe*, *Du*, and *SP*), a dual-axis subplot is used, with probability density shown on the left y-axis and frequency shown on the right y-axis. Pearson correlation analysis was further carried out to evaluate possible multicollinearity, and the results are provided in **Figure 2**. [58]. Results show that, except for a few relatively strong correlations among asphalt, aggregate, and fiber internal features, the correlations among the remaining features all satisfy $|R| < 0.7$ [59]. This indicates weak linear relationships and limited redundancy among the variables, confirming that the selected features were appropriate for model training. Because no severe multicollinearity was observed among the selected variables, all mixture design variables were kept in the input set [60].

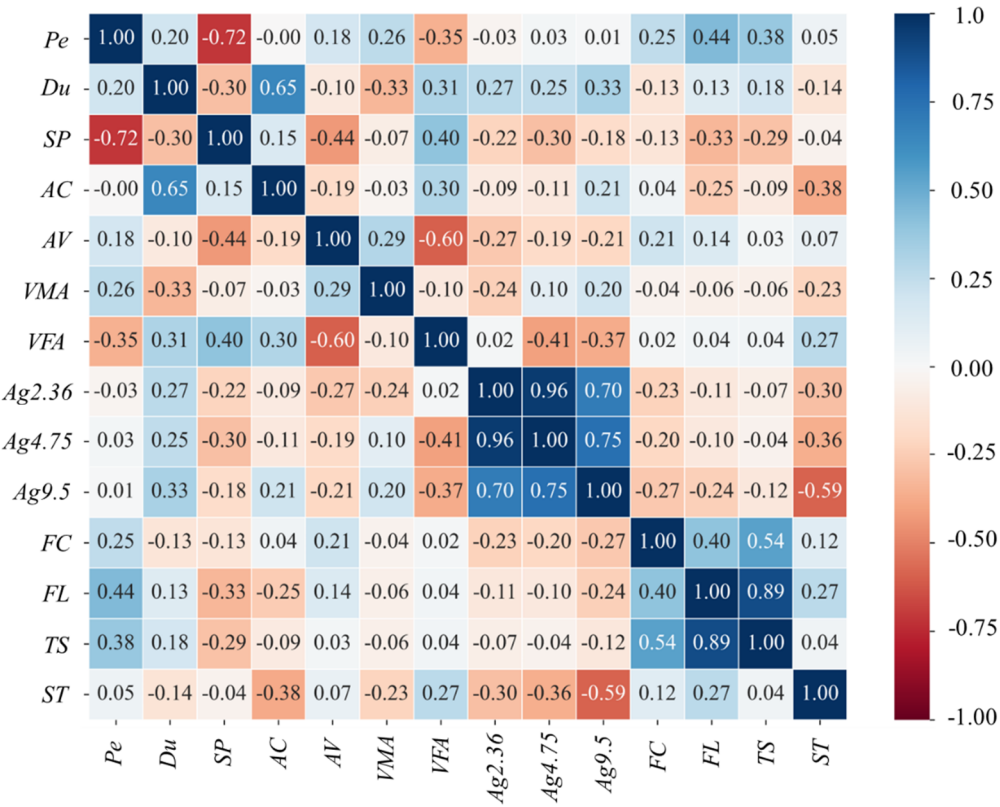


Figure 2. Pearson correlation heatmap of the dataset.

2.2.3. Data Preprocessing

For the categorical variable FT, one-hot encoding was used make it compatible with machine learning models [61]. Numerical features were standardized using Z-score scaling before training to ensure a common scale, improving efficiency and convergence by reducing the influence of large-magnitude variables [62,63]. The formula was shown as follows:

$$z = \frac{x - m}{\sigma}$$

(1)

where x denotes an input variable, m and σ represent its mean and standard deviation, respectively.

The standardized dataset was then randomly reordered to minimize possible sequence-related effects and to enhance the representativeness of the training samples [60]. After that, 80% of the samples were assigned to the training set and the remaining 20% to the testing set [64].

2.2. Machine Learning Models

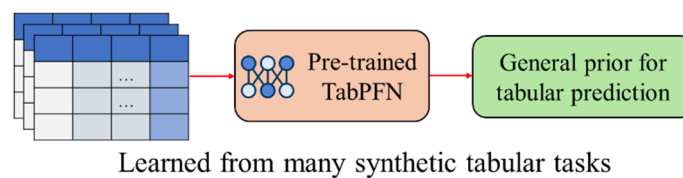
Six machine learning models were employed in this study: Tabular Prior-data Fitted Network (TabPFN) [65], Support Vector Regression (SVR) [66], Random Forest (RF) [67], Extreme Gradient Boosting Trees (XGBoost) [68], Light Gradient Boosting Machine (LightGBM) [69] and Artificial Neural Network (ANN) [70]. The classification of these models is presented in Table 3.

Table 3. Overview of the machine learning models.

No.	Model	Category	Notes
1	TabPFN	Foundation model	Transformer-based prediction
2	ANN	Classical	Nonlinear regression
3	SVR	Classical	Kernel-based regression
4	RF	Ensemble – Bagging	Bagging of decision trees
5	XGBoost	Ensemble – Boosting	Boosting model with regularization

Among the six models, TabPFN deserves particular attention because it differs fundamentally from conventional tabular learning models. As illustrated in **Figure 3(a)**, TabPFN is a pretrained tabular foundation model trained on a large set of synthetic tabular tasks drawn from diverse data-generating processes [71]. Through this pretraining, it learns a general prior for tabular prediction, allowing it to capture transferable patterns instead of being trained from scratch for each dataset. As shown in **Figure 3(b)**, in the present study, the fourteen input descriptors of asphalt concrete, including asphalt-related, aggregate-related, and fiber-related variables, were organized into tabular inputs and then fed into the pretrained TabPFN model to predict the target variable ST. Its underlying architecture is Transformer-based, in which tabular inputs are encoded and processed through stacked Transformer blocks before being passed to the prediction head [65]. This design makes TabPFN particularly suitable for the present ST database, which contains a moderate number of samples and a compact set of descriptors, because it reduces the need for repeated dataset-specific retraining and extensive hyperparameter optimization while still maintaining strong predictive performance. In this study, the open-source TabPFN from PriorLabs was used via its regressor interface for prediction [72].

(a) Pre-training on synthetic tabular tasks



(b) Application to asphalt concrete dataset

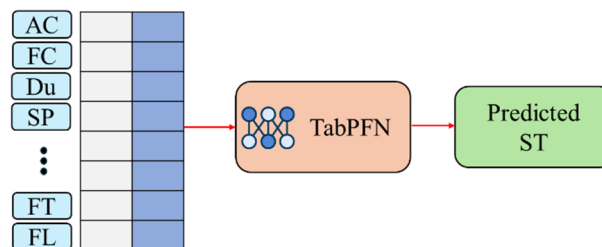


Figure 3. Schematic diagram of TabPFN pre-training and its application to ST prediction.

2.3. Evaluation Metrics

Model performance was evaluated using five indicators: root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), median absolute deviation (MAD) of the residuals, and the coefficient of determination (R^2). Their mathematical definitions are given in Eqs. (2)–(6). Among them, MAE reflects the average error in original units and is easily interpretable. RMSE emphasizes larger errors, while MAPE expresses error as a percentage. MAD is more robust to outliers, and R^2 indicates the proportion of variance explained, with values closer to 1 showing better fit. In general, lower RMSE, MAPE, MAE, and MAD values indicate higher predictive accuracy, whereas a higher R^2 suggests superior model performance.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{i,pre} - y_{i,test})^2} \quad (2)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_{i,pre} - y_{i,test}}{y_{i,test}} \right| \times 100\% \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{i,pre} - y_{i,test}| \quad (4)$$

$$\begin{cases} e_i = y_i - \hat{y}_i \\ MAD = \text{median}(|e_i - \text{median}(e)|) \end{cases} \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{i,test} - y_{i,pre})^2}{\sum_{i=1}^n (y_{i,test} - \bar{y})^2} \quad (6)$$

Note: For each target variable, n represents the number of samples in the testing (or training) set, y_i and $\hat{y}_i$ are the actual and predicted values of the i -th sample, respectively; $\bar{y}$ is the mean of the ground-truth values; and e_i represents the residual for the i -th sample.

To enable an integrated comparison across metrics with different scales and optimization directions, the five metrics were further transformed into dimensionless bounded scores within the interval (0,1), where a larger value consistently indicates better performance. Since RMSE, MAPE, MAE, and MAD are error-type metrics for which smaller values are preferred, whereas a larger R^2 is preferred, the first step was to convert all metrics into a unified “higher-is-better” utility:

$$u_{k,m} = \begin{cases} -x_{k,m}, & k \in \{RMSE, MAPE, MAE, MAD\} \\ x_{k,m}, & k \in R^2 \end{cases} \quad (7)$$

where $x_{k,m}$ denotes the raw value of metric k for model m , and $u_{k,m}$ is the corresponding utility value after directional unification.

Then, for each metric k , z-score standardization was performed across all candidate models:

$$z_{k,m} = \frac{u_{k,m} - \mu_k}{\sigma_k} \quad (8)$$

where μ_k and σ_k are the mean and standard deviation of the utility values of metric k across all models, respectively.

Finally, to obtain a stable bounded score and reduce the dominance of extreme values, the standardized values were mapped into the interval (0,1) using a sigmoid function:

$$s_{k,m} = \frac{1}{1 + \exp(-z_{k,m}/\alpha)} \quad (9)$$

where α is a scaling parameter controlling the steepness of the transformation. In this study, $\alpha=0.5$.

Based on the five normalized metric scores, the composite score of models m was calculated as the arithmetic mean of the normalized scores over the five metrics:

$$\text{Composite Score}_m = \frac{1}{5} \sum_{k \in K} s_{k,m} \quad (10)$$

$$\text{where } K = \{RMSE, MAPE, MAE, MAD, R^2\} \quad (11)$$

A larger composite score indicates better overall predictive performance after jointly considering prediction accuracy, goodness of fit, and residual stability. In this study, the normalized metric scores were used to construct the radar chart in Figure 8(a), while the composite score was used to rank the overall performance of different models in Figure 8(b).

2.4. Hyperparameter Tuning Through Objective Optimization

Unlike conventional machine learning models, TabPFN was not subjected to hyperparameter tuning in this study. This is because TabPFN is designed as a pretrained tabular foundation model, whose predictive capability mainly stems from large-scale prior pretraining rather than dataset-

specific parameter adjustment. According to the characteristics of the model itself and the recommendation of its original authors, TabPFN is intended to be used largely in its default configuration, thereby avoiding the expensive and often unnecessary hyperparameter optimization process required by many traditional machine learning algorithms. This property is also one of the practical advantages of TabPFN, especially for small-to-medium-sized tabular datasets, as it allows strong predictive performance to be achieved with minimal manual intervention [65].

For the remaining machine learning models, hyperparameter optimization plays an important role in improving predictive performance. Common approaches include Grid Search [73], Random Search [74] and Genetic Algorithms-based method [75]. Grid Search and Random Search were not adopted in this study because they become computationally inefficient when exploring high-dimensional hyperparameter spaces and are less effective in capturing complex interactions among hyperparameters [76]. Therefore, Non-dominated Sorting Genetic Algorithm II (NSGA-II), a Genetic Algorithms-based method, was employed. NSGA-II provides an efficient global search strategy by maintaining population diversity and balancing exploration and exploitation during optimization process [77]. It searches hyperparameter space through an evolutionary process in which each generation contains multiple candidate solutions, and each individual represents a specific hyperparameter setting. In present work, the tuning process aimed to minimize RMSE of target variable ST. The corresponding tuning procedure is illustrated in **Figure 4**.

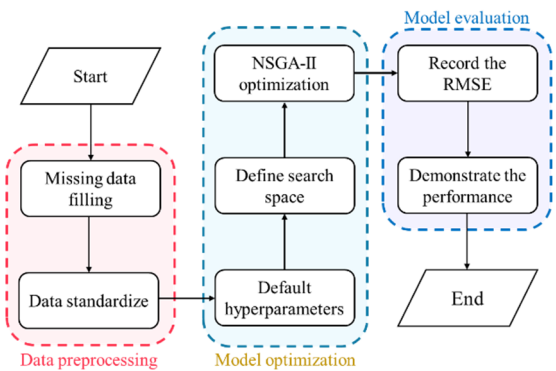


Figure 4. Flowchart of the NSGA-II-based hyperparameter tuning process.

2.5. SHAP-Based Model Explanation

ML models have shown strong predictive capability in materials engineering, yet their practical application is often constrained by limited interpretability because many of them function as “black-box” systems. To address this issue, SHAP [78] can be employed as a post hoc interpretation tool to quantify the contribution of each input variable and thus provide transparent explanations for model outputs. As illustrated in **Figure 5**, the SHAP-based explanation framework can be understood in four parts. In **Figure 5(a)**, the input variables are first fed into the trained ML model, which operates as a black-box predictor and generates the target prediction. **Figure 5(b)** then shows that this prediction can be decomposed into a baseline value together with the contribution of individual input features, expressed as SHAP values. **Figure 5(c)** further illustrates the additive principle of SHAP, in which the final prediction is obtained by adjusting the base value through the positive or negative effects of different variables. In this process, positive SHAP values increase the prediction, whereas negative SHAP values decrease it. Finally, **Figure 5(d)** presents the interpretation results at both local and global levels. Local analysis explains how a single prediction is formed, while global analysis summarizes the overall importance and influence patterns of features across the entire dataset. In this way, SHAP converts the original black-box prediction into an interpretable explanation framework, thereby improving the transparency of the model decision-making process.

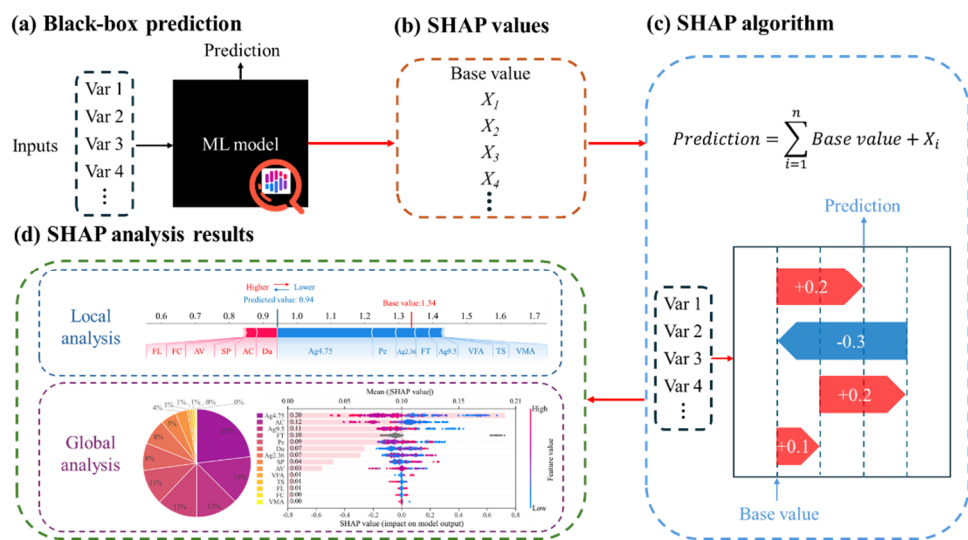


Figure 5. SHAP-based interpretation process for machine learning model predictions.

3. Results and Discussion

3.1. Hyperparameter Optimization Results

Optimal hyperparameter combinations were identified using the pymoo-based NSGA-II algorithm, which was executed with a population size of 16 over 100 generations. For each candidate hyperparameter set, the objective value was defined as the mean RMSE from 5-fold cross-validation on the training data, and the optimization process aimed to minimize this value.

Based on the above optimization settings, the termination criterion was defined as 100 generations. As shown in **Figure 6**, all five models converged well before the preset maximum generation, indicating that this setting provided sufficient search depth while avoiding unnecessary computational expense. Specifically, RF, XGBoost, and LightGBM reached stable validation RMSE values within the early generations, whereas SVR showed only slight improvement after its initial convergence. By contrast, ANN exhibited a relatively slower optimization process, with a pronounced reduction in RMSE during the early generations and a gradual plateau after approximately 30–40 generations, followed by only marginal improvement thereafter. Overall, Figure 6 illustrates the evolution of the best validation RMSE during the NSGA-II optimization process, confirming that the selected generation limit was adequate and that extending the search further would be unlikely to produce substantial performance gains. The resulting hyperparameter combinations for all tuned models are provided in **Appendix B, Table B1**.

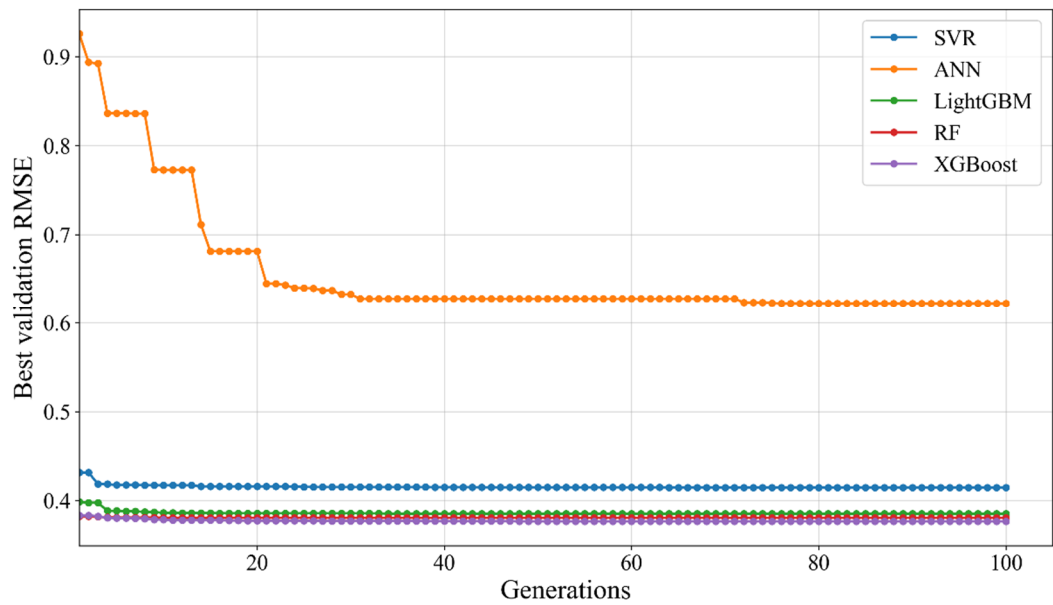
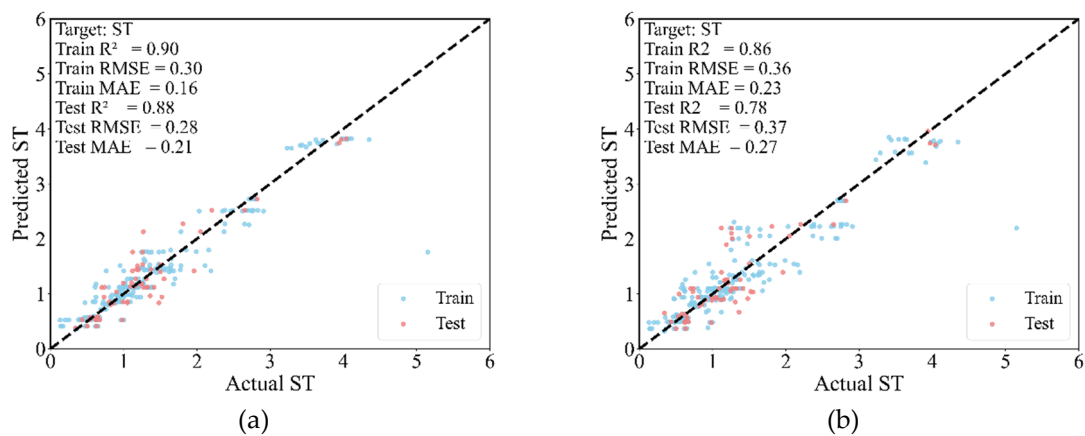


Figure 6. RMSE variation of the models during NSGA-II iterations.

3.2. Prediction Performance

Figure 7 compares the fitting and generalization behavior of the six machine-learning models for ST prediction using the training and testing datasets. The scatter points for the two datasets are plotted against the 1:1 reference line, which represents perfect agreement between model outputs and measured values. Data points located closer to this reference line indicate stronger predictive consistency and smaller deviations. It is also noted that MAE is lower than RMSE for all models, which is consistent with the general expectation for satisfactory machine-learning predictions [79]. Overall, all six models were capable of predicting ST with satisfactory accuracy. Among them, TabPFN achieved the best overall performance on the testing set, with the lowest RMSE (0.28) and the highest R² value of 0.88. The detailed performance metrics of all models on the testing set are provided in Appendix B, Table B2.



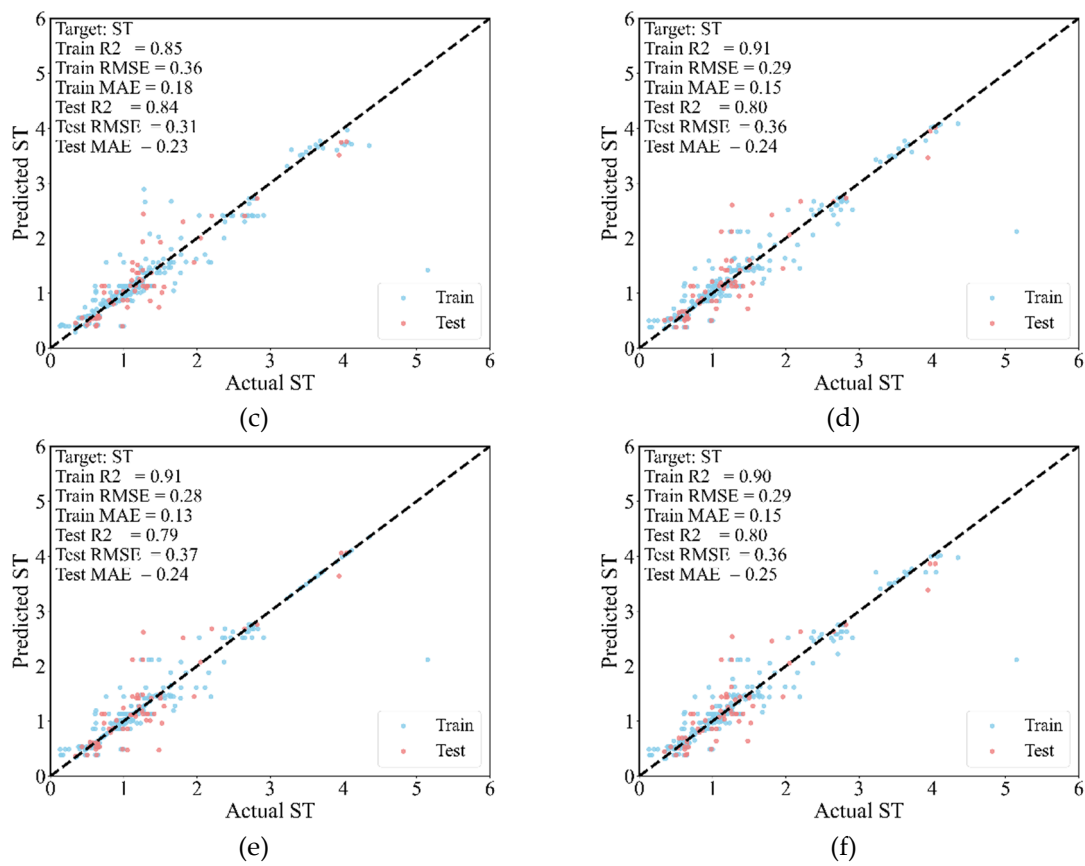


Figure 7. Performance of ML models for ST prediction on the training and testing sets: (a) TabPFN; (b) ANN; (c) SVR; (d) RF (e) XGBoost; (f) LightGBM.

The multi-metric predictive performance of the six models for ST is summarized in **Figure 8**. As shown in **Figure 8(a)**, TabPFN achieved the best overall performance across the five metrics, showing clear advantages in RMSE, MAPE, MAE, MAD, and R². SVR ranked second and also exhibited relatively balanced predictive capability. XGBoost, RF, and LightGBM showed moderate overall performance, while ANN obtained the lowest scores among the six models. This trend is further confirmed by the composite scores in **Figure 8(b)**, where TabPFN reached the highest score of 0.91, followed by SVR (0.66). Overall, the results indicate that TabPFN is the most effective model for ST prediction in the present dataset, whereas the other models still provide acceptable but comparatively weaker performance.

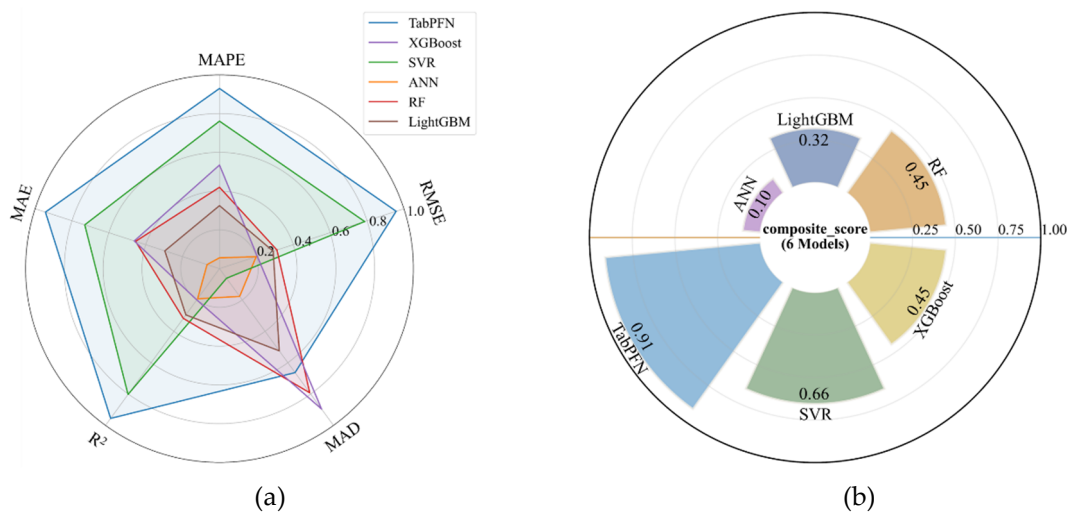
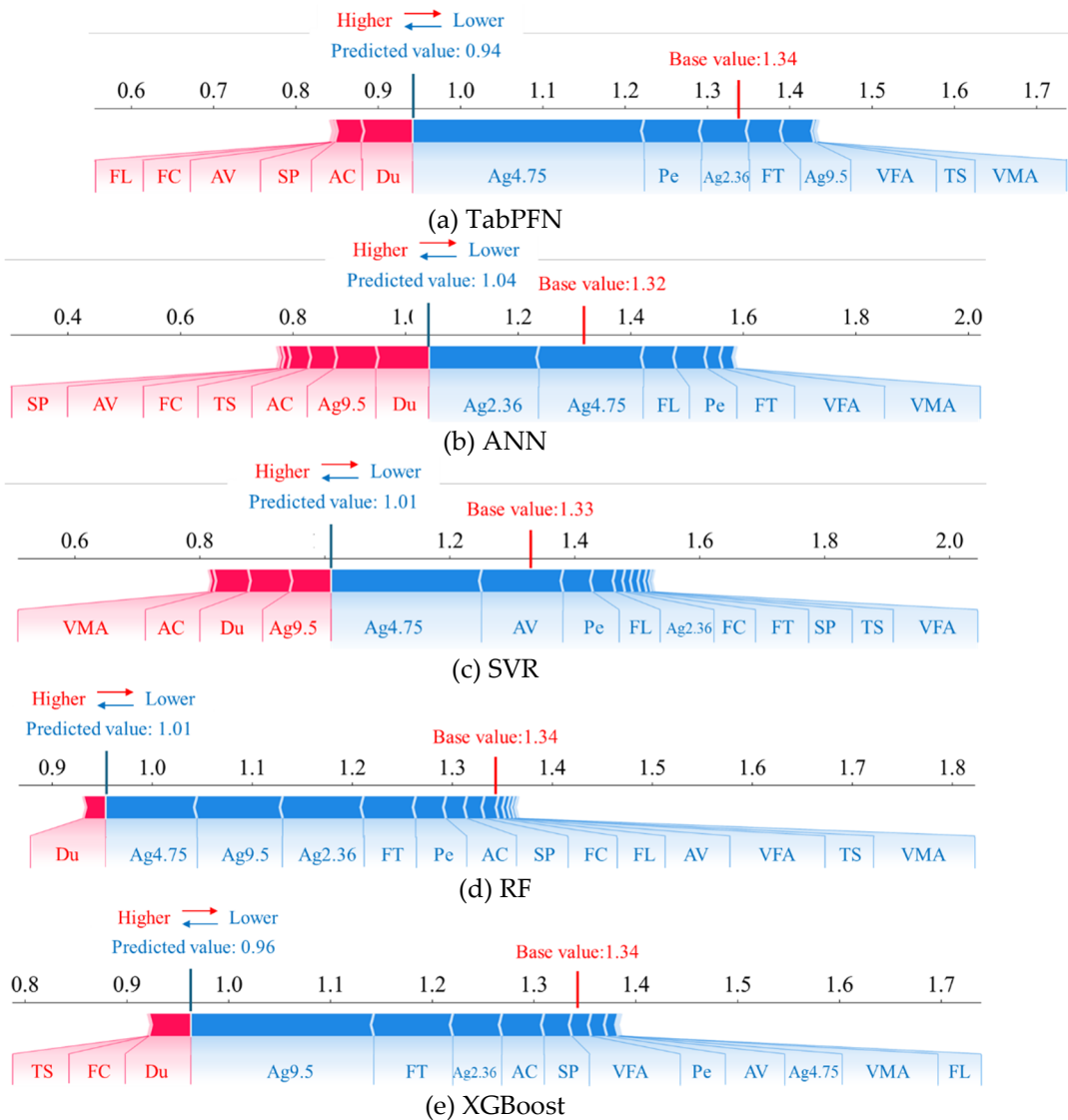


Figure 8. Performance evaluation of different models for ST prediction: (a) radar chart based on five evaluation metrics; (b) composite score integrating the five metrics.

3.3. Local Interpretability Based on SHAP

To provide a case-level interpretation of the prediction behavior of the six machine-learning models, a randomly chosen sample was examined in detail. The corresponding feature values were AC = 4.6 wt.%, Pe = 72 (0.1 mm), Du = 137 cm, SP = 51 °C, AV = 4.48%, VMA = 17.51%, VFA = 71.97%, Ag2.36 = 56%, Ag4.75 = 69%, Ag9.5 = 86%, FT = No_fiber, FC = 0 wt.%, FL = 0 mm, TS = 0 MPa, and ST = 0.93 MPa. As shown in the SHAP force plots in **Figure 9**, all six models predict the selected sample at values lower than their corresponding base values, indicating an overall downward shift from the average prediction. Although the dominant contributing features vary across models, some consistent patterns can still be observed. In particular, FT (No_fiber) shows a negative contribution in all six models, while VMA, Ag4.75, Ag2.36, and Pe also tend to reduce the predicted ST in most cases. By contrast, the positive contributions are more model-dependent, with features such as Du, AC, SP, AV, FC, TS, and FL increasing the predicted ST in some models. The final ST values predicted by TabPFN, ANN, SVR, RF, XGBoost, and LightGBM were 0.94, 1.04, 1.01, 0.96, 0.97, and 0.97 MPa, respectively. Given that the actual ST value of the selected sample is 0.93 MPa, the corresponding relative errors are 1.08%, 11.83%, 8.60%, 3.23%, 4.30%, and 4.30%, respectively. Among the six models, TabPFN yields the closest prediction to the actual value, while RF, XGBoost, and LightGBM also maintain prediction errors within 5%, indicating satisfactory local predictive accuracy for this sample.



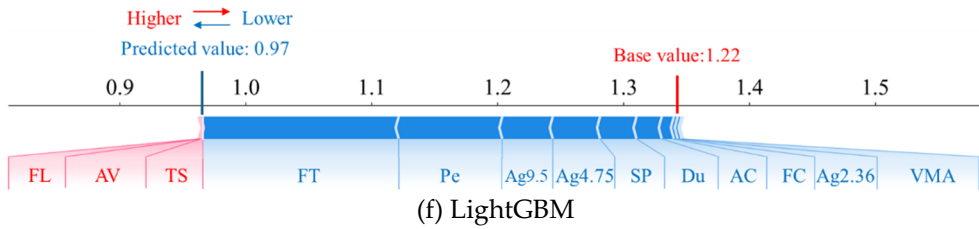


Figure 9. Illustration of the prediction behavior of the six machine-learning models. Red bars represent positive effects, whereas blue bars represent negative effects.

3.4. Global Interpretability Based on SHAP

3.4.1. Contribution of Individual Features

In addition to local explanation for a single sample, SHAP can also be used to quantify the overall influence of input variables on ST prediction across the full dataset. **Figure 10** presents the global SHAP interpretation results of the TabPFN model, selected because of its superior predictive accuracy. The pie chart reports the proportional contribution of each feature based on the sum of absolute SHAP values, while the beeswarm plot further illustrates the distribution and polarity of feature effects for all samples. In the beeswarm plot, each point represents a sample, and the color scale indicates the feature value from low to high. The horizontal axis corresponds to the SHAP value, where positive and negative values denote increasing and decreasing effects on the predicted ST, respectively. Features are ordered by mean absolute SHAP value, allowing direct comparison of their overall importance. The results show that Ag4.75 is the most influential predictor, with AC, Ag9.5, FT, Pe, and Du also contributing substantially, whereas VMA, FC, FL, and TS play relatively minor roles in the model output. The beeswarm plots of the remaining five models are shown in **Fig. C1** of Appendix C.

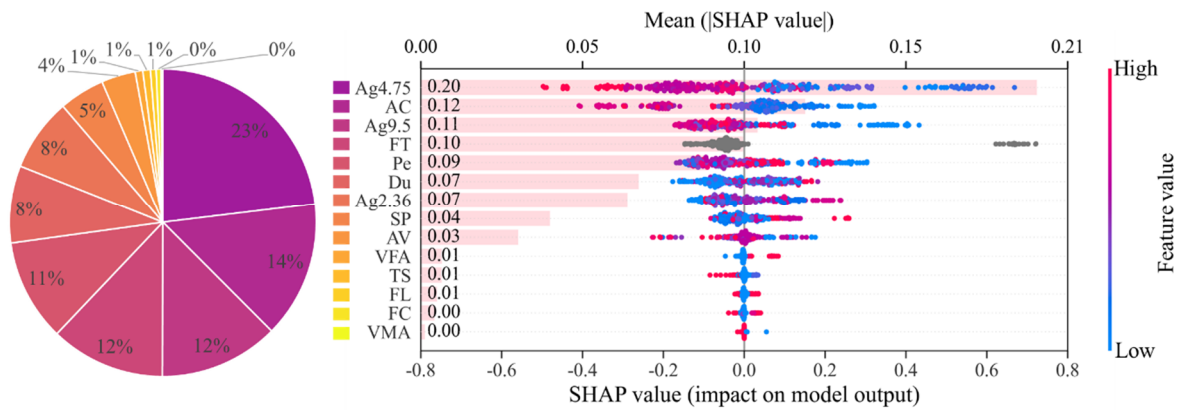


Figure 10. SHAP analysis of the TabPFN predictions.

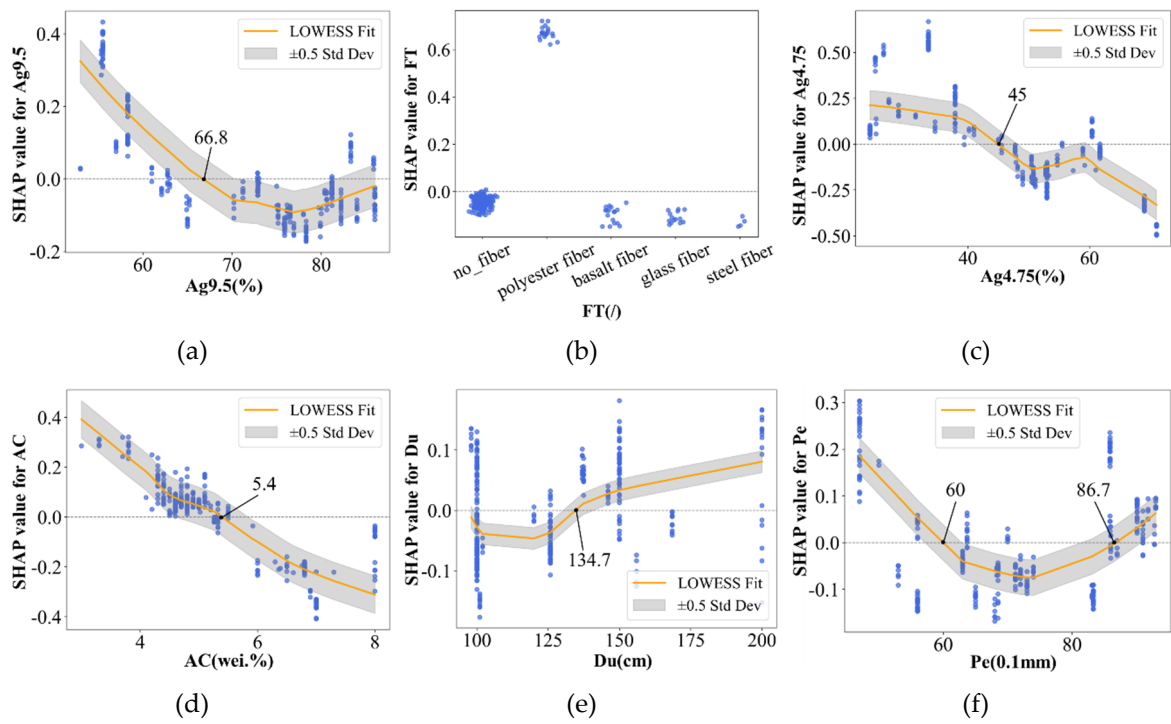
As reflected by the pie charts in **Figure 10** and **Fig. C1**, the relative importance of individual variables is not exactly identical across models; however, a clear overall pattern can still be observed. Based on the averaged percentages summarized in **Table 4**, Ag9.5, FT, Ag4.75, AC, and Du are identified as high-impact variables, each contributing more than 10% on average. Pe, Ag2.36, SP, and AV fall into the medium-impact group, with average contributions between 5% and 10%. In contrast, FL, FC, TS, VFA, and VMA exhibit average contribution percentages below 5%, indicating relatively limited influence on the model output. In aggregate, high-, medium-, and low-impact features account for 65.5%, 26.5%, and 7.9% of the total influence, respectively, highlighting the dominant role of a small subset of variables in determining ST prediction.

Table 4. Ranking of variable importance for ST prediction.

Variables	Proportion of models (%)						Mean (%)
	TabPFN	ANN	SVR	RF	XGBoost	LightGBM	
<i>Ag9.5</i>	12	11	15	22	33	20	18.8
<i>FT</i>	12	7	7	12	14	25	12.8
<i>Ag4.75</i>	23	15	13	11	1	7	11.7
<i>AC</i>	14	15	11	10	8	9	11.2
<i>Du</i>	8	10	9	10	17	12	11
<i>Pe</i>	11	7	9	8	7	11	8.8
<i>Ag2.36</i>	8	6	7	8	3	4	6
<i>SP</i>	5	8	7	6	7	3	6
<i>AV</i>	4	5	7	5	6	7	5.7
<i>FL</i>	1	8	7	2	0	0	3
<i>FC</i>	0	1	4	4	1	1	1.8
<i>TS</i>	1	4	2	1	0	0	1.3
<i>VFA</i>	1	2	1	1	3	0	1.3
<i>VMA</i>	0	1	1	0	0	1	0.5

3.4.2. Feature-Wise Dependence Analysis

To examine the variation patterns and possible threshold behaviors of key variables affecting ST, **Figure 11** displays the SHAP dependence plots of all input features obtained from the TabPFN model, thereby revealing how each parameter influences the ST of asphalt concrete.



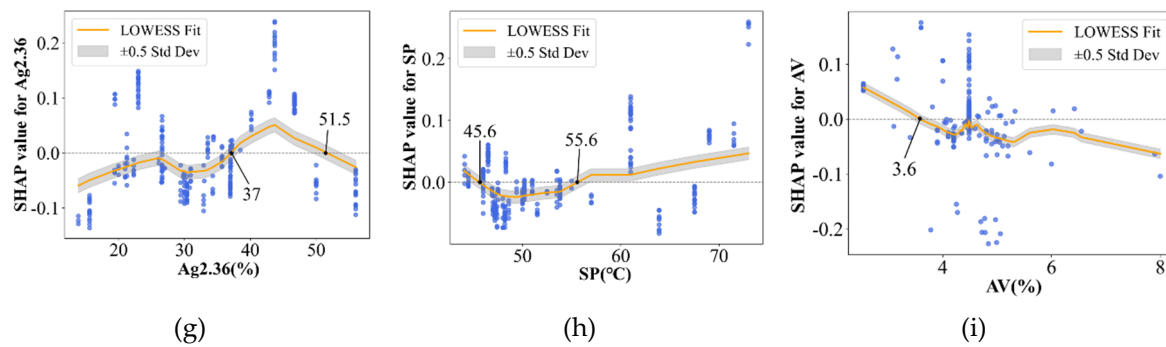


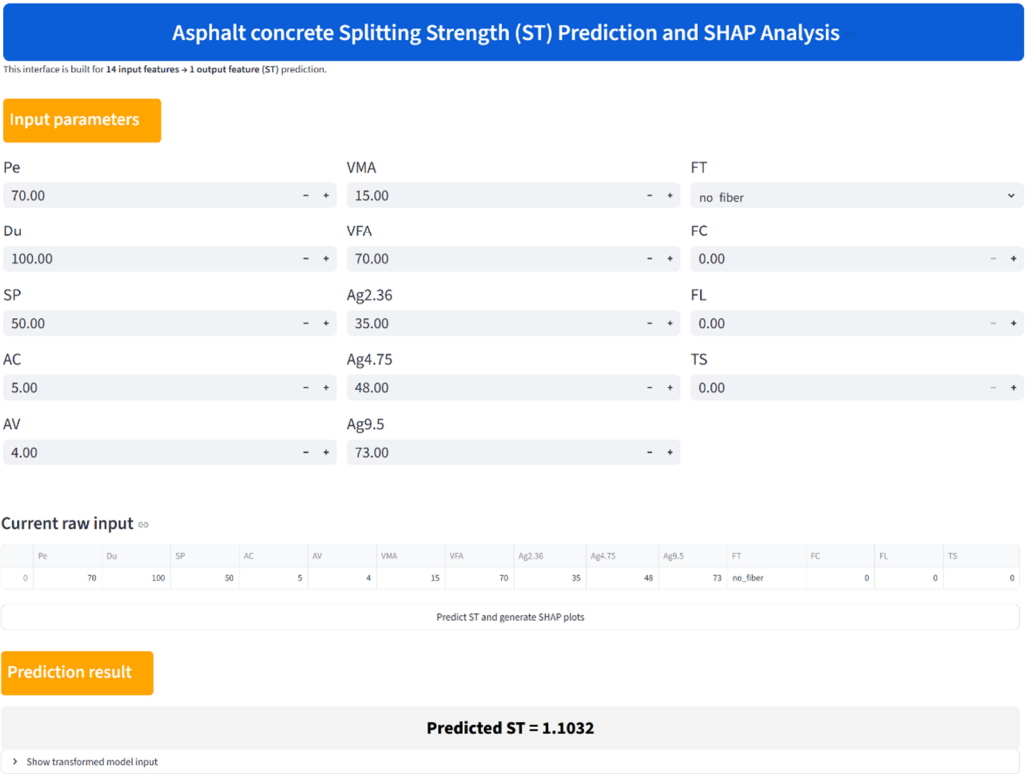
Figure 11. SHAP dependence plots of the nine key input variables for ST prediction based on the TabPFN model: (a) Ag9.5; (b) FT; (c) Ag4.75; (d) AC; (e) Du; (f) Pe; (g) Ag2.36; (h) SP; (i) AV.

Based on **Table 4**, the nine variables shown in **Figure 11** (Ag9.5, FT, Ag4.75, AC, Du, Pe, Ag2.36, SP, and AV) were selected for dependence analysis because they comprise all high- and medium-impact features and jointly explain 92.0% of the total average SHAP contribution for ST prediction. Their SHAP dependence plots were fitted with LOWESS curves and accompanied by ± 0.5 standard deviation error bands to characterize nonlinear trends and local uncertainty. This visualization facilitates parametric interpretation of the relationships between feature values and their corresponding SHAP effects on the predicted splitting strength. In contrast, the five low-impact variables (FL, FC, TS, VFA, and VMA), whose combined contribution is only 7.9%, were excluded from further discussion. Since FT is a categorical descriptor, its plot is presented as grouped scatter distributions for different fiber types rather than as a continuous fitted curve.

When the baseline ST value is 1.34 (see **Figure 9(a)**), the SHAP dependence plots in **Figure 11** show that the nine dominant variables can be classified into three types according to their influence patterns. First, Ag9.5, Ag4.75, AC, and AV exhibit overall negative correlations with ST, as shown in **Figure 11(a), (c), (d), and (i)**, suggesting that larger values of these variables are generally unfavorable for improving splitting strength. From a physical perspective, this pattern can be attributed to the progressive weakening of the internal load-carrying structure of the mixture. Higher Ag9.5 and Ag4.75 passing rates generally indicate a gradation shift toward a less effective aggregate skeleton and weaker interlocking action [80]. Likewise, excessive AC may lead to an overly binder-rich system, in which thick asphalt films reduce the contribution of aggregate interlock to tensile resistance [81]. For AV, its negative effect is more direct, since a higher void content increases internal discontinuities and stress concentration, thus making crack initiation and propagation easier during the splitting process [82]. Second, Du shows an overall positive correlation with ST, as illustrated in **Figure 11(e)**, with its SHAP contribution becoming positive after a certain threshold. This trend is physically reasonable because higher ductility indicates a greater ability of the asphalt binder to accommodate tensile deformation and dissipate fracture energy, thereby reducing stress concentration and delaying crack propagation during the splitting process [83]. Third, Pe, Ag2.36, and SP display non-monotonic effects on ST, as shown in **Figure 11(f), (g), and (h)**, indicating that their contributions vary across different value ranges. This non-monotonic behavior is physically plausible because Pe, Ag2.36, and SP affect ST through structural and binder-property balance rather than through a simple linear mechanism, so only certain value ranges are favorable for resisting splitting failure [84–86]. For the categorical variable FT in **Figure 11(b)**, the SHAP distribution reveals a pronounced category-dependent pattern rather than a uniform fiber-reinforcement effect. Polyester fiber is associated with the highest positive contribution, whereas the other fiber categories show lower or near-neutral SHAP values. Given the imbalance among fiber categories in the present database, these results should be interpreted as dataset-dependent relative model contributions rather than definitive judgments on the intrinsic effectiveness of individual fiber types. According to these relationships, the following thresholds or favorable ranges are suggested for enhancing ST performance: $\text{Ag9.5} < 66.8\%$, $\text{Ag4.75} < 45.0\%$, $\text{AC} < 5.4 \text{ wt.}\%$, $\text{AV} < 3.6\%$, $\text{Du} > 134.7 \text{ cm}$, $\text{Pe} < 60$ or $\text{Pe} > 86.7$ (0.1 mm), $37.0\% < \text{Ag2.36} < 51.5\%$, and $\text{SP} < 45.6^\circ\text{C}$ or $\text{SP} > 55.6^\circ\text{C}$.

4. Graphical User Interface Platform

As shown in **Figure 12**, the graphical user interface (GUI) was developed in Python based on the Streamlit framework. It enables users to enter 14 variables associated with asphalt mixture composition, aggregate gradation, and fiber characteristics for splitting strength (ST) prediction. After entering the required parameters, the current raw input is displayed in tabular form, and the user can click the prediction button to obtain the predicted ST value generated by the deployed pre-trained model. In addition to prediction, the platform provides SHAP-based interpretability analysis for the current sample. As illustrated in **Figure 12**, the GUI presents a waterfall plot to visualize how individual input features contribute positively or negatively to the final ST prediction, thereby improving the transparency and interpretability of the model. Therefore, the developed platform serves not only as a practical prediction tool, but also as an interpretable decision-support interface for asphalt concrete splitting strength evaluation. The platform is available at <https://st-gui-app-nlj7snzfjkvdaqf4yvkv.streamlit.app/>



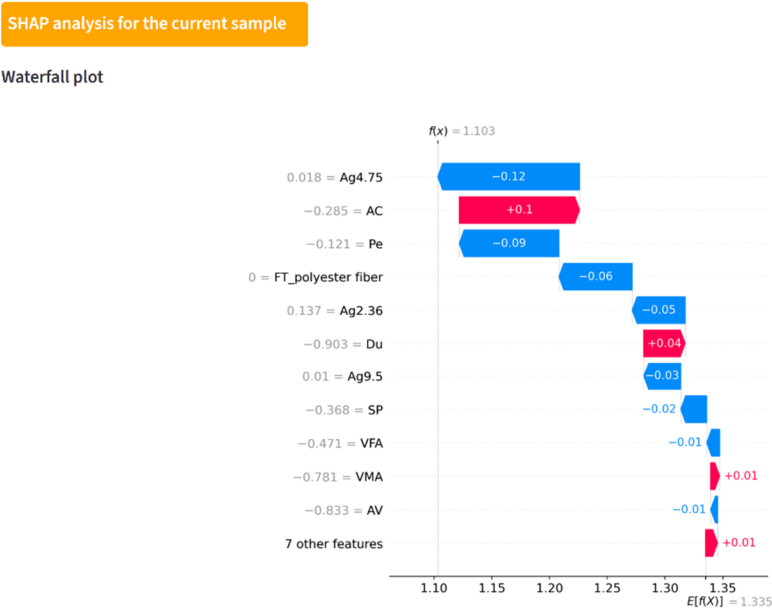


Figure 12. GUI for asphalt concrete splitting strength (ST) prediction and SHAP-based interpretation.

5. Limitations and Future Work

5.1. Overall Effectiveness

This study demonstrates the overall effectiveness of an interpretable data-driven framework for ST prediction of asphalt concrete by integrating dataset preprocessing, multi-model learning, hyperparameter optimization, SHAP-based explanation, and a user-oriented GUI platform. The results presented in Section 3 indicate that all six machine learning models achieved acceptable prediction accuracy. Among them, TabPFN showed the strongest overall performance on the testing set and obtained the highest composite score, demonstrating that the proposed framework can effectively learn the complex nonlinear relationships between asphalt-, aggregate-, and fiber-related variables and the resulting splitting strength. Beyond predictive accuracy, the SHAP analysis further improves the engineering usefulness of the framework by identifying the dominant variables and clarifying their different influence patterns, including negative, positive, non-monotonic, and category-dependent effects. Therefore, the main value of the present work lies not only in accurate ST prediction, but also in providing interpretable parameter-level guidance for mixture design and offering a practical basis for future digital design tools for asphalt concrete.

5.2. Challenges and Limitations

Despite the promising results, several limitations remain. Since the database was assembled from different published studies, variations in material sources, mix design details, specimen preparation, testing conditions, and reporting practices are unavoidable, which may affect the consistency and transferability of the developed models. In addition, the distribution of fiber types is not fully balanced, and this may reduce the reliability of the category-specific patterns identified for FT. It should therefore be emphasized that the SHAP results for FT reflect relative model contributions under the present dataset rather than definitive judgments on the intrinsic effectiveness of each fiber type. Moreover, although SHAP provides useful interpretability, it describes model-learned associations rather than causal mechanisms, meaning that the reported thresholds and favorable intervals should be treated as empirical guidance. Finally, the present framework is limited to ST prediction and has not yet been extended to long-term service behavior, environmental coupling effects, or full engineering life-cycle evaluation.

5.3. Opportunities for Future Research

Future studies should further strengthen the data basis and expand the application range of the current framework. An important direction for future work is to establish a larger and more standardized asphalt concrete database with more balanced fiber categories, more complete reporting of material properties, and more consistent testing protocols, so that model training and interpretation can become more robust and transferable. On this basis, future studies can extend the framework from single-property prediction to multi-objective design by jointly considering strength, cracking resistance, rutting resistance, durability, and sustainability indicators, thereby supporting more comprehensive optimization of asphalt mixtures. In parallel, combining SHAP with stronger causal analysis, uncertainty quantification, and external validation from laboratory or field data would make the design recommendations more credible and engineering-oriented. Finally, the current GUI can be further upgraded into a continuously updated intelligent platform in which new data are periodically incorporated, models are retrained, and users can obtain not only ST predictions but also dynamic recommendation ranges for mixture parameters under different engineering scenarios.

6. Conclusion

This study established an interpretable machine-learning framework to predict the splitting strength (ST) of asphalt concrete using a literature-based database with fourteen input variables associated with asphalt properties, aggregate gradation, and fiber characteristics. Six machine learning models were established and compared, and the framework further integrated hyperparameter optimization, SHAP-based interpretation, and a GUI platform to improve both predictive capability and engineering usability. Based on the findings of this study, the following conclusions can be drawn.

- All six machine learning models demonstrated satisfactory capability for ST prediction, confirming that ML is effective in capturing the nonlinear relationships between mixture design variables and splitting strength. Among them, TabPFN achieved the best overall predictive performance on the testing set, with the lowest RMSE of 0.28, the highest R^2 of 0.88, and the highest composite score of 0.91. SVR ranked second overall, while XGBoost, RF, and LightGBM showed moderate but still acceptable predictive performance. These results indicate that TabPFN is the most suitable model for ST prediction in the present dataset.
- The SHAP analysis showed that the prediction of ST is mainly governed by a limited number of dominant variables. Based on the average feature contributions across the six models, Ag9.5, FT, Ag4.75, AC, and Du were identified as high-impact variables, while Pe, Ag2.36, SP, and AV were classified as medium-impact variables. Together, these nine variables accounted for 92.0% of the total average SHAP contribution, whereas FL, FC, TS, VFA, and VMA had relatively minor influence. In addition, the SHAP force-plot analysis for a representative sample showed that TabPFN provided the closest prediction to the actual ST value, further confirming its strong local interpretability and predictive reliability.
- The SHAP dependence analysis further revealed that the dominant variables exhibit different influence patterns on ST, including overall negative correlations, positive correlations, non-monotonic effects, and category-dependent effects. Specifically, Ag9.5, Ag4.75, AC, and AV showed overall negative correlations with ST; Du showed an overall positive correlation; and Pe, Ag2.36, and SP exhibited non-monotonic relationships. For the categorical feature FT, polyester fiber showed a comparatively stronger positive contribution to ST than the other fiber types in the present dataset under the current data conditions. Based on the dependence analysis, the favorable ranges for improving ST were identified as $\text{Ag9.5} < 66.8\%$, $\text{Ag4.75} < 45.0\%$, $\text{AC} < 5.4 \text{ wt.}\%$, $\text{AV} < 3.6\%$, $\text{Du} > 134.7 \text{ cm}$, $\text{Pe} < 60 \text{ or } > 86.7 (0.1 \text{ mm})$, $37.0\% < \text{Ag2.36} < 51.5\%$, and $\text{SP} < 45.6 \text{ }^\circ\text{C} \text{ or } > 55.6 \text{ }^\circ\text{C}$.

- Beyond model construction and interpretation, this study also established a GUI platform to enhance the accessibility and applicability of the developed framework. By integrating prediction and SHAP-based explanation into a user-oriented interface, the platform provides a practical tool for estimating ST and understanding the role of individual design variables. Overall, the proposed framework offers not only accurate prediction of splitting strength, but also interpretable guidance for mixture design, thereby demonstrating the potential of explainable artificial intelligence in the intelligent design and optimization of asphalt concrete.

Author Contributions: Conceptualization: **Xiao Tan, Jianglei Xing**; methodology: **Xiao Tan, Jianglei Xing**; software: **Jianglei Xing**; validation: **Pengwei Guo, Dongzhao Jin, Yuhuan Wang**; formal analysis: **Jianglei Xing**; investigation: **Jianglei Xing, Xiao Tan**; resources: **Xiao Tan, Yihao Li**; data curation: **Jianglei Xing, Yihao Li**; writing—original draft preparation: **Xiao Tan, Jianglei Xing**; writing—review and editing: **Xiao Tan, Dongzhao Jin, Pengwei Guo**; visualization: **Jianglei Xing**; supervision: **Xiao Tan**; project administration: **Xiao Tan, Dongzhao Jin**; funding acquisition, **Xiao Tan, Huiya Niu**. All authors have read and agreed to the published version of the manuscript.

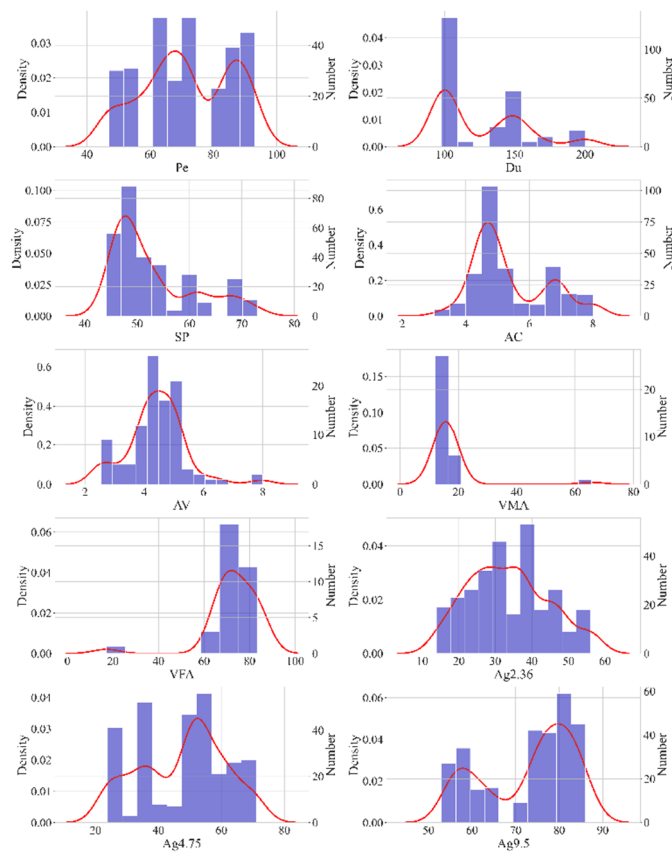
Funding: This research was funded by the National Natural Science Foundation of China (Grant No: 52508343), Basic Research Program of Jiangsu (Grant No: BK20251486), Fundamental Research Funds for the Central Universities (Grant No: B250201004), and Shanghai Sailing Program (Program ID: 23YF1437700).

Ethical Approval: The results/data/figures in this manuscript have not been published elsewhere, nor are they under consideration by another publisher.

Data Availability Statement: Data will be made available on request.

Declaration of Competing Interest: The authors declare no competing financial interests or personal affiliations that could have influenced the results presented in this manuscript.

Appendix A. Data description



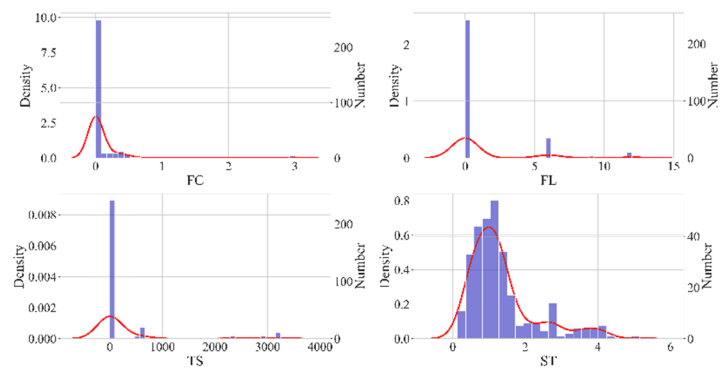


Figure A1. Distribution of fourteen input features and one output feature of the dataset.

Appendix B. Model Configuration and Performance Evaluation

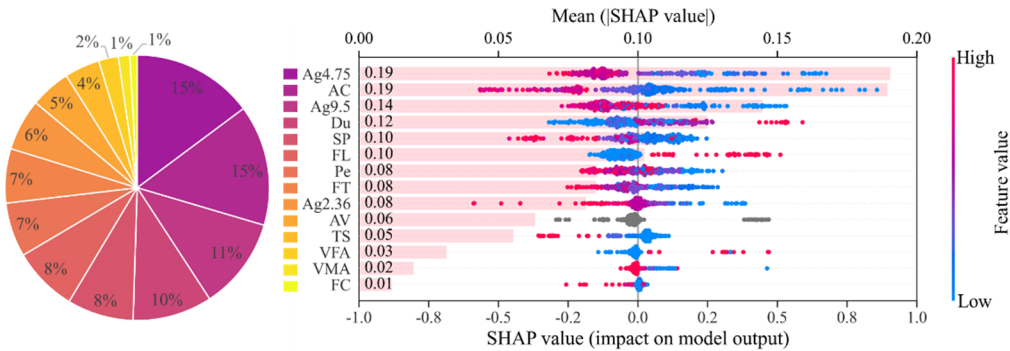
Table B1. Hyperparameter combinations adopted for each model.

Models	Hyperparameters
TabPFN	default hyperparameters
ANN	n= 2 hidden_layer_sizes = (128, 64) learning_rate_init = 0.01 batch_size = 151 activation = relu solver = adam validation_fraction = 0.1 early_stopping=True
SVR	C = 5.48 gamma = 0.14 epsilon = 0.24 kernel = rbf
RF	n_estimators = 370 max_depth = 13 min_samples_split = 2 min_samples_leaf = 1 max_features = log2 bootstrap = True
XGBoost	n_estimators = 207 learning_rate = 0.09 max_depth = 6 objective = reg:squarederror tree_method = hist
LightGBM	n_estimators = 477 learning_rate = 0.05 max_depth = 7 min_child_samples= 12 reg_alpha = 0.07 reg_lambda = 0.04 num_leaves = 57

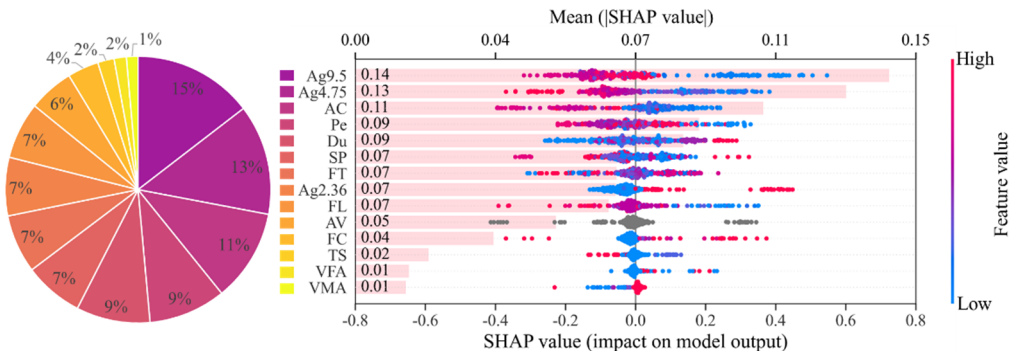
Table B2. Performance metrics of the models for ST prediction on the testing set.

Model	Metrics				
	RMSE	MAE	MAPE	MAD	R ²
TabPFN	0.28	0.21	18.01	0.14	0.88
ANN	0.37	0.27	24.87	0.16	0.78
SVR	0.31	0.23	19.81	0.17	0.84
RF	0.36	0.24	21.69	0.14	0.80
XGBoost	0.37	0.24	21.11	0.13	0.79
LightGBM	0.36	0.25	22.21	0.14	0.80

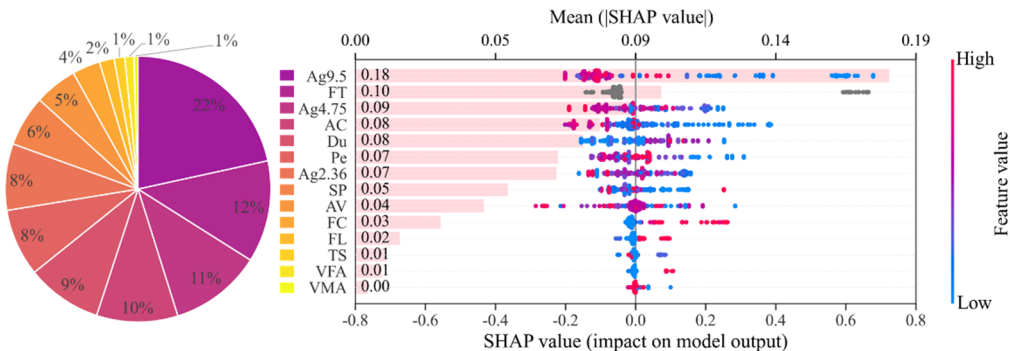
Appendix C. SHAP Analysis Demonstration



(a) ANN



(b) SVR



(c) Random forest

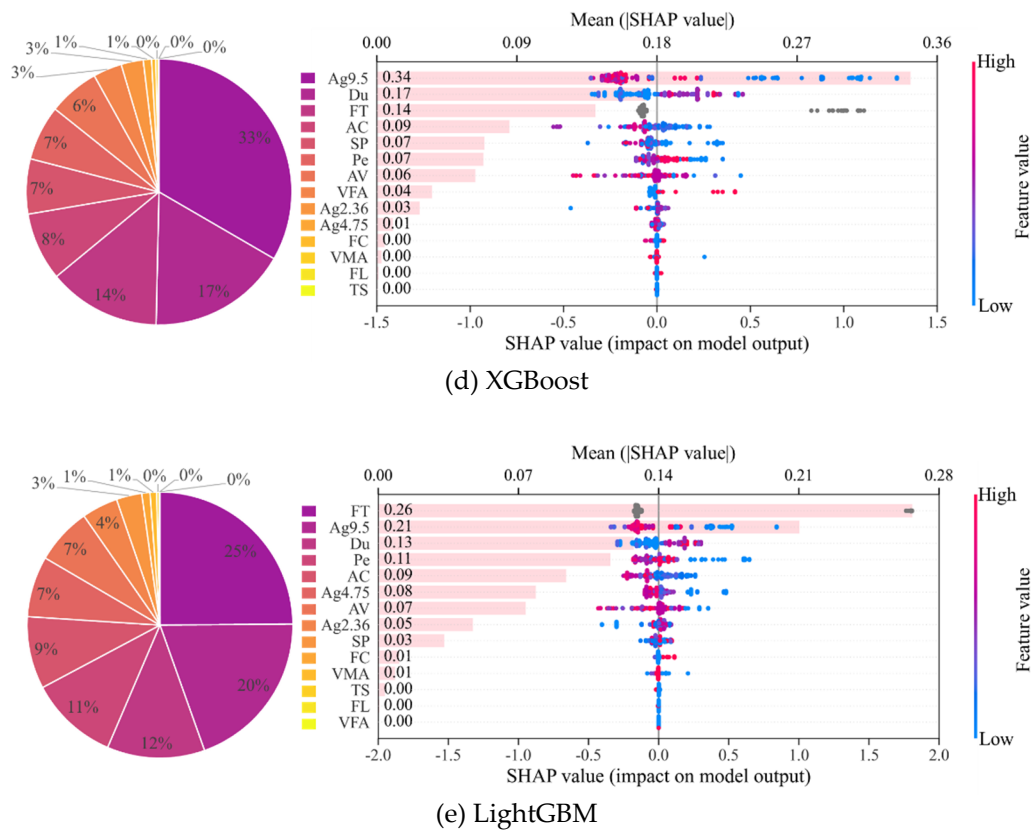


Figure C1. Summary plot of SHAP-based interpretation of the outputs from the remaining five machine learning models: (a) ANN; (b) SVR; (c) RF; (d) XGBoost; (e) LightGBM.

Abbreviation List

Abbreviation	Full name
AC	Asphalt content
Ag2.36	2.36 mm aggregate passing rate
Ag4.75	4.75 mm aggregate passing rate
Ag9.5	9.5 mm aggregate passing rate
AV	Air voids
Du	Ductility
FC	Fiber content
FL	Fiber length
FT	Fiber type
Pe	Penetration
SP	Softening point
ST	Splitting strength
TS	Tensile strength
VFA	Voids filled with asphalt
VMA	Voids in mineral aggregate

References

1. Wang, F., Hoff, I., Yang, F., Wu, S., Xie, J., Li, N., and Zhang, L. (2021). Comparative assessments for environmental impacts from three advanced asphalt pavement construction cases. *Journal of Cleaner Production*, 297, p.126659. <https://doi.org/10.1016/j.jclepro.2021.126659>
2. AlKheder, S., AlKandari, D., and AlYatama, S. (2022). Sustainable assessment criteria for airport runway material selection: A fuzzy analytical hierarchy approach. *Engineering, Construction and Architectural Management*, 29(8), pp.3091-3113. <https://doi.org/10.1108/ECAM-01-2021-0052>
3. James, W., and Thompson, M. K. (2021). Contaminants from four new pervious and impervious pavements in a parking-lot. *Advances in Modeling the Management of Stormwater Impacts*, pp. 207-222. CRC Press. <https://doi.org/10.1201/9781003208945>
4. Ning, Z., Sun, Z., Liu, Y., Dong, J., Meng, X., Wang, Q., and Wei, Y. (2024). Evaluating the impervious performance of hydraulic asphalt concrete in embankment dams: A study of crack evolution at different temperatures. *Construction and Building Materials*, 440, p.137247. <https://doi.org/10.1016/j.conbuildmat.2024.137247>
5. Bieliatynskiy, A., Yang, S., Pershakov, V., Shao, M., and Ta, M. (2022). Features of the hot recycling method used to repair asphalt concrete pavements. *Materials Science-Poland*, 40(2), pp.181-195. <https://doi.org/10.2478/msp-2022-0021>
6. Yao, H., Wang, Y., Ma, P., Li, X., and You, Z. (2023). A literature review: asphalt pavement repair technologies and materials. *Proceedings of the Institution of Civil Engineers-Engineering Sustainability*, 177(5), pp. 259-273. <https://doi.org/10.1680/jensu.22.10000>
7. Rivera-Pérez, J., Talebpour, A., and Al-Qadi, I. L. (2023). Prediction of asphalt concrete flexibility index and rut depth utilising deep learning and Monte Carlo Dropout simulation. *International Journal of Pavement Engineering*, 24(1), p.2253964. <https://doi.org/10.1080/10298436.2023.2253964>
8. Ma, R., Li, Y., Cheng, P., Chen, X., and Cheng, A. (2024). Low-temperature cracking and improvement methods for asphalt pavement in cold regions: A review. *Buildings*, 14(12), p.3802. <https://doi.org/10.3390/buildings14123802>
9. Al-Atroush, M. E. (2022). Structural behavior of the geothermo-electrical asphalt pavement: A critical review concerning climate change. *Heliyon*, 8(12). <https://doi.org/10.1016/j.heliyon.2022.e12107>
10. Arabzadeh, A., Ceylan, H., Kim, S., Gopalakrishnan, K., and Sassani, A. (2016). Superhydrophobic coatings on asphalt concrete surfaces: Toward smart solutions for winter pavement maintenance. *Transportation Research Record*, 2551(1), pp.10-17. <https://doi.org/10.3141/2551-02>
11. Dias, J. F., Picado-Santos, L. G., and Capitão, S. D. (2014). Mechanical performance of dry process fine crumb rubber asphalt mixtures placed on the Portuguese road network. *Construction and Building Materials*, 73, pp.247-254. <https://doi.org/10.1016/j.conbuildmat.2014.09.110>
12. Zaumanis, M., Mallick, R. B., and Frank, R. (2016). 100% hot mix asphalt recycling: Challenges and benefits. *Transportation Research Procedia*, 14, pp.3493-3502. <https://doi.org/10.1016/j.trpro.2016.05.315>
13. Liu, Q. T., and Wu, S. P. (2014). Effects of steel wool distribution on properties of porous asphalt concrete. *Key Engineering Materials*, 599, pp.150-154. <https://doi.org/10.4028/www.scientific.net/KEM.599.150>
14. García, A., Norambuena-Contreras, J., Bueno, M., and Partl, M. N. (2014). Influence of steel wool fibers on the mechanical, thermal, and healing properties of dense asphalt concrete. *Journal of Testing and Evaluation*, 42(5), pp.1107-1118. <https://doi.org/10.1520/JTE20130197>
15. Pasandín, A. R., and Pérez, I. (2015). Overview of bituminous mixtures made with recycled concrete aggregates. *Construction and Building Materials*, 74, pp.151-161. <https://doi.org/10.1016/j.conbuildmat.2014.10.035>
16. Wang, L., Zhang, J., Song, M., Tian, B., Li, K., Liang, Y., Han, J., and Wu, Z. (2017). A shell-crosslinked polymeric micelle system for pH/redox dual stimuli-triggered DOX on-demand release and enhanced antitumor activity. *Colloids and Surfaces B: Biointerfaces*, 152, pp.1-11. <https://doi.org/10.1016/j.colsurfb.2016.12.032>
17. Hejazi, S. M., Abtahi, S. M., Sheikhzadeh, M., and Semnani, D. (2008). Introducing two simple models for predicting fiber-reinforced asphalt concrete behavior during longitudinal loads. *Journal of Applied Polymer Science*, 109(5), pp.2872-2881. <https://doi.org/10.1002/app.28349>

18. Karanam, G. D., and Underwood, B. S. (2024). Mechanical characterization and performance prediction of fiber-modified asphalt mixes. *International Journal of Pavement Research and Technology*, pp.1-19. <https://doi.org/10.1007/s42947-024-00445-9>
19. Khan, A. R., Fareed, A., Ali, A., Pandya, H., Ali, A., and Mehta, Y. (2023). Seasonal performance prediction comparison of unreinforced and fiber-reinforced Asphalt mixtures for airfield pavements. In *Airfield and Highway Pavements 2023* (pp. 374-384). <https://doi.org/10.1061/9780784484890.034>
20. Tan, X., Xing, J., Wang, Y., Qiu, H., Mahjoubi, S., and Guo, P. (2026). Explainable machine learning for predicting compressive strength of rubberized concrete: SHAP interpretation, lifecycle assessment, and design recommendations. *Journal of Cleaner Production*, 538, p.147338. <https://doi.org/10.1016/j.jclepro.2025.147338>
21. Tan, X., Xing, J., Mahjoubi, S., Guo, P., Wei, Z., Wang, Y., Ren, J., Ai, L., and Bao, Y. (2026). Explainable machine learning and life cycle assessment for sustainable design of fiber-reinforced asphalt concrete. *Journal of Cleaner Production*, 547, p.147759. <https://doi.org/10.1016/j.jclepro.2026.147759>
22. Upadhyaya, A., Thakur, M. S., Al Ansari, M. S., Malik, M. A., Alahmadi, A. A., Alwetaishi, M., and Alzaed, A. N. (2022). Marshall stability prediction with glass and carbon fiber modified asphalt mix using machine learning techniques. *Materials*, 15(24), p.8944. <https://doi.org/10.3390/ma15248944>
23. Upadhyaya, A., Thakur, M. S., and Sihag, P. (2024). Predicting Marshall stability of carbon fiber-reinforced asphalt concrete using machine learning techniques. *International Journal of Pavement Research and Technology*, 17(1), pp.102-122. <https://doi.org/10.1007/s42947-022-00223-5>
24. Phung, B. N., Le, T. H., Nguyen, M. K., Nguyen, T. A., and Ly, H. B. (2023). Practical numerical tool for marshall stability prediction based on machine learning: an application for asphalt concrete containing basalt fiber. *Journal of Science and Transport Technology*, pp.26-43. <https://doi.org/10.58845/jstt.utt.2023.en.3.3.26-43>
25. Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3), pp.31-57. <https://doi.org/10.1145/3233231>
26. Wang, Z. (2020). Evaluation method of adhesion between aggregate and bitumen and long-term moisture susceptibility evaluation of hydraulic asphalt concrete. PhD Thesis, Chinese Research Institute of Water Resources and Hydropower Engineering. <https://doi.org/10.27646/d.cnki.gsldy.2020.000035>
27. He, J., Zhu, X., Yang, H., and Wang, W. (2014). Experimental research on the gravel aggregate water stability performance of asphalt concrete core wall. *China Rural Water and Hydropower*, (11), pp.109-112. <https://irrigate.whu.edu.cn/CN/Y2014/V0/I11/109>
28. Lin, P. (2025). Research on Mixture Proportion Design and Road Performance of Asphalt Concrete in Hot Area. *Engineering and Technological Research*, 10 (03), pp.125-127. <https://doi.org/10.19537/j.cnki.2096-2789.2025.03.041>
29. Tu, Y., Chen, G., Cheng, Z., and Cheng, S. (2022). Effect of nano-SiO₂ on properties of recycled aggregate asphalt mixture. *Materials Reports*, 36 (S1), pp.220-224. <http://www.mater-rep.com/CN/Y2022/V36/IZ1/22030139>
30. Lin, Z., and Wang, F. (2020). The Crack Resistance Experiment of Composite Fiber Asphalt Concrete. *Journal of Shenyang University (Natural Science)*, 36 (03), pp.500-506. <https://doi.org/10.11717/j.issn:2095-1922.2020.03.14>
31. Qin, L. (2019). Study on the Effects of Aging on the Volume and Water Stability Properties of Steel Slag and Its Asphalt Concrete. *Journal of China & Foreign Highway*, 39 (06), pp.264-270. <https://doi.org/10.14048/j.issn.1671-2579.2019.06.056>
32. Kong, Z., Zhang, Y., and Zhang, A. (2014). The influence of granite powder filler on the water stability of asphalt concrete. *Journal of China & Foreign Highway*, 34 (05), pp.287-290. <https://doi.org/10.14048/j.issn.1671-2579.2014.05.065>
33. Wang, A., Jiao, C., Han, C., and Kaung, Q. (2025). Study on the effect of salt corrosion on the performance of cold mixpermeable asphalt concrete. *New Building Materials*, 52 (07), pp.67-70. <https://www.cnki.com.cn/Article/CJFDTOTAL-XXJZ202507013.htm>

34. Fan, T. (2020). Study on performance of calcium sulfate whisker - polyester fiber compound modified asphalt and asphalt mixture. PhD Thesis, Chang'an University. <https://doi.org/10.26976/d.cnki.gchau.2020.000034>

35. Ma, Z. (2024). Design and application research of electrically heated ice and snow melting paving structure based on conductive rubber composite material. PhD Thesis, Jilin University. <https://doi.org/10.27162/d.cnki.gjlin.2024.000312>

36. Shu, J., Xv, K., Liu, S., Wan, P., Liu, Q., and Wu, S. Effects of calcium alginate/Fe3O4 composite self-healing capsules on road performance of asphalt concrete. *Journal of Wuhan University of Technology (Transportation Science & Engineering)*, pp.1-17. <https://link.cnki.net/urlid/42.1824.U.20250325.1426.0>

37. Zhu, C. (2018). Research on road performance and mechanical properties of diatomite-basalt fiber compound modified asphalt mixture. PhD Thesis, Jilin University. <https://cdmd.cnki.com.cn/Article/CDMD-10183-1018213539.htm>

38. Ge, Q., Wu, H., and Wang, G. (2019). The mix design of graphite steel and carbon fiber modified conductive asphalt mixture. *Technology & Economy in Areas of Communications*, 21 (06), pp.51-54. <https://doi.org/10.19348/j.cnki.issn1008-5696.2019.06.009>

39. Zhang, Z., Chai, Z., and Tao, Z. (2019). Research on the road performance of Dense-Mix Concrete Waste Ash Asphalt Mixture. *Journal of Highway and Transportation Research and Development*, 15 (08), pp.72-75. <https://www.cnki.com.cn/Article/CJFDTOTAL-GLJJ201908025.htm>

40. Zhu, T. (2017). Structural analysis and design for recycled asphalt pavement based on the performance characteristics of recycled asphalt mixture. PhD Thesis, Southeast University. <https://cdmd.cnki.com.cn/Article/CDMD-10286-1017171128.htm>

41. Zhao, H. (2016). Microwave absorbing properties and road performance of asphalt mixture doped with natural magnetite. PhD Thesis, Chang'an University. <https://cdmd.cnki.com.cn/Article/CDMD-10710-1017804015.htm>

42. Tang, J. (2013). Experimental research on composition and performance of fiber reinforced asphalt mixture. PhD Thesis, Zhengzhou University. <https://cdmd.cnki.com.cn/Article/CDMD-10459-1013257892.htm>

43. Liang, X. (2011). The Research of oil-stone interface adhesive based on modified surface of asphalt and stone. PhD Thesis, Jilin University. <https://cdmd.cnki.com.cn/Article/CDMD-10183-1012257811.htm>

44. Gao, C. (2012). Microcosmic analysis and performance research of basalt fiber asphalt concrete. PhD Thesis, Jilin University. <https://cdmd.cnki.com.cn/Article/CDMD-10183-1012365736.htm>

45. Shen, F. (2012). Research on composite steel bridge deck pavement of cement-emulsifying asphalt and waterborne epoxy. PhD Thesis, Wuhan University of Technology. <https://cdmd.cnki.com.cn/Article/CDMD-10497-1012442305.htm>

46. Luo, S. (2012). Pavement disease environment and dynamic coupling analysis of asphalt pavement in high temperature and rainy area. PhD Thesis, Central South University. <https://cdmd.cnki.com.cn/Article/CDMD-10533-1012475004.htm>

47. Chen, M. (2012). Research on snow melting and solar energy collection for thermal conductive asphalt pavement. PhD Thesis, Wuhan University of Technology. <https://cdmd.cnki.com.cn/Article/CDMD-10497-1012442416.htm>

48. Wei, G. (2020). Study on Crack Resistance of fiberglass-polyester paving mat in asphalt pavement. PhD Thesis, Chongqing Jiaotong University. <https://doi.org/10.27671/d.cnki.gcjtc.2020.001066>

49. Li, C. (2020). Study on the self-healing performance and mechanism of asphalt concrete under microwave radiation. PhD Thesis, Wuhan University of Technology. <https://doi.org/10.27381/d.cnki.gwlg.2020.000075>

50. Zhang, Q. (2020). Study on water damage mechanism of asphalt mixture in multi-factor environment of the south coast. PhD Thesis, Zhejiang University. <https://doi.org/10.27461/d.cnki.gzjdx.2020.002487>

51. Liu, D. (2018). Study on road performance and structural characteristics of asphalt stabilized macadam with high modulus. PhD Thesis, Wuhan University of Technology. <https://doi.org/10.27381/d.cnki.gwlg.2018.000025>

52. Liu, W. (2018). Study on enhancement mechanism and healing evaluation of microwave absorption of asphalt mixture. PhD Thesis, Southeast University. <https://cdmd.cnki.com.cn/Article/CDMD-10286-1019650164.htm>
53. Xv, C. (2010). Research on performance of glass fiber-diatomite composite modified asphalt concrete. PhD Thesis, Jilin University. <https://cdmd.cnki.com.cn/Article/CDMD-10183-2011014225.htm>
54. Xiao, J. (2011). Study on structure formation mechanism and features of cement emulsified asphalt mixture. PhD Thesis, Chang'an University. <https://cdmd.cnki.com.cn/Article/CDMD-10710-1016327595.htm>
55. Ai, C. (2008). Characteristics and design methods of asphalt pavement in plateau-cold region PhD Thesis, Southwest Jiaotong University. <https://cdmd.cnki.com.cn/Article/CDMD-10613-2008177745.htm>
56. Zhao, P., Li, M., He, W., Liu, Z., and Gao, Y. (2018). Application research of reinforced PAN fiber in color asphalt bus pavement. *Construction Technology*, 47 (20), pp.19-21+25. <https://www.cnki.com.cn/Article/CJFDTOTAL-SGJS201820009.htm>
57. Phung, B. N., Le, T. H., Nguyen, T. A., Hoang, H. G. T., and Ly, H. B. (2023). Novel approaches to predict the Marshall parameters of basalt fiber asphalt concrete. *Construction and Building Materials*, 400, p.132847. <https://doi.org/10.1016/j.conbuildmat.2023.132847>
58. Benesty, J., Chen, J., Huang, Y., and Cohen, I. (2009). Pearson correlation coefficient. *Noise Reduction in Speech Processing*, pp.1-4. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-00296-0_5
59. Guo, P., Meng, W., and Bao, Y. (2024). Knowledge-guided data-driven design of ultra-high-performance geopolymer (UHPG). *Cement and Concrete Composites*, 153, p.105723. <https://doi.org/10.1016/j.cemconcomp.2024.105723>
60. Zhou, C., Wang, W., and Zheng, Y. (2024). Data-driven shear capacity analysis of headed stud in steel-UHPC composite structures. *Engineering Structures*, 321, p.118946. <https://doi.org/10.1016/j.engstruct.2024.118946>
61. Hancock, J. T., and Khoshgoftaar, T. M. (2020). Survey on categorical data for neural networks. *Journal of Big Data*, 7(1), p.28. <https://doi.org/10.1186/s40537-020-00305-w>
62. Singh, D., and Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97, p.105524. <https://doi.org/10.1016/j.asoc.2019.105524>
63. Jain, A., Nandakumar, K., and Ross, A. (2005). Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12), pp.2270-2285. <https://doi.org/10.1016/j.patcog.2005.01.012>
64. Ejaz, U., Khan, S.M., Jehangir, S., Ahmad, Z., Abdullah, A., Iqbal, M., Khalid, N., Nazir, A. and Svenning, J.C. (2024). Monitoring the Industrial waste polluted stream-Integrated analytics and machine learning for water quality index assessment. *Journal of Cleaner Production*, 450, p.141877. <https://doi.org/10.1016/j.jclepro.2024.141877>
65. Hollmann, N., Müller, S., Purucker, L., Krishnakumar, A., Körfer, M., Hoo, S.B., Schirrmeister, R.T., and Hutter, F., Accurate predictions on small data with a tabular foundation model. *Nature*, 2025. 637(8045): pp.319-326. <https://doi.org/10.1038/s41586-024-08328-6>
66. Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, pp.273-297. <https://doi.org/10.1007/BF00994018>
67. Breiman, L. (2001). Random forests. *Machine learning*, 45(1), pp.5-32. <https://doi.org/10.1023/A:1010933404324>
68. Chen, T., and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.785-794. <https://doi.org/10.1145/2939672.2939785>
69. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q and Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp.3149-3157. <https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf>
70. Kaveh, A. (2024). Applications of Artificial Neural Networks and Machine Learning in Civil Engineering. Springer. <https://doi.org/10.1007/978-3-031-66051-1>

71. Hollmann, N., Müller, S., Eggensperger, K., and Hutter, F. (2022). TabPFN: A transformer that solves small tabular classification problems in a second. arXiv preprint arXiv:2207.01848. <https://doi.org/10.48550/arXiv.2207.01848>
72. PriorLabs. (2026). Tabular Foundation Models. <https://priorlabs.ai/tabpfn>
73. Sun, Y., Ding, S., Zhang, Z., and Jia, W. (2021). An improved grid search algorithm to optimize SVR for prediction. *Soft Computing*, 25, pp.5633-5644. <https://doi.org/10.1007/s00500-020-05560-w>
74. Bergstra, J., and Bengio, Y. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, 13(1), pp.281-305. <https://dl.acm.org/doi/10.5555/2188385.2188395>
75. Yang, L., and Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415, pp.295-316. <https://doi.org/10.1016/j.neucom.2020.07.061>
76. Bergstra, J., and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(2), pp.281-305. <https://dl.acm.org/doi/10.5555/2188385.2188395>
77. Deb, K., Agrawal, S., Pratap, A., and Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In *International Conference on Parallel Problem Solving from Nature*, pp.849-858. Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45356-3_83
78. Lundberg, S. M. and Lee, S. I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp.4768-4777. <https://dl.acm.org/doi/10.5555/3295222.3295230>
79. Khan, M., Lao, J., and Dai, J. G. (2022). Comparative study of advanced computational techniques for estimating the compressive strength of UHPC. *Journal of Asian Concrete Federation*, 8(1), pp.51-68. <https://doi.org/10.18702/acf.2022.6.8.1.51>
80. Shi, C., Qian, G., Yu, H., Zhu, X., Yuan, M., Dai, W., Ge, J., and Zheng, X. (2024). Research on the evolution of aggregate skeleton characteristics of asphalt mixture under uniaxial compression loading. *Construction and Building Materials*, 413, p.134769. <https://doi.org/10.1016/j.conbuildmat.2023.134769>
81. Lin, P., Liu, X., Ren, S., Xu, J., Li, Y., and Li, M. (2023). Effects of bitumen thickness on the aging behavior of high-content polymer-modified asphalt mixture. *Polymers*, 15(10), p.2325. <https://doi.org/10.3390/polym15102325>
82. Zhang, Y., Luo, X., Luo, R., and Lytton, R. L. (2014). Crack initiation in asphalt mixtures under external compressive loads. *Construction and Building Materials*, 72, pp.94-103. <https://doi.org/10.1016/j.conbuildmat.2014.09.009>
83. Guo, M., Yao, X., and Du, X. (2023). Low temperature cracking behavior of asphalt binders and mixtures: A review. *Journal of Road Engineering*, 3(4), pp.350-369. <https://doi.org/10.1016/j.jreng.2023.11.001>
84. Khair, A., Wang, L., Li, H., Han, Y., Lin, Z., Sun, Y., and Zhang, H. (2026). Comparative performance evaluation of asphalt binder modified with high-content pretreated crumb rubber and various additives. *Journal of Road Engineering*. <https://doi.org/10.1016/j.jreng.2025.07.003>
85. Malluru, S., Islam, S. M. I., Saidi, A., Baditha, A. K., Chiu, G., and Mehta, Y. (2025). A state-of-the-practice review on the challenges of asphalt binder and a roadmap towards sustainable alternatives—A call to action. *Materials*, 18(10), p.2312. <https://doi.org/10.3390/ma18102312>
86. Wang, X., Gu, X., Jiang, J., and Deng, H. (2018). Experimental analysis of skeleton strength of porous asphalt mixtures. *Construction and Building Materials*, 171, pp.13-21. <https://doi.org/10.1016/j.conbuildmat.2018.03.116>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.