Gradient boosting decision trees to study laboratory and field performance in pavement management

Berangi, Mohammadjavad; Lontra, Bernardo Mota; Anupam, Kumar; Erkens, Sandra; Van Vliet, Dave; Snippe, Almar; Moenielal, Mahesh

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

RESEARCH ARTICLE

Gradient boosting decision trees to study laboratory and field performance in pavement management

Mohammadjavad Berangi[1] | **Bernardo Mota Lontra[2]** | **Kumar Anupam[1]** |
Sandra Erkens[1] | **Dave Van Vliet[2]** | **Almar Snippe[2]** | **Mahesh Moenielal[2]**

[1]Faculty of Civil Engineering and Geosciences, Delft University of Technology, South Holland, The Netherlands

[2]TNO, The Netherlands Institute for Applied Scientific Research, Delft, The Netherlands

Correspondence
Kumar Anupam, Building 23, Stevinweg 1, 2628CN, Delft, South Holland, The Netherlands.
Email: k.anupam@tudelft.nl

Abstract
Inconsistencies between performance data from laboratory-prepared and field samples have been widely reported. These inconsistencies often result in inaccurate condition prediction, which leads to inefficient maintenance planning. Traditional pavement management systems (PMS) do not have the appropriate means (e.g., mechanistic solutions, extensive data handling facilities, etc.) to consider these data inconsistencies. With the growing demand for sustainable materials, there is a need for more self-learning systems that could quickly transfer laboratory-based information to field-based information inside the PMS. The article aims to present a future-ready machine learning-based framework for analyzing the differences between laboratory and field-prepared samples. Developed on the basis of data obtained from field and laboratory data, the gradient-boosting decision trees-based framework was able to establish a good relationship between laboratory performance and field performance ($R^2_{\text{test}} > 80$ for all models). At the same time, the framework could also show more complex relationships that are often not considered in practice.

1 | INTRODUCTION

Pavement infrastructures are essential in facilitating economic growth for any country (Yao et al., 2020). The construction and maintenance of pavement infrastructures require substantial budget allocation and planning (Peraka & Biligiri, 2020). For example, more than $400 billion is invested globally each year in pavement construction and maintenance (Torres-Machí et al., 2015). In developed countries, more focus is now on maintaining the existing pavement infrastructures than constructing new ones, as the road networks are often already saturated. However, considering the limited budget available to road

agencies, a primary challenge is to optimize investment impacts on serviceability (Yao et al., 2024).

Road agencies generally determine appropriate maintenance strategies and budget allocation by predicting optimum pavement performance over time (Pan et al., 2011). The traditional approach to performance prediction is to appropriately choose the laboratory mix design method based on its empirical relationship with in-field experience (Polaczyk et al., 2021; Zaumanis et al., 2018; Zhou et al., 2021). However, the performance variation in laboratory mix design methods stems from differences in the evaluation process and the distinct construction process (e.g., compaction techniques; Gartner, 1989). Hence,

laboratory samples should be prepared to closely match field conditions (Iskender & Aksoy, 2012; Tia, 2005, 2005). Furthermore, care should also be taken to maintain consistency in compaction processes and other factors, such as temperature in both environments (Airey & Collop, 2016).

Laboratory conditions offer a more controlled environment with smaller material quantities and stricter controls over temperature (Airey & Collop, 2016). Therefore, the relationship between independent variables (e.g., properties of asphalt mixture) and dependent variables (e.g., functional performance indicator) could be significantly different in field conditions. Such a difference can be statistically represented by examining the variability or discrepancies between the independent and dependent variables, which may be intricate and manifest nonlinearly. Furthermore, the rapid change in mixing compositions resulting from the scarcity of raw materials, weather fluctuation, changes in traffic characteristics, and other technological advances are changing asphalt concrete beyond the narrow margins for which traditional mix design criteria might not hold true (Mousavi Rad et al., 2022; P. Pereira & Pais, 2017; Zaumanis et al., 2018). The research studies concluded that these aspects have not been well studied. In contrast, traditional pavement performance prediction models (PPPMs), such as mechanistic-empirical models, assume that the laboratory performance sufficiently reflects the field performance. These models often require advanced calibration methods due to the variations in materials, construction methods, and environmental conditions that can impact the performance of the pavement in the field (Dong et al., 2020).

With advancements in computational power, machine learning (ML) methods have been utilized intensively. Their ability to handle vast amounts of data enabled researchers to overcome some of the limitations inherent in traditional approaches. However, even in ML-based PPPMs, it is often presumed that laboratory performance sufficiently reflects field performance. Furthermore, ML models often struggle with the lack of enough interpretability of the results, even the sophisticated methods with improved predictive capabilities (Fan et al., 2023). The lack of interpretability can be a significant drawback in the practical application of ML-based pavement performance prediction models. Without a clear understanding of the model's prediction process, practitioners may be hesitant to trust the results and make decisions based on them. Considering the limitations highlighted, the objective and the scope of the research are presented in the following subsection.

1.1 | Research objective and scope

The objective of the research is to develop a future-ready ML-based pavement performance prediction framework to study the reliability of lab-prepared samples on the actual field performance. Before using the proposed framework to study the importance of the parameters of asphalt mixture on functional performance indicators, the validity of the models is first checked against multilinear regression (MLR), support vector machine (SVM), and random forest (RF). The proposed framework will provide better insights into pavement behavior in the field by using the information obtained in the laboratory with the end goal of enabling better-informed decision-making, optimizing maintenance strategies, and ultimately leading to cost savings and enhanced road safety.

To achieve the objective of the research, the scope of the research is identified as follows:

1. Comparing the performance of mixtures prepared in-lab and in-field conditions.
2. Developing the gradient boosting decision tree (GBDT) model for predicting functional performance indicators such as stiffness, fatigue resistance, rutting, and water sensitivity.
3. Validating the performance of the GBDT model with the MLR, SVM, and RF.
4. Improving the interpretability of the GBDT model using Shapley additive explanations (SHAP).
5. Evaluating hypotheses that are commonly deemed correct in pavement construction practices.

1.2 | Research framework

In order to have an easier understanding, the methodology of this research is divided into four stages (see Figure 1). In the first stage, data from both the laboratory and the field were collected. In Stage 2, the data from Stage 1 were preprocessed for Stage 3. In Stage 3, ML models were developed. In Stage 4, the developed models are utilized to evaluate the reliability of the lab-prepared samples on the field performance. In the following subsections, each of the stages is explained. It is noted that the research framework is based on the following assumptions:

1. The selected features adequately represent the key factors influencing pavement performance.
2. The collected data from the six construction projects are representative of broader industry practices.
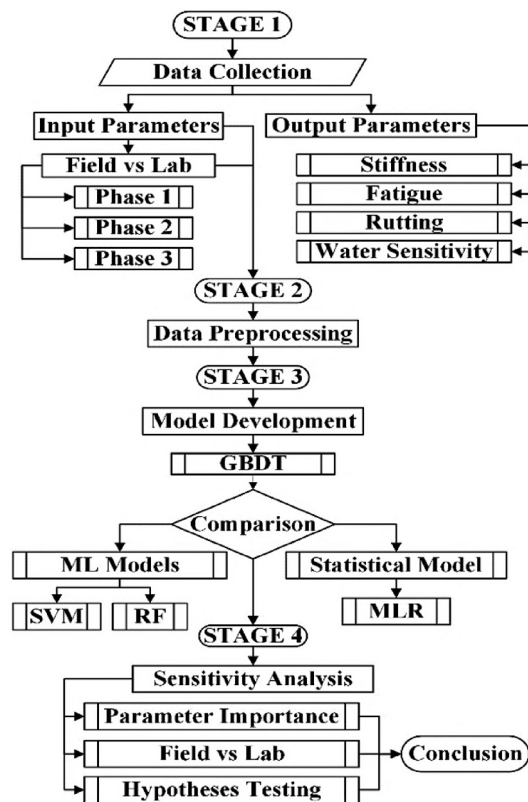
FIGURE 1 Research framework. GBDT, gradient boosting decision tree; MLR, multilinear regression; RF, random forest; SVM, support vector machine.

1.3 | The novelty of the research

The novelty of the presented paper depends on the following key points:

1. *The scope of the problem*: As discussed in the previous sub-sections, none of the existing approaches provide a framework for automatically identifying the difference between the functional performance of laboratory and field-prepared samples with the usage of already existing ML technologies.
2. *The completeness of the solution*: From the reviewed literature (see Section 2), it can be concluded that none of the previous studies presents a full-fledged methodology encompassing processes from (field and laboratory) data collection strategy to providing interpretable results.
3. *The emulation of experts′ knowledge*: With the proposed ML-based innovative framework, this research will enable the emulation of experts' knowledge into ML modeling to solve complex issues in pavement engineering. This can only be achieved by developing state-of-the-art processes to build a framework combining modern techniques and expertise built over decades.

This paper paves the way for further research to solve other relevant issues (apart from the ones highlighted in this research)

4. *Discovering hidden patterns in target parameters*: The proposed framework can discover hidden patterns in the material's behavior and provide deeper insights into the material composition and mixture optimization that were previously unknown.

1.4 | Paper organization

This paper is organized to showcase the novel approach to studying the performance difference between samples prepared in the laboratory and the field. It begins with an overview of PPPM models, which sets the stage for the methodological part of the study, where the development process of GBDT models, including training and testing based on the collected data, is meticulously detailed. Following this, the results and discussions section interprets the findings and contextualizes them within the existing body of knowledge. Last, conclusions and future research sections are presented.

2 | LITERATURE REVIEW OF PERFORMANCE PREDICTION METHODS

With the advancement in pavement asset management, many road agencies utilize various PPPMs incorporated within a dedicated pavement management system (PMS; Uddin et al., 2013). Road agencies use PPPMs to plan their budget allocation well in advance (Haas & Hudson, 2015). PPPMs mainly evaluate the connection between pavement performance indicators and pertinent influencing parameters in the PMS. These models are utilized to investigate the degradation process and predict future pavement conditions (Khattak et al., 2013; Marcelino et al., 2021; J. Yang et al., 2003).

PPPMs can be divided into four categories: empirical, mechanistic, mechanistic-empirical, and probabilistic, based on their underlying approach to modeling and predicting pavement performance (Yehia & Swei, 2020). Each category represents a different methodology and assumptions for analyzing and predicting pavement behavior and deterioration over time. In Table 1 a summary of the key elements and limitations of these methods, which are explained in the following paragraphs, are presented.

Empirical models such as the Highway Development and Management Model (HDM-4; Kerali et al., 2000) mainly rely on observed data to explore the relationship between pavement performance and various influencing parameters (Hu et al., 2022). The advantages of empirical

TABLE 1 Summary of reviewed literature on PPPMs: Key elements, limitations.

Category	References reviewed for key elements and limitation	Key elements	Limitation
Empirical models	(Abu Al-Rub & Darabi, 2012; Deng & Shi, 2023; Hu et al., 2022; Kerali et al., 2000)	−Rely on observed data to explore the relationship between pavement performance and various influencing parameters −Simple model development process	−Exclude mechanistic behaviors −Limited understanding of distress propagation −Poor adaptability to changes (traffic, materials, climate) −Inaccurate long-term predictions −Cannot explore material behavior under varied conditions
Mechanistic models	(Abu Al-Rub & Darabi, 2012; Darabi et al., 2012; Deng & Shi, 2023; Haddad et al., 2022; B. Huang et al., 2001; W. Huang et al., 2020; Hunter et al., 2007; Kettil et al., 2007; T. Ma et al., 2018; W. Wang et al., 2017; Yao et al., 2022; Y. Zhang et al., 2017)	−Predict distress by integrating specific material behaviors −Consider different loading conditions (pulse, moving, equivalent loads) −Use finite element and discrete element approaches −Offer reliable predictions when adequately calibrated	−Overreliance on the theoretical behavior of materials −Performance prediction can be complex and time-consuming −Temporal and spatial uncertainties in stress and environmental conditions
Mechanistic-empirical models	(Dong et al., 2020; Guo et al., 2022; Q. Li et al., 2011)	−Combine mechanistic approaches and empirical methods	−Require intensive calibration and validation procedures −Practicability can be limited
Probabilistic models	(Alimoradi et al., 2022; Basnet et al., 2023; Ghahramani, 2015; Hong & Prozzi, 2006, Jiménez & Mrawira, 2012; Liu et al., 2022b; Mizutani & Yuan, 2023; Saha et al., 2017; Surendrakumar et al., 2013)	−Predict pavement condition transitions with predefined probability distributions −Account for deterioration, environmental conditions, loading, and maintenance histories	Require simplification when data are limited −Possible inaccuracies if the chosen probability distribution does not reflect actual data −Often do not consider the influence of initial state and timely variations in condition transitions
Machine learning (ML)	(Adeli & Hung, 1994; Adeli & Yeh, 1989; Arrieta et al., 2020; Belle & Papantonis, 2021; Bergstra et al., 2013; Deng et al., 2024; Gong, Sun, Hu, et al., 2019; Guo et al., 2022; Hou et al., 2021; Justo-Silva et al., 2021; Kargah-Ostadi et al., 2023; Liang et al., 2022; Liu et al., 2022a; Lundberg & Lee, 2017; Nghiem et al., 2023; Pasupunuri et al., 2024; Peraka & Biligiri, 2020; D. R. Pereira et al., 2020; Rafiei & Adeli, 2017; Ridley, 2022; Sarkhani Benemaran et al., 2023; Yang Song et al., 2022; Tamagusko & Ferreira, 2023; C. Wang et al., 2023, 2024; L. Yang & Shami, 2020; Yao et al., 2019, 2022; M. Zhang et al., 2020)	−Can be improved with more data −Model intricate input–output relationships −Handle large volumes and complex data effectively −Good in decision-making by providing clear information on future trends and challenges −Interpretability of ML models can be enhanced using explainable artificial intelligence (XAI) methods −Incorporating physics-guided ML (PIML) can enhance the generalizability and interpretability of the models −PIML often requires less data	−Often lack the capability to identify the most influential factors in an explanatory way −Prone to misinterpretation and difficult to check the validity −Enhancing interpretability using XAI methods often comes at the cost of reduced accuracy −XAI techniques can be computationally intensive −Incorporating PIML is often complex and requires deep understating of both the physical domain and advanced ML techniques −High-quality and representative data are crucial for training PGML models

models lie in their simple model development process and the established relationships between pavement performance and various influencing parameters (Hu et al., 2022). However, these models often exclude the fundamental mechanistic behaviors occurring within asphalt mixture layers and their interactions (Liu et al., 2022b). Such an exclusion can result in a limited understanding of the diverse parameters involved in distress propagation and restrict the model's ability to account for complex variations. As Deng and Shi (2023) reported empirical models may encounter difficulties adapting to changes such as traffic pattern shifts, material properties, or climate variations. Besides, the lack of adaptability can lead to inaccuracies when making predictions over extended periods. Furthermore, while empirical models provide straightforward insights into pavement performance, they cannot explore the mechanistic intricacies that define material behavior under varying conditions. The gap sets the stage for mechanistic models, which offer a more detailed analysis by accounting for material-specific behaviors and dynamic loading scenarios (Abu Al-Rub & Darabi, 2012).

Mechanistic models predict distress by integrating specific material behaviors as the models developed based on viscoplastic (Hunter et al., 2007) and elastic-viscoplastic behaviors (B. Huang et al., 2001). Hence, mechanistic models can consider different loading conditions, including pulse loads (Hunter et al., 2007), moving loads (Saleeb et al., 2005), and equivalent loads more effectively (Kettil et al., 2007). Methods such as finite element (Haddad et al., 2022) and discrete element approach (T. Ma et al., 2018) are commonly adopted in mechanistic models. Mechanical models for pavement analysis rely on lab-measured material properties and calibrated model coefficients. Properly calibrated models with generalized theories offer reliable predictions for new experimental data (Abu Al-Rub & Darabi, 2012; Y. Zhang et al., 2017). However, for pavement materials that suffer from significant variations, overreliance on the materials' expected theoretical behavior can make the performance prediction process complex and time-consuming (Yao et al., 2022). For example, the stress state of the pavement and environmental conditions often show temporal and spatial uncertainties, which can lead to time-consuming dynamic analyses to alleviate these uncertainties (Deng & Shi, 2023; W. Huang et al., 2020; W. Wang et al., 2017). Therefore, implementing such a time-consuming analysis would hinder the ability to make prompt pavement maintenance and rehabilitation decisions. This limitation underscores the need for another approach that can benefit from the advantages of both mechanistic and empirical models.

Mechanistic-empirical models adopt mechanistic approaches to compute the critical pavement responses (e.g., tensile strain) and empirical methods to pavement

distress based on statistical relations between road structures and field observations (Q. Li et al., 2011). However, despite their acceptable prediction performance, their practicability could be limited. This limitation arises because they require material input parameters from local laboratory test results and often neglect complex environmental and context-related variables (Guo et al., 2022). Hence, like mechanistic models, developing a reliable mechanistic-empirical model necessitates both a calibration procedure to estimate the parameters of the empirical model and a validation procedure to evaluate the prediction accuracy of the calibrated model (Dong et al., 2020). Intensive calibration can underscore the need for probabilistic prediction models, which offer a dynamic perspective on pavement conditions by incorporating historical, environmental, and loading data (Ghahramani, 2015).

Probabilistic prediction models can predict the likelihood of a pavement's condition from a different state to the current state by employing a predefined probability distribution (Liu et al., 2022b). These models can include the dynamic characteristics of pavements concerning deterioration, environmental and loading conditions, and maintenance histories (Alimoradi et al., 2022). For example, Markov chain Monte Carlo is a well-known probabilistic simulation technique that different researchers in pavement engineering have used (Jiménez & Mrawira, 2012; Saha et al., 2017; Surendrakumar et al., 2013). However, when data availability is limited, these models are forced to simplify and choose a probability distribution, which can lead to inaccuracies if the selected distribution does not accurately reflect the actual data (Basnet et al., 2023). Moreover, in probabilistic models like Markov chain Monte Carlo, while the subsequent state depends solely on the current state, the influence of the initial state (e.g., initial pavement material properties) and timely variations in pavement condition transition trends is generally not considered (Liu et al., 2022b; Mizutani & Yuan, 2023).

With the rapid development of computational techniques, ML (Adeli & Hung, 1994) and deep learning have been widely adopted in civil engineering. These advanced methods have helped address some of the limitations of previously discussed techniques (Tamagusko & Ferreira, 2023). Adeli and Yeh (1989) were among the researchers who pioneered applying ML in the civil engineering domain. The popularity of ML stems from its capacity to model intricate input–output relationships (Justo-Silva et al., 2021; Yao et al., 2022). Additionally, the capability of ML to improve over time with mode data can make the resulting models more adaptable, compared to traditional approaches, which are dependent on a set of equations and assumptions (Tamagusko & Ferreira, 2023). ML algorithms can predict various functional performance indicators of

asphalt mixture (Liu et al., 2022a). By training on extensive pavement data, ML models discern patterns and relationships that can enhance PPPMs and PMS (D. R. Pereira et al., 2020; Rafiei & Adeli, 2017).

Among various ML models, artificial neural networks (Adeli & Yeh, 1989) and SVM (Cortes & Vapnik, 1995) models have been extensively utilized in developing PPPMs (Hou et al., 2021). However, despite the competence of these models in accurately predicting pavement performances, they often lack the capability to identify which influencing factors have the most significant impact on the final outcomes in an explanatory way. Researchers Guo et al. (2022) and Yang Song et al. (2022) reported that the literature on the application of ML for PPPMs mainly focuses on model accuracy or applies the primary methods for parameter importance estimation (Yao et al., 2019). Moreover, ML-based PPPMs often interpret results by assessing parameter importance or sensitivity, but they usually do not explore deeply the effects of parameters (Gong, Sun, Hu, et al., 2019; Yao et al., 2019).

Since today's PMS handles growing data volume and complexity, it is possible to consider diverse parameter formulations (Peraka & Biligiri, 2020; Yao et al., 2022). However, without comprehensive explanatory modeling (Rosé et al., 2019), these formulations might yield inaccurate or misleading results, potentially leading to suboptimal maintenance decisions and inefficient resource allocation (Yao et al., 2022). Therefore, explanatory modeling is essential to understand specific parameter impacts on model outputs for practical applications. Otherwise, the reliability of decisions based on ML model predictions cannot be guaranteed (Belle & Papantonis, 2021)

In order to provide more reliable and explainable models, the GBDT framework has become a popular choice in recent years (Liang et al., 2022). Gong, Sun, and Huang (2019) developed a GBDT-based model to enhance the predictive performance of fatigue cracking. Input parameters such as pavement thickness, resilient modulus of subgrade, and climatic conditions were considered in this model. It was highlighted that the predictive performance of GBDT models significantly outperformed the transfer functions presented in the mechanistic-empirical pavement design guide (Hallin, 2004). In another study, M. Zhang et al. (2020) developed a GBDT model to predict asphalt overlay performance, achieving satisfactory predictive performance ($R^2_{test} > 80\%$) for roughness, rutting, transverse cracking, and non-wheel path longitudinal cracking. Researchers Guo et al. (2022) and C. Wang et al. (2023) proposed a GBDT-based model to predict the international roughness index and rut depth while considering various influencing parameters such as temperature and wind speed. The proposed GBDT models in these studies outperformed other models. Sarkhani Benemaran et al.

(2023) used GBDT to predict the resilient modulus of subgrade in flexible pavement foundations. The GBDT model had a significant prediction accuracy ($R^2 = 0.991$), compared to the other models.

The studies reviewed above indicate that although using GBDT models can help identify the most influential parameters, it is often difficult to comprehensively explain the interactions or contributions of these parameters to individual predictions. Therefore, to improve the explainability of the ML-based models, pavement researchers steered more toward explainable artificial intelligence (XAI) methodologies (Ridley, 2022) by incorporating methods such as SHAP (Lundberg & Lee, 2017). For example, Guo et al. (2022) and Yang Song et al. (2022) employed a specialized version of GBDT to model the international roughness index and determine parameter importance utilizing SHAP. While outcomes were noteworthy, the model's hyperparameter optimization relied on grid search, which might lack computational and performance efficiency, compared to Bayesian optimization (Bergstra et al., 2013; L. Yang & Shami, 2020). Yao et al. (2022) enhanced the clarity of the Bayesian neural network model developed for predicting transverse cracks by implementing SHAP. Although the obtained results are promising, the scope of the study is limited to transverse cracks, and the mechanistic properties of the asphalt mixture (e.g., phase angle) were less considered. Despite the promising results obtained by incorporating XAI methodologies, the number of research that has been carried out in developing PPPMs using these methodologies is limited due to issues such as prioritizing accuracy over interpretability (Yang Song et al., 2022), the computationally intensive nature, and the complexity of XAI implementation (Arrieta et al., 2020).

In response to the issues mentioned in the previous paragraph, researchers Deng et al. (2024), Kargah-Ostadi et al. (2023), and Pasupunuri et al. (2024) tried to incorporate pavement-domain-specific physical knowledge into the learning process of ML models through a framework called physics-guided ML (PIML; Nghiem et al., 2023). PIML aims to improve not only the explainability in the ML models but also the data efficiency, the generalizability, and the physical plausibility of predictions (G. Wang et al., 2024). However, regardless of the lack of sufficient studies about PIMLs in PPPMs, there are some challenges, such as finding proper physical laws matching the available dataset, complexity of physical laws (Fuks & Tchelepi, 2020), data efficiency and quality (Meng et al., 2022) and computational costs (Shukla et al., 2022).

In summary, field performance indicators are frequently evaluated in controlled environments of laboratories. Consequently, it is vital to identify the discrepancies between the properties of mixtures prepared in the lab and those

in the field. However, differentiating between laboratory and field properties can introduce an additional layer of complexity to PPPMs as it requires the consideration of more parameters. Due to the recent advancement in computational facilities, ML-based models have become more powerful, which could help pavement researchers establish such connections (Marcelino et al., 2021; X. Yang et al., 2021). However, to the best of the author's knowledge, such a framework is not yet available for the pavement community.

3 | DEVELOPING GBDT MODELS

The following section presents the application of GBDT models as adopted in this research. At first, a brief introduction to GBDT is presented. Then, an overview of the data collection and preprocessing stages is described. Subsequently, a discussion of findings based on the preliminary data analysis of the proposed models is presented.

3.1 | Brief overview of GBDT

The GBDT models have become a popular choice in recent years as they often outperform artificial neural network models (Lundberg et al., 2020). This is because GBDT draws insights and methods from both statistical and ML methods (Barua et al., 2021; X. Ma et al., 2017). There are other benefits to using GBDT, such as being effective in handling datasets with high cardinality (Hancock & Khoshgoftaar, 2020) and missing values (Barua et al., 2021; Ding et al., 2018; Friedman, 2001).

Since the foreseen dataset of this research (see Section 3.2) is expected to have many input parameters and missing values, the GBDT framework was selected. Gradient boosting is a method that generates a powerful learning model by combining multiple so-called "weak learners" (Friedman, 2001). It is noted that since gradient boosting utilizes regression decision trees (Breiman, 2017) to fit the gradient descent algorithm (Hastie et al., 2009), it is also known as a gradient-boosting decision tree. Detailed information about the mathematical representation of GBDT can be found in Friedman (2001). In the next section, a concise overview of the process involved in developing GBDT models tailored to the specific goals of this research is presented.

3.2 | Overview of data collection and preparation

In order to mitigate the bias in the models, data were collected from different contractors and construction sites
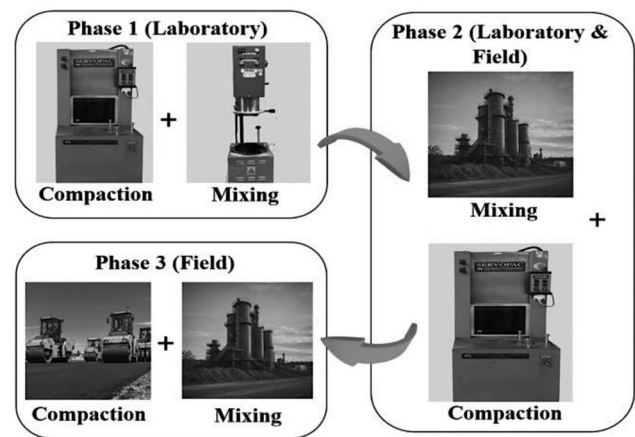

FIGURE 2 Test specimens are made with three different combinations of laboratory and field mixing and compacting setups.

with similar construction conditions to represent typical construction practices in the Netherlands. In total, six different construction projects were selected for the data collection. It is noted that these six projects are not test sections or specifically built for this project; instead, they are road sections that were constructed for public usage. The data from the projects were collected at distinct stages (so-called "phases") throughout each project as shown in Figure 2 and described in Table 2.

As shown in Figure 2, in "Phase 1," different asphalt samples corresponding to the field mix design were mixed and compacted in a controlled laboratory setup. In "Phase 2," asphalt samples were collected directly from the asphalt plant. Subsequently, the mixtures were compacted in the laboratory. "Phase 3" is classified as a phase in which the samples were obtained from the construction sites. This means the obtained samples are mixed and compacted at the site.

As one of the objectives of this research is to compare the laboratory data with field observations, key laboratory tests were performed. Table 3 presents the performed laboratory tests together with the functional performance indicators that will be used in developing the models. Additional tests were carried out after 6 and 12 months of construction (categorized as "P3-Y0") to measure the influence of bitumen aging following the procedures outlined in EN 12607-1 (2014). Moreover, to reflect the intermediate and long-term effect of aging at the mixture scale, the tests were repeated after 2, 3, and 6 years of construction ("P3"Y6). It is noted that all the samples were aged inside the laboratory in controlled conditions at the constant temperature of $13 \pm 2°C$. This implies that there was an absence of UV exposure and moisture conditioning.

TABLE 2 Overview of the classification of mixture conditioning.

Code:	Description	Activity	Mixing setup	Mixing context	Compaction setup	Compaction context
Phase 1 (P1-Y0)	At the time of construction	Asphalt mixture and bitumen analysis	Forced action mixer	Lab	–Gyrator comp –Hand roller –Mini roller –Segment comp –Shear box	Lab
			Planetary mixer	Lab	–Gyrator comp –Mini roller	Lab
Phase 2 (P2-Y0)	At the time of construction	Asphalt mixture and bitumen analysis	Asphalt plant	Field	–Gyrator comp –Hand roller –Mini roller –Segment comp –Shear box	Lab
Phase 3	At the time of construction, from the road	Asphalt mixture and bitumen analysis	Asphalt plant	Field	–Field roller	Field
Phase 3 (P3-Y0)	Six and 12 months after construction	Bitumen analysis				
Phase 3 (P3-Y2)	Two years after construction	Asphalt mixture and bitumen analysis				
Phase 3 (P3-Y3)	Three years after construction	Asphalt mixture and bitumen analysis				
Phase 3 (P3-Y6)	Six years after construction	Asphalt mixture and bitumen analysis				

TABLE 3 Tests and standards for determining asphalt mixture functional performance indicator and bitumen functional properties.

Functional performance indicator of asphalt mixture/bitumen	Laboratory test	Standard
Stiffness ($E^*[\mathbf{MPa}]$)	Four-point bending	(*EN 12697-26, 2012*), method B
Resistance to fatigue ($\varepsilon_6 \left[\frac{\mu m}{m}\right]$)	Four-point bending	(*EN 12697-24, 2018*), method D
rutting ($f_c \left[\frac{\mu\varepsilon}{cycle} \cdot 10^6\right]$)	cyclical triaxial test with signal	(*EN 12697-25, 2016*), method B
Water sensitivity (unitless, the ration of $\frac{ITS_{Wet}\ [MPa]}{ITS_{Dry}\ [MPa]}$)	Indirect tensile strength (ITS)	(*EN 12697-12, 2018*; *EN 12697-25, 2016*), method A, and (*EN 12697-23, 2017*)
Bitumen penetration[a]	Penetration test	(*NEN-EN 1426, 2015*)
Softening temperature[a]	Ring and ball test	(*NEN-EN 1427, 2015*)
Phase angle[a]	Dynamic shear rheometer	(*NEN-EN 14770, 2022*) (from −10 to 60°C)

[a]Parameters related to bitumen properties.

3.2.1 | Preprocessing of the collected dataset

Several studies (Alexandropoulos et al., 2019; Budach et al., 2022; García et al., 2015) have reported that proper data pre-processing can significantly improve the robustness and accuracy of ML models. Therefore, as a first step, the collected data were preprocessed because the performance of

GBDT models, like any other ML models, heavily relies on the data quality (Zaki & Meira, 2014).

This research uses several preprocessing steps, such as data integration, cleaning, and transformation. It is noted that the dataset was split into training (80%) and testing (20%) sets before data transformation in order to mitigate the risks of information leakage (Cawley & Talbot, 2010;

Kaufman et al., 2012). An overview of the dataset resulting from the preprocessing step is presented at the end of this subsection.

Data transformation

In order to facilitate the learning process of ML models, data transformation techniques are commonly used to convert a non-linear relationship between an input and target parameter into a linear relation. Some of the ML models, such as SVM, cannot handle categorical variables directly (Jung & Kim, 2023; Lee & Kim, 2010). In this research, non-linear relationships exist in the dataset, including those introduced by categorical parameters (e.g., compaction setups and friction reduction systems). Therefore, categorical parameters were transformed into numerical formats to get a linear relationship. Prokhorenkova et al. (2018) reported that ordered target encoding is effective when dealing with high cardinality categorical parameters (i.e., categorical parameters with many levels or categories). Ordered target encoding can reduce the dimensionality of the data while still preserving important information.

Since the dataset in this research contained numerous categorical parameters with high cardinality, ordered targeting encoding was applied to limit the categorical parameters into more manageable forms. The ordered targeting encoding method was chosen over other methods, such as one-hot encoding (Okada et al., 2019) because one-hot encoding can result in high-dimensional feature spaces (Kunanbayev et al., 2021). This can increase computational complexity and memory usage (Cerda & Varoquaux, 2022).

In the first step of the ordered target encoding method, the mean target value is calculated for each instance in the dataset using only the instances that precede it. In the second step, numerous permutations are carried out to mitigate the significant variation in the initial values caused by the procedure in the first step. The final value is calculated as the average of all permutations in Step 2. Detailed information about ordered target encoding can be found in research by Prokhorenkova et al. (2018). It is noted that the original format of categorical parameters is presented in Table 4 to facilitate a better understanding.

3.2.2 | Overview of preprocessed dataset

An overview of the dataset obtained after the preprocessing steps is presented in Table 5. The total number of data points for each output parameter and the maximum and minimum values are presented. As can be seen, stiffness and fatigue resistance parameters possess more data points and a broader range (i.e., the difference between maximum and minimum values) in comparison to rutting and indirect tensile strength (ITS).

TABLE 4 Different categorical variables in the dataset with their subcategories.

Compaction setup (CS)	Friction reduction system (FRS)[a]	Mixing setup (MS)
Field roller (FR)	Two-layer rubber with silicon grease (TRSG)	Asphalt Plant (AP)
Gyrator compactor (GCOM)	2x Marshallpaper (2MP)	Forced Action Mixer (FAM)
Hand roller (HR)	PTFE gecoat vlies (PTFE)	Planetary Mixer (PM)
Mini roller (MR)	Acre system (AC)	
Segment compactor (SCOM)		
Shear box (SB)		

[a]FRS is designed to reduce frictional losses during the triaxial test to ensure uniform and consistent test results.

TABLE 5 Overview of output parameters and the number of data points.

Output (functional performance indicators of asphalt mixture)	Minimum value	Maximum value	Total data points
Stiffness	6144	17,866.47	425
Resistance to fatigue	37.142	177.47	407
Rutting	0.02	0.95	120
ITS	1.25	4.6	192

In Table 6, an overview of the input parameters, along with their measurement unit and minimum and maximum values, is presented. The data in Table 6 show that various input parameters, such as "age" and "target density," were considered for the model development. Each parameter represents a property of the asphalt mixture. For example, parameters "Target density" and "compaction degree" represent the density standards for the mixture and the degree of compaction achieved. The composition of the asphalt mixture is mainly represented by parameters such as the "Target Mass Composition" of stone, sand, filler, and bitumen. The physicochemical properties of bitumen are presented with parameters such as "Bitumen Penetration" and "Bitumen Phase Angle."

3.3 | Preliminary analysis

A preliminary analysis was done to fine-tune the hyperparameters of the models and compare the GBDT models with statistical and ML models. A description of the tasks mentioned above is provided in the following subsections.

TABLE 6 Overview of the input parameters in the dataset and their corresponding minimum and maximum values.

Input parameter	Unit	Minimum value	Maximum value
Age (Y)	Year	0	6
Target density (TD)	kg/m^3	2360	2399
Compaction degree (CD)	%	97.35	103.31
Bitumen Penetration (BP)	0.1 mm	11	55
Bitumen Phase Angle (δ)	Degree	39.96	66.57
Target Stone Mass Percentage (TST)	%	52.58	57.9
Target Sand Mass Percentage (TSA)	%	32.12	36.81
Target Filler Mass Percentage (TF)	%	5.72	7.64
Target Bitumen Mass Percentage (TB)	%	4.25	5.4
Reclaimed Asphalt Pavement (RAP)	%	50	65
Mixing setup (MS)	n/a	n/a	n/a
Compaction setup (CS)	n/a	n/a	n/a
Friction Reduction System (FRS)	n/a	n/a	n/a
Sample condition (SC)	n/a	Dry (0)	Wet (1)

TABLE 7 Predictive performance of the gradient boosting decision tree (GBDT) models without hyperparameter optimization.

Metric	Stiffness	$\varepsilon 6$	fc lin	ITS
R^2_{test}	0.94	0.72	0.77	0.75

3.3.1 | GBDT hyperparameters optimization

After the preprocessing step, the dataset was used to train and test the GBDT models. In Table 7, the R^2_{test} of the models without any hyperparameter optimization is presented. As can be seen from the table, the performance of the GBDT model is overall satisfactory for all functional performance indicators. However, since GBDT contains different hyperparameters (Chen et al., 2019) throughout the learning process, hyperparameter optimization is essential to improve the performance of the models. Therefore, in this research, the hyperparameters such as "learning rate" (Hastie et al., 2009), the "max depth" (Friedman, 2001), "l2 leaf reg" (Tian & Zhang, 2022), and "minimum child samples" (Breiman, 2017; Yanyan Song &

Lu, 2015) were considered to avoid overfitting and improve the performance of the GBDT model.

In order to find the optimum values and combinations for the hyperparameters, the Bayesian optimization technique (L. Yang & Shami, 2020), in conjunction with 10-fold cross-validation (CV; Anguita et al., 2012), was used. Bayesian optimization is a highly effective algorithm that aims to identify the global optimal solution within the parameter space (K. Li et al., 2024). The possible range of values for hyperparameters (see Table 8) was considered based on the literature review, the scikit-learn library (Sklearn.ensemble, 2024), and the need to balance model complexity with computational efficiency.

As shown in Table 8, the value range for the "learning rate" hyperparameter was considered between 0.001 and 0.3, and different optimized values for stiffness (0.016), fatigue resistance (0.01), rutting (0.02), and ITS (0.009) models were obtained. The optimized learning rate values for each model indicate the rate of weight updates needed for the best performance. Furthermore, the impact of the chosen hyperparameters on the GBDT models is illustrated in Figure 3. As can be seen, the "min_child_samples" is the most influential hyperparameter, which could suggest its role in controlling overfitting. In second place is the "learning_rate," which was expected as its role in controlling the magnitude of the update to the model's weights or parameters was emphasized in previous studies (Friedman, 2001).

3.3.2 | Predictive performance

After tuning the GBDT models, R^2 and root mean squared error (RMSE) are employed to represent their predictive accuracy in Table 9. The comparison between the results presented for the R^2_{test} in Tables 7 and 9 shows that the hyperparameter optimization led to an average improvement of approximately 7.6% across all models. This improvement can highlight the significant positive impact of hyperparameter optimization on the performance of the models.

According to the results presented in Table 9, the differences between the mean R^2 values from CV and those from the training dataset were found to be negligible. This observation suggests a consistent model performance across different data distributions. Besides, the high R^2 values obtained for the testing datasets imply that the models maintain a significant predictive accuracy when applied to new datasets. Furthermore, in order to provide a way to compare the model's error relative to the range of performance indicators, the normalized RMSE (NRMSE) was also utilized.

TABLE 8 Optimized hyperparameters of GBDT for different functional performance indicators.

Hyperparameter	Possible range	Optimized value			
		Stiffness	$\varepsilon 6$	fc lin	ITS
Learning rate (LR)	[0.001–0.3]	0.016	0.01	0.02	0.009
Max depth (MD)	[3–20]	10	10	13	11
l2 leaf reg (L2)	[0–10]	2.5	1	1	4
Min child sample (MCS)	[1–100]	8	32	1	32
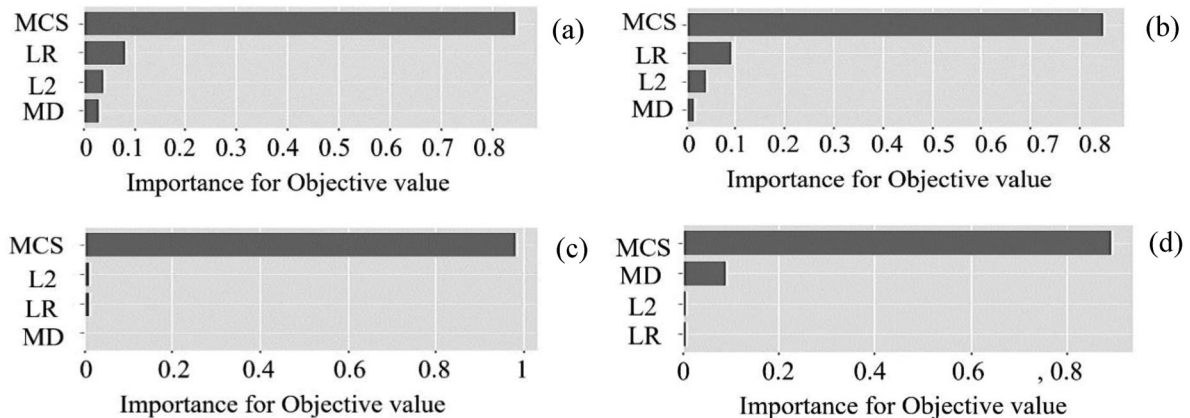
FIGURE 3 Hyperparameter importance for: (a) stiffness, (b) fatigue resistance, (c) rutting, and (d) indirect tensile strength (ITS) (for acronyms, refer to Table 8).

TABLE 9 Predictive performance on testing data for different functional performance indicators after hyperparameter optimization.

| Metrics	Stiffness	$\varepsilon 6$	fc lin	ITS
Root mean squared error (RMSE)	518.89	9.13	0.063	0.23
Normalized RMSE (NRMSE)	0.0443	0.0651	0.0677	0.0687
R^2_{train}	0.954	0.829	0.895	0.900
R^2_{test}	0.951	0.802	0.812	0.842
R^2_{CV}	0.919	0.762	0.794	0.823

As shown in Table 9, the GBDT model shows high predictive accuracy for stiffness with an R^2_{test} value of 0.951. Given the scale of stiffness values (see Table 5), the model's predictive deviation is relatively moderate ($NRMSE = 0.0443$), which suggests the effectiveness of the model in predicting stiffness. The model also shows acceptable predictive capability for fatigue resistance with an R^2_{test} of 0.802. Considering the range of fatigue resistance values, the error in the model's predictions is seen as reasonably small ($NRMSE = 0.0651$), which reflects a rea-

sonably accurate model. The predictive performance of the model for rutting is also reasonable, with an R^2_{test} of 0.812 and $NRMSE$ of 0.0677. Last, the model for ITS shows an R^2_{test} of 0.842, and the $NRMSE$ of 0.0687 signifies that it is reliably accurate.
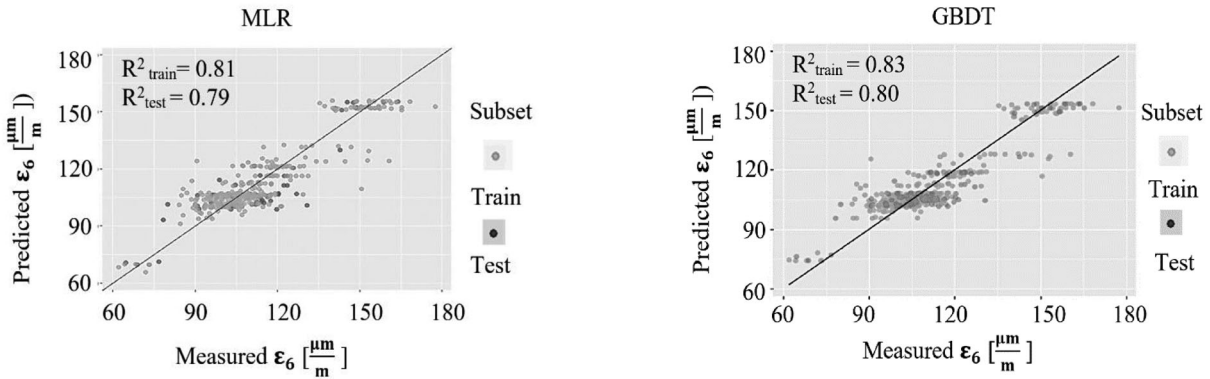
3.3.3 | Comparison with multiple linear regression

In order to evaluate the optimized GBDT models, MLR models were developed for each functional performance indicator. The MLR was chosen as a baseline because it was frequently used by previous researchers (Nyirandayisabye et al., 2022; Tran et al., 2023; Zhao et al., 2022). Detailed information about MLR can be found in (Box et al., 2005; Eberly, 2007).
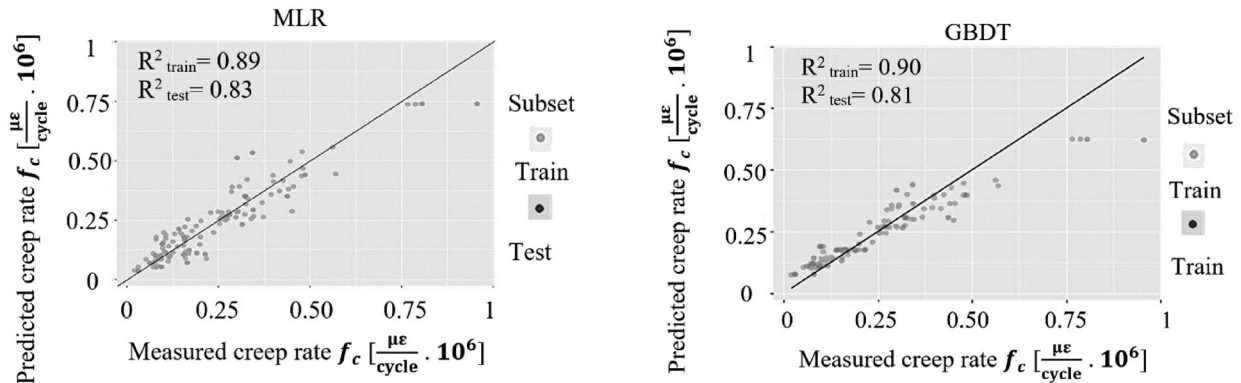
Figure 4 shows that both MLR and GBDT were able to establish the correlation between input and output parameters. In general, it can be observed that R^2 values are very similar to each other. As can be seen from Figure 4a,b,d, the GBDT was found to have a slightly better R^2 value. This could be because of GBDT's ability to recognize nonlinear relationships between input and output parameters. It is noted that the observed step-like pattern in Figure 4b is due to the nature of the fatigue dataset. As Figure 4c
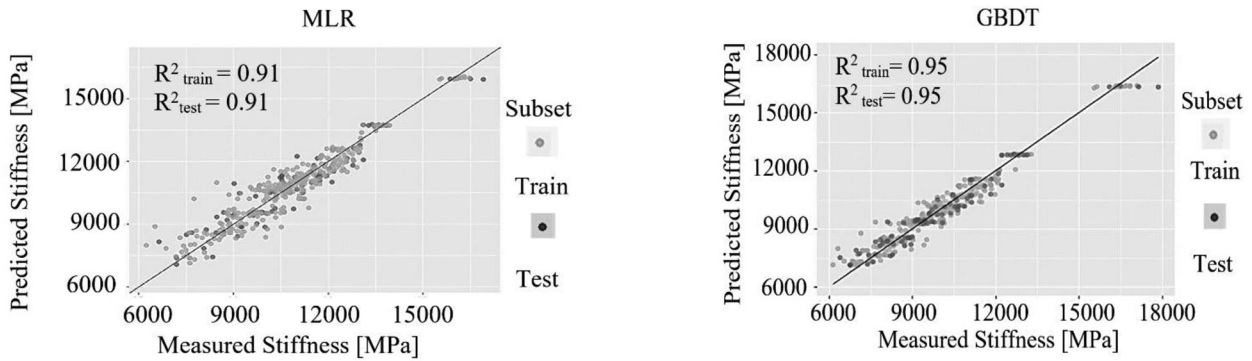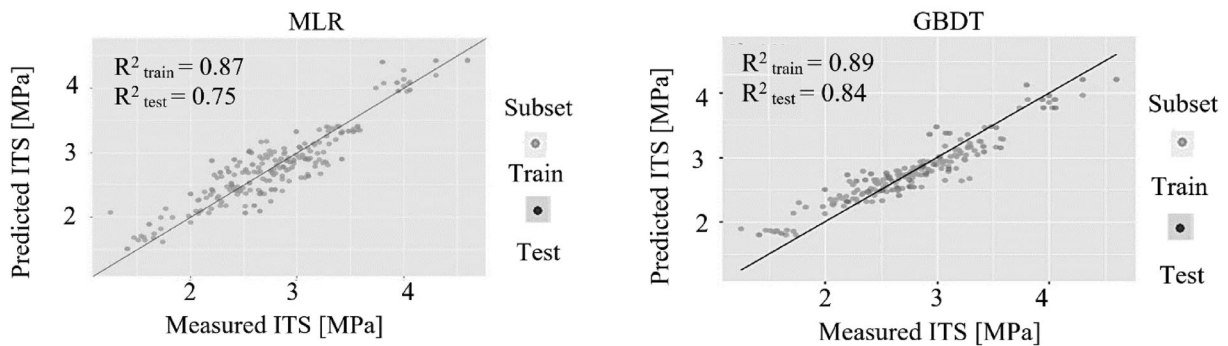
FIGURE 4 Performance comparison of gradient boosting decision tree (GBDT) and multilinear regression (MLR) models. (a) Predictive versus actual measured stiffness; (b) predictive versus actual measured resistance to fatigue, (c) predictive versus actual measured rutting, and (d) predictive versus actual measurements of ITS.

TABLE 10 The optimized values for the support vector machine (SVM) models' hyperparameters.

Hyperparameter	Stiffness	$\varepsilon 6$	fc lin	ITS
Kernel	Radial basis function	Radial basis function	Radial basis function	Radial basis function
C	1.5	1.5	1.5	10
Gamma	0.1	0.2	0.3	0.1
ε	0.1	0.1	0.2	0.1

TABLE 11 The optimized values for the random forest (RF) model's hyperparameters.

	Stiffness	$\varepsilon 6$	fc lin	ITS
Max depth	80	80	80	80
Max features	5	2	5	5
Min samples leaf	3	3	3	3
Min samples split	8	8	8	8
Number of estimators	1000	1000	200	1000

shows, GBDT did not improve the predictive performance in the rutting model. From the author's perspective, this is because the creep rate is the linear part of the creep curve, and an MLR model could fit the linear relationships more accurately. Furthermore, the scarcity of data points for creep rate and ITS might hinder the learning efficacy of the GBDT model.

3.3.4 | Comparison with other ML methods

After evaluating the proposed GBDT models with the MLR models, to evaluate the efficiency and effectiveness, the models were compared with ML methods such as SVM (Cortes & Vapnik, 1995) and RF (Breiman, 2001). The reason behind selecting SVM and RF is that these models are among the frequently used models by several researchers in the pavement domain (see Section 2). The authors established SVM and RF models with the hyperparameters presented in Tables 10 and 11. These hyperparameters were selected based on the standard practices in ML. All of the hyperparameters were optimized by the Bayesian optimization and 10-fold CV (as for the proposed model), and the optimized values are listed in Tables 10 and 11. It is noted that the description of the hyperparameters for SVM and RF can be found in (Random Forest Regressor, 2024; SVR- Sklearn, 2024).

The comparative evaluation of the prediction performance for three ML methods is presented in Table 12. Based on the R^2_{test}, the GBDT models slightly outperformed other algorithms. Furthermore, the proposed GBDT framework consistently offers balanced results

across different performance indicators by maintaining the R^2_{test} values above 0.8. In contrast, the SVM model exhibited a notably low R^2_{test} value of 0.216 for predicting rutting. Similarly, while the RF model performed reasonably well, it showed variability with R^2_{test} values across different functional performance indicators. Furthermore, the RMSE values further emphasize the efficiency of the GBDT model. Although the RF model showed a slightly better RMSE for stiffness (515.71), compared to GBDT's (518.89), the GBDT model maintained lower RMSE values for other indicators, underscoring its overall accuracy and stability.

4 | RESULTS AND DISCUSSION

In order to achieve a general overview of the influencing input parameters in the developed GBDT models, SHAP values were incorporated. SHAP is based on ideas from game theory (Lundberg & Lee, 2017), which aims to distribute the contributions of each input parameter (see input features X_1 to X_{12} in Figure 5a as a symbolic representation of such parameters) with equal consideration and collectively use them to make a fair prediction (Z. Li, 2022). As shown in Equation (1), the quantification of the contribution of each parameter is obtained by adding the parameter "X_j" to different groups of parameters on the prediction of the model.

$$\text{Shapley} \ (X_j) = \sum_{S \subseteq N \setminus \{j\}} \frac{k! \, (p - k - 1)!}{p!} \left(f \left(S \subseteq N \{j\} \right) \right. \\ \left. - f \left(S \right) \right) \quad (1)$$

where "k" is the number of parameters, "p" is the total number of parameters, "$N \setminus \{j\}$" is a set of all possible combinations of parameters excluding "X_j," is a parameter set in "$N \setminus \{j\}, f(S)$" is the model prediction with parameters in "S," and "$f(S \cup \{j\})$" is the model prediction with parameters in S plus parameter "X_j."

The following subsections present the results of the SHAP analysis for each input parameter for stiffness, fatigue resistance, rutting, and ITS.

4.1 | Stiffness

The SHAP summary plot (Lundberg et al., 2019), as shown in Figure 5, provides a visualization toolkit to study the impact of each input parameter (or feature) on predicting stiffness values. Since the input features are arranged in order based on their average SHAP value, the "compaction degree," "bitumen phase angle," and "mix setup" are found to be of significant importance.

TABLE 12 Performance comparison of the GBDT models with SVM and RF based on the R^2_{test} and RMSE test.

Models	Stiffness		$\varepsilon6$		fc lin		ITS	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
SVM	0.938	545.17	0.800	8.82	0.216	0.194	0.816	0.27
RF	0.943	515.71	0.756	9.74	0.729	0.114	0.743	0.32
GBDT	0.951	518.89	0.802	9.13	0.812	0.063	0.842	0.23
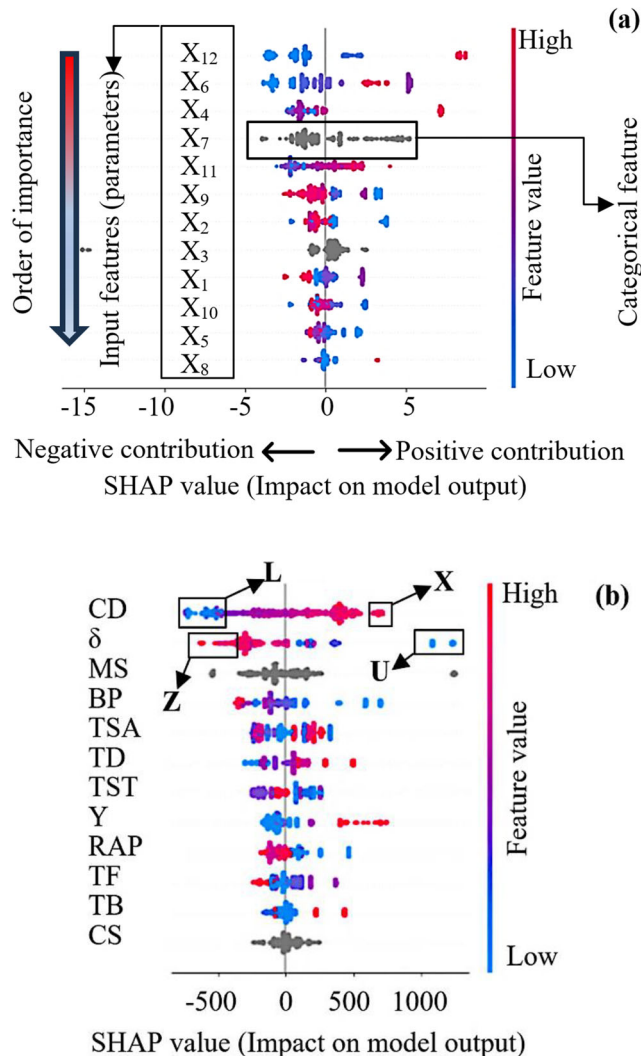
FIGURE 5 (a) Legend of the summary Shapley additive explanations (SHAP) plot, (b) summary SHAP plot for stiffness. (For acronyms, refer to Table 6). *Note*: Each point represents the contribution of a feature to an observation, with the SHAP value determining the position on the horizontal axis. Negative SHAP values indicate negative contributions (left side), and positive SHAP values indicate positive contributions (right side). Feature values are color-coded: blue for low values and red for high values.

As seen in Figure 5b, the higher values of the "compaction degree," highlighted within the rectangle marked "X," are located on the right side of the center line. This implies that higher compaction degrees will increase stiff-
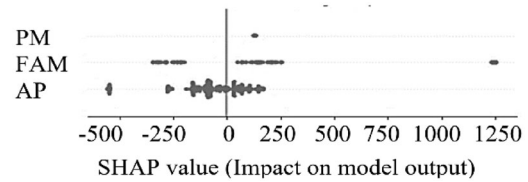

FIGURE 6 SHAP plot of mixing setup for stiffness (For acronyms, refer to Table 4).

ness. Conversely, lower values of the "compaction degree," indicated by the rectangle marked "L," appear on the left side of the center line, indicating that lower compaction degrees will decrease stiffness value. The observed relationship between compaction degree and stiffness aligns with the findings of previous research (Brown, 1990).

For the data points corresponding to "phase angle," the highest values highlighted within the rectangle marked "Z" are situated on the left side of the middle line. This indicates that higher "phase angle" values have a decreasing effect on stiffness value. On the contrary, lower values of the "phase angle," indicated by the rectangle marked "U," are found on the right side of the center line. This implies that a lower "phase angle" correlates with increased stiffness. This could be associated with the fact that a higher phase angle denotes a material's propensity to become more viscous, which would reduce its elastic behavior.

As seen in Figure 5b, the third input feature in order, that is, "mixing setup," does not have important values. Hence, such a feature in the plot does not provide meaningful information. To solve this issue, further investigations are necessary. Figure 6 shows a deeper understanding of the various correlations. The central idea behind these plots remains the same as the actual SHAP summary plot with the difference in the y-axis, which does not represent the importance of the categories. This plot is used to organize the categories for visual clarity and display the dimension of data related to the categorical variable.

As seen from Figure 6, the available data points for the "planetary mixer" are less to draw any definitive conclusion. However, considering that the available data lie on the right side of the central line, it gives an indication that using the "planetary mixer" can lead to more stiffness. The results for the "forced action mixer" are inconclusive as its
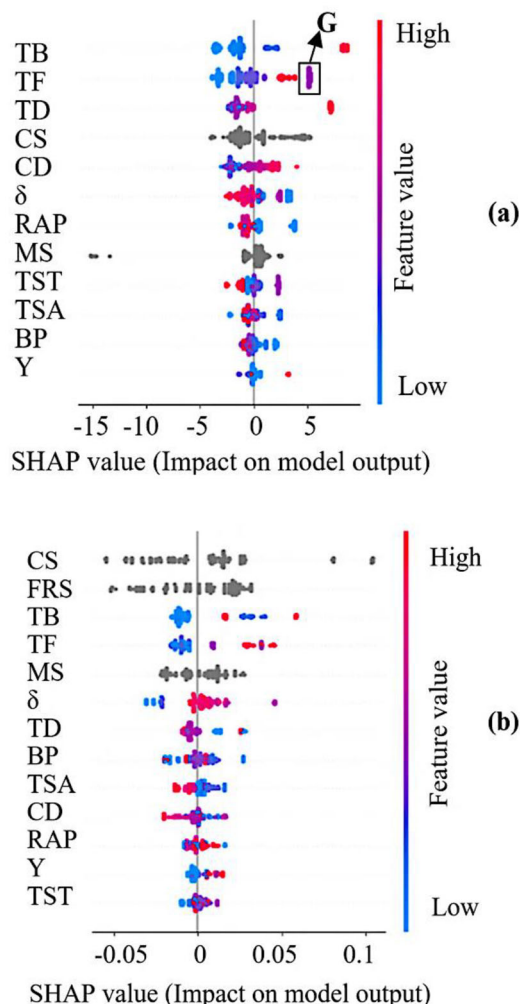
FIGURE 7 Summary SHAP plot for (a) fatigue resistance and (b) for rutting (For acronyms, refer to Table 6).
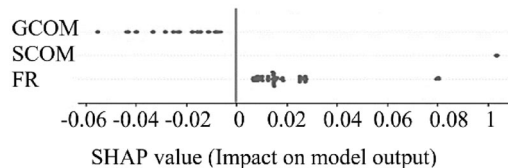

FIGURE 8 SHAP plot of compaction setup of rutting (For acronyms, refer to Table 4).

get bitumen mass percentage" values have an increasing effect on fatigue value. This is in line with findings from previous research (Erkens & van Vliet, 2014), which indicates that mixtures with a lower bitumen content tend to be more susceptible to cracking, thereby reducing their fatigue resistance.

A similar trend to "target bitumen mass percentage" is observed for data points of "target filler mass percentage," where its lower values decrease fatigue resistance values. However, for the data of this feature, the high values do not lie at the rightmost side of the horizontal axis. In contrast, purple values (i.e., lower than the red color in the feature value bar), as marked by the rectangle "G," lie at the end of the horizontal axis. This shows that adding more filler will not increase the fatigue resistance beyond the critical limit, which is consistent with the findings of the previous studies (B. Huang et al., 2007).

Data points of "target density" show behavior similar to "target bitumen mass percentage." The high values of "target density" are on the right side of the central line, which implies that its higher values have an increasing effect on fatigue resistance values. This observation is consistent with the previous research by Mogawer et al. (2011), which postulated that higher densities can influence the air void content and elevate the initial stiffness of the mixture, potentially extending its fatigue life.

4.3 | Rutting

As illustrated in Figure 7b, the top three features that influence rutting prediction are "compaction setup," "friction reduction system," and "target bitumen mass percentage." Given that "compaction setup" is a categorical variable, Figure 8 was preferred to investigate the relative impact of various compaction setups on rutting values. Since the "gyratory compactor" data points are located on the left side of the central line, it can indicate that using "gyratory compactor" could result in a lower likelihood of rutting. The available data points for the "segment compactor" are less to draw any definitive conclusion. However, considering that the available data lie on the right side of the central line, it gives an indication that by using the "segment compactor," more rutting can be observed.
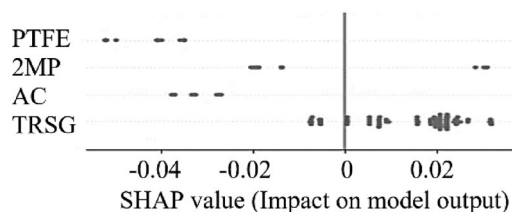
data points are almost equally distributed on both sides of the central line. In contrast, the data points associated with "asphalt plants" are slightly leaning toward the left side of the central line, which indicates that by using this mixing setup, less stiffness can be observed. On the basis of observation from Figure 6, the importance of the mixing process in achieving the desired stiffness of asphalt can be concluded.

4.2 | Fatigue resistance

As shown in Figure 7a, the top three features that influence fatigue resistance prediction are "target bitumen mass percentage," "target filler mass percentage," and "target density."

It can be observed that data points with high "target bitumen mass percentage" values are placed on the right side of the center line. This indicates that higher "tar-

FIGURE 9 SHAP summary plot of friction reduction system for rutting (For acronyms, refer to Table 4).

FIGURE 10 Summary SHAP plot for ITS values (For acronyms, refer to Table 6).

FIGURE 11 SHAP summary plot of sample condition for ITS values.

In contrast to the "gyratory compactor," the data points associated with the "field roller" are located on the right side of the central line, which raises the possibility that more rutting could be seen if this compaction setup was used. This observation indicates that the samples prepared in the laboratory using the "gyratory compactor" result in lower rutting as compared with the field rollers. It is noted that these observations are valid if the same condition is maintained in this research.

As "friction reduction system" is a categorical variable, Figure 9 was used to investigate the relative impact of different friction reduction setups on rutting values. It can be seen from the figure that data points related to " PTFE gecoat vlies" and " Acre system" are located on the left side of the central line. This can suggest that using these friction reduction systems could lead to less rutting. The data points related to "2 maal marshallpapier" cannot indicate a conclusive result as they are distributed almost equally on both sides of the central line. The observation of these three setups does not align with the purpose of using "friction reduction systems" as they are designed to reduce frictional losses during the triaxial test and are expected to increase rutting measurements (Seleridis, 2017). However, as shown in Figure 9, such an expectation is only observed in the case of "two-layer rubber treated with silicon grease," as its data points are clustered on the right side of the central line.

Regarding the third influencing feature on rutting, the high values of "target bitumen mass percentage" are located on the right side of the central line (see Figure 7b). This can indicate that a higher amount of bitumen content has an increasing effect on rutting values. This finding is consistent with previous research (Erkens et al. 2014), which has reported that increased bitumen content within the mix elevates the propensity for rutting

As Figure 10 shows, the key features influencing ITS values are "sample condition," "bitumen phase angle," and "compaction degree." Observing the "sample condition" in the first place was expected as it affects the binding strength between bitumen and aggregates, consequently influencing water sensitivity (O Abd & Ramanu, 2016).

Considering that "sample condition" is a categorical variable, Figure 11 was used to investigate the relative
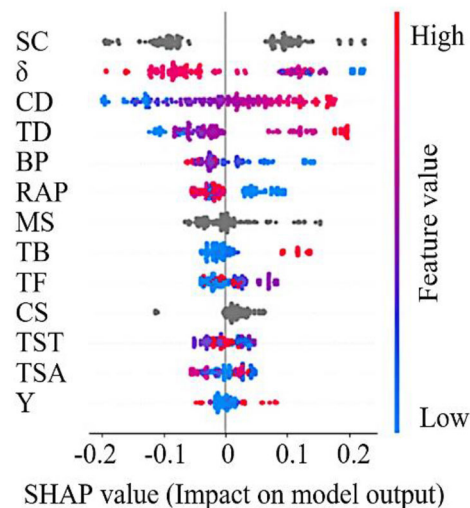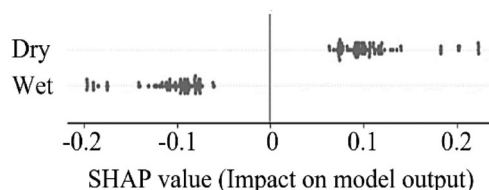
impact of different conditions on ITS values. As can be observed from the figure, data points related to "dry samples" are clustered on the right side of the central line, which signifies that higher ITS values could be observed by using dry samples. On the contrary, data points related to the "wet samples" are clustered on the left side of the central line, which indicates that lower ITS values could be expected by using wet samples. Such an observation is aligned with the previous studies (Sulejmani et al., 2019; Tarefder & Ahmad, 2015, 2017). Dry samples typically have higher ITS values due to the strength between bitumen and aggregates, which can result in a better resistance to water sensitivity. Conversely, wet samples tend to have lower ITS values as moisture weakens the bond between bitumen and aggregates and makes the mixture more susceptible to damage.

Regarding data points related to the "bitumen phase angle" (see Figure 10), although high values can be observed on both sides of the central line, the highest values are more toward the left side. This indicates that higher values of "bitumen phase angle" decrease the ITS values. Similar to what was observed with stiffness, this can be attributed to the fact that a higher phase angle indicates the material's tendency toward a more viscous state, thus

resulting in diminished elastic behavior. In other words, the material is less able to bounce back to its original shape after being subjected to tensile stress, which can lead to a lower ITS value.

For data points related to "compaction degree" (see Figure 10), it can be observed that the highest and lowest values are located on the right and left sides of the central line, respectively. This can indicate that high values of the "compaction degree" have an increasing effect, and its low values have a decreasing effect on the ITS values. This observation is aligned with the previous finding (Brown, 1990), as a lower compaction degree leads to more void content and increases the risk of cracking and water damage. Conversely, a higher compaction degree can increase ITS values since it helps in enhancing mix cohesion.

Sections 4.1 to 4.4 have provided insights into the impact of individual features on the stiffness, fatigue resistance, rutting, and ITS of asphalt mixtures through SHAP value analysis, respectively. While these insights have highlighted the significant influence of specific features under controlled conditions, further examination is necessary to understand the interaction of these features with the broader aspects of mixture conditioning, aging, and compaction processes. Section 4.5 will extend this analysis by introducing the "Phase-Year" feature, which captures the cumulative effects of mixture conditioning, aging, and compaction processes. This additional layer of analysis aims to discern between the effect of laboratory and field conditions.

4.4 | Effect of mixture conditioning on SHAP values

This section outlines the combined effect of mixture preparation, compaction process, and aging (see Table 2) as an independent variable ("Phase-Year") in the modeling of four functional performance indicators. The rationale for considering "Phase-Year" is to study the differences between the effect of laboratory and field conditions. Hence, in the following subsection, the effect of such a feature on four functional performance indicators will be studied.

4.4.1 | Mixture conditioning: Stiffness

As shown in Figure 12a, the "compaction degree" and "bitumen phase angle" are the most influencing features on stiffness. However, compared to Figure 5, the "Phase-Year" emerges as the third influencing feature,
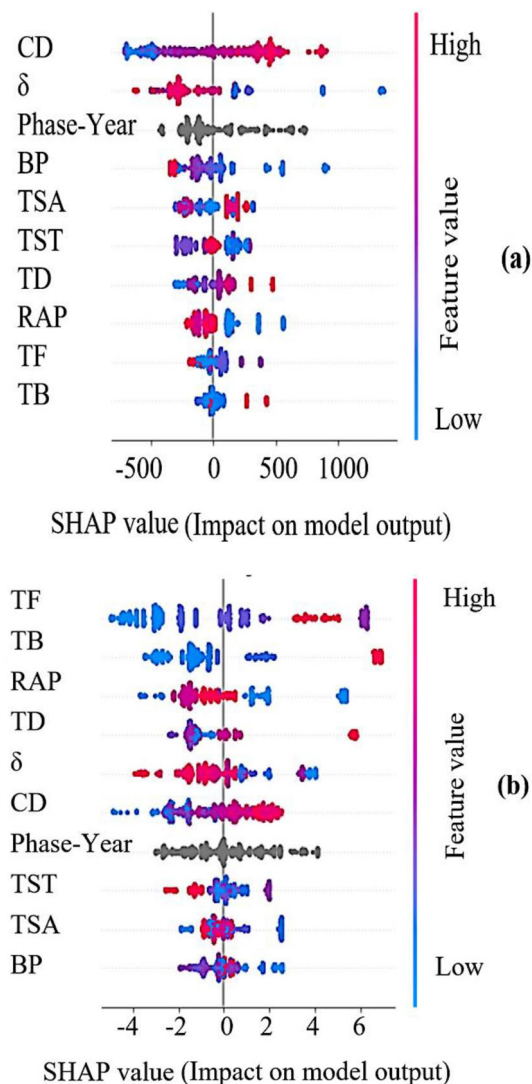

FIGURE 12 Summary SHAP plot by incorporating "Phases-Year" feature, (a) stiffness, (b) fatigue resistance (For acronyms, refer to Table 6).

which can underscore the importance of considering the combined impact of conditioning and aging on stiffness.

As "Phase-Year" is a categorical variable, Figure 13 was used to investigate the relative impact of this feature on stiffness values. As can be seen from the figure, most of the data points for "Phase 1-Year 0" and "Phase 3-Year 0" that are located on the left side of the central line have a similar range (indicated by rectangle "S"). The observed similarity seems to suggest that laboratory outcomes are consistent with field observations. Besides, since the values in "Phase 3-Year 0" to "Phase 3-Year 6" lie from the extreme left to the extreme right, it might indicate a significant contribution of aging in stiffness.
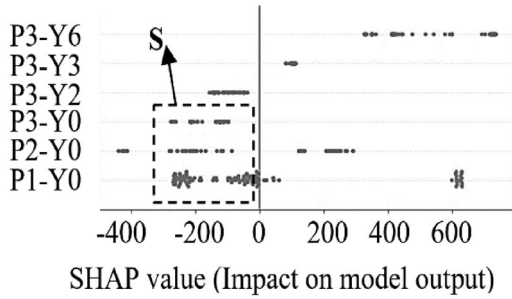
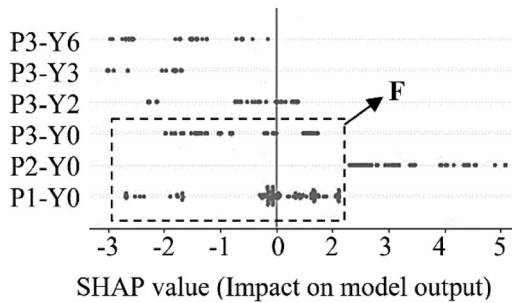**FIGURE 13** SHAP plot for different "Phases-Year" in stiffness model (For acronyms, refer to Table 2).

FIGURE 14 SHAP plot for different "Phases-Year" in fatigue resistance model (For acronyms, refer to Table 2).

4.4.2 | Mixture conditioning: Fatigue resistance

It is observed from Figure 12b that "Phase-Year" did not appear among the top three influencing features on fatigue resistance. However, like stiffness, a similar trend was observed between "Phase 1-Year 0" and "Phase 3-Year 0" (indicated by rectangle "F" in Figure 14), which can imply that laboratory outcomes are consistent with field observations. However, in contrast to stiffness, the observed trend from "Phase 3-Year 0" to "Phase 3-Year 6" can imply the decreasing effects of aging on fatigue resistance. Moreover, it can be observed that the data points across different years in Phase 3 are clustered within similar ranges, which can indicate a minor role of aging in influencing fatigue resistance.

4.4.3 | Mixture conditioning: Rutting

As can be seen from Figure 15a, "Phase-Year" is the second most influencing feature on the rutting value. Further evaluation of this feature in Figure 16 shows divergent patterns from "Phase 1-Year 0" through "Phase 3-Year 0", which can imply that laboratory results do not mirror field observations. However, it can be observed that the data points from "Phase 3-Year 0" to "Phase 3-Year 6" are clustered
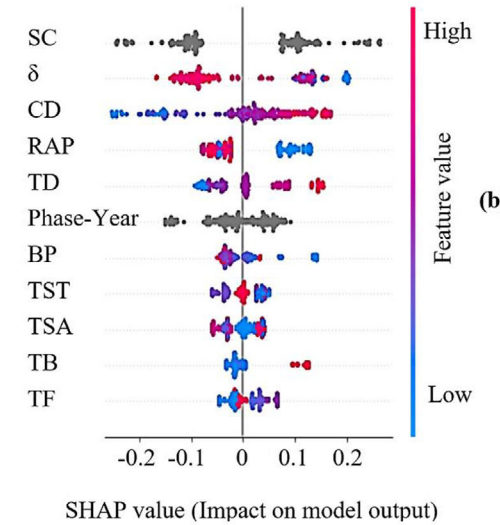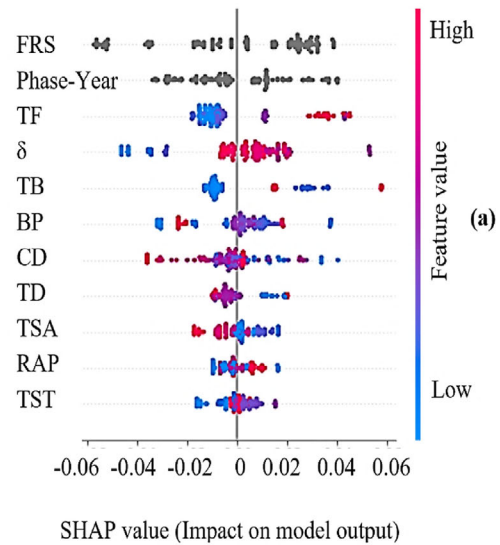

FIGURE 15 Summary SHAP plot by incorporating the "Phases-Year" features, (a) rutting, (b) ITS (For acronyms, refer to Table 6).
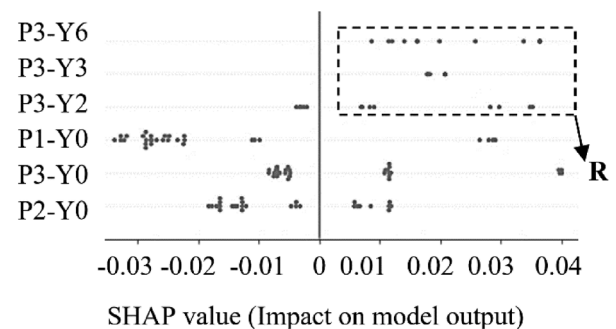

FIGURE 16 SHAP plot for different "Phases-Year" in the rutting model (For acronyms, refer to Table 2).
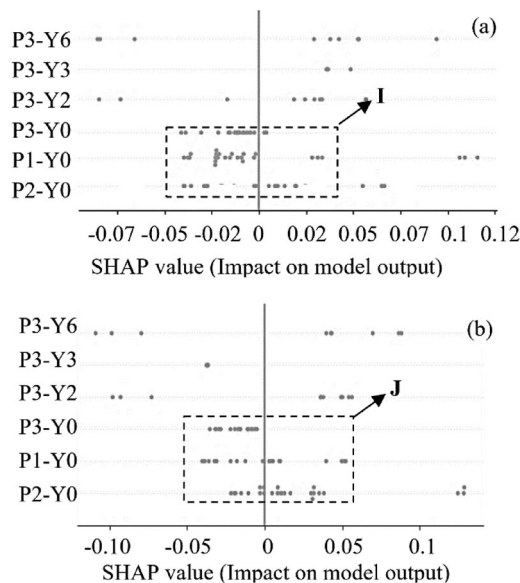
FIGURE 17 SHAP plot for ITS values, (a) dry samples, (b) wet samples (For acronyms, refer to Table 2).

on the right side of the central line (see rectangle "R"). Such an observation can imply that aging increases rutting. However, previous research (Papagiannakis & Masad, 2008) shows that aging correlates with increased material stiffness and reduced rutting. Moreover, it can be observed that the data points across different years in "Phase 3" are clustered within similar ranges, which can indicate a minor role of aging in influencing rutting.

4.4.4 | Mixture conditioning: ITS

Figure 15b shows that "Phase-Year" is not among the top three features impacting ITS value. A closer look at the information provided in Figure 17 shows that the range of data points from "Phase 1 - Year 0" to "Phase 3 - Year 0" is similar (see rectangles "I" and "J" in Figure 17a,b). Akin to stiffness and fatigue, such a similarity might suggest that laboratory results are comparable with field observation for ITS values. Besides, since no clear trend is observed from "Phase 3-Year 0" to "Phase 3-Year 6," the effects of aging on ITS value become inconclusive, and thus further data are required. Moreover, no clear difference in the effect of sample condition (i.e., dry and wet) on ITS was observed.

The section presents a connection between some key performance indicators and the effect of conditions in which the corresponding samples were prepared. However, in industry, slight deviations from the ideal recommendations are realized, leading to various uncertainties. The following section aims to explore some of the intricacies by setting hypotheses.

4.5 | Hypotheses testing

As well known within the pavement community, in practice, several assumptions have to be made due to the uncertainty in the actual condition of construction. These assumptions are often made without proven scientific support, which might cause slight deviations from standard practices and result in decreased pavement performance. Therefore, it is important for field experts to constantly evaluate and improve their practices to ensure the longevity and durability of the pavement. In this research, since the data were collected from the field and laboratory (under a controlled environment), the developed GBDT models were used to test some of these key assumptions in the form of hypotheses (see Table 13). These hypotheses were established using discussion with field experts in the Netherlands to assess whether variation in material/design characteristics leads to differences in functional performance indicators. As mentioned above, these hypotheses may not be deemed true since they are not scientifically examined.

The first hypothesis was set to study the effect of more and softer bitumen on the four functional performance indicators. In order to statistically test this hypothesis, three corresponding features, "bitumen content," "bitumen penetration," and "bitumen phase angle," are considered.

In the second hypothesis, the effect of aging on the functional performance of the mixture is evaluated using "year" as the corresponding feature. Although the second hypothesis is not directly related to the construction practices, it is important for the maintenance activities during the service life of the mixture. In the last hypothesis, the effect of density on the function performance indicators is investigated via "compaction degree" as the corresponding feature.

4.5.1 | Hypothesis testing of stiffness-related effects

It is expected that exceeding the prescribed quantity of bitumen in design or utilizing bitumen with higher penetration and higher phase angle than the designated specification will reduce stiffness modulus. In order to show the results of the analysis of the hypothesis regarding stiffness, Figure 18 was created using the SHAP dependence plot (Lundberg et al., 2019). In the SHAP dependence plot, the horizontal axis displays the values of features, and the vertical axis shows the impact of each data point on the model (i.e., SHAP value). It is noted that the zero value on the vertical axis represents the average SHAP value, and deviations from the zero point indicate either a positive or negative impact.

TABLE 13   Tested hypotheses.

Functional performance indicator	Hypothesis	Corresponding feature	Expected to be
Stiffness modulus	H1–More and softer bitumen	–%Bitumen –Bitumen penetration, –Bitumen phase angle	Decreased
	H2-Aging	–Year	Increased
	H3–Higher density	–Compaction degree	Increased
Resistance to fatigue	H4–More and Softer bitumen	–%Bitumen –Bitumen penetration, –Bitumen phase angle	Increased
	H5–Aging	–Year	Decreased
	H6–Higher density	–Compaction degree	Increased
Resistance to rutting	H7–More and Softer bitumen	–%Bitumen –Bitumen penetration, –Bitumen phase angle	Decreased
	H8–Aging	–Year	Increased
	H9–Higher density	Compaction degree	Increased
ITS	H10–More and Softer bitumen	–%Bitumen –Bitumen penetration, –Bitumen phase angle	Increased
	H11–Aging	–Year	Decreased
	H12–Higher density	–Compaction degree	Increased

The dashed curve "X" in Figure 18a indicates higher bitumen penetration (softer bitumen) reduces stiffness value, which is also aligned with the previous findings (Papagiannakis & Masad, 2008). Similar patterns were also observed for the phase angle since the higher phase angle values (more viscosity) reduce stiffness (see the dashed curve "Y" in Figure 18b). The behavior of bitumen content was also aligned with the expectation. Except for anomalies (indicated with rectangle "Z" in Figure 18c), the trend indicates that higher bitumen content decreases stiffness. It is noted that these anomalies were found to be associated with polymer-modified bitumen samples. Overall, the data shown in the Figure 18a–c proves hypothesis "H1."

From Figure 18d, it is observed that as the "Year" on the horizontal axis increases, the stiffness also increases (indicated by the dashed curve "T"). The data in the figure confirms hypothesis "H2," which also matches the previous findings that reported aging causes bitumen to become stiffer (Aguiar-Moya et al., 2017), thereby increasing the overall stiffness.

Regarding the effect of density on stiffness, as indicated with dashed curve "B" in Figure 18e, a denser mixture results in stiffer properties until a critical limit. Beyond this critical limit (marked with a red dashed line "Q"), an average decrease in stiffness can be observed. The trend of compaction degree impact on stiffness remains consistent when considering maximum density (indicated with the dashed curve "A" in Figure 18f), which means that there

is a possible maximum contribution from density to the stiffness (Mogawer et al., 2011). The pattern observed in the compaction degree plot proves the hypothesis " H3." Moreover, the presented results in Figure 18e give a deeper insight into the effect of compaction degree on stiffness and might help to identify the optimal density for mixtures. Applying an appropriate trend line could pinpoint the density level where further increases cease to augment stiffness. However, additional experiments with denser samples are required to fully comprehend the effects of exceptionally high densities on asphalt mixtures' stiffness.

4.5.2 | Hypothesis testing of fatigue-related effects

As shown by the dashed curve "O" in Figure 19a, increased bitumen penetration (i.e., softer bitumen) reduces fatigue resistance. This observation does not match the expectation that using softer bitumen would positively affect fatigue resistance. The data presented in the corresponding figure disprove the expected effect of bitumen penetration on fatigue resistance, which is also consistent with the findings of previous researchers. (Papagiannakis & Masad, 2008). Similarly, the phase angle showed contradictory results to the expectation, as a higher phase angle was associated with reduced fatigue resistance (indicated with dashed curve" L" in Figure 19b).
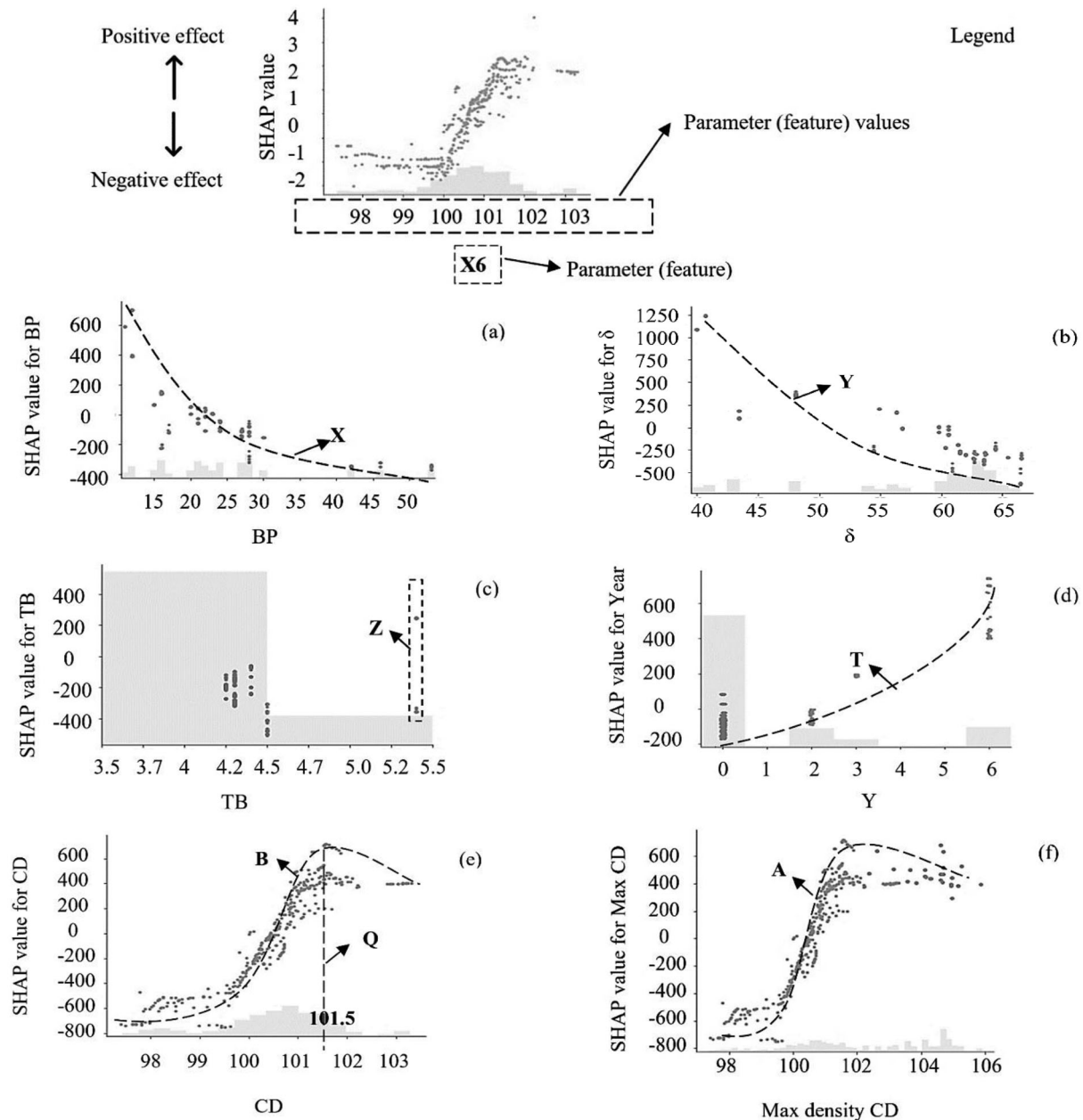
FIGURE 18 The SHAP dependence plot presents the results of testing hypotheses for stiffness (For acronyms, refer to Table 6. It is noted that max density CD is compaction degree considering maximum density.)

The pattern indicated with an arrow "A" in Figure 19c proves the expected effect of bitumen content on fatigue resistance. The presented data in the corresponding figure are also aligned with the previous studies (Erkens & van Vliet, 2014). The higher bitumen content can make a more durable and flexible pavement. As a result, asphalt mixtures with higher bitumen content could have higher fatigue resistance as they can withstand repeated loading.

From Figure 19d, it was seen that as the "Year" increases, the fatigue resistance decreases (indicated with arrow "B"), except for anomalies in the sixth year (indicated with the rectangle "Y"). The overall observed trend proves the

hypothesis "H5," which is also aligned with the previous literature (Baek et al., 2012; Papagiannakis & Masad, 2008), which reported that bitumen becomes stiffer with aging, thereby decreasing fatigue resistance.

Figure 19-e,f show results related to the effect of density on fatigue resistance. As the dashed curves "C" and "D" show in such figures, an increased compaction degree leads to higher fatigue resistance. Similar to the finding related to stiffness, fatigue resistance reaches its maximum at 101.5% compaction (marked with the dashed line "X"). The observed trend in Figure 19-e,f proves the hypothesis "H6." Additionally, the results provide a deeper
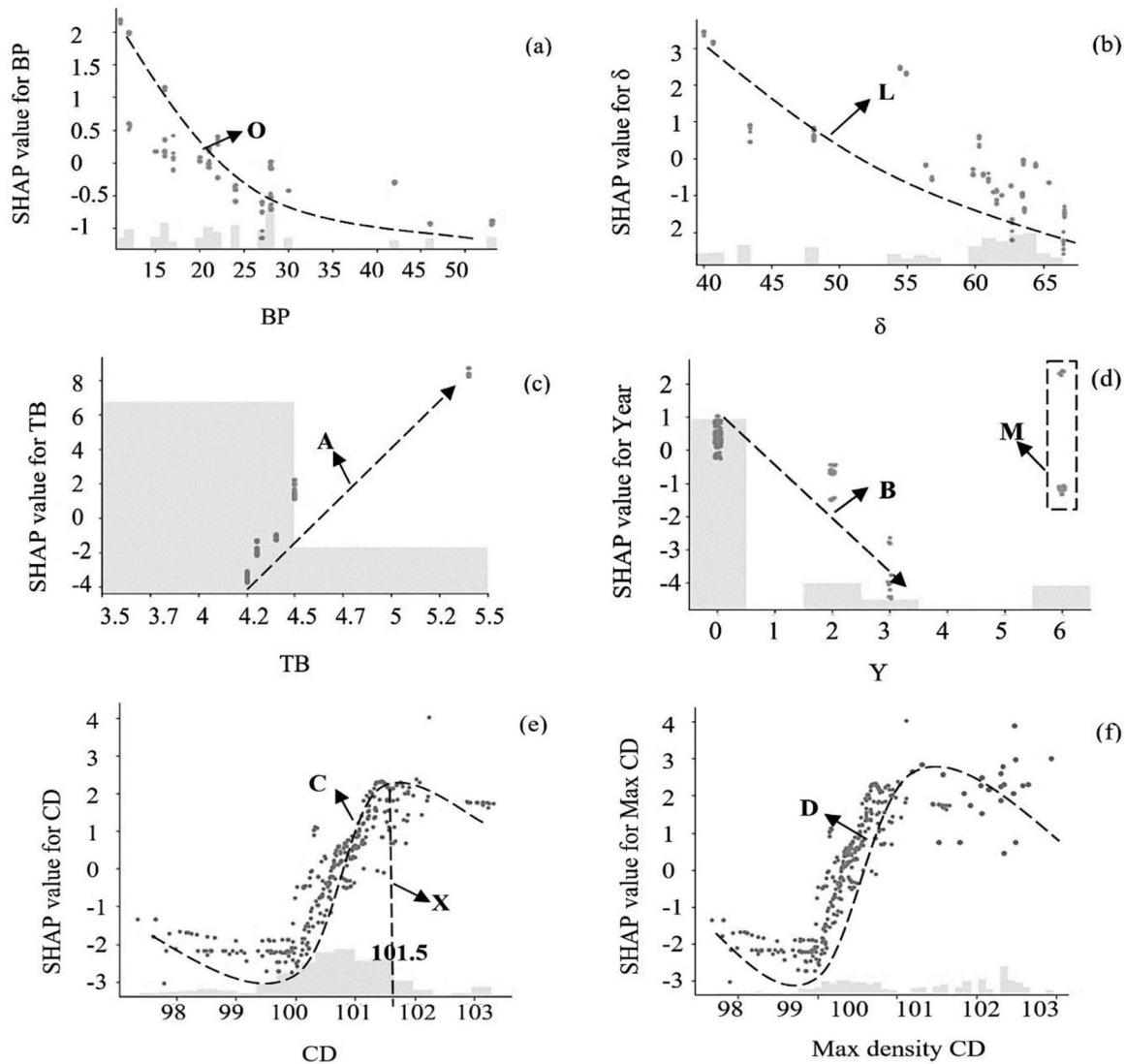
FIGURE 19 SHAP dependence plot for showing the results of testing hypotheses for fatigue resistance (For acronyms, refer to Table 6).

understanding of the impact of compaction degree on fatigue resistance and may assist in determining the optimal density for achieving maximum fatigue resistance.

4.5.3 | Hypothesis testing of rutting-related effects

Figure 20a–c shows results related to the effect of more and softer bitumen on rutting. As there is no observable clear pattern, data shown in the corresponding figures cannot either prove or disprove the hypothesis "H7." The results of studying the hypothesis regarding the effect of aging on rutting are shown in Figure 20d.

The data shown in the figure disprove the hypothesis "H8," which is also in accordance with the previous research findings (Babadopulos et al., 2016). Aging is

expected to stiffen the bitumen and improve its rutting resistance. However, from the figure, it can be seen that an increase in the "Year" feature has a positive correlation with rutting (indicated with the dashed arrow "F"). The findings of studying the impact of higher density on rutting are presented in Figure 20e,f. Since no clear pattern can be seen from the figures, the hypothesis "H9" can be neither proved nor disproved. Therefore, further research should be carried out.

4.5.4 | Hypothesis testing of ITS effects

The findings from the study on the impact of more and softer bitumen on ITS values are illustrated in Figure 21a–c. In contrast to the expectation, the higher penetration and phase angle decrease the ITS values as marked with dashed
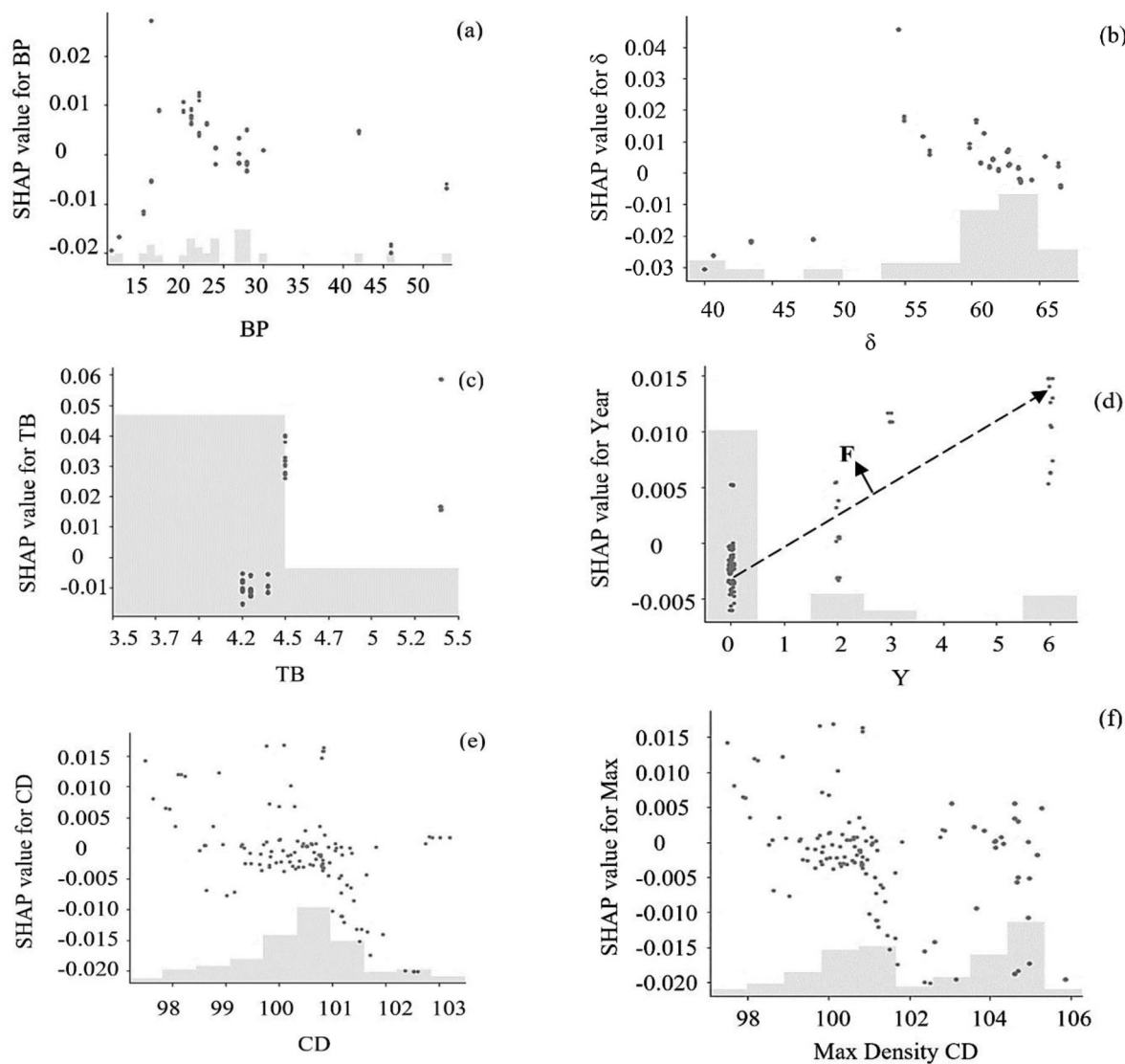
FIGURE 20 SHAP dependence plot for showing the results of testing hypotheses for rutting (For acronyms, refer to Table 6).

cure "H" and "P" in Figure 21a,b, respectively. However, the results for the effect of bitumen content were aligned with the expectation. As indicated with the dashed arrow "K" in Figure 21c, higher bitumen content increases ITS values. The findings regarding testing the hypothesis about the impact of aging on ITS values can be observed in Figure 21d. Since no clear pattern can be seen through the years, the hypothesis "H11" cannot be proved or disproved.

Figure 21e,f shows findings about the impact of density on ITS values, and the observed trend in the figures proves the hypothesis " H12." Like stiffness and fatigue, the increasing effect of compaction degree on ITS values is limited. The maximum ITS value can be reached at the 102% compaction degree (indicated with the dashed line "U") with no further increment beyond this critical point. It is noted that the variation in SHAP values among wet and dry

conditions appears minimal, suggesting water sensitivity in samples is low.

4.6 | Summary of the results

The ML-based models identify the relative importance of features in studying the reliability of lab-prepared samples, which were consistent with the findings of previous research studies. It was found that the results obtained for stiffness, fatigue, and ITS show consistent results from the samples obtained in the field and the samples prepared in the laboratory. These indicate that the compaction and mixing setups used in the laboratory simulate well in-field compaction and mixing practices.

It is noted that the results regarding rutting were found to be inconclusive, which means that the data did not
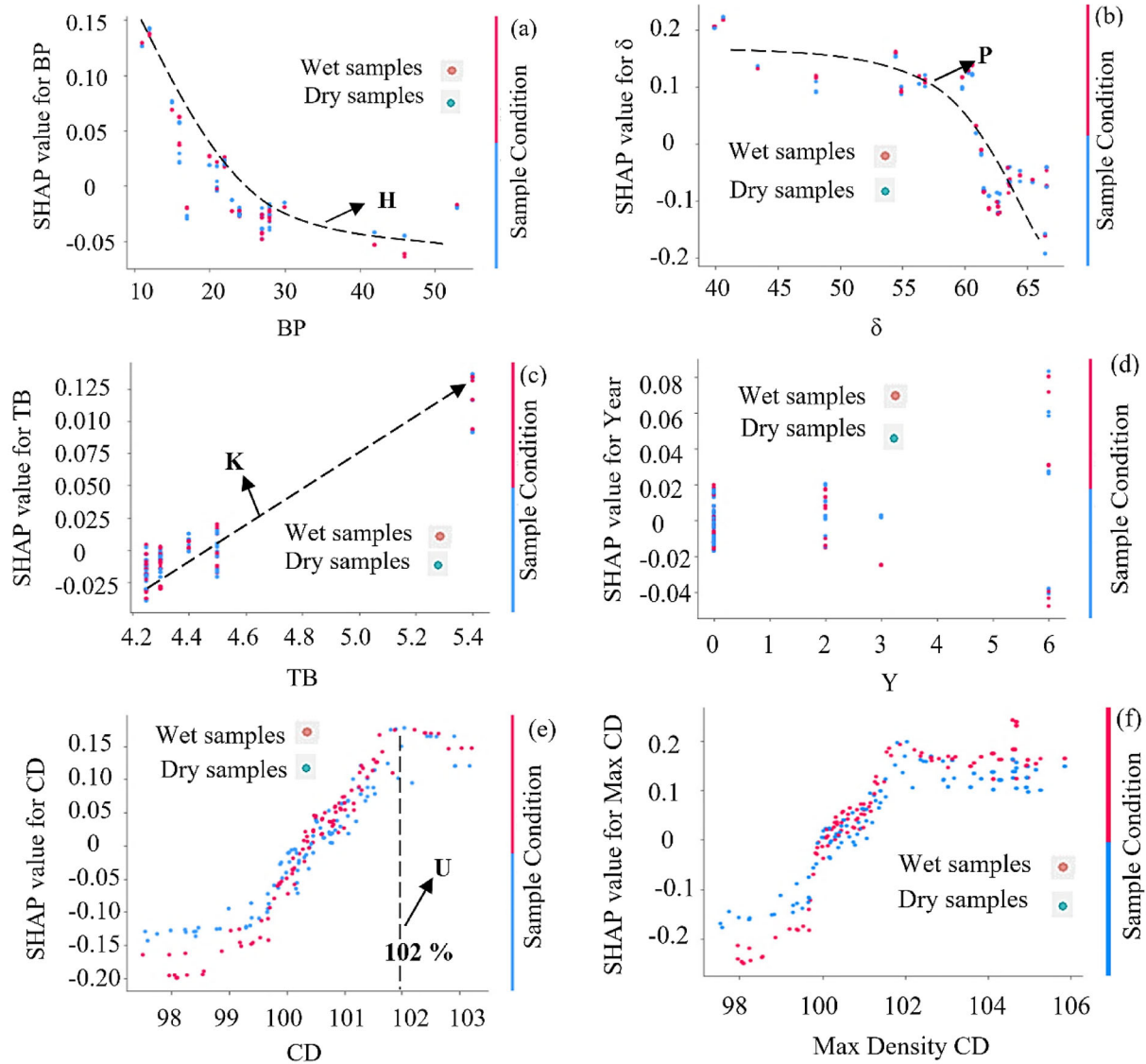
FIGURE 21 SHAP dependence plot for illustrating the results of testing hypotheses for ITS (For acronyms, refer to Table) Radial basis function.

convincingly show whether the laboratory compaction and mixing setup could simulate the field condition. Therefore, further data and investigation are required.

Since, in practice, it is extremely difficult to control various conditions compared to laboratory environments, these variations may result in functional properties that are different from expectations. To further study the differences between field and laboratory setups, various hypotheses were tested. The formulations of hypotheses were done based on the notions of experts from their field experiences. The key highlight of the hypothesis testing is as follows:

1. Higher bitumen content or using softer bitumen type leads to reduced stiffness;
2. Samples with higher density result in higher stiffness;
3. Using softer bitumen does not lead to higher fatigue resistance and ITS values;
4. Aged samples result in lower fatigue resistance;
5. Denser samples show higher fatigue resistance;
6. The results of this research could neither prove nor disprove the effect of aging on ITS values. Hence, further research is required.
7. No clear indication is found if higher bitumen content or using softer bitumen results in improved rutting resistance.

5 | CONCLUSION

The main ambition of the research was to propose an ML-based future-ready toolkit to support better pavement

maintenance strategies considering modern-day challenges. It was discussed that due to some of the modern-day challenges, the current experience-based pavement maintenance strategies may become unreliable. To solve these issues, the pavement community is carrying out research in multi-fold directions (sustainable materials, green materials, improved design standards, etc.), which mostly rely on laboratory-based information. Researchers have identified that the gap between in-field performance and laboratory measurements is one of the key hurdles. Hence, the aim of this research is to better understand the difference between the functional performance of samples prepared in the laboratory and the field. Since a future-ready toolkit is expected to utilize big data, in this research, the ML-based GBDT approach was followed.

Datasets containing features such as material properties, mixing and compaction setups, and so forth from six actual construction projects were collected. Using the dataset as input, GBDT toolkits were trained and optimized through CV and Bayesian optimization. Since a limited dataset was obtained, as the first step, the results of the GBDT model were compared with the statistical model as a benchmark and subsequently with other ML models to validate its performance. The GBDT-based toolkit is expected to provide a better interpretation of the results (particularly with big data).

GBDT-based outcomes were summarized in this article.

The results demonstrated that the GBDT model achieved high predictive performance with R^2 values of 0.951 for stiffness, 0.802 for fatigue resistance, 0.812 for rutting, and 0.842 for ITS on the testing data. These results indicate a strong correlation between predicted and actual values, suggesting the effectiveness of the GBDT approach in capturing the underlying patterns in the data. Furthermore, the normalized RMSE values were low, with 0.0443 for stiffness, 0.0651 for fatigue resistance, 0.0677 for rutting, and 0.0687 for ITS, indicating that the predictions across different functional performance indicators were fairly accurate. A few critical observations showed that some of the commonly accepted notions might be incorrect.

The overall framework proposed in this research is expected to be an important pathway for different transportation practitioners because the framework can be adapted to solve further complex issues. The adaptability of the framework is important because the particular tests conducted on the samples may not be fully replicable in different countries because of varying standard requirements. A simple solution could be to tailor the model parameters according to the needs of different agencies. Another application of the proposed framework is to improve pavement maintenance strategies via implementation inside the existing PMS of different agencies.

Better managerial insights can obtained after such an implementation using user-friendly software.

6 | FUTURE RESEARCH DIRECTION AND RECOMMENDATIONS

With the contributions of this research, as explained in the previous sections, there are opportunities to extend this research. Exploring the impact of other features, such as traffic and environmental conditions, can provide a more comprehensive understanding of the effects on pavement functional performance.

In future research, it is recommended that the stability of the proposed framework is assessed under various sets of input conditions. This could be achieved for example by incorporation of dynamic ensemble learning techniques (Alam et al., 2020), which dynamically allows for the parameters to be adjusted. Moreover, utilizing self-supervised learning methods (Rafiei et al., 2022) can also provide better initialization for parameters like weights, improving upon random initialization used in this research. Incorporating such newly developed methods may lead to more robust and adaptive GBDT architecture models in non-stationary environments where the data distribution continuously changes over time.

REFERENCES

Abu Al-Rub, R. K., & Darabi, M. K. (2012). A thermodynamic framework for constitutive modeling of time- and rate-dependent materials. Part I: Theory. *International Journal of Plasticity*, *34*, 61–92. https://doi.org/10.1016/j.ijplas.2012.01.002

Adeli, H., & Hung, S.-L. (1994). *Machine learning: Neural networks, genetic algorithms, and fuzzy systems*. John Wiley & Sons, Inc.

Adeli, H., & Yeh, C. (1989). Perceptron learning in engineering design. *Computer-Aided Civil and Infrastructure Engineering*, *4*(4), 247–256. https://doi.org/10.1111/j.1467-8667.1989.tb00026.x

Aguiar-Moya, J. P., Salazar-Delgado, J., García, A., Baldi-Sevilla, A., Bonilla-Mora, V., & Loría-Salazar, L. G. (2017). Effect of ageing on micromechanical properties of bitumen by means of atomic force microscopy. *Road Materials and Pavement Design*, *18*(sup2), 203–215. https://doi.org/10.1080/14680629.2017.1304249

Airey, G. D., & Collop, A. C. (2016). Mechanical and structural assessment of laboratory- and field-compacted asphalt mixtures. *International Journal of Pavement Engineering*, *17*(1), 50–63. https://doi.org/10.1080/10298436.2014.925551

Alam, K. M. R., Siddique, N., & Adeli, H. (2020). A dynamic ensemble learning algorithm for neural networks. *Neural Computing and Applications*, *32*(12), 8675–8690. https://doi.org/10.1007/s00521-019-04359-7

Alexandropoulos, S.-A. N., Kotsiantis, S. B., & Vrahatis, M. N. (2019). Data preprocessing in predictive data mining. *The Knowledge Engineering Review*, *34*, e1. https://doi.org/10.1017/S026988891800036X

Alimoradi, S., Golroo, A., & Asgharzadeh, S. M. (2022). Development of pavement roughness master curves using Markov chain. *International Journal of Pavement Engineering*, *23*(2), 453–463. https://doi.org/10.1080/10298436.2020.1752917

Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., & Ridella, S. (2012). The 'K' in K-fold Cross Validation. *ESANN*, *102*, 441–446.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., & Benjamins, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115.

Babadopulos, L. F. D. A. L., Ferreira, J. L. S., & Soares, J. B. (2016). An approach to couple aging to stiffness and permanent deformation modeling of asphalt mixtures. *Materials and Structures*, *49*(12), 4929–4945. https://doi.org/10.1617/s11527-016-0834-4

Baek, C., Underwood, B. S., & Kim, Y. R. (2012). Effects of oxidative aging on asphalt mixture properties. *Transportation Research Record: Journal of the Transportation Research Board*, *2296*(1), 77–85. https://doi.org/10.3141/2296-08

Barua, L., Zou, B., Noruzoliaee, M., & Derrible, S. (2021). A gradient boosting approach to understanding airport runway and taxiway pavement deterioration. *International Journal of Pavement Engineering*, *22*(13), 1673–1687. https://doi.org/10.1080/10298436.2020.1714616

Basnet, K. S., Shrestha, J. K., Shrestha, R. N., Basnet, K. S., Shrestha, J. K., & Shrestha, R. N. (2023). Pavement performance model for road maintenance and repair planning: A review of predictive techniques. *Digital Transportation and Safety*, *2*(4), Article DTS-2023-0021. https://doi.org/10.48130/DTS-2023-0021

Belle, V., & Papantonis, I. (2021). Principles and practice of explainable machine learning. *Frontiers in Big Data, 4*, 688969. https://www.frontiersin.org/articles/10.3389/fdata.2021.688969

Bergstra, J., Yamins, D., & Cox, D. (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. *International Conference on Machine Learning*, Atlanta, GA (pp. 115–123).

Box, G. E., Hunter, J. S., & Hunter, W. G. (2005). Statistics for experimenters. In Wiley series in probability and statistics. Hoboken, NJ: Wiley.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L. (2017). *Classification and regression trees*. Routledge.

Brown, E. (1990). Density of asphalt concrete—How much is needed?

Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., Naumann, F., & Harmouch, H. (2022). *The effects of data quality on machine learning performance*. arXiv. https://doi.org/10.48550/arXiv.2207.14529

Cawley, G. C., & Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research*, *11*, 2079–2107.

Cerda, P., & Varoquaux, G. (2022). Encoding high-cardinality string categorical variables. *IEEE Transactions on Knowledge and Data Engineering*, *34*(3), 1164–1176. https://doi.org/10.1109/TKDE.2020.2992529

Chen, J., de Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., Van Donkelaar, A., Hvidtfeldt, U. A., & Katsouyanni, K. (2019). A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. *Environment International*, *130*, 104934.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1023/A:1022627411411

Darabi, M. K., Al-Rub, R. K. A., Masad, E. A., & Little, D. N. (2012). Thermodynamic-based model for coupling temperature-dependent viscoelastic, viscoplastic, and viscodamage constitutive behavior of asphalt mixtures. *International Journal for Numerical and Analytical Methods in Geomechanics*, *36*(7), 817–854. https://doi.org/10.1002/nag.1030

Deng, Y., & Shi, X. (2023). Modeling the rutting performance of asphalt pavements: A review. *Journal of Infrastructure Preservation and Resilience*, *4*(1), 17. https://doi.org/10.1186/s43065-023-00082-9

Deng, Y., Wang, H., & Shi, X. (2024). Physics-guided neural network for predicting asphalt mixture rutting with balanced accuracy, stability and rationality. *Neural Networks*, *172*, 106085. https://doi.org/10.1016/j.neunet.2023.12.039

Ding, C., (Jason) Cao, X., & Næss, P. (2018). Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. *Transportation Research Part A: Policy and Practice*, *110*, 107–117. https://doi.org/10.1016/j.tra.2018.02.009

Dong, S., Yuan, X.-X., & Hao, P. (2020). An advanced local calibration method for mechanistic-empirical pavement design. *Computer-Aided Civil and Infrastructure Engineering*, *35*(11), 1276–1290. https://doi.org/10.1111/mice.12574

Eberly, L. E. (2007). Multiple linear regression. In W. T. Ambrosius (Ed.), *Topics in biostatistics* (pp. 165–187). Humana Press. https://doi.org/10.1007/978-1-59745-530-5_9

EN 12607-1.2014—Bitumen and bituminous binders—Determination of the resistance to hardening under influence of heat and air—Part 1: RTFOT method (Standard EN 12607-1). (2014). *EN 12607-1.2014—Bitumen and bituminous binders—Determination of the resistance to hardening under influence of heat and air—Part 1: RTFOT method (Standard EN 12607-1)*. European Committee for Standardization. https://standards.iteh.ai/catalog/standards/cen/5e1f0c0a-99bc-4c99-a191-cf07c1482527/en-12607-1-2014

EN 12697-12.2018—Bituminous mixtures—Test methods—Part 12: Determination of the water sensitivity of bituminous specimens (Standard EN 12697-12; Version 2018). (2018). *EN 12697-12.2018—Bituminous mixtures—Test methods—Part 12: Determination of the water sensitivity of bituminous specimens (Standard EN 12697-12; Version 2018)*. European Committee for Standardization. https://standards.iteh.ai/catalog/standards/sist/7e1c3a7f-fa4d-4a39-942f-45bd488c5f86/sist-en-12697-12-2018

EN 12697-23.2017—Bituminous mixtures—Test methods—Part 23: Determination of the indirect tensile strength of bituminous specimens (Standard EN 12697-23; Version 2017). (2017). *EN 12697-23.2017—Bituminous mixtures—Test methods—Part 23: Determination of the indirect tensile strength of bituminous specimens (Standard EN 12697-23; Version 2017)*. European Committee for Standardization. https://standards.iteh.ai/catalog/standards/cen/4920c3b0-a083-4d95-acb7-1c9a3fc1e0c5/en-12697-23-2017

COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING **WILEY**

EN 12697-24.2018—Bituminous mixtures—Test methods—Part 24: Resistance to fatigue (Standard EN 12697-24; Version 2018). (2018). *EN 12697-24.2018—Bituminous mixtures—Test methods—Part 24: Resistance to fatigue (Standard EN 12697-24; Version 2018)*. European Committee for Standardization. https://standards.iteh.ai/catalog/standards/cen/bc8b8967-9d2c-4a6e-8738-8978c77da5db/en-12697-24-2018

EN 12697-25.2016—Bituminous mixtures—Test methods—Part 25: Cyclic compression test (Standard EN 12697-25; Version 2016). (2016). *EN 12697-25.2016—Bituminous mixtures—Test methods—Part 25: Cyclic compression test (Standard EN 12697-25; Version 2016)*. European Committee for Standardization. https://standards.iteh.ai/catalog/standards/cen/cffade27-8ed5-4960-bd3a-ad6bb201b9ec/en-12697-25-2016

EN 12697-26.2012 -Bituminous mixtures—Test methods for hot mix asphalt—Part 26: Stiffness (Standard EN 12697-26; Version 2012). (2012). *EN 12697-26.2012 -Bituminous mixtures—Test methods for hot mix asphalt—Part 26: Stiffness (Standard EN 12697-26; Version 2012)*. European Committee for Standardization. https://standards.iteh.ai/catalog/standards/cen/6e19a3f7-d3aa-4906-b631-882b28846217/en-12697-26-2012

Erkens, S., Stigter, J., Sluer, B., Khedoe, RN., van der Wall, A., & de Bondt, A. (2014). NL-LAB: onderzoek naar de voorspellende waarde van proef 62. In CROW InfraDagen (pp. 1). CROW. https://pure.tudelft.nl/ws/files/17792328/2014_4_065_A_NL_LAB_proef_62_infradagen_2014_DEF.pdf

Erkens, S., & van Vliet, D. (2014). *De meerwaarde van structureel, langjarig bemonsteren*. In CROW Infradagen. https://research.tudelft.nl/files/4614805/2014_6_NL_LAB_infradagen_2014DEF_meerwaarde_structureel_meerjarig_bemonsteren.pdf

Fan, C., Xu, J., Natarajan, B. Y., & Mostafavi, A. (2023). Interpretable machine learning learns complex interactions of urban features to understand socio-economic inequality. *Computer-Aided Civil and Infrastructure Engineering*, *38*(14), 2013–2029. https://doi.org/10.1111/mice.12972

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Fuks, O., & Tchelepi, H. A. (2020). Limitations of physics informed machine learning for nonlinear two-phase transport in porous media. *Journal of Machine Learning for Modeling and Computing*, *1*(1), 19–37. https://doi.org/10.1615/JMachLearnModelComput.2020033905

García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining* (Vol. 72). Springer International Publishing. https://doi.org/10.1007/978-3-319-10247-4

Gartner, W. (1989). Asphalt concrete mix design: Development of more rational approaches. ASTM International.

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, *521*(7553), Article 7553. https://doi.org/10.1038/nature14541

Gong, H., Sun, Y., Hu, W., Polaczyk, P. A., & Huang, B. (2019). Investigating impacts of asphalt mixture properties on pavement performance using LTPP data through random forests. *Construction and Building Materials*, *204*, 203–212. https://doi.org/10.1016/j.conbuildmat.2019.01.198

Gong, H., Sun, Y., & Huang, B. (2019). Gradient Boosted Models for Enhancing Fatigue Cracking Prediction in Mechanistic-Empirical Pavement Design Guide. *Journal of Transportation Engineering*, Part B: Pavements, *145*(2), 04019014. https://doi.org/10.1061/jpeodx.0000121

Guo, R., Fu, D., & Sollazzo, G. (2022). An ensemble learning model for asphalt pavement performance prediction based on gradient boosting decision tree. *International Journal of Pavement Engineering*, *23*(10), 3633–3646. https://doi.org/10.1080/10298436.2021.1910825

Haas, R., & Hudson, W. R. (2015). *Pavement asset management*. John Wiley & Sons.

Haddad, A. J., Chehab, G. R., & Saad, G. A. (2022). The use of deep neural networks for developing generic pavement rutting predictive models. *International Journal of Pavement Engineering*, *23*(12), 4260–4276. https://doi.org/10.1080/10298436.2021.1942466

Hallin, J. (2004). *Development* of the 2002 guide for the design of new and rehabilitated pavement structures: Phase II. Transportation Research Board. http://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=218

Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: An interdisciplinary review. *Journal of Big Data*, *7*(1), 94. https://doi.org/10.1186/s40537-020-00369-8

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Vol. 2). Springer.

Hong, F., & Prozzi, J. A. (2006). Estimation of pavement performance deterioration using bayesian approach. *Journal of Infrastructure Systems*, *12*(2), 77–86. https://doi.org/10.1061/(ASCE)1076-0342(2006)12:2(77)

Hou, Y., Li, Q., Zhang, C., Lu, G., Ye, Z., Chen, Y., Wang, L., & Cao, D. (2021). The state-of-the-art review on applications of intrusive sensing, image processing techniques, and machine learning methods in pavement monitoring and *analysis. Engineering*, *7*(6), 845–856. https://doi.org/10.1016/j.eng.2020.07.030

Hu, A., Bai, Q., Chen, L., Meng, S., Li, Q., & Xu, Z. (2022). A review on empirical methods of pavement performance modeling. *Construction and Building Materials*, *342*, 127968. https://doi.org/10.1016/j.conbuildmat.2022.127968

Huang, B., Mohammad, L. N., & Rasoulian, M. (2001). Three-dimensional numerical simulation of asphalt pavement at Louisiana Accelerated Loading Facility. *Transportation Research Record*, *1764*(1), 44–58. https://doi.org/10.3141/1764-06

Huang, B., Shu, X., & Chen, X. (2007). Effects of mineral fillers on hot-mix asphalt laboratory-measured properties. *International Journal of Pavement Engineering*, *8*(1), 1–9. https://doi.org/10.1080/10298430600819170

Huang, W., Liang, S., & Wei, Y. (2020). Surface deflection-based reliability analysis of asphalt pavement design. *Science China Technological Sciences*, *63*(9), 1824–1836.

Hunter, A. E., Airey, G. D., & Harireche, O. (2007). Numerical modeling of asphalt mixture wheel tracking experiments. *International Journal of Pavement Engineering and Asphalt Technology*, *8*, 52–71.

Iskender, E., & Aksoy, A. (2012). Field and laboratory performance comparison for asphalt mixtures with different moisture conditioning systems. *Construction and Building Materials*, *27*(1), 45–53. https://doi.org/10.1016/j.conbuildmat.2011.08.019

Jiménez, L. A., & Mrawira, D. (2012). Bayesian Regression in Pavement Deterioration Modeling: Revisiting the AASHO Road Test Rut Depth Model. *Infraestructura Vial*, *14*(25), Article 25.

Jung, T., & Kim, J. (2023). A new support vector machine for categorical features. *Expert Systems with Applications*, *229*, 120449. https://doi.org/10.1016/j.eswa.2023.120449

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

Kargah-Ostadi, N., Vasylevskyi, K., Ablets, A., & Drach, A. (2023). Reconciling pavement condition data from connected vehicles with the International Roughness Index from standard monitoring equipment using physics-integrated machine learning. *Transportation Research Record*, *2678*(2), 416–429. https://doi.org/10.1177/03611981231174406

Kaufman, S., Rosset, S., Perlich, C., & Stitelman, O. (2012). Leakage in data mining: Formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data*, *6*(4), 15:1–15:21. https://doi.org/10.1145/2382577.2382579

Kerali, H. G., Odoki, J. B., & Stannard, E. E. (2000). *Overview of HDM-4. The highway development and management series* (Vol. 1). World Road Association, PIARC. World Bank.

Kettil, P., Lenhof, B., Runesson, K., & Wiberg, N.-E. (2007). Simulation of inelastic deformation in road structures due to cyclic mechanical and thermal loads. *Computers & Structures*, *85*(1), 59–70. https://doi.org/10.1016/j.compstruc.2006.08.060

Khattak, M. J., Landry, C., Veazey, J., & Zhang, Z. (2013). Rigid and composite pavement index-based performance models for network pavement management system in the state of Louisiana. *International Journal of Pavement Engineering*, *14*(7), 612–628. https://doi.org/10.1080/10298436.2012.715643

Kunanbayev, K., Temirbek, I., & Zollanvari, A. (2021). Complex encoding. In *2021 International Joint Conference on Neural Networks (IJCNN)*, (pp. 1–6). IEEE. https://doi.org/10.1109/IJCNN52387.2021.9534094

Lee, D., Derrible, S., & Pereira, F. C. (2018). Comparison of Four Types of Artificial Neural Network and a Multinomial Logit Model for Travel Mode Choice Modeling. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(49), 101–112. https://doi.org/10.1177/0361198118796971

Lee, N., & Kim, J.-M. (2010). Conversion of categorical variables into numerical variables via Bayesian network classifiers for binary classifications. *Computational Statistics & Data Analysis*, *54*(5), 1247–1265. https://doi.org/10.1016/j.csda.2009.11.003

Li, K., Pan, L., Guo, X., & Wang, Y. F. (2024). Hybrid random aggregation model and Bayesian optimization-based convolutional neural network for estimating the concrete compressive strength. *Computer-Aided Civil and Infrastructure Engineering*, *39*(4), 559–574. https://doi.org/10.1111/mice.13096

Li, Q., Xiao, D. X., Wang, K. C. P., Hall, K. D., & Qiu, Y. (2011). Mechanistic-empirical pavement design guide (MEPDG): A bird's-eye view. *Journal of Modern Transportation*, *19*(2), 114–133. https://doi.org/10.1007/BF03325749

Li, Z. (2022). Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. *Computers, Environment and Urban Systems*, *96*, 101845. https://doi.org/10.1016/j.compenvurbsys.2022.101845

Liang, M., Chang, Z., He, S., Chen, Y., Gan, Y., Schlangen, E., & Šavija, B. (2022). Predicting early-age stress evolution in restrained concrete by thermo-chemo-mechanical model and active ensemble learning. *Computer-Aided Civil and Infrastructure Engineering*, *37*(14), 1809–1833. https://doi.org/10.1111/mice.12915

Liu, J., Liu, F., Zheng, C., Zhou, D., & Wang, L. (2022a). Optimizing asphalt mix design through predicting effective asphalt content and absorbed asphalt content using machine learning. *Construction and Building Materials*, *325*, 126607. https://doi.org/10.1016/j.conbuildmat.2022.126607

Liu, J., Liu, F., Zheng, C., Zhou, D., & Wang, L. (2022b). Optimizing asphalt mix design through predicting the rut depth of asphalt pavement using machine learning. *Construction and Building Materials*, *356*, 129211. https://doi.org/10.1016/j.conbuildmat.2022.129211

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2019). *Consistent individualized feature attribution for tree ensembles*. arXiv. https://doi.org/10.48550/arXiv.1802.03888

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, Long Beach, CA.

Ma, T., Zhang, D., Zhang, Y., Wang, S., & Huang, X. (2018). Simulation of wheel tracking test for asphalt mixture using discrete element modelling. *Road Materials and Pavement Design*, *19*(2), 367–384. https://doi.org/10.1080/14680629.2016.1261725

Ma, X., Ding, C., Luan, S., Wang, Y., & Wang, Y. (2017). Prioritizing influential factors for freeway incident clearance time prediction using the gradient boosting decision trees method. *IEEE Transactions on Intelligent Transportation Systems*, *18*(9), 2303–2310. https://doi.org/10.1109/TITS.2016.2635719

Marcelino, P., de Lurdes Antunes, M., Fortunato, E., & Gomes, M. C. (2021). Machine learning approach for pavement performance prediction. *International Journal of Pavement Engineering*, *22*(3), 341–354. https://doi.org/10.1080/10298436.2019.1609673

Meng, C., Seo, S., Cao, D., Griesemer, S., & Liu, Y. (2022). *When physics meets machine learning: A survey of physics-informed machine learning*. arXiv. https://doi.org/10.48550/arXiv.2203.16797

Mizutani, D., & Yuan, X.-X. (2023). Infrastructure deterioration modeling with an inhomogeneous continuous time Markov chain: A latent state approach with analytic transition probabilities. *Computer-Aided Civil and Infrastructure Engineering*, *38*(13), 1730–1748. https://doi.org/10.1111/mice.12976

Mogawer, W. S., Austerman, A. J., Daniel, J. S., Zhou, F., & Bennert, T. (2011). Evaluation of the effects of hot mix asphalt density on mixture fatigue performance, rutting performance and MEPDG distress predictions. *International Journal of Pavement Engineering*, *12*(2), 161–175. https://doi.org/10.1080/10298436.2010.546857

Mousavi Rad, S., Kamboozia, N., Anupam, K., & Saed, S. A. (2022). Experimental evaluation of the fatigue performance and self-healing behavior of nanomodified porous asphalt mixtures containing RAP Materials under the aging condition and freeze–thaw cycle. *Journal of Materials in Civil Engineering*, *34*(12), 04022323. https://doi.org/10.1061/(ASCE)MT.1943-5533.0004488

NEN-EN 1426. 2015—Bitumen en bitumineuze bindmiddelen—Bepaling van de naaldpenetratie (Standard NEN-EN 1426; Version 2015). (2015). *NEN-EN 1426:2015—Bitumen en bitumineuze bindmiddelen—Bepaling van de naaldpenetratie (Standard NEN-EN 1426; Version 2015*. Nederlands Normalisatie-instituut (NEN). https://www.nen.nl/nen-en-1426-2015-en-208859

NEN-EN 1427. 2015—Bitumen en bitumineuze bindmiddelen—Bepaling van het verzachtingspunt—Ring- en kogelmethode (Standard NEN-EN 1427; Version 2015). (2015). *NEN-EN 1427:2015—Bitumen en bitumineuze bindmiddelen—Bepaling van het verzachtingspunt—Ring- en kogelmethode (Standard NEN-EN 1427; Version 2015)*. Nederlands Normalisatie-instituut (NEN). https://www.nen.nl/nen-en-1427-2015-en-208860

NEN-EN 14770. 2022 -Bitumen and bitumineuze bindmiddelen—Bepaling van complexe afschuifmodulus en fasehoek—Dynamische Afschuif Rheometer (DSR) (Standard NEN-EN 14770; Version 2022). (2022). *NEN-EN 14770:2022 -Bitumen and bitumineuze bindmiddelen—Bepaling van complexe afschuifmodulus en fasehoek—Dynamische Afschuif Rheometer (DSR) (Standard NEN-EN 14770; Version 2022)*. Nederlands Normalisatie-instituut (NEN). https://www.nen.nl/nen-en-14770-2022-ontw-en-293420

Nghiem, T. X., Drgoňa, J., Jones, C., Nagy, Z., Schwan, R., Dey, B., Chakrabarty, A., Di Cairano, S., Paulson, J. A., & Carron, A. (2023). Physics-informed machine learning for modeling and control of dynamical systems. *2023 American Control Conference (ACC)*, San Diego, CA (pp. 3735–3750). https://ieeexplore.ieee.org/abstract/document/10155901/

Nyirandayisabye, R., Li, H., Dong, Q., Hakuzweyezu, T., & Nkinahamira, F. (2022). Automatic pavement damage predictions using various machine learning algorithms: Evaluation and comparison. *Results in Engineering*, 16, 100657. https://doi.org/10.1016/j.rineng.2022.100657

O Abd, E. H., & Ramanu, M. (2016). Stripping distress on hot mixed asphalt pavement. GRD Journals | Global Research and Development Journal for Engineering | Recent Advances in Civil Engineering for Global.

Okada, S., Ohzeki, M., & Taguchi, S. (2019). Efficient partition of integer optimization problems with one-hot encoding. *Scientific Reports*, 9(1), 13036. https://doi.org/10.1038/s41598-019-49539-6

Pan, N.-F., Ko, C.-H., Yang, M.-D., & Hsu, K.-C. (2011). Pavement performance prediction through fuzzy regression. *Expert Systems with Applications*, 38(8), 10010–10017. https://doi.org/10.1016/j.eswa.2011.02.007

Papagiannakis, A. T., & Masad, E. A. (2008). *Pavement design and materials*. John Wiley & Sons. https://books.google.com/books?hl=en&lr=&id=wipJ0qBvv40C&oi=fnd&pg=PA1&dq=Pavement+Design+and+Materials&ots=ApwWeK6ZIF&sig=-qfE5m4M5r70iPj136UouoJGOuE

Pasupunuri, S. K., Thom, N., & Li, L. (2024). Roughness prediction of jointed plain concrete pavement using physics informed neural networks. *Transportation Research Record*. Advance online publication. https://doi.org/10.1177/03611981241245991

Peraka, N. S. P., & Biligiri, K. P. (2020). Pavement asset management systems and technologies: A review. *Automation in Construction*, 119, 103336. https://doi.org/10.1016/j.autcon.2020.103336

Pereira, D. R., Piteri, M. A., Souza, A. N., Papa, J. P., & Adeli, H. (2020). FEMa: A finite element machine for fast learning. *Neural Computing and Applications*, 32(10), 6393–6404. https://doi.org/10.1007/s00521-019-04146-4

Pereira, P., & Pais, J. (2017). Main flexible pavement and mix design methods in Europe and challenges for the development of an European method. *Journal of Traffic and Transportation Engineering (English Edition)*, 4(4), 316–346. https://doi.org/10.1016/j.jtte.2017.06.001

Polaczyk, P., Ma, Y., Xiao, R., Hu, W., Jiang, X., & Huang, B. (2021). Characterization of aggregate interlocking in hot mix asphalt by mechanistic performance tests. *Road Materials and Pavement Design*, 22(Suppl 1), S498–S513. https://doi.org/10.1080/14680629.2021.1908408

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 31, Montreal, Canada.

Rafiei, M. H., & Adeli, H. (2017). A new neural dynamic classification algorithm. *IEEE Transactions on Neural Networks and Learning Systems*, 28(12), 3074–3083.

Rafiei, M. H., Gauthier, L. V., Adeli, H., & Takabi, D. (2022). Self-supervised learning for electroencephalography. *IEEE Transactions on Neural Networks and Learning Systems*, 35(2), 1457–1471. https://ieeexplore.ieee.org/abstract/document/9837871/

Random Forest Regressor. (2024). *scikit-learn*. https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html

Ridley, M. (2022). Explainable artificial intelligence (XAI): Adoption and advocacy. *Information Technology and Libraries*, 41(2), Article 2. https://doi.org/10.6017/ital.v41i2.14683

Rosé, C. P., McLaughlin, E. A., Liu, R., & Koedinger, K. R. (2019). Explanatory learner models: Why machine learning (alone) is not the answer. *British Journal of Educational Technology*, 50(6), 2943–2958. https://doi.org/10.1111/bjet.12858

Saha, P., Ksaibati, K., & Atadero, R. (2017). Developing pavement distress deterioration models for pavement management system using Markovian probabilistic process. *Advances in Civil Engineering*, 2017. https://doi.org/10.1155/2017/8292056

Saleeb, A., Liang, R. Y., Qablan, H. A., & Powers, D. (2005). Numerical simulation techniques for HMA rutting under loaded wheel tester. *International Journal of Pavement Engineering*, 6(1), 57–66. https://doi.org/10.1080/10298430500068704

Sarkhani Benemaran, R., Esmaeili-Falak, M., & Javadi, A. (2023). Predicting resilient modulus of flexible pavement foundation using extreme gradient boosting based optimised models. *International Journal of Pavement Engineering*, 24(2), 2095385. https://doi.org/10.1080/10298436.2022.2095385

Seleridis, G. (2017). *Evaluation of the current test methods of water sensitivity and permanent deformation*. https://repository.tudelft.nl/islandora/object/uuid:dfb8113a-43c9-4dec-9f18-8da85eedf922

Shukla, K., Xu, M., Trask, N., & Karniadakis, G. E. (2022). Scalable algorithms for physics-informed neural and graph networks. *Data-Centric Engineering*, 3, e24. https://doi.org/10.1017/dce.2022.24

Sklearn.ensemble. (2024). *scikit-learn*. https://scikit-learn/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html

Song, Y., & Lu, Y. (2015). Decision tree methods: Applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2), 130. https://doi.org/10.11919/j.issn.1002-0829.215044

Song, Y., Wang, Y. D., Hu, X., & Liu, J. (2022). An efficient and explainable ensemble learning model for asphalt pavement condition prediction based on LTPP dataset. *IEEE Transactions on Intelligent Transportation Systems*, 23(11), 22084–22093. https://doi.org/10.1109/TITS.2022.3164596

Sulejmani, P., Said, S., Agardh, S., & Ahmed, A. (2019). Moisture sensitivity of asphalt mixtures using cycling pore pressure conditioning. *Transportation Research Record: Journal of the Transportation Research Board*, 2673(2), 294–303. https://doi.org/10.1177/0361198118823496

Surendrakumar, K., Prashant, N., & Mayuresh, P. (2013). Application of Markovian probabilistic process to develop a decision support system for pavement maintenance management. *International Journal of Scientific & Technology Research*, 2(8), 295–303.

SVR-sklearn. (2024). *scikit-learn*. https://scikit-learn/stable/modules/generated/sklearn.svm.SVR.html

Tamagusko, T., & Ferreira, A. (2023). Machine learning for prediction of the international roughness index on flexible pavements: A review, challenges, and future directions. *Infrastructures*, *8*(12), Article 12. https://doi.org/10.3390/infrastructures8120170

Tarefder, R. A., & Ahmad, M. (2015). Evaluating the relationship between permeability and moisture damage of asphalt concrete pavements. *Journal of Materials in Civil Engineering*, *27*(5), 04014172. https://doi.org/10.1061/(ASCE)MT.1943-5533.0001129

Tarefder, R. A., & Ahmad, M. (2017). Evaluation of pore structure and its influence on permeability and moisture damage in asphalt concrete. *International Journal of Pavement Engineering*, *18*(3), 274–283. https://doi.org/10.1080/10298436.2015.1065995

Tia, M. (2005). Fundamentals and practice of asphalt mixture design procedures to assure adequate performance. *13th Conference on Pavement Engineering*, Hsin Chu, Taiwan.

Tian, Y., & Zhang, Y. (2022). A comprehensive survey on regularization strategies in machine learning. *Information Fusion*, *80*, 146–166.

Torres-Machí, C., Chamorro, A., Pellicer, E., Yepes, V., & Videla, C. (2015). Sustainable pavement management: Integrating economic, technical, and environmental aspects in decision making. *Transportation Research Record: Journal of the Transportation Research Board*, *2523*(1), 56–63. https://doi.org/10.3141/2523-07

Tran, H., Robert, D., Gunarathna, P., & Setunge, S. (2023). Multi-time step deterioration prediction of freeways using linear regression and machine learning approaches: A case study. *International Journal of Pavement Research and Technology*. Advance online publication. https://doi.org/10.1007/s42947-023-00376-x

Uddin, W., Hudson, W. R., & Haas, R. (2013). P*ublic infrastructure asset management*. https://scholar.google.com/scholar_lookup?hl=en&publication_year=2013&author=W.+Uddin&author=W.R.+Hudson&author=R.+Haas&title=Public+infrastructure+asset+management

Wang, C., Xiao, W., & Liu, J. (2023). Developing an improved extreme gradient boosting model for predicting the international roughness index of rigid pavement. *Construction and Building Materials*, *408*, 133523. https://doi.org/10.1016/j.conbuildmat.2023.133523

Wang, G., Fang, Q., Wang, J., Li, Q. M., Chen, J. Y., & Liu, Y. (2024). Estimation of load for tunnel lining in elastic soil using physics-informed neural network. *Computer-Aided Civil and Infrastructure Engineering*. Advance online publication. https://doi.org/10.1111/mice.13208

Wang, W., Qin, Y., Li, X., Wang, D., & Chen, H. (2017). Comparisons of faulting-based pavement performance prediction models. *Advances in Materials Science and Engineering*, *2017*, e6845215. https://doi.org/10.1155/2017/6845215

Yang, J., Lu, J. J., Gunaratne, M., & Xiang, Q. (2003). Forecasting overall pavement condition with neural networks: Application on Florida Highway Network. *Transportation Research Record*, *1853*(1), 3–12. https://doi.org/10.3141/1853-01

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, *415*, 295–316.

Yang, X., Guan, J., Ding, L., You, Z., Lee, V. C. S., Mohd Hasan, M. R., & Cheng, X. (2021). Research and applications of artificial neural network in pavement engineering: A state-of-the-art review. *Journal of Traffic and Transportation Engineering (English Edition)*, *8*(6), 1000–1021. https://doi.org/10.1016/j.jtte.2021.03.005

Yao, L., Dong, Q., Jiang, J., & Ni, F. (2019). Establishment of prediction models of asphalt pavement performance based on a novel data calibration method and neural network. *Transportation Research Record*, *2673*(1), 66–82. https://doi.org/10.1177/0361198118822501

Yao, L., Dong, Q., Jiang, J., & Ni, F. (2020). Deep reinforcement learning for long-term pavement maintenance planning. *Computer-Aided Civil and Infrastructure Engineering*, *35*(11), 1230–1245. https://doi.org/10.1111/mice.12558

Yao, L., Leng, Z., Jiang, J., & Ni, F. (2022). Modelling of pavement performance evolution considering uncertainty and interpretability: A machine learning based framework. *International Journal of Pavement Engineering*, *23*(14), 5211–5226. https://doi.org/10.1080/10298436.2021.2001814

Yao, L., Leng, Z., Jiang, J., & Ni, F. (2024). A multi-agent reinforcement learning model for maintenance optimization of interdependent highway pavement networks. *Computer-Aided Civil and Infrastructure Engineering*. Advance online publication. https://doi.org/10.1111/mice.13234

Yehia, A., & Swei, O. (2020). Probabilistic infrastructure performance models: An iterative-methods approach. *Transportation Research Part C: Emerging Technologies*, *111*, 245–254.

Zaki, M. J., & Meira, W. (2014). *Data mining and analysis: Fundamental concepts and algorithms*. Cambridge University Press.

Zaumanis, M., Poulikakos, L. D., & Partl, M. N. (2018). Performance-based design of asphalt mixtures and review of key parameters. *Materials & Design*, *141*, 185–201. https://doi.org/10.1016/j.matdes.2017.12.035

Zhang, M., Gong, H., Jia, X., Xiao, R., Jiang, X., Ma, Y., & Huang, B. (2020). Analysis of critical factors to asphalt overlay performance using gradient boosted models. *Construction and Building Materials*, *262*, 120083. https://doi.org/10.1016/j.conbuildmat.2020.120083

Zhang, Y., Gu, F., Birgisson, B., & Lytton, R. L. (2017). Viscoelasticplastic–fracture modeling of asphalt mixtures under monotonic and repeated loads. *Transportation Research Record*, *2631*(1), 20–29. https://doi.org/10.3141/2631-03

Zhao, J., Wang, H., & Lu, P. (2022). Impact analysis of traffic loading on pavement performance using support vector regression model. *International Journal of Pavement Engineering*, *23*(11), 3716–3728. https://doi.org/10.1080/10298436.2021.1915493

Zhou, F., Steger, R., & Mogawer, W. (2021). Development of a coherent framework for balanced mix design and production quality control and quality acceptance. *Construction and Building Materials*, *287*, 123020. https://doi.org/10.1016/j.conbuildmat.2021.123020