

Combining deep neural networks and Gaussian processes for asphalt rheological insights

Khadijeh, Mahmoud; Kasbergen, Cor; Erkens, Sandra; Varveri, Aikaterini

DOI

[10.1016/j.rineng.2025.105629](https://doi.org/10.1016/j.rineng.2025.105629)

Publication date

2025

Document Version

Final published version

Published in

Results in Engineering

Citation (APA)

Khadijeh, M., Kasbergen, C., Erkens, S., & Varveri, A. (2025). Combining deep neural networks and Gaussian processes for asphalt rheological insights. *Results in Engineering*, 26, Article 105629. <https://doi.org/10.1016/j.rineng.2025.105629>

Important note

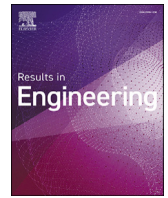
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Research paper

Combining deep neural networks and Gaussian processes for asphalt rheological insights

Mahmoud Khadijeh^{*}, Cor Kasbergen, Sandra Erkens, Aikaterini Varveri

Department of Engineering Structures, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, Delft, 2628 CN, Netherlands

ARTICLE INFO

Keywords:

Multiscale modeling
Asphalt binder
Asphalt mastic
Machine learning
Neural networks
Gaussian process

ABSTRACT

Asphalt binders are critical for asphalt pavement performance, and understanding their rheological behavior is essential for designing durable roadways. The complex shear modulus (G^*) and phase angle (δ) are primary parameters characterizing binder rheology. This study introduces a novel hybrid machine learning model combining deep neural networks (DNN) and Gaussian process regression (GPR) to predict G^* and δ for bituminous binders and binder-filler systems (mastics). DNN excel at capturing complex, nonlinear relationships among eleven binder and thirteen mastic input parameters, including aging conditions, chemical and physical properties, and test parameters. However, standalone DNN struggle with small datasets, common at the binder scale, and lack inherent uncertainty quantification, limiting reliability in engineering applications. GPR improves DNN by refining predictions through probabilistic modeling, while providing uncertainty estimates, and enhancing accuracy with limited or noisy data. The hybrid model leverages DNN's feature extraction capabilities and GPR's ability to smooth predictions, significantly improving performance over standalone DNN. The hybrid model achieves high prediction accuracy, with R^2 values of 0.997 for G^* and 0.947 for δ for binders, and 0.993 for G^* and 0.972 for δ for mastics, reducing G^* prediction error from 22.7% to 0.031% for fresh asphalt binder compared to standalone DNN. Feature importance analysis using random forest and SHAP techniques identifies test temperature, aging conditions, and penetration as key influencers of G^* and δ . This hybrid approach enhances the characterization of complex asphalt materials, offering pavement engineers a robust, reliable tool for predicting material behavior under diverse conditions.

1. Introduction

1.1. Fundamental aspects of asphalt binder: from composition to performance

Asphalt binder, also known as bitumen, plays a crucial role in pavement construction. It acts as a viscoelastic adhesive, binding the aggregate particles together. The unique characteristics of asphalt binder allow it to behave both as a viscous fluid and an elastic solid, which is essential to efficiently distribute stresses across the pavement surface. This capability helps prevent permanent deformations, such as rutting, ensuring the durability of the pavement [1]. Additionally, the strong ad-

hesion properties of asphalt binders create a resilient bond between the binder and the aggregate, preventing issues such as raveling [2,3].

The viscoelastic properties of asphalt binder are typically characterized by two key parameters: the complex shear modulus (G^*) and phase angle (δ). These parameters are measured using the dynamic shear rheometry test (DSR), which applies oscillatory shear stresses to a thin film of asphalt binder sandwiched between two parallel plates [4].

G^* represents the total resistance of the binder to deformation when repeatedly sheared, while δ indicates the lag between the applied shear stress and the resulting shear strain, reflecting the relative amounts of recoverable and non-recoverable deformation [5]. A lower δ indi-

Abbreviations: ML, Machine Learning; ANN, Artificial Neural Network(s); DNN, Deep Neural Network(s); GPR, Gaussian Process Regression; FEM, Finite Element Model; DSR, Dynamic Shear Rheometer; SARA, Saturates, Aromatics, Resins, and Asphaltenes; RTFOT, Rolling Thin Film Oven Test; PAV, Pressure Aging Vessel; SHAP, SHapley Additive exPlanations; ANOVA, Analysis of Variance; MSE, Mean Squared Error; ReLU, Rectified Linear Unit; Adam, Adaptive Moment Estimation; SGP, Sparse Gaussian Process; RBF, Radial Basis Function; TLC-FID, Thin-Layer Chromatography with Flame Ionization Detection; SHRP, Strategic Highway Research Program; GSCM, Generalized Self-Consistent Model.

^{*} Corresponding author.

E-mail address: m.khadijeh@tudelft.nl (M. Khadijeh).

<https://doi.org/10.1016/j.rineng.2025.105629>

Received 1 April 2025; Received in revised form 26 May 2025; Accepted 3 June 2025

Available online 6 June 2025

2590-1230/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

cates more elastic (solid-like) behavior, while a higher δ suggests more viscous (liquid-like) behavior. These measurements are necessary for predicting pavement performance under various loading and environmental conditions.

The rheological properties of asphalt binders are not static and can change over time due to various environmental factors. One of the most significant processes affecting these properties is oxidative aging, which occurs when asphalt binders are exposed to environmental conditions. Oxidative aging can lead to cracking and other forms of distress that can change the behavior of the pavement. Oxidative aging typically occurs in two phases: short-term aging, which occurs during the mixing and construction process, and long-term aging, which takes place over the pavement's service life [6]. To simulate these aging processes in laboratory settings, researchers employ two standardized tests. The rolling thin film oven test (RTFOT) replicates short-term aging by exposing the binder to heat and air flow, mimicking the conditions during mixing and paving [7]. For long-term aging simulation, the pressure aging vessel (PAV) test subjects the RTFOT-aged binder to high temperature and pressure for an extended period, representing years of in-service aging. These tests allow engineers to evaluate how the binder's properties evolve over time and predict long-term pavement performance.

From a chemical point of view, asphalt binder is characterized based on polarity into saturates, aromatics, resins, and asphaltenes (often referred to as SARA fractions). The specific distribution and relative proportions of these fractions within the binder play a crucial role in determining the overall behavior and performance characteristics of the binder. For instance, it has been indicated that asphaltenes enhance the stiffness and elasticity of the asphalt binder, especially at high temperatures, and increase the high performance grade temperature of the binder by increasing the polar fraction content [8]. In addition, the impact of the asphaltene content on the stiffness and viscoelastic properties of asphalt binders has been analyzed, showing significant changes with varying asphaltene levels [9].

1.2. Asphalt mastic: the interplay between binder and mineral

The complexity of the asphalt binder increases further when fillers are introduced to form the mastic, which represents the next level of composite material in asphalt mixtures. Mastic, a combination of asphalt binder and fine mineral particles, exhibits unique rheological and mechanical properties that differ from those of pure binder. The interaction between the binder and filler particles creates a complex system that requires detailed investigation. It has been revealed that asphalt binder aging is influenced by more than just the type of binder. Fillers can slow the aging process of asphalt binder through two main mechanisms: the physical presence of fillers blocks oxygen diffusion, and the chemically active fillers can adsorb polar functional groups from binders, further mitigating the aging effects [10–12]. Additionally, a framework study was developed to examine the interactions between different filler materials and asphalt binder within asphalt mastic to investigate how these interactions affect the stiffness of the mastic and the behavior under shear stress [13]. Furthermore, it was highlighted how different filler types and the rheological properties of asphalt mastics influence the bonding between asphalt and aggregate. The findings suggest that filler characteristics, such as pore volume and specific surface area, significantly affect the rheological behavior of asphalt mastics [14].

1.3. Machine learning applications in asphalt research

As the complexity of asphalt binder and mastic behavior becomes increasingly apparent, researchers are turning to advanced computational methods to improve their understanding and predictive capabilities. Machine learning (ML), a subset of artificial intelligence, has emerged as a powerful tool in this field by offering new ways to analyze data, identify patterns, and make predictions that were previously challenging with traditional methods [15,16]. ML algorithms process large amounts of

data from various sources, including rheological measurements, composition analysis, and performance tests. They are capable of identifying complex relationships between variables and predicting outcomes with high accuracy. This approach is particularly valuable in asphalt research, where multiple factors interact in complex ways to influence material behavior [17].

Although most of the research has concentrated on mixture scales due to data availability, some studies have explored the use of artificial neural networks (ANNs), a widely recognized ML approach, at smaller scales. Building on this foundation, deep neural networks (DNN), a more advanced form of ANNs characterized by multiple hidden layers, have emerged as powerful tools for modeling complex patterns. The depth of these networks allows them to learn hierarchical representations of data, making them particularly effective in scenarios where subtle, multi-scale interactions are present [18]. In small-scale modeling, DNN can capture fine-grained spatial variability and temporal dynamics that simpler models might overlook. DNN excel in capturing highly nonlinear and multifaceted interactions among a large number of input parameters, such as chemical compositions (SARA fractions), physical properties (penetration, softening point), aging conditions, and test parameters. The deep architecture of DNN enables superior feature extraction, automatically learning hierarchical representations of data without extensive manual feature engineering.

For example, one study developed an ANN model that predicts asphalt binder recovery and nonrecoverable compliance at high stress levels [19]. This model utilized five inputs: test temperature, frequency, storage modulus, loss modulus, and viscosity. The highly accurate predictions indicate that these models could help design asphalt mixtures that perform well under various conditions. Another study created an ANN model to predict the performance of a geopolymer-modified asphalt binder [20]. This research tested various combinations of parameter, algorithm, and network architectures, identifying the best performing model through statistical analysis. The results demonstrated both the potential of ANN to predict complex material behaviors and the importance of selecting the appropriate model architecture and training algorithms for accuracy. Beyond the binder scale, research has shown that ANN can predict the complex modulus (G^*) of asphalt mastic samples using inputs, such as temperature, frequency and the filler-asphalt ratio [21]. This investigation confirmed that ANN models can effectively evaluate the properties of asphalt mastic.

However, the majority of ML models have been developed for larger scales, such as asphalt mixtures and pavements. This trend is primarily driven by the increased availability of comprehensive datasets on these scales, which are often collected through routine quality control processes and long-term pavement monitoring programs. Some studies have developed ANN models to predict the effects of aging on the properties of asphalt mixture, demonstrating the effectiveness of neural networks in identifying complex nonlinear relationships between the mixture aging index (AMI) and various input factors [22]. Other research efforts have used ANN with large datasets, utilizing mean square logarithmic error as the loss function to predict the fatigue life of asphalt concrete (AC), focusing on factors such as strain level, binder content, and air void content [23].

Gaussian process regression (GPR) has emerged as another powerful method in predictive modeling. As a probabilistic approach to regression, GPR models the relationship between inputs and outputs using a Gaussian process, which provides a measure of uncertainty for predictions. This flexibility and ability to quantify uncertainty make GPR highly effective in various fields [24–26]. For large datasets, such as those encountered at the mastic scale, Sparse Gaussian Process (SGP) models offer a computationally efficient alternative to full GPR by approximating the posterior distribution with a reduced set of inducing points, maintaining predictive accuracy while significantly reducing computational complexity [27]. Applications in pavement engineering include automatic estimation of the resilient modulus of different pavement materials, demonstrating the effectiveness of GPR in predicting

Table 1
Overview of key studies on machine learning applications in asphalt research.

Reference	Study Summary
(Seitllari et al., 2019) [22]	Research Object: Asphalt mixture properties Research Objective: Predict aging effects using artificial neural networks (ANNs) Key Parameters: Asphaltene content as a primary indicator
(Houlfik et al., 2024) [23]	Research Object: Asphalt concrete Research Objective: Predict fatigue life using ANN models Key Parameters: Strain level, binder content, air-void content
(Hamid et al., 2022) [19]	Research Object: Asphalt binder Research Objective: Predict recovery and nonrecoverable compliance under high stress using ANN Key Parameters: Test temperature, test frequency, storage modulus, loss modulus, viscosity
(Alas et al., 2019) [20]	Research Object: Geopolymer-modified asphalt binder Research Objective: Develop accurate ANN models for performance prediction Key Parameters: Various parameter combinations, algorithms, and network architectures
(Yan et al., 2014) [21]	Research Object: Asphalt mastic Research Objective: Predict G^* using ANN models Key Parameters: Temperature, frequency, filler-asphalt ratio
(Ghanizadeh et al., 2021) [28]	Research Object: Pavement materials Research Objective: Pavement materials (stabilized base materials) Key Parameters: Wetting-drying cycles, free lime to silica ratio, alumina and iron oxide compounds in additives, maximum dry density to optimum moisture content ratio, deviator stress, confining stress

pavement properties [28]. Table 1 summarizes recent studies on ML applications in asphalt research.

1.4. Objectives of this study

This study proposes an advanced hybrid approach that combines DNN with GPR to predict G^* and δ of asphalt binders and mastics. The integration of DNN and GPR draws on the strengths of both methods to address their individual shortcomings. DNN are highly effective at identifying complex, nonlinear patterns within high dimensional data, making them ideal for modeling the detailed relationships between input parameters such as aging conditions, chemical composition, test parameters, and rheological outcomes. However, DNN can struggle with small datasets, common at the binder scale, and they do not naturally provide uncertainty estimates, which are important for dependable engineering predictions. In contrast, GPR offers a probabilistic framework that excels at quantifying uncertainty and refining predictions, particularly with limited or noisy data. Yet, GPR can become computationally impractical for large datasets, such as those at the mastic scale, and may not easily handle strongly nonlinear relationships without extensive tuning. This hybrid model utilizes the capacity of DNN to process raw inputs and extract key features, which GPR refines to improve accuracy and provide uncertainty estimates.

The binder and mastic models are hierarchically interconnected. The binder model (11 inputs) predicts intrinsic binder properties, while the mastic model (13 inputs, including filler percentage and stiffness) builds on these, using binder data augmented by FEM to generate synthetic mastic data. Both share the DNN-GPR methodology, enabling a multi-scale analysis of binder to mastic behavior.

The study also applies model-independent techniques and model-dependent methods to identify critical variables that influence asphalt behavior, shedding light on the model decision making process. The rest of this paper is structured as follows: Section 2 focuses on the detailed methodology, including data collection, the development of hybrid model, and the integration of finite element modeling to simulate mastic properties. Section 3 presents the results and discussion, evaluating the predictive performance of the hybrid model for the rheological properties of binder and mastic, along with a comparative analysis of feature importance and uncertainty contributions.

2. Methodology

2.1. Data mining

The selection of input parameters was guided by their expected impact on the rheological properties of asphalt binder and asphalt mastic, while also taking into account data availability. For instance, the viscosity of asphalt binder is known to significantly influence the rheological behavior and the overall performance of pavements. Binders with lower viscosity, such as modified binders, tend to enhance rutting resistance due to their improved flow characteristics and ability to distribute loads. Conversely, highly viscous binders, like those found in reclaimed asphalt pavement (RAP), may negatively affect fatigue performance, potentially leading to premature cracking [29,30]. However, despite the recognized importance of viscosity in asphalt binder behavior, it was not included as an input parameter in this study. This decision was primarily due to insufficient availability of data across the range of samples and conditions required for a comprehensive analysis. The lack of consistent viscosity measurements across the dataset would have limited the reliability and applicability of any conclusions drawn from the inclusion of the viscosity. As a result, the focus was placed on other influential parameters for which more complete and reliable data were available.

In light of these considerations, input parameters representing various aspects of asphalt binder and mastic behavior were considered, including physical properties and test parameters, chemical compositions, aging conditions, and filler properties:

- **Aging conditions:** These were represented as categorical variables corresponding to distinct aging states that represent fresh (unaged), short-term (simulated by RTFOT) and long-term (simulated by PAV) rather than continuous values of aging time, pressure, or temperature. This categorical approach reflects the standardized nature of aging tests that ensures the model captures the rheological responses for each state, which are critical for understanding binder property evolution under different environmental conditions.
- **Physical properties:** Softening point and penetration were included as key indicators of binder consistency. The softening point indicates the temperature at which the binder begins to soften, while penetration measures the binder's hardness at a standard temperature.

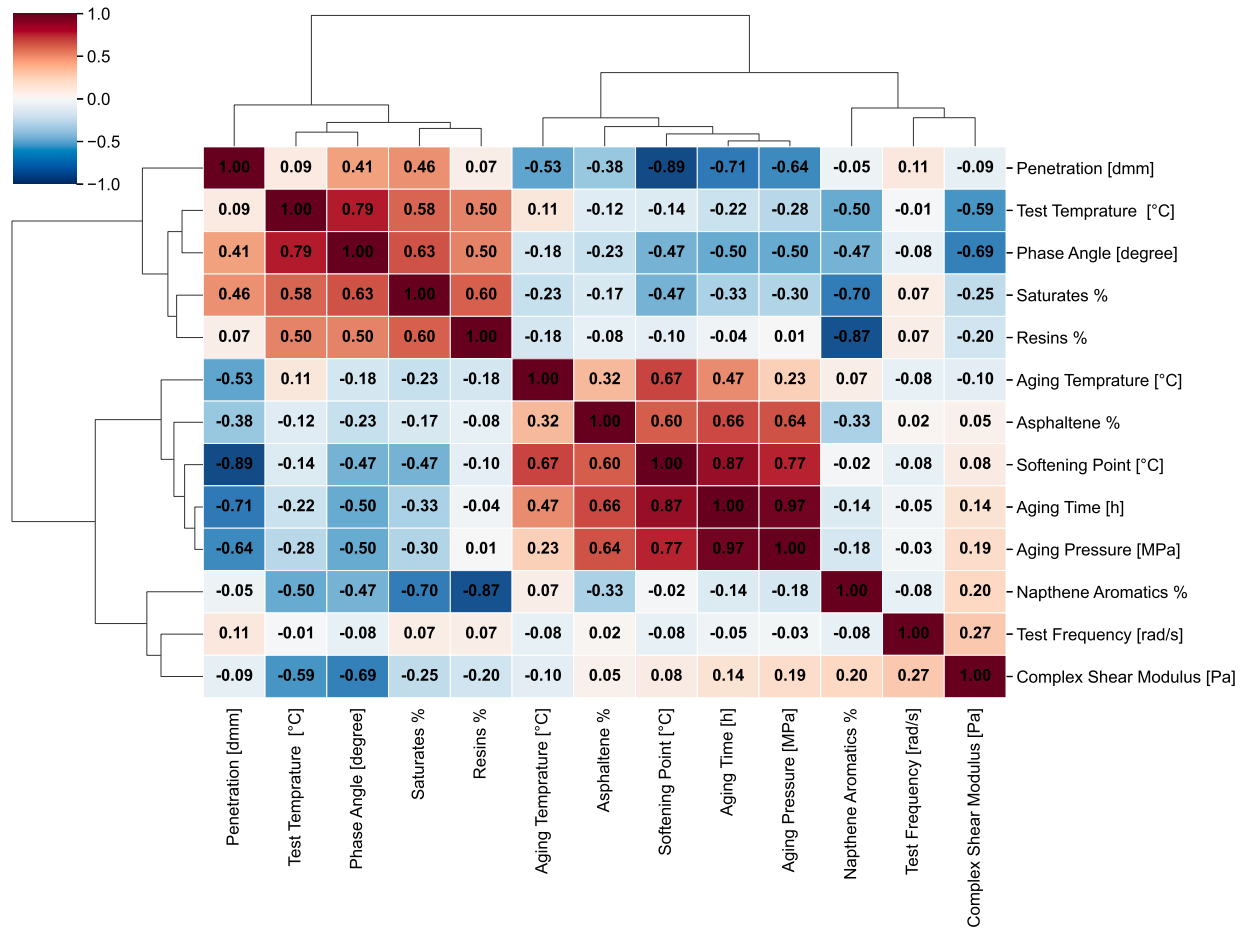


Fig. 1. Hierarchically clustered correlation heatmap for asphalt binder.

- DSR test parameters: Raw experimental measurements of G^* and δ were obtained for each combination of test temperature and frequency using the DSR test. These measurements reflect the direct rheological response of the binder at each temperature-frequency combination, without the use of master curve derivations or predictive modeling.
- Chemical composition: This was represented by the volume percentages of asphaltenes, saturates, resins, and aromatics (SARA fractions). These components significantly influence the rheological properties and aging characteristics of asphalt binders.
- For asphalt mastic, similar inputs were considered with the addition of filler percentage and filler stiffness in the binder matrix. These parameters account for the influence of mineral fillers on the overall performance of the mastic, including its stiffening effect and impact on rheological properties.

The analysis yielded 11 input variables for the binder scale and 13 for the mastic scale. This approach provided a thorough understanding of these levels and showed how fillers influence and interact with the binder's characteristics. To further investigate the relationships between the selected input parameters, a correlation analysis was performed, visualized through a heatmap (Fig. 1).

This analysis revealed interrelationships among the chosen parameters, providing valuable insights into the complex nature of asphalt binder and mastic behavior. The hierarchical clustering approach reorganized the correlation matrix to group variables with similar correlation patterns together, creating distinct clusters that reveal the underlying structure of relationships within the dataset. This methodology transforms the traditional correlation matrix into a more interpretable format where related variables appear adjacent to each other, forming

coherent blocks that represent natural groupings of aging parameters, chemical composition variables, and testing conditions. The following key observations from the correlation heatmap were observed:

- 1- Strong negative correlations were observed between penetration and aging conditions (-0.53, -0.71, -0.64 for aging temperature, time, and pressure, respectively). These relationships underscore the interconnected nature of the physical properties and environmental conditions in asphalt materials.
- 2- The SARA fractions (saturates, aromatics, resins, and asphaltenes) showed varying degrees of correlation with physical properties, emphasizing the importance of chemical composition in determining overall binder behavior.
- 3- The DSR test parameters (temperature and frequency) demonstrated moderate correlations with several other variables, indicating their broad influence on measured rheological properties.

The presence of strong correlations between certain input parameters indicates a degree of multicollinearity in the dataset. In traditional statistical modeling, this might be a concern. However, given the complex, nonlinear nature of asphalt behavior and the capabilities of neural networks in handling such data, we chose to retain all parameters for our analysis. This decision was supported by the subsequent high performance of our hybrid model, which achieved high R^2 values and low mean squared error (MSE).

2.2. Data collection

The model's development relied on data from three carefully selected sources, chosen to ensure a comprehensive, high-quality, and represen-

Table 2
Summary of the statistical data used to develop the models.

Input Parameters	Mean	Standard Deviation	Min	Max
Aging Conditions*	-	-	-	-
Asphaltenes [%]	15.17	3.23	4	26.1
Aromatics [%]	45.98	7.56	24.3	53.3
Resins [%]	4.96	2.76	1.9	13.2
Saturates [%]	33.88	5.15	24.12	54.9
Penetration [dmm]	73.57	27.43	36.26	160
Softening Point [°C]	52.15	4.64	42.78	58.16
Test Temperature [°C]	19.40	18.03	-10.54	60
Test Frequency [rad/s]	17.52	28.36	0.1	100
Fillers [%]	18.2	7.2	8	28
Fillers Stiffness [GPa]	34.5	70.7	30	140

* Aging Conditions were modeled as categorical variables representing standardized test states: Fresh (no aging, 0 hours, 0 MPa), Short-Term (RTFOT, 5 hours at 163 °C, 0 MPa), and Long-Term (PAV, 20 hours at 100 °C, 2.1 MPa). These reflect discrete experimental conditions rather than a continuous range.

tative dataset to predict the rheological properties of asphalt binders and mastics. These sources were selected based on their informativeness (the depth and relevance of measured parameters), representativeness (coverage of diverse binder types, aging conditions, and test scenarios), and adherence to standardized testing protocols, which enable robust model training and validation. The first source, experimental data from Delft University of Technology, provides high-quality measurements from standardized tests, capturing a wide range of binder behaviors [31]. The second source, experiments from the University of Antwerp, augments the data of the first source by adding variability in binder compositions and aging states [32]. The third source, the SHRP Materials Reference Library, offers a well-established, industry standard dataset for benchmarking and enhancing dataset diversity. Together, these sources ensure that the dataset reflects the complex interplay of chemical, physical, and environmental factors that influence the behavior of asphalt material [33]. To ensure consistency across all sources, the SARA fractions for the majority of samples were determined using the same method: Thin-Layer Chromatography with Flame Ionization Detection (TLC-FID), following the IP 469 (2001) standard [34].

To address the scarcity of experimental mastic data, FEM simulations were employed to generate synthetic data for the mastic scale, resulting in a dataset of 5000 data points compared to 400 for the binder scale. The FEM simulations were conducted using ABAQUS by employing a Maxwell viscoelasticity model with Prony series parameters to represent the time-dependent behavior of asphalt mastics. The models incorporated binder properties from the aforementioned experimental datasets, augmented with filler characteristics (percentage: 8–28%, stiffness: 30–140 GPa), and were tested in a range of frequencies (0.1–100 rad/s) and temperatures (−10.54 to 60 °C).

The FEM assumed a random distribution of filler particles within a homogeneous binder matrix, neglecting particle agglomeration to simplify computational complexity. The Maxwell model captured viscoelastic responses but assumed linear viscoelasticity, which may not fully represent nonlinear behaviors under extreme conditions (Section 2.4). Filler-binder interactions were modeled based on physicochemical bonding, with stiffness values derived from typical mineral fillers used in asphalt mastics. The FEM-generated data were validated against both analytical and experimental benchmarks to ensure reliability (Appendix A).

Due to the comprehensive nature of the mastic testing, the resulting dataset was significantly larger than of the asphalt binder model. Specifically, the mastic model incorporated 5000 data points, while the asphalt binder model utilized 400 data points. This substantial difference in dataset size reflects the complexity and expanded testing requirements associated with the mastic level. 80% of the collected data was allocated to the training set, with the remaining 20% reserved for the testing set. This division ensures robust model training while providing a rep-

resentative sample for model validation and testing. Summary of the statistical data used to develop the models is listed in Table 2.

2.3. Hybrid model development

2.3.1. Deep neural network (DNN) component

The DNN were developed using the TensorFlow library based on feedforward backpropagation [35]. These networks process input features through multiple hidden layers of neurons, enabling the learning of complex patterns. The output of each neuron is computed as follows:

$$y = f \left(\sum_{i=1}^n w_i x_i + b \right) \quad (1)$$

where x_i are inputs, w_i are weights, b is the bias, and f is the activation function. The overall network function, F , representing the feedforward operation through multiple layers, is then given by:

$$F(x) = f^{(L)} \left(f^{(L-1)} \left(\dots f^{(1)}(x) \right) \right) \quad (2)$$

where L denotes the layer and $f^{(L)}$ is the activation function applied at the L -th layer. The weights are updated through backpropagation to minimize prediction errors and allow the model to generalize from data [36].

A five-fold cross-validation method was used to determine optimal architecture and prevent overfitting [37]. It was found that a network with three hidden layers (64, 32, and 16 neurons, respectively) provided optimal performance, balancing model complexity and performance. Activation functions introduce nonlinear properties to the network. Without activation functions, a DNN, regardless of its depth, would behave just like a single-layer perceptron because summing these layers would still result in linear operations. In this study, the Rectified Linear Unit (ReLU) function was chosen as the activation function. ReLU is defined as $f(x) = \max(0, x)$, which simply outputs the input if it is positive and zero otherwise. This simplicity makes ReLU computationally efficient and less prone to vanishing gradient problems during training, where gradients can become too small for effective learning. Thus, the use of ReLU contributes to faster convergence during training and improved overall performance of the deep neural network [38].

The Adam optimizer was used with a learning rate of 0.001 for stable and adaptive gradient updates [39]. A batch size of 128 balanced convergence stability with memory efficiency. To improve the robustness and generalization of the model, early stopping (patience = 50), L_1 regularization (0.01), and dropout (rate = 0.2) were applied.

Two scaling techniques were used to prepare the data for neural network modeling. First, a logarithmic transformation was applied to the target variables (G^* and δ) to stabilize the variance and normalize their distribution, addressing the issue of skewed data (observed at the binder

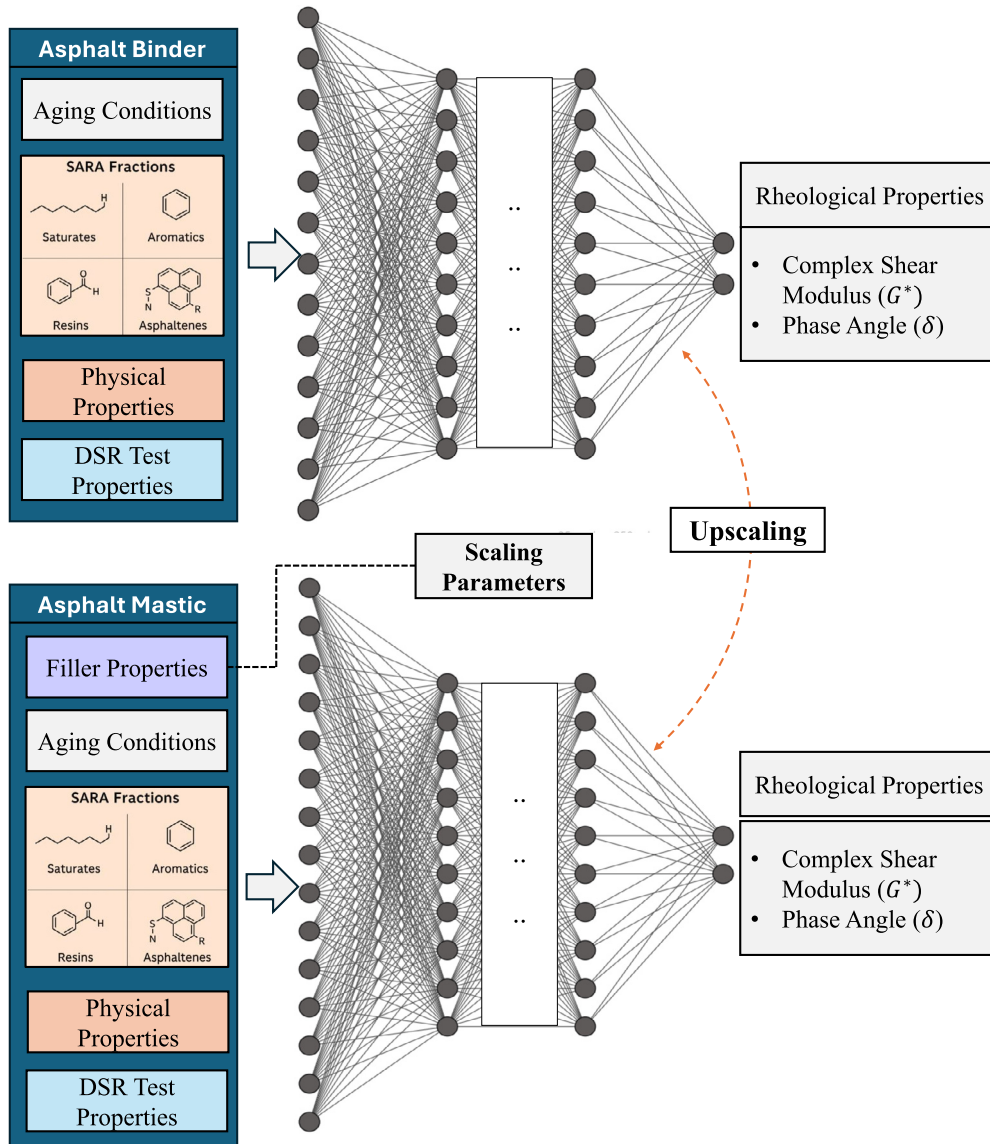


Fig. 2. Schematic representation of the final DNN models used for the first predictions.

scale) with a wide range of values. This ensures more meaningful model error metrics and better generalization. Secondly, standard scaling was used for both features and targets to center the data with a mean of zero and a standard deviation of one, making the learning process more efficient and preventing any single feature from dominating due to its scale. Details of the hyperparameters and the final DNN models are provided in Table 3 and Fig. 2, respectively.

2.3.2. Gaussian process regression (GPR) model

The GPR model was developed using the GPy library [40]. GPR models are nonparametric probabilistic models that are particularly suited for regression tasks. They provide not only predictions, but also associated uncertainty estimation, which is advantageous in assessing the reliability of the predictions.

Given a set of input features X and corresponding target values y , the GPR model assumes a prior distribution over functions and is given by:

$$y = \mathcal{GP}(m(x), k(x, x')) \quad (3)$$

where $m(x)$ is the mean function, often assumed to be zero, and $k(x, x')$ is the covariance function or kernel, such as the Radial Basis Function

(RBF) used here. The GPR predictive distribution at a new input x^* is then normally distributed, given by:

$$y^* | x^*, X, y \sim \mathcal{N}(\mu(x^*), \sigma^2(x^*)) \quad (4)$$

where $\mu(x^*)$ and $\sigma^2(x^*)$ are the mean and variance of the predictive distribution, respectively.

The Radial Basis Function (RBF) kernel was chosen for the GPR model. This kernel is a popular choice for GPR models due to its flexibility and capability to handle nonlinear relationships in the data. The kernel parameters (variance and lengthscale) were initialized as 0.5 and 1, respectively. The GPR optimization process adjusts the kernel parameters to best capture the underlying patterns in the data.

For the mastic-scale model, which comprises 5,000 data points due to the inclusion of filler properties and varied test conditions, a Sparse Gaussian Process (SGP) model was employed to enhance computational efficiency. Full GPR has a computational complexity of $O(n^3)$, which, for 5,000 data points, led to significant training times during our experiments, particularly when optimizing kernel parameters. By contrast, SGP reduces this complexity to $O(nm^2)$, where $m = 200$ inducing points were used [41].

Table 3
Results of hyperparameter tuning and validation for DNN models.

Hyperparameter	Value
No. of Inputs	11 ^a - 13 ^b
No. of Hidden Layers	3.0
No. of Hidden Neurons	64 32 16
Activation Function	ReLU ReLU
Optimizer	Adam
Learning Rate	0.001
Dropout Rate	0.2
L_1 Regularization	0.01
Batch Size	128
Epoch	1000 ^c

Validation Results for Hidden Layer Configurations (Binder - Average)			
No. of Layers	Training MSE	Testing MSE	R ²
1 Layer	0.35	0.68	0.85
2 Layers	0.22	0.55	0.88
3 Layers ^d	0.08	0.32	> 0.9
4 Layers	0.08	0.38	0.9

Validation Results for Hidden Layer Configurations (Mastic - Average)			
No. of Layers	Training MSE	Testing MSE	R ²
1 Layer	0.40	0.70	0.84
2 Layers	0.25	0.58	0.87
3 Layers ^e	0.10	0.10	> 0.9
4 Layers	0.10	0.15	0.9

^a Asphalt binder model.

^b Asphalt mastic model.

^c Early stopping with a patience of 50 epochs.

^d Selected configuration for the binder and mastic models due to lowest testing MSE and highest R², balancing performance and complexity.

^e Selected configuration for the mastic model due to lowest testing MSE and highest R², balancing performance and complexity.

Table 4
Results of the hyperparameter tuning for the GPR models.

Hyperparameter	Value
No. of Inputs	5.0
Kernel Type	RBF
Initial Variance	0.5
Initial Lengthscale	1.0
Number of Inducing Points	200 ^a

^a Only in the mastic model.

These inducing points help approximate the posterior distribution of the GPR, making the model computationally efficient without significantly sacrificing accuracy. For the SGP model, the same RBF kernel was used, and the number of inducing points m was chosen based on a trade-off between computational efficiency and prediction accuracy. The inducing points were initialized using a subset of the training data, and the optimization process adjusted their positions to best represent the underlying data distribution. Table 4 lists the hyperparameter used to train the GPR models.

2.3.3. Hybrid modeling: integrating DNN and GPR for improved accuracy and uncertainty assessment

To develop the hybrid DNN-GPR model, five-fold cross-validation was employed to determine the optimal DNN architecture, specifically the number of hidden layers and neurons. This process identified a configuration with three hidden layers (64, 32, and 16 neurons) as optimal, balancing performance and complexity (Table 3). Subsequently, five DNN were independently trained using this fixed architecture and identical hyperparameters. The random seed was not fixed, allowing each DNN to start with different initial weights and experience variations in

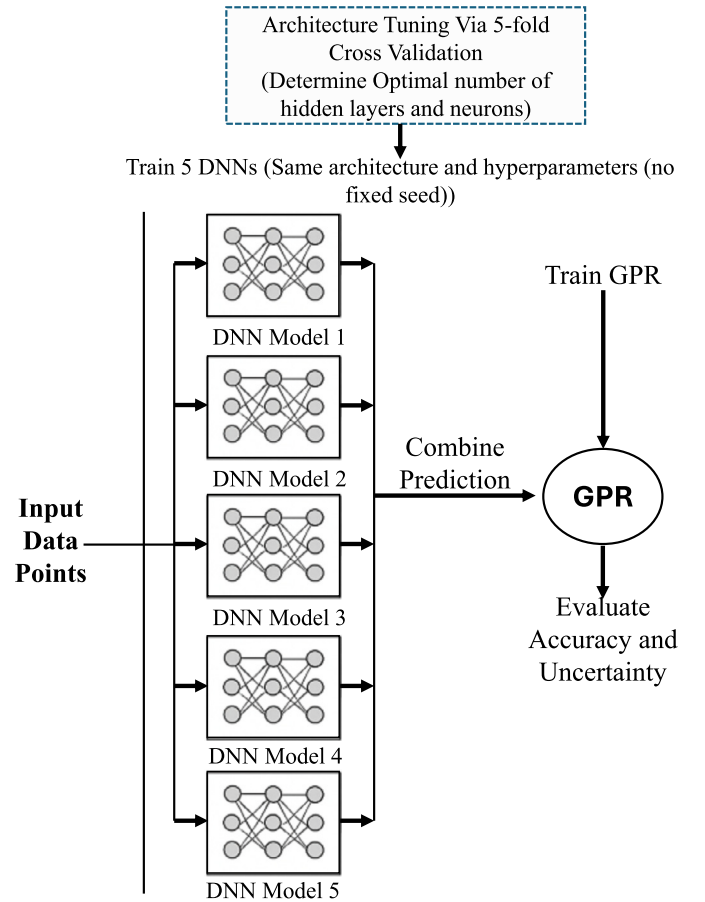


Fig. 3. Schematic representation of the hybrid approach combining five DNN and GPR for asphalt binder and asphalt mastic predictions.

batch sampling during training. This approach ensures that the five DNN are neither cross-validation folds nor models with different hyperparameter configurations but rather instances of the same model with varying weight initializations due to random initialization and stochastic training processes.

The predictions of the five independently trained DNN, denoted as $p_1, p_2, \dots, p_5$, are generated for the entire dataset and aggregated to serve as inputs for the GPR model. Each DNN produces predictions for G^* and δ , capturing variability due to random weight initialization and batch sampling. To form the GPR input, these predictions are concatenated into a 5-dimensional feature vector, $[p_1, p_2, p_3, p_4, p_5]$, without additional weighting or transformation, preserving the raw predictive information. The GPR model, implemented with an RBF kernel, refines these predictions and quantifies their uncertainty:

$$y^* = \mathcal{GP}(p_1, p_2, \dots, p_5) \quad (5)$$

This methodology aims to use the robust feature extraction capabilities of DNN and the probabilistic strength of GPR to improve accuracy and reliability of the prediction [42]. Fig. 3 shows a schematic representation of the combined modeling approach using DNN and GPR, detailing the process flow from initial data input to the integration of predictions and assessment of model accuracy.

2.4. Finite element model: incorporating mastic properties

Studying the mastic level is essential for several reasons. When fillers are mixed with asphalt binder to create mastic, they alter the physical and chemical properties of binders, which in turn affects pavement performance [10]. These fillers can also modify the binder's viscosity and

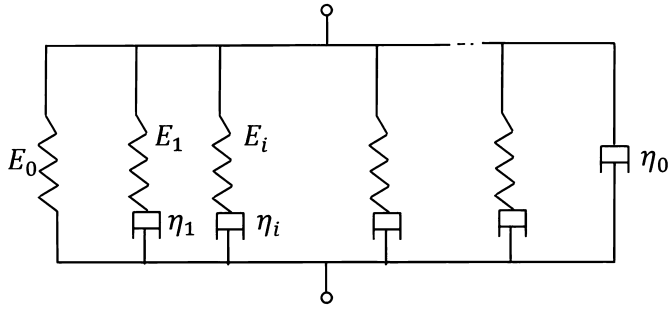


Fig. 4. Schematic of the Maxwell Model for viscoelastic materials, showing E_1 and E_2 as springs and η_1 and η_2 as dashpots.

stiffness, influencing the pavement's ability to resist deformation and cracking. However, data at the mastic scale is scarce and often unavailable. Even when such data exists, it is challenging to find datasets that comprehensively cover chemical and physical properties along with various aging conditions. This scarcity can be attributed to the complexity of mastic-level testing, the high costs and time requirements for comprehensive analysis. Additionally, proprietary concerns often limit the sharing of detailed formulation data.

The primary objective of the FEM is to generate synthetic training data at the mastic scale. This synthetic data will then be used to train the hybrid model. To simulate the mastic level, the Maxwell model available in ABAQUS, a widely-used FEM software, was employed [43]. The Maxwell model is a viscoelastic model that represents material behavior as a combination of elastic and viscous components, consisting of a spring and a dashpot connected in series (Fig. 4). The basic equation for the Maxwell model is:

$$\sigma(t) = E * \epsilon(t) * \exp\left(-\frac{t}{\tau}\right) \quad (6)$$

where $\sigma(t)$ is stress, E is the elastic modulus, $\epsilon(t)$ is strain, t is time, and τ is the relaxation time, which represents the characteristic time scale over which stress decays in a viscoelastic material under constant strain. In ABAQUS, this model is implemented using the parameters of the Prony series, which allow for the representation of multiple relaxation times to more accurately capture complex viscoelastic behavior [44]. The Prony series expresses the relaxation modulus as a sum of exponential terms:

$$E(t) = E_{\infty} + \sum \left(E_i * \exp\left(-\frac{t}{\tau_i}\right) \right) \quad (7)$$

where E_{∞} is the long-term modulus, E_i are relaxation strengths, and τ_i are relaxation times. This formulation enables a more detailed representation of the material's time-dependent response.

For these simulations, binder-level data collected from available sources served as a starting point. To generate data at the mastic level, fillers were incorporated into the model and treated as elastic materials. This approach allowed for the simulation of the composite behavior of the mastic, accounting for both the viscoelastic properties of the binder and the elastic properties of the fillers.

In the FE model, the top and bottom plates were designed as rigid bodies. A cohesive stiffness interaction was established between the mastic sample and these plates. While the bottom plate remained stationary, the upper plate was considered to oscillate in a logarithmic sine sweep pattern, replicating the conditions of a laboratory test. The fillers were randomly distributed throughout the asphalt binder, with tie contacts assigned between the fillers and the binder to represent their interaction. The mesh size was carefully chosen to be sufficiently small, balancing computational efficiency with precision in representing the mastic's structure and behavior. Fig. 5 shows the process of integrating asphalt binder and fillers to generate new datasets using the FEM model.

2.5. Feature importance analysis

2.5.1. Pre-modeling feature selection techniques

Pre-Modeling feature selection uses statistical and machine learning methods as preliminary steps before constructing the final model. These preliminary techniques assist in selecting relevant features, conducting significance tests, and generating initial predictions about the most important features. In this study, three methods were employed including the Chi-Squared test, ANOVA test, Random forest.

Chi-Squared method. The Chi-Squared method is used to test the independence between variables. In the context of feature selection for machine learning, the Chi-Squared test can be applied to identify features that are most relevant to the target variable [45]. The Chi-Squared statistic is calculated as follows:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (8)$$

where O_i is the observed frequency for category i , E_i is the expected frequency for category i . The expected frequency E_i is calculated as:

$$E_i = \frac{(\text{row total}) \times (\text{column total})}{\text{grand total}} \quad (9)$$

ANOVA test. The Analysis of Variance (ANOVA) test is a statistical method used to compare the means of three or more samples to understand if at least one sample mean is significantly different from the others. In machine learning, ANOVA is useful for feature selection by identifying features that contribute significantly to the variance in the target variable [46]. The F-statistic used in ANOVA is calculated as follows:

$$F = \frac{\text{Mean Square Between Groups (MSB)}}{\text{Mean Square Within Groups (MSW)}} \quad (10)$$

where MSB (Mean Square Between Groups) is the variance between the group means, MSW (Mean Square Within Groups) is the variance within the groups. The formulas for MSB and MSW are:

$$\text{MSB} = \frac{\sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2}{k - 1} \quad (11)$$

$$\text{MSW} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}{N - k} \quad (12)$$

where k is the number of groups, n_i is the number of observations in group i , $\bar{X}_i$ is the mean of group i , $\bar{X}$ is the overall mean, N is the total number of observations.

Random forest. Random forest is a powerful ensemble learning technique that builds multiple decision trees and merges their outputs to improve predictive accuracy and control overfitting [47]. Each tree in the Random forest is trained on a different subset of the data, and the final prediction is made by averaging the predictions of all trees (for regression) or by majority vote (for classification) [48]. This method is particularly effective for feature selection as it provides feature importance scores based on the contribution of each feature to the model's accuracy. The importance score for feature f is:

$$I(f) = \sum_{t \in T} \frac{N_t}{N} \Delta i_t(f) \quad (13)$$

where T is the set of all trees in the forest, N_t is the number of samples reaching node t , N is the total number of samples, $\Delta i_t(f)$ is the decrease in impurity at node t when split by feature f .

2.5.2. Post-modeling feature importance evaluation

Post-modeling feature importance approach involves techniques that are applied after the machine learning models have been trained. These methods use the outputs of the models to understand the importance of different features.

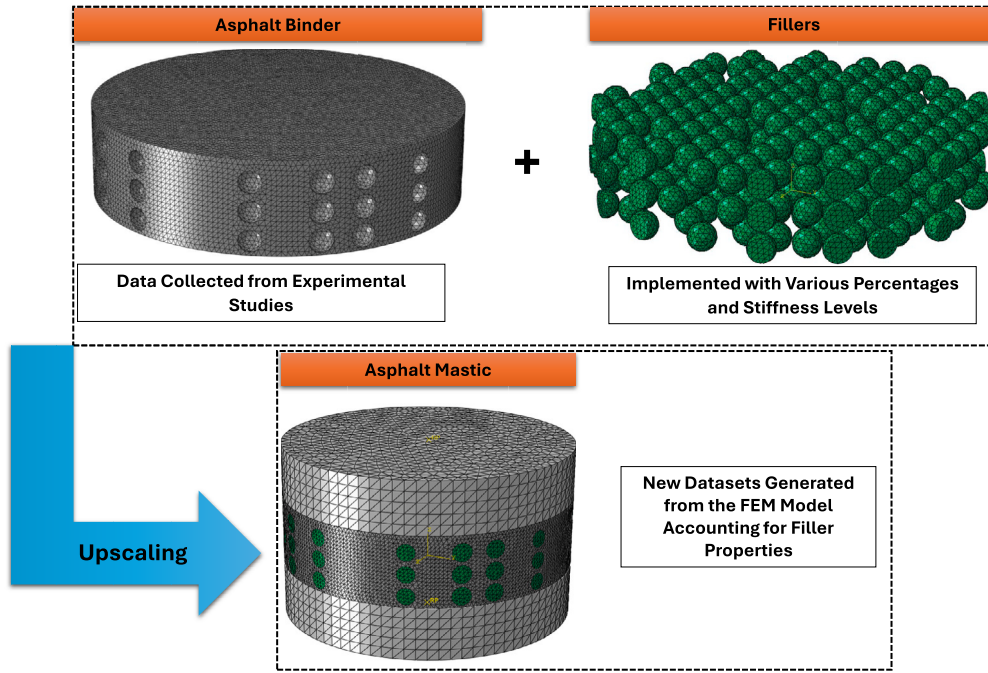


Fig. 5. Integrating asphalt binder and fillers to generate new datasets using the FEM model.

SHAP (SHapley Additive exPlanations) method. SHAP is a unified framework for interpreting machine learning models by assigning each feature an importance value for a particular prediction [49,50]. Based on cooperative game theory, SHAP values provide a way to fairly distribute the prediction among the features according to their contribution. This method can help in understanding the impact of each feature on the predictions and identify which features are driving the model's decisions [51,52].

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \quad (14)$$

where F is the set of all features, S is a subset of features that does not include i , $F(S)$ is the model prediction for the subset S , $|S|$ is the size of subset S . This equation sums the marginal contributions of feature j across all possible subsets S of the feature set F .

3. Results and discussion

3.1. Prediction of binder rheological properties

Figs. 6 to 9 illustrate the performance of five DNN models in predicting the complex modulus G^* and phase angle δ of asphalt binder, as well as the performance of the GPR model using the predicted results from the DNN. The training set predictions versus true values for G^* show a high R^2 value ranging from 0.89 to 0.96, indicating strong model performance. Similarly, for δ , the R^2 values are consistently high, around 0.97, demonstrating that the DNN models have effectively captured the relationship in the training data. In the testing set, the performance for G^* remains robust, with R^2 values between 0.92 and 0.96. However, the scatter plots indicate some deviations from the ideal fit, pointing to potential uncertainties. For δ , the R^2 values are lower, around 0.83 to 0.84, suggesting higher uncertainty and less precise predictions compared to the training set.

The MSE plots for both training and testing datasets show a decreasing trend with increasing epochs, indicating that the models are learning and improving their predictions over time (Fig. 10). The uncertainties in the DNN models can be attributed to the complex relationship between the input features and the targets. DNN, despite their power, may not always capture these complex relationships perfectly. The low number

of the dataset at the binder scale might also lead to higher uncertainty in the predictions.

The GPR model, using the DNN predicted results, shows an improvement in precision with an R^2 value of 0.997 for G^* and 0.9471 for δ (Fig. 11). The GPR model also provides a range of uncertainty ($\pm 2\sigma$ range), clearly visible in the shaded areas around the predictive mean. This range helps in quantifying the confidence in the predictions, which was not provided by the DNN models alone.

However, for δ , the uncertainty range is notably wider, particularly at lower values. For instance, at a predicted δ of 20° , the $\pm 2\sigma$ range spans approximately $\pm 10^\circ$, representing a relative uncertainty of about 50%. This level of uncertainty can be significant in practical asphalt applications, where precise phase angle measurements are critical for assessing the viscoelastic behavior of binders, such as distinguishing between elastic and viscous responses under varying environmental conditions. The wider uncertainty in δ predictions is likely influenced by the limited binder-scale dataset, which may not fully capture the complex, nonlinear relationships between input features (e.g., aging conditions, test temperature, and frequency) and δ .

Additionally, δ is highly sensitive to aging conditions (e.g., long-term PAV aging), which introduces greater variability in the predictions, especially in regions with sparse data.

Section 3.4 provides more examples to further illustrate the benefits of combining GPR and DNN models for asphalt binder property prediction.

3.2. Prediction of mastic rheological properties

At the mastic level, five DNN were also developed, each taking 13 input parameters that represent properties of both the binder and filler. The training results show high R^2 values for both G^* and δ , with δ showing slightly lower values compared to G^* (Fig. 12 and Fig. 13). The testing phase demonstrates similar trends, with R^2 values of 0.97 for G^* and 0.95 to 0.96 for δ (Fig. 14 and Fig. 15). Fig. 16 shows that both the training and testing MSEs stabilize towards the end of the training process, suggesting the convergence of the model without overfitting. However, prediction uncertainty is observed in both training and testing phases, indicating a need for further investigation and management of this uncertainty.

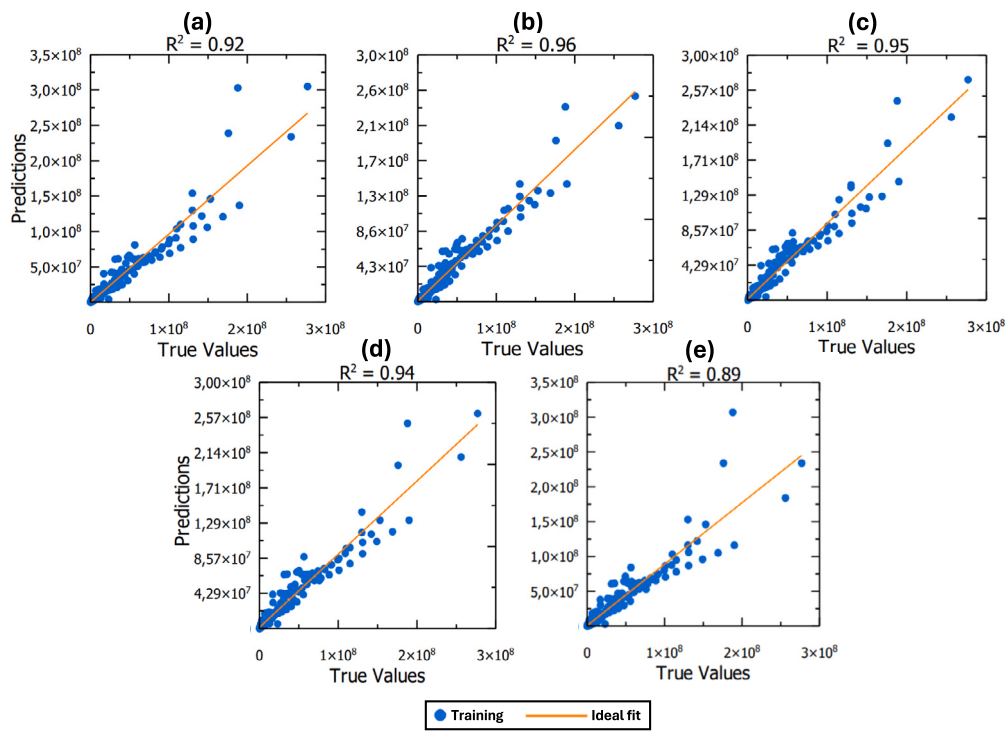


Fig. 6. Training performance of five DNN models in predicting G^* (a-e) of asphalt binder. The true values correspond to experimentally measured data.

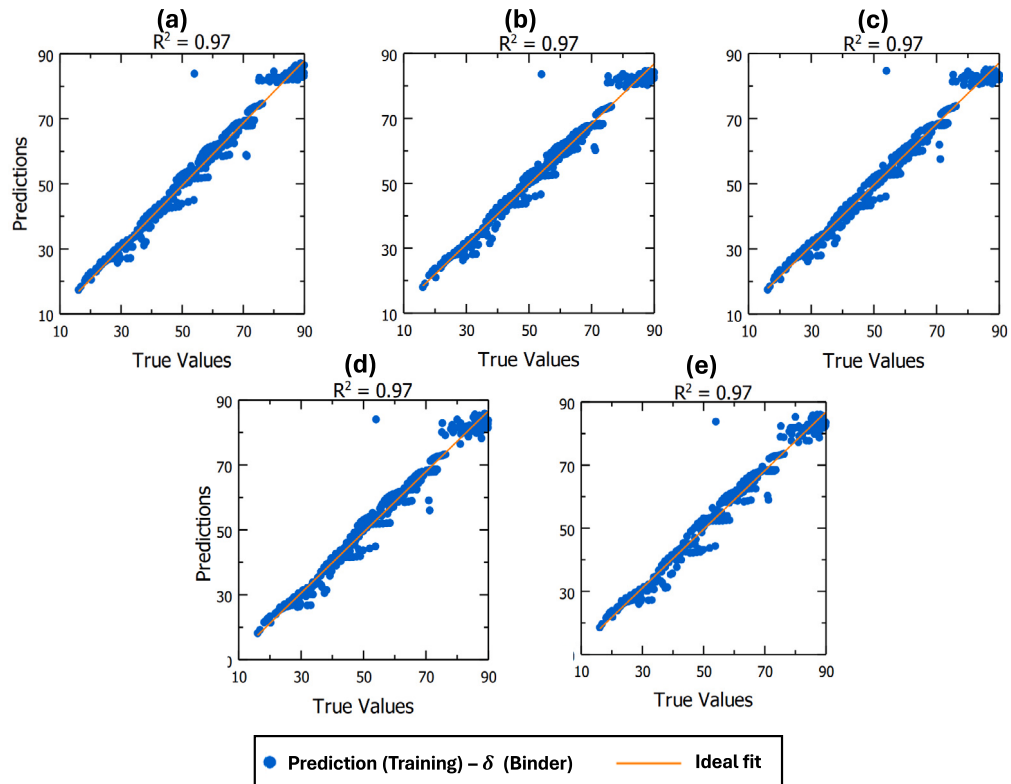


Fig. 7. Training performance of five DNN models in predicting δ (a-e) of asphalt binder. The true values correspond to experimentally measured data.

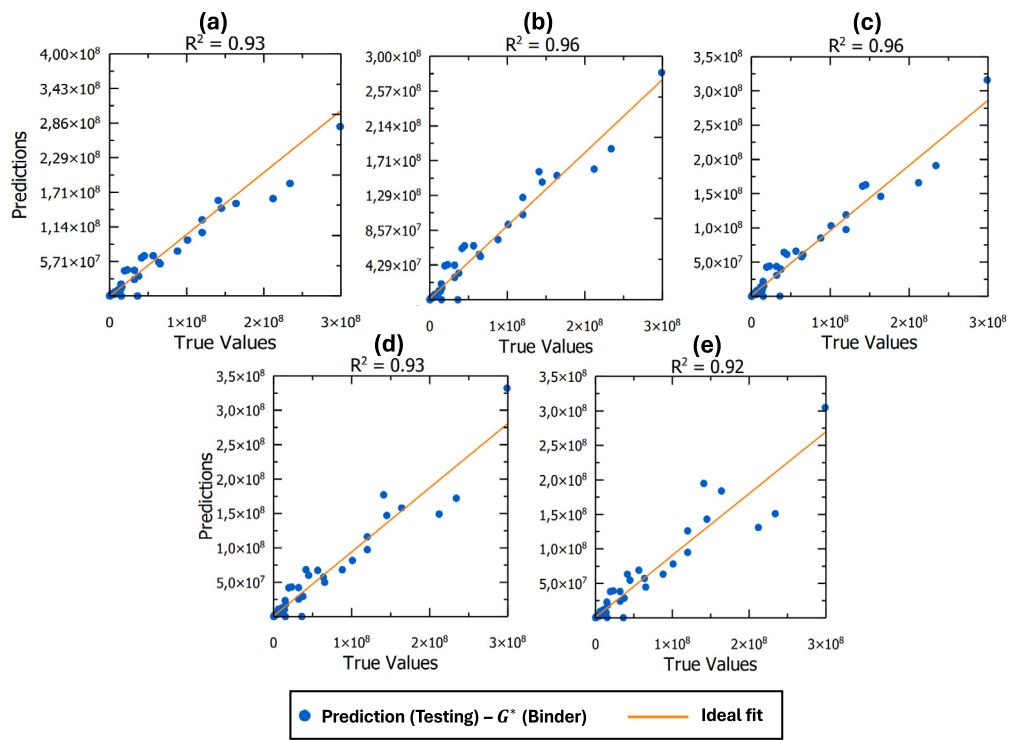


Fig. 8. Testing performance of five DNN models in predicting G^* (a-e) of asphalt binder. The true values correspond to experimentally measured data.

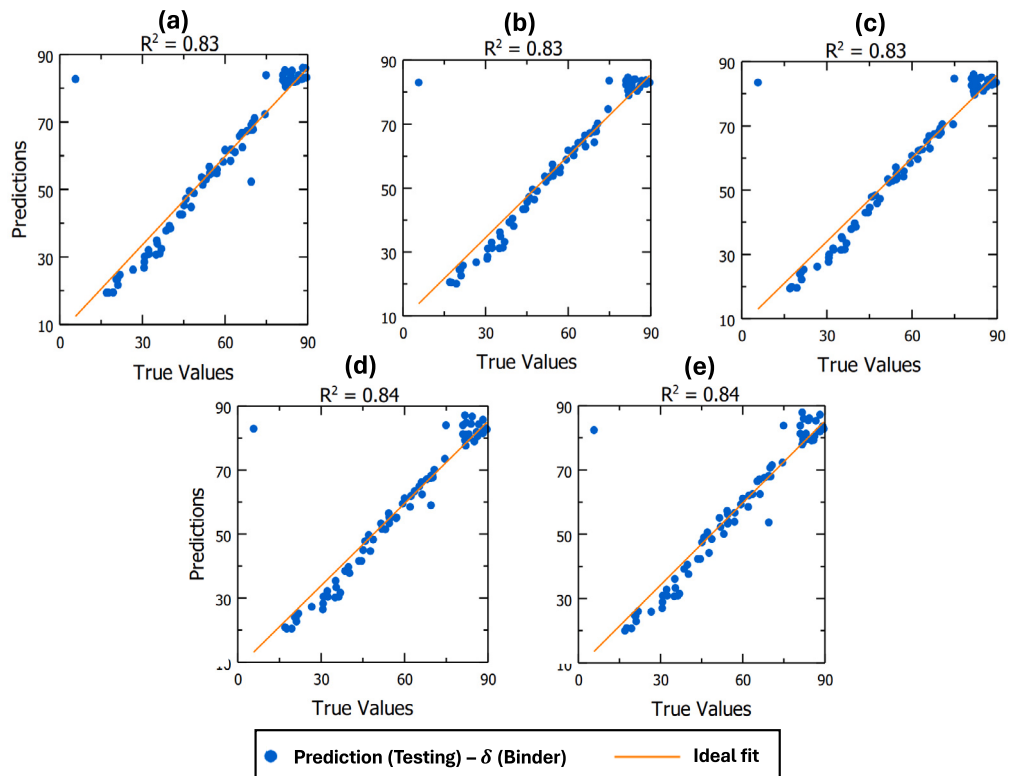


Fig. 9. Testing performance of five DNN models in predicting δ (a-e) of asphalt binder. The true values correspond to experimentally measured data.

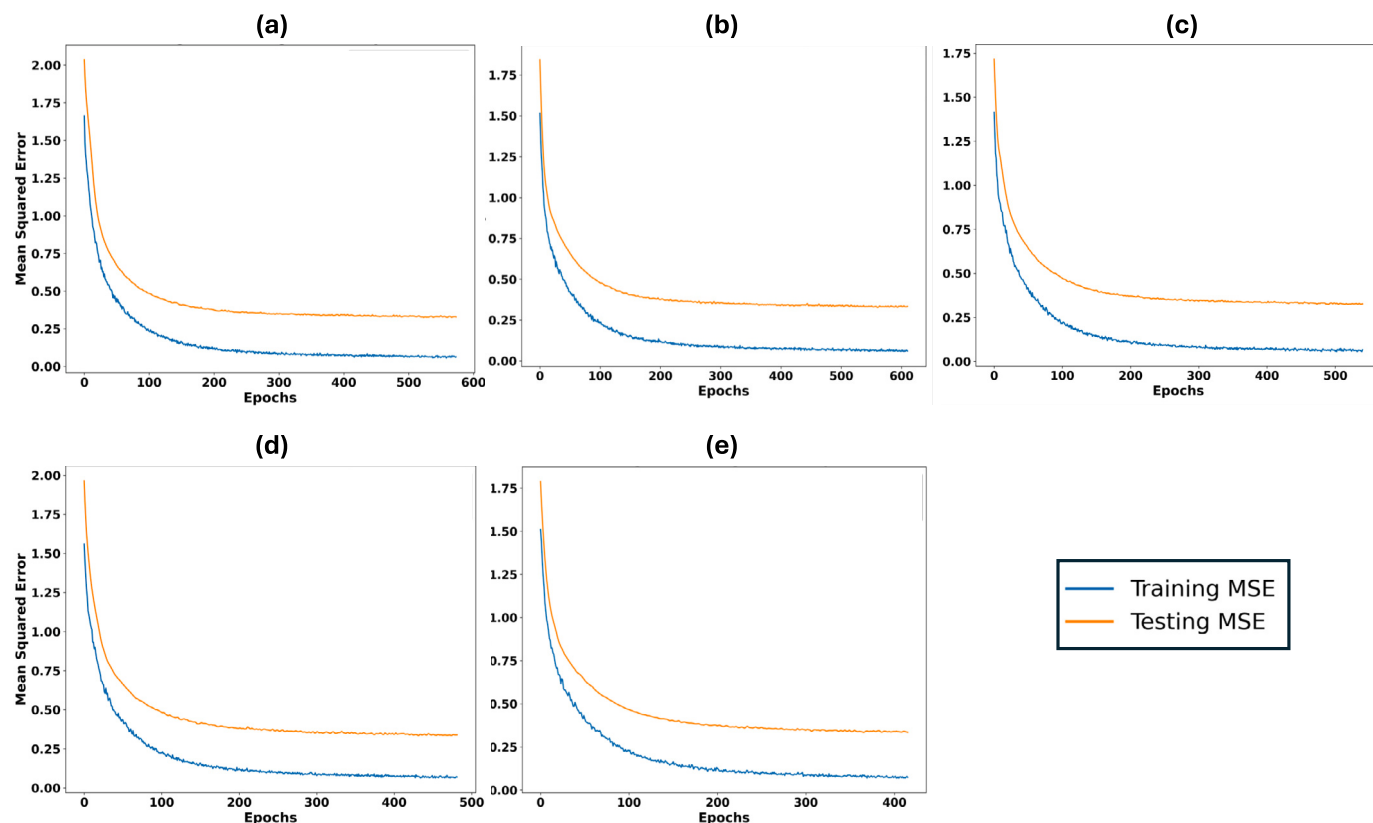


Fig. 10. Comparative view of training and testing MSE across epochs for five deep neural networks used for predicting asphalt binder rheological properties. Subfigures (a) to (e) show training MSE (blue) and testing MSE (orange) curves for each network.

For the GPR model, G^* shows a high R^2 value of 0.9930, indicating excellent predictive performance with well-quantified uncertainty (Fig. 17). The model for δ performs slightly less well, with an R^2 value of 0.9726. The narrow uncertainty bands around the predictions for both models suggest a high level of confidence in the results.

3.3. Key influential properties

3.3.1. Comparative analysis of feature importance

The selection of statistical approaches can significantly affect conclusions about which features have the greatest impact. This highlights the importance of carefully choosing methods that suit both the characteristics of the dataset and align with current understanding and expectations in the field of study. Fig. 18 illustrates the normalized percentage contributions of various features to the outputs G^* (left) and δ (right) analyzed via three different approaches: ANOVA, Random forest, and Chi-squared.

Multiple methods indicate varying feature importance for G^* . Key factors such as test temperature and penetration consistently emerge as significant, alongside the categorical aging state, highlighting its role in determining G^* . For δ , the aging state particularly the long-term PAV condition stands out as a consistently important feature across all methods, with aging pressure showing a pronounced effect in the random forest model.

However, while ANOVA, Chi-squared, and random forest methods provide valuable insights, they also have limitations. ANOVA assumes normally distributed data and equal variances, which may not always be true. Chi-square tests require categorical data and may not capture complex interactions between features. Random forest, although flexible, can be computationally intensive for large datasets and may overfit without careful tuning [53,54].

When dealing with large datasets, these traditional methods can struggle with scalability and may not efficiently handle high-dimensional data. They often fail to account for feature interactions adequately, leading to potentially misleading conclusions [55].

3.3.2. SHAP analysis for feature importance analysis

The SHAP method uses the outputs of the ML models to accurately identify the most important features, providing a more robust and scalable approach to feature selection. In this study, SHAP analysis was performed using the predictions from the five DNN models, which form the core of the hybrid DNN-GPR framework. This choice aligns with the primary modeling approach of the study. SHAP was used to classify the most important features that impact G^* and δ on both the binder and the mastic scales, effectively handling larger datasets specifically at the mastic scale and offering detailed information on the significance of the features on different scales [56].

Fig. 19 presents bar charts that compare the importance of features for the binder scale and mastic scale using mean absolute SHAP values. A mechanistic interpretation of the key variables highlights their physical influence on asphalt rheology. Test temperature and aging temperature emerge as dominant factors for both G^* and δ on both scales. Elevated test temperatures increase molecular mobility, reducing G^* by transitioning the asphalt toward a more viscous state, which in turn increases δ as the material exhibits greater lag in strain response. Conversely, aging temperature, particularly during long-term aging such as PAV at 100 °C, accelerates oxidative aging by forming polar functional groups like ketones and sulfoxides. This process increases binder stiffness, elevating G^* and reducing δ by enhancing elastic recovery, which is crucial for resisting permanent deformation such as rutting in pavements [57,58].

Penetration significantly influences G^* . Lower penetration values, indicative of harder binders, correspond to higher G^* due to increased

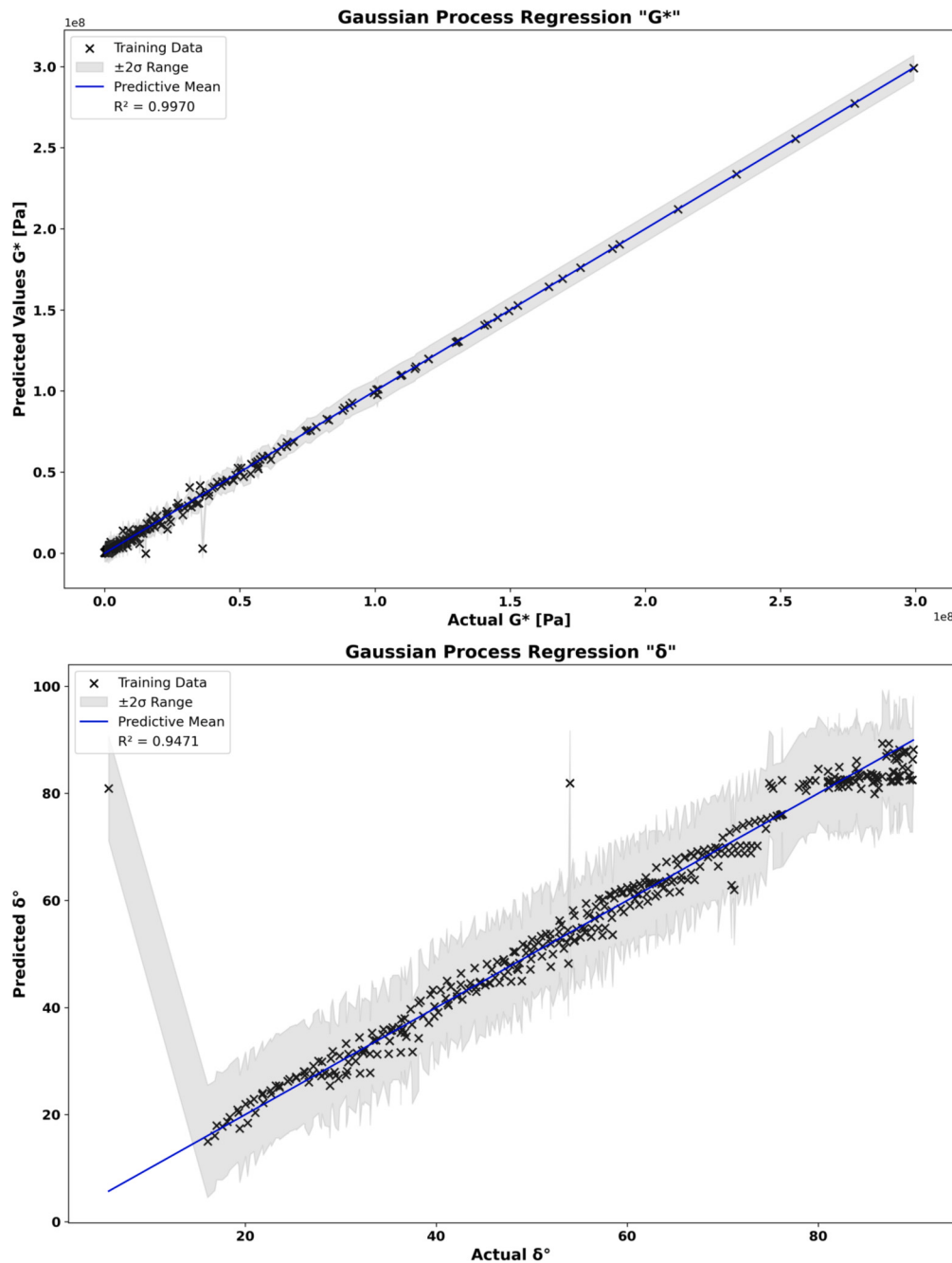


Fig. 11. Gaussian process model performance in predicting (G^*) and (δ) of asphalt binder.

resistance to shear deformation. This relationship stems from the binder's molecular structure, where reduced penetration often results from higher asphaltene content or aging-induced crosslinking, amplifying the elastic component of the viscoelastic response. For δ , penetration exerts a moderate effect, as harder binders exhibit less viscous lag, resulting in lower phase angles. Aging conditions play an important role in altering δ . Long-term aging through PAV, with 20 h exposure at 2.1 MPa, increases asphaltene and resin fractions while depleting aromatics, stiffening the binder and reducing δ by favoring elastic behavior over viscous flow.

Test frequency also affects both G^* and δ . Higher frequencies reduce relaxation time, limiting molecular rearrangement and increasing G^* as the binder exhibits greater stiffness. Simultaneously, higher frequencies decrease δ by reducing viscous dissipation, reflecting the binder's time-dependent viscoelastic properties under dynamic loading conditions like

traffic-induced stresses. At the mastic scale, filler stiffness surpasses asphaltene importance due to strong physicochemical interactions between mineral fillers and the binder. High stiffness fillers increase G^* by reinforcing the composite structure, minimizing deformation under load, and decrease δ by enhancing elastic recovery, which improves rutting resistance.

The chemical composition, represented by SARA fractions, provides further insight. Asphaltenes and naphthene aromatics exhibit moderate importance. Asphaltenes increase G^* by enhancing the polar fraction content, which stiffens the binder at high temperatures, while naphthene aromatics contribute to viscosity and ductility. Saturates and resins show lower importance, as their effects are overshadowed by temperature and aging, though they influence the binder's flexibility and aging susceptibility.

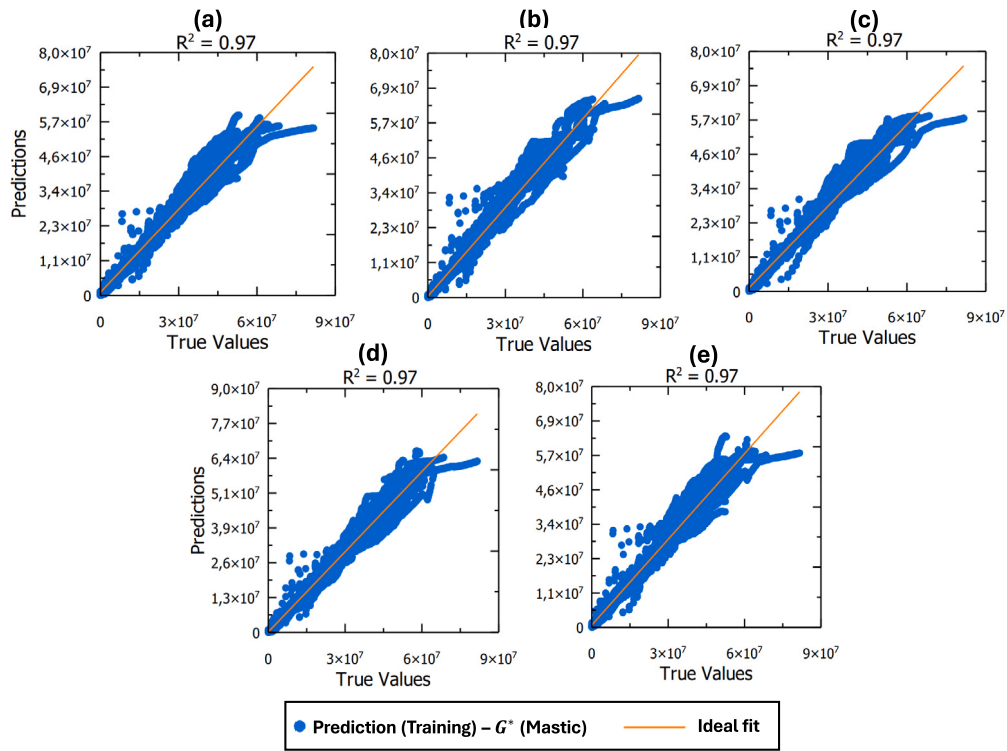


Fig. 12. Training performance of five DNN models in predicting G^* (a-e) of asphalt mastic. The true values are derived from FEM simulations, which are compared against experimental measurements.

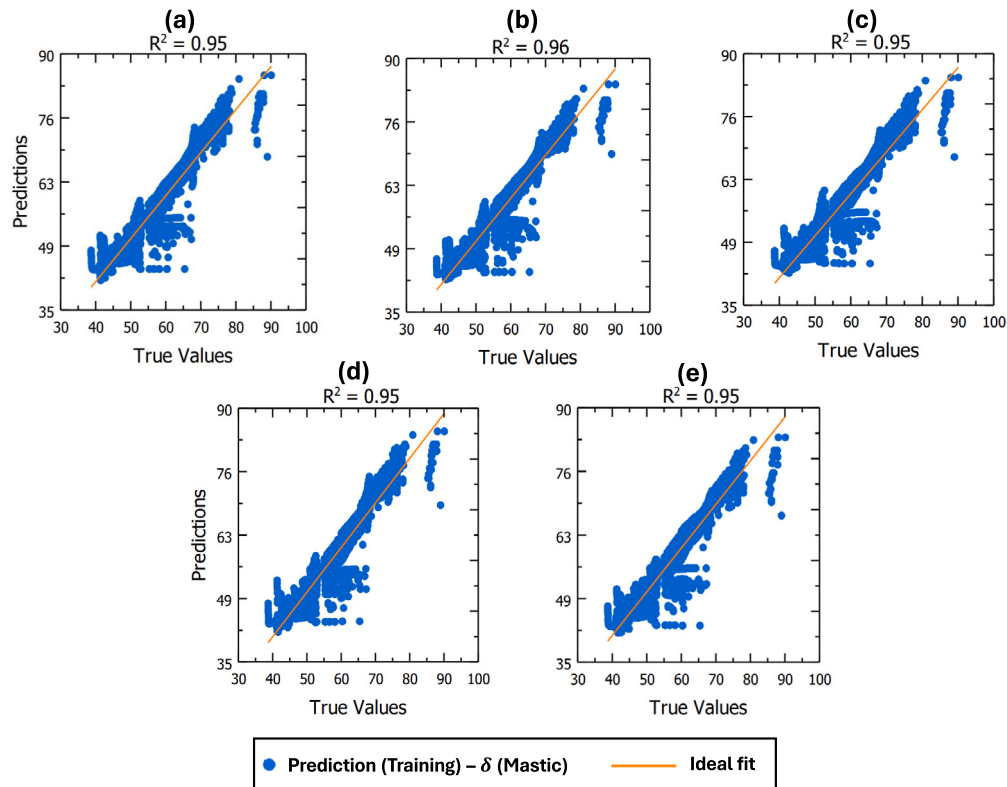


Fig. 13. Training performance of five DNN models in predicting δ (a-e) of asphalt mastic. The true values are derived from FEM simulations, which are compared against experimental measurements.

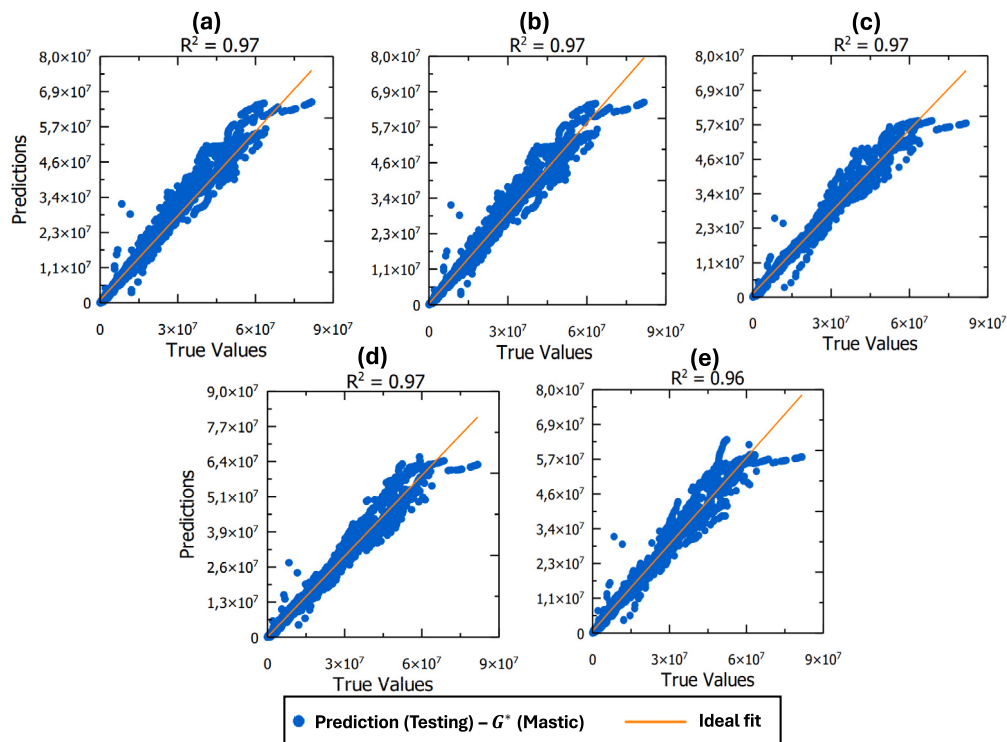


Fig. 14. Testing performance of five DNN models in predicting G^* (a-e) of asphalt mastic. The true values are derived from FEM simulations, which are compared against experimental measurements.

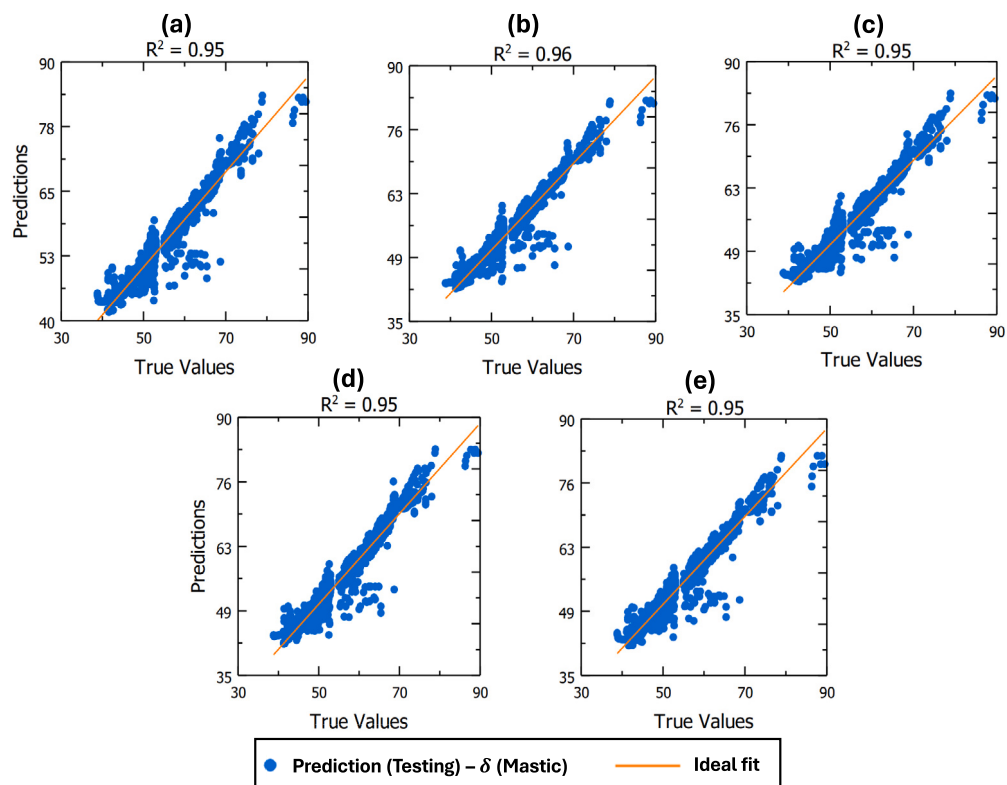


Fig. 15. Testing performance of five DNN models in predicting δ (a-e) of asphalt mastic. The true values are derived from FEM simulations, which are compared against experimental measurements.

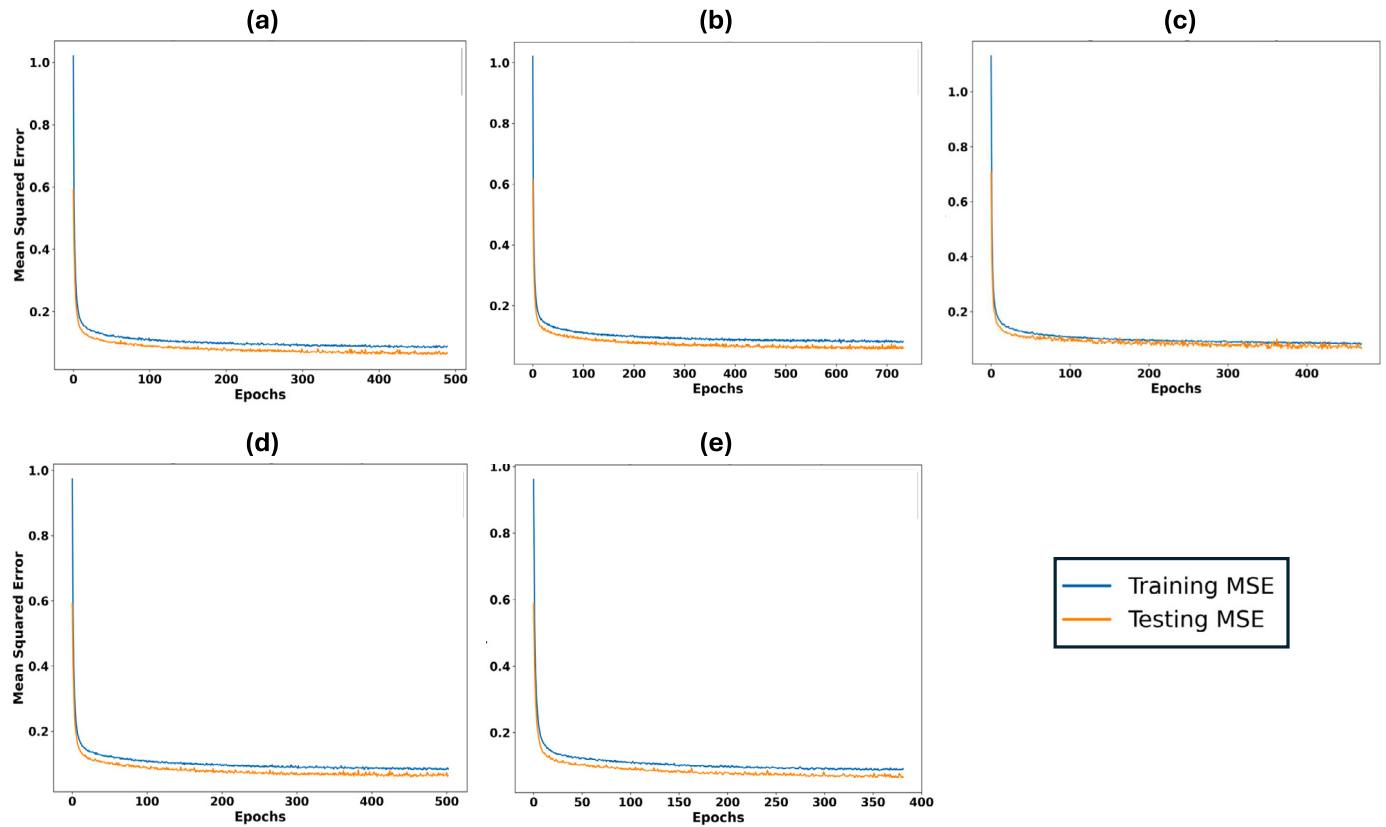


Fig. 16. Comparative view of training and testing MSE across epochs for five deep neural networks used for predicting asphalt mastic rheological properties. Subfigures (a) to (e) show Training MSE (blue) and Testing MSE (orange) curves for each network.

Table 5

Training and inference times for DNN and DNN-GPR models.

Model	Training Time (s)	Inference Time (s)
DNN (Binder)	120.5	5.1
DNN-GPR (Binder)	142.3	8.8
DNN (Mastic)	185.7	15.4
DNN-GPR (Mastic)	225.9	55.6

Table 6

Input parameters of two bitumen samples used to evaluate the accuracy of the hybrid model.

Parameter	Fresh Bitumen	LTA Aged Bitumen
Aging Temperature [°C]	0	100
Aging Time [h]	0	20
Aging Pressure [MPa]	0	2.1
Asphaltene (%)	12.8	18.6
Naphthene aromatics (%)	53.3	43.9
saturates (%)	3.6	3.6
resins (%)	30.3	33.9
Penetration [dmm]	91	44.2
Softening Point [°C]	48	58.1576736
Test Temperature [°C]	0	0
Test Frequency [rad/s]	100	0.2

3.4. Comparison between the DNN and the DNN-GPR hybrid model

A comparative analysis was performed to assess the improvements of the proposed hybrid model. The five DNN were compared to the hybrid DNN-GPR model. In addition to predictive accuracy, computational costs were evaluated (Table 5). The DNN-GPR model increases training time by 18.1% (binder) and 21.6% (mastic) due to GPR's kernel optimization, with inference times rising by 33.3% and 35.3%, respectively. These modest increases are acceptable for engineering design, as GPR's uncertainty quantification enhances prediction reliability (e.g., reducing G^* error from 22.7% to 0.031% for fresh asphalt binder).

The analysis focused on two extreme cases of binders:

- 1- Fresh asphalt binder: Where all aging parameters are set to zero.
- 2- Long-term aged (LTA) binder: Subjected to maximum aging conditions.

Table 6 provides the specific input parameters for both fresh and LTA bitumen samples used in this comparison. Fig. 20 shows the prediction deviations for five DNN models alongside the hybrid DNN-GP model in estimating G^* and δ . For fresh binder, the largest prediction error was noted in DNN 1, approximately 22.7% for both G^* and δ . Conversely, the hybrid model exhibited significant enhancements, reducing errors to roughly 0.031% for G^* and 6.2% for δ .

In the case of long-term aged (LTA) binder, DNN 3 demonstrated the highest error rates, with 82.3% for G^* and 13.7% for δ . The hybrid DNN-GP model, however, showed a marked improvement, achieving relative errors of 12.7% for G^* and 1.4% for δ . These results highlight the potential advantages of integrating GPR with DNN in managing the complex aging processes of materials like asphalt binder.

A comparative study was also conducted to evaluate the improvements of the proposed hybrid model at the mastic scale. The analysis focused on a STA binder sample containing high stiffness fillers at low concentrations within the matrix. The input parameters of the mastic sample used to assess the precision of the hybrid model are listed in Table 7. Fig. 21 shows the prediction deviations for five DNN models compared to the hybrid DNN-GP model. Similar to the case of asphalt binder, the DNN+GP model demonstrates a significant improvement by reducing the error percentage for G^* by approximately 50% compared to DNN 2. The high errors and negative predictions in the standalone DNN stem from their sensitivity to the stochastic nature of training. The

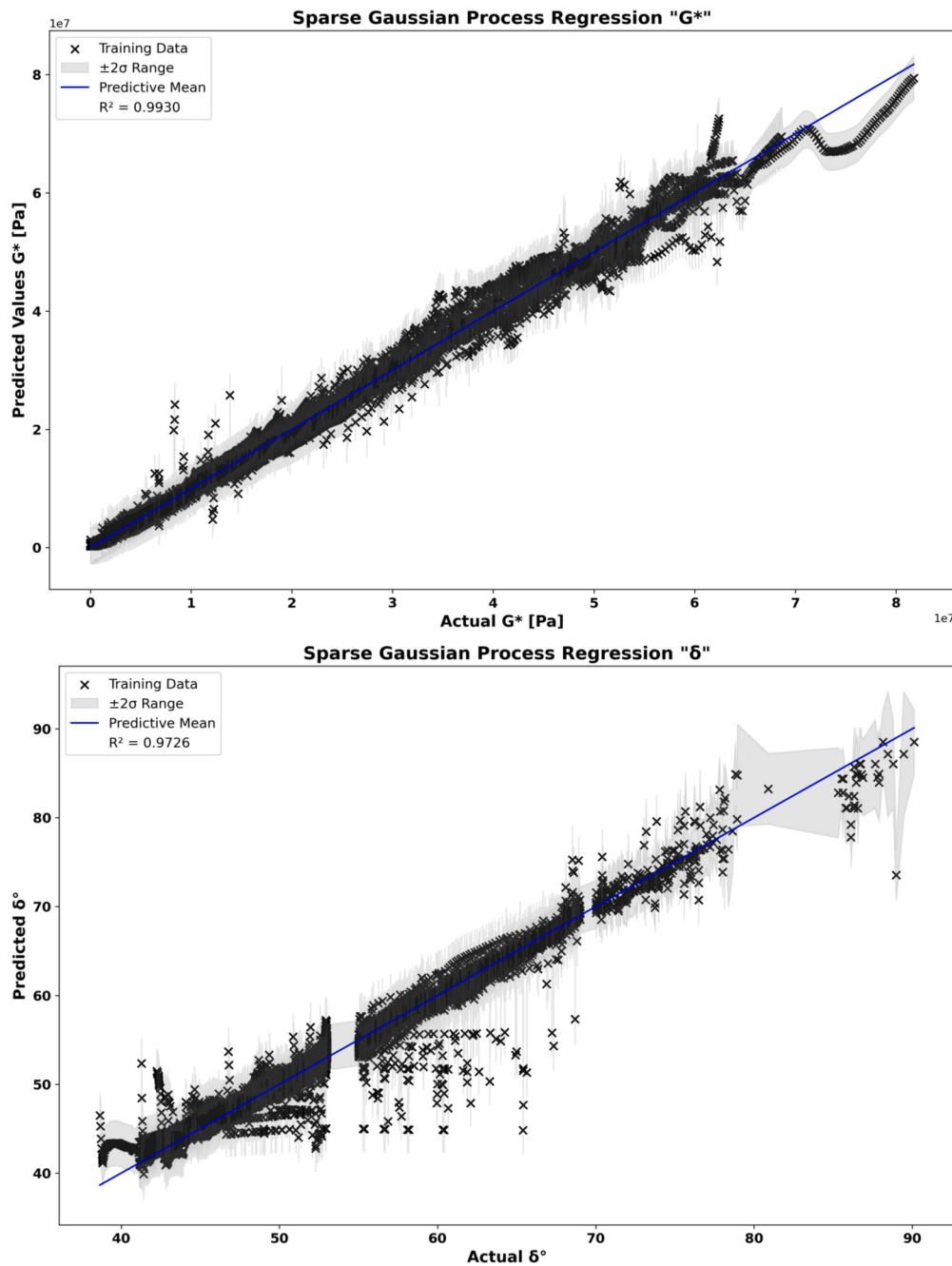


Fig. 17. Gaussian process model performance in predicting (G^*) and (δ) of asphalt mastic.

five DNN were trained independently with different random weight initializations. This variability can lead to outliers, including physically implausible negative values for G^* , which is a positive quantity that represents G^* . The hybrid DNN-GPR model mitigates these issues by using the GPR to refine the DNN predictions.

3.5. Feature contributions to prediction uncertainty

The feature importance results also suggest differential impacts tied to binder evolution, particularly through aging conditions. For instance, long-term aging (PAV) consistently ranks as a critical factor for δ across methods (Fig. 18), reflecting its pronounced effect on binder viscoelasticity as it transitions from fresh to aged states. This aligns with the higher uncertainty observed in standalone DNN predictions for aged

binders (e.g., Fig. 20), where chemical and physical changes amplify variability, subsequently reduced by the GP's probabilistic smoothing. While this study focuses on a generalized feature impact across binders, these trends hint at how aging-sensitive binders may exhibit greater dependence on environmental and test parameters, an aspect warranting further exploration with binder-specific datasets.

To further explain how specific features influence prediction uncertainty, Fig. 22 quantifies their contributions to the variance of the GPR model for G^* and δ . Features such as test temperature and long-term aging emerge as the primary drivers of uncertainty, consistent with their high importance (Fig. 18), reflecting their sensitivity to change binder state. Comparison with SHAP values reveals that while importance and uncertainty often align, features such as test frequency contribute more to uncertainty in δ than their overall importance suggests. This analysis

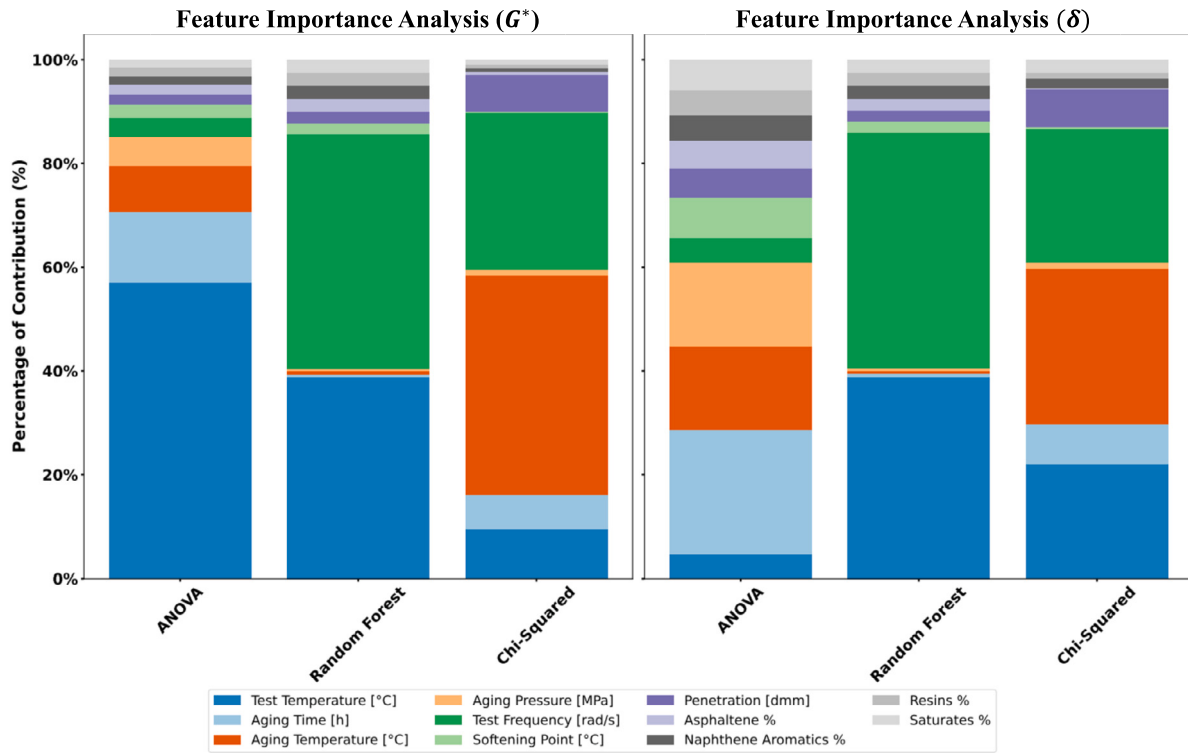


Fig. 18. Comparative Analysis of Feature Importance using ANOVA, Random forest and Chi-squared.

Table 7

Input parameters of the mastic sample used to evaluate the accuracy of the hybrid model.

Parameter	Mastic (STA Bitumen + Fillers)
Aging Temperature [°C]	163
Aging Time [h]	5
Aging Pressure [MPa]	0
Asphaltene (%)	14.6
Naphthene aromatics (%)	51.6
saturates (%)	3.6
resins (%)	30.2
Penetration [dmm]	70.2
Softening Point [°C]	53.76
Filler (%)	8
Filler Stiffness [GPa]	70
Test Temperature [°C]	0
Test Frequency [rad/s]	100

underscores the hybrid model's ability to enhance reliability by addressing feature-driven uncertainty in all conditions.

3.6. Limitations and future directions

The hybrid DNN-GPR model achieves high predictive accuracy for asphalt binder and mastic rheological properties, yet several limitations must be considered to fully understand its applicability.

First, the model was trained and validated using laboratory data from controlled experiments and standardized aging tests (RTFOT and PAV), achieving high R^2 values. However, the complexities of real-world field conditions are not fully represented in the existing datasets., including variable environmental factors (e.g., humidity, UV exposure), dynamic traffic loading, and long-term pavement performance. The standardized aging protocols simplify oxidative aging processes, which may not fully represent the nonlinear aging dynamics in in-service pavements. Consequently, the generalizability of the model to field conditions remains untested, and the predictions may deviate under diverse scenarios. Future research should validate the model against field data from oper-

ational pavements to ensure robustness in various environmental and loading conditions.

Secondly, the mastic-scale model heavily relies on synthetic data generated through FEM simulations. To improve reliability and better reflect real-world behavior, future studies should focus on collecting comprehensive experimental mastic datasets, thereby minimizing reliance on synthetic inputs and enhancing the ability of the model to capture diverse mastic responses.

Finally, the binder-scale model is constrained by a relatively small dataset compared to the mastic scale, which may limit the ability to capture the full spectrum of binder behaviors, particularly for modified or unconventional binders. The exclusion of viscosity as an input parameter, due to insufficient data availability, further restricts the model's ability to predict rheological behaviors influenced by this critical property.

4. Conclusions

This study introduces a hybrid modeling framework that combines DNN with GPR to predict G^* and δ of asphalt binders and mastics. The main findings are:

- The hybrid DNN-GPR model achieved R^2 values up to 0.997 for G^* and 0.972 for δ , outperforming standalone DNN models in terms of predictive accuracy.
- Finite Element Modeling (FEM) was effectively utilized to generate synthetic mastic data, addressing the scarcity of experimental datasets and enabling multiscale model training.
- Test temperature, aging conditions, and penetration were consistently identified as the most significant factors influencing rheological behavior, based on SHAP, ANOVA, random forest, and chi-squared analyses.
- The proposed model provides a reliable and interpretable tool for asphalt material design, with potential to improve the durability and sustainability of pavements.

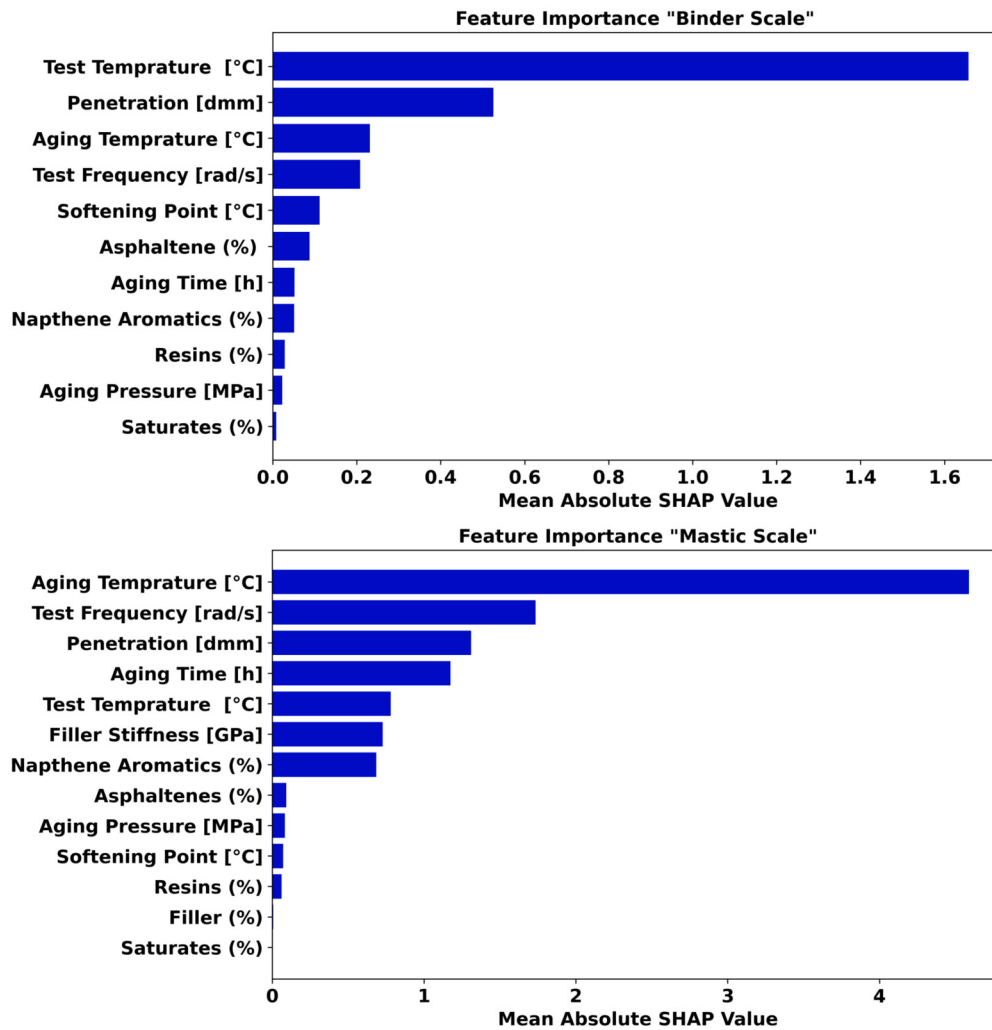


Fig. 19. Identifying the most important features affecting G^* and δ simultaneously using SHAP method.

- Future work should explore model performance under diverse field conditions and investigate binder-specific aging mechanisms as well as additional environmental variables.

CRedit authorship contribution statement

Mahmoud Khadijeh: Writing – original draft, Validation, Resources, Methodology, Investigation, Conceptualization. **Cor Kasbergen:** Supervision. **Sandra Erkens:** Supervision. **Aikaterini Varveri:** Writing – review & editing, Supervision, Methodology, Investigation.

Declaration of competing interest

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

Acknowledgement

This publication is part of the project ‘A multiscale approach towards future road infrastructure: How to design sustainable paving materials?’ (project number 18148) of the research program NWO Talent Programme Veni AES 2020, which is awarded to A.V. and financed by the Dutch Research Council (NWO).

Appendix A. Validation of FEM results

The FEM simulation incorporated several critical elements to accurately reproduce the experimental conditions. An implicit time step was utilized for the analysis, with a logarithmic frequency sweep ranging from 0.0159 to 16 Hz that replicates the experimental test conditions. Proper load transfer between model components was ensured through tie contacts. Additionally, a cohesive stiffness contact was established between the mastic sample and both the top and bottom plates.

The Genetic algorithm parameters were selected to balance thorough exploration of the solution space with computational efficiency [59]. Tables A.1 and A.2 summarize the key parameters used in this study, including the filler properties, binder properties, Genetic algorithm and Prony series parameters.

The FEM results are compared against the GSCM, and experimental test data. Two representative examples are provided: (1) a comparison of FEM and GSCM predictions across different filler percentages, and (2) a validation of FEM against experimental tests under fresh and long-term aged (LTA) conditions. More details on the model development and the input parameters used in these validations can be found in [60].

A.1. Validation against GSCM across different filler percentages

The GSCM is an analytical approach commonly used to predict the effective properties of composite materials, such as asphalt mastics, by assuming a homogeneous matrix with embedded inclusions. In this

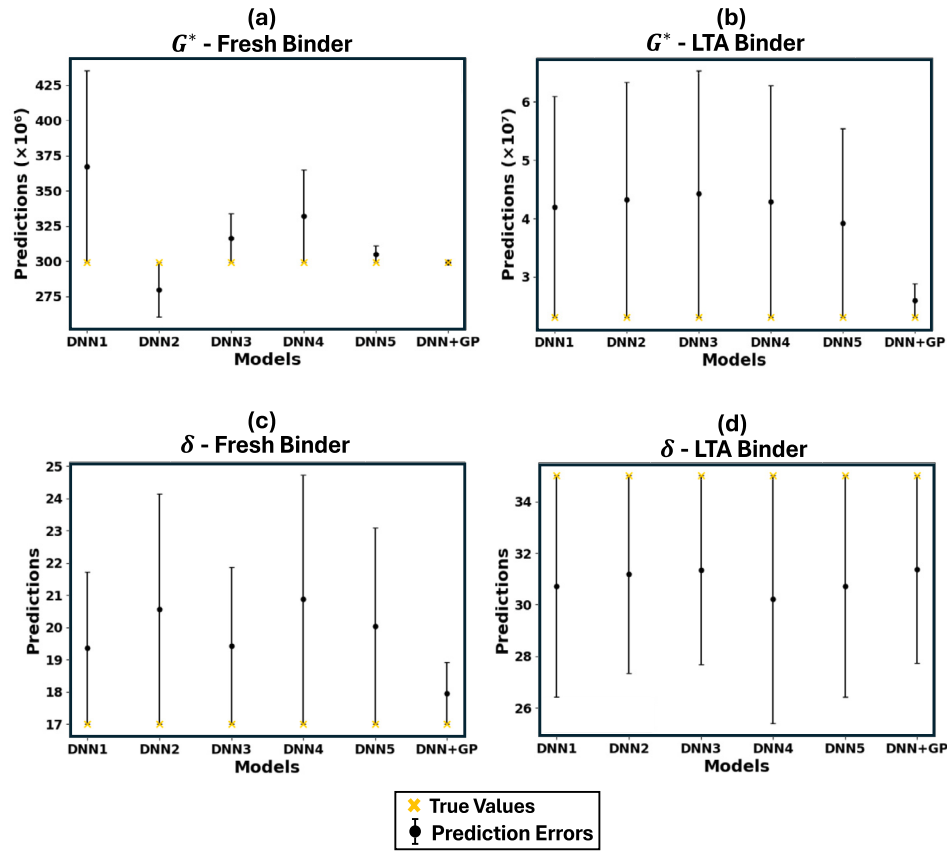


Fig. 20. Prediction deviations for five DNN models alongside the hybrid DNN-GP model in estimating G^* and δ for (a - b) fresh and (c - d) LTA asphalt binders.

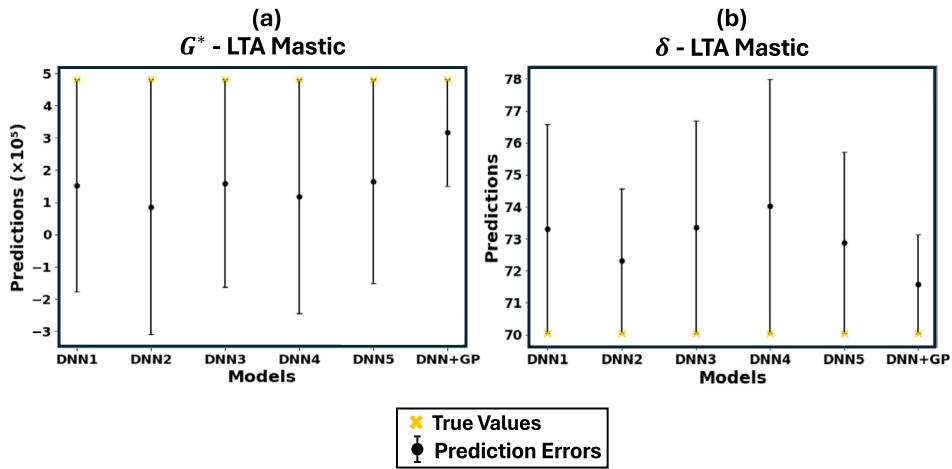


Fig. 21. Prediction deviations for five DNN models alongside the hybrid DNN-GP model in estimating (a) G^* and (b) δ for LTA asphalt mastic.

Table A.1

Parameters used in the FEM simulation.

Parameter	Value
Temperature [$^{\circ}\text{C}$]	20 $^{\circ}\text{C}$
Frequency [Hz]	[0.0159-16]
Population Size	15
Recombination	0.7
Mutation Rate	[0.5, 1]
Number of Prony Parameters	10
Fillers Stiffness [GPa]	70
Fillers [%]	[8-28]
Instantaneous Young's Modulus (E_0) [MPa]	240.5
Poisson's ratio (ν)	0.35

study, the FEM predictions of the G^* are compared with GSCM across a range of frequencies (0 to 16 Hz) and varying filler percentages.

Fig. A.10 illustrates the comparison for six different filler percentages, ranging from low to high volume fractions. The FEM results show a consistent trend with the GSCM predictions across all frequencies, particularly at lower filler percentages. However, as the filler percentage increases, the GSCM predictions deviate more from the FEM results, especially at higher frequencies. This deviation arises due to the limitations of the GSCM, which does not fully account for particle interactions at higher filler concentrations. At higher filler percentages, the interactions between filler particles and the binder become more significant, leading to a stiffer mastic response that the GSCM underestimates. The FEM, by contrast, explicitly models these interactions through the

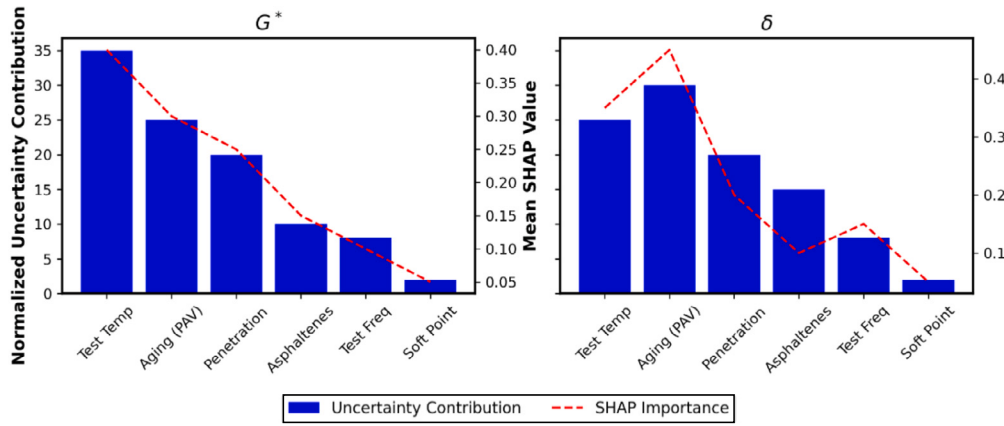


Fig. 22. Bar plot illustrating the normalized contribution of key input features to uncertainty in predicting G^* (left) and δ (right) using the hybrid DNN-GP model.

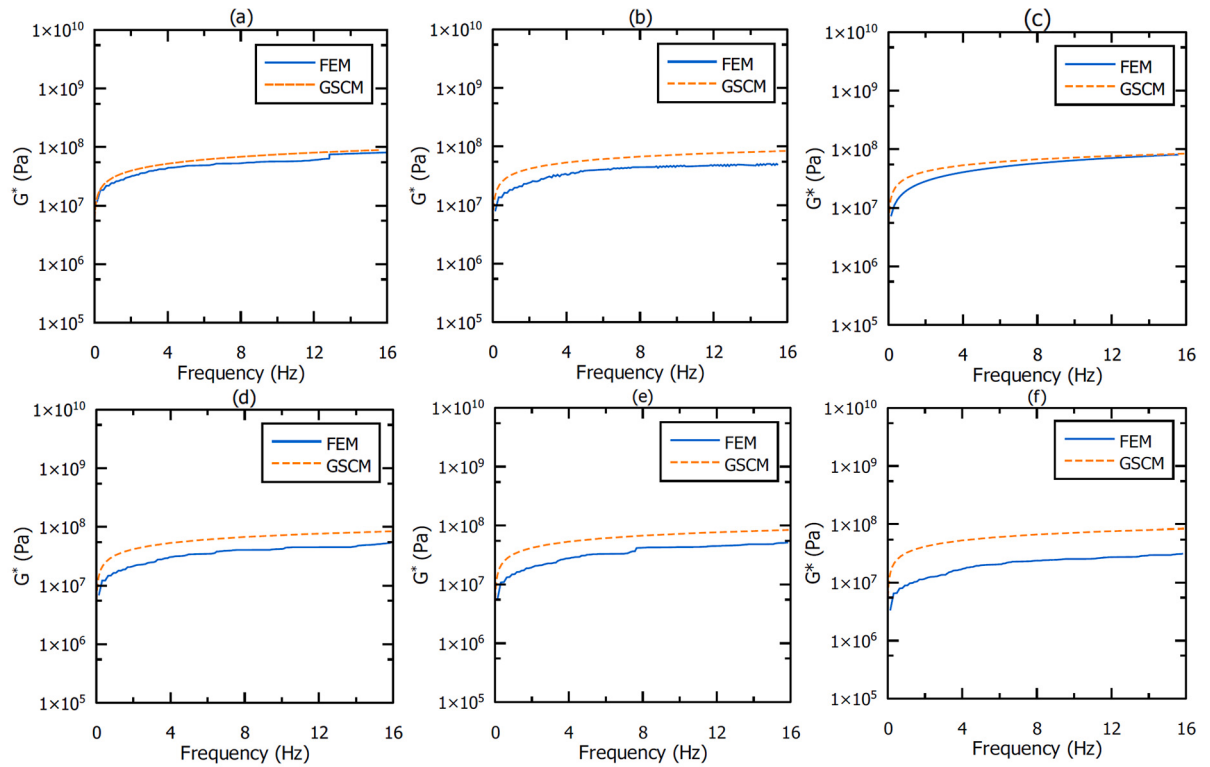


Fig. A.10. Comparison of FEM and GSCM predictions for complex shear modulus (G^*) across different frequencies and filler percentages. Originally published in [60] under a Creative Commons Attribution (CC BY) license.

Table A.2

Prony Series Parameters used in the FEM simulation.

i	Relaxation times (τ_i)	Weights (g_i)
1	0.004694686	0.093632628
2	0.115054235	0.117440582
3	0.012508845	0.097836541
4	0.004675019	0.098564454
5	0.033004199	0.093474197
6	0.039371971	0.034233612
7	0.004888031	0.157408419
8	4.005918554	0.051016550
9	0.004949636	0.149722400
10	0.568111227	0.086711249

viding a more accurate representation of the mastic's behavior under such conditions.

A.2. Validation against experimental tests under fresh and LTA conditions

The FEM predictions are further validated by comparing them with experimental test data obtained from the DSR tests conducted on asphalt mastic samples under two aging conditions: fresh (unaged) and long-term aged (LTA). Fig. A.11 (a) and (b) present the comparison of G^* predictions across a frequency range of 0 to 20 Hz for these two conditions.

In Fig. A.11, the fresh mastic sample (aging parameters set to zero) shows excellent agreement between the FEM predictions and the experimental data. The FEM accurately captures the viscoelastic behavior of the mastic, with G^* increasing steadily with frequency, reflecting the stiffening response. Similarly, in Fig. A.11 (b), the LTA binder sample demonstrates strong alignment between the FEM and experimental re-

Maxwell viscoelastic model and the random distribution of fillers, pro-

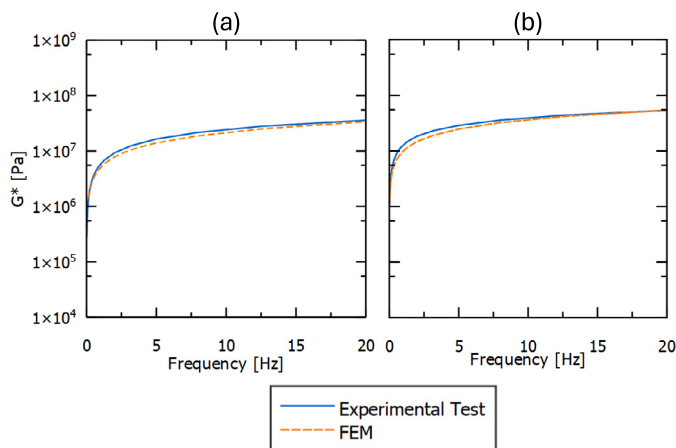


Fig. A.11. Comparison of experimental and FEM results for G^* of asphalt mastic at 20 °C. (a) Fresh asphalt mastic, (b) LTA asphalt mastic.

sults. The LTA sample exhibits a higher G^* due to the increased stiffness from oxidative aging, a trend that the FEM effectively replicates across the frequency range.

The close match between the FEM and the experimental data in both fresh and LTA conditions underscores the reliability of the FEM in capturing the rheological behavior of asphalt mastic under varying aging states. The use of the Maxwell model in ABAQUS, combined with Prony series parameters to represent the viscoelastic response, enables the FEM to accurately simulate the time-dependent behavior even as its properties evolve due to aging.

Appendix B. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.rineng.2025.105629>.

Data availability

The FEM results that support the findings of this study are available from the corresponding author upon reasonable request.

References

- [1] J.G. Speight, *Asphalt Materials Science and Technology*, Springer, 2016.
- [2] R. Jing, A. Varveri, X. Liu, A. Scarpas, S. Erkens, Laboratory and field aging effect on bitumen chemistry and rheology in porous asphalt mixture, *Transp. Res. Rec.* 2673 (3) (2019) 365–374, <https://doi.org/10.1177/0361198119833362>.
- [3] A. Nasertork, S. Ranjbar, M. Rahai, F. Moghadas Nejad, Pavement raveling inspection using a new image texture-based feature set and artificial intelligence, *Adv. Eng. Inform.* 62 (2024) 102665, <https://doi.org/10.1016/j.aei.2024.102665>.
- [4] G.D. Airey, Use of black diagrams to identify inconsistencies in rheological data, *Road Mater. Pavement Des.* 3 (4) (2002) 403–424, <https://doi.org/10.1080/14680629.2002.9689933>.
- [5] N.I.M. Yusoff, M.T. Shaw, G.D. Airey, Modelling the linear viscoelastic rheological properties of bituminous binders, *Constr. Build. Mater.* 25 (5) (2011) 2171–2189, <https://doi.org/10.1016/j.conbuildmat.2010.11.086>.
- [6] W.A. Ogbon, W. Jiang, D. Yuan, J. Xiao, C. Xing, S.C.J. Moya, Comparative analysis of several short-term ageing simulation methods and their impact on long-term aging performance of asphalt binders, *Constr. Build. Mater.* 437 (2024) 136969, <https://doi.org/10.1016/j.conbuildmat.2024.136969>.
- [7] B. Hofko, A. Cannone Falchetto, J. Grenfell, L. Huber, X. Lu, L. Porot, L. Poulikakos, Z. You, Effect of short-term ageing temperature on bitumen properties, *Road Mater. Pavement Des.* 18 (sup2) (2017) 108–117, <https://doi.org/10.1080/14680629.2017.1304268>.
- [8] A. Ghasemirad, N. Bala, L. Hashemian, High-temperature performance evaluation of asphaltene-modified asphalt binders, *Molecules* 25 (15) (2020), <https://doi.org/10.3390/molecules25153326>.
- [9] N.A.A. Mazalan, M.K.I. Mohd Satar, A. Mohamed, M.N. Mohd Warid, Rheological properties of asphaltene-modified asphalt binder and mastic, *Phys. Chem. Earth Parts A/B/C* 131 (2023) 103422, <https://doi.org/10.1016/j.pce.2023.103422>.
- [10] F. Mastoras, A. Varveri, M. van Tooren, S. Erkens, Effect of mineral fillers on ageing of bituminous mastics, *Constr. Build. Mater.* 276 (2021) 122215, <https://doi.org/10.1016/j.conbuildmat.2020.122215>.
- [11] M. Steineder, B. Hofko, Assessing the impact of filler properties, moisture, and aging regarding fatigue resistance of asphalt mastic, *Road Mater. Pavement Des.* 24 (12) (2023) 2811–2826, <https://doi.org/10.1080/14680629.2023.2172066>.
- [12] A. Akbari, A. Akbari, A. Ghanbari, S. Hesami, Investigating the influence of aging and filler type on the fatigue behavior of bitumen mastics, *Constr. Build. Mater.* 269 (2021) 121254, <https://doi.org/10.1016/j.conbuildmat.2020.121254>.
- [13] N. Mukhtar, M.R. Mohd Hasan, M.F.H. Mohd Ghazali, Z. Mohd Zin, K.A. Shariff, A. Sani, Influence of concentration and packing of filler particles on the stiffening effect and shearing behaviour of asphalt mastic, *Constr. Build. Mater.* 295 (2021) 123660, <https://doi.org/10.1016/j.conbuildmat.2021.123660>.
- [14] G. E. J. Zhang, Q. Shen, P. Ji, J. Wang, Y. Xiao, Influence of filler type and rheological properties of asphalt mastic on the asphalt mastic–aggregate interaction, *Materials* 16 (2) (2023), <https://doi.org/10.3390/ma16020574>.
- [15] Q. Wang, D. Wu, G. Li, Z. Liu, J. Tong, X. Chen, W. Gao, Machine learning aided uncertainty quantification for engineering structures involving material-geometric randomness and data imperfection, *Comput. Methods Appl. Mech. Eng.* 423 (2024) 116868, <https://doi.org/10.1016/j.cma.2024.116868>.
- [16] K. Veasna, Z. Feng, Q. Zhang, M. Knezevic, Machine learning-based multi-objective optimization for efficient identification of crystal plasticity model parameters, *Comput. Methods Appl. Mech. Eng.* 403 (2023) 115740, <https://doi.org/10.1016/j.cma.2022.115740>.
- [17] I.N. Usanga, C.C. Ikeagwuani, R.K. Etim, I.C. Attah, Predictive modeling of modified asphalt mixture rutting potentials: machine learning approach, *Iran. J. Sci. Technol. Trans. Civ. Eng.* 47 (6) (2023) 4087–4101, <https://doi.org/10.1007/s40996-023-01192-w>.
- [18] J. Rivera-Pérez, I.L. Al-Qadi, Asphalt concrete mix design optimization using autoencoder deep neural networks, *Transp. Res. Rec.* 2678 (1) (2024) 426–438, <https://doi.org/10.1177/03611981231171153>.
- [19] A. Hamid, H. Baaj, M. El-Hakim, Predicting the recovery and nonrecoverable compliance behaviour of asphalt binders using artificial neural networks, *Processes* 10 (2022) 1–20, <https://doi.org/10.3390/pr10122633>.
- [20] M. Alas, S. Ali, Prediction of the high-temperature performance of a geopolymer modified asphalt binder using artificial neural networks, *Int. J. Technol.* 10 (2019) 417–427, <https://doi.org/10.14716/ijtech.v10i2.2421>.
- [21] K. Yan, L. You, Investigation of complex modulus of asphalt mastic by artificial neural networks, *Indian J. Eng. Mater. Sci.* 21 (2014) 445–450.
- [22] A. Seitlari, Y.S. Kumbargeri, K.P. Biligiri, I. Boz, A soft computing approach to predict and evaluate asphalt mixture aging characteristics using asphaltene as a performance indicator, *Mater. Struct.* 52 (5) (2019) 100, <https://doi.org/10.1617/s11527-019-1402-5>.
- [23] J. Houlik, J. Valentin, V. Nežerka, Predicting the fatigue life of asphalt concrete using neural networks, *arXiv preprint arXiv:2406.01523*, 2024, <https://doi.org/10.48550/arXiv.2406.01523>.
- [24] A. Venkatraman, D. Montes de Oca Zapiain, H. Lim, S.R. Kalidindi, Texture-sensitive prediction of micro-spring performance using Gaussian process models calibrated to finite element simulations, *Mater. Des.* 197 (2021) 109198, <https://doi.org/10.1016/j.matdes.2020.109198>.
- [25] X. Xu, Y. Zhang, Price forecasts of ten steel products using Gaussian process regressions, *Eng. Appl. Artif. Intell.* 126 (2023) 106870, <https://doi.org/10.1016/j.engappai.2023.106870>.
- [26] H. Zhang, Y. Wu, S. Yang, Probabilistic analysis of tunnel convergence in spatially variable soil based on Gaussian process regression, *Eng. Appl. Artif. Intell.* 131 (2024) 107840, <https://doi.org/10.1016/j.engappai.2023.107840>.
- [27] M. Bauer, M. van der Wilk, C.E. Rasmussen, Understanding probabilistic sparse Gaussian process approximations, in: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, vol. 29, Curran Associates, Inc., 2016.
- [28] A.R. Ghanizadeh, N. Heidarabadizadeh, F. Heravi, Gaussian process regression (gpr) for auto-estimation of resilient modulus of stabilized base materials, *J. Soft Comput. Civ. Eng.* 5 (1) (2021) 80–94, <https://doi.org/10.22115/SCCE.2021.269187.1273>.
- [29] S. Girimath, D. Singh, E. Manthos, W. Mampearachchi, Effects of reclaimed asphalt binder on rheological properties and cohesion energy of crumb rubber modified binder, *Innov. Infrastruct. Solut.* 3 (2018) 1–8, <https://doi.org/10.1007/s41062-018-0164-1>.
- [30] Y. Li, F. Xu, Y. Wang, H. Liu, L. Peng, Y. Xiao, Q. Liang, X. Li, Study on viscoelastic properties of various fiber-reinforced asphalt binders, *Materials* 17 (5) (2024), <https://doi.org/10.3390/ma17051085>.
- [31] S. Ren, X. Liu, P. Lin, R. Jing, S. Erkens, Toward the long-term aging influence and novel reaction kinetics models of bitumen, *Int. J. Pavement Eng.* 24 (2) (2023) 2024188, <https://doi.org/10.1080/10298436.2021.2024188>.
- [32] G. Pipintakos, H. Soenen, H.V. Ching, C.V. Velde, S.V. Doorslaer, F. Lemièrre, A. Varveri, W. Van den bergh, Exploring the oxidative mechanisms of bitumen after laboratory short- and long-term ageing, *Constr. Build. Mater.* 289 (2021) 123182, <https://doi.org/10.1016/j.conbuildmat.2021.123182>.
- [33] D.R. Jones, et al., *SHRP Materials Reference Library: Asphalt Cements: A Concise Data Compilation*, Strategic Highway Research Program, vol. 1, National Research Council, Washington, DC, 1993.

- [34] E. Publications, IP 469: Determination of Saturated, Aromatic and Polar Compounds in Petroleum Products by Thin Layer Chromatography and Flame Ionization Detection, 2016.
- [35] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: large-scale machine learning on heterogeneous systems, software available from <https://www.tensorflow.org/>, 2015.
- [36] J. Schmidhuber, Deep learning in neural networks: an overview, *Neural Netw.* 61 (2015) 85–117, <https://doi.org/10.1016/j.neunet.2014.09.003>.
- [37] P. Refaailzadeh, L. Tang, H. Liu, *Cross-Validation*, Springer US, Boston, MA, 2009.
- [38] A.F. Agarap, Deep learning using rectified linear units (relu), arXiv preprint arXiv:1803.08375, 2018, <https://doi.org/10.48550/arXiv.1803.08375>.
- [39] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014, <https://doi.org/10.48550/arXiv.1412.6980>.
- [40] GPy, GPy: a Gaussian process framework in python, <http://github.com/SheffieldML/GPy>, 2012.
- [41] A.C. Ogren, B.T. Feng, K.L. Bouman, C. Daraio, Gaussian process regression as a surrogate model for the computation of dispersion relations, *Comput. Methods Appl. Mech. Eng.* 420 (2024) 116661, <https://doi.org/10.1016/j.cma.2023.116661>.
- [42] J.N. Fuhg, M. Marino, N. Bouklas, Local approximate Gaussian process regression for data-driven constitutive models: development and comparison with neural networks, *Comput. Methods Appl. Mech. Eng.* 388 (2022) 114217, <https://doi.org/10.1016/j.cma.2021.114217>.
- [43] M. Smith, ABAQUS/Standard User's Manual, Version 6.9, Dassault Systèmes Simulia Corp, United States, 2009.
- [44] M.H.R. Ghoreishy, Determination of the parameters of the Prony series in hyper-viscoelastic material models using the finite element method, *Mater. Des.* 35 (2012) 791–797, <https://doi.org/10.1016/j.matdes.2011.05.057>.
- [45] K. Pearson, X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Lond. Edinb. Dublin Philos. Mag. J. Sci.* 50 (302) (1900) 157–175, <https://doi.org/10.1080/14786440009463897>.
- [46] A.W. Edwards, R.A. Fischer, statistical methods for research workers, (1925), in: *Landmark Writings in Western Mathematics 1640–1940*, Elsevier, 1925, pp. 856–870.
- [47] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32, <https://doi.org/10.1023/A:1010933404324>.
- [48] Y. Xiao, M. Jin, G. Qi, W. Shi, K.X. Li, X. Du, Interpreting the influential factors in ship detention using a novel random forest algorithm considering dataset imbalance and uncertainty, *Eng. Appl. Artif. Intell.* 133 (2024) 108369, <https://doi.org/10.1016/j.engappai.2024.108369>.
- [49] T. Yan, X. Xing, T. Xia, D. Wang, Relation between fault characteristic frequencies and local interpretability Shapley additive explanations for continuous machine health monitoring, *Eng. Appl. Artif. Intell.* 136 (2024) 109046, <https://doi.org/10.1016/j.engappai.2024.109046>.
- [50] C. Joo, H. Park, J. Lim, H. Cho, J. Kim, Machine learning-based heat deflection temperature prediction and effect analysis in polypropylene composites using cat-boost and Shapley additive explanations, *Eng. Appl. Artif. Intell.* 126 (2023) 106873, <https://doi.org/10.1016/j.engappai.2023.106873>.
- [51] S.M. Lundberg, S. Lee, A unified approach to interpreting model predictions, *CoRR*, arXiv:1705.07874, 2017, arXiv:1705.07874, <https://doi.org/10.48550/arXiv.1705.07874>.
- [52] M.V. García, J.L. Aznarte, Shapley additive explanations for no2 forecasting, *Ecol. Inform.* 56 (2020) 101039.
- [53] J. Li, K. Cheng, S. Wang, F. Morstatter, R.P. Trevino, J. Tang, H. Liu, Feature selection: a data perspective, *ACM Comput. Surv.* 50 (6) (2017) 1–45, <https://doi.org/10.1145/3136625>.
- [54] J. Cai, J. Luo, S. Wang, S. Yang, Feature selection in machine learning: a new perspective, *Neurocomputing* 300 (2018) 70–79, <https://doi.org/10.1016/j.neucom.2017.11.077>.
- [55] A.-C. Haury, P. Gestraud, J.-P. Vert, The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures, *PLoS ONE* 6 (12) (2011) e28210, <https://doi.org/10.1371/journal.pone.0028210>.
- [56] E.O. Abiodun, A. Alabdulatif, O.I. Abiodun, M. Alawida, A. Alabdulatif, R.S. Alkhawaldeh, A systematic review of emerging feature selection optimization methods for optimal text classification: the present state and prospective opportunities, *Neural Comput. Appl.* 33 (22) (2021) 15091–15118.
- [57] J. Jiang, X. Guo, P. Hao, X. Guo, S. Huang, K. Chen, Effects of sara fractions on the rheological properties and microscopic behavior of hard asphalt, *Constr. Build. Mater.* 438 (2024) 137055, <https://doi.org/10.1016/j.conbuildmat.2024.137055>.
- [58] J. Wang, T. Wang, X. Hou, F. Xiao, Modelling of rheological and chemical properties of asphalt binder considering sara fraction, *Fuel* 238 (2019) 320–330, <https://doi.org/10.1016/j.fuel.2018.10.126>.
- [59] M. Khadijeh, A. Varveri, C. Kasbergen, S. Erkens, Prony series fit for viscoelastic data using a genetic algorithm, <https://doi.org/10.5281/zenodo.13735908>, Sep. 2024.
- [60] M. Khadijeh, C. Kasbergen, S. Erkens, A. Varveri, Exploring the roles of numerical simulations and machine learning in multiscale paving materials analysis: applications, challenges, best practices, *Comput. Methods Appl. Mech. Eng.* 433 (2025) 117462, <https://doi.org/10.1016/j.cma.2024.117462>.