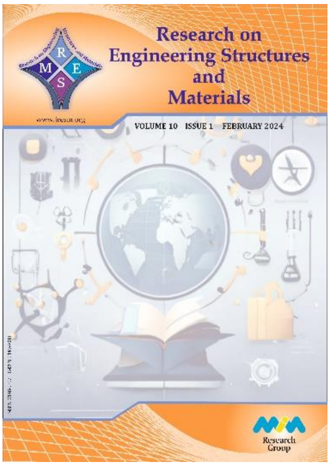
Research on Engineering Structures & Materials
www.jresm.org
Research Group



Predicting flexural-creep stiffness in bending beam rheometer (BBR) experiments using advanced super learner machine learning techniques

Alireza Roshan, Magdy Abdelrahman

Online Publication Date: 09 February 2024
URL: http://www.jresm.org/archive/resm2024.58me1027rs.html
DOI: http://dx.doi.org/10.17515/resm2024.58me1027rs

Journal Abbreviation: *Res. Eng. Struct. Mater.*

To cite this article
Roshan A, Abdelrahman M. Predicting flexural-creep stiffness in bending beam rheometer (BBR) experiments using advanced super learner machine learning techniques. *Res. Eng. Struct. Mater.*, 2024; 10(3): 1195-1208.

Research on Engineering Structures & Materials
www.jresm.org
MIM Research Group

Research Article

Predicting flexural-creep stiffness in bending beam rheometer (BBR) experiments using advanced super learner machine learning techniques

Alireza Roshan[*,1,a], Magdy Abdelrahman[2,b]

[1]Department of Civil, Architectural and Environmental Engineering, Missouri University of Science and Technology, Rolla, MO 65409, U.S.A
[2]Missouri Asphalt Pavement Association (MAPA) Endowed Professor, Department of Civil, Architectural and Environmental Engineering, Missouri University of Science and Technology, Rolla, MO 65409, U.S.A

Article Info

Article history:

Received 27 Oct 2023
Accepted 09 Feb 2024

Keywords:

Flexural-creep stiffness; Bending beam rheometer; Long-term pavement performance; Machine learning super learner methods

Abstract

BBR test is commonly used to assess the low-temperature performance grade (PG) of asphalt binders, with the flexural-creep stiffness being a critical parameter calculated through this test. However, it has notable limitations that demand attention. The significant amount of asphalt binder needed for test specimens increases costs and resource consumption. Additionally, the complex and time-consuming specimen preparation process hinders testing efficiency and introduces result variability, affecting the accuracy and reliability of PG determinations. In recent years, machine learning (ML) has emerged as a promising substitute for predicting various engineering values. In this study, the primary focus was on harnessing super learner (SL) techniques to predict the creep stiffness of asphalt binders. The SL approach combines multiple ML algorithms to enhance predictive accuracy and reduce individual model biases. Bagging, boosting, and stacking algorithms were employed in the construction of these prediction models. To conduct the investigation, data from 1350 samples sourced from the Long-Term Pavement Performance (LTPP) website were utilized to explore the influence of six crucial variables on the creep stiffness of asphalt binders. The proposed method demonstrated high accuracy, nearing 90% in the coefficient of determination. The Stacking model achieved a low Mean Absolute Percentage Error of 2.86% and robust Prediction Accuracy of 97.14% for randomly selected data points. Furthermore, the sensitivity analysis highlighted the significance of distinct input variables in influencing the creep stiffness of asphalt binders. Notably, the test temperature emerged as the most influential factor affecting creep stiffness, according to the conducted study.

1. Introduction

The asphalt pavements undergo significant impacts from climatic changes, which, in turn, profoundly impact the performance of asphalt mixtures and the characteristics of asphalt binder. As temperatures fluctuate, these materials experience significant transformations that directly influence their properties and response to external stresses. As temperatures decrease, the asphalt binder experiences a critical transition from a plastic state to a solid state. In colder conditions, the binder becomes much stiffer and exhibits a lower resistance to flow, resulting in increased viscosity. This change in viscosity affects the overall flexibility and resilience of the asphalt mixture, making it less capable of accommodating dynamic loads and stresses induced by traffic. The reduced flexibility and increased stiffness of the asphalt binder in colder climates can lead to several performance issues for

*Corresponding author: alireza.roshan@mst.edu
[a]orcid.org/0000-0002-9399-3501; [b]orcid.org/0000-0002-8722-0203
DOI: http://dx.doi.org/10.17515/resm2024.58me1027rs

the asphalt pavement. It becomes more susceptible to cracking and fracture due to its limited ability to absorb the energy from vehicular loads. Additionally, the reduced flowability of the binder makes it challenging for the pavement to self-heal or recover from minor damages caused by traffic, further contributing to the overall deterioration of the pavement structure [1].

When asphalt pavements experience elevated tensile stress surpassing the movement resistance of the asphalt binder, it results in the formation of cracks within the binder, which eventually spread across the pavement surface. This phenomenon is known as low-temperature cracking, leading to significant functional and structural failures in the pavements [2]. To study and address these cracks that may occur during the asphalt pavement's design life, a specialized test is conducted on the asphalt binder material. This specific test is conducted to assess the creep stiffness value at low temperatures, serving as a crucial indicator of the asphalt binder's resistance to low-temperature cracking. The creep stiffness measurement provides valuable insights into how well the binder can withstand tensile stresses caused by low temperatures without undergoing excessive deformation or cracking. This property is crucial for ensuring the long-term performance and durability of asphalt pavements, especially in regions with colder climates or significant temperature variations [3]. To assess the asphalt binder's resistance to low-temperature cracking, a specialized test is conducted using an asphalt binder prismatic beam. The beam possesses specific dimensions, measuring 125 mm in length, 6.25 mm in height, and 12.5 mm in width. In the test, the prismatic beam is horizontally positioned in a cold fluid bath, and a constant load of 980 mN is applied at the midpoint of the beam. The deflection of the specimen is measured, and the creep stiffness is calculated using the actual load and specimen dimensions. The test is performed at various constant temperatures within the low-temperature range. At a loading time of 60 seconds, the creep stiffness and the corresponding m-value are derived. The m-value represents the absolute value of the slope of the stiffness versus time curve on a double logarithmic scale [4].

BBR has become the predominant testing method for assessing the low-temperature characteristics of asphalt binders, particularly those subjected to prolonged aging before undergoing BBR testing. Despite its widespread use, several authors have highlighted significant drawbacks associated with BBR testing, particularly concerning specimen preparation and testing conditions. One of the main drawbacks is the complexity and labour-intensive nature of preparing BBR test specimens. The process requires meticulous attention to detail and precise measurements to ensure accurate results. This can be time-consuming and may introduce variability in the test outcomes due to inconsistencies in specimen preparation. Additionally, the testing conditions in BBR may not always fully represent the real-world environmental conditions that the asphalt binders encounter in the field. For instance, the test temperatures in BBR may not accurately mimic the wide range of temperature fluctuations experienced by pavements in different geographic locations and climates. This limitation can impact the relevance and applicability of BBR results to actual pavement performance [5-7]. The direct grading process to determine the PG of asphalt pavements involves using devices like the Dynamic Shear Rheometer (DSR) and BBR to directly measure pavement performance. However, this approach requires a substantial budget, presenting challenges for researchers and pavement laboratories in certain countries. To address these limitations, researchers have explored alternative methods for indirectly assessing pavement performance, which are more accessible and cost-effective. One such indirect estimation method involves leveraging weather data, particularly the maximum and minimum air temperatures in construction regions, as a basis for assessing the PG grade. By utilizing weather data, researchers can infer the performance characteristics of the pavement under varying temperature conditions. Additionally, some researchers have proposed an indirect estimation of the PG grade based on pavement characteristics under "real-world conditions" while adhering to the specifications of traditional grading systems used in developed countries. This approach aims to find a balance between cost-effectiveness and accuracy in assessing the PG grade [8]. During the Strategic Highway Research Program (SHRP), there was a consideration to use the DSR with parallel plate geometry for the low-temperature PG system. However, it was eventually not selected for this purpose due to a significant challenge. Researchers have acknowledged that DSR measurements at temperatures below approximately 5 °C led to compliance errors in dynamic responses, particularly evident when employing the standard thin film binder geometry. To overcome this limitation and accurately measure the low-temperature rheological properties of asphalt binder, SHRP developed the BBR. The BBR test was specifically designed to assess the stiffness and creep behaviour of asphalt binders under low-temperature conditions. By using the BBR, researchers were able to obtain more reliable and consistent data related to the performance of asphalt binders at lower temperatures [9]. Recognizing the challenges and limitations associated with the BBR for low-temperature binder evaluation, some researchers have sought an alternative approach. As a result, they have endeavored to shift from BBR-based testing to a method exclusively reliant on DSR. In this new approach, the DSR is utilized for these evaluations, three distinct geometries were employed: torsion bar, 8-mm Parallel Plate, and 4-mm Parallel Plate. However, it is important to highlight that testing asphalt binder at extremely low temperatures, such as -30°C, can only be accomplished using the 4-mm plate geometry. This limitation arises from the torque capacity constraints of motors in typical commercially available DSRs. The use of the 4-mm plate geometry allows researchers to accurately measure the rheological properties of asphalt binders under these extreme low-temperature conditions, enabling a more comprehensive assessment of their performance in challenging environments [10-12]. Recently, the approach to determining the BBR equivalent low performance grade has centered on converting the complex shear modulus ($G^*(\omega)$) to creep compliance ($D(t)$). This conversion involves transforming data acquired in the DSR frequency domain into the BBR time domain. Various interconversion methods have been employed in these studies, and they are grounded in linear viscoelastic theory. On the other hand, approximation-based methods offer simplified procedures to approximate the interconversion from DSR frequency domain to BBR time domain [13,14]. Indeed, while the approximation-based interconversion methods may not offer the same level of precision as rigorous methods in recent years, there has been a growing trend among researchers to explore the application of ML for evaluating the rheological parameters of asphalt binders.

ML techniques, such as regression, neural networks, and ensemble methods, have shown great promise in various engineering applications due to their ability to handle complex datasets and identify patterns that might not be easily discernible through traditional methods. In a research study, artificial neural network and self-validated ensemble modeling techniques were used to predict low-temperature fracture energy of asphalt mixtures and both methods showed high prediction accuracy [15]. In another study, new predictive models were developed to estimate the dynamic modulus and phase angle of asphalt concrete accurately. The models considered temperature and loading frequency as key factors, and statistical analysis revealed their effectiveness in providing precise estimations for these properties [16]. In other study, the Extreme Learning Machine (ELM) algorithm, optimized by Genetic Algorithm (GA), was employed to rapidly predict the low-temperature rheological properties of styrenic block copolymer modified asphalt based on the raw material properties. The GA-ELM model outperformed traditional models, reducing errors by 68.97-81.48% [17]. Another research introduced a data-driven Convolutional Neural Network (CNN) model to forecast the phase angle behavior of asphalt concrete mixtures. The proposed CNN model achieves an impressive R^2 score of 0.90, indicating high accuracy in its predictions [18].

From the existing literature, it can be observed that ML methods have not been extensively used to predict the creep stiffness, considering various factors such as test temperature, penetration, kinematic viscosity, and absolute viscosity (dynamic) also the occurrence of physical hardening during the storage of BBR specimens at constant low temperatures has been observed in multiple studies [19-20]. Because physical hardening changes the rheological properties of asphalt binders the inclusion of the mentioned parameters contributes to the attainment of more precise predictions of creep stiffness prediction. Additionally, most of the mentioned algorithms in previous studies are individual learning algorithms, whereas new SL techniques, which are more accurate, powerful, and robust, are gaining popularity. The core concept of SL techniques involves training multiple weak learners with the training data and then combining them to create a strong learner. These weak learners are based on individual learning algorithms. As a result, group learning models (strong learners) significantly enhance prediction accuracy and model robustness. Three primary groups of algorithms for group learning are bagging, boosting, and stacking, and their distinctions can be found in a review article [21]. Great potential for improving the prediction accuracy and reliability of creep stiffness in asphalt binders is offered through the utilization of these advanced SL methods. This is attributed to the presence of non-linearities and interactions between various factors on asphalt binders. In high-dimensional feature spaces, a preference for nonlinear models such as bagging and boosting may arise for feature selection, regularization, and prediction, while simpler models like linear regression may suffice for fewer features [22].

The primary aim of this research is to develop a predictive model capable of forecasting creep stiffness in BBR Experiments using data collected from the LTPP dataset. The proposed model incorporates various factors that influence the rheological properties of the asphalt binder, primarily related to test conditions and binder properties. These influential factors include test temperature, penetration, kinematic viscosity, absolute viscosity (dynamic viscosity) and specific gravity. The model is constructed by training SL techniques with the collected data, resulting in a strong learner that can accurately predict the value of creep stiffness. Additionally, the research investigates the impact of key factors in the best model approach, such as the amount of training data, sensitivity, and the number of input variables. This comprehensive approach seeks to enhance the accuracy and reliability of predicting creep stiffness, contributing to a better understanding of asphalt binder behavior, and facilitating the design of more durable and resilient asphalt pavements.

2. Methods

ML techniques have proven highly advantageous in civil engineering, providing rapid and precise outcomes with minimal error rates. These ML methods can be categorized into three main groups: supervised learning, unsupervised learning (including clustering algorithms and Principal Component Analysis), and reinforcement learning. Supervised learning involves feeding the algorithm substantial volumes of labeled data containing input and output variables. By identifying patterns and learning from observations, the algorithm generates predictions until the error reaches an acceptable level. This type of learning can be further divided into two categories: classification and regression [23-24]. The diverse range of ML methods empowers civil engineering researchers to enhance efficiency and accuracy in their analyses and decision-making processes. In the context of predicting creep stiffness, a model is constructed using SL techniques trained with the collected data. This results in a robust learner capable of accurately forecasting creep stiffness values. Additionally, the research explores the impact of key factors in the best model approach, such as the amount of training data, sensitivity, and the number of input variables.

In contrast to the less powerful regression models previously used, this research focuses on harnessing the predictive capabilities of the SL approach to forecast creep stiffness in the BBR test. By combining and optimizing different base learners, the SL models aim to improve the accuracy and reliability of predictions for this important rheological property of asphalt binders. In this study, SL models were developed based on several ensemble methods, including Random Forest, Gradient Boosting Machine (GBM), Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), Categorical gradient Boosting (CatBoost), and stacking, to estimate the Creep stiffness. Through a process of hyper-parameter optimization, the optimal conditions for each ensemble algorithm were obtained using a grid search method to achieve the best performance. The effectiveness of the SL models was built using datasets collected from the LTPP database. A comparison was made regarding the performance of all models. Lower, Mean Squared Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the higher R-squared (R^2 score) indicated the superior performance of the SL models compared to other approaches, showcasing their robust predictive capabilities for estimating the Creep stiffness of asphalt binders.

2.1. Super Learner Machine Learning Techniques

The SL represents an ensemble machine learning algorithm that combines and utilizes various ensemble algorithms to achieve the best prediction model. Ensemble learning methods can be categorized into three distinct groups: bagging, boosting, and stacking, as depicted in Figure 1. All these methods were employed in our study. Initially, the bagging method, specifically the Random Forest algorithm, was employed. Subsequently, the boosting method, including GBM, AdaBoost, XGBoost, and Catboost, was incorporated. Ultimately, a stacking ensemble model was utilized to address any weaknesses and leverage the inherent strengths of each individual model. The objective was to combine predictions from the contributing methods through a meta-model. The implementation of these ensemble learning strategies aimed to bolster the precision and resilience of our predictive models.

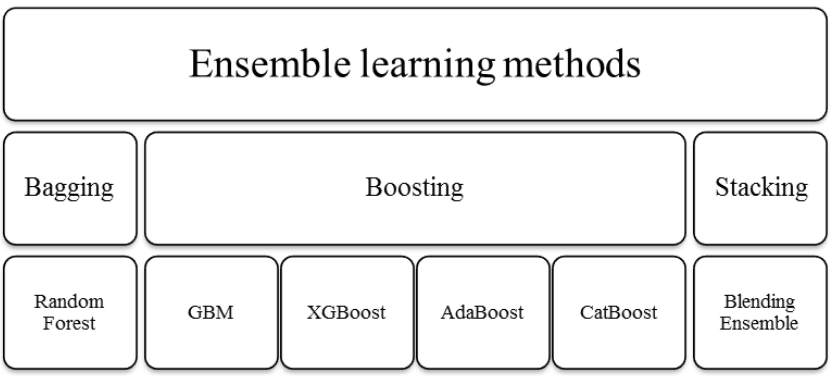


Fig. 1. Flowchart of ensemble methods

2.1.1 Random Forest

Random Forest represents an ensemble learning technique based on bagging. It serves purposes in both regression and classification tasks. Random Forest entails the creation of multiple individual binary decision trees, each incorporating an element of randomness. This stochastic element encourages the trees to produce independent estimates, despite being constructed using a deterministic algorithm and a training dataset [25].

2.1.2 Gradient Boosting Machine (GBM)

GBM utilizes the gradient descent technique to construct models, taking into account the negative partial derivatives of the loss function. The initial model is adapted to fit the original data more effectively and is subsequently refined to address residuals and overcome limitations of the preceding model. This iterative process continues until a convergence criterion is satisfied [26].

2.1.3 Extreme Gradient Boosting (XGBoost)

XGBoost, a prominent boosting technique, expands on the principles of GBM. It involves the sequential development of regression trees, with each successive tree trained on the residuals of the preceding one. This approach effectively mitigates overfitting and enhances computational efficiency. Employing a level-wise learning strategy, XGBoost prioritizes splits that result in the most significant reduction in loss at each leaf [27].

2.1.4 Adaptive Boosting (AdaBoost)

AdaBoost employs multiple decision tree regressors as weak learners, extracting insights from diverse attributes within the dataset. The core idea behind AdaBoost revolves around iteratively updating parameters linked to a specific set of functions. This incremental incorporation of new trees fosters the creation of a resilient learner with improved predictive abilities [28].

2.1.5 Categorical Gradient Boosting (CatBoost)

CatBoost utilizes the entire dataset for training and introduces random permutations to each example. It introduces a novel method for computing leaf values during the selection of tree structures, effectively tackling the biased gradient challenges commonly faced by traditional boosting algorithms. Through these enhancements, CatBoost notably enhances model performance and the capacity to generalize [29].

2.1.5 Stacking

In contrast to bagging and boosting, stacking combines several classifiers or regressors produced by different machine learning algorithms, functioning across various levels or layers. Given the potential for the stacking ensemble model to generate various permutations via different ML algorithms, this research prioritized the application of this SL method. Here, linear regression was utilized as the meta-learner to amalgamate different algorithms, with the aim of achieving heightened accuracy [30].

3. Data Collection and Processing

The data utilized in this study were sourced from the LTPP website, a component of the Strategic Highway Research Program (SHRP). The chosen factors for investigation encompass test temperature, penetration, kinematic viscosity, absolute viscosity, and specific gravity. A comprehensive dataset consisting of 1350 data points was gathered, encompassing the specified input variables. To ensure the dataset's quality, missing data and outliers were filtered out, resulting in a final dataset of 1202 records of Creep stiffness values. The descriptive statistics of the influential parameters used for modelling are presented in Table 1.

Table 1. Statistical properties of dataset

ID	Data	Unit	mean	std	min	max
1	Flexural Creep Stiffness	MPa	265.92	148.56	36.00	775.00
2	M_Value	-	0.32	0.06	0.15	0.5
3	BBR Test Temprature	°C	-17.71	4.62	-30.00	-6.00
4	PENETRATION_77F	mm	34	19	1.00	114.00
5	PENETRATION_115F	mm	182	89.45	5.00	449.00
6	ABSOLUTE_VISC_140F	cP	5.28	1.42	5.50	1.24e+06
7	KINEMATIC_VISC_275F	cSt	1087.72	790.62	212.00	4553.00
8	SPECIFIC_GRAVITY	g/cm3	1.04	0.01	1.00	1.09

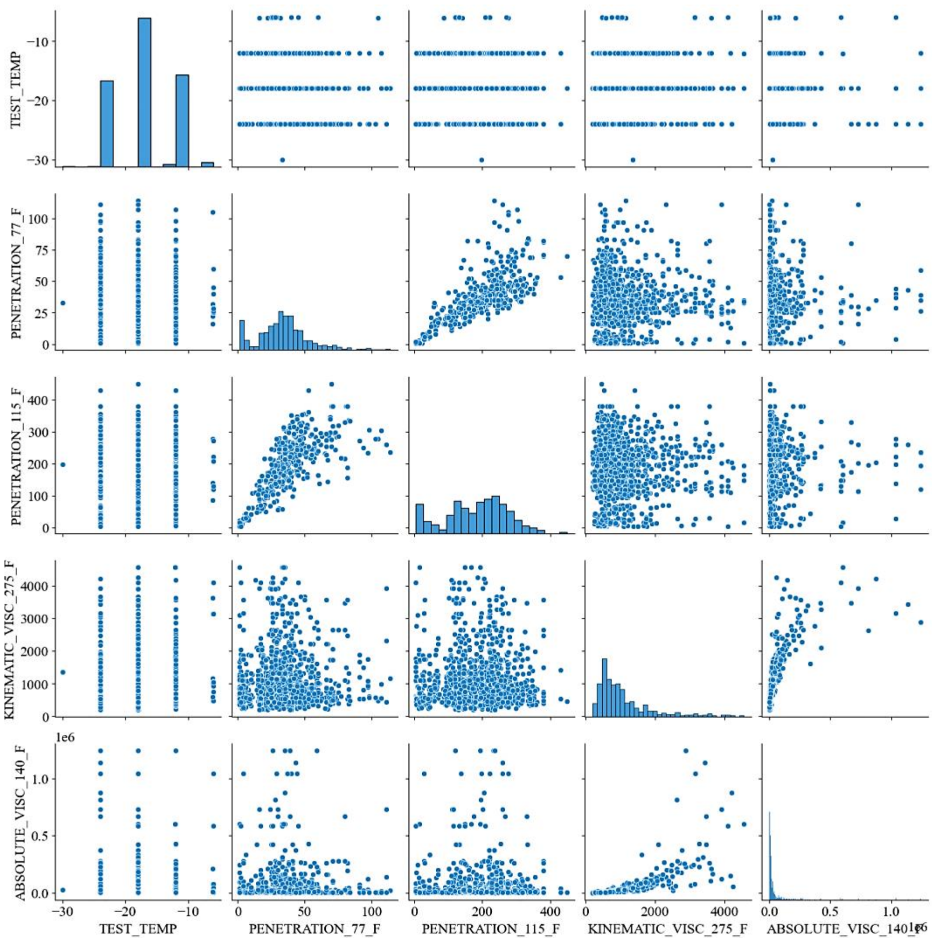


Fig. 2. Pair plot of some variable in dataset

Data visualization is a powerful and essential tool for gaining insights into qualitative data. By utilizing visualization techniques, we can effectively extract valuable information from datasets and detect patterns, outliers, and other irregularities. pair plots were utilized to visualize the data distribution of the dataset in this paper. The data distribution was visualized through these pair plots, as depicted in Figure 2, providing intriguing insights into the relationships between different variables and contributing to a deeper understanding of the data's structure. Informed decisions could be made and potential trends or anomalies that might influence the analysis and modeling process were identified through data visualization. A data splitting strategy was adopted to develop and assess our ML model. The dataset was divided into two segments: a training sample comprising 75 percent of the data and a test sample containing the remaining 25 percent. The training sample was used to construct and train the ML model using various SL techniques. These methods were utilized to uncover the underlying patterns and relationships between the input features, such as test temperature, m-value, penetration, kinematic viscosity, absolute viscosity, specific gravity, and the Creep stiffness. Once the model was trained, its performance was evaluated using the test sample, which had been withheld during the training phase. By the trained model, the ability to accurately predict the Creep stiffness based on the provided input data was assessed. Valuable insights into the model's generalization and its performance on previously unseen data were provided by this evaluation. This rigorous testing ensured that new data could be effectively handled by our model, enabling accurate predictions in real-world scenarios.

4. Results and Discussion

The main objective of this study was the development of accurate prediction models for estimating the Creep stiffness of BBR tests. To accomplish this, supervised ML algorithms were employed, with a specific emphasis on various ensemble models. Six ensemble models, including random forest, GBM, AdaBoost, XGBoost CatBoost and a stacking method combining elements of boosting and bagging, were implemented using the scikit-learn library in Python. The use of these SL models aimed to predict the Creep stiffness based on the rheological properties of various asphalt binders. By training each model with the labeled dataset, intricate relationships between the input features and their corresponding Creep stiffness values were learned.

For the evaluation of the accuracy and effectiveness of these ML models, four widely used performance metrics, namely MSE, RMSE, MAE, and the R^2 score, were employed. valuable insights into the models' performance in accurately predicting the Creep stiffness values and capturing the underlying patterns in the data were provided by these metrics. The MSE quantifies the average squared difference between the actual and predicted values, providing an overall measure of prediction accuracy. The RMSE, derived from MSE, represents the square root of the average squared error, offering a measure of prediction deviation relative to the actual values. The MAE calculates the average absolute difference between the actual and predicted values, providing a straightforward measure of the model's predictive errors. Furthermore, the R^2 score statistic serves as a crucial indicator of how well the influence of an independent variable explains the variance in a dependent variable. It assists in determining the extent to which the variability in the Creep stiffness values can be attributed to the variations in the input features.

In Figure 3, the visualization of the distribution of predicted results against actual results for all models presented and accompanied by the best fit line for the prediction distribution. Remarkably, R^2 scores of 0.89, 0.886, and 0.885 were achieved by the Stacking, Random Forest, and GBM models, respectively, suggesting credible prediction outcomes. Deviation from the fit line was observed in some data points, notably in the calculations of AdaBoost and CatBoost. It is noteworthy that the dataset includes abrupt changes in specific values, negatively affecting the accuracy of sensitive algorithms such as AdaBoost and CatBoost. This resulted in lower R^2 scores and higher MSE and RMSE for these models. In contrast, more accurate predictions on the test data were demonstrated by other algorithms, mainly due to their adeptness in capturing the nonlinear nature of the dataset. Moreover, certain algorithms, especially those incorporating randomness (e.g., random forest), may yield slightly different results in each training iteration. This inherent variability can contribute to deviations. The reduced deviation observed in stacking can be attributed to its capacity to leverage the capabilities of a variety of well-performing models, resulting in predictions that surpass those of any single model in the ensemble. The non-linear connection between Creep stiffness and other variables is indicated by this finding, considering the characteristics of the dataset and the intricate interplay of various factors.

The metrics for all algorithms used in this study are presented in Table 2. The good performance of the stacking, Random Forest, and GBM models can be attributed to their adeptness in handling intricate and non-linear relationships between Creep stiffness and other variables. These models excel in addressing abrupt changes and nonlinearity within the dataset, resulting in more precise and accurate predictions. The lower MSE and RMSE values achieved by these two models further affirm their performance in capturing the intricate relationships between Creep stiffness and other variables.

Table 2. SL models metrics

Model	MAE	MSE	RMSE	R^2 score
Random forest	37.62	2795.42	52.87	0.886
GBM	37.56	2829.75	53.19	0.885
AdaBoost	47.07	3690.37	60.74	0.850
XGBoost	42.13	3402.20	58.32	0.862
CatBoost	39.10	2913.44	53.97	0.882
Stacking	36.36	2479.14	49.79	0.899

The accuracy results of all the models reinforce the notion that the transition from linear regression to ensemble models, specifically Boosting methods, enhances the capability to capture non-linear relationships present in the data. This, in turn, leads to significantly improved prediction accuracy for the targeted problem of estimating the Creep stiffness of asphalt binders.

The lower R^2 score of approximately 0.85 for the AdaBoost algorithm can be attributed to its sensitivity to outliers and noise in the data. AdaBoost, being an ensemble learning method, aims to sequentially fit weak learners to the data, with each subsequent model giving more weight to the misclassified points by the previous ones. This sensitivity to outliers and noise can lead to an overemphasis on capturing individual data points, causing the model to try fitting the noise in the data rather than generalizing the underlying patterns. As a result, the model may exhibit a lower R^2 score, indicating that it does not explain as much of the variance in the dependent variable as desired. Robust preprocessing techniques, such as outlier removal or data cleaning, could potentially improve the performance of AdaBoost in scenarios where outliers and noise have a significant impact on the model's fitting process.

A remarkable improvement is observed in all results obtained by the SL methods compared to the referencing work. The relationship between the predicted and actual values is closely aligned with the best fit line for the dataset in Figure 3. The radar plot depicted in Figure 4 provides a visual representation of the R^2 score values attributed to each method. The shape exhibited in the radar plot notably demonstrates the closely clustered values of R^2 score for each respective method, highlighting the similarity in their predictive performance.

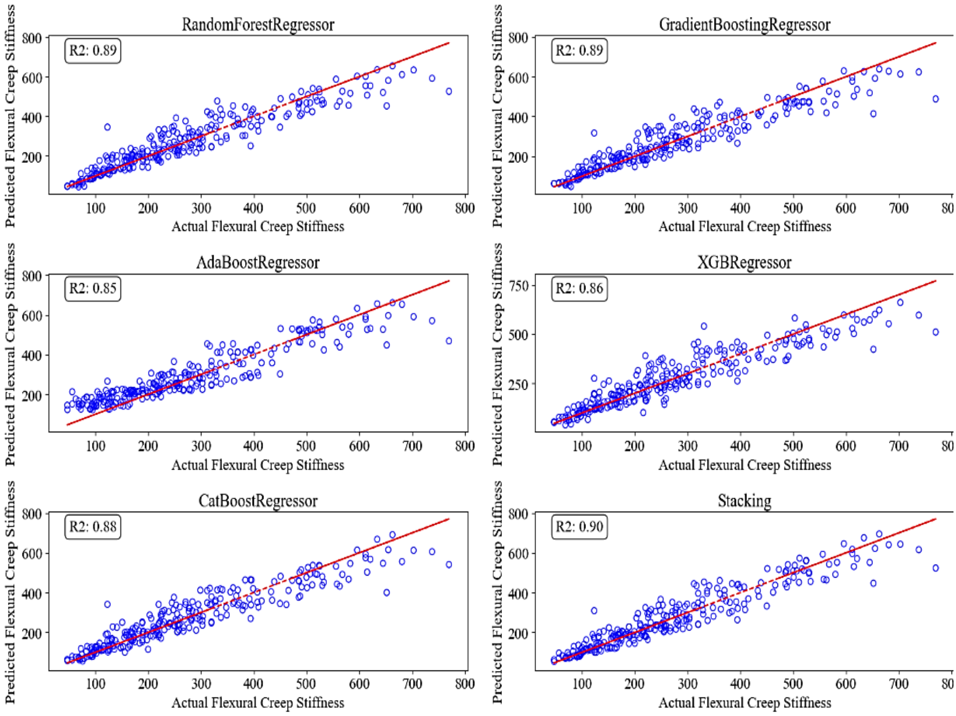


Fig. 3. Predicted vs actual values of the flexural-creep stiffness for the different methods

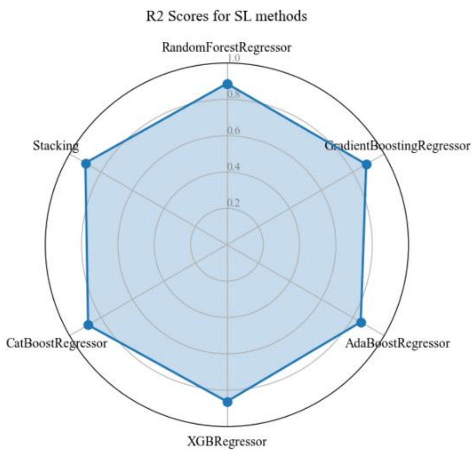


Fig .4. Radar plot of R^2 score different methods

In Figure 5, the feature importance of various factors selected for analysis was demonstrated. Given the best performance of GBM, it was selected to assess the feature importance of each factor. It is evident that the importance of the test temperature outweighs that of other factors, followed by penetration at 115°F, kinematic viscosity, and specific gravity, which exhibited higher significance in predicting the creep stiffness.

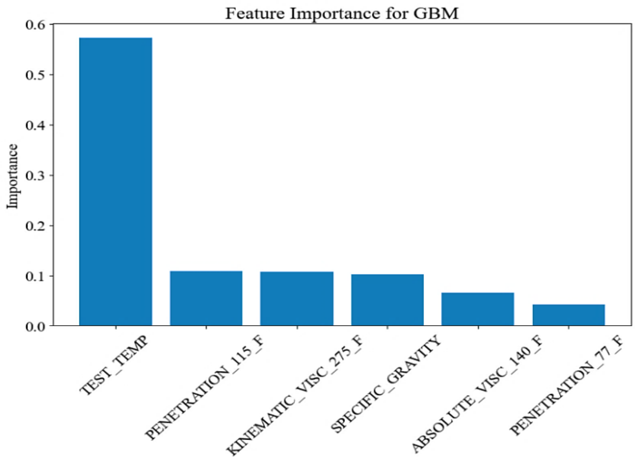


Fig. 5. Feature importance of GBM method

Table 3. Random selected actual vs prediction values of flexural-creep stiffness (MPa)

Actual Value	Random Forest	GBM	AdaBoost	XGBoost	CatBoost	Stacking
189	216	202	235	191	218	192
75.5	63	56	52	80	66	79
76	72	63	61	79	66	73
174	175	174	205	160	178	172
249	262	224	245	203	233	235
130	139	147	177	169	138	132
199	221	201	214	230	202	202

Table 3 presents the random display and comparison of the predicted values of the flexural-creep stiffness, obtained by various methods employed in this research, with the actual values from the test data. It is apparent that, in most cases, the prediction results are deemed acceptable; however, certain instances reveal that certain algorithms have generated values that are not considered satisfactory. This discrepancy could be attributed to the underfit of the algorithm for those specific values. The Stacking model achieved a low Mean Absolute Percentage Error (MAPE) of 2.86% and a high prediction accuracy of 97.14%. The GBM model, while slightly less accurate, still demonstrated a respectable MAPE of 10.56% and a prediction accuracy of 89.44% for the selected data points. This performance is attributed to the intrinsic nature and characteristics of these algorithms, allowing them to converge towards the accurate prediction value through the creation of multiple sub-branches and hidden layers.

5. Conclusions

In this study, the application of various SL machine learning models for predicting the Creep stiffness of asphalt binders in BBR experiments was explored. The aim was to develop accurate and robust prediction models that could provide a better understanding of the rheological behavior of asphalt binders under low-temperature conditions.

Through an extensive analysis using a diverse dataset and various ensemble learning methods, including Random Forest, GBM, CatBoost, XGBoost, Adaboost, and Stacking, prediction models were successfully constructed that achieved high accuracy in predicting the Creep stiffness. Along with the best fit line for the prediction distribution. Notably, the Stacking, Random Forest, and GBM models achieved R^2 scores of 0.89, 0.886, and 0.885, respectively, indicated the strong ability of these methods to capture complex and non-linear relationships between the input variables and the target variable.

The comparison of randomly selected actual vs. predicted values of flexural-creep stiffness across different methods reveals compelling performance metrics. The Stacking model stands out with a notably low Mean Absolute Percentage Error (MAPE) of 2.86% and an impressive prediction accuracy of 97.14%. on the other hand, the GBM model, while slightly less accurate, maintains a respectable performance, showcasing a MAPE of 10.56% and a prediction accuracy of 89.44% for the specific data points under consideration.

The observed lower R^2 score of approximately 0.85 in the AdaBoost algorithm can be ascribed to its susceptibility to outliers and noise within the dataset. This sensitivity to irregularities in the data poses a challenge, potentially leading to overemphasis on individual data points and thereby influencing the algorithm's overall performance

Additionally, the importance of transitioning from conventional linear regression methods to ensemble learning techniques, particularly boosting methods, was emphasized in our research. These ensemble models were capable of handling non-linearity and abrupt changes within the dataset, resulting in improved prediction accuracy compared to traditional approaches like Linear Regression. The results of feature selection revealed that test temperature has an important effect on creep stiffness, leading to higher values.

In conclusion, this research successfully applied ML techniques to predict the low-temperature rheological properties of asphalt binders. The developed models showed promising results in accurately estimating the Creep stiffness, providing valuable insights for the design and engineering of more durable and resilient asphalt pavements. These findings contribute to the advancement of asphalt binder testing and characterization methods, paving the way for more efficient and sustainable infrastructure development in the field of civil engineering. Future studies can build upon this research by exploring additional factors and testing conditions to further enhance the predictive capabilities of ML models in asphalt binder evaluations.

References

[1] Basu A, Marasteanu MO, Hesp SA. Time-temperature superposition and physical hardening effects in low-temperature asphalt binder grading. Transportation research record. 2003;1829(1):1-7. https://doi.org/10.3141/1829-01

[2] Roshan A, Ghasemi A. Laboratory Investigation of Micronized Lomashell Powder Effects on Asphalt Binder and Mix Performance. Journal of Engineering & Technological Sciences. 2023 May 1;55(3). https://doi.org/10.5614/j.eng.technol.sci.2023.55.3.9

[3] Cannone Falchetto A, Moon KH, Wang D, Riccardi C. Investigation on the cooling medium effect in the characterization of asphalt binder with the bending beam rheometer (BBR). Canadian Journal of Civil Engineering. 2018;45(7):594-604. https://doi.org/10.1139/cjce-2017-0586

[4] AASHTO T. 125; Standard Method of Test for Determining the Creep stiffness of Asphalt Mixtures Using the Bending Beam Rheometer (BBR). American Association of State Highway and Transportation Officials: Washington, DC, USA. 2016.

[5] Wang D, Cannone Falchetto A, Riccardi C, Westerhoff J, Wistuba MP. Investigation on the effect of physical hardening and aging temperature on low-temperature rheological properties of asphalt binder. Road Materials and Pavement Design. 2021 May 4;22(5):1117-39. https://doi.org/10.1080/14680629.2019.1665089
[6] Wang D, Falchetto AC, Riccardi C, Poulikakos L, Hofko B, Porot L, Wistuba MP, Baaj H, Mikhailenko P, Moon KH. Investigation on the combined effect of aging temperatures and cooling medium on rheological properties of asphalt binder based on DSR and BBR. Road Materials and Pavement Design. 2019 Apr 30;20(sup1):S409-33. https://doi.org/10.1080/14680629.2019.1589559
[7] Lu X, Uhlback P, Soenen H. Investigation of bitumen low temperature properties using a dynamic shear rheometer with 4 mm parallel plates. International Journal of Pavement Research and Technology. 2017 Jan 1;10(1):15-22. https://doi.org/10.1016/j.ijprt.2016.08.010
[8] Komaragiri S, Filonzi A, Masad A, Hazlett D, Mahmoud E, Bhasin A. Using the dynamic shear rheometer for low-temperature grading of asphalt binders. Journal of Testing and Evaluation. 2022 May 1;50(3):1622-33. https://doi.org/10.1520/jte20210277
[9] Bahia HU, Anderson DA, Christensen DW. The bending beam rheometer; a simple device for measuring low-temperature rheology of asphalt binders (with discussion). Journal of the Association of Asphalt Paving Technologists. 1992;61. https://doi.org/10.1520/stp18187s
[10] Carret JC, Falchetto AC, Marasteanu MO, Di Benedetto H, Wistuba MP, Sauzeat C. Comparison of rheological parameters of asphalt binders obtained from bending beam rheometer and dynamic shear rheometer at low temperatures. Road Materials and Pavement Design. 2015 May 25;16(sup1):211-27. https://doi.org/10.1080/14680629.2015.1029696
[11] Zeng ZA, Underwood BS, Castorena C. Low-temperature performance grade characterisation of asphalt binder using the dynamic shear rheometer. International Journal of Pavement Engineering. 2022 Feb 23;23(3):811-23. https://doi.org/10.1080/10298436.2020.1774766
[12] Sui C, Farrar MJ, Tuminello WH, Turner TF. New technique for measuring low-temperature properties of asphalt binders with small amounts of material. Transportation research record. 2010 Jan;2179(1):23-8. https://doi.org/10.3141/2179-03
[13] Lu X, Uhlback P, Soenen H. Investigation of bitumen low temperature properties using a dynamic shear rheometer with 4 mm parallel plates. International Journal of Pavement Research and Technology. 2017 Jan 1;10(1):15-22. https://doi.org/10.1016/j.ijprt.2016.08.010
[14] Yin H, Chehab GR, Stoffels SM, Kumar T, Premkumar L. Use of creep compliance interconverted from complex modulus for thermal cracking prediction using the M–E pavement design guide. International Journal of Pavement Engineering. 2010 Apr 1;11(2):95-105. https://doi.org/10.1080/10298430802621531
[15] Mirzaiyanrajeh D, Dave EV, Sias JE, Ramsey P. Developing a prediction model for low-temperature fracture energy of asphalt mixtures using machine learning approach. International Journal of Pavement Engineering. 2022 Jan 13:1-2. https://doi.org/10.1080/10298436.2021.2024185
[16] Rahman AS, Mannan UA, Tarefder A. Binder rheology based dynamic modulus and phase angle predictive models for asphalt concrete. Airfield and Highway Pavements. 2017 Aug 27. https://doi.org/10.1061/9780784480939.019
[17] Chen Q, Wang C, Song L. Prediction of low-temperature rheological properties of SBS modified asphalt. Advances in Civil Engineering. 2020 Dec 2;2020:1-8. https://doi.org/10.1155/2020/8864766
[18] Hussain F, Ali Y, Irfan M, Ashraf M, Ahmed S. A data-driven model for phase angle behaviour of asphalt concrete mixtures based on convolutional neural network.

Construction and Building Materials. 2021 Feb 1;269:121235. https://doi.org/10.1016/j.conbuildmat.2020.121235
[19] Xu J, Fan Z, Lin J, Liu P, Wang D, Oeser M. Study on the effects of reversible aging on the low temperature performance of asphalt binders. Construction and Building Materials. 2021 Aug 9;295:123604. https://doi.org/10.1016/j.conbuildmat.2021.123604
[20] Tan X, Huang G, Li X, Zhang M, Ma X, Zhang J. Research on the response of bitumen binder and bitumen mastic to physical hardening effect based on rheological properties. Construction and Building Materials. 2024 Jan 19;412:134875. https://doi.org/10.1016/j.conbuildmat.2024.134875
[21] Sapkota SC, Saha P, Das S, Meesaraganda LP. Prediction of the compressive strength of normal concrete using ensemble machine learning approach. Asian Journal of Civil Engineering. 2023 Jul 8:1-4. https://doi.org/10.1007/s42107-023-00796-x
[22] Malashin, I.P.; Tynchenko, V.S.; Nelyub, V.A.; Borodulin, A.S.; Gantimurov, A.P. Estimation and Prediction of the Polymers' Physical Characteristics Using the Machine Learning Models. *Polymers* 2024 Jan, *16*, 115. https://doi.org/10.3390/polym16010115
[23] Ahmadi M, Lonbar AG, Sharifi A, Beris AT, Nouri M, Javidi AS. Application of segment anything model for civil infrastructure defect assessment. arXiv preprint arXiv:2304.12600. 2023 Apr 25. https://doi.org/10.48550/arXiv.2304.12600
[24] Upadhya A, Thakur MS, Sharma N, Sihag P. Assessment of soft computing-based techniques for the prediction of marshall stability of asphalt concrete reinforced with glass fiber. International Journal of Pavement Research and Technology. 2022 Nov;15(6):1366-85. https://doi.org/10.1007/s42947-021-00094-2
[25] Singh B, Sihag P, Singh K. Modelling of impact of water quality on infiltration rate of soil by random forest regression. Modeling Earth Systems and Environment. 2017 Sep;3:999-1004. https://doi.org/10.1007/s40808-017-0347-3
[26] Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of statistics. 2001 Oct 1:1189-232. https://doi.org/10.1007/s40808-017-0347-3
[27] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. InProceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining 2016 Aug 13 (pp. 785-794). https://doi.org/10.1145/2939672.2939785
[28] Freund Y, Schapire RE. Experiments with a new boosting algorithm. Inicml 1996 Jul 3 (Vol. 96, pp. 148-156). https://doi.org/10.1145/238061.238163
[29] Zhang F, Fleyeh H. Short term electricity spot price forecasting using catboost and bidirectional long short term memory neural network. In2019 16th International Conference on the European Energy Market (EEM) 2019 Sep 18 (pp. 1-6). IEEE. https://doi.org/10.1109/eem.2019.8916412
[30] Wolpert DH. Stacked generalization. Neural networks. 1992 Jan 1;5(2):241-59. https://doi.org/10.1016/s0893-6080(05)80023-1