TriDurLE

National Center for Transportation
Infrastructure Durability & Life-Extension

Project ID: 2021-MST-04

AI-Based Prediction Models for Transportation Infrastructure Asset Management Data Hub – Phase I

Final Report

By

Xianbiao Hu, Yang Song, and Chenxi Chen

Pennsylvania State University

Jenny Liu and David Yizhuan Wang

Missouri University of Science and Technology

for

National University Transportation Center TriDurLE

Department of Civil & Environmental Engineering

405 Spokane Street PO Box 642910

Washington State University Pullman, WA 99164-2910

Date 8/31/2024

Table of Contents

Contents

List of Figures

List of Tables

Executive Summary

Transportation infrastructure system networks are essential to sustain our economy, society, and quality of life. The Federal Highway Administration (FHWA) indicates the data-driven asset management (inspection, maintenance, and emergency response) of urban infrastructure (building, transportation, energy, and cyber) represents the future of work and identifies it as one of their strategic objectives. An accurate prediction of asphalt pavement condition is important to guide pavement maintenance practices. The existing models for pavement condition predictions are predominantly based on linear regressions or simple machine learning technics (e.g., the artificial neural network (ANN) and the random forest models). However, additional work on these models is needed to improve their basic assumptions, training efficiency, and interpretability. To this end, a new modeling approach is proposed in this project, which includes a Light Gradient Boosting Machine (LightGBM)-based ensemble learning model, coupled with the Shapley Additive Explanation (SHAP) method, to predict the International Roughness Index (IRI) of asphalt pavements. The SHAP method was applied to interpret the underlying influencing factors and their interactions. 20 features were initially identified as the model inputs, and 1,706 observations were extracted from the Long-Term Pavement Performance (LTPP) database. Two benchmark models, namely the Mechanistic-Empirical Pavement Design Guide (MEPDG) model and the ANN model, were used for comparison. The results showed that the developed model outperformed both benchmark models in terms of R-square, root mean square error, and the absolute error. Feature interpretation was performed to identify the top influencing factors of IRI, and the interaction effects between two features were also analyzed. The 20-feature model was further simplified based on the analysis result. The simplified model only required five features to efficiently and effectively predict IRI using the proposed LightGBM-based approach, which can reduce the workload in data collection and management for pavement engineers.

Chapter 1. Introduction

1.1 Problem Background

Transportation infrastructure system networks are essential to sustain our economy, society, and quality of life. The Federal Highway Administration (FHWA) indicates the data-driven asset management (inspection, maintenance, and emergency response) of urban infrastructure (building, transportation, energy, and cyber) represents the future of work and identifies it as one of their strategic objectives (FHWA 2018). In the broad field of traffic operations, representative examples include the data-driven optimization of traffic signal infrastructure using reinforcement learning methods [1-5], the application of automated vehicles for highway infrastructure maintenance [6-9], and the development of electric vehicle charging infrastructure to promote sustainable transportation modes [10-12].

In the field of roadway infrastructure, predicting pavement conditions requires digitization and the integration of multidisciplinary data from the entire infrastructure system. Establishing a data hub equipped for the ongoing technological revolution in transportation infrastructure engineering is critical, with systematic data mining serving as a key step toward achieving this goal. The 4th Industrial Revolution has introduced advanced technologies such as the Internet of Things (IoT), high-speed mobile internet, artificial intelligence (AI), cloud computing, and big data analytics, offering unprecedented capabilities to support this effort.

Numerous data in the transportation / pavement area have been collected over years. One of the most well-known databases is the Long-Term Pavement Performance (LTPP) database, managed by FHWA. The database includes information from 2,581 pavement sections in ninety U.S. states, territories, and Canadian provinces. The information with respect to climate, traffic, pavement structure, and pavement performance has been recorded. Like LTPP, there are other open-source pavement databases such as WesTrack, MnRoad, and the National Center for Asphalt Technology (NCAT) Test Track database, managed by different research institutes and state agencies. With the pavement management system (PMS) databases at state agencies taken account, these collections have provided considerable and even overwhelming amount of data to pavement engineers. However, the current research in pavement and transportation engineering has not fully taken advantage of the existing data.

The most used data analysis method in pavement engineering is simple linear regression. Models developed from regressions requires low computational effort and can be applied regardless of the sample size, but regression models usually provide limited prediction accuracy. Regression models have been widely implemented in pavement industry, including in the American Association of State Highway and Transportation Officials (AASHTO) Pavement Mechanistic-Empirical (ME) Design Program. With the development of the advanced computation technologies, the existing data can be better interpreted, and new information can be generated to benefit the pavement community. Multiple research studies have recently tried to implement data-driven intelligence in pavement data analysis. However, there have been limitations in these early studies, including: the data were extracted from limited pavement sections or climate regions, the contributing factors were not clearly identified, and the research findings were difficult to implement. Given the problems in the current research, we propose a research study focusing on data mining using the massive data that have been collected in the pavement and transportation area. The advanced computational methods will be wisely selected and utilized. The outcomes will primarily cover the following four aspects: pavement performance prediction, the effects of changes in traffic patterns in pavement performance, prediction of properties of paving materials, and decision makings in pavement preservation, repair, and rehabilitation.

1.2 Research Statement

Roughness is an important indicator for pavement condition, representing the pavement riding quality. To quantify the pavement roughness, the World Bank developed the International Roughness Index (IRI) in the 1980s, which was defined as “the accumulated suspension vertical motion divided by the distance travelled as obtained from a mathematical model of a simulated quarter-car traversing a measured profile at 80 km/h” [13]. Pavements with high roughness can cause the increase of fuel consumption, cost of vehicle maintenance, traffic crash, and greenhouse gas emissions [14]. It is cost-intensive for highway agencies to manually survey and monitor the roughness of pavements frequently [15]. An accurate prediction of pavement roughness is important to guide the practices of pavement management and maintenance.

Researchers have developed several IRI predictive models with different contributing factors.

Those prediction models can be generally categorized into conventional regression-based models and machine learning-based models. The data used in most of these studies were collected from the Long-Term Pavement Performance (LTPP) database or the local Department of Transportations (DOTs) databases [16, 17]. The regression-based models usually adopted deterministic and explicit model formulations. For example, in the well-known Mechanistic-Empirical Pavement Design Guide (MEPDG), the IRI was assumed to be in a linear relationship with performance measurements (i.e., initial IRI after construction, area of fatigue cracking, length of transverse cracking, and average rut depth), environment conditions (i.e., average annual freezing index, average annual precipitation and rainfall), pavement structure (i.e., percent plasticity index of the soil) and pavement age. The MEPDG model was regressed based on 1,926 samples from the LTPP database, with a R^2 (coefficient of determination) of 0.56, which meant only 56% of the variation in the dataset could be explained by the model. Similarly, Paterson proposed an incremental linear model with independent variables including structural information (structural number), performance conditions (i.e., area of cracking, rut depth, area and protrusion of patch repairs, and volume of open potholes), traffic information (i.e., equivalent single axle loads (ESALs)), environment information (moisture and temperature), and pavement age [18]. Later on, this model became the base upon which the HDM-III and HDM-4 models were developed by the World Bank with more complex formulae [19, 20]. More research using linear regression models can be found in Pérez-Acebo et al. [20], Al-Omari and Darter [21], Mactutis et al. [22].

However, linear regression models were usually limited in prediction accuracy due to a few methodological issues. The method assumed linear relationships between the predictor and the variables, which can fall short to describe the complex non-linear relationships. In addition, linear regressions made strong assumptions of multivariate normality and homoscedasticity. The former stated that the residuals should be normally distributed while the latter assumed the error levels were similar across the range of each independent variable. However, it is known that pavement performance and its independent variables vary significantly both spatially and temporally, which means one universal model developed using the linear regression method can yield significantly different levels of errors in predicting pavement conditions for different climate regions. Other regression-based models for IRI prediction were in the forms of polynomial, exponential or hybrid for relations between pavement roughness and their independent variables. Although some improvements in prediction accuracy were made in those regression models, the issues associated with the unrealistic assumptions in the regression-based models have not yet been addressed [16, 23-25]. Moreover, the regression modeling approach was ineffective in revealing the impact of each influencing factor due to the limitations of the model forms and the dependence among predictors. Previous studies indicated that the multi-collinearity among the features in the MEPDG model concealed the impact of each individual feature on IRI, as the principle of independency in regression analyses was violated [26, 27].

To address the issues of regression-based models, machine learning models, such as the artificial neural network (ANN) and the random forest models (RFM), have been employed to predict IRI. Those models utilized the mechanisms that mapped the nonlinear relationships among the multiple modeling parameters with implicit and flexible model formations. Among all, ANN has been one of the most used techniques. The ANN models were more flexible and scalable than the conventional regression models in terms of formation, and as a result, they can use sufficient input features and learn the hidden relationships without imposing any explicit relationships. ANN can also learn and model the non-linear relationships to address the dependence issue among predictors. For example, two ANN models (i.e., dot product and quadratic function ANN) with R^2 of 0.69 were developed based on the data collected by Kansas DOT considering features including performance (i.e., rutting, fatigue cracking, transverse cracking, blocking cracking) and traffic (i.e., EASL) [16]. Lin et al. [28] developed a three-layer back propagation neural network to estimate IRI, with functional class and pavement distress (i.e., rutting, alligator cracking, patching, and bleeding) as the input features. Yet, only simple attribution analysis was conducted based on the node weight without discussions of the interactions among predictors. Choi et al. [17] utilized more variables including structure information (i.e., structure number, top layer thickness), material properties (i.e., percent of air void, viscosity and percent of aggregate gradation passing on 200 sieve) and traffic (i.e., EASL), and demonstrated the superiority of ANN over linear models in terms of prediction accuracy. Kaya et al. [29] developed an ANN prediction model based on pavement age and pavement performance (i.e., rutting, longitudinal cracking, transverse cracking and IRI values measured in the previous two years) with R^2 reaching 0.991. Similar research could also be found in other studies [30-32].

Attempts have also been made to adopt other machine learning models. Kargah-Ostadi and Stoffels [33] trained Generalized Radial Basis Function (RBF) networks and Gaussian Support Vector

Machines (GSVM) to estimate IRI. The result showed that the RBF and GSVM model outperformed ANN with the early stopping approach while underperformed ANN model with the Bayesian inference regularization technique. Likewise, a Group Method of Data Handling (GMDH) model was established with nine variables by Ziari et al. [34]. Recently, the ensemble learning technics were applied to predict pavement performance, and the learning algorithm demonstrated its advantages over other machine learning models. For example, as a bagging-based ensemble algorithm, a random forest regression (RFM) was developed to model IRI by generating base parallel and independent regression trees with the output being the average of each individual weak regressor [35]. Compared to ANN, the model managed to improve the generalization capabilities and reduce the overfitting effect with R^2 of t 0.974 in the testing set [36, 37]. However, in their studies, instead of using the measured data directly as the "target IRI", the measured data were pre-smoothed using interpolation with a linear function, which challenged the model soundness. Other research built on machine learning approaches can also be found in Zhang et al. [38] and Gong et al. [39]. However, the models became complicated as the number of layers in ANN or the number of trees in RFM increased. The model training could be slow and inefficient. In addition, the interpretability of results from such models was usually low since most models functioned like black boxes.

1.3 Objectives

To address these issues, in this study, a Light Gradient Boosting Machine (LightGBM)-based ensemble learning model coupled with the Shapley Additive Explanation (SHAP) method was proposed. The LightGBM model was utilized to predict the IRI of asphalt pavements, and the SHAP method was applied to interpret the impact of underlying influencing factors and their interaction effects. 1,706 observations have been extracted from the LTPP database. The dataset was divided into two subsets, with 80% for training and 20% for testing. Two benchmark models, namely the MEPDG model and an ANN model, were used for comparison. Feature interpretation was performed to identify the top influencing factors of IRI evolution. Further, the interaction effects between two features were analyzed to uncover some implicit but important attributions to IRI value.

Chapter 2. Methodology

2.1 LightGBM-based IRI prediction model

The LightGBM method, different from the bagging-based models (e.g., RFM), applies a boosting-based learning algorithm, and produces low bias by training new weak learner from the output of previous learners. It was firstly proposed by Ke et al. [40] and was developed based on the gradient boosting decision tree (GBDT) algorithm. GBDT is a typical boosting-based model with the base leaners being decision trees [41]. Because it only utilizes the first derivative of the loss function during base model training, the training efficiency of GBDT declines when processing datasets with large sample size and multiple dimensions. Compared with GBDT, LightGBM applies the second order gradient statistics to optimize the objective, and introduces two novel technics, namely the gradient-based one-side sampling (GOSS) algorithm and the exclusive feature bundling (EFB) algorithm, to improve prediction accuracy and efficiency.

In LightGBM, the training dataset is designated as $O = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, where x_i is an array representing the defined features, and y_i is the IRI value for the i^{th} observation out of total n training samples (i.e., 1,364 in this study). The predictor set and the response set are designated as $X = \{x_1, x_2, \dots x_n\}$ and $Y = \{y_1, y_2, \dots y_n\}$, respectively. The target of the learning process is to find the function that minimizes the objective function, which is defined as the summation of the loss function and the regularization term as presented in (1).

$$\tilde{F} = arg\min_{F} E_{X,Y}\,(Obj) = arg\min_{F} E_{X,Y}\left(\sum_{i=1}^{n} l(y_i, F(x_i)) + \sum_{k=1}^{K} \Omega(f_k)\right) \qquad (1)$$

where l is the loss function and Ω is a regularization term. F is the combination of a total number of K base regression trees with each being f_k, as presented in (2).

$$F = \sum_{k=1}^{K} f_k(X) \qquad (2)$$

Therefore, for the k^{th}tree, the objective function can be formulated as (3).

$$Obj_k = \sum_{i=1}^{n} l(y_i, F_{k-1}(x_i) + f_k(x_i)) + \sum_{k=1}^{K} \Omega(f_k) \tag{3}$$

After Taylor Expansion with constants being removed (as constant components don't play a role in the optimization problem), the first term of (3) is presented as (4).

$$\sum_{i=1}^{n} l(y_i, F_{k-1}(x_i) + f_k(x_i)) \cong \sum_{i=1}^{n} (g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i)) \tag{4}$$

where g_i and h_i denote the first and second order gradient statistics of l, respectively. Then (4) can be reformulated by the leaf nodes of the trees, as shown in (5).

$$\sum_{i=1}^{n} (g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i)) = \sum_{j=1}^{T} \left(\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i \right) w_j^2 \right) \tag{5}$$

where T represents the total number of leaf node; I_j is the sample set on the node j; w_j is the weight of the leaf node j. Similarly, the second term of (3) can be presented by the leaf nodes as in (6).

$$\sum_{k=1}^{K} \Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} \omega_j^2 \tag{6}$$

where γ is a penalty parameter to control the complexity of tree structure and λ controls the regularization level. By combining (5) and (6), the objective function is expressed as (7):

$$Obj_k = \sum_{j=1}^{T} \left(\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i \right) w_j^2 \right) + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} \omega_j^2 \tag{7}$$

Apparently, in (7), the objective function is a quadratic function of w_j. Its minimum is reached when the first derivative $\frac{\partial Obj_k}{\partial \mathrm{w_j}} = 0$. Hence, the optimal leaf node weight w_j^* and the minimum objective value can be solved via (8) and (9).

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{8}$$

$$Obj^{*} = -\frac{1}{2}\sum_{j=1}^{T}\frac{\left(\sum_{i\in I_j} g_i\right)^2}{\sum_{i\in I_j} h_i + \lambda} + \gamma T \qquad (9)$$

According to the above-mentioned process, during the optimization of each base regression tree, the key step is to find the optimal split points at tree nodes. The EFB and GOSS algorithms are introduced below to help define the split points efficiently. In a nutshell, EFB is employed to speed up the training process of LightGBM via bundling many exclusive features into a few dense features. On the other hand, GOSS improves the computational efficiency by optimizing the observation sampling. The rationale behind the algorithm is that an instance can be determined to be well-trained if its first order gradient is low. By excluding the observations with low first-order gradient, the number of split points between candidate instances can be reduced; thus, reducing the computational cost. The GOSS algorithm is implemented in the following procedures. The training samples are first sorted by their gradients in descending order. Then, a subset A containing the top $a * 100\%$ samples is constructed. From the remaining sample set A^c consisting of $(1 - a) * 100\%$ samples, the number of $b * |A^c|$ samples are randomly selected to form a set B, which maintains the original distribution of training samples and thus preserves the accuracy of the predictive model. In the end, the instances are split on $A \cup B$ according to the estimated optimal variance gain $\widetilde{V_p}(d)$ presented in (10).

$$\widetilde{V_p}(d) = \frac{1}{n}\left(\frac{\left(\sum_{x_i\in A_l}\ddot{g}_i + \frac{1-a}{b}\sum_{x_i\in B_l}\ddot{g}_i\right)^2}{n_l^j(d)} + \frac{\left(\sum_{x_i\in A_r}\ddot{g}_i + \frac{1-a}{b}\sum_{x_i\in B_r}\ddot{g}_i\right)^2}{n_r^p(d)}\right) \qquad (10)$$

where p denotes the p^{th} feature, d is the optimal split point between instances, $\ddot{g}_i$ is the negative gradient of the loss function l, i.e., $\ddot{g}_i = -g_i$, $\frac{1-a}{b}$ is a parameter to normalize the summation of gradients, $A_l = \{x_i \in A: x_{ip} \leq d\}$, $A_r = \{x_i \in A: x_{ip} > d\}$, $B_l = \{x_i \in B: x_{ip} \leq d\}$, $B_r = \{x_i \in B: x_{ip} > d\}$, and x_{ip} is the value of feature p from instance i. Studies have shown that GOSS outperforms the random sampling method adopted by other GBDT algorithms in terms of prediction accuracy [40].

2.2 SHAP method for results interpretation

SHapley Additive exPlanation (SHAP) is one of the additive feature attribution methods to explain individual predictions based on the game theoretically optimal Shapley Values [42]. While other global feature attribution methods that include gain, split count, and feature permutation may produce inconsistent results by lowering the assigned importance of a feature when the true impact of that feature increases, SHAP is a consistent feature attribution method. The SHAP values are calculated by averaging all possible orderings of input feature permutations; in contrast, other methods only consider a single ordering of independent variables. Hence, the SHAP method can attribute consistent feature importance and recover influential features. In this study, the SHAP method is adopted to explain the ensemble learning results.

In the SHAP algorithm, the IRI value of the proposed ensemble model can be represented by the sum of attribution value of each input feature, as shown in (11).

$$\widehat{y_i} = f(x_i) = \phi_0 + \sum_{p=1}^{|P|} \phi_p z_{ip} \tag{11}$$

where $\widehat{y_i}$ and $f(x_i)$ are the predicted IRI value of the i^{th} sample. ϕ_0 is a constant which equals the average of predicted IRI value from all observations, as shown in (12), where N is the number of total IRI observations (i.e., 1,706 in this study).

$$\phi_0 = \frac{1}{N}\sum_{i=1}^{N} \widehat{y_i} \tag{12}$$

In the last term of (11), P is the set of all IRI-influencing features with its dimension being $|P|$. ϕ_p is the attribution values of the p^{th} input feature, as shown in (13). z_{ip} is a binary variable, i.e., $z_{ip} \in \{0,1\}$, which denotes whether the p^{th} feature exists ($z_{ip} = 1$) or not ($z_{ip} = 0$) on the decision path of decision trees in the ensemble learning model.

$$\phi_p = \sum_{S \subseteq (P \setminus \{p\})} \frac{|S|!\,(|P| - |S| - 1)!}{|P|!} [f(S \cup \{p\}) - f(S)] \tag{13}$$

In (13), S is a subset of the input features excluding the p^{th} feature, with its dimension being $|S|$. $f(S)$ denotes the average model output of all samples with the feature set S. $f(S \cup \{p\})$ represents the average model output of all samples with the union of the feature set S and the p^{th} feature utilized. $\frac{|S|!(|P|-|S|-1)!}{|P|!}$ denotes the weight of the subset S. To be specific, the denominator $|P|!$ represents the number of permutations for the set of all input features. The left part of the numerator $|S|!$ denotes the number of permutations for the subset S, and the right part of the numerator $(|P|-|S|-1)!$ denotes the number of permutations for the subset $P \backslash (S \cup \{p\})$. Hence, the sum of weight of all possible S equals 1.

Further, the attribution of each feature can be considered as the summation of its main effects and the intersection effects from other features. In other words, if the attribution of all pairs of independent variables are calculated, a symmetric matrix consisting of pairwise interactions can be obtained. The interaction effects between two features (e.g., the p^{th} and q^{th} features) can be obtained by (14).

$$\phi_{p,q} = \sum_{S \subseteq (P \backslash \{p,q\})} \frac{|S|!\,(P-|S|-2)!}{2(P-1)!} \nabla_{ij}(S) \tag{14}$$

where $p \neq q$ and

$$\begin{aligned} \nabla_{pq}(S) == f(S \cup \{p,q\}) - f(S \cup \{p\}) \\ - f(S \cup \{q\}) + f(S) \\ = f(S \cup \{p,q\}) \\ - f(S \cup \{q\}) - [f(S \cup \{p\}) \\ + f(S)] \end{aligned} \tag{15}$$

The main effect of the p^{th} feature equals the intersection effects subtracting from the SHAP value for the p^{th} feature, as shown in (16).

$$\phi_{p,p} = \phi_p - \sum_{q \neq p} \phi_{p,q} \tag{16}$$

Since the SHAP intersection values are defined as the difference between the SHAP values for the p^{th} feature when the q^{th} feature is present and the SHAP values for the p^{th} feature when the q^{th} feature is absent, such decomposition is capable of unveiling significant interactions captured by the tree ensembled model.

Chapter 3. Data Preparation

3.1 Data extraction

In this study, the LTPP Standard Data Release (SDR) Number 34, released in July 2020, was utilized as the major data resource. To develop the IRI predictive model using LightGBM, 246 road sections with 1,706 observations were retrieved from Structural Factors for Flexible Pavements (SPS-1) experiment. The extracted sections covered all four climate regions (i.e., dry freeze, dry no-freeze, wet freeze, and wet no-freeze) across the U.S. (Figure 1). The Specific Pavement Study (SPS) programs were selected due to the higher data quality compared with other programs since SPS only contained new construction sections [43, 44].

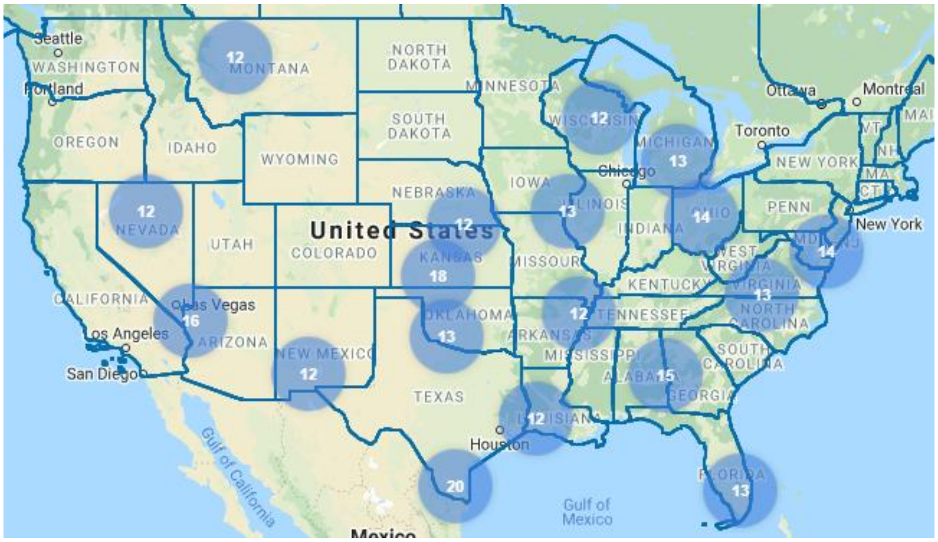
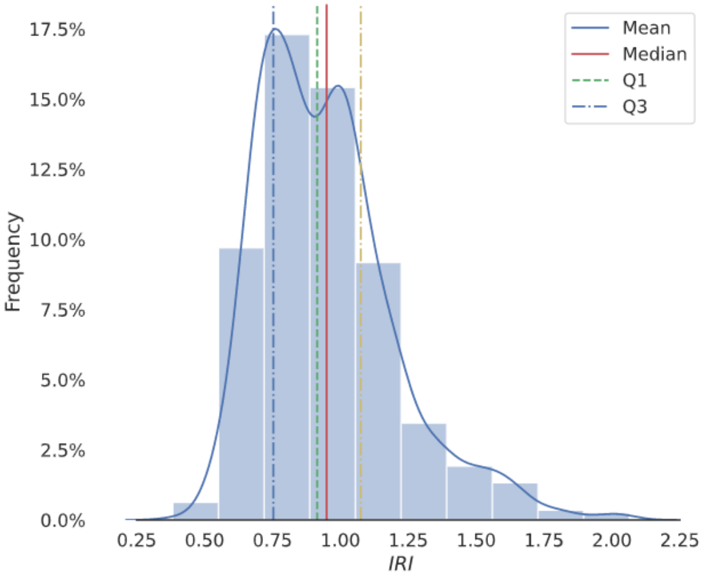


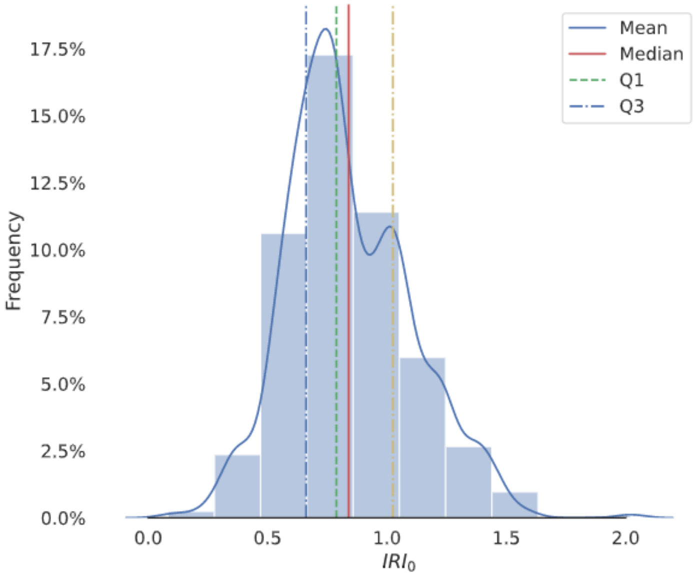
Figure 1: SPS-1 sections locations for IRI prediction

3.2 Definitions of response variable and features

In this section, the response variable (i.e., IRI), and 20 features (i.e., independent variables) that may contribute to the pavement deterioration were defined. In machine learning, features were individual independent variables that function as inputs in the system. Therein, the response variable for the model was the IRI measurement at a specific age after the traffic open date. The distribution of IRI value was shown in Figure 2-(a). An approximate normal distribution that was slightly skewed to left can be observed. Most of the sections were in good roughness condition with the mean IRI being 0.95 m/km. The third quartile was 1.08 m/km, indicating that nearly a quarter of the samples failed to pass the 1 m/km IRI limit.



(a)



(b)

Figure 2: Distribution of a) IRI and b) Initial IRI_0

TABLE I summarized all the 20 features. The features were divided into categories such as pavement performance (multiple types of distresses, rutting, and friction), pavement structure, drainage, climate conditions, traffic history and material properties. The definition and source of the response variable were also provided in the table.

TABLE I. Definition Of All Features And Response

Category	Notation	Definition and Unit	LTPP Module
Response variable	IRI	The IRI value measured at specific age since traffic open date. (m/km)	MON
Pavement performance	IRI_0	The IRI value measured when age was 0. (m/km)	MON
	Cr_{Gator}	Area of alligator cracking in square meters. (m^2)	
	Cr_{Lwp}	Length of longitudinal cracks within the defined wheel paths in meters. (m)	
	Cr_{Lnwp}	Length of longitudinal cracks not in the defined wheel paths in meters. (m)	
	Pt_A	Area of patches in square meters. (m^2)	
	Pt_N	Number of patches in square meters. (m^2)	
	Cr_{Wp}	Length of wheelpath cracks in meters. (m)	
	Cr_{Gt183}	Total length of transverse cracks greater than 1.83. (m)	
	Rt	The depth of rutting in millimeters. (mm)	
	Fr	Friction number between the vehicle wheel tire and the pavement	
Pavement structure	Tk_{Sb}	Layer thickness measurement for surface coarse and binder course. (in)	SPS
	Md_s	Average backcalculated elastic modulus of the surface layer.(psi)	FWD
Drainage	$Hydr$	Average measured hydraulic conductivity of the	DRAIN

		specimen. (cm/sec)	
Climate	$Prcp$	Average monthly precipitation in millimeters. (mm)	AWS
	Fz	Average freeze index. (℃/day)	
Traffic	$Esal$	Annual average ESAL ($kESAL$)	TRF
	$Esal_q$	quadratic form of Kesal ($kESAL^2$)	
	Age	Time duration between new construction to roughness survey date. ($year$)	
Material properties	Gr	Mean specific gravity of asphalt cement	TST
	Pt_{Ca}	Coarse aggregate amount percent by total weight of aggregate in percentage. (%)	

Among all the performance-related predictors, the initial IRI value after construction, IRI_0, was considered as the most important feature by existing studies [37, 45]. Since the first survey of pavement roughness was not always conducted right after construction, linear regression models were used to infer the initial IRI value when the initial IRI were not available in some sections. The distribution of the initial IRI was presented in Figure 2-(b). The mean value of IRI_0 was 0.84 m/km. Other predictors were selected because they were considered to have impacts on pavement smoothness. The two pavement structural parameter (Tk_{Sb} and Md_s) were assumed to influence the pavement roughness via the structure strength and deformation [46]. The predictors related to drainage ($Hydr$), climate ($Prcp$ and Fz), material properties (Gr and Pt_{Ca}), and traffic ($Esal$, $Esal_q$ and Age) were also reasonably presumed to be associated with roughness.

It should be noted that while all data were extracted from the LTPP database, some level of data processing was performed to support the modeling work. For example, in the LTPP database, the amount of cracking and patch, were stored with three categories: low, medium, and high. Therefore, for each type of distress, the lengths or areas were calculated by summing up the value of all three categories. As for the rutting and IRI variables, the values were the average of the measurements at the left and right wheel paths. In addition, pavement performance measurements including IRI, rutting, and other distresses, might not be collected during the same time interval.

Hence, the rutting and other performance data were interpolated by regression to match the observation time of IRI.

Chapter 4. MODELLING RESULTS

4.1 Model training

To train the machine learning model, the entire dataset was randomly divided into two subsets, 80% for training and 20% for testing. The training set was used to train the model parameters while the testing set was saved to evaluate the model performance. To obtain the optimal parameter set for the model, a grid search with a 10-fold cross validation was implemented. To be specific, the training set was randomly divided into ten equally sized subsets. With one subset left out iteratively as the validation set, the searching for the best combination of parameters was conducted on the other nine subsets, thus generating a performance score, i.e., mean absolute error (MAE) as defined in Eq. (17), on the validation set. After the parameters of interest were generated exhaustively from a grid of parameter values within the predefined space, the optimal parameter set was obtained if its average performance score on all validation sets reached the minimum.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y_i}| \qquad (17)$$

TABLE II listed the parameters used in the grid search, including their definitions, ranges, and the optimal values. The parameter tuning began with randomly selecting a relatively small n_estimators and a large learning_rate to ensure a high initial search speed. The basic parameters (i.e., max_depth and num_leaves) to establish the tree for each iteration can then be determined. Meanwhile, the system selected other regularization-related and Input/Output parameters to minimize overfitting and improve the computation efficiency. Those parameters included max_bin, min_data_in_leaf, feature_fraction, bagging_fraction, bagging_freq, lambda_l1, lambda_l2, and min_gain_to_split. Finally, the optimal n_estimators and learning_rate were searched jointly to further improve accuracy.

TABLE II. Parameters For The Grid Search

Parameter	Description	Space Range	Optimum
max_depth	Maximum height allowed for each tree (≥1)	min=3, max=11, step= 2	9

num_leaves	Maximum number of leaves in one tree (>1)	min=5, max=145, step= 10	35
max_bin	Maximum number of bins that feature values will be bucketed in (>1)	min=5, max=585, step= 20	265
min_data_in_leaf	Minimal number of data in one leaf (≥0)	min=0, max=50, step= 10	0
feature_fraction	Proportion of features used in each round (0<, ≤1)	min=0.6, max=1, step= 0.1	0.8
bagging_fraction	Proportion of samples used in each round (0<, ≤1)	min=0.6, max=1, step= 0.1	0.6
bagging_freq	Frequency for bagging	min=0, max=80, step= 10	0
lambda_l1	Lasso (L1) regularization term on weight (≥0)	[0, 1e-4, 0.01, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0]	0.01
lambda_l2	Ridge (L2) regularization term on weight (≥0)	[0,1e-4, 0.01, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0]	0.01
min_gain_to_split	The minimal gain to perform split (≥0)	min=0, max=1, step= 0.1	0
learning_rate	Shrinkage rate, i.e., how fast does the algorithm move in one step (>0)	[0.010, 0.015, 0.1, 0.2, 0.3]	0.015
n_estimators	Number of boosting iterations, i.e., number of trees to build in the learning algorithm (≥0)	min=0, max=5000, step=1	3,986

When all parameters reached their optimal values, the model was trained for the whole training set. Figure 3 illustrated the learning curve of the convergence process. It can be observed that as the number of trees (i.e., n_estimators) increased, the MAE value reduced significantly at the beginning and remained stable when over 1,000 base trees were fitted, indicating a fast convergence property.

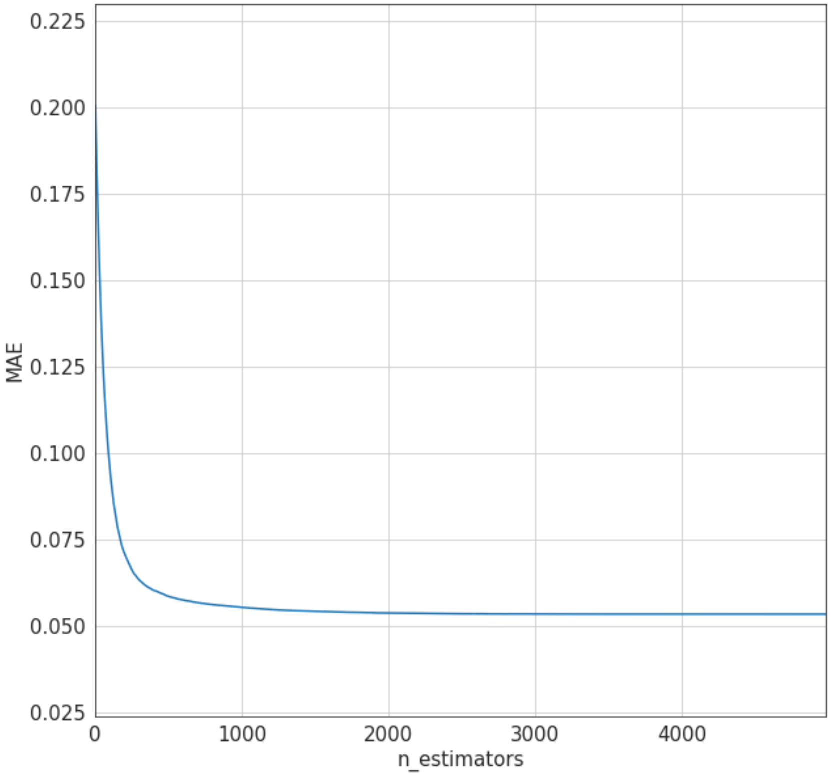
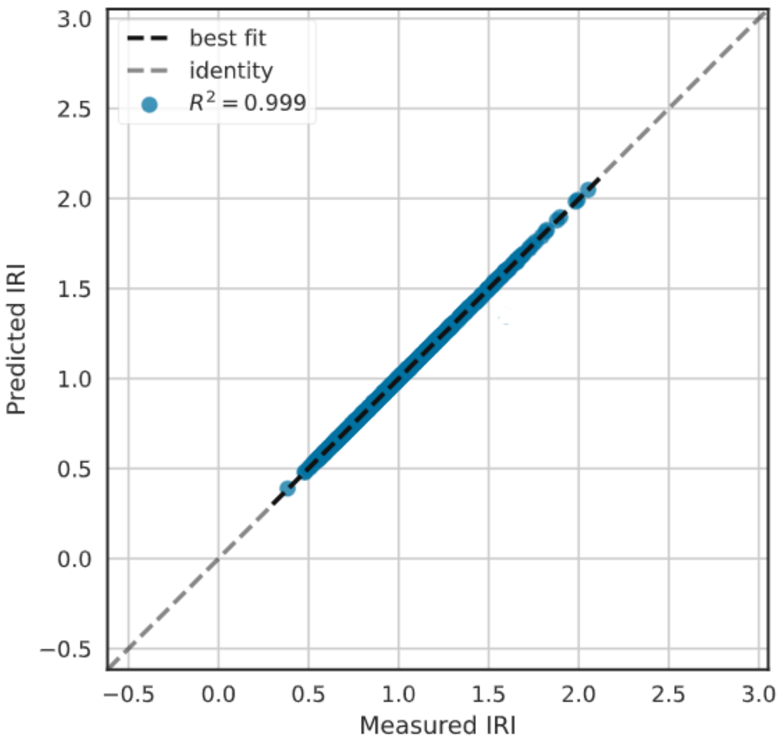


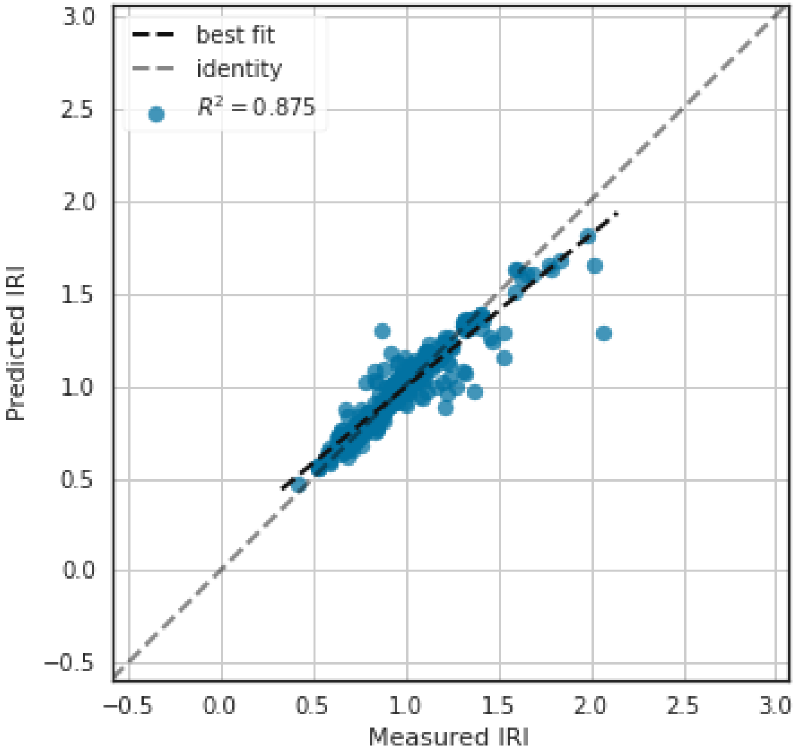
Figure 3: Learning curve of the convergence process (MAE vs. n_estimators).

4.2 Prediction accuracy

The proposed ensemble model was trained on the training set and evaluated using the testing set for its prediction accuracy. As shown in Figure 4, the R^2 value of the proposed model reached 0.999 at training set (Figure 4-a) and 0.875 at testing set (Figure 4-b), indicating a satisfactory prediction accuracy.


best fit
identity
$R^2 = 0.999$
Predicted IRI
Measured IRI
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0

(a)


best fit
identity
$R^2 = 0.875$
Predicted IRI
Measured IRI
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0

(b)

Figure 4: Comparison between the predicted and measured IRI: a) training dataset and b) testing dataset.

The residual plot and the residual distribution of the model on training and testing data were presented in Figure 5 and Figure 6, respectively. It was clear from Figure 5 that the training residuals were closer to the horizontal line than the testing residuals, which was consistent with the results presented in Figure 5. It can be seen from Figure 6 that for the training dataset, the residuals followed a normal distribution and were evenly distributed on the negative and positive sides. But for the testing residuals, the plot was slightly skewed to the negative side. In other words, the proposed model may have slightly underestimated the IRI in some cases in the testing dataset.

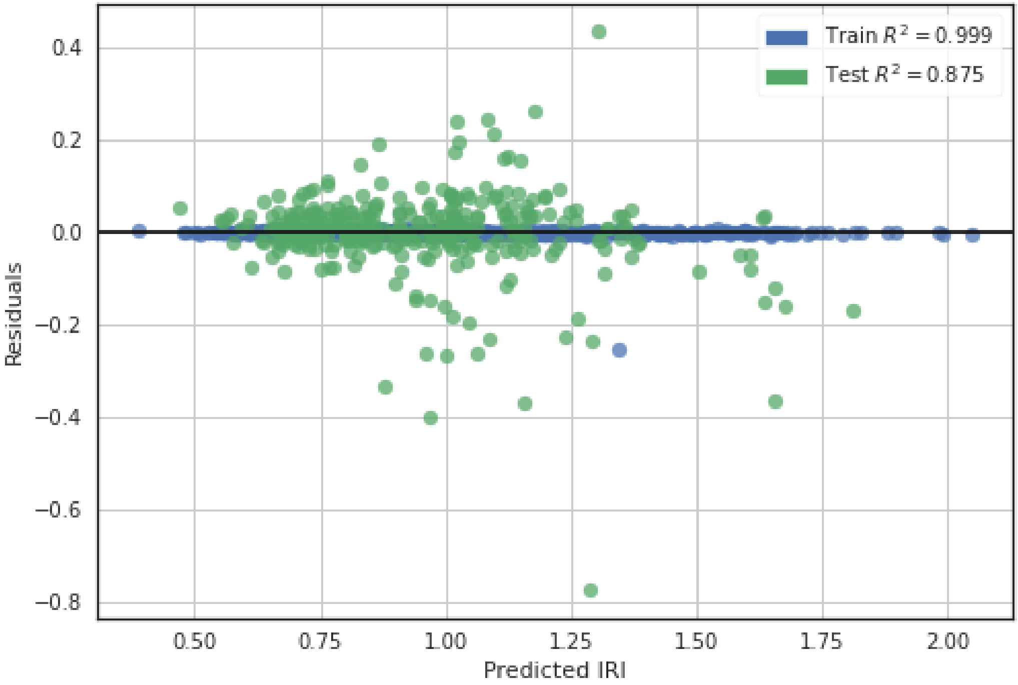
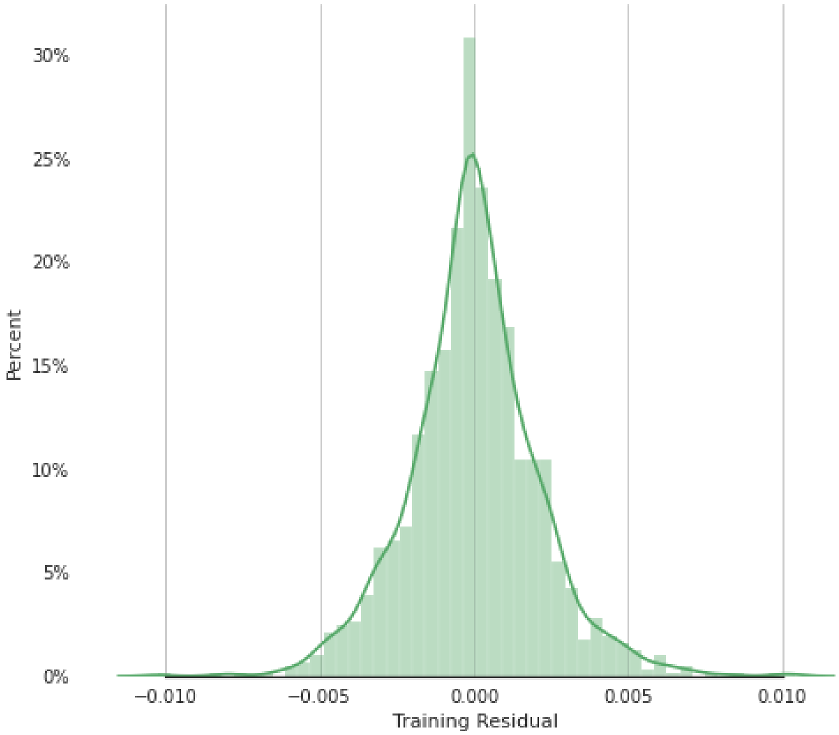
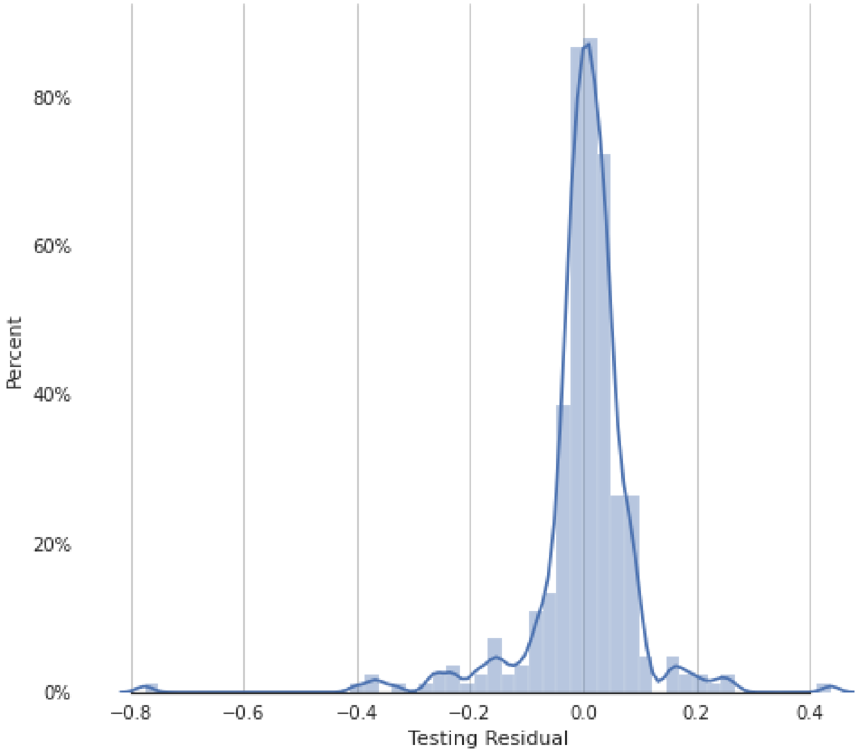


Figure 5: Residual plot of prediction results


30%
25%
20%
15%
10%
5%
0%
Percent
−0.010
−0.005
0.000
0.005
0.010
Training Residual

(a)


80%
60%
40%
20%
0%
Percent
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
Testing Residual

(b)

Figure 6: Residual distribution of a) training and b) testing.

4.3 Comparison with benchmark models

The prediction results were compared with the two benchmarked models, i.e., the MEPDG model and an ANN model. The MEPDG model is a multiple linear model considering initial IRI value, cracking, and rutting and is shown in ASSHTO [47].

$$IRI = IRI_0 + 0.0150(SF) + 0.400(FC) + 0.00080(TC) + 40.0(RD) \tag{18}$$

where IRI_0 is the initial IRI value after construction, SF is the site factor that consists of pavement age, plasticity index of the soil, average annual freezing index, and average annual precipitation, FC is the area of fatigue cracking, TC is the length of transverse cracking, and RD is the average rut depth. The other benchmarked model is the representative neural network model from Chandra et al. [32]. This ANN model is a three-layer back-propagation neural networks with the input layer consisting of five nodes of rutting, cracking, potholes, patch work, and raveling. The size of hidden layer is 15, and the output node is the IRI value.

TABLE III presents the comparison results. The comparison showed that the proposed model outperforms the other two benchmarked models in terms of predictive performance. The R^2 value was the highest among the three (32% higher than MEPDG model, and 12% higher than the ANN model). Its RMSE value was the lowest at 0.09, which was superior to the MEPDG model at 0.30 and the ANN model at 0.66. In terms of MAE values, the proposed model was at 0.05 and much lower than ANN model at 0.39. The MAE value of the MEPDG model was not available in the literature. These results demonstrated the superiority of the proposed LightBGM-based ensemble model in terms of prediction accuracy, over the simple parametric MEPDG model and the representative machine learning ANN model.

TABLE III. Performance Comparison Among The sThree Models

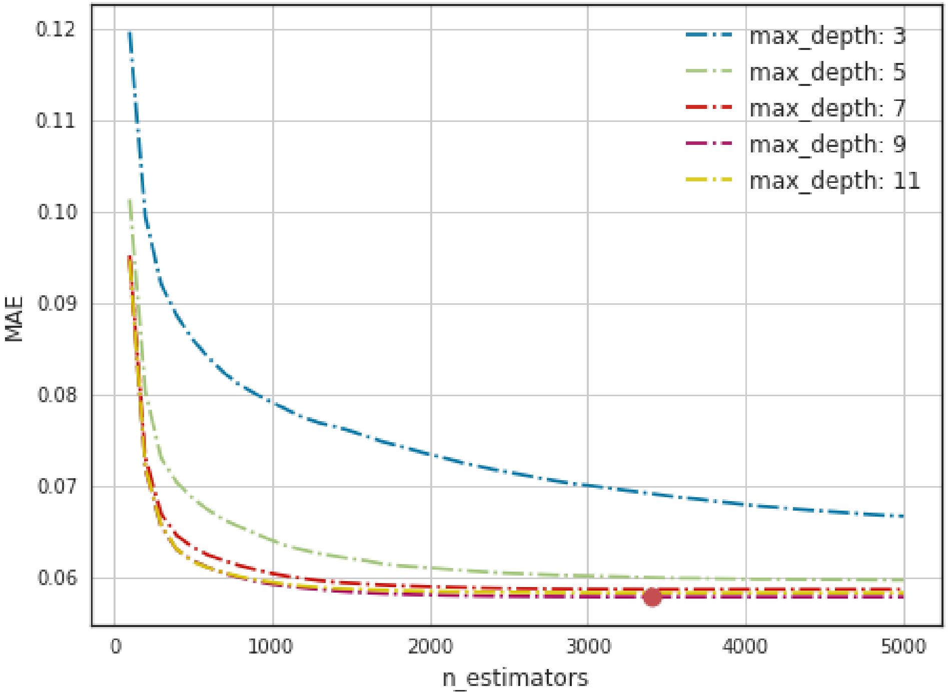
	R Square	RMSE	MAE
MEPDG	0.56	0.30	NA
Neural network	0.76	0.66	0.39

Proposed LightGBM	0.88	0.09	0.05

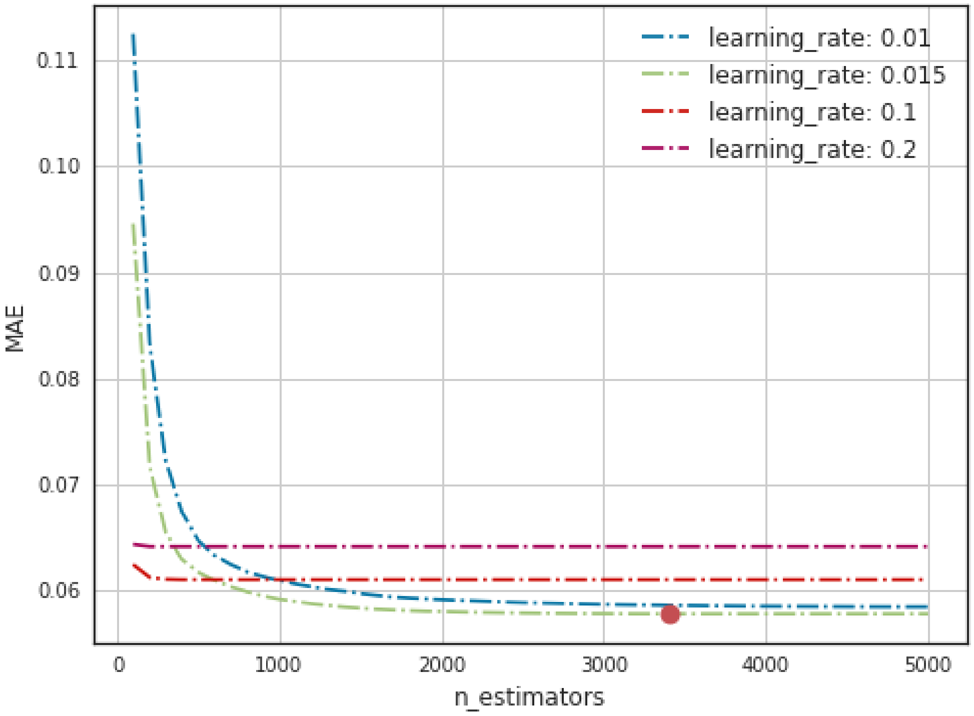
4.4 Parameter impact analysis

According to the result of the model training, two important parameters, i.e., max_depth and learning_rate, were selected for the impact analysis. Max_depth was the maximum height of each tree in the learning algorithm, and learning_rate represented the step size in reducing training loss and thus, controlled how fast the algorithm moved in one step.

For parameter impact analysis, only one parameter was allowed to change in a predefined range each time while the others were fixed with their optimal values presented in TABLE II. For example, max_depth increased from 3 to 11 with a step size of 2 while learning_rate was fixed at 0.015 and num_leaves was fixed at 35. The results were illustrated in Figure 7.



(a)



(b)

Figure 7: Parameter impact results of a) max_depth and b) learning_rate

Figure 7-a presents how the max_depth impacted the predictive performance, in which the X-axis is the number of trees trained, i.e., the number of boosting iterations, and the Y-axis is the model performance measured by MAE. The results showed that the MAE value was relatively high when max_depth was low (i.e., 3). The reason was that when base trees with only 3 levels were used, the model was not sufficient to capture the complex system dynamics, and thus, suffered from underfitting. When the max_depth increased from 3 to 5, and then to 7 and 9, the MAE value dropped accordingly. However, when max_depth reached 11, the MAE value started to increase again, indicating too many tree levels may have overfitting issues (i.e., the model worked well in the training dataset but fell short in the testing dataset). When max_depth equaled 9, the prediction capability of the model reached the optimum. In Figure 7-(b), when the learning_rate decreased from 0.2 to 0.1, and then to 0.015, the MAE value dropped accordingly because more searches can be conducted with a smaller learning rate before the model reached convergence. However, when learning_rate further reduced to 0.01, the MAE value increased again. When the learning_rate was set to 0.015, the model generated the best performance. The optimum of model parameters was considered consistent and robust because the cross validation technique was applied in the training

dataset and random sampling method was adopted to obtain the testing dataset.

4.5 Feature interpretation

The SHAP value of each feature was calculated to measure its explicit impact on IRI. For each feature, the SHAP values of all training samples were averaged. The mean SHAP value of one feature represented the average impact of this feature on the model output, which was considered as the corresponding feature importance. To illustrate the comparison results, all the average SHAP values of the twenty features were scaled with a summation value of 100 (i.e., the summation of them equaled to 100), denoted as the importance value. The importance values of the top 10 features of the ensemble learning model were presented in TABLE IV. A higher importance value meant this feature was more closely related with IRI. As listed in Table 4, these top 10 features corresponded to 87.25% of the total contribution.

TABLE IV. Importance of top 10 features of the ensemble learning model

Features	Importance
IRI_0	34.35
Cr_{Lwp}	10.08
$Esal$	9.32
Rt	7.66
Cr_{Wp}	5.07
Pt_A	4.53
Cr_{Lnwp}	4.30
Age	4.10
Cr_{Gt183}	4.03
Gr	3.80

TABLE IV showed that the top five features, i.e., IRI_0, Cr_{Lwp},$Esal$, Rt, and Cr_{Wp}, had more significant impact on IRI values than the others. Among them, IRI_0 was the most important feature, with its contribution more than triple of the second significant feature. This finding was consistent with prior studies [37, 45]. The other two pavement performance features, Cr_{Lwp} and Rt, ranked the second and the fourth, accounted for 17.74% of the total contribution. It was notable that pavement performance features, especially the ones measured at the wheelpath, i.e., Cr_{Lwp}, Rt and Cr_{Wp}, had strong impacts on the development of IRI. It was worth mentioning that the top 5 significant features except $Esal$ were in agreement with the important input variables of the MEPDG model, which validated the effectiveness of the proposed model. This ranking showed that the IRI value was mostly affected by the pavement performance features.

Figure 8 illustrates the beeswarm plot of the 10 most important features to further understand the relationship between each feature and IRI. In each row, the contribution of that feature to an observation was represented by a single dot. The position of each dot on the X-axis was determined by the SHAP value of that feature for the corresponding observation. In other words, a negative SHAP value would place the dot on the left side of the figure, and a positive SHAP value would place it on the right side. The density of samples was represented by the locations of the piles of the dots; namely, when more dots were located together, the pile was larger. The color represented the original value of the feature. Red color was used when the feature value was high, and blue color represented the low feature values.

In Figure 8, it can be observed that for observations with high IRI_0 values (i.e., the red dots of the first row), their contributions to IRI value were positive (i.e., the red dots were mostly located on the right side of the figure), indicating that a higher initial IRI value at the construction phase lead to a higher IRI. Similar trends can also be found for the features related to pavement performance, including Cr_{Lwp}, Rt, Cr_{Wp}, Pt_A, Cr_{Lnwp}, and Cr_{Gt183}. This finding suggested that there were positive relationships between the IRI value and the amount of pavement distress; in other words, the more distress, the lower pavement smoothness. This pattern was reversed for $Esal$, in which high feature values (red color dots) were located on the left-side of the diagram, which indicated that observations with higher $Esal$ values generated low negative SHAP values. It might seem counterintuitive that heavy traffic was negatively correlated with IRI. One possible reason was that the construction quality of the pavements with high design traffic volume was better than those of the roads with lower functional classes. A similar negative trend between pavement performance deterioration and the design traffic volume has been reported in other studies [48], which pointed out that the 1993 AASTHO pavement design method tended to over-design the pavement thickness for the given design traffic volume.

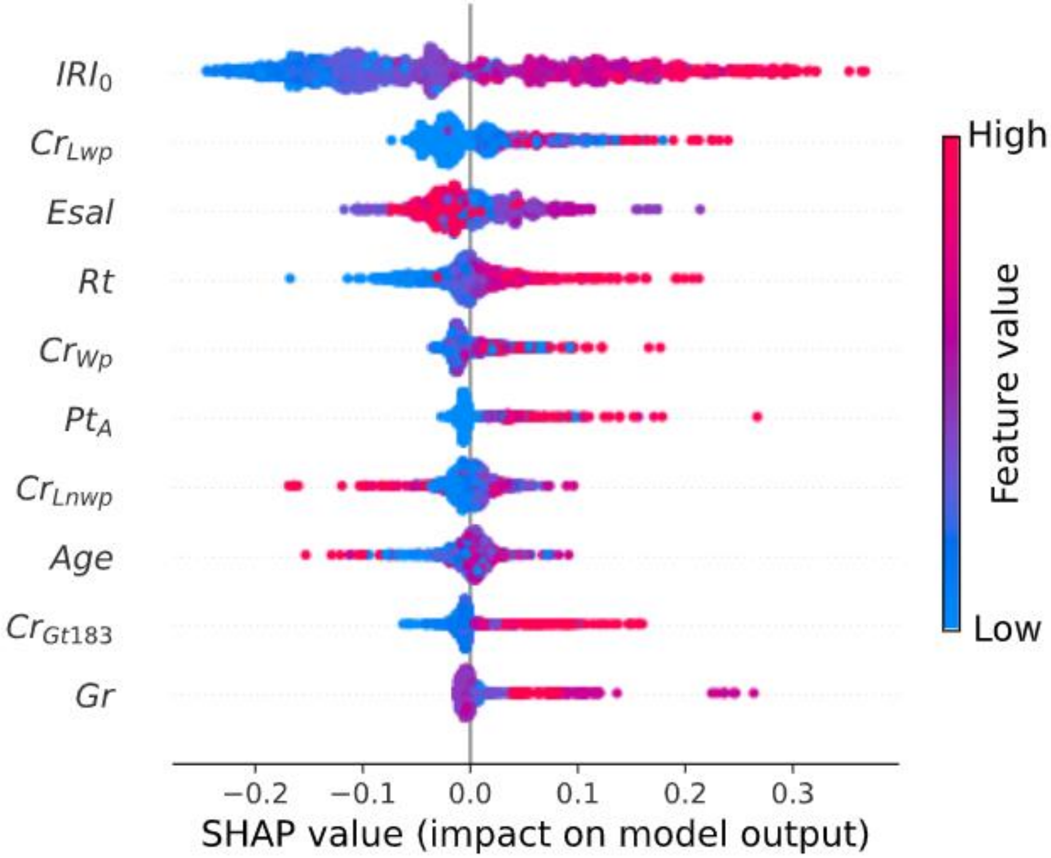


Figure 8: SHAP beeswarm plot of the top 10 important features

As discussed in the methodology section, the SHAP value can also interpret the interaction effects between two features, which can uncover some implicit but important attributions to the IRI values. For example, Figure 9 presents the SHAP dependence plot of IRI_0 and Pt_A. The X-axis is the value of I_0, and the Y-axis is the SHAP value for IRI_0. The color of the dots represents the value of the patch area. To be specific, the red color represents large patch areas, and the blue color is used for the data with small patch areas. It can be observed that when IRI_0 was low (i.e., in the range of 0.0~1.0) and the patch area was small (i.e., those with blue color), most SHAP values were negative between -0.2~0.0, meaning IRI_0 had lower impact on the pavement deterioration. This was because these pavements were in relatively good initial condition and thus, did not deteriorate significantly. When IRI_0 increased (i.e., in the range of 1.0~1.5), the observations with large patch areas (red dots) had lower SHAP values compared with ones with smaller patch areas or with less maintenance work. This trend suggested that for the road sections with high IRI_0, if better maintenance with more patches were performed, the road condition would deteriorate much slower than those without proper maintenance. Hence, more maintenance efforts should be devoted to the road sections with unsatisfactory smoothness, which can help slow the deterioration of the riding quality at a low cost.

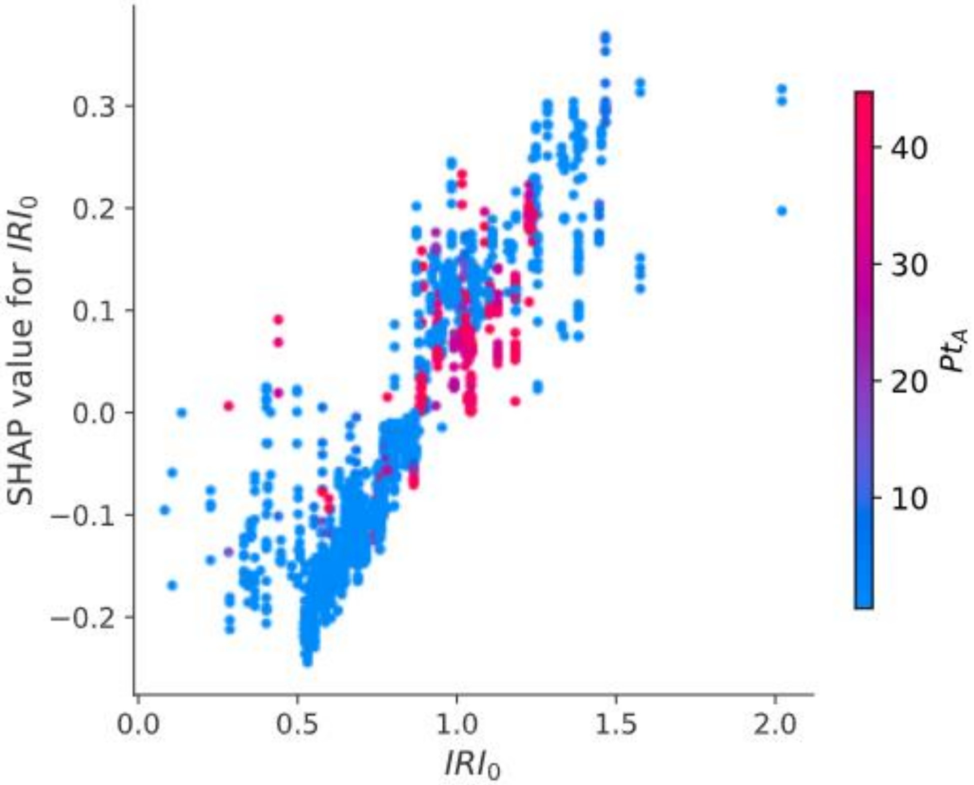
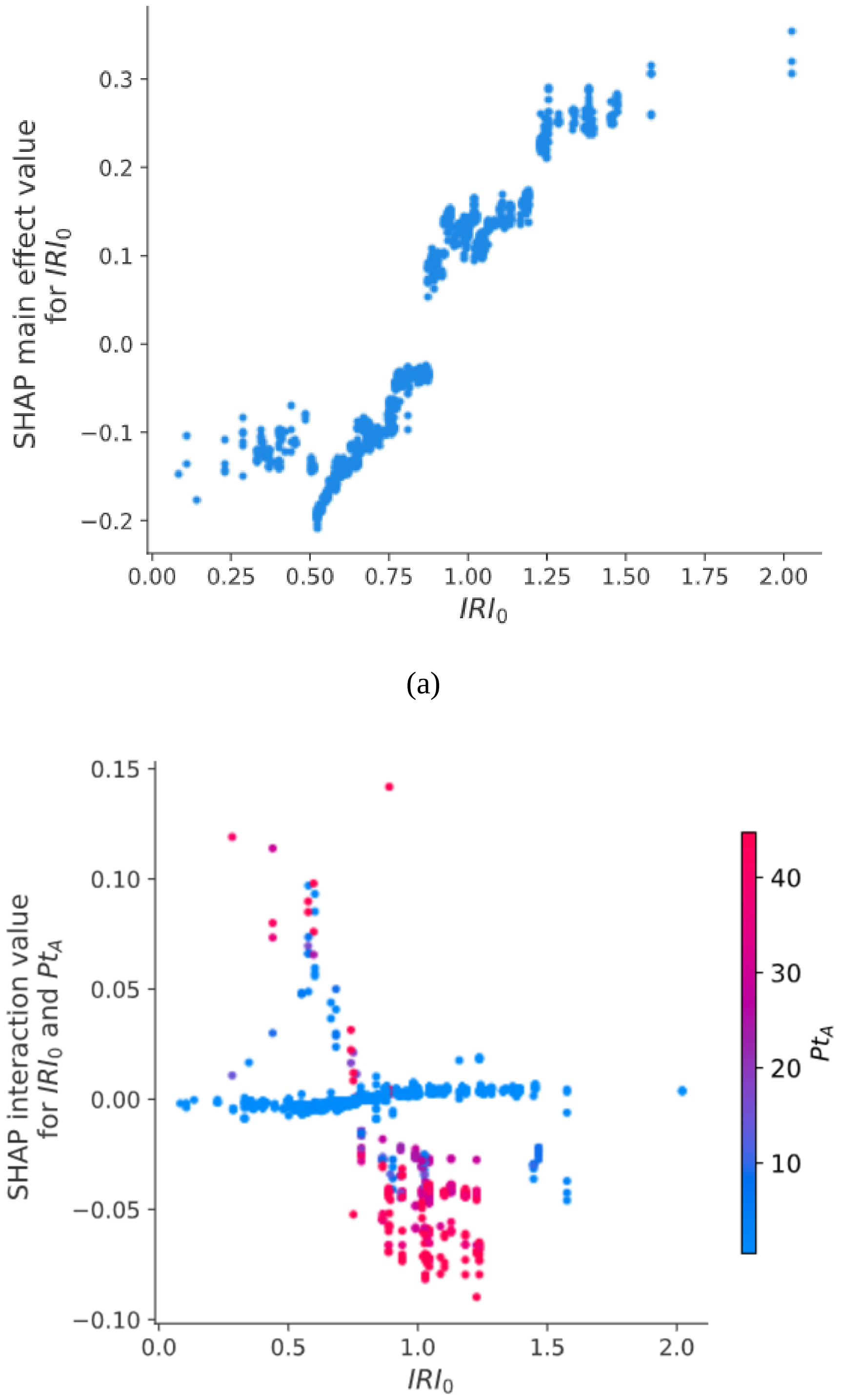


Figure 9: SHAP dependence plot of IRI_0 and Pt_A

The SHAP dependence plot in Figure 9 can be decomposed into the main effects of IRI_0 in Figure 10-(a) and the interaction effects between IRI_0 and Pt_A in Figure 10-(b) to unveil their interaction effects. The main effect was calculated by (16), which consistently presented the positive correlation between IRI_0 and IRI value, as suggested by Figure 10-(a). The interaction effects were obtained by (14) and shown in Figure 10-(b), which accounted for most of the vertical dispersion in Figure 9. It can be observed that when the patching area was low (i.e., blue dots), the interaction between the two variables had minor effects on pavement roughness regardless of the value of the IRI_0. However, when both the patching area the IRI_0 value were high, the final IRI was negatively affected by the interaction. The analysis results indicated that when the initial pavement condition was poor, the patching helped to improve pavement roughness.

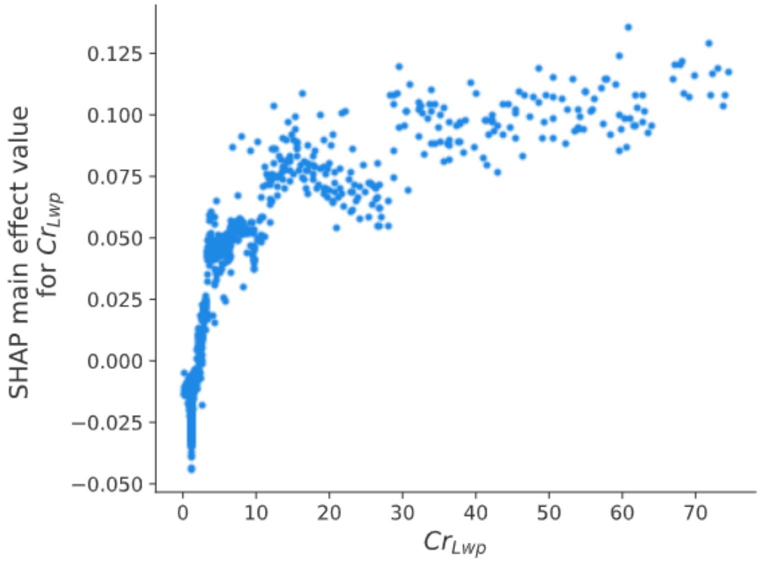


(a)

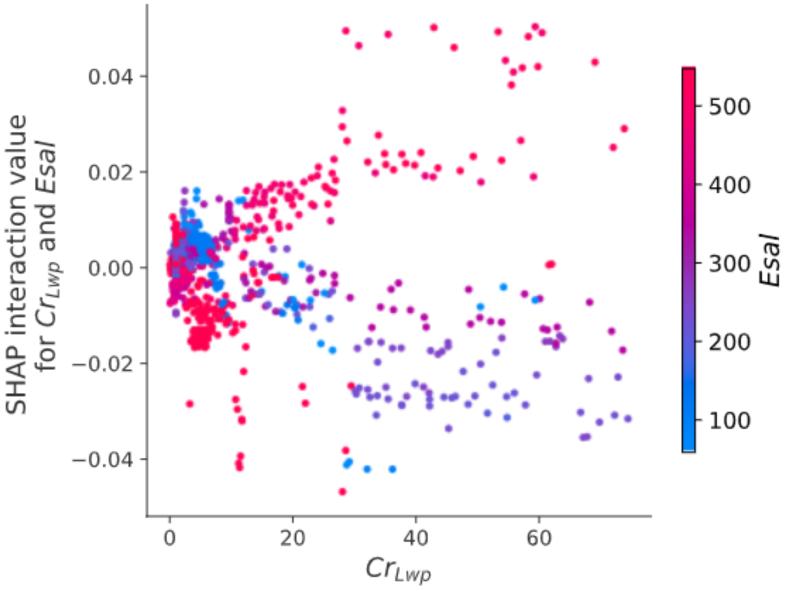
(b)

Figure 10: Separation of SHAP values of IRI_0 and Pt_A with a) Main effects and b) Interaction effects

Similarly, interaction effects between other independent variables are shown in Figure 11 and Figure 12. Figure 11-(a) and Figure 12-(a) present the positive relationships between IRI value and two features, i.e., Cr_{Lwp} and Rt, respectively, which was consistent with the finding from Figure 8. Further, Figure 11-(b) indicated that when Cr_{Lwp} was high, a greater number of $Esals$ would lead to a higher IRI value. Figure 12-(b) showed that when the rutting was high, the increases in Cr_{Wp} lead to a higher IRI value.

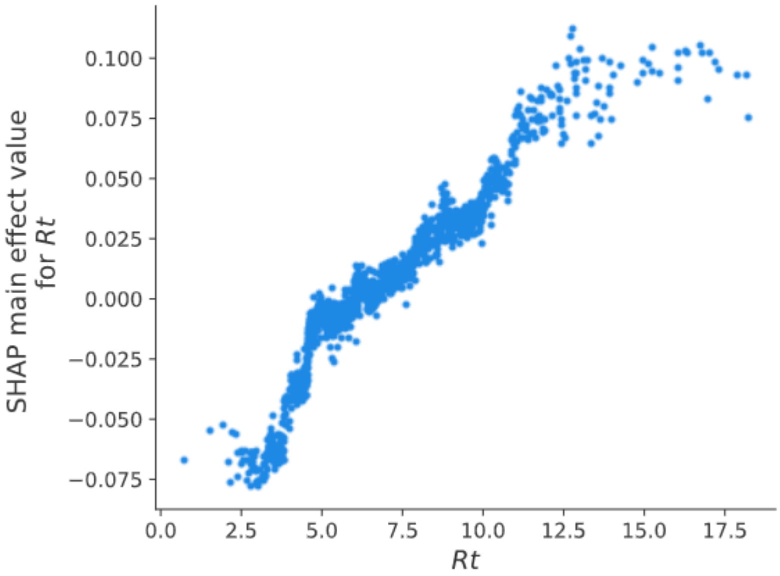


(a)

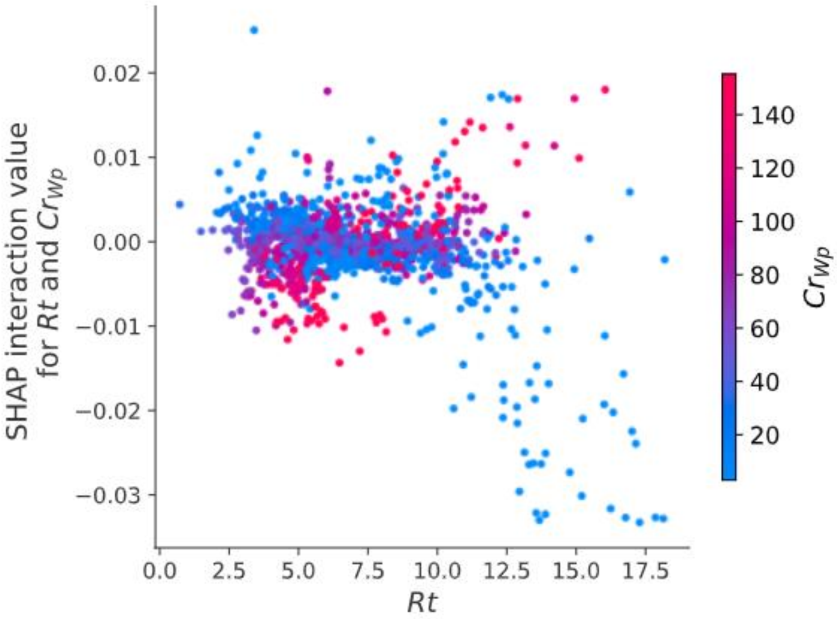


(b)

Figure 11 Separation of SHAP values of Cr_{Lwp} and $Esal$ with a) Main effects and b) Interaction effects



(a)



(b)

Figure 12 Separation of SHAP values of Rt and Cr_{Wp} with a) Main effects and b) Interaction effects

4.6 Model Simplification

With the importance of each feature interpreted, the researchers have trained the model with only the top five features, i.e., IRI_0, Cr_{Lwp}, $Esal$, Rt, and Cr_{Wp}. The model was simplified so that it can be easily implemented by pavement engineers and highway agencies. The R^2 in the training dataset remained high at 0.99, and the R^2 in verification reduced slightly from 0.88 to 0.84. These results indicated proper training using the proposed LightGBM-based ensemble learning model can provide sufficient accuracy in the IRI prediction. This finding suggested that the proposed model was able to reduce the workload dramatically for pavement engineers in terms of data collection compared with other predictive models.

Chapter 5. CONCLUSIONS

In this study, a LightGBM-based ensemble learning model was developed to predict IRI value using data from the LTPP database. 20 features related to pavement performance, structure, drainage, climate, traffic, and material properties were defined as independent variables. 1,706 observations extracted from the LTPP database were used to train and validate the predictive model. The R^2 of the model reached 0.999 on the training dataset and 0.875 on the testing dataset. The RMSE of the model was 0.09, and the MAE was 0.05. The result indicated that the LightGBM-based ensemble learning model outperformed the MEPDG model and the ANN-based model.

The modelling results were further interpreted with the SHAP method, which produced consistent feature attribution by averaging all possible orderings of input feature permutations considering all samples. The result showed that the IRI_0 was the most critical feature in estimating IRI value. Strong positive relationships were found between IRI and the performance-related significant predictors measured at wheelpath, including Cr_{Lwp}, Rt, and Cr_{Wp}. In addition, the SHAP values were used to interpret the interaction effects between two features, which provided valuable insights for pavement maintenance practices. Moreover, a simplified predictive model with the top five features, i.e., IRI_0, Cr_{Lwp}, $Esal$, Rt, and Cr_{Wp} was developed. The model was accurate and effective to predict the pavement IRI values and practical for further implementations.

Chapter 6. DATA AVAILABILITY

All data used during the study can be downloaded from https://infopave.fhwa.dot.gov/Data/StandardDataRelease/.

Chapter 7. References

1. Fan, W. and T. Yang, *Impact of Connected and Autonomous Vehicles on Signalized Intersections With Transit Signal Priority*. 2022, University of North Carolina at Charlotte. Center for Advanced Multimodal ….
2. Yang, T. and W. Fan, *Evaluation of transit signal priority at signalized intersections under connected vehicle environment.* Transportation Planning and Technology, 2023. **46**(2): p. 145-159.
3. Yang, T. and W. Fan, *Transit Signal Priority under Connected Vehicle Environment: Deep Reinforcement Learning Approach.* Journal of Intelligent Transportation Systems, 2024: p. 1-13.
4. Yang, T. and W. Fan, *Enhancing Robustness of Deep Reinforcement Learning Based Adaptive Traffic Signal Controllers in Mixed Traffic Environments Through Data Fusion and Multi-Discrete Actions.* IEEE Transactions on Intelligent Transportation Systems, 2024: p. 1-13.
5. Yang, T., *Transit Signal Priority Control with Connected Vehicle Technology: Deep Reinforcement Learning Approach*. 2024, The University of North Carolina at Charlotte.
6. Tang, Q., Y. Cheng, X. Hu, C. Chen, Y. Song, and R. Qin, *Evaluation Methodology of Leader-Follower Autonomous Vehicle System for Work Zone Maintenance.* Transportation Research Record: Journal of the Transportation Research Board, 2020.
7. Tang, Q., X. Hu, and R. Qin, *Development of Operation Guidelines for Leader-Follower Autonomous Maintenance Vehicles at Work Zone Locations.* Transportation Research Record Journal of the Transportation Research Board, 2021.
8. Tang, Q., X. Hu, and H. Yang, *Identification of Operational Design Domain for Autonomous Truck Mounted Attenuator System on Multilane Highways.* Transportation Research Record, 2021: p. 03611981211061555.
9. Tang, Q. and X. Hu, *A multi-state merging based analytical model for an operation design domain of autonomous vehicles in work zones on two-lane highways.* Journal of Intelligent Transportation Systems, 2022: p. 1-14.
10. Chen, C., Y. Song, X. Hu, and I.G. Guardiola, *Analysis of Electric Vehicle Charging Behavior Patterns with Function Principal Component Analysis Approach.* Journal of Advanced Transportation, 2020. **2020**: p. 8850654.
11. Song, Y. and X. Hu, *Learning-based demand-supply-coupled charging station location problem for electric vehicle demand management.* Transportation Research Part D: Transport and Environment, 2023. **125**: p. 103975.
12. Song, Y. and X. Hu, *Learning electric vehicle driver range anxiety with an initial state of charge-oriented gradient boosting approach.* Journal of Intelligent Transportation Systems, 2023. **27**(2): p. 238-256.
13. ARA, *Guide for mechanistic-empirical design of new and rehabilitated pavement structures*, In *NCHRP 1-37A Final Report*. 2004: Washington, D.C.

14. Robbins, M.M. and N.H. Tran, *A synthesis report: value of pavement smoothness and ride quality to roadway users and the impact of pavement roughness on vehicle operating costs*, In *NCAT Report 16-03* 2016, National Center for Asphalt Technology: Auburn, Alabama.

15. Wang, K.C. and W. Gong. *Automated pavement distress survey: a review and a new direction*. In *Pavement evaluation conference*. 2002.

16. Khattak, M.J., M.A. Nur, M.R.-U.-K. Bhuyan, and K. Gaspard, *International roughness index models for HMA overlay treatment of flexible and composite pavements.* International Journal of Pavement Engineering, 2014. **15**(4): p. 334-344.

17. Choi, J.H., T.M. Adams, and H.U. Bahia, *Pavement Roughness Modeling Using Back-Propagation Neural Networks.* Computer-Aided Civil and Infrastructure Engineering, 2004. **19**(4): p. 295-303.

18. Paterson, W.D., *Road deterioration and maintenance effects: Models for planning and management*. 1987.

19. Watanatada, T., W. Paterson, A. Bhandari, C. Harral, A.M. Dhareshwar, and K. Tsunokawa, *The highway design and maintenance standards model. volumes 1 and 2.* 1987.

20. Pérez-Acebo, H., A. Linares-Unamunzaga, E. Rojí, and H. Gonzalo-Orden, *IRI performance models for flexible pavements in two-lane roads until first maintenance and/or rehabilitation work.* Coatings, 2020. **10**(2): p. 97.

21. Al-Omari, B. and M. Darter, *Effect of pavement deterioration types on IRI and rehabilitation.* Transportation research record, 1995. **1505**: p. 57.

22. Mactutis, J.A., S.H. Alavi, and W.C. Ott, *Investigation of relationship between roughness and pavement surface distress based on WesTrack project.* Transportation research record, 2000. **1699**(1): p. 107-113.

23. George, K., *MDOT pavement management system: prediction models and feedback system*. 2000, Mississippi. Dept. of Transportation.

24. Hozayen, H.A. and F. Alrukaibi, *Development of acceptance measures for long term performance of BOT highway projects.* Efficient Transportation and Pavement Systems: Characterization, Mechanisms, Simulation, and Modeling; Sayed, A.-Q., Ed, 2008: p. 335-348.

25. Al-Suleiman, T.I. and A.M. Shiyab, *Prediction of pavement remaining service life using roughness data—case study in Dubai.* International Journal of Pavement Engineering, 2003. **4**(2): p. 121-129.

26. Hoel, L., J. Garber, and A.W. Sadek, *Transportation Infrastructure Engineering–A Multimodal Integration SI Edition.* Cengage Learning, United States of America, 2011: p. 112-115.

27. Ahmed, N.G., G.J. Awda, and S.E. Saleh, *Development of pavement condition index model for flexible pavement in Baghdad City*. Journal of Engineering, 2008. **14**(1): p. 2120-2135.

28. Lin, J.-D., J.-T. Yau, and L.-H. Hsiao. *Correlation analysis between international roughness index (IRI) and pavement distress by neural network*. In *82nd Annual Meeting of the Transportation Research Board*. 2003.

29. Kaya, O., H. Ceylan, S. Kim, D. Waid, and B.P. Moore, *Statistics and Artificial Intelligence-Based Pavement Performance and Remaining Service Life Prediction Models for Flexible and Composite Pavement Systems.* Transportation Research Record, 2020. **2674**(10): p. 448-460.

30. Gupta, A., P. Kumar, and R. Rastogi, *Pavement deterioration and maintenance model for low volume roads.* International Journal of Pavement Research and Technology, 2011. **4**(4): p. 195.

31. Thube, D.T., *Artificial neural network (ANN) based pavement deterioration models for low volume roads in India.* International Journal of Pavement Research and Technology, 2012. **5**(2): p. 115.

32. Chandra, S., C.R. Sekhar, A.K. Bharti, and B. Kangadurai, *Relationship between pavement roughness and distress parameters for Indian highways.* Journal of transportation engineering, 2013. **139**(5): p. 467-475.

33. Kargah-Ostadi, N. and S.M. Stoffels, *Framework for development and comprehensive comparison of empirical pavement performance models.* Journal of Transportation Engineering, 2015. **141**(8): p. 04015012.

34. Ziari, H., J. Sobhani, J. Ayoubinejad, and T. Hartmann, *Prediction of IRI in short and long terms for flexible pavements: ANN and GMDH methods.* International journal of pavement engineering, 2016. **17**(9): p. 776-788.

35. Sagi, O. and L. Rokach, *Ensemble learning: A survey.* Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018. **8**(4): p. e1249.

36. Marcelino, P., M.d. Lurdes Antunes, E. Fortunato, and M. Castilho Gomes, *Machine learning approach for pavement performance prediction.* International Journal of Pavement Engineering, 2019: p. 1-14.

37. Gong, H., Y. Sun, X. Shu, and B. Huang, *Use of random forests regression for predicting IRI of asphalt pavements.* Construction and Building Materials, 2018. **189**: p. 890-897.

38. Zhang, M., H. Gong, X. Jia, R. Xiao, X. Jiang, Y. Ma, and B. Huang, *Analysis of critical factors to asphalt overlay performance using gradient boosted models.* Construction and Building Materials, 2020. **262**: p. 120083.

39. Gong, H., Y. Sun, W. Hu, and B. Huang, *Neural networks for fatigue cracking prediction using outputs from pavement mechanistic-empirical design.* International Journal of Pavement Engineering, 2021. **22**(2): p. 162-172.

40. Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, *Lightgbm: A highly efficient gradient boosting decision tree.* Advances in neural information processing systems, 2017. **30**: p. 3146-3154.

41. Friedman, J.H., *Greedy function approximation: a gradient boosting machine.* Annals of statistics, 2001: p. 1189-1232.

42. Fujimoto, K., I. Kojadinovic, and J.-L. Marichal, *Axiomatic characterizations of probabilistic and cardinal-probabilistic interaction indices.* Games and Economic Behavior, 2006. **55**(1): p. 72-99.

43. Jia, Y., X. Dai, S. Wang, Y. Gao, J. Wang, and W. Zhou, *Evaluation of long-term effectiveness of preventive maintenance treatments using LTPP SPS-3 experiment data.* Construction and Building Materials, 2020. **247**: p. 118585.

44. Chen, X., H. Zhu, Q. Dong, and B. Huang, *Optimal thresholds for pavement preventive maintenance treatments using LTPP data.* Journal of Transportation Engineering, Part A: Systems, 2017. **143**(6): p. 04017018.

45. Von Quintus, H., A. Yau, M. Witczak, D. Andrei, and W. Houston, *Appendix OO-1: Background and preliminary smoothness prediction models for flexible pavements. Guide for mechanistic empirical design of new and rehabilitated pavement structures.* Transportation Research Board, 2001.

46. Vaillancourt, M., L. Houy, D. Perraton, and D. Breysse, *Variability of subgrade soil rigidity and its effects on the roughness of flexible pavements: a probabilistic approach.* Materials and Structures, 2015. **48**(11): p. 3527-3536.

47. ASSHTO, *Mechanistic-empirical Pavement Design Guide: A Manual of Practice*. 2008: American Association of State Highway and Transportation Officials.

48. Wang, Y.D., J. Jeong, and Y.R. Kim, *Comparison of Treatment Timing between Aggregate Base and Full-Depth Asphalt Roads.* Journal of Transportation Engineering, Part B: Pavements, 2020. **146**(4): p. 04020057.