Machine Learning-Based Prediction of International Roughness Index for Continuous Reinforced Concrete Pavements

Ragaa T. Abd El-Hakim
Ahmed N. Awaad
Sherif M. El-Badawy

Follow this and additional works at: https://mej.researchcommons.org/home

Part of the Architecture Commons, Engineering Commons, and the Life Sciences Commons
ORIGINAL STUDY

Machine Learning-based Prediction of International Roughness Index for Continuous Reinforced Concrete Pavements

Ragaa T. Abd El-Hakim a,*, Ahmed N. Awaad b, Sherif M. El-Badawy b

a Department of Public Works Engineering, Faculty of Engineering, Tanta University, Tanta, Egypt
b Department of Public Works Engineering, Faculty of Engineering, Mansoura University, Mansoura, Egypt

Abstract

The International Roughness Index (IRI) serves as a crucial indicator for ride quality and user comfort. As road roughness escalates, road serviceability diminishes, resulting in reduced vehicle speed and increased travel time, and consequently higher carbon dioxide emissions. Predicting the IRI is of utmost importance for pavement management systems and sustainable development overall. While numerous studies have forecasted the IRI of flexible pavements, there is a notable scarcity of research focusing on rigid pavement performance prediction. This study addresses the gap in predicting IRI for Continuous Reinforced Concrete Pavements (CRCP), an understudied aspect of pavement engineering. Leveraging the Long-Term Pavement Performance database, different machine learning techniques were applied to different input parameter representations. There are 90 measurements for the data points of the IRI. The input variables include the initial IRI, counts of medium-severity and high-severity transverse cracks, counts of medium-severity and high-severity punchouts, the percentage of pavement surface with patching (ranging from medium to high severity in both flexible and rigid pavements), pavement age, freezing index, and the percentage of subgrade material passing through the No. 200 US sieve. Through data analysis and machine learning algorithms, an accurate IRI prediction model for CRCP is developed. The results of this study show that the adaptive boosting algorithm model for CRCP yielded very good prediction accuracy ($R^2 = 0.90$ and $0.83$ for training and testing datasets, respectively) with low bias. The study findings offer valuable insights into CRCP IRI prediction, benefiting pavement management and maintenance strategies.

Keywords: Adaptive boosting algorithm, Continuous reinforced concrete pavements, Ensemble learning, International roughness index, Machine learning, Modeling, Transverse cracks

1. Introduction

Roads play an important role in the economic growth and development of any country. Due to their importance and the high cost of construction and maintenance, pavement performance evaluation has been a primary concern of road designers, practitioners, and researchers. The smoothness of road surface does not only provide ideal conditions for ride quality and user comfort, but also for vehicle-operating conditions. Rough roads cause turbulence in vehicle movement, consequently leading to higher fuel consumption, motor exhaustion, and higher carbon dioxide and greenhouse emissions (Mirtabar et al., 2022).

The accurate performance evaluation of pavement condition enables the correct selection, decision, and prioritization of maintenance, rehabilitation, and reconstruction activities. Riding comfort, structural capacity, safety, and aesthetics are the main aspects of pavement condition at any time during its service life (Gadiya et al., 2015). It would be advantageous to achieve the previously mentioned objectives and furthermore, sustainably protect the...
environment and reverse the adverse effects of rough roads. Rigid pavements, when properly designed, are durable and their maintenance cost is lower comparatively to flexible pavements. They are best-known for rutting resistance. However, they still experience roughness when subjected to loads and environmental conditions. International Roughness Index (IRI) is an important indicator of pavement performance since it reflects not only pavement condition, but also user comfort and ride quality of the road. The accurate prediction of IRI is very crucial to pavement management system practitioners.

Considerable attention has been dedicated to modeling and predicting the IRI due to its significance as a crucial performance metric in pavement condition evaluation systems, playing a pivotal role in prioritizing maintenance activities. A substantial and expanding body of literature explores various modeling techniques for IRI predictions. The following subsections will show the research efforts dedicated for IRI predictions both in flexible and rigid pavements.

1.1. International roughness index predictions of flexible pavements

In Canada, a linear regression model based on the Long-Term Pavement Performance (LTPP) database established a correlation between IRI and pavement surface distresses under distinct climatic conditions. These developed models underwent calibration, validation, and were found to be statistically significant (Patrick and Soliman, 2019). A different regression model for the IRI pertaining to thin hot mix asphalt overlay was created using data from the US LTPP database. The key factors influencing this model included the initial IRI (measured just after construction) and the shape factor of the IRI deterioration curve. The findings suggested that the deterioration of IRI in high temperatures was primarily influenced by structural strength and equivalent single-axle loads, while in low temperatures, it was predominantly affected by the average annual precipitation (Qian et al., 2018). IRI for flexible pavements was forecasted based on the initial IRI and the age of the pavement. The model took into consideration the influences of climate, subgrade, treatment type, pavement type, traffic loading, and functional classification. Data for this analysis were gathered from a 10-year pavement management database spanning from 2005 to 2014, provided by the Texas Department of Transportation (TxDOT) (Dalla Rosa et al., 2017). IRI of rigid and flexible pavements using Laos Road Management System database was predicted by multiple linear regression (MLR) technique (Gharieb and Nishikawa, 2021).

Recently, there has been increasing interest in machine learning (ML) and Artificial Neural Network (ANN) predictions in pavement engineering problems (Ceylan et al., 2014). Thus, IRI modeling is considered in many research using pavement characteristics as feeding features and ML algorithms such as MLR, Random Forest Regression, and Gradient Boosting techniques (Abdelaziz et al., 2020; Marcelino et al., 2019; Guo et al., 2022). Another study focused on the analysis of a data set from the LTPP database to quantify the IRI of asphalt concrete using a back-propagation neural network (Choi et al., 2004). ANN was also used to develop time-dependent roughness prediction models for three types of pavements: Portland cement concrete pavement, asphalt overlay over concrete pavement, and asphalt pavement using data from Indiana PMS database to develop rational pay factor limits and construction smoothness specifications (Chou and Pellinen, 2005). A predictive model for IRI was formulated utilizing ANN for flexible pavements across four climatic zones (wet-freeze, dry-freeze, wet no-freeze, and dry no-freeze) within the US LTPP database. The model's inputs encompassed climate and traffic data, specifically considering annual average temperature, freezing index, maximum humidity, minimum humidity, precipitation, average daily traffic, and average daily truck traffic. The ANN model, characterized by a 7-9-9-1 architecture with a hyperbolic tangent sigmoid transfer function, demonstrated the most favorable performance with the lowest root mean square error (RMSE) recorded at 0.01 (Hossain et al., 2019). Some other models used both MLR analysis and ANNs to predict IRI for flexible pavements as a function of pavement age, initial IRI, transverse cracks, alligator cracks, and SD of the rut depth based on LTPP (Abdelaziz et al., 2020).

More recent models adopted different ML and data mining techniques. IRI was predicted based on data mining techniques for the flexible pavement in Indonesia ($R^2 \geq 0.7$). The model allows for the identification of the input parameters that control the behavior of the IRI, including equivalent single-axle load, longitudinal cracking, pothole, age, crack, $\text{IRI}_{\mu}$, and rutting (Rifai et al., 2015). ML algorithms like random forest have been used to develop an IRI regression model of flexible pavements based on distress measurements, traffic patterns, climatic conditions, maintenance records, and structural data extracted from the LTPP program database (Gong et al., 2018). An IRI prediction model was also
developed based on fuzzy-trend time-series forecasting and particle swarm optimization techniques using the LTPP database. The model indicated very good results in terms of the RMSE (0.191) and relative error (6.37%) (Li et al., 2019). IRI was predicted for 5 and 10 years, based on LTPP data using different ML algorithms based on previous IRI measurements, structural, climatic, and traffic data (Marcelino et al., 2019). Other models predicted pavement roughness using the ANN by means of smartphone measurements (Alatoom et al., 2021).

1.2. International roughness index predictions of rigid pavements

One of the most important multilinear regression prediction models to predict rigid pavement IRI was the Continuous Reinforced Concrete Pavements (CRCP) smoothness model for the NCHRP 1-37 study as shown in equation (1) (NCHRP, 1999):

\[
\text{IRI} = \text{IRI}_0 + 0.003 \times \text{TC} + 0.008 \times \text{PUNCH} + 0.45 \times \text{SF} + 0.2 \times \text{PATCH} \tag{1}
\]

where

- \( \text{IRI}_0 \) = initial International Roughness Index, measured in meters per kilometer.
- \( \text{TC} \) = number of medium- and high-transverse cracks per kilometer.
- \( \text{PUNCH} \) = number of medium- and high-severity punchouts per kilometer.
- \( \text{PATCH} \) = percentage of pavement surface with patching (medium-to-high severity, applicable to both flexible and rigid pavements).
- \( \text{SF} \) = site factor, calculated as Age \((1 + \text{FI})\) \((1 + \text{P200})/1\ 000\ 000\).
- \( \text{Age} \) = pavement age in years.
- \( \text{FI} \) = freezing index, measured in degrees celsius days.
- \( \text{P200} \) = percentage of subgrade material passing through the 0.075-mm sieve.

The earlier CRCP model achieved a coefficient of determination \((R^2)\) of 0.60.

MLR was utilized to estimate the IRI of rigid and flexible pavements using Laos Road Management System database (Gharieb and Nishikawa, 2021).

IRI of Taiwanese roads was predicted as a function of pavement distresses based on a back-propagation ANN. The ANN showed that IRI may be predicted accurately from distress, which shows that IRI may completely reflect pavement distress conditions (Lin et al., 2003).

An ANN model was created to forecast the IRI for sections of Jointed Plain Concrete Pavement. The model took into account various factors, including the initial IRI value, pavement age, transverse cracking, percentage of joints spalled, areas with flexible and rigid patching, total joint faulting, freezing index, and the percentage of subgrade material passing through the No. 200 US sieve. The dataset utilized for constructing the model comprised 184 data points obtained from the LTPP database (Abd El-Hakim and El-Badawy, 2013). Another model for Jointed Plain Concrete Pavement IRI prediction was created using a hybrid wavelet optimally pruned extreme learning machine based on the LTPP database (Kaloop et al., 2022).

As represented in the literature, most of the previous studies were focused on the prediction of the IRI of flexible pavement and there were only limited studies that addressed the prediction of the IRI of CRCP. Moreover, the previous studies on predicting IRI on CRCP were focused on utilizing ANN, hence, it is important to investigate different ML techniques and algorithms to find out the best algorithms for predicting the IRI of CRCP.

2. Objectives and contributions

From the aforementioned literature studies, the key problem with these studies is that, despite the large number of research papers and the efforts spent to model and predict IRI for flexible pavements, there are very few studies that focused on the prediction of IRI for rigid pavements and especially CRCP. Therefore, the main objective of this study is to develop an accurate IRI prediction model for CRCP pavements based on the LTPP database using different types of ML techniques with different compositions of input parameters. By achieving highly accurate IRI predictions for CRCP, this research will significantly contribute to:

1. Enhanced pavement management: CRCP owners and maintenance agencies can gain valuable insights for proactive maintenance and rehabilitation strategies.
2. Advanced data-driven decision-making: this framework establishes a data-driven approach for CRCP IRI prediction, paving the way for future research and innovation in pavement engineering.

3. Data collection and analysis

The dataset employed in this study is identical to the dataset utilized for the rigid pavement smoothness model within the Mechanistic Empirical
Pavement Design Guide (NCHRP, 1999). These data are a part of the LTPP General Pavement Studies GPS-5 experiment for CRCP pavements. The data set consists of 90 data points corresponding to IRI and distress measurements. The dataset features are estimated initial IRI after construction (IRI0), counts of medium-severity and high-severity transverse cracks, counts of medium-severity and high-severity punchouts, percentage of pavement surface with patching (ranging from medium to high severity in both flexible and rigid pavements), pavement age in years, freezing index, and the percentage of subgrade material passing through the No. 200 US sieve. Table 1 shows the descriptive statistics for this data set.

Fig. 1 describes the distribution of the dependent output IRI and the other input-independent variables in the dataset. The figure demonstrates that the distributions of input parameters and output IRI are not fully normal, which denotes that the relationship between the input variables and IRI is nonlinear. This implies that multilinear regression analysis is not the optimum prediction model. The punchouts histogram shows that there are only three records that are not 0 in the dataset, therefore, punchouts were dropped and were not used as input features for the models.

To measure the collinearity between the variables, the Pearson correlation coefficient matrix was developed as in Fig. 2, the Pearson correlation coefficient of 0.69 indicates that IRI is highly correlated to the initial IRI value. Punch comes second in correlation with Pearson correlation of 0.43, while Patch and the %Passing sieve No. 200 come in the third place with a Pearson correlation coefficient of 0.33. Furthermore, the relationship between the IRI (the output-dependent variable) and the input-independent variables was investigated to point out any noteworthy trends as depicted in Fig. 3. It is noticeable that there are no visibly recognized patterns between the IRI and the independent variables, except with IRIo.

4. Research methodology

To detect the model that best describes the relationship between input parameters and IRI, a collection of ML modeling techniques, from simple to complex models, were developed using Python programming language. The lazy Regressor module from Lazypredict Python package developed by Shankar et al. (Welcome to Lazy Predict), which has a set of popular ML linear and nonlinear algorithms, was utilized to find out the most efficient algorithm for data representation. This module consists of 40 ML models from simple models such as simple linear regression, and decision trees to complex nonlinear models such as bagging and boosting ensembles and support vector machine models. The module trains all these ML techniques on the training dataset, after that, it uses the test data to sort the algorithms based on its results, which represents a decent way to select the appropriate technique for the dataset. The module was employed in different ML studies such as predicting the shear capacity of steel, cancer studies, and detecting diabetes (Garai et al., 2023; Nguyen et al., 2022; Elamary et al., 2023).

On the other hand, ANN were not the appropriate technique due to the small number of records’ dataset because ANNs often require a large amount of data for effective training and generalization (Kaplan et al., 2020).

It is important to realize that IRI depends on the pavement type and the kind of distress, the relative influence of material qualities, and distresses on IRI can change. However, in general, distresses have a higher effect on IRI than material features. This is because although material qualities only indirectly affect IRI through their influence on pavement smoothness, distresses have the ability to directly induce bumps and dips on the pavement surface (Mücka, 2019). Thus, in this study, the focus was to predict the IRI of CRCP centered on the pavement distresses and slightly consider the material characteristics using P#200 that was accessible in the dataset.

Table 1. Descriptive statistics of International Roughness Index and input parameters for the prediction model.

<table>
<thead>
<tr>
<th>Input parameters</th>
<th>Age (years)</th>
<th>IRI0 (m/km)</th>
<th>TC</th>
<th>Punch</th>
<th>Patch</th>
<th>FI</th>
<th>P200</th>
<th>SF</th>
<th>IRI (m/km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>90 points</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>30.6</td>
<td>1.71</td>
<td>303.6</td>
<td>13.2</td>
<td>4</td>
<td>1095.9</td>
<td>97.3</td>
<td>0.88</td>
<td>2.60</td>
</tr>
<tr>
<td>Minimum</td>
<td>3.9</td>
<td>0.75</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.70</td>
<td>0</td>
<td>0.79</td>
</tr>
<tr>
<td>Mean</td>
<td>16.78</td>
<td>1.22</td>
<td>40.99</td>
<td>0.37</td>
<td>0.21</td>
<td>225.40</td>
<td>49.60</td>
<td>0.17</td>
<td>1.45</td>
</tr>
<tr>
<td>Median</td>
<td>17.65</td>
<td>1.20</td>
<td>16.50</td>
<td>0</td>
<td>0</td>
<td>94.7</td>
<td>49.30</td>
<td>0.05</td>
<td>1.43</td>
</tr>
<tr>
<td>SD</td>
<td>6.43</td>
<td>0.28</td>
<td>62.09</td>
<td>1.81</td>
<td>0.59</td>
<td>262.84</td>
<td>27.81</td>
<td>0.23</td>
<td>0.36</td>
</tr>
</tbody>
</table>

FI, freezing index (°C days); IRI0, initial International Roughness Index (m/km); P200, percent subgrade material passing the 0.075-mm sieve; Patch, percentage pavement surface with patching (M–H severity flexible and rigid); Punch, number of medium-severity and high-severity punchouts/km; TC, number of medium-transverse and high-transverse cracks/km.
The study began as shown in Fig. 4 by identifying two types of input data representation. The first one was with six parameters and the second one was using four input features. The input parameters in the first representation were IRI0 (m/km), TC (cracks/km), PATCH (percentage pavement surface with patching), age (years), FI (°C days), and P#200 (%), while the second data representation style contained IRI0 (m/km), TC (cracks/km), PATCH, and site factor (SF).

In the second phase, the input parameters were standardized to have a mean equal to 0 and a SD equal to 1. Following this, the data was then randomly divided with stratifying on the values of measured IRI into two sets: training set (70%), and cross-validation set (30%).

After that, the Lazy Predict Python library was utilized to obtain the best algorithms to describe the link between input parameters and the output feature. At this time, the first five best algorithms were estimated using goodness-of-fit measurements and the best model was selected.

Finally, a deeper analysis was conducted on the selected model to measure its performance accurately.

To be mentioned, the most accurate algorithms in both data distributions were Decision Tree Regressor, Adaptive Boosting Regressor (AdaBoost), Extra Trees Regressor (XT), Extreme Gradient Boosting Regressor (XGB), and Random Forest Regressor (RF).

4.1. Machine learning techniques

4.1.1. Decision Tree Regressor

Decision Trees are multipurpose ML algorithms that can handle classification and regression. They
Fig. 2. Pearson’s correlation between the parameters.

Fig. 3. Scatterplot of IRI and IRI0, TC punch, patch, FI, P200, age, and SF. IRI, International Roughness Index; SF, site factor.
are sophisticated algorithms that can fit complex datasets. The Decision Tree Regressor attempts to forecast a continuous target variable by dividing the feature variables into tiny zones, with one prediction in each zone (Hamel, 2009). The algorithm begins with the initial decision node as in Fig. 5a, referred to as the root node. It represents the full dataset, which is partitioned further into two or more homogeneous sets. The decision nodes represent the characteristics of the dataset, the branches indicate the decision rules, and each leaf node represents the conclusion (Introduction to Decision Trees: Why Should You Use Them? | 365 Data Science and (n.d.)).

4.1.2. Adaptive boosting regressor

Freund and Schapire (1997) invented the first boosting algorithm in practice, AdaBoost, a supervised ensemble learning method, in 1995. AdaBoost's most-used estimator is decision trees with one level, which is decision trees with just one split. These trees are often referred to as Decision Stumps (Rojas, 2009). Each following model in the series (Decision Tree) seeks to improve the estimates provided by the previous model. This is accomplished by weighting the training dataset to focus more on instances of training in which past models made prediction errors as illustrated in Fig. 5b.

4.1.3. Extra trees regressor

The XT is also an ensemble technique for regression tasks based on Extremely Randomized Trees. Random Forest algorithm and Extra Trees algorithm are so similar, nonetheless, XT introduces additional randomness in the tree-building process. The model starts by data sampling for each tree in the ensemble, a random subset of the training data is sampled with replacement (bootstrapping). Following that, a random subset of features is considered for splitting the decision tree node. Unlike Random Forests, which use the best split among the randomly selected features, Extra Trees select random splits at decision tree nodes as shown in Fig. 5. Finally, majority voting as each tree in the ensemble predicts the target value, and the final prediction is obtained by averaging the predictions from all the trees (Geurts et al., 2006).
4.1.4. Extreme gradient boosting regressor

XGBoost (Extreme Gradient Boosting) is a strong gradient boosting-based ensemble learning algorithm developed by Tianqi Chen and Carlos Guestrin. The XGBoost method is a highly optimized and efficient variant of gradient boosting that uses regularization techniques to prevent overfitting and manage missing data. It gradually constructs an ensemble of weak prediction models, often decision trees, with each new model aiming to repair the faults of the preceding ones. The ensemble is first initialized by the model using a basic model, often a decision tree with a single node (a leaf). It estimates the negative gradient (residuals) of the loss function with regard to the current predictions for each following iteration (boosting round). This stage finds the samples where the model fails. To simulate the mistakes, the model applies a new decision tree to the negative gradients. The ensemble is then updated by adding the new tree with a learning rate to prevent overfitting. These procedures are repeated until the desired accuracy is achieved or the stopping condition is met. Finally, predictions are made by summing the predictions of all trees in the ensemble.

4.1.5. Random forest regressor

As illustrated in Fig. 5c, Random Forest is an also ensemble learning method. It builds multiple decision trees during training and combines their predictions to improve overall performance and reduce overfitting. The idea behind Random Forest is to create an ensemble of diverse and uncorrelated trees, which collectively provide more robust and accurate predictions. The learning process is similar to Extra Trees, except that the RF uses the best split, but XT randomizes the split to make more generalizations.

4.2. Model performance assessment

The goodness of fit and comparison between models was conducted using coefficient of determination ($R^2$), RMSE, Pearson’s correlation ($r$), and mean absolute error (MAE). Both RMSE and MAE measure the spread or errors. However, MAE is more robust in case there are outliers. The following equations illustrate the prementioned indicators:

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum (IRI_i - \overline{IRI_i})^2}{\sum (IRI_i - \overline{IRI})^2}$$  \hspace{1cm} (2)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (IRI_i - \overline{IRI_i})^2}{N}}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |IRI_i - \overline{IRI_i}|$$

$$r = \frac{\sum (IRI_i - \overline{IRI})(\overline{IRI_i} - \overline{IRI})}{\sqrt{\sum (IRI_i - \overline{IRI})^2 \sum (\overline{IRI_i} - \overline{IRI})^2}}$$  \hspace{1cm} (5)

where $\overline{IRI}$, and $IRI$ are the predicted and measured values of IRI, respectively, while $\overline{IRI}$, and $\overline{IRI}$ are the mean values of predicted and measured values of IRI; N is the number of data points.

5. Results and analysis

5.1. Proposed machine learning algorithm and data representation

Many highly employed ML modeling techniques from simple models like decision trees, linear regression, and k-nearest neighbors to complex and ensemble models, which consist of a set of parallel or sequences models, like Random Forest, were trained by lazy regressor. The performance for training and testing data for each ML technique was evaluated by the coefficient of determination ($R^2$) and RMSE, as shown in Tables 2 and 3 for six and four features, respectively, for the first five high-accuracy models.

To summarize the results, Fig. 6 is plotted to illustrate the $R^2$ for the top achiever algorithms in the Lazy Regressor training and testing phase. As depicted in the figure, the results for the four-feature representation, which has $R^2$ more than 0.8, were higher than the six-feature representation. It can also be noticed that Decision Tree Regressor, XT, and XGB techniques in both representations tend to overfit the data with $R^2$ equal to 1.0 for training datasets. The RF algorithm also has some overfitting as the gap between training and testing

<table>
<thead>
<tr>
<th>Model</th>
<th>Training dataset</th>
<th>Testing dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>RMSE</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>0.92</td>
<td>0.10</td>
</tr>
<tr>
<td>Extra Trees</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>XGB</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.95</td>
<td>0.08</td>
</tr>
</tbody>
</table>

MAE, mean absolute error; RMSE, root mean square error.

Figure 5c: Bar graph showing the performance of different models with $R^2$ values.

Table 2. Continuous reinforced concrete pavements roughness index six-feature best five models’ goodness-of-fit results.
datasets is wide, conversely, the results’ variation between training and testing datasets of AdaBoost algorithm was the smallest among the investigated techniques. Moreover, the \( R^2 \) values for test datasets for AdaBoost in four-feature and six-feature representations were 0.85 and 0.80, respectively, which represents the predomination of four-feature representation and AdaBoost Algorithm.

In addition to the previously mentioned comparison between ML techniques, the Taylor plot was developed for the same algorithms as represented in Fig. 7. In Taylor’s diagram, there are multiple indicators in one plot. In the \( x \) and \( y \) axes, SD is presented like quadrants as demonstrated in cyan lines, and correlation coefficients were depicted by radial cyan lines. While the RMSE is represented in dotted magenta lines. The closer the model is to the observation point, the lowest the SD is, the higher its correlation, and minimum RMSE with the test data. In Fig. 7a, most algorithms were far from observations, while in Fig. 7b, which represents the four-feature input parameters, the models’ points were nearest to the observation point, therefore, the four-feature input parameters were better in the data representation as the SF parameter played as a dimensionality reduction technique to reduce some of the existing collinearity.

**Fig. 6. Training and testing datasets \( R^2 \) for the most accurate models.**

(a) Training and testing datasets \( R^2 \)-squared for 6-feature input parameter models.

(b) Training and testing datasets \( R^2 \)-squared for 4-feature input parameter models.

**Table 3. Continuous reinforced concrete pavements roughness index four-feature best five models’ goodness-of-fit results.**

<table>
<thead>
<tr>
<th>Model</th>
<th>Training dataset</th>
<th></th>
<th>Testing dataset</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( R^2 )</td>
<td>RMSE</td>
<td>MAE</td>
<td>( R^2 )</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.91</td>
</tr>
<tr>
<td>AdaBoost</td>
<td>0.92</td>
<td>0.09</td>
<td>0.08</td>
<td>0.85</td>
</tr>
<tr>
<td>Extra Trees</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.89</td>
</tr>
<tr>
<td>XGB</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.86</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.95</td>
<td>0.08</td>
<td>0.06</td>
<td>0.83</td>
</tr>
</tbody>
</table>

MAE, mean absolute error; RMSE, root mean square error.
5.2. Detailed analysis of the most accurate machine learning technique

The AdaBoost model with a four-feature input data style was adopted as the optimized modeling technique, due to its high accuracy in comparison with other models. The model was retrained, after hyperparameter optimization to ensure the model generalization, using 63 records (70% of the dataset), and tested by 27 records (70% of the dataset). Fig. 8 illustrates the relation between the observed and predicted data of the AdaBoost algorithm for

---

Fig. 7. Taylor plot for four-feature and six-feature representation results. a: Six-feature representation results. b: for feature representation results.
training and testing datasets. In Fig. 8a, the training dataset achieved $R^2$ of 0.9, and in Fig. 8b, the testing dataset achieved $R^2$ of 0.83.

Fig. 9 depicts the results of all LTPP dataset records on the generated AdaBoost model and literature-based NCHRP 1-37 equation. In Fig. 9a, the AdaBoost model achieved $R^2$ of 0.88, while the NCHRP achieved $R^2$ of 0.48 in Fig. 9b. Moreover, the RMSE of AdaBoost and NCHRP were 0.13 and 0.26 m/km, respectively. In addition to the quantitative results, it is obvious that from Fig. 9a and b, the points of the AdaBoost model were near the equality line contrary to literature-based model.

Moreover, due to the small amount of dataset, Leave One Out Cross-Validation technique was used to assess model performance as shown in Fig. 10a as the model training with the whole dataset, except one record for test and so on. This technique presents a less-biased indicator of the test data MAE compared with using only one test dataset. It is noticed that in Fig. 10b, the Leave One Out Cross-Validation test MAEs were less than 0.3 m/km, except for some outliers, which indicates the robustness of the model.

In addition to measuring the model accuracy, the study of residual analysis was conducted to ascertain there were not any trends in the residuals. Residuals are the deviations between the model's predicted value and the measured value in the test dataset. In Fig. 11, it is clear that the residuals were randomly dispersed and there was not any pattern,
Fig. 10. AdaBoost leave one out cross-validation. a: LOOCV technique (Cross Validation and Model Selection and (n.d.). b: AdaBoost LOOCV.

Fig. 11. AdaBoost residual analysis.

Fig. 12. Feature importance of AdaBoost model.
hence, the used model was appropriate to capture this relationship.

\[
\text{Residual} = \text{observed Value} - \text{predicted Value} \quad (6)
\]

Finally, a feature importance study was conducted on the proposed model to find out which input parameter is the most important than others. AdaBoost's feature importance is obtained from the feature importance provided by its base estimator. By using the Decision Tree as a base regressor, then the AdaBoost feature importance is established by the average feature importance provided by each Decision Tree. In Decision Trees, feature importance is computed as the decrease in node impurity weighted by the probability of reaching that node (Random Forest Regression). In Fig. 12, initial IRI was the most important feature in estimating the IRI value for CRCP, followed by SF with relative importance less than 25%, then the number of medium-and-high-transverse cracks with relative importance near 20%, and the final parameter was the percentage pavement surface with patching with slightly more than 10%.

6. Summary and conclusions

In summary, this study aimed to fill the gap in the literature by predicting the IRI for CRCP. By employing various ML techniques and data representations using the LTPP database, the following conclusions are established.

The comparison of ML algorithms revealed that the four-feature representation (IRI₀, TC, PATCH, and SF) outperforms the six-feature one (IRI₀, TC, PATCH, age, FI, and P#200), with AdaBoost consistently demonstrating strong generalization performance.

(1) The Taylor plot reinforced the effectiveness of the four-feature representation, reducing collinearity and enhancing model predictions.
(2) AdaBoost was identified with a four-feature input representation (IRI₀, TC, PATCH, and SF) as the most accurate model, outperforming other algorithms with high R² values of 0.9 and 0.83 for both training and testing datasets, respectively.
(3) Residual analysis confirmed the appropriateness of the AdaBoost model, as the residuals were randomly dispersed.
(4) The study's findings indicate that the input parameter 'initial IRI' is the most critical predictor of CRCP Roughness Index, followed by the 'SF,' 'number of medium-and-high-transverse cracks,' and 'percentage pavement surface with patching.'
(5) The study proposes a robust prediction model for CRCP Roughness Index, benefiting pavement management and maintenance strategies. The model's accuracy and feature importance insights provide valuable tools for optimizing infrastructure management and enhancing pavement performance assessment.

The simplicity and availability of the model make it rational to be applied by practitioners and researchers to estimate the IRI for CRCP. However, the limitation of the study was the limited number of records in the LTPP database, thus, for future work, it is important to train and validate the model on other datasets to ensure generalization.

Author credit statement

The authors confirm contribution to the paper as follows: study conception and design: Sherif M. El-Badawy and Ragaa T. Abd El-Hakim; data manipulation and Python coding: Ahmed N. Awaad; analysis and interpretation of results: Ahmed N. Awaad, Sherif M. El-Badawy, and Ragaa T. Abd El-Hakim; draft manuscript preparation: Ahmed N. Awaad, Sherif M. El-Badawy, and Ragaa T. Abd El-Hakim. All authors reviewed the results and approved the final version of the paper.

Conflict of interest

The authors declare no conflict of interest.

References