Journal of Science and Transport Technology
Journal homepage: https://jstt.vn/index.php/en

Accurate and Interpretable Prediction of Marshall Stability for Basalt Fiber Modified Asphalt Concrete using Ensemble Machine Learning

Huong Giang Thi Hoang[1], Ngoc Kien Bui[2,*], Thanh Hai Le[1], Thi Diep Phuong Bach[1], Hoa Van Bui[1], Tai Van Nguyen[3]
[1]University of Transport Technology, Hanoi 100000, Vietnam
[2]Graduate School of Engineering, The University of Tokyo, 113-8656, Tokyo, Japan
[3]The Management Authority for Southern Area Development of Ho Chi Minh city, Ho Chi Minh city, Vietnam

Article info
Type of article:
Original research paper

DOI:
https://doi.org/10.58845/jstt.utt.2025.en.5.3.1-24

***Corresponding author:**
Email address:
kienbui@g.ecc.u-tokyo.ac.jp

Received: 14/04/2025
Received in Revised Form: 18/06/2025
Accepted: 03/07/2025

Abstract: Marshall Stability (MS), a parameter that reflects the load-bearing capacity and deformation resistance of asphalt concrete, is critical for pavement performance and durability. This study assesses the predictive capability of five tree-based machine learning (ML) algorithms – Decision Tree Regression, CatBoost Regressor, Random Forest Regression, Extreme Gradient Boosting Regression, Light Gradient Boosting Machine – in estimating the MS of basalt fiber – modified asphalt concrete (BFMAC). A compiled database of 128 samples was used for model training. Models were optimized with GridSearchCV and 5-fold cross-validation (CV), assessed via multiple statistical metrics, while SHAP analysis provided model interpretability. Among the tested models, Random Forest Regression (RFR) demonstrated the highest predictive accuracy ($R^2 \approx 0.922$, RMSE ≈ 0.748 on the test set) and exhibited strong generalization capability. Interpretability analysis revealed that aggregate gradation (specifically, percentage of aggregate passing 2.36 mm and 4.75 mm sieves) and binder penetration were the most significant factors influencing MS prediction, followed by fiber content. This research underscores the potential of interpretable ML models, such as RFR, in accurately predicting MS, offering a viable alternative to conventional experimental methods for pavement material assessment.

Keywords: Marshall Stability; Machine Learning; Basalt fiber.

1. Introduction

Asphalt concrete (AC) is a commonly used material for surfacing high-grade highways and urban roads [1,2]. AC pavements offer numerous advantages, including smoothness, evenness, minimal dust and noise, relatively high strength, impermeability to prevent water infiltration into the subgrade and base course, complete mechanization from production to construction, immediate trafficability post-construction, and ease of maintenance and repair [3,4]. However, this type of pavement has fundamental drawbacks: temperature sensitivity, susceptibility to deformation at high temperatures, cracking at low

temperatures, cumulative non-recoverable rutting under repeated traffic loads, stripping under moist conditions, and shoving under high lateral forces [5]. These disadvantages are particularly evident in the operational performance of AC pavements using conventional asphalt binder under high ambient temperatures, heavy traffic volumes, and heavy axle loads.

To enhance the mechanical properties of AC, various additives have been incorporated. Traditional additives such as polymers [6] and styrene-butadiene-styrene (SBS) [7,8], are used to improve elasticity and ductility, enabling the pavement to better withstand deformation and minimizing rutting and cracking. The incorporation of fibers contributes to the enhancement of the mechanical characteristics of AC. For instance, synthetic materials such as polyester and aramid, and mineral-based fibers like glass, carbon, cellulose, and basalt are introduced to elevate the tensile strength and toughness of the asphalt mixture [9–11]. These fibers facilitate stress distribution, mitigate cracking, and augment the load-bearing capability of the pavement structure. Utilizing both additives and fibers in combination allows for the optimization of AC's mechanical properties, consequently potentially improving the overall quality and service duration of the pavement.

Among the fibers discussed, basalt fiber (BF) is identified as a natural, environmentally compatible, and non-toxic material [12–14]. For this reason, BF is increasingly regarded as a potential substitute for other fibers, like glass fiber, in reinforcement applications. Nevertheless, experimental investigations concerning BF-modified asphalt concrete (BFMAC) are comparatively limited. This limitation introduces challenges associated with mixture design and the determination of optimal binder and fiber proportions. Within the studied properties of AC containing BF, Marshall Stability (MS) is a notable mechanical parameter applied in the Marshall mix design method. The MS value is defined by the maximum load a test specimen sustains before failure occurs [15,16]. This parameter is employed to evaluate the load-bearing capacity of AC under high-temperature conditions induced by traffic loading. Currently, MS is determined through laboratory testing. Although this approach provides accurate results and control over test input variables, it involves considerable expense, lengthy procedures, and requires technicians with specialized skills. Consequently, an alternative method is sought for the rapid determination of MS for BFMAC, which would provide a foundation for assessing other properties of this composite material.

Progress in artificial intelligence (AI) over recent years has significantly broadened the applicability of machine learning (ML) algorithms across various domains, including complex engineering challenges [17–22]. In the context of pavement materials, ML has demonstrated exceptional efficacy in predicting the mechanical properties of AC. Gong et al. [23] utilized deep learning algorithms to estimate the MS of hot mix asphalt, while Behnood [24] applied biogeography-based programming (BBP) and M5P model trees [25]. Similarly, Barugahare [26] employed bagged tree ensembles for MS prediction. These studies consistently highlight the robust predictive capabilities of AI models, achieving high coefficients of determination R^2 ranging from 0.918 [25] to 0.9781 [24]. Another study successfully predicted the Marshall parameters of stone matrix asphalt (SMA) using ANN techniques [27]. Xiao et al. [28] explored ANN applications in predicting the fatigue life of rubberized AC with reclaimed asphalt pavement (RAP), demonstrating superior performance compared to traditional statistical methods. Phung et al. [14] employed the XGBoost (XGB) algorithm to predict the MS and Marshall Flow (MF) of BFMAC, achieving impressive results with a Pearson correlation coefficient R = 0.9758 for MS and equal 0.9085 for MF. Given these findings, ML algorithms, particularly decision tree-based models, offer a powerful framework for

analyzing sensitivity and predicting the MS of BFMAC, paving the way for more efficient and cost-effective pavement design and maintenance strategies.

This study evaluates the performance of five tree-based algorithms-Decision Tree Regression (DTR), Random Forest Regression (RFR), CatBoost Regressor (CBR), Extreme Gradient Boosting Regression (XGBR), and Light Gradient Boosting Machine (LGBM)-in predicting the MS of BFMAC. For this purpose, a database comprising 128 experimental results sourced from international studies was compiled, covering diverse conditions and material compositions. The models' performances were assessed using various statistical metrics, including the coefficient of determination (R^2), mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), and the A20 index. These metrics were calculated subsequent to hyperparameter optimization aimed at improving model predictive capabilities. Lastly, a sensitivity analysis was conducted to examine the influence of input parameters on the MS of BFMAC. This analysis helps determine the relative importance of each parameter regarding its effect on BFMAC performance.

2. Materials and Methods

2.1. Engineering background

2.1.1. Marshall Stability of AC

The Marshall mix design method is an empirical procedure for determining the optimal asphalt binder content in AC mixes. A principal parameter obtained from this method is MS, which is measured in accordance with standard protocols such as ASTM D6927. The MS value represents the maximum load that a cylindrical specimen, either 6-inch or 4-inch in diameter, can sustain at a constant 60°C temperature before failure. This value indicates the resistance of the mixture to permanent deformation, like rutting, when subjected to traffic loads. A higher MS value generally corresponds to a stiffer, more stable asphalt mixture capable of withstanding heavy loading. Since the laboratory determination of MS is a time-consuming process, the development of accurate predictive models provides an efficient alternative for optimizing mix design.

2.1.2. The role of basalt fiber in AC

Basalt fiber (BF) is a material derived from molten volcanic rock that exhibits high tensile strength, thermal stability, and durability. When incorporated into AC, BF forms a three-dimensional reinforcing network within the asphalt binder and aggregate matrix (Fig. 1). This network serves to enhance binder cohesion, improve the tensile strength and fracture resistance of the mixture by bridging micro-cracks, and increase overall stiffness. The addition of BF is therefore expected to increase the mixture's resistance to rutting and fatigue cracking, which is reflected by an increase in the MS value. The efficacy of the fiber reinforcement is a function of parameters such as fiber content, length, and dispersion within the mixture.

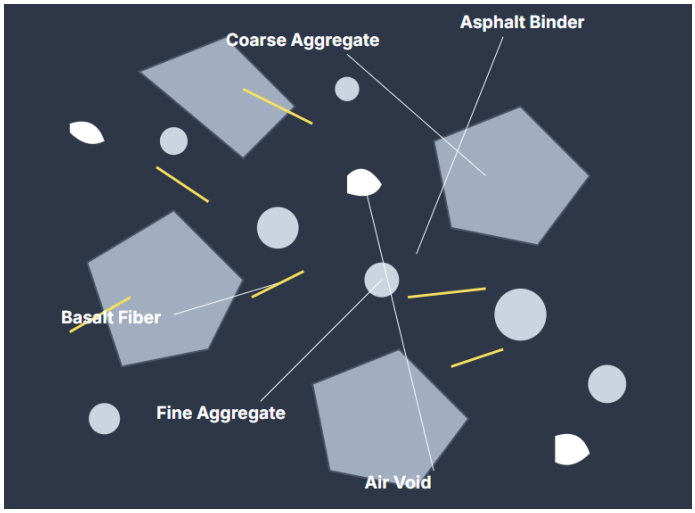

Fig. 1. Schematic illustration of BFMAC components

2.2. Database acquisition

A primary consideration when compiling a database from multiple sources is the consistency of the experimental procedures. To maintain uniformity, the source studies [29–47] were selected based on their adherence to standardized testing methods, principally ASTM D6927 or equivalent national standards. This selection process ensures that critical testing parameters,

such as the Marshall test temperature (60 °C) and compaction effort (75 blows), were consistent across the dataset. Minor variations in factors not explicitly modeled, for instance, aggregate mineralogy, are unavoidable and are treated as inherent noise within the data. The ten input parameters were selected due to their established significance in the asphalt mix design and pavement engineering literature, representing the principal properties of the constituent materials (BF, asphalt binder, and aggregates) that influence the mechanical performance of AC. This database consists of 128 test samples and includes ten input parameters: fiber tensile strength (X1), fiber content (X2), fiber length (X3), fiber diameter (X4), binder penetration (X5), binder softening point (X6), binder content (X7), and the percentage by weight of aggregate passing the 2.36 mm (X8), 4.75 mm (X9), and 9.5 mm (X10) sieves. A detailed statistical summary for all input parameters and the output parameter (MS), including symbol, unit, minimum, median, average, maximum, standard deviation (StD), and skewness (Sk), is presented in Table 1. The specific ranges covered by the input parameters in the database are: fiber tensile strength (X1) from 0 to 4425 MPa; fiber content (X2) from 0 to 2%; fiber length (X3) from 0 to 24 mm; fiber diameter (X4) from 0 to 17 μm; binder penetration (X5) from 55 to 92.3; binder softening point (X6) from 44.5 to 81; aggregate passing 2.36 mm (X8) from 14.9% to 58.62%; aggregate passing 4.75 mm (X9) from 20.08% to 97.88%; and aggregate passing 9.5 mm (X10) from 50% to 100%.

Table 1. Input and output parameters employed for ML model development

	Unit	Min	Median	Average	Max	Std	Sk
Tensile strength (X1)	MPa	0	2800	2800.66	4425	1092.16	0.92
Content of fiber (X2)	%	0	0.34	0.34	2	0.25	3.11
Length of fiber (X3)	mm	0	6	6.81	24	4.12	1.63
Diameter of fiber (X4)	μm	0	13	12.74	17	3.88	2.36
Penetration (X5)	0.1mm	55	67.3	72.11	92.3	11.95	0.5
Softening Point (X6)	°C	44.5	51.9	56.89	81	10.38	0.46
Content of binder (X7)	%	4	5.2	5.43	10.39	1.06	2.3
Aggregate 2.36 mm (X8)	%	14.9	33.02	30.64	58.62	11.64	0.12
Aggregate 4.75 mm (X9)	%	20.08	49.3	43.22	97.88	16.9	0.65
Aggregate 9.5 mm (X10)	%	50	70.1	71.34	100	9	0.58
Marshall stability (MS)	kN	5.69	10.26	10.65	17.7	2.58	0.39

Sk=Skewness; Std=Standard deviation

An examination of the correlation matrix for the input and output variables within the dataset was performed to analyze the characteristics of the collected data. Understanding the interrelationships among input variables, as well as their relationships with the output variable, is relevant for evaluating the relative contribution of the inputs. Accordingly, the distributions of the parameters and the correlations between them were analyzed, with the results presented graphically in Fig. 2. In this figure, positive correlation coefficients denote a positive linear relationship, while negative coefficients indicate a negative linear relationship. The magnitude and direction of these correlations are visually represented by the size and color of the circles; red indicates strong positive correlations and blue indicates strong negative correlations. The analysis indicates that certain pairs of input parameters exhibit relatively high correlation coefficients. For instance, X9 and X8 have a very high positive correlation coefficient of 0.95,

suggesting that these variables are closely related. Similarly, X5 shows a strong negative correlation with X3 (-0.75). Although high multicollinearity can negatively affect certain regression models, such as linear regression, by inflating the variance of coefficient estimates, a decision was made to retain all original input variables. This approach was adopted for two reasons. First, the study's objective is not limited to predicting MS but also includes interpreting the underlying material relationships. The retention of the original, physically meaningful engineering parameters is necessary for the interpretability of the SHAP analysis results. In contrast, variable reduction methods like Principal Component Analysis (PCA) generate abstract features, which obscures the direct connection to mix design parameters. Second, tree-based ensemble models, including RFR, are robust to multicollinearity. While multicollinearity may influence the allocation of feature importance between correlated variables, it generally does not degrade the model's overall predictive accuracy. Therefore, all ten input variables were retained to preserve interpretability without a substantive loss in predictive performance.

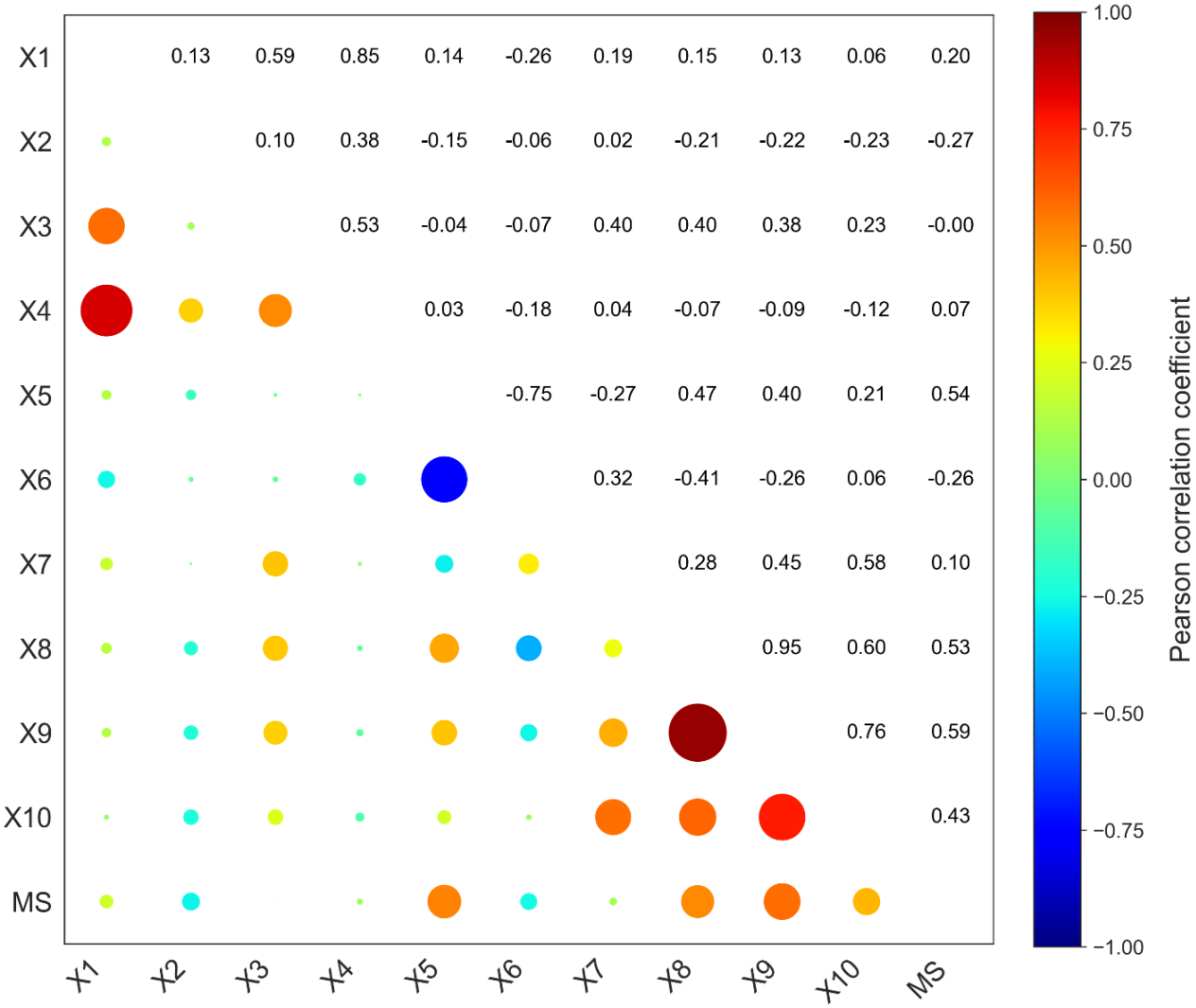

Fig. 2. Correlation matrix of the input and output variables

Fig. 3 shows the frequency distributions for the input and output parameters, comprising fiber tensile strength, fiber content, fiber length, fiber diameter, binder penetration, binder softening point, binder content, aggregate passing 2.36 mm, aggregate passing 4.75 mm, and aggregate passing 9.5 mm. A significant number of fibers have very low tensile strength (<1000 MPa),

suggesting variability in material quality or composition. The data suggests that the content of fiber is generally low, with a narrow range of variation. The most fibers fall within a specific diameter range, contributing to consistent material characteristics. Besides, there is significant variability in penetration values, suggesting differences in material consistency or composition. And the most materials soften within a specific temperature range (50–70°C), with some outliers reaching up to 80°C. Content of binder is generally low (4-6%) and tightly distributed among samples. MS values are distributed across a wide range, from 7.5 kN to 17.5 kN, with smaller peaks observed at 12.5 kN and 15 kN. This suggests variability in material stability, with most samples exhibiting moderate stability values. Furthermore, the dataset contains samples where the fiber content (X2) is zero. These data points represent conventional AC without fiber modification and were included to serve as a baseline for the models. The presence of this control group enables the quantification of the incremental effect of fiber addition relative to a zero-content reference. This approach improves the model's generalization capability across a wider range of mix designs and allows for a more comprehensive sensitivity analysis of the fiber content parameter.

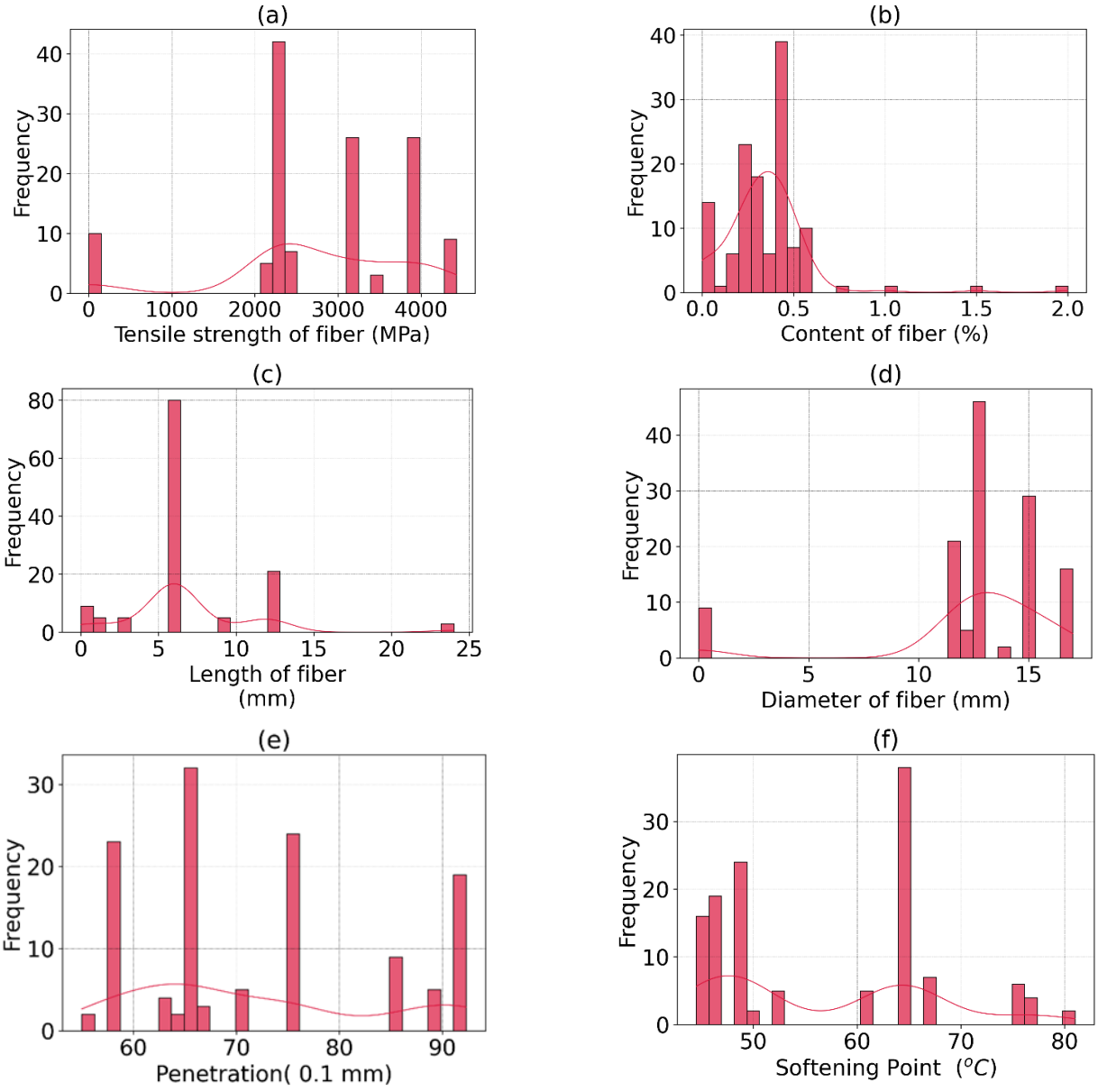

Fig. 3. Distribution of Inputs and Outputs

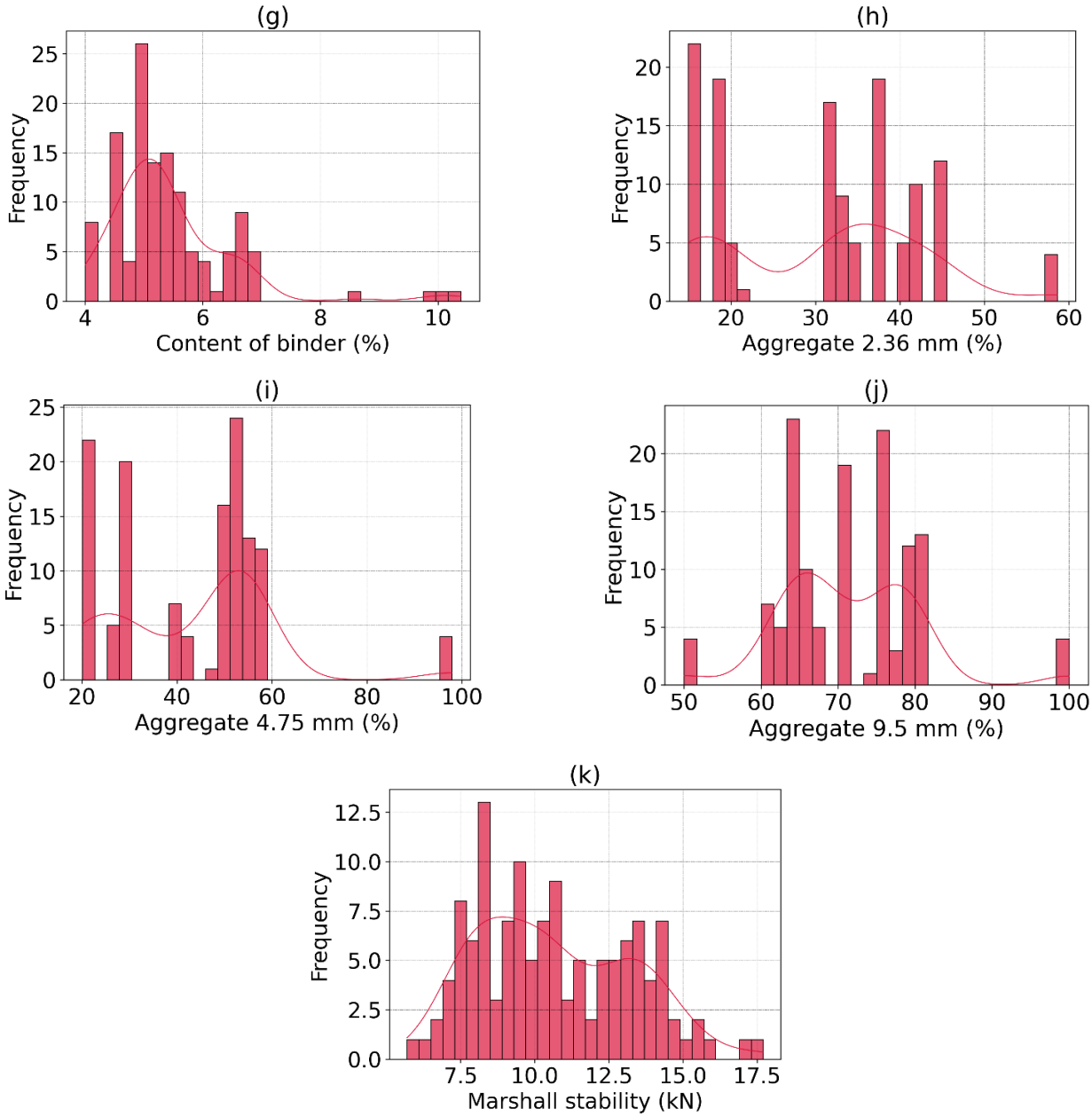

Fig. 3. (continued)

Each database was partitioned into two subsets: a training set, comprising 70% of the data, allocated for developing the ML models, and a testing set, containing the remaining 30%, reserved for evaluating model accuracy. Maintaining separation between the training and testing datasets is standard practice. Specifically, the testing data are not utilized during the model development process and remain unexposed to the models until the final evaluation phase. This methodological separation facilitates an objective assessment of the models' predictive performance on previously unobserved data. Additionally, feature scaling was not applied to the input variables. This preprocessing step was omitted due to the robustness of the selected tree-based models to the scale and distribution of the input features. These models function by partitioning data based on threshold values, a process that is unaffected by monotonic transformations of the variables.

2.3. ML methods

In this study, five tree-based ML algorithms were selected to predict the MS of BFMAC. These models were chosen for several reasons. First, they are effective for processing tabular data and

can model the complex, non-linear relationships between input features and the target output that are common in materials engineering. Second, compared to other algorithms such as neural networks or support vector machines, they exhibit low sensitivity to the scale of input variables and do not require data preprocessing like normalization or standardization. Finally, tree-based models, particularly when combined with techniques like SHAP, are highly interpretable, permitting an examination of the factors driving the predictions. This characteristic aligns with the study's objective to not only predict MS but also to understand the underlying material behavior.

2.3.1. Decision Tree Regression

Decision Tree Regression (DTR) is a fundamental non-parametric supervised learning algorithm that predicts a target value by learning simple decision rules inferred from the data features [48]. The model recursively partitions the dataset into smaller, more homogeneous subsets based on the input features. At each node, the algorithm selects the feature and split point that results in the greatest reduction in variance (for regression tasks). This process continues until a stopping criterion, such as a maximum tree depth or a minimum number of samples per leaf, is met. The final prediction for a given data point is the average of the target values in the terminal node (leaf) it falls into. While DTR models are highly interpretable and easy to visualize, their primary limitation is a strong tendency to overfit the training data, capturing noise rather than the true underlying signal. This high variance can be mitigated by using them as base learners in ensemble methods.

2.3.2. Random Forest Regression

RFR is a powerful ensemble learning technique designed to overcome the overfitting problem of individual DT [49]. It operates by constructing a multitude of DT at training time using a method called bagging, or bootstrap aggregating. The process involves two key steps of randomization:

1. Row sampling: Each tree is trained on a different random sample (a "bootstrap sample") drawn with replacement from the original training dataset.
2. Feature sampling: At each split in a tree, only a random subset of the total features is considered for finding the best split.

By averaging the predictions of all the individual trees, the RFR model reduces variance and produces a more stable and accurate prediction. This diversification strategy makes RFR robust to noise and less prone to overfitting, while still retaining the ability to capture complex interactions and provide feature importance rankings.

2.3.3. CatBoost Regressor

CatBoost Regressor (CBR) is a state-of-the-art gradient boosting algorithm that is particularly powerful for datasets containing categorical features [50,51]. While it is built on the same boosting principles as other gradient boosting methods, where models are trained sequentially to correct the errors of their predecessors, CBR introduces several unique innovations. Its most significant advantage is its sophisticated handling of categorical data. It employs a special technique called ordered boosting, a permutation-based approach, to process categorical features without causing "target leakage," a common issue in other gradient boosting implementations that can lead to overfitting. Furthermore, CBR grows symmetric (or oblivious) trees, meaning the same splitting criterion is used across an entire level of the tree. This structure acts as a form of regularization, helps prevent overfitting, and allows for very fast model prediction. These features make CBR a robust and highly effective algorithm.

2.3.4. Extreme Gradient Boosting Regression

Extreme Gradient Boosting Regression (XGBR) is an advanced and highly efficient implementation of the gradient boosting framework [52]. While it follows the same core principle of sequential error correction as GBR, XGBR introduces several key improvements that enhance

both performance and computational speed [53]:

1. Regularization: XGBR incorporates both L1 (Lasso) and L2 (Ridge) regularization terms into its objective function. This penalizes model complexity and helps prevent overfitting, which is a significant advantage over standard GBR.
2. Optimized algorithm: It employs a more efficient, cache-aware algorithm with parallel and distributed computing capabilities, leading to significantly faster training times on large datasets.
3. Handling of missing values: XGBR has a built-in routine for handling missing data, which simplifies the data preprocessing pipeline.

These enhancements have made XGBR a dominant algorithm in many ML competitions and real-world applications.

2.3.5. Light Gradient Boosting Machine

The Light Gradient Boosting Machine (LGBM) is a more recent gradient boosting framework designed for even greater speed and efficiency, particularly with large datasets [54]. It introduces two novel techniques to achieve this:

1. Gradient-based one-side sampling (GOSS): Instead of using all data instances to compute the information gain for a split, GOSS retains all instances with large gradients (i.e., data points that are poorly predicted) and performs random sampling on those with small gradients. This focuses the training process on the more "difficult" data points without losing accuracy.
2. Exclusive feature bundling (EFB): It bundles mutually exclusive features (i.e., features that rarely take non-zero values simultaneously) together, reducing the number of features and speeding up training.

Furthermore, LGBM grows trees leaf-wise rather than level-wise (as in most other tree algorithms), which allows it to converge more quickly and often results in lower loss for the same number of leaves. These optimizations make LGBM an extremely fast and memory-efficient choice.

2.3.6. Cross-validation

In the field of ML, the phenomenon of overfitting occurs when a model learns the training data too well, capturing its inherent noise and specific patterns. This overfitting can lead to reduced generalization ability, manifesting as inaccurate predictions and diminished predictive accuracy on unseen validation datasets. CV is a technique frequently employed to address this problem. When training ML models with K-fold CV, the dataset is randomly partitioned into subsets. The dataset was partitioned into a training set, containing 70% of the data, and a testing set, comprising the remaining 30%. To facilitate an objective assessment of the model's generalization performance, the testing set was kept separate and was not involved in the model training procedures. Thus, the model had no exposure to the test set data during the training phase. The training dataset itself was further processed using K-fold CV. This involved dividing the training data into K subsets, or folds, of approximately equal size. The model training process was then repeated K times. In each iteration, one distinct fold was utilized as a validation set, for purposes such as hyperparameter tuning, while the model was trained on the combined data from the other K-1 folds. The final performance evaluation is determined by averaging the evaluation outcomes from each of the K training runs. It is generally recommended that K not be set to an excessively high value. A large K can result in a validation set that is substantially smaller than the training set [55]. In such a case, the evaluation results may not accurately reflect the true generalization capacity of the ML methodology, especially when dealing with large-scale datasets. For this study, K-fold CV with K = 5 is implemented.

2.3.7. Grid SearchCV

The hyperparameters of the ML algorithm affect the accuracy of the model [56]. Hyperparameters cannot be estimated directly

from the learning data, they must be established before training a model because these hyperparameters define the architecture of an ML model [57]. Therefore, to fit an ML model for different problems, the hyperparameters need to be tuned. GridSearchCV is a fundamental technique used for hyperparameter optimization in ML.

The fundamental operating principle of GridSearchCV involves a systematic, exhaustive search across a user-defined, discrete hyperparameter space, commonly referred to as a "grid." This grid is constructed by specifying a finite set of candidate values for each hyperparameter under investigation. The algorithm proceeds by methodically generating and evaluating every unique combination of these hyperparameter values. For each generated combination, one ML model is instantiated, trained, and subsequently assessed. Crucially, the performance evaluation for every parameter set is conducted using a k-fold CV technique. This integration of CV serves to yield a robust estimate of the model's generalization performance for the given hyperparameters.

2.3.8. Shap values

A unified strategy for deciphering the results of any ML model is provided by the SHAP (SHapley Additive exPlanations) values paradigm, developed by Lundberg and Lee [58]. For a given prediction, each feature is assigned an importance score. This is accomplished by assessing all potential feature sets and calculating each feature's effect on the disparity between the model's prediction with and without it. For sensitivity analysis, the SHAP values technique is extremely beneficial, yielding a lucid and reliable method to grasp how input variables affect the model's output. When forecasting the MS of AC using BF, one can employ SHAP values to identify the primary determinants that affect the model's forecasts, thus delivering deep insight into the complex connections linking mixture attributes and performance.

2.3.9. Metrics

To assess the predictive accuracy of ML models, five standard statistical metrics are employed, including R^2, MAE, RMSE, MAPE, and A20 index. R^2, which ranges from 0 to 1, indicates the correlation between actual and predicted experimental values, with a value of 1 signifying a perfectly accurate model and 0 an imperfect one [59]. Conversely, RMSE, MAE, MAPE, and A20 quantify the average deviation between observed and anticipated results. Specifically, a ML model exhibits greater quantitative precision when RMSE, MAE, MAPE, and A20 values are lower, and R^2 approaches one. The formula for determining the above performance indicators can be found in research [60].

3. Results and discussions

The outcomes and analysis of the ML model development are presented in this section. Initially, the selection of hyperparameters for the different models under consideration is detailed. Following this, the performance of the developed models is evaluated, and a comparison is made to ascertain the most effective methodology.

3.1. ML hyperparameter selection

The specific hyperparameters and their corresponding value ranges that were explored during the development and optimization of the ML models are outlined in Table 2. This systematic exploration ensured a thorough search for the optimal model configurations. For the learning_rate parameter, a search was conducted over the range of 0.01 to 0.3 using nine equal increments, with an additional value of 0.5 included to evaluate model performance at a higher rate. Tree complexity was controlled by varying the maximum depth between 3 and 10, and the number of leaves (for applicable models) from 15 to 63. The number of trees in the ensemble, specified as n_estimators or iterations, was varied from 50 to 500 to assess the effect of ensemble size on model performance. These selected ranges are suitable for developing robust ML models as they include commonly suggested values and allow for a broad investigation of different model structures and learning dynamics.

This comprehensive search space increases the probability of identifying hyperparameter combinations that produce optimal predictive performance for the given dataset, as explored through methods such as GridSearchCV mentioned in the methodology.

Table 2. Input and output parameters utilized for model development

Hyperparameter	Values in grid
'learning_rate'	0.01, 0.0422, 0.0744, 0.1067, 0.1389, 0.1711, 0.2033, 0.2356, 0.2678, 0.3, 0.5
'max_depth' / 'depth' (Max tree depth)	3, 4, 5, 6, 7, 8, 9
'n_estimators', 'iterations' (Number of Trees)	10, 20, 30, 40, 50, 60, 70 ,80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500
'num_leaves'	15, 20, 26, 31, 36, 42, 47, 52, 58, 63

Table 3 summarizes the results of the hyperparameter optimization process conducted using 5-fold CV. The optimal set of hyperparameters ('Best_Params') for each evaluated model (CBR, XGBR, RFR, LGBM, DTR) was identified using GridSearchCV. The selection criterion was the minimization of the mean RMSE across the validation folds. This score, denoted as the 'Mean CV Score (RMSE)', provides a robust estimate of a model's generalization performance on the training data. It was the primary criterion for selecting the final model configuration prior to evaluation on the independent test set. The results indicate the optimal configurations identified within the search space defined; for instance, the CBR model demonstrated the best performance with a depth of 5, 500 iterations, and a learning rate of 0.3, achieving the lowest CV score of 1.046. The varying optimal parameters across models emphasizes the need for algorithm-specific tuning. Presenting these hyperparameter search results represents a standard step in thorough ML model development, demonstrating the process of selecting the most promising configurations for each algorithm prior to final performance evaluation on unseen data.

Table 3. Hyperparameter optimization results

Model	**Best_Params**	**Mean CV Score (RMSE)**
CBR	'depth': 5, 'iterations': 500, 'learning_rate': 0.3	1.046
XGBR	'learning_rate': 0.5, 'max_depth': 3, 'n_estimators': 250	1.072
RFR	'max_depth': 8, 'max_features': 1.0, 'n_estimators': 150	1.122
LGBM	'learning_rate': 0.3, 'max_depth': 3, 'num_leaves': 15	1.308
DTR	'max_depth': 6, 'max_leaf_nodes': 20, 'min_samples_split': 3	1.331

Fig. 4 provides a comparative visualization of the prediction error, specifically the RMSE, for the five evaluated ML models (CBR, XGBR, RFR, LGBM, and DTR). The comparison is illustrated across three distinct datasets: the training set (Train Data), the CV results (Validation Data), and the independent test set (Test Data). The height of the bars indicates the mean RMSE, while the error bars shown for the CV data represent the variability of the RMSE across the different CV folds. The results depicted were obtained subsequent to the identification of the optimal hyperparameters for each model through a search process. Each optimized model was trained using the designated training dataset. Its performance was then evaluated using the RMSE metric on the training data itself, through a 5-fold CV procedure (implied by CV5) using the training data partitions, and finally on the unseen test dataset to assess generalization capability.

General trends in performance (RMSE) across the three data partitions are evident. Lower

RMSE values indicate improved model performance. Regarding training performance, the CBR and XGBR models achieved notably low RMSE values, significantly lower than the other models. The RFR model also exhibited a low training RMSE, while the DTR and LGBM models showed higher training errors. In terms of CV performance, the RFR, CBR, and XGBR models demonstrated comparable average RMSE values (around 1.0-1.1), which were lower than those for the LGBM and DTR models (around 1.3). However, the error bars for the CV results were considerably large for all models, indicating substantial performance variation across the different data folds. Concerning test performance, the RFR and CBR models demonstrated the lowest RMSE (approximately 0.75) on the independent test set, suggesting the best generalization performance among the models tested. The XGBR, DTR, and LGBM models showed higher RMSE values on the test set (around 1.07-1.14). Comparisons reveal a significant difference between the training RMSE and the CV/test RMSE for the CBR and XGBR models, suggesting that these models may be overfitting the training data. The RFR model showed a smaller gap between training and test/CV performance, indicating potentially better generalization.

Fig. 4 effectively illustrates the critical trade-off between model fitting and generalization. While the CBR and XGBR models demonstrated a strong capacity to learn the training data, their higher errors on the CV and test sets suggest overfitting. The DTR model, being a single DT, likely has high variance, reflected in its relatively poorer performance. The RFR and CBR models emerged as the top performers on the unseen test data, achieving the lowest RMSE. The RFR model, in particular, appears to achieve a favorable balance, avoiding the extreme overfitting observed in the CBR/XGBR models while still achieving strong predictive accuracy on the test set. The large variance observed during CV (error bars) for all models might indicate sensitivity to the specific training data subsets, suggesting that while the RFR and CBR models perform best on average, performance could fluctuate depending on the data encountered. The advantage of ensemble methods such as RFR, CBR, and XGBR over a single DTR is evident in the generally lower errors.

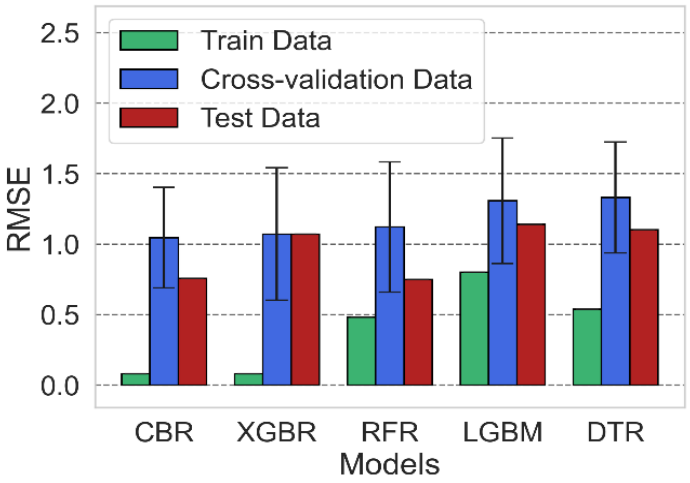

Fig. 4. Comparison of model RMSE on Train, CV, and Test sets

The application of ML models, including ensemble techniques such as RFR, Gradient Boosting variants (XGBR, LGBM, CBR), and DTR, to predict MS and other AC properties aligns with recent research trends in the literature. Studies support the effectiveness of these approaches compared to traditional methods. A direct comparison of the RMSE values obtained here (~0.75 for the best models on the test set) with specific prior studies is challenging due to variations in datasets (e.g., different aggregate types, use of fibers such as carbon, basalt, or waste tire metal), input features, and reported performance metrics. While this study focused on tree-based models due to their interpretability and performance on tabular data, the results can be contextualized with findings from other ML families, such as Artificial Neural Networks (ANN) and Support Vector Regression (SVR). For instance, Upadhya et al. [15] reported an R^2 value of 0.86 for an ANN model predicting the MS of carbon fiber-reinforced asphalt. Nguyen et al. [27] achieved high predictive accuracy for Marshall parameters of stone matrix asphalt with hybrid AI models, and

other studies have demonstrated the utility of ANNs for predicting various asphalt properties [28]. The best-performing RFR model from the current work, with an R^2 of 0.922 on the test set, exhibits a level of accuracy competitive with, and in some cases exceeding, that of these alternative approaches. This result indicates that the focus on tree-based ensembles yielded high predictive accuracy while maintaining model interpretability, a primary objective of this investigation.

Table 4. Model evaluation results

	Metrics \ Models	RFR	CBR	XGBR	DTR	LGBM
Train	MAE	0.291	0.026	0.028	0.422	0.469
	RMSE	0.482	0.079	0.079	0.539	0.801
	R^2	0.964	0.999	0.999	0.955	0.901
	MAPE	2.907	0.294	0.309	4.175	4.813
	A20	98.039	100	100	100	96.078
Test	MAE	0.53	0.423	0.516	0.837	0.758
	RMSE	0.748	0.757	1.071	1.103	1.14
	R^2	0.922	0.92	0.84	0.83	0.818
	MAPE	5.598	4.096	5.006	8.425	8.268
	A20	92.308	92.308	92.308	88.462	84.615
All	MAE	0.339	0.106	0.127	0.507	0.528
	RMSE	0.547	0.348	0.488	0.692	0.881
	R^2	0.955	0.982	0.964	0.928	0.883
	MAPE	3.453	1.066	1.263	5.038	5.515
	A20	96.875	98.438	98.438	97.656	93.75

Table 4 presents a detailed quantitative assessment of the predictive performance for the five ML models (RFR, CBR, XGBR, DTR, LGBM). It expands on the evaluation shown in Fig. 4 by utilizing a broader suite of statistical metrics: MAE, RMSE, R^2, MAPE, and the A20 index. Consistent with the trend observed in Fig. 4's RMSE, RFR and CBR achieved the highest R^2 scores on the test set (0.922 and 0.920), indicating that they account for the largest proportion of variance in the unseen data. XGBR, DTR, and LGBM followed with lower test R^2 values. When considering average error magnitude, CBR demonstrated slightly lower MAE (0.423) and MAPE (4.096%) on the test set compared to RFR (MAE: 0.530, MAPE: 5.598%). This suggests that CBR may have slightly smaller average prediction errors, even though RFR exhibited marginally better RMSE and R^2. The A20 index indicated high reliability for most models on the test set, with RFR, CBR, and XGBR all predicting over 92% of instances within a 20% error margin. LGBM had the lowest A20 score at 84.6%. The table clearly quantifies the overfitting discussed previously; for example, CBR's R^2 decreased from 0.999 (Train) to 0.920 (Test), while RFR showed better stability (0.964 to 0.922). Similar trends were observed for MAE and MAPE values between the Train and Test sets. Although both CBR and RFR were the two highest-performing models on the test set, RFR was selected as the optimal model because of its generalization capability and resistance to overfitting. As indicated, a considerable discrepancy exists between the training and testing performance of the boosting models, particularly CBR. The CBR model yielded an R^2 of 0.999 and an MAE of 0.026 on the training data, an outcome that suggests it overfitted the training samples. This is supported by the subsequent decrease in performance on the independent test data (R^2=0.920, MAE=0.423). In contrast, the RFR model exhibited greater stability, with a smaller difference between its performance on the training set (R^2=0.964, MAE=0.291) and the test set

(R^2=0.922, MAE=0.530). While the test MAE and MAPE for the CBR model were slightly lower, the RFR model's resistance to overfitting indicates greater reliability and a higher likelihood of consistent performance on new data. For practical engineering applications, this robustness is preferable to a minor improvement in a single error metric. Consequently, RFR was identified as the most suitable overall model.

Fig. 5 comprises the bar charts that provide a visual comparison of key performance metrics for the five ML models. The bar chart format facilitates rapid comparisons between models and datasets for the crucial metrics. It renders performance differences and the extent of overfitting (the train-test performance gap) readily apparent. By visualizing all the metrics, the figure supports a balanced assessment, confirming that while RFR and CBR show some indications of higher training performance compared to testing, they generalize significantly better than the other models and produce the most accurate predictions on unseen data based on both explained variance and average error magnitude.

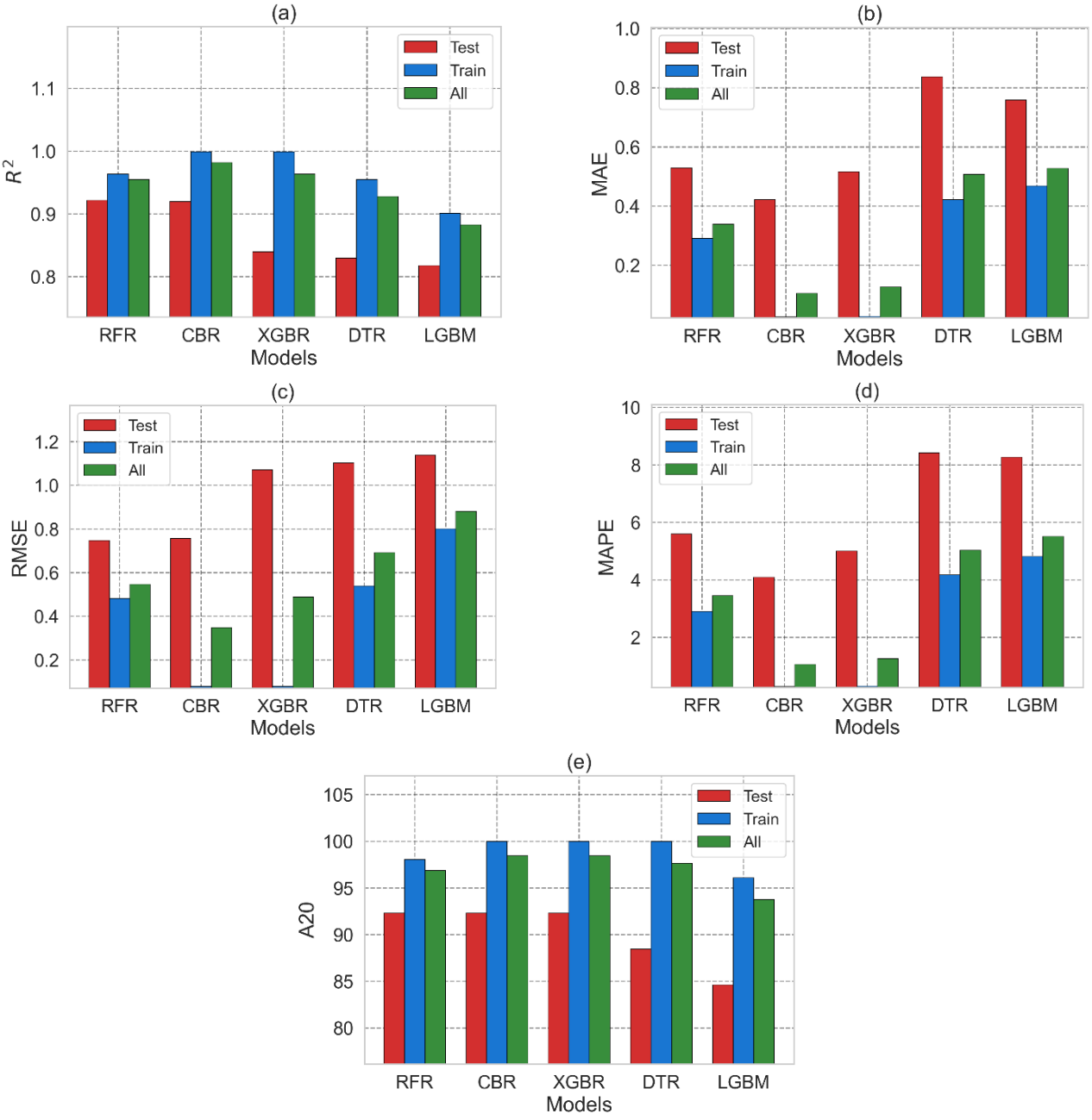

Fig. 5. Comparison of models' performance

3.2. Model performance

This subsection provides a visual evaluation of the performance and predictive accuracy of the RFR model, which was determined to be the best-performing model based on the quantitative metrics. This visual assessment complements the numerical results by demonstrating the model's behavior through regression and prediction comparison plots, specifically Figs. 5 and 6.

Fig. 6 presents regression plots that compare the RFR model's predicted MS values with the actual experimental values. Separate plots are provided for the Training, Testing, and the combined All datasets, facilitating a visual inspection of both model fit and its capacity to generalize to unseen data. Based on the expected content, these plots are anticipated to show data points clustering closely around the line of perfect agreement (y=x), particularly for the testing dataset, visually corroborating the high R^2 value previously reported for the RFR model. Any observable differences in the scatter of points between the training and testing plots would further illustrate the model's generalization capability.

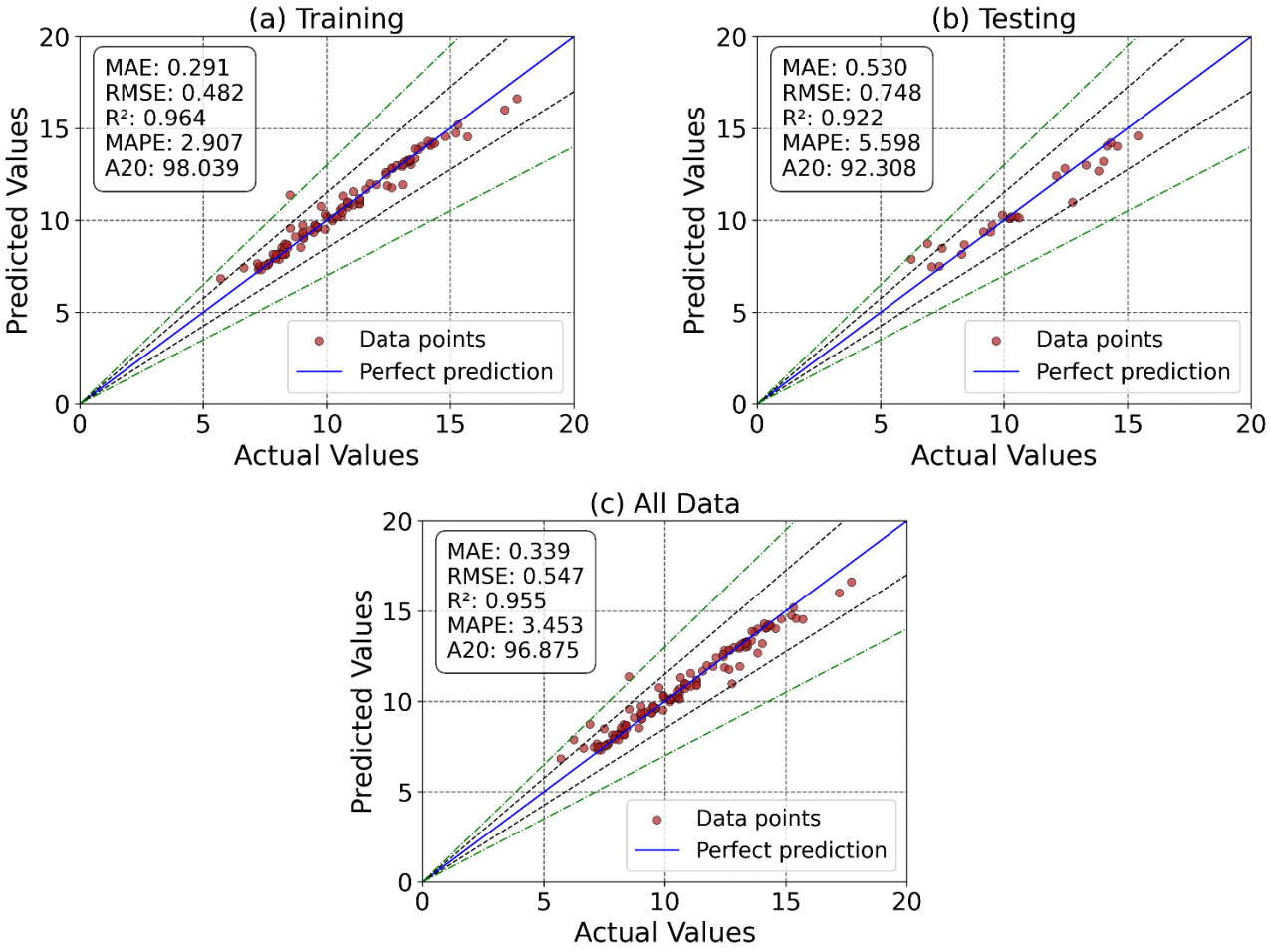

Fig. 6. Regression analysis for Training, Testing and All dataset

Further assessment is provided in Fig. 7, which directly compares the predicted MS values from the RFR model against the actual experimental values for the dataset samples. Based on the expected content, this plot is anticipated to demonstrate how effectively the predicted values track the actual measurements across the data range. It allows for a visual check of the magnitude of prediction errors (the deviations between predicted and actual points) and helps identify any potential systematic error patterns, such as consistent over- or under-prediction in specific value ranges. The distribution of residuals for the Training set (a) exhibits a tight concentration around zero, visually reflecting the very low training errors reported. For the Testing

set (b), the spread is noticeably wider, indicating larger prediction errors when the model generalizes to unseen data; most test residuals appear to fall within approximately -1.5 kN and +1.5 kN. The overall shape in all plots is roughly unimodal and somewhat bell-shaped, although potentially exhibiting slight skewness. The green cumulative percentage lines illustrate the accumulation of errors. For instance, on the test set (b), the curve indicates that a large majority (likely around 80-90%) of the residuals fall below +1 kN.

In summary, the visual evidence presented is expected to support the quantitative findings. These visualizations confirm the high accuracy and robust generalization ability of the selected RFR model for predicting MS from the given input features, thereby confirming its suitability for this engineering prediction task.

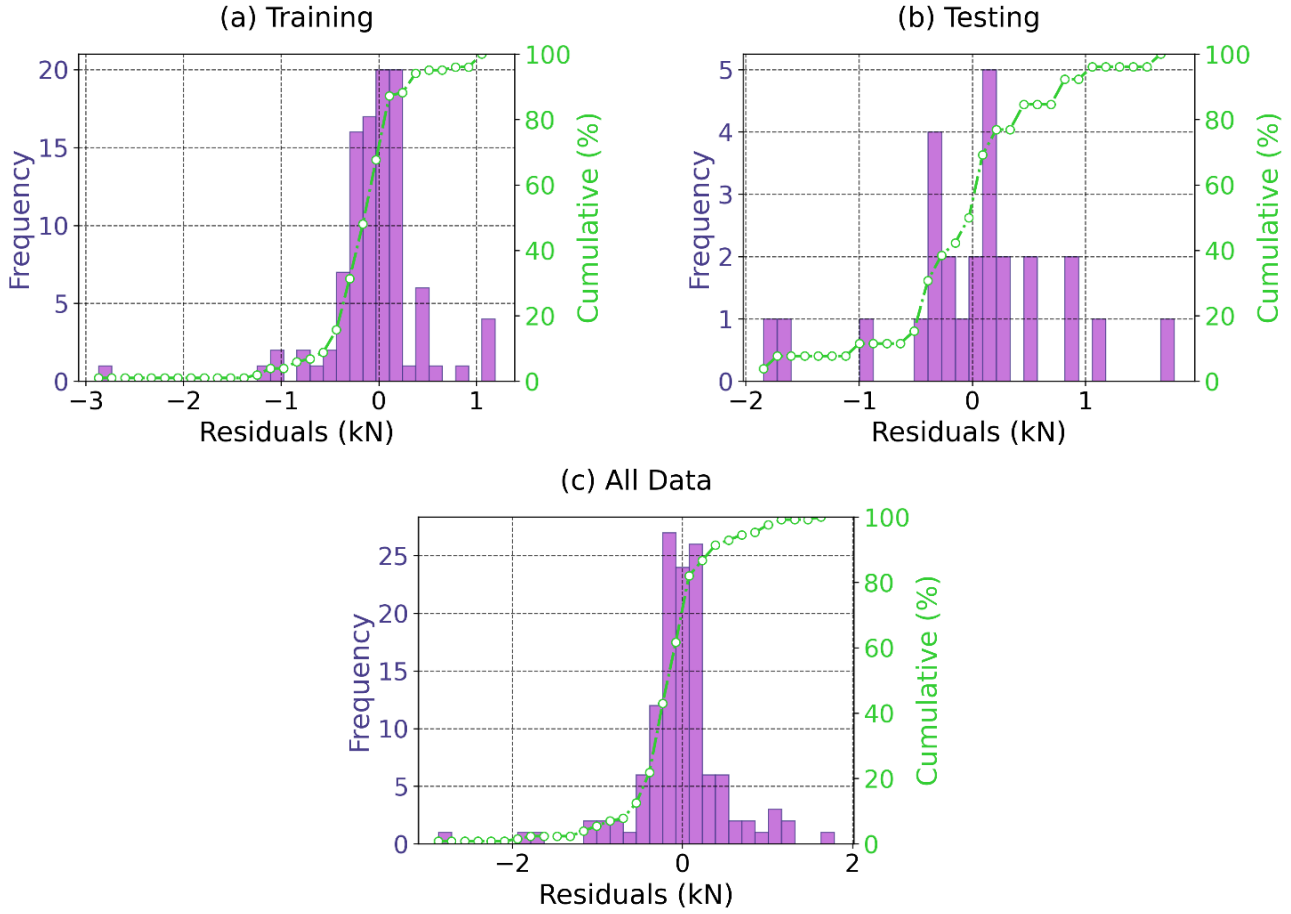

Fig. 7. Histogram showing the comparison between predicted and actual values

3.3. Sensitivity analysis

To interpret the predictions of the RFR model and examine the factors influencing MS, a sensitivity analysis was conducted using SHAP (SHapley Additive exPlanations). Local (per-instance) and global (overall) views of the feature impacts are presented in Fig. 8 and Fig. 9. Fig. 8 presents a SHAP summary plot, designed to elucidate the output of the ML model, presumably the best-performing RFR model. It illustrates the importance and impact of each input feature (X1 through X10) on the prediction of MS. Features are ranked vertically based on their overall importance, and for each feature, the plot displays the distribution of SHAP values across all samples, where each dot represents a single prediction. The color of each dot indicates the original value of that feature for that prediction (red for high values, blue for low values), while its horizontal position shows the SHAP value – the impact of that feature on the model output for that specific prediction (positive values increase the prediction, negative values decrease it).

The analysis indicates that aggregate

gradation features, specifically X8 (Aggregate 2.36 mm) and X9 (Aggregate 4.75 mm), are the most influential factors. These are followed in descending order of importance by X5 (Penetration), X2 (Content of fiber), and X7 (Content of binder). Conversely, features related to fiber properties, such as X1 (Tensile strength) and X4 (Diameter of fiber), had the least influence on the model's predictions.

The color-coding reveals how feature values relate to prediction impact. For the most important features, X8 and X9, higher values (red dots) are generally associated with positive SHAP values, meaning higher percentages of these aggregates tend to increase the predicted stability. For X5 (Penetration), the relationship appears inverse: higher penetration values (red dots, softer binder) correspond to negative SHAP values, decreasing predicted stability. For X2 (Content of fiber), higher fiber content (red dots) generally leads to positive SHAP values, increasing predicted stability. X7 (Content of binder) shows that higher values (red) tend to have positive SHAP values, increasing stability.

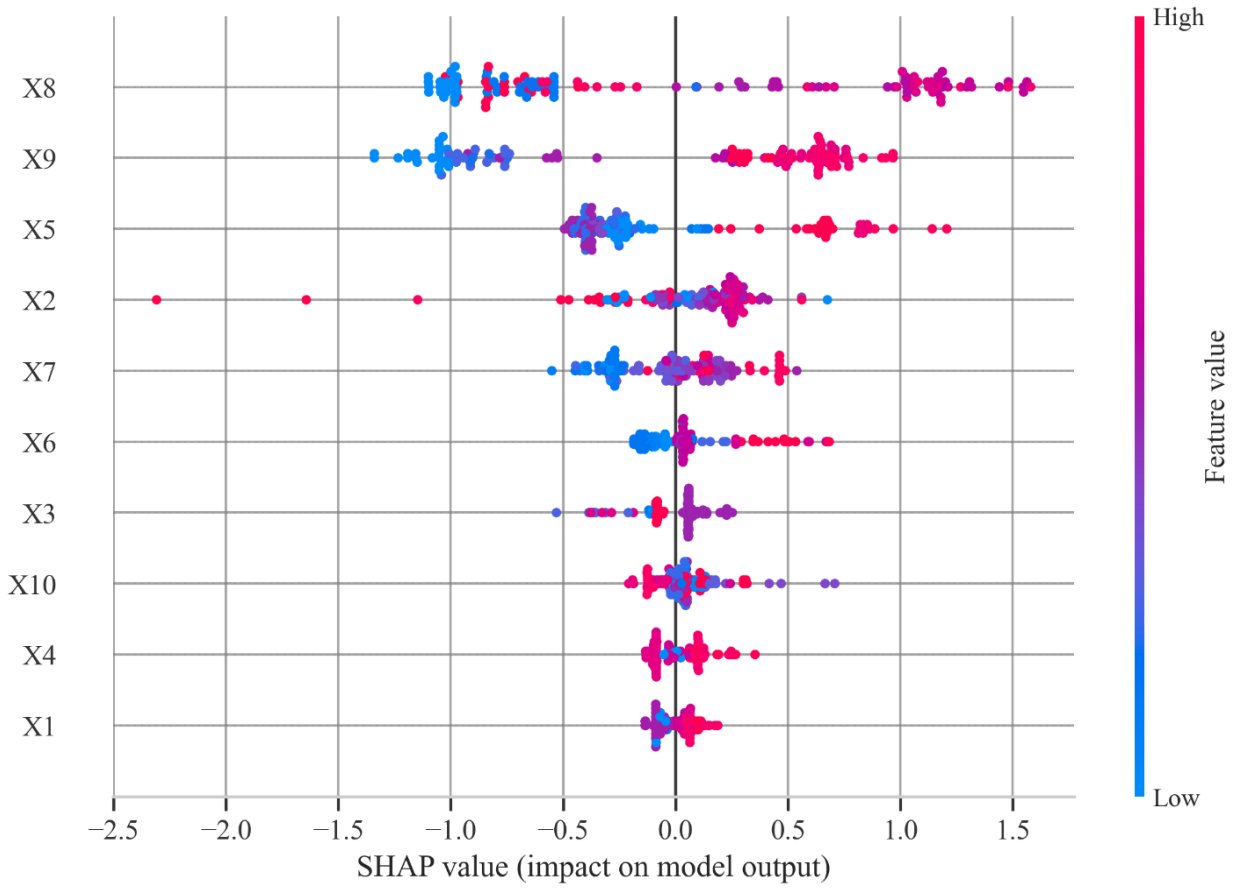

Fig. 8. SHAP summary plot illustrating feature impacts on MS prediction

The SHAP summary plot provides the RFR model's internal logic, enhancing its interpretability beyond simple accuracy metrics. These findings are consistent with established principles of asphalt pavement engineering. The high importance of X8 and X9 indicates that the aggregate skeleton and the degree of interlock between particles are principal contributors to the mixture's load-bearing capacity and resistance to deformation, which are properties quantified by the MS value. The model shows that higher percentages of these aggregate fractions are associated with positive SHAP values, thereby increasing the predicted MS. The importance of Penetration is also consistent with existing knowledge; high penetration values correspond to negative SHAP values, indicating that softer binders reduce the stiffness and stability of the

mixture. The SHAP analysis for X2 (Content of fiber) indicates what may appear to be a counter-intuitive relationship. The model shows that higher fiber content is generally associated with negative SHAP values, suggesting a decrease in the predicted MS. This finding seems to contradict the expected reinforcing role of fibers; however, the result can be explained by the concept of an optimal fiber content in composite materials. Although adding fibers can improve mixture properties up to a certain threshold, excessive quantities can negatively affect performance. High fiber volumes may lead to clumping, poor dispersion within the binder, and an increase in air voids, all of which can disrupt the integrity of the aggregate skeleton and reduce the mixture's stability. It is likely that the model reflects instances in the training data where fiber content exceeded this optimal threshold, thus identifying its negative impact on stability in those cases.

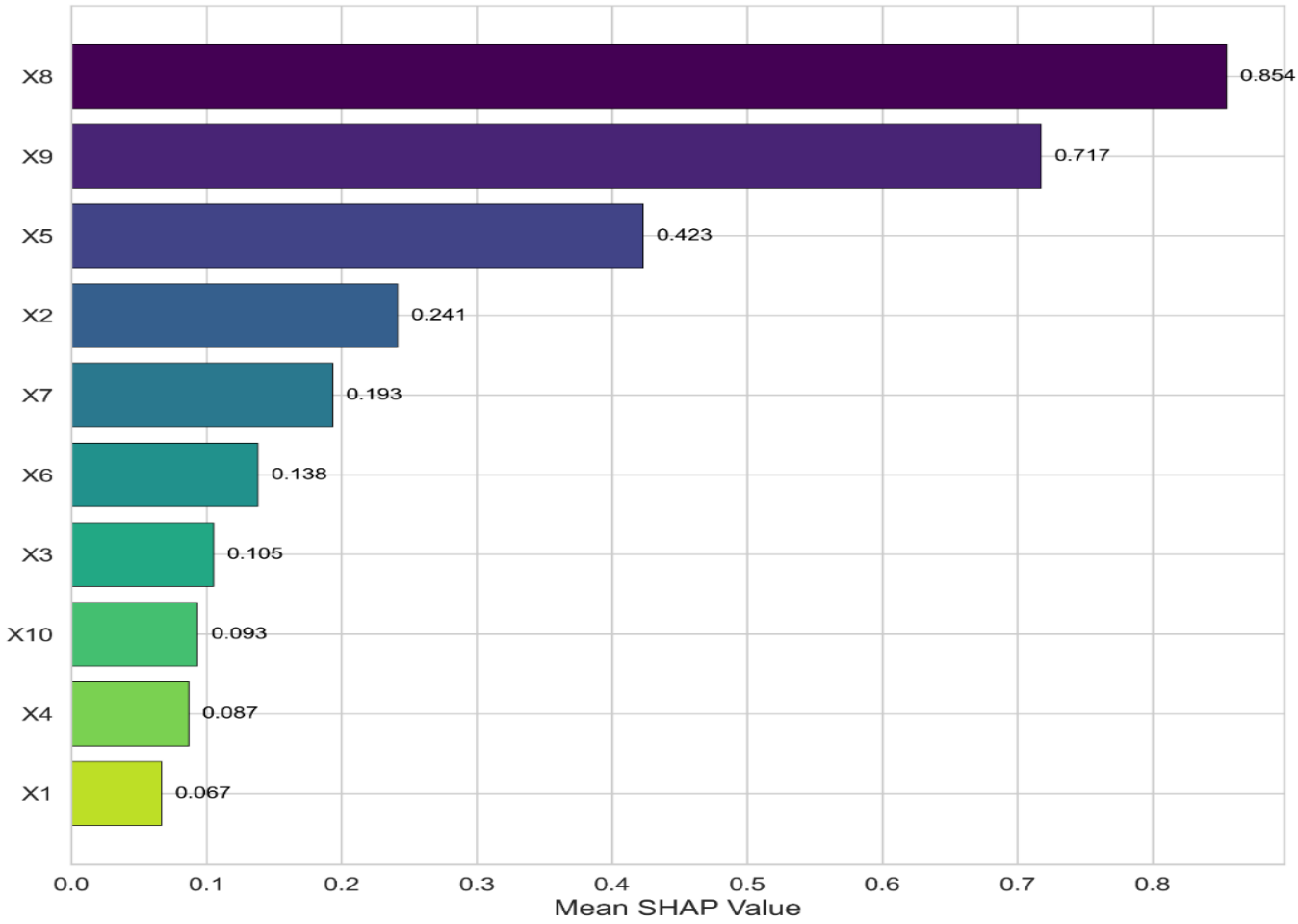

Fig. 9. Global feature importance based on mean absolute SHAP values

Fig. 9 presents a horizontal bar chart summarizing the global importance of each input feature (X1 through X10) based on the mean absolute SHAP value derived from the ML model, presumably the RFR model. The features are ranked vertically in descending order of importance, with the most important feature at the top. The length of each bar, along with the corresponding numerical value displayed, quantifies the average absolute impact of that feature on the model's prediction of MS across all samples in the dataset. The bar chart provides a direct ranking of feature importance. It quantitatively confirms that X8 (Aggregate 2.36 mm) has the highest mean absolute SHAP value (≈0.854), establishing it as the most influential predictor. It is followed in order by X9 (Aggregate 4.75 mm) (≈0.717), X5 (Penetration) (≈0.423), and X2 (Content of fiber) (≈0.241). The plot visually and numerically illustrates the relative differences in

importance. For instance, the impact of X8 is substantially greater than all other features, and a noticeable decrease in importance occurs after the top four features (X8, X9, X5, X2). Features X1, X4, and X10 exhibit the lowest average impact on the predictions. This ranking, based on mean absolute SHAP values, is consistent with the overall importance suggested by the visual spread in the previous SHAP summary plot.

Fig. 9 effectively summarizes the global feature importance findings from the SHAP analysis into a clear, easily interpretable format. By averaging the absolute SHAP values, it provides a definitive ranking of which input variables exert the most influence on the RFR model's MS predictions. The results strongly suggest that aggregate grading, specifically the percentages of 2.36 mm and 4.75 mm aggregates, and binder penetration are the primary factors the model relies upon, followed by fiber content. This concise summary is useful for identifying the key drivers learned by the model and aligns well with the principles of asphalt mix design, where aggregate structure and binder properties are known to be critical for stability. Additionally, while a sensitivity analysis could be performed by removing low-impact features (e.g., X1, X4) and retraining the model, the SHAP analysis provides a more detailed assessment, as it quantifies the marginal contribution of each feature to every individual prediction. Given the low mean SHAP values for X1 and X4 (0.067 and 0.087, respectively), their contribution to the model's predictive accuracy is negligible. Therefore, the removal of these features would be unlikely to substantively alter the results, and the present analysis is sufficient to demonstrate their limited influence on the model's predictions.

4. Conclusions and Perspectives

This investigation successfully developed and evaluated several ML models for the prediction of MS of AC, based on key material properties and mix design parameters. Through a rigorous comparison employing multiple performance metrics and CV, the RFR model was identified as the most effective, demonstrating high accuracy (R^2 ≈ 0.922 on the test set) and strong generalization capabilities. Visual assessments through regression analysis and residual distribution plots further corroborated the RFR model's reliability and unbiased predictive performance. Interpretability analysis utilizing SHAP revealed the key factors influencing the RFR model's predictions. Aggregate characteristics, specifically the percentages passing the 2.36 mm (X8) and 4.75 mm (X9) sieves, along with binder penetration (X5) and fiber content (X2), were identified as the most influential variables. These findings are consistent with fundamental principles of asphalt mix design, enhancing confidence in the model's learned relationships.

The results emphasize the considerable potential of ML, particularly ensemble methods such as RFR combined with interpretability techniques like SHAP, to provide rapid, cost-effective, and reliable estimations of critical asphalt performance parameters such as MS. This approach can function as a valuable tool to complement or potentially streamline traditional, time-consuming laboratory testing procedures.

Future research directions arising from this work include expanding the database with more varied experimental data encompassing different material sources and environmental conditions to enhance the model's robustness and applicability range. Additionally, exploring other advanced ML architectures, including deep learning techniques, could potentially yield further improvements in predictive accuracy.

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and material: Data will be made available on request.

References

[1] S. Magar, F. Xiao, D. Singh, B. Showkat. (2022). Applications of reclaimed asphalt

pavement in India – A review. *Journal of Cleaner Production*, 335, 130221. https://doi.org/10.1016/j.jclepro.2021.130221

[2] T. Baghaee Moghaddam. (2019). Development of high modulus asphalt concrete mix design technology for use on Ontario's highways. *Theses, University of Waterloo*.

[3] S. Wu, O. Tahri. (2021). State-of-art carbon and graphene family nanomaterials for asphalt modification. *Road Materials and Pavement Design*, 22(4), 735-756. https://doi.org/10.1080/14680629.2019.1642946

[4] J.G. Speight. (2015). Asphalt materials science and technology. Butterworth-Heinemann.

[5] Eng. Ahmed Mohamed Abdl-Latif Eraky. (2022). Improving the Performance of Highways and Airports Flexible Pavement to Resist the Effect of Octahedral Shear Stresses. *PhD Thesis, Helwan University, Egypt.*

[6] U. Isacsson, X. Lu. (1995). Testing and appraisal of polymer modified road bitumens—state of the art. *Materials and Structures*, 28, 139-159. https://doi.org/10.1007/BF02473221

[7] G.D. Airey. (2003). Rheological properties of styrene butadiene styrene polymer modified road bitumens☆. Fuel, 82(14), 1709-1719. https://doi.org/10.1016/S0016-2361(03)00146-7

[8] A. Behnood, J. Olek. (2017). Rheological properties of asphalt binders modified with styrene-butadiene-styrene (SBS), ground tire rubber (GTR), or polyphosphoric acid (PPA). *Construction and Building Materials*, 151, 464-478.https://doi.org/10.1016/j.conbuildmat.2017.06.115

[9] M. Ameri, M. Nemati, H. Shaker, F. Jafari. (2019). Experimental and numerical investigation of the properties of the Hot Mix Asphalt Concrete with basalt and glass fiber. *Fracture and Structural Integrity*, 13(50), 149-162. https://doi.org/10.3221/IGF-ESIS.50.14

[10] D. Zheng, W. Song, J. Fu, G. Xue, J. Li, S. Cao. (2020). Research on mechanical characteristics, fractal dimension and internal structure of fiber reinforced concrete under uniaxial compression. *Construction and Building Materials*, 258, 120351. https://doi.org/10.1016/j.conbuildmat.2020.120351

[11] B.J. Putman, S.N. Amirkhanian. (2004). Utilization of waste fibers in stone matrix asphalt mixtures. *Resources, Conservation and Recycling*, 42(3), 265-274. https://doi.org/10.1016/j.resconrec.2004.04.005

[12] N. Morova. (2013). Investigation of usability of basalt fibers in hot mix asphalt concrete. *Construction and Building Materials*, 47, 175-180. https://doi.org/10.1016/j.conbuildmat.2013.04.048

[13] C. Wu, L. Li, Y. Cheng, Z. Gu, Z. Lv, R. Wang, B. Guan. (2020). Effect of Diatomite and Basalt Fibers on Pavement Performance and Vibration Attenuation of Waste Tires Rubber-Modified Asphalt Mixtures. *Mathematical Problems in Engineering*, 2020, 8853428, 1-13. https://doi.org/10.1155/2020/8853428

[14] B.-N. Phung, T.-H. Le, T.-A. Nguyen, H.-G.T. Hoang, H.-B. Ly. (2023). Novel approaches to predict the Marshall parameters of basalt fiber asphalt concrete. *Construction and Building Materials*, 400, 132847. https://doi.org/10.1016/j.conbuildmat.2023.132847

[15] A. Upadhya, M.S. Thakur, P. Sihag. (2024). Predicting Marshall Stability of Carbon Fiber-Reinforced Asphalt Concrete Using Machine Learning Techniques. *International Journal of Pavement Research and Technology*, 17, 102-122. https://doi.org/10.1007/s42947-022-00223-5

[16] H.-G.T. Hoang, H.-L. Nguyen, H.-B. Ly. (2024). Evaluation of the performance enhancement of asphalt concrete via graphene oxide incorporation: A multi-test approach. *Proceedings of the Institution of Mechanical*

Engineers, Part L: Journal of Materials: Design and Applications, 239(6), 1156-1169. https://doi.org/10.1177/14644207241282113

[17] H.-B. Ly, L.M. Le, H.T. Duong, T.C. Nguyen, T.A. Pham, T.-T. Le, V.M. Le, L. N.-Ngoc, B.T. Pham. (2019). Hybrid artificial intelligence approaches for predicting critical buckling load of structural members under compression considering the influence of initial geometric imperfections. *Applied Sciences*, 9(11), 2258. https://doi.org/10.3390/app9112258

[18] H.-B. Ly, B.T. Pham, D.V. Dao, V.M. Le, L.M. Le, T.-T. Le. (2019). Improvement of ANFIS model for prediction of compressive strength of manufactured sand concrete. *Applied Sciences*, 9(18), 3841. https://doi.org/10.3390/app9183841

[19] H.-G.T. Hoang, T.-A. Nguyen, H.-L. Nguyen, H.-B. Ly. (2022). Neural network approach for GO-modified asphalt properties estimation. *Case Studies in Construction Materials*, 17, e01617. https://doi.org/10.1016/j.cscm.2022.e01617

[20] V.-S. Ha, H.N.T. Bao. (2024). Machine Learning Models for Real-Time Traffic Prediction: A Case Study in Urban Traffic Management. *Journal of Science and Transport Technology*, 4(2), 1-12. https://doi.org/10.58845/jstt.utt.2024.en.4.2.1-12

[21] N.T. Kien. (2023). Prediction of the surface rutting of unpaved road–Case study from B&T Quang Binh windfarm project. *Journal of Science and Transport Technology*, 3(3), 6-12. https://doi.org/10.58845/jstt.utt.2023.en.3.3.6-12

[22] T.Q. Ngo, L.Q. Nguyen, V.Q. Tran. (2022). Predicting tensile strength of cemented paste backfill with aid of second order polynomial regression. *Journal of Science and Transport Technology*, 2(4), 43-51. https://doi.org/10.58845/jstt.utt.2022.en.2.4.43-51

[23] H. Gong, Y. Sun, Y. Dong, B. Han, P. Polaczyk, W. Hu, B. Huang. (2020). Improved estimation of dynamic modulus for hot mix asphalt using deep learning. Construction and Building Materials, 263, 119912. https://doi.org/10.1016/j.conbuildmat.2020.119912

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

[25] A. Behnood, D. Daneshvar. (2020). A machine learning study of the dynamic modulus of asphalt concretes: An application of M5P model tree algorithm. *Construction and Building Materials*, 262, 120544. https://doi.org/10.1016/j.conbuildmat.2020.120544

[26] J. Barugahare, A.N. Amirkhanian, F. Xiao, S.N. Amirkhanian. (2020). Predicting the dynamic modulus of hot mix asphalt mixtures using bagged trees ensemble. *Construction and Building Materials*, 260, 120468. https://doi.org/10.1016/j.conbuildmat.2020.120468

[27] H.-L. Nguyen, T.-H. Le, C.-T. Pham, T.-T. Le, L.S. Ho, V.M. Le, B.T. Pham, H.-B. Ly. (2019). Development of hybrid artificial intelligence approaches and a support vector machine algorithm for predicting the marshall parameters of stone matrix asphalt. *Applied Sciences*, 9(15), 3172. https://doi.org/10.3390/app9153172

[28] F. Xiao, S.A. Amirkhanian, C.H. Juang. (2009). Prediction of fatigue life of rubberized asphalt concrete mixtures containing reclaimed asphalt pavement using artificial neural networks. *Journal of Materials in Civil Engineering*, 21(6), 253-261. https://doi.org/10.1061/(ASCE)0899-1561(2009)21:6(253)

[29] Y. Zheng, Y. Cai, G. Zhang, H. Fang. (2014). Fatigue property of basalt fiber-modified asphalt mixture under complicated environment. *Journal of Wuhan University of Technology-Mater. Sci. Ed.*, 29, 996-1004. https://doi.org/10.1007/s11595-014-1033-1

[30] N. Morova. (2013). Investigation of usability of basalt fibers in hot mix asphalt concrete. *Construction and Building Materials*, 47, 175-180.https://doi.org/10.1016/j.conbuildmat.2013.04.048.

[31] H. Zhao, B. Guan, R. Xiong, A. Zhang. (2020). Investigation of the performance of basalt fiber reinforced asphalt mixture. *Applied Sciences*, 10(5), 1561. https://doi.org/10.3390/app10051561

[32] Y. Cheng, D. Yu, Y. Gong, C. Zhu, J. Tao, W. Wang. (2018). Laboratory evaluation on performance of eco-friendly basalt fiber and diatomite compound modified asphalt mixture. *Materials*, 11(12), 2400. https://doi.org/10.3390/ma11122400

[33] W. Wang, Y. Cheng, G. Tan. (2018). Design optimization of SBS-modified asphalt mixture reinforced with eco-friendly basalt fiber based on response surface methodology. *Materials*, 11(8), 1311. https://doi.org/10.3390/ma11081311

[34] Y. Cheng, L. Li, P. Zhou, Y. Zhang, H. Liu. (2019). Multi-objective optimization design and test of compound diatomite and basalt fiber asphalt mixture. *Materials*, 12(9), 1461. https://doi.org/10.3390/ma12091461

[35] W. Wang, Y. Cheng, P. Zhou, G. Tan, H. Wang, H. Liu. (2019). Performance evaluation of styrene-butadiene-styrene-modified stone mastic asphalt with basalt fiber using different compaction methods. *Polymers*, 11(6), 1006. https://doi.org/10.3390/polym11061006

[36] W.-X. Fan, H.-G. Kang, Y.-X. Zheng. (2010). Experimental study of pavement performance of basalt fiber-modified asphalt mixture. *Journal of Southeast University (English Edition)*, 26(4), 614-617.

[37] Y.Z. Chen, Z.X. Li. (2012). Study of road property of basalt fiber asphalt concrete. *Applied Mechanics and Materials, Trans Tech Publications*, 238, 22-25. https://doi.org/10.4028/www.scientific.net/amm.238.22

[38] W.X. Fan, S.F. Zhang, L.Q. Liu. (2013). Laboratory study of marshall of basalt fiber-modified asphalt mixture. *Applied Mechanics and Materials, Trans Tech Publ*ications, 256-259, 1851-1857. https://doi.org/10.4028/www.scientific.net/AMM.256-259.1851

[39] C. Chai, Y. Cheng, Y. Zhang, B. Zhu, H. Liu. (2020). Mechanical properties of crumb rubber and basalt fiber composite modified porous asphalt concrete with steel slag as aggregate. *Polymers*, 12(11), 2552. https://doi.org/10.3390/polym12112552

[40] Y. Cheng, C. Chai, Y. Zhang, Y. Chen, B. Zhu. (2019). A new eco-friendly porous asphalt mixture modified by crumb rubber and basalt fiber. *Sustainability*, 11(20), 5754. https://doi.org/10.3390/su11205754

[41] W. Wang, Y. Cheng, H. Chen, G. Tan, Z. Lv, Y. Bai. (2019). Study on the performances of waste crumb rubber modified asphalt mixture with eco-friendly diatomite and basalt fiber. *Sustainability*, 11(19), 5282. https://doi.org/10.3390/su11195282

[42] Y.X. Zheng, Y.C. Cai, Y.M. Zhang. (2011). Laboratory study of pavement performance of basalt fiber-modified asphalt mixture. *Advanced Materials Research, Trans Tech Publications*, 266, 175-179. https://doi.org/10.4028/www.scientific.net/AMR.266.175

[43] L. Zhao, J. Chen, S. Wang. (2010). Using mineral fibers to improve asphalt and asphalt mixture behavior. *Traffic and Transportation Studies 2010*, 1352-1360. https://doi.org/10.1061/41123(383)129

[44] M.H. Hao, P.W. Hao. (2014). Natural mineral fiber improved asphalt mixture

performance. *Applied Mechanics and Materials, Trans Tech Publications*, 638-640, 1166-1170. https://doi.org/10.4028/www.scientific.net/AMM.638-640.1166

[45] T.T. Nguyen, N.H. Tran, H.K. Pham, X.C. Bui, N.K. Bui. (2021). Research on the effects of sulfur on characteristics of sulfur bituminous binder (SBB) and hot mix asphalt–sulfur (HMAS). *Transport and Communications Science Journal*, 72(1), (01/2021), 33-45. https://doi.org/10.47869/tcsj.72.1.5

[46] F. Liu, A. Dong, C. Liu, W. Wu. (2018). Mix design of asphalt mixture used for the waterproof and anti-cracking layer in the rainy area of South China. *Journal of Applied Biomaterials & Functional Materials*, 16, 112-118.https://doi.org/10.1177/2280800017753052

[47] T. Huang, J. Chen, M. Li, Y. Tang, S. Lv, H. Yu, H. Liu, W. Yan. (2021). Investigation on three-dimensional failure criterion of asphalt mixture considering the effect of stiffness. *Construction and Building Materials*, 285, 122431.https://doi.org/10.1016/j.conbuildmat.2021.122431

[48] S. Pathak, I. Mishra, A. Swetapadma. (2018). An assessment of decision tree based classification and regression algorithms. *2018 3rd International Conference on Inventive Computation Technologies (ICICT), IEEE*, 2018, pp. 92-95. https://doi.org/10.1109/ICICT43934.2018.9034296

[49] L. Breiman. (2001). Random forests, *Machine Learning*, 45, 5-32. https://doi.org/10.1023/A:1010933404324

[50] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin. (2018). CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*.

[51] A.V. Dorogush, V. Ershov, A. Gulin. (2018). CatBoost: gradient boosting with categorical features support. arXiv:1810.11363 (2018). https://doi.org/10.48550/arXiv.1810.11363

[52] H.-G.T. Hoang, T.-A. Nguyen, H.-B. Ly. (2025). Interpretable machine learning models for concrete compressive strength prediction. *Innovative Infrastructure Solutions,* 10, 5. https://doi.org/10.1007/s41062-024-01808-8

[53] T. Chen, C. Guestrin. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACMSIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. Association for Computing Machinery*.https://doi.org/10.1145/2939672.2939785

[54] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*.

[55] B.G. Marcot, A.M. Hanea. (2021). What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?. *Computational Statistics*, 36, 2009-2031. https://doi.org/10.1007/s00180-020-00999-9

[56] G. Rong, S. Alu, K. Li, Y. Su, J. Zhang, Y. Zhang, T. Li. (2020). Rainfall induced landslide susceptibility mapping based on bayesian optimized random forest and gradient boosting decision tree models—a case study of Shuicheng county, China. *Water*, 12(11), 3066. https://doi.org/10.3390/w12113066

[57] M. Kuhn, K. Johnson. (2013). Applied Predictive Modeling. *Springer*.

[58] S.M. Lundberg, S.-I. Lee. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*.

[59] A.C. Cameron, F.A.G. Windmeijer. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, 77(2), 329-342. https://doi.org/10.1016/S0304-4076(96)01818-0

[60] H.-G.T. Hoang, T.-A. Nguyen, H.-B. Ly.

(2025). Interpretable machine learning models for concrete compressive strength prediction. Innovative Infrastructure Solutions, 10, 5. https://doi.org/10.1007/s41062-024-01808-8