Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2026 (44) 4** : 1–8.
ISSN-0037-8712
doi:10.5269/bspm.80219

Development of Pavement Condition Index Modelling Using Machine Learning Techniques

Dipanwita Mahata, Madhu Kundur Anand, Kanimozhee Sudalai, Abhoy Mahato, Praveena Kumara Kammasandra Munivenkataswamy

ABSTRACT: The Pavement performance modelling is critical for sustainable transportation infrastructure, since deterioration caused by traffic loads, weather conditions, and structural distress has a direct impact on serviceability and maintenance costs. This research compared two prediction approaches for predicting Pavement Condition Index (PCI): Artificial Neural Network (ANN) and Random Forest (RF) using key deterioration variables: patches, potholes, temperature, depressions and cumulative Equivalent Single Axle Loads (ESAL). The model demonstrated strong predictive performance, with RF achieving $R^2 = 0.916$ (RMSE = 3.42), and ANN attaining the highest precision with $R^2 = 0.961$ (RMSE = 2.34). Sensitivity analysis revealed that temperature, traffic loading, and potholes were the most important indicators, with patching and depressions having little significance. The ANN model improved better predictive capability and RF balanced accuracy with shifting importance. Collectively, technologies enhance improve data-driven pavement management by allowing for accurate PCI forecasting and permitting proactive, cost-effective maintenance planning for resilient transportation networks.

Keywords: Artificial neural network, random forest algorithm, pavement condition index.

Contents

1. Introduction

India's road network is rapidly expanding due to urbanization and economic expansion. In order to guarantee long-term performance, safety, and sustainability, this growth has increased the demand for dependable and effective pavement maintenance. Because they facilitate travel, trade, and social interaction, roads are essential to the growth of a country. However, surface distresses like cracking, rutting, potholes, and patching are brought on by aged pavements, environmental pressures, and growing traffic volumes. These deteriorations endanger travel safety, increase maintenance costs, and hasten structural failures if left unchecked.

The Pavement Condition Index (PCI) is a commonly accepted indicator used to assess the structural and functional condition of pavements. Poor pavement conditions raise vehicle operating costs, fuel consumption, and accident risks, maintaining optimal PCI levels is critical for both ride quality and economic efficiency. Manual field surveys, which are subjective, labour-intensive and resource-intensive, constitute the foundation of traditional PCI evaluation techniques. Pavement performance evaluation could be greatly enhanced by recent developments in machine learning (ML). Faster, more objective, and more cost-effective evaluation is made possible by machine learning techniques that can analyse historical and distress data to find intricate links and anticipate PCI. The researchers investigated at a variety of pavement and traffic conditions in [1], to construct predictive models for pavement performance, considering traffic intensity, distress types and environmental variables. In [2], following model prediction, employed morphological methods and a triple-thresholding technique were used to improve detection accuracy and fracture segmentation. The study in [3] also evaluated the relative importance

2020 *Mathematics Subject Classification:* 68T07, 92B20.
Submitted November 18, 2025. Published February 26, 2026

of various variables to demonstrate how weather conditions interact with seal coat applications in order to better understand the impact of environmental factors on pavement maintenance practices. In [4] the researchers included essential input parameters such as pavement serviceability index, pavement age, surface roughness, and additional traffic and structural aspects in order for the models to accurately depict the complex interdependencies influencing pavement behaviour. Additional variables were added in [5], including distress density, severity level, and the presence of manholes, in order to identify the main factors influencing pavement condition. A Physics-Guided Neural Network (PGNN) framework was presented, which integrates physical laws into the training process of Artificial Neural Networks (ANN) to predict the International Roughness Index (IRI) in [6]. The study in [7] intended to discover the most effective modelling technique capable of properly portraying pavement deterioration and assisting maintenance decision-making by conducting a thorough investigation of these variables. Researchers emphasized the automated detection and classification of typical pavement problems, such as alligator cracking, potholes, and other surface defects in [8]. In [9] a predictive paradigm was designed to help infrastructure managers adopt data-driven maintenance programs, particularly under unfavourable environmental conditions. The study in [10] investigated the potential of machine learning models in capturing subtle relationships between traffic loading, pavement structure, and roadway parameters in order to more accurately forecast pavement performance.

Our study was carried out on selected road stretches located in the state of Karnataka, India. Two road sections were identified for detailed analysis: The study areas selected are Kethuhalli to T-05 of Tumkur Taluk, Tumkur district, Pandarahalli to Godabanahal of Chitradurga Taluk, Chitradurga district. These stretches totalling to 11km (104 data) were chosen as representative samples for pavement condition assessment due to their exposure to mixed traffic, including heavy commercial vehicles, and varying environmental conditions.

2. Equivalent Single Axle Load (ESAL)

Traffic in terms of Equivalent Single Axle Load (ESAL) shall be considered for the design of pavement. If sufficient data are available at the stretch with respect to the wheel load distribution of commercial vehicles or the vehicle damage factor and their transverse placement, the cumulative standard axles may be worked out based on actual data. The design procedure is followed as per the guidelines provided in IRC SP-72. Procedure followed is as shown below:

1. The average annual daily traffic is calculated from following formula

$$\text{AADT} = \frac{T + (S \times T \times 0.6 \times t)}{365} \tag{2.1}$$

   Where,

$$T = \text{Total traffic for lean season}$$
$$S = \text{Number of harvesting season}$$
$$t = \text{Duration of single harvesting season}$$

2. Cumulative ESAL application over 10 years at 6% growth rate is calculated from following formula.

$$\text{Cumulative ESAL} = T' \times 4811 \times L \tag{2.2}$$

   Where,

$$T' = \text{ESAL application per day}$$
$$L = \text{Lane distributor factor}$$

3. From the calculated ESAL, category of traffic is found as per SP-72 guidelines

4. From Traffic category value and obtained CBR values of the soil samples from the lab tests crust thickness of the pavement is determined from the following chart as per IRC.SP-72:2007.

3. Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) are a subset of machine learning algorithms inspired by the functioning of the human brain. They are designed to recognize patterns, model complex nonlinear relationships, and make predictions based on input data.Figure 1 illustrates the architecture of the Artificial Neural Network (ANN) model developed for predicting the Pavement Condition Index (PCI). The diagram shows a feed-forward neural network consisting of three main layers: an input layer, two hidden layers, and an output layer. The connections between nodes represent the flow of information and the strength of the weights assigned during model training.

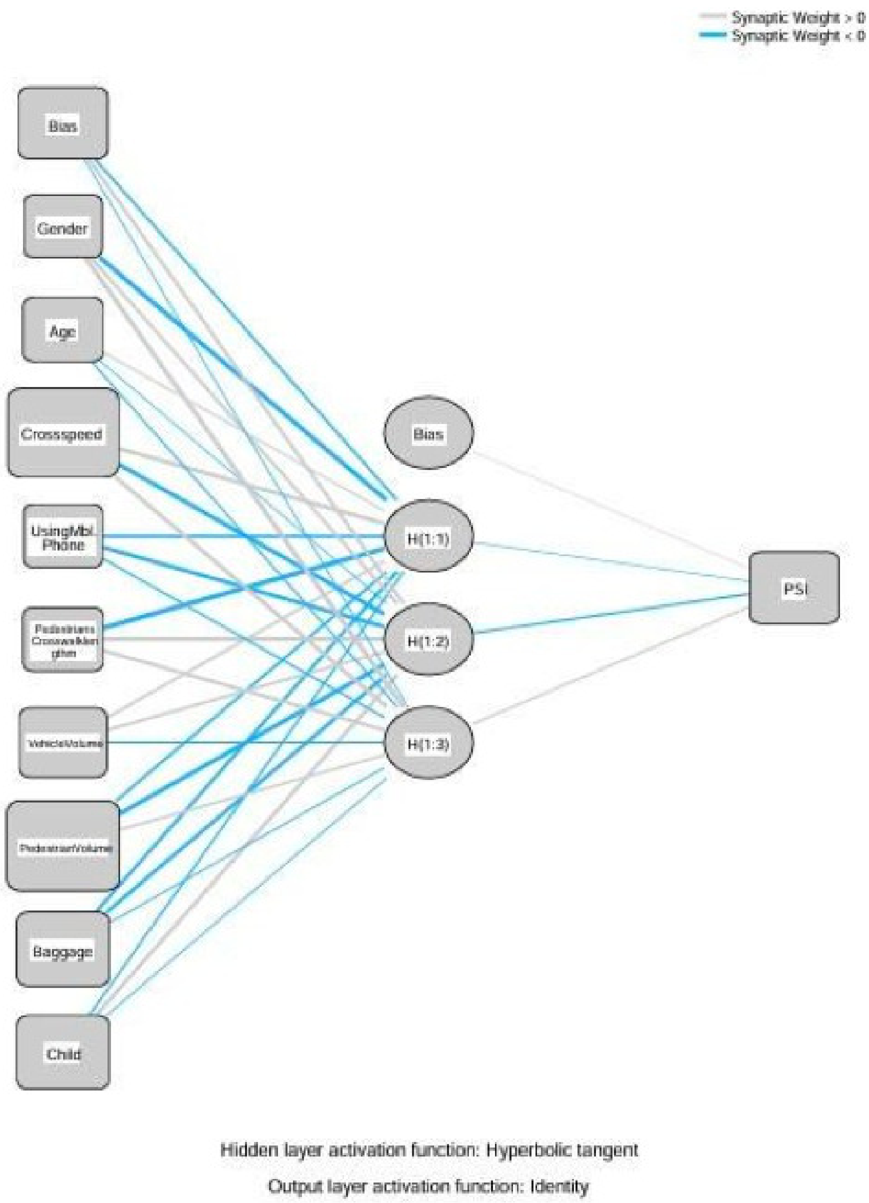

Figure 1: ANN modelling showing the hidden layers and output layers function.

A sensitivity analysis was conducted to evaluate the relative influence of the input variables on PCI prediction in the ANN model. The analysis evaluates the relative influence of each distress and environmental factor on PCI variation. Among all parameters, Cumulative ESAL (0.128) exhibits the highest sensitivity, indicating that traffic loading is the most dominant factor affecting pavement deterioration. Pothole (0.027) and Rutting (0.025) display moderate sensitivity, reflecting their significant impact on surface roughness and ride quality. Temperature (0.011) also has a noticeable influence, as high temperatures can soften asphalt binders and cause deformation, while low temperatures may induce thermal cracking. Depression (0.014) and Patch (0.007) show smaller effects, suggesting localized surface irregularities that slightly influence overall pavement condition. Cracking (0.005) records the lowest sensitivity, indicating minimal direct impact within this dataset, though it often initiates more severe distresses over time.

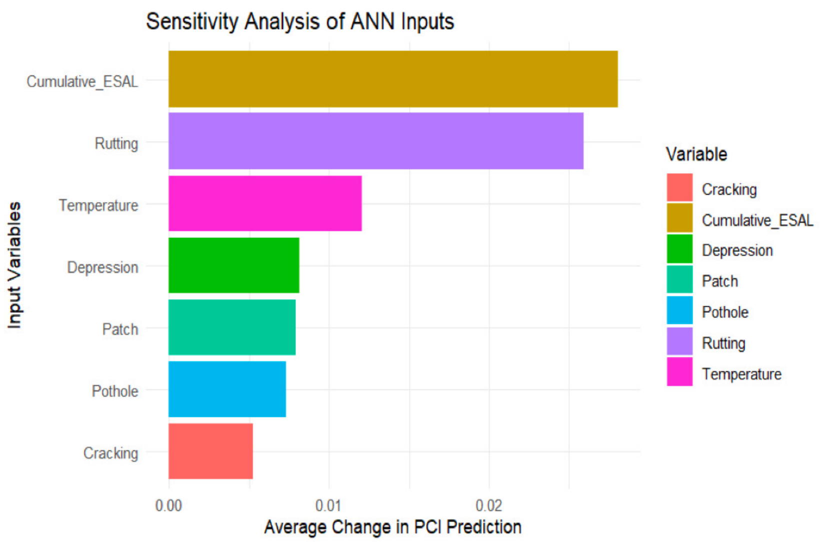

Figure 2: Sensitivity bar chart.

The results confirm that Cumulative ESAL is the most critical variable governing PCI deterioration, followed by pothole, rutting, and temperature, while minor surface defects contribute less. This finding highlights the ANN model's ability to capture both structural and environmental influences, providing a realistic understanding of pavement performance. From the figure 3, it can be observed that the ANN model demonstrates excellent predictive performance. The majority of the green points lie very close to the line of equality, showing minimal deviation between predicted and actual PCI values. This confirms that the ANN has successfully captured the underlying deterioration patterns in the dataset with very high accuracy.

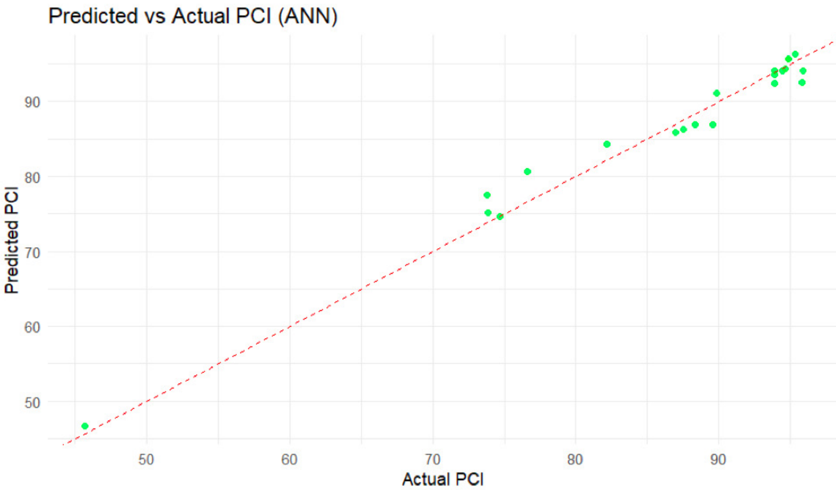

Figure 3: Predicted vs Actual PCI graph for Artificial Neural Network Model.

In our study, an Artificial Neural Network model was developed to predict the Pavement Condition Index based on key deterioration factors including pothole, rutting, patch, depression, cracking, temperature, and cumulative traffic loading expressed as Equivalent Single Axle Loads (ESALs). The ANN model demonstrated outstanding predictive capability, achieving an R^2 value of 0.96, which indicates that it was able to explain 96 percentage of the variation in PCI. This exceptionally high coefficient of determination highlights the strong relationship between the input variables and PCI, confirming the ability of ANN to capture complex nonlinear interactions that traditional models may overlook. Furthermore, the model yielded a Root Mean Squared Error (RMSE) of 2.34, signifying that prediction errors were minimal when compared to the PCI's scale of 0-100. The Mean Squared Error (MSE) of 5.47 was likewise the lowest among the evaluated models, including Random Forest (RF), confirming that the ANN provided the most accurate predictions.

4. Random Forest (RF) Model

Random Forest (RF) is an ensemble-based machine learning algorithm that has emerged as one of the most powerful and extensively used techniques for both regression and classification.

The variable importance analysis confirms that the Random Forest model not only provides accurate PCI predictions but also offers meaningful insights into the relative influence of deterioration factors. This information is particularly valuable for pavement management practices, as it helps to prioritize monitoring and maintenance activities on the most critical variables affecting pavement performance. The overall Gini impurity (0.388) confirmed acceptable classification purity, although some overlap across PCI classes was observed.

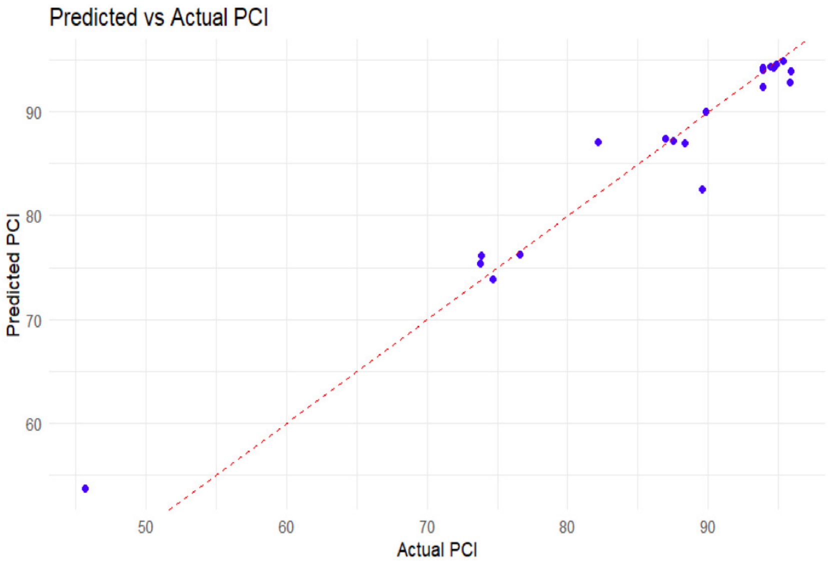

Figure 4: Predicted vs Actual PCI Plot for Random Forest Model.

Figure 5 presents the scatter plot of predicted PCI values against the actual PCI values obtained from the Random Forest regression model. The red dashed line represents the line of equality, where the predicted PCI would exactly match the observed PCI. The distribution of points around this line provides a visual indication of the model's predictive accuracy.

The visual distribution of the points, combined with the low error metrics, confirms that the Random Forest model is highly effective for PCI prediction. The results indicate that the model not only captures overall deterioration trends accurately but also provides sufficiently precise estimates for practical use in pavement management systems. The performance of the Random Forest model was evaluated based on three key statistical indicators: the Coefficient of Determination (R^2), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). These metrics provide a quantitative assessment of how well the model predicts the Pavement Condition Index (PCI) given input variables such as patches, depressions, potholes, temperature, and cumulative ESAL. The obtained R^2 value of 0.916 shows that approximately 91.6 percentage of the variation in PCI can be explained by the independent variables used in the model. This high number demonstrates a strong relationship between the predicted and observed PCI values, showing that the Random Forest model accurately captures the complex, nonlinear interactions among pavement distresses and influencing factors. The Root Mean Squared Error (RMSE) was calculated to be 3.42, representing the average deviation between the predicted and actual PCI values. This low RMSE indicates that the Random Forest model's predictions are close to the genuine observed values and have few prediction mistakes. Similarly, the Mean Squared Error (MSE) of 11.69482 supports the model's effectiveness by suggesting that the squared differences between predicted and real PCI values remain low across the dataset.

5. Conclusion

The present study set out to evaluate and compare the performance of two predictive modelling approaches, Random Forest (RF) and Artificial Neural Networks (ANN) for Pavement Condition Index (PCI) prediction. The motivation stemmed from the growing demand for accurate, reliable, and interpretable models to assist proactive pavement management, in which maintenance and rehabilitation strategies are driven by data-driven insights. Pavement deterioration is a complicated process driven by a variety of elements such as structural distresses, traffic loading, and environmental conditions, necessitating models capable of capturing both linear and nonlinear interactions. The comparison research demonstrates that while ANN provides the highest accuracy and provides a good blend of robustness and explanatory power. All models consistently achieved R^2 values above 0.90, indicating that the specified factors (potholes, patches, depression, temperature, cumulative ESAL, etc.) are good predictors of PCI.

Acknowledgments

The authors would like to thank the referees for their valuable comments and suggestions, which helped improve the work

References

1. Zhang, J., Guo, W., Song, B., Zhuo, Y., Zhang, Y., Performance prediction of asphalt pavement based on random forest. *Journal of Beijing University of Technology* 47(11), 1256–1263 (2021).
2. Peng, C., Yang, M., Zheng, Q., Zhang, J., Wang, D., Yan, R., Wang, J., Li, B., A triple-thresholds pavement crack detection method leveraging random structured forest. *Construction and Building Materials* 263, 120080 (2020).
3. Luo, X., Wang, F., Bhandari, S., Wang, N., Qiu, X., Effectiveness evaluation and influencing factor analysis of pavement seal coat treatments using random forests. *Construction and Building Materials* 282, 122688 (2021).
4. Hanandeh, S., Introducing mathematical modelling to estimate pavement quality index of flexible pavements based on genetic algorithm and artificial neural networks. *Case Studies in Construction Materials* 16, e00991 (2022).
5. Issa, A., Samaneh, H., Ghanim, M., Predicting pavement condition index using artificial neural networks. *Ain Shams Engineering Journal* 12(4), 4261–4271 (2021).
6. Chen, K., Jiang, Y., Xu, J., Zhang, J., Physics-guided neural network for predicting international roughness index of flexible pavements. *Engineering Applications of Artificial Intelligence* 129, 109922 (2024).
7. Ali, A. A., Milad, A., Hussein, A., Yusoff, N. I. M., Heneash, U., Predicting pavement condition index based on the utilization of machine learning techniques: a case study. *Journal of Road Engineering* 3(2), 63–73 (2023).
8. Elhadidy, A. A., El-Badawy, S. M., Elbeltagi, E. E., A simplified pavement condition index regression model for pavement evaluation. *International Journal of Pavement Engineering* 22(5), 643–652 (2019).
9. Ramachandraiah, S. T., Kumar, P., Pasupunuri, S. K., Shinganmakki, J. R., Evaluation of pavement surface distress using image processing and artificial neural network. *Journal of Testing and Evaluation* 51(4), 2041–2057 (2023).
10. Elseifi, M. A., Sarkar, M. T. A., Paudel, R., Abohamer, H., Mousa, M. R., A machine learning–based framework for predicting pavement roughness and aggregate segregation during construction. *Journal of Transportation Engineering, Part B: Pavements* 150(3), 04024011 (2024).

Dipanwita Mahata,
PG Scholar, School of Civil Engineering,
REVA University,
India.
E-mail address: mahatadipanwita44@gmail.com

and

Madhu Kundur Anand,
Assistant Professor,
School of Civil Engineering,
REVA University,
India.
E-mail address: madhu.ka@reva.edu.in

and

Kanimozhee Sudalai,
Assistant Professor,
School of Civil Engineering,
REVA University,
India.
E-mail address: Kanimozhee.s@reva.edu.in

and

Abhoy Mahato,
Consultant,
Deloitte Consulting India Pvt. Ltd,
Hyderabad
India.
E-mail address: Abhoym12@gmail.com

and

Praveena Kumara Kammasandra Munivenkataswamy,
Associate Professor,
Department of Mathematics,
School of Applied Sciences,
REVA University,
India..
E-mail address: praveenkumara.km@reva.edu.in