Proceedings of 14th IOE Graduate Conference
Peer Reviewed
Year: 2023 Month: December Volume: 14
ISSN: 2350-8914 (Online), 2350-8906 (Print)

IRI Prediction Model using Random Forest Regression

Malika Awal [a], Krishna Singh Basnet [b]

[a] *Department of Civil Engineering, Pulchowk Campus, IOE, Tribhuwan University, Nepal*
[b] *Road Boards Nepal*
[a] 078mscom008.malika@pcampus.edu.np, [b] krishnasinghbasnet@gmail.com

Abstract
Nepal, being a landlocked country, road network forms the major transport means. More than 60% of around 11,000 kms of national highway road network is blacktopped. Abundant resources are utilized during road construction and thus the road networks have an asset value representing a significant proportion of national wealth. In lack of timely maintenance, we are losing a great value of road assets. If pavement maintenance are done timely, over half the repair costs can be avoided. For proper planning of maintenance strategy, prediction of pavement performance is necessary. Pavement roughness relates directly to pavement performance reflecting pavement serviceability to the road users and is measured in terms of International Roughness Index (IRI). The pavements' IRI increases over time, leading to pavement deterioration. IRI prediction model represents the IRI increment rate due to the passage of time. In this study, IRI prediction model is developed using Random Forest Regression considering initial IRI, no. of years, rainfall, maximum and minimum temperature and traffic as input features affecting IRI. The model is trained using dataset from three highways and the model is evaluated with R-squared score of 0.811664 for test values and predicted values.The result of this model can be used to plan maintenance and rehabilitation strategy for the road sections and to optimize the road maintenance with maximum performance and minimum cost.

Keywords
road asset, IRI, Random Forest Regression

1. Introduction

As a landlocked country, road network forms the major transportation network. Abundant resources are utilized during road construction and thus the road networks have an asset value representing a significant proportion of national wealth. More than 60 percentage of around 11,000 kms of national highway road network is blacktopped. Based on survey of 2013/14, 5800 km of the SRN infers 57% in good condition, 31% in fair condition and 11% in poor condition, while standards requires 95% of SRN roads in good condition [1]. A study by [1] states that Nepal is losing the value of 1 billion USD due to the lack of proper maintenance work. The condition of road impacts the level of service, road users cost, accidents, environmental degradation and national development. Thus, road network requires the timely maintenance to preserve and keep the serviceable conditions of road assets nearly as possible in its original condition.

The functional condition of road refers to the safe and smooth riding surface. The riding quality of the pavement can be assessed through the roughness of the surface. Pavement roughness relates directly to the pavement performance reflecting pavement serviceability to the road users. Roughness measurements are made in terms of International Roughness Index (IRI). IRI is a geometric path characteristic since its value is proportional to the pavement surface. IRI is calculated over equally spaced intervals along the road profile and is expressed as the units of displacement over units of length. It is measured in meter per kilometer [2]. If the value of IRI is 10 m per km, the amplitude (up and down) of the road is 10 m in each road length km [3]. Pavement deterioration prediction models is a major component for pavement maintenance system. Pavement condition can be assessed by the IRI value. As pavement condition deteriorates, its value increases. Many agencies use IRI as a performance measure for assessing the pavement condition of its National Highway System. The [4] made a study regarding the effectiveness of IRI compared to SDI system for assessing the quality and performance and its findings indicate that though there is no significant difference between the value generated by IRI or SDI, IRI can be used singly to replace SDI. The IRI increment is affected by the initial IRI, no. of years passed without maintenance, rainfall, maximum and minimum temperature [5]. IRI prediction model is developed in this study based on these parameters and use of random forest regression.

2. IRI Prediction Model

An ideal pavement management aims to maintain all pavement sections at a sufficiently high level of service and structural conditions. The pavements' IRI increases over time, leading to pavement deterioration. IRI prediction model represents the IRI increment rate due to the passage of time, which is affected by various factors like initial IRI value, climatic conditions, traffic loading, structural design of pavement, etc. Numerous studies have been made in this sector of pavement performance, with structural, climatic and traffic parameters as independent variables and IRI as performance indicator.

The study made by [6] considered the pavement thickness, structural number, precipitation, temperature, freeze index, humidity and traffic as the factors that affect the IRI of the pavement and IRI prediction model is developed using random forest regression. Likewise, the study made by [7]

indicated the optimum maintenance strategy of road based on IRI value considering the cracking, patching, potholes, shoving, rutting and raveling. This study employed sigmoidal function regression for the purpose. Based on age of pavement after last rehabilitation,the study made by [8] assessed the performance of pavement considering the ride index, rut index, fatigue index, transverse index, longitudinal index and maintenance index. A regression model is developed by [6] for IRI as dependent variable and subgrade, climate, pavement type, traffic loading, urban or rural road classification as independent variables. The study made by [9] used random forest regression to predict IRI based on pavement thickness, fatigue, initial IRI, traffic, temperature, precipitation, cracking (transverse, edge), rut, ravel, pothole and age of pavement.

Machine learning models have proved to be a valid alternative to classical models that uses linear statistical models for forecasting future values of an observed time series. Numerous studies have been made in this field and the results show that machine learning models can outperform traditional methods.

3. Methodology

In this study, an IRI prediction model is developed by the use of random forest regression algorithm and by considering base IRI, no. of years without maintenance, precipitation, temperature and traffic as independent variables which affects the resulting IRI value.

3.1 Data Collection

The required data for the study were collected from different secondary sources. The IRI and traffic data were obtained from Department of Roads (DoR) Nepal, Highway Management Information System (HMIS) unit database. The HMIS database has the IRI data for the national highways of Nepal from the year 2012 to 2021. However, the data for the year 2017 is missing and so the period 2012 to 2016 is considered for the study. The climatic data (precipitation, maximum temperature and minimum temperature) were obtained from [10]. It has the district wise data from 1981 to 2019. The climatic data for the period 2012 to 2016 were distributed district wise for the highway chainage. AADT in pcu for various chainage are considered for the traffic data. The required data are obtained from DoR, HMIS unit database. The data are available for some chainages only, so the data are shared in the nearby chainages. For the chainages with no traffic count, the traffic data from its successive station is considered, assuming that no other link roads join the successive station. Three highways; Mahendra highway, Tribhuwan highway and Araniko highway are considered in this study.

3.2 Random Forest Regression

Random forest is an advanced form of decision tree based method which uses the mean or mode as the prediction for the observations in the split regions. It is an advancement of the single classification and regression trees (CART) method, which follows a simple non-parametric regression approach [9]. In this paper, random forest regression algorithm is used using scikit-learn, with the use of parameters n estimators of 50 and random- state of 2. n estimators represents the number of trees in the forest. Higher number of trees gives better learning while slowing the training process considerably. Random state controls the randomness of the bootstraping of the samples used when building trees. Random forest regression for IRI prediction of pavement assists in prioritizing road maintenance and optimizing resource allocation for maintenance. For developing the IRI prediction model, the following independent variables are used.

Table 1: Sample dataset used for model development

Base IRI	No.	Tmax	Rainfall	Tmin	Traffic	IRI
4.978	1	27.12	1736.5	17.58	7506	5.66
4.21	1	27.53	1404.37	18.36	9115	6.6
4.25	1	27.55	1283.17	17.92	7394	4.4
4.6019	1	27.12	1736.5	17.58	7394	5
4.6019	2	27.55	1283.17	18.36	9115	6.3
4.6019	3	27.53	1804.37	17.58	7506	4.87
5.7701	1	30.37	1311.65	19.16	5198	5.86
5.7701	2	30.78	1244.19	19.69	7665	6.2
5.7701	3	31.25	1226.51	20.30	5960	6.2
5.7701	4	31.71	1136.79	20.83	11053	8.0

Initial IRI The IRI value for the base year is considered as initial IRI. It represents the minimum of the consecutive increasing IRI value for the particular section.

No. of years This value is the number of years after the base year with increment in IRI value. If the IRI value for year "T+1" is higher than for the year "T" in particular section, it is considered that no maintenance works are carried out in that section.

Rainfall Annual rainfall in mm is considered in the study and district wise distribution of rainfall is made.

Temperature Average of annual maximum and minimum temperature with district wise distribution is considered in this study.

Traffic AADT in pcu is considered for the traffic.

IRI This is the resulting IRI measured after no. of years of initial IRI.

Based on the above database with 2,778 datas of which 80% is considered the training dataset and 20% is the testing dataset, maximum value for no. of years being 4, the model is trained.

3.3 Model Training

The Random Forest Regression is carried out through coding in Jupyter notebook in python environment. In this model developed, base IRI, no of years, rainfall, maximum temperature, minimum temperature and traffic are the input features while IRI is the target variable. train_test_split from scikitlearn is used to split the data into training and testing

sets. The testing set contains 20% of the data and the training set contains the remaining 80%. A Random Forest Model is developed importing the RandomForestRegressor class from scikitlearn with 50 trees and a fixed random seed of 2. The model is trained on the training data using the 'fit' method. 'y_train.values.ravel()' is used to convert the target variable into a 1D array. It then predicts the target variable for the test data.

3.4 Model Evaluation

The model is evaluated based on the Rsquared score for training data. It measures how well the model fits the training data. The Mean Absolute Error(MAE) and Mean Squared Error (MSE) is calculated for the training data. The Rsquared score for actual test data and predicted test data is calculated to assess the accuracy of prediction of the model. The feature importances for each input feature is also calculated.

4. Result and Discussion

The R-squared score for training data is calculated to be 0.950357. The value of R-squared score nearing to 1 indicates that the model explains the variance in the dependent variable with a good fit. A R-squared score of approximately 0.950357 is very close to 1 suggesting that the model is fit for predicting the variability in the dependent variable using the given independent features. It shows that the prediction of resulting IRI based on initial IRI, no. of years, rainfall, max and min temperature and traffic is significant and 95% of the variability in IRI can be explained based on the above input features. Likewise, the R-squared score for test value and predicted value is found to be 0.811664.

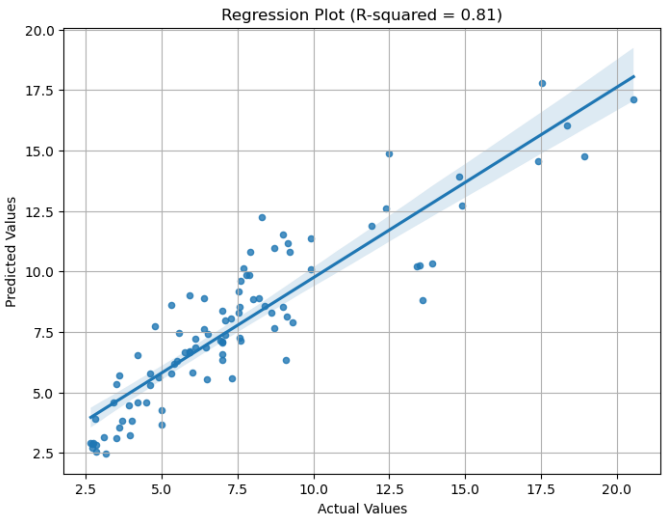

Figure 1: regression for predicted and actual values

The mean absolute error (MAE) for test data and predicted data is calculated to be 0.504261, which shows that, on average, the model's IRI predictions vary roughly by 0.504261 m per km. Likewise, the Mean Squared Error (MSE) for test data and predicted data is calculated to be 0.558705. The feature importance for the input features are found to be 0.708, 0.069, 0.052, 0.049, 0.035 and 0.086 respectively indicating initial IRI is the most crucial feature for predicting IRI in this model. The current state of IRI is a critical factor in determining what IRI will be in the following year. The other parameters have a relatively low importance.

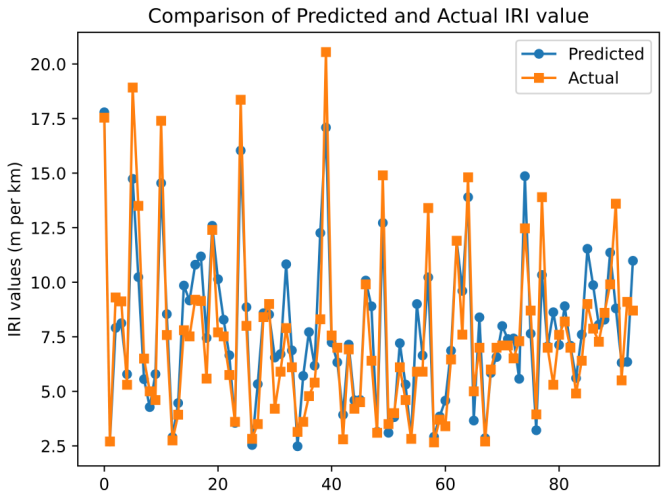

Figure 2: Predicted vs Actual IRI

5. Conclusion

An IRI prediction model is developed using the random forest regression considering base IRI, no. of years, rainfall, maximum and minimum temperature and traffic as input features. The developed model is able to predict IRI value to accuracy of regression coefficient of 0.81. The result of this model can be used to plan maintenance and rehabilitation strategy for the road sections and to optimize the road maintenance with maximum performance under minimum cost.

6. Limitations and Recommendations

A major limitation of this study is only initial IRI, no. of years, rainfall, maximum and minimum temperature and traffic are considered to impact IRI value. However, other than these, there are various factors like thickness of pavement, subgrade condition, structural number of pavement, geography, etc. which affects the IRI value. These factors are not considered in this study. Also, the pavement deterioration for both the newly constructed road and older roads after maintenance are considered the same in this study. Considering these parameters and training the model with larger number of data from the other national highways can yield better result.

Acknowledgments

The authors acknowledge the support of Department of Roads for the database and Department of Civil Engineering for the proper guidance, including all the helping hands for the successful completion of the work.

References

[1] Abhiman Das Mulmi et al. Assessment of performance based road maintenance practices in nepal. *Open Journal of Civil Engineering*, 6(02):225, 2016.

[2] Muhammad Mubaraki. Highway subsurface assessment using pavement surface distress and roughness data. *International Journal of Pavement Research and Technology*, 9(5):393–402, 2016.

[3] Orlean G Dela Cruz, Christian A Mendoza, and Kristel D Lopez. International roughness index as road performance indicator: A literature review. In *IOP Conference Series: Earth and Environmental Science*, volume 822, page 012016. IOP Publishing, 2021.

[4] Hary Agus Rahardjo et al. The effectiveness of iri compared to sdi system for assessing the quality and performance of materials used in flexible pavement in java, indonesia. In *IOP Conference Series: Materials Science and Engineering*, volume 241, page 012007. IOP Publishing, 2017.

[5] Francisco Dalla Rosa, Litao Liu, and Nasir G Gharaibeh. Iri prediction model for use in network-level pavement management systems. *Journal of transportation engineering, part B: pavements*, 143(1):04017001, 2017.

[6] Pedro Marcelino, Maria de Lurdes Antunes, Eduardo Fortunato, and Marta Castilho Gomes. Machine learning approach for pavement performance prediction. *International Journal of Pavement Engineering*, 22(3):341–354, 2021.

[7] Amr A Elhadidy, Emad E Elbeltagi, and Mohammad A Ammar. Optimum analysis of pavement maintenance using multi-objective genetic algorithms. *HBRC Journal*, 11(1):107–113, 2015.

[8] Promothes Saha, Khaled Ksaibati, Rebecca Atadero, et al. Developing pavement distress deterioration models for pavement management system using markovian probabilistic process. *Advances in Civil Engineering*, 2017, 2017.

[9] Hongren Gong, Yiren Sun, Xiang Shu, and Baoshan Huang. Use of random forests regression for predicting iri of asphalt pavements. *Construction and Building Materials*, 189:890–897, 2018.

[10] Open Data Nepal. District-wise daily climate data for nepal, 2018.