EPiC Series in Built Environment
Volume 7, 2026, Pages 713–722
Proceedings of Associated Schools of Construction 62nd Annual International Conference

Evaluating the Impact of Trend-Smoothing on Pavement Condition Classification Models

Pham Anh Thu Nguyen[1], Frederick Bongwon Chung[1], Houman Kamran Habibkhani[1], and Luke Lee[1]
[1]University of the Pacific

Machine learning (ML) has become an increasingly important component of pavement management systems (PMS), where historical condition data are used to support maintenance and rehabilitation planning. However, these datasets frequently contain sensor noise and undocumented field activities, which can introduce abrupt and unrealistic improvements in condition scores. Such anomalies disrupt the expected gradual deterioration patterns and can reduce the predictive reliability of data-driven models. To address this, a trend-consistent adjustment method based on typical annual deterioration rates is applied to smooth the data and restore continuity. This study identifies the optimal degree of smoothing to apply during data preprocessing to maximize generalization accuracy when training data contain noise and unrecorded maintenance events. Model evaluation is performed using the original, uncorrected test data to reflect real-world prediction scenarios. Prediction performance improved significantly with correction, with the best model (Voting Classifier) reaching an F1-Score of 81.98% at the 10% correction level, representing a 7.35 percentage point increase over the raw data baseline. The optimal correction range was found to be 10-20% of imperfect sections corrected, confirming that light, selective smoothing balances trend fidelity with real-world variability better than raw or heavily smoothed data, producing a more reliable prediction of pavement deterioration.

Keywords: Pavement Management, Artificial Intelligence, Data Imputation, Predictive Modeling, Pavement Deterioration

Introduction

The adoption of ML techniques has become central to optimizing modern PMS worldwide. Accurate predictive models, trained on historical condition data, are essential for determining when and where to allocate limited maintenance and rehabilitation (M&R) resources. These models allow transportation agencies to forecast the future condition of road assets, moving management from a reactive to a proactive strategy (Hosseini et al., 2020; Yang et al., 2025). The Georgia Department of Transportation (GDOT), like many agencies, relies on collected data, such as the Overall Condition Index (OCI), to make these critical planning decisions.

However, a fundamental challenge in leveraging large real-world OCI datasets is the prevalence of data anomalies. These inconsistencies arise from a combination of sensor noise during data collection, errors in manual reporting, and undocumented field events (such as unrecorded minor maintenance or spot repairs) that incorrectly show a sudden improvement in pavement condition (FHWA, 2013; Chang et al., 2020; Abukhalil et al., 2022). These unexpected score increases violate the underlying physical principle of gradual pavement deterioration, introducing noise that significantly weakens the generalization and predictive power of ML models (Aranha et al., 2023). Without reliable trend-following data, future OCI scores cannot be accurately predicted, leading to suboptimal planning and resource waste.

Consequently, this study addresses the essential but under-examined question of data preparation: How much historical pavement condition data should be smoothed to maximize predictive model performance? The primary objective is to systematically investigate the relationship between the degree of data correction and the resulting accuracy of ML forecasts on untouched future data. This research utilized historical OCI data from 2017 to 2021 and employed a median-based, trend-consistent adjustment based on typical annual decline rates. The proportion of noise records in the training set corrected via median adjustment was systematically varied, ranging from 10% up to 50% of the inconsistent observations. This approach allows for a direct assessment of how smoothing intensity affects model reliability and provides a foundation for establishing practical data preprocessing guidelines.

The key contributions of this study are as follows:

- The substantial and previously under-recognized impact of smoothing intensity on pavement deterioration forecasting accuracy was demonstrated, and a structured approach was introduced to systematically vary this intensity by adjusting the proportion of corrected inconsistent records.

- A median-based adjustment anchored to typical annual deterioration rates was applied to restore realistic pavement performance trajectories.

- Smoothing effects were quantified using multiple ML models, and predictive accuracy on unseen future-year data was evaluated to derive evidence-based recommendations for selecting smoothing levels that balance model stability with the preservation of genuine deterioration signals.

The remainder of this paper is structured as follows. The Background and Related Work section reviews the literature on ML in PMS and the challenges associated with data quality and noise. The Methodology section details the data source, the anomaly correction mechanism, the systematic experimental design, and the suite of ML models deployed. The Results and Discussion section presents the prediction error metrics (F1-score) across the varied correction levels and ML models and interprets the findings. Finally, the paper concludes with a summary of the study’s contributions and practical recommendations for pavement engineers followed by future work.

Background and Related Work

The prediction of pavement performance has been studied for decades, with early approaches relying on mechanistic-empirical models and regression-based curve fitting. These methods established deterioration as a largely monotonic process, but they often struggled to capture local variability and diverse influencing factors (Prozzi & Madanat, 2004). More recently, the rise of PMS has generated large historical datasets of condition indices such as Pavement Condition Index and OCI, enabling data-driven methods (Lu & Tolliver, 2012; Prozzi & Madanat, 2004). However, these datasets frequently contain noise, measurement errors, or undocumented maintenance, which introduce sudden increases that do not reflect true pavement behavior. The treatment of such irregularities remains an open challenge in the literature (Abukhalil et al., 2022).

ML has emerged as a powerful alternative to traditional methods for pavement deterioration, especially for classifying pavement sections into condition categories (e.g., excellent, good, fair, poor). Studies have applied tree-based ensembles such as Random Forests (Zhang et al., 2018) and boosting methods like XGBoost and LightGBM (Wang et al., 2020) due to their ability to handle non-linearities and heterogeneous features. Other works have explored distance-based classifiers such as K-Nearest Neighbors (KNN), as well as linear baselines like Logistic Regression and Support Vector Classifiers (SVC) (Gopalakrishnan, 2018). More recently, ensemble strategies including majority voting and stacking have been shown to further improve predictive stability (Chen & Guestrin, 2016; Ke et al., 2017). Despite these advances, model accuracy is still highly dependent on the quality of the input dataset, particularly the treatment of anomalous condition records (Aranha et al., 2023; Erfani et al., 2025).

The preprocessing of historical pavement data has typically focused on outlier detection (e.g., Z-score methods, Hampel filters) or simple averaging to fill missing or inconsistent values (Kheirati et al., 2022). While effective for point-level cleaning, these methods rarely enforce the expected monotonic deterioration trend. As a result, models trained on original data may learn unstable patterns, especially near condition class thresholds.

To address this, several studies have introduced trend-smoothing techniques that adjust reversals to better align with expected declines (Prozzi & Madanat, 2004). A key methodological choice in smoothing is whether to use the mean or the median as the basis for correction. The mean can be skewed by extreme drops or measurement errors, while the median provides a more robust measure of central tendency for year-to-year declines (Hyndman & Athanasopoulos, 2018). Although median-based corrections have been widely studied in general time-series preprocessing, their application to pavement deterioration datasets remains limited, particularly in the context of selective smoothing rather than full replacement.

Existing work shows both the promise of ML for pavement condition classification and the need to preprocess noisy records. Yet most studies either train on raw data or apply uniform smoothing to every irregularity. Far less attention is paid to the amount of smoothing, and how many records should be adjusted before training. This study addresses that gap by varying the share of adjusted records under a fixed, median-based trend rule and evaluating on a separate set of road sections that never appear in training (to avoid overlapping that could inflate accuracy). Across Random Forest, XGBoost, LightGBM, KNN, Logistic Regression, SVC, and ensemble methods (Voting, Stacking), moderate adjustment (~10–20%) outperforms both no adjustment and heavy adjustment, offering a practical guideline for balancing data fidelity and predictive performance.

Methodology

Methodology Overview

This study follows a structured, step-by-step methodology to evaluate how selective trend-smoothing of historical pavement condition data affects ML classification performance (see Figure 1). First, historical OCI data and associated traffic, climate, and structural variables are collected and organized. Next, anomalous OCI records that violate expected deterioration trends are identified using longest strictly decreasing subsequences. These anomalies are selectively corrected using median-based annual decline rates, with the proportion of corrected records systematically varied. ML classification models are then trained on each corrected dataset, while evaluation is performed on unmodified test data to reflect real-world prediction conditions. Finally, model performance is compared across correction levels using the F1-score to identify the optimal degree of smoothing.

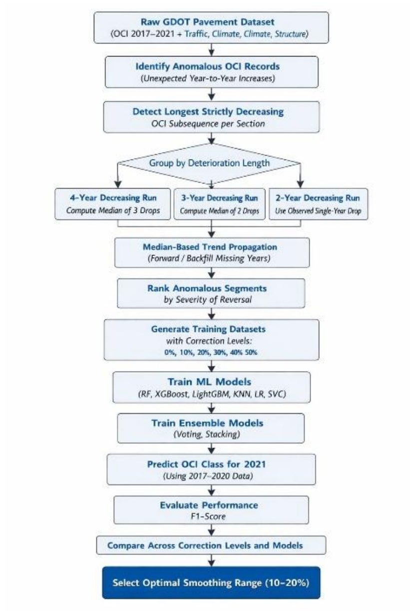

Figure 1. Visual overview of the complete workflow

Data Source and Scope

A dataset from the GDOT was used to develop predictive models. The dataset contains predefined maintenance sections observed annually from 2017 to 2021, so each pavement section (unique by its name) carries a five-point condition history together with traffic, climate, and structural descriptors. GDOT reports condition using the OCI, a 0–100 index computed from distress severity and extent. In this study, the prediction target is the end-of-period OCI score. The GDOT transportation asset management plan recommends different maintenance types depending on the range of OCI, which can be expressed in classes (see Table 1). In addition to the five OCI values, the feature set includes average annual daily traffic (AADT), climate summaries (precipitation; mean/min/max temperature), and structural fields (pavement type, lanes, lane miles, mileage). These covariates are used as predictors for the OCI score. Descriptions of variables used in this study are described in Table 2.

Table 1. Pavement condition classes by OCI_{2021}

Condition Classes	OCI
0	[90,100]
1	[80,90)
2	[70,80)
3	[60,70)
4	[50,60)
5	[0,50)

Table 2. Variable Descriptions

Variables	Descriptions
OCI	Score: Numerical, ranging from 44.86 to 100.00
Average annual daily traffic	Count: Numerical, ranging from 90 to 216,000
Number of lanes	Count: Numerical, ranging from 1 to 4
Average Annual Precipitation	Inches: Numerical, ranging from 45.25 to 78.49
Average Temperature	Fahrenheit: Numerical, ranging from 57.08 to 69.57
Minimum Temperature	Fahrenheit: Numerical, ranging from 45.97 to 59.35
Maximum Temperature	Fahrenheit: Numerical, ranging from 67.98 to 80.78

Anomaly Identification and Correction

Anomaly Definition

Anomalies in pavement condition data are defined as unexpected year-to-year increases in OCI values that contradict the physical principle of gradual pavement deterioration. For a pavement section with OCI values an anomaly is identified when:

$$OCI_{t+1} > OCI_t$$

Such increases may arise from sensor noise, reporting inconsistencies, or undocumented maintenance activities. These anomalies are identified consistently across all pavement sections and observation years.

Longest Strictly Decreasing Subsequence (LSDS)–Based Correction

The core principle underlying the proposed data correction framework is the identification of the LSDS of OCI values for each pavement section. The LSDS is interpreted as the most plausible representation of the section's natural deterioration trend, as it preserves physically consistent year-over-year declines while excluding non-physical increases attributable to noise or undocumented interventions.

Based on the length of the identified LSDS, pavement sections are categorized into distinct deterioration run-length groups. This grouping determines how a representative median annual decline is computed and subsequently propagated to correct missing or inconsistent OCI values. Specifically, sections exhibiting longer monotonic deterioration runs provide more reliable information for estimating typical decline rates, whereas shorter runs require more conservative extrapolation.

The distribution of pavement sections by LSDS run length is summarized in Table 3, which illustrates the prevalence of 4-year, 3-year, and 2-year deterioration patterns across the dataset.

Table 3. Distribution of Pavements by Longest Observed Deterioration Run Length

History Length Group	Number of Sections (N)
4-Year	438
3-Year	982
2-Year	1339
Total number of sections	3004

After identifying the longest strictly decreasing subsequence (LSDS), pavement sections are grouped based on the length of this monotonic deterioration run. Each group is corrected using a group-specific strategy, as summarized below:

- 4-year run (four consecutive decreases): The median of the three observed year-to-year OCI drops is computed and used to correct a single missing boundary year. If the run ends in 2020, the 2021 OCI value is forward filled; if the run ends in 2021, the 2017 value is backfilled.

- 3-year run (three consecutive decreases): The median of the two observed year-to-year OCI drops is computed and propagated outward to correct the remaining end years, preserving a consistent deterioration trend.

- 2-year run (single decrease): The observed year-to-year OCI drop is used as the representative decline and propagated forward and backward to correct all remaining years.

These rules ensure that corrections are consistent with the observed deterioration behavior while avoiding arbitrary smoothing. The complete LSDS identification, grouping logic, and median-decline correction process are illustrated in Figure 2.

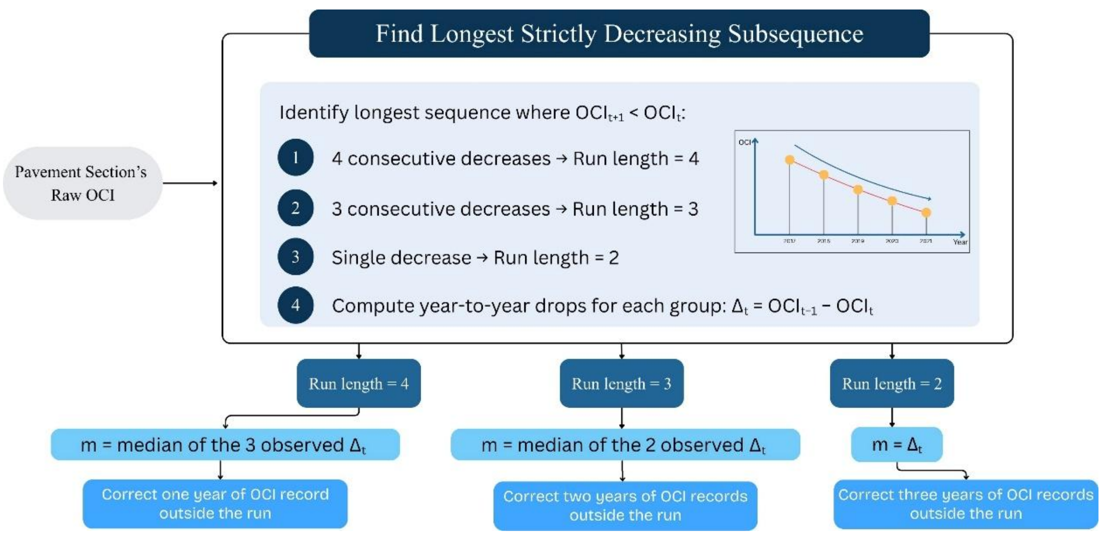

Figure 2. LSDS-based classification and median-decline correction of OCI

Comparison with Standard Anomaly Detection Methods

Standard anomaly detection approaches, such as threshold-based filtering or statistical outlier detection, typically identify and remove observations that deviate significantly from the overall distribution. While effective for point-level noise reduction, such methods are less suitable for short pavement condition time series, as they disrupt temporal continuity and reduce the already limited number of training observations. In contrast, the proposed trend-based smoothing approach corrects anomalous OCI increases by enforcing a physically meaningful monotonic deterioration pattern, rather than discarding entire observations. This preserves complete temporal sequences for each pavement section, maintains sample size, and prevents bias introduced by selectively removing years with atypical behavior. Moreover, correcting anomalous observations allows machine learning models to learn stable deterioration trends while still reflecting realistic year-to-year variability, which is critical for generalization to real-world pavement management scenarios.

Experimental Design and Systematic Variation

To test the core research hypothesis, a series of experimental training datasets were constructed. The total count of identified anomalous OCI increases established the population of noisy records. Correction levels were systematically varied by imputing the calculated median trend decline into 10%, 20%, 30%, 40%, and 50% of the total identified anomalous records. The dataset was partitioned using a 70/30 training–testing split, which was applied consistently across all ML models. All anomaly correction and trend-smoothing procedures were performed exclusively on the training set. The testing set was left entirely unmodified and retained its original OCI values to reflect real-world prediction conditions. For each correction level, a unique corrected training dataset was generated, while the same fixed test set was used for evaluation. This design ensured that model predictive performance was evaluated strictly against the raw, original condition data, effectively isolating the impact of the imputation proportion on model generalization.

Predictive Modeling and Evaluation Metrics

To establish the robustness of the findings, a comprehensive suite of ML models was deployed. All models were used for the classification task. This suite included individual classifiers (Random Forest, XGBoost, LightGBM, KNN, Logistic Regression, and SVC). Additionally, two sophisticated ensemble techniques, a Voting Classifier and a Stacking Classifier, were utilized to potentially enhance predictive performance. Each classifier was trained on the six prepared datasets (the five corrected sets plus the raw, uncorrected set). The models were tasked with predicting the OCI condition class for the year 2021 based on four historical data from year 2017 up to year 2020.

Performance across the different correction levels and ML models was primarily evaluated using Classification F1-Score defined as:

$$F1 - Score = \frac{2 \, . Precision \, . Recall}{Precision + Recall}$$

Where precision measures the proportion of correctly predicted condition classes among all predicted instances, and recall measures the proportion of correctly predicted instances among all true instances. The F1-score balances these two quantities and is particularly suitable for multi-class pavement condition classification.

The correction level that yielded the maximum classification metric across different models was designated as the optimal imputation dose for enhancing predictive reliability.

Results and Discussion

The classification performance metrics, specifically the F1-Score (see Table 4), demonstrates a significant, uniform improvement when training data is corrected, confirming that data refinement dramatically enhances the predictive capabilities of ML models in PMS.

Table 4. F1-Score Metrics Across All Models and Correction Levels

Models	**0% (Raw)**	**10%**	**20%**	**30%**	**40%**	**50%**
Random Forest	73.57	81.21	80.50	80.07	80.30	80.68
XGBoost	73.41	81.86	80.84	79.53	79.14	79.40
LightGBM	72.61	81.08	81.14	81.07	80.84	79.41
KNN	71.07	79.53	79.07	78.82	77.69	78.01
Logistic Regression	72.23	80.26	79.32	79.32	79.22	79.74
SVC	71.54	80.13	79.85	79.74	79.60	80.04
Voting Classifier	74.63	81.98	81.25	80.94	80.44	80.25
Stacking Classifier	74.58	81.90	80.64	80.69	80.25	79.41

Predictive Performance Across Correction Levels

As shown in Table 4 and visualized in Figure 1, the effect of introducing data correction is highly positive. Performance scores across all models jumped significantly from the raw data F1-Score average of approximately 73% to average of approximately 81% for the 10% correction level. This improvement confirms the foundational hypothesis: correcting a small fraction of noisy records fundamentally stabilizes the learning process by enforcing the expected deterioration principle.

The peak performance for the OCI classification task was concentrated in the range of 10% to 20% correction.

- The Voting Classifier achieved the single highest F1-Score of 81.98% at the 10% correction level, representing a substantial increase over the raw data baseline of 74.63%.

- High scores were consistently observed at the 20% level, where the Voting Classifier reached 81.25% and LightGBM reached 81.14%.

Critically, the performance stability was high across the entire range from 10% up to 50% correction. The median trend correction method applied here proved highly resilient; although the maximum generalization accuracy occurred at light smoothing (10–20%), the performance remained robust, often exceeding 79% F1-Score, even when applied liberally. This suggests that trend-consistent adjustment is very effective at capturing the naturally deterioration signal.

Predictive Performance across ML Models

The evaluation confirmed the continued superiority of ensemble methods for complex pavement data classification. Ensemble methods, including the Voting Classifier and the Stacking Classifier, consistently delivered the highest performance across all correction levels, demonstrating that combining the outputs of diverse models is a reliable strategy for OCI class prediction. Among the individual classifiers, XGBoost, LightGBM, and Random Forest performed the best, routinely achieving F1-Scores above 80% in the optimal correction range. Distance-based and linear classifiers (KNN, Logistic Regression, and SVC) showed slightly lower, but still significant, improvement following correction. These results provide strong, quantifiable evidence that an optimal degree of trend-smoothing exists and offers practitioners an evidence-based guideline: to optimize classification models for future maintenance and rehabilitation (M&R) planning, one should target light, selective correction (10%–20% of anomalies). This level balances trend fidelity with real-world variability better than raw or heavily smoothed data

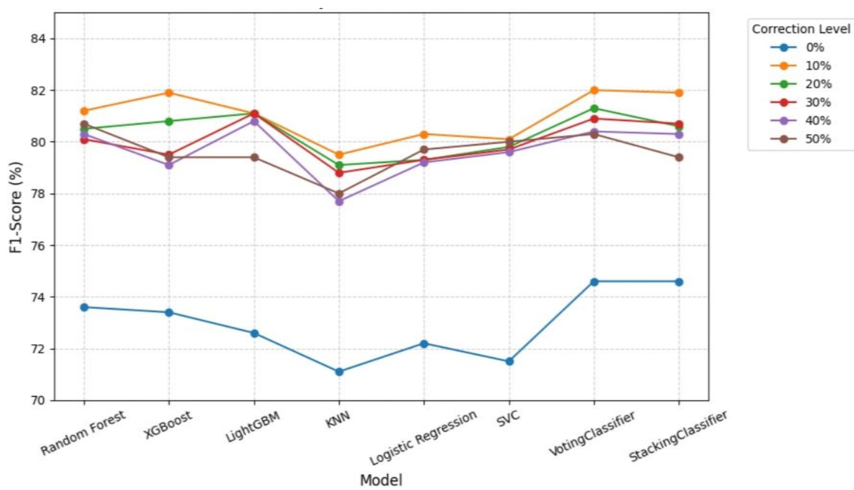

Figure 3. Models' F1-Scores across Different Correction Levels

Conclusion

This study addressed the critical data quality challenge in PMS by systematically investigating the optimal proportion of anomaly correction for maximizing the generalization of ML classification models. Utilizing a comprehensive suite of classifiers on GDOT OCI data, the research successfully quantified the impact of median trend imputation across a range of correction levels. The findings establish clear and practical guidelines for data preparation. Applying correction to 10%-20% of the noisy training data yields a substantial increase in predictive performance (an approximately 8-point F1-Score increase across the suite). The predictive power of the classifiers peaked at the 10% and 20% correction levels. Crucially, performance stability was high across the entire 10% to 50% range, suggesting that the median decline correction is a robust and resilient method. Ensemble methods, particularly the Voting and Stacking Classifiers, consistently delivered the highest performance, demonstrating that combining the outputs of diverse models is the most reliable strategy for OCI class prediction. The consistency of the results advises practitioners to target light, selective correction (10%–20% of anomalies) to optimize their classification models for future maintenance and rehabilitation planning. Future work should focus on automating the identification of the optimal correction level using adaptive algorithms rather than fixed percentages. This would involve developing a data-driven approach, such as utilizing meta-learning or self-optimizing statistical methods, to dynamically determine the most effective smoothing dose based on the specific characteristics and inherent noise level of each new dataset.

References

Abukhalil, Y., Halim, R., & Kumar, A. (2022). Detection of unreported treatments in pavement management systems using machine-learning classification. Iowa Department of Transportation. https://doi.org/10.1061/JPEODX.0000400

Aranha, A. L., Bernucci, L. L. B., & Vasconcelos, K. L. (2023). Effects of different training datasets on machine learning models for pavement performance prediction. Transportation Research Record: Journal of the Transportation Research Board, 2677(8), 196–206. https://doi.org/10.1177/03611981231155902

Chang, G., Gilliland, A., & Rada, G. R. (2020). Successful practices for quality management of pavement surface condition data. Federal Highway Administration. https://rosap.ntl.bts.gov/view/dot/55792

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Gopalakrishnan, K. (2018). Deep learning in data-driven pavement image analysis and automated distress detection: A review. Data, 3(3), 28. https://doi.org/10.3390/data3030028

Hosseini, S. A., Alhasan, A., & Smadi, O. (2020). Use of deep learning to study modeling deterioration of pavements: A case study in Iowa. Infrastructures, 5(11), 95. https://doi.org/10.3390/infrastructures5110095

Hyndman, R., & Athanasopoulos, G. (2021). Forecasting: Principles and practice (3rd ed.). OTexts. https://otexts.com/fpp3

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Microsoft Research, Peking University, & Microsoft Redmond. (2017). LightGBM: A highly efficient gradient boosting decision tree. In 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

Kheirati, A., & Golroo, A. (2022). Machine learning for developing a pavement condition index. Automation in Construction, 139, 104296. https://doi.org/10.1016/j.autcon.2022.104296

Prozzi, J. A., & Madanat, S. M. (2004). Development of pavement performance models by combining experimental and field data. https://escholarship.org/uc/item/6cf8v5cw

Wang, X., Zhao, J., Li, Q., Fang, N., Wang, P., Ding, L., & Li, S. (2020). A hybrid model for prediction in asphalt pavement performance based on support vector machine and grey relation analysis. Journal of Advanced Transportation, 2020, 1–14. https://doi.org/10.1155/2020/7534970

Zhang, A., Wang, K. C. P., Fei, Y., Liu, Y., Tao, S., Chen, C., Li, J. Q., & Li, B. (2018). Deep learning–based fully automated pavement crack detection on 3D asphalt surfaces with an improved CrackNet. Journal of Computing in Civil Engineering, 32(5). https://doi.org/10.1061/(asce)cp.1943-5487.0000775