An Efficient and Explainable Ensemble Learning Model for Asphalt Pavement Condition Prediction Based on LTPP Dataset

Yang Song, Yizhuang David Wang, Xianbiao Hu, and Jenny Liu

***Abstract*— 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 techniques. 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 manuscript, which includes a ThunderGBM-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. Twenty features were initially identified as the model inputs, and 2,699 observations were extracted from the Long-Term Pavement Performance (LTPP) database. Three benchmark models, namely the Mechanistic-Empirical Pavement Design Guide (MEPDG) model, the ANN model and the RF model, were used for comparison. The results showed that the developed model achieved a satisfactory result with a R-squared (R^2) value of 0.88 and Root Mean Square Error (RMSE) of 0.08, both better than three benchmark models. It ran 86 times and 2.3 times faster than the ANN and RF model, respectively. Feature interpretation was performed to identify the top influencing factors of IRI. The 20-feature model was further simplified based on the analysis result. The simplified model only required six features to efficiently and effectively predict IRI using the proposed ThunderGBM-based approach, which can reduce the workload in data collection and management for pavement engineers.**

***Index Terms*— Asphalt pavement, performance prediction, international roughness index (IRI), ensemble learning, long-term pavement performance (LTPP).**

I. Introduction

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" [1]. Pavements with high roughness can cause the increase of fuel consumption, cost of vehicle maintenance, traffic crash, and greenhouse gas emissions [2]. It is cost-intensive for highway agencies to manually survey and monitor the roughness of pavements frequently [3]. 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 [4], [5]. 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-squared (R^2) 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 [6]. 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 [7], [8]. More research using linear regression models can be found in Pérez-Acebo *et al.* [8], Al-Omari and Darter [9], Mactutis *et al.* [10].

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

Manuscript received 3 August 2021; revised 18 January 2022; accepted 22 February 2022. Date of publication 13 April 2022; date of current version 7 November 2022. This work was supported in part by the Center for Intelligent Infrastructure (CII), Missouri University of Science and Technology, under Seed Grant "Preparing Multidisciplinary Datahub for Digital and Intelligent Infrastructure System." The Associate Editor for this article was Y. Hou. *(Corresponding author: Jenny Liu.)*

Yang Song and Xianbiao Hu are with the Department of Civil and Environmental Engineering, The Pennsylvania State University, State College, PA 16801 USA.

Yizhuang David Wang and Jenny Liu are with the Department of Civil, Architectural, and Environmental Engineering, Missouri University of Science and Technology, Rolla, MO 65409 USA (e-mail: jennyliu@mst.edu).

Digital Object Identifier 10.1109/TITS.2022.3164596

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 [11], [12]. 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 [13], [14].

To address the issues of regression-based models, machine learning models, such as the artificial neural network (ANN) and the random forest (RF) models, 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.* [15] 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.* [5] 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.* [16] 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 [17]–[19].

Attempts have also been made to adopt other machine learning models. Kargah-Ostadi and Stoffels [20] 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.* [21]. 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 RF was developed to model IRI by generating base parallel and independent regression trees with the output being the average of each individual weak regressor [22]. 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 [23], [24]. 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.* [25] and Gong *et al.* [26]. However, the models became complicated as the number of layers in ANN or the number of trees in RF 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.

To address these issues, in this study, a ThunderGBM-based ensemble learning model coupled with the Shapley Additive Explanation (SHAP) method was proposed. The ThunderGBM model was utilized to predict the IRI of asphalt pavements, and the SHAP method was applied to interpret the impact of underlying influencing factors. 2,699 observations have been extracted from the LTPP database. The dataset was divided into two subsets, with 80% for training and 20% for testing. Three benchmark models, namely the MEPDG model, an ANN model and a RF model, were used for comparison. Feature interpretation was performed to identify the top influencing factors of IRI evolution.

II. Methodology

A. ThunderGBM-Based IRI Prediction Model

The ThunderGBM method, different from the bagging-based models (e.g., RF), 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 Wen *et al.* [27] and was developed based on the gradient boosting decision tree (GBDT) algorithm. GBDT is a typical boosting-based model with the base learners being decision trees [28]. 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 the primary GBDT, ThunderGBM applies the second order gradient statistics to optimize the objective to improve prediction accuracy and efficiency. Meanwhile, as the latest implementation of gradient boosting algorithm, ThunderGBM is Graphics Processing Units (GPU)-oriented that adopts optimization techniques including high dimensional data histogram building and feature-based data partitioning to maximize the usage of the GPU. Thus it outperforms other existing boosting based algorithms like XGBoost [29], LightGBM [30] and CatBoost [31] in terms of computational efficiency while producing similar models [27].

The training dataset is designated as $O = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, 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., 2,159 in this study). The predictor set and the response set are designated as $X = \{x_1, x_2, \ldots, x_n$ and $Y = \{y_1, y_2, \ldots, y_n$, 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).

$$\begin{aligned}\tilde{F} &= arg \min_F E_{X,Y}(Obj) \\ &= arg \min_F E_{X,Y}\left(\sum_{i=1}^{n} l(y_i, F(x_i)) + \sum_{k=1}^{K} \Omega(f_k)\right) \quad (1)\end{aligned}$$

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).

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

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

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

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).

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

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).

$$\begin{aligned}&\sum_{i=1}^{n} \left(g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i)\right) \\ &\quad = \sum_{j=1}^{T} \left(\left(\sum_{i \in \epsilon I_j} g_i\right) w_j + \frac{1}{2}\left(\sum_{i \in \epsilon I_j} h_i\right) w_j^2\right) \quad (5)\end{aligned}$$

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).

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

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):

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

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

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

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

In the above-mentioned process, the two key components that determine learning speed are (1) computing the gradient g_i and the second order derivative h_i, and (2) constructing trees. Therefore, ThunderGBM performs GPU-aware techniques to optimize the implementation of the two parts. To speed up the former, each GPU thread is dedicated to the prediction of an instance and computing its g_i and h_i, thus the massive parallelism of GPUs can be fully utilized. For the acceleration of the latter component, the feature histogram at each node is contructed by strable sort without the need to reserve memory, so the memory efficiency can be improved especially for high dimensional data. In addition, when determining the best candidate node split point, the feature-based data partitioning strategy enables accessing all values of a feature in one GPU block, not requiring to consider the histogram of other features, which reduces communication cost and increases the computational efficiency. Also, each GPU thread is dedicated to calculating the gain of every candidate split point so that GPU resources could be efficiently exploited. With all the above optimizations for the GPU-oriented training, ThunderGBM outperformed existing gradient boosting based algorithms in terms of computational speed.

B. SHAP Method for Results Interpretation

SHAP is one of the additive feature attribution methods to explain individual predictions based on the game theoretically optimal Shapley Values [32]. 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 (10).

$$\hat{y}_i = f(x_i) = \phi_0 + \sum_{p=1}^{|P|} \phi_p z_{ip} \tag{10}$$

where $\hat{y}_i$ 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 (11), where N is the number of total IRI observations (i.e., 2,699 in this study).

$$\phi_0 = \frac{1}{N} \sum_{i=1}^{N} \hat{y}_i \tag{11}$$

In the last term of (10), 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 (12). 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.

$$\phi_p = \sum_{S \subseteq (P \setminus \{p\})} \frac{|S|!\,(|P| - |S| - 1)!}{|P|!} [f(S \cup \{p\}) - f(S)] \tag{12}$$

In (12), 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 $P(S \cup \{p\})$. Hence, the sum of weight of all possible S equals 1. To conclude, the contribution of with the p^{th} input feature on the target IRI value can be measured consistently by the ϕ_p value.

III. Data Preparation

A. Data Extraction

In this study, the LTPP Standard Data Release (SDR) Number 35, released in July 2021, was utilized as the major data resource. To develop the IRI predictive model using ThunderGBM, 246 road sections with 2,699 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 [33], [34].

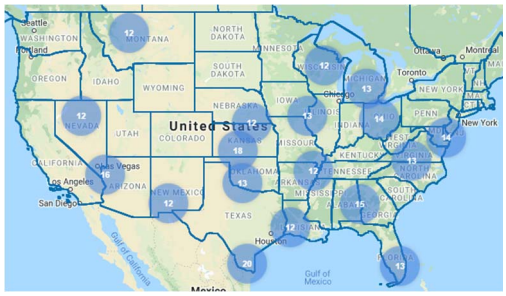


Fig. 1. SPS-1 sections locations for IRI prediction.

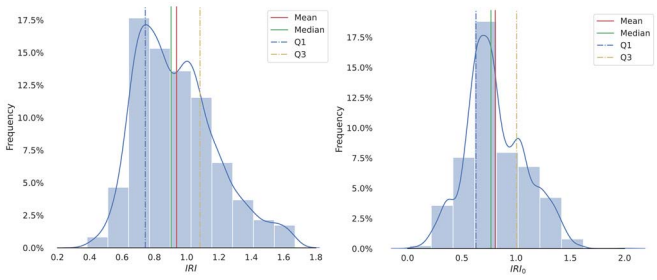


Fig. 2. Distribution of a) IRI and b) initial IRI_0.

B. 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 0.94 m/km. The third quartile was 1.08 m/km, indicating that nearly a quarter of the samples were below the 1 m/km IRI limit.

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.

Among all the performance-related predictors, the initial IRI value after construction, IRI_0, was considered as the most important feature by existing studies [24], [35]. 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

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 specimen. (cm/sec)	DRAIN
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. (%)	

in Figure 2-(b). The mean value of IRI_0 was 0.81 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 [36]. 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 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.

IV. MODELLING RESULTS

A. 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 5-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 four subsets, thus generating a performance score, i.e., Root Mean square Error (RMSE) as defined in Eq. (13), 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.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{13}$$

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_trees and a large learning_rate to ensure a high initial search speed. The basic parameters (i.e., depth) 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 min_child_weight, max_num_bin, gamma and lambda_tgbm. Finally, the optimal n_trees and learning_rate were searched jointly to further improve accuracy.

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_trees) increased, the RMSE value reduced significantly at the beginning and remained stable when over 1,000 base trees were fitted, indicating a fast convergence property.

B. 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 0.881 at testing set (Figure 4-b), indicating a satisfactory prediction accuracy.

The residual plot and the residual distribution of the model on training and testing data were presented in Figure 5 and

TABLE II
PARAMETERS FOR THE GRID SEARCH

Parameter	Description	Space Range	Optimum
depth	Maximum height allowed for each tree (⩾1)	min=3, max=7, step= 1	6
min_child_weight	Minimum sum of instance weight in a child node(⩾0)	min=1, max=4, step= 1	3
max_num_bin	Maximum number of bins that feature values will be bucketed in (>1)	min=5, max=585, step= 20	85
gamma	Minimum loss reduction needed to make a split on a leaf node (⩾0)	[0, 0.05, 0.1, 0.2, 0.3, 0.5, 1]	0
lambda_tgbm	L2 regularization term on weights (⩾ 0)	[0, 0.01, 0.1, 0.2, 0.3, 0.5, 0.7, 0.9, 1]	0.9
learning_rate	Shrinkage rate, i.e., how fast does the algorithm move in one step (>0)	[0.001, 0.002, 0.005, 0.1, 0.2]	0.1
n_trees	Number of boosting iterations, i.e., number of trees to	min=100, max=5000, step=100	3,600
	build in the learning algorithm (⩾0)		

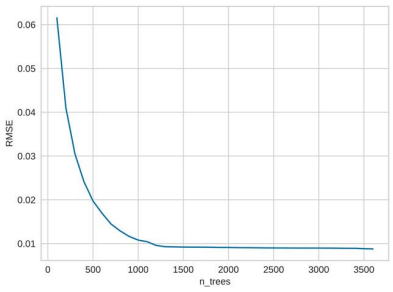


Fig. 3. Learning curve of the convergence process.

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 and testing dataset, the residuals followed a normal distribution and were evenly distributed on the negative and positive sides. In other words, the proposed model achieved good balance between underestimation and overestimation in both the training and testing dataset.

C. Comparison With Benchmark Models

The prediction results were compared with the two benchmarked models, i.e., the MEPDG model, an ANN model and a RF model. The MEPDG model is a multiple linear model considering initial IRI value, cracking, and rutting and is shown in ASSHTO [37].

$$IRI = IRI_0 + 0.0150\,(SF) + 0.400\,(FC) + 0.00080\,(TC) + 40.0(RD) \quad (14)$$

where IRI_0 is the initial IRI value after construction, $SF = Age(0.02003\,(PI+1) + 0.007947\,(Rain+1) + 0.000636(FI+1))$, is the site factor considering climate and soil variations with Age being pavement age, PI being plasticity index of the soil, $Rain$ being average annual precipitation and FI being average annual freezing index, FC is the area of fatigue cracking, TC is the length of transverse cracking, and RD is the average rut depth. The other two benchmarked models are an ANN model and a RF model. They are trained and tested on the same data set as the proposed ThunderGBM model, i.e., the input is the 20 features and output is the IRI value. The two models were fully tuned with their optimized parameters are given in TABLE III.

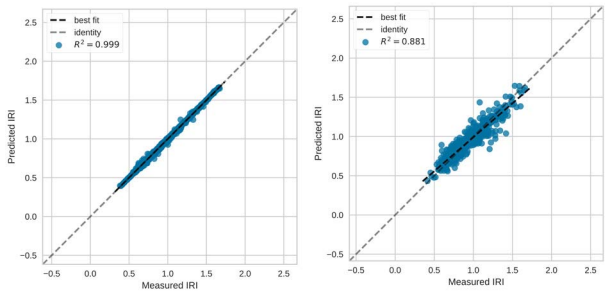


Fig. 4. Comparison between the predicted and measured IRI: a) training dataset and b) testing dataset.

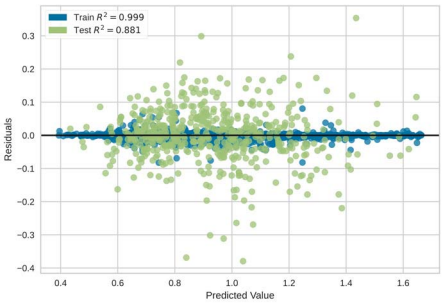


Fig. 5. Residual plot of prediction results.

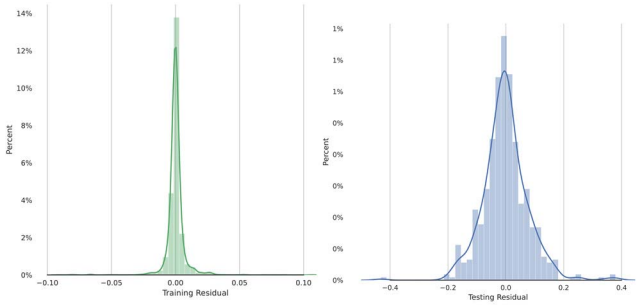


Fig. 6. Residual distribution of a) training and b) testing.

TABLE III

PARAMETERS FOR THE BENCHMARKED ANN AND RF MODELS

Model	Parameter	Optimum
ANN	hiddden_layer_sizes	300
	alpha	0.0001
	solver	adam
	learning_rate	0.001
	epoch	2500
RF	max_features	16
	max_depth	28
	n_estimator	2400

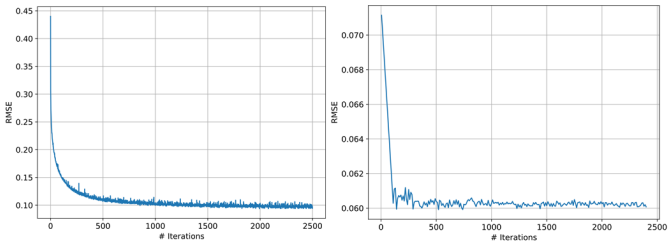
TABLE IV

PERFORMANCE COMPARISON AMONG FOUR MODELS

	R Square	RMSE	Training Time (sec)
MEPDG	0.560	0.300	0.01
ANN	0.792	0.108	451.03
RF	0.862	0.087	17.12
Proposed ThunderGBM	0.881	0.081	5.14

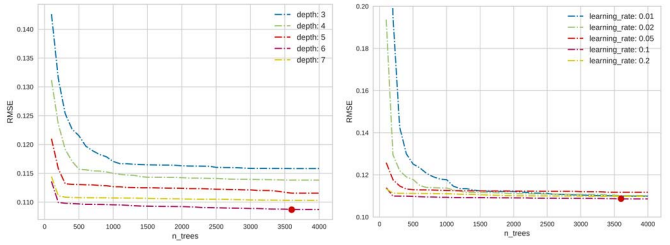
TABLE IV presents the comparison results. The comparison showed that the proposed model outperforms the other three benchmarked models in terms of predictive performance. The R^2 value was the highest among the three (32% higher than MEPDG model, 9% higher than the ANN model, 2% higher than the RF model). Its RMSE value was the lowest at 0.081, which was superior to the MEPDG model at 0.300, the ANN model at 0.108 and the RF model at 0.087. These results demonstrated the superiority of the proposed ThunderGBM-based ensemble model in terms of prediction accuracy, over the simple parametric MEPDG model, the representative machine learning ANN model and bagging-based RF model.

In terms of computational efficiency, the proposed model and the three benchemarked models were trained on a desktop computer with NVIDIA (R) Quadro (TM) RTX 5000 GPU and Intel (R) Xeon (TM) W-2295 CPU respectively. TABLE IV suggests that the linear MEPDG model ran the fastest due to its simplicity in modeling. When compared with the two machine learning models, the proposed ThunderGBM model could finish training with just 5 seconds, whereas the ANN model took over 450 seconds and RF model needed over 15 seconds. It could be demonstrated that the proposed model was 86 times and 2.3 times faster than the ANN and RF models, respectively. Aside from the ability of ThunderGBM to utilize GPU computing power, such difference in computational speed could explained by the different ways of training error reduction in each iteration. The update of node weight of the ANN model may increase training error in certain iteration if learning rate was larger than optimum. The bagging-based RF model trained each weak regression tree independently so some underfitted trees may be generated in some iterations. On the other hand, the proposed boosting-based ThunderGBM model sequentially trained each sub-learner based on the residual of the previous regression tree, thus always reducing training error when a new base tree was aggregated. Since the training loss was keeping decrease as iteration grew, the model converged faster than the other two machine learning



(a) (b)

Fig. 7. Learning curve of the convergence process (RMSE vs. # iterations). a) ANN and b) RF.



(a) (b)

Fig. 8. Parameter impact results of a) depth and b) learning_rate.

benchmarked models. Such characteristics could be illustrated by the learning curves given in Figure 3 and Figure 7.

In Figure 7-a and Figure 7-b, as the increase of iterations, the training loss, i.e., RMSE, of the ANN and RF model overall dropped until convergence, though in certain round, the training error may increase slightly due to the loss descending characteristics of the two models. In contrast, in Figure 3, the training error of the boosting-based model always kept decreasing smoothly without any oscillation when the optimal number of base learners were trained. Hence, compared with the two benchmarked machine learning models, the proposed model could ensure a faster and stabler training progress until the optimum was reached.

D. Parameter Impact Analysis

According to the result of the model training, two important parameters, i.e., depth and learning_rate, were selected for the impact analysis. 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, depth increased from 3 to 1 with a step size of 1 while learning_rate was fixed at 0.1 and min_child_weight was fixed at 3. The results were illustrated in Figure 8.

Figure 8-a presents how the 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 RMSE. The results showed that the RMSE value was relatively high when depth was low (i.e., 3). The reason was that when base trees with

TABLE V
IMPORTANCE OF TOP 10 FEATURES OF THE ENSEMBLE LEARNING MODEL

Features	Importance
IRI_0	25.17
Rt	11.94
Cr_{Lwp}	8.48
Pt_A	7.99
Fr	7.97
Cr_{Lnwp}	6.88
Cr_{Wp}	5.28
Cr_{Gator}	5.18
Age	4.95
$Esal$	4.20

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 4, and then to 5, the RMSE value dropped accordingly. However, when max_depth reached 7, the RMSE 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 6, the prediction capability of the model reached the optimum. In Figure 8-(b), when the learning_rate decreased from 0.2 to 0.1, the RMSE 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 0.05, the RMSE value increased again. When the learning_rate was set to 0.1, 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.

E. 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 V. A higher importance value meant this feature was more closely related with IRI. As listed in TABLE V, these top 10 features corresponded to 88.51% of the total contribution.

TABLE V showed that the top three features, i.e., IRI_0, Rt, and Cr_{Lwp}, had more significant impact on IRI values than the others, with their important over 8.00%. Among them, IRI_0 was the most important feature, with its contribution more than double of the second significant feature. This finding was consistent with prior studies [24], [35]. The other two pavement performance features, Rt and Cr_{Lwp}, ranked the second and the third, accounted for 20.42% of the total contribution. It was notable that pavement performance features, especially the ones measured at the wheelpath, i.e., Rt, Cr_{Lwp}, Fr and Cr_{Wp}, had strong impacts on the development of IRI. It was worth mentioning that the top three significant features 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 9 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 9, 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 other features related to pavement distress, including Rt, Cr_{Lwp}, Pt_A, Cr_{Lnwp}, Cr_{Wp} and Cr_{Gator}. This finding suggested that there were positive relationships between the IRI value and the amount of pavement distresses; 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. This has been validated by the fact that the samples represented by red color dots on the left-side of the diagram had an average layer thickness of surface course and binder course of 5.63 inch, while that for samples represented by blue color dots on the right-side of the diagram was 4.96 inch. Thus generally there was a negative correlation between Esals and IRI due to the reason that pavement with heavy traffic was better designed and constructed with thicker layer. A similar negative trend between pavement performance deterioration and the design traffic volume has been reported in other studies [38], which pointed out that the 1993 AASHTO pavement design method tended to over-design the pavement thickness for the given design traffic volume.

Based on the relationship between each important feature and IRI explained by SHAP values, effective measures could be suggested to guide pavement maintenance practices. Give the positive impact of pavement performance related features, especially the ones measured at the wheelpath on

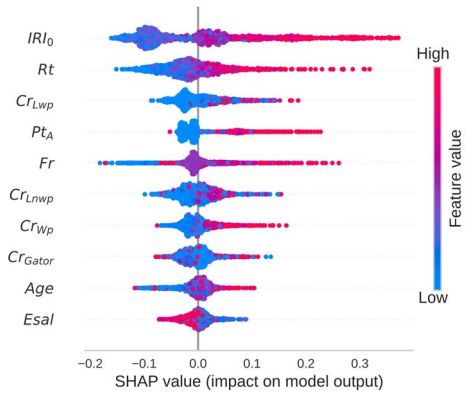


Fig. 9. SHAP beeswarm plot of the top 10 important features.

IRI, more maintenance efforts should be devoted to distresses at wheelpath. For example, cracks at wheelpath should be address timely, preferably with sealing than patches since Cr_{Lwp}, Cr_{Wp} and Pt_A are found to be significantly positively related to IRI. Besides, microsurfacing treatments may be adopted for it, which was found to be capable to reduce rut depth and cracks, thus decreasing IRI value [39]. In addition, as suggested by the relationship between Esals and IRI, if the design index like pavement thickness is more elevated with respect to design traffic volume, the deterioration of pavement roughness will also be slowed.

F. Model Simplification

With the importance of each feature interpreted, the researchers have trained the model with only six features from the top 10 features, including the basic information, i.e., Age, and wheelpath related features, i.e., IRI_0, Rt, Cr_{Lwp}, Fr and Cr_{Wp}. The model was simplified so that it can be easily implemented by pavement engineers and highway agencies. To specify, in each survey, only evaluating wheelpath condition would be enough to support the estimation of IRI value. The R^2 in the training dataset remained high at 0.99, and the R^2 in verification reduced slightly from 0.88 to 0.83. These results indicated proper training using the proposed ThunderGBM-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.

V. Conclusion

In this study, a ThunderGBM-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. 2,699 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 0.881 on the testing dataset. The RMSE of the model was 0.09 with training time being 5.14 sec. The result indicated that the ThunderGBM-based ensemble learning model outperformed the MEPDG, the ANN, and RF models in terms of prediction performance, and ran 86 and 2.3 times faster than the latter two machine learning benchmarked models, respectively.

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 Rt, Cr_{Lwp}, Fr and Cr_{Wp}. The findings from SHAP methods provided valuable insights for pavement maintenance practices. For highway agencies, preventive maintenance such as cracking sealing early-stage cracks along wheel path would be cost-effective to remain a low IRI value throughout the pavement service life, while microsurfacing treatments may be adopted if rutting and cracks occur at the same time. Moreover, a simplified predictive model with the six features, i.e., Age, IRI_0, Rt, Cr_{Lwp}, Fr and Cr_{Wp}, was developed, in which the last four features are wheelpath condition-related and can be predicted by any pavement performance analysis program. The model was able to provide pavement IRI predictions with improved accuracy and effectiveness, which will better guide pavement engineers to plan their maintenance treatments accordingly with the consideration of the predicted IRI values.

Data Availability

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

Acknowledgment

Preliminary results were presented at the Center for Intelligent Infrastructure (CII) research meeting on December 15, 2020.

References

[1] ARA Inc, "Guide for mechanistic-empirical design of new and rehabilitated pavement structures," Transp. Res. Board Nat. Acad., Washington, DC, USA, Final Rep. NCHRP 1-37A, 2004.

[2] M. M. Robbins 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," Nat. Center Asphalt Technol., Auburn, AL, USA, NCAT Rep. 16-03, 2016.

[3] K. C. Wang and W. Gong, "Automated pavement distress survey: A review and a new direction," in *Proc. Pavement Eval. Conf.*, 2002, pp. 21–25.

[4] M. J. Khattak, M. A. Nur, M. R.-U.-K. Bhuyan, and K. Gaspard, "International roughness index models for HMA overlay treatment of flexible and composite pavements," *Int. J. Pavement Eng.*, vol. 15, no. 4, pp. 334–344, Apr. 2014.

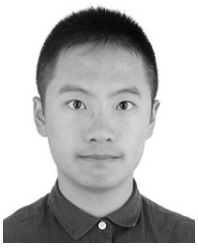
[5] J.-H. Choi, T. M. Adams, and H. U. Bahia, "Pavement roughness modeling using back-propagation neural networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 19, no. 4, pp. 295–303, Jul. 2004.

[6] W. D. Paterson, *Road Deterioration and Maintenance Effects: Models for Planning and Management*. Washington, DC, USA: World Bank, 1987.

[7] T. Watanatada *et al.*, *The Highway Design and Maintenance Standards Model*, vols. 1–2. Washington, DC, USA: World Bank, 1987.

[8] H. Pérez-Acebo, 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*, vol. 10, no. 2, p. 97, Jan. 2020.

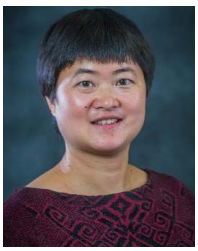
[9] B. Al-Omari and M. Darter, "Effect of pavement deterioration types on IRI and rehabilitation," *Transp. Res. Rec.*, vol. 1505, no. 136, p. 57, 1995.
[10] J. A. Mactutis, S. H. Alavi, and W. C. Ott, "Investigation of relationship between roughness and pavement surface distress based on WesTrack project," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1699, no. 1, pp. 107–113, Jan. 2000.
[11] H. A. Hozayen and F. Alrukaibi, "Development of acceptance measures for long term performance of BOT highway projects," in *Efficient Transportation and Pavement Systems: Characterization, Mechanisms, Simulation, and Modeling*, A.-Q. Sayed, Ed. London, U.K.: Taylor & Francis. 2008, pp. 335–348.
[12] T. I. Al-Suleiman and A. M. S. Shiyab, "Prediction of pavement remaining service life using roughness data—Case study in Dubai," *Int. J. Pavement Eng.*, vol. 4, no. 2, pp. 121–129, Jun. 2003.
[13] L. Hoel, *Transportation Infrastructure Engineering: A Multimodal Integration SI Edition*. Boston, MA, USA: Cengage Learning, 2011, pp. 112–115.
[14] N. G. Ahmed, "Development of pavement condition index model for flexible pavement in Baghdad City," *J. Eng.*, vol. 14, no. 1, pp. 2120–2135, 2008.
[15] J.-D. Lin, J.-T. Yau, and L.-H. Hsiao, "Correlation analysis between international roughness index (IRI) and pavement distress by neural network," in *Proc. 82nd Annu. Meeting Transp. Res. Board*, 2003, pp. 1–21.
[16] O. Kaya, 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," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2674, no. 10, pp. 448–460, Oct. 2020.
[17] A. Gupta, "Pavement deterioration and maintenance model for low volume roads," *Int. J. Pavement Res. Technol.*, vol. 4, no. 4, p. 195, 2011.
[18] D. T. Thube, "Artificial neural network (ANN) based pavement deterioration models for low volume roads in India," *Int. J. Pavement Res. Technol.*, vol. 5, no. 2, p. 115, 2012.
[19] S. Chandra, C. R. Sekhar, A. K. Bharti, and B. Kangadurai, "Relationship between pavement roughness and distress parameters for Indian highways," *J. Transp. Eng.*, vol. 139, no. 5, pp. 467–475, May 2013.
[20] N. Kargah-Ostadi and S. M. Stoffels, "Framework for development and comprehensive comparison of empirical pavement performance models," *J. Transp. Eng.*, vol. 141, no. 8, Aug. 2015, Art. no. 04015012.
[21] H. Ziari, J. Sobhani, J. Ayoubinejad, and T. Hartmann, "Prediction of IRI in short and long terms for flexible pavements: ANN and GMDH methods," *Int. J. Pavement Eng.*, vol. 17, no. 9, pp. 776–788, Oct. 2016.
[22] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 8, no. 4, Jul. 2018, Art. no. e1249.
[23] P. Marcelino, "Machine learning approach for pavement performance prediction," *Int. J. Pavement Eng.*, vol. 22, no. 3, pp. 341–354, 2021.
[24] H. Gong, Y. Sun, X. Shu, and B. Huang, "Use of random forests regression for predicting IRI of asphalt pavements," *Construct. Building Mater.*, vol. 189, pp. 890–897, Nov. 2018.
[25] M. Zhang *et al.*, "Analysis of critical factors to asphalt overlay performance using gradient boosted models," *Construct. Building Mater.*, vol. 262, Nov. 2020, Art. no. 120083.
[26] H. Gong, Y. Sun, W. Hu, and B. Huang, "Neural networks for fatigue cracking prediction using outputs from pavement mechanistic-empirical design," *Int. J. Pavement Eng.*, vol. 22, no. 2, pp. 162–172, Jan. 2021.
[27] Z. Wen, "ThunderGBM: Fast GBDTs and random forests on GPUs," *J. Mach. Learn. Res.*, vol. 21, no. 108, pp. 1–5, 2020.
[28] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
[29] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.
[30] G. Ke, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 3146–3154.
[31] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," 2017, *arXiv:1706.09516*.
[32] K. Fujimoto, I. Kojadinovic, and J.-L. Marichal, "Axiomatic characterizations of probabilistic and cardinal-probabilistic interaction indices," *Games Econ. Behav.*, vol. 55, no. 1, pp. 72–99, Apr. 2006.
[33] Y. Jia, 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," *Construct. Building Mater.*, vol. 247, Jun. 2020, Art. no. 118585.
[34] X. Chen, H. Zhu, Q. Dong, and B. Huang, "Optimal thresholds for pavement preventive maintenance treatments using LTPP data," *J. Transp. Eng. A, Syst.*, vol. 143, no. 6, Jun. 2017, Art. no. 04017018.
[35] H. Von Quintus, A. Yau, M. W. Witczak, D. Andrei, and W. N. Houston, "Appendix OO-1: Background and preliminary smoothness prediction models for flexible pavements. Guide for mechanistic empirical design of new and rehabilitated pavement structures," Transp. Res. Board, Washington, DC, USA, Final Rep. NCHRP 1-37A Project, 2001.
[36] M. Vaillancourt, L. Houy, D. Perraton, and D. Breysse, "Variability of subgrade soil rigidity and its effects on the roughness of flexible pavements: A probabilistic approach," *Mater. Struct.*, vol. 48, no. 11, pp. 3527–3536, Nov. 2015.
[37] *Mechanistic-Empirical Pavement Design Guide: A Manual of Practice*. Washington, DC, USA: American Association of State Highway and Transportation Officials, 2008.
[38] Y. D. Wang, J. Jeong, and Y. R. Kim, "Comparison of treatment timing between aggregate base and full-depth asphalt roads," *J. Transp. Eng. B, Pavements*, vol. 146, no. 4, Dec. 2020, Art. no. 04020057.
[39] S. Labi, G. Lamptey, and S.-H. Kong, "Effectiveness of microsurfacing treatments," *J. Transp. Eng.*, vol. 133, no. 5, pp. 298–307, May 2007.



Yang Song received the B.S. degree in civil engineering from Chang'an University, Xi'an, China, in 2015, and the M.S. degree in transportation engineering from Tongji University, Shanghai, China, in 2018. He is currently pursuing the Ph.D. degree in civil engineering with The Pennsylvania State University, State College, USA. His research interests include transportation infrastructure systems and electric vehicle.

Yizhuang David Wang received the Ph.D. degree from North Carolina State University in 2019. He is currently a Research Associate with the Missouri University of Science and Technology. His current research is focused on paving materials characterization; multiscale modelling of bituminous materials; pavement performance evaluation and testing; and pavement preservation, repair, and rehabilitation.

Xianbiao Hu received the Ph.D. degree from The University of Arizona, Tucson, AZ, USA, in 2013. He is currently an Assistant Professor with the Civil and Environmental Engineering Department, The Pennsylvania State University. His current research is focused on smart mobility systems, connected and automated vehicles, electric vehicles, mobility behavior management, transportation big data analytics, and traffic flow and systems modeling.



Jenny Liu received the Ph.D. degree from Texas A&M University, College Station, TX, USA, in 2006. She is currently a Professor with the Department of Civil, Architecture, and Environmental Engineering, Missouri University of Science and Technology. Her current research interests include engineering characterization and modeling of infrastructural materials, pavement design, pavement performance evaluation and testing, pavement preservation, non-destructive testing and condition assessment, and pavement management systems.