SEBHA UNIVERSITY JOURNAL OF PURE & APPLIED SCIENCES VOL.22 NO. 3 2023
DOI: 10.51984/JOPAS.V22I3.2733

جامعة سبها للعلوم البحتة والتطبيقية مجلة
Sebha University Journal of Pure & Applied Sciences

Journal homepage: www.sebhau.edu.ly/journal/index.php/jopas

Application of Machine Learning Techniques for Asphalt Pavement Performance Prediction

*Abdualmtalab Abdualaziz Ali[1], Abdalrhman Milad [2]

[1] Lecturer, Azzaytuna University, Civil Engineering Departement, Tarhuna, Libya
[2] Assistant Professor, University of Nizwa, Civil Engineering Departement, Oman

Keywords:
Machine learning
Multiple linear regression
Pavement condition index
Pavement management systems
Libya

ABSTRACT

Pavement management systems (PMS) and maintaining the quality of pavement roads are crucial to human and societal well-being. However, maintaining asphalt pavement quality is complex due to various factors, such as climate change, traffic volume, material properties, and pavement age. This research aims to develop pavement condition index (PCI) models in three U.S. states (California, Hawaii, and New Mexico) using Multiple Linear Regression (MLR) and compared with four additional machine learning (ML) algorithms which are: Random Forest (R.F.), Decision Tree (D.T.), Gradient Boosting (B.G.), and Adaboost were trained. The data obtained was employed for predicting the PCI model as a function of pavement distress and traffic volume. The inputs related to pavement distress and traffic volume variables' effects: pavement age, fatigue cracking, longitudinal cracking, transverse cracking, Cumulative Equivalent Single Axle Load (ESAL), Annual Average Daily Truck Traffic (AADTT), and Annual Average Daily Traffic (AADT). According to the statistical evaluation results, all the ML models exhibited excellent prediction capabilities, as evidenced by their high coefficient of determination (R^2) values of 96.8%,96.6%,97.1%, and 97.4% and low Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Square Error values of 1.888%, 1.874%,1.830, and 1.556%, and 2.529%,2.613%,2.391%, and 2.545% and 6.348%,6.828%,5.716%, and 5.081% and 9.98%, respectively. Furthermore, the results indicate that the ML models demonstrated superior prediction accuracy compared to the (MLR) models developed under the same data.

تتطبيق تقنيات التعلم الآلي لتوقع أداء رصف الأسفلت

*عبدالمطلب عبدالعزيز يخلف علي[1] و عبدالرحمن ميلاد[2]

[1] جامعة الزيتونة ، قسم الهندسة المدنية ، ترهونة ، ليبيا
[2] جامعة نزوى ، قسم الهندسة المدنية ، سلطنة عمان

الكلمات المفتاحية:
التعلم الآلي
الانحدار الخطي المتعدد
مؤشر حالة الرصف
أنظمة إدارة الرصف
ليبيا

الملخص

تعد أنظمة إدارة الرصف (PMS) والحفاظ على جودة طرق الرصف أمرًا بالغ الأهمية لرفاهية الإنسان والمجتمع. ومع ذلك، فإن الحفاظ على جودة رصف الأسفلت أمر معقد بسبب عوامل مختلفة، مثل تغير المناخ، وحجم حركة المرور، وخصائص المواد، وعمر الرصيف. يهدف هذا البحث إلى تطوير نماذج مؤشر حالة الرصف (PCI) في ثلاث ولايات في الولايات المتحدة (كاليفورنيا وهاواي ونيو مكسيكو) باستخدام الانحدار الخطي المتعدد (MLR) ومقارنتها بأربع خوارزميات إضافية للتعلم الآلي (ML) وهي: الغابة العشوائية، (RF)، وأشجار القرار (DT)، التعزيز الاشتقاقي (GB)، وتم استخدام البيانات التي تم الحصول عليها للتنبؤ بنموذج PCI كدالة لاستغاثة الرصيف وحجم حركة المرور. المدخلات المتعلقة بضيق الرصيف وتأثيرات متغيرات حجم المرور: عمر الرصف، التشقق الناتج عن التعب، والتشقق الطولي والعرضي، الحمولة التراكمية المكافئة أحادية المحور (ESAL)، والمتوسط السنوي لحركة الشاحنات اليومية (AADTT)، والمتوسط السنوي لحركة المرور اليومية (AADT). وفقًا لنتائج التقييم الإحصائي، أظهرت جميع نماذج ML قدرات تنبؤ ممتازة، وفقًا لنتائج التقييم الإحصائي، أظهرت جميع نماذج ML قدرات تنبؤ ممتازة، كما يتضح من قيم معامل التحديد (R^2) العالية بنسبة 96.8٪ و96.6٪ و97.1٪ و97.4٪ ومتوسط الخطأ المطلق المنخفض

*Corresponding author:
E-mail addresses: aayali@azu.edu.ly, (A. Milad) a.milad@unizwa.edu.om

Article History : Received 10 April 2023 - Received in revised form 31 August 2023 - Accepted 02 October 2023

(MAE) وخطأ متوسط الجذر التربيعي (RMSE) ومتوسط قيم الخطأ التربيعي 1.888٪ و1.874٪
و1.830 و1.556٪ و2.529٪ و2.613٪ و2.391٪ و 2.545٪ و 6.348٪ و 6.828٪ و 5.716٪ و
5.081٪ و 9.98٪ على التوالي. علاوة على ذلك، تشير النتائج إلى أن نماذج ML أظهرت دقة تنبؤ فائقة مقارنة
بنماذج الانحدار الخطي المتعدد (MLR) التي تم تطويرها وفقًا لنفس البيانات.

Introduction

The Pavement Management System (PMS) is a system used to control, assess, and monitor pavements designed to minimize maintenance costs, reduce environmental impacts, and provide long-term performance [1]. PMS utilizes a variety of data sources, such as traffic volume, weather, and pavement condition to track pavement performance over time [2-4]. The most widely used techniques in the PMS are the Present Serviceability Index (PSI), International Roughness Index (IRI), and Pavement Condition Index (PCI). The PSI is a measure of pavement condition and is defined as the difference between a pavement's present serviceability and its initial serviceability [5].

The PCI is a measure of a pavement's overall condition and is derived by combining the PSI and IRI. However, each pavement condition rating system is helpful for different purposes [6]. For instance, the PSI is useful for determining pavement maintenance needs, and the IRI is useful for assessing ride quality and safety [7,8].

In recent years, there has been a noteworthy growth in the use of ML methods for predicting pavement performance [9]. ML algorithms are well suited for this task, as they can effectively capture the complex interactions between pavement properties and performance indicators [10]. Ali et al. offered a technique for evaluating the pavement performance of 19 roads in St. John's, Newfoundland, Canada, where the PCI and IRI were the main indicators in characterizing the overall pavement performance of asphalt pavement [11]. Sagheer et al. developed a knowledge-based technique for pavement distress categorization using logic programming and the Prolog language to assess distress in flexible pavements [12]. Relatively few studies have been conducted in recent years to predict the PCI of flexible pavements using ML approaches [13,14].

In the literature, several studies have highlighted the importance of assessment programs that include PCI testing and distress surveys to determine the structural conditions of the pavement [15-17]. Nowadays, the assessment of pavement performance using PCI is a fundamental component of any PMS, and ML models have been used to predict pavement performance. In 2010, Bianchini and Bandini established a model to predict pavement performance using neuro-fuzzy. Hence, the outcomes and precision of the established model were superior to those of the linear regression one [18]. Similarly, Terzi (2006) demonstrated the PSI of flexible pavement by using ANNs. Thus, the regression value of the ANN-developed model was higher than the AASHTO model [19]. Moreover, in 2020, Yamany et al. offered individual performance models for each state based on performance data obtained from its own road network. On the other hand, the random parameter regression model was superior in some cases when considering individual states [20].

The motivation of this research is to integrate PMS with machine learning (ML), which could be used to predict the maintenance needs of a pavement over time. ML technologies integrated with artificial intelligence (A.I.) technologies, including Random Forest (R.F.), Decision Tree (D.T.), Gradient Boosting (B.G.), and Adaboost, can predict various situations in the PMS, which deals with the rehabilitation and maintenance of flexible pavement.

This study aims to employ ML techniques to predict the PCI to provide insights into the future performance of the pavement. Pavement management also helps to identify current deficiencies and distresses, such as cracking and rutting. In addition, base failure or subgrade instability were more serious structural issues.

PROPOSED METHODOLOGY

For evaluating and predicting the pavement condition index (PCI) of asphalt pavements, 61 major roads with various operational conditions were selected from three U.S. states., namely California, Hawaii, and New Mexico. Figure (1) presents the research methodology used, which consists of the following:

- Data collection.
- Data Preprocessing
- Model Development
  - Conventional Technique (using the Statistical Package for the Social Sciences (SPSS) software
  - Four Machine Learning (ML) algorithms: Random Forest (R.F.), Decision Tree (D.T.), Gradient Boosting (B.G.), and Adaboost.
- Comparison and Validation.

Data Description and Preprocessing

The data set was collected from Long Term Pavement Performance (LTPP). The data set consists of 61 rows and eight columns. The data collected from 61 roads involved three states in the U.S. (California, Hawaii, and New Mexico). To achieve the aims of this research, the data obtained was employed for predicting the PCI model as a function of pavement distress and traffic volume. The inputs related to pavement distress and traffic volume variables' effects: pavement age, fatigue cracking, longitudinal cracking, transverse cracking, Cumulative Equivalent Single Axle Load (ESAL), Annual Average Daily Truck Traffic (AADTT), and Annual Average Daily Traffic (AADT). These pavement problems negatively impact travel times, accidents, and the environment. Moreover, accidents tend to increase in areas with longer travel times due to drivers attempting to avoid poor road conditions. Developing a predictive model to determine pavement performance would be extremely useful for the competent authorities in selecting the most appropriate pavement maintenance system. Table (1) presents the collected data. The LTPP can access the data for free on its website. https://infopave.fhwa.dot.gov.

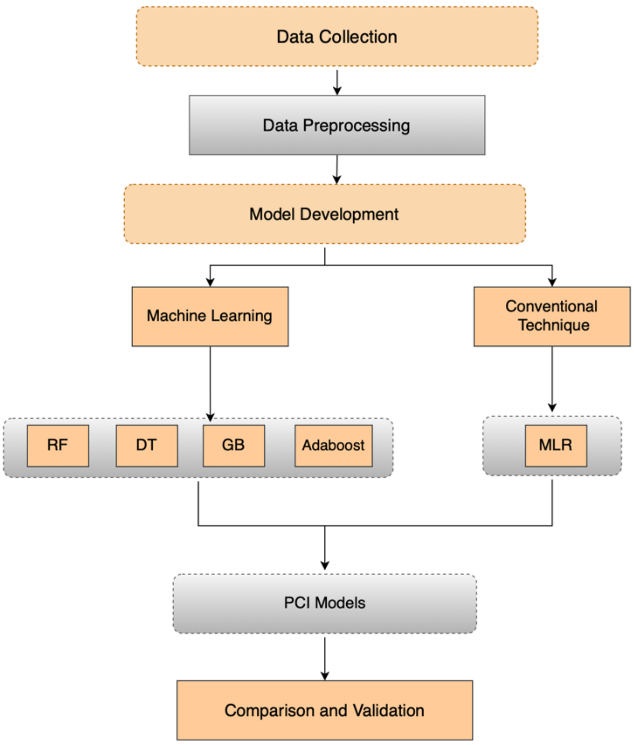

Fig 1. Methodological framework depicted.

Pavement Condition Index (PCI)

The ASTM D6433-18 method was employed to determine the PCI values for the 61 road sections using the data acquired from the LTPP dataset.

Table 1: Gathered pavement distress and traffic volume.

	Minimum	Maximum	Mean	Std. Deviation
Age	3	34	15.31	7.254
Fatigue cracking	0	304.8	6.595	39.2646
Longitudinal Cracking	0	305.6	66.354	96.5415
Transverse Cracking	0	140	13.36	25.525
ESAL	4851	1085824	221457.61	307000.387
AADTT	11	3538	766.07	994.947
AADT	4015	1294908	262620.95	348434.257
PCI	50	100	73.93682623	14.18856648

Performance evaluation metrics

A model validation process was employed to assess the predictive capabilities of the MLR and ML models. This process involved evaluating the models' ability to make accurate predictions. To verify the effectiveness of the models, 15% of the data was reserved for testing purposes. This data was used to predict PCI values, which were then compared to the actual PCI values. Various validation and performance measures are typically used to evaluate the validity and performance of statistical models. This study used the R^2, mean absolute error (MAE), root mean square error (RMSE), and mean square error (MSE), to assess the validity and compare the performance of the MLR and ML models. Table (2) presents these measures are calculated.

Table 2. Mathematical representation of the performance metrics

Measure Models	Formula	Variables Description
Determination Coefficient	$R^2 = 1 - \frac{\sum_i \quad (t_i - o_i)^2}{\sum_i \quad (o_i - o)^2}$	o_i=Actual value observation I,
Mean Absolute Error	MAE=$\frac{1}{n}\sum_i^n \quad \lvert t_i - o_i \rvert$	t_i = Predicted value of observation I,
Root Mean Squared Error	RMSE=$\sqrt{\frac{\sum_i \quad (t_i - o_i)^2}{n}}$	n = Number of observations.
Mean Square Error	MSE=$\frac{1}{n}\sum_i \quad (t_i - o_i)^2$	

Research Analysis Approaches

Developing Conventional Techniques Models

The MLR technique has proven reliable for creating precise and efficient models. Its purpose is to forecast the PCI value for flexible pavement based on seven types of pavement damage. To evaluate the model's effectiveness, the R^2 values, as well as the RMSE and MAE methods, were used. Table (1) shows the MLR technique for the PCI model. The results of the MLR prediction for the PCI models are shown in Figure (2). Equation (1) in Table (3) depicts the regression models and the connection between input variables (pavement distress and traffic volume) and PCI as follows:

$$PCI = 102.55 - 1.92 \times Age + 0.049 \times Fatigue\ Cracking - 0.004 \times \text{Longitudinal Cracking} + 0.091 \times Transverse\ Cracking + 1.04 \times 10^{-5} \times \text{ESAL} - 0.002 \times AADTT - 3.48 \times 10^{-6} \times AADT \quad (1)$$

Table 3: The PCI model summary.

Model	Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients β	t-stat
(Constant)	102.55	1.649	-	62.207
Age	-1.92	0.092	-0.982	-20.916
Fatigue Cracking	0.049	0.018	0.134	2.745
Longitudinal Cracking	-0.004	0.007	-0.029	-0.594
Transverse Cracking	0.091	0.025	0.163	3.627
ESAL	1.04E-05	0	0.226	1.708
AADTT	-0.002	0.003	-0.17	-0.932
AADT	-3.48E-06	0	-0.085	-0.549
R^2(%)	90.3			
MAE	3.351			
RMSE	4.381			
MSE	19.193			

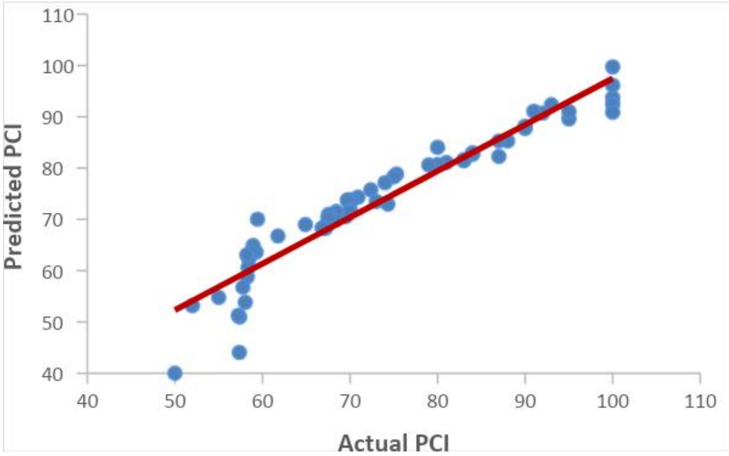

Fig 2. PCI Measured versus PCI Predicted plot.

Equation (1) shows the result of the regression analysis for PCI. The PCI had negatively correlated with age, longitudinal cracking, AADTT, and AADTT, while PCI had positively correlated with fatigue cracking, Transverse Cracking, and ESAL. The regression model developed using the mathematical method mentioned above was validated using statistical error measures, namely R^2, MAE, RMSE, and MSE. The results indicated that R^2 was strong, while the values of MAE, RMSE, and MSE were deemed acceptable.

Developing Machine Learning (ML) Models

Developing ML Techniques Models four techniques were used in this research (R.F., D.T., G.B., and Adaboost) to predict PCI based on four types of pavement distress and three traffic volume variables for asphalt pavement. Table (4) presents the modelling results for this study's three machine-learning techniques. Figures (3), (4), (5), and (6) present the (R.F.), (D.T.), (G.B.), and (Adaboost) prediction results for PCI models.

Table 4: Performance of PCI models ML techniques based on pavement distress and traffic volume.

ML Technique	Statistical Error Measures (%)			
	R^2	**MAE**	**RMSE**	**MSE**
RF	96.8	1.888	2.529	6.348
DT	96.6	1.874	2.613	6.828
GB	97.1	1.830	2.391	5.716
Adaboost	97.4	1.556	2.254	5.081

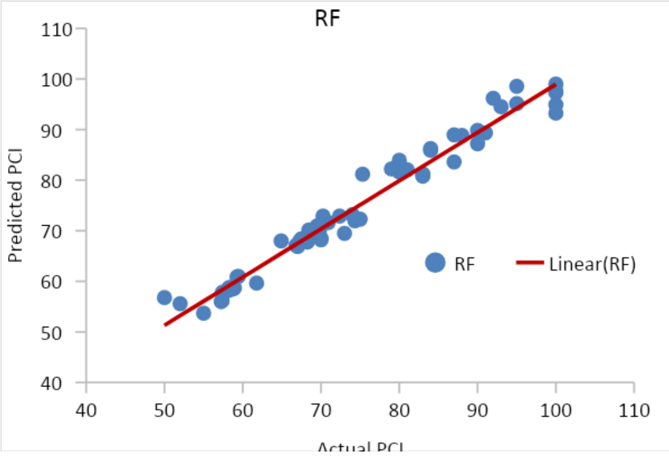

Fig 3. R.F. prediction results for PCI model.

Table (4) and Figures (3) to (6) present the modelling results for this research's four **ML** techniques as follows:

Random Forest (R.F.): This technique performing model scored R^2 =96.6%, **MAE** = 1.874%, **RMSE** =2.613%, and **MSE** =6.828
Decision Tree (D.T.): This technique performing model scored R^2= 96.8%, **MAE** = 1.888%, **RMSE** = 2.529%, and **MSE** =6.348.
Gradient Boosting (G.B.): This Technique performing model scored R^2 =97.1%, and **MAE** = 1.830%, **RMSE** =2.391%, and **MSE** =5.716.
Adaboost: This technique performing model scored R^2 =97.4%, **MAE** = 1.556%, **RMSE** =2.254%, and **MSE** =5.081.
Based on the results, the (**Adaboost**) model was more accurate than the (**R.F.**), (**D.T.**), and (**G.B.**) models, respectively.

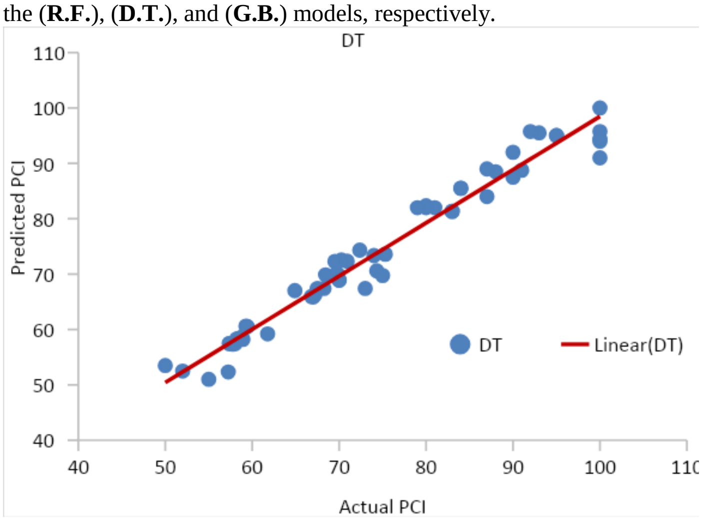

Fig 4. D.T. prediction results for PCI model.

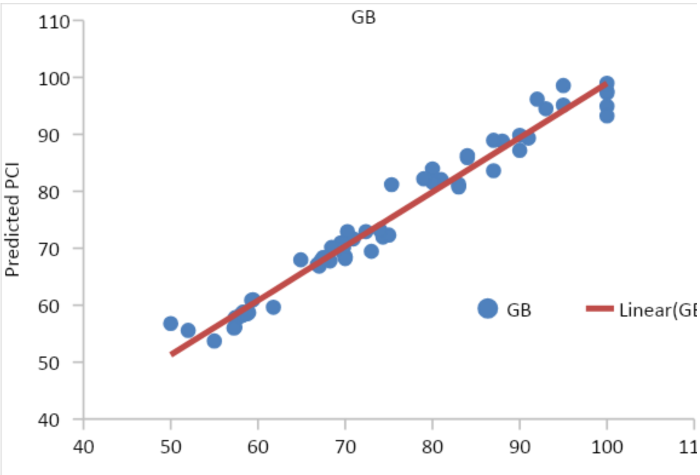

Fig 5. G.B. prediction results for PCI model.

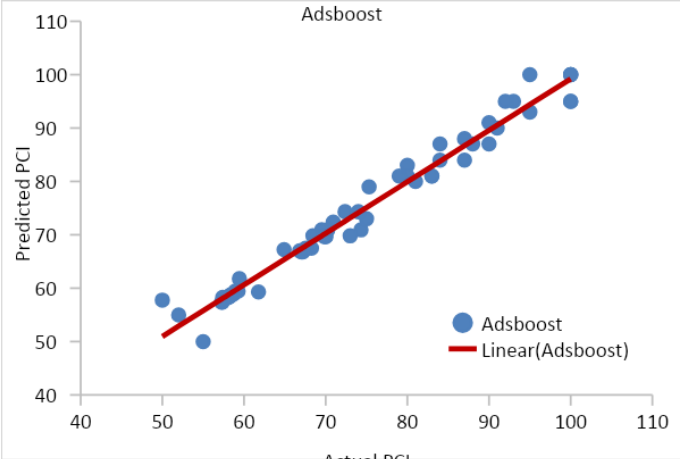

Fig 6. Adaboost prediction results for PCI model.

Comparison of Machine Learning with Conventional Techniques

Several metrics, including R^2, MAE, RMSE, and MSE were employed to evaluate the effectiveness of machine learning techniques compared to conventional methods. The analysis of the results obtained from both approaches revealed that all models exhibited high accuracy. A comparative study between traditional and ML techniques is presented in Table (5), while Figures (7) to (10) depict the contrast between machine learning and conventional methods.

Table 5: Comparison of the ML and MLR techniques.

Technique	Statistical Error Measures			
	R^2	**MAE**	**RMSE**	**MSE**
MLR	90.3	3.351	4.381	19.193
RF	96.8	1.888	2.529	6.348
DT	96.6	1.874	2.613	6.828
GB	97.1	1.830	2.391	5.716
Adaboost	97.4	1.556	2.254	5.081

Table (5) and Figures (7) to (10) compare traditional techniques and machine learning methods in terms of their capacity to predict **PCI** values using pavement distress and traffic volume data. The following observations can be drawn:

- Both traditional and machine learning approaches can accurately forecast **PCI** values based on pavement distress and traffic volume data.
- Results indicated that all (**ML**) techniques demonstrated a higher accuracy compared to the traditional approach (**MLR**), exhibiting only a slight error between actual and predicted values.
- Although the **ML** techniques did not provide equations, the **ML** approaches illustrated an excellent correlation among **PCI** and input variables (pavement distress and traffic volume).
- The results indicated that (**Adaboost**) model was more accurate than the (**R.F.**), (**D.T.**), and (**G.B.**) models, respectively.
- The results indicated that the R^2 of the (**Adaboost**) model improved by 7.29% compared to the **MLR** approach.
- The results indicated that the **MAE** value of the (Adaboost) model was reduced by 53.57% compared to the MLR method.
- The results showed that the **RMSE** and **MSE** values of the (**Adaboost**) model were reduced by 48.55% and 73.53% compared **MLR** method, respectively.
- The results indicated that the R^2 of the (Adaboost) model improved by 0.62%,0.82%, and 0.31% compared to the **R.F., D.T.**, and **G.B.** approach, respectively.
- The results showed that the **MAE** value of the (**Adaboost**) model was reduced by 17.58%, 16.97%, and 14.97%, compared to the **R.F., D.T.**, and **G.B.** approach, respectively.
- The results indicated that the **RMSE** value of the (Adaboost) approach was reduced by 10.87%, 13.74%, and 5.73%,

compared to the **R.F.**, **D.T.**, and **G.B.** approach, respectively.

- The results indicated that the **MSE** value of the (**Adaboost**) approach was reduced by 19.96%, 25.59%, and 11.11%, compared to the **R.F.**, **D.T.**, and **G.B.** approach, respectively.

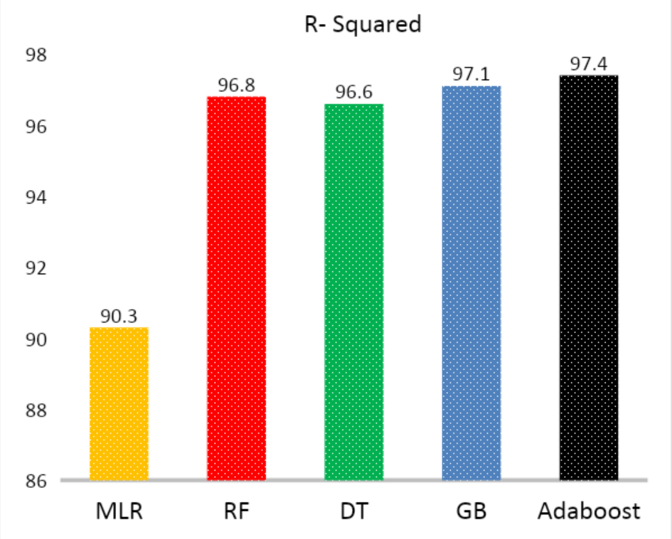

Fig 7. Comparison among the ML and MLR techniques (R-Squared).

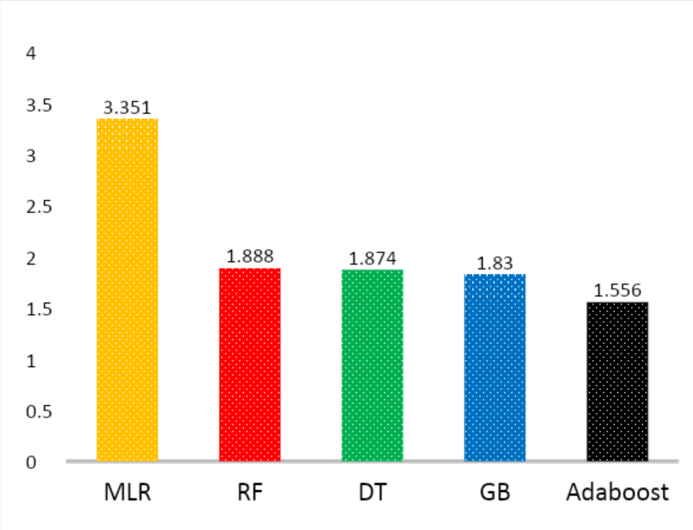

Fig 8. Comparison among the ML and MLR techniques (MAE).

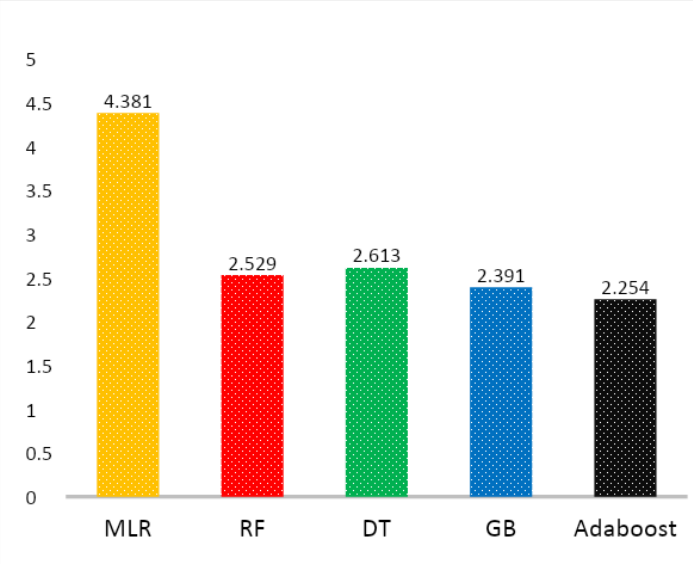

Fig 9. Comparison among the ML and MLR techniques (RMSE).

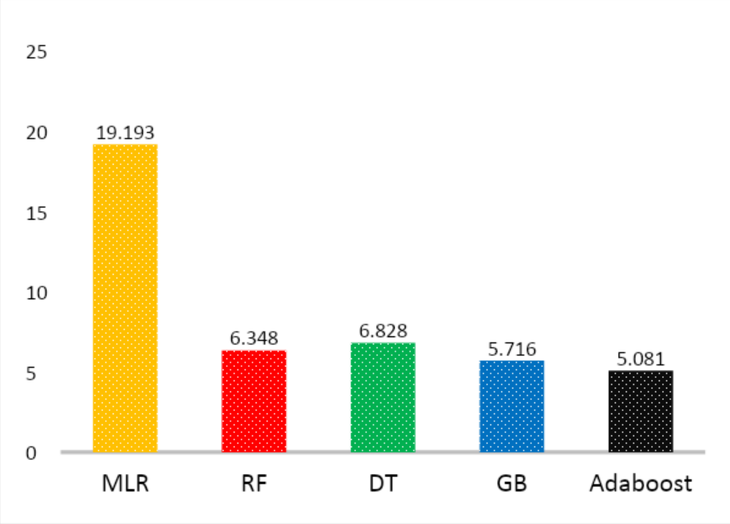

Fig 10. Comparison among the ML and MLR techniques (MSE).

Conclusions

In this study, the prediction of the PCI is based on pavement distress and traffic volume for three U.S. states. California, Hawaii, and New Mexico were developed using the conventional Technique (MLR) and four ML techniques R.F., D.T., G.B., and Adaboost. Based on the light of the study's findings, the following conclusions can be drawn:

- This study focused on predicting PCI value by analysing pavement distress and traffic volume using sixty-one road sections for flexible pavements selected in three U.S. states from the LTPP dataset.
- The assessment of ML techniques indicated that Adaboost was more precise in predicting PCI values compared to R.F., D.T., and G.B. techniques, respectively.
- Based on the study's findings, both ML and MLR techniques were viable models for predicting PCI values based on pavement distress and traffic volume values.
- The study demonstrated that both techniques could accurately predict PCI values, decrease the need for visual examination of PCI values, and save budgets and time.

Data availability statement

The submitted article contains all the models and data used or generated during the study.

Declaration of Competing Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Data availability statement

The submitted article contains all the models and data used or generated during the study.

Declaration of Competing Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1]- Miah, M. T., Oh, E., Chai, G., & Bell, P. (2020). An overview of the airport pavement management systems (APMS). International Journal of Pavement Research and Technology, 13, 581-590.

[2]- Shon, H., Cho, C. S., Byon, Y. J., & Lee, J. (2022). Autonomous condition monitoring-based pavement management system. Automation in Construction, 138, 104222.

[3]- Meegoda, J. N., & Gao, S. (2014). Roughness progression model for asphalt pavements using long-term pavement performance data. Journal of Transportation Engineering, 140(8), 04014037.

[4]- Shtayat, A., Moridpour, S., Best, B., & Rumi, S. (2022). An overview of pavement degradation prediction models. Journal of Advanced Transportation, 2022.

[5]- Wolters, A., Zimmerman, K., Schattler, K., & Rietgraf, A. (2011). Implementing pavement management systems for local agencies.

[6]- Shah, Y. U., Jain, S. S., Tiwari, D., & Jain, M. K. (2013). Development of overall pavement condition index for urban road network. Procedia-Social and Behavioral Sciences, 104, 332-341.

[7]- Aleadelat, W., & Ksaibati, K. (2017). Estimation of pavement serviceability index through android-based smartphone application for local roads. Transportation Research Record, 2639(1), 129-135.

[8]- Rahman, M. M., Uddin, M. M., & Gassman, S. L. (2017). Pavement performance evaluation models for South Carolina. KSCE Journal of Civil Engineering, 21(7), 2695.

[9]- Justo-Silva, R., Ferreira, A., & Flintsch, G. (2021). Review on machine learning techniques for developing pavement performance prediction models. Sustainability, 13(9), 5248.

[10]- Cano-Ortiz, S., Pascual-Munoz, P., & Castro-Fresno, D. (2022). Machine learning algorithms for monitoring pavement performance. Automation in Construction, 139, 104309.

[11]- Ali, S., O'Sullivan, J., El-Diraby, T.E. and Valadan Zoej, M.J., 2017. Evaluation of pavement performance using mechanistic-empirical models and the international roughness index. Transportation Research Record, 2620(1), pp.41-49.

[12]- Sagheer, M., El-Diraby, T.E. and O'Sullivan, J., 2005. A knowledge-based system for flexible pavement distress categorization. Journal of Computing in Civil Engineering, 19(2), pp.144-156.

[13]- O'Sullivan, J., El-Diraby, T.E. and Valadan Zoej, M.J., 2017. Evaluation of pavement performance using mechanistic-empirical models and the international roughness index. Transportation

[14]- Mousa, M.M., Abdel-Wahab, A.M., Mostafa, R.M. and Hamza, M.A., 2018. Performance prediction of reclaimed asphalt pavement and virgin aggregate blends using multi-layer elastic analysis. Transportation Research Part B: Methodological, 111, pp.1-20.

[15]- Shahin, M. Y. (2005). Pavement management for airports, roads, and parking lots (Vol. 501). New York: Springer.

[16]- Zhang, Z., Claros, G., Manuel, L., & Damnjanovic, I. (2003). Development of structural condition index to support pavement maintenance and rehabilitation decisions at network level. Transportation research record, 1827(1), 10-17.

[17]- Park, B., Cho, S., Rahbar-Rastegar, R., Nantung, T. E., & Haddock, J. E. (2022). Prediction of critical responses in full-depth asphalt pavements using the falling weight deflectometer deflection basin parameters. Construction and Building Materials, 318, 126019.

[18]- Bianchini, A., & Bandini, P. (2010). Prediction of pavement performance through neuro-fuzzy reasoning. Computer-Aided Civil and Infrastructure Engineering, 25(1), 39-54.

[19]- Terzi, S. (2006). Modeling the pavement present serviceability index of flexible highway pavements using data mining. Journal of Applied Sciences, 6(1), 193-197.

[20]- Yamany, M. S., Saeed, T. U., Volovski, M., & Ahmed, A. (2020). Characterizing the performance of interstate flexible pavements using artificial neural networks and random parameters regression. Journal of Infrastructure Systems, 26(2), 04020010.