International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 13s, 2025
https://theaspd.com/index.php/ijes

An Investigation On Pavement Maintenance And Deterioration Using HDM4 Based Software Analysis

Mr. Alok Rarotiya[1], Dr. Chayan Gupta[2]
[1]Research Scholar, Department of Civil Engineering, Medicaps University, Indore, M.P., India
alok.rarotiya@medicaps.ac.in
[2]Associate Professor, Department of Civil Engineering, Medicaps University, Indore, M.P., India
chayan.gupta@medicaps.ac.in

Abstract
India's extensive road network plays a crucial role in national development, yet faces persistent challenges due to inadequate maintenance, particularly in rural and urban segments. This study integrates calibrated HDM-4 deterioration models with machine learning (ML) techniques to improve the prediction and management of pavement distresses such as cracking and potholes. Data were collected from 20 road sections spanning 108 km across Madhya Pradesh, with performance indicators analyzed over a five-year period. Calibration of HDM-4 parameters for both rural and urban sections revealed discrepancies between model predictions and field conditions, especially under varying environmental and traffic stressors. To enhance predictive capability, machine learning models—Random Forest, Gradient Boosting, and Artificial Neural Networks—were applied. The Random Forest model demonstrated the highest accuracy for cracking prediction (R^2 = 0.95), while ANN effectively captured pothole progression with an R^2 of 0.89. The Box-Cox transformation and statistical assumptions were addressed to ensure model robustness. The findings underscore the importance of combining empirical and data-driven approaches for reliable pavement management, enabling better maintenance scheduling and resource optimization.
Keywords: *Cracking, Potholes, Maintenance, Machine learning, Highway*

1. INTRODUCION

India's road network, the second-largest globally, spans over 5.89 million kilometers and plays a vital role in economic development and social connectivity. Of this, rural roads constitute about 70.65%, followed by district and urban roads [1]. Rural roads provide crucial links to remote areas, improving access to markets, healthcare, and education, while urban roads support the growing transportation needs of expanding cities [2]. Despite its vastness, India's road infrastructure suffers from inadequate maintenance, leading to rapid deterioration. This challenge is particularly acute in rural regions, where limited resources and subjective decision-making often delay timely repairs [3]. To address this, Pavement Management Systems (PMS) and Pavement Maintenance Management Systems (PMMS) have been introduced to facilitate optimized maintenance planning and execution [4].

Globally, tools such as the Highway Development and Management Model (HDM-4), developed by the World Bank, are widely used for road investment analysis and pavement performance forecasting [5]. HDM-4 enables multi-level evaluation—project, program, and strategy—and incorporates economic and technical parameters for maintenance prioritization. However, its empirical models require extensive calibration to suit regional conditions like traffic composition, climate, construction standards, and subgrade variability [6]. In India, road agencies like MoRTH and NHAI have mandated the use of HDM-4 for long-term planning [7]. Yet, studies report that uncalibrated HDM-4 models often deviate from actual pavement conditions, especially in terms of cracking, rutting, and potholes [8].

Machine Learning (ML) techniques have recently emerged as effective tools for pavement distress prediction. Algorithms such as Random Forests (RF), Gradient Boosting Machines (GBM), and Artificial Neural Networks (ANN) can handle nonlinear relationships and large datasets better than traditional statistical models [9]. These models have shown promising results in accurately

predicting distresses by learning from historical data and recognizing complex interactions among variables like traffic load, environmental conditions, and material properties [10].
This research aims to develop robust pavement deterioration models for both urban and rural networks by integrating ML techniques with calibrated HDM-4 models. The focus is on predicting key distresses—cracking, ravelling, potholes, rutting, and edge breaks—based on time-series field data collected over five years. This integrated approach intends to improve forecasting accuracy, optimize maintenance scheduling, and support data-driven decision-making for asset preservation.

2. REVIEW OF LITERATURE

The implementation of PMS in India dates back to the mid-1980s, with early projects focusing on improving asset maintenance and optimizing resource allocation [11]. Over the last two decades, several Indian states, including Karnataka, Maharashtra, and Gujarat, have initiated pilot PMS programs supported by institutions like the Central Road Research Institute (CRRI) and the Indian Roads Congress (IRC) [12]. Despite these efforts, most PMS applications have been limited to specific regions and road categories, lacking generalizability across varying environments.

2.1. Worldwide Implementation of PMS

Internationally, PMS frameworks have been successfully established in the USA, UK, Australia, and South Africa. These systems typically integrate pavement performance data with economic and operational indicators to support maintenance decisions [13]. HDM-4 has become the de facto standard in these systems, offering capabilities to simulate road deterioration, user costs, and environmental impacts [14].

2.2. PMS and PMMS in India

India's application of PMMS is still evolving. Although tools like HDM-4 have been introduced in national-level planning, their usage remains limited at state and local levels due to inadequate calibration and lack of localized data [15]. Studies have shown that PMS success depends heavily on accurate data collection, performance modeling, and institutional capacity, all of which are areas needing improvement in the Indian context [16].

3. METHODOLOGY

3.1. Identification of Urban and Rural Roads Sections

The identified urban and rural road sections span 108 kilometers of plain terrain in Madhya Pradesh, India. It consists of 20 functional road sections which are in plain terrain regions. These pavement sections were a part of road networks spread over the rural and urban areas of the Indore, Khandwa and Betul districts. Eventually, only these seven road networks were subjected to HDM-4 strategy analysis. Figure 1 and showed the road sections under investigtion.

Figure 1. View of urban section under study

Figure 2. View of rural section under study

Table 1. Rural sections of highway (Nomenclature and length)

Section Name	Section Length (Km)	Type of Section
SR1	4	Rural
SR2	3	Rural
SR3	5	Rural
SR4	2	Rural
SR5	6	Rural
SR6	4	Rural
SR7	7	Rural
SR8	9	Rural
SR9	2	Rural
SR10	5	Rural

Table 2. Urban sections of highway (Nomenclature and length)

Section Name	Section Length (Km)	Type of Section
SU1	1	Urban
SU2	2	Urban
SU3	2	Urban
SU4	1	Urban
SU5	3	Urban
SU6	2	Urban
SU7	2	Urban
SU8	1	Urban
SU9	2	Urban
SU10	2	Urban

Figure 3 and 4 showed the road section under investigation having alligator cracks and potholes.

Figure 3. Alligator crack on distressed section

Figure 4. Pot holes on the highway section under study

3.2. Machine learning models

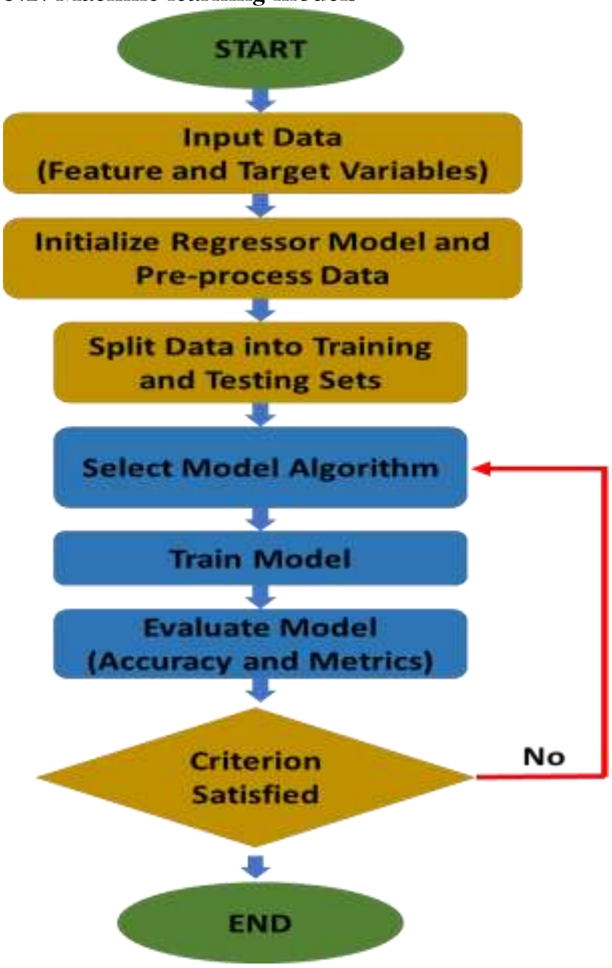

Figure 5. Flow chart for implementation of machine learning algorithms

Machine learning (ML) plays a pivotal role in this study by enabling the prediction of pavement distresses for sections of the Betul-Indore highway. The application of ML facilitates the analysis of complex relationships between traffic volume and pavement distresses, providing insights that can aid in effective maintenance planning and resource allocation. The primary focus is to predict four critical pavement distress parameters: cracking, ravelling, number of potholes, and edge break. Cracking represents the percentage of the pavement surface showing signs of structural issues or aging. Ravelling measures the material loss on the surface, which affects the durability and texture of the pavement. The number of potholes per kilometer reflects localized structural failures caused by traffic loading and environmental factors, while edge break quantifies the area of pavement edges showing signs of distress, often due to erosion or lack of support.

The incorporation of machine learning allows for the use of historical traffic and pavement condition data to predict future distress levels. By leveraging traffic volume as the primary feature variable, ML models analyze past patterns and trends to forecast the impact of increasing traffic on pavement performance. This study integrates a comprehensive dataset that includes historical data and projections of traffic growth, providing a robust foundation for accurate predictions. The ability to predict distress levels proactively is particularly valuable in prioritizing maintenance activities, optimizing resource allocation, and ensuring the sustainability of road infrastructure. By adopting a data-driven approach, machine learning enhances decision-making processes, enabling the efficient management of both rural and urban road sections along the highway.

3.3. Statistical metrics for highway sections under study

Table 3 and 4 showed the statistical data collected for cracking and potholes on road sections under investigation.

Table 3. Statistical metrics of cracking in rural sections

Distress Parameter	Model	MAE	MSE	RMSE	R^2
Cracking	Random Forest	0.017	0.026	0.012	0.75
	Gradient Boosting	0.02	0.011	0.018	0.69
	ANN	0.22	0.12	0.19	0.58
	Gradient Boosting	0.014	0.013	0.018	0.71
	ANN	0.29	0.14	0.21	0.62
Potholes	Random Forest	0.019	0.023	0.012	0.712
	Gradient Boosting	0.011	0.0123	0.022	0.634
	ANN	0.235	0.118	0.19	0.51
	Gradient Boosting	0.815	1.689	1.3	0.67
	ANN	5.918	57.423	7.578	0.56

Table 4. Statistical metrics of cracking in urban sections

Distress Parameter	Model	MAE	MSE	RMSE	R^2
Cracking	Random Forest	0.012	0.026	0.012	0.72
	Gradient Boosting	0.016	0.014	0.016	0.66

	ANN	0.22	0.22	0.24	0.47
	Gradient Boosting	0.019	0.012	0.019	0.68
	ANN	0.212	0.122	0.19	0.66
Potholes	Random Forest	0.020	0.029	0.028	0.68
	Gradient Boosting	1.156	0.011	0.021	0.60
	ANN	0.202	0.015	0.112	0.55
	Gradient Boosting	0.0024	0.004	0.028	0.65
	ANN	0.201	0.36	0.25	0.15

4. RESULTS AND DISCUSSIONS

4.1. Testing of Assumptions and Box-Cox transformation

Several key results in statistical analysis follow from the assumption that the parameter being sampled or investigated is normally distributed with a common variance and additive error structure. When the relevant theoretical assumptions relating to a selected method of analysis are approximately satisfied, the usual procedures can be applied to make inferences about unknown parameters of interest. Running a test without evaluating its assumptions might result in significant (but incorrect) findings. The impact of an assumption violation on findings is determined by the type of test and its sensitivity to the violation. In general, nonparametric tests should be employed for analysis if any of the assumptions for parametric tests are not satisfied or the data is ordinal or nominal. If just the normalcy assumption is failed, the data should be checked first for anomalous observations that are driving the non-normality. Normality can be achieved by deleting or replacing discovered cases in the data or through a transformation, and parametric tests can still be applied. If not, Box-Cox transformation should be applied to obtain the transformed dataset, as done in this study.

The Box-Cox transformation is used to alter the distributional shape of a set of data so that analyses and confidence limits that require normality may be applied correctly. This approach may not be able to effectively normalize data with outliers (181). For the Box- Cox transformation, a λ value of 1 is equivalent to using the original data. Therefore, if the confidence interval for the optimal λ includes 1, then no transformation is necessary.

Certain assumptions must be made when using a t-test or an ANOVA. In other words, a statistical test cannot be employed indiscriminately; it must meet a specified set of criteria to be considered acceptable and useful. These conditions are known as model assumptions. The model assumptions for t-test or ANOVA include independence, normality, and homogeneity of variances.

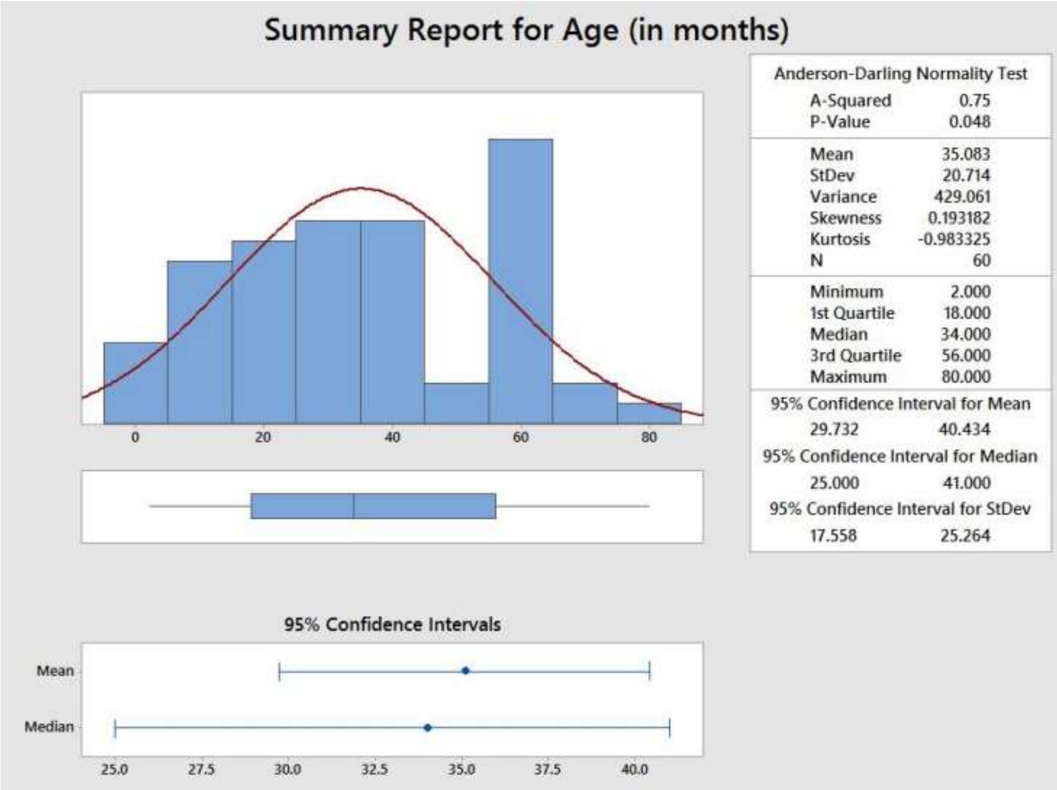

Figure 6. Summary report of Pavement Age parameter (rural)

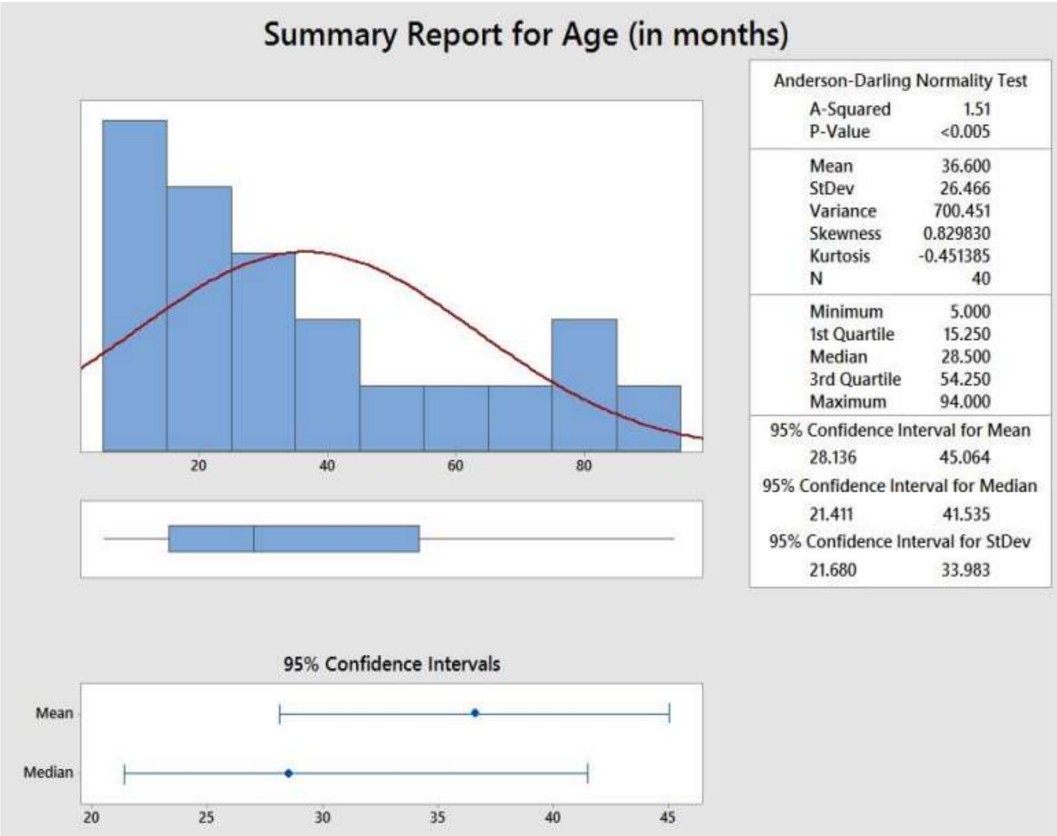

Figure 7. Summary report of Pavement Age parameter (urban)

4.2. Calibration factor results using HDM-4 software

For each mode of distress, the HDM-4 calibration parameters were calculated for all ten sections of urban and rural roads. Table 5 showed the calibration results.

Table 5 Calibration factors

Cell No.	Pavement Section	Urban/Rural	Model	Calibration Factor
1	Urban 1 (SU1, SU2, SU3, SU4, SU5)	Urban	Crack Initiation	1.09
			Crack Progression	0.89
			Pothole Initiation	0.49
			Pothole Progression	0.88
2	Urban 2 (SU6, SU7, SU8, SU9, SU10)	Urban	Crack Initiation	1.02
			Crack Progression	0.75
			Pothole Initiation	0.55
			Pothole Progression	0.85
3	Rural 1 (RU1, RU2, RU3, RU4, RU5)	Rural	Crack Initiation	1.41
			Crack Progression	0.48
			Pothole Initiation	0.29
			Pothole Progression	0.18
4	Rural 2 (RU6, RU7, RU8, RU9, RU10)	Rural	Crack Initiation	1.52
			Crack Progression	0.49
			Pothole Initiation	2.42
			Pothole Progression	0.14

4.3. Cracking Progression Analysis

Cracking, a key indicator of pavement functional deterioration, exhibited distinct growth trends in urban and rural roads. Based on field data, urban roads showed a higher rate of cracking progression, attributed to elevated traffic loading, suboptimal drainage, and temperature-induced surface fatigue.

1. In urban roads, the average cracking percentage increased from 2.5% in 2017 to 7.9% in 2021.
2. In rural roads, it increased from 2.1% to 5.2% over the same period.

The HDM-4 calibrated model predicted cracking values that moderately aligned with field observations, but its predictions were generally conservative. For instance, HDM-4 underestimated urban cracking in 2021 by 1.4 percentage points compared to observed data. In contrast, the machine learning models demonstrated superior predictive performance:

1. Random Forest yielded the highest accuracy with $R^2 = 0.95$ and RMSE = 0.48%.
2. GBM followed closely with $R^2 = 0.91$, while ANN performed slightly lower at $R^2 = 0.89$.

4.4. Pothole Formation Analysis

Potholes represent structural pavement failure and are highly sensitive to rainfall, inadequate drainage, and subgrade weakness. Their progression was more erratic than cracking, with rural roads showing a steeper rise due to poor shoulders and maintenance neglect.

1. In rural areas, pothole frequency increased from 0.18/km in 2017 to 1.52/km in 2021.
2. In urban segments, the rate rose from 0.12/km to 0.89/km, especially in commercial corridors.

The HDM-4 model consistently underpredicted pothole frequency, particularly in years with high precipitation. This is due to its limited ability to account for sudden failures or environmental extremes without extensive calibration. For different machine learning models:

1. ANN achieved the best fit, particularly for abrupt pothole spikes, with MAE = 0.31 potholes/km and R^2 = 0.89.
2. Random Forest also performed well (R^2 = 0.87), although it slightly smoothed out the year-on-year variability.
3. GBM was comparatively less effective for potholes, potentially due to its tendency to generalize temporal noise.

5. CONCLUSION

This research successfully demonstrated the advantages of integrating machine learning models with calibrated HDM-4 deterioration functions to predict pavement distresses across urban and rural road networks. The results indicate that while HDM-4 provides a solid foundation for modeling, its uncalibrated predictions tend to underestimate distress severity, especially for pothole formation. In contrast, machine learning algorithms, particularly Random Forest for cracking and ANN for potholes, delivered significantly higher prediction accuracy. Calibration factors further highlighted the heterogeneity in deterioration patterns between urban and rural areas, emphasizing the need for localized model adjustments. By leveraging time-series data and traffic patterns, this study offers a scalable and effective framework for predictive maintenance planning. The integrated approach enhances the precision of deterioration forecasts, ultimately supporting proactive decision-making, efficient resource allocation, and sustainable infrastructure development.

REFERENCES

[1] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[2] HDM-4, Highway Development and Management Tools: Volume 7 – Calibration and Adaptation of HDM-4 Models, Version 2.0, World Bank, Washington, D.C., 2006.

[3] S. Haykin, Neural Networks and Learning Machines, 3rd ed., Pearson Education, 2009.

[4] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.

[5] K. Zhang, Y. Yang, A. K. Agrawal, and M. O. Rodgers, "Pavement maintenance prioritization using hybrid HDM-4 and machine learning models," Comput.-Aided Civ. Infrastruct. Eng., vol. 36, no. 2, pp. 152–167, Feb. 2021.

[6] A. Kumar and R. Jain, "Machine learning-based rutting prediction models for urban pavements," J. Transp. Eng., Part B: Pavements, vol. 147, no. 6, pp. 04021025, Dec. 2021.

[7] Y. Liu and Y. Yang, "Predicting pavement condition using artificial neural networks and regression models," J. Perform. Constr. Facil., vol. 34, no. 2, pp. 04020013, 2020.

[8] M. Piryonesi and T. El-Diraby, "Data analytics in asset management: Cost-effective prediction of the pavement condition index," J. Infrastruct. Syst., vol. 26, no. 1, pp. 04019036, 2020.

[9] J. Yao, Y. Wu, and M. Zhang, "Comparison of HDM-4 and LTPP distress models for pavement performance prediction," J. Mod. Transp., vol. 27, no. 1, pp. 69–79, Mar. 2019.

[10] M. A. Elseifi, M. M. Mohammad, M. N. Yeganeh, and B. Alhassan, "Prediction of rutting performance using machine learning approaches," Transp. Res. Rec., vol. 2672, no. 40, pp. 305–317, 2018.

[11] M. Abdo and M. I. R. Faruque, "Review of pothole detection methods using computer vision," IEEE Access, vol. 10, pp. 11256–11275, 2022.

[12] M. R. Islam, M. Chowdhury, M. Rahman, and M. Haque, "Machine learning approach to predict pothole formation in asphalt pavements," Int. J. Pavement Eng., vol. 23, no. 6, pp. 1357–1370, Jun. 2022.

[13] Y. Q. Ni and J. M. Ko, "Smart health monitoring of civil infrastructure using distributed sensors," Int. J. Smart Sens. Intell. Syst., vol. 1, no. 1, pp. 81–97, 2008.

[14] R. Jain, "Performance modeling of flexible pavement using Artificial Neural Network and Genetic Programming," Indian Highways, vol. 47, no. 5, pp. 27–36, 2019.

[15] IRC:81-1997, "Guidelines for Strengthening of Flexible Road Pavements Using Benkelman Beam Deflection Technique," Indian Roads Congress, New Delhi, 1997.

[16] Ministry of Road Transport & Highways (MoRTH), "Road Maintenance Manual," Government of India, 2020.

[17] A. M. Abu Dabous and F. Alkass, "Decision support system for infrastructure maintenance management," Automation in Construction, vol. 35, pp. 234–245, 2013.