Titre: Title:	Enhanced prediction and uncertainty modeling of pavement roughness using machine learning and conformal prediction
Auteurs: Authors:	Sadegh Ghavami, Hamed Naseri, & Farzad Safi Jahanshahi
Date:	2025
Type:	Article de revue / Article
Référence: Citation:	Ghavami, S., Naseri, H., & Safi Jahanshahi, F. (2025). Enhanced prediction and uncertainty modeling of pavement roughness using machine learning and conformal prediction. Infrastructures, 10(7), 166 (24 pages). https://doi.org/10.3390/infrastructures10070166

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/66424/
Version:	Version officielle de l'éditeur / Published version Révisé par les pairs / Refereed
Conditions d'utilisation: Terms of Use:	Creative Commons Attribution 4.0 International (CC BY)

Document publié chez l'éditeur officiel
Document issued by the official publisher

Titre de la revue: Journal Title:	Infrastructures (vol. 10, no. 7)
Maison d'édition: Publisher:	Multidisciplinary Digital Publishing Institute
URL officiel: Official URL:	https://doi.org/10.3390/infrastructures10070166
Mention légale: Legal notice:	©2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Article

Enhanced Prediction and Uncertainty Modeling of Pavement Roughness Using Machine Learning and Conformal Prediction

Sadegh Ghavami [1,*] , Hamed Naseri [2] and Farzad Safi Jahanshahi [3,*]

1 Faculty of Civil Engineering, Sahand University of Technology, Tabriz 51335/1996, Iran
2 Department of Civil, Geological, and Mining Engineering, Polytechnique Montréal, Montréal, QC H3T 1J4, Canada; hamed.naseri@polymtl.ca
3 Department of Civil Engineering, Sirjan University of Technology, Sirjan 7813733385, Iran
* Correspondence: ghavamijamal@sut.ac.ir (S.G.); farzad_jahanshahi@yahoo.com (F.S.J.)

Abstract

Pavement performance models are considered a key element in pavement management systems since they can predict the future condition of pavements using historical data. Several indicators are used to evaluate the condition of pavements (such as the pavement condition index, rutting depth, and cracking severity), and the international roughness index (IRI), which is the most widely employed worldwide. This study aimed to develop an accurate IRI prediction model. Ten prediction methods were trained on a dataset of 35 independent variables. The performance of the methods was compared, and the light gradient boosting machine was identified as the best-performing method for IRI prediction. Then, the SHAP was synchronized with the best-performing method to prioritize variables based on their relative influence on IRI. The results suggested that initial IRI, mean annual temperature, and the duration between data collections had the strongest relative influence on IRI prediction. Another objective of this study was to determine the optimal uncertainty model for IRI prediction. In this regard, 12 uncertainty models were developed based on different conformal prediction methods. Gray relational analysis was performed to identify the optimal uncertainty model. The results showed that Minmax/80 was the optimal uncertainty model for IRI prediction, with an effective coverage of 93.4% and an average interval width of 0.256 m/km. Finally, a further analysis was performed on the outcomes of the optimal uncertainty model, and initial IRI, duration, annual precipitation, and a few distress parameters were identified as uncertain. The results of the framework indicate in which situations the predicted IRI may be unreliable.

Keywords: international roughness index; ensemble learning; conformal prediction interval; uncertainty; gray relational analysis

1. Introduction

Transportation infrastructure is an essential part of mobility and economic development in modern societies, being a pillar of the economy [1]. Pavement, as a component of transportation infrastructure, plays a significant role in roadway safety and performance. In the United States alone, the paved road infrastructure covers a distance of approximately 5 million kilometers, enabling the transport of people and large volumes of freight [2]. While pavements are an unavoidable aspect of infrastructure, they deteriorate over time due to various reasons, such as aging, increasing traffic volumes, and severe weather events. Therefore, it is essential to keep pavements in operational condition. Still, in 2023, around

39% of roads in the United States were in poor or mediocre condition, and only 45% were classified as good condition. Such inappropriate road conditions imposed drivers with additional operational costs and driving on deteriorated pavements costs around $1400 for each driver annually due to vehicle operating costs and lost time [3]. In 2023, the number of fatalities due to vehicle traffic crashes was around 41 thousand people, and one of the reasons could be the poor condition of pavements [4]. Improving pavement conditions can address these issues.

Maintenance and rehabilitation (M&R) practices are required to maintain pavements in acceptable performance conditions. Effective M&R planning is based on precise forecasting of future pavement conditions, achievable through the use of pavement condition indicators [5]. Global pavement surface roughness is one of the most widely used indicators, including the international roughness index (IRI), and it has been extensively utilized in previous research to predict the IRI in the planning horizon of M&R scheduling. That is, the M&R scheduling models minimize the IRI of pavement sections in a planning horizon [6]. Due to the importance of the IRI, many research papers have employed various models to predict its trend over time [7]. Although various IRI prediction models have been developed in previous studies, IRI prediction uncertainty modeling has not received enough attention. This study applies conformal prediction intervals to cover this research gap.

The primary objectives of this study are as follows:

- Accurate IRI prediction: all the vital variables are applied in the modeling; ten prediction methods (i.e., ensemble learning, machine learning, and deep learning methods) are used to identify the best-performing prediction method for IRI prediction.
- Identify the top determinant of IRI prediction: After detecting the best-performing prediction method, it is synchronized with an interpretation technique to capture the relative influence of variables on IRI.
- Optimal uncertainty model: Various conformal prediction methods at different target coverage levels are compared by a multi-criteria decision-making method to detect the optimal uncertainty model.
- Uncertainty parameters in IRI prediction: A further analysis is performed on the results of the optimal uncertainty model to determine the uncertain parameters in IRI prediction.

2. Literature Review

Many prediction models have been applied to predict IRI. Most of these models involve several factors to investigate their influence on pavement roughness. Age is one of the primary factors widely used in previous studies. Al-Suleiman and Shiyab [8] presented two IRI prediction models based on regression analysis for slow and fast traffic lanes separately, with pavement age as the primary independent variable. The coefficient of determination (R^2) was calculated at 0.80 for the slow lane model and 0.61 for the fast lane model. The results suggested that there is a strong correlation between age and IRI of pavement sections. Similarly, Tsunokawa and Schofer [9] proposed an IRI model with the initial IRI and pavement age as independent variables. IRI modeling that only uses initial conditions and pavement age may not yield the most accurate predictions. Many studies have incorporated traffic loads and structural features as additional parameters to improve IRI prediction models.

Traffic-associated variables like equivalent single axle load (ESAL) and cumulative ESAL (CESAL) have been commonly used in past investigations of IRI prediction models [10–16]. Along with the traffic-associated variables, structural characteristics have also been considered as an independent variable using structural number (SN), which has emerged as a major factor in many IRI prediction models [10–12,14,15]. For example, Choi et al. [11] proposed IRI prediction models based on CESAL, asphalt content, and SN using

multiple linear regression (MLR) and a backpropagation neural network. The MLR model had a correlation coefficient (R) of 0.46, and the backpropagation neural network R was 0.87, indicating a higher level of predictive performance.

In another study, Jaafar and Fahmi [14] collected data from 34 long-term pavement performance (LTPP) observations and applied initial IRI, pavement age, ESAL, construction number, and SN as independent variables to develop two IRI prediction models. They applied MLR and artificial neural network (ANN) methods to develop the models and achieved R^2 values of 0.25 and 0.80, respectively.

Furthermore, in addition to structural and traffic parameters, climatic characteristics have also been included in IRI modeling. Common climatic parameters used in IRI modeling are average annual temperature, the number of freeze–thaw cycles per year, freezing index, total annual snowfall, and annual precipitation [12,13,16]. For instance, using average annual precipitation, as well as ESAL and SN, Albuquerque and Nuñez [12] developed two IRI models using MLR and achieved an R^2 of 0.87. A regression-based IRI model for flexible pavements was developed by Khattak et al. [13] using climatic parameters (i.e., precipitation and temperature), as well as initial IRI, CESAL, and pavement thickness, on a dataset of 623 observations. The developed model reached an R^2 of 0.47. Dalla Rosa et al. [17] developed a regression-based IRI prediction model that included climatic conditions, treatment type, subgrade properties, traffic loadings, and pavement type as independent variables. The root mean square error (RMSE) was 0.21 m/km.

Multiple studies have highlighted that pavement distress ought to be included as an independent variable in IRI modeling [18–21]. Hence, some studies considered pavement distress variables in the IRI prediction models. For example, Owolabi et al. [19] developed an IRI prediction function with multiple linear regression (MLR) solely with three distress variables, including rut depth, patches, and longitudinal cracks length.

Although some research has focused strictly on pavement distress in their studies, others have included other attributes in addition to the distress indicators for the IRI model. For example, Abdelaziz et al. [21] combined initial IRI, pavement age, along with three pavement distress indicators—fatigue cracking, transverse cracking, and the standard deviation of rut depth—to create two IRI models, one using MLR and the other using ANN. The R^2 values for the two models were 0.57 for the MLR-based model and 0.75 for the ANN model.

As can be seen from the above-mentioned studies, MLR and ANN have been widely used algorithms for IRI prediction. However, there are other machine learning algorithms that can outperform MLR and ANN when comparing prediction accuracy. To this end, researchers have begun to apply other prediction techniques to improve prediction performance. For example, Marcelino et al. [22] used random forest regression to develop a long-term IRI prediction framework to predict IRI for up to 10 years. The model included climatic, structural, and traffic data, resulting in prediction errors of 6.95%. In another publication, Marcelino et al. [23] used the AdaBoost algorithm to develop yet another IRI model for flexible pavements, considering the variables pavement thickness, average annual daily traffic, SN, and climate, including average annual temperature and precipitation. The AdaBoost-based model resulted in the highest predictive accuracy with an R^2 value of 0.986. Damirchilo et al. [24] applied extreme gradient boosting, random forest regression (RFR), and support vector machine (SVM) algorithms, which produced R^2 of 0.70, 0.66, and 0.44, respectively. The independent variables of their models included ESAL, pavement age, precipitation, level of freeze–thaw days, freeze index, and the number of hot days for IRI prediction.

More in-depth literature review shows that only a select number of input features were used in earlier research studies to avoid increasing model complexity. Although the integration of all significant categorized variables, such as stripping of initial IRI, lane

width, traffic loading, structure properties of links in the models, climate conditions, and pavement distresses, is needed, the previous models reported in the literature have rarely employed all of these variables simultaneously. To address this, this study aims to collect a comprehensive dataset including all the vital parameters to maximize the performance of IRI prediction models. Also, there are powerful machine learning models whose applications to predict IRI have been limited, such as categorical boosting and light gradient boosting machine. Accordingly, this study employs many machine learning algorithms and compares their prediction performance using various performance indicators. Moreover, to the best of the authors' knowledge, all the previous studies applied deterministic prediction models to predict IRI, and uncertainty has not been modeled in this research domain. To address this, this study applied a conformal prediction interval framework to investigate the uncertainty in IRI prediction and identify which factors can increase such uncertainty. Unlike previous IRI prediction models, the proposed uncertainty framework can indicate which data observations and in which situations the predicted IRI may be unreliable. This feature improves the interpretability and practical usefulness of the model by allowing decision makers to assess the trustworthiness of individual predictions.

3. Methodology

In this section, the applied dataset for the modeling is first described. Then, the prediction methods are briefly introduced. Subsequently, the performance metrics used to identify the best-performing prediction technique are presented. Afterward, the uncertainty methods are described. Finally, the further analysis used to detect uncertain parameters is described. The methodology flowchart is shown in Figure 1.
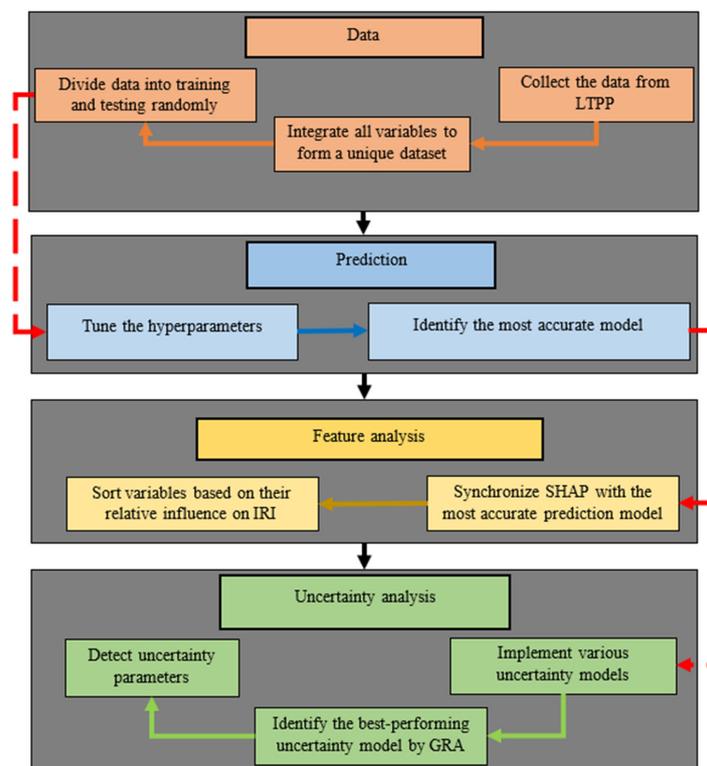

Figure 1. The methodology flowchart.

3.1. Datasets

The long-term pavement performance (LTPP) database, containing condition data for 526 flexible pavement sections in the 50 states of the United States, supplied the data required to build the IRI-prediction model. The flexible pavement sections are from intercity

highways, which are roads that physically connect cities, but are outside the urban centers of those cities.

One of the goals of this research is to identify which features are most effective in predicting IRI over time. The input variables selected spanned multiple variable categories, including initial pavement condition, pavement age, lane width, traffic data, structural data, climate characteristics, and pavement surface distress. Traffic loading was represented by cumulative equivalent single axle loads (CESAL), and the structural data was represented by structural number (SN). We selected these variables for two reasons: first, they have been used in previous studies (please see the literature review), though not all together; second, they represent all the available variables in the LTPP dataset.

Climatic features included in the model were average annual temperature, annual freezing index, freeze–thaw cycles in each year, and annual amounts of precipitation. Furthermore, the statistical value of parameters is presented in Table 1. In this table, the final row represents the statistical summary of the dependent variable (IRI) used for developing and evaluating prediction models. Although IRI is the predicted outcome, including its descriptive statistics in the methodology section is essential for understanding the characteristics of the dataset and the modeling context.

Table 1. Statistical value of parameters.

Parameters	Statistical Value			
	Min	**Max**	**Mean**	**StD**
Initial IRI (m/km)	0.54	3.57	1.13	0.47
Structural number	0	10	4.63	1.6
Equivalent single axle loads (ESALs) (number of loads)	0	4,103,546.08	391,899.25	513,441.69
Surface area of low severity alligator cracking (m^2)	0	488	16.16	52.14
Surface area of medium severity alligator cracking (m^2)	0	218.8	4.56	22.1
Surface area of high severity alligator cracking (m^2)	0	411.5	2.91	28.02
Length of low severity longitudinal cracking in the wheel path (m)	0	161.7	4.12	16.42
Length of medium severity longitudinal cracking in the wheel path (m)	0	61.7	0.99	6.18
Length of high severity longitudinal cracking in the wheel path (m)	0	4.4	0.01	0.2
Length of low severity longitudinal cracking in the non-wheel path (m)	0	329.1	37.62	72.47
Length of medium severity longitudinal cracking in the non-wheel path (m)	0	305	11.56	38.01
Length of high severity longitudinal cracking in the non-wheel path (m)	0	305	6.03	31.82
Length of sealed low severity longitudinal cracks in the non-wheel path (m)	0	329.1	1.44	17.67
Length of sealed medium severity longitudinal cracks in the non-wheel path (m)	0	7.2	0.04	0.48
Number of low severity transverse cracks	0	152	5.43	17.28
Number of medium severity transverse cracks	0	62	1.36	6.23
Number of high severity transverse cracks	0	12	0.2	1
Length of low severity transverse cracks (m)	0	134	4.44	14.68
Length of medium severity transverse cracks (m)	0	140.5	2.32	10.21
Length of high severity transverse cracks (m)	0	27.7	0.55	2.68
Length of low severity sealed transverse cracks (m)	0	106	0.33	4.79
Length of medium severity sealed transverse cracks (m)	0	6.5	0.03	0.42
Length of high severity sealed transverse cracks (m)	0	7	0.02	0.33
Area of binder bleeding (bleeding) (m^2)	0	320.3	20.9	63.73
Surface degradation due to aggregate loss (raveling) (m^2)	0	564.3	24.4	93.66
Number of occurrences where water/fines are ejected under traffic loads (pumping)	0	10	0.13	0.78
Total length of pavement affected by pumping (m)	0	186.80	1.78	14.91
Width of the surveyed pavement section (m)	3.00	4.3	3.63	0.17
Total cracked length in the wheel path (m)	0	305	45.04	83.92
Length of transverse cracks longer than 183 cm (often used as a threshold for major cracks) (m)	0	114.6	3.23	12.67
Total annual precipitation (mm/year)	122.3	1923.4	902.86	447.67
Mean annual temperature (°C)	1.3	24.4	14.20	6.12
Annual freezing index (cumulative degree-days below freezing) (°C.days)	0	1744	275.43	422.50
Number of freeze–thaw cycles per year	0	179	67.95	41.17
The duration between data collections (days)	106	2651	728.27	476.94
IRI (m/km)	0.59	4.37	1.26	0.56

In total, 35 independent variables were used to build the IRI prediction model. We extracted data from the LTPP database of pavement sections with complete lists of the

36 variables for which the sections received no maintenance and rehabilitation (M&R) treatments. The data reviewed represents a wide range of observation periods on pavement sections from 106 days to 7 years of IRI measurements [25]. Table 1 summarizes descriptive statistics including the maximum, minimum, average, and standard deviation for selected variables.

From this dataset, we developed two sample sets to construct the models: the training set and the testing set. The training set was applied to train and tune the machine learning algorithms. The testing set used pavement section data that had not been seen in the model in the training process, and they were applied to evaluate the predictive performance of the developed models [26]. In this study, we randomly split the data sets into 80% training and 20% testing data. Further, we applied Optuna and k-fold cross-validation (considering k = 5) to tune the hyperparameters of machine learning techniques.

3.2. Prediction Techniques

Ten prediction techniques are applied to predict IRI and identify the best-performing technique on IRI prediction. These methods are briefly introduced in this section.

3.2.1. K-Nearest Neighbor

K-nearest neighbor (KNN) is a black-box machine learning method that has been widely employed for modeling and prediction since the 1970s. This method estimates the dependent variable of a given observation in testing data based on the dependent variable of similar observations in the training set. This method identifies K similar observations based on their independent variables. That is, the nearest observations are determined by a distance function and the values of independent variables. Finally, the K nearest observations are selected, and their dependent variables are used to predict the dependent variable of the testing data [27].

3.2.2. Linear Regression

Linear regression (LR) is a common and well-known statistical method typically used to measure relationships between independent (input) variables and a dependent (output) variable. In the last several years, LR has gained much favor as a machine-learning algorithm [28]. LR aims to generate linear models whereby the dependent variable (y) is represented as a linear combination of any number of independent variables (x). On the one hand, when the model has one independent variable, we label it simple linear regression, and if we have two or more independent variables to consider, we have multiple linear regression. The simple linear regression model can be generally represented as:

$$y = b_0 + b_1 x_1. \tag{1}$$

Equation (1) is considered the mathematical description of a straight line; b_0 is the y-intercept and b_1 designates the slope of the line and b_1 represents the slope of the line.

If we think about multiple linear regression, we can see that the model not just extends from a line to a plane, but also to a hyperplane in more than two dimensions. The general representation of the multiple linear regression model is:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n \tag{2}$$

In Equation (2), the model consists of n predictor variables, represented as $x_1, x_2, \ldots, x_n$.

3.2.3. Multi-Layer Perceptron

The multi-layer perceptron (MLP) can be defined as a deep learning method that introduces a complex architecture in artificial neural networks [29]. MLPs are mainly used

for supervised learning and can be thought of as a feedforward neural network, which maps a single or a set of inputs to an output or a set of outputs. The MLP consists of multiple layers of perceptrons across its architecture; the input nodes of each layer are connected, like a directed graph.

In the MLP, both the adaline (adaptive linear neuron) rule and the perceptron rule have been used to train single feedforward neural networks, where the connections (weights) to the units are adjusted according to a unit function applied to a linearization of the inputs. An MLP is a fully connected neural network, typically including three types of layers: input, hidden, and output. The input layer receives the information and transfers it to the hidden layer. The hidden layer(s) apply non-linear transformations to the inputs using activation functions. The output layer produces the final prediction [30].

3.2.4. Decision Tree

Decision tree regression is a type of supervised learning that predicts continuous numerical outcomes based on learned feature representations in a tree structure. It works by following a series of recursive binary judgements about the feature values, whereby each judgement continues to refine subsets of the dataset, going from larger, increasingly smaller, and homogenous subsets until the final predictive output is a continuous variable. This process allows the model to predict a meaningful output for the numeric continuous variable according to the learning experience of the model [31].

The tree is made up of internal decision nodes and terminal leaves. With an internal node representing a split based on either a single feature or the two levels of the feature, while the topmost internal node, called a root node, is the node based on the feature that best predicts the dependent variable. The search is carried out by top-down recursive partitioning, where the splits try to minimize the variance of the whole dataset into subsets. Specifically, the metric of measurement that is often used in terms of variance is the number of values (standard deviation) represented through partitions: a zero standard deviation indicates the subset is homogeneous, while larger groups are more heterogeneous [32].

3.2.5. Random Forest

The random forest (RF) algorithm is an ensemble method that can be utilized for classification and regression problems. It generates a predictive model by combining multiple decision trees as its base learners [33]. The method uses a technique called bootstrap aggregating, or bagging, in order to produce the model.

With bagging, for the original training dataset X, the algorithm samples B random samples from X, with replacement, to form *B* training datasets. Each of those *B* training datasets is then used to train a single decision tree. These decision trees are generated in parallel and do not share any information in the training process. After the training phase is complete, the ensemble of trees may be used to predict unseen instances of "x" created based on outputs from each of the B trees combined with the following equation:

$$\hat{f} = \frac{1}{B} \sum_{b=1}^{B} f_b\left(\acute{x}\right) \qquad (3)$$

Decision trees independently have high variance, but when combined in a random forest, the variance of the model is much lower. In regression problems, random forest uses the average value of all of the predictions from the trees to generate the final output [34].

3.2.6. Adaptive Boosting

Adaptive boosting (AdaB) is an ensemble learning method that improves the prediction performance by combining multiple weak learners into a single strong predictor.

Usually, AdaB only uses decision stumps—single-level decision trees—as weak learners [35]. The algorithm assigns weights to instances of the training samples according to how difficult they are to predict. That is, the harder to predict samples are assigned a higher weight, while correctly classified samples are assigned a lower weight. This provides the model with the ability to focus on the more difficult cases more than other samples in each iteration of training.

AdaB for regression is a meta-estimator. It starts by fitting a base regressor onto the original dataset, and then fits additional regressors iteratively to the data. Each time a new regressor is fitted, the model will update the weights of the training samples, depending on the prediction errors for the previous regressor. This means that the model focuses on the hardest examples with the fitted new regressor to improve the model's predictive performance.

3.2.7. Gradient Boosting

Gradient boosting (GB) is another ensemble learning method used for prediction purposes. Gradient boosting uses the boosting paradigm, where multiple weak learners (usually decision trees) are sequentially combined to produce a stronger predictive model. A simple first tree is built, often a one-node stump. The models of each additional tree are trained on the previous tree ensemble's errors (residuals).

A learning rate is used to regulate the impact of each tree on the model's final prediction, which lowers the impact of each individual tree on the correction. The approach is to add trees over time to tune the model. This process is continued until a predetermined number of iterations or an improvement in predictive accuracy is no longer observed [36].

3.2.8. Light Gradient Boosting

Light gradient boosting (LGB) is another ensemble learning method applied to predict IRI. Like other ensemble learning techniques, LGB produces a given number of weak learners and syncs them to develop a more accurate prediction model. This method is an updated version of GB, which was introduced by Microsoft. LGBM is an efficient method, and it can reduce memory usage by parallel learning. A leaf-wise leaf growth strategy is used by LGB, which controls the depth growth of decision trees. In this strategy, multiple leaves at the same depth can be split simultaneously, enabling multi-threaded optimization during training [37].

3.2.9. Extreme Gradient Boosting

Extreme gradient boosting (XGB) is a powerful supervised learning algorithm for regression problems in machine learning [38]. It is computationally efficient and scalable, and provides much higher prediction ability than normal decision trees, while it is less interpretable. It works by learning a series of base learners, normally decision trees, to build a strong ensemble.

The objective function of XGB has two pieces: a loss function and a regularization term. The loss function measures the difference between predicted target values and actual target values. The second piece, regularization, penalizes model complexity to minimize overfitting. XGBoost iteratively minimizes its objective function during training, allowing it to create successive base learners to strengthen its prediction ability.

3.2.10. Categorical Boosting

Categorical boosting (CatB) is another efficient and powerful ensemble learning method applied in this study to predict IRI. CatB uses Bayesian estimators to reduce the likelihood of overfitting, and it replaces categorical variables with binary variables in the training phase, which can reduce computational complexity. Like other ensemble learning methods, CatB generates various decision trees and combines their outcomes

to create a powerful ensemble model. However, this method has a symmetric structure and the depth of decision trees across all nodes is uniform, which is different from other ensemble learning methods, like XGB.

3.3. Performance Metrics

To compare the performance of prediction techniques and identify the most accurate one, a number of statistical metrics are used, including the coefficient of determination (R^2), mean absolute error (MAE), mean square error (MSE), mean absolute percentage error (MAPE), the Akaike information criterion (AIC), and the explained variance (EV). These metrics are calculated as follows:

$$R^2 = \frac{\left(n\sum EXP_i MDL_i - \sum EXP_i \sum MDL_i\right)^2}{\left(n\sum EXP_i^2 - (\sum EXP_i)^2\right)\left(n\sum MDL_i^2 - (\sum MDL_i)^2\right)} \tag{4}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|EXP_i - MDL_i| \tag{5}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(EXP_i - MDL_i) \tag{6}$$

$$MAPE = \frac{\sum_{i=1}^{n}|EXP_i - MDL_i|}{\sum_{i=1}^{n}EXP_i} \tag{7}$$

$$EV = 1 - \frac{Var(EXP_i - MDL_i)}{VarEXP_i} \tag{8}$$

Where EXP_i and MDL_i represent the actual and predicted values of data observation i, respectively. Also, var signifies that variance.

AIC will normally be used to compare the quality of the statistical models for a given dataset; that is, a lower value indicates better fitting (but a penalty is given for increasing the number of parameters).

$$AIC = 2k - 2ln(\hat{L}) \tag{9}$$

k shows the number of the model's parameters and $\hat{L}$ is the maximum likelihood of the model. For linear regression models (with normally distributed errors), the equivalent form is as follows:

$$AIC = n.ln\left(\frac{RSS}{n}\right) + 2k \tag{10}$$

Residual sum of squares is presented by RSS and n is the number of observations. In terms of evaluating model performance, a higher R^2 and EV signify that the model has a better fit to the observed data. In contrast, it is preferable that MAE, MSE, and MAPE yield a lower value. Moreover, if a model has the lowest negative AIC value, it has better performance for predicting.

3.4. Interpretation Method

Another objective of this study is to determine the relative influence of variables on IRI. In this regard, after detecting the best-performing machine learning method by performance metrics, it is synchronized with the SHapley Additive exPlanation (SHAP) to interpret its results and identify the top determinants of IRI prediction. SHAP is a game theory-based method that applies local explanations to prioritize independent variables based on their relative influence on the dependent variable. To this end, this method uses a unitless metric, called the SHAP value, to compare the relative influences. We used SHAP since it performs better than other interpretation techniques in terms of consistency and computational performance [39].

3.5. Uncertainty Modeling

As mentioned, one of the main objectives of this study is to model uncertainty in IRI prediction. To this end, the best-performing machine learning method (identified by performance metrics) is used in the uncertainty analyses. For modeling uncertainty, we applied the model agnostic prediction interval estimator (MAPIE). MAPIE implements conformal prediction techniques to estimate uncertainties, offering strong theoretical guarantees on marginal coverage under mild assumptions about the model and the underlying data distribution. For more details about MAPIE, refer to Taquet et al. [40].

MAPIE relies on three main elements: the base prediction model, the method for constructing prediction intervals (called conformal prediction methods), and the target coverage level. As mentioned, we used the best-performing machine learning technique (among the ten methods applied) as the base prediction model. We used four conformal prediction methods: Naïve, Jackknife, Jackknife+, and Minmax, since they are the conventional methods to create prediction intervals for regression prediction problems [40]. Rather than predicting a single value, these methods evaluate the lower bound and upper bound of the predicted value for each data observation and provide us with a predicted interval. These intervals are calculated based on the residuals. The target coverage level is the desired proportion of observations for which the true values fall within the predicted intervals [40]. In this study, three target coverage levels are considered: 80%, 90%, and 95%. The conventional values for the target coverage level are 80%, 90%, 95%, and 99% [41]. We excluded 99% since one of the methods could obtain an effective coverage of 100% with an effective target coverage level of 95%. That is, increasing the target coverage level after this threshold (95%) was not effective since the width of intervals was increased without any improvement in the effective coverage.

Therefore, 12 uncertainty models are developed by combinations of four conformal prediction methods and three target coverage levels. To identify the optimal combination (out of 12), a multi-criteria decision-making method, called grey relational analysis (GRA), is used. For more details about GRA, refer to Panda et al. [42]. This method compares the combinations of prediction interval methods and target coverage levels using two criteria: the average interval width and the actual coverage level of testing data (i.e., effective coverage). Each method predicts an interval (i.e., range) for every observation. The average interval width is calculated as the mean difference between the upper and lower bounds of these intervals. The actual coverage level is defined as the percentage of testing observations whose true values fall within the predicted intervals. An optimal uncertainty model refers to a model with a lower average interval width and a higher actual coverage level.

After detecting the optimal uncertainty model, it is applied to identify uncertainty parameters. In this regard, the independent variables of the IRI prediction models (shown in Table 1) were considered independent variables. The interval width of data observations calculated by the optimal uncertainty model is considered the dependent variable. Then, an MLR model is applied to determine the significant variables on the uncertainty of IRI prediction.

4. Results and Discussion

4.1. Performance of Prediction Methods

As mentioned, ten prediction methods were applied to predict IRI and identify the best-performing method based on various metrics. In this part, the results of the developed models used for predicting IRI are presented. The value of predicted IRI and measured IRI for testing data is presented in Figure 2. This figure contains ten developed models against the measured IRI. An equality line is also drawn. The proximity of the data to the equality line provides insight into the accuracy of the developed models. Looking at the

details, among various models, LGB produces the least variance in data, with predictions close to the equality line representing the power of LGB for predicting IRI. The statistical parameters for evaluating developed models on testing data are presented in Figures 3–8.
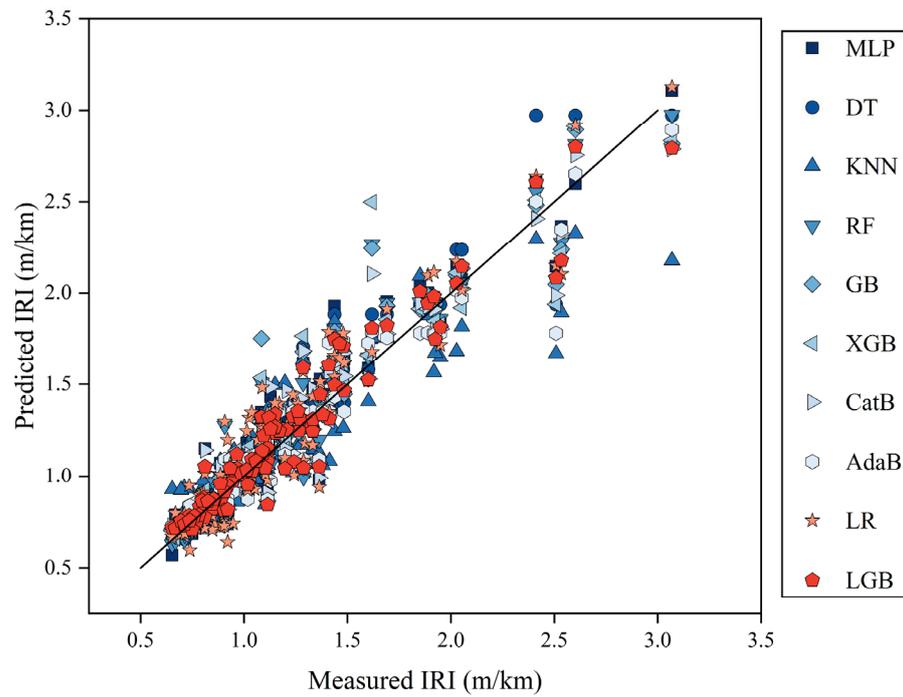

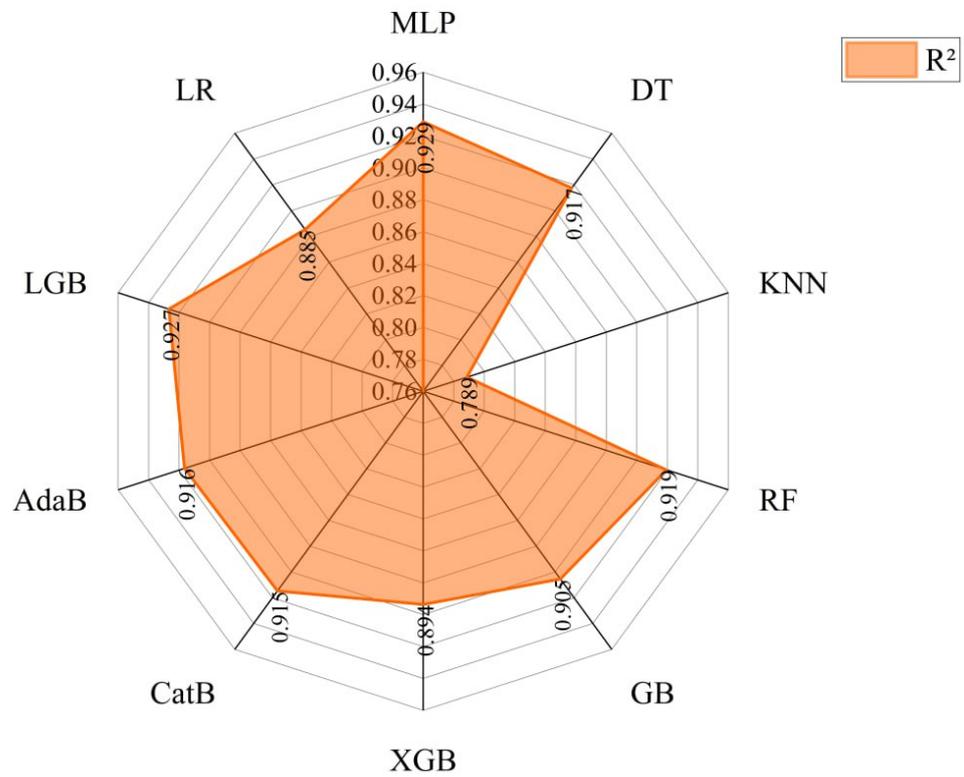**Figure 2.** Performance of the developed models for predicting IRI on testing data.

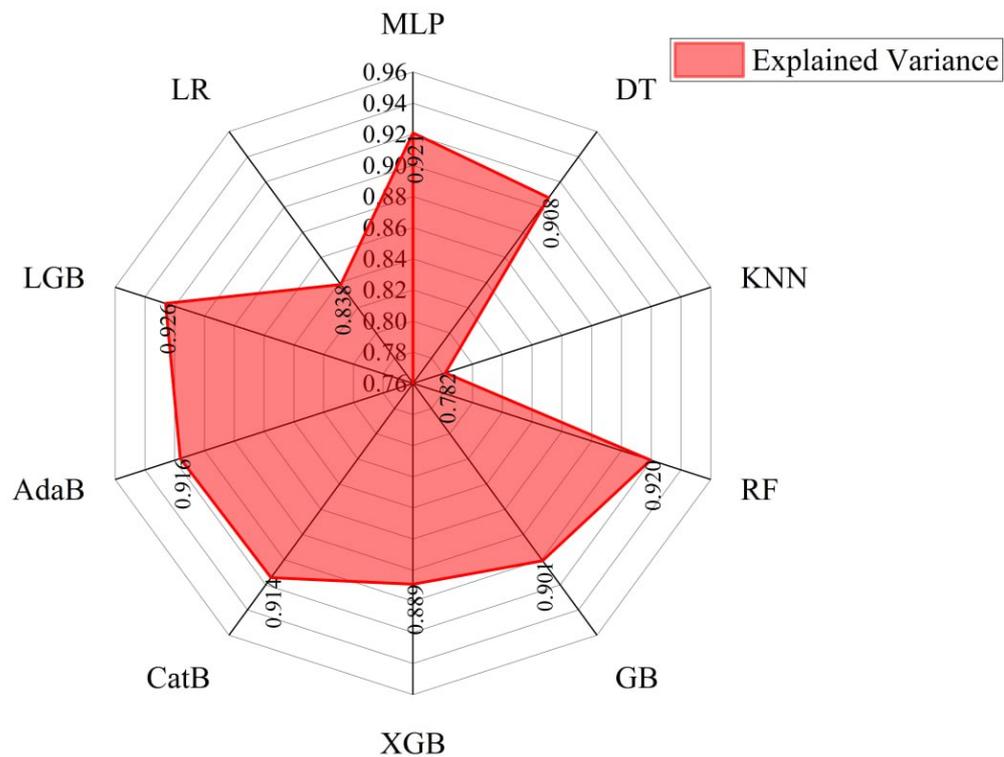**Figure 3.** The result of the R^2 value for various models.
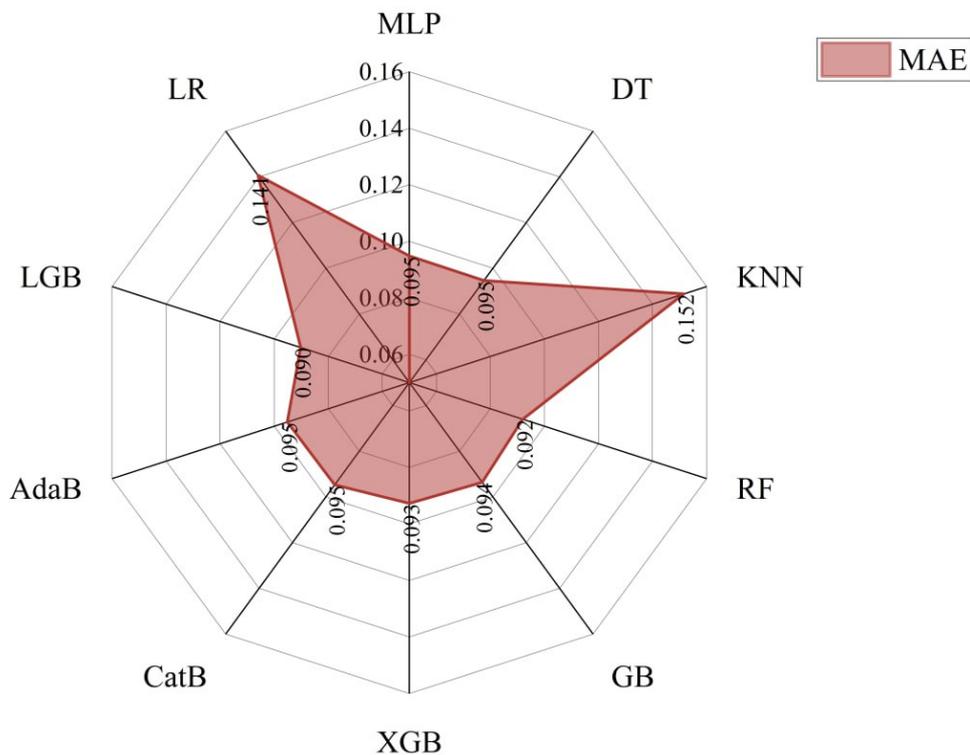
Figure 4. The result of the EV value for various models.

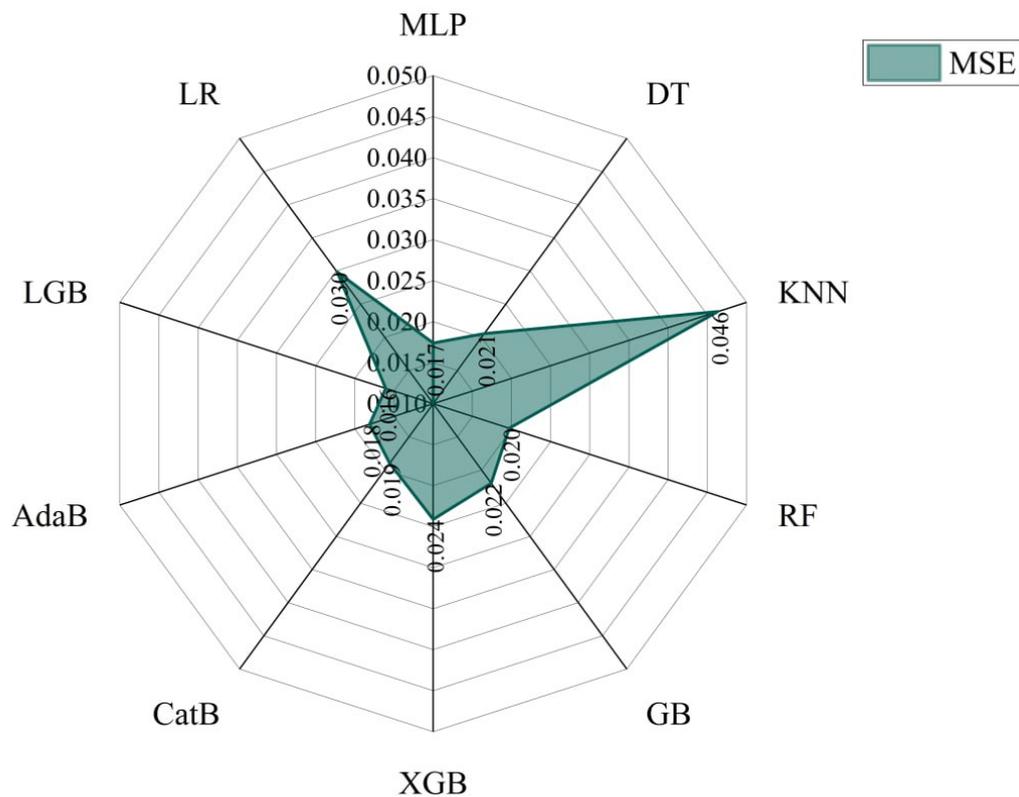**Figure 5.** The result of the MAE value for various models.

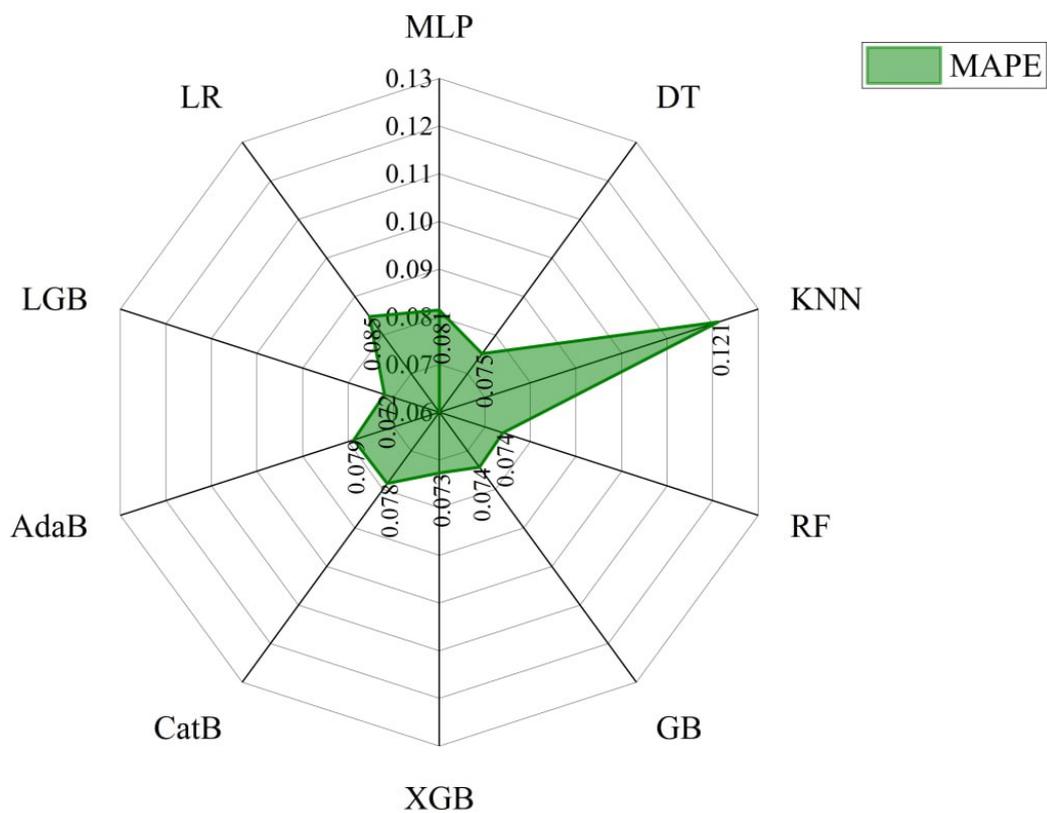**Figure 6.** The result of the MSE value for various models.

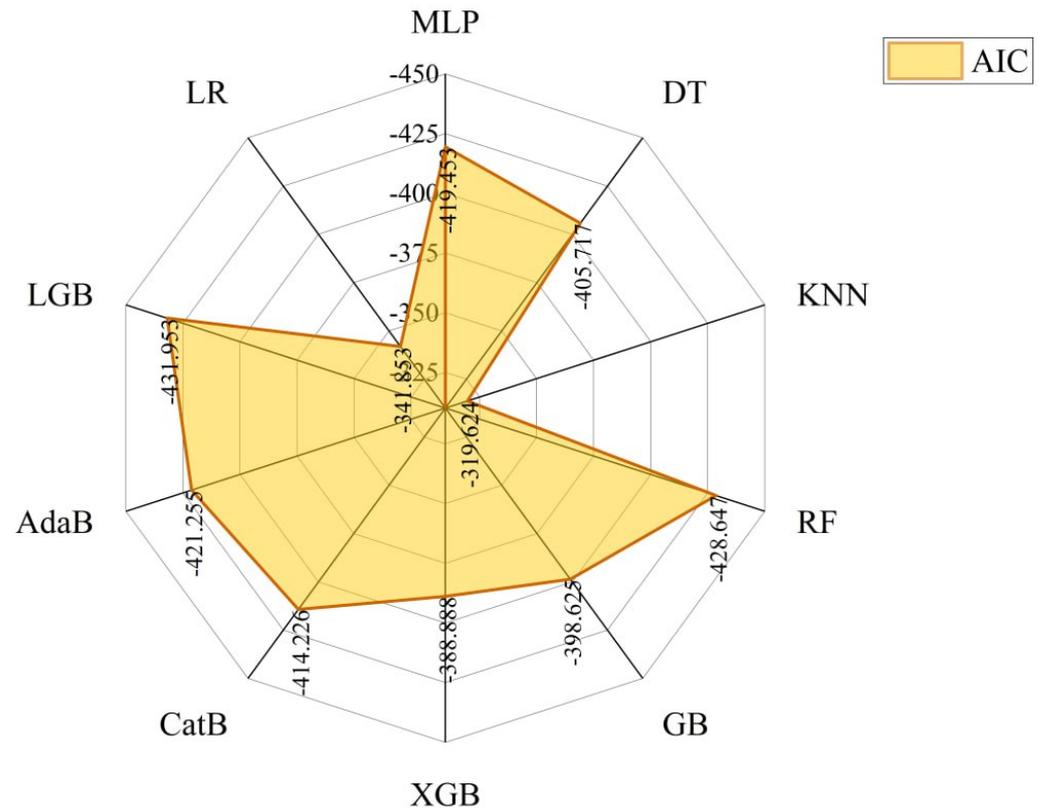**Figure 7.** The result of the MAPE value for various models.

Figure 8. The result of the AIC value for various models.

Figure 3 shows the R^2 value of various methods on testing data, with the highest value implying better performance. What stands out from the figure is that the R^2 of MLP, DT, RF, GB, CatB, AdaB, and LGB was over 0.9. Among these techniques, the highest testing data R^2 is associated with MLP, followed by LGB and RF. On the other hand, KNN was ranked as the last model. EV is illustrated in Figure 4 and has the same approach as R^2. LGB had superior precision in comparison with other methods. Regarding EV, only LGB and MLP obtain a value of over 0.92. Similarly, KNN was the worst-performing model based on EV.

In prediction methods, MAE, MSE, and MAPE parameters should have the lowest values (preferred 0). Figures 5–7 show that the LGB method has the lowest value of these three parameters on testing data, showing that this method outperforms other techniques. AIC had an inverse approach, and a model can predict a reliable and accurate IRI if it has the lowest negative value. The AIC of various methods on testing data is presented in Figure 8. As can be seen, like other metrics (except R^2), LGB provided a superior performance.

Figure 9 illustrates the error distribution of the developed models. If we look at the details, in terms of the LGB model, the overall distribution of values approaches zero, suggesting that values are not biased. In other words, the difference between the measured and predicted values of the model is similar to zero. In addition, despite the other methods having a wide range of errors, approximately from −65% to 40%, the error range of LGB is within the range from −20% to 30%, implying its superior performance on IRI prediction. To sum up, LGB outperforms other prediction methods when comparing EV, MAE, MSE, MAPE, AIC, and the error histogram. However, LGB was the second-best model when comparing R^2 after MLP. We consider LGB to be the best-performing method for IRI prediction for two reasons. First, its performance is better than other prediction methods in 6 metrics (out of 7). Second, we aim to minimize the prediction accuracy, while R^2 is not a well-suited metric for evaluating prediction accuracy. Also, the difference between the R^2 of LGM and MLP (maximum R^2) is not significant (i.e., 0.002). Therefore, LGB is employed

for further analysis, presented in the following parts (i.e., SHAP and uncertainty). The optimal hyperparameters of all methods are presented in Table 2. Table 3 summarizes the performance of prediction models.
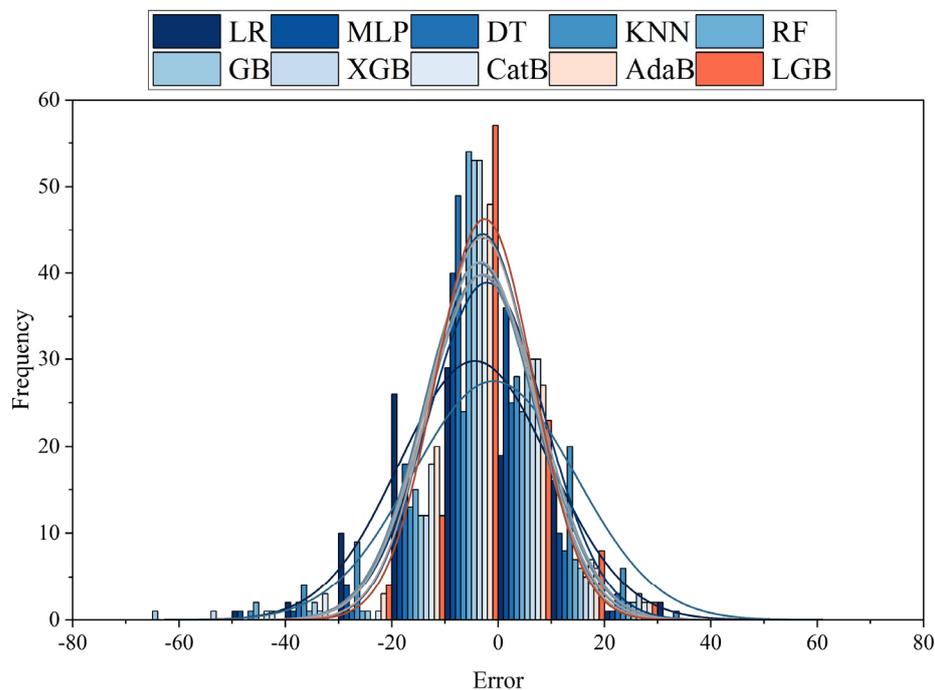

Figure 9. The frequency and normal distribution of models.

Table 2. The optimal hyperparameters.

Model	Hidden_Layer_Sizes	Activation	Solver	Alpha	Learning_Rate
MLP	365	tanh	sgd	0.001	adaptive
	learning_rate	max_depth	n_estimators		
GB	0.01739	3	470		
XGB	0.03579	3	436		
CatB	0.03977	3	261		
AdaB	0.0127	-	70		
LGB	0.07662	3	72		
	max_depth	min_samples_split	n_estimators		
RF	23	2	155		
	n_neighbors	weights	p		
KNN	4	distance	2		
	fit_intercept				
LR	TRUE				

Table 3. The performance of prediction models.

Model	R^2	MAE (m/km)	MSE (m/km)^2	AIC	EV	MAPE (%)
MLP	0.929	0.095	0.017	−419.453	0.921	0.081
DT	0.917	0.095	0.021	−405.717	0.908	0.075
KNN	0.789	0.152	0.046	−319.624	0.782	0.121
RF	0.919	0.092	0.020	−428.647	0.920	0.074
GB	0.905	0.094	0.022	−398.625	0.901	0.074
XGB	0.894	0.093	0.024	−388.888	0.889	0.073
CatB	0.915	0.095	0.019	−414.226	0.914	0.078
AdaB	0.916	0.095	0.018	−421.255	0.916	0.079
LGB	0.927	0.090	0.016	−431.953	0.926	0.072
LR	0.885	0.141	0.030	−341.853	0.838	0.085

4.2. Relative Influence of Variables

The relative influence of independent variables on IRI is calculated by SHAP, and the top 20 variables are presented in Figure 10. As shown, the initial IRI has the top-ranked determinant of IRI prediction, followed by mean annual temperature and the duration between data collections. Song et al. [43] applied a ThunderGBM-based ensemble learning model and SHAP to identify the relative influence of variables on IRI, and the results suggested that initial IRI was by far the top determinant of IRI, followed by the depth of rutting and the area of patches. Although their results were in line with the results of this study regarding the top-ranked determinant (initial IRI), the ranking of duration was significantly different (ninth in that study) from the current study (third). Erfani et al. [44] used XGBoost and SHAP to capture the top determinants of IRI. The outcomes suggested that initial IRI was first, and average annual temperature was the fourth top variable, which is in harmony with the results of the current study.
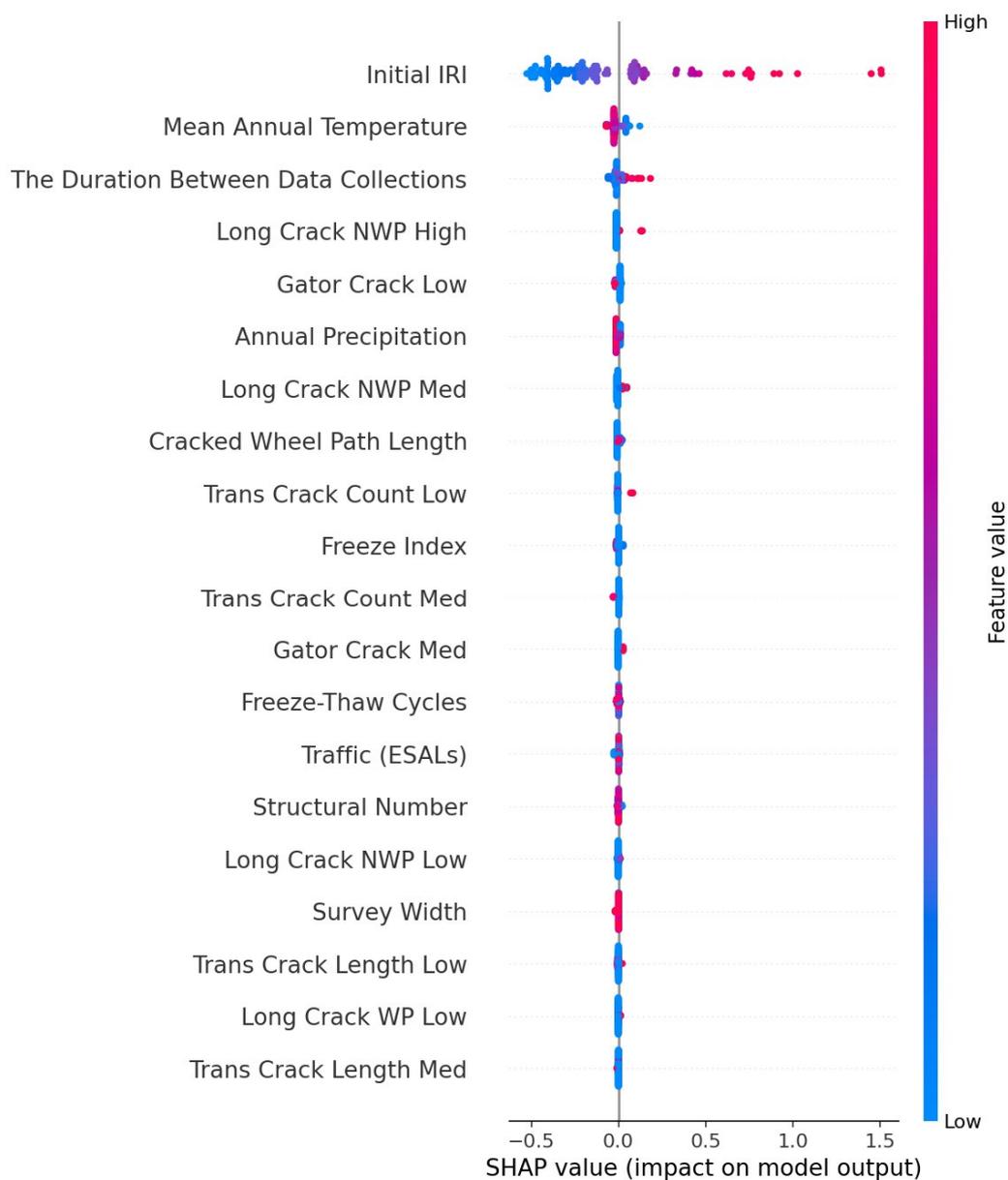

Figure 10. The SHAP value of variables on IRI.

Clearly, a higher initial IRI value leads to a higher IRI value in the next data collection. Further, the IRI values are higher in colder regions. In regions with frost-sensitive soil and availability of water, subfreezing temperatures can lead to frost heaving and lifting of stones in the pavement [45]. The freeze and thaw cycles in colder regions are expected to be higher, and freeze and thaw cycles deteriorate the mechanistic properties of asphalt mixes, and as a result, the durability of pavement is reduced, and IRI is increased [46]. In cold regions, it is more likely to snow. The process of snow plowing and anti-icing can damage the pavements, and as a result, IRI is likely to increase. These might be some of the reasons why IRI is higher in colder regions [47]. The third variable is the duration between data collections. IRI is a function of time, and as time passes, it increases exponentially [9]. Regarding distress, when the length of high severity longitudinal cracking in the non-wheel path was very high, IRI was expected to be higher. The presence of high-severity longitudinal cracking in the non-wheel path is often correlated with other forms of distress, such as transverse cracking, block cracking, edge cracking, and patching. Also, the NWP longitudinal cracking can represent the condition of pavement in terms of environmental conditions, age, layer thickness, and kinematic viscosity, and it might be one reason why this variable is a strong predictor of IRI.

4.3. Uncertainty Analysis

As mentioned, we applied a conformal prediction interval framework with the best-performing prediction technique (LGM) to investigate uncertainty in predicting IRI. Four conformal prediction methods were used for constructing prediction intervals: Naïve, Jackknife, Jackknife+, and Minmax. Three target coverage levels were considered: 80%, 90%, and 95%. The effective coverage (i.e., actual coverage level for testing data) of various conformal prediction methods is presented in Figure 11. As shown, the effective coverage of Minmax was higher than that of other methods at all target coverage levels, followed by Jackknife+, Jackknife, and Naïve. The effective coverage of Minmax, Jackknife, and Jackknife+ was higher than the target coverage level at all levels.
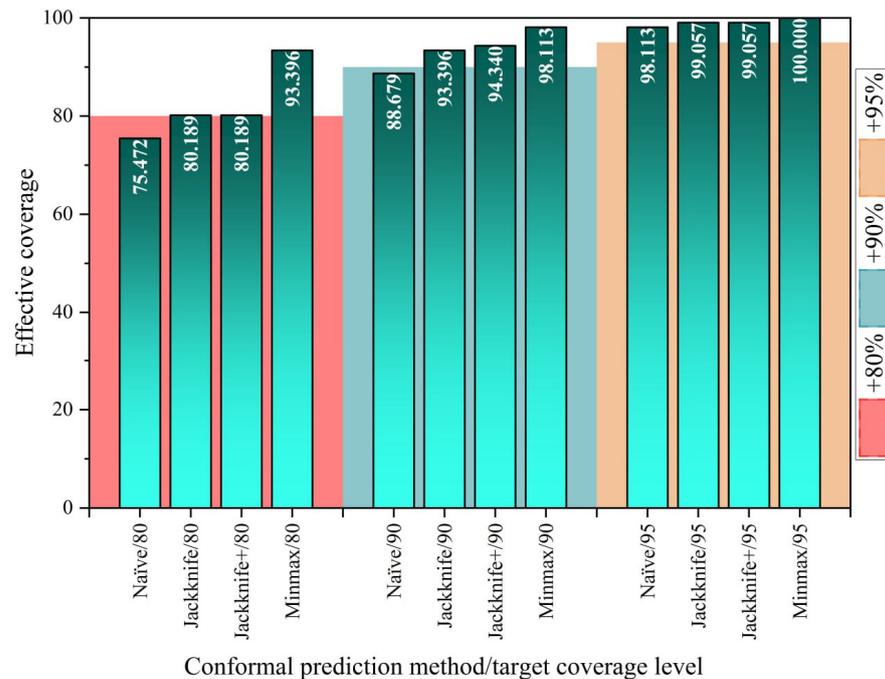

Figure 11. The effective coverage of conformal prediction methods.

The average interval width of these techniques is presented in Figure 12. As can be seen, the average interval width is minimum for Naïve at all levels, followed by Jackknife, Jackknife+, and Minmax. Hence, it can be theorized that Minmax ensures reliable coverage at the expense of wider intervals. Jackknife+ and Jackknife provide a favorable trade-off between precision and coverage. Naïve tries to minimize the average interval width by sacrificing the coverage.
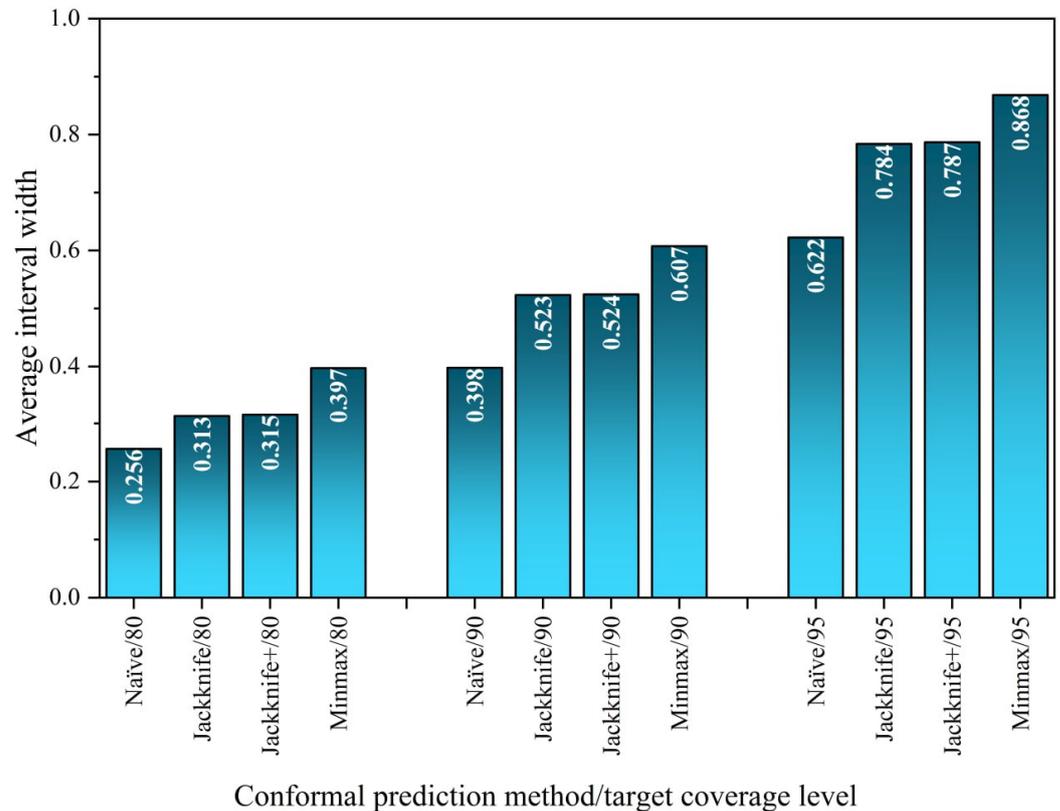

Figure 12. The average interval width of conformal prediction methods.

Figure 13 presents the prediction intervals of IRI along with the corresponding actual values for the testing data. In this figure, the gray bars and yellow dots imply that the actual IRI was within the prediction interval, and the predicted interval crossed the equity line (e.g., the actual IRI was 1.5 m/km, and the predicted interval was from 1.4 m/km to 1.55 m/km). On the other hand, the red error dots and bars denoted that the actual value was outside the prediction interval (e.g., the actual IRI was 1.5 m/km, and the predicted interval was from 1.25 m/km to 1.45 m/km).

As can be perceived, by increasing the target coverage level (going from top to bottom), the error was reduced, but the width of intervals was increased. Further, Minmax has the highest coverage, but the largest intervals. For example, Minmax/95 can correctly cover all the testing data, but the interval widths are extremely large. On the other hand, Naïve has the lowest coverage and smallest intervals. Jackknife and Jackknife+ provide a trade-off between width and coverage.
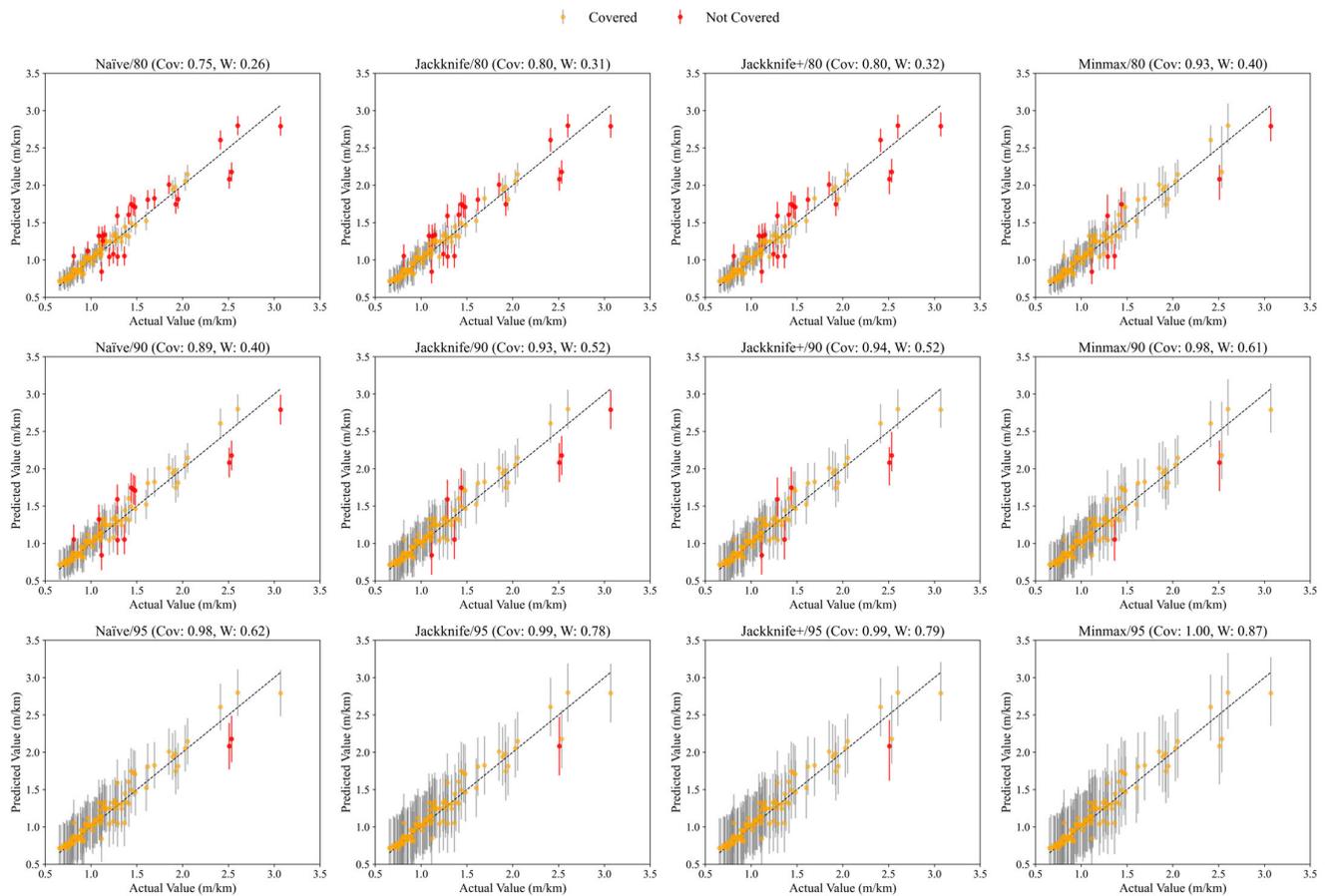
Figure 13. The prediction intervals generated by conformal prediction methods. (Cov: effective coverage, W: average interval width).

4.4. Optimal Uncertainty Model

As shown in the previous section, we developed 12 uncertainty models based on four conformal prediction methods and three target coverage levels. To find the optimal uncertainty model, a GRA is performed. GRA uses a metric (gray relational grade (GRG)) to compare various alternatives based on multiple criteria. The GRG ranges between 0.33 (the worst possible option) and 1 (the best possible option) [42]. The criteria in GRA are effective coverage and average interval width. That is, GRA aims to find the optimal uncertainty model with the highest effective coverage and minimum average interval width. The GRG of uncertainty models is presented in Figure 14. Drawing on the results, Minmax/80 is the optimal uncertainty model for IRI prediction with a GRG of 0.667, followed by Minmax/95, Naïve 80, and Minmax/90. Hence, the Minmax technique outperforms other conformal prediction methods when comparing effective coverage and average interval width simultaneously. The Minmax/80, as the optimal uncertainty model, has an effective coverage of 93.4% with an average interval width of 0.256 m/km. In other words, Minmax/80 provides a prediction range for testing data with an average width of 0.256 m/km, and for 93.4% of observations, the actual IRI is within the range.
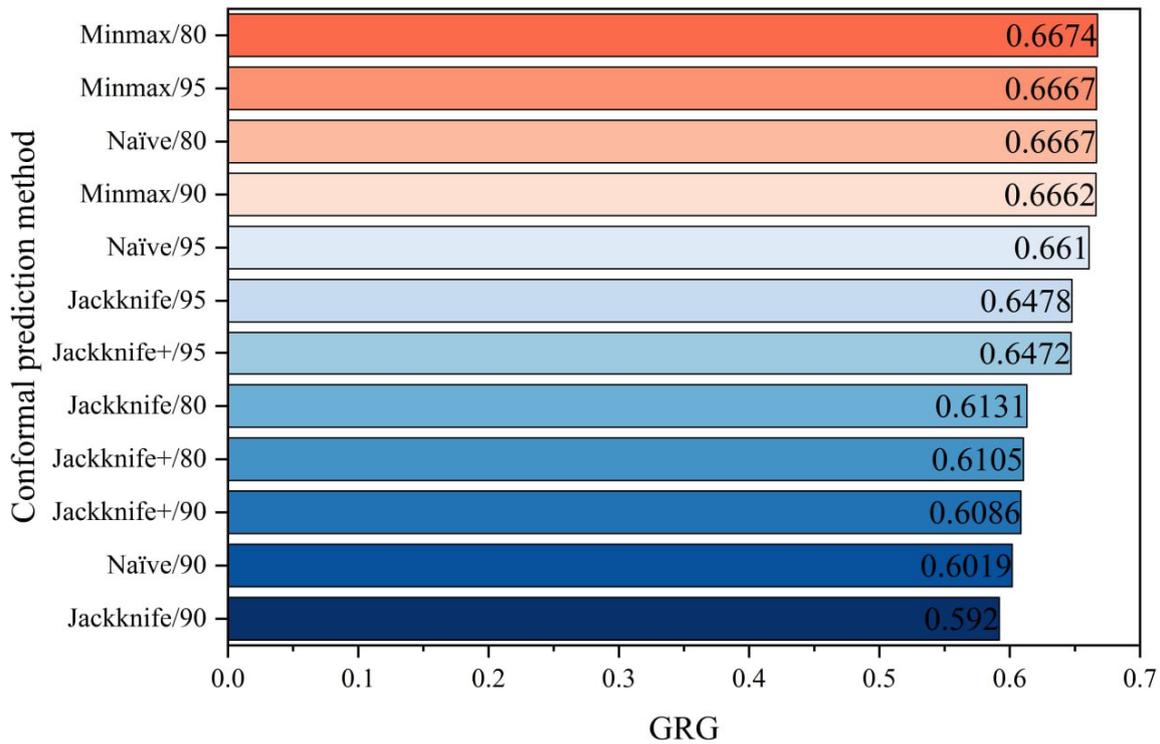
Figure 14. The GRG of various uncertainty models.

4.5. Uncertainty Parameters

After detecting the optimal uncertainty model (Minmax/80), its results were used to identify uncertainty parameters. To this end, the interval width of testing data observations is considered the dependent variable, and other variables shown in Table 1 are independent variables. Then, an MLR model is performed, and the variables with a *p*-value of less than 0.05 are considered uncertainty parameters in IRI prediction. The uncertainty parameters and their coefficients in the MLR analysis are presented in Table 4. As shown, the Initial IRI is an uncertain parameter, and a higher initial IRI increases the uncertainty in IRI prediction. The duration between data collections is another uncertainty parameter, and predicting IRI for longer durations increases the uncertainty. Further, lower ESALs increase the uncertainty in IRI prediction. The lower length of low severity transverse cracks and medium severity sealed transverse cracks, and the higher length of transverse cracks longer than 183 cm are associated with higher uncertainty levels. Moreover, in regions with higher total annual precipitation, the certainty about the predicted IRI is less than in other regions.

Table 4. Uncertainty parameters in IRI prediction.

Variable	Coef.	Std. Err.	t	$p > |t|$
Initial IRI (m/km)	7.83×10^{-2}	0.015	5.176	0.000
The duration between data collections (days)	1.88×10^{-7}	0.000	6.550	0.000
ESALs (number of loads)	-3.76×10^{-8}	0.000	-2.479	0.016
Length of low severity transverse cracks (m)	-5.01×10^{-3}	0.002	-2.277	0.026
Length of medium severity sealed transverse cracks (m)	-6.98×10^{-2}	0.035	-2.004	0.049
Length of transverse cracks longer than 183 cm (m)	1.18×10^{-2}	0.006	2.126	0.037
Total annual precipitation (mm)	3.05×10^{-5}	0.000	2.268	0.026

5. Conclusions

This study investigated IRI prediction and the associated uncertainty with these models. Previous studies generally applied a few independent variables to predict IRI. However, this study collects a comprehensive dataset including all the essential features to maximize the performance of IRI prediction models, and a dataset of over 500 observations and 35 independent variables was used for modeling. Another contribution of this study is to apply powerful machine learning models that can be widely applied to predict IRI, such as LGB and CatB. Another novelty of this investigation is the application of a conformal prediction interval framework to analyze the uncertainty in IRI prediction and to identify the factors that contribute to increased uncertainty. In the following paragraphs, a summary of the objectives and results is presented.

The first objective of this study was to detect the best-performing prediction method for IRI prediction. Ten prediction methods were tuned and trained on the dataset, and their performance was compared based on different metrics. The results showed that the light gradient boosting machine (LGB) is the best-performing method, and it could predict the IRI of testing data observations with an R^2 of 0.927, an MAE of 0.09 m/km, an MSE of 0.016 $(m/km)^2$, an EV of 0.926, and an MAPE of 7.2%.

The second objective of this investigation was to analyze the relative influence of independent variables on IRI prediction. In this regard, SHAP was synchronized with the best-performing method (LGB) to sort variables based on their relative influence on IRI. The results suggested that initial IRI was the top-ranked variable on IRI, followed by mean annual temperature, the duration between data collections, and the length of high-severity longitudinal cracking in the non-wheel path. A more detailed look at the SHAP results revealed that higher initial IRI, lower mean annual temperature, longer durations, and higher length of high-severity longitudinal cracking in the non-wheel path led to higher IRI values.

The third goal of this study was to determine the optimal uncertainty model for IRI prediction. To this end, 12 uncertainty models are developed by combinations of four conformal prediction methods and three target coverage levels. These methods were compared by effective coverage and average interval width. Then, a multi-criteria decision-making method (GRA) was utilized to identify the optimal uncertainty model. The outcomes of GRA suggested that Minmax/80 was the optimal model with a GRG of 0.667. This method could cover 93.4% of testing data observation with an average interval width of 0.256 m/km. The Minmax/80 can be applied to predict the upper bound of IRI. Therefore, practitioners can apply the upper bound of predicted IRI in the pavement M&R planning for the pavement sections where the IRI is critical.

The fourth objective of the current study was to determine which parameters and conditions lead to an increase in the uncertainty of IRI prediction. A further analysis was performed on the outcomes of the optimal uncertainty model, and the results showed that higher initial IRI, longer duration of IRI prediction, lower ESALs, lower length of low severity transverse cracks, lower medium severity sealed transverse cracks, higher length of transverse cracks longer than 183 cm, and higher total annual precipitation were all associated with increased uncertainty in the predicted IRI.

These results offer actionable insights for pavement engineers and decision-makers. First, they can apply LGB for IRI prediction in pavement M&R planning as a reliable tool, and therefore, the accuracy for pavement M&R planning can be increased. Second, the results of SHAP can help researchers and practitioners identify the most influential variables on IRI. Therefore, they can use this information to monitor and manage the most impactful variables and develop powerful pavement management systems. We also present the optimal uncertainty model (Minmax/80) that can be used in conservative M&R planning,

particularly in the sections where safety and service quality are significant concerns. This study introduces the parameters that can increase uncertainty in IRI prediction. Hence, the decision-makers can be informed of situations where the uncertainty of IRI prediction is higher, and they need to apply stochastic programming to plan M&R activities for pavements under such conditions.

One of the limitations of this study is that it checks only three target coverage levels. To address this, future studies can consider more target coverage levels to detect the optimal uncertainty methods. Another limitation of this study is that it does not check the influence of data size on the uncertainty of IRI prediction. Hence, it is recommended that future studies use different datasets of different sizes to check the relationship between dataset size and the level of uncertainty. Another important limitation is the lack of geographic diversity in the data used. The findings may not generalize well to regions with different road conditions or construction practices. Hence, future studies are encouraged to apply the proposed technique in this study to other case studies and compare their results with the findings of the current study.

Author Contributions: Conceptualization, S.G., H.N., and F.S.J.; methodology, S.G., H.N., and F.S.J.; software, S.G., H.N., and F.S.J.; formal analysis, S.G., H.N., and F.S.J.; investigation, S.G., H.N., and F.S.J.; data curation, S.G., H.N., and F.S.J.; writing—original draft preparation, S.G., H.N., and F.S.J. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: We used LTPP data, which is available at: https://infopave.fhwa.dot.gov/ (accessed on 10 June 2025).

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

References

1. Zhang, N.; Alipour, A. A Two-Level Mixed-Integer Programming Model for Bridge Replacement Prioritization. *Comput. Civ. Infrastruct. Eng.* **2020**, *35*, 116–133. [CrossRef]
2. Zalama, E.; Gómez-García-Bermejo, J.; Medina, R.; Llamas, J. Road Crack Detection Using Visual Features Extracted by Gabor Filters. *Comput. Civ. Infrastruct. Eng.* **2014**, *29*, 342–358. [CrossRef]
3. Infrastructure Report Card. 2025. Available online: https://infrastructurereportcard.org/ROADS (accessed on 10 June 2025).
4. AASHTO NHTSA: Motor Vehicle Traffic Fatalities Declined in 2023. Available online: https://aashtojournal.transportation.org/nhtsa-motor-vehicle-traffic-fatalities-declined-in-2023/ (accessed on 22 May 2025).
5. Naseri, H.; Shokoohi, M.; Jahanbakhsh, H.; Golroo, A.; Gandomi, A.H. Evolutionary and Swarm Intelligence Algorithms on Pavement Maintenance and Rehabilitation Planning. *Int. J. Pavement Eng.* **2021**, *23*, 4649–4663. [CrossRef]
6. Naseri, H.; Aliakbari, A.; Javadian, M.A.; Aliakbari, A.; Waygood, E.O.D. A Novel Technique for Multi-Objective Sustainable Decisions for Pavement Maintenance and Rehabilitation. *Case Stud. Constr. Mater.* **2024**, *20*, e03037. [CrossRef]
7. Wu, Y.; Zhang, Q.; Wang, Y.; Zhu, X. Advanced Hybrid CNN-GRU Model for IRI Prediction in Flexible Asphalt Pavements. *J. Transp. Eng. Part B Pavements* **2025**, *151*, 04025003. [CrossRef]
8. Al-Suleiman (Obaidat), T.I.; Shiyab, A.M.S. Prediction of Pavement Remaining Service Life Using Roughness Data—Case Study in Dubai. *Int. J. Pavement Eng.* **2003**, *4*, 121–129. [CrossRef]
9. Tsunokawa, K.; Schofer, J.L. Trend Curve Optimal Control Model for Highway Pavement Maintenance: Case Study and Evaluation. *Transp. Res.-A* **1994**, *28*, 151–166. [CrossRef]
10. George, K.P. *MDOT Pavement Management System Prediction Models and Feedback System*; by The University of Mississippi MDOT Pavement Management System: Prediction Models and MS-DOT-RD-00-119; University of Mississippi Mississippi Department of Transportation: Jackson, MS, USA, 2000.
11. Choi, J.-H.; Adams, T.M.; Bahia, H.U. Pavement Roughness Modeling Using Back-Propagation Neural Networks. *Comput. Civ. Infrastruct. Eng.* **2004**, *19*, 295–303. [CrossRef]
12. Albuquerque, F.; Núñez, W. Development of Roughness Prediction Models for Low-Volume Road Networks in Northeast Brazil. *Transp. Res. Rec.* **2011**, *2205*, 198–205. [CrossRef]
13. Khattak, M.J.; Nur, M.A.; Bhuyan, M.R.U.K.; Gaspard, K. International Roughness Index Models for HMA Overlay Treatment of Flexible and Composite Pavements. *Int. J. Pavement Eng.* **2014**, *15*, 334–344. [CrossRef]

14. Jaafar, M.; Fahmi, Z. Asphalt Pavement Roughness Modeling Using the Artificial Neural Network and Linear Regression Approaches for LTPP Southern Region. In Proceedings of the Transportation Research Board 95th Annual Meeting. (No. 16-4191), Washington DC, USA, 10–14 January 2016.

15. Mazari, M.; Rodriguez, D.D. Prediction of Pavement Roughness Using a Hybrid Gene Expression Programming-Neural Network Technique. *J. Traffic Transp. Eng. (Engl. Ed.)* **2016**, *3*, 448–455. [CrossRef]

16. Gong, H.; Sun, Y.; Shu, X.; Huang, B. Use of Random Forests Regression for Predicting IRI of Asphalt Pavements. *Constr. Build. Mater.* **2018**, *189*, 890–897. [CrossRef]

17. Dalla Rosa, F.; Liu, L.; Gharaibeh, N.G. IRI Prediction Model for Use in Network-Level Pavement Management Systems. *J. Transp. Eng. Part B Pavements* **2017**, *143*, 04017001. [CrossRef]

18. Lin, J.; Yau, J.-T.; Hsiao, L.-H. Correlation Analysis Between International Roughness Index (IRI). In Proceedings of the By. Transportation Research Board 82th Annual Meeting, Washington, DC, USA, 12–16 January 2003; pp. 1–21.

19. Owolabi, A.O.; Sadiq, O.M.; Abiola, O.S. Development of performance models for a typical flexible road pavement in Nigeria. *Int. J. Traffic Transp. Eng.* **2012**, *2*, 178–184. [CrossRef]

20. *AASHTO Mechanistic-Empirical Pavement Design Guide: A Manual of Practice*, 2nd ed.; American Association of State Highway and Transportation Officials; American Association of State Highway and Transportation Officials: Washington, DC, USA, 2015.

21. Abdelaziz, N.; Abd El-Hakim, R.T.; El-Badawy, S.M.; Afify, H.A. International Roughness Index Prediction Model for Flexible Pavements. *Int. J. Pavement Eng.* **2020**, *21*, 88–99. [CrossRef]

22. Marcelino, P.; de Lurdes Antunes, M.; Fortunato, E.; Gomes, M.C. Machine Learning Approach for Pavement Performance Prediction. *Int. J. Pavement Eng.* **2021**, *22*, 341–354. [CrossRef]

23. Marcelino, P.; de Lurdes Antunes, M.; Fortunato, E.; Gomes, M.C. Transfer Learning for Pavement Performance Prediction. *Int. J. Pavement Res. Technol.* **2020**, *13*, 154–167. [CrossRef]

24. Damirchilo, F.; Hosseini, A.; Mellat Parast, M.; Fini, E.H. Machine Learning Approach to Predict International Roughness Index Using Long-Term Pavement Performance Data. *J. Transp. Eng. Part B Pavements* **2021**, *147*, 04021058. [CrossRef]

25. Long-Term Pavement Performance (LTPP) Dataset. 2025. Available online: https://infopave.fhwa.dot.gov/ (accessed on 10 June 2025).

26. Naseri, H.; Waygood, E.O.D.; Patterson, Z.; Alousi-Jones, M.; Wang, B. Travel Mode Choice Prediction: Developing New Techniques to Prioritize Variables and Interpret Black-Box Machine Learning Techniques. *Transp. Plan. Technol.* **2024**, *48*, 582–605. [CrossRef]

27. Naseri, H.; Waygood, E.O.D.; Wang, B.; Patterson, Z.; Daziano, R.A. A Novel Feature Selection Technique to Better Predict Climate Change Stage of Change. *Sustainability* **2021**, *14*, 40. [CrossRef]

28. Montgomery, D.; Peck, E.; Vining, G. Introduction to Linear Regression Analysis–Google Books. Available online: https://books.google.ca/books?hl=en&lr=&id=tCIgEAAAQBAJ&oi=fnd&pg=PR13&dq=Montgomery,+D.C.;+Peck,+E.A.;+Vining,+G.G.+Introduction+to+Linear+Regression+Analysis&ots=lgvgXwk0Sp&sig=pea9aztuHDbS7SoAIsj_vvtRCDw&redir_esc=y#v=onepage&q=Montgomery%252CD.C.%253 (accessed on 10 June 2025).

29. Murtagh, F. Multilayer Perceptrons for Classification and Regression. *Neurocomputing* **1991**, *2*, 183–197. [CrossRef]

30. Naseri, H.; Jahanbakhsh, H.; Khezri, K.; Shirzadi Javid, A.A. Toward Sustainability in Optimizing the Fly Ash Concrete Mixture Ingredients by Introducing a New Prediction Algorithm. *Environ. Dev. Sustain.* **2022**, *24*, 2767–2803. [CrossRef]

31. Song, Y.Y.; Lu, Y. Decision Tree Methods: Applications for Classification and Prediction. *Shanghai Arch. Psychiatry* **2015**, *27*, 130–135. [CrossRef] [PubMed]

32. Myles, A.J.; Feudale, R.N.; Liu, Y.; Woody, N.A.; Brown, S.D. An Introduction to Decision Tree Modeling. *J. Chemom.* **2004**, *18*, 275–285. [CrossRef]

33. Naseri, H.; Shokoohi, M.; Jahanbakhsh, H.; Karimi, M.M.; Waygood, E.O.D. Novel Soft-Computing Approach to Better Predict Flexible Pavement Roughness. *Transp. Res. Rec. J. Transp. Res. Board* **2023**, *2677*, 246–259. [CrossRef]

34. Biau, G.; Scornet, E. A Random Forest Guided Tour. *Test* **2016**, *25*, 197–227. [CrossRef]

35. Solomatine, D.P.; Shrestha, D.L. AdaBoost.RT: A Boosting Algorithm for Regression Problems. In Proceedings of the IEEE International Conference on Neural Networks–Conference Proceedings, Budapest, Hungary, 25–29 July 2004; Volume 2, pp. 1163–1168.

36. Friedman, J.H. Stochastic Gradient Boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [CrossRef]

37. Naseri, H.; Waygood, E.O.D.; Patterson, Z.; Wang, B. Who Is More Likely to Buy Electric Vehicles? *Transp. Policy* **2024**, *155*, 15–28. [CrossRef]

38. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; pp. 785–794.

39. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 2017, pp. 4766–4775.

40. Taquet, V.; Blot, V.; Morzadec, T.; Lacombe, L.; Brunel, N. MAPIE: An Open-Source Library for Distribution-Free Uncertainty Quantification. *arXiv* **2022**, arXiv:2207.12274.

41. Dietterich, T.G.; Hostetler, J. Conformal Prediction Intervals for Markov Decision Process Trajectories. *arXiv* **2022**, arXiv:2206.04860.

42. Panda, A.; Sahoo, A.K.; Rout, A.K. Multi-Attribute Decision Making Parametric Optimization and Modeling in Hard Turning Using Ceramic Insert through Grey Relational Analysis: A Case Study. *Decis. Sci. Lett.* **2016**, *5*, 581–592. [CrossRef]

43. Song, Y.; Wang, Y.D.; Hu, X.; Liu, J. An Efficient and Explainable Ensemble Learning Model for Asphalt Pavement Condition Prediction Based on LTPP Dataset. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 22084–22093. [CrossRef]

44. Erfani, A.; Shayesteh, N.; Adnan, T. Data-Augmented Explainable AI for Pavement Roughness Prediction. *Autom. Constr.* **2025**, *176*, 106307. [CrossRef]

45. Alzubaidi, H.; Magnusson, R. Deterioration and Rating of Gravel Roads: State of the Art. *Road Mater. Pavement Des.* **2002**, *3*, 235–260. [CrossRef]

46. El-Hakim, M.; Tighe, S.L. Impact of Freeze-Thaw Cycles on Mechanical Properties of Asphalt Mixes. *Transp. Res. Rec.* **2014**, *2444*, 20–27. [CrossRef]

47. Klein-Paste, A.; Dalen, R. The Fundamentals of Plowing, Anti-icing, De-icing and Sanding. In *Sustainable Winter Road Operations*; Wiley: Hoboken, NJ, USA, 2018; pp. 82–100, ISBN 9781119185161.