KSCE Journal of Civil Engineering
DOI 10.1007/s12205-024-0093-9

pISSN 1226-7988, eISSN 1976-3808
www.springer.com/12205

Highway Engineering

Influence Analysis of Pavement Distress on International Roughness Index Using Machine Learning

Kibeom Kwon[a], **Hangseok Choi**[a], **Khanh Pham**[b,c], **Sangwoo Kim**[a], **and Abraham Bae**[d]

[a]*School of Civil, Environmental and Architectural Engineering, Korea University, Seoul 02841, Korea*
[b]*School of Civil Engineering and Management, International University, Ho Chi Minh 700000, Vietnam*
[c]*Vietnam National University, Ho Chi Minh 700000, Vietnam*
[d]*Future and Fusion Lab of Architectural, Civil, and Environmental Engineering, Korea University, Seoul 02841, Korea*

ARTICLE HISTORY

Received 17 January 2024
Revised 18 April 2024
Accepted 13 May 2024
Published Online 13 July 2024

KEYWORDS

Influence analysis
International roughness index
Pavement distress
Severity
Interpretable machine learning

ABSTRACT

The International Roughness Index (IRI) is closely related to pavement distress. However, previous studies employing statistics and machine learning approaches would present challenges in comprehensively analyzing the influence of pavement distress on IRI considering their severities. This study introduces interpretable machine learning to investigate the influence of pavement distress on IRI, with a particular focus on the severity of pavement distress. The pavement distress and IRI data for flexible pavements obtained from the long-term pavement performance (LTPP) program were meticulously preprocessed. The developed random forest (RF) model demonstrated satisfactory predictive performance, with an RMSE of 0.2191 and an R^2 of 0.7874. The relationship between pavement distress and IRI, as captured by the developed model, was further analyzed using the SHapley Additive exPlanations (SHAP) method. The model interpretation identified the transverse crack, rutting, and alligator crack as the key factors influencing IRI. Notably, both transverse and alligator cracks exhibited significant contributions to IRI increment at medium and high severity levels, highlighting the importance of proactive maintenance for these distress types at lower severity levels. Additionally, a threshold in rutting depth was observed, which could increase IRI. A comparative analysis with the AASHTO MEPDG smoothness model demonstrated that the predictive performance of the RF model was notably superior.

1. Introduction

The International Roughness Index (IRI)–a primary criterion for evaluating pavement quality–quantifies pavement smoothness by analyzing the longitudinal profile of the pavement surface. Longitudinal elevation data is collected at 25 cm intervals using an inertial profiler mounted on a test vehicle. This profile data is then processed using an IRI filter, which accentuates specific wavelengths perceived by a standard passenger-type vehicle traveling at 80 km/h (Sayers, 1986). Within the IRI filter's sensitive range, elevations corresponding to the wavelengths ranging from approximately 1.5 – 3.0 m/cycle and 8 – 20 m/cycle are amplified, while those outside these ranges are attenuated. Consequently, the IRI tends to increase with encountering pavement distress within these amplified wavelength ranges.

The IRI and pavement distress exhibit two common characteristics. First, they manifest physically on the pavement surface, making them both valuable pavement performance measures. Second, they result from the combined interaction of four major pavement design factors: material properties, climatic conditions, traffic volume, and structural design features of the pavement. For assessing pavement performance, there is no mechanical basis for predicting the IRI corresponding to the design factors, even though the IRI serves as a concise index, reflecting ride quality and pavement serviceability. By contrast, pavement distress can be a weighted combination of the design factors with explicit grounds, which allows engineers to effectively identify the primary cause of pavement distress, thereby improving pavement performance.

Four pavement design factors are often used as predictor

CORRESPONDENCE Abraham Bae abebae@korea.ac.kr Future and Fusion Lab of Architectural, Civil and Environmental Engineering, Korea University, Seoul 02841, Korea

variables for IRI prediction. However, interpreting IRI prediction results in terms of numerous pavement design input variables presents challenges. Despite the widespread use of design input data in the mechanistic–empirical pavement design guide (MEPDG) provided by the American Association of State Highway and Transportation Officials (AASHTO), model development for IRI prediction has primarily focused on three critical pavement distress types: transverse crack, rutting, and fatigue cracks (AASHTO, 2008; AASHTO, 2015; AASHTO, 2020). The underlying principle behind this pavement distress-based approach is that these physical manifestations on the pavement surface can establish a more direct and evident relationship with the IRI. Although an empirical statistical linear regression methodology is utilized in AASHTO, several other distress types—in addition to the major ones—have yet to be accurately captured by this linear relationship, despite their significance. Furthermore, their interpretation is expected to be complicated during model formulation. A machine learning approach could offer a potential solution for developing an IRI prediction model that considers the complexities of diverse pavement distresses.

This study introduces interpretable machine learning to perform the influence analysis of pavement distress on IRI. The IRI prediction models that employ two machine learning algorithms: random forest (RF) and extreme gradient boosting (XGB) was developed. The prediction results were interpreted using the Shapley additive explanation (SHAP), thereby demonstrating the detailed influences of the pavement distresses on the IRI with respect to their severities.

2. Related Works

2.1 Statistics Modeling for IRI Prediction

Rougher pavement surfaces exhibit a positive correlation with higher IRI values. In particular, when the roughness falls within the sensitive wavelengths as previously discussed, it may result in a substantial increase in IRI. Given that pavement distress directly contributes to the roughness of pavement surfaces, numerous studies have concentrated on IRI prediction by utilizing pavement distress information. Many of these predictions have been made using statistical regression models falling within the category of statistical approaches.

Paterson (1989) pioneered the development of an IRI incremental model that incorporated various pavement distresses, including crack, rutting, patching, and pothole. In addition to these pavement distress variables, factors such as pavement structural number, pavement age, and traffic volume were also considered. The empirical model achieved the coefficient of determination (R^2) values ranging from 0.55 – 0.59. Bashar and Darter (1995) aimed to establish a quantitative relationship between the pavement service rate (PSR) and pavement distress, attempting to correlate PSR with IRI; specifically, assuming a high severity for the distress. Mactutis et al. (2000) associated IRI with two primary distresses (fatigue crack and rutting) and incorporated the initial IRI as a significant variable for IRI prediction. Taking Indian roads as an example, Sandra and Sarkar (2013) utilized four pavement distresses (rutting, crack, pothole, and raveling) and a single maintenance variable (patching) to derive a linear regression model for IRI prediction, categorizing distress severity into three levels. For similar variables without severity categorization on different Indian highways, Chandra et al. (2013) derived linear and nonlinear IRI prediction models. Abdelaziz et al. (2020) developed a linear regression model for IRI prediction using the major pavement distresses, initial IRI, and pavement age obtained from the long-term pavement performance (LTPP) program. This model accounted for both new pavements and pavements after overlaying, yielding an R^2 of 0.57. Focusing specifically on pavement distress in urban road pavements, Al-Mansour and Shokri (2022) derived linear regression models for IRI prediction, achieving R^2 ranging from 0.33 to 0.40. Moreover, comprehensive IRI regression models were developed during the establishment of MEPDG in 2004 (ARA, 2004). These models have been updated multiple times (AASHTO, 2008; AASHTO, 2015; AASHTO, 2020). The AASHTO models primarily focused on three major pavement distresses (fatigue cracks, transverse crack, and rutting), while also incorporating variables such as the initial IRI, pavement age, and other site-specific factors in model development.

The previous studies using the statistical approach effectively analyzed IRI concerning pavement distress. Nevertheless, these studies primarily focused on a restricted range of distress types within the local context. For example, the AASHTO models may face challenges in adequately addressing the implications of distress severity.

2.2 Machine Learning for IRI Prediction

With the advancement of computer technologies, researchers have extensively employed machine learning algorithms, including various innovative approaches such as artificial neural networks (ANNs), support vector machines (SVMs), and random forest (RF), to develop IRI prediction models for flexible pavements. Many of these studies have relied on extensive instances extracted from the LTPP database.

A first subset of these studies, focusing on the direct relationship between pavement distress and the longitudinal profile roughness of road surfaces, used pavement distress as the predictor variable. Consequently, this approach excluded pavement design factors such as pavement structural features, climatic conditions, material properties, and traffic volume, as pavement distress was considered the combined outcome of these design factors. In a pioneering study, Lin et al. (2003) selected 13 types of pavement distresses and utilized a three-layer ANN model for IRI prediction. Chandra et al. (2013) and Abdelaziz et al. (2020) compared the performance of their ANN models with those of conventional regression models based on the same database and concluded that the ANN models outperformed the conventional regression models in terms of IRI prediction accuracy. These preceding studies successfully demonstrated the applicability of machine learning techniques in IRI prediction based on pavement distress. However, a more detailed interpretation of the individual contribution of pavement

distresses to IRI is still necessary, focusing on pavement distress severity as speculated in this study.

A second subset of studies (Kargah-Ostadi and Stoffels, 2015; Ziari et al., 2016; Sollazzo et al., 2017; Zeiada et al., 2019; Marcelino et al., 2021) focused on pavement management systems at the network level, adopting IRI as a comprehensive measure of overall pavement serviceability. In these studies, pavement distress was excluded as a predictor variable because each specific pavement distress was considered as a component contributing to the overall pavement performance. By adopting IRI as the pavement serviceability measure, the impact of multiple types of pavement distress could be captured in an integrated manner. Consequently, the focus of these studies was on pavement design factors, including traffic volume, climatic conditions, and material properties.

A third set of studies pursued a more comprehensive approach by incorporating both pavement distresses and the four pavement design factors into the model development framework (Gong et al., 2018; Zhang et al., 2020; Damirchilo et al., 2021; Kaloop et al., 2022). While these studies utilized all the available data to build machine-learning models, only a few pavement design factors were selected for model development. In other words, a limited portion of the available data was utilized to construct prediction models rather than comprehensively using the entire variables.

The previous machine learning-based studies concerning pavement distress would focus on the overall IRI prediction performance, not the impact of each pavement distress with respect to their severities despite its importance. This study introduces interpretable machine learning to investigate the effect of various pavement distresses on IRI further at a detailed level that the empirical AASHTO model has performed on three major pavement distresses.

3. Database Construction

3.1 LTPP Program

3.1.1 Experiment Sections

The LTPP program was initiated in 1987 as part of the Strategic Highway Research Program (SHRP) and has been under the administration of the Federal Highway Administration (FHWA) since 1992. Over the years, this program has collected a substantial volume of pavement-related data, which has been organized into four main categories and drawn from 2,509 pavement sections, starting in 1989.

The LTPP program comprised two sets of experiments: General Pavement Studies (GPSs) and Specific Pavement Studies (SPSs). The GPS test sections were established on pre-existing pavements, totaling nearly 800 sections in active road service. In contrast, the SPS experiments involved the construction and evaluation of approximately 1,600 new test sections. The GPSs and SPSs were further divided into 20 and 16 subprojects, respectively. Each subproject was designed to assess pavement performance based on specific factors such as pavement types (e.g., flexible and rigid pavements), maintenance types, overlay rehabilitation, and other relevant variables.

3.1.2 Pavement Performance

The LTPP program collected pavement performance data, including pavement distresses and IRI. The designated GPS and SPS test sections covered a span of 152 m. Data collection commenced from the date when traffic was first allowed on these test sections. Regular field visits were conducted as part of these data collection efforts, enabling continuous monitoring and assessment of pavement performance over time, accounting for different traffic loads and environmental conditions.

Pavement distresses, including alligator crack, transverse crack, longitudinal crack, and pothole, could be evaluated through two methods. The first method involves engineers conducting distress surveys during field visits. The second method entails distress evaluation via images with a dimension of 35 mm, which captures the condition of pavement surfaces. Regardless of the method employed, the criteria for distress evaluation are based on the Distress Identification Manual for the LTPP Project (Miller and Bellinger, 2003). Rutting is measured using a 1.4 m straight edge on the wheel path or by digitizing the transverse profile of the road to calculate the maximum depth. IRI values are computed and recorded based on a longitudinal profile collected at 25 mm intervals. These profiles are obtained by profiling the surface of the test section using an inertial profiler mounted on a vehicle and then filtering the data with a moving average of 0.305 m.

3.2. Data Preprocessing

3.2.1 LTPP Experimental Study Selection

The objective of this study is to demonstrate the influence of pavement distress on IRI, considering their severities by introducing interpretable machine learning. Accordingly, this study developed a machine learning model to predict IRI considering pavement distresses (Section 4) and analyzed their influences on IRI through a model interpretation method (Section 5).

This study performed data preprocessing on the data obtained from the LTPP program to construct the database for model development. The data were extracted from specific sections of the LTPP experimental study, namely GPS-1, GPS-2, and SPS-1, which were specifically associated with flexible pavements. GPS-1 and GPS-2 represent pre-existing pavement sections constructed with asphalt on granular and bound bases, respectively. On the other hand, SPS-1 represents new pavement sections designed to examine the effects of structural factors on flexible pavements. However, certain sections corresponding to these studies were excluded from the analysis for various reasons, including incomplete distress data (e.g., missing rutting data), unrealistic initial IRI after extrapolation, as discussed later in Section 3.2.3, and insufficient data. Table 1 presents details on the number of pavement sections analyzed and the corresponding instances used in each experimental study.

Table 1. LTPP Experimental Studies for Analysis

Study	Pavement type	Number of sections		Number of instances	
		LTPP program	This study	LTPP program	This study
GPS-1	Asphalt concrete on granular base	216	127	4,425	581
GPS-2	Asphalt concrete on bound base	144	60	2,400	247
SPS-1	Strategic study of the structural factors of flexible pavements	227	190	5,844	1,860
Total		587	377	12,669	2,688

Table 2. Data Collected from LTPP Experimental Studies

Data	LTPP source data name	Unit	Severity	Number of variables
Rutting	MAX_MEAN_DEPTH_1_8	mm	-	1
Alligator crack	GATOR_CRACK	m^2	L, M, H	3
Longitudinal crack in the wheel path	LONG_CRACK_WP	m	L, M, H	3
Longitudinal crack in the non-wheel path	LONG_CRACK_NWP	m	L, M, H	3
Transverse crack	TRANS_CRACK	m	L, M, H	3
Pothole	POTHOLES	m^2	L, M, H	3
Shoving	SHOVING	m^2	-	1
Bleeding	BLEEDING	m^2	-	1
Polished aggregates	POLISH_AGG	m^2	-	1
Raveling	RAVELING	m^2	-	1
Pumping	PUMPING	m	-	1
Initial IRI	MRI	m/km	-	1
Pavement age	CONSTRUCTION_DATE	day	-	1

3.2.2 Variable Selection

In the LTPP database, IRI is recorded for both the right and left wheel paths during each site visit. The average IRI value for both wheel paths is determined as the mean roughness index (MRI). In this study, the MRI value was used as the target variable to represent each IRI.

Meanwhile, this study considered 11 pavement distress types as predictor variables. Alligator crack, transverse crack, longitudinal crack in both the wheel and non-wheel paths, and pothole were categorized into three severity levels, namely low [L], medium [M], and high [H]. These severity levels were incorporated into the model development. For instance, a alligator crack with medium severity was labelled as GATOR_CRACK_A_M, where [A] denoted the area unit and [M] indicated medium severity.

Two predictor variables unrelated to pavement distress, namely the initial IRI and pavement age, were included in the modelling process. However, these variables were not directly recorded in the LTPP database. Therefore, the initial IRI was estimated by extrapolating the recorded IRI values over the service period (refer to Section 3.2.3), while the pavement age was computed using construction date information. A total of 23 predictor variables, consisting of 21 pavement distress variables and two non-distress variables (i.e., initial IRI, and pavement age), were adopted in this study and summarized in Table 2.

3.2.3 Data Interpolation and Extrapolation

The dates of pavement distress and IRI data collection may not always coincide or be simultaneously recorded for a specific section. To maintain the consistency of the utilized database, pavement distress data that were not synchronized with IRI data were generated through linear interpolation based on the IRI data, which was recorded immediately before and after recoding pavement distress. However, no extrapolation of pavement distress data was carried out to ensure data reliability. In these cases, the final IRI values were removed from the analysis.

The initial IRI is a critical variable that significantly impacts long-term pavement performance. Although it is widely recognized that pavements with lower initial IRIs tend to have longer service periods, only limited initial IRI data were recorded in the LTPP database. In GPSs, a substantial time gap of at least five years typically exists between the construction date and the initial IRI recording. Therefore, the absence of initial IRI data poses challenges for model development. Fortunately, GPSs generally encompass sufficiently extended service periods, making them suitable and reliable for extrapolating the initial IRI. Accordingly, the initial IRI values for all GPS sections were extrapolated using an exponential relationship. This study considered pavement sections with an initial IRI value greater than 0.6 m/km, representing a realistic minimum initial IRI threshold. This threshold was applied to filter out unrealistic or erroneous estimates for the initial IRI (Perera and Kohn, 2001). For certain sections, the extrapolated initial IRI value exceeded the IRI value during the service period. In such cases, these sections were excluded from the analysis, even though the initial IRI satisfied the minimum requirement of

Table 3. Statistical Description of the Adopted Variables

No	Variable	Unit	Min	Q1	Median	Q3	Max	COV
1	IRI	m/km	0.613	0.837	1.050	1.387	4.292	0.409
2	Initial IRI	m/km	0.600	0.701	0.818	1.028	1.844	0.276
3	Pavement Age	day	31	1,393	2,840	4,886	40,372	0.888
4	GATOR_CRACK_A_L	m^2	0	0	0.100	9.784	292.149	2.244
5	GATOR_CRACK_A_M	m^2	0	0	0	0.091	405.600	3.419
6	GATOR_CRACK_A_H	m^2	0	0	0	0	487.618	7.066
7	LONG_CRACK_WP_L_L	m	0	0	0	2.799	179.857	2.977
8	LONG_CRACK_WP_L_M	m	0	0	0	0	214.162	4.892
9	LONG_CRACK_WP_L_H	m	0	0	0	0	119.957	9.672
10	LONG_CRACK_NWP_L_L	m	0	0	1.213	34.702	322.611	1.810
11	LONG_CRACK_NWP_L_M	m	0	0	0	0.816	223.811	2.896
12	LONG_CRACK_NWP_L_H	m	0	0	0	0	289.700	5.475
13	TRANS_CRACK_L_L	m	0	0	0.148	8.848	208.957	2.129
14	TRANS_CRACK_L_M	m	0	0	0	0.884	139.303	2.823
15	TRANS_CRACK_L_H	m	0	0	0	0	78.699	3.937
16	POTHOLES_A_L	m^2	0	0	0	0	0.390	19.466
17	POTHOLES_A_M	m^2	0	0	0	0	0.600	26.312
18	POTHOLES_A_H	m^2	0	0	0	0	0.418	18.752
19	SHOVING_A	m^2	0	0	0	0	152.439	22.789
20	BLEEDING	m^2	0	0	0	0	480.495	3.832
21	POLISH_AGG_A	m^2	0	0	0	0	325.000	13.368
22	RAVELING	m^2	0	0	0	0	568.514	3.023
23	PUMPING_L	m	0	0	0	0	179.611	6.587
24	MAX_MEAN_DEPTH_1_8[a]	mm	0	3.981	5.440	7.587	27.000	0.579

[a]Rutting depth

0.6 m/km.

In contrast, with regard to SPSs, there was virtually no difference between the construction date and the first visit date for IRI measurements. Consequently, there was no practical requirement to calculate the initial IRI. However, to ensure consistency between the GPS and SPS data, the same extrapolation approach that was used to determine the initial IRI in the GPSs was also applied to the SPSs.

3.2.4 Data Elimination

The LTPP database comprehensively records the history of pavement maintenance and rehabilitation, irrespective of whether they are major or minor operations, all of which can influence pavement smoothness. Throughout the data mining process in this study, localized maintenance activities, such as crack sealing and pothole patching, demonstrated minimal impact on the IRI. While certain sections experienced a slight IRI decrease following maintenance procedures like fog seal coating or aggregate seal coating, this reduction was not consistently observed across all sections. Even when there was a decrease, the IRI typically rebounded to its original state within one to two years of service. Therefore, data from sections undergoing these specific maintenance activities were included in this study.

In contrast, in the case of maintenance involving overlaying after milling, the IRI showed a significant and consistent decrease in almost all sections. Accordingly, data collected after overlaying maintenance were excluded from this study, aligning with established principles in pavement design. Notably, the design manual for overlaying pavements differs from that for new pavement designs due to the distinct behaviors of new and overlaying pavements.

Consequently, Table 3 presents statistical descriptions of the 24 variables adopted in the constructed database using six indices: the minimum and maximum values, the lower and upper quartiles (Q1 and Q3), the median, and the coefficient of variation (COV). Among these, 21 pavement distress variables, initial IRI, and pavement age were used as the predictor variables, while the IRI served as the target variable. These 24 variables were derived from a comprehensive database, which incorporated 2,688 instances, as shown in Table 1. Machine learning models were constructed based on this extensive database.

4. Model Development and Evaluation

4.1 Overview of Machine Learning Techniques

4.1.1 Random Forest (RF)

Random Forest (RF) is an ensemble learning algorithm that

Table 4. Search Space of Hyperparameters

Algorithm	Hyperparameter	Description	Search space
RF	*n_estimators*	Number of tree models	10 – 500
	max_depth	Maximum depth for tree models	1 – 20
	min_samples_split	Minimum number of samples required to split a node	2 – 10
	min_samples_leaf	Minimum number of samples required in a leaf	1 – 10
XGB	*n_estimators*	Number of tree models	10 – 500
	max_depth	Maximum depth for tree models	1 – 20
	min_child_weight	Minimum sum of instance weight required in each leaf node	0 – 10
	gamma	Minimum loss reduction to make a further partition on a leaf node	0 – 1
	subsample	Ratio of the training data	0 – 1
	learning_rate	Step size at which the algorithm makes updates to the model weights	0.01 – 1

consists of a collection of decision trees. Each decision tree is trained using a bootstrap sample randomly selected from the training set with replacement, allowing specific data to appear in multiple bootstrap samples simultaneously. As a result, n bootstrap samples and their corresponding n decision trees are constructed. Consequently, the final prediction is determined as the mean of the predictions from the individual decision trees.

4.1.2 eXtreme Gradient Boosting (XGB)

eXtreme Gradient Boosting (XGB) is a boosting-type ensemble learning algorithm designed to create a strong predictor by sequentially adding weak predictors. It achieves this by establishing multiple weak learners and applying weights to each one to reduce the errors introduced by the previous weak learner. This process is iteratively performed on all the weak learners to build a robust prediction model. XGB employs an objective function that integrates both the loss function and regularization terms, enabling the algorithm to effectively obtain a balance between model accuracy and complexity. This approach helps prevent overfitting and enhances the algorithm's generalization capabilities.

4.1.3 Bayesian Optimization (BO)

Bayesian optimization (BO) aims to find an optimal solution, which can be a hyperparameter combination, for minimizing or maximizing a given objective function (e.g., performance metric), thereby improving model performance. In this study, Gaussian Process (GP) was adopted as a surrogate model, which allows to estimate a multi-dimensional Gaussian distribution (known as a posterior distribution) for the objective function using Bayes' theorem and previous observations. The acquisition function, Expected Improvement (EI) in this case, guides the selection of a new hyperparameter combination. The corresponding objective function value is then obtained, and this new observation updates the GP by refining the posterior distribution of the objective function. These processes are iterated until the termination condition, typically a maximum number of iterations, is satisfied. Compared to other hyperparameter tuning methods such as grid search and random search, the BO efficiently explores the hyperparameter search space.

4.2 Model Implementation

This study aims to perform an in-depth influence analysis of pavement distress on IRI with respect to their severities by interpreting their relationships captured within a machine learning model. To this end, the IRI prediction models with RF and XGB algorithms were developed using the Python programming language. In this study, the representative ensemble learning algorithms (i.e., RF and XGB) was employed owing to their robustness to outliers, ability to generate better predictive performance, and capacity to prevent overfitting. The database was divided into training and test sets, with 70 % allocated for training and the remaining 30 % for test. The BO and five-fold cross-validation were employed on the training set to explore the optimal hyperparameter combination from each search space, as listed in Table 4.

This study evaluated the predictive performance using two standard metrics: root mean square error (RMSE) and coefficient of determination (R^2). RMSE is the square root of the average of the squared differences between predicted and actual values, and R^2 measures the proportion of variation in the target variable explained by the independent input variables. These metrics are expressed in Eqs. (1) and (2), where Nrepresents the number of data points, y stands for the actual value, $\hat{y}$ means the predicted value, and $\bar{y}$ denotes the mean of the data.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2} \tag{1}$$

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

Figure 1 illustrates the test results of the developed RF and XGB models, indicating a correlation between the actual and predicted IRI values. Both the RF and XGB models exhibited excellent predictive performances, with the former yielding an RMSE of 0.2191 and an R^2 of 0.7874, and the latter resulting in an RMSE of 0.2268 and an R^2 of 0.7722. Upon comparing the derived performance of the two models, the RF model was proposed as the optimal IRI prediction model, utilizing the following

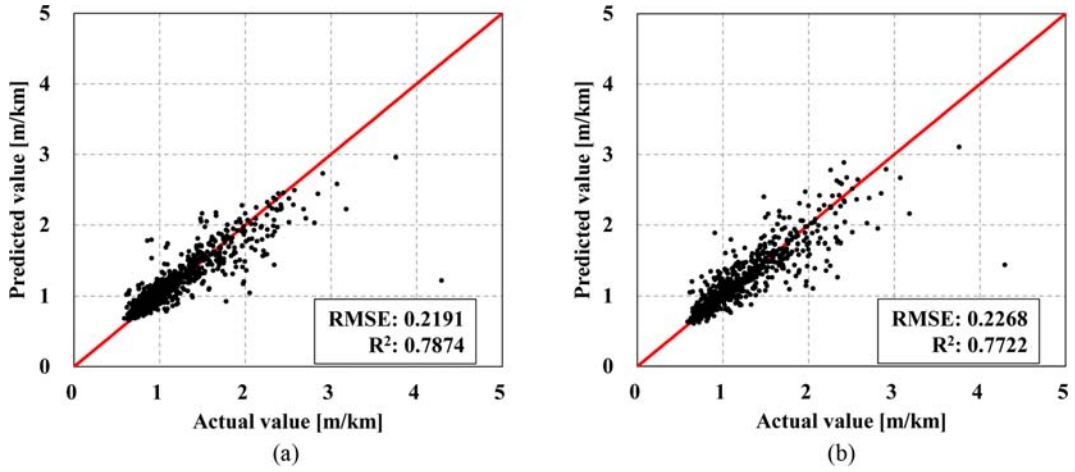

Fig. 1. Predictive Performance of RF and XGB Models: (a) RF Model, (b) XGB Model

hyperparameter combination: *n_estimators* = 484, *max_depth* = 16, *min_samples_leaf* = 1, *min_samples_split* = 2.

These results demonstrated that the developed RF model outperformed the previous statistics-based models (as discussed in Section 2.1). Notably, the significance of these findings was attributed to the model's exclusive focus on pavement distress without considering pavement design factors, confirming the viability of the adopting pavement distresses as predictor variables. Moreover, it can be inferred that the relationship between pavement distresses and IRI were appropriately captured within the developed RF model.

5. Influence Analysis by Model Interpretation

5.1 Variable Importance Ranking

With the proposed RF model, the variable's influence on the IRI predictions was assessed based on the SHapley Additive exPlanations (SHAP) method using the Python programming language in Section 5.1 and 5.2. This method, initially developed by Lundberg and Lee (2017), is rooted in cooperative game theory and adapted for interpreting machine learning models. It enables a fair allocation of contributions to a prediction among the model's input variables.

Figure 2(a) illustrates the resulting SHAP values corresponding to each predictor variable, where positive and negative values signify whether the variable increases or decreases IRI values, respectively. Data values of each predictor variable are represented in red or blue, indicating high or low values, respectively. Notably, the high-severity alligator crack (i.e., GATOR_CRACK_A_H) yielded positive SHAP values, depicted in red, suggesting that higher values of this variable correspond to higher IRI values. Fig. 2(a) also shows that most pavement distresses, initial IRI, and pavement age tend to increase IRI, indicating a positive correlation with these variables.

In some instances, as shown in Fig. 2(a), the data represented in red with a negative SHAP value indicate that high values of the variable contributed to decreasing IRI, while data in blue with

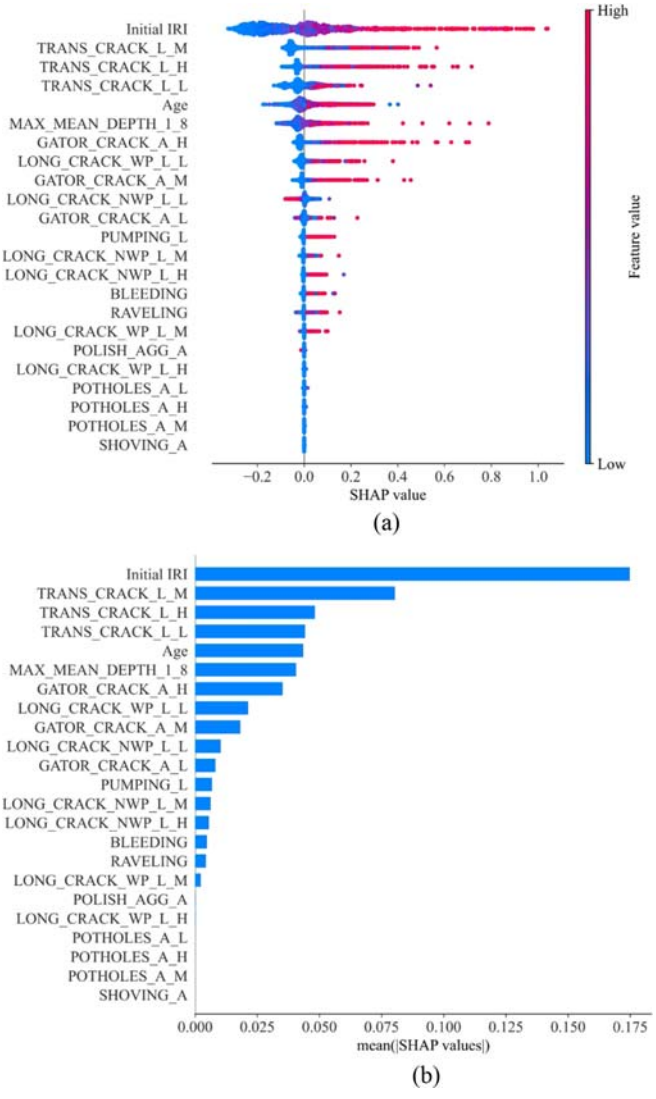

Fig. 2. Variable Importance Obtained by SHAP Method: (a) SHAP Summary Plot – Dot Graph, (b) SHAP Summary Plot – Bar Graph

positive SHAP values indicate that low values of the variable contributed to increasing IRI. Accordingly, both red and blue colors can be obtained in the negative and positive SHAP value ranges.

Figure 2(b) shows the mean SHAP value, representing the overall contribution of each predictor variable to IRI prediction. This mean SHAP value is an average of the absolute SHAP values of the individual distress variables. A longer bar in the graph signifies a greater contribution to the prediction, allowing for a ranking of each variable's importance. However, it is important to note that the mean SHAP value does not convey the direction the prediction, i.e., whether it leads to an increase or decrease in IRI.

In addition, to confirm the feasibility of the variable importance derived from the SHAP method, the maximal information coefficient (MIC) was calculated using the constructed database, assessing the mutual correlations among the IRI and predictor variables. MIC quantifies the maximum shared information between two variables, considering all potential relationships without restriction to specific types (e.g., linear or exponential). The MIC values range from 0 to 1, where 0 indicates no relationship between the paired variables, and 1 indicates a perfect relationship. It is determined as the highest maximum mutual information and further details about MIC can be found in Reshef et al. (2011).

Figure 3 displays the MIC values of the 23 predictor variables. Notably, initial IRI showed the strongest relationship with IRI, consistent with the conventional understanding that smoother pavements tend to have better long-term performance. Rutting and transverse crack ranked prominently, highlighting their significant influence on IRI. These two pavement distresses were distinctively incorporated into the MEDPG IRI prediction regression model. Transverse crack with high severity ranked lower compared to those with low or medium severity. Interestingly, according to Figs. 2 and 3, the longitudinal crack with low severity showed a similar ranking to alligator crack, despite the less common consideration of longitudinal crack in traditional IRI prediction models. In the AASHTO MEPDG, longitudinal crack was initially

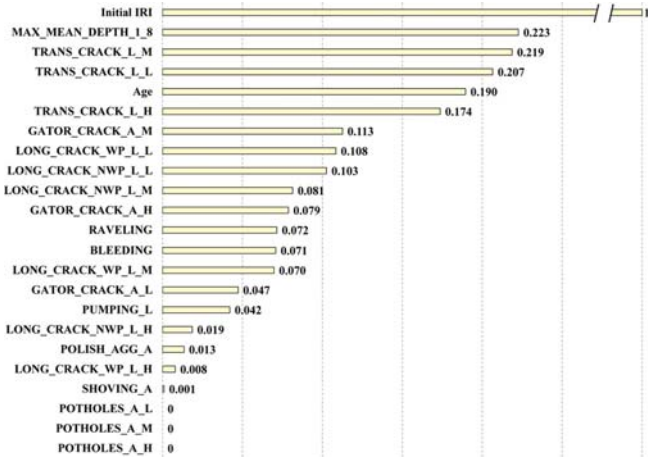

Fig. 3. MIC Values between the Adopted Predictor Variables and IRI

part of predictor variables in the 2004 version (ARA, 2004), but was excluded in the revised 2008 version. In contrast, pothole, despite its well-known status as a pavement distress of engineering concern, had minimal impact on IRI, as shown in Fig. 3. This finding may be attributed to their infrequent occurrence, often not consistently observed in road surface profile measurements conducted for IRI assessments, particularly in the right and left wheel paths.

It can be noted that the order of contribution inferred from the SHAP values (Fig. 2(b)) was consistent with that obtained from the MIC values. Table 5 compares the contribution orders derived from the MIC and SHAP values for the proposed model. Initial IRI, major pavement distresses (i.e., transverse crack, rutting, and alligator crack), and pavement age were consistently identified as the most influential variables for IRI prediction. Initial IRI was found to have the most significant influence based on the results from both the MIC and SHAP value analyses, emphasizing the importance of ensuring the quality and accurate recording of initial pavement conditions.

As indicated in Table 5, the six variables with the highest and lowest contributions consistently maintain their rankings in both the MIC- and SHAP-based contribution orders. Here, the six variables with the lowest contributions (POLISH_AGG_A, LONG_ CRACK_WP_L_H, SHOVING_A, POTHOLES_A_L,

Table 5. Contribution Order of Predictor Variables by MIC and SHAP Value Analyses

Contribution order	Criteria	
	MIC value	SHAP value
1	Initial IRI	Initial IRI
2	MAX_MEAN_DEPTH_1_8	TRANS_CRACK_L_M
3	TRANS_CRACK_L_M	TRANS_CRACK_L_H
4	TRANS_CRACK_L_L	TRANS_CRACK_L_L
5	Age	Age
6	TRANS_CRACK_L_H	MAX_MEAN_DEPTH_1_8
7	GATOR_CRACK_A_M	GATOR_CRACK_A_H
8	LONG_CRACK_WP_L_L	LONG_CRACK_WP_L_L
9	LONG_CRACK_NWP_L_L	GATOR_CRACK_A_M
10	LONG_CRACK_NWP_L_M	LONG_CRACK_NWP_L_L
11	GATOR_CRACK_A_H	GATOR_CRACK_A_L
12	RAVELING	PUMPING_L
13	BLEEDING	LONG_CRACK_NWP_L_M
14	LONG_CRACK_WP_L_M	LONG_CRACK_NWP_L_H
15	GATOR_CRACK_A_L	BLEEDING
16	PUMPING_L	RAVELING
17	LONG_CRACK_NWP_L_H	LONG_CRACK_WP_L_M
18	POLISH_AGG_A	POLISH_AGG_A
19	LONG_CRACK_WP_L_H	LONG_CRACK_WP_L_H
20	SHOVING_A	POTHOLES_A_L
21	POTHOLES_A_L	POTHOLES_A_H
22	POTHOLES_A_M	POTHOLES_A_M
23	POTHOLES_A_H	SHOVING_A

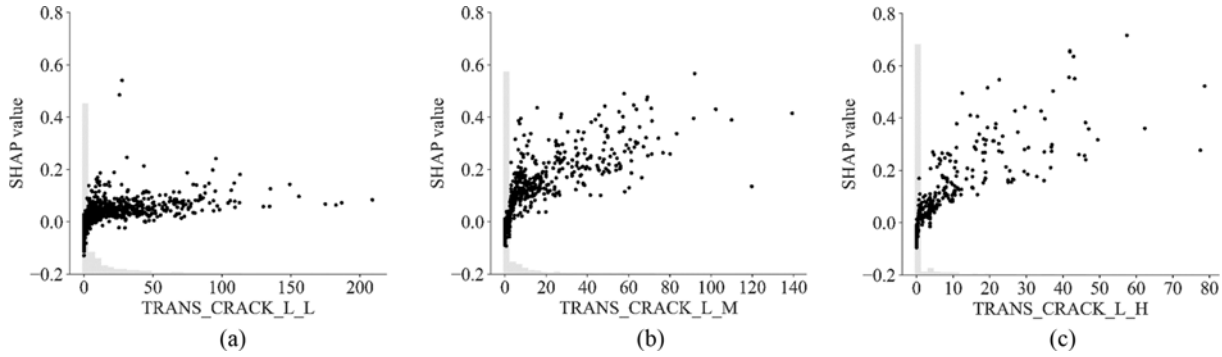

Fig. 4. Interaction between Data and SHAP values of Transverse Crack: (a) Low Severity, (b) Medium Severity, (c) High Severity

POTHOLES_A_M, and POTHOLES_A_H) exhibited a COV of more than 10 (as noted in Table 2), indicating their minimal influence on IRI due to their uniform distribution over a wide range, regardless of IRI.

5.2 Influence of Major Pavement Distress on IRI

5.2.1 Transverse Crack

As presented in Table 5, transverse crack is a significantly influential variable on IRI. The ranking of its contribution based on severity levels was in the order of medium, high, and low. This ranking order is consistent, as shown in Fig. 4. The existence of low-severity transverse crack was found to have an impact on IRI increment in model prediction since most of their SHAP values were positive (Fig. 4(a)); but these SHAP values did not appear to increase with longer length of the low-severity transverse crack. In contrast, for the medium- and high-severity transverse crack, an increase in total crack length led to higher SHAP values. Furthermore, the SHAP values for the medium- and high-severity transverse crack were significantly greater than those for the low-severity transverse crack (Figs. 4(b) and 4(c)). Considering the relatively low occurrence frequency of the medium- and high-severity transverse crack, their impact on IRI increment can be considered more significant. This implies that proactive maintenance for the low-severity transverse crack is important to prevent them from further propagating and to maintain pavement smoothness.

5.2.2 Rutting

Figure 5 illustrates the influence of rutting depth on the IRI predictions. A rutting depth above approximately 7 mm had a significant influence on the IRI values. Significantly, the degree of influence notably increased for rutting depths, which exceeded 20 mm. Conversely, for rutting depths lower than 7 mm, most of the SHAP values were close to zero, indicating that the influence of these low rutting depths was marginal. Consequently, a threshold (Rutting depth = 7 mm) was determined, which showed a significant influence on IRI increase.

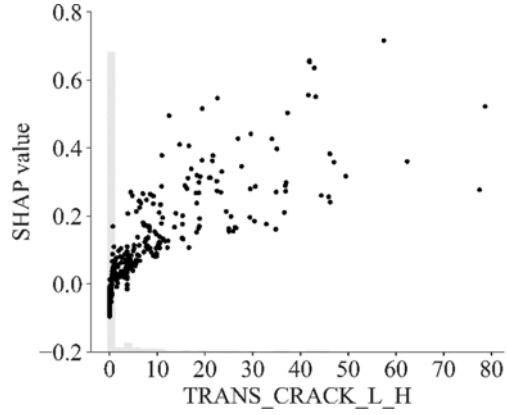

Fig. 5. Interaction between Data and SHAP values of the Rutting

The rutting depth measured in the transverse direction may not significantly influence IRI if the variations in rutting depth are small along the vehicle's driving distance in the longitudinal direction. It can be inferred that at greater rutting depths, the variation is sufficiently large to affect IRI. Accordingly, many studies have attempted to determine the standard deviation of the

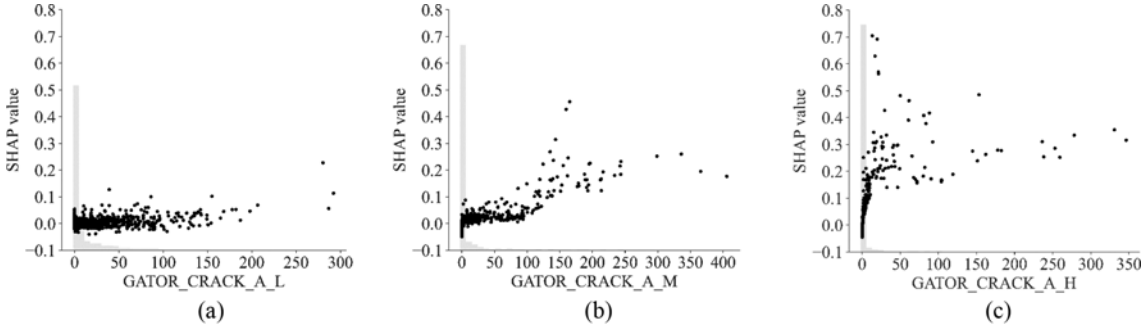

Fig. 6. Interaction between Data and SHAP values of the Alligator Crack: (a) Low Severity, (b) Medium Severity, (c) High Severity

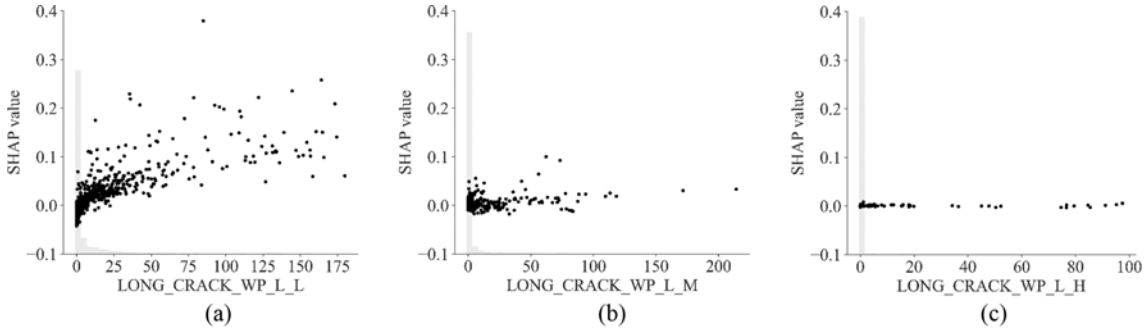

Fig. 7. Interaction between Data and SHAP values of Longitudinal Crack in the Wheel Path: (a) Low Severity, (b) Medium Severity, (c) High Severity

rutting depth rather than the mean of the rutting depth (Paterson, 1989; Bashar and Darter, 1995).

5.2.3 Alligator Crack

Figure 6 presents the contribution of alligator crack to the IRI, considering both their quantity and severity. Similar to the transverse crack, the alligator crack with medium and high severity had a greater influence on IRI than those with low severity. Particularly, at high severity, even minor alligator crack produced noticeably elevated SHAP values. These observations highlight the importance of proactive maintenance to prevent the progression of alligator crack.

5.2.4 Longitudinal Crack in the Wheel and Non-Wheel Paths

As shown in Fig. 7, longitudinal cracks in the wheel path, regardless of their severity, produced positive SHAP values. Interestingly, the low-severity longitudinal crack in the wheel path resulted in relatively higher SHAP values than the medium- and high-severity longitudinal cracks, which exhibited minimal influence on IRI. The low-severity one was determined to be the eighth most significant contributor to IRI prediction based on the contribution order derived from both the MIC and SHAP values (see Table 5).

Regarding the longitudinal crack in the non-wheel path, it would be intricate to analyze the influence on IRI. Fig. 2 indicates that a low-severity crack was associated with a decrease in IRI. In contrast, the medium- and high-severity cracks tended to increase IRI, although the degrees of their influences were relatively marginal. This observation may be attributed to the fact that longitudinal profiles for IRI computation are typically obtained on both wheel paths.

In the first version of the AASHTO MEPDG, longitudinal crack in both the wheel and non-wheel paths were incorporated as predictor variables in the IRI prediction model (ARA, 2004). However, these cracks were excluded in the subsequent version of the MEPDG (AASHTO, 2008). The analysis based on the interpretable machine learning approach suggested that the longitudinal crack in the wheel path could serve to advance the IRI prediction and management.

Table 6. Predictive Performance Comparison between Partial and Full Models

Model	Algorithm	Training		Test	
		RMSE	R^2	RMSE	R^2
Partial Model	RF	0.0875	0.9683	0.2217	0.7824
	XGB	0.0006	1.0000	0.2351	0.7553
Full Model	RF	0.0893	0.9669	0.2191	0.7874
	XGB	0.0006	1.0000	0.2268	0.7722

5.2.5 Minor Pavement Distress

Pothole of any severity, polished aggregates, and shoving were identified as pavement distresses with a minimal influence (i.e., minor pavement distresses) on IRI, as presented in Fig. 2(b). In addition to employing a machine learning model that incorporated all the variables (i.e., the full model), an additional model was developed utilizing only a subset of variables, excluding the minor pavement distresses (i.e., the partial model). However, it is important to note that the high-severity longitudinal crack in the wheel path (i.e., LONG_CRACK_WP_L_H) was not excluded. This decision was made to consider the significant impact of the low- and medium-severity longitudinal crack in the wheel path and to maintain simplicity in variable selection, despite its classification as a minor pavement distress.

Table 6 presents a comprehensive comparison of the predictive performances of the partial and full models, indicating minimal disparities between the two results. Upon examining the results in Fig. 2(a), the SHAP values for the pothole, polished aggregates, and shoving do not exhibit distinctive positive or negative contributions, suggesting a limited effect on the model predictions. Therefore, the exclusion of these variables did not significantly affect model accuracy, as previously mentioned.

5.2.6 Comparative Analysis with AASHTO MEPDG Smoothness Model

Based on the same database used to establish the machine learning models, IRI was predicted using the AASHTO MEPDG smoothness model, as presented in Eq. (3). The MEPDG smoothness model is an empirical statistical model that focuses on major distress factors such as rutting, fatigue crack, and transverse crack (AASHTO, 2020). In

Table 7. Comparative Analysis between RF Model and AASHTO Model

Model	R^2	Number of sections	Number of instances
AASHTO model	0.5184	258	1,556
RF model	0.7874	377	2,688

addition, the model incorporates environmental factors and subgrade material properties as site-specific factors. It can be noted that the model was developed based on the US standard unit system.

$$IRI = IRI_0 + 40.0*RD + 0.4*FC_{total} + 0.008*TC + 0.015*SF, \quad (3)$$

where

FC_{total} = % area of alligator, longitudinal, and reflective cracks in wheel path
IRI_0 = Initial IRI after construction [in/mi]
RD = Average rutting depth [in]
SF = Site factor considering precipitation, freezing index, and subgrade properties
TC = Length of transverse crack [ft/mi]

Table 7 summarizes the comparison of prediction performance between the developed RF model and the AASHTO MEPDG smoothness model. The AASHTO model achieved an R^2 of 0.5184, indicating lower predictive performance compared to the RF model (R^2 = 0.7874). Some sections used in establishing the RF model lacked site-specific factor data, resulting in reduced dataset utilization for the application of the AASHTO model. Despite this limitation, the predictive performance of the RF model was significantly superior to that of the latest AASHTO model. The enhanced prediction potential of machine learning, facilitated by the comprehensive utilization of pavement distress data considering severity, suggests that it could outperform statistical predictions that incorporate site-specific factors where pavement is constructed.

5.3 Use of Developed RF Model

The machine learning model developed in this study is applicable to flexible pavement roads without overlaying. Additionally, it is not limited by specific climatic zones or subgrades, allowing for application across various environmental conditions and soil types.

However, it should be noted that the RF model is built using pavement distress data collected from highway pavement roads, where vehicles travel at a constant speed. In urban areas, frequent stopping situations and lower speeds of vehicles increase the likelihood of minor distress preceding major issues, such as potholes and shoving, leading to irregular maintenance activities. Therefore, the application of the RF model and the corresponding influence analysis results are not recommended for urban pavement roads. Future studies could focus on developing an IRI prediction model and conducting influence analysis using extensive urban pavement road data.

6. Conclusions

This study demonstrated the influence of pavement distress on IRI through the utilization of a meticulously preprocessed LTPP database. The developed IRI prediction model, incorporating interpretable machine learning, facilitated a deeper comprehension of the influences, with particular emphasis on the severities of pavement distress. The key findings and contributions of this study are as follows:

1. The developed RF model achieved superior predictive performance with an RMSE of 0.2191 and an R^2 of 0.7874, representing the validity of the relationship between pavement distresses and IRI captured within the developed model.
2. The model interpretation results indicated that the major pavement distresses, including transverse crack, rutting, and alligator crack, were the most influential factors. Additionally, the low-severity longitudinal crack in the wheel path exhibited a similar impact on IRI as alligator crack. In contrast, pothole, polished aggregates, and shoving were found to have minimal influences.
3. Both transverse and alligator cracks at medium and high severities significantly contributed to IRI increment, demonstrating the requirement for proactive maintenance of these distresses at the low severity level. Moreover, the existence of a threshold in rutting depth inducing significant IRI increment was observed.
4. A comparative analysis demonstrated that the RF model outperformed the AASHTO MEPDG smoothness model. The enhanced predictive performance of machine learning, achieved through the comprehensive incorporation of pavement distress data that accounts for severity, indicates its potential to surpass statistical predictions that consider the site conditions of pavement.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (No. 2019R1A2C2086647, 2020R1A6A1A03045059, and RS-2023-00245022).

ORCID

Kibeom Kwon http://orcid.org/0000-0001-8889-1386
Hangseok Choi http://orcid.org/0000-0002-2040-8850
Khanh Pham http://orcid.org/0000-0001-6566-6776
Sangwoo Kim http://orcid.org/0009-0005-3880-2214
Abraham Bae http://orcid.org/0000-0002-3355-0653

References

Abdelaziz N, Abd El-Hakim RT, El-Badawy SM, Afify HA (2020) International Roughness Index prediction model for flexible pavements. *International Journal of Pavement Engineering* 21(1):88-99, DOI:

10.1080/10298436.2018.1441414
Al-Mansour AI, Shokri AA (2022) Correlation of pavement distress and roughness measurement. *Applied Sciences* 12(8):3748, DOI: 10.3390/app12083748
American Association of State Highway and Transportation Officials (AASHTO) (2008) Mechanistic-empirical pavement design guide: A manual of practice. Interim Edition, Washington DC, USA
American Association of State Highway and Transportation Officials (AASHTO) (2015) Mechanistic-empirical pavement design guide: A Manual of Practice, second ed, Washington, District of Columbia
American Association of State Highway and Transportation Officials (AASHTO) (2020) Mechanistic-empirical pavement design guide: A Manual of Practice, third ed, Washington, District of Columbia
Applied Research Associates (ARA) (2004) Inc., Guide for Mechanistic-Empirical Design of New and Rehabilitated Pavement Structures. National Cooperative Highway Research Program Project 1-37A, Transportation Research Board, National Research Council, Washington, District of Columbia
Bashar A, Darter I (1995) Effect of pavement deterioration types on IRI and rehabilitation, Transportation Research Record, 1505, 57
Chandra S, Sekhar CR, Bharti AK, Kangadurai B (2013) Relationship between pavement roughness and distress parameters for Indian highways. *Journal of Transportation Engineering* 139(5):467-475, DOI: 10.1061/(ASCE)TE.1943-5436.0000512
Damirchilo F, Hosseini A, Mellat Parast M, Fini EH (2021) Machine learning approach to predict international roughness index using long-term pavement performance data. *Journal of Transportation Engineering, Part B: Pavements* 147(4):04021058, DOI: 10.1061/JPEODX.0000312
Gong H, Sun Y, Shu X, Huang B (2018) Use of random forests regression for predicting IRI of asphalt pavements. *Construction and Building Materials* 189:890-897, DOI: 10.1016/j.conbuildmat.2018.09.017
Kaloop MR, El-Badawy SM, Ahn J, Sim HB, Hu JW, Abd El-Hakim RT (2022) A hybrid wavelet-optimally-pruned extreme learning machine model for the estimation of international roughness index of rigid pavements. *International Journal of Pavement Engineering* 23(3):862-876, DOI: 10.1080/10298436.2020.1776281
Karballaeezadeh N, Mohammadzadeh SD, Moazemi D, Band SS, Mosavi A, Reuter U (2020) Smart structural health monitoring of flexible pavements using machine learning methods. *Coatings* 10(11):1100, DOI: 10.3390/coatings10111100
Kargah-Ostadi N, Stoffels SM (2015) Framework for development and comprehensive comparison of empirical pavement performance models. *Journal of Transportation Engineering* 141(8):04015012, DOI: 10.1061/(ASCE)TE.1943-5436.0000779
Lin JD, Yau JT, Hsiao LH (2003) Correlation analysis between international roughness index (IRI) and pavement distress by neural network. *In 82nd Annual Meeting of the Transportation Research Board* 12(16): 1-21
Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30
Mactutis JA, Alavi SH, Ott WC (2000) Investigation of relationship between roughness and pavement surface distress based on WesTrack project. *Transportation Research Record* 1699:107-113, DOI: 10.3141/1699-15
Marcelino P, de Lurdes Antunes M, Fortunato E, Gomes MC (2021) Machine learning approach for pavement performance prediction. *International Journal of Pavement Engineering* 22(3):341-354, DOI: 10.1080/10298436.2019.1609673
Miller JS, Bellinger WY (2003) Distress identification manual for the long term performance program. Office of Infrastructure Research and Development, Federal Highway Administration, Report No. FHWA-RD-03-031
Paterson W (1989) A transferable causal model for predicting roughness progression in flexible pavements. Transportation Research Record, 1215
Perera RW, Kohn SD (2001) LTPP Data analysis: Factors Affecting Pavement Smoothness, Transportation Research Board, National Research Council, Washington, District of Columbia
Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, Lander ES, Mitzenmacher M, Sabeti PC (2011) Detecting novel associations in large data sets. *Science* 334(6062):1518-1524, DOI: 10.1126/science.1205438
Sandra AK, Sarkar AK (2013) Development of a model for estimating International Roughness Index from pavement distresses. *International Journal of Pavement Engineering* 14(8):715-724, DOI: 10.1080/10298436.2012.703322
Sayers MW (1986) The international road roughness experiment: Establishing correlation and a calibration standard for measurements. University of Michigan, Ann Arbor, Transportation Research Institute
Sollazzo G, Fwa TF, Bosurgi G (2017) An ANN model to correlate roughness and structural performance in asphalt pavements. *Construction and Building Materials* 134:684-693, DOI: 10.1016/j.conbuildmat.2016.12.186
Zeiada W, Hamad K, Omar M, Underwood BS, Khalil MA, Karzad AS (2019) Investigation and modelling of asphalt pavement performance in cold regions. *International Journal of Pavement Engineering* 20(8):986-997, DOI: 10.1080/10298436.2017.1373391
Zhang M, Gong H, Jia X, Xiao R, Jiang X, Ma Y, Huang B (2020) Analysis of critical factors to asphalt overlay performance using gradient boosted models. *Construction and Building Materials* 262: 120083, DOI: 10.1016/j.conbuildmat.2020.120083
Ziari H, Sobhani J, Ayoubinejad J, Hartmann T (2016) Analysing the accuracy of pavement performance models in the short and long terms: GMDH and ANFIS methods. *Road Materials and Pavement Design* 17(3):619-637, DOI: 10.1080/14680629.2015.1108218