buildings

MDPI

Article

An Interpretable Pretrained Tabular Modeling Framework for Predicting IRI Across Multiple Pavement Structural Configurations

Liang Qin [1], Tong Liu [1,*], Qianhui Sun [1] and Mingxin Tang [2]

[1] School of Management, Shenyang Jianzhu University, Shenyang 110168, China; qinliang@stu.sjzu.edu.cn (L.Q.)
[2] School of Building and Environment, The Hong Kong Polytechnic University, Hong Kong 999077, China
* Correspondence: liu202277@163.com; Tel.: +86-159-4024-3278

Abstract

With increasing traffic loads and increasingly complex climate conditions, accurate prediction of the International Roughness Index (IRI) of asphalt pavements is crucial for developing effective maintenance plans. However, traditional regression models have limitations in capturing the coupled effects of traffic, structure, and environmental factors. To overcome this limitation, this study constructed a dataset containing 10,836 samples based on the Long-Term Pavement Performance (LTPP) database, integrating traffic load, pavement structure parameters, and climate variables. The variance inflation factor (VIF) and correlation analysis were used to validate the effectiveness of feature selection. We trained nine machine learning models and optimized the hyperparameters using a Bayesian optimization method with five-fold cross-validation to ensure good generalization ability. Results show that the TabPFN model, based on prior information, achieved the best overall performance with a coefficient of determination R^2 = 0.9474 and a low prediction error (RMSE = 0.138) on the test set. Paired *t*-tests based on cross-validation further confirmed that TabPFN's predictive performance is statistically superior to the baseline model. SHAP and generalized additive model (GAM) analyses indicate that traffic load is the main driver of IRI growth, while structural layer thickness, within a certain range, can mitigate pavement roughness. Climatic factors have indirect long-term effects through cumulative environmental exposure. Although the main drivers differ slightly among different pavement structures, traffic load consistently plays a dominant role. To enhance the model's practical applicability, we also developed a user-friendly graphical interface (GUI) for fast and accurate IRI prediction.

Keywords: International Roughness Index; asphalt concrete; machine learning; LTPP data; interpretable prediction

Check for updates

Academic Editor: Erwin Oh

Received: 12 February 2026
Revised: 19 March 2026
Accepted: 25 March 2026
Published: 29 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Against the backdrop of increasingly severe global climate change and continuously growing traffic demand, transportation infrastructure is facing increasingly complex operating environments characterized by the combined effects of environmental conditions and mechanical loads. Therefore, the long-term performance evolution of pavement systems has become a core research topic in the fields of infrastructure and pavement structure engineering [1,2]. The combination of long-term climate fluctuations and continuously increasing traffic loads leads to significant nonlinear degradation patterns and cumulative damage effects in pavement structures, thus imposing more stringent requirements on

maintenance and asset management strategies [3,4]. However, existing pavement performance assessment methods still mainly rely on single factors or weakly coupled indicators, such as service life, climate conditions, or traffic intensity. These methods have inherent limitations in characterizing the real degradation mechanisms resulting from the interaction of multiple factors [5,6]. This limitation is particularly pronounced under harsh operating conditions involving extreme temperatures, frequent freeze–thaw cycles, and heavy traffic loads. In such situations, environmental factors and load effects can have a synergistic effect, significantly accelerating structural damage and surface roughness deterioration. Therefore, the applicability of traditional methods under complex operating conditions is greatly limited [7–9].

Accurate prediction of pavement performance indicators, particularly the International Roughness Index (IRI), is a critical prerequisite for improving the scientific basis of pavement design and maintenance decision-making [10,11]. Traditional pavement performance prediction approaches are largely derived from small-sample empirical statistical analyses, typically employing linear or nonlinear regression models to describe the effects of traffic loading, structural parameters, and environmental factors on in-service performance [12,13]. However, as traffic conditions, climatic environments, and pavement structural configurations become increasingly complex, the associated influencing factors exhibit high dimensionality, strong nonlinearity, and long-term service dependence. Under such multi-variable coupled conditions, conventional models struggle to accurately represent pavement performance evolution, thereby limiting their applicability and predictive accuracy in complex structural systems [14,15]. In recent years, machine learning (ML) techniques have been widely introduced into the field of pavement and infrastructure engineering for performance prediction and assessment [16,17]. Existing studies have demonstrated that ML methods offer significant advantages in handling multi-source, nonlinear, and high-dimensional data, making them effective tools for predicting key performance indicators of pavement materials and structures [18,19]. These include the mechanical properties of asphalt mixtures [20–22], rutting depth [23,24], durability [25,26], and surface roughness or unevenness [27–29]. Compared with traditional empirical and mechanistic models [30,31], ML approaches can uncover complex relationships among variables directly from experimental or monitoring data, with fewer prior assumptions, thereby improving both predictive accuracy and modeling efficiency [32].

Nevertheless, the predictive performance of machine learning (ML) models is highly sensitive to factors such as sample size, hyperparameter configuration, and the underlying model architecture [33,34]. This issue is particularly pronounced in road and pavement materials engineering, where experimental datasets are often limited, especially for emerging pavement materials and innovative structural configurations. The restricted sample size and narrow feature coverage substantially constrain the prediction accuracy and generalization capability of ML models [35,36]. To better exploit the advantages of machine learning in modeling complex nonlinear systems, some researchers have extended ML-based prediction frameworks by integrating optimization algorithms for intelligent design of pavement materials and structural parameters [37,38]. In such approaches, material compositions, mechanical properties, and engineering constraints are embedded into ML surrogate models, which are then coupled with heuristic or intelligent optimization techniques, including particle swarm optimization (PSO) [39,40], genetic algorithms (GA) [41,42], and Bayesian optimization (BO) [43,44]. These combined methods enable multi-objective optimization of asphalt mixture proportions, pavement structural parameters, and design schemes, aiming to achieve improved performance, resource efficiency, and cost control simultaneously [45]. Within this "prediction–optimization" coupled framework, the development of ML surro-

gate models with high predictive accuracy and strong generalization capability constitutes the essential foundation for efficient and reliable optimization-based design [46].

On the other hand, the complexity of service environments and structural characteristics leads to significant nonlinearity and discreteness in the International Roughness Index (IRI) under different service stages and operating conditions, thus placing higher demands on the predictive capabilities of machine learning models [47,48]. Individual machine learning models often have inherent limitations in IRI prediction. Simple models are prone to underfitting and struggle to capture the potential interactions between traffic loads, structural parameters, material properties, and environmental factors; while more complex models may overfit and exhibit reduced generalization ability under unseen pavement conditions [49]. Furthermore, many existing studies focus primarily on comparing and selecting individual algorithms, largely neglecting the collaborative modeling of multi-source feature coupling and high-order nonlinear relationships [50]. To address these challenges, pre-trained tabular models have emerged in recent years as a novel paradigm for structured data analysis. Unlike traditional boosting or shallow neural networks, pre-trained models can learn transferable feature interaction priors from large-scale synthetic or real-world tabular datasets, thus better adapting to downstream tasks under limited supervision [51].

Representative frameworks include Transformer-based architectures such as TabTransformer and FT-Transformer, which introduce self-attention mechanisms to model contextual feature embeddings; and probabilistic meta-learning methods such as TabPFN, which approximate Bayesian posterior inference through large-scale prior data simulation [47]. These models are theoretically based on representation learning, attention-based feature interaction modeling, and Bayesian decision theory. Their training paradigm typically follows large-scale pre-training followed by task-specific inference without extensive hyperparameter tuning [48]. These characteristics give them an advantage in capturing complex nonlinear couplings and reducing the burden of manual optimization. However, pre-trained tabular models may still face limitations on highly heterogeneous engineering datasets, including sensitivity to distribution shifts and difficulty in extrapolating predictions beyond the learned prior space [49]. Despite these limitations, their ability to combine ensemble learning principles with probabilistic inference makes them particularly promising for small-sample, multi-factor road performance prediction. When combined with interpretability analysis such as SHAP and GAM, these models can not only improve prediction accuracy but also promote mechanism-oriented insights, thereby strengthening the theoretical basis and practical motivation for applying pre-trained table models in IRI prediction [50].

Numerous studies have applied machine learning techniques to predict pavement roughness using Long-Term Road Performance (LTPP) databases, including algorithms such as random forests, gradient boosting, support vector regression, and deep neural networks. However, due to differences in feature selection, data preprocessing, and model architecture, the reported prediction accuracy varies significantly. To better place this study within the existing literature, Table 1 summarizes representative LTPP-based IRI prediction studies, including their datasets, modeling methods, and reported performance metrics (e.g., R^2, RMSE). In most previous studies using traditional machine learning models to predict IRI, R^2 values typically ranged from 0.79 to 0.93. Despite these advances, many existing studies still focus on specific pavement structures or limited climatic conditions, making it difficult to generalize the results to diverse traffic, structural, and environmental scenarios.

Table 1. Research on pavement performance prediction.

Reference	Data Source	Sample	Model	Performance Index	R^2
[51]	LTPP	2111	GPR	IRI	0.89
[52]	LTPP	1725	ANN	IRI	0.75
[53]	LTPP	3238	LSTM-Attention	IRI	0.79
[54]	LTPP	4782	CNN-GRU hybrid model	IRI	0.89
[55]	LTPP	395	PSO-SVR	IRI	0.91
[56]	LTPP	1414	ANN	IRI	0.92

Furthermore, a large portion of existing methods remains in the black-box prediction stage, lacking a deep understanding of the mechanisms underlying IRI evolution under multi-factor coupling conditions. To overcome these limitations, this study develops an interpretable pavement roughness index (IRI) prediction framework based on long-term pavement performance test (LTPP) data. This framework simultaneously considers multiple climate zones, pavement structure configurations, and traffic conditions. A pre-trained tabular model (TabPFN) is introduced to learn from prior information of feature interactions, improving prediction accuracy and stability in heterogeneous scenarios. The model achieves an R^2 of 0.9474 and an RMSE of 0.138 on the test set, demonstrating competitive performance compared to previously reported methods.

In addition to prediction accuracy, this study further integrates SHAP-based feature attribution and generalized additive model (GAM) analysis to identify key driving factors and quantify the nonlinear threshold effects controlling pavement roughness evolution. The combination of the pre-trained tabular model and interpretable analysis not only improves predictive capability but also deepens the understanding of the mechanisms underlying pavement performance under traffic–structure–climate coupling interactions, thereby enhancing the practical application value of the proposed framework in pavement management and maintenance decisions. The remainder of this paper is organized as follows: Section 1 introduces the research background and objectives; Section 2 describes the construction and feature system of the LTPP dataset; Section 3 introduces the development of machine learning models, hyperparameter optimization, and performance evaluation methods; Section 4 compares the predictive performance of different models under multi-factor pavement conditions and reveals the key influencing factors and mechanisms of IRI evolution through interpretability analysis; Section 5 discusses the limitations of this study and future research directions; finally, Section 6 summarizes the main conclusions. The entire research process is shown in Figure 1.

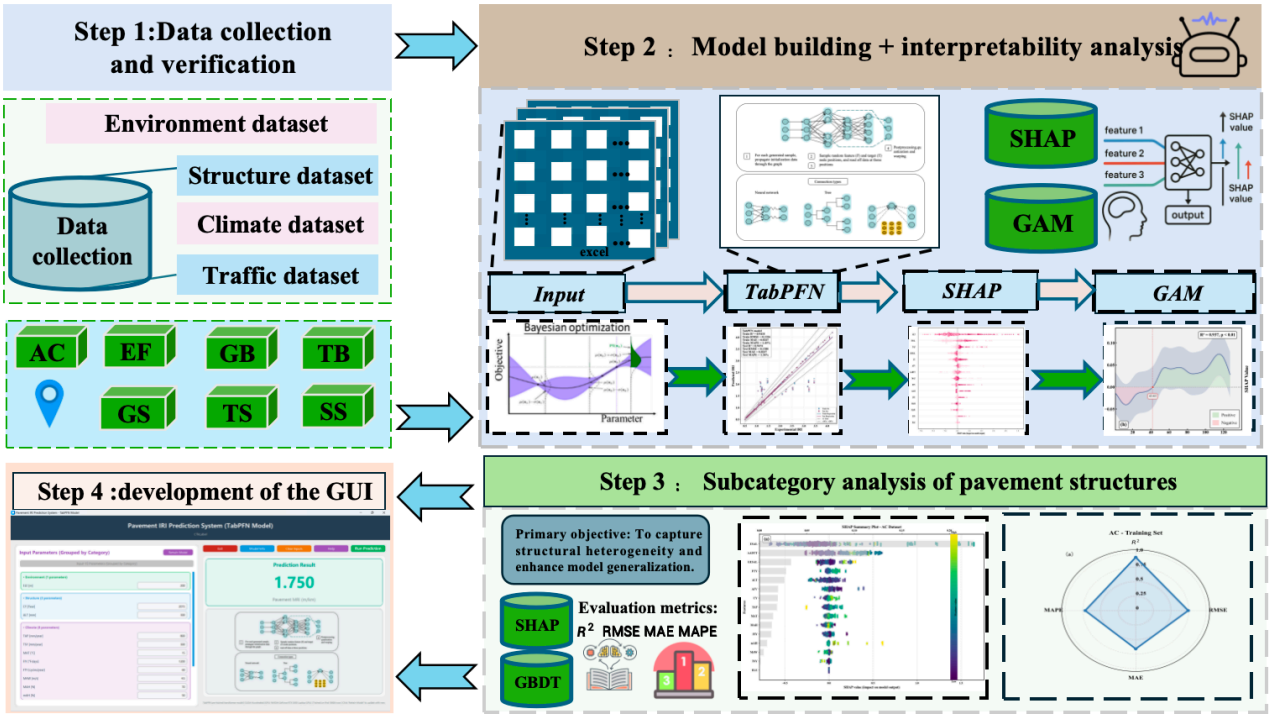


Figure 1. Workflow of this paper.

2. Database Construction and Feature Development

2.1. Data Collection and Preprocessing

This study is based on the Long-Term Pavement Performance (LTPP) database maintained by the Federal Highway Administration (FHWA) of the U.S. Department of Transportation. The LTPP project was originally part of the Strategic Highway Research Program (SHRP) and has been continuously updated since 1987 [57]. The database provides long-term field performance data collected from test sections in multiple climate zones in North America, including pavement structure characteristics, traffic load information, climate conditions, and damage indices. These comprehensive datasets have been widely used in pavement performance modeling and infrastructure management research. The data used in this study came from the official LTPP InfoPave platform, which provides standardized access to the LTPP database. After data extraction, cleaning, and integration, 10,836 segment-year observations were obtained for subsequent analysis. For the detailed key list and other information, please see Appendix A.

To ensure data quality and reproducibility, data preprocessing followed a consistent and transparent workflow. The initial dataset extracted from the LTPP database contained 224,801 segment-year observations. Preprocessing comprised four main steps:

(1) Records with missing values in the target variable, the International Roughness Index (IRI), were removed, as they could not be used for supervised model training.

(2) For the remaining observations, continuous predictor variables with a small number of missing values were imputed using the global mean of each variable, while categorical variables contained no missing values and required no imputation.

(3) These preprocessing steps resulted in a final modeling dataset of 10,836 complete observations.

(4) To avoid potential data leakage in time-dependent prediction tasks, the dataset was split into training and test sets according to observation year: records prior to 2000 were used for training, and observations from 2000 onwards were used for testing. All imputation values were computed solely from the training set to ensure consistency and reproducibility.

As shown in Table 2, flexible pavements constitute the majority of the dataset (8412 samples), followed by rigid pavements (1724 samples) and composite pavements (700 samples). Since flexible pavement segments dominate the dataset, the trained model may primarily reflect the performance patterns of flexible pavement structures. Therefore, caution should be exercised when generalizing the results to rigid pavement systems with different structural properties.

Table 2. Number of samples of different road surfaces.

Pavement Type	Sample Size
Flexible pavement	8412
Rigid pavement	1724
Composite pavement	700

After handling missing values, the dataset used for modeling contained 10,836 valid segment-year observations, which constituted the final modeling samples. Categorical variables describing pavement layer types (AC, EF, GB, GS, SS, TB, and TS) are encoded using one-hot coding to preserve their nominal features and avoid introducing artificial ordinal relationships. Continuous predictor variables are standardized using Z-score normalization to eliminate scale differences between features. Normalization parameters are estimated only using the training set and then applied to the test set to prevent information leakage. The target variable (IRI) retains its original scale to maintain its engineering interpretability.

2.2. Pavement Layer Materials and Feature Representation

As shown in Table 3 and Figure 2, the material categories or functional types of the pavement layers included Asphalt Concrete (AC), Elastic Foam (EF), Granular Base (GB), Granular Subbase (GS), Subgrade Soil (SS), Treated Base (TB), and Treated Subgrade (TS). These categories correspond to standardized material or layer-type identifiers recorded in the LTPP database to describe the structural composition of pavement sections.

Table 3. Source of Samples and Description of Function and Features for Each Layer.

Layer Type	Primary Function	Pavement Layer
Asphalt Concrete (AC)	Provides structural strength and durability; supports traffic loads and ensures ride comfort.	Surface Layer
Elastic Foam (EF)	Absorbs stress and energy; reduces vibration and noise; enhances fatigue performance.	Base Layer
Granular Base (GB)	Provides bearing capacity; distributes loads and ensures drainage.	Base Layer
Treated Base (TB)	Strengthens structure via stabilization; reduces permanent deformation and rutting.	Base Layer
Granular Subbase (GS)	Serves as transitional support; improves stability, drainage, and frost resistance.	Subbase Layer
Treated Subbase (TS)	Enhances sublayer load bearing; mitigates soil deformation and improves long-term performance.	Subbase Layer
Subgrade Soil (SS)	Forms natural foundation; determines bearing capacity and deformation control	Subgrade

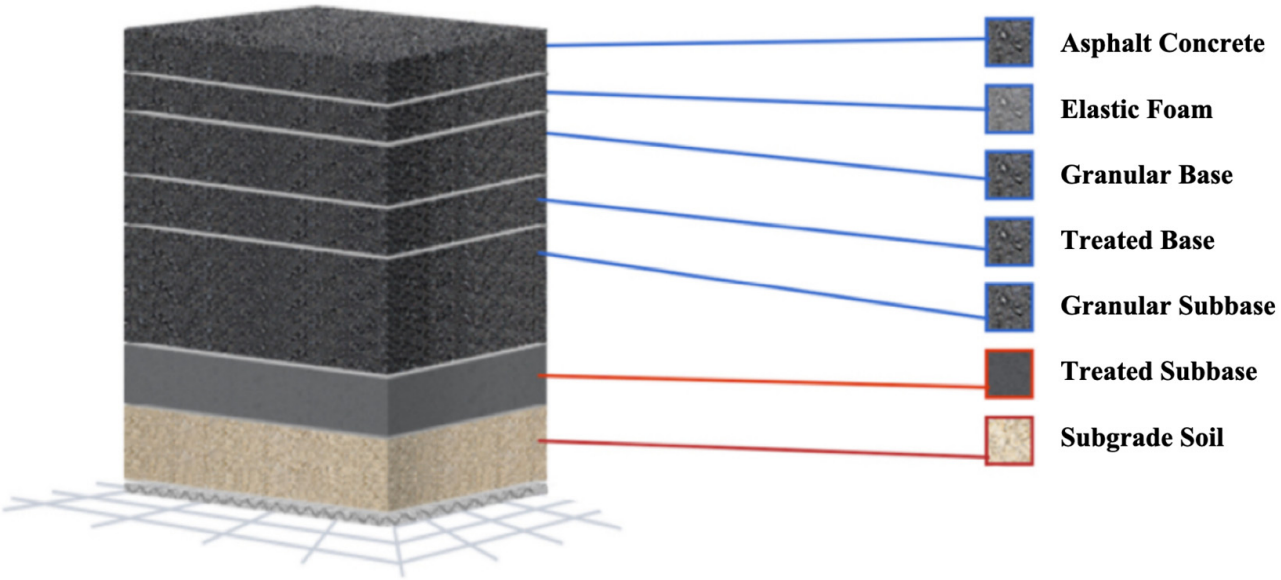


Figure 2. Schematic Diagram of Pavement Layers.

Among them, Elastic Foam (EF) refers to compressible foam-type materials used as stress-absorbing or cushioning layers within the pavement structure. In some pavement systems, such layers are introduced to enhance energy dissipation, reduce vibration transmission, and mitigate fatigue damage caused by repeated traffic loading. Due to their distinct mechanical behavior—characterized by higher compressibility and damping capacity compared with conventional granular or stabilized layers—EF layers may influence load transfer mechanisms and long-term pavement roughness development differently. Therefore, in this study, EF was retained as an independent material category to allow the machine learning models to capture potential differences in structural performance associated with this layer type.

2.3. Data Features

Descriptive Statistics

A multidimensional feature framework was developed by integrating pavement structural attributes, climatic conditions, and traffic loading characteristics to comprehensively characterize pavement performance. Elevation (ELE) characterizes the geographical and environmental context of each road segment. Structural variables, including average layer thickness (ALT) and service life (CY), reflect the physical properties and service history of the pavement structure. Climate variables, encompassing precipitation, temperature, freeze–thaw cycles, and humidity—TAP, TSY, MAT, FIY, FTY, MAW, MAH, and mAH—reflect the impact of environmental factors on pavement performance. Traffic load variables, including AADTT, ATV, ESAL, and GESAL, quantify the strength and cumulative impact of vehicle loads on the pavement structural layers.

Following data preprocessing, descriptive statistical analyses were performed for all features. Metrics, including the mean, median, minimum, maximum, and frequency distribution, were calculated to characterize the dataset. The results are summarized in Table 4, providing a detailed overview of the pavement materials, structural properties, and environmental and loading conditions captured in the study.

Table 4. Descriptive statistical analysis of the variables.

Types of Variables		Variables	Min.	Max.	Avg.	Median.
Input	Pavement Structure	ELE	−34.00	5991.00	419.66	274.00
		ALT	1.81	76.60	7.73	5.83
		CY	1989.00	2012.00	1991.91	2000.00
	Climate	TAP	23.60	2619.70	852.95	865.90
		TSY	0.00	10,286.00	671.16	501.00
		MAT	−4.10	26.60	12.15	11.40
		FIY	0.00	3369.00	384.24	207.00
		FTY	0.00	236.00	78.03	80.00
		MAW	0.70	7.80	3.76	3.80
		MAH	35.00	99.00	85.87	89.00
		mAH	13.00	76.00	46.17	49.00
	Traffic	AADTT	0.00	15,170.00	1191.11	803.00
		ATV	0.00	5,537,050.00	413,002.96	268,640.00
		ESAL	0.00	4,295,722.00	460,667.14	269,629.00
		GESAL	0.00	3,495,555.00	397,360.91	242,292.00
Output		IRI	0.32	5.87	1.49	1.33

2.4. Data Correlation and Collinearity

Figure 3 illustrates the correlation and multicollinearity among the sample dataset features. Panel (a) presents a detailed comparison of pairwise feature correlations using Spearman rank analysis [58], while panel (b) depicts the multicollinearity among variables, with correlation coefficients and variance inflation factors (*VIF*) computed according to Equations (1) and (2).

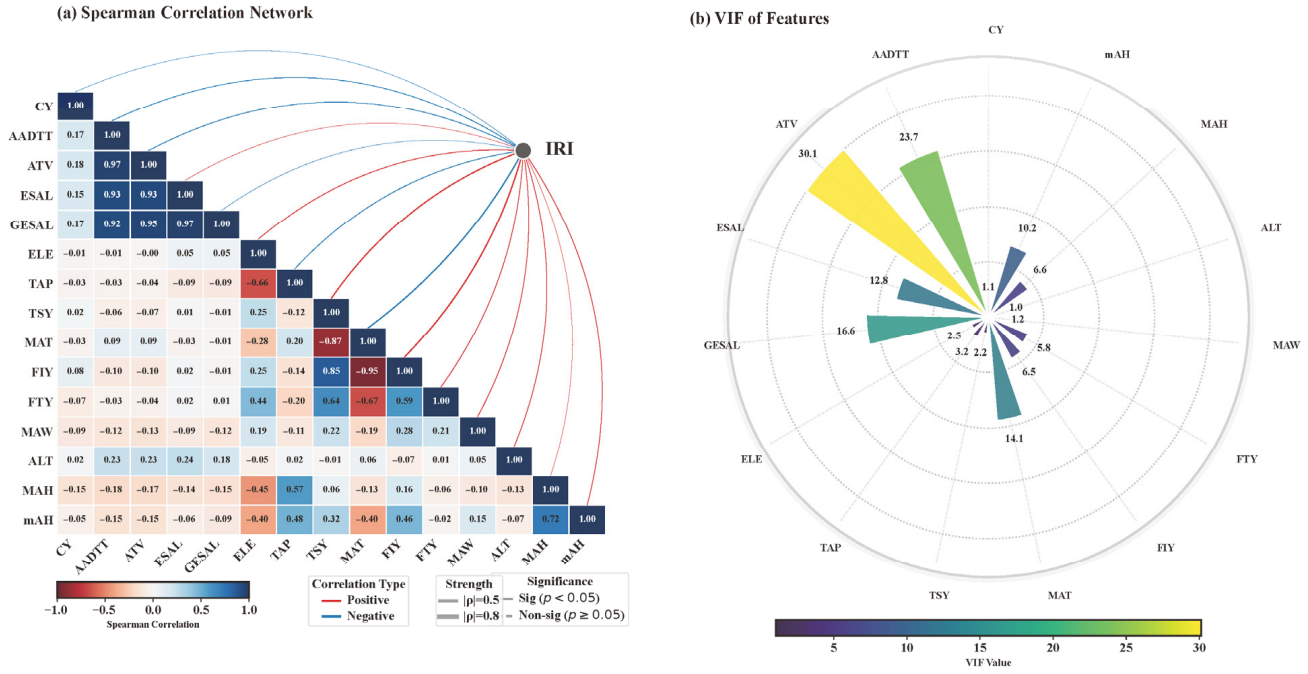


Figure 3. Feature Correlation and Multicollinearity Analysis of the Sample Dataset.

The traffic loading variables AADTT, ATV, ESAL, and GESAL exhibited high statistical correlations. Multicollinearity analysis indicated that several variables (e.g., ATV, AADTT, and GESAL) had VIF values exceeding 15, suggesting significant multicollinearity. This is attributable to the fact that these variables reflect different aspects of traffic dynamics,

including traffic flow trends and cumulative equivalent axle loads. Their correlation and multicollinearity arise primarily from their derivation based on truck traffic volume and cumulative effects, reflecting consistent physical mechanisms and computational logic within the same traffic system at different scales and weights, rather than redundancy or data anomalies. Consequently, these variables were retained in the modeling process to provide a comprehensive characterization of traffic loading effects on pavement structural performance.

Analysis of climatic variables revealed a significant negative correlation between annual mean temperature (MAT) and both annual snowfall (TSY) and freeze index (FIY). This reflects the longer duration and intensity of winter in colder regions, where snow accumulation and freezing processes are more pronounced. The observed negative correlation captures the intrinsic coupling of climatic factors across temperature levels and cold severity, rather than indicating any data irregularities.

$$VIF_j = \frac{1}{1 - R_j^2} \tag{1}$$

Here, VIF_j denotes the variance inflation factor of the j-th independent variable, and R_j^2 is the coefficient of determination obtained by regressing this variable against all other independent variables. Higher VIF values indicate stronger correlations with other predictors and a greater risk of multicollinearity. Spearman correlation coefficients were computed as

$$\rho(x, y) = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(R(x_i) - \overline{R}(x)\right)\left(R(y_i) - \overline{R}(y)\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(R(x_i) - \overline{R}(x)\right) \cdot \frac{1}{n}\sum_{i=1}^{n}\left(R(y_i) - \overline{R}(y)\right)}} \tag{2}$$

where $R(x_i)$ and $R(y_i)$ represent the ranks of x_i and y_i, and $\overline{R}(x)$ and $\overline{R}(y)$ are the mean ranks.

3. Methodology

3.1. K-Fold Cross-Validation

K-Fold cross-validation (CV) is a standard technique to evaluate model performance and stability. The dataset is split into K equally sized folds; in each iteration, K–1 folds are used for training and the remaining fold for validation. Repeating this process K times and averaging the results reduces bias from a single split and provides a more reliable estimate of model performance. When multiple observations belong to the same entity, the standard K-Fold may cause data leakage. Grouped K-Fold addresses this by keeping all observations from the same group together (e.g., a pavement section), ensuring no data from the same section appears in both training and validation sets [59].

In this study, the final modeling dataset contained 10,836 complete observations, with no missing values for either the target or predictor variables. The dataset was partitioned chronologically: records prior to 2000 constituted the training set, and observations from 2000 onwards were used for the test set. It is important to note that all data preprocessing steps were performed as described in Section 2.1. In the training set, five-fold cross-validation was performed using SHRP_ID as the grouping key, ensuring all measurements for each pavement segment were kept within the same fold. The mean root mean square error (RMSE) of each fold was used to guide Bayesian hyperparameter optimization. Subgroup analyses were performed separately. First, the complete dataset was stratified by pavement structure type, and then independent pavement segment-level training-test set partitions were performed within each subgroup to create homogeneous subsets for sensitivity analysis.

3.2. Hyperparameter Optimization Methods

In machine learning models, hyperparameter optimization is a critical procedure aimed at identifying the optimal combination of parameters to maximize model performance [60]. Hyperparameters must be defined prior to model training and have a significant impact on predictive accuracy and generalization. Common optimization approaches include grid search and Bayesian optimization. Grid search exhaustively evaluates all possible parameter combinations to identify the best configuration and is suitable when the number of hyperparameters is limited and computational resources are sufficient. In contrast, Bayesian optimization constructs a probabilistic surrogate model to effectively balance exploration and exploitation, making it particularly advantageous when model evaluations are computationally expensive, such as in predictive modeling of pavement structures and pavement materials.

3.3. Machine Learning Model

The pavement dataset compiled in this study encompasses a wide range of structural, traffic, and environmental conditions. To mitigate potential overfitting and comprehensively evaluate predictive performance, nine regression models were employed: TabPFN, TabM, GBDT, MLP, RF, XGBoost, SVR, LightGBM, and KNN. These models were selected to represent diverse learning paradigms, ranging from ensemble and tree-based methods to deep learning and instance-based approaches, thereby ensuring robust modeling of complex, high-dimensional interactions among pavement materials and environmental factors.

Gradient Boosting Decision Trees (GBDT) [61] iteratively construct regression trees using residuals from prior iterations as optimization targets, achieving high-precision fitting of complex nonlinear relationships. TabPFN [62], a pre-trained probabilistic feed-forward network, learns universal inference rules from large-scale synthetic tabular data, enabling near-Bayesian optimal predictions under limited sample conditions. TabM [63] employs a modular deep network tailored for tabular data, enhancing representation of high-dimensional heterogeneous features through feature embeddings and learnable interaction modules. Multi-Layer Perceptions (MLP) [64] approximate arbitrary continuous functions via multi-layer nonlinear mappings, offering strong global modeling capacity but sensitivity to dataset size and regularization. Random Forests (RF) [65] integrate multiple decision trees trained on randomly sampled subspaces of samples and features, effectively reducing variance and improving robustness. XGBoost [66] extends the gradient boosting framework with second-order derivative information and regularization terms, enabling efficient and controllable tree learning. Support Vector Regression (SVR) [67] constructs an optimal regression hyperplane in a high-dimensional feature space using kernel functions, grounded in structural risk minimization. LightGBM [68] leverages histogram-based gradient boosting and leaf-wise growth strategies to enhance computational efficiency without sacrificing accuracy. Finally, K-Nearest Neighbors (KNN) [69], a non-parametric instance-based method, predicts outcomes based on distance-weighted averages of neighboring samples, capturing local structural patterns but demonstrating sensitivity to feature scaling and noise.

3.4. Model Evaluation Metrics

This study used root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and R^2 to evaluate model performance. In this study, the target variable (IRI) was not transformed or standardized during model training and evaluation. Only the predictor variables were normalized using Z-score standardization, while the IRI values of the predicted and observed values were retained at their original

physical scale (IRI range: 0.32–5.87). Therefore, all performance metrics were calculated directly based on the original IRI scale, ensuring consistency and interpretability of the reported results. The difference between RMSE and MAE stems from the distribution of prediction errors; a small number of relatively large deviations increase the squared error term, resulting in a higher RMSE than MAE [70]. The corresponding computational formulas for these metrics are provided in Table 5.

Table 5. Formulas for the 4 Model Evaluation Indicators.

Index	Formula	
R^2	$R^2 = 1 - \frac{\sum_{k=1}^{N}(y_k - \hat{y}_k)^2}{\sum_{k=1}^{N}(y_k - \bar{y})^2}$	(3)
RMSE	$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(y_k - \hat{y}_k)^2}$	(4)
MAE	$MAE = \frac{1}{N}\sum_{k=1}^{N}\lvert y_k - \hat{y}_k \rvert$	(5)
MAPE	$MAPE = \frac{1}{N}\sum_{k=1}^{N}\frac{\lvert y_k - \hat{y}_k \rvert}{y_k} \times 100\%$	(6)

3.5. Key Factor Interpretability and Critical Threshold Analysis

3.5.1. SHAP Feature Interpretation

SHAP (Shapley Additive Explanations) [71] applies Shapley values from cooperative game theory to quantify the contribution of individual features to single-sample predictions. By computing the marginal contribution of each feature across all possible feature combinations and taking a weighted average, SHAP reveals the decision logic of complex nonlinear models. This approach provides an intuitive visualization of how each feature influences the target variable, facilitating the evaluation of model reliability in predicting pavement performance metrics such as IRI.

3.5.2. Critical Thresholds via Generalized Additive Models (GAMs)

To analyze the influence of environmental factors on IRI and inform optimized maintenance timing, the concept of "critical thresholds" was introduced [72]. These thresholds correspond to points where variations in feature values induce significant changes in IRI. Using input feature values, generalized additive models (GAMs) were constructed to perform nonlinear fitting, allowing the identification of key feature thresholds and revealing the underlying response mechanisms of pavement roughness. This approach addresses the limitations of traditional linear models in capturing nonlinear variations. The GAM formulation is expressed as

$$f(x) = \phi_0 + \sum_{i=1}^{|N|}\phi_i = \phi_0 + \sum_{i=1}^{|N|}[g_i(x_i) + \varepsilon_i] \quad (7)$$

Here, $f(x)$ represents the model-predicted IRI for sample x; ϕ_0 denotes the baseline prediction, i.e., the average prediction across all samples; $|N|$ represents the total number of input features; ϕ_i denotes the SHAP value of the i-th feature, which quantifies that feature's contribution to the model's prediction for a given sample; x_i is the observation value of the i-th feature; $g_i(x_i)$ is the smoothing function of the GAM fitting, used to capture the nonlinear relationship between the feature value and its SHAP contribution; and ε_i is the residual term, representing the fitting error in the GAM process.

3.5.3. Statistical Significance Testing Based on Paired *t*-Test

To statistically assess whether the performance differences between TabPFN and the baseline models are significant, a paired *t*-test was conducted based on the results obtained from the cross-validation folds. The paired *t*-test evaluates whether the mean difference between two paired observations is significantly different from zero. In this study, each pair corresponds to the prediction performance (e.g., R^2 or RMSE) obtained by two models on the same cross-validation fold. By comparing the differences across folds, the test determines whether the observed performance gap reflects a systematic improvement rather than random fluctuations. Let d_i denote the difference in performance between TabPFN and a baseline model on the i-th fold, and let n be the number of folds. The test statistic is calculated as

$$t = \frac{\overline{d}}{S_d / \sqrt{n}} \tag{8}$$

where $\overline{d}$ represents the mean of the paired differences, S_d is the standard deviation of the differences, and n is the number of paired observations. Under the null hypothesis that the mean difference equals zero, the statistic follows t distribution with $n - 1$ degrees of freedom. A small *p*-value indicates that the performance difference between the two models is statistically significant.

3.6. Analysis of Structural Subgroups

To evaluate the impact of pavement structural variations on model predictive performance, a structural subgroup analysis was further conducted. Given that asphalt concrete (AC) constituted the predominant surface material in the dataset, subgroups were defined primarily based on the types of base and subbase layers. For each structural subgroup, the pretrained tabular neural network model TabPFN was independently trained to ensure that model parameters were fully adapted to the specific data characteristics of each subgroup, rather than directly adopting the parameter configuration of the full-sample model. This strategy facilitates a more objective and reliable assessment of predictive performance under different structural conditions. Model performance was comprehensively compared using the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE), enabling an analysis of whether variations in structural layers significantly affect prediction accuracy and stability.

4. Results Analysis and Discussion

4.1. Optimal Hyperparameters of ML Model

To evaluate the stability and predictive performance of the model under different pavement structure configurations, we conducted a robustness analysis. Hyperparameter optimization plays a crucial role in IRI prediction because proper tuning of the model architecture and learning parameters can improve prediction accuracy while balancing model complexity and generalization ability (Table 6).

Hyperparameter Optimization Setup: The hyperparameters of all machine learning models (except TabPFN and TabM) were optimized using Bayesian optimization (BO) to ensure reproducibility. A Gaussian process (GP) regressor with a Matérn kernel ($\nu = 2.5$) served as the surrogate model, and the Expected Improvement (EI) acquisition function with an exploration parameter $\xi = 0.01$ guided the search. The GP model was initialized with 10 randomly sampled points, and the optimization budget was set to 50 iterations per model. The objective function to minimize was defined as the average root mean square error (RMSE) from a five-fold grouped cross-validation on the training set, with SHRP_ID as the grouping key. Critically, this grouped CV procedure was strictly followed within each BO iteration to evaluate candidate hyperparameter sets, ensuring that hyperparameter

tuning and model evaluation remained statistically independent and avoiding potential data leakage. The optimization process terminated when the predefined number of iterations was reached. The convergence behavior of the Bayesian optimization process is illustrated in Appendix B.

Table 6. Hyperparameter optimization results of 9 ML models.

ML Model	Hyperparameter	Search Range	Optimal Value
XGBoost	n_estimators	50–1000	405
	learning_rate	0.02–0.2 (log)	0.0384
	max_depth	4–10	9
	subsample	0.6–1.0	0.7663
	colsample_bytree	0.6–1.0	0.6022
RF	n_estimators	50–1000	810
	max_depth	8–25	10
	min_samples_split	5–20	8
	min_samples_leaf	2–8	2
SVR	C	0.1–1000 (log)	0.7201
	epsilon	0.01–1.0 (log)	0.0666
	kernel	[‘rbf’, ‘poly’]	rbf
	gamma	[‘scale’, ‘auto’]	scale
LightGBM	n_estimators	50–1000	522
	learning_rate	0.02–0.2 (log)	0.0529
	max_depth	4–10	8
	num_leaves	25–100	83
	min_child_samples	5–20	17
KNN	n_neighbors	5–30	25
	weights	[‘uniform’, ‘distance’]	uniform
	p	1–3	2
GBDT	n_estimators	50–1000	665
	learning_rate	0.02–0.2 (log)	0.0270
	max_depth	4–10	9
	min_samples_split	1–20	5
	min_samples_leaf	2–8	6
MLP	activation	[‘relu’, ‘tanh’]	tanh
	learning_rate_init	5×10^{-4}–5×10^{-2} (log)	0.0467
	hidden_layer_sizes	[(100, 50), (200, 100, 50), (256, 128, 64)]	(200, 100, 50)
	alpha (L2 regularization)	1×10^{-5}–1×10^{-2} (log)	0.0038
	max_epochs	200–1000	927
TabPFN	device	Default/Fixed	auto
	n_estimators	Default/Fixed	8
	random_state	Default/Fixed	42
TabM	learning_rate	Default/Fixed	0.0033
	epochs	Default/Fixed	51
	batch_size	Default/Fixed	512

For the TabPFN model, as a prior-data fitted network, minimal hyperparameter tuning is required. We employed the TabPFNRegressor from the tabpfn library (version 6.2.0) with default settings, except for enabling ignore_pretraining_limits = True to accommodate the size of our dataset. The device was set to ‘auto’, and all other parameters, such as n_estimators = 8 and softmax_temperature = 0.9, followed the library defaults, ensuring reproducible and efficient training.

For the TabM model, we used the implementation provided by the official pytorch-tabular library (version 1.1.0). The default configuration for regression tasks was applied via the library's TabModel class, including a learning rate of 0.0033, 51 training epochs, a batch size of 512, and the default optimizer, scheduler, and loss function. No additional hyperparameter search was performed for TabM to maintain consistency with its standard benchmarking setup.

4.2. Comparison of Prediction Performance of Different Machine Learning Models

In traditional machine learning approaches, K-Nearest Neighbors (KNN) and Support Vector Regression (SVR) exhibit relatively limited predictive performance on the test set. Specifically, KNN achieves an R^2 value of 0.5841 with an RMSE of 0.3880, while SVR performs even worse with an R^2 of 0.2929 and an RMSE of 0.5058. Although KNN can capture local similarity patterns within the dataset, its prediction reliability is highly sensitive to the latent distribution of samples and the scale of input variables. In contrast, SVR strongly depends on the selection of kernel functions and hyperparameter configurations. When applied to complex engineering datasets with heterogeneous features, SVR often struggles to adequately represent nonlinear relationships, resulting in significant prediction bias, particularly in regions with higher IRI values. These limitations restrict the generalization capability of both models across different pavement structures and surface conditions.

For ensemble learning algorithms, Gradient Boosting Decision Tree (GBDT), XGBoost, LightGBM, and Random Forest (RF) demonstrate considerably stronger predictive performance. On the test set, the ranking based on R^2 values are LightGBM (0.9334) > RF (0.9250) > GBDT (0.9136) > XGBoost (0.8963), while the RMSE values remain relatively low, ranging from 0.1552 to 0.1937. These tree-based ensemble models effectively capture complex nonlinear interactions among traffic loading, pavement structural parameters, and environmental conditions through iterative residual optimization and ensemble aggregation. By integrating multiple decision trees, these algorithms reduce variance and improve robustness, enabling stable predictive performance across diverse pavement structures and operational environments. In deep learning and tabular modeling approaches, the Multilayer Perceptron (MLP) demonstrates relatively weak predictive capability, achieving a test set R^2 of 0.2106 and an RMSE of 0.2345. Although neural networks theoretically possess strong nonlinear representation capacity, their effectiveness often depends on large-scale datasets. Under limited-sample engineering datasets, the model may fail to fully exploit complex feature interactions, leading to reduced prediction accuracy.

The TabM model, designed specifically for tabular data learning, achieves moderate predictive performance with a test set R^2 of 0.8370 and an RMSE of 0.2429. While this model improves the representation of high-dimensional tabular relationships compared with conventional neural networks, prediction bias remains noticeable under extreme value conditions, suggesting that the model still faces challenges in capturing rare or highly nonlinear degradation patterns.

In contrast, the TabPFN model, which is extensively pre-trained on large-scale synthetic tabular datasets, consistently achieves the best predictive performance among all models. All comparative experiments were rerun under the final preprocessing pipeline described in Section 2.1, and the resulting model rankings remained stable despite minor variations in the numerical metrics. On the test set, TabPFN obtains an R^2 value of 0.9474 and an RMSE of 0.1380, significantly outperforming both conventional machine learning and neural network-based models. This model leverages a prior-informed learning mechanism that embeds feature interaction patterns learned during pre-training, enabling efficient inference even under relatively limited real-world engineering datasets. As a

result, TabPFN demonstrates strong predictive accuracy and robust generalization across different pavement structures, traffic loading conditions, and environmental scenarios.

As shown in Table 7 and overall, based on the test set R^2 values, the predictive performance ranking of the models can be summarized as follows: TabPFN > LightGBM > RF > GBDT > XGBoost > TabM > KNN > SVR > MLP. The superior performance of TabPFN can be attributed to its combination of large-scale prior learning and an embedded feature interaction mechanism, which enables the model to effectively capture the complex nonlinear relationships among pavement structures, traffic loading factors, and environmental conditions. This capability allows TabPFN to simultaneously achieve high prediction accuracy, strong stability, and reliable generalization, making it the most effective method for predicting pavement IRI in this study.

Table 7. Evaluation Metrics Results of Nine Machine Learning Models.

ML Model	Model Performance Evaluation Parameters							
	Train Set				Test Set			
	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE
TabPFN	0.9414	0.1336	0.0247	1.4462	0.9474	0.1380	0.0247	1.3649
TabM	0.8168	0.2370	0.1442	11.0573	0.8370	0.2429	0.1485	11.4749
GBDT	0.9523	0.1210	0.0255	1.6560	0.9136	0.1768	0.0348	2.1917
XGBoost	0.9533	0.1196	0.0277	1.9082	0.8963	0.1937	0.0411	2.7267
RF	0.9347	0.1415	0.0668	5.4092	0.9250	0.1684	0.0723	5.8640
LightGBM	0.9496	0.1244	0.0291	1.9523	0.9334	0.1552	0.0345	2.2967
MLP	0.2309	0.2857	0.3410	27.4663	0.2106	0.2345	0.3661	28.6109
KNN	0.6090	0.3463	0.2264	17.3231	0.5841	0.3880	0.2436	18.5068
SVR	0.3319	0.4527	0.2540	17.1353	0.2929	0.5058	0.2719	17.7228

4.2.1. Statistical Significance Test Results

To verify whether the performance advantage of TabPFN in model comparison is statistically significant, we performed paired *t*-tests on the model's predictive performance based on the k-fold cross-validation results. The Table 7 only shows the comparison results between TabPFN and other models. As shown in Table 8, the performance difference between TabPFN and most of the compared models is statistically significant ($p < 0.05$), further confirming its superiority.

Table 8. Statistical significance of TabPFN performance against other models.

Model	t-Statistic	*p*-Value
XGBoost	5.21	4.62×10^{-4}
RF	4.97	6.13×10^{-4}
SVR	8.86	3.44×10^{-5}
LightGBM	4.82	8.01×10^{-4}
KNN	8.03	7.08×10^{-5}
GBDT	4.93	6.57×10^{-4}
MLP	12.47	1.13×10^{-6}
TabM	11.78	2.94×10^{-6}

To further test whether the performance difference between TabPFN and competing models is statistically significant, we performed paired *t*-tests based on the k-fold cross-validation results. Specifically, we applied paired *t*-tests to compare the prediction errors of TabPFN and each baseline model on the same validation fold. This method allows for a rigorous statistical evaluation of the observed performance improvement to determine

whether it is caused by random fluctuations or represents a sustained advantage of the proposed model.

The results are summarized in Table 8. For all baseline models, the paired *t*-tests showed statistically significant differences ($p < 0.001$), indicating that TabPFN's predictive performance is significantly better than other comparative algorithms. Of particular note is that the t-statistic of TabPFN is significantly higher than that of the neural network model (MLP, t = 12.47) and TabM (t = 11.78), indicating a significant performance improvement. Compared with commonly used ensemble learning models (e.g., XGBoost, Random Forest, LightGBM, and GBDT), TabPFN also exhibits a similar statistically significant performance improvement. In conclusion, statistical tests confirm that the superiority of TabPFN is not only reflected in numerical differences but is also supported by statistically significant evidence in cross-validation, thus enhancing the robustness of model comparisons.

Figure 4 illustrates the boxplot distributions of the ratio between predicted and experimental IRI values (IRI_{pre}/IRI_{exp}) across all models for both the training and test sets. Compared with the other models, TabPFN exhibited a more concentrated distribution centered closely around 1, with the smallest interquartile range, indicating superior stability and consistency of its predictions. Considering both the risk of prediction bias and overall generalization performance, TabPFN outperformed all other models in terms of predictive accuracy and reliability. These results confirm TabPFN as the optimal machine learning approach for forecasting IRI in pavement structures, effectively capturing the complex interactions among structural configurations, traffic loading, and environmental factors while maintaining robust performance across diverse datasets.

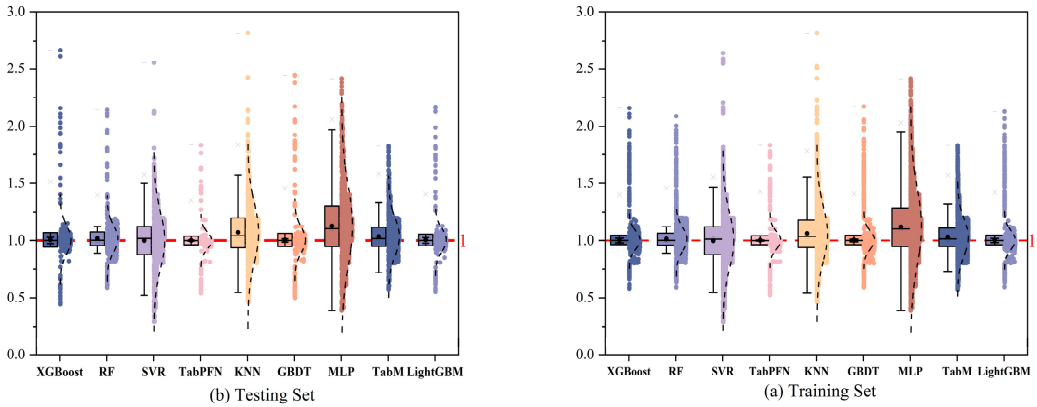


Figure 4. Error comparison between the TabPFN model and other models.

Figure 5 compares the performance of nine machine learning models in predicting the International Roughness Index (IRI) across training and test datasets. Overall, the distribution of prediction points relative to the ideal reference line (y = x) exhibited notable differences among the models. Some models generated predictions that were highly concentrated and predominantly located within a ±10% error band, reflecting strong local accuracy and stability. In contrast, other models showed more pronounced dispersion, indicating reduced precision and variability in capturing complex pavement responses. These results highlight the differential adaptability of the models when forecasting IRI across diverse pavement structures and pavement compositions, emphasizing the necessity of selecting models capable of robustly representing the nonlinear interactions among structural layers, traffic loading, and environmental factors.

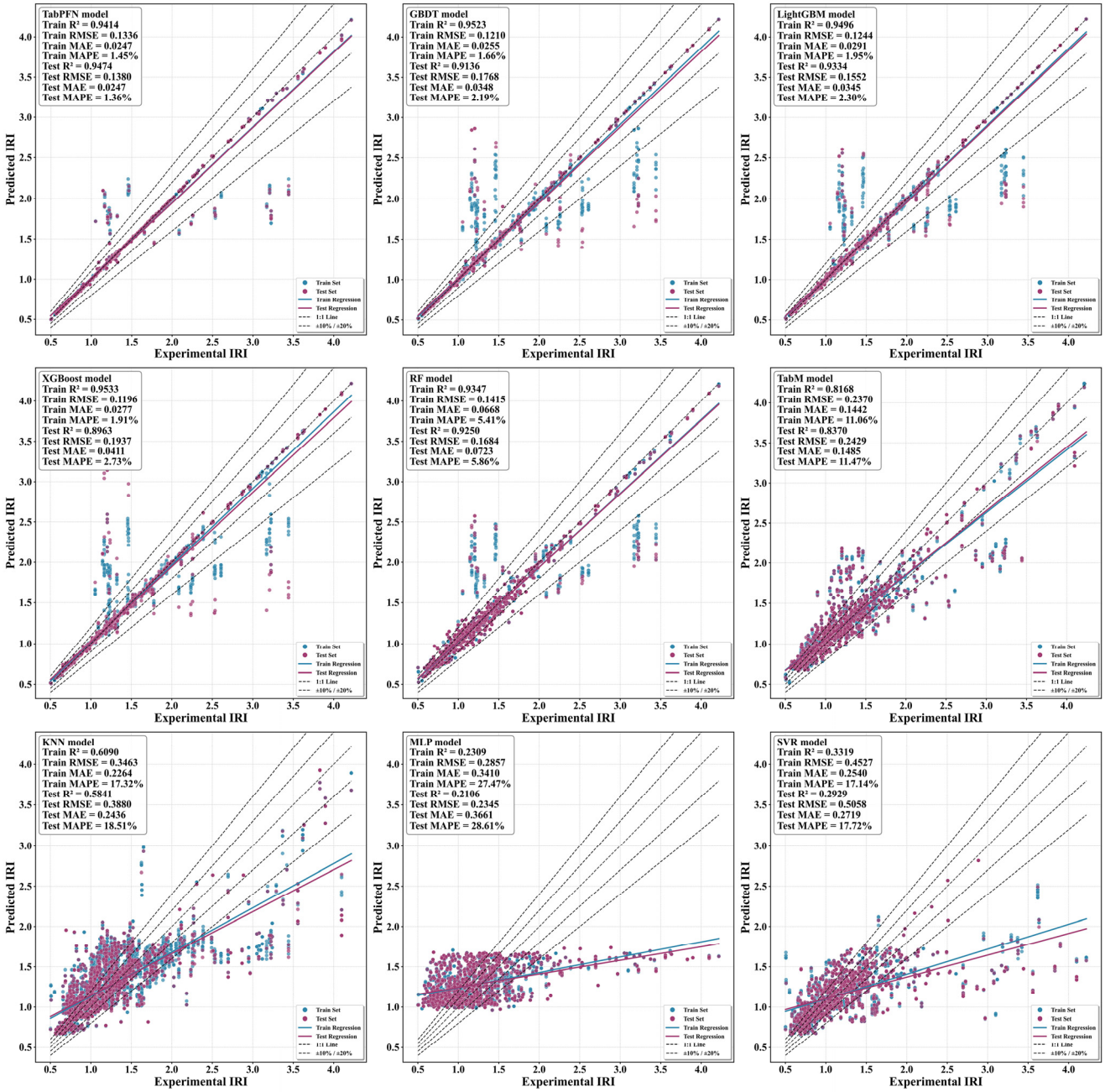


Figure 5. Comparison of predicted and measured IRI values across ML models.

4.2.2. Performance Comparison of Different Road Surface Structures

In Section 4.2, the dataset primarily consists of flexible pavements with asphalt concrete (AC) surface layers, with the remainder being rigid pavements and other structural types. Specifically, out of a total of 10,836 observations, there are 5960 records for AC pavements, 2167 records for GB+TB+EF pavements, 1625 records for GS+TS pavements, and 1084 records for SS pavements. Figure 6 shows the distribution of IRI for these pavement types. These pavement types differ significantly in structural composition and degradation mechanisms. The dominance of flexible pavement samples may introduce structural bias when constructing a unified prediction model applicable to multiple pavement types, thus limiting the model's ability to represent other structural systems.

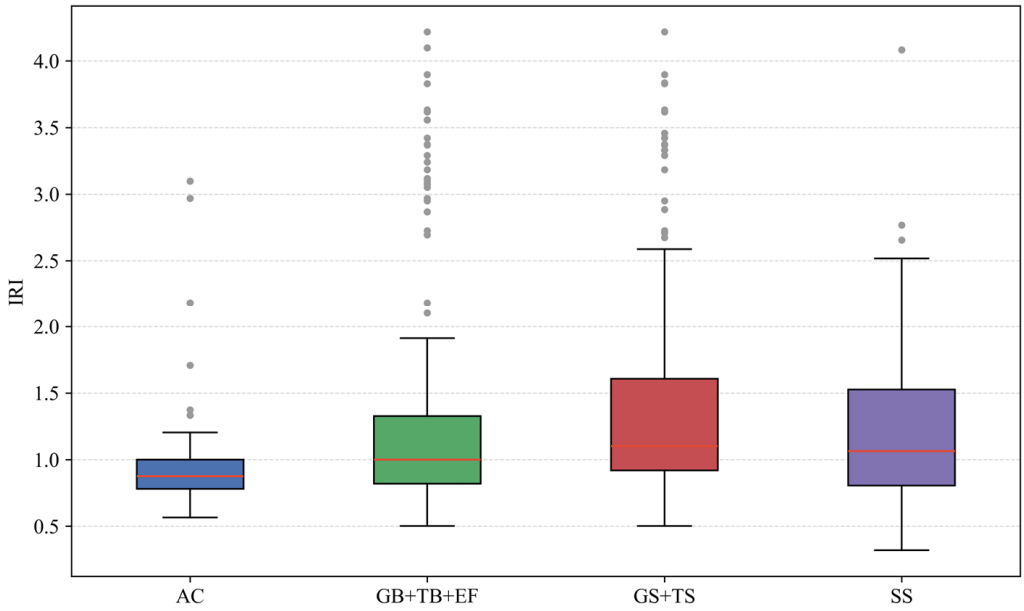


Figure 6. IRI data characteristics.

To investigate the impact of pavement structural configurations on model behaviour, we performed a subgroup analysis by segmenting the dataset based on the presence or absence of specific layer types defined in Section 2.2, creating homogeneous subgroups with similar structural characteristics. Four key subgroups were defined: AC (Asphalt Concrete), where the surface layer is exclusively asphalt concrete and serves as the most prevalent control group; GS+TS, comprising sections that contain either a granular subbase (GS), a treated subbase (TS), or both beneath the surface, representing pavements with an engineered subbase for improved support and drainage; GB+TB+EF, including sections that contain any combination of granular base (GB), treated base (TB), or engineered fill (EF) layers within the base course, representing pavements with a reinforced or stabilized base for enhanced load distribution; and SS (Surface Seal), consisting of sections with a surface seal as the wearing course, representing thinner surface treatments. The predictive performance for each subgroup was evaluated using the same rigorous five-fold grouped cross-validation procedure described in Section 3.1, applied independently to each subgroup, with all observations belonging to the same pavement section (SHRP_ID) kept within the same fold to ensure fair and leakage-free comparisons across structurally distinct pavement types. To ensure a fair and robust evaluation of predictive performance within each structural subgroup, we independently applied the same grouped cross-validation procedure as used for the full dataset (detailed in Section 3.1). Specifically, for each subgroup dataset (AC, GS+TS, etc.), we performed a five-fold grouped cross-validation using the SHRP_ID as the grouping key. This ensures that all observations from the same pavement section are kept within the same fold, preventing data leakage. The performance metrics (R^2, RMSE) reported in Table 9 for each subgroup and model (TabPFN, RF, LR) are the averages across the five CV folds. To quantify the uncertainty of these performance estimates, we employed the bootstrap method. For each model-subgroup combination, we performed 1000 bootstrap resampling iterations on the respective subgroup's dataset, each time recomputing the average performance over a newly drawn five-fold grouped CV split. The 2.5th and 97.5th percentiles of the resulting bootstrap distribution form the reported 95% confidence intervals.

Table 9. Predictive performance of different structural subgroups.

Dataset	Model	Bootstrap CI (95%)			
		Train Set		Test Set	
		R^2 (95% CI)	RMSE (95% CI)	R^2 (95% CI)	RMSE (95% CI)
AC	TabPFN	0.9273 (0.9058–0.9512)	0.0751 (0.0599–0.0896)	0.9243 (0.9036–0.9443)	0.0763 (0.0511–0.0984)
	RF	0.9273 (0.9058–0.9312)	0.0772 (0.0611–0.0914)	0.7578 (0.5778–0.8821)	0.1357 (0.0925–0.1731)
	LR	0.1916 (0.1707–0.2112)	0.2439 (0.2290–0.2596)	0.1997 (0.1605–0.2305)	0.2467 (0.2188–0.2745)
GS+TS	TabPFN	0.9477 (0.9429–0.9508)	0.1616 (0.1575–0.1656)	0.9494 (0.9483–0.9513)	0.1651 (0.1571–0.1726)
	RF	0.9277 (0.9265–0.9301)	0.1912 (0.1868–0.1955)	0.9287 (0.9252–0.9303)	0.1954 (0.1865–0.2033)
	LR	0.5061 (0.4890–0.5255)	0.5078 (0.4942–0.5218)	0.5058 (0.4718–0.5390)	0.5191 (0.4873–0.5468)
GB+TB+EF	TabPFN	0.9478 (0.9440–0.9516)	0.1781 (0.1722–0.1838)	0.9486 (0.9406–0.9557)	0.1793 (0.1668–0.1918)
	RF	0.9479 (0.9445–0.9513)	0.1783 (0.1739–0.1829)	0.9485 (0.9482–0.9687)	0.1795 (0.1709–0.1886)
	LR	0.3661 (0.3462–0.3870)	0.6204 (0.6067–0.6340)	0.3702 (0.3336–0.4048)	0.6278 (0.6014–0.6550)
SS	TabPFN	0.9417 (0.9403–0.9501)	0.1548 (0.1521–0.1573)	0.9495 (0.9424–0.9502)	0.1556 (0.1504–0.1610)
	RF	0.9217 (0.9212–0.9301)	0.1831 (0.1800–0.1860)	0.9246 (0.9214–0.9303)	0.1841 (0.1780–0.1901)
	LR	0.3707 (0.3631–0.3786)	0.5490 (0.5392–0.5580)	0.3705 (0.3542–0.3861)	0.5520 (0.5237–0.5721)

To further evaluate the predictive reliability of different pavement structure subgroups, we compared baseline models and quantified the predictive uncertainty for four representative pavement structures (AC, GS+TS, GB+TB+EF, and SS). Linear regression (LR) and random forest (RF) were used as baseline models. For each model and subgroup, we performed five-fold grouped cross-validation (CV) using SHRP_ID as the grouping key to ensure that observations from the same pavement segment remained in the same fold. We calculated the performance metrics for the training data within each fold and the corresponding validation folds, and reported the average of the five CV folds. To further quantify the uncertainty of these performance estimates, we applied 1000 bootstrap resampling iterations to the fold-level results to estimate the 95% confidence interval (CI). We used percentile bootstrap, with the 2.5th and 97.5th percentiles of the bootstrap distribution as the lower and upper bounds of the confidence intervals. In Table 9, the "Train Set" and "Test Set" columns correspond to the average training performance within each cross-validation fold and the average validation performance across all folds, respectively, rather than the results of a single fixed training/test set partition. The results are summarized in Table 9 and Figure 7.

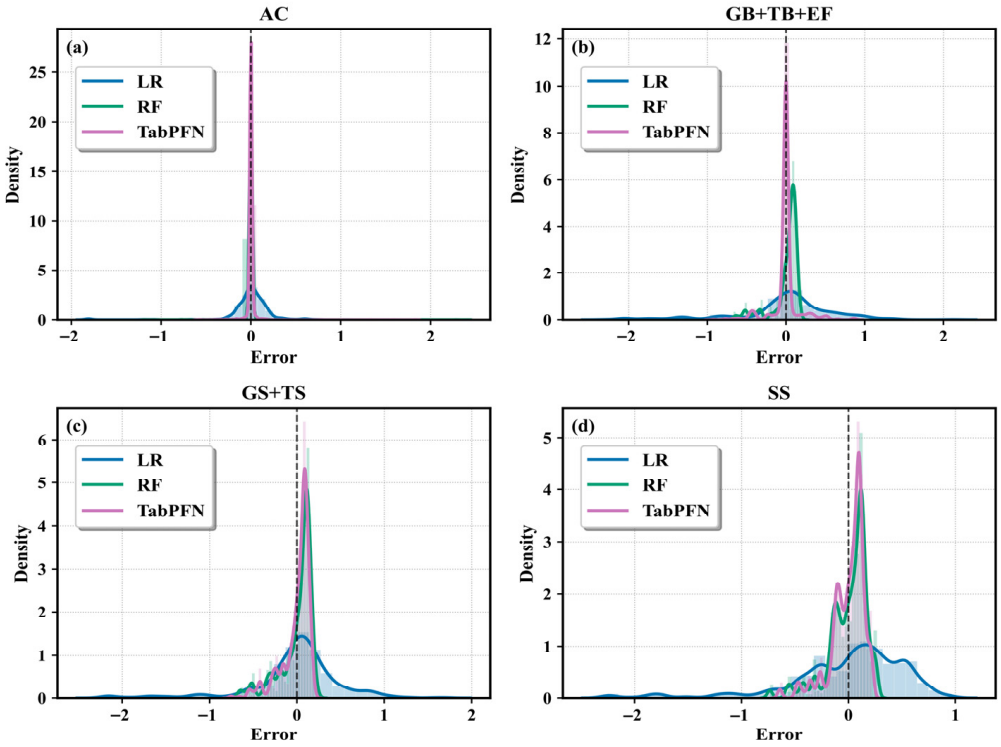


Figure 7. Error distribution of each model.

In all subgroups, TabPFN generally outperformed the baseline models, with narrower bootstrap confidence intervals suggesting relatively stable predictions. For example, in the AC subgroup (5960 records), TabPFN achieved a test set R^2 of 0.924 (95% CI: 0.9036–0.9443), notably higher than Random Forest (R^2 = 0.758) and Linear Regression (R^2 = 0.200). Similar trends were observed in the GS+TS (1625 records), GB+TB+EF (2167 records), and SS (1084 records) subgroups, indicating that TabPFN can effectively capture nonlinear relationships between structural, traffic, and environmental factors, while partially mitigating overfitting issues associated with smaller sample sizes or potential data leakage. Overall, these results suggest that the model provides reasonably robust and reliable predictions across the examined pavement structures, though performance may vary depending on subgroup size and data characteristics.

4.3. Discussion on the Interpretability of Machine Learning Models

Based on the principles of cooperative game theory, SHapley Additive exPlanations (SHAP) can quantify the contribution of each input feature to the model's prediction outcomes. In this framework, the sign of a SHAP value indicates whether the corresponding feature increases (positive) or decreases (negative) the predicted outcome relative to the baseline (mean prediction), while its magnitude reflects the strength of its influence. In this study, SHAP was applied to the performance-optimal TabPFN model to conduct an interpretability analysis, aiming to elucidate the mechanistic effects of input variables on IRI predictions [73].

4.3.1. SHAP Global Analysis

Figure 8 presents the SHAP summary plot, in which each row corresponds to a feature, ordered from top to bottom according to its importance. Each point represents the SHAP value for a specific sample. The color gradient, ranging from low to high, reflects the actual value of the corresponding feature, allowing visual interpretation of how feature magnitude relates to its impact on predictions. The results indicate that the Average Layer Thickness (ALT) of the pavement structure exerted the greatest influence on IRI, as it

governs the structural stiffness and load-dispersion capacity of the pavement system. An excessively thin layer tends to induce stress concentration, whereas an overly thick layer may be more susceptible to interlayer shear and temperature gradients, thereby accelerating surface roughness development. The top five key features influencing IRI were ALT, ESAL, TAP, GESAL, and CY. Among these, ESAL and GESAL represent traffic loading factors, TAP corresponds to climatic conditions, while ALT and CY are associated with pavement structural characteristics.

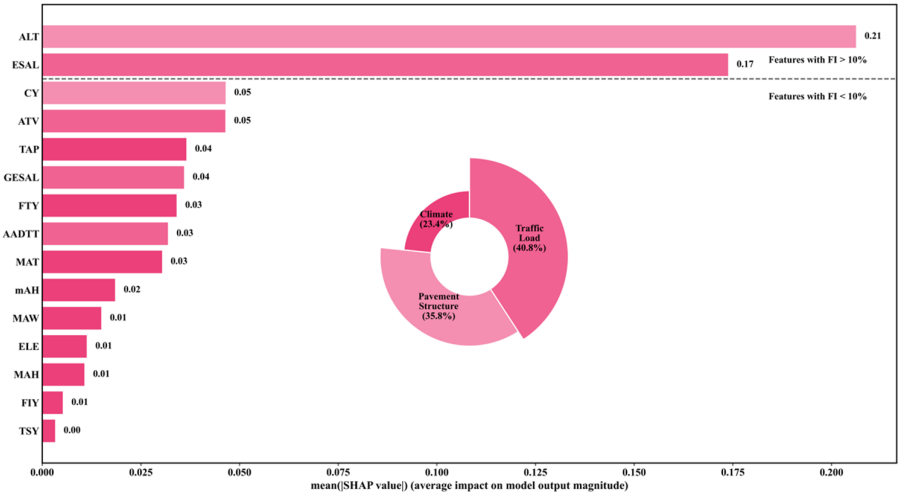


Figure 8. Local feature contributions to model predictions visualized using SHAP values.

Average contribution of each feature to the model's output quantified using the mean absolute SHAP values (mean |SHAP|) (Figure 9). The horizontal axis represents feature importance, and the vertical axis lists the features in descending order of contribution. Traffic loading variables accounted for 39.5% of the total influence, primarily through axle load magnitude and repeated loading cycles that accelerate pavement roughness. Climatic variables contributed 24.5%, affecting the performance of pavement materials and the rate of structural deterioration through temperature variation, precipitation, humidity, and freeze–thaw cycles. Pavement structural factors accounted for 35.9% of the total influence, among which ALT played a dominant role by controlling load dispersion and structural stiffness within the pavement system.

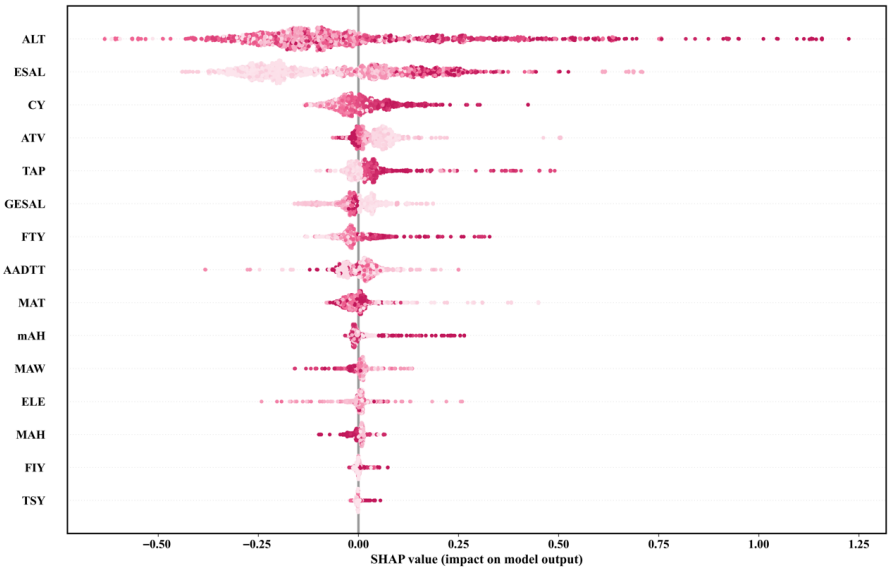


Figure 9. Ranking of global feature importance in the road roughness (IRI) prediction model.

4.3.2. Generalized Additive Model (GAM) Analysis

In Section 4.3.1, we first used SHAP analysis to quantify the contribution of each variable to the IRI. To further characterize the nonlinear relationship and identify potential transition intervals, we fitted a generalized additive model (GAM) to characterize the relationship between SHAP values and the corresponding input feature values (Figure 10a–l). This GAM employs a penalized spline smoother with automatically selected smoothing parameters. Potential thresholds were identified by the inflection points of the fitted curves, defined as the points where the first derivative changes sign or the SHAP contribution crosses zero. To reduce sampling variability, we applied bootstrap resampling to estimate confidence intervals; therefore, the resulting thresholds represent approximate transition points rather than strict deterministic limits. Related threshold reports are available in Appendix C.

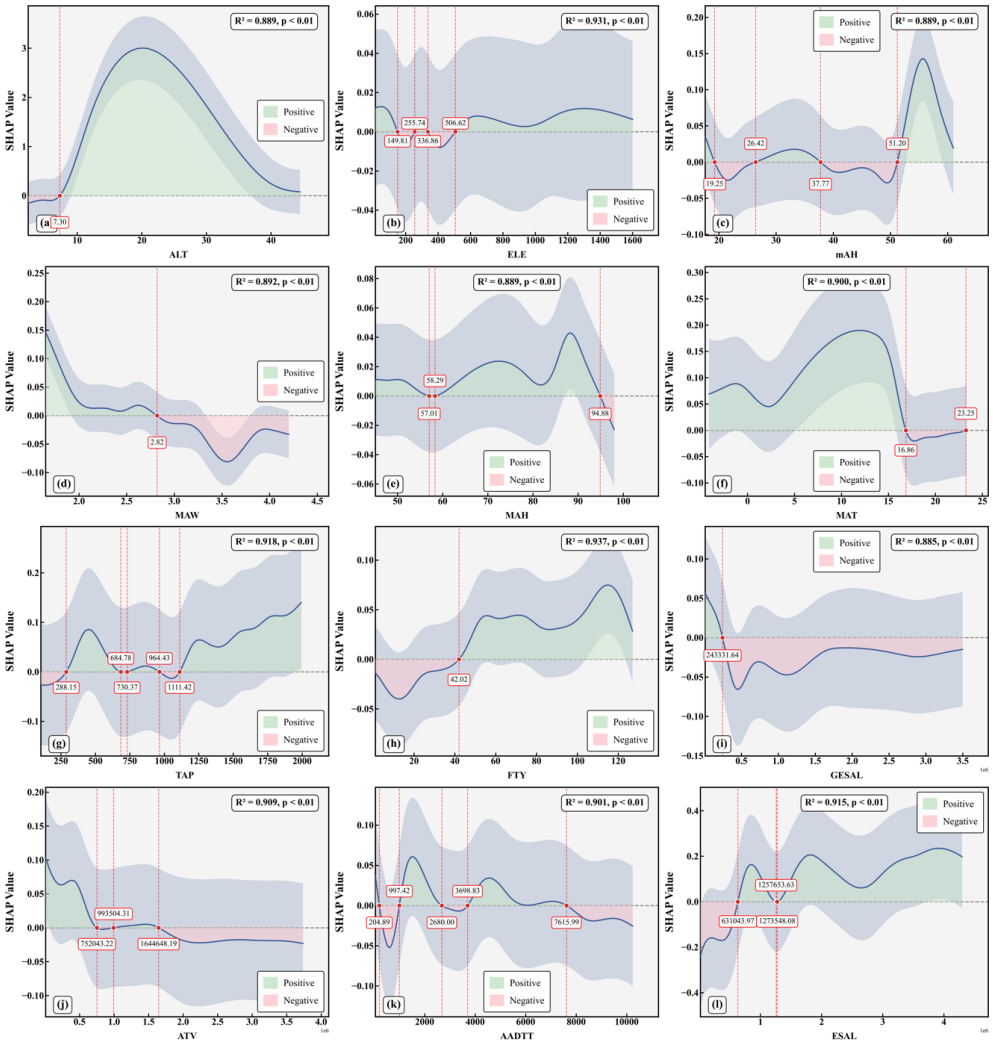


Figure 10. Identification of critical points for key factors in each category.

To improve engineering interpretability, the critical transition intervals identified from the SHAP–GAM analysis are summarized in Table 10, which reports the corresponding variables, threshold ranges, and their practical implications for pavement deterioration. As shown in Figure 10.

Table 10. SHAP-GAM Thresholds and Engineering Implications for Pavement Structures.

Variable	Threshold	Observed Effect on IRI	Engineering Implication
ALT	>7.30 cm	IRI increases with higher SHAP contribution	Pavement performance becomes more sensitive to structural thickness under repeated traffic loading
ELE	<149.81 m or >506.62 m	Higher SHAP contributions at extreme elevations	Extreme elevation conditions may intensify climate-related pavement deterioration
mAH	26.42–37.77%	Increased SHAP contribution	Moderate humidity ranges may promote material deterioration processes
MAW	<2.82 m/s	Increased SHAP contribution to IRI	Low wind speed environments may facilitate moisture retention and accelerate pavement degradation
MAH	58.29–94.88%	Increased IRI risk	High humidity weakens pavement materials and accelerates roughness development
MAT	<16.8 °C	Higher SHAP contribution	Lower temperature environments intensify pavement deterioration
TAP	Around 1111.42 mm	Second transition point in SHAP response	High precipitation may accelerate moisture-related pavement damage
FTY	>42 cycles/year	Significant increase in IRI	Repeated freeze–thaw cycles accelerate pavement structural damage
GESAL	<243,331.64	Increased SHAP contribution	Rapid traffic growth accelerates cumulative load effects and pavement deterioration
ATV	<752,043	Higher SHAP contribution observed with increasing values	Higher truck traffic intensity accelerates cumulative traffic loading effects on pavement roughness
AADTT	997–2680; 3699–7616	Elevated SHAP contribution	Moderate to high truck traffic levels accelerate pavement fatigue accumulation and intensify pavement deterioration.
ESAL	>6.31 × 10^5	Rapid increase in SHAP contribution	Cumulative heavy traffic loading accelerates pavement roughness growth

For structural factors, IRI shows a clear sensitivity to pavement structure in subfigure (a), where the average layer thickness (ALT) exceeds approximately 7.30 cm, suggesting that structural responses to repeated traffic loading become more pronounced beyond this point. Subfigure (b) shows that altitude (ELE) also exhibits nonlinear behavior, with higher SHAP contributions observed below 149.81 m and above 506.62 m, indicating that extreme elevation conditions may intensify climate-related deterioration mechanisms.

For environmental and climatic factors, several transition intervals associated with accelerated roughness growth are identified. Subfigure (d) indicates that IRI risk increases when the annual mean wind speed (MAW) falls below 2.82 m/s. Subfigures (c) and (e) show that higher SHAP contributions are observed when the annual mean minimum humidity (mAH) ranges between 26.42% and 37.77% and the annual mean maximum humidity (MAH) ranges between 58.29% and 94.88%, suggesting moisture-related weakening of pavement materials. Subfigure (f) further demonstrates that lower annual mean temperature (MAT < 16.86 °C), while subfigure (g) highlights precipitation transition ranges around 288.15 mm and 1111.42 mm, both indicating climatic conditions under

which pavement structures become more vulnerable to degradation. In addition, subfigure (h) shows that IRI deterioration accelerates when the number of freeze–thaw cycles (FTY) exceeds approximately 42 per year, highlighting the damaging effects of repeated freeze–thaw processes.

For traffic-related variables, the SHAP–GAM analysis reveals clear loading thresholds associated with accelerated roughness growth. Subfigure (l) shows that IRI increases significantly when the cumulative equivalent single axle load (ESAL) exceeds approximately 6.31×10^5, indicating that pavement deterioration becomes more pronounced beyond this traffic loading level. Subfigures (k) and (j) further indicate that elevated SHAP contributions are observed when AADTT ranges between 997–2680 and 3699–7616, and when ATV reaches higher levels, suggesting that sustained heavy truck traffic accelerates pavement fatigue accumulation. Similarly, subfigure (i) shows the effect of GESAL on IRI growth. These quantitative thresholds provide potential reference indicators for pavement monitoring and maintenance decision-making.

4.3.3. Interpretability Analysis of Different Pavement Structures

Section 4.2.1 utilizes a unified modeling and interpretation framework to analyze the overall impact pattern of the entire dataset on IRI. However, different types of pavements have different structural compositions and degradation mechanisms. Therefore, this study applies the TabPFN model to each subgroup and uses the SHAP summary plot in Figure 11 to reveal the directional influence of key features.

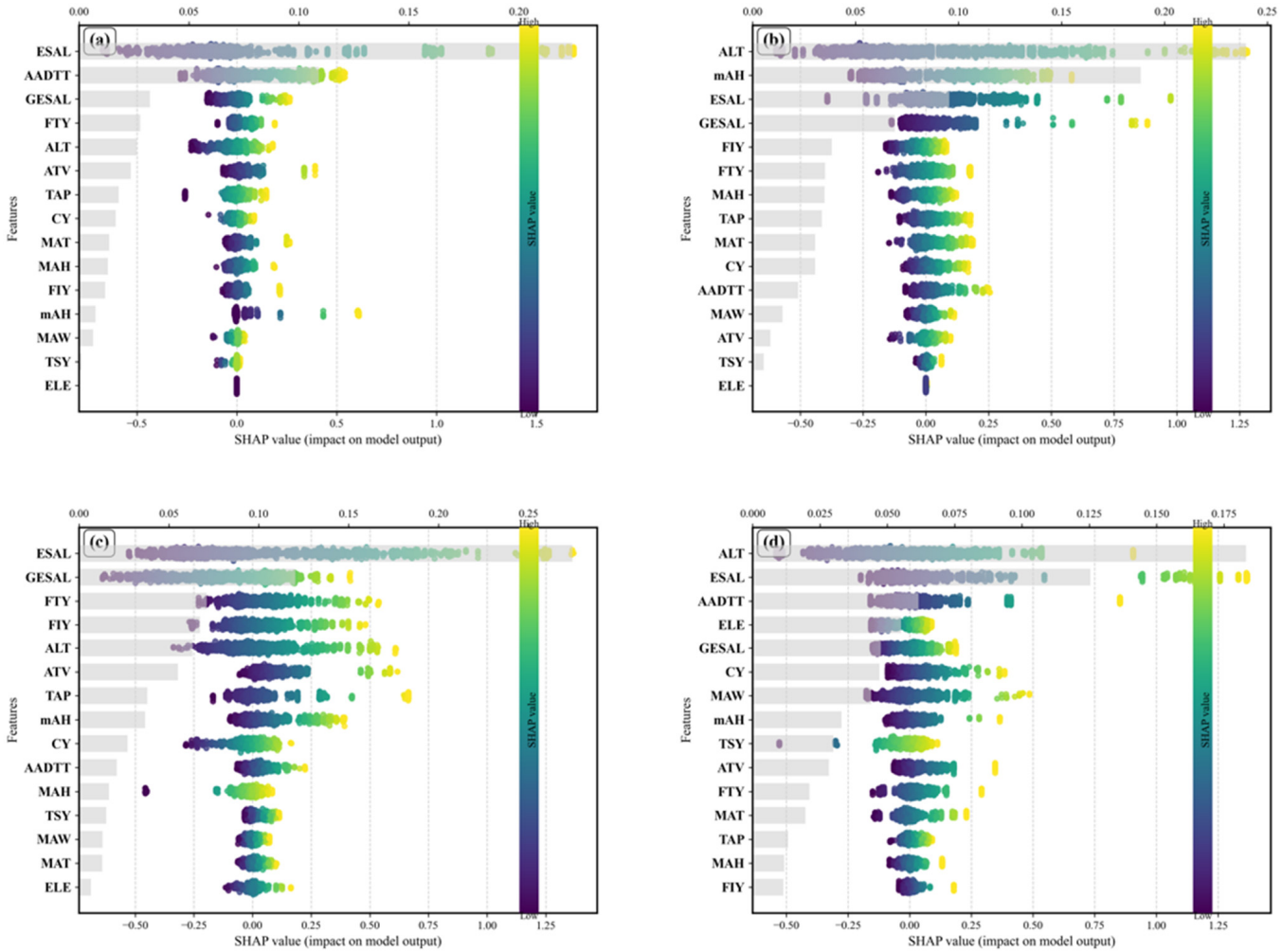


Figure 11. SHAP global analysis by pavement structure: (AC; GS+TS; GB+TB+EF; SS).

To investigate how feature contributions vary across different pavement structural configurations, we performed subgroup-level interpretability analysis using SHAP values. For each subgroup, the mean absolute SHAP value (mean |SHAP|) was computed for each feature to quantify its overall influence on model predictions. Table 11 lists the four most influential variables in each subgroup based on mean |SHAP|. The results indicate that although several core variables affect IRI predictions across all groups, their relative importance varies with pavement structural characteristics. In some configurations, traffic-related variables play a more significant role, whereas in others, structural parameters and environmental factors are more influential. These differences suggest heterogeneous mechanisms of pavement performance degradation under different structural conditions.

Table 11. Key Contribution Features |SHAP| Values in Structural Subgroups.

Dataset	Feature	Mean |SHAP| Values
AC	ESAL	0.2244
	AADTT	0.1091
	GESAL	0.0322
	FTY	0.0278
GS+TS	ALT	0.2397
	mAH	0.1885
	ESAL	0.0956
	GESAL	0.0691
GB+TB+EF	ESAL	0.2745
	GESAL	0.1204
	FTY	0.0711
	FIY	0.0674
SS	ALT	0.1832
	ESAL	0.1254
	AADTT	0.0616
	ELE	0.0551

For asphalt concrete (AC) pavements (a), traffic load dominates changes in the International Roughness Index (IRI), with equivalent single-axle load (ESAL) and average daily truck traffic trend (AADTT) having significant positive effects, while average layer thickness (ALT) and service time (CY) reflect cumulative structural degradation; the influence of climate is relatively limited. For GS+TS pavements (b), structural factors still dominate; excessively high average layer thickness (ALT) may promote reflective cracking, while cold, humid, and high-altitude environments further accelerate pavement roughness growth. For GB+TB+EF pavements (c), the traffic–environment coupling effect is significant: average layer thickness (ALT) is a key structural factor, while higher annual equivalent single-axle load (GESAL) and average daily truck traffic trend (AADTT), as well as frequent freeze–thaw cycles (FTY), exacerbate IRI degradation, whereas warm, dry environments mitigate deterioration. For SS sections (d), insufficient average layer thickness (ALT) limits the structural bearing capacity, while high annual equivalent uniaxial load (GESAL) and frequent freeze–thaw cycles (FTY) accelerate subgrade deterioration and pavement roughness evolution. Overall, pavement maintenance strategies should focus on traffic load management, structural thickness optimization, drainage improvement, and climate-adaptive antifreeze measures.

4.4. Graphical IRI Prediction Platform Using TabPFN

To facilitate practical application, we developed a lightweight graphical user interface (GUI) using Python 3.12, with the pre-trained TabPFN model serving as the core prediction

module. The interface was implemented using a standard Python GUI framework and is designed as a simple decision-support tool for researchers and pavement engineers who wish to rapidly estimate IRI values under different structural and environmental conditions without directly running the machine learning model.

This interface allows users to input key parameters related to traffic load, pavement structure characteristics, and climate conditions to quickly obtain IRI predictions. As shown in Figure 12, the GUI consists of a parameter input module and a result display module. Prior to prediction, basic input validation procedures are performed to ensure that the user-provided values fall within reasonable engineering ranges and that missing or inconsistent inputs are avoided. These checks help improve the reliability and stability of the prediction process.

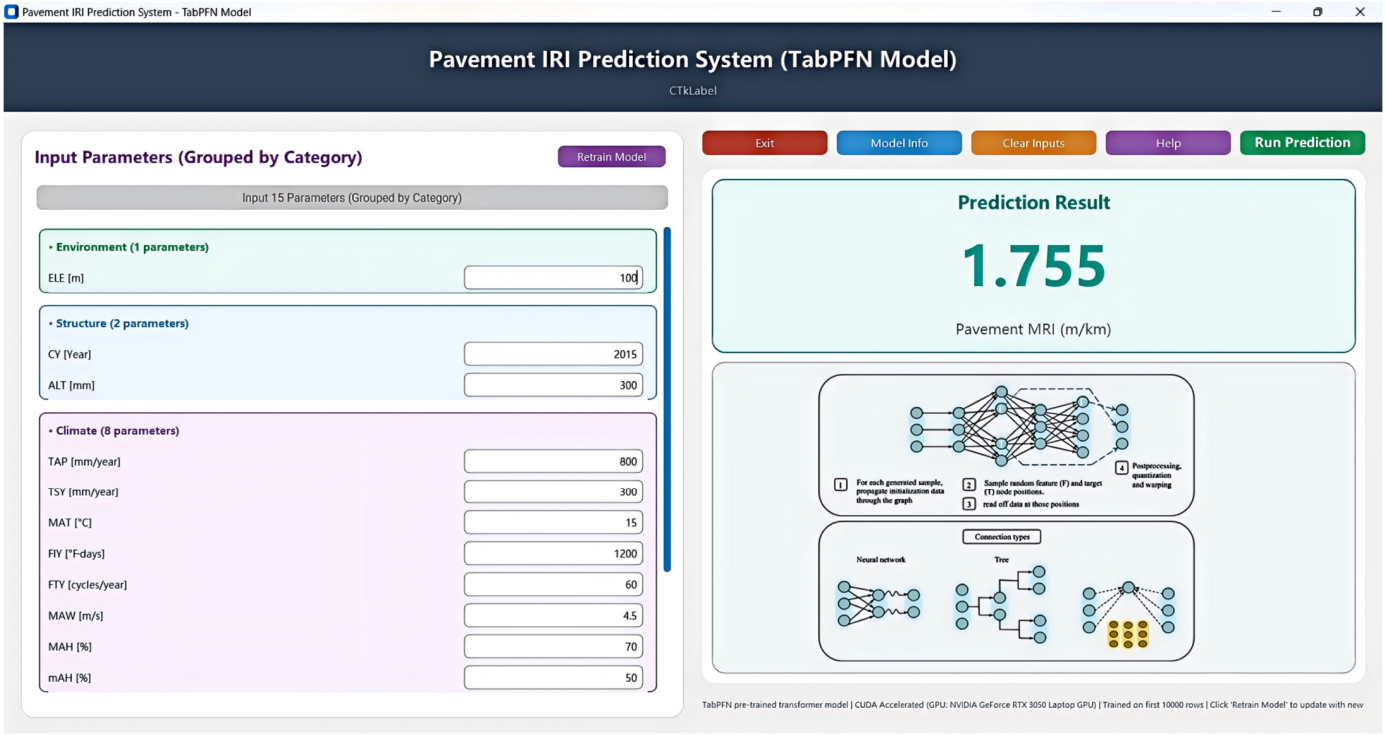


Figure 12. Application of TabPFN model in road performance prediction.

Once the required parameters are specified, users can execute the prediction by clicking the "Run Prediction" button. The predicted IRI value will then be automatically displayed in the output panel. The predictions generated through the GUI are directly produced by the trained TabPFN model used in the experimental analysis, ensuring consistency between the reported modeling results and the practical application tool.

5. Limitations and Future Research

This study introduced the prior-knowledge-driven tabular prediction model TabPFN, combined with SHAP and generalized additive models (GAMs), to establish an IRI analysis and modeling framework that balances predictive accuracy and interpretability. The framework was designed to explore the performance evolution of pavement structures under the coupled effects of traffic, structural, and climatic factors. However, due to constraints related to data conditions, model assumptions, and analytical scale, the current approach still exhibits limitations in applicability and interpretive depth. These limitations mainly involve data sources, modeling strategies, interpretability analysis, and adaptability across different pavement structures, and thus require further refinement in future work.

Limitations and Future Perspectives

From the data perspective, this study primarily relied on the North American LTPP database, whose regional coverage is limited, potentially restricting model generalization under other climatic, traffic, and pavement structure conditions. In addition, the lack of long-term continuous time-series observations constrained the representation of dynamic performance evolution, while high correlations among traffic variables may introduce training redundancy and reduce the robustness of feature attribution. Future studies could incorporate incremental learning with annual observation sequences, multi-source data fusion, and transfer learning to simulate long-term pavement evolution and improve regional adaptability.

From the modeling perspective, although TabPFN achieved high predictive accuracy, it remained sensitive to data distributions and outliers, and the absence of time-series features limited its ability to capture long-term performance trends. Future work could integrate temporal modeling techniques or dynamic updating frameworks to enhance predictions of long-term deterioration, thereby improving the reliability of mid- and long-term maintenance decisions.

SHAP values were computed using a model-agnostic Kernel SHAP interpreter, with the TabPFN model serving as the surrogate prediction function. This approach ensures consistent, model-independent interpretation of feature contributions. A background reference set randomly sampled from the training data was used to estimate Shapley values, enabling both global and local attribution analysis. However, strong multicollinearity (VIF > 15) exists among traffic-related variables, which may lead to attribution redistribution among relevant predictors, potentially affecting the stability of the SHAP interpretation. Since multicollinearity can still introduce uncertainty into feature attribution, future research could further improve interpretability by introducing multivariate interaction models, causal inference frameworks, and threshold identification methods, as well as extending the analysis to other pavement performance indicators such as rutting depth and flexural strength.

From the structural perspective, sample sizes for certain stratified pavement structure types were limited, and special structures were underrepresented, while the study mainly focused on flexible asphalt pavements and did not include rigid pavements. Future research could expand to additional pavement structure types, such as rigid and composite systems, to improve the applicability and engineering relevance of the findings.

6. Conclusions

This study utilizes the LTPP dataset to construct an interpretable machine learning framework for predicting road surface roughness (IRI) under traffic–structure–climate coupled conditions. Regarding model accuracy, this framework integrates a pre-trained tabular model, ensemble evaluation, and interpretable analysis to capture the nonlinear evolution of road surface performance. The optimized pre-trained tabular model (TabPFN) achieved the best overall predictive performance (test R^2 = 0.9474; RMSE = 0.1380). To clearly summarize the empirical results and their engineering implications, the key evidence and quantitative results of this study are presented in a structured format in Table 12.

Regarding influencing factors, SHAP analysis further quantifies the relative contributions of the main factor groups. The results show that traffic factors contribute 39.5% to IRI evolution, structural factors 35.9%, and climate factors 24.5%. Furthermore, the SHAP-GAM coupled analysis identified several nonlinear transition thresholds, including ESAL > 6.31×10^5, MAT < 16.86 °C, and ALT > 7.30 cm, beyond which the rate of pavement roughness increase accelerates. These results highlight the combined impact of traffic load, structural configuration, and climatic conditions on pavement degradation.

Table 12. Summary of Key Evidences and Quantitative Results in This Study.

Evidence Category	Key Result	Quantitative Evidence	Engineering Implication
Model Performance	TabPFN achieved best prediction accuracy	Test $R^2 = 0.9474$; RMSE = 0.1380	Reliable IRI prediction under complex conditions
Model stability	Ensemble and cross-validation improved generalization	Stable performance across 5-fold CV	Reduced overfitting and improved transferability
Factor Contribution	Traffic, structure, and climate jointly drive IRI evolution	Traffic: 39.5%; Structure: 35.9%; Climate: 24.5%	Multi-factor coupling must be considered in maintenance
GAM nonlinear threshold	Critical turning points identified	ESAL > 6.31×10^6; MAT < 16.86 °C; ALT > 7.30 cm	Supports threshold-based preventive maintenance
Pavement-Type Sensitivity	Influencing factors vary by pavement structure	AC: ALT, ESAL/AADTT; GB+TB+EF: ALT, GESAL/AADTT; GS+TS: ALT, FTY; SS: ALT, GESAL	Enables structure-specific lifecycle management

At different pavement structure levels, while the average layer thickness (ALT) affects the International Roughness Index (IRI) for all pavement types, the importance of traffic variables (such as ESAL/AADTT and GESAL) and climatic indicators (such as FTY, MAT, and MAH) varies depending on the pavement structure category (asphalt concrete, GB+TB+EF, GS+TS, and SS). The SHAP analysis confirmed these structural differences and demonstrated that predictive models for specific pavement types can improve interpretability and support targeted maintenance and lifecycle management strategies.

Author Contributions: Conceptualization, L.Q. and T.L.; methodology, L.Q.; software, L.Q.; validation, T.L. and Q.S.; formal analysis, L.Q.; investigation, Q.S.; resources, Q.S.; data curation, M.T.; writing—original draft preparation, L.Q.; writing—review and editing, T.L. and Q.S.; visualization, M.T.; supervision, T.L.; project administration, T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National College Student Innovation and Entrepreneurship Training Program (Liaoning Province), grant number 202510153016.

Data Availability Statement: The dataset used in this study was derived from the Long-Term Pavement Performance (LTPP) database. The processed dataset and the scripts used for data extraction and preprocessing are publicly available at the following GitHub repository: (https://github.com/INKED2/Experimental-data-and-code.git) (accessed on 15 March 2026). The repository includes the workflow for merging relevant LTPP tables and generating the final modeling dataset. The original raw data can be obtained from the official LTPP database, while the provided code allows full reproduction of the data preparation and analysis procedures described in this study.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (GPT-5) for the purposes of translating and polishing the text. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

ALT	AVG_LAYER_THICKNESS
IRI	International Roughness Index
ATV	ANNUAL_TRUCK_VOLUME_TREND
GESAL	ANNUAL_GESAL_TREND
TAP	TOTAL_ANN_PRECIP
MAT	MEAN_ANN_TEMP_AVG
FTY	FREEZE_THAW_YR
MAH	MAX_ANN_HUM_AVG
CY	CONSTRUCTION_ YEAR
AADTT	AADTT_ALL_TRUCKS_TREND
ESAL	ANNUAL_ESAL_TREND
TSY	TOTAL_SNOWFALL_YR
FIY	FREEZE_INDEX_YR
MAW	MEAN_ANN_WIND_AVG
mAH	MIN_MON_HUM_AVG
LTPP	Long-Term Pavement Performance Database
R^2	Coefficient of Determination
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Square Error
BO	Bayesian Optimization
SHAP	SHapley Additive exPlanations
GAM	Generalized Additive Model
GPR	Gaussian Process Regression
ANN	Artificial Neural Network
LSTM	Long Short-Term Memory network
CNN	Convolutional Neural Network
GRU	Gated Recurrent Unit
PSO	Particle Swarm Optimization
SVR	Support Vector Regression

Appendix A

Table A1. LTPP Data Keyword Merging.

Source Table	Key Fields Used for Join	Derived Variables
TRF_TREND	STATE_CODE, STATE_CODE_EXP, SHRP_ID, CONSTRUCTION_NO, YEAR	AADTT, ATV, ESAL, GESAL
CLM_VWS_PRECIP_ANNUAL	STATE_CODE, STATE_CODE_EXP, SHRP_ID, YEAR	TAP, FTY, TSY, FIY
CLM_VWS_TEMP_ANNUAL	STATE_CODE, STATE_CODE_EXP, SHRP_ID, YEAR	MAT
CLM_VWS_WIND_ANNUAL	STATE_CODE, STATE_CODE_EXP, SHRP_ID, YEAR	MAW
CLM_VWS_HUMIDITY_ANNUAL	STATE_CODE, STATE_CODE_EXP, SHRP_ID, YEAR	MAH, mAH
IRI	STATE_CODE, STATE_CODE_EXP, SHRP_ID, YEAR	IRI
CLM_OWS_LOCATION	STATE_CODE, STATE_CODE_EXP, SHRP_ID	ALT, ELE
TST_L05B	STATE_CODE, STATE_CODE_EXP, SHRP_ID	LAYER_CNT, AVG_LAYER_THICKNESS

Appendix B

The optimal hyperparameters identified in Section 4.1 were used to train the nine selected machine learning models. Figure A1 presents the learning curves of each model, which illustrate their fitting behavior. It can be observed that, as the number of iterations increased, the loss metric (RMSE) consistently decreased, indicating strong model reliability without evidence of overfitting. Notably, convergence curves are not shown for Support Vector Regression (SVR) and K-Nearest Neighbors (KNN) due to differences in their training mechanisms: SVR solves a convex optimization problem, while KNN predicts based on distance metrics. Similarly, Random Forest (RF), as an ensemble model, aggregates independently trained decision trees and does not involve iterative training, so no convergence curve is available. TabPFN and TabM, as pretrained or single-pass tabular models, also do not generate RMSE sequences suitable for plotting.

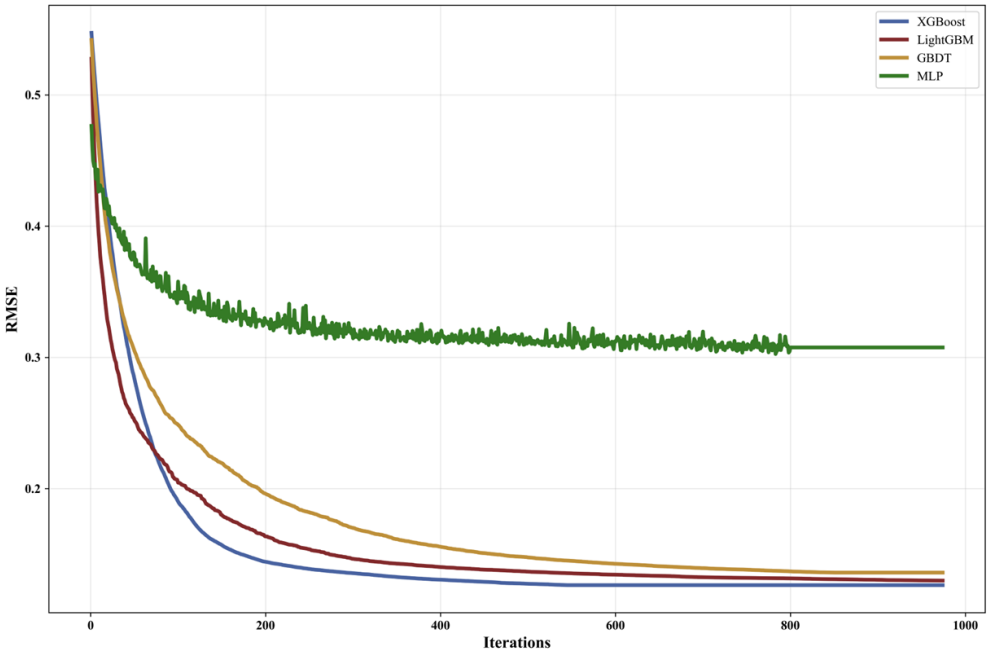


Figure A1. Bayesian hyperparameter optimization process.

Appendix C

Table A2. GAM threshold uncertainty range.

Feature	Root_Index	Threshold	CI_Lower	CI_Upper	Bootstrap_N	CI_Level
AADTT	1	204.8883	195.9284	224.6359	1000	95
AADTT	2	997.4238	975.4175	1017.9421	1000	95
AADTT	3	2679.9996	2326.5961	7664.1427	1000	95
AADTT	4	3698.8270	2699.7695	4093.3364	797	95
AADTT	5	7615.9906	5511.1347	7810.9969	797	95
ATV	1	752,043.2239	720,251.4289	1,647,631.1538	1000	95
ATV	2	993,504.3082	855,922.5474	1,485,723.5174	882	95
ATV	3	1,644,648.1935	1,059,912.8113	1,672,045.2853	882	95
ESAL	1	631,043.9718	607,211.3041	652,347.0988	1000	95
ESAL	2	1,257,653.6293	1,187,152.0557	1,249,907.8346	334	95
ESAL	3	1,273,548.0831	1,268,513.9118	1,358,175.1102	334	95

Table A2. *Cont.*

Feature	Root_Index	Threshold	CI_Lower	CI_Upper	Bootstrap_N	CI_Level
GESAL	1	243,331.6419	227,324.5664	262,316.1842	1000	95
ELE	1	149.8131	142.0602	159.6350	1000	95
ELE	2	255.7434	238.5566	299.9143	1000	95
ELE	3	336.8588	319.3378	375.6019	981	95
ELE	4	506.6202	449.5647	561.7288	981	95
TAP	1	288.1520	279.3273	301.5358	1000	95
TAP	2	684.7846	631.0919	999.6476	999	95
TAP	3	730.3740	725.6052	1119.7333	999	95
TAP	4	964.4261	914.9548	1005.2250	434	95
TAP	5	1111.4170	1093.5076	1119.8389	434	95
MAT	1	16.8558	16.7521	17.0527	1000	95
MAT	2	23.2547	22.6807	23.1803	988	95
FTY	1	42.0207	39.9520	42.4907	1000	95
MAW	1	2.8155	2.7987	2.8485	1000	95
ALT	1	7.2961	7.1690	7.4257	1000	95
MAH	1	57.0082	41.0763	94.9602	1000	95
MAH	2	58.2863	56.1962	94.7142	833	95
MAH	3	94.8817	58.2606	95.0576	792	95
mAH	1	19.2503	19.1112	19.4262	1000	95
mAH	2	26.4241	24.8695	51.2960	1000	95
mAH	3	37.7706	34.8460	38.5218	895	95
mAH	4	51.2016	50.9470	51.4338	884	95

References

1. Öztürk, İ.; Lehtonen, E.; Madigan, R.; Lee, Y.M.; Aittoniemi, E.; Merat, N. Cross-country differences in willingness to use conditionally automated driving systems: Impact of technology affinity, driving skills, and perceived traffic climate. *Technol. Soc.* **2025**, *82*, 102903. [CrossRef]
2. Cao, R.; Yao, L.; Leng, Z. Synthesising climate and traffic factors for pavement regionalisation in China. *Int. J. Pavement Eng.* **2025**, *26*, 2450098. [CrossRef]
3. Zhong, J.; Li, Y.; Bloss, W.J.; Harrison, R.M. Street-scale black carbon modelling over the West Midlands, United Kingdom: Sensitivity test of traffic emission factor adjustments. *Environ. Int.* **2025**, *196*, 109265. [CrossRef] [PubMed]
4. Gu, Z.; Peng, B.; Xin, Y. Higher traffic crash risk in extreme hot days? A spatiotemporal examination of risk factors and influencing features. *Int. J. Disaster Risk Reduct.* **2025**, *116*, 105045. [CrossRef]
5. Ghadi, M.Q. Investigating the impact of climate change on traffic accidents in Jordan. *Sustainability* **2025**, *17*, 2161. [CrossRef]
6. Ge, Y.; Zhao, H.; Liu, T. Prediction and analysis of the severity of road traffic accidents at traffic signs under rainstorm conditions. In *Proceedings Volume 13422, Proceedings of the Fourth International Conference on Intelligent Traffic Systems and Smart City (ITSSC 2024), Xi'an, China, 23–25 August 2024*; SPIE: Bellingham, WA, USA; Volume 2025, pp. 422–427.
7. Khamkhanpom, P.; Nguyen, Q.P.; Le, X.Q.; Nguyen, Q.T. Mechanistic-Empirical pavement design method and applicability in Laos. *Transp. Res. Procedia* **2025**, *85*, 18–25. [CrossRef]
8. Ahmed, T.; Isied, M.; Souliman, M.I. Leveraging physics with deep learning: Physics-informed neural networks (PINN) for IRI prediction in flexible pavements. *Can. J. Civ. Eng.* **2025**, *52*, 1885–1899. [CrossRef]
9. Kwon, K.; Yeom, Y.; Shin, Y.; Bae, A.; Choi, H. Machine learning models for predicting the International Roughness Index of asphalt concrete overlays on Portland cement concrete pavements. *Comput.-Aided Civ. Infrastruct. Eng.* **2025**, *40*, 3385–3402. [CrossRef]
10. Purwanto, A.; Tjendani, H.T.; Witjaksana, B. Analysis of Pavement Condition Using The International Roughness Index (IRI) Method With The Roadroid Application on The Genengan–Lembeyan Road Section In Magetan District. *J. Soc. Res.* **2025**, *4*, 337–346. [CrossRef]
11. Ban, I.; Bonari, J.; Paggi, M. A computational framework for evaluating tire-asphalt hysteretic friction including pavement roughness. *arXiv* **2025**, arXiv:2504.01511. [CrossRef]
12. Ali, A.A.; Hussein, A.; Heneash, U. Performance of soft computing technique in predicting the pavement international roughness index: Case study. *Int. J. Pavement Res. Technol.* **2025**, *18*, 346–364. [CrossRef]

13. Al-Mahamid, H.; Al-Nabulsi, D.; Torok, A. Developing safety performance functions incorporating pavement roughness using Poisson regression and Machine learning models on Jordan's Desert Highway. *Transp. Res. Interdiscip. Perspect.* **2025**, *34*, 101659. [CrossRef]
14. Jagadeesh, A.; Premarathna, W.; Kumar, A.; Kasbergen, C.; Erkens, S. Finite element modelling of jointed plain concrete pavements under rolling forklift tire. *Eng. Struct.* **2025**, *328*, 119705. [CrossRef]
15. Chau, A.D.; Hoang, H.T.; Nguyen, L.D. Web-Based Decision Support System for Automated Pavement Design and Life-Cycle Cost Analysis. *J. Constr. Eng. Manag.* **2025**, *151*, 05025004. [CrossRef]
16. Ayesh, D.; Ruixuan, Z.; Yilin, G.; Jinjiang, Z.; Chaminda, G. Deploying machine learning for long-term road pavement moisture prediction: A case study from Queensland, Australia. *J. Road Eng.* **2025**, *5*, 184–201. [CrossRef]
17. Rizelioğlu, M. An extensive bibliometric analysis of pavement deterioration detection using sensors and machine learning: Trends, innovations, and future directions. *Alex. Eng. J.* **2025**, *112*, 349–366. [CrossRef]
18. Wang, S.; Xia, P.; Gong, F.; Liu, J.; Huang, B.-T. Incremental Update Framework for Multi-Source Low Carbon Concrete Prediction Based on Boctgan-Adaptive Weight and Experience Replay. 2025. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5138018 (accessed on 5 March 2026).
19. Rajan, K.; Aryal, M.; Sharma, K.; Bhandary, N.P.; Pokhrel, R.; Acharya, I.P. Development of a framework for the prediction of slope stability using machine learning paradigms. *Nat. Hazards* **2025**, *121*, 83–107. [CrossRef]
20. Sanij, H.K.; Babagoli, R.; Elyasi, R.M. Enhancing stone matrix asphalt performance with sugarcane bagasse ash: Mechanical properties and machine learning-based predictions using XGBoost and random forest. *Case Stud. Constr. Mater.* **2025**, *23*, e05186. [CrossRef]
21. Zhong, C.; Qian, G.; Gong, X.; Yu, H.; Jun, C.; Zhong, Y.; Ma, J.; Gu, H. Study on mesostructure evolution behavior of asphalt mixture compaction process based on deep learning image processing. *Constr. Build. Mater.* **2025**, *458*, 139650. [CrossRef]
22. Liu, J.; Cheng, C.; Wang, Z.; Yang, S.; Wang, L. Intelligent Asphalt Mixture Design: A Combined Supervised Machine Learning and Deep Reinforcement Learning Approach. *Transp. Res. Rec.* **2025**, *2679*, 03611981251320382. [CrossRef]
23. Golanbari, B.; Mardani, A.; Farhadi, N.; Nazari Chamki, A. Applications of machine learning in predicting rut depth in off-road environments. *Sci. Rep.* **2025**, *15*, 5486. [CrossRef] [PubMed]
24. Grube, G.; Grigolato, S.; Ala-Ilomäki, J.; Routa, J.; Lindeman, H.; Astrup, R.; Talbot, B. Modelling machine-induced soil deformation in forest soils using stump proximity and machine learning. *Biosyst. Eng.* **2025**, *258*, 104255. [CrossRef]
25. Jalota, S.; Suthar, M. Machine learning approach for estimating durability of modified bituminous mixes under wet-dry cycle condition. *Eur. J. Environ. Civ. Eng.* **2025**, *29*, 2818–2844. [CrossRef]
26. Al-Khateeb, G.; Alnaqbi, A.; Zeiada, W. Predictive modeling of punchouts in continuously reinforced concrete pavement: A machine learning approach. *AI Civ. Eng.* **2025**, *4*, 15. [CrossRef]
27. Alnaqbi, A.; Zeiada, W.; Al-Khateeb, G. A hybrid machine learning method of support vector regression with particle swarm optimization for predicting IRI in continuously reinforced concrete pavement. *J. Eng. Appl. Sci.* **2025**, *72*, 128. [CrossRef]
28. Xu, C.; Chen, X.; Zeng, Q.; Cai, M.; Zhang, W.; Yu, B. A framework of integrating machine learning model and pavement life cycle assessment to optimize asphalt mixture design. *Constr. Build. Mater.* **2025**, *469*, 140481. [CrossRef]
29. Sharma, A.; Sachdeva, S.N.; Aggarwal, P. Predicting IRI using machine learning techniques. *Int. J. Pavement Res. Technol.* **2023**, *16*, 128–137. [CrossRef]
30. Rajender, A.; Samanta, A.K.; Paral, A. Comparative study of corrosion-based service life prediction of reinforced concrete structures using traditional and machine learning approach. *Int. J. Struct. Integr.* **2025**, *16*, 591–621. [CrossRef]
31. Lashkov, I.; Yuan, R.; Zhang, G. Machine learning-based vehicle detection and tracking based on headlight extraction and GMM clustering under low illumination conditions. *Expert Syst. Appl.* **2025**, *267*, 126240. [CrossRef]
32. Zhou, G.; Gao, H.; Cai, Y.; Guo, J.; Zhao, X. A Filter Method for Vehicle-Based Moving LiDAR Point Cloud Data for Removing IRI-Insensitive Components of Longitudinal Profile. *Remote Sens.* **2026**, *18*, 240. [CrossRef]
33. A Ilemobayo, J.; Durodola, O.; Alade, O.; JAwotunde, O.; TOlanrewaju, A.; Falana, O.; Ogungbire, A.; Osinuga, A.; Ogunbiyi, D.; Ifeanyi, A.; et al. Hyperparameter tuning in machine learning: A comprehensive review. *J. Eng. Res. Rep.* **2024**, *26*, 388–395. [CrossRef]
34. Morales-Hernández, A.; Van Nieuwenhuyse, I.; Rojas Gonzalez, S. A survey on multi-objective hyperparameter optimization algorithms for machine learning. *Artif. Intell. Rev.* **2023**, *56*, 8043–8093. [CrossRef]
35. Hanifi, S.; Cammarono, A.; Zare-Behtash, H. Advanced hyperparameter optimization of deep learning models for wind power prediction. *Renew. Energy* **2024**, *221*, 119700. [CrossRef]
36. Shu, J.; Meng, D.; Xu, Z. Learning an explicit hyper-parameter prediction function conditioned on tasks. *J. Mach. Learn. Res.* **2023**, *24*, 1–74.
37. Elghaish, F.; Matarneh, S.; Abdellatef, E.; Rahimian, F.; Hosseini, M.R.; Farouk Kineber, A. Multi-layers deep learning model with feature selection for automated detection and classification of highway pavement cracks. *Smart Sustain. Built Environ.* **2025**, *14*, 511–535. [CrossRef]

38. Chen, K.; Torbaghan, M.E.; Thom, N.; Garcia-Hernández, A.; Faramarzi, A.; Chapman, D. A Machine Learning based approach to predict road rutting considering uncertainty. *Case Stud. Constr. Mater.* **2024**, *20*, e03186. [CrossRef]
39. Haoran, Z.; Zidong, Z.; Min, W.; Xin, Y.; Yongxin, W.; Chen, C.; Jun, Q. Integrated design optimization method for pavement structure and materials based on further development of finite element and particle swarm optimization algorithm. *Constr. Build. Mater.* **2024**, *426*, 136080. [CrossRef]
40. Xiao, M.; Luo, R.; Chen, Y.; Ge, X. Prediction model of asphalt pavement functional and structural performance using PSO-BPNN algorithm. *Constr. Build. Mater.* **2023**, *407*, 133534. [CrossRef]
41. Chiou, Y.-S.; Ho, M.-C.; Song, P.-Y.; Lin, J.-D.; Lu, S.-H.; Ke, C.-Y. A Study on the Application of Genetic Algorithms to the Optimization of Road Maintenance Strategies. *Appl. Sci.* **2025**, *15*, 10094. [CrossRef]
42. Altarabsheh, A.; Altarabsheh, I.; Ventresca, M. A hybrid genetic algorithm to maintain road networks using reliability theory. *Struct. Infrastruct. Eng.* **2023**, *19*, 810–823. [CrossRef]
43. Li, H.; Zhang, J.; Yang, X.; Ye, M.; Jiang, W.; Gong, J.; Tian, Y.; Zhao, L.; Wang, W.; Xu, Z. Bayesian optimization based extreme gradient boosting and GPR time-frequency features for the recognition of moisture damage in asphalt pavement. *Constr. Build. Mater.* **2024**, *434*, 136675. [CrossRef]
44. Sheikh, I.R.; Ming, Z.; Xiaohui, S.; Changqing, C.; Xiangsheng, C.; Zijun, D.; Foci, C. Interpretable machine learning framework for resilient modulus estimation using LTPP data for pavements. *Case Stud. Constr. Mater.* **2025**, *23*, e05403. [CrossRef]
45. Shafiee, M.; Fattahi, M.; Roshani, E.; Popov, P. Enhanced prediction of urban road pavement performance under climate change with machine learning. *J. Civ. Eng. Constr.* **2024**, *13*, 159–169. [CrossRef]
46. Fares, A.; Abdelkader, E.M.; Faris, N.; Zayed, T. Using machine learning for road performance modelling and influential factors investigation. *Int. J. Struct. Civil Eng. Res.* **2023**, *12*, 154–159. [CrossRef]
47. Breejen, F.d.; Bae, S.; Cha, S.; Yun, S.-Y. Fine-tuned In-Context Learning Transformers are Excellent Tabular Data Classifiers. *arXiv* **2024**, arXiv:2405.13396.
48. Gardner, J.; Perdomo, J.C.; Schmidt, L. Large scale transfer learning for tabular data via language modeling. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 45155–45205.
49. Zhang, X.; Wang, Z.; Jiang, L.; Gao, W.; Wang, P.; Liu, K. TFWT: Tabular Feature Weighting with Transformer. *arXiv* **2024**, arXiv:2405.08403. [CrossRef]
50. Peroni, M.; Le, F.; Sheinin, V. Robust Tabular Foundation Models. *arXiv* **2025**, arXiv:2512.03307.
51. Sharma, A.; Aggarwal, P. IRI prediction using machine learning models. *WSEAS Trans. Comput. Res.* **2023**, *11*, 111–116. [CrossRef]
52. Khalifah, R.; Isied, M.; Souliman, M. Asphalt pavement bleeding prediction model using LTPP database. In *Bituminous Mixtures and Pavements VIII*; CRC Press: Boca Raton, FL, USA, 2024; pp. 879–885.
53. Chen, L.; Li, H.; Wang, S.; Shan, F.; Han, Y.; Zhong, G. Imporved model for pavement performance prediction based on recurrent neural network using LTPP database. *Int. J. Transp. Sci. Technol.* **2024**, *19*, 128–138. [CrossRef]
54. Song, Y.; Wang, Y.D.; Hu, X.; Liu, J. An efficient and explainable ensemble learning model for asphalt pavement condition prediction based on LTPP dataset. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 22084–22093. [CrossRef]
55. Ahmed, T.; Isied, M.; Souliman, M. Enhancing road safety: Developing a neural network-based model for predicting skid resistance in asphalt pavements using LTPP data. *Discover Civ. Eng.* **2025**, *2*, 226. [CrossRef]
56. Suliman, A.M.; Awed, A.M.; Abd El-Hakim, R.T.; El-Badawy, S.M. International roughness index prediction for jointed plain concrete pavements using regression and machine learning techniques. *Transp. Res. Rec.* **2024**, *2678*, 235–250. [CrossRef]
57. Elkins, G.E.; Ostrom, B. *Long-Term Pavement Performance Information Management System User Guide*; U.S. Department of Transportation, Federal Highway Administration: Washington, DC, USA, 2021.
58. Wei, J.; Shen, T.; Wang, K.; Liu, J.; Wang, S.; Hu, W. Transfer learning framework for the wind pressure prediction of high-rise building surfaces using wind tunnel experiments and machine learning. *Build. Environ.* **2025**, *271*, 112620. [CrossRef]
59. Zaidi, S.A.; Chouvatut, V.; Phongnarisorn, C.; Praserttitipong, D. Deep learning based detection of endometriosis lesions in laparoscopic images with 5-fold cross-validation. *Intell.-Based Med.* **2025**, *11*, 100230. [CrossRef]
60. Dada, B.A.; Nwulu, N.I.; Olukanmi, S.O. Bayesian Optimization with Optuna for Enhanced Soil Nutrient Prediction: A Comparative Study with Genetic Algorithm and Particle Swarm Optimization. *Smart Agric. Technol.* **2025**, *12*, 101136. [CrossRef]
61. Hu, H.; Xia, X.; Jian, L.; Zhang, Y.; Tang, Y.; Wang, L.; Liao, Y.; Pan, Z. *Analysis of the Impact of Street Physical Morphology on Thermal Environment Based on GBDT-SHAP Machine Learning Model: A Case Study of the Five Central Districts of Chengdu*; Research Square: Asheville, NC, USA, 2025.
62. Ye, H.-J.; Liu, S.-Y.; Chao, W.-L. A closer look at tabpfn v2: Strength, limitation, and extension. *arXiv* **2025**, arXiv:2502.17361. [CrossRef]
63. Qu, J.; HolzmÃžller, D.; Varoquaux, G.; Morvan, M.L. Tabicl: A tabular foundation model for in-context learning on large data. *arXiv* **2025**, arXiv:2502.05564.
64. Plati, C.; Armeni, A.; Kyriakou, C.; Asoniti, D. AI for Predicting Pavement Roughness in Road Monitoring and Maintenance. *Infrastructures* **2025**, *10*, 157. [CrossRef]

65. Arshad, M.; Hamza, H.M. Applications of machine learning and traditional prediction techniques for estimating resilient modulus values of unbound granular materials incorporating reclaimed asphalt pavement (RAP) as a primary component. *Constr. Build. Mater.* **2025**, *473*, 140900. [CrossRef]
66. Zhu, J.; Yin, Y.; Ma, T.; Wang, D. A novel maintenance decision model for asphalt pavement considering crack causes based on random forest and XGBoost. *Constr. Build. Mater.* **2025**, *477*, 140610. [CrossRef]
67. Luo, C.; Zhu, S.-P.; Keshtegar, B.; Niu, X.; Taylan, O. An enhanced uniform simulation approach coupled with SVR for efficient structural reliability analysis. *Reliab. Eng. Syst. Saf.* **2023**, *237*, 109377. [CrossRef]
68. Khan, A.; Zhang, W.; Wu, Y.; She, X.; Jiang, X.; Huang, W.; Wang, H.; Bao, W.; Feng, C. Prediction of pavement maintenance quality and performance indicators using particle swarm optimized gradient boosting decision trees. *Int. J. Pavement Eng.* **2025**, *26*, 2595129. [CrossRef]
69. Yamany, M.S.; Elshaboury, N.; Abdelaty, A.; Smadi, O.; Ksaibati, K. Pavement Roughness Prediction on Local Roads: Machine Learning Models and Classification Granularity. *Int. J. Pavement Res. Technol.* **2025**, *18*, 1–20. [CrossRef]
70. Zhang, Y.; Peng, J.; Wang, Z.; Xi, M.; Liu, J.; Xu, L. Machine learning-assisted sustainable mix design of waste glass powder concrete with strength–cost–CO2 emissions trade-offs. *Buildings* **2025**, *15*, 2640. [CrossRef]
71. Jamil, M.H.; Jagirdar, R.; Kashem, A.; Ali, M.N.; Deb, D. Modeling of Marshall Stability of plastic-reinforced asphalt concrete using machine learning algorithms and SHAP. *Hybrid Adv.* **2025**, *10*, 100483. [CrossRef]
72. Bordt, S.; von Luxburg, U. From Shapley values to generalized additive models and back. In Proceedings of the International Conference on Artificial Intelligence and Statistics, Valencia, Spain, 2–4 May 2024; PMLR: 2023, pp. 709–745.
73. Xu, Q.; Li, J.; Fan, Y.; Gao, Z.; Wang, Z.; Xu, L.; Wang, S.; Liu, J. Intelligent prediction framework for axial compressive capacity of FRP-RACFST columns. *Mater. Today Commun.* **2024**, *41*, 110999. [CrossRef]

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.