A step toward a micromechanics-informed neural network for predicting asphalt mixture stiffness

Anupam, Kumar; Berangi, Mohammadjavad; Camargo, Juan Camilo; Kasbergen, Cor; Erkens, Sandra

DOI
10.1111/mice.70000

Publication date
2025

Document Version
Final published version

Published in
Computer-Aided Civil and Infrastructure Engineering

Accepted: 4 June 2025

DOI: 10.1111/mice.70000

RESEARCH ARTICLE

COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING WILEY

A step toward a micromechanics-informed neural network for predicting asphalt mixture stiffness

Kumar Anupam | **Mohammadjavad Berangi** | **Juan Camilo Camargo** | **Cor Kasbergen** | **Sandra Erkens**

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

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

Funding information
Ministry of Infrastructure and Water Management, Grant/Award Number: 31164321

Abstract

Asphalt mixtures show complex mechanical behavior due to their heterogeneous structure. Traditionally, the mechanical characterization of asphalt mixture is done through laboratory testing or micromechanical modeling. While laboratory tests and micromechanical models provide reliable measurements and physical interpretability, they are often resource-intensive and demand extensive calibration. Recent advances in machine learning address some of the above issues by enabling accurate predictions, though often lacking physical interpretability and stability. Hence, this study aims to present a novel micromechanics-infused neural network (MINN) framework for predicting asphalt mixture stiffness. The framework embeds micromechanical principles derived from the modified Hirsch model into the neural network's loss function, allowing the model to learn from experimental data while adhering to micromechanics-based constraints. In this study, feature selection is performed using BorutaShap, and Bayesian optimization is applied for hyperparameter tuning. Results show that MINN improves prediction accuracy, interpretability, and robustness.

1 | INTRODUCTION

In the pavement community, asphalt mixtures are generally treated as viscoelastic, which means that their mechanical characteristics are affected by loading rate and temperature (Cao et al., 2017). Therefore, to develop an appropriate mixture design, it is often necessary to estimate its mechanical properties, such as stiffness, accurately. While laboratory tests are typically carried out to determine these properties, they are often time- and resource-intensive (Elseifi et al., 2006; W. Huang et al., 2019). As an attractive alternative, several studies over the past few decades have explored theoretical techniques to predict mechanical properties (Rafiei et al., 2017; Tang et al., 2019; H. Zhang et al., 2018a). One promising approach, as proposed by researchers (Mura, 2013; J. Zhang, 2020; H. Zhang et al., 2020), is the utilization of micromechanical models. Researchers (H. Wang et al., 2021; H. Zhang et al., 2018a) have highlighted that these micromechanical models can effectively replicate laboratory results without requiring extensive sample preparation, thereby reducing resource consumption.

Among diverse categories of micromechanical models, numerical micromechanical models (Zohdi & Wriggers, 2008), which are developed using the finite element method (FEM; Belytschko et al., 2014) and/or the discrete element method (DEM; Cundall & Strack, 1979), have gained popularity among researchers (Dhundup et al.,

2025; B. Li et al., 2025). However, accurately representing the complex microstructures of the asphalt mixtures by these methods demands the generation of a large number of elements, which necessitates significant computational resources and extended processing times (Jin et al., 2021; Yu & Adeli, 1993). Moreover, despite the ability of FEM and DEM to realistically model asphalt mixtures, the accuracy of the predictions heavily depends on the careful calibration of specific input parameters (Kaewnuratchadasorn et al., 2024; Zeng et al., 2020). In contrast, analytical micromechanical models are anticipated to offer accurate predictions of a composite's mechanical properties without requiring significant computational resources (H. Wang et al., 2021).

Analytical micromechanical models simplify the representation of asphalt mixtures' microstructure (e.g., aggregates are often modeled as uniform spherical or ellipsoidal particles; Zeng et al., 2020). The simplification often enables the derivation of analytical solutions for the mechanical properties of the mixtures (Yin et al., 2008). As a result, these models have a reliable capability to precisely predict the mechanical properties of asphalt mixtures (Underwood & Kim, 2013; H. Zhang, 2022). However, despite their accuracy, analytical models often fall short in providing a realistic depiction of the actual microstructure (Luo & Lytton, 2011; H. Zhang et al., 2020).

The trade-off between simplicity and realism, as discussed above, has led the research on analytical micromechanical modeling to be often concentrated on continuum-based micromechanical models (CBMM; Zeng et al., 2020). These models treat the heterogeneous asphalt mixture as an equivalent homogeneous material with effective properties (i.e., averaged mechanical properties that reflect the combined behavior of all constituent materials; Wiśniewska et al., 2019; C. H. Yang et al., 2005). In the CBMMs, homogenization techniques (Charalambakis, 2010) are often employed to predict the macroscopic properties of materials by analyzing their microstructural features. The technique "averages out" the microscale variations to predict the behavior of the material on a larger scale (Torquato, 2002).

In homogenization techniques, the micromechanical properties of a mixture are derived from the properties of its phases (i.e., the distinct material components such as aggregates, binders, and voids) (Charalambakis, 2010). The technique allows researchers to calculate the stress and strain values inside each phase based on the applied loading conditions (Bornert et al., 2001). Simple homogenization models such as Hirsch (1962) and Christensen models (D. W. Christensen et al., 2003) have been successfully used by previous researchers (Dongre et al., 2005; Pellinen et al., 2007) for predicting mechanical properties of different asphalt mixtures.

The Hirsch model (Hirsch, 1962) was initially developed to study the modulus of concrete by considering the moduli of aggregates and cement. In this model, the different phases of the composite are assumed to be arranged in a combination of parallel and series configurations (R. M. Christensen, 2005). The modulus of the composite is determined by the moduli and volume fractions of its constituent phases (R. M. Christensen, 2005). Later, D. W. Christensen et al. (2003) modified the original Hirsch model to be used for asphalt concrete.

Although simplified homogenization models offer an intuitive representation of the stiffness via measurable engineering properties of the mixture (Gong et al., 2022), many researchers (Behnood & Mohammadi Golafshani, 2021; Ceylan et al., 2009; D. W. Christensen & Bonaquist, 2015) raised concerns regarding their limited ability to capture the complex interactions in asphalt mixtures. To address this, several studies (Barugahare et al., 2022; Moussa & Owais, 2021; Shu & Huang, 2008) have explored the use of various machine learning (ML; Adeli & Hung, 1994) methods such as neural networks (NNs; Adeli & Yeh, 1989) as alternatives for predicting stiffness. ML methods offer significant potential for tackling scientific problems where the underlying processes are not fully understood (K. Li et al., 2024; Rafiei & Adeli, 2017). They are also efficient when running mechanistic models at the desired spatial and temporal resolutions is computationally impractical (Willard et al., 2022a). However, the implementation of even advanced black-box ML models often faces restricted efficacy in scientific fields (Kaewnuratchadasorn et al., 2024; Willard et al., 2022a). The limitation is due to several challenges, such as large data requirements (Karpatne et al., 2017), producing physically consistent results (Willard et al., 2020), poor generalizability (Caldwell et al., 2014), scalability (Alam et al., 2019), and interpretability (Doshi-Velez & Kim, 2017).

In order to alleviate the limitation of ML models, several potential approaches have been recently proposed for integrating physics within ML (Karpatne et al., 2017). These approaches are residual modeling, physics-guided loss function, physics-informed initialization, and physics-based design of architecture that can be categorized as ″physics-informed machine learning (PIML; Karpatne et al., 2017). More information about PIML and its application in science and engineering can be found elsewhere (Willard et al., 2022a).

In the last few years, extensive research has been carried out in various disciplines using PIML, such as computational physics (Tanaka et al., 2021), material discovery (Cang et al., 2018; Schleder et al., 2019), geotechnical engineering (G. Wang et al., 2024), climate science (Faghmous

& Kumar, 2014; O'Gorman & Dwyer, 2018), turbulence modeling (Mohan & Gaitonde, 2018; Xiao et al., 2019), and hydrology (Xu & Valocchi, 2015). In the pavement engineering domain, Deng et al. (2024) used a physics-guided loss function approach, an autoencoder for feature selection, and a feedforward NN for rut depth prediction. The results showed that implementing physics information improves the model's stability and rationality while maintaining high accuracy. However, the authors concluded the need for further research to effectively model variables that show varying levels of influence on rutting performance. In another study (Kargah-Ostadi et al., 2023), researchers used the physics-informed initialization approach, in which they pre-trained a model using data derived from the simulation and then optimized specific parameters of the model based on the international roughness index measurements. The authors highlighted that their proposed model had high accuracy and generalizability for pavement sections only with a low level of roughness.

In a similar study, K. Chen et al. (2025) proposed a framework for predicting the international roughness index by integrating domain knowledge into NN in two complementary approaches. The first approach was augmenting the input space of an NN with outputs from FEM, while the second approach modified the loss function of the NN to penalize deviations from physically consistent behavior. The second approach showed improvements in long-term prediction accuracy and model stability. However, the first approach, while adding computational complexity, resulted in only marginal gains, particularly in short-term predictions. The author concluded that the selection of the integration approach is critical in achieving tangible benefits from PIML.

Xiong et al. (2024) proposed a novel approach for integrating micromechanical knowledge with ML, in which the model is trained without relying on any labeled datasets. The proposed framework was applied to various non-linear micromechanical problems like anisotropic microplasticity. While the proposed framework by the authors achieved good results, the focus of the study was still on solving micromechanical problems outside of the pavement engineering domain.

Given the wide range of approaches used in the reviewed literature to model asphalt mixture properties, Table 1 summarizes their typical methods, along with their strengths and limitations, to provide a clearer overview. To the best of the author's knowledge, no existing studies in the pavement field have proposed an approach to utilize micromechanical models as the loss function within ML to predict pavement mechanical properties.

1.1 | Research objective and scope

The goal of the research is to develop a framework to integrate micromechanical knowledge with ML to predict the mechanical properties of the asphalt mixture. The physics-guided loss function approach is used to develop a framework that achieves an appropriate equilibrium between accuracy, stability, and interpretability. The framework is expected to provide better insights into asphalt mixture characteristics by infusing the mechanical properties of asphalt mixtures into modeling processes.

The scope of the study is defined as follows:

1. The stiffness of asphalt mixtures is considered as the target (output) of the predictive model.
2. The NN is considered as the ML model.
3. The modified Hirsch model is implemented to leverage micromechanical principles into ML-based predictive modeling.
4. The stability and interpretability of the model are evaluated through comparisons with the micromechanical model.

1.2 | The novelty of the research

The novelty of the study is as follows:

1. The scope of the problem: As discussed in the previous sub-sections, none of the existing approaches provide a framework to integrate micromechanical knowledge with NN.
2. Micromechanically consistent model: Ensuring the model is rooted in micromechanics principles for reliable and scientifically sound predictions.
3. Enhanced transparency and interpretability: Emphasizing clear model explanations to improve trust and understanding of the results.
4. Application: To the best of the author's knowledge, none of the previous research has proposed a framework to integrate micromechanical knowledge with ML for predicting visco-elastic material properties.

2 | RESEARCH FRAMEWORK

In order to present an easier understanding, the research framework is presented in Figure 1. As the corresponding figure shows, the framework is categorized into four

TABLE 1 Overview of existing approaches for predicting the asphalt mixture properties.

Modeling family	Common method(s)	Key advantages	Main limitations	Representative refs.
Laboratory testing	Dynamic modulus tests, four-point bending, and so forth	Direct measurement; high fidelity	Time- and cost-intensive; limited temperature/frequency range	(Cao et al., 2017; Elseifi et al., 2006)
Analytical micromechanics	Hirsch, Christensen, and so forth	Closed-form; fast; interpretable	Simplifying phase geometry	(Charalambakis, 2010; R. M. Christensen, 2005; D. W. Christensen & Bonaquist, 2015; D. W. Christensen et al., 2003; Dongre et al., 2005; Hirsch, 1962; Luo & Lytton, 2011; Shu & Huang, 2008; Torquato, 2002; Underwood & Kim, 2013; Wiśniewska et al., 2019; C. H. Yang et al., 2005; Yin et al., 2008; Zeng et al., 2020; H. Zhang et al., 2018a)
Numerical micromechanics	FEM/DEM RVEs, and so forth	Captures realistic geometry; flexible	High computational cost; parameter calibration	(Belytschko et al., 2014; Cundall & Strack, 1979; Jin et al., 2021; Yu & Adeli, 1993; Zeng et al., 2020; Zohdi & Wriggers, 2008)
ML	NN	Learns complex patterns; no constitutive assumptions	Requires large, well-labeled data; poor physical interpretability	(Adeli & Yeh, 1989; Barugahare et al., 2022; Behnood & Mohammadi Golafshani, 2021; Moussa & Owais, 2021)
PIML	Residual modeling, physics-guided loss function, physics-informed initialization, and physics-based design of architecture	Combine data with physics; better generalisation and stability	Extra loss term may raise training cost; needs an analytical or partial differential equation model	(K. Chen et al., 2025; Deng et al., 2024; Kaewnuratchadasorn et al., 2024; Kargah-Ostadi et al., 2023; Willard et al., 2020; Xiong et al., 2024)

Abbreviations: DEM, discrete element method; FEM, finite element method; PIML, physics-informed machine learning.

stages: (1) setting up the micromechanical model, (2) data preparation, (3) development of the micromechanics-infused framework, and (4) model evaluation. Stage 1, as it will be discussed in Section 2.1, provides a process of selecting an appropriate micromechanical model. In Stage 2 (see Section 2.2), data are prepared, and key features are selected as input parameters for the model. Using the data from the previous stage, a micromechanics-infused NN model (MINN) is set up in Stage 3 (see Section 2.3). In Stage 4, defined metrics are implemented to evaluate the MINN.

2.1 | Stage 1: Setting up the micromechanical model

Based on the research needs described in the previous section, the first stage of the research framework is selecting an appropriate micromechanical model. A key aspect of the selection process is evaluating whether the constitutive relationship can be incorporated as a constraint condition. A brief discussion of the aspects is provided below.

2.1.1 | Stiffness tensor

In the context of linear elasticity, the fundamental constitutive equation describing a material's behavior is typically expressed as Equation (1) (Mase et al., 2009; H. Zhang et al., 2020).

$$\sigma = \mathbf{C} : \varepsilon \tag{1}$$

where "σ" represents the second-order stress tensor, "ε" denotes the second-order strain tensor, and "C" corresponds to the fourth-order stiffness tensor (H. Zhang et al., 2020). The sign ":" denotes the double dot product operation between two tensors. For materials that are isotropic, the stiffness tensor can be calculated using the bulk modulus "K" and shear modulus "G" as shown in Equation (2) (Bonet & Wood, 1997)

$$C = 3KI^{v} + 2GI^{d} \tag{2}$$

where "I^{v}" represent volumetric tensor, and "I^{d}" stands for deviatoric tensors (H. Zhang et al., 2020). These

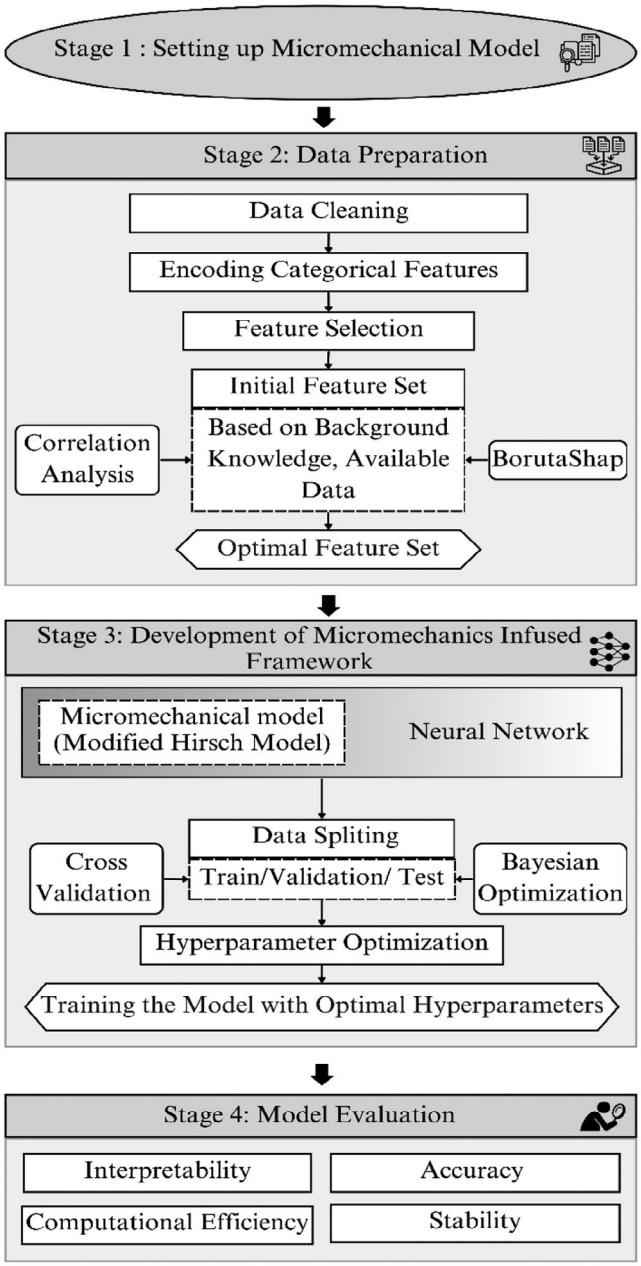

FIGURE 1 Research framework.

tensors ensure that "C" correctly captures material behavior in volumetric and shear deformation. Furthermore, for any fourth-order tensors, "B_1" and "B_2" are given as Equations (3) and (4), and their double dot product can be directly given by Equation (5).

$$B_1 = B_1^v I^v + B_1^d I^d \tag{3}$$

$$B_2 = B_2^v I^v + B_2^d I^d \tag{4}$$

$$B_1 : B_2 = B_1^v B_2^v I^v + B_1^d B_2^d I^d \tag{5}$$

The presented equations are essential for analyzing stiffness behavior in complex materials, such as composites, where multiple phases with different stiffness properties interact.

2.1.2 | Effective stiffness of a composite

In composite materials, the effective stiffness tensor"C_{eff}" is calculated by the average stress and strain of the composite as Equation (6) (Hill, 1963)

$$\langle \sigma \rangle_c = C_{\text{eff}} \langle \varepsilon \rangle_c \tag{6}$$

where "$\langle \sigma \rangle_c$" and "$\langle \varepsilon \rangle_c$" represent the volume-averaged stress and strain, respectively. For a representative volume element (RVE;Charalambakis, 2010), these averages are given by Equation (7).

$$\langle \sigma \rangle_c = \frac{1}{V} \int_V \sigma dV, \quad \langle \varepsilon \rangle_c = \frac{1}{V} \int_V \varepsilon dV \tag{7}$$

At the phase level, the same relations hold for phase "r" over the volume of this phase "V_r" as shown in Equation (8).

$$\langle \sigma \rangle_r = \frac{1}{V_r} \int_{V_r} \sigma dV, \quad \langle \varepsilon \rangle_r = \frac{1}{V_r} \int_{V_r} \varepsilon dV \tag{8}$$

By inserting Equation (8) into Equation (7), the expressions for "$\langle \sigma \rangle_c$" and "$\langle \varepsilon \rangle_c$" can be reformulated in Equations (9) and (10).

$$\langle \sigma \rangle_c = \sum_{r=1}^{N} \phi_r \langle \sigma \rangle_r \tag{9}$$

$$\langle \varepsilon \rangle_c = \sum_{r=1}^{N} \phi_r \langle \varepsilon \rangle_r \tag{10}$$

where "$\phi_r \varepsilon$ represents the volume fraction of phase "r" within the RVE, given by Equation (11) (H. Zhang, 2022).

$$\phi_r = \frac{V_r}{V} \tag{11}$$

Since each phase obeys the constitutive law (Bonet & Wood, 1997):

$$\langle \sigma \rangle_r = C_r \langle \varepsilon \rangle_r \tag{12}$$

where "C_r" is the stiffness tensor of phase "r" (H. Zhang, 2022). By substituting Equations (6) and (11) into Equation (9), "C_{eff}" can be expressed in terms of "C_r" as shown in Equation (13).

$$C_{\text{eff}} : \langle \varepsilon \rangle_c = \sum_{r=1}^{N} \phi_r C_r \langle \varepsilon \rangle_r \tag{13}$$

The strain localization tensor "A_r" of phase "r" can then be introduced in Equation (14) to express the relation between the composite strain and strain.

$$\langle \varepsilon \rangle_r = A_r : \langle \varepsilon \rangle_c \tag{14}$$

The value of "C_{eff}" can be further explained by combining Equations (14) and (13)

$$C_{\text{eff}} = \sum_{r=1}^{N} \phi_r C_r : A_r \tag{15}$$

By replacing Equation (14) into Equation (10), Equation (16) can be obtained as follows (H. Zhang, 2022):

$$\sum_{r=1}^{N} \phi_r A_r = I \tag{16}$$

The determination or estimation of "ϕ_r, ε "C_r," and "A_r" values are essential in order to compute "C_{eff}" (H. Zhang, 2022). Typically, the components of a composite are known, allowing "$\phi_r \varepsilon$ and "C_r" to be experimentally obtained by laboratory tests. Although "A_r" can be measured using techniques such as digital image processing (Gonzales & Wintz, 1987) and smart sensors, these methods are not widely proven by pavement engineers and researchers (H. Zhang et al., 2020, 2021). Therefore, the primary objective of CBMM is to determine the value of such a parameter (H. Zhang et al., 2020).

Researchers (Hashin, 1983; Nemat-Nasser & Hori, 2013) proposed several different approaches to calculate "C_{eff}," which can be broadly categorized as geometry-based (R. M. Christensen & Lo, 1979) and bound-based (Hashin & Shtrikman, 1963; Paul, 1960). Since the geometry-based approach is used in the modeling stage in this study, a brief description of this approach is presented in the following subsection. It is noted that further information about the bound-based approach can be found elsewhere (Hashin, 1965; Milton, 1981, 1982; Walpole, 1966).

2.1.3 | Geometry-based approach

The geometry-based approach utilizes a predefined geometric model (R. M. Christensen & Lo, 1979) that relatively establishes the spatial arrangement of the individual phases in composites. By using such an arrangement, a closed-form solution of "C_{eff}" can be obtained. The geometry-based approach can be further classified into models where the individual phases are arranged in parallel, series, or a combination of both. A notable example of such models is the Hirsch model (Hirsch, 1962), which is briefly described in the following subsection.

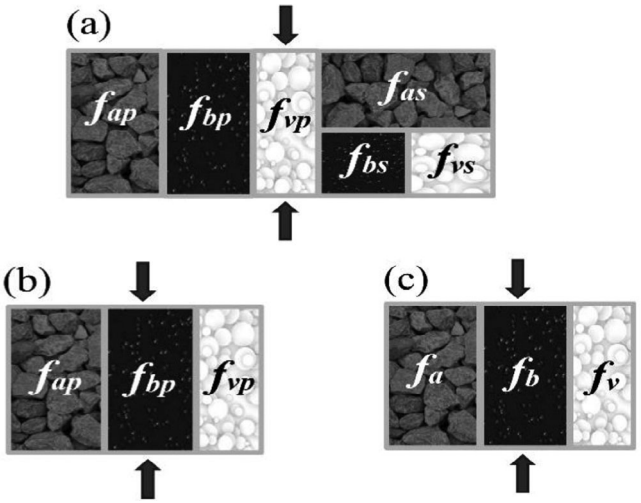

FIGURE 2 Evolution of the Christensen model configuration for asphalt mixtures: (a) The original configuration showing sub-phases arranged in parallel (f_{ap}, f_{bp}, f_{vp}) and series ($f_{as}, f_{bs} f_{vs}$); (b) a simplified version using only parallel sub-phases (f_{ap}, f_{bp}, f_v); (c) the proposed modification by Zhang et al. (2018b), where each phase (aggregates f_a, binder f_b, air voids f_v) is directly modeled without sub-phase separation.

Hirsch model

In order to modify the original Hirsch model for applicability to asphalt mixes, D. W. Christensen et al. (2003) proposed several phase arrangements (H. Zhang, 2022). After extensively evaluating the arrangements, the researchers concluded that the most accurate predictions of the asphalt modulus were obtained with the version where sub-phases were arranged in parallel and series (H. Zhang et al., 2018b). These sub-phases represent asphalt binders, aggregates, and air voids (see Figure 2a).

As given in Equation (17), the dynamic Young's modulus of an asphalt mixture, denoted as "$|E^*|_{mix}$," is calculated based on the volume fractions and moduli of the asphalt binder, aggregates, and air voids (H. Zhang et al., 2018b).

$$|E^*|_{mix}(f) = f_{ap} E_a + 3 f_{bp} |G^*|_b + (f_{as} + f_{bs} + f_{vs}) \left[\frac{f_{as}}{E_a} + \frac{(f_{bs} + f_{vs})^2}{3 f_{bs} |G^*|_b (f)} \right]^{-1} \tag{17}$$

where

1. The terms "p" and "s" refer to the parallel and series components, respectively.
2. "a," "b," and "c" represent aggregate, binder, and air void phases, respectively,
3. "f_a," "f_b," and "f_v" are the aggregate phase volume fraction, asphalt binder phase volume fraction, and air void phase volume fraction, respectively,

4. "E_a" and "$|G^*|_b$" are Young's modulus of the aggregate phase and the dynamic shear modulus of the asphalt binder phase, respectively.

D. W. Christensen et al. (2003) observed that the series sub-phases had a much less significant impact on an asphalt mix's estimated modulus than the parallel sub-phases. The observation indicates that the behavior of an asphalt mixture is well represented by a parallel configuration of its constituent phases. Therefore, D. W. Christensen and Bonaquist (2015) further simplified the original model to a straightforward parallel configuration as shown in Figure 2b and described by Equation (18).

$$|E^*|_{mix}\ (f) = f_{ap}E_a + 3\,f_{bp}|G^*|_b\,(f) \tag{18}$$

Since obtaining "f_{ap}" and "f_{bp}" through laboratory and field tests is challenging (Anderson et al., 2011; Underwood et al., 2012), researchers D. W. Christensen et al. (2003) have proposed a further simplification of the equation by introducing the contact factor (P_c; see Equation 19).

$$|E^*|_{mix}\ (f) = P_c\,(f)\left(f_{ap}E_a + 3\,f_{bp}|G^*|_b\,(f)\right) \tag{19}$$

In the equation, "P_c" represents the proportion of the parallel component within the total volume of the composite, which consequently ranges between 0 and 1 (H. Zhang, 2022).

Equation (19) is derived from Equation (18), assuming the parallel component is equally distributed across all phases. This assumption implies that the value of "P_c" remains constant across all composite phases without varying between different phases. According to past studies (D. W. Christensen et al., 2003), the values of "P_c" can be determined through laboratory tests, and it was reported that the predictions of the model are found to be aligned well with the laboratory test results (D. W. Christensen & Bonaquist, 2015; D. W. Christensen et al., 2003). However, H. Zhang et al. (2018a) pointed out that the physical interpretation of "P_c" is challenging for several reasons.

The "P_c" factor specifies the aggregate contact factor, referring to the contribution of aggregate particles that are in close contact with one another (H. Zhang et al., 2018b). Furthermore, it was interpreted by D. W. Christensen et al. (2003) that higher "P_c" values at elevated frequencies or lower temperatures suggest increased interaction between aggregate particles (H. Zhang, 2022). However, this interpretation is inconsistent with the physical behavior of aggregates, as fewer aggregate particles are expected in close contact under high-frequency or low-temperature conditions (H. Zhang et al., 2018b). This behavior is likely due to the stiffening of the asphalt binder under these conditions. Conversely, at low frequencies or high temperatures, when the asphalt binder becomes softer, aggregate particles are expected to move more freely, leading to more pronounced contact among the particles.

Considering the inconsistency mentioned above, H. Zhang (2022) proposed a revised parallel arrangement considering the total volume of bitumen, aggregates, and air voids (see Figure 2c and Equation 20). Additionally, the author introduced the term "aggregate organization factor" (P_a) to represent the temperature- and frequency-dependent influence of the aggregate phase on the prediction of mixture stiffness (see Equation 21; H. Zhang 2022). To identify P_a, a regression model as shown in Equation (22) was presented (H. Zhang, 2022).

$$|E^*|_{mix}\ (f) = f_aE_a + 3\,f_b|G^*|_b\,(f) \tag{20}$$

$$|E^*|_{mix}\ (f) = P_a\,(f)\,f_a\,E_a + 3\,f_b|G^*|_b\,(f) \tag{21}$$

where "f_a," "f_b," "E_a," and "$|G^*|_b$ " are described in Equation (17) (H. Zhang et al., 2018b).

$$P_a = a + (1-a)\frac{\exp\left(b + \frac{cln(f_b)}{f_b+f_v|G^*|_b+d(f_b+f_v)}\right)}{1+\exp\left(b+\frac{cln(f_b)}{f_b+f_v|G^*|_b}+d\,(f_b+f_v)\right)} \tag{22}$$

where "a," "b," "c," and "d" are the regression coefficients.

The Hirsch model has been widely adopted in pavement engineering (Pellinen et al., 2007; Shen et al., 2013; Yan et al., 2024) due to its reliable performance across diverse asphalt mixtures and loading conditions. Its formulation links meso-scale material properties, such as binder and aggregate characteristics to macro-scale stiffness. Owing to its simplicity and computational efficiency, the Hirsch model is well-suited for integration into data-driven modeling frameworks, where interpretability and low computational overhead are essential. As discussed in this section, the Hirsch model was further modified by researchers to better capture the micromechanical behavior of asphalt mixtures. Therefore, in this research, the modified Hirsch model, presented in Equations (20) and (21), serves as the foundation for developing the MINN. It is noted that more complicated models, such as the Mori–Tanaka model (Mori & Tanaka, 1973) and self-consistent model (Hill, 1965), can be implemented, but it is out of the scope of the research.

2.2 | Stage 2: Data preparation

In order to develop the MINN framework, it is essential to collect a relevant and representative dataset. The dataset used for model development and evaluation in this study was derived from the authors' previous work

TABLE 2a Overview of the numerical features in the dataset.

Feature name	Unit	Description
Target density (TD)	kg/m^3	The bulk density of the asphalt mixture based on the mix design
Bitumen penetration (BP)	0.1 mm	Depth of penetration under standard loading using penetration test (NEN-EN 1426, 2015)
Bitumen phase angle (δ)	Degree (°)	Phase lag between stress and strain (NEN-EN 14770, 2022)
Target sand volume percentage (TSA)	%	Mass percentage of sand in the mix design
Target filler volume percentage (TF)	%	Mass percentage of filler in the mixture
Target stone mass percentage (TST)	%	Mass percentage of coarse aggregate
Target bitumen volume percentage (fb)	%	Design target bitumen volume fraction
Reclaimed asphalt pavement (RAP)	%	Percentage of reclaimed material used
Age	Year	Time since the sample preparation
Bitumen softening point (SP)	°C	The temperature at which bitumen softens using the ring and ball test (NEN-EN 1427, 2015)
Complex modulus (G^*)	Pa	Representing the total resistance of the bitumen to deformation under sinusoidal loading, measured using a dynamic shear rheometer (NEN-EN 14770, 2022)
Density (D)		Measured density of the compacted sample
Air void (fv)	%	Percentage of air volume in the compacted mixture
Stiffness[a]	MPa	Four-point bending stiffness (EN 12697-26, 2012), method B

[a]The target feature.

TABLE 2b Overview of the categorical features in the dataset.

Categorical features					
Feature name	**Subcategories**				
Compaction setup (CD)	Field roller (FR)	Hand roller (HR)	Mini roller (MR)	Segment compactor (SC)	Shear box (SB)
Mixing setup (MS)	Planetary mixer (PM)	Asphalt plant (AP)		Forced action mixer (FAM)	
Polymer-modified bitumen (PMB)	Yes			No	

(Berangi et al., 2025) and is summarized in Table 2a,b. It comprises data points collected from six different road construction projects to reduce potential model bias. These projects were not experimental test sections but rather real-world road segments constructed for public use. All samples were aged under controlled laboratory conditions at a constant temperature of 13 $\pm$ 2°C, without exposure to ultraviolet radiation or moisture. Given the relatively small geographical size of the Netherlands, it is reasonable to assume limited variability due to spatial differences.

The quality of the collected dataset plays a critical role in determining the accuracy, robustness, and generalizability of the resulting model (Domingos, 2012). In order to enhance the quality of the data, key important preprocessing steps such as data cleaning (Rahm & Do, 2000), encoding categorical features (Kuhn & Johnson, 2013), feature selection (Guyon & Elisseeff, 2003), and feature scaling (Kotsiantis et al., 2006) should be carried out (Zheng & Casari, 2018). The following subsections provide a brief description of the key preprocessing steps.

2.2.1 | Data cleaning

In the data cleaning step, detecting and correcting errors or inconsistencies in raw data should be carried out to improve its quality and reliability (Chu et al., 2016).

Common tasks in data cleaning are handling missing data, reducing noise in the data, eliminating outliers, and resolving discrepancies in the dataset (Dasu & Johnson, 2003). Since the dataset in this study was collected from different construction projects, carrying out the cleaning step is essential to ensure that subsequent analysis and modeling are based on accurate and consistent information.

2.2.2 | Encoding categorical features

In encoding categorical data, categorical features are converted into a numerical format that ML algorithms can process (Kotsiantis et al., 2006). Since MINN requires numeric input, categorical features should be encoded as numbers.

In this research, one-hot encoding (Okada et al., 2019) is selected since it can directly map each category to a unique vector with a binary representation. The binary representation is particularly well-suited for NN input layers, where the model can easily interpret the presence or absence of a category, such as compaction and mixing setups, using individual neurons (Poslavskaya & Korolev, 2023).

After data cleaning and transformation of categorical features, reducing the dimensionality of the dataset becomes essential. High-dimensional data such as this dataset that include bitumen properties (e.g., penetration, phase angle, and softening point) and mix composition variables (e.g., sand, filler, and stone volume percentages) may lead to overfitting and reduced model interpretability (Guyon & Elisseeff, 2003). Therefore, maintaining a lean and relevant input feature set is critical to reducing the above-mentioned issues (Guyon & Elisseeff, 2003). To reduce dimensionality, feature selection methods are an effective solution (Liu & Motoda, 2007). However, prior to implementing future selection methods, the dataset should be split into training and testing subsets (Brownlee, 2020).

2.2.3 | Feature selection

Feature selection methods can be broadly classified into three categories (Pandey et al., 2019): filter methods (Guyon & Elisseeff, 2003), wrapper methods (Kohavi & John, 1997), and embedded methods (Zou & Hastie, 2005). In this study, an approach called BorutaShap (Keany, 2020), which integrates elements of both wrapper and embedded methods, was used to identify the optimal feature set. The BorutaShap was selected with the expectation that it can handle the complex and inter-correlated nature of the dataset as discussed previously. A brief description of the BorutaShap methods is presented in the following subsection.

BorutaShap

BorutaShap extends the original Boruta algorithm (Kursa & Rudnicki, 2010) by incorporating SHapley Additive exPlanations (SHAP; Lundberg & Lee, 2017), which enhances its ability to capture global feature importance, particularly when dealing with interdependent and nonlinear relationships. Compared to the traditional Boruta method, BorutaShap offers a more computationally efficient and reliable approach while maintaining robust feature selection performance (Yao et al., 2022).

In the BorutaShap, while the Boruta algorithm is used to determine the relevant features, SHAP is implemented to rank them according to their importance. In the Boruta algorithm, a set of shadow features (Ahmed et al., 2022; Qiao et al., 2023) is randomly created by permuting each feature in the dataset. Then, a base model, which is traditionally a random forest model (Breiman, 2001), is trained on the original dataset alongside the shadow features. The importance score of each feature is compared to the scores of its corresponding shadow features. If a feature is found to be "more important" than all its shadow features, it is labeled as "confirmed" and retained in the feature set. Conversely, if a feature is not more important than any of its shadow features, it is labeled as "rejected" and removed from the feature set. It is noted that in a scenario where the feature is more important than some of its shadow features but not all, it is deemed "tentative," and the process is repeated with additional iterations.

Although Boruta is a robust feature selection method, it relies on the accuracy of feature importance calculations, which may be biased or insufficient for certain datasets (Kursa & Rudnicki, 2010). By incorporating SHAP values as the feature selection method in Boruta, the interpretability of SHAP's feature explanations is combined with the robustness of the Boruta algorithm. The combination ensures that only the most significant features with the highest contribution are retained in the dataset (Ejiyi et al., 2024).

Selection of base models for BorutaShap

While SHAP is a "model-agnostic" method (Nordin et al., 2023; Ribeiro et al., 2016), its effectiveness depends on the ability of the base model to properly learn feature interactions and importance (Lundberg & Lee, 2017). Different base models capture feature importance differently, especially when the dataset exhibits non-linear relationships or varying noise levels. Therefore, it is required to compare the performance of different models to determine the most effective model for selecting the feature based on the dataset. Since tree-based models (Friedman, 2001) are effective for managing non-linear relationships and high-dimensional data, they often produce more

```
While not terminate_feature_selection:
    # Perform feature selection steps
    previous_accepted_features = accepted_features.copy()

    # Update accepted_features and tentative_features

    if accepted_features:
        if (len(accepted_features) ==
len(previous_accepted_features)) or (len(tentative_features)
== 0):
            terminate_feature_selection = True
```

FIGURE 3 Termination criteria.

meaningful and stable SHAP values (Kazemitabar et al., 2017), compared to models like linear regression (Seber & Lee, 2003) or support vector machine (Cortes & Vapnik, 1995). In this study, the comparison of different tree-based models, including (a) XGBoost (T. Chen & Guestrin, 2016), (b) CatBoost (Prokhorenkova et al., 2018), and (c) extra-trees regressor (Geurts et al., 2006) is proposed.

Evaluation of the impact of seed values

In order to avoid overfitting and data leakage, feature selection should be considered to be carried out entirely within the training set. The consideration ensures that the test set remains a truly independent evaluation set, reserved only for final performance assessment after feature selection is complete.

The random seed value (Bengio, 2012) used in the train-test split process (Keany, 2020) influences which data points end up in the training set, potentially affecting which features are selected. To assess the stability of BorutaShap under different train-test splits, feature selection is advised to be repeated across multiple seeds, and the resulting feature sets should be compared. The comparison helps to verify that the selected features are not overly sensitive to a particular split.

Establishing a stopping criterion

In order to prevent unnecessary iterations and premature termination, a stopping criterion should be implemented. In this study, the feature selection in the BorutaShap is required to be terminated only when the conditions presented in Figure 3 are met. Incorporating this condition can maintain a balance between relevance and stability of features.

Evaluation of base models via prediction variance

The stability of each based model should be evaluated after the termination of the feature selection process. There are several prominent metrics, such as the Jaccard index (Yeh & Yang, 2016), adjusted Rand index (Santos & Embrechts, 2009), stability index (Nogueira et al., 2018), and selection frequency (Beinrucker et al., 2012) for stability evaluation.

As summarized in Table 3, pairwise indices such as "Jaccard" and "adjusted Rand" measure overlap between two feature sets, so a "k-seed" experiment requires "$\binom{k}{2}$" comparisons that are sensitive to the outliers (Kalousis et al., 2006; J. Li et al., 2018). "Nogueira's stability index" aggregates all seeds but penalizes even a single feature's absence, leading to deceptively low scores when BorutaShap iterates for many rounds (Jiang et al., 2022; Nogueira et al., 2018). Likewise, "*Selection-frequency*" treats each feature independently, ignoring the "set" structure that matters when a wrapper selects features jointly (Nouraie & Muller, 2025). Therefore, none of the existing metrics are suitable to be used in BorutaShap as it focuses on the stability of confirmed features across all runs. Therefore, a metric is introduced in this study so-called "average common important feature ratio (ACIF)" (see Equation 21). The ACIF evaluates the variability of each base model by quantifying the ratio of retained acceptable features across various random seeds. The ACIF varies from "0" (no features retained) to "1" (all features retained), with higher values signifying less variation.

$$\mathrm{ACIF}_i = \frac{1}{n}\sum_{j=1}^{n}\frac{\#CIF}{\#IF_{i,j}} \tag{23}$$

where

1. CIF_i refers to the proportion of features that are frequently identified as important by model "i,"
2. ACIF: average common important feature,
3. IF denotes the set of important features,
4. n is the total number of random seed iterations used for the base model,
5. $IF_{i,j}$ represents the count of features marked as important by model "I" during the "jth" seed iteration.

2.2.4 | Feature scaling

Feature scaling is required because MINN assumes uniform feature scales, and features that vary significantly in the range may skew the performance of the model or cause instability. Additionally, scaling makes it easier to compare and combine features of different units in a unified manner. Since the features in the collected dataset (Berangi et al., 2025) have different ranges, scaling these features within a specific range should be carried out. In this study, MinMaxScaler (Brownlee, 2016) is chosen

TABLE 3 Comparison of existing stability metrics and their limitations in the BorutaShap context.

Metric	Measurement index	Limitation in BorutaShap context
Jaccard index	The proportion of overlap between feature sets obtained from two different selection runs	Evaluates only pairwise similarity; does not reflect overall agreement across multiple iterations or seeds
Adjusted Rand index	Similarity measure that quantifies the agreement between two sets, adjusted for chance	Limited to pairwise comparisons; does not generalize well to assessing ensemble-level stability
Stability index (Nogueira et al., 2018)	Measures agreement across multiple runs using an entropy-based formulation	Highly sensitive to small differences in feature sets; may over-penalize minor variations in iterative methods
Selection frequency	The proportion of runs in which each individual feature is selected	Selection frequency evaluates features independently and does not capture the collective stability of the feature set across all runs

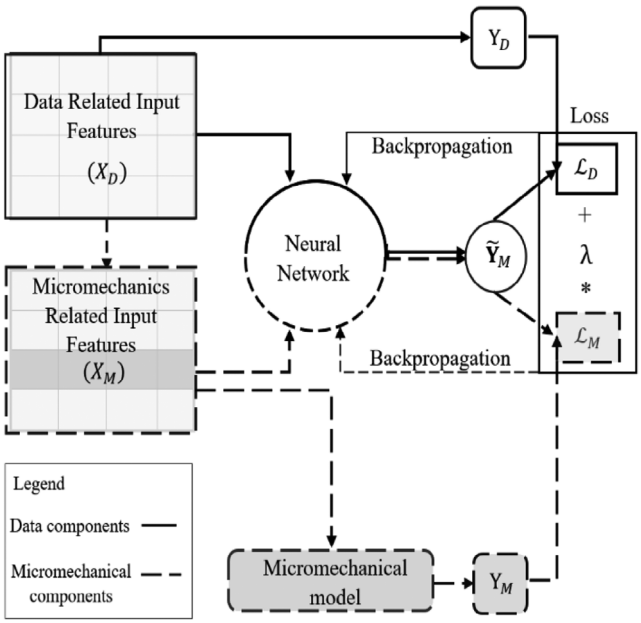

FIGURE 4 Architecture of the proposed neural network framework incorporating both data-driven and micromechanics-informed input representations.

due to its ability to preserve binary attributes generated through one-hot encoding.

2.3 | Stage 3: Development of micromechanics-infused framework

Figure 4 shows the architecture for developing MINN. The architecture encompasses different components, which will be explained in the following subsections.

2.3.1 | NN

The NNs have emerged as a fundamental component in contemporary ML because they can effectively represent intricate and non-linear relationships within datasets (LeCun et al., 2015). During the training process of NNs, the adjustments of weights and biases are achieved by utilizing a backpropagation technique (Rumelhart et al., 1986). The technique calculates error gradients (g_t) through the use of the chain rule (Rumelhart et al., 1986) to evaluate the loss ($\mathcal{L}$) (Rumelhart et al., 1986) of the output data with respect to each weight and bias (see Equations 24 and 25). The process discussed above supports NNs gradually diminishing the disparity between the predicted and observed stiffness metrics.

$$g_t = \frac{\partial \mathcal{L}}{\partial \theta} |_{\theta_{t-1}} \tag{24}$$

The loss function "$\mathcal{L}$" is formulated as follows:

$$\mathcal{L}\left(\widetilde{Y_M}, Y_D\right) = \frac{1}{M} |\widetilde{Y_M} - Y_D|_2^2 \tag{25}$$

where

1. θ_t represents the model parameters being updated, including both weights and biases;
2. θ_{t-1} indicates the values of the parameters at the previous iteration (i.e., at time step $t-1$)
3. $\widetilde{Y_M}$ and Y_D refer to the predicted and actual stiffness based on the data, respectively;
4. M is the number of samples;
5. The loss function employs the mean squared error (MSE) to reduce prediction error and adjust the model's weights and biases accordingly.

Since the loss function in Equation (25) only depends on "Y_D," the input data (X_D) plays a crucial role in the entire learning process. Thus, any issues related to data quality can directly impact the accuracy and reliability of the model's predictions (Deng et al., 2024). As it was discussed in the introduction, relying solely on data-driven learning without incorporating physical knowledge may lead to issues such as overfitting or physically meaningless results (Karpatne et al., 2017). To address these issues, this study integrates micromechanics into the NN model, ensuring that predictions are both data-driven and physics-informed.

2.3.2 | Micromechanical model integration

In order to ensure that the predictions are both data and micromechanical-informed, the interactions between aggregates and bitumen should be captured. To capture the interactions, a dataset (X_M) that represents the micromechanical features linked to the micromechanical stiffness (Y_M) should be created. The "Y_M" should be calculated directly using the selected micromechanical model. It is noted that the "X_M" and "*Ym*" works together with "X_D" and "Y_D" in NN.

Creating "X_M" should be followed by making it dimensionally compatible with "X_D" via applying a transformation process using feature masking (Daw et al., 2021; Willard et al., 2020). The transformation process used in this study preserves the actual values of each required micromechanical feature (*f*) while averaging all other features across samples (Deng et al., 2024). The averaging process removes unnecessary variation from irrelevant features, ensuring that the dimensionality of "X_M," matches "X_D" without introducing unnecessary noise (Deng et al., 2024). The general formulation of the process mentioned above is shown in Equation (26).

$$\mathrm{X_{M,\,f}} = \begin{bmatrix} \frac{1}{k}\sum_{i=1}^{K} x_{o,i} & \cdots & \frac{1}{k}\sum_{i=1}^{K} x_{o,i} \\ \vdots & \ddots & \vdots \\ x_{f,1} & \cdots & x_{f,k} \\ \frac{1}{k}\sum_{i=1}^{K} x_{o,i} & \cdots & \frac{1}{k}\sum_{i=1}^{K} x_{o,i} \end{bmatrix} \cdot \begin{bmatrix} \overline{X_o} \\ \vdots \\ X_f \\ \overline{X_o} \end{bmatrix} \tag{26}$$

where

1. $X_{M.f}$: The transformed micromechanical dataset, where each feature (*f*) is handled appropriately,
2. *"f"*: Any required feature for the micromechanical mode,
3. X_f: The actual values of the required feature (f) across all samples (retained as-is),
4. $\overline{X_o}$: Other features, which are averaged across all samples,
5. *"k"*: The total number of samples in the dataset,
6. *"i"*: The index used to represent each individual sample, where $i \in \{1,2,\ldots,k\}$.

2.3.3 | Defining the micromechanics-guided loss function

In addition to including micromechanically meaningful features (X_M), a customized loss function should be designed to evaluate both the prediction accuracy and the consistency with the micromechanical relationship. The customized loss function requires the MINN model to minimize both the prediction error derived from the data and the deviation from the micromechanically consistent behavior described by the micromechanical model (Karpatne et al., 2017). In order to create such a loss function, data-driven error terms such as MSE should be combined with a micromechanical penalty term, which penalizes predictions that violate micromechanical principles. The loss function used in this research is presented in Equation (27).

$$Loss = \mathcal{L}_D + \lambda \mathcal{L}_M, \tag{27}$$

where the training loss $\mathcal{L}_D$ (see Equation 25), and $\mathcal{L}_M$ is the micromechanics-based loss. $\mathcal{L}_M$ is weighted by "λ which is a hyperparameter (J. Chen et al., 2019). The optimum value for "λ and other hyperparameters in MINN should be determined using optimization techniques. In the following section, a brief description of model training and optimization is presented.

2.3.4 | Model training and optimization

Hyperparameters control the level of complexity or regularization (J. Chen et al., 2019) in MINN (Xia et al., 2017). In this study, different hyperparameters such as the number of hidden layers, the number of neurons per layer, the dropout rate (Srivastava et al., 2014), the learning rate (Kingma & Ba, 2014), and the micromechanics regularization parameter (λ) were considered. These hyperparameters were considered due to their influence on model complexity (Bianchini & Scarselli, 2014), regularization (Girosi et al., 1995), and convergence (Y. Li & Yuan, 2017), which are critical for balancing predictive accuracy (Poernomo & Kang, 2018; Sun et al., 2021). The number of hidden layers and the number of neurons per layer control model complexity, allowing the network to learn intricate patterns. However, increasing them excessively can lead to overfitting. To mitigate overfitting, the dropout hyperparameter acts as a regularization mechanism, randomly

TABLE 4 Possible ranges for hyperparameters of the *MINN*.

Hyperparameter	Possible range
Dropout	[0,0.5]
Learning rate	$[10^{-7},10^{-1}]$
Hidden layers	[2,4]
Neurons per layer	[4,128]
λ	$[10^{-12},10^{-6}]$

deactivating neurons during training. In parallel, the learning rate influences convergence efficiency, where a high value may cause instability, while a low value slows down learning.

Since MINN has several hyperparameters as discussed above, optimization (see Equation 28) plays a critical role in enhancing model performance (L. Yang & Shami, 2020).

$$x^* = \text{argmin}_{x \in X} f(x) \tag{28}$$

where

1. *"f(x)"* represents the objective function that needs to be minimized,
2. *"x*"* denotes the set of parameters that produce the minimum value of *"f(x),"*
3. *"x"* refers to a hyperparameter that can take any value from the search space *"X."*

There are several methods for hyperparameter optimizations, such as grid search (Liashchynskyi & Liashchynskyi, 2019) and random search (Bergstra & Bengio, 2012). In this research, Bayesian optimization is selected for hyperparameter optimization due to its ability to identify optimal hyperparameters efficiently (L. Yang & Shami, 2020). Unlike grid search and random search, Bayesian optimization is an informed optimization technique (Kim et al., 2023), which means that it utilizes prior evaluation knowledge to guide subsequent searches. In Bayesian optimization, prior knowledge is leveraged by building a probabilistic model to select the optimum values for hyperparameters. Detailed information about Bayesian optimization can be found elsewhere (Snoek et al., 2012).

A critical step in applying Bayesian optimization is selecting a search space by defining the possible range of values for hyperparameters (Snoek et al., 2012). The selection of an appropriate search space is essential, as a narrow or broad range may exclude optimal solutions (Snoek et al., 2012). In this study, the ranges for hyperparameters (see Table 4) are determined based on insights from the literature (Bengio, 2012; Smith, 2017) and the need for mitigating overfitting. Bayesian optimization then explores these defined ranges, using probabilistic modeling to refine and narrow the search for optimal hyperparameter values.

Once the search space is defined, the selected hyperparameters must be evaluated on a properly trained model. To ensure that the obtained hyperparameters are not biased by a single train-test split, cross-validation (Anguita et al., 2012) is recommended. Cross-validation improves the reliability of hyperparameter selection by averaging performance across multiple train-test splits. While hyperparameter optimization aims to enhance model efficiency and generalization, it is crucial to evaluate whether these optimizations translate into improved predictive accuracy and robustness. Therefore, a comprehensive model evaluation is necessary to determine the effectiveness of the trained MINN.

2.4 | Stage 4: Model evaluation

In order to evaluate the accuracy of the MINN, different accuracy metrics such as the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute percentage error (MAPE) are advised to be used (Provost & Fawcett, 2013). Although these metrics are useful for assessing accuracy, they are not sufficient for evaluating the interpretability and stability of the model.

Interpretability is widely recognized as a crucial aspect of predictive modeling, particularly in scientific and engineering applications where trust in model predictions is essential (Brundage et al., 2020; Lo Piano, 2020; Thiebes et al., 2021). In the ML approaches, interpretability is often assessed using methods such as SHAP (Kazemitabar et al., 2017) and local interpretable model-agnostic explanations (LIME; Koli et al., 2024; Zafar & Khan, 2021). These methods help to quantify the way individual inputs influence a prediction of the model. While SHAP and LIME methods provide valuable insights into model behavior, they are primarily designed for purely data-driven models and rely on mathematical relationships within the dataset (Misheva & Osterrieder, 2023; Salih et al., 2025). As a result, these methods do not explicitly evaluate whether the predictions of a model remain consistent with underlying physical principles. In models where ML models integrate domain-specific knowledge like MINN, interpretability should be assessed by verifying whether its predictions align with "Y_M."

Stability refers to the extent to which MINN produces consistent predictions when trained repeatedly under different initializations (Bousquet & Elisseeff, 2002). Initialization is the process in which the initial parameters of the MINN (such as weights and biases) are randomly assigned before training begins. These parameters serve as the starting point for learning and are typically initialized

using methods like Xavier's uniform distribution (Glorot & Bengio, 2010). Since the initialization is random, different training runs can start from different initial values, leading to variations in the final trained parameters (Semmelrock et al., 2023). The randomness can result in the model converging to different local minima, potentially affecting its generalization and performance.

To control the randomness, a random seed (Bouthillier et al., 2019) is often set before training. A random seed is a fixed starting point for the random number generator used in parameter initialization. When the same seed is used, the model follows the same initialization process, ensuring reproducibility. However, using only a single seed might lead to misleading results, as it may represent only a specific set of initial conditions, and the performance of the model might not be generalizable. When different seeds are used, the randomly initialized parameters vary, which can cause fluctuations in the predictions, highlighting the need to evaluate the stability across multiple initializations to ensure consistent performance.

The above issue has been reported in multiple recent studies using different ML models. Reimers and Gurevych (2017) reported that random seed variability can lead to statistically significant differences in the performance of natural language processing models, while Henderson et al. (2018) showed fluctuations in learning outcomes caused by stochasticity in the learning process (e.g., random weight initialization) in reinforcement learning settings. Similarly, Zech et al. (2019) found considerable instability in the prediction of their image-processing model using different random seeds. These findings collectively emphasize that evaluating model stability across multiple initializations is essential to ensure reproducibility and robustness.

In order to evaluate the stability of the model, the average coefficient of variation "$\overline{CV}$" (Bindu et al., 2019), as presented in Equation (29), should be calculated over different random seeds.

$$\overline{CV} = \frac{1}{N}\sum_{i=1}^{N}\frac{\sigma_i}{\mu_i} \tag{29}$$

where

- "i" and "j" are values from 1 to "N" and satisfy "$i \neq j$,"
- "N" refers to the total count of predicted output vectors,
- $\varepsilon\sigma_i\varepsilon$ indicates the standard deviation of the "ith" element across all predicted outputs,
- "$\mu_i\varepsilon$ denotes the average value of the "ith" element calculated from all predicted output vectors.

In addition to evaluating predictive accuracy and stability, it is important to assess the computational efficiency of the proposed MINN framework. Several recent studies reported that PIML models showed longer training time, compared to traditional ML models (Grossmann et al., 2023; A. J. Huang & Agarwal, 2023; Kaewnuratchadasorn et al., 2024). To evaluate the computational cost associated with the micromechanics-informed loss in this study, the convergence time of MINN is quantitatively compared to a traditional NN trained under otherwise identical conditions.

3 | RESULTS AND DISCUSSIONS

This section presents the results using the dataset presented in Table 2 and the framework explained in the previous section. For the ease of understanding of the readers, the discussions are structured into subsections, maintaining the same order as the framework presented in Figure 1.

3.1 | Data preparation

An overview of the dataset after carrying out cleaning and encoding categorical features is summarized in Table 5a,b, which resulted in 425 unique data points. Table 5a presents the numerical input features, including their respective measurement units, minimum and maximum values, and standard deviations. The dataset comprises a diverse set of features, such as "age" and "target density," which were incorporated into the model development process. Each feature represents a specific characteristic of the asphalt mixture, contributing to a comprehensive understanding of its properties. The physicochemical properties (Le Guern et al., 2010) of bitumen are described through features such as bitumen penetration and "bitumen phase angle," which influence the mechanical behavior of the material. Additionally, the composition of the asphalt mixture is captured by key features such as the target mass composition of stone, sand, filler, and bitumen, which determine the overall mixture structure.

The presented dataset in Table 5a,b required minimal cleaning, as it was generally well-structured. Primary cleaning steps were standardizing all numerical features (e.g., converting kPa to MPa) and uniformly labeling categorical variables such as mixing and compaction methods to facilitate encoding. Duplicate records and inconsistent entries were then identified and removed using sample ID tracking. Carrying out the above-mentioned steps, the dataset contained a few missing values, which were removed rather than imputed to prevent introducing potential biases.

Following the removal of the missing values, since the dataset did not have the required feature of the modified

TABLE 5a Overview of the numerical features after preprocessing steps.

Numerical features					
Feature name	**Unit**	**Mean**	**Min**	**Max**	**Standard deviation**
Target density (TD)	kg/m^3	2380.85	2360.00	2399.00	10.71
Bitumen penetration (BP)	0.1 mm	25.54	11.00	53.00	10.54
Bitumen phase angle (δ)	Degree (°)	58.35	39.96	66.57	8.15
Target sand volume percentage (TSA)	%	43.24	38.94	46.63	2.19
Target filler volume percentage (TF)	%	6.30	5.76	7.65	0.60
Target stone mass percentage (TST)	%	39.54	36.42	42.48	2.10
Target bitumen volume percentage (fb)	%	0.11	0.10	0.13	0.01
Reclaimed asphalt pavement (RAP)	%	58.28	50.00	65.00	6.19
Age	Year	0.54	0.00	6.00	1.23
Bitumen softening point (SP)	°C	63.32	55.80	82.60	8.02
Complex modulus (G*)	Pa	1,140,421.61	194,139.51	6,632,724.39	1,648,476.80
Density (D)		2391.09	2307.81	2453.60	31.67
P_a[a]	Unitless	0.51	0.29	0.87	0.16
Volume aggregate fraction (fa)[a]		0.85	0.82	0.89	0.01
Air void (fv)	%	3.61	0.51	6.93	1.60
Stiffness	MPa	9737.44	6144.00	17,866.47	2092.53

[a]Required features for the micromechanical model.

TABLE 5b Overview of the categorical features in the dataset after one-hot encoding.

Feature name	**FR**[a]	**HR**[a]	**MR**[a]	**SC**[a]	**SB**[a]	**PM**[a]	**AP**[a]	**FAM**[a]	**PMB_Yes**[a]	**PMB_No**[a]
CD[a]	1[b]	0[b]	0	0	0	–	–	–	–	–
MS[a]	–	–	–	–	–	1	0	0	–	–
PMB[a]	–	–	–	–	–	–	–	–	1	0

[a]For acronyms, please refer to Table 2b.
[b]Each row in the dataset contains binary values (1 or 0) for each encoded feature, indicating the presence (1) or absence (0) of a particular category.

Hirsch model, two features, "P_a" (see Equation 22) and "f_a" (see Equation 17), were developed. Furthermore, it was previously discussed that categorical features cannot be directly used by MINN. Therefore, to make these features machine-interpretable, they were encoded using one-hot encoding (Okada et al., 2019) as shown in Table 5b.

As discussed in the methodology, a key aspect of this study is selecting key features to be incorporated into the model development. As can be seen from Table 6, four different base models (i.e., random forest, XGBoost, CatBoost, and extra trees) were compared.

Figure 5 shows the significance of each feature quantified by "Z-scores" (Salkind, 2006). The "Z-scores" are derived from the SHAP values assigned to each feature in the BorutaShap, indicating their relative contribution to the output of the model. Therefore, features with higher "Z-scores" contributed more to the output of the model.

Although the Z-score highlights the importance of features, it does not indicate the stability of their selection. The boxplots in Figure 5 show the distribution of Z-scores across multiple BorutaShap runs for each feature. Narrower boxplots indicate features that are consistently important, whereas wider boxplots suggest variability in their importance. These boxplots were further subcategorized as "A," "R," and "S" to represent the accepted, rejected, and shadow features, respectively. As it was discussed in the methodology, features are being selected in BorutaShap that perform better than pure randomness represented by shadow features.

As shown in Figure 5, "Pa" and "fa" are among the top five most important features, as indicated by the dashed rectangle labeled "M." These five features exhibit the highest Z-score values and relatively narrow box plots, signifying their strong relevance and stability in the analysis. In contrast, "FR" (highlighted by another dashed

TABLE 6 BorutaShap results across different base models and seeds.

Base model	Seed	Iterations[a]	R^2 test	Average R^2	IF[b]	CIF[b]	ACIF[b]
Random forest	0	1600	0.77	0.79	12	9	0.70
	7	6400	0.80		15		
	31	1600	0.84		12		
	42	1600	0.77		13		
XGBoost	0	200	0.77	0.76	7	6	0.76
	7	400	0.68		8		
	31	400	0.86		8		
	42	400	0.76		9		
CatBoost	0	1600	0.77	0.83	18	15	0.88
	7	400	0.88		17		
	31	400	0.84		16		
	42	1600	0.83		17		
Extra-trees	0	1600	0.82	0.84	21	21	0.98
	7	1600	0.86		21		
	31	1600	0.88		21		
	42	400	0.83		23		

Abbreviations: ACIF, average common important feature ratio; R^2, coefficient of determination.

[a]The number of iterations listed in the table refers to the final run where the count of selected features remained unchanged between the two latest iterations.

[b]For acronyms, refer to Equation (23).

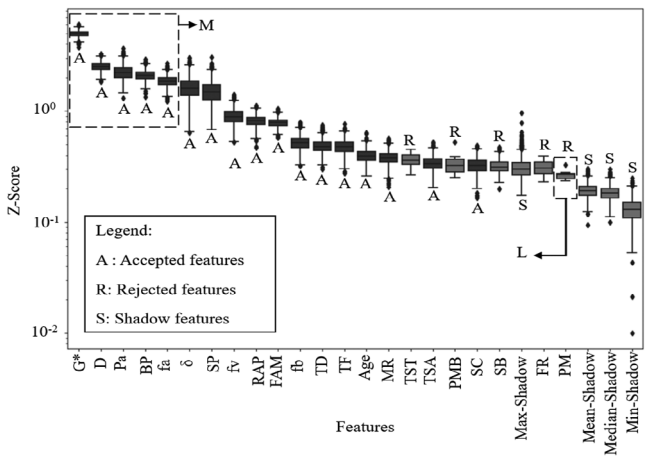

FIGURE 5 Results of the BorutaSHAP feature selection (for acronyms, refer to Table 5).

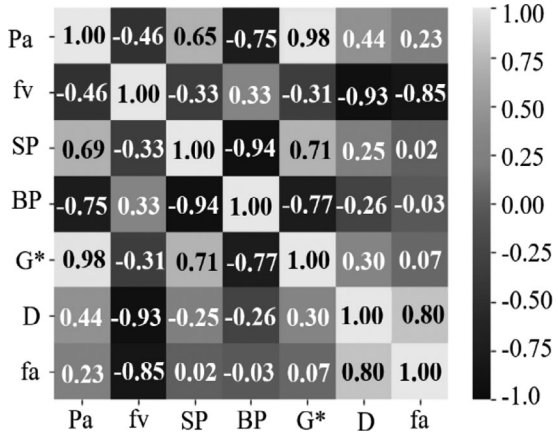

FIGURE 6 Results of the correlation analysis (for acronyms, refer to Table 2).

rectangle) has the lowest Z-score, suggesting it contributes the least to the predictive model.

The prominent position of "Pa" and "fa" underscores the critical role of micromechanical properties in governing asphalt mixture behavior. Their high importance suggests that the internal mechanical interactions at the microscale significantly influence the overall material performance. Furthermore, the observation from the figure is consistent with the concept of physics-aware feature importance, which argues that the features the model relies on most should be those that impose the governing physical mechanisms (Willard et al., 2022a). Furthermore, the results shown in Figure 5 support the necessity of incorporating micromechanical features into the evaluation and design of asphalt mixtures, as these features serve as key indicators of overall pavement performance and enhance the interpretability of the MINN framework.

Among the top five features in Figure 5, "G*" showed the highest "Z-score," which signifies its high influence on stiffness. This finding aligns with the common knowledge that a higher "G*" value corresponds to increased stiffness and enhances the ability of mixtures to endure deformation under traffic-induced stresses and environmental conditions (Leng et al., 2021; Sathvik et al., 2024). As can be observed from the figure, "D" was the second

most important feature. This observation is consistent with prior studies demonstrating that higher compaction results in higher aggregate interlocking and reduced air voids, leading to enhanced stiffness and durability (Brown et al., 2001; Tran et al., 2016).

In contrast to the features discussed above, "FR" had the lowest Z-score, indicating minimal impact on asphalt stiffness, which is aligned with the previous studies (Berangi et al., 2025; Plati et al., 2016). These studies reported that different compaction setups influence the mechanical performance of the asphalt mixture but do not significantly dictate final stiffness when compared to other factors such as mix design and temperature control.

While BorutaShap identifies the most relevant features based on their individual contribution to the model, it does not account for redundancy or multicollinearity (Guyon & Elisseeff, 2003; Kursa & Rudnicki, 2010). To reduce multicollinearity, a correlation analysis using Spearman's correlation (De Winter et al., 2016) was carried out as shown in Figure 6.

As can be seen from the figure, since "f_v" and "D" had high correlations ($|\rho| > 0.8$) with "f_a," they were not considered in MINN for further analysis. Additionally, the feature "SP" was not selected due to its high correlation ($|\rho| = 0.94$) with "BP."

3.2 | Micromechanics-infused framework

The architecture of the MINN model is presented in Figure 7a. To obtain the optimum values for the hyperparameters presented in the figure, Bayesian optimization combined with 10-fold cross-validation was applied to the search space defined in Table 4.

As shown in the corresponding figure, the architecture of the MINN starts from an input layer consisting of 14 neurons, each representing a feature selected from the feature selection process. These inputs are then propagated through three fully connected hidden layers, each comprising 32 neurons. The rectified linear unit (Y. Li & Yuan, 2017) activation function is applied to each hidden layer, enabling the model to capture complex patterns while mitigating the vanishing gradient (Hochreiter, 1998) problem, thereby enhancing training at the end of the architecture. MINN generates its prediction through a single-neuron output layer, which is dedicated to estimating the stiffness. Since the goal is to predict a continuous value, the linear activation function (Sharma et al., 2017) is used in this layer to allow unrestricted output values.

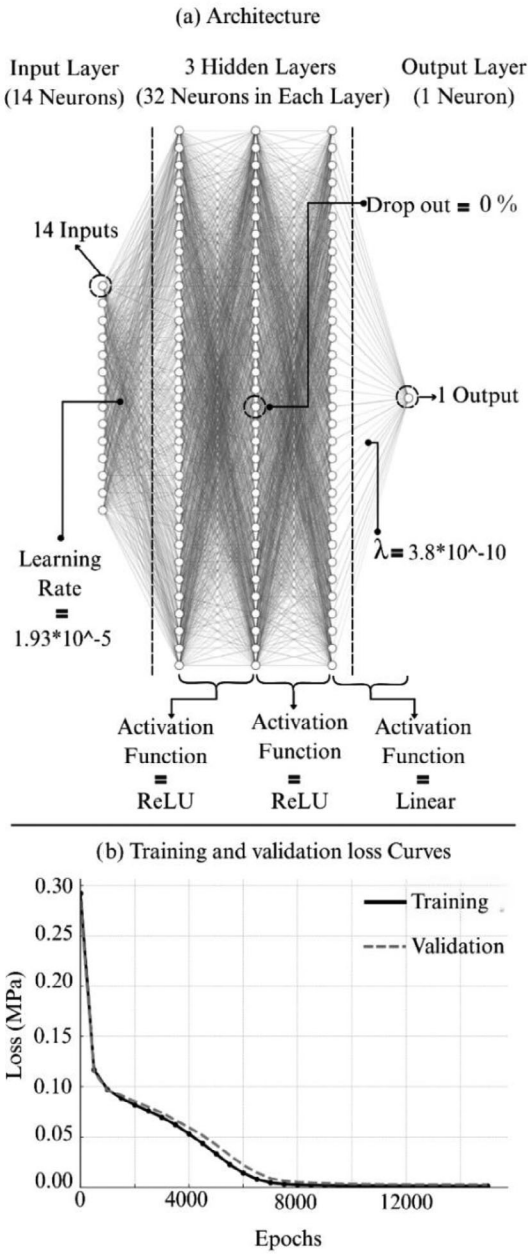

FIGURE 7 (a) Micromechanics-infused neural network (MINN) architecture and optimized values for hyperparameters and (b) training and validation loss versus epochs. ReLU, rectified linear unit.

3.3 | Model evaluation

As discussed in the methodology, the performance of MINN should be evaluated by considering aspects such as accuracy, interpretability, stability, and computational efficiency. However, before evaluating these aspects, the individual effect of key components of the research framework, such as feature selection and optimization on the accuracy, is evaluated, and the results are presented in the following subsection.

TABLE 7 Ablation results showing the effects of feature selection and hyperparameter tuning on MINN performance.

Experiment	Feature selection	Hyperparameter tuning	R^2_{Train}	$R^2_{Validation}$	R^2_{Test}	RMSE (MPa)	MAPE (%)
E_1 (baseline)	All input features	Manual	0.82	0.74	0.59	999.36	8.51
E_2	All input features	Bayesian	0.91	0.88	0.74	739.02	6.57
E_3	BurotaShap + correlation	Manual	0.89	0.87	0.77	741.42	6.61
E_4	BurotaShap + correlation	Bayesian	0.96	0.94	0.82	672.30	5.17

Abbreviations: MAPE, mean absolute percentage error; RMSE, root mean squared error.

3.3.1 | Ablation analysis: Evaluating framework components

Table 7 presents the results of an ablation study (Meyes et al., 2019) conducted to quantify the individual and combined contributions of two key components within the proposed MINN framework: (i) feature selection using BorutaShap and correlation analysis and (ii) hyperparameter tuning via Bayesian optimization on the accuracy of MINN. Four experimental configurations (E_1 to E_4) were evaluated, systematically varying the use of selected features and optimization strategies.

In "E_1," all features presented in Table 5 were used without any hyperparameter tuning. As can be seen from the table, the model yielded relatively lower accuracy, which can support the negative impact of high dimensionality on the model's performance (Guyon & Elisseeff, 2003). Incorporating Bayesian optimization while retaining all features in "E_2," led to a noticeable improvement in predictive accuracy, increasing the "R^2_{Test}" by 25% and reducing MAPE by 22%. The improvements can underscore the importance of hyperparameter tuning (Snoek et al., 2012).

In "E_3," feature selection without Bayesian optimization yielded marginally better generalization, compared to E_2, achieving a test R^2 of 0.77 and MAPE of 6.61%. The results show the role of BorutaShap in removing unnecessary inputs, which is in line with prior findings (Kursa & Rudnicki, 2010; Sebastián & González-Guillén, 2024).

The full framework is presented in "E_4," which integrates both feature selection and Bayesian optimization, achieving the highest accuracy across all subsets, with an R^2 of 0.96, 0.94, and 0.82 for training, validation, and testing sets, respectively. Considering the wide range of stiffness values presented in Table 5, the RMSE value of 672.30 MPa for the test set is relatively low, further confirming the accuracy of the model. Additionally, the MAPE values consistently ranged from 3.21% to 5.17% across datasets. The consistency can underscore the stability and reliability of the model in making predictions.

The results in "E_4" show the synergistic advantage of integrating dimensionality reduction with optimization methodologies reported by previous studies (Binder et al., 2020). The information presented in Table 7 can validate the design choice of embedding both dimensionality-reduction and adaptive-learning modules in the MINN framework.

3.3.2 | Predictive accuracy and interpretability

As shown in Figure 7b, the validation loss tracks the training loss without significant divergence and does not show signs of overfitting. The loss values were stabilized after approximately 8000 epochs, indicating that the model has reached convergence. To further evaluate generalization performance, Figure 8a presents a comparison between the measured and predicted stiffness values on the test set. As the corresponding figure shows, the alignment of the data points along the equality line indicates that the model learned patterns from the collected data, and it can make accurate predictions. It is noted that accuracy metrics alone do not guarantee the interpretability of the model. As discussed in Section 2.4, traditional interpretability methods assess the influence of input features but do not verify consistency with physical principles. Therefore, in the MINN framework, interpretability is assessed by comparing "$\widetilde{Y_M}$" against the "Y_M" (see Figure 8b). The comparison helps to ensure that the predictions remain anchored to interpretable domain knowledge principles.

In Figure 8b, the relationship "$\widetilde{Y_M}$" against the "Y_M" can be expressed as an Equation (30):

$$\widetilde{Y_M} = Y_M + \Delta_{NN} \tag{30}$$

where "Δ_{NN}" represents the residual correction produced by the NN. As shown in Figure 8b, the strong correlation ($R^2 \sim 0.98$) indicates that "Δ_{NN}" remains unbiased and proportionally small across the full stiffness range. The small value of the "Δ_{NN}" suggests that the majority of the predictions can be directly traced back to known physical relationships, thereby maintaining a high level of interpretability. Furthermore, as shown in Figure 5, the micromechanical-related features were among the most influential features, which means that "Δ_{NN}" is guided by the physically interpretable features.

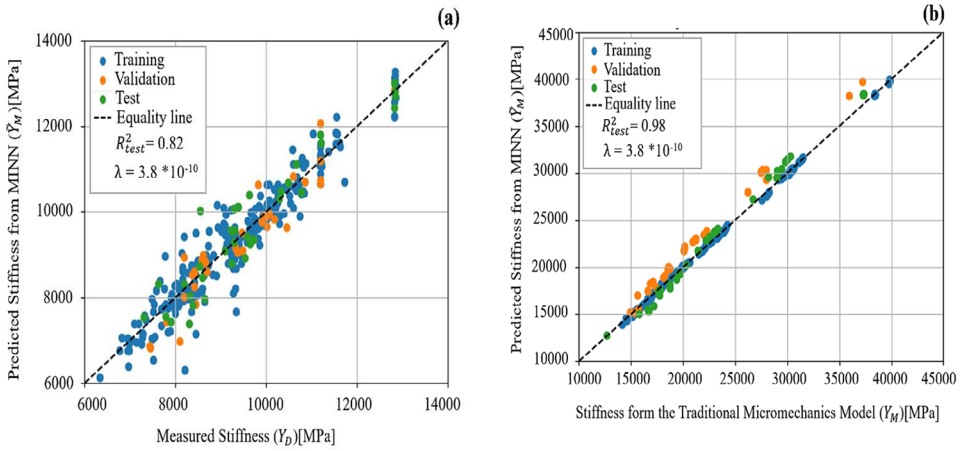

FIGURE 8 (a) Comparison between "Y_D" and "$\widetilde{Y_M}$"; (b) Comparison between "Y_M" and "$\widetilde{Y_M}$."

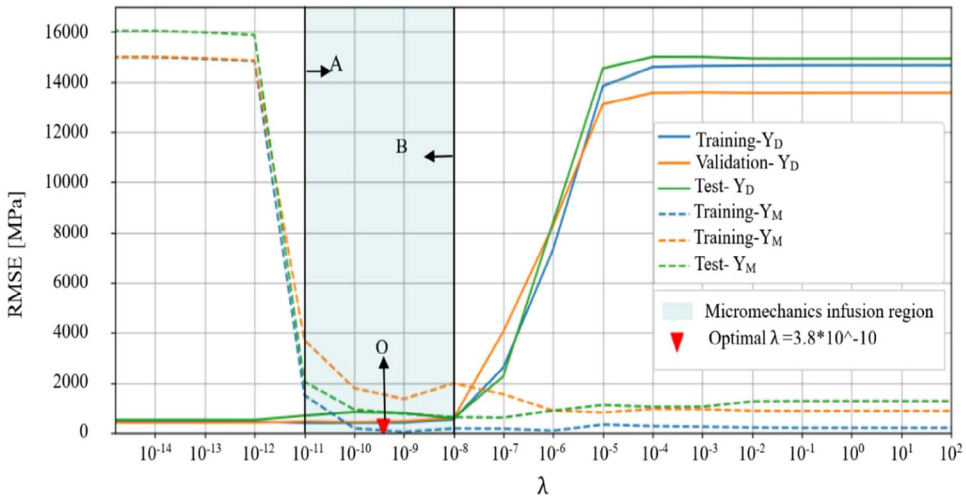

FIGURE 9 The effect of λ on root mean squared error (RMSE) for predicting Y_D and Y_M.

3.3.3 | Effect of the micromechanics infusion weighing factor on the accuracy and interpretability

In the previous section, the impact of incorporating micromechanics, using the optimized weighing factor (i.e., "λ"), on the accuracy and interpretability of MINN was presented. However, as discussed in the methodology, the "λ" can vary from small to large values, which can affect the accuracy and interpretability of the model. This section provides a brief explanation of the way different values of "λ" influence the model's accuracy and interpretability.

In order to assess the effect of "λ" on the accuracy and the interpretability of the proposed model, RMSE values were plotted for the measured stiffness and the stiffness obtained from the traditional micromechanical model (see Figure 9). Various curves in the plot represent the RMSE obtained from different datasets and output targets, including training, validation, and test sets.

As the corresponding figure shows, a region between lines "A" and line "B" is labeled as the "Micromechanics Infusion Region," which spans the interval between 10^{-11} and 10^{-8}. This region highlights the range of "λ" values where a balanced model is obtained that integrates micromechanical principles without compromising its predictive capability. Within the range, "RMSE" values remain relatively low and stable across all datasets for "Y_D" and "Y_M." The lowest "RMSE" value is achieved at 3.8×10^{-10}, marked by "O," which is obtained from the optimization process. It is noted that the obtained value reflects the best trade-off between learning from data and respecting micromechanical constraints.

A deeper investigation of Figure 9 shows that on the left side of the infusion region, where "λ" falls below 10^{-11}, the

model falls into an "under-regularized" condition (Engl et al., 1996; Hastie et al., 2009). In this condition, the model prioritizes fitting the empirical data closely, potentially at the expense of violating micromechanical law (Raissi et al., 2019). Therefore, as can be seen from the figure, the "RMSE" values for the "Y_M" increase while the "RMSE" values for "Y_D" remain low under this condition.

In contrast, on the right side of the infusion region, where "λ" increases beyond 10^{-8}, the model shifts into an "over-regularized" condition (Engl et al., 1996; Hastie et al., 2009). When the model is over-regularized, the RMSE values for "Y_M" remain low, while the "RMSE" for "Y_D" rises significantly. The above observation could be explained by the fact that the proposed model is increasingly constrained by the micromechanical law to the point where it no longer fits the data.

Based on the results shown in Figure 9, it can be concluded that incorporating micromechanical information into a data-driven model has the potential to enhance its interpretability. This finding aligns with the conclusions of a previous study (Karniadakis et al., 2021), which reported that balancing the relative strength of data and physics constraints, which are typically controlled by hyperparameters or regularization parameters, is key for model interpretability and physical consistency.

3.3.4 | Stability evaluation

The stability of the MINN was evaluated using the "$\overline{CV}$" and the results are presented in Figure 10. As shown in Figure 10a, the "$\overline{CV}$" for the base data-driven model (i.e., without incorporating micromechanics; $\lambda = 0$) is relatively low at 6.49%. However, when micromechanics is infused into the model (see Figure 10b), the "$\overline{CV}$" significantly decreases by "56%," resulting in a "$\overline{CV}$" of "2.86%." This reduction reflects enhanced stability and convergence consistency, which can be attributed to the regularizing role of the micromechanics-based loss. The micromechanical constraint acts as a regularizer, narrowing the solution space and flattening sharp directions in the optimization landscape, thus reducing sensitivity to random initializations (Karniadakis et al., 2021; Urbán et al., 2025).

Further evaluation of Figure 10a shows that when "$\lambda = 0$," the model relies purely on data-driven correlations, resulting in predictions that deviate noticeably from the line of equality (as highlighted by arrow "D"). Slightly higher deviation can be observed at higher stiffness values, suggesting that the model predictive power might be reduced at the upper end of the input values. This observation is typical behavior for data-driven models when attempting to extrapolate beyond well-represented

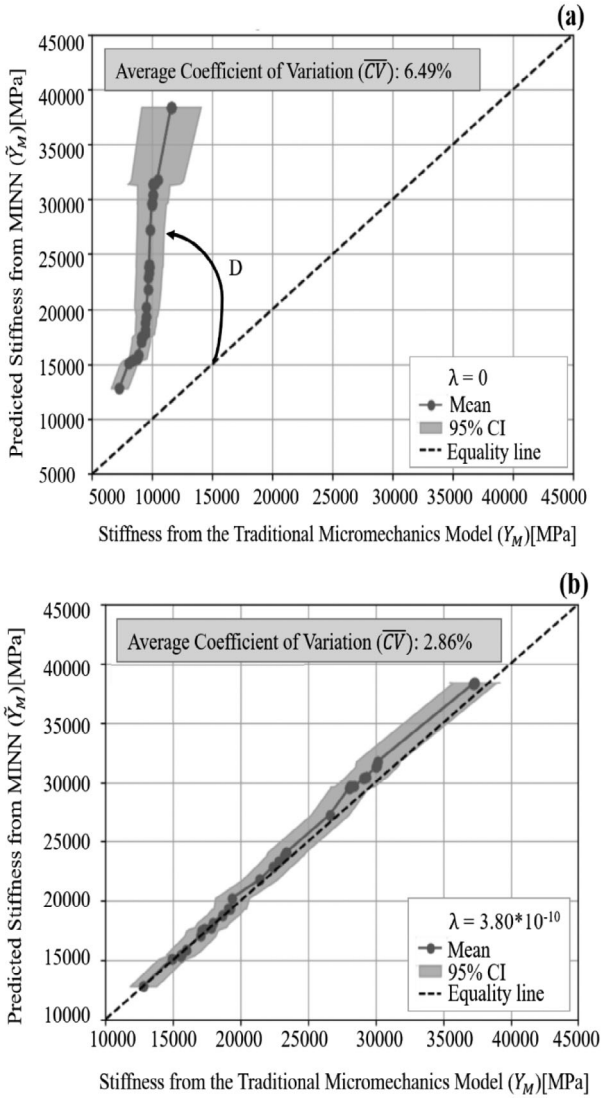

FIGURE 10 Predictions (Y_M) versus target (Y_M), (a) without infusing micromechanics and (b) with the micromechanics infusion.

regions of input space (Karpatne et al., 2017). In contrast, Figure 10b shows that when "$\lambda = 3.810^{-10}$," the predictions are closer to the line of equality, with a lower "$\overline{CV}$," suggesting improved generalization and overfitting mitigation. The improvement occurs because the micromechanical penalty injects physically plausible gradients even in sparsely sampled regions, limiting parameter (e.g., weights and biases) drift and effectively controlling model complexity (Niu et al., 2025; Willard et al., 2022a).

3.3.5 | Computational convergence time evaluation

This section aims to provide a quantitative comparison of MINN and traditional NN in terms of computational

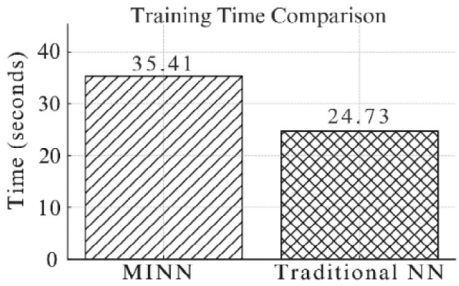

FIGURE 11 Comparison of training time between the proposed MINN and a traditional neural network.

convergence time. The convergence times were obtained based on early stopping, where training was halted if the validation loss did not improve for 500 consecutive epochs. The models were trained and evaluated on a workstation equipped with an AMD Ryzen 16-Core Processor (3.90 GHz) and a single NVIDIA Quadro RTX 8000 GPU. To ensure a fair comparison, all architectural and training components of the MINN framework were kept identical for the traditional NN, except that the "λ" was set to 0, effectively disabling the micromechanics-informed loss. The comparative results are presented in Figure 11.

As shown in Figure 11, MINN took approximately 35.41 s to reach convergence, while the traditional NN (with the micromechanics-informed component disabled by setting $\lambda = 0$) completed training in 24.73 s. The difference in the training time can be related to the extra micromechanical-based loss term incorporated in MINN. This is because micromechanical-informed loss needs additional forward passes through the micromechanical model and gradient calculation, resulting in an overhead in training time. While the observation from the figure is aligned with the findings in the literature (Grossmann et al., 2023; A. J. Huang & Agarwal, 2023; Kaewnuratchadasorn et al., 2024), researchers (Daw et al., 2021; Willard et al., 2022b) argued that PIML frameworks trade off an increase in training complexity for improved model reliability and interpretability. In this study, the 10-s increase in training time can be considered as a justifiable expense for integrating micromechanical consistency.

4 | CONCLUSION AND OUTLOOK

Reliable prediction of mechanical characteristics is essential for more accurately predicting the remaining lifetime of civil engineering infrastructures, which ultimately will help researchers and decision-makers to select more adequate materials without the need for conducting extensive laboratory experiments (Adeli, 2019; Freitag et al., 2009). In this regard, both micromechanics-based models and ML-based models seem to be viable alternatives. However, each comes with inherent strengths and limitations, which restrict their standalone applicability. This study proposed a novel computational framework (MINN) that aims to utilize the benefits of both approaches while seeking to reduce the need for significant data and laboratory experiments. The proposed framework infuses complex micromechanical modeling principles into an NN for predicting the mechanical characteristics of a visco-elastic material. The infusion enables the NN to learn from available data while being implicitly guided by micromechanical knowledge. This research focuses on using MINN to better predict the stiffness of asphaltic materials (i.e., viscoelastic in nature), which will help in better mixture design, structural design, performance evaluation, and maintenance planning.

Based on the information provided in the previous sections, the key conclusions are summarized as follows:

1. Integrating the BorutaShap algorithm for automated feature selection with correlation analysis was shown to effectively identify the most relevant features for pavement performance prediction.
2. The proposed MINN model provides high predictive accuracy across datasets, with consistent R^2, RMSE, and MAPE values, demonstrating strong generalization capabilities and stable predictive capability.
3. In this research, the optimized value obtained for "λ" from the Bayesian optimization enhanced the accuracy and interpretability of the model. Furthermore, the stability of the model is improved, as highlighted by the "$\overline{CV}$" value, which highlights enhanced consistency across different random seeds.
4. The model without micromechanics showed wider confidence intervals and deviations from equality, whereas incorporating micromechanics resulted in more consistent, physically aligned predictions.
5. Incorporating the micromechanics-informed loss resulted in a slight increase in computational convergence time relative to the conventional NN.

In summary, the analysis indicated that the proposed MINN has the capability to synergize the strengths of both micromechanical modeling and data-driven learning. In the engineering domain, where datasets on mechanical properties are often limited in size, MINN, by integrating micromechanical models, seems to produce better results with the availability of moderate datasets (see Section 2.2). This conclusion is in line with the conclusion of previous papers in other domains (Kaewnuratchadasorn et al., 2024; Taghizadeh et al., 2025; G. Wang et al., 2024).

Compared to traditional micromechanical models alone, MINN enables adaptation to variations in real-world

conditions (e.g., different aggregate types, binder sources, or compaction methods) by learning subtle patterns from available data without abandoning the underlying physics. Compared to purely data-driven models, MINN improves generalization, interpretability, and stability, particularly when data are limited, heterogeneous, or noisy. The advantages mentioned above are expected to make the MINN framework well-suited for practical deployment in pavement management systems, where reliable predictions must often be made across diverse conditions with limited site-specific calibration data.

4.1 | Future research direction

The modified Hirsch model has been widely used in various domains (D. W. Christensen & Bonaquist, 2015; Díaz-Mendoza et al., 2022; Lezgy-Nazargah et al., 2018) due to its reliable performance across a variety of materials. However, its embedding in the MINN framework assumes an idealized combination of parallel/series configuration of asphalt mixture phases. While this assumption enhances interpretability, it might restrict the capacity of the model for superior prediction accuracy, particularly in cases where real material behavior deviates from the simplified micromechanical assumptions. Although the data-driven component of the MINN partially mitigates this restriction, the embedded physics term continues to act as a soft prior, potentially constraining the model from reaching maximum accuracy.

The MINN framework can be adapted (developed) for other engineering materials. The key requirement is the availability of an analytical micromechanical relation that links phase-level properties to macro-scale properties. Any composite for which a closed-form homogenization formula exists can be accommodated by replacing the Hirsch term and supplying the requisite phase descriptors.

In terms of improvement in the model framework, another future direction could be to explore the incorporation of self-supervised learning techniques (Rafiei et al., 2024) to enhance the synergy between data-driven learning and physical reasoning. Such integration has the potential to improve model generalization, particularly when validated against independent datasets generated under different experimental conditions. Furthermore, the proposed MINN framework in this study can be compared with other physics-informed models in the literature to provide a benchmark for researchers.

ACKNOWLEDGMENTS

This research is part of the NL-LAB project (Nederlands Langjarige Asfalt Bemonstering), carried out under the Knowledge-based Pavement Engineering (KPE) research program framework. The KPE framework is supported by the Ministry of Infrastructure and Water Management in the Netherlands (project Number: 31164321).

REFERENCES

Adeli, H. (2019). Four decades of computing in civil engineering. In C. Ha-Minh, D. Dao, F. Benboudjema, S. Derrible, D. Huynh, & A. Tang. (Eds.), *Lecture notes in civil engineering: Vol. 54. CIGOS 2019, Innovation for sustainable infrastructure*. Springer. https://doi.org/10.1007/978-981-15-0802-8_1

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

Ahmed, H., Soliman, H., & Elmogy, M. (2022). Early detection of Alzheimer's disease using single nucleotide polymorphisms analysis based on gradient boosting tree. *Computers in Biology and Medicine*, *146*, 105622. https://doi.org/10.1016/j.compbiomed.2022.105622

Alam, K. M. R., Siddique, N., & Adeli, H. (2019). A dynamic ensemble learning algorithm for neural networks. *Neural Computing and Applications*, *32*(12), Article12. https://doi.org/10.1007/s00521-019-04359-7

Anderson, R. M., King, G. N., Hanson, D. I., & Blankenship, P. B. (2011). Evaluation of the relationship between asphalt binder properties and non-load related cracking. *Journal of the Association of Asphalt Paving Technologists*, *80*, 615–664. https://trid.trb.org/View/1135500

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

Barugahare, J., Amirkhanian, A. N., Xiao, F., & Amirkhanian, S. N. (2022). ANN-based dynamic modulus models of asphalt mixtures with similar input variables as Hirsch and Witczak models. *International Journal of Pavement Engineering*, *23*(5), 1328–1338. https://doi.org/10.1080/10298436.2020.1799209

Behnood, A., & Mohammadi Golafshani, E. (2021). Predicting the dynamic modulus of asphalt mixture using machine learning techniques: An application of multi biogeography-based programming. *Construction and Building Materials*, *266*, 120983. https://doi.org/10.1016/j.conbuildmat.2020.120983

Beinrucker, A., Dogan, Ü., & Blanchard, G. (2012). A simple extension of stability feature selection. In A. Pinz, T. Pock, H. Bischof, & F. Leberl (Eds.), *Lecture notes in computer science: Vol. 7476. Pattern recognition. DAGM/OAGM 2012* (pp. 256–265). Springer. https://doi.org/10.1007/978-3-642-32717-9_26

Belytschko, T., Liu, W. K., Moran, B., & Elkhodary, K. (2014). *Nonlinear finite elements for continua and structures*. John Wiley & Sons.

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. In G. Montavon, G. B. Orr, & K.-R. Müller (Eds.), *Neural networks: Tricks of the trade* (Vol. 7700, pp. 437–478). Springer. https://doi.org/10.1007/978-3-642-35289-8_26

Berangi, M., Lontra, B. M., Anupam, K., Erkens, S., Van Vliet, D., Snippe, A., & Moenielal, M. (2025). Gradient boosting decision trees to study laboratory and field performance in pavement management. *Computer-Aided Civil and Infrastructure Engineering*, *40*(1), 3–32. https://doi.org/10.1111/mice.13322

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*(2), 281–305.

Bianchini, M., & Scarselli, F. (2014). On the complexity of neural network classifiers: A comparison between shallow and deep architectures. *IEEE Transactions on Neural Networks and Learning Systems*, *25*(8), 1553–1565.

Binder, M., Moosbauer, J., Thomas, J., & Bischl, B. (2020). Multi-objective hyperparameter tuning and feature selection using filter ensembles. *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, Cancún, Mexico (pp. 471–479). https://doi.org/10.1145/3377930.3389815

Bindu, K. H., Morusupalli, R., Dey, N., & Rao, C. R. (2019). *Coefficient of variation and machine learning applications*. CRC Press.

Bonet, J., & Wood, R. D. (1997). *Nonlinear continuum mechanics for finite element analysis*. Cambridge University Press.

Bornert, M., Bretheau, T., & Gilormini, P. (2001). Homogénéisation en mécanique des matériaux, Tome, 2: Comportements non linéaires et problèmes ouverts. Hermes Science. https://hal.science/hal-00112721

Bousquet, O., & Elisseeff, A. (2002). Stability and generalization. *Journal of Machine Learning Research*, *2*, 499–526.

Bouthillier, X., Laurent, C., & Vincent, P., (2019). Unreproducible research is reproducible. *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, CA. https://proceedings.mlr.press/v97/bouthillier19a.html

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

Brown, E. R., Kandhal, P. S., & Zhang, J. (2001). Performance testing for hot mix asphalt. *NCAT report*, *1*(05).

Brownlee, J. (2016). Machine learning mastery with Python: Understand your data, create accurate models, and work projects end-to-end. Machine Learning Mastery.

Brownlee, J. (2020). Data preparation for machine learning: Data cleaning, feature selection, and data transforms in Python. Machine Learning Mastery.

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., Khlaaf, H., Yang, J., Toner, H., Fong, R., Maharaj, T., Koh, P. W., Hooker, S., Leung, J., Trask, A., Bluemke, E., Lebensold, J., O'Keefe, C., Koren, M., … Anderljung, M. (2020). *Toward trustworthy AI development: Mechanisms for supporting verifiable claims*. (No. arXiv:2004.07213). arXiv. https://doi.org/10.48550/arXiv.2004.07213

Caldwell, P. M., Bretherton, C. S., Zelinka, M. D., Klein, S. A., Santer, B. D., & Sanderson, B. M. (2014). Statistical significance of climate sensitivity predictors obtained by data mining. *Geophysical Research Letters*, *41*(5), 1803–1808. https://doi.org/10.1002/2014GL059205

Cang, R., Li, H., Yao, H., Jiao, Y., & Ren, Y. (2018). Improving direct physical properties prediction of heterogeneous materials from imaging data via convolutional neural network and a morphology-aware generative model. *Computational Materials Science*, *150*, 212–221. https://doi.org/10.1016/j.commatsci.2018.03.074

Cao, P., Jin, F., Changjun, Z., & Feng, D. (2017). Investigation on statistical characteristics of asphalt concrete dynamic moduli with random aggregate distribution model. *Construction and Building Materials*, *148*, 723–733. https://doi.org/10.1016/j.conbuildmat.2017.05.012

Ceylan, H., Schwartz, C. W., Kim, S., & Gopalakrishnan, K. (2009). Accuracy of predictive models for dynamic modulus of hot-mix asphalt. *Journal of Materials in Civil Engineering*, *21*(6), 286–293. https://doi.org/10.1061/(ASCE)0899-1561(2009)21:6(286)

Charalambakis, N. (2010). Homogenization techniques and micromechanics. *A Survey and Perspectives. Applied Mechanics Reviews*, *63*(3), 030803. https://doi.org/10.1115/1.4001911

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.

Chen, K., Torbaghan, M. E., Thom, N., & Faramarzi, A. (2025). Physics-guided neural network for predicting international roughness index on flexible pavements considering accuracy, uncertainty and stability. *Engineering Applications of Artificial Intelligence*, *142*, 109922.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA (pp. 785–794).

Christensen, D. W., & Bonaquist, R. (2015). Improved Hirsch model for estimating the modulus of hot-mix asphalt. *Road Materials and Pavement Design*, *16*(sup2), 254–274. https://doi.org/10.1080/14680629.2015.1077635

Christensen, D. W. Jr., Pellinen, T., & Bonaquist, R. F. (2003). Hirsch model for estimating the modulus of asphalt concrete. *Journal of the Association of Asphalt Paving Technologists*, *72*, 97–121.

Christensen, R. M. (2005). *Mechanics of composite materials*. Courier Corporation.

Christensen, R. M., & Lo, K. H. (1979). Solutions for effective shear properties in three phase sphere and cylinder models. *Journal of the Mechanics and Physics of Solids*, *27*(4), 315–330. https://doi.org/10.1016/0022-5096(79)90032-2

Chu, X., Ilyas, I. F., Krishnan, S., & Wang, J. (2016). Data cleaning: Overview and emerging challenges. *Proceedings of the 2016 International Conference on Management of Data*, San Francisco, CA (pp. 2201–2206). https://doi.org/10.1145/2882903.2912574

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

Cundall, P. A., & Strack, O. D. L. (1979). A discrete numerical model for granular assemblies. *Géotechnique*, *29*(1), 47–65. https://doi.org/10.1680/geot.1979.29.1.47

Dasu, T., & Johnson, T. (2003). *Exploratory data mining and data cleaning*. John Wiley & Sons.

Daw, A., Karpatne, A., Watkins, W., Read, J., & Kumar, V. (2021). *Physics-guided neural networks (PGNN): An application in lake temperature modeling*. (No. arXiv:1710.11431). arXiv. https://doi.org/10.48550/arXiv.1710.11431

De Winter, J. C., Gosling, S. D., & Potter, J. (2016). Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological Methods*, *21*(3), 273.

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

Dhundup, G., Zhou, J., Bekoe, M., Sun, L., Mao, S., & Yan, Y. (2025). Integrating a mortar model into discrete element simulation for enhanced understanding of asphalt mixture cracking. *Computer-Aided Civil and Infrastructure Engineering*. Advance online publication. https://doi.org/10.1111/mice.13425

Díaz-Mendoza, J. M., Valles-Rosales, D. J., Park, Y. H., & Sabo, R. C. (2022). Micromechanical modeling for tensile properties of wood plastic composites: Use of pruned waste from pecan orchards as sustainable material for reinforcement of thermoplastic composite. *Polymers*, *14*(3), 504. https://doi.org/10.3390/polym14030504

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78–87.

Dongre, R., Myers, L., D'Angelo, J., Paugh, C., & Gudimettla, J. (2005). Field evaluation of Witczak and Hirsch models for predicting dynamic modulus of hot-mix asphalt (with discussion). *Journal of the Association of Asphalt Paving Technologists*, *74*, 381–442. https://trid.trb.org/View/776082

Doshi-Velez, F., & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning*. (No. arXiv: 1702.08608). arXiv. https://doi.org/10.48550/arXiv.1702.08608

Ejiyi, C. J., Qin, Z., Ukwuoma, C. C., Nneji, G. U., Monday, H. N., Ejiyi, M. B., Ejiyi, T. U., Okechukwu, U., & Bamisile, O. O. (2024). Comparative performance analysis of Boruta, SHAP, and Borutashap for disease diagnosis: A study with multiple machine learning algorithms. *Network: Computation in Neural Systems*. Advance online publication. https://doi.org/10.1080/0954898X.2024.2331506

Elseifi, M. A., Al-Qadi, I. L., & Yoo, P. J. (2006). Viscoelastic modeling and field validation of flexible pavements. *Journal of Engineering Mechanics*, *132*(2), 172–178. https://doi.org/10.1061/(ASCE)0733-9399(2006)132:2(172)

EN 12697-26. *2012—Bituminous mixtures—Test methods for hot mix asphalt—Part 26: Stiffness*. (Standard No. EN 12697-26; Version 2012). (2012). European Committee for Standardization.

Engl, H. W., Hanke, M., & Neubauer, A. (1996). *Regularization of inverse problems* (Vol. 375). Springer Science & Business Media.

Faghmous, J. H., & Kumar, V. (2014). A big data guide to understanding climate change: The case for theory-guided data science. *Big Data*, *2*(3), 155–163. https://doi.org/10.1089/big.2014.0026

Freitag, S., Beer, M., Graf, W., & Kaliske, M. (2009). Lifetime prediction using accelerated test data and neural networks. *Computers & Structures*, *87*(19), 1187–1194. https://doi.org/10.1016/j.compstruc.2008.12.007

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Girosi, F., Jones, M., & Poggio, T. (1995). Regularization theory and neural networks architectures. *Neural Computation*, *7*(2), 219–269.

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Sardinia, Italy (pp. 249–256). http://proceedings.mlr.press/v9/glorot10a

Gong, H., Sun, Y., Dong, Y., Hu, W., Han, B., Polaczyk, P., & Huang, B. (2022). An efficient and robust method for predicting asphalt concrete dynamic modulus. *International Journal of Pavement Engineering*, *23*(8), 2565–2576. https://doi.org/10.1080/10298436.2020.1865533

Gonzales, R. C., & Wintz, P. (1987). *Digital image processing*. Addison-Wesley Longman Publishing Co., Inc. https://dl.acm.org/doi/abs/10.5555/22881

Grossmann, T. G., Komorowska, U. J., Latz, J., & Schönlieb, C.-B. (2023). *Can physics-informed neural networks beat the finite element method?* (No. arXiv:2302.04107). arXiv. https://doi.org/10.48550/arXiv.2302.04107

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*, 1157–1182.

Hashin, Z. (1965). On elastic behaviour of fibre reinforced materials of arbitrary transverse phase geometry. *Journal of the Mechanics and Physics of Solids*, *13*(3), 119–134. https://doi.org/10.1016/0022-5096(65)90015-3

Hashin, Z. (1983). *Analysis of composite materials—A survey*. http://vucoe.drbriansullivan.com/wp-content/uploads/Hashin-Analysis-of-Composites-A-Survey-downloaded-original-1.pdf

Hashin, Z., & Shtrikman, S. (1963). A variational approach to the theory of the elastic behaviour of multiphase materials. *Journal of the Mechanics and Physics of Solids*, *11*(2), 127–140. https://doi.org/10.1016/0022-5096(63)90060-7

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Citeseer.

Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., & Meger, D. (2018). Deep reinforcement learning that matters. *Proceedings of the AAAI Conference on Artificial Intelligence*, New Orleans, LA. https://ojs.aaai.org/index.php/AAAI/article/view/11694

Hill, R. (1963). Elastic properties of reinforced solids: Some theoretical principles. *Journal of the Mechanics and Physics of Solids*, *11*(5), 357–372.

Hill, R. (1965). A self-consistent mechanics of composite materials. *Journal of the Mechanics and Physics of Solids*, *13*(4), 213–222. https://doi.org/10.1016/0022-5096(65)90010-4

Hirsch, T. J. (1962). Modulus of elasticity of concrete affected by elastic moduli of cement paste matrix and aggregate. *Proceeding of the American Concrete Institute*, *59*, 427–452.

Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *06*(02), 107–116. https://doi.org/10.1142/S0218488598000094

Huang, A. J., & Agarwal, S. (2023). On the limitations of physics-informed deep learning: Illustrations using first-order hyperbolic conservation law-based traffic flow models. *IEEE Open Journal of Intelligent Transportation Systems*, *4*, 279–293.

Huang, W., Wang, H., Yin, Y., Zhang, X., & Yuan, J. (2019). Microstructural modeling of rheological mechanical response for asphalt mixture using an image-based finite element approach. *Materials*, *12*(13), 2041.

Jiang, L., Haiminen, N., Carrieri, A.-P., Huang, S., Vázquez-Baeza, Y., Parida, L., Kim, H.-C., Swafford, A. D., Knight, R., & Natarajan, L. (2022). Utilizing stability criteria in choosing feature selection methods yields reproducible results in microbiome data. *Biometrics*, *78*(3), 1155–1167.

Jin, C., Zhang, W., Liu, P., Yang, X., & Oeser, M. (2021). Morphological simplification of asphaltic mixture components for micromechanical simulation using finite element method. *Computer-Aided Civil*

and Infrastructure Engineering, *36*(11), 1435–1452. https://doi.org/10.1111/mice.12696

Kaewnuratchadasorn, C., Wang, J., & Kim, C.-W. (2024). Physics-informed neural operator solver and super-resolution for solid mechanics. *Computer-Aided Civil and Infrastructure Engineering*, *39*(22), 3435–3451. https://doi.org/10.1111/mice.13292

Kalousis, A., Prados, J., & Hilario, M. (2006). Stability of feature selection algorithms: A study on high-dimensional spaces. *Knowledge and Information Systems*, *12*(1), Article 1. https://doi.org/10.1007/s10115-006-0040-8

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

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*(6), 422–440. https://doi.org/10.1038/s42254-021-00314-5

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., & Kumar, V. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, *29*(10), 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168

Kazemitabar, S. J., Amini, A., Bloniarz, A., & Talwalkar, A. (2017). Variable importance using decision trees. *Neural Information Processing Systems*, Long Beach, CA.

Keany, E. (2020). *BorutaShap: A wrapper feature selection method which combines the Boruta feature selection algorithm with Shapley values*. Zenodo.

Kim, T., Ha, B., & Hwangbo, S. (2023). Online machine learning approach for system marginal price forecasting using multiple economic indicators: A novel model for real-time decision making. *Machine Learning with Applications*, *14*, 100505. https://doi.org/10.1016/j.mlwa.2023.100505

Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization*. arXiv Preprint arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, *97*(1), 273–324. https://doi.org/10.1016/S0004-3702(97)00043-X

Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering*, *30*(1), 25–36.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer. https://doi.org/10.1007/978-1-4614-6849-3

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software*, *36*, 1–13.

Le Guern, M., Chailleux, E., Farcas, F., Dreessen, S., & Mabille, I. (2010). Physico-chemical analysis of five hard bitumens: Identification of chemical species and molecular organization before and after artificial aging. *Fuel*, *89*(11), 3330–3339.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Leng, Z., Tan, Z., Cao, P., & Zhang, Y. (2021). An efficient model for predicting the dynamic performance of fine aggregate matrix. *Computer-Aided Civil and Infrastructure Engineering*, *36*(11), 1467–1479. https://doi.org/10.1111/mice.12706

Lezgy-Nazargah, M., Emamian, S. A., Aghasizadeh, E., & Khani, M. (2018). Predicting the mechanical properties of ordinary concrete and nano-silica concrete using micromechanical methods. *Sādhanā*, *43*(12), 196. https://doi.org/10.1007/s12046-018-0965-0

Li, B., Zhang, Y., Wei, D., Yao, T., Hu, Y., & Dou, H. (2025). Evolution of clogging of porous asphalt concrete in the seepage process through integration of computer tomography, computational fluid dynamics, and discrete element method. *Computer-Aided Civil and Infrastructure Engineering*, *40*(12), 1652–1674. https://doi.org/10.1111/mice.13419

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2018). Feature selection: A data perspective. *ACM Computing Surveys*, *50*(6), 1–45. https://doi.org/10.1145/3136625

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, Y., & Yuan, Y. (2017). Convergence analysis of two-layer neural networks with ReLU activation. *Advances in Neural Information Processing Systems*, 30, Long Beach, CA.

Liashchynskyi, P., & Liashchynskyi, P. (2019). *Grid search, random search, genetic algorithm: A big comparison for NAS*. (No. arXiv:1912.06059). arXiv. https://doi.org/10.48550/arXiv.1912.06059

Liu, H., & Motoda, H. (2007). *Computational methods of feature selection*. CRC Press. https://books.google.com/books?hl=en&lr=&id=N1ViHNWZeQ0C&oi=fnd&pg=PP1&dq=Liu,+H.,+%26+Motoda,+H.+(2007).+Computational+methods+of+feature+selection.+CRC+Press.+ISBN:+978-1-58488-879-3&ots=pLWdBzybwO&sig=xeYgQIBWcs_V11-PcL4QUQVl0zY

Lo Piano, S. (2020). Ethical principles in machine learning and artificial intelligence: Cases from the field and possible ways forward. *Humanities and Social Sciences Communications*, *7*(1), 1–7.

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

Luo, R., & Lytton, R. L. (2011). Self-consistent micromechanics models of an asphalt mixture. *Journal of Materials in Civil Engineering*, *23*(1), 49–55. https://doi.org/10.1061/(ASCE)MT.1943-5533.0000007

Mase, G. T., Smelser, R. E., & Mase, G. E. (2009). *Continuum mechanics for engineers* (3rd ed.). CRC Press. https://doi.org/10.1201/9781420085396

Meyes, R., Lu, M., de Puiseau, C. W., & Meisen, T. (2019). *Ablation studies in artificial neural networks*. (No. arXiv:1901.08644). arXiv. https://doi.org/10.48550/arXiv.1901.08644

Milton, G. W. (1981). Concerning bounds on the transport and mechanical properties of multicomponent composite materials. *Applied Physics A Solids and Surfaces*, *26*(2), 125–130. https://doi.org/10.1007/BF00616659

Milton, G. W. (1982). Bounds on the elastic and transport properties of two-component composites. *Journal of the Mechanics and Physics of Solids*, *30*(3), 177–191. https://doi.org/10.1016/0022-5096(82)90022-9

Misheva, B. H., & Osterrieder, J. (2023). *A hypothesis on good practices for AI-based systems for financial time series forecasting: Towards*

domain-driven XAI methods. (No. arXiv:2311.07513). arXiv. https://doi.org/10.48550/arXiv.2311.07513

Mohan, A. T., & Gaitonde, D. V. (2018). *A deep learning based approach to reduced order modeling for turbulent flow control using LSTM neural networks*. (No. arXiv:1804.09269). arXiv. https://doi.org/10.48550/arXiv.1804.09269

Mori, T., & Tanaka, K. (1973). Average stress in matrix and average elastic energy of materials with misfitting inclusions. *Acta Metallurgica*, *21*(5), 571–574. https://doi.org/10.1016/0001-6160(73)90064-3

Moussa, G. S., & Owais, M. (2021). Modeling Hot-Mix asphalt dynamic modulus using deep residual neural networks: Parametric and sensitivity analysis study. *Construction and Building Materials*, *294*, 123589. https://doi.org/10.1016/j.conbuildmat.2021.123589

Mura, T. (2013). *Micromechanics of defects in solids*. Springer Science & Business Media.

Nemat-Nasser, S., & Hori, M. (2013). *Micromechanics: Overall properties of heterogeneous materials*. Elsevier.

NEN-EN 1426. (2015). *2015—Bitumen en bitumineuze bindmiddelen—Bepaling van de naaldpenetratie*. (Standard No. NEN-EN 1426; Version 2015). Nederlands Normalisatie-instituut (NEN).

NEN-EN 1427. (2015). *2015—Bitumen en bitumineuze bindmiddelen—Bepaling van het verzachtingspunt—Ring- en kogelmethode*. (Standard No. NEN-EN 1427; Version 2015). Nederlands Normalisatie-instituut (NEN). https://www.nen.nl/nen-en-1427-2015-en-208860

NEN-EN 14770. (2022). *2022—Bitumen and bitumineuze bindmiddelen—Bepaling van complexe afschuifmodulus en fasehoek—Dynamische Afschuif Rheometer (DSR)*. (Standard No. NEN-EN 14770; Version 2022). Nederlands Normalisatie-instituut (NEN).

Niu, P., Guo, J., Chen, Y., Zhou, Y., Feng, M., & Shi, Y. (2025). Improved physics-informed neural network in mitigating gradient-related failures. *Neurocomputing*, *638*, 130167. https://doi.org/10.1016/j.neucom.2025.130167

Nogueira, S., Sechidis, K., & Brown, G. (2018). On the stability of feature selection algorithms. *Journal of Machine Learning Research*, *18*(174), 1–54.

Nordin, N., Zainol, Z., Mohd Noor, M. H., & Chan, L. F. (2023). An explainable predictive model for suicide attempt risk using an ensemble learning and Shapley Additive Explanations (SHAP) approach. *Asian Journal of Psychiatry*, *79*, 103316. https://doi.org/10.1016/j.ajp.2022.103316

Nouraie, M., & Muller, S. (2025). *On the selection stability of stability selection and its applications*. (No. arXiv:2411.09097). arXiv. https://doi.org/10.48550/arXiv.2411.09097

O'Gorman, P. A., & Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, *10*(10), 2548–2563. https://doi.org/10.1029/2018MS001351

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

Pandey, A. C., Rajpoot, D. S., & Saraswat, M. (2019). Feature selection method based on hybrid data transformation and binary binomial cuckoo search. *Journal of Ambient Intelligence and Humanized Computing*, *11*(2), Article 2. https://doi.org/10.1007/s12652-019-01330-1

Paul, B. (1960). Prediction of elastic constants of multiphase materials. *Transactions of AIME*, *218*, 36–41.

Pellinen, T., Zofka, A., Marasteanu, M., & Funk, N. (2007). Asphalt mixture stiffness predictive models (with discussion). *Journal of the Association of Asphalt Paving Technologists*, *76*, 575–625. https://trid.trb.org/View/839508

Plati, C., Georgiou, P., & Loizos, A. (2016). Influence of different roller compaction modes on asphalt mix performance. *International Journal of Pavement Engineering*, *17*(1), 64–70. https://doi.org/10.1080/10298436.2014.925552

Poernomo, A., & Kang, D.-K. (2018). Biased Dropout and Crossmap Dropout: Learning towards effective dropout regularization in convolutional neural network. *Neural Networks*, *104*, 60–67.

Poslavskaya, E., & Korolev, A. (2023). *Encoding categorical data: Is there yet anything "hotter" than one-hot encoding?* https://doi.org/10.48550/ARXIV.2312.16930

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, Montréal, Canada.

Provost, F., & Fawcett, T. (2013). *Data Science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media, Inc.

Qiao, Q., Yunusa-Kaltungo, A., & Edwards, R. E. (2023). Developing a machine learning based building energy consumption prediction approach using limited data: Boruta feature selection and empirical mode decomposition. *Energy Reports*, *9*, 3643–3660. https://doi.org/10.1016/j.egyr.2023.02.046

Rafiei, M. H., & Adeli, H. (2017). A novel machine learning-based algorithm to detect damage in high-rise building structures. *Structural Design of Tall and Special Buildings*, *26*(18), e1400. https://doi.org/10.1002/tal.1400

Rafiei, M. H., Gauthier, L. V., Adeli, H., & Takabi, D. (2024). Self-supervised learning for near-wild cognitive workload estimation. *Journal of Medical Systems*, *48*(1), 107. https://doi.org/10.1007/s10916-024-02122-7

Rafiei, M. H., Khushefati, W. H., Demirboga, R., & Adeli, H. (2017). Novel approach for concrete mixture design using neural dynamics model and virtual lab concept. *ACI Materials Journal*, *114*(1), 117–127.

Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Engineering Bulletin*, *23*(4), 3–13.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707.

Reimers, N., & Gurevych, I. (2017). *Reporting score distributions makes a difference: Performance study of LSTM-networks for sequence tagging*. (No. arXiv:1707.09861). arXiv. https://doi.org/10.48550/arXiv.1707.09861

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA (pp. 1135–1144).

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

Salih, A. M., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S. E., Lekadir, K., & Menegaz, G. (2025). A perspective on explainable artificial intelligence methods: SHAP and LIME. *Advanced Intelligent Systems*, *7*(1), 2400304. https://doi.org/10.1002/aisy.202400304

Salkind, N. J. (2006). *Encyclopedia of measurement and statistics*. SAGE Publications.

Santos, J. M., & Embrechts, M. (2009). On the use of the adjusted rand index as a metric for evaluating supervised classification. In C. Alippi, M. Polycarpou, C. Panayiotou, & G. Ellinas (Eds.), *Lecture notes in computer science, Vol. 5769. Artificial neural networks—ICANN 2009. ICANN 2009* (pp. 175–184). Springer. https://doi.org/10.1007/978-3-642-04277-5_18

Koli, S., Bhat, K., Korade, P., D. M. A. M. O. K. (2024). Unlocking machine learning model decisions: a comparative analysis of LIME and SHAP for enhanced interpretability. *Journal of Electrical Systems*, *20*(2s), 598–613. https://doi.org/10.52783/jes.1480

Sathvik, S., Kumar, G. S., Bahrami, A., Nitin, G. C., Singh, S. K., Althaqafi, E., & Özkılıç, Y. O. (2024). Evaluation of asphalt binder and mixture properties utilizing fish scale powder as a biomodifier. *Case Studies in Construction Materials*, *20*, e03238. https://doi.org/10.1016/j.cscm.2024.e03238

Schleder, G. R., Padilha, A. C. M., Acosta, C. M., Costa, M., & Fazzio, A. (2019). From DFT to machine learning: Recent approaches to materials science—A review. *Journal of Physics: Materials*, *2*(3), 032001. https://doi.org/10.1088/2515-7639/ab084b

Sebastián, C., & González-Guillén, C. E. (2024). A feature selection method based on Shapley values robust for concept shift in regression. *Neural Computing and Applications*, *36*(23), 14575–14597. https://doi.org/10.1007/s00521-024-09745-4

Seber, G. A. F., & Lee, A. J. (2003). *Linear regression analysis*. John Wiley & Sons.

Semmelrock, H., Kopeinik, S., Theiler, D., Ross-Hellauer, T., & Kowald, D. (2023). *Reproducibility in machine learning-driven research*. (No. arXiv:2307.10320). arXiv. https://doi.org/10.48550/arXiv.2307.10320

Sharma, S., Sharma, S., & Athaiya, A. (2017). Activation functions in neural networks. *Towards Data Science*, *6*(12), 310–316.

Shen, S., Yu, H., Willoughby, K. A., DeVol, J. R., & Uhlmeyer, J. S. (2013). Local practice of assessing dynamic modulus properties for Washington State mixtures. *Transportation Research Record: Journal of the Transportation Research Board*, *2373*(1), 89–99. https://doi.org/10.3141/2373-10

Shu, X., & Huang, B. (2008). Dynamic modulus prediction of HMA mixtures based on the viscoelastic micromechanical model. *Journal of Materials in Civil Engineering*, *20*(8), 530–538. https://doi.org/10.1061/(ASCE)0899-1561(2008)20:8(530)

Smith, L. N. (2017). Cyclical learning rates for training neural networks. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Santa Rosa, CA (pp. 464–472). https://ieeexplore.ieee.org/abstract/document/7926641/

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 25, Lake Tahoe, NV. https://proceedings.neurips.cc/paper/2012/hash/05311655a15b75fab86956663e1819cd-Abstract.html

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(1), 1929–1958.

Sun, C., Sharma, J., & Maiti, M. (2021). *Investigating the relationship between dropout regularization and model complexity in neural networks*. (No. arXiv:2108.06628). arXiv. https://doi.org/10.48550/arXiv.2108.06628

Taghizadeh, M., Zandsalimi, Z., Nabian, M. A., Shafiee-Jood, M., & Alemazkoor, N. (2025). Interpretable physics-informed graph neural networks for flood forecasting. *Computer-Aided Civil and Infrastructure Engineering*. Advance online publication. https://doi.org/10.1111/mice.13484

Tanaka, A., Tomiya, A., & Hashimoto, K. (2021). *Deep learning and physics*. Springer. https://doi.org/10.1007/978-981-33-6108-9

Tang, T., Anupam, K., Kasbergen, C., Scarpas, A., & Erkens, S. (2019). A finite element study of rain intensity on skid resistance for permeable asphalt concrete mixes. *Construction and Building Materials*, *220*, 464–475. https://doi.org/10.1016/j.conbuildmat.2019.05.185

Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, *31*(2), 447–464. https://doi.org/10.1007/s12525-020-00441-4

Torquato, S. (2002). *Random heterogeneous materials* (Vol. 16). Springer. https://doi.org/10.1007/978-1-4757-6355-3

Tran, N., Turner, P., & Shambley, J., (2016). *Enhanced compaction to improve durability and extend pavement service life: A literature review*. NCAT Report 16-02, NCAT, Auburn University.

Underwood, B. S., Baek, C., & Kim, Y. R. (2012). Simplified viscoelastic continuum damage model as platform for asphalt concrete fatigue analysis. *Transportation Research Record: Journal of the Transportation Research Board*, *2296*(1), 36–45. https://doi.org/10.3141/2296-04

Underwood, B. S., & Kim, Y. R. (2013). Effect of volumetric factors on the mechanical behavior of asphalt fine aggregate matrix and the relationship to asphalt mixture properties. *Construction and Building Materials*, *49*, 672–681. https://doi.org/10.1016/j.conbuildmat.2013.08.045

Urbán, J. F., Stefanou, P., & Pons, J. A. (2025). Unveiling the optimization process of physics informed neural networks: How accurate and competitive can PINNs be? *Journal of Computational Physics*, *523*, 113656. https://doi.org/10.1016/j.jcp.2024.113656

Walpole, L. J. (1966). On bounds for the overall elastic moduli of inhomogeneous systems—I. *Journal of the Mechanics and Physics of Solids*, *14*(3), 151–162. https://doi.org/10.1016/0022-5096(66)90035-4

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*, *39*(17), 2701–2718. https://doi.org/10.1111/mice.13208

Wang, H., Zhang, H., Liu, X., Skarpas, A., Erkens, S., & Leng, Z. (2021). Micromechanics-based complex modulus prediction of crumb rubber modified bitumen considering interparticle interactions. *Road Materials and Pavement Design*, *22*(sup1), S251–S268. https://doi.org/10.1080/14680629.2021.1899965

Willard, J. D., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2022). Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys*, *55*(4), 66. https://doi.org/10.1145/3514228

Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2020). *Integrating physics-based modeling with machine learning: A survey*. arXiv Preprint arXiv:2003.04919, 1(1). https://arxiv.org/pdf/2112.12979

Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2022). Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys*, *55*(4), 1–37.

Wiśniewska, A., Hernik, S., Liber-Kneć, A., & Egner, H. (2019). Effective properties of composite material based on total strain energy equivalence. *Composites Part B: Engineering*, *166*, 213–220. https://doi.org/10.1016/j.compositesb.2018.11.094

Xia, Y., Liu, C., Li, Y., & Liu, N. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications*, *78*, 225–241. https://doi.org/10.1016/j.eswa.2017.02.017

Xiao, D., Heaney, C. E., Mottet, L., Fang, F., Lin, W., Navon, I. M., Guo, Y., Matar, O. K., Robins, A. G., & Pain, C. C. (2019). A reduced order model for turbulent flows in the urban environment using machine learning. *Building and Environment*, *148*, 323–337. https://doi.org/10.1016/j.buildenv.2018.10.035

Xiong, Z., Yang, P., & Zhao, P. (2024). Physics-driven neural networks for nonlinear micromechanics. *International Journal of Mechanical Sciences*, *273*, 109214. https://doi.org/10.1016/j.ijmecsci.2024.109214

Xu, T., & Valocchi, A. J. (2015). Data-driven methods to improve baseflow prediction of a regional groundwater model. *Computers & Geosciences*, *85*, 124–136. https://doi.org/10.1016/j.cageo.2015.05.016

Yan, X., Zhou, Z., Liu, Z., & Zhou, Y. (2024). Investigation of dynamic viscoelastic characteristics of permeable asphalt. *Materials*, *17*(12), 2984. https://doi.org/10.3390/ma17122984

Yang, C. H., Huh, H., & Hahn, H. T. (2005). Investigation of effective material properties in composites with internal defect or reinforcement particles. *International Journal of Solids and Structures*, *42*(24), 6141–6165. https://doi.org/10.1016/j.ijsolstr.2005.06.094

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

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

Yeh, C.-C., & Yang, M.-S. (2016). A Generalization of Rand and Jaccard indices with its fuzzy extension. *International Journal of Fuzzy Systems*, *18*(6), 1008–1018. https://doi.org/10.1007/s40815-016-0263-0

Yin, H. M., Buttlar, W. G., Paulino, G. H., & Benedetto, H. D. (2008). Assessment of existing micro-mechanical models for asphalt mastics considering viscoelastic effects. *Road Materials and Pavement Design*, *9*(1), 31–57. https://doi.org/10.1080/14680629.2008.9690106

Yu, G., & Adeli, H. (1993). Object-oriented finite element analysis using EER model. *Journal of Structural Engineering*, *119*(9), 2763–2781. https://doi.org/10.1061/(ASCE)0733-9445(1993)119:9(2763)

Zafar, M. R., & Khan, N. (2021). Deterministic local interpretable model-agnostic explanations for stable explainability. *Machine Learning and Knowledge Extraction*, *3*(3), 525–541. https://doi.org/10.3390/make3030027

Zeng, H., Xu, W., Zhang, M., & Yang, p. (2020). Calibration of DEM-FEM model parameters for traction performance analysis of an off-road tire on gravel terrain. *Powder Technology*, *362*, 350–361.

Zhang, H. (2022). *Micromechanical modelling of porous asphalt mixtures* [Doctoral dissertation, TU Delft]. TU Delft Repository. https://repository.tudelft.nl/record/uuid:35a16fd8-6cde-49c6-9455-042aa4ac4d0a

Zhang, H., Anupam, K., Scarpas, A., & Kasbergen, C. (2018a). Comparison of different micromechanical models for predicting the effective properties of open graded mixes. *Transportation Research Record*, *2672*(28), 404–415. https://doi.org/10.1177/0361198118794713

Zhang, H., Anupam, K., Scarpas, A., & Kasbergen, C. (2018b). Issues in the prediction of the mechanical properties of open graded mixes. *Transportation Research Record*, *2672*(40), 32–40. https://doi.org/10.1177/0361198118792117

Zhang, H., Anupam, K., Scarpas, T., Kasbergen, C., & Erkens, S. (2021). Contact mechanics based solution to predict modulus of asphalt materials with high porosities. *Materials & Design*, *206*, 109752. https://doi.org/10.1016/j.matdes.2021.109752

Zhang, H., Anupam, K., Scarpas, T., Kasbergen, C., Erkens, S., & Al Khateeb, L. (2020). Continuum-based micromechanical models for asphalt materials: Current practices & beyond. *Construction and Building Materials*, *260*, 119675. https://doi.org/10.1016/j.conbuildmat.2020.119675

Zhang, J. (2020). Advances in micromechanical modelling of asphalt mixtures: A review. *European Journal of Environmental and Civil Engineering*, *24*(5), 583–602. https://doi.org/10.1080/19648189.2017.1410727

Zheng, A., & Casari, A. (2018). *Feature engineering for machine learning: Principles and techniques for data scientists*. O'Reilly Media, Inc.

Zohdi, T. I., & Wriggers, P. (2008). *An introduction to computational micromechanics*. Springer Science & Business Media.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

How to cite this article: Anupam, K., Berangi, M., Camargo, J. C., Kasbergen, C., & Erkens, S. (2025). A step toward a micromechanics-informed neural network for predicting asphalt mixture stiffness. *Computer-Aided Civil and Infrastructure Engineering*, 1–28. https://doi.org/10.1111/mice.70000