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.

Toward Climate-Resilient Road Infrastructure: A Machine Learning Approach for Predicting Moisture Damage in Asphalt Pavements

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: July 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10093583/v1

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

Additional Declarations: No competing interests reported.

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, are primarily reactive and lack predictive capability, limiting their effectiveness 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 influencing moisture damage. Similarly, the pavement structural capacity contributed significantly to improving resistance against moisture-induced deterioration. The developed models were integrated into a web-based decision-support tool, enabling transportation agencies to perform predictive assessments and implement proactive maintenance strategies.

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].

Numerous studies [6] have 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 [7]. However, despite these advancements, limited studies have focused on the development of predictive tools capable of predicting the initiation and progression of moisture damage over time. In particular, there remains a gap in the ability to predict when moisture damage is likely to occur after construction and to quantify its expected severity throughout the pavement's service life. 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 time-dependent predictions over a 20-year post-construction period, enabling the estimation of when moisture damage is likely to initiate and how its severity is expected to evolve under varying project conditions. 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 Ground Penetrating Radar (GPR) to detect moisture damage in asphalt pavement layers [8]. 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 [9] 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 [10–11] 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 [12]. 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 [13]. 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 [14]. However, individual decision trees may suffer from limited predictive performance when applied to complex, high-dimensional datasets. To overcome this limitation, ensemble methods such as the Random Forest algorithm have been introduced. Random Forest constructs multiple decision trees using bootstrap sampling and aggregates their predictions, resulting in improved model stability, reduced overfitting, and enhanced predictive accuracy. This makes it particularly suitable for modeling complex interactions among pavement, traffic, and environmental variables. Yet, its use in predicting the occurrence and severity of moisture damage in asphalt pavements remains limited. Therefore, this study evaluates the capability of the Random Forest algorithm in predicting moisture damage occurrence and severity using Long-Term Pavement Performance (LTPP) data.

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, 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 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 over multiple years of service, starting from the year of construction (age zero) and extending forward. This resulted in a total of 8094 records. Each of the 8094 records represented a unique combination of pavement section and monitoring year forming

a section–year panel data structure. 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 m².
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 m².
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 [15]. To quantify moisture damage in this study, statistical analysis of pothole area (PA) and raveling area (RA) was conducted to identify representative threshold values. Prior to statistical analysis, observations with zero values were excluded to avoid biasing the percentile calculations toward non-damaged conditions. Percentile thresholds were then computed using only non-zero values to better represent the severity distribution of observed distress. For pothole area (PA), the 25th percentile (PA25) was 0.045 m², and the 50th percentile (PA50) was 0.114 m². For raveling area (RA), the 25th percentile (RA25) was 34 m², and the 50th percentile (RA50) was 162.6 m². 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 [0 < PA< PA25] AND [0 < RA< RA25]
- Moderate moisture damage: if [PA25 ≤ PA < PA50] OR [RA25 ≤ RA < RA50]

- 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, the modeling approach was divided into two sequential stages. In the first stage, a binary classification model was developed to detect the presence of moisture damage. In the second stage, a separate multi-class classification model was developed to determine the severity of moisture damage.

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 Eq. 1.

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

A Random Forest (RF) classification model was employed to approximate this function. About 80% of the records were used for training, while the remaining 20% were used to test the developed model.

5.1.2 Model Training

Default hyperparameter settings were used to train the Random Forest model. The use of default hyperparameter provided a robust and computationally efficient approach while allowing the evaluation of the inherent predictive capability of the algorithm without additional model complexity. The trained Random Forest model achieved a training accuracy of 99%, indicating a strong fit to the training data. However, model performance was primarily evaluated using the testing dataset to assess generalization capability, as described in the following section.

Relative Importance is a statistical measure defined as the percentage contribution of each input variable to the model, with the total contribution of all variables summing to 100%. A higher value of Relative Importance indicates that the variable is more significant. Figure 1 presents the Relative 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 Fig. 1, the asphalt layer thickness demonstrated high importance, emphasizing the significance of structural capacity in mitigating moisture-induced damage and enhancing pavement durability. Traffic loading, represented by AADTT, also emerged as one of the most influential variables affecting the prediction of moisture damage, highlighting its critical role in accelerating pavement deterioration under repeated loading conditions. In addition, environmental factors, particularly annual precipitation and temperature, contributed substantially to the model, reflecting their strong influence on the initiation and progression of damage mechanisms. Additionally, age-related variables, particularly higher age classes, showed notable importance, confirming that older pavements are more susceptible to moisture damage.

5.1.3 Model Testing

Figure 2 presents the confusion matrix for the testing dataset. As shown, the model achieved high classification accuracy for the no-damage class of 97% ($\frac{1326}{1326+32} \times 100$), and for the damage class of 70% ($\frac{176}{176+85} \times 100$), with an overall accuracy of 93% demonstrating strong performance in detecting moisture damage.

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

5.2.2 Model Training

The model was trained using a Random Forest classification algorithm considering the default hyperparameter settings. Figure 3 presents the Relative Importance of the input variables for the trained model, highlighting the 15 most influential features. As show, annual temperature demonstrated the highest importance, highlighting its critical role in accelerating damage progression. Traffic loading, represented by AADTT, also showed substantial importance, indicating its strong contribution to the worsening of damage severity under repeated loading. In addition, moisture-related environmental factors, particularly annual precipitation and humidity, contributed significantly to the model, reflecting their influence on moisture infiltration, accumulation, and the intensification of damage mechanisms. Furthermore, asphalt layer thickness exhibited notable importance, emphasizing the role of structural capacity in mitigating the progression and severity of moisture-induced damage.

5.2.3 Model Testing

Figure 4 presents the confusion matrix for the testing dataset. As shown, the model achieved a classification accuracy 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 results indicated that the model performed best in identifying high-severity moisture damage, while the moderate severity class exhibited the lowest accuracy 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.

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.

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.

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 Fig. 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.

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 performance, durability, sustainability of asphalt pavements. The conventional method to detect moisture damage in asphalt pavements is through core extraction. Yet, coring is destructive, time-consuming, costly, and labor-intensive prompting a shift towards NDE methods like Ground Penetrating Radar, Rolling Wheel Deflectometer, Traffic Speed Deflection Device, and advanced thermal imaging. Yet, NDE methods remain largely reactive and do not provide predictive capabilities limiting their ability to implement proactive design and maintenance strategies. Consequently, this study addressed this gap through extracting 8094 records from the LTPP database, and using these records to train two machine learning–based classification models based on the Random Forest Algorithm. The first model predicted the occurrence of moisture damage (present or absent) with an overall testing accuracy of 93%, while the second model predicted the severity level of moisture damage (low, moderate, or high) with an overall testing accuracy of 82%. Feature importance analysis indicated that environmental variables and traffic loading, play a dominant role in governing both the occurrence and severity of moisture damage. It was also indicated that the pavement structural capacity, represented by asphalt thickness contributes significantly to improving resistance against moisture-induced deterioration. Eventually, the two developed Random Forest models were 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.

Declarations

CRediT authorship contribution statement

Youssef Mousa: Formal analysis, Visualization, Investigation, Writing – original draft

Momen Mousa: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Supervision.

Egehan Koyuncu: Data curation, Software, Writing – original draft.

Mahasen Abdelhafez: Conceptualization, Methodology, Writing – review & editing.

Min Suh: Methodology, Writing – review & editing.

Iftekhar Basith: Methodology, Writing – review & editing.

Funding Declaration

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

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

3. Hicks RG (1991) NCHRP Synthesis 175: Moisture Damage in Asphalt Concrete. Transportation Research Board, Washington, D.C.
4. Elseifi MA, Mousa MR, Gaspard K (2022) Impact of the great flood of 2016 on the asphaltic concrete road infrastructure in Louisiana. Transp Res Rec 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. Omar HA, Yusoff NIM, Mubaraki M, Ceylan H (2020) Effects of moisture damage on asphalt mixtures. J Traffic Transp Eng (Engl Ed) 7(5):600–628
7. Gorkem C, Sengoz B (2009) Predicting stripping and moisture induced damage of asphalt concrete prepared with polymer modified bitumen and hydrated lime. Constr Build Mater 23:2227–2236. https://doi.org/10.1016/j.conbuildmat.2008.12.001
8. Rmeili E, Scullion T (1997) Detecting Stripping in Asphalt Concrete Layers Using Ground Penetrating Radar. Transp Res Rec 1568(1):165–174
9. Celaya M, Nazarian S (2005) Stripping Detection in Asphalt Pavements with Seismic Methods. Transportation Research Record: Journal of the Transportation Research Board, Vol. No. 1, pp. 64–74, 2007
10. Gorkem C, Sengoz B (2009) Predicting stripping and moisture induced damage of asphalt concrete prepared with polymer modified bitumen and hydrated lime. Constr Build Mater 23(6):2227–2236. https://doi.org/10.1016/j. conbuildmat.2008.12.001
11. Chen C, Chandra S, Han Y, Seo H (2022) Deep learning-based thermal image analysis for pavement defect detection and classification considering complex pavement conditions. Remote Sens 14(1):106. https://doi.org/10.3390/rs14010106
12. Elseifi MA, Dhakal N, Abohamer H, Ma Y, Zihan ZUA (2021) Cost-effective detection and repair of moisture damage in pavements
13. Mousa MR, Mousa SR, Hassan M, Carlson P, Elnaml IA (2021) Predicting the retroreflectivity degradation of waterborne paint pavement markings using advanced machine learning techniques. Transp Res Rec 2675(9):483–494
14. Kashani AT, Shariat-Mohaymany A, Ranjbari AA (2011) Preliminary Communication Saf Secur Traffic 23(1):11–17Data Mining Approach to Identify Key Factors of Traffic Injury Severity
15. Fromm HJ (1974) The mechanisms of asphalt stripping from aggregate surfaces, Proc. Assoc. Asphalt Paving Technol. 43 191–223

Figures

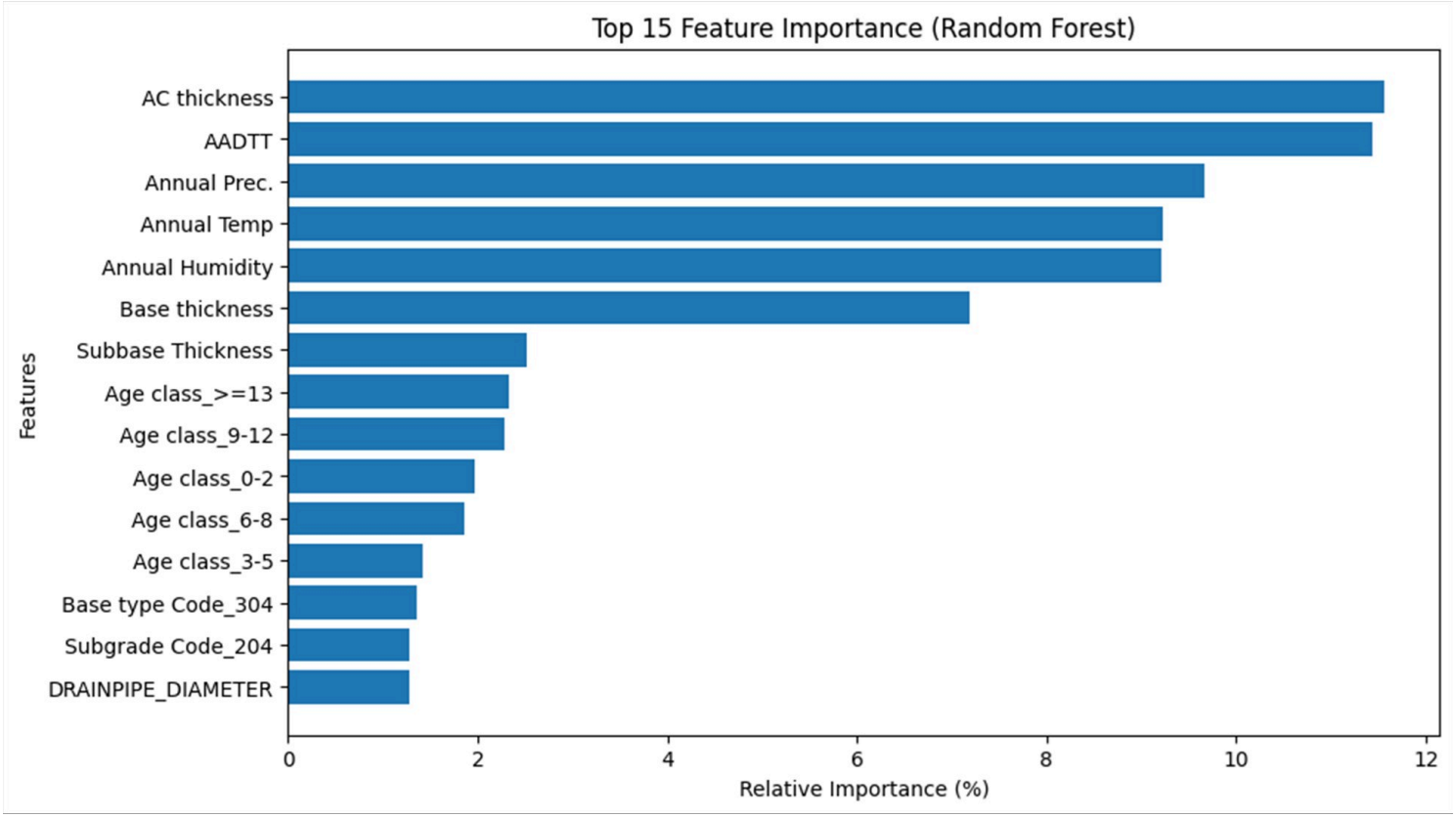


Figure 1

Realtive Important of top 15 features in the trained detection model

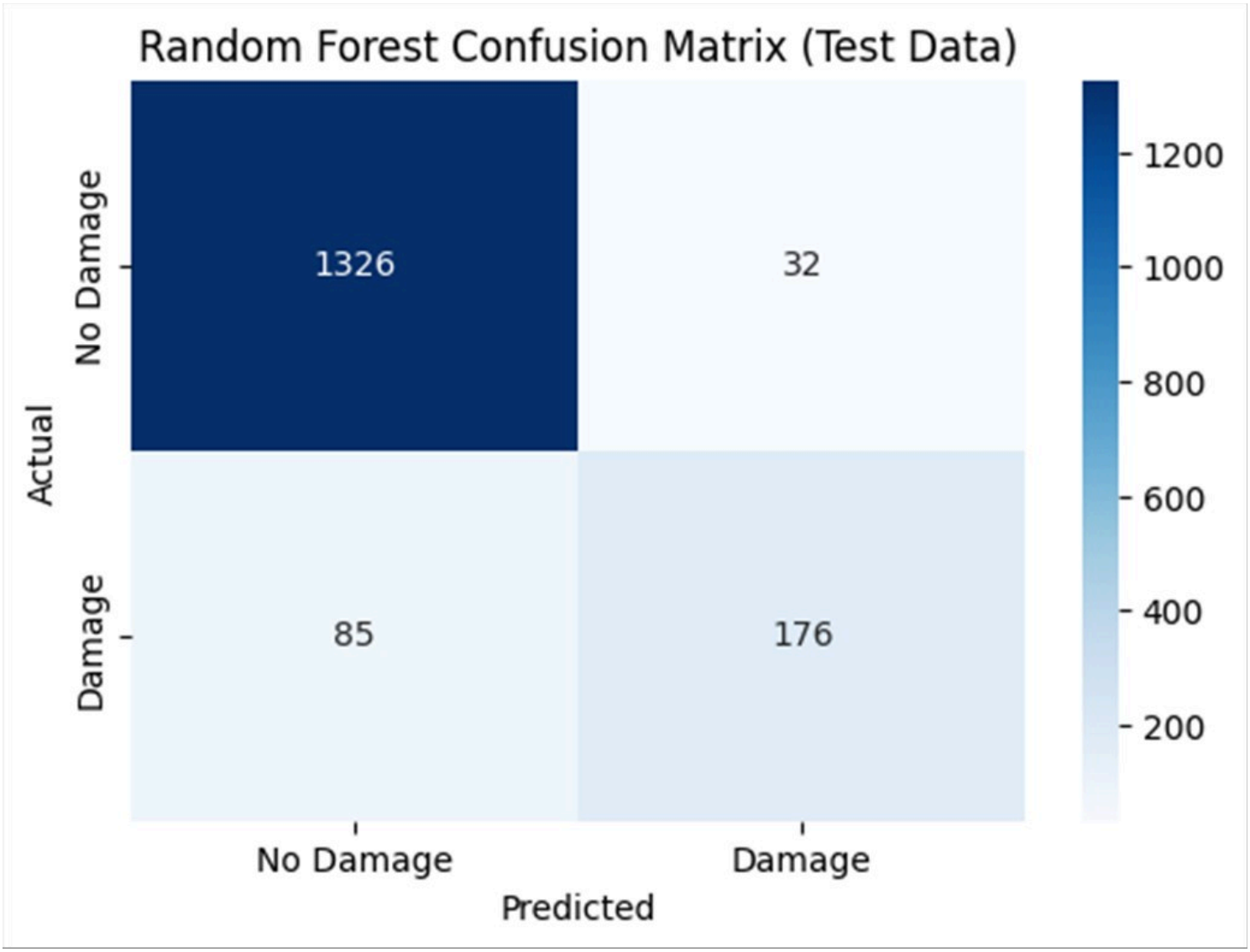


Figure 2

Confusion matrix for testing data of the detection model

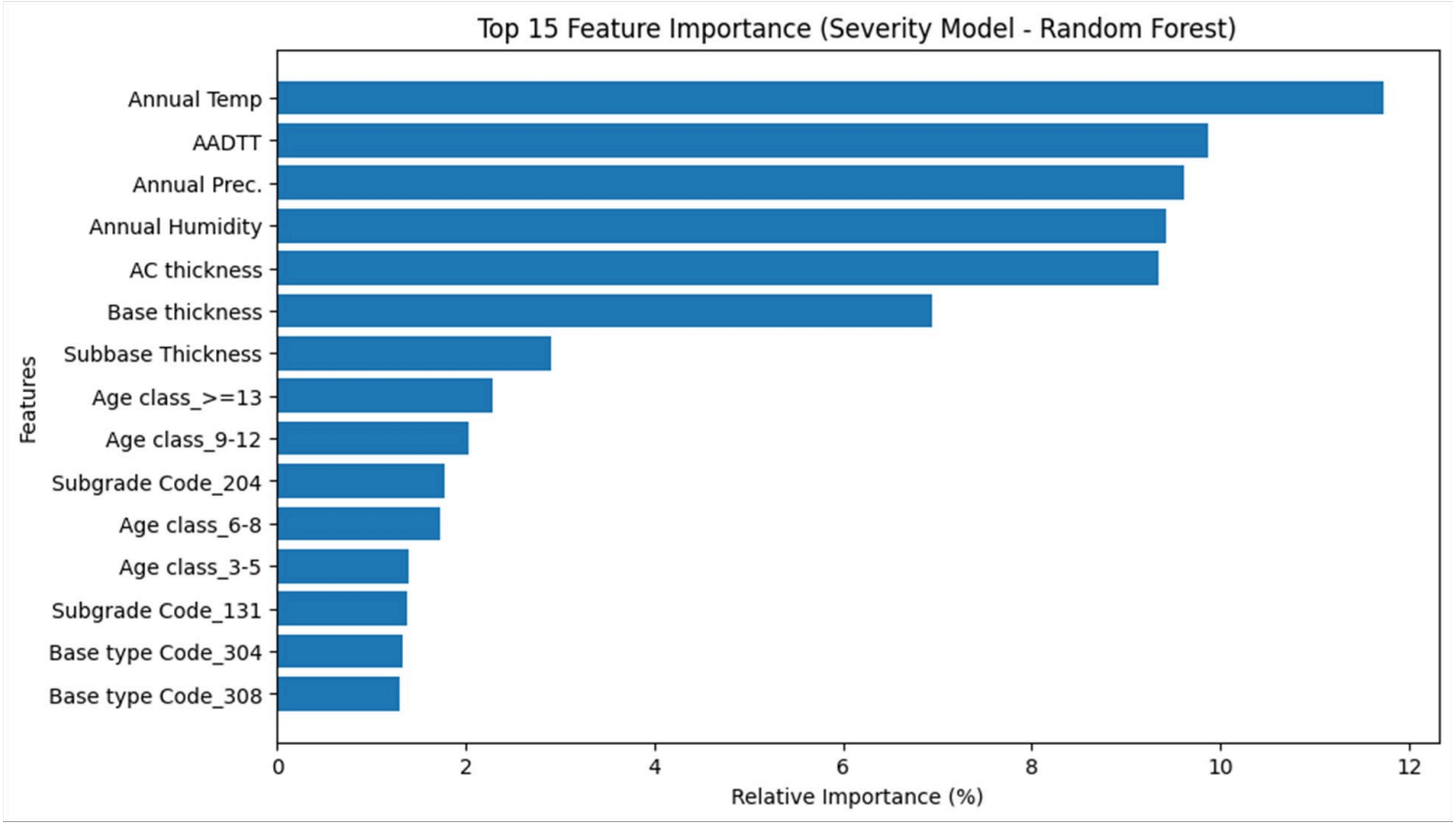


Figure 3

Realtive Important of top 15 features in the trained severity model

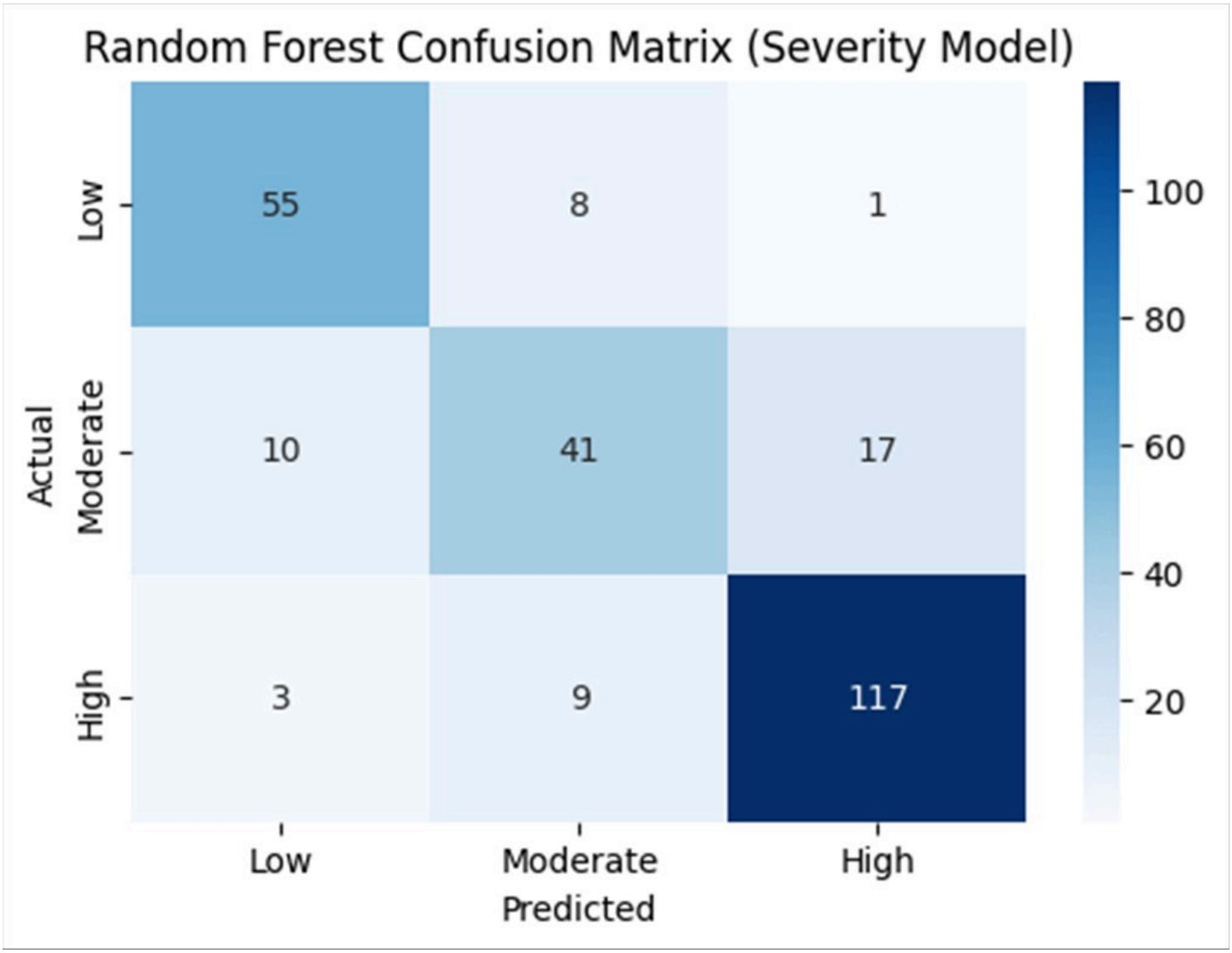


Figure 4

Confusion matrix for testing data of the severity model

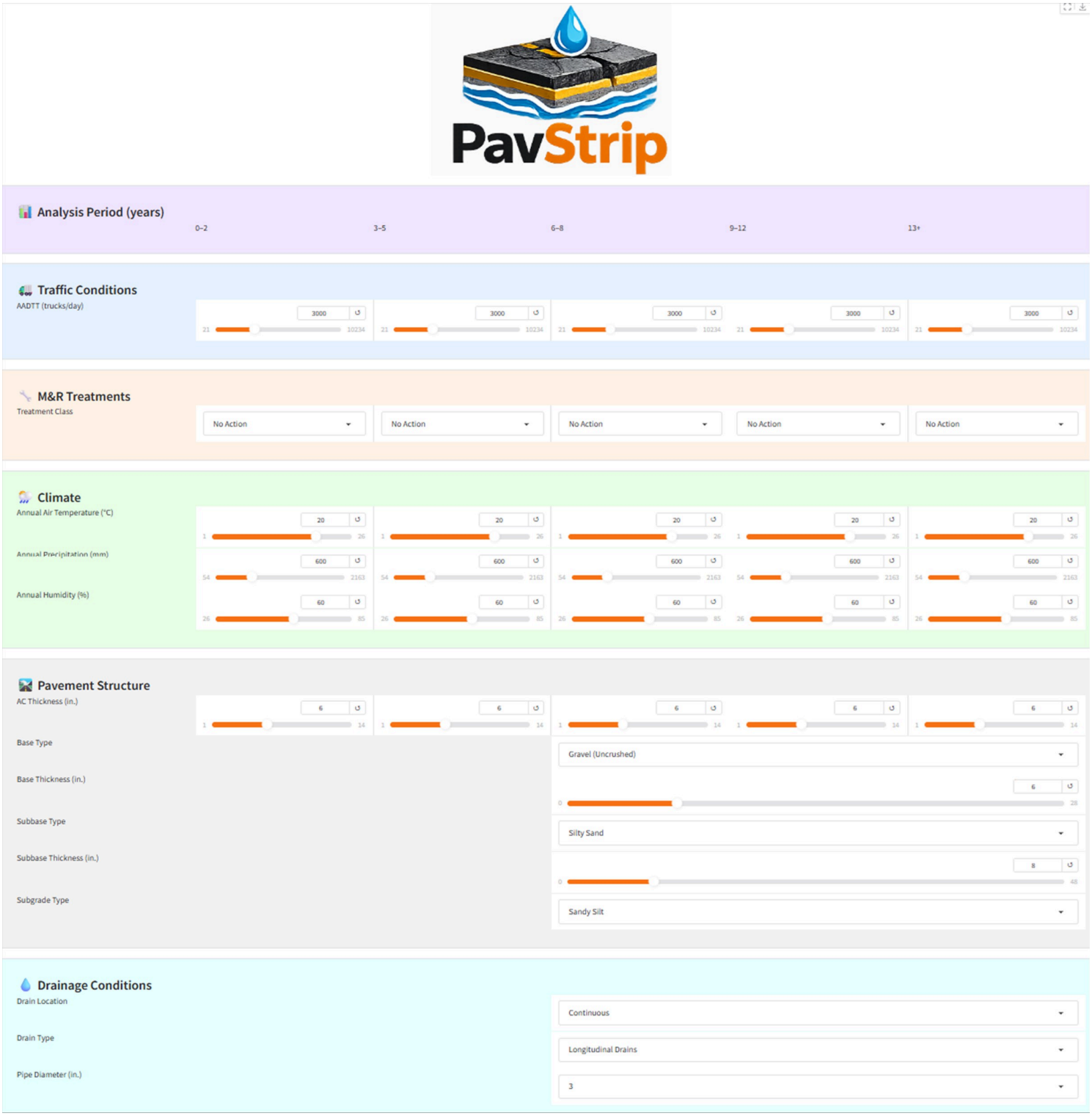


Figure 5

User interface of the PavStrip tool illustrating input parameter structure

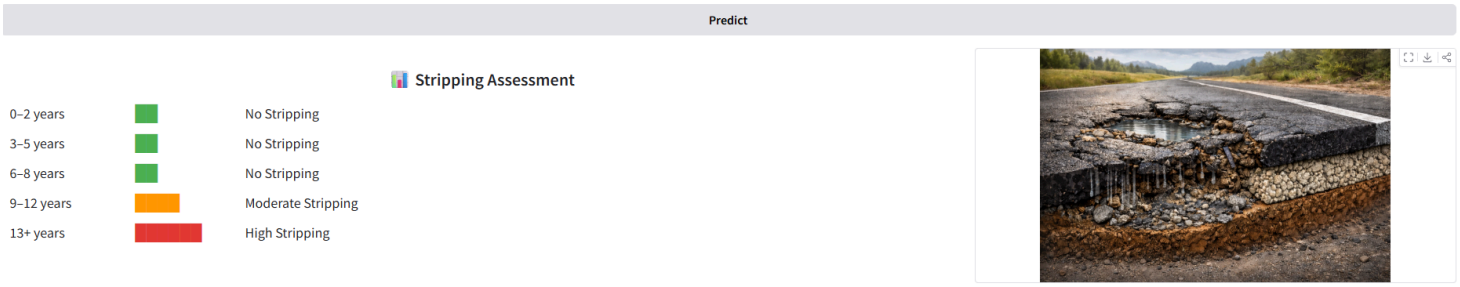


Figure 6

User interface of the PavStrip tool illustrating tool output