Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Machine Learning-Based Classification of Moisture Damage Occurrence and Severity in Asphalt Pavements Using LTPP Data

Youssef Mousa
Texas A&M University
Momen R. Mousa
mxm363@shsu.edu
Sam Houston State University
Egehan Koyuncu
Sam Houston State University
Mahasen Abdelhafez
Sam Houston State University
Min Suh
Sam Houston State University
Iftekhar Basith
Sam Houston State University

Research Article

Keywords: Moisture damage, Pavement performance prediction, Stripping, Pavement management System, Long-Term Pavement Performance

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11750027/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Machine Learning-Based Classification of Moisture Damage Occurrence and Severity in Asphalt Pavements Using LTPP Data

Abstract

Moisture damage is a critical distress mechanism that significantly compromises the durability, structural performance, and sustainability of asphalt pavements. Conventional detection methods, such as core extraction and non-destructive evaluation (NDE) techniques, provide valuable and reliable information regarding existing moisture damage in asphalt pavements. However, these methods are primarily used after deterioration has developed and therefore offer limited capability for predicting the future occurrence and progression of moisture damage, which is important for proactive pavement management. Consequently, this study aimed to develop a data-driven framework to predict both the occurrence and severity of moisture damage in asphalt pavements using project-level characteristics. A total of 8094 records were extracted from the Long-Term Pavement Performance (LTPP) database and used to train two Random Forest–based classification models: (i) a binary model to predict the presence of moisture damage, and (ii) a multi-class model to predict damage severity (low, moderate, or high). The models were developed using an 80/20 training–testing split with default hyperparameter settings. The binary occurrence model achieved an overall testing accuracy of 93%, while the multi-class severity model achieved an overall accuracy of 82%. Feature importance analysis showed that environmental variables, particularly temperature and precipitation, along with traffic loading, played a dominant role in predicting moisture damage. Similarly, the pavement structural capacity contributed significantly to predicting moisture damage in asphalt pavements. The developed models were integrated into a web-based decision-

support tool that enables transportation agencies to perform predictive assessments, optimize the timing of inspections and coring activities, and support proactive maintenance decision-making. The tool is intended to complement, rather than replace, conventional assessment methods by helping practitioners identify the pavement ages at which moisture damage is most likely to occur and when verification activities should be conducted.

Keywords: Moisture damage; Pavement performance prediction; Stripping; Pavement management System; Long-Term Pavement Performance

1. Introduction

Roads are a critical component of the national infrastructure used in transporting people, freight, and services safely, and creating prosperity and welfare. However, road infrastructure systems are exposed to the accelerating impacts of climate change, including heat waves, intense precipitation, flooding, erosion, and frequent freeze-thaw cycles. These significantly affect pavement durability, structural integrity, and long-term functionality, creating urgent challenges for the design, preservation, and management of resilient transport networks. Water is widely recognized as one of the most detrimental factors influencing pavement performance. Moisture ingress into asphalt pavements triggers moisture damage which is a primary cause of premature pavement deterioration. Moisture damage can be defined as the loss of bond between the asphalt cement or the mastic (asphalt cement plus the mineral filler—74 μm and smaller aggregate) and the fine and coarse aggregate. Moisture damage also occurs because moisture permeates and weakens the mastic itself under repeated traffic loading [1].

Moisture damage typically initiates at critical locations such as the bottom of the pavement layer, where tensile stresses are highest and moisture presence is significant.

As traffic loading continues, the loss of adhesion between the binder and aggregates progressively spreads through the mixture, leading to upward propagation of damage toward the surface [2]. Over time, this internal deterioration manifests at the pavement surface in the form of visible distresses, such as raveling and pothole formation [3] reducing pavement durability. This loss of durability directly translates into a reduction in pavement service life. Previous studies have shown that moisture damage can reduce the pavement service life by up to 11 years [4]. The implications of moisture damage extend beyond durability to encompass broader sustainability considerations. For example, pavements that last for 15 years instead of 7 years can have dramatically smaller carbon footprints over their lifecycles. Therefore, moisture resilience is a sustainability imperative, not just a performance target [5].

Recent studies have also focused on improving asphalt pavement performance and sustainability through the use of innovative and recycled materials, including rubber-modified asphalt mixtures, tire-derived aggregates, recycled tire fibers, waste glass powder modifiers, hydrated lime, warm-mix asphalt technologies, and other waste-based composite materials. These studies have demonstrated the potential of such materials to enhance pavement durability, reduce moisture susceptibility, and improve long-term performance under varying traffic and environmental conditions [6-12]. While these efforts primarily focus on material enhancement, they further emphasize the importance of understanding and predicting pavement deterioration mechanisms, including moisture-induced damage.

A previous study [13] has been conducted to understand the mechanisms of moisture damage in asphalt pavements, with particular emphasis on the adhesion phenomena

between asphalt binder and aggregate surfaces. Adhesion theories were developed to explain the bonding mechanisms within the bitumen–aggregate system; however, it has been widely recognized that no single theory can fully capture this complex interaction due to the multiple physical and chemical factors involved. In parallel, significant research efforts have been devoted to the detection and assessment of moisture damage in the field, utilizing advanced technologies such as Ground Penetrating Radar (GPR) [14]. However, despite these advancements, limited studies have focused on the development of predictive tools capable of estimating the occurrence and severity of moisture damage as a function of project characteristics and pavement age. In particular, there remains a gap in the ability to estimate the age classes in which moisture damage is likely to be present after construction and to quantify its expected severity. Addressing this gap is critical for proactive pavement management, and it highlights the need for data-driven approaches that can support timely decision-making and improve long-term pavement performance.

2. Scope and Objective

This study aims to develop a user-friendly web-based application that can predict the occurrence (present or absent) and severity of moisture damage in asphalt pavements based on project-level characteristics, such as climate, traffic, pavement structure, etc. The web-based application is intended to provide age-class-based estimates of the occurrence and severity of moisture damage (age classes of 0–2, 3–5, 6–8, 9–12, and ≥13 years) for the specified project conditions. The estimates are made for a given set of project characteristics and age class; the framework does not forecast the future condition of a specific section from its own observed history. To achieve this objective, a total of

8094 records were extracted from the Long-Term Pavement Performance (LTPP) database and used to train two machine learning–based classification models: (i) a model to predict the occurrence of moisture damage (present or absent), and (ii) a model to predict the severity level of moisture damage (low, moderate, or high), both based on project-level inputs. The developed models were subsequently integrated into a hierarchical (two-stage) prediction framework and implemented within a web-based application. This integration enables transportation agencies and practitioners to effectively utilize the predictive capabilities of the models without requiring prior expertise in machine learning, thereby supporting data-driven decision-making in pavement management and design.

3. Background

The conventional method to detect moisture damage in asphalt pavements is through core extraction which enables visual inspection of pavement layers where moisture damage can be clearly identified. Yet, coring is destructive, time-consuming, costly, and labor-intensive. Moreover, cores only represent discrete locations and may not accurately represent the continuous conditions throughout the pavement section. Consequently, moisture-induced damage is rarely addressed during pavement preservation activities, which may result in loss of state funds due to unsatisfactory performance after repair. To address this issue, numerous research studies have been conducted to detect moisture damage in asphalt pavements through non-destructive evaluation (NDE) techniques to allow for cost-effective maintenance activities to address the detected damage. For instance, the Texas Department of Transportation investigated the ability of GPR to detect moisture damage in asphalt pavement layers [15]. The study demonstrated that

GPR could successfully detect moisture damage, particularly when deterioration had reached moderate to severe levels. It was observed that pavement sections affected by moisture damage exhibited higher moisture content, increased air voids, or both, resulting in anomalously elevated dielectric constants in the damaged layers. Additionally, intermittent negative peaks within the surface layer of GPR scans were identified as typical indicators of moisture presence.

Building on the application of wave-based NDE methods, Celaya and Nazarian [16] used Portable Seismic Property Analyzer (PSPA) with ground truth data to evaluate the capabilities of both Ultrasonic Surface Waves (USW) and Impact Echo (IE) in detecting moisture damage in asphalt pavements. USW was found to be superior to the IE method in identifying the severity and extent of the damage in terms of asphalt layer stiffness. In parallel with these developments, recent research [17-18] has explored the integration of advanced thermal imaging techniques with deep learning algorithms to enhance the detection of moisture-related damage. By analyzing thermal variations across the pavement surface, this approach has demonstrated strong potential for accurately identifying subsurface moisture conditions. When combined with RGB imagery, thermal imaging methods further improved detection capabilities, offering a promising tool for comprehensive pavement condition assessment. Furthermore, other recent studies have successfully employed deflection-based measurements obtained from devices such as the Rolling Wheel Deflectometer (RWD) and the Traffic Speed Deflection Device (TSDD) to detect moisture damage in asphalt pavements [19]. These methods provide continuous, network-level assessment capabilities, thereby supporting more efficient identification and characterization of moisture-induced deterioration.

Despite the demonstrated effectiveness of these NDE techniques in identifying moisture-induced damage, their application remains largely reactive, as they primarily detect moisture damage after it has already occurred. This limitation is significant because moisture-induced damage develops gradually and is governed by complex interactions among traffic loading, environmental conditions, material properties, and drainage characteristics. The capability to predict both the occurrence and severity of moisture damage throughout the pavement service life is essential for optimizing material selection, improving structural design, and enhancing maintenance planning. Limited research efforts have focused on developing such predictive frameworks. Therefore, there is a critical need for new approaches that can integrate multiple influencing factors and provide forward-looking predictions of moisture damage in asphalt pavements.

In this context, data-driven modeling approaches, particularly supervised machine learning techniques, offer a promising pathway to address this research gap. Unlike traditional parametric methods, which rely on predefined functional forms and assumptions regarding variable relationships, machine learning models are capable of capturing complex, non-linear interactions among a large number of influencing factors without requiring strict assumptions about data structure. This capability is especially important in pavement performance prediction, where datasets often exhibit high dimensionality and multicollinearity, limiting the predictive accuracy of conventional approaches [20].

Various machine learning (ML) algorithms have been successfully applied in transportation and pavement engineering applications, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Trees (DT), Gradient

Boosting Machines (GBM), and ensemble learning techniques. Compared with traditional statistical approaches, ML methods offer several advantages, including the ability to capture complex nonlinear relationships, analyze large and high-dimensional datasets, and identify interactions among multiple influencing factors without requiring predefined functional relationships. Such capabilities are particularly beneficial in pavement performance studies, where traffic, environmental, structural, and maintenance variables interact in a complex manner. However, ML models may also present certain challenges, including the risk of overfitting, sensitivity to data quality, computational requirements, and reduced interpretability for some advanced algorithms. These limitations can be mitigated through appropriate data preprocessing, feature selection, cross-validation procedures, regularization techniques, and ensemble learning approaches, which improve model robustness and generalization capability [21]. One of the widely used machine learning techniques among the computer science community due to its interpretability and ease of implementation is the decision tree algorithm [22]. However, individual decision trees may suffer from instability and limited predictive performance when applied to complex, high-dimensional datasets. To overcome these limitations, ensemble methods such as the Random Forest (RF) algorithm [23] have been developed. RF constructs multiple decision trees using bootstrap sampling and aggregates their predictions, resulting in improved model stability, reduced overfitting, and enhanced predictive accuracy. In addition, RF can effectively handle both numerical and categorical variables and provides measures of feature importance, making it particularly suitable for modeling complex interactions among pavement, traffic, and environmental variables.

Despite the growing application of machine learning techniques in pavement engineering, existing studies have primarily focused on predicting pavement performance measures such as roughness, cracking, rutting, structural condition, and other pavement distresses. To the best of the authors' knowledge, no previous study has developed a machine-learning framework specifically designed to predict both the occurrence and severity of moisture damage in asphalt pavements at the network level. Furthermore, the proposed framework integrates a comprehensive set of project-level variables representing traffic, environmental, structural, drainage, maintenance, and material-related characteristics obtained from 497 pavement sections across 48 states. Therefore, this study introduces a novel data-driven approach for predicting moisture damage occurrence and progression and implementing the resulting models within a practical web-based decision-support tool for pavement management applications.

4. Data Description and Preprocessing

4.1 Data Retrieval and Compilation

The data used in this study were obtained from the Long-Term Pavement Performance (LTPP) database [24], a federally sponsored program administered by the Federal Highway Administration (FHWA). The LTPP program was established to systematically collect long-term field performance data for in-service pavement sections across the U.S. In the LTPP database, each pavement section represents a predefined roadway segment with a standard length of approximately 152 m (500 ft). Pavement sections are periodically surveyed to collect detailed information on pavement distress, structural characteristics, material properties, traffic loading, and environmental exposure.

In this study, the LTPP database was mined, and only asphalt pavement sections constructed in or after 1985 were selected to ensure that distress surveys were available since the time of construction, thereby capturing pavement performance throughout the pavement's service life starting from construction. This filtering process resulted in a total of 497 pavement sections distributed across 48 different states capturing a wide range of climatic conditions, traffic characteristics, and pavement structure.

For each of the 497 pavement sections, data was retrieved for multiple years of service, beginning at the year of construction (age zero) and extending forward. As a result, the same pavement section could appear multiple times in the dataset, with each record corresponding to a different observation year. This process resulted in a total of 8,094 records. Therefore, each record represented a unique combination of pavement section and monitoring year, forming a section-year panel dataset that captures the temporal evolution of pavement condition over time. For each of the 8094 records, the following was retrieved from the LTPP database:

1. Construction year: numerical variable; ranged from 1985 to 1999.
2. Distress survey year: numerical variable; ranged from 1985 to 2024.
3. Pothole area (PA): Total area of potholes observed within the pavement section in a given year; numerical variable; ranged from 0 to 6 m^2.
4. Raveling area (RA): Total area of raveling distress observed within the pavement section in a given year; numerical variable; ranged from 0 to 550 m^2.
5. Maintenance and rehabilitation (M&R) treatment: Type of M&R applied to the pavement section in a given year; categorical variable; included multiple categories such as no treatment, crack sealing, patching, mill-and-overlay, etc.

6. Base material (BM): Type of material used in the base layer; categorical variable; included 21 different materials such as crushed stone, crushed gravel, etc.
7. Subbase material (SM): Type of material used in the subbase layer; categorical variable; included 23 different material types.
8. Subgrade soil type (SG): Type of the underlying subgrade soil; categorical variable; included more than 30 soil classes such as silty sand, clayey sand, etc.
9. Subsurface drainage type (DT): Configuration of the drainage system beneath the pavement; categorical variable that included four categories; no drainage, longitudinal drains, drainage blanket, and drainage blanket with longitudinal drains.
10. Drainage location (DL): spatial distribution of the subsurface drainage system along the pavement section; categorical variable that included three categories; continuous drainage, intermittent drainage, or no drainage.
11. Drainpipe diameter (DD): Diameter of the subsurface drainage pipe; numerical variable; measured in inches and included discrete values of 0, 3, 4, 6, 8, and 18, where 0 indicates absence of drainage pipe.
12. Average Annual Daily Truck Traffic (AADTT): numerical variable; ranged from 21 to 10234 trucks/day.
13. Annual air temperature (T): Average air temperature over a year, calculated from daily mean temperatures; numerical variable; ranged from 0 to 26.5 °C.
14. Annual precipitation (P): Water equivalent of total surface precipitation over year time period; numerical variable; ranged from 53 mm to 2,162 mm.

15. Annual humidity (H): Average relative humidity for the year; numerical variable; ranged from 26% to 85%.
16. Asphalt layer thickness (AT): Total thickness of asphalt layers in the pavement structure; numerical variable; ranged from 1 to 14.7 inches.
17. Base thickness (BT): Thickness of the base layer beneath the asphalt surface; numerical variable; ranged from 0 to 28 inches.
18. Subbase thickness (ST): Thickness of the subbase layer; numerical variable; ranged from 0 to 48 inches.

4.2 Feature Engineering and Data Transformation

The pavement age (A) was computed for each of the 8094 records as the difference between the distress survey year and the construction year. However, the distribution of pavement age (A) across the dataset was highly unbalanced. For example, only a single record had an age of 37 years, whereas 482 records had an age of 5 years. Such imbalance in data distribution could negatively affect model performance and generalization. Hence, pavement age "A" was grouped into five categories: 0–2 years, 3–5 years, 6–8 years, 9–12 years, and ≥13 years. This categorization improved the balance of observations across age groups and provided a more stable representation of pavement aging for modeling purposes.

Similarly, the M&R treatments in the dataset included more than 10 treatments. These treatments were grouped into four functional categories for modeling purposes: F0 representing no treatment, F1 representing preventive maintenance treatments, F2 representing corrective maintenance treatments, and F3 representing rehabilitation

treatments. This grouping provided a simplified yet meaningful representation of treatment effects.

In this study, moisture damage was characterized based on the occurrence and severity of associated surface distresses, specifically raveling and potholes. This is because raveling and potholes are widely recognized in literature as key indicators of moisture-induced deterioration in asphalt pavements [25]. To quantify moisture damage in this study, statistical analysis of pothole area (PA) and raveling area (RA) was conducted to identify representative threshold values. The thresholds were computed using only the training subset of the data (80%, see Section 5), so that no testing information influenced the target definition. Within this training subset, observations with zero values were excluded to avoid biasing the percentile calculations toward non-damaged conditions. For pothole area (PA), the 25th percentile (PA25) was 0.045 m^2, and the 50th percentile (PA50) was 0.114 m^2. For raveling area (RA), the 25th percentile (RA25) was 34 m^2, and the 50th percentile (RA50) was 162.6 m^2. These thresholds were then applied unchanged to the testing subset. Based on these percentile thresholds, moisture damage severity was classified into four categorical levels as follows:

- No moisture damage: if PA= 0 AND RA = 0
- Low moisture damage: if [PA< PA25] AND [RA< RA25] AND NOT [PA = 0 AND RA = 0]
- Moderate moisture damage: if [PA25 ≤ PA < PA50] OR [RA25 ≤ RA < RA50] AND NOT HIGH MOISTURE DAMAGE CLASS
- High moisture damage: if [PA ≥ PA50] OR [RA ≥ RA50]

Out of the total 8094 observations, 6791 were classified as no moisture damage, 321 as low moisture damage, 337 as moderate moisture damage, and 645 as high moisture damage. This distribution showed that the no moisture damage class was dominant within the dataset. This was expected, as the dataset included a larger proportion of records from earlier stages of pavement service life, where moisture damage is less likely to develop. Moisture damage typically emerges and progresses at later stages, which were less represented in the dataset.

4.3 Encoding of Categorical Variables

The categorical variables in the dataset were encoded to enable their use in ML models. Specifically, the following variables were treated as categorical: pavement age (A), M&R treatment, BM, SM, SG, DT, DL. These variables were converted into numerical form using one-hot encoding, whereby each category was represented as a binary feature. In addition, the moisture damage classes (no, low, moderate, and high stripping) were encoded using label encoding to serve as the target variable in the ML models. All remaining variables were retained as numerical inputs without transformation.

5. Model Development

To address the issue of data imbalance, modeling framework was intentionally structured as a two-stage prediction system. In the first stage, a binary RF classification model was developed to predict the occurrence of moisture damage (damage versus no damage). Subsequently, a second RF model was developed using only the 1,303 observations exhibiting moisture damage to predict the severity level (low, moderate, or high). By isolating the severity prediction task from the dominant no-damage class, the second model avoided the substantial imbalance that would have existed if all four classes had

been modeled simultaneously. This hierarchical modeling strategy was specifically adopted to improve model robustness and predictive capability for moisture-damaged sections.

5.1 Moisture Damage Detection Using Binary Classification Model

5.1.1 Model Overview

To develop this model, all the 8094 records were used but the moisture damage classes were aggregated into two categories: no damage and moisture damage (combining the low, moderate, and high classes). This resulted in a more balanced dataset (6791 no damage and 1303 with moisture damage). Accordingly, moisture damage was modeled as a function of key variables influencing it, including age class (A), maintenance and rehabilitation treatment (M&R), type of base material (BM), type of subbase material (SM), type of subgrade soil (SG), subsurface drainage type (DT), drainage location (DL), drainpipe diameter (DD), average annual daily truck traffic (AADTT), annual temperature (T), annual precipitation (P), annual humidity (H), asphalt thickness (AT), base thickness (BT), and subbase thickness (ST). The relationship is expressed as shown in Equation 1.

$$\text{Moisture Damage} = f(A, M\&R, BM, SM, SG, DT, DL, DD, AADTT, T, P, H, AT, BT, ST) \quad (1)$$

A RF classification model was employed to approximate this function. The dataset was split at the pavement-section level: approximately 80% of the sections (and all their records) were used for training and the remaining 20% of sections for testing, so that no section appeared in both sets.

5.1.2 Model Training

The RF algorithm incorporates several hyperparameters that can influence model performance and generalization capability. In this study, three key hyperparameters were considered:

a) Number of trees (N): the total number of decision trees constructed within the ensemble. Increasing the number of trees generally improves model stability but increases computational effort.

b) Maximum tree depth (D): the maximum allowable depth of each decision tree, which controls model complexity and helps prevent overfitting.

c) Minimum number of samples required for splitting (S): the minimum number of observations required to further split an internal node, thereby controlling tree growth and model complexity.

To identify the optimal hyperparameter configuration, a combination of grid search and five-fold cross-validation was employed. The hyperparameter search space included: $N \in [100, 200, 300]$, $D \in [\text{None}, 10, 20, 30]$, and $S \in [2, 5, 10]$. The grid search procedure systematically evaluated all possible combinations within the predefined search space.

The grid search was guided by a five-fold cross-validation procedure, in which the training dataset was partitioned into five folds using grouped cross-validation, with pavement section as the grouping variable. All records from a given section were therefore assigned to the same fold, so that no section appeared in both the training and validation data within an iteration. This prevents information about a section's repeated characteristics from leaking into validation. During each iteration, four folds were used for model training while the remaining fold was used for validation. This process was repeated five times so that each fold served once as the validation set. For each hyperparameter combination, the classification accuracy was calculated for all validation folds, and the average cross-validation accuracy was computed to provide a robust estimate of model performance.

The results indicated that the optimal hyperparameter combination corresponded to N = 100, D = None, and S = 2, which are the default settings used by the Random Forest algorithm. The corresponding average five-fold cross-validation accuracy was approximately 92%. The trained RF model achieved a training accuracy of 99%, indicating a strong fit to the training data. This high training accuracy is expected for a Random Forest with unrestricted tree depth, which closely fits the in-bag training records. The relevant indicator of generalization is that the section-grouped cross-validation accuracy (approximately 92%) and the accuracy on the unseen test sections (93%) are consistent, which does not indicate a substantial generalization gap.

Feature Importance is the impurity-based (mean decrease in impurity) feature importance of the Random Forest model, computed on the training data and normalized so that the contributions of all features sum to 100%. It indicates how much each input variable is used by the model in making its predictions. It does not represent statistical significance, causal influence, or the direction of a variable's effect. In addition, this measure can favor continuous or high-cardinality variables and can understate categorical variables whose importance is divided among many one-hot columns. A higher value of Feature Importance indicates that the variable is more significant. Figure 1 presents the Feature Importance of the input variables for the trained model, highlighting the 15 most influential features. It should be noted that the model included a total of 115 input features, including 8 numerical features and 107 categorical features. The large number of features is a result of the application of one-hot encoding to the categorical variables, where each category within a variable is represented as a separate binary feature.

As shown in Figure 1, asphalt layer thickness had the highest importance value in the developed model, indicating that the model relied heavily on this variable when predicting moisture damage occurrence. Similarly, traffic loading (AADTT), annual precipitation, annual temperature, and pavement age were among the most influential predictors. These findings are consistent with the established understanding that structural, traffic, environmental, and aging-related factors are associated with pavement deterioration.

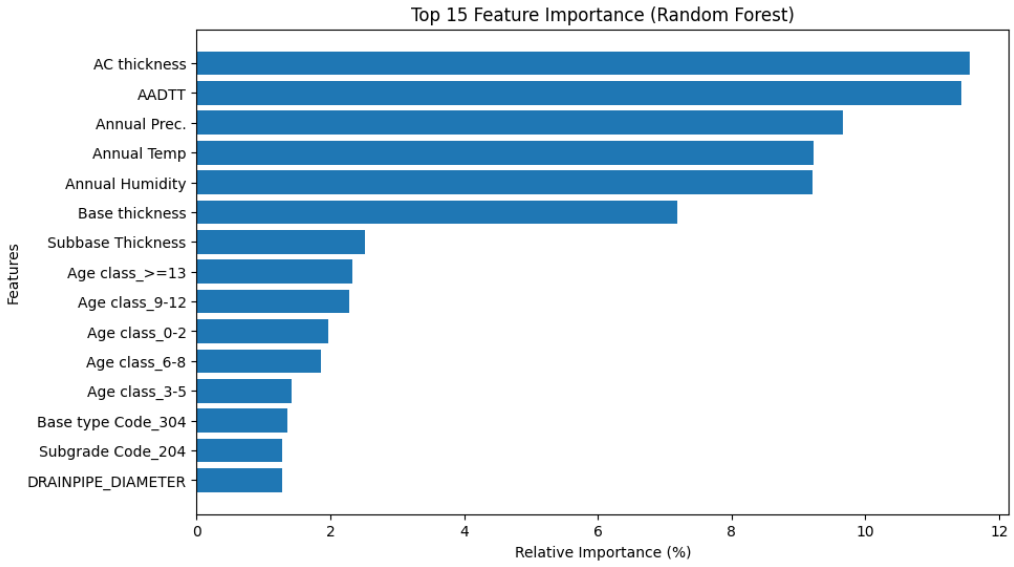

Fig. 1. Feature Importance of top 15 features in the trained detection model

5.1.3 Model Testing

Figure 2 presents the confusion matrix for the testing dataset. As shown, the model achieved a classification accuracy of approximately 97% for the no-damage class [1326/(1326+32) × 100] and 67% for the moisture-damage class [176/(176+85) × 100], resulting in an overall accuracy of approximately 93%. The model achieved a precision of 84.6%, recall of 67.4%, F1-score of 75.0%, and specificity of 97.6%. These results indicate that the model effectively distinguished between moisture-damaged and non-damaged

pavement sections while maintaining strong performance despite the class imbalance present in the dataset. To place these results in context, a classifier that always predicted the majority (no-damage) class would achieve 83.9% accuracy on the test set, compared with 92.8% for the developed model. Because accuracy is influenced by class imbalance, class-sensitive metrics are also reported in Table 1: balanced accuracy of 82.5%, a Matthews correlation coefficient (MCC) of 0.715, and Cohen's kappa of 0.709. The damage recall of 67.4% indicates that 85 of the 261 damaged test records (32.6%) were not detected. The model may therefore under-detect moisture damage, and its predictions should be used to prioritize verification activities such as coring rather than to exclude sections from inspection.

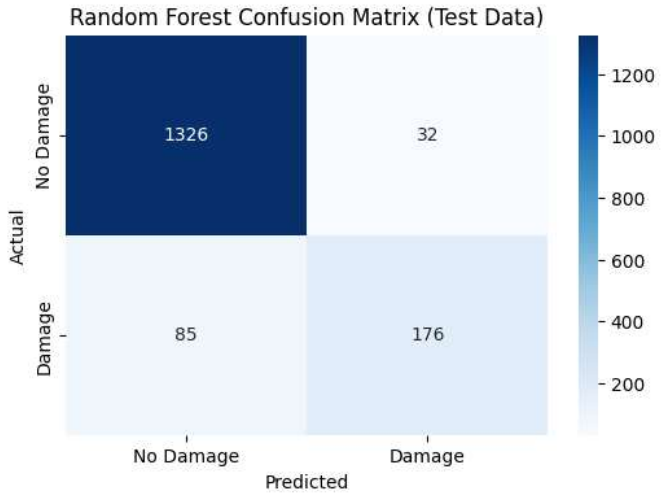

Fig. 2. Confusion matrix for testing data of the detection model

5.2 Moisture Damage Severity Using Multi-Class Classification Model

5.2.1 Model Overview

To develop the severity model, only a subset of the dataset was utilized, consisting exclusively of records exhibiting moisture damage. Specifically, 1303 records were selected, including 321, 337, and 645 observations associated with low, moderate, and high

420 moisture damage, respectively. This approach was adopted to address the class imbalance
421 present in the original dataset and to ensure that the model focuses on distinguishing
422 between different severity levels of moisture damage without being influenced by the
423 majority no-damage class. A RF classification model was employed to perform multi-class
424 prediction of moisture damage severity. In this model, the output was defined as a
425 categorical variable with three classes representing severity levels (low, moderate, and
426 high), while the model inputs were consistent with those defined in Equation 1. Similar to
427 the previous model, 80% of the records were used for training, while the remaining 20%
428 were reserved for testing the developed model.

429 **5.2.2 Model Training**

430 The model was trained using a RF classification algorithm. To optimize model
431 performance, a grid-search approach combined with five-fold cross-validation was
432 employed. The hyperparameter tuning process focused on the same three key parameters
433 (N, D, and S). Multiple combinations of these hyperparameters were systematically
434 evaluated using five-fold cross-validation, where the training dataset was partitioned into
435 five section-grouped folds (grouped by pavement section), and each fold was used once for
436 validation. The results indicated that the optimal hyperparameter configuration
437 corresponded to N = 100, D = None, and S = 2. The corresponding average cross-validation
438 accuracy was 76% across the five folds. Therefore, these hyperparameter values were
439 retained for the final model development.

440 Figure 3 presents the Feature Importance of the input variables for the trained model,
441 highlighting the 15 most influential features. As shown, the annual temperature had the
442 highest importance value within the severity model, indicating that the model relied heavily

on this variable when distinguishing among severity levels. Traffic loading (AADTT) also had a high importance value within the model. In addition, moisture-related environmental factors, particularly annual precipitation and humidity, as well as the asphalt layer thickness were among the highest-ranked predictors.

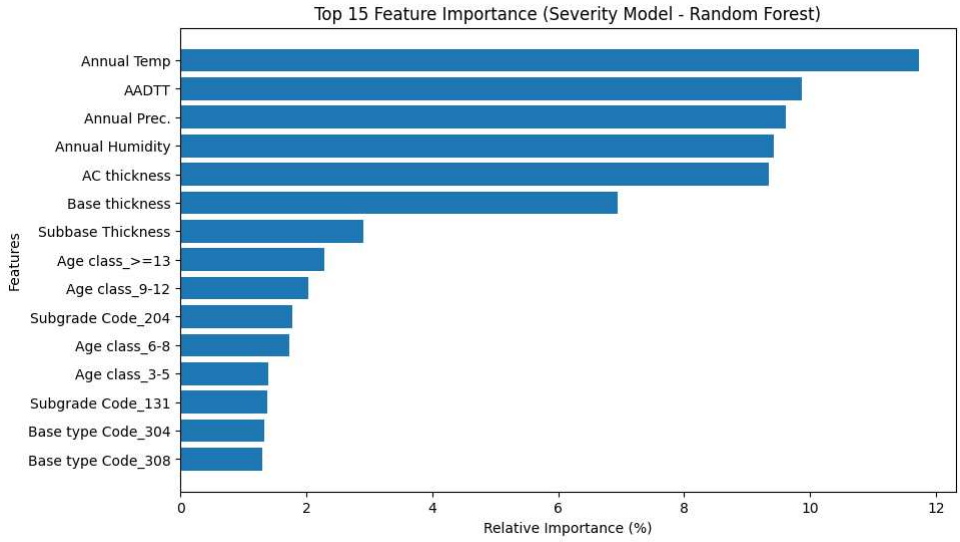

Fig. 3. Feature Importance of top 15 features in the trained severity model

5.2.3 Model Testing

Figure 4 presents the confusion matrix for the testing dataset. As shown, the model achieved a classification recall of approximately 86% for the low severity class [55/(55+8+1)×100], 60% for the moderate severity class [41/(41+10+17)×100], and 91% for the high severity class [117/(117+3+9)×100], with an overall accuracy of approximately 82%. The low severity class achieved a precision of 80.9% and an F1-score of 83.3%, while the moderate severity class achieved a precision of 70.7% and an F1-score of 65.1%. The high severity class achieved the highest performance, with a precision of 86.7% and an F1-score of 88.7%. The results indicate that the model performed best in identifying high-severity moisture damage, whereas the moderate severity class exhibited

the lowest performance due to confusion with adjacent classes (low and high). This behavior is expected given the inherent similarity between intermediate severity levels. Overall, the classification performance is considered satisfactory for multi-class severity prediction, particularly given the increased complexity of distinguishing between closely related damage levels. A majority-class classifier would achieve 49.4% accuracy on the severity test set, compared with 81.6% for the developed model. Class-sensitive metrics (Table 1) show a macro-F1 of 79.0%, a balanced accuracy of 79.0%, and a Cohen's kappa of 0.705. The lower recall of the moderate class (60.3%) reflects confusion with adjacent classes: of the 68 moderate test records, 10 were predicted as low and 17 as high.

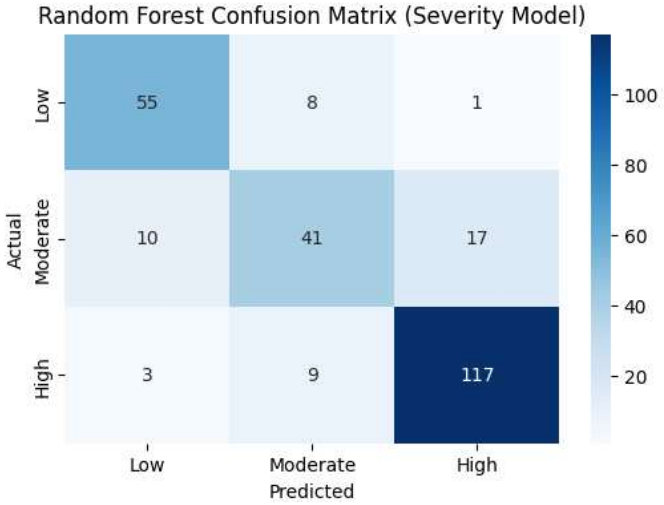

Fig. 4. Confusion matrix for testing data of the severity model

Table 1. Class-sensitive performance metrics of the occurrence and severity models on the test data (unseen sections).

Metric	Occurrence Model	Severity Model
Overall accuracy	92.8%	81.6%
Majority-class baseline accuracy	83.9%	49.4%
Balanced accuracy	82.5%	79.0%
Macro-F1	-	79.0%
MCC	0.715	-
Cohen's kappa	0.709	0.705

Damage recall / false-negative rate	67.4%/32.6%	-
Recall (low / moderate / high)	-	85.9%/60.3%/90.7%

6. Decision-Making Tool

The two developed ML models were integrated into a web-based application, named PavStrip, to function as a practical decision-support tool for transportation agencies. The PavStrip application was developed using the Python programming language and deployed through the Hugging Face Spaces platform. The web-based graphical user interface was implemented using the Gradio framework, which enables interactive communication between users and the prediction models. The trained RF models were integrated into the Python backend and connected to the user interface, allowing users to input project-specific traffic, environmental, structural, and drainage characteristics and receive real-time predictions of moisture damage occurrence and severity. This deployment framework provides a practical and user-friendly decision-support tool that can be readily accessed through a web browser by transportation agencies and pavement practitioners.

6.1 Tool Input

Figure 5 presents the graphical interface of the PavStrip tool, illustrating the structure and layout of the input parameters. The interface is organized to allow users to systematically enter the inputs related to traffic loading, M&R treatments, environmental conditions, pavement structure, and drainage conditions. The input parameters are defined for multiple analysis periods, enabling users to provide time-dependent data corresponding to different stages of the pavement life cycle. For each analysis period, the user inputs the relevant values, allowing the tool to evaluate the progression of moisture damage over time. To ensure the reliability and consistency of predictions, all input values are

constrained within predefined ranges derived from the observed dataset used in model development.

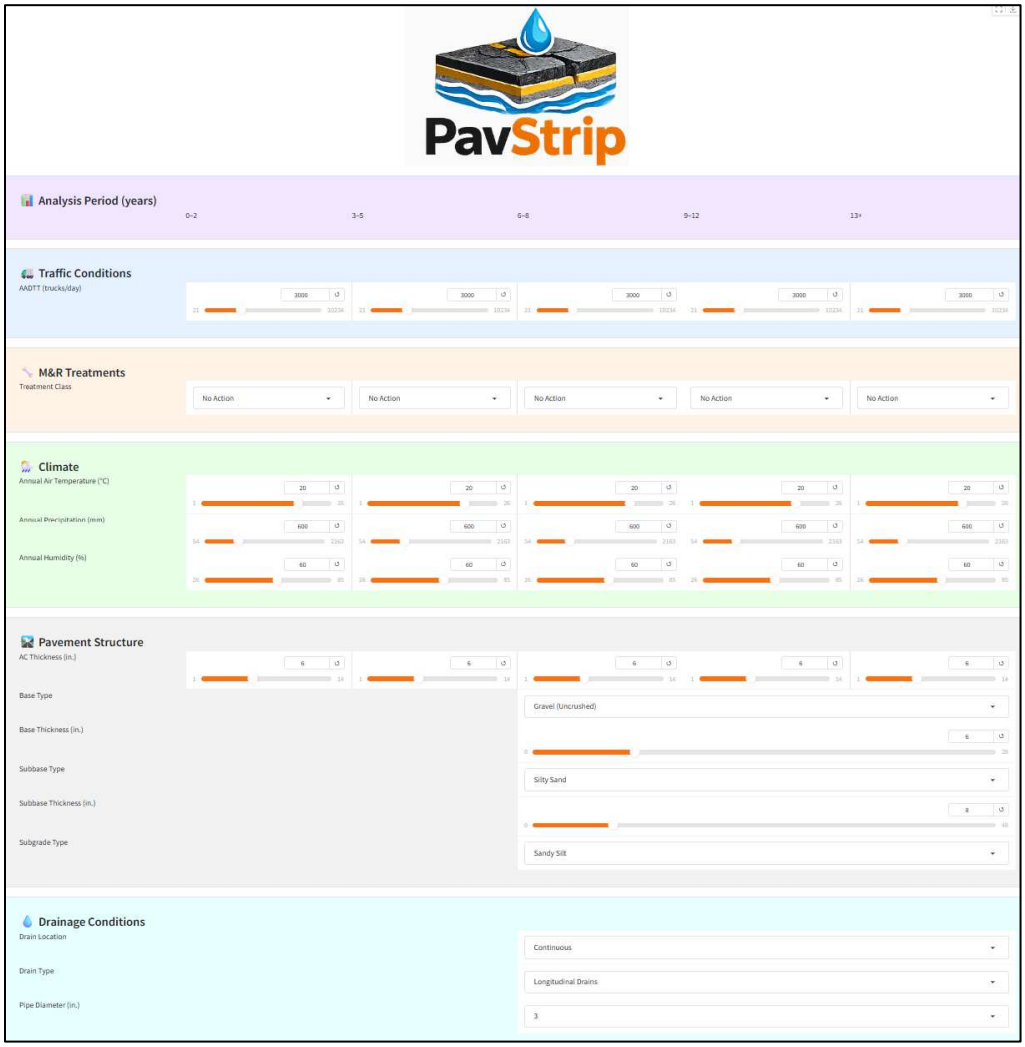

Fig. 5. User interface of the PavStrip tool illustrating input parameter structure

6.2 Tool Output

Upon user input, the tool sequentially evaluates the likelihood of moisture damage occurrence and, if present, predicts its severity across multiple analysis periods. The predicted damage levels are presented in a compact visual format, see Figure 6, facilitating rapid interpretation of damage progression. This design allows transportation

practitioners to assess moisture damage risk effectively and to support informed decision-making regarding M&R strategies in a user-friendly and efficient manner.

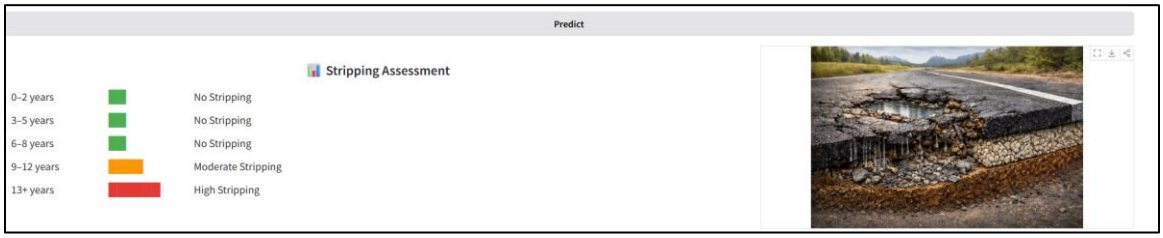

Fig. 6. User interface of the PavStrip tool illustrating tool output

6.3 Illustrative Application of PavStrip

To demonstrate the practical application of PavStrip, consider a transportation agency designing a new asphalt pavement section. The agency enters the project-level characteristics into the application (shown in Figure 5). Based on these inputs, PavStrip predicts the occurrence and severity of moisture damage over multiple service-life periods. Assume that the model generates the results in Figure 6:

- Age 0-2 years: No stripping
- Age 3-5 years: No stripping
- Age 6-8 years: No stripping
- Age 9-12 years: Moderate stripping
- Age ≥13 years: High stripping

These results indicate that moisture-related deterioration is unlikely to occur during the first eight years of service but is expected to emerge between years 9 and 12 and continue progressing thereafter. Based on this prediction, the transportation agency may elect to implement preventive maintenance treatments, such as a thin overlay or surface sealing during years 6-8, to reduce moisture infiltration and potentially delay the onset of moisture damage.

Alternatively, if the agency decides not to perform preventive maintenance during years 6-8, the PavStrip predictions can still provide valuable guidance for pavement monitoring and inspection planning. Rather than conducting costly coring operations or detailed field investigations throughout the pavement service life, the agency can focus moisture-damage assessment activities during the period where damage is predicted to initiate (years 9-12). For example, coring or other diagnostic evaluations may be scheduled during this period to verify the presence and extent of stripping. If moderate stripping is confirmed, corrective maintenance actions such as localized patching, milling and replacement of affected areas, or other appropriate rehabilitation measures can then be implemented before the pavement progresses to the high-damage condition predicted after year 13. In this way, PavStrip supports proactive, data-driven decision-making by helping agencies optimize both maintenance timing and inspection resources while reducing unnecessary testing during periods when moisture damage is not expected to occur.

PavStrip application is intended to be made publicly available to transportation agencies, practitioners, and researchers. Furthermore, as additional data become available through future releases of the LTPP database, the underlying dataset may be expanded and the machine learning models retrained. Such updates are expected to improve model robustness, predictive accuracy, and applicability for moisture damage prediction and pavement management decision-making.

7. Summary and Conclusions

Moisture damage is a significant distress that affects the overall performance of asphalt pavements. Moisture-induced damage has detrimental consequences on the long-term

549 performance, durability, sustainability of asphalt pavements. The conventional method to
550 detect moisture damage in asphalt pavements is through core extraction. Yet, coring is
551 destructive, time-consuming, costly, and labor-intensive prompting a shift towards NDE
552 methods like GPR, Rolling Wheel Deflectometer, Traffic Speed Deflection Device, and
553 advanced thermal imaging. Yet, NDE methods remain largely reactive and do not provide
554 predictive capabilities limiting their ability to implement proactive design and
555 maintenance strategies. Consequently, this study addressed this gap through extracting
556 8094 records from the LTPP database, and using these records to train two machine
557 learning–based classification models based on the RF Algorithm. The first model
558 predicted the occurrence of moisture damage (present or absent) with an overall testing
559 accuracy of 93%, while the second model predicted the severity level of moisture damage
560 (low, moderate, or high) with an overall testing accuracy of 82%. Feature importance
561 analysis indicated that environmental variables and traffic loading, play a dominant role
562 in predicting both the occurrence and severity of moisture damage. It was also indicated
563 that the pavement structural capacity, represented by asphalt thickness contributes
564 significantly to predicting moisture damage in asphalt pavements. Eventually, the two
565 developed RF models were integrated into a hierarchical (two-stage) prediction
566 framework and implemented within a web-based application. This integration enables
567 transportation agencies and practitioners to effectively utilize the predictive capabilities
568 of the models without requiring prior expertise in machine learning, thereby supporting
569 data-driven decision-making in pavement management and design.

570 **8. Future Work**

While the proposed framework demonstrated promising performance in predicting the occurrence and severity of moisture damage, several opportunities exist for future improvements as follows:

- In this study, moisture damage was inferred from raveling and pothole distresses, which are widely recognized manifestations of moisture-induced damage in asphalt pavements. However, future studies could incorporate direct indicators of moisture damage obtained through field coring, laboratory testing, dielectric measurements, thermal imaging, GPR, or other NDE techniques.
- In this study, moisture damage severity classes were established using percentile-based thresholds derived from the statistical distribution of pothole and raveling measurements in the LTPP database. Future research should focus on establishing engineering-based severity thresholds to further enhance the practical interpretation of the predicted severity categories and facilitate broader implementation of the proposed framework within pavement management systems.
- Future studies should investigate the incorporation of asphalt mixture characteristics, such as binder grade, air void content, aggregate mineral composition, anti-stripping additives, mixture type, etc.
- Future research should investigate the incorporation of more detailed climatic indicators, such as freeze-thaw cycles, rainfall intensity, extreme weather events, and the duration of moisture exposure.

- Although the LTPP program provides a large and diverse dataset covering a wide range of climatic, traffic, and pavement conditions across the United States, future studies should evaluate the transferability and generalizability of the developed models using independent pavement datasets. External validation using data from state agencies and other pavement monitoring programs would further strengthen confidence in the model predictions and support broader implementation of the framework.

- M&R treatment timing: M&R treatments were included as model inputs whenever they were applied before the distress survey date within the survey year. Only the very few treatments dated after the survey date were excluded, to avoid using post-outcome information. Because treatments were represented by type group and year-level timing, the effects of specific treatment types, treatment timing, and treatment effectiveness on distress progression were not modeled explicitly. Future studies should use event-level maintenance data to examine these effects.

- Future studies should apply nested section-grouped cross-validation to separate hyperparameter tuning from performance estimation.

- Future studies should include chronological (out-of-time) evaluation, training on earlier observations and testing on later ones.

- Future studies should evaluate geographic transferability using state- or climate-region holdout validation.

- Future studies should report uncertainty estimates at the section level, such as bootstrap confidence intervals over sections, to account for repeated records from the same section.
- Future studies should compute grouped permutation importance (or SHAP values) on held-out sections, with one-hot columns grouped by their parent variable, to obtain importance estimates that are less affected by variable cardinality and one-hot splitting.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of Data and Materials

The dataset supporting the conclusions of this article is available in the Long-Term Pavement Performance (LTPP) Standard Data Release (SDR) repository, https://infopave.fhwa.dot.gov/Data/StandardDataRelease.

References

1. Little DN, Jones DR. Chemical and mechanical processes of moisture damage in hot-mix asphalt pavements. In: Moisture Sensitivity of Asphalt Pavements – A National Seminar; 2003. California Department of Transportation; Federal Highway Administration; National Asphalt Pavement Association; California Asphalt Pavement Alliance; Transportation Research Board.
2. Q. Lu, Investigation of conditions for moisture damage in asphalt concrete and appropriate laboratory test methods, University of California, Berkeley, 2005.

3. Hicks, R. G. NCHRP Synthesis 175: Moisture Damage in Asphalt Concrete. Transportation Research Board, Washington, D.C., 1991.
4. Elseifi MA, Mousa MR, Gaspard K. Impact of the great flood of 2016 on the asphaltic concrete road infrastructure in Louisiana. Transp Res Rec. 2022;2676(8):463–474.
5. Sripath Technologies. How water affects asphalt sustainability and durability. https://sripath.com/water-affects-asphalt-sustainability-and-durability/. Accessed 19 Jun 2026.
6. Jin, D., Xin, K., Yin, L., Mohammadi, S., Cetin, B., & You, Z. (2024). Performance of rubber modified asphalt mixture with tire-derived aggregate subgrade. Construction and Building Materials, 449, 138261.
7. Jin, D., Wang, J., You, L., Ge, D., Liu, C., Liu, H., & You, Z. (2021). Waste cathode-ray-tube glass powder modified asphalt materials: Preparation and characterization. *Journal of Cleaner Production*, *314*, 127949.
8. Jin, D., Yin, L., Nedrich, S., Boateng, K. A., & You, Z. (2024). Resurface of rubber modified asphalt mixture with stress absorbing membrane interlayer: From laboratory to field application. *Construction and Building Materials*, *441*, 137452.
9. Jin, D., Mohammadi, S., Xin, K., Yin, L., & You, Z. (2024). Laboratory performance and field demonstration of asphalt overlay with recycled rubber and tire fabric fiber. *Construction and Building Materials*, *438*, 136941.
10. Jin, D., Meyer, T. K., Chen, S., Boateng, K. A., Pearce, J. M., & You, Z. (2022). Evaluation of lab performance of stamp sand and acrylonitrile styrene acrylate

waste composites without asphalt as road surface materials. *Construction and Building Materials*, *338*, 127569.

11. Albayati, A. H., Al-Mosawe, H. M., Allawi, A. A., & Oukaili, N. (2018). Moisture susceptibility of sustainable warm mix asphalt. *Advances in civil engineering*, *2018*(1), 3109435.
12. Al-Tameemi, A. F., Wang, Y., Albayati, A., & Haynes, J. (2019). Moisture susceptibility and fatigue performance of hydrated lime–modified asphalt concrete: Experiment and design application case study. *Journal of Materials in Civil Engineering*, *31*(4), 04019019.
13. Omar HA, Yusoff NIM, Mubaraki M, Ceylan H. Effects of moisture damage on asphalt mixtures. J Traffic Transp Eng (Engl Ed). 2020;7(5):600–628.
14. C. Gorkem, B. Sengoz, Predicting stripping and moisture induced damage of asphalt concrete prepared with polymer modified bitumen and hydrated lime, Constr. Build. Mater. 23 (2009) 2227–2236. https://doi.org/10.1016/j.conbuildmat.2008.12.001.
15. Rmeili, E., and T. Scullion. Detecting Stripping in Asphalt Concrete Layers Using Ground Penetrating Radar. Transportation Research Record, Vol. 1568, No. 1, 1997, pp. 165–174.
16. Celaya, M., and S. Nazarian. Stripping Detection in Asphalt Pavements with Seismic Methods. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2005, No. 1, pp. 64–74, 2007.
17. C. Gorkem, B. Sengoz, Predicting stripping and moisture induced damage of asphalt concrete prepared with polymer modified bitumen and hydrated lime,

Constr. Build. Mater. 23 (6) (2009) 2227–2236, https://doi.org/10.1016/j.conbuildmat.2008.12.001.

18. C. Chen, S. Chandra, Y. Han, H. Seo, Deep learning-based thermal image analysis for pavement defect detection and classification considering complex pavement conditions, Remote Sens 14 (1) (2022) 106, https://doi.org/10.3390/rs14010106.
19. Elseifi MA, Dhakal N, Abohamer H, Ma Y, Zihan ZUA. Cost-effective detection and repair of moisture damage in pavements. 2021.
20. Mousa MR, Mousa SR, Hassan M, Carlson P, Elnaml IA. Predicting the retroreflectivity degradation of waterborne paint pavement markings using advanced machine learning techniques. Transp Res Rec. 2021;2675(9):483–494.
21. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science. 2015;349(6245):255-260. https://doi.org/10.1126/science.aaa8415.
22. Kashani A. T., Shariat-Mohaymany A., Ranjbari A. A Data Mining Approach to Identify Key Factors of Traffic Injury Severity. Preliminary Communication Safety and Security of Traffic, Vol. 23, No. 1, 2011, pp. 11–17.
23. Breiman L. Random forests. Machine Learning. 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324
24. Miller JS, Bellinger WY. Distress Identification Manual for the Long-Term Pavement Performance Program (Fifth Revised Edition). Report No. FHWA-HRT-13-092. Federal Highway Administration, McLean, VA; 2014. https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/13092/13092.pdf

704 25. H.J. Fromm, The mechanisms of asphalt stripping from aggregate surfaces,
705 Proc. Assoc. Asphalt Paving Technol. 43 (1974) 191–223.