



Novel base predictive model of resilient modulus of compacted subgrade soils by using interpretable approaches with graphical user interface

Loai Alkhattabi^a, Kiran Arif^{b,c,*}

^a Department of Civil and Environmental Engineering, College of Engineering, University of Jeddah, Jeddah 23890, Saudi Arabia

^b Department of Computer Science, COMSATS University Islamabad, Wah Campus, Islamabad 47040, Pakistan

^c Western Caspian University, Baku, Azerbaijan



ARTICLE INFO

Keywords:

Resilient modulus
Subgrade
Graphical user interface
Machine learning
Statistical measures

ABSTRACT

The flexible pavement system consists of layers that are made up of a blend of aggregates and bitumen. To design flexible pavement systems that are both safe and environmentally sustainable, it is essential to have an accurate understanding of the resilient modulus (M_r) of the compacted subgrade soil. M_r refers to the ability of the soil to resist deformation under repeated loads. Thus, a critical parameter affects the performance and longevity of pavement systems. This study employs machine learning (ML) algorithms such as individual and ensemble learners using an extensive database of 2813 data points. These include dry unit weight, weighted plasticity index, confining stress, deviator stress, moisture content, and the number of freeze-thaw cycles (FT). The individual or weak learners were incorporated to create strong and robust ensemble learners by employing techniques such as bagging, adaptive boosting, and random forest (RF). Ensemble learning methods were used to improve the performance of individual learners, such as support vector machine (SVM) and decision tree (DT), by combining their predictions. To achieve the highest R^2 value, a total of twenty bagging and boosting submodels were trained and optimized. The validation of the test data was carried out through K-Fold cross-validation, utilizing metrics such as R^2 , MAE, and RMSE. The developed models were rigorously tested using statistical indices (MAE, MSE, RMSE, and RMLSE) to verify their predictive accuracy, reliability, and trustworthiness. The findings indicate that the integration of bagging and boosting techniques improves the efficiency of individual machine learning (ML) models. The combination of RF and DT utilizing bagging resulted in the most reliable performance, achieving an R^2 value of 0.9 and demonstrating minimum errors. In general, the implementation of the ensemble algorithm in ML improved the overall prediction accuracy of the model. Sensitivity analysis reveals that the prediction of the resilient modulus (M_r) of the subgrade is primarily influenced by dry density, confining stress, and deviator stress. Moreover, a graphical user interface (GUI) is developed for practical implantation.

1. Introduction

The pavement system consists of several layers having different material properties. These layers typically include the surface, base, subbase, and subgrade [1–3]. The strength and stiffness of these layers play an important role in designing the pavement system [4–6]. The subgrade layer is called the foundation of the pavement system [7]. It transfers applied loads to the ground. However, seasonal fluctuations and cyclic or dynamic loads on pavement layers cause stresses in these layers [8]. Moreover, the primary support for all pavements is due to the underlying subgrade layer. This layer typically consists of varying soil

combinations is considered as the fundamental base of the pavement structure, and is responsible for transferring applied loads to the ground [9]. Thus, resilient modulus (M_r) serves as a vital parameter for assessing subgrade material's behaviour under various environmental and load conditions [10–13]. Additionally, it explains the material's inelastic response to traffic loads. As pavement layers endure repetitive traffic loads, the subgrade soil undergoes both reversible and irreversible strains with each load cycle [12]. With an increasing number of load repetitions, the extent of plastic deformation diminishes until it becomes recoverable. Therefore, M_r is defined as the ratio of the applied deviator stress to the recoverable strain. This ratio serves as a descriptor for the

* Corresponding author.

E-mail addresses: laalkhattabi@uj.edu.sa (L. Alkhattabi), kiranarif12345@gmail.com (K. Arif).

Table 1

Equations for predicting resilient modulus (M_r) of soils through generalized regression.

References	Soil	Equation
Kim [34]	A-4 and A-6 soils	$\frac{M_r}{P_a} = k_1 \left[\frac{P_a \cdot \sigma_{oct}}{\tau_{oct}^2} \right]^{k_2}$
	A-7–6 soils	$M_r = k_1 \cdot P_a \left[\frac{9P_a}{2} \left(\frac{1}{3\sigma_d} + \frac{\sigma_3}{\sigma_d^2} \right) \right]^{k_2}$
Universal model [34,35]	All types of soils	$M_r = k_1 \cdot P_a \left(\frac{\theta}{P_a} \right)^{k_2} \left(\frac{\tau_{oct}}{P_a} + 1 \right)^{k_3}$
Pezo and Hudson [36]	Fine grain materials	$M_r = k_1 \sigma_3^{k_2} \nu_d^{k_3}$

M_r = resilient modulus; τ_{oct} = octahedral shear stress
 P_a = atmosphere pressure; θ = bulk stress
 σ_{oct} = octahedral normal stress σ_d = deviator stress
 k_1 = regression coefficients $k_2 \& k_3$ = regression coefficients

material's behaviour, providing insights into its deformation and recovery characteristics under stress. The M_r of the soil has a direct relationship with both fatigue cracking and permanent deformation. Hence, considered an essential property for understanding pavement behaviour and its design [14–17].

The M_r of compacted pavement subgrade soils depends upon multiple factors. These factors encompass a range of characteristics and conditions that affect the soil's ability to withstand and recover from applied loads [12]. Some key factors include the soil's composition, density, moisture content, stress history, temperature, and loading frequency [18–20]. Each of these factors contributes to the overall resilient modulus of the subgrade soils and plays a significant role in determining their response to traffic loading and subsequent pavement performance [21]. Moreover, the seasonal environmental effects are also important in regions with seasonal freezing. Typically, subgrade soils are compacted at their maximum dry unit weight and optimal moisture content. The moisture content can vary seasonally due to environmental factors, such as precipitation, surface infiltration, plant transpiration, evaporation, temperature fluctuation, and groundwater table variation [22–24]. In regions where seasonal freezing occurs, FT cycles can cause substantial harm to pavements. Thus, resulting in pavement surface spalling and cracking [25]. This can also lead to severe settlement caused by the influx of water during the spring thaw. These issues stem from the decline in subgrade soil strength and stiffness, as well as an increase in water content caused by FT. The M_r and strength of pavement materials degrade during FT processes. This results in a significant increase in tensile strain on the surface layer of the pavement when subjected to traffic loading [26]. Moreover, the increase in tensile strain can cause fatigue cracks to form on the pavement surface, which can lead to further damage if left unaddressed. Thus, permanent deformation can develop, resulting in rutting of the pavement.

Furthermore, when pavement materials experience desiccation (drying out) and thermal shrinkage (reduction in volume due to temperature changes), longitudinal and transverse cracks may appear. The M_r refers to a measure of the elastic modulus (EM) that defines the non-linear stress-strain properties of subgrade materials. This is an important factor in the structural response of pavements [27–30]. M_r considers the effects of various stress states, including traffic loading, confining stress, cyclic stress, and deviator stress. In addition, M_r can be determined through various laboratory testing methods, which include in-situ testing, cyclic tri-axial load testing, resonant column testing, and torsional shear testing [31–33]. Despite their effectiveness, these methods are expensive and complex. Nevertheless, incorporating M_r in the design and structural analysis of multi-layer pavement systems has been recommended by various codes, including ME PDG, AASHTO, and NCHRP. Moreover, it is imperative to consider multiple properties of the materials in the unbound layers for describing the resilient behaviour of pavement, including several soil hydrological and physical characteristics. Thus, it is critical to gain an accurate understanding of the factors

that precisely affect M_r . As a result, many studies have been conducted that led to the proposal of several constitutive models that estimate M_r of soil properties, applied loads, and stress states. Typically, these models are developed using regression analysis, as demonstrated in Table 1. It provides a collection of equations that are used for the estimation of the M_r of soils.

These equations rely on key variables influenced by different stress states, such as σ_2 and σ_3 , which vary across soil samples. These stress states vary among soil samples in the laboratory and play a crucial role in the determination of the updated M_r value for each soil sample. The values of k_1 , k_2 , and k_3 are obtained by fitting the prediction model to the laboratory-generated M_r test data, using either linear or nonlinear regression analyses. It is important to note that k_1 cannot be negative, as M_r cannot have negative values. On the other hand, k_2 should always be positive, as an increase in the confining stress typically results in a stiffening effect on the material. This will lead to a higher value for M_r . Conversely, k_3 must be negative, as an increase in deviator or shear stress results in a softening effect on the material. Finally, a second set of regression analyses is conducted to establish the relationship between these k -coefficients and various soil physical properties, such as moisture content, dry density, plasticity index, liquid limit, coefficient of curvature, uniformity coefficient, and percent passing #200 sieve [37–39]. The details on how these models are obtained can be found in previously published literature. Moreover, it is clear that the aforementioned analysis method is complex, and there are other drawbacks to using regression analysis. Although the model developed through regression analysis performs well on selected data sets. Therefore, its effectiveness is limited to the range of datasets used in the analysis. Additionally, existing models lack validation and testing on new data [40–42]. Given the shortcomings of these models and the intricate behaviour of M_r , there is a need to develop more effective prediction models for estimating M_r in pavement subgrade soils. Several researchers have used traditional linear and nonlinear equations based on statistical analysis to provide prediction measures of resilient modulus. However, accurate prediction is difficult, and more research is required to overcome these challenges. In recent years, neural-based ML techniques have been used to overcome challenges and improve the accuracy of predictions in various applications. [43–52].

The use of ML techniques is becoming increasingly popular in civil engineering, particularly in the prediction of mechanical properties of materials like concrete, asphalt, and resilient modulus (M_r) value [53–60]. These methods rely heavily on extensive datasets to build accurate models. The success of these models is contingent upon both the quality and quantity of the data used during the model development process [61–63]. In other words, having a substantial and high-quality dataset is crucial for the effectiveness of these ML methods [64–68]. Javed et al. [47] utilized gene expression programming (GEP) to forecast the compressive strength (CS) of sugarcane bagasse ash (SCBA) concrete. Likewise, Aslam et al. [46] used the GEP approach to predict the strength of high-strength concrete (HSC). Their findings indicated a strong correlation between the predicted values and target values. In addition, researchers have employed artificial neural networks (ANNs) to forecast the strength of concrete made of waste and recycled materials, and their finding reveals that ANNs have shown a strong relationship with fewer errors [69]. Zaman et al. [70] conducted a study exploring the construction of different ANNs. The goal was to establish a relationship between M_r and standard subgrade soil characteristics, along with stress states encountered in pavement design. Similarly, Zou et al. [71] employed GEP and ANNs for the estimation of compressed subgrade soils and reported robust results. Deswal and Pal [72] evaluated the performance of two machine-based extreme learning regression models for calculating the resilient modulus (M_r) of cohesive soils. The author compared their results with those obtained from SVM methods and demonstrated a robust performance by these models. In addition, Sadrossadat et al. [73] used a neuro-fuzzy adaptive interface system to forecast the behaviour of subgrade soils in flexible pavements. ML

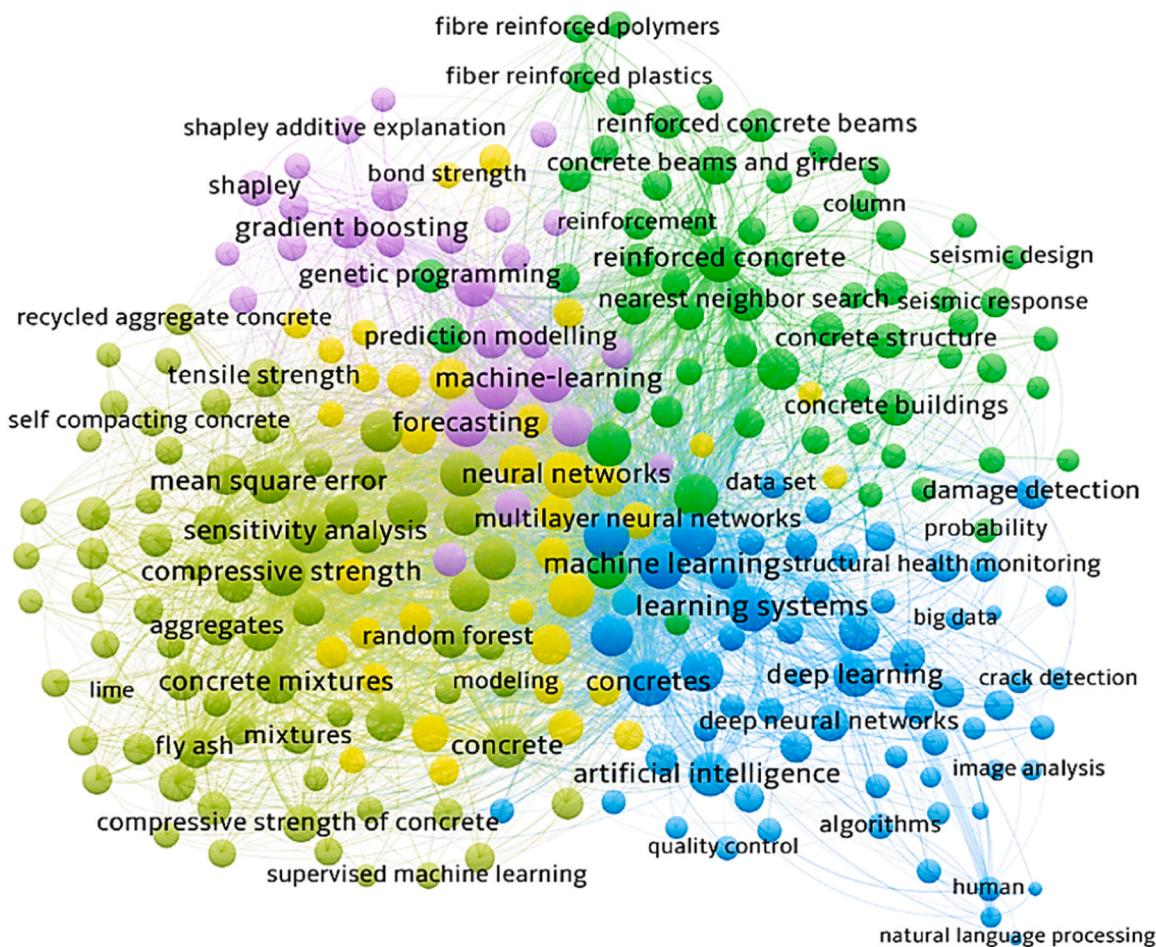


Fig. 1. MLA Bibliometric Analysis.

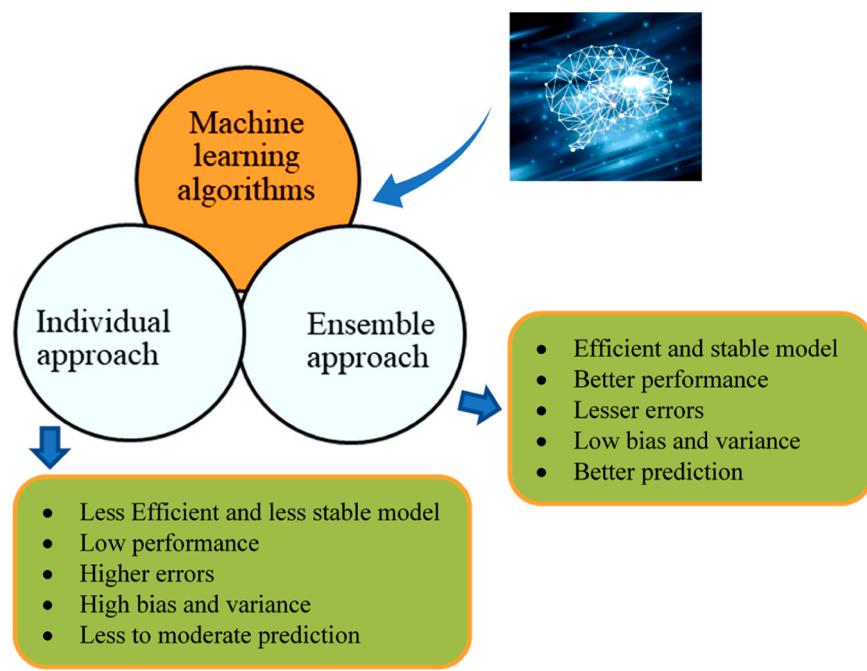


Fig. 2. Comparison between approaches.

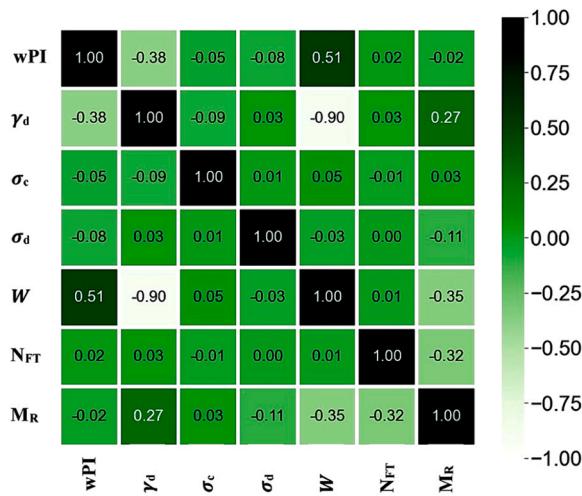


Fig. 3. Heat map correlation.

algorithms are versatile tools that can be utilized to predict a range of responses in various engineering or data science fields, beyond the compressive or tensile nature of concrete. These examples highlight the promising potential of ML algorithms in various domains. Moreover, the significance of MLAs in civil engineering domain is illustrated in Fig. 1.

Recently, there has been an increasing inclination towards the adoption of ensemble modelling methods to improve the overall effectiveness of models. This is achieved by combining individual or weak learners to build much stronger predictive learners [74]. Feng et al. [75] used ensemble learning techniques to predict the properties of reinforced concrete (RC) structural elements and observed a strong and robust performance of the models. Bui et al. [76] also used a modified firefly algorithm in combination with an ANN to forecast the behaviour of high-performance concrete (HPC). The results indicate that the hybrid model performed better and gave a strong performance. Similarly, Salami et al. [77] utilized the RF approach and achieved a good accuracy of $R^2 = 0.9867$. Cai et al. [78] predicted the penetration of chloride in RC structures located in marine environment using several supervised machine ensemble algorithms.

Halil et al. [58] applied three ensemble modelling techniques, using a decision tree (DT) as the base learner. The hybrid model demonstrated superior predictive performance for HPC strength compared to other models, resulting in a robust R^2 value of 0.9368. In a study conducted by Kermani et al. [59], the efficiency of five soft computing base learners was systematically assessed with the aim of predicting concrete corrosion in sewer systems. The study incorporated both tree-based and network-based learners. The notable outcome was the superior performance of Random Forest (RF) ensemble learners compared to the other models under consideration, demonstrating a substantial R^2 value of 0.872. This suggests that ensemble modelling approaches, such as RF, demonstrated enhanced impact and robust performance in predicting concrete corrosion. The study indicates that ensemble learning models, which combine multiple models, exhibit more desirable characteristics and yield superior results compared to individual learning models. This is further illustrated in Fig. 2, highlighting the distinction between individual and ensemble models.

Complex real-world challenges often necessitate the presence of intelligent systems that exhibit human-like expertise within a particular domain. These systems should also possess the capability to adapt to dynamic environments and provide explanations regarding their decision-making processes and actions. The objective of this study is to evaluate the accuracy of predicting resilient modulus (M_r) of compacted subgrade by employing different techniques, including individual and ensemble approaches. To accomplish this, data points from the literature were utilized. The programming tools used for modelling were

Anaconda Spyder and Jupyter Notebook. The model parameters included weighted plasticity index, confining stress, dry unit weight, deviator stress, moisture content, and the number of FT cycles, with M_r as the output parameter. Contour graphs were generated to demonstrate the correlation between input and output parameters. Moreover, this study has made a notable contribution to the advancement of a novel and efficient methodology for predicting the resilient modulus. By introducing a potential pathway in the form of a graphical user interface (GUI), integrated strategies within the field of transportation can be effectively utilized. Additionally, Shapley analysis was employed to determine the effectiveness of each variable in achieving the desired output. Furthermore, statistical metrics were employed to assess the accuracy of the model.

2. Data description

Data description is an important step in the process of building machine-learning models [79]. It involves thoroughly understanding the dataset and its characteristics before delving into the modelling phase. Thus, the M_r of compacted subgrade soil has been modelled using data obtained from previously published research (Supplementary file). Furthermore, instances (data points) used for modelling will have higher reliability and accuracy of the model. The database consisted of a total of 2813 experimental test results with resilient modulus as the response parameter and six explanatory variables namely as the weighted plasticity index (wPI), deviator stress (σ_d in kPa), moisture content (w in %), dry unit weight (γ_d in kN/m³), the number of FT cycles (N_{FT}), and confining stress (σ_c in kPa).

2.1. Database presentation

Python programming based on Anaconda version 3.7 was employed to represent the database and the relation of input parameters to its output model. Fig. 3 represents the heat map correlations between the resilient modulus and the input variables. It can be seen that γ_d shows direct and positive relation to M_r . A higher dry unit weight indicates a denser and more compacted subgrade, which tends to exhibit higher stiffness and load-bearing capacity. However, moisture content and the number of FT cycles have adverse effects. This is because excess moisture softens the subgrade, reduces cohesion, and induces swelling or shrinkage, while FT cycles cause frost heave, thaw weakening, and cracking as illustrated in Fig. 3. Moreover, Fig. 4 depicts the distribution of input parameters and their corresponding outputs, illustrating their relationship.

The contour graphs depict the density of each input parameter concerning M_r , with darker regions indicating a higher concentration linked to concrete strength in the database. The weighted plasticity index ranges from 5.82 to 31.08, with mean and median values around 13.88 and 382.50, respectively. Confining concentrations vary from 0 KPa to 41.4 KPa, representing extremes in the dataset. Analyzing the distribution of explanatory variables aids in constructing a generalized model [40]. The dataset's highest and lowest values illustrate its extremes, while mean, median, and mode serve as indicators of central tendencies. Standard deviation (SD) measures data variability, with a smaller SD indicating a more concentrated distribution and a larger SD suggesting a broader and potentially more diverse distribution. This statistical metric provides insights into the spread or dispersion of data points, aiding in the understanding of the data's overall distribution pattern. Skewness and kurtosis evaluate normal probability distribution regularity and profile [61]. Skewness is a statistical metric used to evaluate the asymmetry within a probability distribution, with zero suggesting perfect symmetry, positive values indicating a longer right tail, and negative values a longer left tail.

Kurtosis serves as a statistical metric characterizing the form of a probability distribution or the extent of "tailedness" in the distribution curve of a dataset. Positive kurtosis (leptokurtic) and zero kurtosis

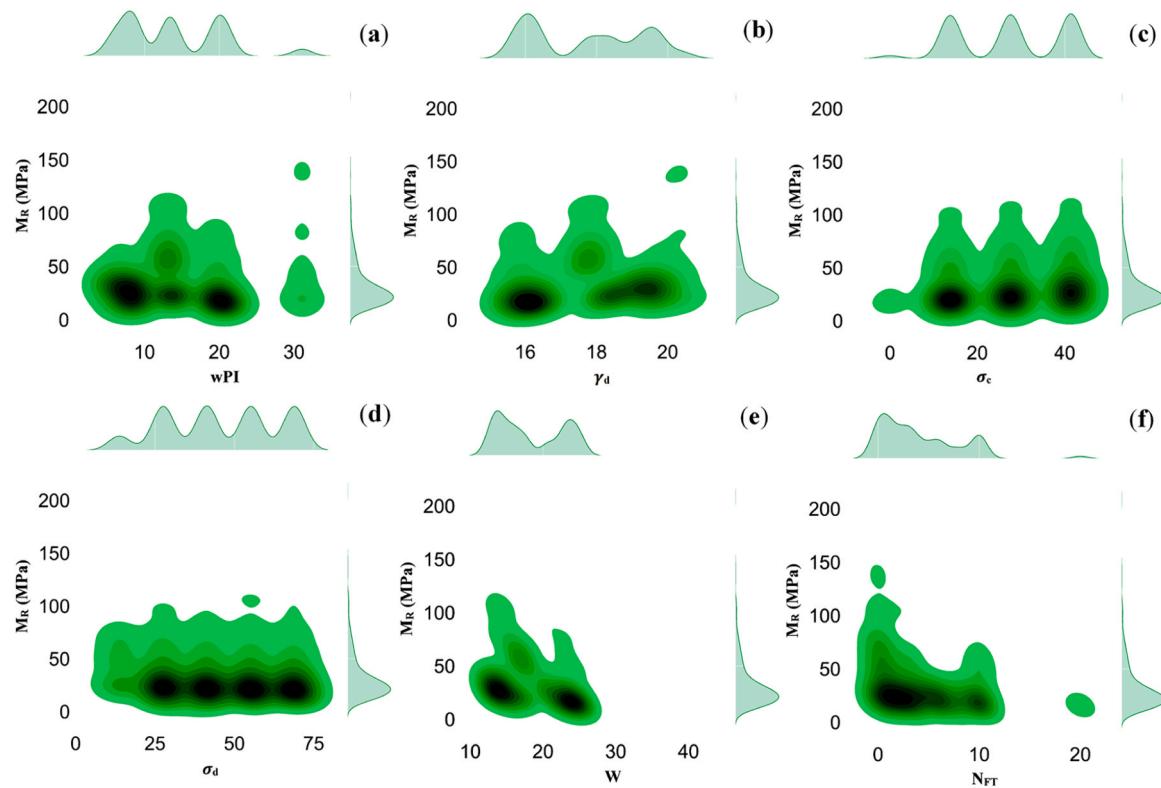


Fig. 4. Relation of parameters to resilient modulus; (a) weighted plasticity index; (b) dry unit weight; (c) confining stress; (d) deviator stress; (e) moisture content; (f) number of FT cycles.

Table 2
Data Statistics for Model Development.

	wPI	γ_d (kN/m ³)	σ_c (kPa)	σ_d (kPa)	w (%)	N_{FT}
count	2813.00	2813.00	2813.00	2813.00	2813.00	2813.00
mean	13.88	17.73	27.17	45.64	18.36	4.13
std	6.43	1.56	11.85	17.34	4.52	3.93
min	5.82	15.50	0.00	13.80	12.30	0.00
25 %	8.28	16.16	13.80	27.60	13.90	1.00
50 %	13.16	17.77	27.60	41.40	17.30	3.00
75 %	19.58	19.15	41.40	55.20	23.10	6.00
Standard error	0.12	0.029	0.22	0.32	0.08	0.07
Sample Variance	41.40	2.42	140.4	300.6	20.4	15.4
Kurtosis	0.00	-1.5	-1.19	-1.11	-0.99	1.9
Skewness	0.71	0.08	-0.14	-0.13	0.35	1.20
Range	25.26	4.9	41.4	55.1	29.2	20
max	31.08	20.40	41.40	68.90	41.54	20.00

Table 3
Pearson Correlation Matrix for Utilized Data in Models.

	wPI	γ_d (kN/m ³)	σ_c (kPa)	σ_d (kPa)	W (%)	N_{FT}	M_R (MPa)
wPI	1						
γ_d (kN/m ³)	-0.37589	1					
σ_c (kPa)	-0.04884	-0.08613	1				
σ_d (kPa)	-0.07923	0.03045	0.01414	1			
W (%)	0.51245	-0.90425	0.05001	-0.03178	1		
N_{FT}	0.02234	0.02995	-0.01262	6.72759E-4	0.00606	1	
M_R (MPa)	-0.02486	0.26998	0.03292	-0.11429	-0.35437	-0.32427	1

(mesokurtic) are characteristics associated with a normal distribution, signifying a certain level of peakedness in the distribution curve. Positive kurtosis (leptokurtic) suggests a more peaked curve, indicating heavier tails and a distribution with more extreme values than a normal distribution. In contrast, negative kurtosis (platykurtic) suggests a flatter

data curve, indicating a distribution with lighter tails compared to a normal distribution [62]. Understanding kurtosis provides valuable insights into the shape and tails of a probability distribution, contributing to a comprehensive analysis of the data's characteristics. Table 2 shows that selected inputs and outputs align with suggested skewness and

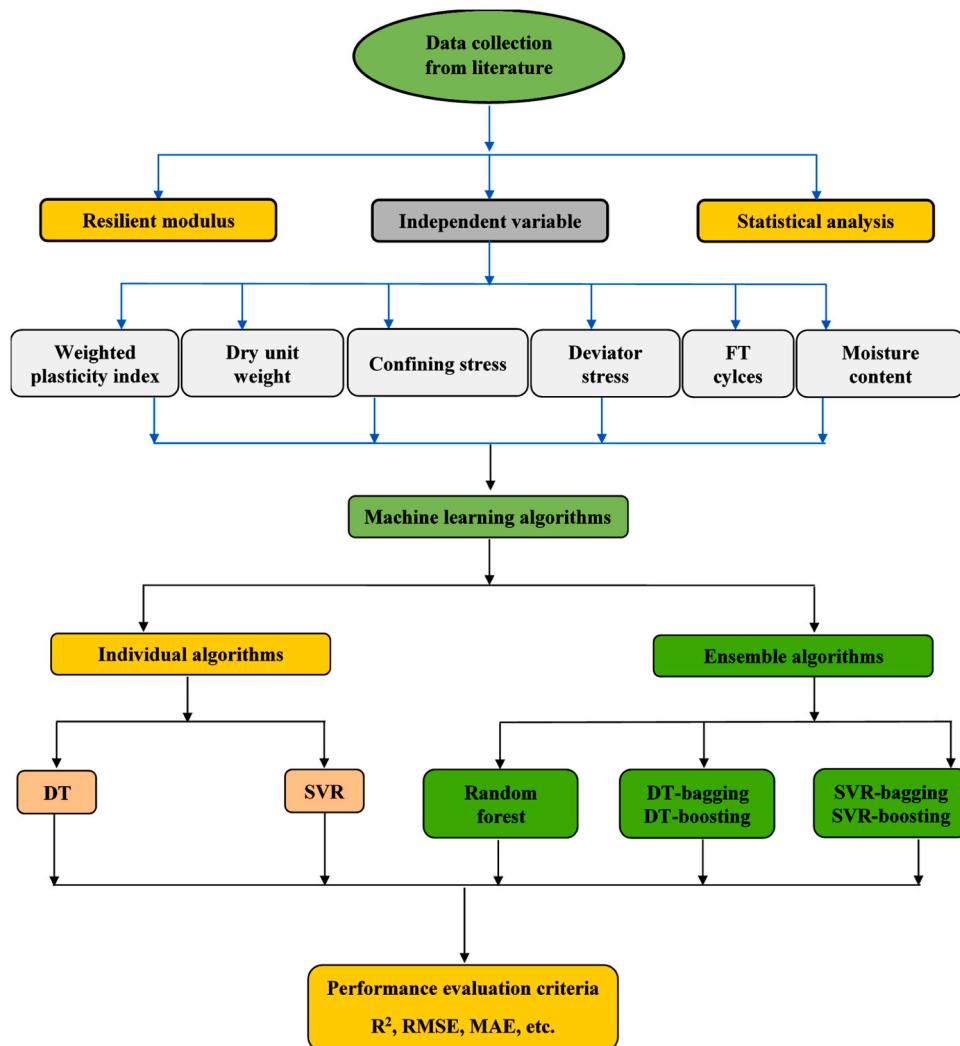


Fig. 5. Research flow chart with details used in making models.

kurtosis ranges, enhancing data distribution. Table 3 represents the Pearson correlation coefficient (r), an important statistic in the context of artificial intelligence (AI) modeling. This coefficient quantifies the correlation between input and output variables, shedding light on the strength of their association. Ranging from -1 to 1 , a value of 0 indicates no correlation, while -1 and $+1$ signify strong negative and positive correlations, respectively. The correlation matrices in Table 3 show a strong connection, without the issue of multi-collinearity. It's crucial to note that in this study, AI methods successfully handle multi-collinearity when modeling M_r . Python programming, along with Seaborn, is employed to implement ML algorithms and generate graphs. It provides a clearer understanding of how changes in input variables relate to changes in output strength, which ultimately helps in making predictions about the mechanical behavior of a system.

3. Methodology

Currently, numerous industries utilize machine-learning algorithms (MLAs) to understand and estimate material behavior [80–84]. In this research, various ML-based approaches, namely support vector machines (SVM), decision trees (DT), and random forests (RF), are employed to assess the resilient modulus (M_r) of subgrade. These particular techniques were chosen due to their widespread application [85–87]. Moreover, the rationale behind selecting these approaches was driven by their distinctive advantages in handling the complexities.

DT approach has the ability to model non-linear relationships and interpret interactions. Similarly, RF improves prediction accuracy by aggregating multiple decision trees, and SVM capturing intricate non-linear patterns using kernel functions, making it adaptable to the diverse complex nature. The addition of ensemble benefits the models by variance reduction, model stability, predictive accuracy, effective handling of non-linear relationships, and robust adaptability. Thus, their use proves accuracy in predicting results in similar studies, and recognized efficiency. Moreover, an ensemble learning technique (ELT) is implemented to forecast the strength of M_r . The ensemble learning method combines the predictions from multiple ML models to enhance overall prediction performance [88]. Fig. 5 depicts the comprehensive flow diagram of the ML algorithm process that is used in our research.

4. Machine learning (ML) methods

4.1. ML Using SVM algorithm

The support vector regression (SVR) technique was first introduced by Vapnik in 1995 [89]. Since then, it has gained widespread use for performing classification, prediction, and regression tasks. Numerous studies have demonstrated the high learning capacity of SVMs in civil and structural engineering [90–92]. The approach utilizes labelled training data to generate support vector machines (SVMs). SVMs can be utilized for either binary classification, where there are two possible

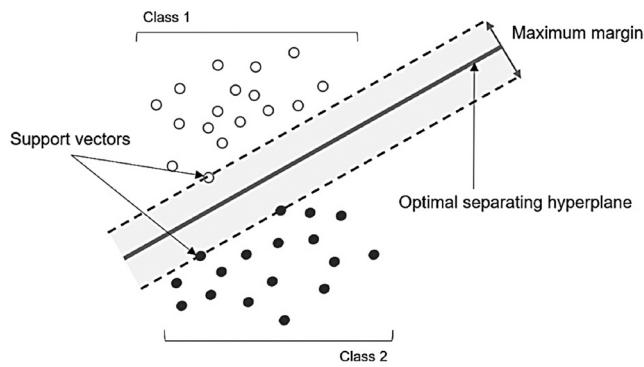


Fig. 6. SVM representation.

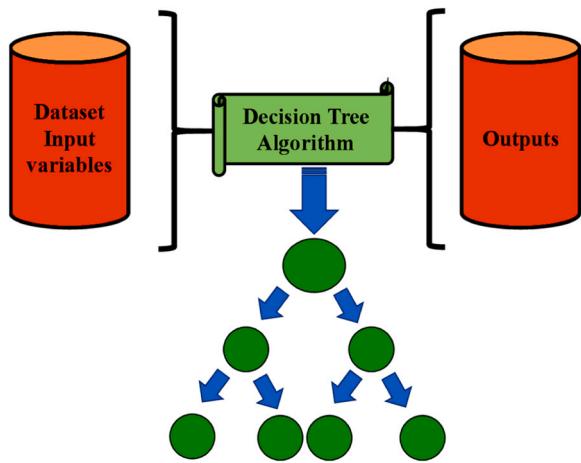


Fig. 7. DT schematic diagram.

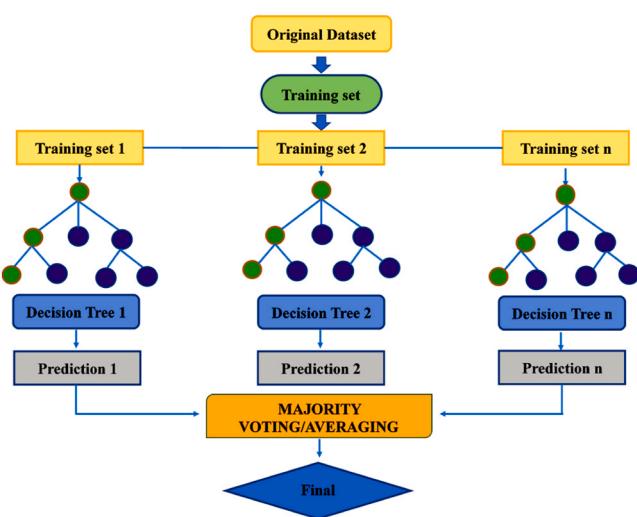


Fig. 8. RF algorithm schematic diagram.

outcomes (0 and 1), or for regression tasks, where the goal is to predict a continuous actual value. Given its proficiency in effectively addressing nonlinear regression challenges, the SVM regression model is frequently employed in input-output analysis. The first step in SVM regression involves mapping the data to an n -dimensional function space in a fixed manner. Nonlinear activation functions are subsequently applied to transform the input data into a more powerful space, creating a clear distinction from the initial space. The expression $f(x, w)$ is employed to

Table 4
Hyper parameters and their values used in constructing models.

Algorithms used	Hyper parameters		
	Title	Values considered	Optimal
Decision Tree	min samples split max depth min samples leaf max features	2-5 none, 1-5 1-5 auto, sqrt	4 None 3 auto
Support Vector Regression	Gamma	0.5, 0.1, 0.01	0.5
	Kernel	Linear, rbf, poly, sigmoid	rbf
Random Forest	C	1-2500	2250
	min samples leaf	1-5	1
	min samples split	2-5	2
	max depth	none, 1-5	None
	n estimators	20-200	60
	max features Bootstrap	auto, sqrt true, false	Auto True

denote the linear function within this transformed space.

$$f(x, w) = \sum_{j=1}^n w_j g_j(x) + b \quad (1)$$

In SVM regression, the function "g_j(x)" is employed to express the transformations of the nonlinear input space, the bias term "b," and the weight vector "w". The primary aim is to acquire optimal values for these parameters by maximizing the regularized risk function. This function considers the delicate balance between model accuracy and decision boundary complexity, making it a crucial aspect of achieving an effective and well-balanced model. The estimate's quality can also be measured using the loss function L, which can be expressed as follows.

$$L_e = L_e(y, f(x, w)) = \begin{cases} 0 & \text{if } |y - f(x, w)| \leq \epsilon \\ |y - f(x, w)| & \text{otherwise} \end{cases} \quad (2)$$

SVM regression is unique in its ability to accommodate a variety of loss functions, which enables the creation of a linear regression function with an extensive feature set while simultaneously reducing model complexity by minimizing $\|w\|^2/2$. This feature provides data scientists with a double benefit. The function uses nonnegative slack variables, denoted by ξ_i for $i=1, \dots, n$, representing the samples present in the π -insensitive area. This simplifies the function as follows, leading to the construction of SVM regression:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (3)$$

$$\text{subject to } \begin{cases} y_i - f(x_i, w) \leq \epsilon + \xi_i^* \\ f(x_i, w) - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, n \end{cases} \quad (4)$$

The optimization problem can be rephrased as a dual problem, and it can be addressed utilizing the following function.

$$f(x) = \sum_{i=1}^{n_{SV}} (\alpha_i + \alpha_i^*) K(x, x_i) \text{ subject to } 0 \leq \alpha_i^* \leq C, 0 \leq \alpha_i \leq C \quad (5)$$

where n_{SV} represents the number of support vectors. The kernel function is expressed as follows:

$$K(x, x_i) = \sum_{i=1}^m (g_i(x) + g_i(x_i)) \quad (6)$$

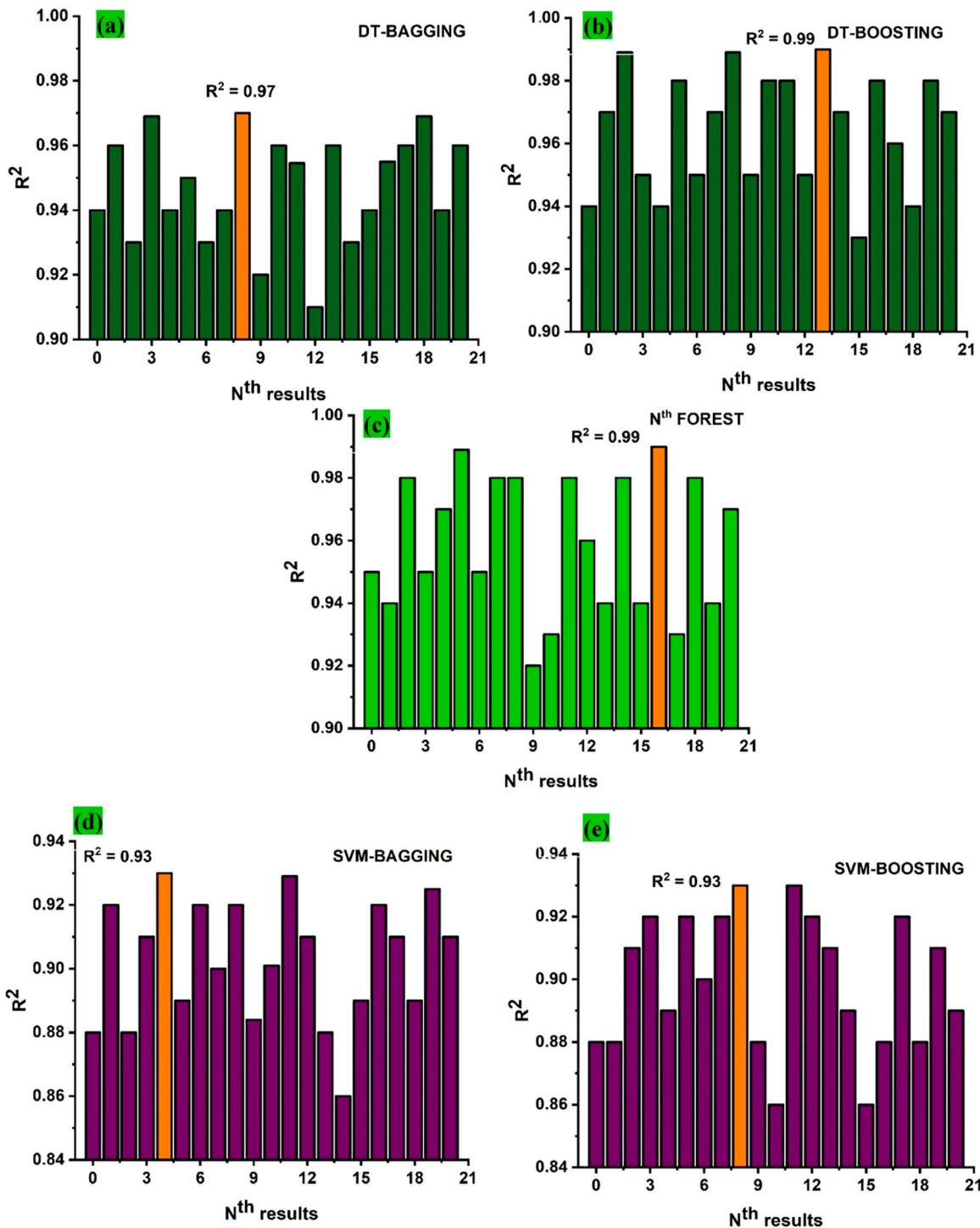


Fig. 9. Model outcomes with bagging and boosting; (a) DT-bagging approach; (b) DT-boosting approach; (c) Modified RF approach; (d) SVR-bagging approach; (e) SVR-boosting approach.

Various mathematical functions, including linear, radial basis, polynomial, or sigmoid functions, are employed during the training of SVM to detect support vectors on the function surface. The selection of the kernel function plays a crucial role in determining the complexity and flexibility of the decision boundary created by the SVM model [93]. Each kernel function has unique properties and is appropriate for different types of data, depending on the data set's specific characteristics and the modelling task's requirements. Moreover, Fig. 6 depicts the standardized SVM for the prediction of the properties.

4.2. Decision tree

A supervised learning technique used in ML, that constructs a tree-like model to make decisions based on a set of input variables [94]. It is commonly used for classification and regression tasks and has gained widespread popularity. In this method the data structure resembles a tree, having inner nodes and leaves. The "inner nodes" are the ones that have one or more branches leading to other nodes. Similarly, "leaves" are the nodes located at the end of a particular branch of the tree, and they do not have any outgoing branches. During the training phase of a

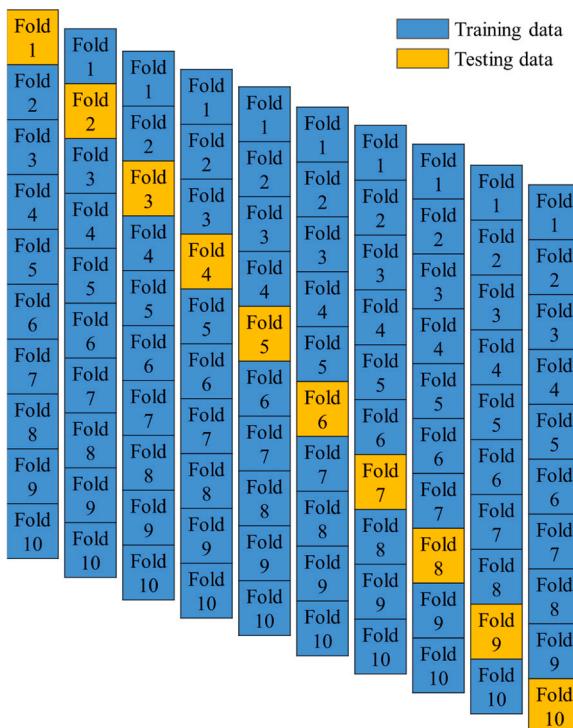


Fig. 10. Model Validation by K-fold cross-validation algorithm.

decision tree, each input variable is given a unique importance, and an "inner node" in the tree splits the input data into multiple classes using a function that is determined at this time. The decision tree algorithm generates the tree by minimizing a fitness function on a given dataset, identifying the optimal tree as illustrated in Fig. 7 [43]. In the absence of classes within the dataset, the algorithm adjusts by constructing a regression model for the dependent variable, taking into account different factors. At each division point in the tree, the algorithm evaluates the disparity between expected and actual fitness function values, distributing multiple points to each variable in the dataset. When all variables exhibit equal errors at the split point, the algorithm selects the variable with the lowest fitness function values as the decisive split point. Iterating through this process, the algorithm systematically constructs the optimal tree, ensuring a comprehensive evaluation of factors and minimizing errors in the model's predictive capabilities [95].

4.3. Random forest regression

Researchers have taken an interest in the RF model because of its classification and regression methodology that is used for predicting the mechanical properties of materials [96]. Shaqadan et al. [97] utilized the RF regression model to predict the compressive strength (CS) of concrete and noticed a reliable performance. Unlike DT, which produces only one tree, RF grows multiple trees to form a forest where different data subsets are randomly selected and distributed across each tree as demonstrated in Fig. 8 [97]. Every tree consists of data arranged in rows and columns with variable dimensions that can be defined. The development of every tree involves the following phases:

1. For each tree, a data frame is generated by selecting a random subset, constituting two-thirds of the entire dataset, through a technique called bagging. Predictor variables are then randomly chosen and used to split the nodes of the tree in the most precise manner possible.
2. The remaining data that was not selected is referred to as "out-of-bag" (OOB) data, which is used to evaluate the performance of each tree, which is known as the out-of-bag error. Following the estimation of the out-of-bag error for each tree, the errors from all trees are aggregated to determine the overall out-of-bag error rate. This is achieved by summing up the individual errors of every tree within the random forest model.
3. After each tree in the random forest algorithm generates its own regression model, the model selects a subset of trees based on the number of correct predictions, which are counted as 1, while incorrect predictions are counted as 0. The trees with the highest number of votes are then chosen for the final model. The predictions made by these selected trees are combined to obtain the overall prediction for the random forest model. This approach of combining predictions from multiple models helps to improve the accuracy and robustness of the final prediction.

4.4. Hyper parameter tuning

Choosing the right hyper parameters is vital for obtaining optimal performance when building ML models [54]. In this study, multiple parameter configurations for each algorithm are used based on previous literature [98–101], as shown in Table 4. This is done to improve the predictive accuracy of the models. Moreover, choosing the correct parameters is crucial when developing non-linear models, and evaluating different combinations to find the optimal configuration for accurate and reliable performance can be a time-consuming process.

Table 5
Statistical measures.

Equation of statistical indicator	Acceptable range	Reference
$R^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2}{\sum_{i=1}^n (Y_i - X_i)^2}$	Close to 1	[102]
$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - E_i)^2}{N}}$	MAE < RMSE	[103]
$MAE = \frac{1}{n} \sum_{i=1}^n E_i - P_i $		[104]
$NSE = 1 - \frac{\sum_{i=1}^n (E_i - P_i)^2}{\sum_{i=1}^n (E_i - \bar{E}_i)^2}$	Higher than 0.65 for very good model	
$RMSLE = \sqrt{\frac{1}{n} \sum_{i=1}^n [\log(P_i + 1) - \log(E_i + 1)]^2}$	Approaches 0 for a good model	[102]
$RSE = \frac{\sum_{i=1}^n (E_i - P_i)^2}{\sum_{i=1}^n (\bar{E}_i - P_i)^2}$		
$RRMSE(\%) = \frac{1}{ \bar{e} } \sqrt{\frac{\sum_{i=1}^n (E_i - P_i)^2}{n}} \times 100$	For excellent model (0–10) %; good model (11–20) %	[104]

n = data points, E_i = Experimental data, P_i = predicted data, \bar{E}_i = average experiment values, \bar{P}_i = average predicted values

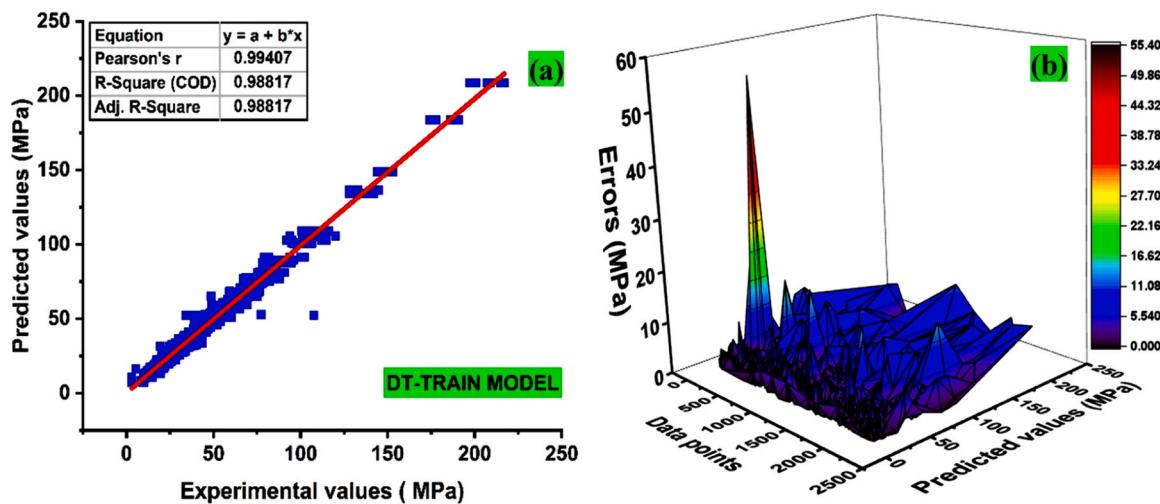


Fig. 11. DT outcomes; (a) train data; (b) train data discrepancies.

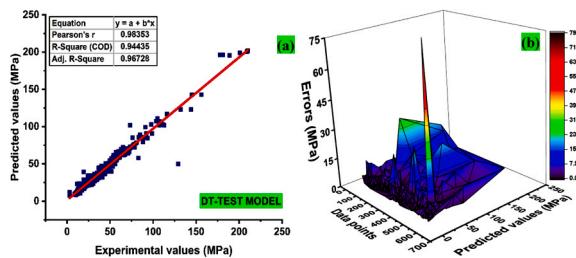


Fig. 12. DT outcomes; (a) test set; (b) test set discrepancies.

5. Ensemble techniques using bagging and boosting

Ensemble algorithm is an effective way to improve the prediction accuracy of ML models. The technique of combining weaker prediction models or component sub-models is known as ensemble learning. It involves using multiple models to make predictions on a given dataset. Although these individual models may be less accurate when used alone. Thus, combining them can help minimize over-fitting issues associated with the training data. By aggregating the predictions of multiple models, the ensemble can benefit from the strengths of each model and produce a more reliable and accurate overall prediction. One of the most widely used ensemble algorithms is known as bagging, which utilizes the bootstrap resampling and aggregation procedures. In bagging, component models replace the original training set, and bootstrap tests are conducted up to the size of the training set. This introduces a level of randomness, allowing certain data points to be utilized multiple times in the final models. The insights drawn from bagging involve averaging the results obtained from all individual models, providing a robust and aggregated prediction. Boosting, on the other hand, is an alternative method that assembles a cumulative model similar to bagging. It facilitates the creation of a larger number of components, achieving enhanced accuracy compared to a single model. In this method the weighted averages of dependent sub-models are incorporated to optimize their contribution to the overall model, offering a sophisticated approach to improve predictive performance. Therefore, this study utilizes SVM, DT, and RF regression to forecast the resilient modulus of subgrade, and various hyper-parameter configurations were incorporated to improve the models' accuracy.

5.1. Fine-tuning parameters for ensemble learners

The parameters that need to be tuned may include the number of

base models (or "learners") in the ensemble, the learning rate (which controls how quickly the model adapts to new data), and other significant features that have a substantial impact on the performance of individual base models within the ensemble. In this study, bagging and boosting ensemble models with varying numbers of component sub-models (ranging from 1 to 20) were built for each learner. Correlation coefficients were employed to determine the best structures for each model. The findings revealed that the ensemble model with boosting demonstrated a higher correlation coefficient for prediction accuracy compared to the individual models, as shown in Fig. 9. Furthermore, Fig. 9(a) and Fig. 9(b) present the outcomes of DT. In addition, Fig. 9(c) displays the results of the modified RF approach. Moreover, Fig. 9(d) and Figure (e) represents SVM models in the context of ensemble models, while

6. K-fold cross-validation approach

The k-fold cross-validation technique is often employed to overcome the bias of random sampling in training data. According to Kohavi's research, ten-fold validation tests provide both reliable variance and reasonable computation time for validation of the models. In this study, the effectiveness of a classification model aimed at categorizing a specified number of data samples into ten distinct subgroups was thoroughly assessed using stratified ten-fold cross-validation. This rigorous evaluation involved dividing the dataset into ten subsets, ensuring a balanced representation of each subgroup in both training and testing phases. During each of the ten iterations of model development and assessment, a fresh data subset was exclusively employed for testing, while the remaining subsets were utilized for training the model. As shown in Fig. 10, the model was validated using the test subset, and the mean precision value obtained by ten models across ten validation iterations was calculated to assess the accuracy of the method.

6.1. Statistical analysis for evaluating ML models

Furthermore, statistical measurements are employed to analyze the performance, accuracy, and dependability of models. Therefore, a total of eight statistical gauges were employed to evaluate the performance of the finalized models. These indicators include Root Mean Squared Error (RMSE), Nash-Sutcliffe Efficiency (NSE), Root Mean Square Logarithmic Error (RMSLE), Root Squared Error (RSE), Mean Absolute Error (MAE), Performance Index (PI), Percentage of Relative Root Mean Square Error (RRMSE%), and Determination Coefficient (R^2). The Nash-Sutcliffe Efficiency or, NSE, is a statistical metric that spans a range from negative infinity to one. A value of 1 signifies a complete alignment between the

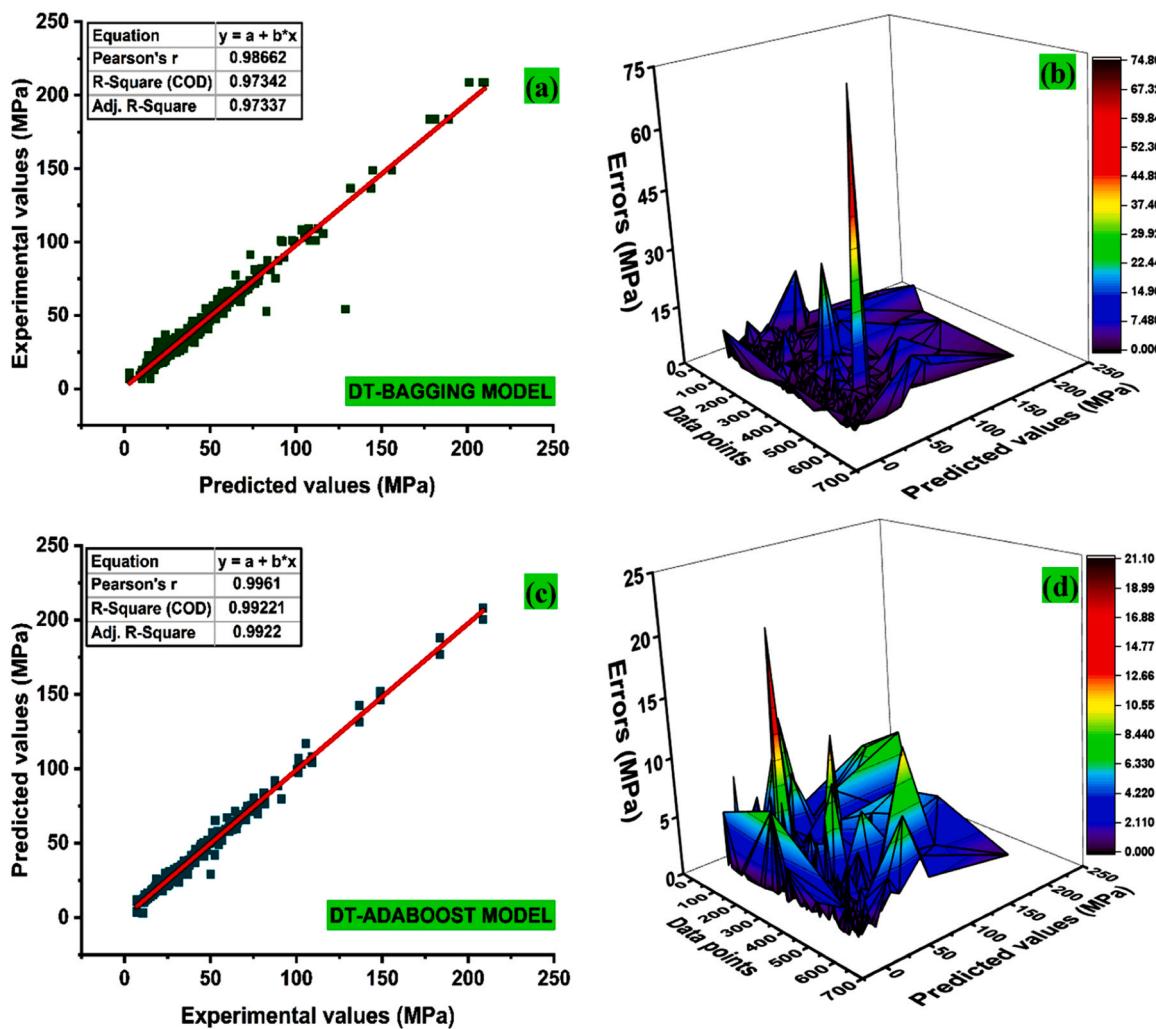


Fig. 13. DT ensemble; (a) bagging result; (b) bagging error results; (c) boosting results; (d) boosting set discrepancies.

model and the data. A correlation is deemed to be strong when the Nash-Sutcliffe Efficiency (NSE) surpasses 0.65, and the coefficient of determination (R^2) should reach 0.8. Typically, a smaller Mean Absolute Error (MAE) is considered more favorable than a higher Root Mean Square Error (RMSE), since it indicates a stronger predictive capability of the models. Furthermore, as the Root Mean Squared Logarithmic Error (RMSLE) approaches zero, it signifies a high level of proficiency exhibited by the model. To guarantee optimal model performance, it is advised to maintain the proportional-integral (PI) value below 0.2. Furthermore, the Root Relative Mean Squared Error (RRMSE) is utilized as a metric to evaluate the performance of a model. A RRMSE value ranging from 0 % to 10 % is considered indicative of exceptional and exceptional model performance. In order to classify a model as satisfactory, it is necessary for the Root Relative Mean Squared Error (RRMSE%) to fall within the range of 10–20 %. Table 5 presents the permissible intervals for the mathematical equations pertaining to the statistical assessment measures.

7. Results

7.1. DT ensembles-bagging and boosting

The outcomes of applying nonlinear regression to the dataset of resilient modulus are shown in Fig. 11. The DT model predictions yield more accurate and effective results by using ensemble algorithms like bagging and boosting (AdaBoost) as compared to the individual

approach. DT can capture non-linear relationships between input features and the target variable. They can learn complex decision boundaries by splitting the data into different regions based on feature values. To capture the results of the DT approach, the model was trained and evaluated using available data. The training process involved using 80 % of the available data. Hence, resulting in an impressive R^2 value of 0.98 (Fig. 11(a)). The R^2 value is a measure of how well the model fits the data. This indicates a stronger correlation with higher values. Moreover, the evaluation of a model is done by using its error distribution is illustrated in Fig. 11(b). Thus, provides valuable insights into the model's predictive performance and the quality of its predictions. It was observed that the model exhibited a small amount of error, with a maximum value of 55.20 MPa. The majority of the training data (99.15 %) had errors of less than 10 MPa. This indicates a good agreement between the predicted and actual values. Additionally, 0.66 % of the data exhibited errors in the range of 10–15 MPa, while 0.17 % had errors surpassing 15 MPa, as illustrated in Fig. 11(b).

In the training phase, the model learns to recognize patterns and relationships within the training data. It adjusts its internal parameters to optimize its ability to make accurate predictions. Testing the model after training is a vital stage in evaluating its performance. The training phase involves the model learning from labelled data, while the testing phase assesses the model's ability to make accurate predictions on new, unseen data. The testing phase is crucial for determining the model's generalization capabilities and overall performance. The result of the DT model (without ensemble techniques) using the test data (20 %) is

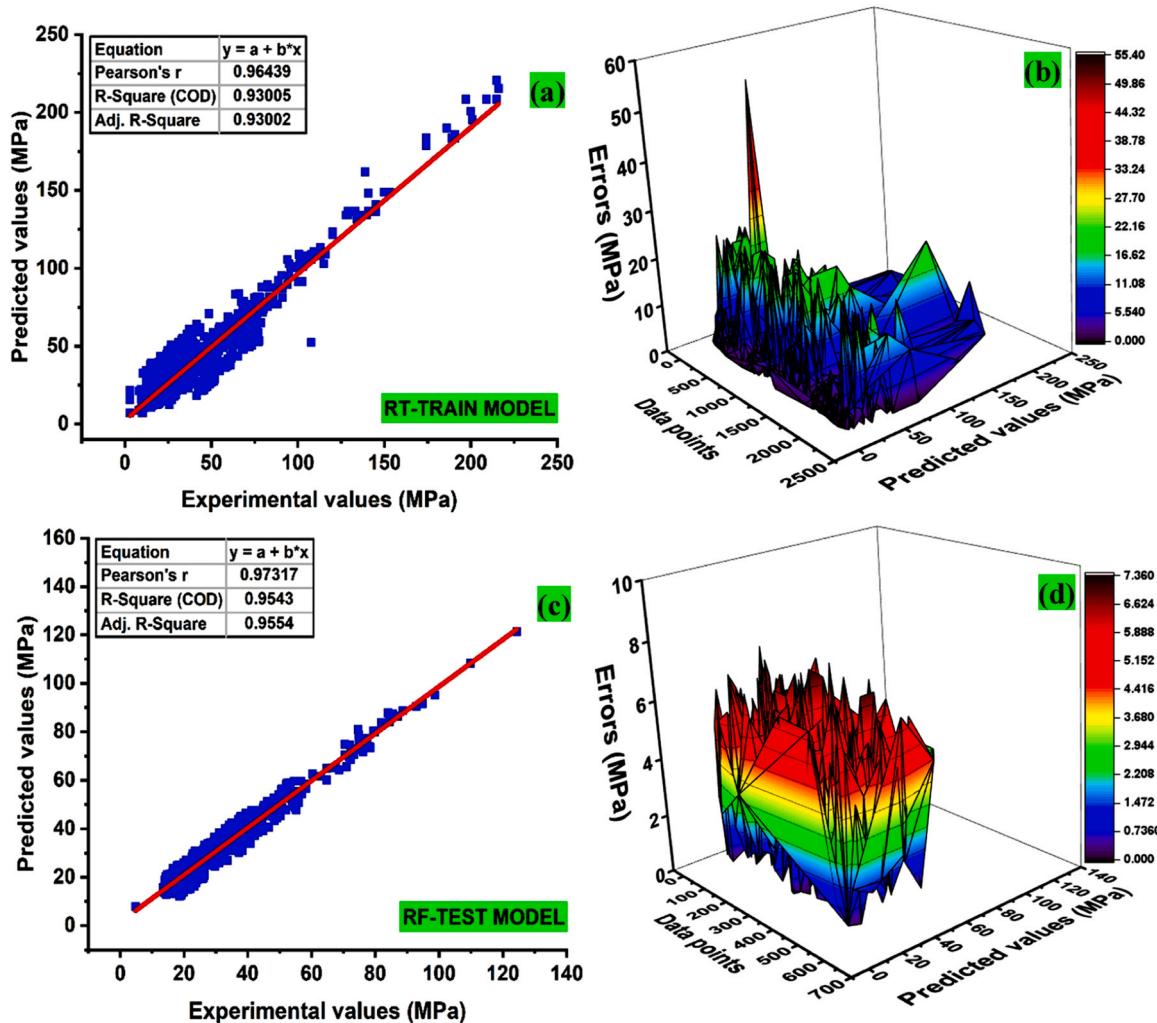


Fig. 14. RF model; (a) bagging result; (b) bagging error results; (c) boosting results; (d) boosting set discrepancies.

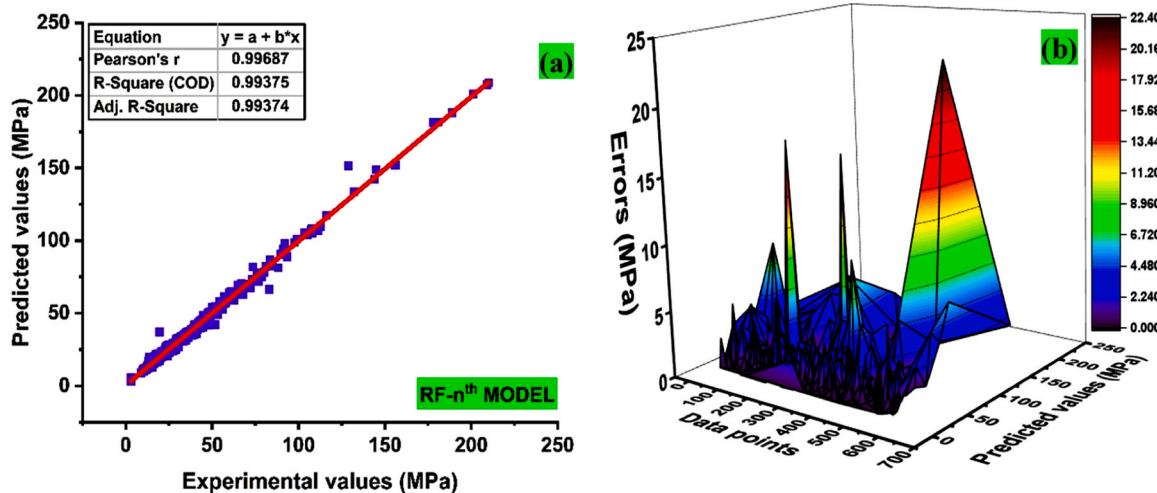


Fig. 15. RF model with n^{th} estimators; (a) n^{th} RF result; (b) n^{th} RF error results.

displayed in Fig. 12(a). The graph indicates a strong correlation to the output with an R^2 value of 0.944. Upon assessing the test set after training the model, it was found that the model demonstrated a low level of error, with the maximum error reaching 78.96 MPa. The majority of the testing data (98.04 %) exhibited errors smaller than 10 MPa. Thus,

indicating a high level of agreement between the predicted values and the actual values. A small proportion of the data (0.71 %) had errors falling between 10 and 15 MPa. Furthermore, a mere 1.24 % of the data exhibited errors exceeding 15 MPa as illustrated in Fig. 12(b).

These findings demonstrate the effectiveness of the individual DT

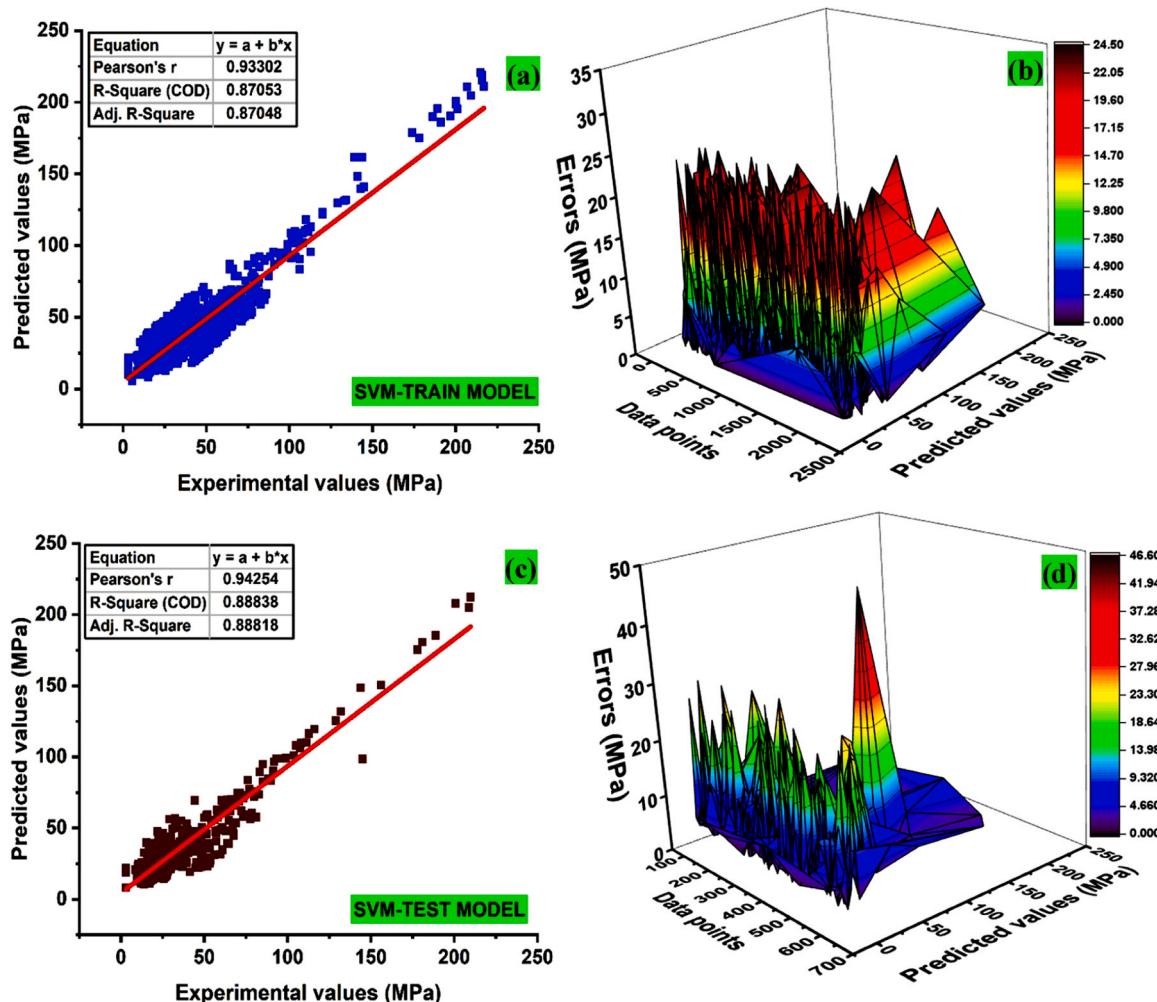


Fig. 16. SVM outcomes; (a) train set; (b) train set discrepancies; (c) test set; (d) test set discrepancies.

model in accurately predicting the output variable. The high R^2 value on the test data suggests that the DT model generalized well beyond the training data. Therefore, indicating its robustness and ability to capture the underlying patterns in the dataset. Similarly, individual models show bias, variance, and overfitting within the model. These can be reduced by ensemble approaches as they help in the mitigation of overfitting and other errors by predictions from multiple models. This collectively provides a more balanced and robust estimate of the target variable. The individual DT method is ensemble with the bagging method as depicted in Fig. 13.

The ensemble nature of algorithms makes it less sensitive to outliers in the data, if it exist in the model. Outliers that may heavily have a significant impact on the predictions of the model are likely to have less influence. As a result, the overall predictions tend to be more robust and less influenced by extreme data points. Fig. 13(a) illustrates the benefits of using ensemble techniques such as bagging, which resulted in more accurate predictions and error distribution (Fig. 13(b)). The utilizing of the bagging as an ensemble technique yielded a highly favourable outcome with a low testing data error, and a substantial R^2 value of 0.97 as depicted in Fig. 13 (a). Similarly, residual error distribution indicates that 98.40 % of the dataset has an error distribution ranging from 0 to 10 MPa. Whereas, 1.06 % has an error between 10 and 15 MPa, and another 0.53 % has an error values greater than between 15 MPa. Similarly, utilization of adaboost on DT harvests prominent outcome as depicted in Fig. 13(c). The DT model combined with adaboost technique exhibited a strong correlation in prediction. This is attributed to the impact of strong learners on the prediction component, as shown in

Fig. 13(c) with its residual errors as shown in Fig. 13(d). DT adaboost exhibited higher accuracy than DT bagging due to its training process with each subsequent model aiming to correct the mistakes made by the previous models with $R^2=0.99$. Moreover, the results show that 99.19 % of test data errors were less than 10 MPa. While, only 0.71 % has an error distribution in the range of 10 and 15 MPa, and 0.17 % has an error values greater than between 15 MPa as depicted in Fig. 13(d).

7.2. Random forest outcome

Random forests (RFs) are a powerful type of EMLAs that combines the principles of bagging and random feature selection. This combination results in prediction models that are highly effective in terms of accuracy and generalization. While also being user-friendly and easy to interpret. The combination of bagging and random feature selection in RFs yields several advantages. First, it helps to reduce the variance of the model and prevent overfitting. By training multiple DTs on different subsets of data and averaging their predictions, RFs achieve better generalization and are less prone to overfitting compared to individual DTs. Similarly, to aforementioned models, the RFs model undergoes training on the data to optimize its performance as illustrated in Fig. 14. Training is crucial for the model to generate accurate predictions on new and unseen data. During training, the model learns the patterns and relationships present in the data. Thus, enabling it to make predictions from these patterns. It also allows for the identification and correction of any errors or biases within the model. It can be seen in Fig. 14(a) that a significant relationship with a value of approximately 0.93 is observed.

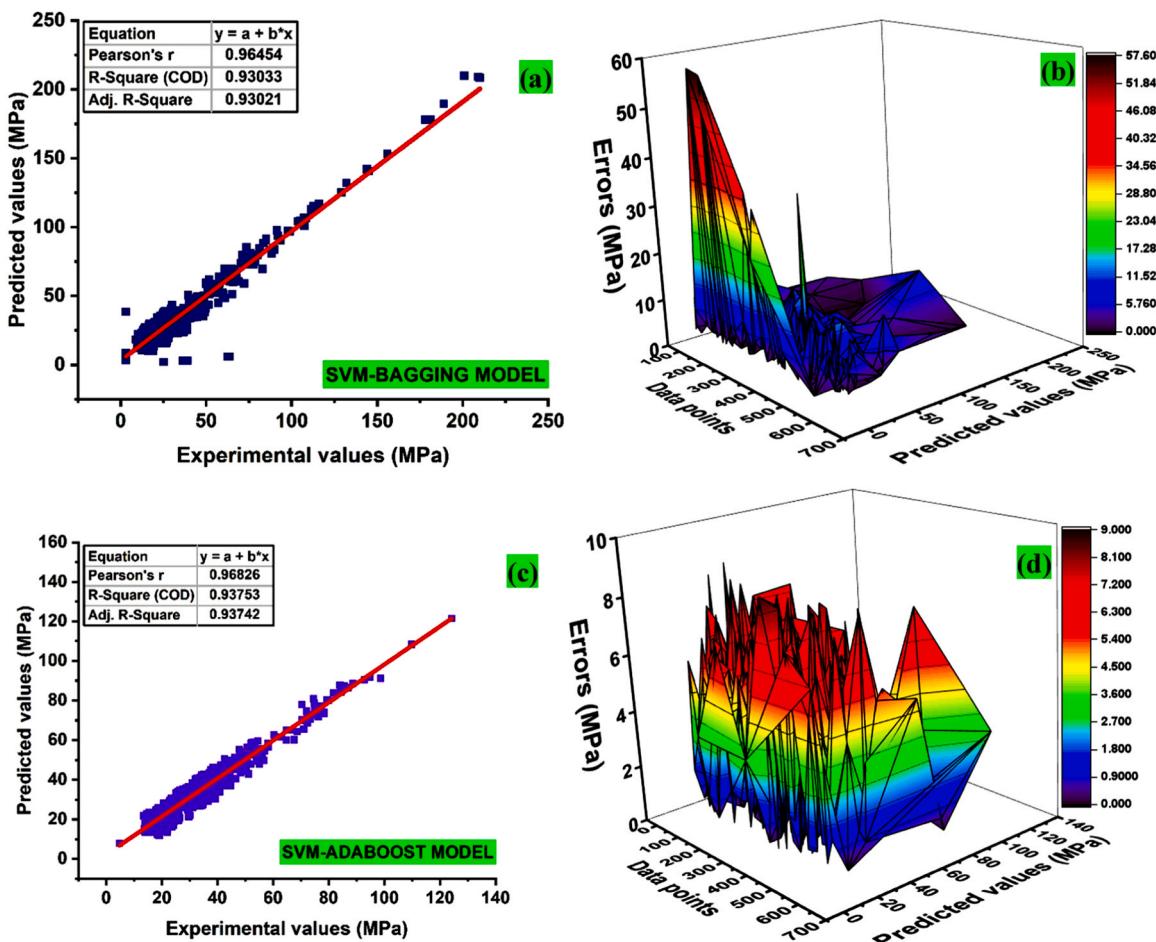


Fig. 17. RF model; (a) bagging result; (b) bagging error results; (c) boosting results; (d) boosting set discrepancies.

Table 6
Statistical analysis of the models.

Statistical measures	DT	DT-bagging	DT-boosting	RF	RF- N th	SVM	SVM-bagging	SVM-bagging		
R ²	Train 0.99	Test 0.97	Test 0.97	Test 0.99	Train 0.93	Test 0.95	Test 0.99	Test 0.93	Test 0.93	
R	0.99	0.98	0.99	1.00	0.96	0.98	1.00	0.93	0.96	0.97
NSE	0.99	0.97	0.97	0.99	0.93	0.95	0.99	0.87	0.89	0.93
MAE	1.71	2.58	2.30	1.52	4.05	3.01	1.30	6.47	7.10	5.48
RMSE	2.85	5.09	4.59	2.45	6.43	3.61	2.23	8.53	9.43	7.53
RMSLE	0.09	0.13	0.12	0.09	0.22	0.13	0.07	0.30	0.32	0.34
RSE	0.11	0.18	0.16	0.09	0.27	0.22	0.08	0.36	0.34	0.27
RRMSE%	8.47	14.81	13.36	7.15	20.12	10.52	6.49	28.05	27.43	21.89
PI	0.04	0.08	0.07	0.04	0.10	0.05	0.03	0.15	0.11	0.06

This indicates a strong correlation between the model's predictions and the training data. However, it is important to note that a strong correlation in the training data does not guarantee good performance on test data. When the model is overly complex, it may overfit the training data. This leads to poor generalization. Therefore, it is recommended to assess the model's generalization ability by testing it on unseen data, as illustrated in Fig. 14(c). The test data helps evaluate whether the model is overfitting or under fitting the training data. It provides insights into the model's actual performance. A good model should exhibit similar performance on both the training and test data, with a minimal difference in their performance metrics. Therefore, Fig. 14(c) shows a stronger model relationship with an R² value of approximately 0.95, which is close to the relationship observed in the training set. This indicates that the model performs well on the test data. The residual error of the train and test set is illustrated in Fig. 14(b) and Fig. 14(d). The residual error

configuration of the test set reveals important insights about the model's performance. An average error of approximately 3.01 MPa is observed. This indicates a typical deviation between the model's forecast values and the experimental values in the test set. Additionally, the maximum error of 7.34 MPa signifies the largest discrepancy between the predicted and actual values as the model's prediction deviates significantly from the true value in the test set. On the other hand, the minimum error of 0.0093 MPa represents the smallest deviation observed, indicating a case where the model's prediction aligns closely with the actual value. Moreover, 100 % of error values lie in the range of 0–10 MPa as illustrated in Fig. 14(d).

The parameter "n_estimators" in random forest regression (RFR) specifies the number of decision trees included in the ensemble model. Increasing the number of trees enhances the model's capacity to accurately represent the data. Thus, a total of (nth = 20 sub-models) were

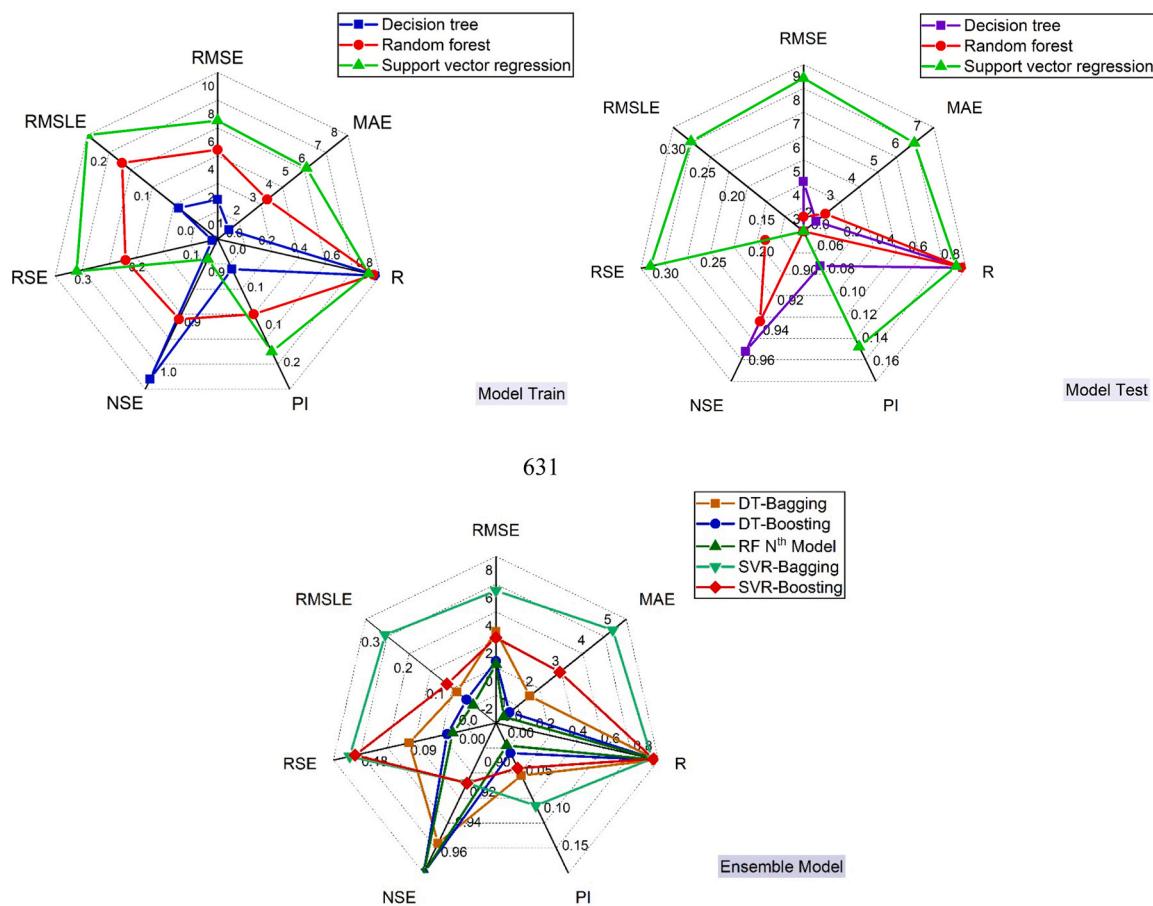


Fig. 18. Radar diagram; (a) Train set; (b) Test set; (c) Ensemble model.

created, and is assessed by their correlation coefficients. All sub-models exhibited a correlation coefficient exceeding $R^2 = 0.84$. These findings reveal a robust correlation between the predicted and measured results for each sub-model. Notably, the sub-model constructed with ($n^{th} = 16$ th sub-models) achieved the highest R^2 value of 0.99, as depicted in Fig. 15 (a). The ensemble model shows more accuracy in predicing the outcome variables with a strong and stable relationship between the predicted and targeted values, as indicated by its R^2 value of 0.99, and a mean error distribution of 1.29 MPa. Moreover, 99.46 % of error values lie in the range of 0–10 MPa, 0 % in the range of 10–15, and 0.53 % greater than 15 MPa as demonstrated in Fig. 15(b).

7.3. Support vector regressor algorithm with ensemble learner

The results obtained from using the SVM model alone and in combination with the ensemble approach are presented in Fig. 16. The training outcome of the SVR with R^2 value of 0.87 is shown in Fig. 16(a), and its error distribution is demonstrated in Fig. 16(b). The SVM model in the test set exhibits a stronger correlation with the target results, achieving an R^2 value of 0.88. This is evident by its average error value of 7.09 MPa, as illustrated in Fig. 16(c) and Fig. 16(d).

In order to improve the efficiency of the SVM model, it was combined with supervised hybrid algorithms, including bagging and boosting. The bagging models of SVM show a robust correlation by demonstrating an R^2 value of 0.93 as indicated in Fig. 17(a). Similarly, Fig. 17(b) displays the error distribution values, with a maximum, minimum and average values of approximately 57.56 MPa, 0.02 MPa, and 5.47 MPa, respectively. The SVM model with boosting also demonstrates a strong correlation between input and output variables by displaying an R^2 value of 0.937 as shown in Fig. 17(c). In Fig. 17(d), the error distribution of the

boosting model is depicted, showcasing an average error of 3.45 MPa, a maximum error of 8.98 MPa, and a minimum error of 0.009 MPa.

7.4. Model validation using statistical analysis

Model evaluation is a dire step in ML for assessing the performance of the models. Thus, K-fold and statistical approaches were employed for the evaluation of the models. The statistical assessment of all ML methods based is displayed in Table 6. Therefore, clearly shows that the ensemble methods outperformed the individual approaches, with a significant reduction in error distribution values for all models. In addition, Fig. 18 depicts the visual demonstration of all models by radar diagram.

8. K-fold cross-validation test results

To assess the accuracy of prediction models, it is essential to conduct a thorough evaluation. Validation plays a pivotal role in this assessment, determining the model's accuracy. One effective technique for this purpose is the K-fold validation test, which involves rearranging and analysing experimental data in distinct parts to evaluate reliability. In this method the data samples are divided into ten equal subgroups, using one subgroup for validation and the remaining for training. The 10-fold cross-validation process repeats this cycle ten times, calculating the average precision to minimize the impact of inherent bias from random data selection. This ensures an accurate representation of the model's generalizability and reliability. Model accuracy is then measured using three metrics: mean absolute error (MAE), root mean squared error (RMSE), and (R^2) values. Fig. 19 (a-i) presents an analysis of the cross-validation data using the MAE, RMSE and R^2 values, which shows

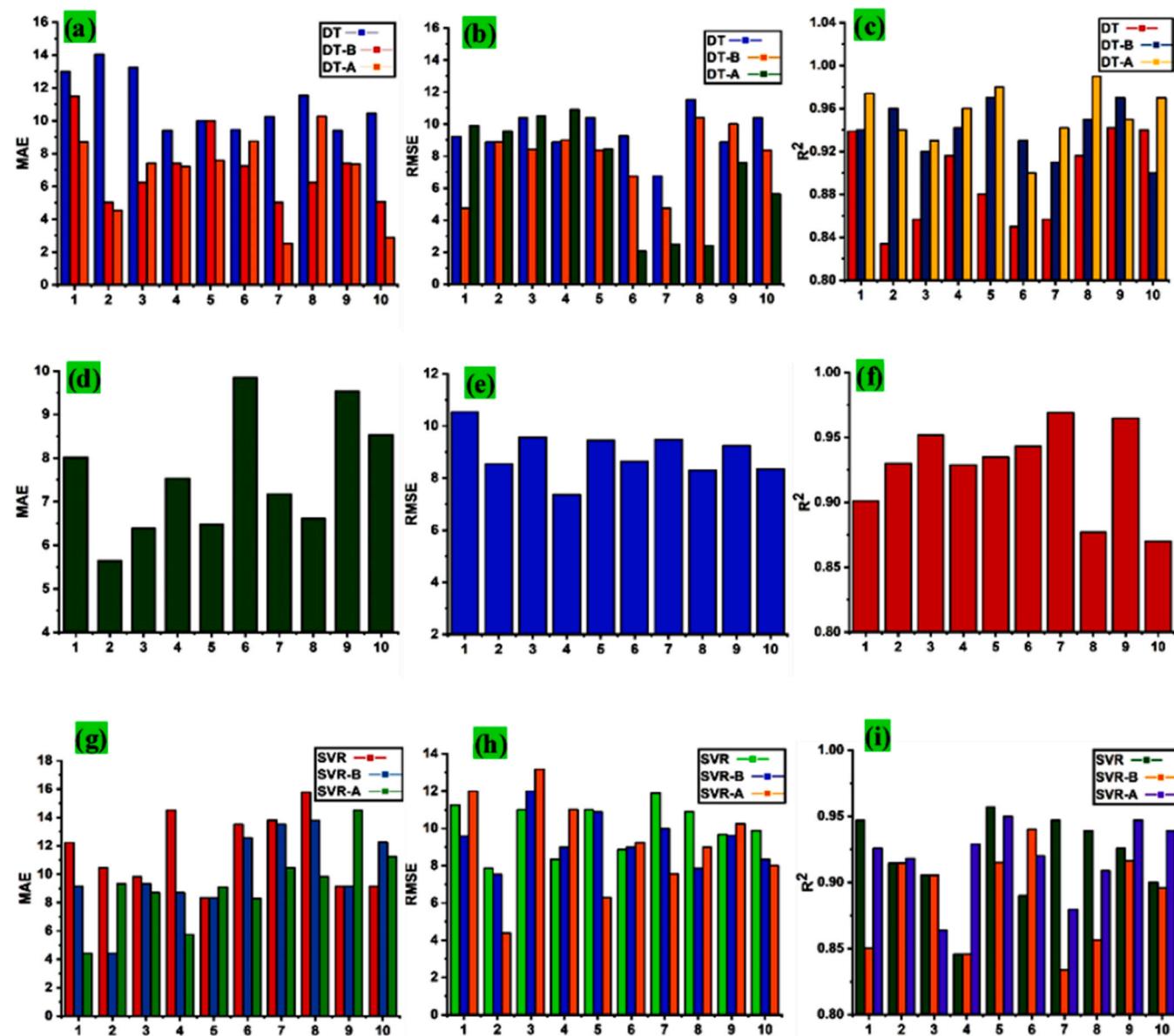


Fig. 19. Validation of ML models; (a) DT approach with MAE; (b) DT bagging approach with RMSE; (c) DT boosting approach with R^2 ; (d) RF method with MAE; (e) RF method with RMSE; (f) RF method with R^2 ; (g) SVM approach with MAE; (h) SVM bagging approach with RMSE; (i) SVM boosting approach with R^2 .

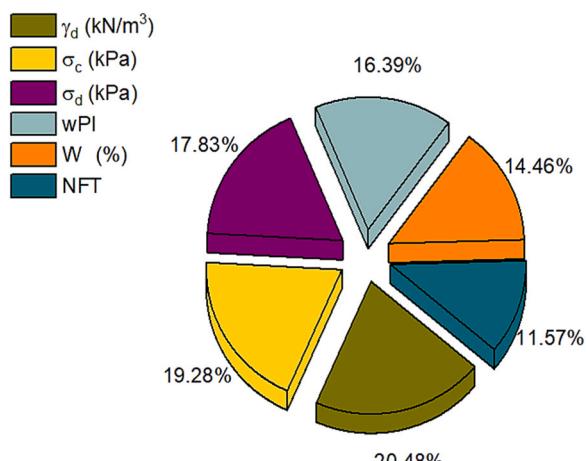


Fig. 20. Parameter contribution to M_r .

moderate to substantial correlation of the nonlinear and ensemble models. It demonstrates that despite some variations, a high level of precision is achieved in predicting results. For instance, the R^2 values of all models range from 0.5 to 0.6 for their lowest scores, while the highest values are found between 0.80 and 0.99. Similarly, the MAE and RMSE values indicate minimal errors for ensemble models as compared to individual models.

9. Importance of permutation variable

A sensitivity analysis is conducted to assess the individual impact of independent variables on the outcomes as illustrated in Fig. 20. It can be seen that γ_d , σ_c , and σ_d have major contribution of about 20.48 %, 19.28 %, and 17.83 to resilient modulus. Thus, an increase in dry unit weight lead to a higher resilient modulus due to denser and more compacted material resulting in increase stiffness. Similarly, confining and deviator stress demonstrates materials ability to resist deformation and material responds to external loads. This densification typically results in an increase in stiffness, leading to a higher resilient modulus.

Graphical User Interface for Prediction of Concrete Design

Input Variables

Weighted plasticity index	wPI	
Dry unit weight	γ_d (kN/m ³)	
Confining stress	σ_c (kPa)	
Deviator stress	σ_d (kPa)	
Moisture content	w (%)	
Number of FT cycles	N_{FT}	

Output

Predict Result

Compressive Strength	M_R (MPa)		MPa
----------------------	-------------	--	-----

Fig. 21. GUI of M_r .

Denser soils can better resist deformation under repeated loading, contributing to a more robust and resilient material.

10. Graphical user interface (GUI)

In recent times, the swift advancement of AI technology has expedited the integration of intelligence and informatization in traditional industries, including the field of concrete. This surge in AI development provides ample motivation for the application of AI techniques in the concrete industry. As depicted in Fig. 21, the graphical user interface (GUI) enhances model flexibility and functions as a user-accessible gateway, simplifying operational complexities. Consequently, it proves highly beneficial for both academic research and industrial applications. This user-friendly interface enables users to seamlessly interact with the model, inputting the necessary parameters effortlessly. Such ease of interaction facilitates the generation of desired results, making it a practical and versatile tool for utilizing the model's capabilities across a diverse range of scenarios.

11. Limitation and future work

This research work delivers a robust methodology for estimating the M_r of subgrade using three models namely as SVR, DR, and RF with 2813 data set. Nevertheless, there might be some limitations regarding

the models' applicability to subgrades with significantly different characteristics or coming from different regions since the dataset might not include all possible soil types and environmental conditions. Furthermore, the models depend greatly on some input parameters and may not reflect fully the actual loading conditions and the long-term environmental variations.

To enhance the effectiveness and applicability of the models, it is suggested that future works should include more data sets from various locations and types of subgrade materials. Moreover, including time-dependent data on the subgrade performance under varying climatic conditions may help us better understand the nature of subgrade materials and increase the reliability of our models. Extending the hybrid modelling techniques that incorporate the strengths of SVM, DT, and RF with other machine learning techniques including neural networks or gradient boosting could enhance the predictive performance. Integrating more environmental factors and realistic traffic loading patterns would improve the robustness of the models. This project's GUI has to be tested and validated in real-life scenarios. The feedback from the practitioners will be used to refine the tool and make it more user-friendly. Through the combination of the graphical user interface (GUI) and Geographical Information Systems (GIS), the users will be able to use the spatial analysis feature to provide a graphical representation and a forecast of the robust modulus in different geographical locations. Future studies may explore for employing AI techniques to improve the

transparency of complex algorithms so that practitioners can understand and trust the outcomes.

12. Conclusions

The primary goal of this study was to evaluate the effectiveness of both ensemble and individual learning approaches in predicting the M_r of the subgrade. Robust models were constructed using ensemble methods such as bagging and boosting, along with individual models like decision trees (DT), random forests (RF), and support vector machines (SVM). The reliability of these models was assessed through predictive capacity and sensitivity analysis, leading to the following conclusions:

1. Ensemble learning algorithms, utilizing bagging and boosting techniques, yielded efficient results and showcased superior performance compared to individual learning methods. Both of these techniques exhibited R^2 values exceeding 0.90. Specifically, employing the boosting approach with decision trees (DT) and random forests (RF) led to a significant improvement in the outcomes.
2. Twenty different models with varying n-estimators in ensemble methods (bagging and boosting) were employed. The RF and DT with boosting achieved a notable R^2 score of 0.99. Thus, demonstrating a superior and reliable performance compared to bagging and stand-alone approaches.
3. Exceptional alignment between predicted and observed outcomes was accomplished by implementing a modified bagging technique with a random forest. This success is credited to the incorporation of a specific machine-based learner within the random forest, highlighting the significant R^2 value of 0.99.
4. Furthermore, to assess the predictive accuracy of the model, a K-Fold cross-validation process was conducted. In this procedure, various statistical indicators were employed, including the coefficient of correlation (R^2), mean absolute errors (MAE), and root mean square error (RMSE). The analysis of these statistical indicators revealed very good and promising results. All the models exhibited lower MAE and RMSE errors while maintaining a high R^2 correlation.
5. Permutation analysis indicated that dry density, confining stress, and deviator stress played pivotal roles in the parametric results for resilient modulus.
6. GUI is developed that can be used for practical implementation. Thus, utilization will save cost and time.

Declaration of Competing Interest

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

Data Availability

Data will be made available on request.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.mtcomm.2024.109764](https://doi.org/10.1016/j.mtcomm.2024.109764).

References

- [1] X. Chu, I. Campos-Guereta, A. Dawson, N. Thom, Sustainable pavement drainage systems: Subgrade moisture, subsurface drainage methods and drainage effectiveness, *Constr. Build. Mater.* 364 (2023) 129950, <https://doi.org/10.1016/j.conbuildmat.2022.129950>.
- [2] A. Hayat, A. Hussain, H.F. Afzidi, Determination of in-field temperature variations in fresh HMA and corresponding compaction temperatures, *Constr. Build. Mater.* 216 (2019) 84–92, <https://doi.org/10.1016/j.conbuildmat.2019.04.262>.
- [3] W. Zhang, K.N. Ahmad, Z. Tong, Z. Hu, H. Wang, M. Wu, K. Zhao, S. Yang, H. Farooq, L.N. Mohammad, In-Time Density Monitoring of In-Place Asphalt Layer Construction via Intelligent Compaction Technology, *J. Mater. Civ. Eng.* 35 (2023), [https://doi.org/10.1061/\(asce\)mt.1943-5533.0004558](https://doi.org/10.1061/(asce)mt.1943-5533.0004558).
- [4] G. Bilgen, O.F. Altuntas, Sustainable re-use of waste glass, cement and lime treated dredged material as pavement material, *Cast Stud. Constr. Mater.* 18 (2023) e01815, <https://doi.org/10.1016/j.cscm.2022.e01815>.
- [5] H.F. Afzidi, M.J. Khattak, Self-Healing Characteristics of Polyvinyl Alcohol-Fiber-Reinforced Hot Mix Asphalt for Enhanced Pavement Durability, *Transp. Res. Rec.* (2024), <https://doi.org/10.1177/03611981241231969>.
- [6] X. Xie, Y. Gao, F. Hou, T. Cheng, A. Hao, H. Qin, Fluid Inverse Volumetric Modeling and Applications from Surface Motion, *IEEE Trans. Vis. Comput. Graph.* (2024), <https://doi.org/10.1109/TVCG.2024.3370551>.
- [7] D. Sarker, J.X. Wang, Moisture influence on structural responses of pavement on expansive soils, *Transp. Geotech.* 35 (2022) 100773, <https://doi.org/10.1016/j.trgeo.2022.100773>.
- [8] Y. Luo, H. Wu, W. Song, J. Yin, Y. Zhan, J. Yu, S. Abubakar Wada, Thermal fatigue and cracking behaviors of asphalt mixtures under different temperature variations, *Constr. Build. Mater.* 369 (2023) 130623, <https://doi.org/10.1016/j.conbuildmat.2023.130623>.
- [9] Z. Zimar, D. Robert, A. Zhou, F. Giustozzi, S. Setunge, J. Kodikara, Application of coal fly ash in pavement subgrade stabilisation: A review, *J. Environ. Manag.* 312 (2022) 114926, <https://doi.org/10.1016/j.jenvman.2022.114926>.
- [10] A.R. Gabr, B. Roy, M.R. Kaloo, D. Kumar, A. Arisha, M. Shiha, S. Shwally, J. W. Hu, S.M. El-Badawy, A novel approach for resilient modulus prediction using extreme learning machine-equilibrium optimiser techniques, *Int. J. Pavement Eng.* 23 (2022) 3346–3356, <https://doi.org/10.1080/10298436.2021.1892109>.
- [11] S. Hao, T. Pabst, Prediction of CBR and resilient modulus of crushed waste rocks using machine learning models, *Acta Geotech.* 17 (2022) 1383–1402, <https://doi.org/10.1007/s11440-022-01472-1>.
- [12] R. Sarkhani Benemaran, M. Esmaeili-Falak, A. Javadi, Predicting resilient modulus of flexible pavement foundation using extreme gradient boosting based optimised models, *Int. J. Pavement Eng.* (2022), <https://doi.org/10.1080/10298436.2022.2095385>.
- [13] D. Hu, Y. Li, X. Yang, X. Liang, K. Zhang, X. Liang, Experiment and Application of NATM Tunnel Deformation Monitoring Based on 3D Laser Scanning, *Struct. Control Heal. Monit.* 2023 (2023), <https://doi.org/10.1155/2023/3341788>.
- [14] A. Fouad, R. Hassan, A. Mahmood, Developing Resilient Modulus Prediction Models Based on Experimental Results of Crushed Hornfels Mixes with Different Gradations and Plasticity, *Int. J. Pavement Res. Technol.* 15 (2022) 124–137, <https://doi.org/10.1007/s42947-021-00005-5>.
- [15] B. Ghorbani, E. Yaghoubi, P.L.P. Wasantha, R. van Staden, M. Guerrieri, S. Fragomeni, Machine learning-based prediction of resilient modulus for blends of tire-derived aggregates and demolition wastes, *Road. Mater. Pavement Des.* (2023), <https://doi.org/10.1080/14680629.2023.2222176>.
- [16] L. Sadik, Developing Prediction Equations for Soil Resilient Modulus Using Evolutionary Machine Learning, *Transp. Infrastruct. Geotechnol.* (2023), <https://doi.org/10.1007/s40515-023-00342-x>.
- [17] W. Zhen, Z. Wang, Q. Wang, W. Sun, R. Wang, W. Zhang, Y. Zhang, W. Qin, B. Li, Q. Wang, B. Hong, Y. Yang, J. Xu, S. Ma, M. Da, L. Feng, X. Zang, X. Mo, X. Sun, M. Wu, J. Xu, J. Xu, Y. Huang, H. Zhang, Cardiovascular disease therapeutics via engineered oral microbiota: Applications and perspective, *IMeta* (2024), <https://doi.org/10.1002/int.2197>.
- [18] T. Lin, T. Ishikawa, J. Yang, T. Tokoro, Evaluation of climate effect on resilient modulus of granular subgrade material, *Cold Reg. Sci. Technol.* 194 (2022), <https://doi.org/10.1016/j.coldregions.2021.103452>.
- [19] S.Y. Kim, D.J. Kim, J.S. Lee, T.H.K. Kang, Y.H. Byun, Resilient modulus estimation using in-situ modulus detector: performance and factors, *Int. J. Pavement Eng.* (2022), <https://doi.org/10.1080/10298436.2022.2096883>.
- [20] C.C. Ikeagwuani, D.C. Nwontu, C.C. Nweke, Resilient modulus descriptive analysis and estimation for fine-grained soils using multivariate and machine learning methods, *Int. J. Pavement Eng.* 23 (2022) 3409–3424, <https://doi.org/10.1080/10298436.2021.1895993>.
- [21] X. Liu, X. Zhang, H. Wang, B. Jiang, Laboratory testing and analysis of dynamic and static resilient modulus of subgrade soil under various influencing factors, *Constr. Build. Mater.* 195 (2019) 178–186, <https://doi.org/10.1016/j.conbuildmat.2018.11.061>.
- [22] J. Peng, J. Zhang, J. Li, Y. Yao, A. Zhang, Modeling humidity and stress-dependent subgrade soils in flexible pavements, *Comput. Geotech.* 120 (2020), <https://doi.org/10.1016/j.compgeo.2019.103413>.
- [23] J. Zhang, J. Peng, L. Zeng, J. Li, F. Li, Rapid estimation of resilient modulus of subgrade soils using performance-related soil properties, *Int. J. Pavement Eng.* 22 (2021) 732–739, <https://doi.org/10.1080/10298436.2019.1643022>.
- [24] J. Zhang, A. Zhang, C. Huang, H. Yu, C. Zhou, Characterising the resilient behaviour of pavement subgrade with construction and demolition waste under Freeze-Thaw cycles, *J. Clean. Prod.* 300 (2021), <https://doi.org/10.1016/j.jclepro.2021.126702>.
- [25] T. Ishikawa, T. Lin, S. Kawabata, S. Kameyama, T. Tokoro, Effect evaluation of freeze-thaw on resilient modulus of unsaturated granular base course material in pavement, *Transp. Geotech.* 21 (2019), <https://doi.org/10.1016/j.trgeo.2019.100284>.
- [26] M.A. Al-Mahbashi, M. Al-Shamrani, Long-term and immediate effects of freeze-thaw cycles on the resilient modulus of treated expansive subgrades, *Road.*

- Mater. Pavement Des. 24 (2023) 2411–2424, <https://doi.org/10.1080/14680629.2022.2146603>.
- [27] A.J. Puppala, L.R. Hoyos, A.K. Potturi, Resilient Moduli Response of Moderately Cement-Treated Reclaimed Asphalt Pavement Aggregates, J. Mater. Civ. Eng. 23 (2011) 990–998, [https://doi.org/10.1061/\(asce\)mt.1943-5533.0000268](https://doi.org/10.1061/(asce)mt.1943-5533.0000268).
- [28] X. Hu, T. Tang, L. Tan, H. Zhang, Fault Detection for Point Machines: A Review, Challenges, and Perspectives, Vol. 12, Page 391. 12, Actuators 2023 (2023) 391, <https://doi.org/10.3390/ACT12100391>.
- [29] J. Cao, S.-T. Quek, H. Xiong, Z. Yang, Comparison of Constrained Unscented and Cubature Kalman Filters for Nonlinear System Parameter Identification, J. Eng. Mech. 149 (2023), <https://doi.org/10.1061/j.emengt.2023.7091>.
- [30] U. Asif, M.F. Javed, M. Abuhussain, M. Ali, W.A. Khan, A. Mohamed, Predicting the mechanical properties of plastic concrete: An optimization method by using genetic programming and ensemble learners, Case Stud. Constr. Mater. 20 (2024) e03135, <https://doi.org/10.1016/j.cscm.2024.e03135>.
- [31] S.Q. Yang, J.Z. Tang, S.S. Wang, D.Sen Yang, W.T. Zheng, An Experimental and Modeling Investigation on Creep Mechanical Behavior of Granite Under Triaxial Cyclic Loading and Unloading, Rock. Mech. Rock. Eng. 55 (2022) 5577–5597, <https://doi.org/10.1007/s00603-022-02920-w>.
- [32] Y. Wang, T. Mao, Y. Xia, X. Li, X. Yi, Macro-meso fatigue failure of bimrocks with various block content subjected to multistage fatigue triaxial loads, Int. J. Fatigue 163 (2022), <https://doi.org/10.1016/j.ijfatigue.2022.107014>.
- [33] K.M. Islam, S.L. Gassman, Effect of the Field-Stress State on the Subgrade Resilient Modulus for Pavement Rutting and IRI, Geotechnics 3 (2023) 360–374, <https://doi.org/10.3390/geotechnics3020021>.
- [34] M. Dong-Gyu Kim, DEVELOPMENT OF A CONSTITUTIVE MODEL FOR RESILIENT MODULUS, 2004. (<https://search.proquest.com/openview/2ab4e5c12c5e4375da7d320542623ef/1?pq-orignal=gscholar&cbl=18750&diss=y>) (accessed October 25, 2023).
- [35] A.L. Smart, D.N. Humphrey, Determination of resilient modulus for Maine roadway soils, (1999). (<https://trid.trb.org/view/647964>) (accessed October 25, 2023).
- [36] R. Pezo, W.R. Hudson, Prediction models of resilient modulus for nongranular materials, Geotech. Test. J. 17 (1994), <https://doi.org/10.1520/gt10109>.
- [37] Z. Han, S.K. Vanapalli, J. ping Ren, W. lie Zou, Characterizing cyclic and static moduli and strength of compacted pavement subgrade soils considering moisture variation, Soils Found. 58 (2018) 1187–1199, <https://doi.org/10.1016/j.sandf.2018.06.003>.
- [38] H. Huang, Y. Yuan, W. Zhang, M. Li, Seismic behavior of a replaceable artificial controllable plastic hinge for precast concrete beam-column joint, Eng. Struct. 245 (2021), <https://doi.org/10.1016/j.engstruct.2021.112848>.
- [39] Y. Su, J. Wang, D. Li, X. Wang, L. Hu, Y. Yao, Y. Kang, End-to-end deep learning model for underground utilities localization using GPR, Autom. Constr. 149 (2023), <https://doi.org/10.1016/j.autcon.2023.104776>.
- [40] M. Alzara, M.F. Rehman, F. Farooq, M. Ali, A.A.A. Beshr, A.M. Yosri, S.B. A El Sayed, Prediction of building energy performance using mathematical gene-expression programming for a selected region of dry-summer climate, Eng. Appl. Artif. Intell. 126 (2023) 106958, <https://doi.org/10.1016/j.engappai.2023.106958>.
- [41] H. Huang, M. Guo, W. Zhang, M. Huang, Seismic Behavior of Strengthened RC Columns under Combined Loadings, J. Bridg. Eng. 27 (2022) 05022005, [https://doi.org/10.1061/\(asce\)be.1943-5592.0001871](https://doi.org/10.1061/(asce)be.1943-5592.0001871).
- [42] D. Lu, J. Liang, X. Du, C. Ma, Z. Gao, Fractional elastoplastic constitutive model for soils based on a novel 3D fractional plastic flow rule, Comput. Geotech. 105 (2019) 277–290, <https://doi.org/10.1016/j.compgeo.2018.10.004>.
- [43] M.A. Khan, F. Farooq, M.F. Javed, A. Zafar, K.A. Ostrowski, F. Aslam, S. Malazdrewicz, M. Małysiak, Simulation of depth of wear of eco-friendly concrete using machine learning based computational approaches, Mater. (Basel) 15 (2022) 58, <https://doi.org/10.3390/ma15010058>.
- [44] F. Farooq, S.K.U. Rahman, A. Akbar, R.A. Khushnood, M.F. Javed, R. Alyousef, H. Alabduljabbar, F. aslam, A comparative study on performance evaluation of hybrid GNP/CNTs in conventional and self-compacting mortar, Alex. Eng. J. 59 (2020) 369–379, <https://doi.org/10.1016/j.aej.2019.12.048>.
- [45] J. Xin, W. Xu, B. Cao, T. Wang, S. Zhang, A deep-learning-based MAC for integrating channel access, rate adaptation and channel switch, (2024). [http://arxiv.org/abs/2406.02291](https://arxiv.org/abs/2406.02291) (accessed July 13, 2024).
- [46] F. Aslam, F. Farooq, M.N. Amin, K. Khan, A. Waheed, A. Akbar, M.F. Javed, R. Alyousef, H. Alabduljabbar, Applications of Gene Expression Programming for Estimating Compressive Strength of High-Strength Concrete, Adv. Civ. Eng. 2020 (2020) 1–23, <https://doi.org/10.1155/2020/8850535>.
- [47] M.F. Javed, M.N. Amin, M.I. Shah, K. Khan, B. Iftikhar, F. Farooq, F. Aslam, R. Alyousef, H. Alabduljabbar, Applications of gene expression programming and regression techniques for estimating compressive strength of bagasse ash based concrete, Crystals 10 (2020) 737, <https://doi.org/10.3390/cryst10090737>.
- [48] F. Abbassi, F. Ahmad, Behavior analysis of concrete with recycled tire rubber as aggregate using 3D-digital image correlation, J. Clean. Prod. 274 (2020) 123074, <https://doi.org/10.1016/j.jclepro.2020.123074>.
- [49] F. Farooq, S. Czarnecki, P. Niewiadomski, F. Aslam, H. Alabduljabbar, K. A. Ostrowski, K. Śliwa-Wieczorek, T. Nowobilski, S. Malazdrewicz, A comparative study for the prediction of the compressive strength of self-compacting concrete modified with fly ash, Mater. (Basel) 14 (2021) 4934, <https://doi.org/10.3390/ma14174934>.
- [50] M.A. Khan, S.A. Memon, F. Farooq, M.F. Javed, F. Aslam, R. Alyousef, Compressive Strength of Fly-Ash-Based Geopolymer Concrete by Gene Expression Programming and Random Forest, Adv. Civ. Eng. 2021 (2021), <https://doi.org/10.1155/2021/6618407>.
- [51] M.N. Amin, K. Khan, A.M. Abu Arab, F. Farooq, S.M. Eldin, M.F. Javed, Prediction of sustainable concrete utilizing rice husk ash (RHA) as supplementary cementitious material (SCM): Optimization and hyper-tuning, J. Mater. Res. Technol. 25 (2023) 1495–1536, <https://doi.org/10.1016/j.jmrt.2023.06.006>.
- [52] C. Chen, D. Han, C.C. Chang, MPCCT: Multimodal vision-language learning paradigm with context-based compact Transformer, Pattern Recognit. 147 (2024), <https://doi.org/10.1016/j.patcog.2023.110084>.
- [53] H. Jiao, Y. Wang, L. Li, K. Arif, F. Farooq, A. Alaskar, A novel approach in forecasting compressive strength of concrete with carbon nanotubes as nanomaterials, Mater. Today Commun. 35 (2023) 106335, <https://doi.org/10.1016/j.mtcomm.2023.106335>.
- [54] G. Xu, G. Zhou, F. Althoei, H.M. Hadidi, A. Alaskar, A.M. Hassan, F. Farooq, Evaluation of properties of bio-composite with interpretable machine learning approaches: optimization and hyper tuning, J. Mater. Res. Technol. 25 (2023) 1421–1446, <https://doi.org/10.1016/j.jmrt.2023.06.007>.
- [55] Y. Hu, A.A. Alaskar, F. Althoei, M.A. Abuhussain, G. Alfallah, F. Farooq, Strength evaluation sustainable concrete with waste ingredients at elevated temperature by employing interpretable algorithms: Optimization and hyper tuning, Mater. Today Commun. 36 (2023) 106467, <https://doi.org/10.1016/j.mtcomm.2023.106467>.
- [56] A. Zaman, R.U.D. Nassar, M. Alyami, S. Shah, M.F. Rehman, I.Y. Hakeem, F. Farooq, Forecasting the strength of micro/nano silica in cementitious matrix by machine learning approaches, Mater. Today Commun. 37 (2023) 107066, <https://doi.org/10.1016/j.mtcomm.2023.107066>.
- [57] Z. Liu, X. Gu, H. Ren, Rutting prediction of asphalt pavement with semi-rigid base: Numerical modeling on laboratory to accelerated pavement testing, Constr. Build. Mater. 375 (2023), <https://doi.org/10.1016/j.conbuildmat.2023.130903>.
- [58] Y. Zhang, M. Ling, F. Kaseer, E. Arambula, R.L. Lytton, A.E. Martin, Prediction and evaluation of rutting and moisture susceptibility in rejuvenated asphalt mixtures, J. Clean. Prod. 333 (2022), <https://doi.org/10.1016/j.jclepro.2021.129980>.
- [59] C. Chen, D. Han, X. Shen, CLVIN: Complete language-vision interaction network for visual question answering, Knowl. -Based Syst. 275 (2023), <https://doi.org/10.1016/j.knosys.2023.110706>.
- [60] R. Fei, Y. Guo, J. Li, B. Hu, L. Yang, An Improved BPNN Method Based on Probability Density for Indoor Location, IEICE Trans. Inf. Syst. E106.D (2023) 773–785, <https://doi.org/10.1587/transinf.2022DLP0073>.
- [61] Y. Zhao, J. Wang, G. Cao, Y. Yuan, X. Yao, L. Qi, Intelligent Control of Multilegged Robot Smooth Motion: A Review, IEEE Access 11 (2023) 86645–86685, <https://doi.org/10.1109/ACCESS.2023.3304992>.
- [62] D.L. Chen, J.W. Zhao, S.R. Qin, SVM strategy and analysis of a three-phase quasi-Z-source inverter with high voltage transmission ratio, Sci. China Technol. Sci. 66 (2023) 2996–3010, <https://doi.org/10.1007/s11431-022-2394-4>.
- [63] S. Meng, F. Meng, H. Chi, H. Chen, A. Pang, A robust observer based on the nonlinear descriptor systems application to estimate the state of charge of lithium-ion batteries, J. Franklin Inst. 360 (2023) 11397–11413, <https://doi.org/10.1016/j.jfranklin.2023.08.037>.
- [64] M. Shi, W. Hu, M. Li, J. Zhang, X. Song, W. Sun, Ensemble regression based on polynomial regression-based decision tree and its application in the in-situ data of tunnel boring machine, Mech. Syst. Signal Process. 188 (2023), <https://doi.org/10.1016/j.ymssp.2022.110022>.
- [65] X. Zhou, D. Lu, Y. Zhang, X. Du, T. Rabczuk, An open-source unconstrained stress updating algorithm for the modified Cam-clay model, Comput. Methods Appl. Mech. Eng. 390 (2022), <https://doi.org/10.1016/j.cma.2021.114356>.
- [66] D. Lu, X. Zhou, X. Du, G. Wang, 3D Dynamic Elastoplastic Constitutive Model of Concrete within the Framework of Rate-Dependent Consistency Condition, J. Eng. Mech. 146 (2020), [https://doi.org/10.1061/\(asce\)jem.1943-7889.0001854](https://doi.org/10.1061/(asce)jem.1943-7889.0001854).
- [67] S. Shi, D. Han, M. Cui, A multimodal hybrid parallel network intrusion detection model, Conn. Sci. 35 (2023), <https://doi.org/10.1080/09540091.2023.2227780>.
- [68] J. Hong, L. Gui, J. Cao, Analysis and Experimental Verification of the Tangential Force Effect on Electromagnetic Vibration of PM Motor, IEEE Trans. Energy Convers. 38 (2023) 1893–1902, <https://doi.org/10.1109/TEC.2023.3241082>.
- [69] D. Van Dao, H.-B.B. Ly, H.-L.L.T. Vu, T.-T.T. Le, B.T. Pham, Investigation and optimization of the C-ANN structure in predicting the compressive strength of foamed concrete, Mater. (Basel) 13 (2020) 1072, <https://doi.org/10.3390/mati30501072>.
- [70] M. Zaman, P. Solanki, A. Ebrahimi, L. White, Neural Network Modeling of Resilient Modulus Using Routine Subgrade Soil Properties, Int. J. Geomech. 10 (2010) 1–12, [https://doi.org/10.1061/\(asce\)1532-3641\(2010\)10:1\(1\)](https://doi.org/10.1061/(asce)1532-3641(2010)10:1(1)).
- [71] W. lie Zou, Z. Han, L. qiang Ding, X. qun Wang, Predicting resilient modulus of compacted subgrade soils under influences of freeze-thaw cycles and moisture using gene expression programming and artificial neural network approaches, Transp. Geotech. 28 (2021) 100520, <https://doi.org/10.1016/j.tgeo.2021.100520>.
- [72] M. Pal, S. Deswal, Extreme Learning Machine Based Modeling of Resilient Modulus of Subgrade Soils, Geotech. Geol. Eng. 32 (2014) 287–296, <https://doi.org/10.1007/s10706-013-9710-y>.
- [73] E. Sadrossadat, A. Heidaripanah, S. Osouli, Prediction of the resilient modulus of flexible pavement subgrade soils using adaptive neuro-fuzzy inference systems, Constr. Build. Mater. 123 (2016) 235–247, <https://doi.org/10.1016/j.conbuildmat.2016.07.008>.
- [74] N. Kardani, M. Aminpour, M. Nouman Amjad Raja, G. Kumar, A. Bardhan, M. Nazem, Prediction of the resilient modulus of compacted subgrade soils using ensemble machine learning methods, Transp. Geotech. 36 (2022), <https://doi.org/10.1016/j.tgeo.2022.100827>.

- [75] D.C. Feng, Z.T. Liu, X.D. Wang, Z.M. Jiang, S.X. Liang, Failure mode classification and bearing capacity prediction for reinforced concrete columns based on ensemble machine learning algorithm, *Adv. Eng. Inform.* 45 (2020) 101126, <https://doi.org/10.1016/j.aei.2020.101126>.
- [76] D.K. Bui, T. Nguyen, J.S. Chou, H. Nguyen-Xuan, T.D. Ngo, A modified firefly algorithm-artificial neural network expert system for predicting compressive and tensile strength of high-performance concrete, *Constr. Build. Mater.* 180 (2018) 320–333, <https://doi.org/10.1016/j.conbuildmat.2018.05.201>.
- [77] B.A. Salami, S.M. Rahman, T.A. Oyehan, M. Maslehuddin, S.U. Al Dulaijan, Ensemble machine learning model for corrosion initiation time estimation of embedded steel reinforced self-compacting concrete, *Meas. J. Int. Meas. Confed.* 165 (2020) 108141, <https://doi.org/10.1016/j.measurement.2020.108141>.
- [78] R. Cai, T. Han, W. Liao, J. Huang, D. Li, A. Kumar, H. Ma, Prediction of surface chloride concentration of marine concrete using ensemble machine learning, *Cem. Concr. Res.* 136 (2020) 106164, <https://doi.org/10.1016/j.cemconres.2020.106164>.
- [79] J.S. Chou, C.F. Tsai, A.D. Pham, Y.H. Lu, Machine learning in concrete strength simulations: Multi-nation data analytics, *Constr. Build. Mater.* 73 (2014) 771–780, <https://doi.org/10.1016/j.conbuildmat.2014.09.054>.
- [80] A. Upadhyay, M.S. Thakur, P. Sihag, R. Kumar, S. Kumar, A. Afeeza, A. Afzal, C. Ahamed Saleel, Modelling and prediction of binder content using latest intelligent machine learning algorithms in carbon fiber reinforced asphalt concrete, *Alex. Eng. J.* 65 (2023) 131–149, <https://doi.org/10.1016/j.aej.2022.09.055>.
- [81] X. Wang, P. Pan, J. Li, Real-time measurement on dynamic temperature variation of asphalt pavement using machine learning, *Meas. J. Int. Meas. Confed.* 207 (2023), <https://doi.org/10.1016/j.measurement.2022.112413>.
- [82] Q.A.T. Bui, D.D. Nguyen, M. Iqbal, F.E. Jalal, I. Prakash, B.T. Pham, Prediction of Interface Shear Stiffness Modulus of Asphalt Pavement using Bagging Ensemble-based Hybrid Machine Learning Model, *Arab. J. Sci. Eng.* (2023), <https://doi.org/10.1007/s13369-023-08014-1>.
- [83] J. Zhao, H. Wang, Machine learning based pavement performance prediction for data-driven decision of asphalt pavement overlay, *Struct. Infrastruct. Eng.* (2023), <https://doi.org/10.1080/15732479.2023.2258498>.
- [84] J. Zhao, H. Wang, P. Lu, Machine learning analysis of overweight traffic impact on survival life of asphalt pavement, *Struct. Infrastruct. Eng.* 19 (2023) 606–616, <https://doi.org/10.1080/15732479.2021.1961827>.
- [85] B. Kalpana, A.K. Reshma, S. Senthil Pandi, S. Dhanasekaran, OESV-KRF: Optimal ensemble support vector kernel random forest based early detection and classification of skin diseases, *Biomed. Signal Process. Control.* 85 (2023) 104779, <https://doi.org/10.1016/j.bspc.2023.104779>.
- [86] J. Tu, L. Hu, K.J. Mohammed, B.N. Le, P. Chen, E. Ali, H.E. Ali, L. Sun, Application of logistic regression, support vector machine and random forest on the effects of titanium dioxide nanoparticles using macroalgae in treatment of certain risk factors associated with kidney injuries, *Environ. Res.* 220 (2023) 115167, <https://doi.org/10.1016/j.envres.2022.115167>.
- [87] A. Hola, S. Czarnecki, Random forest algorithm and support vector machine for nondestructive assessment of mass moisture content of brick walls in historic buildings, *Autom. Constr.* 149 (2023) 104793, <https://doi.org/10.1016/j.autcon.2023.104793>.
- [88] A. Ahmad, F. Farooq, P. Niewiadomski, K. Ostrowski, A. Akbar, F. Aslam, R. Alyousef, A.A. Hk, materials Prediction of Compressive Strength of Fly Ash Based Concrete Using Individual and Ensemble Algorithm Prediction of Compressive Strength of Fly Ash Based Concrete Using Individual and Ensemble Algorithm, *Mdpip. Com.* (2021), <https://doi.org/10.3390/ma14040794>.
- [89] M. Jalal, P. Arabali, Z. Grasley, J.W. Bullard, H. Jalal, Behavior assessment, regression analysis and support vector machine (SVM) modeling of waste tire rubberized concrete, *J. Clean. Prod.* 273 (2020) 122960, <https://doi.org/10.1016/j.jclepro.2020.122960>.
- [90] A.S. Ahmad, M.Y. Hassan, M.P. Abdullah, H.A. Rahman, F. Hussin, H. Abdullah, R. Saidur, A review on applications of ANN and SVM for building electrical energy consumption forecasting, *Renew. Sustain. Energy Rev.* 33 (2014) 102–109, <https://doi.org/10.1016/J.RSER.2014.01.069>.
- [91] A. Seifi, M. Ehteram, V.P. Singh, A. Mosavi, Modeling and uncertainty analysis of groundwater level using six evolutionary optimization algorithms hybridized with ANFIS, SVM, and ANN, *Sustain.* 12 (2020) 4023, <https://doi.org/10.3390/SU12104023>.
- [92] M. Azimi-Pour, H. Eskandari-Naddaf, A. Pakzad, Linear and non-linear SVM prediction for fresh properties and compressive strength of high volume fly ash self-compacting concrete, *Constr. Build. Mater.* 230 (2020) 117021, <https://doi.org/10.1016/j.conbuildmat.2019.117021>.
- [93] J. Zeng, P.C. Roussis, A.S. Mohammed, C. Maraveas, S.A. Fatemi, D.J. Armaghani, P.G. Asteris, Prediction of peak particle velocity caused by blasting through the combinations of boosted-chaid and svm models with various kernels, *Appl. Sci.* 11 (2021) 3705, <https://doi.org/10.3390/app11083705>.
- [94] N. Dahiya, B. Saini, H.D. Chalak, Gradient boosting-based regression modelling for estimating the time period of the irregular precast concrete structural system with cross bracing, *J. King Saud. Univ. - Eng. Sci.* (2021), <https://doi.org/10.1016/j.jksues.2021.08.004>.
- [95] T. Han, A. Siddique, K. Khayat, J. Huang, A. Kumar, An ensemble machine learning approach for prediction and optimization of modulus of elasticity of recycled aggregate concrete, *Constr. Build. Mater.* 244 (2020) 118271, <https://doi.org/10.1016/j.conbuildmat.2020.118271>.
- [96] S.S. Rizvon, K. Jayakumar, Strength prediction models for recycled aggregate concrete using Random Forests, ANN and LASSO, *J. Build. Pathol. Rehabil.* 7 (2022), <https://doi.org/10.1007/s41024-021-00145-y>.
- [97] A. Shaqadan, Prediction of concrete mix strength using random forest model, 2016, <http://www.ripulation.com> (accessed July 2, 2020).
- [98] M. Subramanian, K. Shanmugavadivel, P.S. Nandhini, On fine-tuning deep learning models using transfer learning and hyper-parameters optimization for disease identification in maize leaves, *Neural Comput. Appl.* 34 (2022) 13951–13968, <https://doi.org/10.1007/s00521-022-07246-w>.
- [99] D. Kolar, D. Lisjak, M. Pajak?, M. Gudlin, Intelligent fault diagnosis of rotary machinery by convolutional neural network with automatic hyper-parameters tuning using bayesian optimization, *Sensors* 21 (2021) 2411, <https://doi.org/10.3390/s21072411>.
- [100] T.D. Nguyen, T.H. Tran, N.D. Hoang, Prediction of interface yield stress and plastic viscosity of fresh concrete using a hybrid machine learning approach, *Adv. Eng. Inform.* 44 (2020) 101057, <https://doi.org/10.1016/j.aei.2020.101057>.
- [101] F. Bagherzadeh, T. Shafaghfar, Ensemble Machine Learning approach for evaluating the material characterization of carbon nanotube-reinforced cementitious composites, *Case Stud. Constr. Mater.* 17 (2022) e01537, <https://doi.org/10.1016/j.cscm.2022.e01537>.
- [102] W. Dong, Y. Huang, B. Lehane, G. Ma, Multi-objective design optimization for graphite-based nanomaterials reinforced cementitious composites: A data-driven method with machine learning and NSGA-II, *Constr. Build. Mater.* 331 (2022) 127198, <https://doi.org/10.1016/j.conbuildmat.2022.127198>.
- [103] M.F. Iqbal, M.F. Javed, M. Rauf, I. Azim, M. Ashraf, J. Yang, Q. feng Liu, Sustainable utilization of foundry waste: Forecasting mechanical properties of foundry sand based concrete using multi-expression programming, *Sci. Total Environ.* 780 (2021) 146524, <https://doi.org/10.1016/j.scitenv.2021.146524>.
- [104] D.N. Moriasi, J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, T.L. Veith, Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Trans. ASABE.* 50 (2007) 885–900. (<https://elibrary.asabe.org/abstract.asp?aid=23153>) (accessed October 9, 2023).