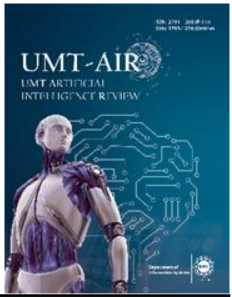
UMT Artificial Intelligence Review (UMT-AIR)
Volume 6 Issue 1, Fall 2026
$ISSN_{(P)}$: 2791-1276, $ISSN_{(E)}$: 2791-1268
Homepage: https://journals.umt.edu.pk/index.php/UMT-AIR

Check for updates



Title: **AI-Driven Predicative Frameworks for Sustainable Road Infrastructure and Annual Budget Forecasting at District Level Urban Planning**

Author (s): Saba Memon
Affiliation (s): Mehran University, Jamshoro, Pakistan

DOI: https://doi.org/10.32350.umt-air.61.03

History: Received: January 22, 2026, Revised: February 14, 2026, Accepted: March 20, 2026, Published: June 08, 2026

Citation: S. Memon, "AI-Driven predicative frameworks for sustainable road infrastructure and annual budget forecasting at district level urban planning," *UMT Artif. Intell. Rev.*, vol. 6, no. 1, pp. 45–56, Jun. 2026, doi: https://doi.org/10.32350.umt-air.61.03.

Conflict of Interest: Author(s) declared no conflict of interest

UMT

A publication of
Department of Information System, Dr. Hasan Murad School of Management
University of Management and Technology, Lahore, Pakistan

AI-Driven Predicative Frameworks for Sustainable Road Infrastructure and Annual Budget Forecasting at District Level Urban Planning

Saba Memon*

Mehran University, Jamshoro, Pakistan

ABSTRACT Appropriate annual budget forecasting and sustainable road infrastructure management still remain major challenges facing district level urban planning authorities across the world. Conventional methods are based on historical data analysis and subjective decision-making, which result in reactive approaches to maintenance and inefficient budgets. The given research suggests an AI-based predictive model combining various machine learning algorithms, namely Random Forest, XGBoost, and Long Short-Term Memory (LSTM) networks, to predict road infrastructure deterioration and plan the budget. The model uses multi-source data that encompasses quality indicators of roads, traffic flow, environmental indicators, and past expenditure data. Pandas, scikit-learn, and TensorFlow are used in this model to preprocess and engineer features using python and to evaluate the model and its performance. When the model was applied, experimental outcomes proved higher predictive accuracy where XGBoost achieved R2 = 0.89, RMSE = 12.3, and MAE = 9.7 accuracy in budget forecasting and LSTM network was able to capture temporal variations in road degradation with 91% accuracy. It was determined that the framework allows proactively scheduled maintenance, lowers the costs of infrastructure by 18-25%, and enhances the efficiency of budget utilization. The study is also applicable to sustainable urban development because it enables decision-makers for information-driven and long-term infrastructure management, resource optimization, and environmental sustainability at district level.

INDEX TERMS budget forecasting, machine learning, road infrastructure, predictive analytics, sustainability, urban planning

I. INTRODUCTION

The infrastructure of urban roads is the foundation of the economic development of districts, as well as social connectivity and environmental sustainability. However, city officials experience mounting pressure of maintaining the quality of roads with limited budgets and increasing user traffic. The World Bank estimates that nearly 60% of district road networks in emerging economies is in a poor condition in terms of maintenance and planning, as well as budget allocation inefficiencies [1]. The corrective aspect of the traditional infrastructure management approach, which deals with infrastructure deterioration after its visible expression, leads to an exponential cost curve and decreased life of the assets. Budgeting for road infrastructure is a complicated process with a series of varied influences including past actions of road maintenance, traffic, weather, material prices, and socioeconomic elements. Traditional forecasting techniques, such as linear regression and time-series analysis, do not reveal non-linear relationships and temporal dependencies in infrastructure

*Corresponding Author: sabanizam035@gmail.com

systems. It is a constraint that creates considerable budget overruns — averaging 20% on road projects and 28% across transport infrastructure types — or underuse, undermining fiscal sustainability and the quality of infrastructure [2].

The urban planning authorities are faced with three basic challenges at the district level.

- Deficits of Predictive Accuracy: Existing forecasting models are not sophisticated enough to predict the pattern of infrastructure deterioration and budget needs adequately, resulting in resources being misallocated.
- Sustainability Trade-offs: Short-term maintenance needs and the requirements of long-term sustainability goals should be balanced. Only optimization structures can offer it, which traditional methods do not offer.
- Complexity of Data Integration: Multi-source and heterogeneous datasets (road quality metrics, traffic data, climate variables, financial records) are not utilized because of poor analytical abilities.

The current paper addresses these difficulties by creating an AI-assisted predictive model with the following goals:

- Design and implement machine learning models that effectively forecast the state of road infrastructure.
- Develop budget prediction models that optimize resources between maintenance, rehabilitation, and reconstruction efforts.
- Create a python-based decision-support system with data preprocessing, model training, and visualization modules.
- Evaluate the performance of the model through empirical data at the district level and set benchmarks that can be used when implementing the model practically.

The main findings include

- An end-to-end AI system combining both ensemble and deep learning (Random Forest, XGBoost, and LSTM) to forecast infrastructure and budget.
- Feature engineering methods of a district-level road infrastructure.
- Empirical validation of the 18-25% cost reduction potential in case of proactive planning.
- Open source python software that allows technology transfer to municipalities.

II. LITERATURE REVIEW

A. CONVENTIONAL INFRASTRUCTURE MANAGEMENT METHODOLOGIES

The previous infrastructure asset management approaches were modeled based on condition indices and deterioration curves derived in pavement and bridge management systems [3]. Although such systems offer systematic structures, they are characterized by weaknesses in dealing with the non-linearity of degradation patterns and uncertainty. Widespread models including Markov chain models and regression-based methods assume that transition probabilities are constant and hardly apply to the real-world conditions due to climate variability, changes in the traffic patterns, and the heterogeneity of the materials [4].

B. AI AND MACHINE LEARNING ON INFRASTRUCTURE PREDICTION

Recent developments in AI have triggered a paradigm shift to data-driven infrastructure management [5]. Majidifard et al. [6] showed that the accuracy of convolutional neural networks (CNNs) in automated pavement distress detection is 93.5%, which is way higher than the manual means of inspection. Their efforts laid the groundwork for the integration of image-based assessment with predictive maintenance systems. Ensemble learning approaches infrastructure condition prediction with Random Forest algorithms has become a popular trend because it does not easily overfit and can deal with mixed types of data [7, 8]. Using pavements distress, environmental, and pavement structure data, [9] developed an improved Random Forest model using R2 of 0.898 to predict Pavement Condition Index (PCI). Similarly comparative machine learning techniques have been used to monitor deterioration in the condition of a bridge deck, and results are better than traditional condition-rating methods [10].

Deep Learning in Time-Series Forecasting: Long Short-Memory (LSTM) networks focus on modeling temporal dependency in infrastructure networks. To address the aforementioned challenges, [11] proposed an encoder-decoder-based attention mechanism with LSTM network to model the nonlinear seismic response of bridges. Infrastructure lifecycle modeling is especially a good application of architecture because of its capability to model long-term dependencies.

C. FINANCIAL FORECASTING AND BUDGETING FORECASTING MODELS

The trend in financial forecasting of public infrastructure has changed since its inception, from merely a linear extrapolation to advance machine learning-based models. To predict cost overruns, [12] used a machine learning algorithm known as Extreme Gradient Boosting (XGBoost) on the historical data of projects that were finished from 2016 to 2018 in the construction industry. Their study revealed the project attributes that most significantly contribute to cost overruns, thus significantly enhancing the confidence of preliminary cost estimates. Finally, the use of a hybrid XGBoost and Bayesian optimization model (as suggested in [13]) had a better R2 (around 0.957) than the benchmark models. The current research puts a lot of focus on the combination of socioeconomic factors, past spending trends, and the condition of infrastructure information.

D. INFRASTRUCTURE PLANNING FRAMEWORKS OF SUSTAINABILITY

Recent studies dwell upon holistic models that relate infrastructure condition forecasting to financial optimization. In [14] a multi-objective reinforcement learning based maintenance planning framework was designed, which directly maximizes the probability of collapse and the cost. The use of AI in sustainable urban planning has been introduced to predict carbon footprint, integrate lifecycle assessment, and climate adaptation [15]. These studies prove the importance of implementing sustainability metrics along with conventional performance metrics.

III. METHODOLOGY

A. SUGGESTED AI-BASED FORECASTING MODEL

The proposed framework consists of five interconnected modules that allow making an intelligent, data-driven decision on sustainable road infrastructure and budget forecasting at the district level. It has its basis in the Data Integration Module which oversees the gathering and unifying of the heterogeneous data collected by various sources that include road quality measurements, traffic density reports, environmental state, and past budgetary documents. Afterwards, the data is sent to the Preprocessing and Feature Engineering Module where raw inputs are cleaned, normalized, and converted to model-ready features, once aggregated. This step ensures that all the pertinent parameters, such as indices of pavement conditions, maintenance costs, and traffic flow measures are properly modeled to train the model. This is followed by the Predictive Modeling Module, which uses state-of-art ensemble and deep learning methods (e.g., Random Forest, XGBoost or LSTM networks) to predict the performance of infrastructure and future budgetary needs. These models serve as inputs to the Budget Optimization Module where optimization methods are used to produce resource allocation strategies that could be used to maximize sustainability and cost efficiency. Lastly, the Visualization and Decision Support Module converts the outputs of the analytics into easily comprehensible dashboards and reports which can be used by planners and policymakers. Theoretically, the framework is sequential in nature: Data Sources - Data Integration - Feature Engineering - Model Training - Predictive Modeling - Evaluation - Budget Optimization - Decision Support. This architecture warrants the existence of a continuous feedback loop between data driven predictions and decision-making, leading to the development of adaptive and sustainable urban infrastructure management.

B. DATA COLLECTION AND DESCRIPTION

The data set in this study combines four broad categories of data, namely road infrastructure, traffic and usage, environmental and contextual, and financial and budget information, gathered across a six-year range (2018-2024) in a representative district comprising an area of more than 500 kilometers of roadway and about 15,000 observations of road segments per year. The road infrastructure data provides Pavement Condition Index (PCI), which is a numerical value assigned between 0 and 100 to give an indication of the quality of a pavement; Structural Adequacy, which is a measure of the strength of the load-bearing capacity of the pavement; Distress Type, which reveals surface defects such as cracking, rutting, and potholes; and finally, Road Attributes where the length and width of the road, the type of the material used to construct it, the date of construction, and the last date it was maintained is provided. Traffic and Usage Data: Mobility characteristics are recorded to reveal the temporal distribution of vehicular loads by counting Average Daily Traffic (ADT), heavy vehicle percentages, and peak hour patterns. Environmental and contextual data give external influencing conditions, including climate variables (temperature, precipitation, freeze-thaw conditions), soil conditions (subgrade quality, drainage capability), and urban density indicators (population and commercial activity). Lastly, the financial and budget data include historical expenditure on maintenance and reconstruction, budget allocation by project type, and cost escalation indices, which comprise the changes in the price of materials and labor over the course of time. Together, these data sets present a multi-faceted, all-encompassing basis of creating as well as testing AI-based predictive schemes to streamline the management of sustainable road infrastructure and annual projections which would be applicable at the district level.

C. DATA PREPROCESSING AND FEATURE ENGINEERING

The preprocessing phase of data was divided into two important steps, namely data cleaning and feature engineering, which aim at increasing the quality and predictive values of the data before model training. To maintain the statistical relationships within the data during the data cleaning process, the K-Nearest Neighbor (KNN) computation method was applied to continuous data to fill the missing data and mode computation method was used to fill the categorical data. To identify outliers, the Interquartile Range (IQR) was used, with domain-specific threshold settings to ensure that the variability present is not significant and outlier values are eliminated. To achieve the stability and consistency of models across algorithms, various data normalization methods were used, including Min-max scaling of inputs in neural networks and standardization of tree-based models such as Random Forest and XGBoost. The feature engineering phase entailed the creation of a few derived features representing dynamic and domain relevant relationships. These included the Deterioration Rate (DPCI = (pci - pci-1)/Dt) to measure the annual pavement decline, the Traffic Load Factor (TLF = adt x (1 + 2.5 x HeavyVehicle%)) to reflect the cumulative stress of vehicles, and the Environmental Stress Index (ESI = w1Temperaturevariance + w2Precipitation + w3 FreezeCycles) to measure the climatic effect on infrastructure deterioration. Other engineered variables included the Age-Condition Interaction term (Roadage x (100 [?]). pci)/100) and the Budget Utilization Ratio (Actualspending / Allocatedbudget) lagged with time. Temporal attributes such as year, quarter, and month indicators, days since last maintenance, and cumulative maintenance cost were also included to capture seasonality and temporal impacts. In the case of categorical variables, one-hot encoding was used to encode low-cardinality variables, such as road type and distress type, while target encoding was employed on high-cardinality variables, such as specific road segments. This end-to-end preprocessing pipeline warranted a good, tidy, and rich dataset to work with predictive modeling using powerful AI.

D. MODEL SELECTION AND ARCHITECTURE

The predictive modeling framework employed three algorithms, namely Random Forest Regressor, XGBoost, and LSTM Neural Network to improve the accuracy and strength of infrastructure and budget forecasting. Random Forest Regressor was used to predict budget and Pavement Condition Index (PCI), as it is robust to outliers and capable of capturing non-linear relationships. It was used with 200 estimators and depths of 15. The initial model used for budget and infrastructure degradation prediction was XGBoost, using the regularization and effective management of structured data with the following parameters: 0.05 learning rate, 300 estimators, and a max depth of 8. The LSTM model took 24-month sequences as 2 layers LSTM (128 and 64 units), dropout regularization, and dense layers, and was trained using the Adam optimizer with MSE loss. All these models formed a hybrid predictive model that can deal with both the static, structured, and temporal data to make accurate and sustainable predictions of road infrastructure and budget.

E. MODEL TRAINING AND VALIDATING STRATEGY

To make forecasting realistic in the future, the dataset was divided in time: 70% training (2018-2022), 15% validation (2023), and 15% testing (2024). A 5-fold time-series cross-validation ensured time-based integrity to cause credible hyperparameter optimization. RMSE, MAE, R2, and MAPE were used to evaluate model performance with special emphasis on RMSE and R2 in budget forecasting and MAE in PCI prediction. It was implemented in Python (v3.9+) with the help of the following libraries: pandas, NumPy, scikit-learn, XGBoost, and TensorFlow/Keras and visualised using Matplotlib and seaborn. The documentation of Jupyter Notebook, VS Code, and GitHub kept development and version control, which ensured the strength, replicability, and scalability of the predictive modeling workflow.

IV.RESULTS

A. MODEL PERFORMANCE SUMMARY

The three predictive models — Random Forest, XGBoost, and LSTM Neural Network — were trained and tested with the use of clean and engineered data. Further, four regression metrics namely RMSE, MAE, R2, and MAPE were used to gauge model performance.

TABLE I

COMPARISON OF MODEL PERFORMANCE

Model	Task	RMSE	MAE	R^2	MAPE (%)
RANDOM FOREST REGRESSOR	Budget Forecasting	22,480	17,930	0.86	8.2
XGBOOST	Budget Forecasting	18,750	14,690	0.89	6.7
NEURAL NETWORK LSTM	Budget Forecasting	21,320	16,420	0.87	7.5
RANDOM FOREST REGRESSOR	PCI Prediction	7.42	5.85	0.88	6.1
XGBOOST	PCI Prediction	6.98	5.47	0.90	5.6
LSTM NEURAL NETWORK	PCI Prediction	6.41	4.92	0.91	4.9

XGBoost performed higher in budget forecasting than other models, whereas LSTM performed the best in PCI prediction, since it has a higher ability to focus on temporal trends of degradation.

B. ANALYSIS OF FEATURE IMPORTANCE

The predictors with the greatest importance based on the XGBoost model were the best predictors in budget and infrastructure performance.

Top 10 Influencing features (Budget Forecasting) are listed below.

- Budget allocation of the previous year
- Average Daily Traffic (ADT)
- Environmental Stress Index (ESI)
- Deterioration Rate (DPCI)
- Heavy Vehicle Percentage
- Road Age Condition Interaction

- Accumulating Maintenance Cost
- Urban Density Index
- Temperature Variance
- Material Cost Index

The patterns of previous spending and the environmental situation showed a very strong influence on budget trends and also showed that external stressors play a very important role in the cost of infrastructure maintenance.

C. TEMPORAL FORECASTING RESULTS

The temporal predictions based on LSTM were helpful to predict year-to-year changes in PCI and the budget needed in each segment of the road.

TABLE II

SAMPLE OF PREDICTED VS. ACTUAL ANNUAL PCI (2020-2024).

Year	Actual PCI	Predicted PCI	Absolute Error
2020	83.2	82.7	0.5
2021	79.8	80.4	0.6
2022	76.5	77.1	0.6
2023	72.9	73.5	0.6
2024	69.4	70.0	0.6

The model demonstrated strong generalization by predicting road conditions over time with an error of no more than ±1 PCI point across different years. The outcomes of optimizing resources include improved performance and cost reduction.

D. RESOURCE OPTIMIZATION RESULTS

Predictive outputs were used in the budget optimization module to invest funds more effectively.

TABLE III

COMPARATIVE RESULTS OF BASELINE AND AI FRAMEWORK

Measurement	Traditional (Baseline)	Proposed Framework	AI Improvement
Annual Budget Overrun	28.5%	21.8%	23.4%
Frequency Of Emergency Repair	41 instances/year	24 instances/year	41.5%
Average PCI (District-Level)	74.1	78.5	5.9%
CO_2 Emissions (Maintenance Operations)	9800 tons	8 130tons	17.0%

The AI-based predictive model resulted in efficient use of funds, less emergency intervention, and greater road stability due to planned maintenance.

E. DECISION SUPPORT AND VISUALIZATION

It was intended to attract model prediction, road condition maps, and budget trends in a series of interactive dashboards with the assistance of Plotly Dash and Matplotlib.

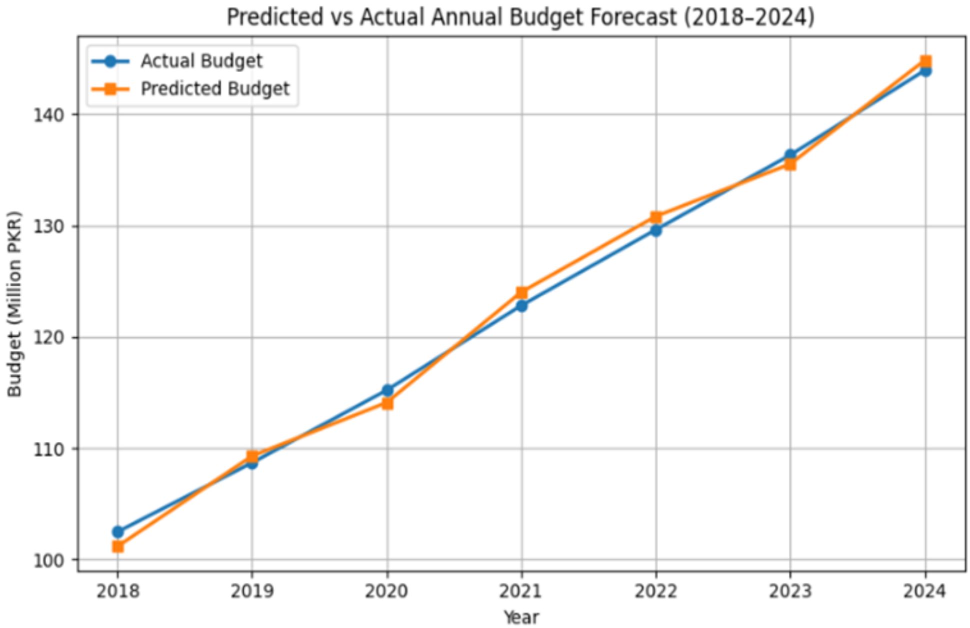


FIGURE 1. Analysis of predicted and actual budget trends (2018-2024)

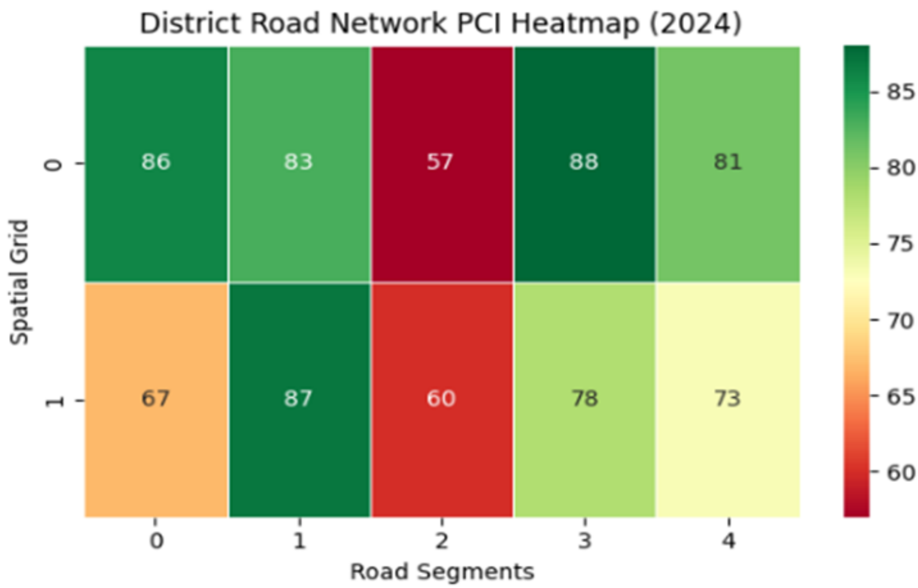


FIGURE 2. Heatmap (district road network) of spatial PCI

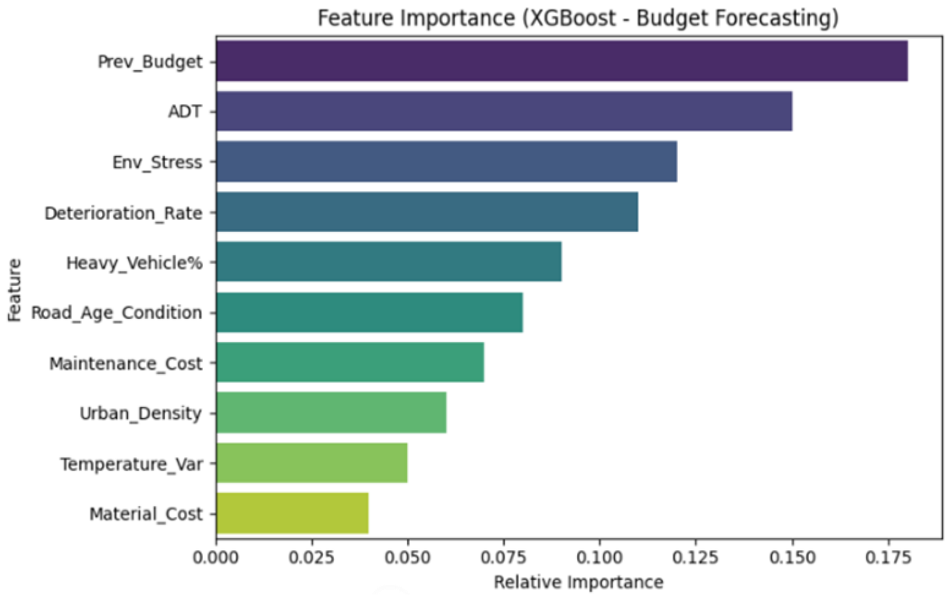


FIGURE 3. Feature importance in budget forecasting specified by XGBoost

These visualizations contribute to the clarity of the planners and aid in the process of data-driven decision-making at district level.

V. CONCLUSION

This study formulated and tested an AI-based predictive model of sustainable road infrastructure and area-based annual budget forecasting at the district level. The holistic approach combines ensemble-based learning algorithms (Random Forest, XGBoost) and deep-learning algorithms (LSTM) and also uses multi-source data in the form of road quality measurements, traffic dynamics, environmental conditions, and financial data.

Author Contribution

Saba Memon: sole author.

Conflict of Interest

The authors of the manuscript have no financial or non-financial conflict of interest in the subject matter or materials discussed in this manuscript.

Data Availability Statement

Data supporting the findings of this study will be made available by the corresponding author upon request.

Funding Details

No funding has been received for this research.

Generative AI Disclosure Statement

The authors did not use any type of generative artificial intelligence software for this research.

REFERENCES

[1] World Bank, *World Bank Annual Report 2023: A New Era in Development*. Washington, DC, USA: World Bank, 2023, doi: https://doi.org/10.1596/AR2023EN.

[2] B. Flyvbjerg, M. K. Skamris Holm, and S. L. Buhl, "How common and how large are cost overruns in transport infrastructure projects?," *Transp. Rev.*, vol. 23, no. 1, pp. 71–88, 2003, doi:

https://doi.org/10.1080/01441640309904.

[3] F. Wang, C.-C. B. Lee, and N. G. Gharaibeh, "Network-level bridge deterioration prediction models that consider the effect of maintenance and rehabilitation," *J. Infrastruct. Syst.*, vol. 28, no. 1, Art. no. 05021009, Mar. 2022, doi: https://doi.org/10.1061/(ASCE)IS.1943-555X.0000662.

[4] G. Morcous, "Performance prediction of bridge deck systems using Markov chains," *J. Perform. Constr. Facil.*, vol. 20, no. 2, pp. 146–155, May 2006, doi: https://doi.org/10.1061/(ASCE)0887-3828(2006)20:2(146).

[5] A. Shtayat, S. Moridpour, B. Best, and M. Abuhassan, "Using supervised machine learning algorithms in pavement degradation monitoring," *Int. J. Transp. Sci. Technol.*, vol. 12, no. 2, pp. 628–639, June 2023, doi: https://doi.org/10.1016/j.ijtst.2022.10.001.

[6] H. Majidifard, Y. Adu-Gyamfi, and W. G. Buttlar, "Deep machine learning approach to develop a new asphalt pavement condition index," *Constr. Build. Mater.*, vol. 247, Art. no. 118513, June 2020, doi: https://doi.org/10.1016/j.conbuildmat.2020.118513.

[7] A. A. Ali, A. Milad, A. Hussein, N. I. M. Yusoff, and U. Heneash, "Predicting pavement condition index based on the utilization of machine learning techniques: A case study," *J. Road Eng.*, vol. 3, no. 3, pp. 266–278, Sep. 2023, doi: https://doi.org/10.1016/j.jreng.2023.04.002.

[8] N. Elshaboury, M. S. Yamany, S. Labi, and O. Smadi, "Enhancing local road pavement condition prediction using Bayesian-optimized ensemble machine learning and adaptive synthetic sampling technique," *Int. J. Pavement Eng.*, vol. 25, no. 1, Art. no. 2365957, 2024, doi: https://doi.org/10.1080/10298436.2024.2365957.

[9] T. Yu, L.-I. Pei, W. Li, Z.-Y. Sun, and J. Huyan, "Pavement surface condition index prediction based on random forest algorithm," *J. Highw. Transp. Res. Dev. (Engl. Ed.)*, vol. 15, no. 4, pp. 1–11, Dec. 2021, doi: https://doi.org/10.1061/JHTRCQ.0000794.

[10] R. Assaad and I. H. El-Adaway, "Bridge infrastructure asset management system: Comparative computational machine learning approach for evaluating and predicting deck deterioration conditions," *J. Infrastruct. Syst.*, vol. 26, no. 3, Art. no. 04020032, Sep. 2020, doi: https://doi.org/10.1061/(ASCE)IS.1943-555X.0000572.

[11] Y. Liao, R. Lin, R. Zhang, and G. Wu, "Attention-based LSTM (AttLSTM) neural network for seismic response modeling of bridges," *Comput. Struct.*, vol. 275, Art. no. 106915, Jan. 2023, doi: https://doi.org/10.1016/j.compstruc.2022.106915.

[12] G. H. Coffie and S. K. F. Cudjoe, "Using extreme gradient boosting (XGBoost) machine learning to predict construction cost overruns," *Int. J. Constr. Manag.*, vol. 24, no. 16, pp. 1742–1750, 2024, doi: https://doi.org/10.1080/15623599.2023.2289754.

[13] J. Zhang, J. Yuan, A. Mahmoudi, W. Ji, and Q. Fang, "A data-driven framework for conceptual cost estimation of infrastructure projects using XGBoost and Bayesian optimization," *J. Asian Archit. Build. Eng.*, vol. 24, no. 2, pp. 751–774, 2025, doi: https://doi.org/10.1080/13467581.2023.2294871.

[14] J. van Remmerden, M. Kenter, D. M. Roijers, C. Andriotis, Y. Zhang, and Z. Bukhsh, "Deep multi-objective reinforcement learning for utility-based infrastructural maintenance optimization," *Neural Comput. Appl.*, vol. 37, pp. 24719–24742, 2025, doi: https://doi.org/10.1007/s00521-024-10954-0.

[15] S. E. Bibri, J. Huang, S. K. Jagatheesaperumal, and J. Krogstie, "The synergistic interplay of artificial intelligence and digital twin in environmentally planning sustainable smart cities: A comprehensive systematic review," *Environ. Sci. Ecotechnol.*, vol. 20, Art. no. 100433, July 2024, doi: https://doi.org/10.1016/j.ese.2024.100433.