




Article

Application of Artificial Intelligence Methods for Predicting the Compressive Strength of Green Concretes with Rice Husk Ash

Miljan Kovačević ^{1,*}, Marijana Hadzima-Nyarko ², Ivanka Netinger Grubeša ³, Dorin Radu ⁴
and Silva Lozančić ²

¹ Faculty of Technical Sciences, University of Pristina, Knjaza Milosa 7, 38220 Kosovska Mitrovica, Serbia

² Faculty of Civil Engineering and Architecture Osijek, Josip Juraj Strossmayer University of Osijek, Vladimira Preloga 3, 31000 Osijek, Croatia; mhadzima@gfos.hr (M.H.-N.); lozancic@gfos.hr (S.L.)

³ Department of Construction, University North, 104. Brigade 3, 42000 Varaždin, Croatia; inetinger@unin.hr

⁴ Faculty of Civil Engineering, Transilvania University of Braşov, 500152 Braşov, Romania; dorin.radu@unitbv.ro

* Correspondence: miljan.kovacevic@pr.ac.rs; Tel.: +381-6061-73801

Abstract: To promote sustainable growth and minimize the greenhouse effect, rice husk fly ash can be used instead of a certain amount of cement. The research models the effects of using rice fly ash as a substitute for regular Portland cement on the compressive strength of concrete. In this study, different machine-learning techniques are investigated and a procedure to determine the optimal model is provided. A database of 909 analyzed samples forms the basis for creating forecast models. The derived models are assessed using the accuracy criteria RMSE, MAE, MAPE, and R. The research shows that artificial intelligence techniques can be used to model the compressive strength of concrete with acceptable accuracy. It is also possible to evaluate the importance of specific input variables and their influence on the strength of such concrete.

Keywords: machine learning; compressive strength; concrete; rice husk ash

MSC: 97R40; 62P30; 62G08



Citation: Kovačević, M.; Hadzima-Nyarko, M.; Grubeša, I.N.; Radu, D.; Lozančić, S. Application of Artificial Intelligence Methods for Predicting the Compressive Strength of Green Concretes with Rice Husk Ash. *Mathematics* **2024**, *12*, 66. <https://doi.org/10.3390/math12010066>

Academic Editors: Kunpeng Liu, Pengfei Wang, Yifeng Gao and Yanjie Fu

Received: 17 November 2023

Revised: 16 December 2023

Accepted: 20 December 2023

Published: 24 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the field of concrete composition, cement is a costly material that is in constant demand worldwide. To effectively manage construction costs, it is therefore essential to utilize modern waste materials, mineral additives and other resources to meet the growing demand for concrete. In 2018, the global production of concrete exceeded 10 billion cubic meters [1], and the production of cement, a key constituent, reached 4 billion tons by 2020 [2].

Reducing reliance on cement, a key component of concrete, offers an opportunity to tackle the pollution associated with the cement industry, which is responsible for around 8–10% of global CO₂ emissions [3]. Rice husk, a notable by-product of agriculture, is generated by removing husks from paddy rice. The total production of milled rice worldwide for 2023 was 512,983 thousand metric tons. The largest producer in the world is China, with a production of 145,846 thousand metric tons; the second is India, with a production of 135,755 thousand metric tons. The total milled rice production in these two countries amounts to almost 55% of world production, according to the United States Department of Agriculture (USDA) report from 2023 [4].

The global market for rice husk ash is expected to witness significant growth over the projected timeframe of 2022 to 2029. According to an analysis by Data Bridge Market Research, the market is registering a compound annual growth rate (CAGR) of 4.9% during this period and will reach a value of USD 21,381.45 thousand by 2029. The major catalyst

for the growth of the global rice husk ash market is its wide applicability in the construction sector, which can be attributed to its high silica content [5].

Numerous agricultural and industrial residues, including rice husk ash (RHA), sugarcane bagasse ash, silica fume and others, are used as supplementary cementitious materials in concrete production, leading to a substantial decrease in the reliance on conventional Portland cement [6]. The study by Ramagiri et al. in 2019 evaluated the shrinkage behavior of three alkali-activated binder (AAB) mixtures activated by sodium hydroxide and sodium silicate, containing varying proportions of fly ash and slag. Multiple linear regression models are developed to predict shrinkage strains, considering age and fly ash percentage. The objective is to establish a generalized equation for predicting shrinkage in different AAB mixes cured at room temperature. The proposed models, ranked by Root Mean Square Error (RMSE), show a high correlation ($R^2 = 0.937$) and offer a reliable tool for estimating shrinkage in similar AAB mixtures [7]. The considered procedure is recommended for similar prediction problems.

The study of Ramagiri et al. in 2021 is focused on understanding factors influencing the compressive strength of AAB concrete and compared the accuracy of different Random Forest (RF) configurations in predicting the compressive strength of ambient-cured AAB concrete. The ranger algorithm with reliefF feature selection exhibited the most accurate predictions based on MAE and RMSE values. The paper recommends the application of the mentioned RF algorithm for the prediction of compressive strength and the determination of important input variables for AAB concrete [8].

RHA comprises amorphous silicon oxide (SiO_2), commonly known as silica. This high pozzolanic characteristic renders it a viable candidate for partially substituting Portland cement in concrete blends [9–12]. Studies have indicated that the incorporation of RHA as a substitute for cement can effectively reduce the hydration temperature in concrete compared to traditional Portland cement concrete [13]. Furthermore, the utilization of RHA enhances various properties of the concrete mixture, including strength, shrinkage, and durability, surpassing the performance achieved with pure Portland cement as a binding agent [13]. In particular, the substitution of up to 50% of Portland cement with RHA demonstrated superior mechanical strength compared to the standard concrete mixture utilizing pure Portland cement within the initial 3 to 7 days of aging [14].

Nasir [15] investigated the variation tendencies of compressive strength (CS) of concrete with different proportions of RHA. The results showed that the strength of the concrete exceeded that of the control group up to 30% RHA and peaked at 15%, after which it began to decrease.

Moreover, existing research has indicated that the compressive strength (CS) of concrete containing RHA is influenced by factors such as age, cement content, water-cement ratio, water content, coarse aggregate content, fine aggregate content, and the presence of superplasticizer [16–18].

While widespread, the conventional practice of open burning of rice husks results in an undesirable increase in crystalline particles unsuitable for concrete applications, as highlighted in the work of Hwang and Chandra [19]. To counter this issue, specialized combustion techniques, as recommended in prior investigations [20,21], regulate temperature and manage carbon content—critical factors influencing the quality of concrete. Research suggests that maintaining combustion temperatures within the range of 500 to 700 °C produces the optimal content of amorphous particles, and the specific surface area of up to 150 m²/g is maximized at this temperature, a conclusion supported by multiple studies [21–26]. The combustion of rice husk yields 20–25% of RHA [27,28].

In the study conducted by Chindaprasirt et al. [29], concrete incorporating RHA demonstrated significant resistance to sulfate attacks. Additionally, Thomas et al. [1], in a comprehensive review, noted that RHA-infused concrete exhibits a dense microstructure, resulting in a substantial reduction of water absorption, potentially by as much as 30 percent. With a silica content exceeding 90%, RHA proves effective as a supplementary cementitious material (SCM) in concrete production [30].

Iqtidar et al. explored 2021 sustainable alternatives for eco-friendly concrete by leveraging machine learning to predict compressive strength in RHA concrete [31]. Employing 192 data points and four machine learning methods (ANN, ANFIS, NLR, and linear regression), the research assesses RHA concrete properties, highlighting the superior performance of ANN and ANFIS. The findings advocate for the application of these ML models in the construction sector to efficiently evaluate material properties and input parameter influences.

Nasir Amin et al. in 2022 investigated the positive influence of RHA in concrete and employed supervised machine learning techniques, including decision trees (DT), bagging regressors, and AdaBoost regressors, to predict the compressive strength of RHA-based concrete [32]. The models were developed using a database containing 192 data points from the available literature. Age, RHA content, cement, superplasticizer, water, and aggregate were the variables employed in the modeling of RHAC. Model evaluation involved metrics such as R^2 , mean absolute error (MAE), root mean square error (RMSE), and root mean square log error (RMSLE), alongside k-fold cross-validation to ensure accuracy. The bagging regressor model outperformed DT and Adaboost, achieving an R^2 value of 0.93.

Amlashi et al. in 2022 introduced machine learning-based models, including Artificial Neural Network (ANN), Multivariate Adaptive Regression Spline (MARS), and M5P Model Tree, to predict compressive strength (CS) in RHA-containing concretes [33]. For this purpose, the models were developed employing 909 data records collected through technical literature. The findings reveal that all models provide reliable CS estimations, with the ANN model outperforming the others. A parametric study identifies factors influencing CS, showing that increasing cement, coarse aggregate, and age while decreasing water, fine aggregate, RHA, and superplasticizer enhance CS. Additionally, sensitivity analysis highlights coarse aggregate content as the most influential parameter affecting CS values.

Bassi et al. in 2023 explored the use of six Machine Learning (ML) algorithms—Linear Regression, Decision Tree, Gradient Boost, Artificial Neural Network, Random Forest, and Support Vector Machines—to predict the compressive strength of RHA-based concrete [34]. With 462 data points and twelve input features, the Decision Tree, Gradient Boost, and Random Forest models exhibit superior accuracy ($R^2 > 0.92$) and minimal errors in predicting compressive strength. Sensitivity analysis highlights the significant impact (more than 95%) of the specific gravity of RHA and water–cement ratio on compressive strength, distinguishing them as key parameters.

The paper written by Li et al. in 2023 introduces an innovative hybrid artificial neural network model optimized using a reptile search algorithm with circle mapping to predict the compressive strength of RHA concrete [35]. The proposed model utilizes 192 concrete data points with six input parameters (age, cement, rice husk ash, superplasticizer, aggregate, and water) and is trained and compared with five other models. Four statistical indices assess predictive performance. The results indicate that the hybrid artificial neural network model excels in prediction accuracy with R^2 (0.9709), Variance Accounted For—VAF (97.0911%), RMSE (3.4489), and MAE (2.6451).

The study conducted by Nasir Amin et al. in 2023 investigated the application of modern machine intelligence techniques, specifically multi-expression programming (MEP) and gene expression programming (GEP), for predicting the compressive strength (CS) of RHA concrete [36]. Additionally, Shapley Additive Explanations (SHAP) analysis is employed to assess the impact and interaction of raw materials on the CS of RHA concrete. Utilizing a comprehensive dataset of 192 data points with six inputs (cement, specimen age, RHA, superplasticizer, water, and fine aggregate), the researchers find that both GEP and MEP models provide reliable CS predictions, aligning closely with actual values. In comparing their performance, MEP, boasting an R^2 of 0.89, outperforms the GEP model, which achieves an R^2 of 0.83. SHAP analysis identifies specimen age as the most crucial factor, followed by cement, positively correlating with the CS of RHA.

With their ability to consider diverse parameters and data from varied concrete samples, machine learning models surpass the limitations of empirical formulas, offering a

robust framework for CS prediction. Despite the widespread application of artificial intelligence algorithms in concrete CS prediction, more studies need to focus on RHA concrete. This study aims to fill this void by analyzing different machine learning algorithms and proposing the best one to enhance CS prediction, specifically in RHA concrete. This contribution aims to advance the field of intelligent optimization models for concrete properties.

The contribution of this research can be seen in the fact that it has attempted to analyze a large number of machine learning algorithms on a basis that is significantly larger than the majority of similar research mentioned in the literature. Each analyzed model was optimized in terms of the hyperparameters and then evaluated using the test dataset. In this way, the models offering the highest accuracy in predicting the compressive strength in the case considered were defined. In addition, the influence of certain variables on the compressive strength of concrete with RHA addition was analyzed.

2. Materials and Methods

The research employs various machine learning methods, including multiple linear regression model, regression trees and created ensemble models (tree bagger, random forests, and boosted trees), support vector machines, neural networks, an ensemble of neural networks, and Gaussian process regression models. The objective is to establish relationships between input variables and the compressive strength of Rice Husk Ash Concrete (RHAC).

2.1. Multiple Linear Regression Model

If the observed problem can be treated as a problem of one dependent and several independent variables, it is a suitable situation for data analysis using the multiple regression method. If the relationship between variables is linear, the case is reduced to a multiple linear model.

Let them be Y dependent variable, x_1, x_2, \dots, x_p independent variables, then the linear model can be written in the following form:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon. \quad (1)$$

In the previous expression, $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ are the unknown parameters to be estimated and ε are the residuals. The Y variable is also called the response variable, that is, the output variable, while the x -variables are called inputs, that is, explanatory variables.

If there are n experiments or measurements, the model can be written as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (2)$$

If the following notation is introduced:

$$Y = \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ 1 & x_{31} & x_{31} & \dots & x_{3p} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \dots \\ \beta_p \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{bmatrix}, \quad (3)$$

a shorter notation in matrix form can be obtained

$$Y = X\beta + \varepsilon, \quad (4)$$

where $X\beta$ is the system component of the model, and ε is the random component of the model.

One of the ways to determine the parameters consists in taking the value for which the sum of the residuals is minimal, which can be expressed as follows:

$$S = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^T \varepsilon = (Y - X\beta)^T (Y - X\beta) \quad (5)$$

$$\frac{\partial S}{\partial \beta} = 0 \quad (6)$$

Applying the previous expressions gives an estimate for β equal to:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (7)$$

and estimate of the system component βX is equal to

$$\hat{Y} = X\hat{\beta} = X(X^T X)^{-1} X^T Y = HY. \quad (8)$$

Assesment of error ε is equal to

$$\hat{\varepsilon} = Y - X\hat{\beta} = (E - H)Y. \quad (9)$$

2.2. Regression Trees Ensembles: Bagging, Random Forest, and Boosted Trees

The core idea underlying regression trees is to partition the input space into distinct regions and allocate predictive values to these segments, facilitating predictions based on pertinent conditions and data characteristics. This machine learning model, applicable to regression and classification tasks, adopts a tree-like structure composed of nodes and branches (Figure 1). Each node represents a specific condition related to the input data, evaluated as data progresses through the tree.

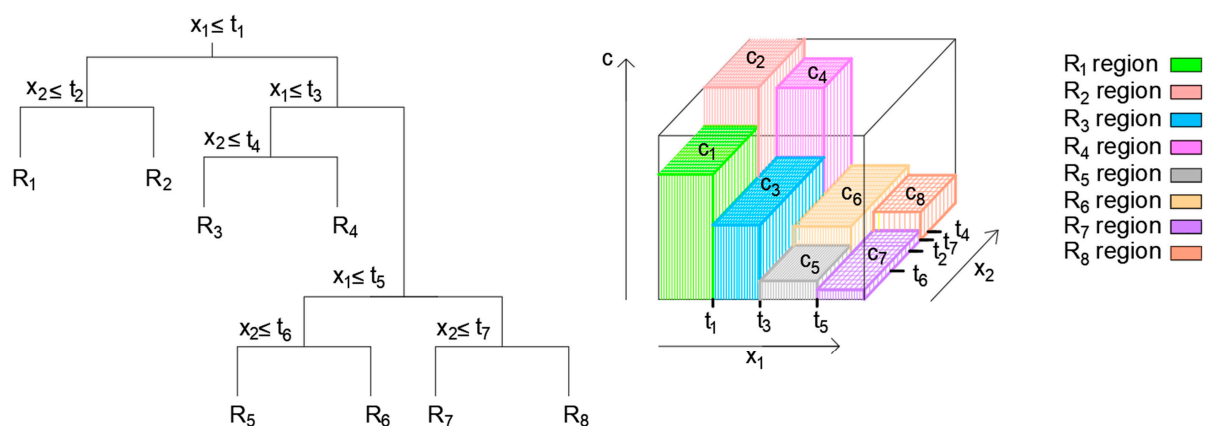


Figure 1. Demonstration of the segmentation of input space into unique regions and the 3D regression surface encapsulated within the structure of a regression tree [37].

The process begins at the root node to predict outcomes, where the initial condition associated with input features is considered. Based on the truth value of this condition, branches are followed to reach subsequent nodes recursively until a leaf node is reached. At the leaf node, a value is obtained, serving as the predicted result, typically numeric for regression tasks.

As the tree is traversed, the input space changes, initially represented by a single set at the root node. The algorithm progressively divides the input space into smaller subsets based on conditions, tailoring predictions to different regions.

Constructing regression trees involves determining optimal split variables and points to partition the input space effectively. These variables are identified by minimizing a spe-

cific expression (Equation (10)) considering all input features, aiming to minimize the sum of squared differences between observed and predicted values in resulting regions [38–41].

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \quad (10)$$

Once identified, the tree-building process iteratively divides regions in a “greedy approach”, emphasizing local optimality at each step [40]. The binary recursive segmentation divides the input space into non-overlapping regions characterized by mean values. Ensemble methods like Bagging, Random Forest, and Boosted Trees further enhance predictive capabilities.

Bagging involves creating subsets of the training dataset through random sampling with replacement, training separate regression tree models, and aggregating predictions to reduce variance (Figure 2). Random Forests introduce diversity by creating multiple regression trees trained using different subsets of variables randomly chosen for splitting on distinct bootstrap samples, decorrelating individual trees, and reducing variance [41]. The ensemble’s final prediction is generated by aggregating predictions from these decorrelated trees, resulting in a robust and high-performing model (Figure 2).

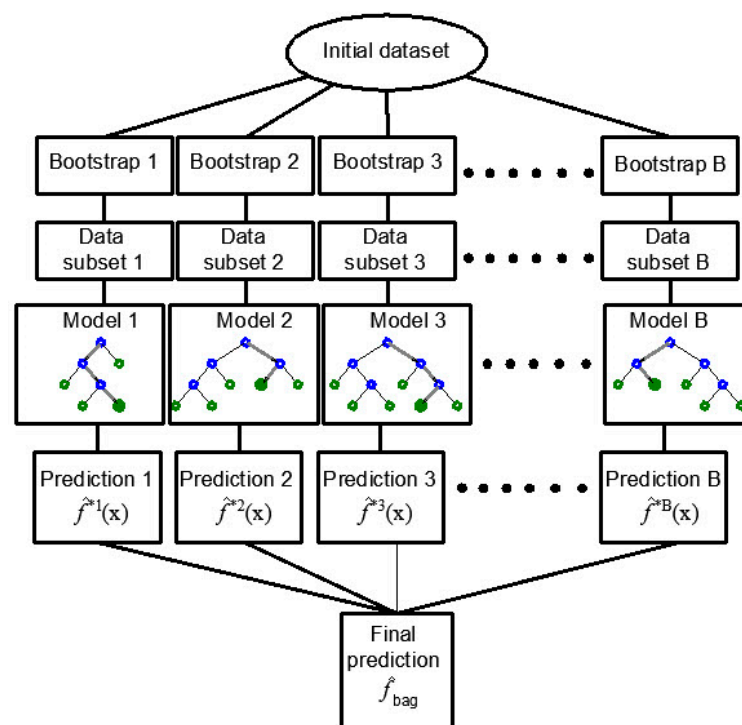


Figure 2. Formation of regression tree ensembles through the utilization of a bagging approach [42].

Utilizing Boosting Trees involves a sequential training approach where each subsequent regression tree added to the ensemble is introduced to improve the overall model’s performance (Figure 3). Within the context of Gradient Boosting, a widely employed technique, submodels are introduced iteratively and chosen based on their effectiveness in estimating the residuals or errors of the prior model in the sequence [43–45]. This iterative process culminates in a conclusive ensemble model comprising multiple submodels that collectively produce highly accurate predictions.

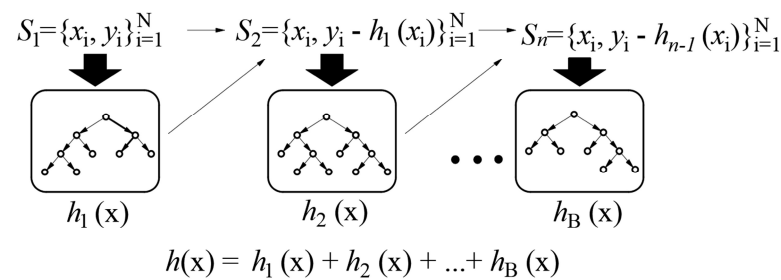


Figure 3. Integration of gradient boosting within the ensemble of regression trees [42].

Within the gradient-boosting tree domain, the learning rate's significance, commonly represented as "lambda" (λ), becomes apparent as a pivotal hyperparameter influencing each tree's impact on the ensemble's ultimate prediction. This learning rate is crucial in dictating how swiftly the model adapts to errors introduced by preceding trees throughout the boosting process.

To evaluate the proportional impact of predictor variables, using the BT method relies on the frequency with which a variable is chosen for splitting. This selection frequency is then weighted by the squared enhancement to the model attributable to each split and averaged across all trees [43–45].

2.3. Support Vector Machine for Regression (SVR)

Consider a training dataset $\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \in \mathbb{R}^n \times \mathbb{R}$, where $x_i \in \mathbb{R}^n$ is the n -dimensional vector representing the model's inputs, and y_i corresponds to the observed responses. The approximation function is given by the expression (11):

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) \mathbf{K}(\mathbf{x}_i, \mathbf{x}) + b. \quad (11)$$

Here, \mathbf{K} denotes the kernel function, and α_i , α_i^* and b are parameters obtained by minimizing the error function. The SVR regression employs the empirical risk function (12) [46–48]:

$$R_{emp}^{\varepsilon}(\mathbf{w}, b) = \frac{1}{l} \sum_{i=1}^l |y_i - f(\mathbf{x}_i, \mathbf{w})|_{\varepsilon}. \quad (12)$$

The SVR algorithm aims to minimize both the empirical risk R_{emp}^{ε} and the $\|\mathbf{w}\|^2$ value concurrently. Vapnik's linear loss function (13) introduces a ε -insensitivity zone, defined by:

$$|y - f(\mathbf{x}, \mathbf{w})|_{\varepsilon} = \begin{cases} 0 & \text{if } |y - f(\mathbf{x}, \mathbf{w})| \leq \varepsilon \\ |y - f(\mathbf{x}, \mathbf{w})| - \varepsilon & \text{otherwise.} \end{cases} \quad (13)$$

This leads to the minimization problem expressed as (14):

$$R = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l |y_i - f(\mathbf{x}_i, \mathbf{w})|_{\varepsilon}. \quad (14)$$

The minimization is equivalent to minimizing following function (15):

$$R_{w, \xi, \xi^*} = \frac{1}{2} \left[\|\mathbf{w}\|^2 + C \left(\sum_{i=1}^l \xi + \sum_{i=1}^l \xi^* \right) \right], \quad (15)$$

where ξ and ξ^* are slack variables. Figure 4 illustrates the nonlinear SVR with a ε -insensitivity zone.

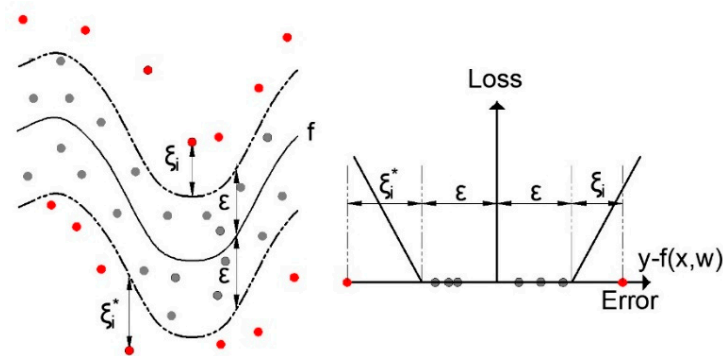


Figure 4. Visualization of Nonlinear Support Vector Regression (SVR) with a ε -insensitivity zone [42].

Linear, RBF, and sigmoid kernels were employed in this study [46–48]. The LIBSVM software, utilizing the SMO optimization algorithm implemented within the MATLAB program, was utilized for this purpose [49,50].

2.4. Artificial Neural Network and Artificial Neural Networks Ensemble

Artificial neural networks emulate the parallel processing observed in the human brain, employing interconnected artificial neurons within a parallel structure. Among neural network architectures, the multilayer perceptron is notable, characterized by its forward signal propagation through three essential layers: input, hidden, and output layers.

In a general configuration, each neuron within one layer connects with every neuron in the subsequent layer (Figure 5a). This connectivity is illustrated in Figure 4, representing a three-layer multilayer perceptron network with n inputs and a single output. The network's behavior is influenced by factors such as the number of neurons and the choice of activation function. To achieve the ability to approximate any function, neural networks must incorporate nonlinear activation functions in the hidden layer. This enables the network to effectively capture and approximate the often complex and nonlinear relationships between input and output variables.

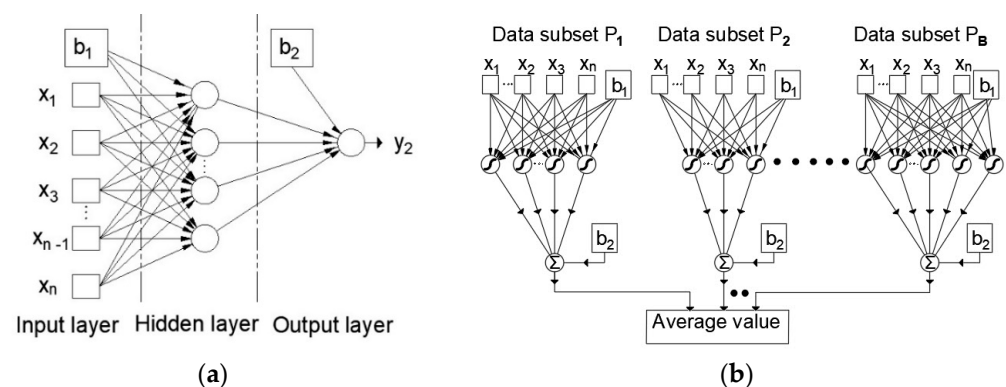


Figure 5. (a) Architecture of a multilayer perceptron artificial neural network; (b) Formation of a neural network ensemble using the Bootstrap Aggregating (Bagging) approach [37].

For instance, a model featuring a single hidden layer with neurons employing a sigmoid activation function, coupled with output layer neurons utilizing a linear activation function, can successfully approximate arbitrary functions given a sufficient number of neurons in the hidden layer.

Determining the optimal neural network structure entails finding the right balance, especially when dealing with multilayer perceptron (MLP) architectures as universal approximators (Figure 5). While an exact method for pinpointing the minimum necessary number of neurons remains elusive, a practical approach involves approximating the upper

limit. This upper limit signifies the maximum number of hidden layer neurons suitable for modeling a system based on a specific dataset.

A precise and reliable method for determining the minimum required number of neurons has not been established. It is recommended to choose a smaller value for N_H from the set of inequalities (16) and (17), where N_i represents the number of neural network inputs, and N_s represents the number of training samples. The proposed criteria are outlined as follows:

$$N_H \leq 2 \times N_i \quad (16)$$

$$N_H \leq \frac{N_s}{N_i + 1} \quad (17)$$

In cases where it is desired to increase the accuracy of the model, the synergistic effect of the combination of a number of models that form an ensemble of neural networks can be examined (Figure 5b).

The individual models that make up the structure of the ensemble are called base models or submodels. To improve model robustness, it is essential to generate a collection of base models using the Bootstrap method for creating different sets of data for training. This involves employing an ensemble approach, training multiple neural networks on these sets, and subsequently averaging their outputs.

2.5. Gaussian Process for Regression (GPR)

Gaussian processes offer a versatile approach to modeling functions in regression tasks, allowing for uncertainty incorporation and effective prediction. The selection of covariance functions and hyperparameters adds adaptability to capture variable relationships [51].

In Gaussian process modeling, the task involves estimating an unknown function, denoted as $f(\cdot)$, in nonlinear regression scenarios. This function adheres to a Gaussian distribution characterized by a mean function $\mu(\cdot)$ and a covariance function $k(\cdot, \cdot)$. The pivotal Gaussian process regression (GPR) component, the covariance matrix K , is influenced by the chosen kernel function (k) [51].

The kernel function (k) is fundamental for assessing covariance or similarity between input data points (x and x'). A frequently employed squared exponential kernel takes the form (18):

$$k(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2l^2}\right) \quad (18)$$

In this expression, σ^2 signifies signal variance, the exponential function “exp” models the similarity between x and x' , and the parameter l , termed the lengthscale, regulates smoothness and spatial extent.

Data observations in a dataset, denoted as $y = \{y_1, \dots, y_n\}$, are viewed as a sample from a multivariate Gaussian distribution (19):

$$(y_1, \dots, y_n)^T \sim N(\mu, K), \quad (19)$$

Gaussian processes are applied to model the connection between input variables x and target variable y , accounting for additive noise $\varepsilon \sim N(0, \sigma^2)$. The primary aim is to estimate the unknown function $f(\cdot)$. The conditional distribution of a test point's response value y^* , given observed data $y = (y_1, \dots, y_n)^T$, is presented, (20), (21) as $N(\hat{y}^*, \hat{\sigma}^{*2})$:

$$\hat{y}^* = \mu(x^*) + K^{*T} K^{-1}(y - \mu), \quad (20)$$

$$\hat{\sigma}^{*2} = K^{**} + \sigma^2 - K^{*T} K^{-1} K^*. \quad (21)$$

Hyperparameters could play a crucial role in unveiling the significance of individual inputs through an approach known as Automatic Relevance Determination (ARD). For instance, the squared exponential covariance function (22), which employs distinct length scale parameters for each input (ARD SE):

$$k(x_p, x_q) = v^2 \exp \left[-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_p^i - x_q^i}{r_i} \right)^2 \right]. \quad (22)$$

In the domain of covariance functions, r_i designates the length scale associated with input dimension i . It is imperative to acknowledge that a significantly large value of r_i results in a diminished importance of the i -th input [51]. The hyperparameters $\{v, r_1, \dots, r_n\}$, in conjunction with the noise variance σ^2 , are subject to estimation through the maximum likelihood method. The log-likelihood of the training data is expressed as (23):

$$L(v, r_1, \dots, r_n, \sigma^2) = -\frac{1}{2} \log \det K - \frac{1}{2} y^T K^{-1} y - \frac{n}{2} \log 2\pi. \quad (23)$$

3. Dataset

A comprehensive dataset concerning concretes incorporating RHA was compiled through extensive documentation, with specific details outlined in Table 1. The compiled dataset encompassed information related to mixture proportions, age, and compressive strength values for 909 concrete samples [33]. The compressive strength values ranged from 2.4 to 118.8 MPa. To ensure consistency in the conditions of all concrete samples, the compressive strength values for samples were converted to equivalent cubic compressive strength using UNESCO conversion factors [52]. This process was carried out according to the methodology presented in the study conducted by Elwell and Fu in 1995 [52].

Table 1. Descriptive statistics for the variables [33].

Num. of Data	Water (kg/m ³)	Cement (kg/m ³)	Fine Aggregate (FA) (kg/m ³)	Coarse Aggregate (CA) (kg/m ³)	RHA (kg/m ³)	Age (Days)	SP (kg/m ³)	CS (MPa)	Ref.
48	215–255	300–2351	580–710	1160–1185	0–134	7–90	0	13.27–47.6	[53]
16	138–207	400–571	578–612	1027–1088	0–171	3–150	10–14.275	31.5–85	[54]
14	153–154	345–385	667–674	1086–1102	0–38	1–365	2.5–3.9	25.34–60.74	[55]
24	128	356–410	786	1044–1062	0–72	3–365	0	42–92	[19]
18	210	280–350	844–870	854–881	0–70	1–180	0	2.4–39.4	[27]
28	178–180	356–600	570–636	906–968	0–153	3–90	0.89–12.48	36.86–106.82	[56]
60	160–170	400–550	540–567	1261–1324	0–110	1–90	5–6.22	18.9–86.8	[57]
36	205–228	228–325	890–900	927–940	0–97	1–180	0–3.67	9.2–45.8	[58]
27	164–204	327–534	690–758	983–1050	0–85	7–91	0.462–4.33	29.22–77.96	[17]
20	150	375	770	1200	0–75	7–90	0–15	30–64	[59]
16	138–173	322–513	720–845	975–1045	0–34	28–90	0–10.26	37–82.1	[16]
20	150–153	392–560	735–764	943–981	0–168	7–180	1.5–9	103.06–118.82	[10]
21	201	266–380	570	1140	0–114	7–28	0	27.22–39.55	[60]
32	203	249–383	561	1148	0–134	7–90	0	27.22–45.98	[61]
32	131.97–202.99	249–383	575	1150	0–134	1–28	0	10.4–46.7	[15]
60	137–190	304–500	745–868	933–995	0–100	1–28	0–3.88	19–68.6	[62]
12	207	313–391	750	994	0–78	1–180	0	19.1–48.1	[26]
16	157–173	492–518	484–510	983–1025	0–52	7–180	0	66.3–93.5	[63]
20	207	313–391	750	994	0–78	1–28	0	17.2–50.2	[20]
24	122.4–150.5	340–430	332–335	1012–1014	0–64.5	7–90	0	30.36–62.5	[14]
20	210	297–396	844	951	0–99	28–360	0	29.7–47.8	[64]
42	212–221	383–783	344–737	933	0–171	1–90	0.3–3.7	19.4–89.725	[65]
76	154–165	280–550	490–560	1200–1345	0–165	3–60	0.87–9.75	30.5–92	[66]
50	166	336–474	433.72–636	1108–1113	0–108	7–180	0–9.4	5.87–73	[67]
60	132.4–178.4	378.8–553.8	543.8–720.7	951.6–1048.3	25–71.7	28	5.3–93.0	51.5–111.8	[68]
5	192	278.7–348.4	573	1189.5	0–69.6	28	0	16.03–29.3	[69]
10	125–156	312–390	713	1079	0–78	7–28	1.85–3.51	30.22–48.53	[70]
5	130	418.5–465	556–562	1268.3–1280.9	0–46.5	14	0	46.2–52.6	[71]
27	185	261.37–461.25	582–623.79	1204–1287.9	0–69.19	7–28	0	19.76–43.16	[72]
8	153	238–340	763	1144	0–102	28	3.4–9.2	28.38–34.98	[73]
35	140–167	456–537	516–669	1055	0–80.6	1–128	5.1–5.37	25–103.5	[74]
12	153	325.1–382.5	482	1394.1	0–57.37	7–28	3.25–3.82	25.9–42.45	[75]
15	138	240–300	660	1290	0–60	7–28	2.4	19.86–31.88	[76]

The input variables encompassed the quantities of fine aggregate, coarse aggregate, Portland cement, RHA, superplasticizer, and water, along with the age of the concrete samples (Table 2). In this study, the division into training and test datasets was accomplished using the randperm function in Matlab. The aim was to create two subsets with similar statistical indicators, adhering to an 80% training and 20% testing ratio. The randperm function was employed, followed by calculations of key statistical parameters, including maximum (Max.), minimum (Min.), average (Mean), mode (Mode), and standard deviation (St.Dev.).

Table 2. Statistical summaries for the variables.

Statistical Analysis of Input and Output Parameters for Training Set						
	Max.	Min.	Mean	Mode	St.Dev.	Count
Water (W) [kg/m ³]	255.00	122.40	175.51	165.00	29.14	727.00
Cement (C) [kg/m ³]	2351.00	228.00	441.95	400.00	262.71	727.00
Sand (S) [kg/m ³]	900.00	332.00	640.99	750.00	129.94	727.00
Coarse aggregate (CA) [kg/m ³]	1394.06	854.00	1090.30	933.00	127.54	727.00
Rice Husk Ash (RHA) [kg/m ³]	171.00	0.00	51.23	0.00	41.28	727.00
Age of samples (A) [days]	365.00	1.00	39.85	28.00	58.90	727.00
Superplasticizer (SP) [kg/m ³]	72.60	0.00	3.46	0.00	6.74	727.00
Compressive strength (CS) [MPa]	118.83	2.40	50.08	52.00	22.66	727.00
Statistical analysis of input and output parameters for Test set						
	Max.	Min.	Mean	Mode	St.Dev.	Count
Water (W) [kg/m ³]	252.00	125.00	176.36	165.00	30.91	182.00
Cement (C) [kg/m ³]	2351.00	228.00	478.52	450.00	357.32	182.00
Sand (S) [kg/m ³]	900.00	335.00	657.25	745.00	124.88	182.00
Coarse aggregate (CA) [kg/m ³]	1394.06	854.00	1088.86	933.00	130.15	182.00
Rice Husk Ash (RHA) [kg/m ³]	171.00	0.00	52.10	0.00	42.58	182.00
Age of samples (A) [days]	365.00	1.00	38.48	28.00	55.26	182.00
Superplasticizer (SP) [kg/m ³]	38.30	0.00	3.55	0.00	5.87	182.00
Compressive strength (CS) [MPa]	118.83	5.87	51.53	65.00	25.19	182.00

Frequency histograms and the mutual correlation between variables are shown in Figure 6.

In the development of models using machine learning methodologies, a supervised learning approach was implemented to ensure that the generated models predict values within the range they acquired during the learning phase. Therefore, it becomes crucial to highlight the boundaries of each individual variable.

Within the entire set of test data, the ratio of RHA to cement ranged from a minimum value of 0% to a maximum value of 53.82%, with a mean value of 13.42%. The ratio of water to binder (cement together with RHA) ranged from a minimum value of 0.0979 to a maximum value of 0.8, with a mean value of 0.3928 for the entire data set.

In assessing the effectiveness of the prediction model, a set of criteria was employed, including the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Pearson's Linear Correlation Coefficient (R), and Mean Absolute Percentage Error (MAPE) [37].

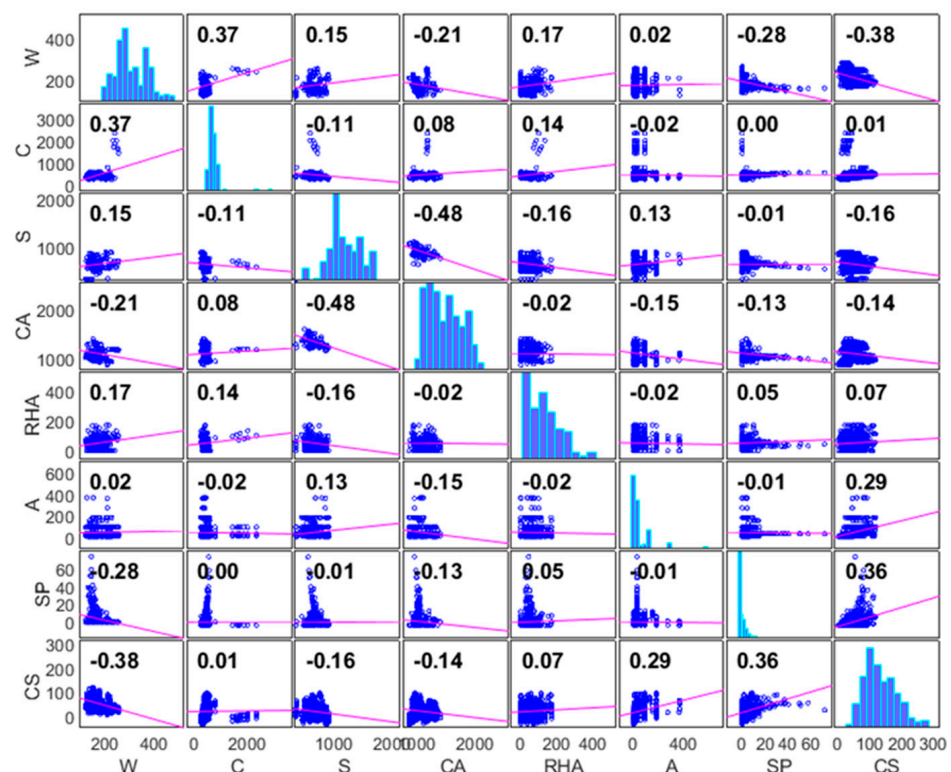


Figure 6. Frequency histogram and correlation between variables.

4. Results and Discussion

An examination of models based on decision trees was conducted, employing Mean Squared Error (MSE) values as the training criterion. The grid-search method was utilized to determine the optimal model parameters. Following the training or calibration phase, the models' generalization characteristics were evaluated on the test dataset using RMSE, MAE, MAPE, and R criteria.

The analysis encompassed the utilization of the following methods:

- Multiple linear regression,
- Bagging method (TreeBagger—TB),
- Random Forests (RF) method,
- Boosted Trees (BT) method,
- Support vector regression,
- Neural networks (standalone and ensemble models),
- Gaussian process regression (GPR).

The mentioned methods were applied because they are appropriate for the amount of data (909 samples) that was considered in the paper. In addition, all the mentioned methods have been shown in the literature to be very suitable for implementation in similar problems of prediction of the continuous value of the compressive strength of concrete, i.e., consideration of regression problems of a similar nature. All calculations were implemented using Matlab software, while the SVR model used Matlab and the LIBSVM Library for Support Vector Machines simultaneously.

In the implementation of multiple linear regression model, the parameters of the model were calculated in the Table 3, as well as the corresponding t statistics and p values for assessing the significance of individual variables.

The estimated intercept term (Intercept: 174.4430) represents the predicted value of the dependent variable (y) when all predictor variables (x_1 to x_7) are zero. The t -statistic (t Stat) tests the hypothesis that the intercept is zero. In this case, a higher t -statistic indicates that the intercept is significantly different from zero. Also, the very small p -value

suggests that the intercept is highly statistically significant. The coefficient for x_1 is highly statistically significant (p -value: 2.0670×10^{-28}), and the negative t -statistic (-11.546) indicates that the relationship is significantly different from zero. Similar interpretations apply to the coefficients for x_2 to x_7 . All coefficients can be considered as statistically significant, suggesting that each predictor variable contributes significantly to the prediction of y . Additionally, the overall model is highly statistically significant, as indicated by the very small p -value associated with the F -statistic (5.04×10^{-74}).

Table 3. Parameters for the multiple linear regression model.

	Coefficient	SE	tStat	p Value
Intercept	174.4430	10.6943	16.3117	3.9661×10^{-51}
x_1 —Water (W) [kg/m ³]	−0.3158	0.0273	−11.5465	2.0670×10^{-28}
x_2 —Cement (C) [kg/m ³]	0.0134	0.0028	4.8252	1.7082×10^{-6}
x_3 —Sand (S) [kg/m ³]	−0.0431	0.0059	−7.2656	9.6885×10^{-13}
x_4 —Coarse aggregate (CA) [kg/m ³]	−0.0512	0.0062	−8.3186	4.4579×10^{-16}
x_5 —Rice Husk Ash (RHA) [kg/m ³]	0.0367	0.0166	2.2092	0.0275
x_6 —Age of samples (A) [days]	0.1108	0.0113	9.7783	2.7601×10^{-21}
x_7 —Superplasticizer (SP) [kg/m ³]	0.6695	0.10561	6.3395	4.0692×10^{-10}

In the implementation of the TreeBagger (TB) and Random Forests (RF) methods, various values of adaptive model parameters were explored, including (Figure 7):

1. Number of generated trees (B). Throughout this analysis, the maximum number of generated trees was constrained to 500.
2. The number of variables utilized for splitting within the tree. TB and RF models operate on a similar mechanism, with the key distinction being that the TB model employs all variables as potential tree split points, while the RF model uses only a specific subset of the entire variable set. Following the recommendation in L. Bryman's paper on Random Forests [39], it is advised that the subset m of variables for splitting should be $p/3$ of the total number of variables or predictors p . In this study, values of m from 1 to 7 (Figure 7) were examined.
3. The minimum number of data or samples assigned to a leaf (min leaf size) within a tree. Consideration was given to values ranging from 1 to 10 samples per tree leaf.

The optimal model from TB and RF models is the TB model with 500 trees and a leaf size equal to 1 (Figure 7).

In the Boosting Trees method, the following model parameters were taken into account (Figure 8):

1. Number of generated trees (B): To prevent overtraining, the maximum limit of base models within the ensemble was set to 100.
2. Learning rate (λ): This parameter, determining the model's training speed, was investigated across various values, including 0.001, 0.01, 0.1, 0.25, 0.5, 0.75, and 1.0.

Number of splits in the tree (d): Trees with a maximum number of splits ranging from $2^0 = 1$ to $2^8 = 256$ were generated.

The optimal model obtained by grid search, highlighted in yellow in Figure 8, featured 100 generated trees, a learning rate of 0.10, and a maximum of 128 splits.

To develop an effective regression model using the support vector regression (SVR) method, the selection of an appropriate kernel function is crucial. Furthermore, determining the parameters of the selected kernel functions and the penalty parameter (C) is essential for optimal performance.

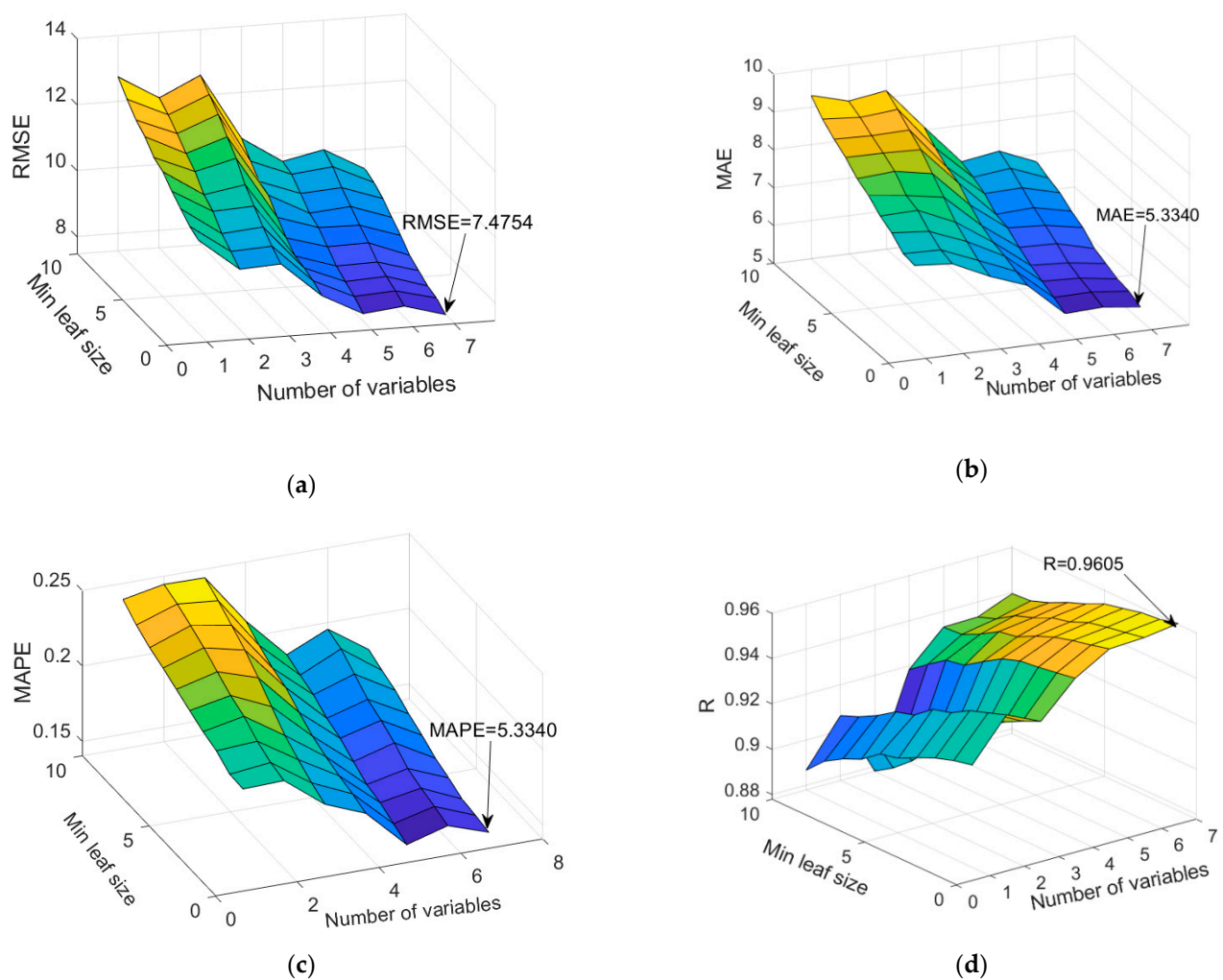


Figure 7. Evaluation of RF and TB models accuracy with respect to the number of randomly selected splitting variables and minimum leaf size: (a) Root Mean Squared Error (RMSE), (b) Mean Absolute Error (MAE), (c) Mean Absolute Percentage Error (MAPE), (d) Coefficient of Correlation (R).

In this study, the investigation involves testing various kernel functions to identify the most suitable one. Linear, RBF and sigmoid kernels for SVR models were analyzed, with input data normalized to the range $[0, 1]$ before model training and testing. The optimal model was identified using the grid search algorithm for all kernels, resulting in specific parameter values for each:

Linear Kernel: $C = 38.27$, $\varepsilon = 0.1012$

RBF Kernel: $C = 94.26$, $\varepsilon = 0.0164$, $\gamma = 3.3364$

Sigmoid Kernel: $C = 121.10$, $\varepsilon = 0.0947$, $\gamma = 0.0085$.

A comparative analysis of different SVR models reveals variations in accuracy based on different criteria, particularly dependent on the kernel function. Linear and sigmoid kernel models demonstrate similar accuracy across various criteria, while the RBF kernel model exhibits significantly higher accuracy in comparison, as indicated in Table 4.

This research undertook an investigation into the architectural configuration of a neural network. The Multilayer Perceptron (MLP) neural network comprises a single input layer, a hidden layer, and an output layer of neurons.

It is important to note that neurons in the output layer are characterized by a linear activation function, while those in the hidden layer are subject to a nonlinear activation function, known for its universal approximator property.

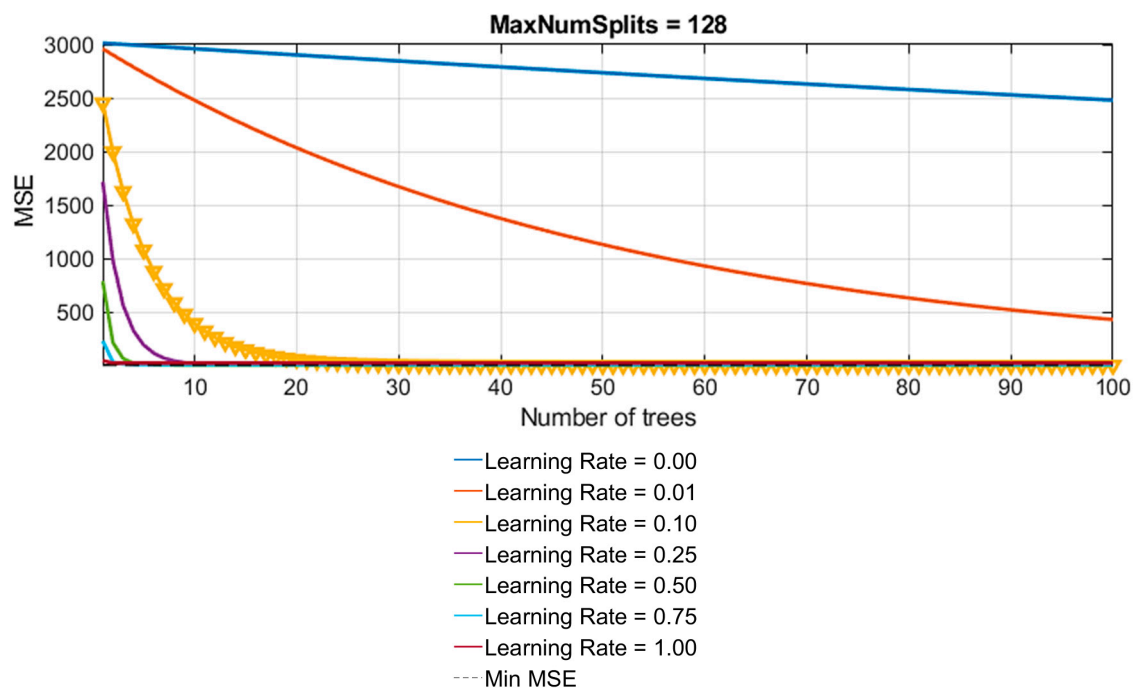


Figure 8. Relationship between Mean Squared Error (MSE) Values, Learning Rate, and the Number of Base Models in the Boosted Trees Method.

Table 4. Parameter settings during model calibration in the MATLAB program.

Parameter	The Value of the Parameter	
	Lower Limit	Upper Limit
Number of epochs	/	1000
The value of MSE (performance)	/	0
Gradient	/	1.00×10^{-7}
The value of the parameter λ_k (Mu)	0.005	1.00×10^{10}

In the specific context of predicting compressive strength, a tailored architecture was implemented, featuring seven neurons in the input layer and one in the output layer. The number of hidden layer neurons is determined experimentally. To offer guidance on establishing the upper limit for neurons in the hidden layer, the research provides recommended expressions (16) and (17), with a preference indicated for smaller values. These recommendations (24), (25) serve as valuable insights for the meticulous optimization of the neural network architecture, ensuring its efficacy in predicting compressive strength [37,42].

$$N_H \leq 2 \times N_i + 1 = 15, \quad (24)$$

$$N_H \leq \frac{N_s}{N_i + 1} = 113. \quad (25)$$

The experimental determination of the optimal number of neurons in the hidden layer constituted a key aspect of this study. Specifically, the analysis commenced with a structural configuration featuring one neuron, followed by a gradual increment in the number of neurons, each architecture systematically evaluated based on RMSE, MAE, R, and MAPE criteria (Figure 9). Model calibration, denoting the adjustment of model parameters or model training, was an integral step in this experimental procedure.

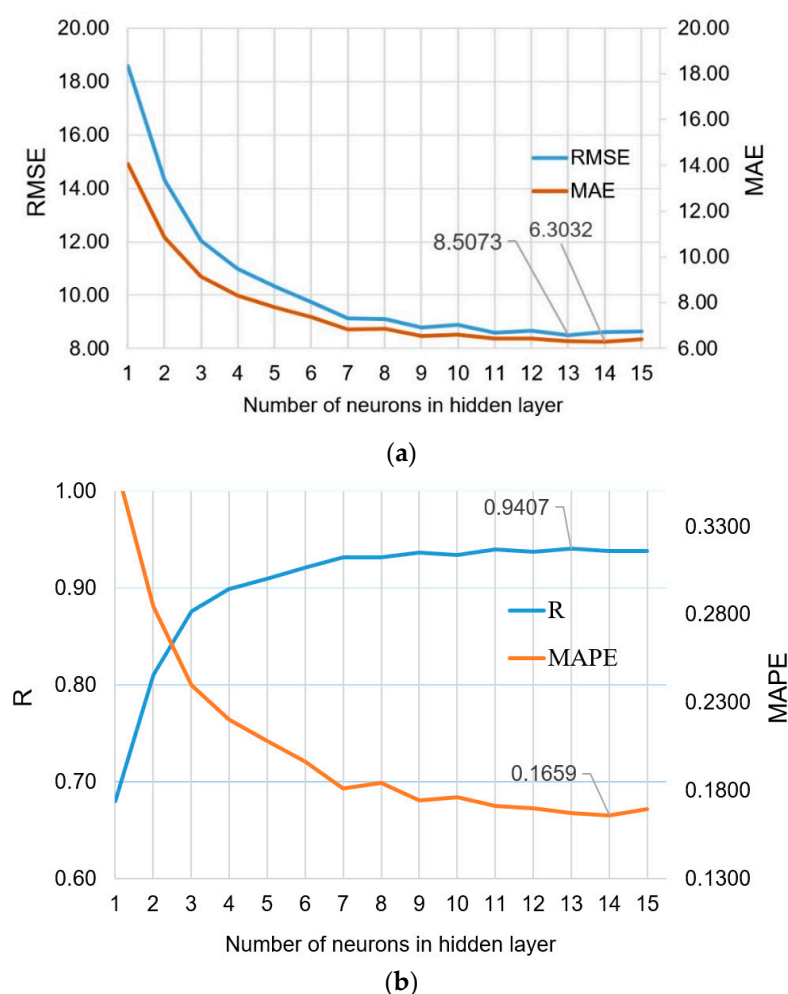


Figure 9. Comparative evaluation of performance metrics across various configurations of multilayer perceptron artificial neural networks (MLP-ANNs): (a) The comparison based on root mean squared error (RMSE) and mean absolute error (MAE), (b) The comparison based on the coefficient of correlation (R) and mean absolute percentage error (MAPE).

Uniformity across all variables was implemented by variable scaling. This precautionary measure stems from the recognition that the absolute size of a variable need not correspond directly to its actual influence. Within the scope of this paper, variables were transformed into the interval $[-1, 1]$. Here, the minimum value was standardized to -1 , the maximum value to 1 , and linear scaling was employed for values falling in between.

Throughout the model training process for all architectures, consistent standard settings (Table 4) were applied within the MATLAB program. This standardized approach ensured a reliable and consistent foundation for the comparative evaluation of different neural network structures. The optimal model for the MAE and MAPE criteria is a model with 14 neurons in the hidden layer and for the RMSE and R criteria a model with 13 neurons in the hidden layer (Figure 9).

In the subsequent phase of the investigation, the study delved into the application of ensemble methods with the primary objective of augmenting model generalization. This examination involved using base models within neural networks, spanning a spectrum of 1 to 15 neurons in their hidden layer (Figure 10). The RMSE value of the ensemble model is indicated by a red circle, while the RMSE values of individual neural networks are represented by blue circles (Figure 10a).

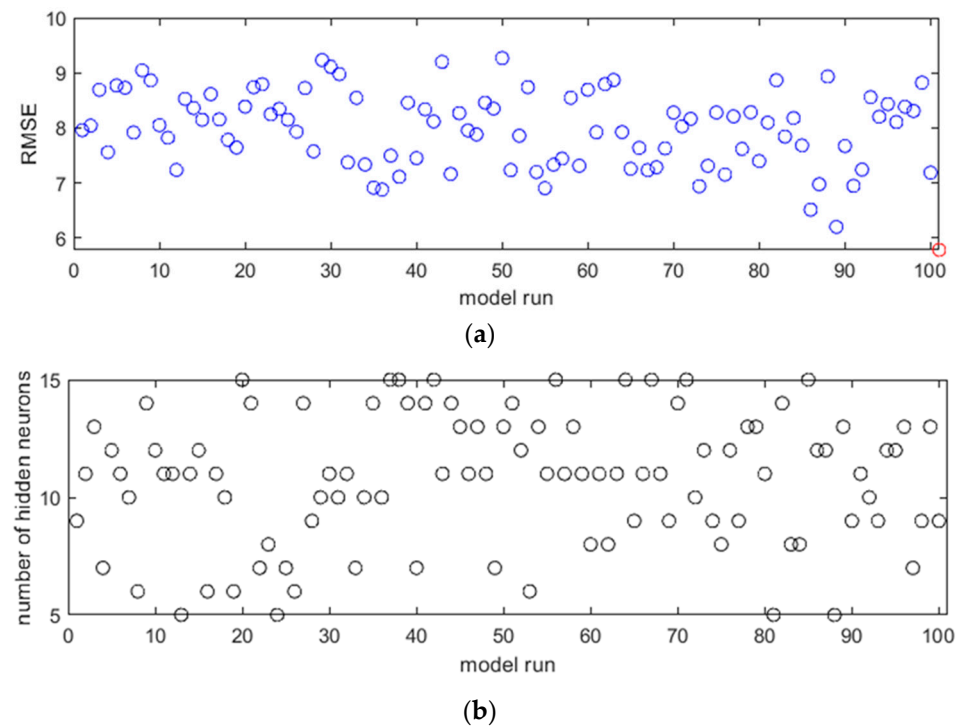


Figure 10. (a) The Root Mean Squared Error (RMSE) values for each iteration are presented alongside the corresponding architectural configurations, (b) the optimal number of neurons in the hidden layer for each iteration.

Notably, each of these base models had the flexibility to incorporate a certain number of neurons into its hidden layer. The dataset used to train the base models in each iteration was carefully formulated using the bootstrap method [38].

This method was used to systematically generate a sample of the same size as the original data set. In each iterative cycle, the base model was determined based on the minimum RMSE (Root Mean Squared Error) value among the 15 generated models (Figure 10). This iterative process continued until the cumulative number of training sets (M) was reached, culminating in the creation of M base models for the ensemble. The comparison between the target values, the ensemble prediction and the individual predictions of the neural network is shown in Figure 11.

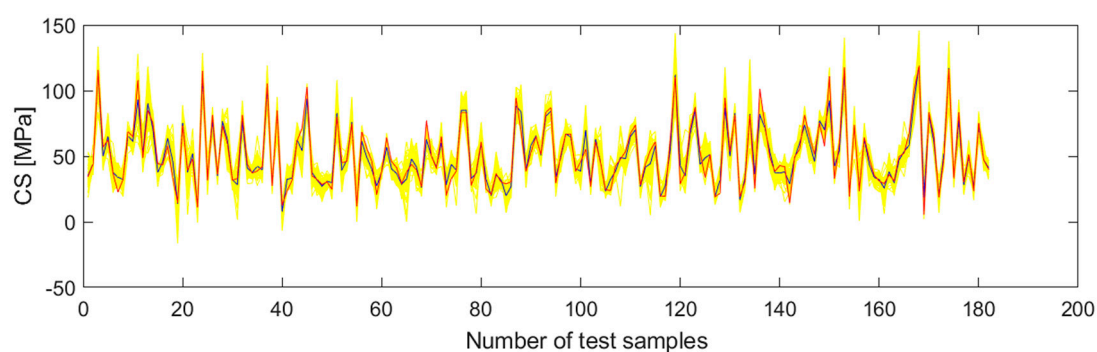


Figure 11. Comparison of target values (red color), ensemble modeled values (dark blue color) and modeled values of individual neural networks (yellow color).

In the following part, a sensitivity analysis was performed for the ensemble composed of individual neural networks. Sensitivity analysis helps assess the impact of small changes in input variables on the model's predictions. In the provided Figure 12, sensitivity analysis is conducted for each input variable in a neural network ensemble model. The mean

sensitivity results are visualized using a bar chart. Positive values suggest an increase in the output, while negative values suggest a decrease. The magnitude of the sensitivity values reflects the strength of the influence of each variable.

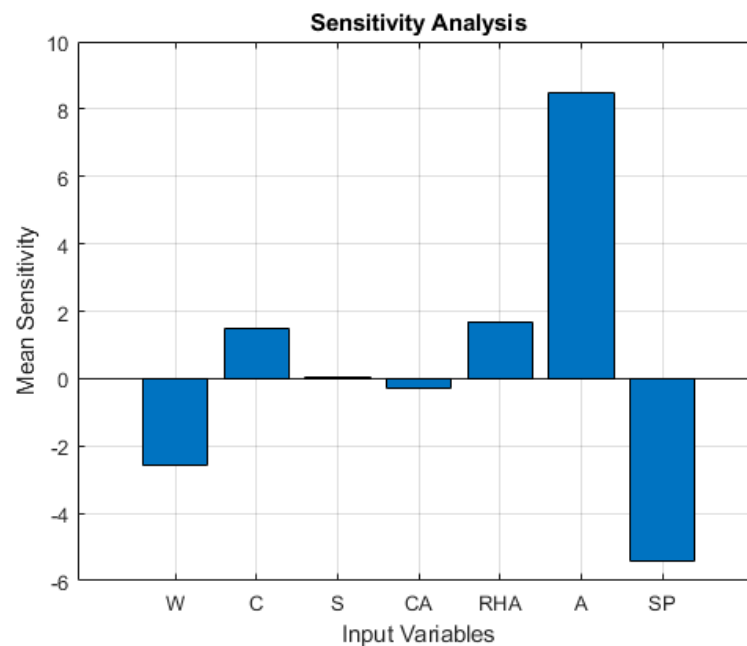


Figure 12. Sensitivity analysis for an ensemble of neural networks.

An examination of various covariance functions was conducted in the process of developing the Gaussian process regression model. These functions included those with a unified length scale parameter (Table 5) for all input variables (such as exponential, square-exponential, Matern 3/2, Matern 5/2 and the rational quadratic covariance function).

Table 5. Optimal parameter values within GPR models utilizing distinct covariance functions.

GP Model Covariance Function	Covariance Function Parameters		
Exponential	$k\left(\left(x_i, x_j \mid \Theta\right)\right)=\sigma_f^2 \exp \left[-\frac{1}{2} \frac{r}{\sigma_l^2}\right]$		
	$\sigma_l = 34.64$		$\sigma_f = 57.90$
Squared Exponential	$k\left(\left(x_i, x_j \mid \Theta\right)\right)=\sigma_f^2 \exp \left[-\frac{1}{2} \frac{\left(x_i-x_j\right)^T\left(x_i-x_j\right)}{\sigma_l^2}\right]$		
	$\sigma_l = 1.78$		$\sigma_f = 29.25$
Matern 3/2	$k\left(\left(x_i, x_j \mid \Theta\right)\right)=\sigma_f^2\left(1+\frac{\sqrt{3} r}{\sigma_l}\right) \exp \left[-\frac{\sqrt{3} r}{\sigma_l}\right]$		
	$\sigma_l = 4.27$		$\sigma_f = 46.81$
Matern 5/2	$k\left(\left(x_i, x_j \mid \Theta\right)\right)=\sigma_f^2\left(1+\frac{\sqrt{5} r}{\sigma_l}+\frac{5 r^2}{3 \sigma_l^2}\right) \exp \left[-\frac{\sqrt{5} r}{\sigma_l}\right]$		
	$\sigma_l = 3.68$		$\sigma_f = 76.43$
Rational Quadratic	$k\left(\left(x_i, x_j \mid \Theta\right)\right)=\sigma_f^2\left(1+\frac{r^2}{2 a \sigma_l^2}\right)^{-a} ; r=0$		
	$\sigma_l = 4.95$	$a = 0.15$	$\sigma_f = 121.35$

Where $r = \sqrt{\left(x_i - x_j\right)^T\left(x_i - x_j\right)}$.

Additionally, exploration extended to equivalent ARD covariance functions, each characterized by a distinct length scale for every input variable. A Z-score transformation

was applied to ensure uniformity in the data, resulting in a mean of zero and a variance of one. The analysis considered models with constant base functions (Table 6).

Table 6. Optimal parameter values within GPR ARD models utilizing distinct covariance functions.

Covariance Function Parameters						
σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7
ARD Exponential:						
$k((x_i, x_j \Theta)) = \sigma_f^2 \exp(-r); \sigma_F = 64.78; r = \sqrt{\sum_{m=1}^d \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}}$						
94.34	120.70	45.17	28.73	1619.22	11.91	147.58
ARD Squared exponential:						
$k((x_i, x_j \Theta)) = \sigma_f^2 \exp\left[-\frac{1}{2} \sum_{m=1}^d \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}\right]; \sigma_f = 30.60$						
1.25	0.98	1.41	0.87	4.60	0.37	3.10
ARD Matern 3/2:						
$k((x_i, x_j \Theta)) = \sigma_f^2 (1 + \sqrt{3}r) \exp[-\sqrt{3}r]; \sigma_f = 33.94$						
2.94	1.82	2.45	1.20	10.02	0.30	5.99
ARD Matern 5/2:						
$k((x_i, x_j \Theta)) = \sigma_f^2 \left(1 + \sqrt{5}r + \frac{5r^2}{3}\right) \exp[-\sqrt{5}r]; \sigma_f = 28.68$						
2.03	1.24	1.67	0.83	6.49	0.19	4.26
ARD Rational quadratic:						
$k((x_i, x_j \Theta)) = \sigma_f^2 \left(1 + \frac{1}{2\alpha} \sum_{m=1}^d \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}\right)^{-\alpha}; \alpha = 0.0228; \sigma_f = 94.38$						
5.36	3.21	4.27	2.17	19.30	0.54	11.01

Where $r = \sqrt{\sum_{m=1}^d \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}}$.

Based on three distinctive criteria RMSE, MAE, and R, the model employing the ARD rational quadratic covariance function emerges as the optimal selection, and ARD Matern 3/2 is optimal according MAPE criteria. As for the MAPE criterion, it secures second place in accuracy, with a marginal difference of 0.0037 compared to the leading model (ARD Rational quadratic) based on this specific criterion. The comparative analysis of all models in the research found that the BT model is optimal in terms of MAE and R criteria. In terms of the RMSE value, the neural network ensemble is the optimal model, while in terms of the MAPE value, the optimal model is the GP ARD matern 3/2 model (Table 6).

As the BT model is optimal regarding two criteria (MAE and R) of prediction accuracy, further analysis was performed only on this model. The procedure for calculating RMSE involves:

- squaring the difference between the target and forecast values,
- calculating the average, and
- subsequently determining the square root of this value.

Since the errors for RMSE are squared before averaging, this measure is particularly sensitive to significant extreme errors. While RMSE gives more weight to the forecast of extreme values of the samples, the MAE criterion gives equal weight to the prediction of all samples, so the BT model can be considered better because it has a better value of the MAE criterion and is therefore considered optimal (Table 7). Optimal values according specific criteria are marked in bold.

Table 7. Comparative analysis of machine learning models according to adopted accuracy criteria.

Model	RMSE	MAE	MAPE/100	R
Mult. linear regression	18.4949	14.4915	0.3806	0.7013
TreeBagger	7.4754	5.3440	0.1404	0.9605
Random Forest	7.6917	5.4204	0.1416	0.9584
Boosted Trees	6.2397	3.9140	0.1067	0.9710
SVM linear	18.3955	14.2280	0.3738	0.6998
SVM RBF	7.0863	4.9661	0.1328	0.9598
SVM Sigmoid	18.4378	14.2290	0.3768	0.6987
NN 8-13- 1	8.5073	6.3133	0.1669	0.9407
NN 8-14- 1	8.6289	6.3032	0.1659	0.9381
Ensamble	5.7858	4.1518	0.1098	0.9705
GP exponential	8.3622	6.0671	0.1713	0.9441
GP Sq.exponential	7.2756	5.4373	0.1492	0.9573
GP matern 3/2	7.3460	5.3685	0.1494	0.9565
GP matern 5/2	7.2130	5.2829	0.1470	0.9581
GP Rat. quadratic	7.2162	5.3151	0.1470	0.9580
GP ARD exponential	8.2035	5.2744	0.1183	0.9505
GP ARD Sq. exponential	8.3531	5.2521	0.1188	0.9483
GP ARD matern 3/2	9.3960	5.0817	0.1045	0.9351
GP ARD matern 5/2	6.9976	4.7886	0.1127	0.9652
GP ARD Rat. quadratic	6.7137	4.6157	0.1082	0.9681

In this regard, an analysis of the significance of individual input variables on the accuracy of the compression strength prediction was performed on BT model (Figure 13).

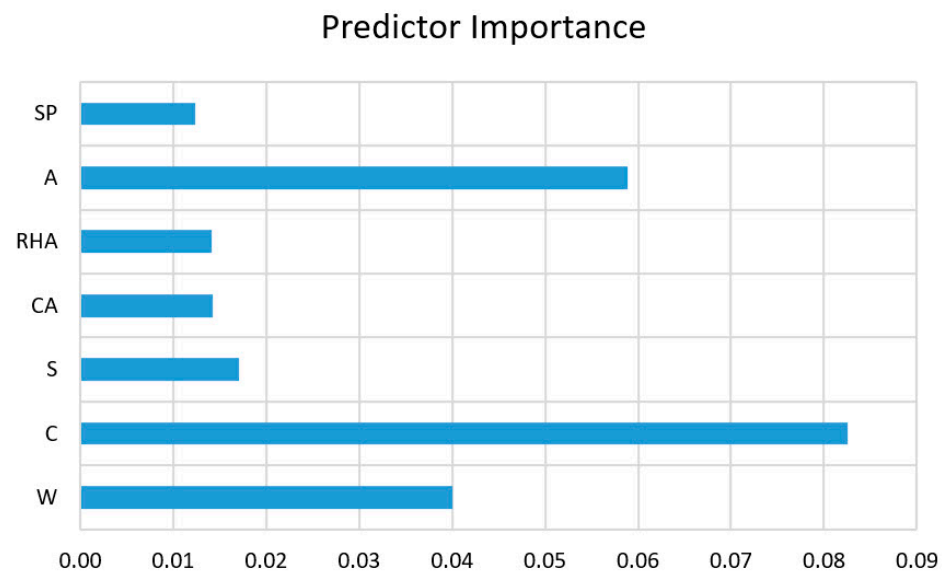
**Figure 13.** Predictor (input variable) importance for optimal Boosted trees model.

Figure 13 shows the dependence of the compressive strength of the concrete on the amount of rice husk ash added to the concrete in kg/m^3 , which was determined using the optimal BT model. The diagram was created so that all RHAC components had an average value from the database in Table 2 and only the RHA content varied from zero to $170 \text{ kg}/\text{m}^3$. The diagram (Figure 14) shows the positive effect of adding RHA to the concrete.

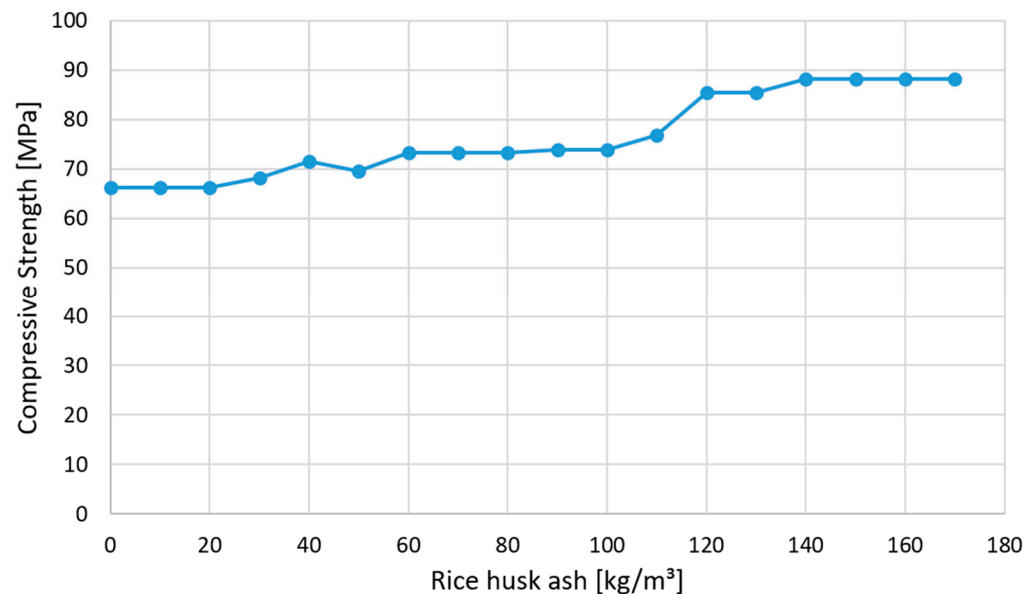


Figure 14. Dependence of compressive strength on RHA content.

The highest compressive strength is in the range of 100–140 kg/m³. Based on the analyzed data, an increase in the RHA content above these values practically does not lead to an increase in compressive strength. The maximum increase of CS concrete with RHA, when all other variables have an average value, is 33% based on the analyzed database (Figure 13).

The obtained results are compatible with a significant number of investigations. Research conducted by Kishore et al. and Ganesan et al. showed that concretes with the addition of RHA increase their compressive strength in the initial 3–7 days and after 28 days [14,15]. Kishore et al. found that this increase is about 30% compared to the concrete control group. This research showed that the prediction of the best BT model compared to the control group gives a maximum increase of 33% when the values of the other variables are equal to their mean values. Almost all researchers [16–18,36] singled out age, cement content, and water content as the most important input variables. At the same time, the increase in age and the amount of cement are more significantly positively correlated with CS, while the amount of water is more significantly negatively correlated with CS, which is by the conducted research [33], where this research confirmed the most significant importance of the age of concrete on the value of CS, while the content of cement and the amount of water is the second or third most important variable for compressive strength. The research also indicates a positive correlation of RHA with concrete compressive strength in the considered range of RHA amounts.

5. Conclusions

This study describes a set of advanced machine learning methods developed for predicting the compressive strength (CS) of RHA-enriched concrete. The study addresses the use of different models, including those based on regression trees, such as TreeBagger (TB), Random Forest (RF) and Boosted Trees (BT). In addition, the study includes Support Vector Machine (SVM) models with linear, RBF and sigmoid kernels, single Neural Network (NN) models and ensembles comprising single NN models. In addition, Gaussian Process Regression (GPR) models are investigated, each characterized by different kernel functions.

The paper describes a meticulous process for determining the optimal parameters for all models considered. It evaluates the accuracy of each model using predefined criteria: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Pearson's Correlation Coefficient (R). The study emphasizes

the importance of the individual input variables for model accuracy and highlights the positive influence of rice husk ash on concrete strength.

The Boosted Trees (BT) model identified as the optimal model based on the predefined criteria (RMSE, MAPE, MAPE and R) showed excellent performance with values of 6.24 MPa, 3.91 MPa, 10.67% and 0.97, respectively. Considering the ability of the model to predict RHAC for different concrete ages (1 to 365 days), the accuracy achieved can be accepted as satisfactory.

The work emphasizes the advantage of using ensembles, especially models with stacked trees. The results facilitate sector-specific modelling and enable efficient mixture composition with less time and financial expenditure.

The investigation uncovered that the correlation between compressive strength (CS) and Rice Husk Ash (RHA) content lacks a strongly expressed peak. Instead, it indicates a range of RHA levels linked to a more significant strength increase. This observation is likely due to the exclusion of RHA's factors like chemical composition, particle size, specific gravity, and other properties in the current research. Addressing these aspects in future studies has the potential to enhance precision and opens up new avenues for exploration.

Author Contributions: Conceptualization, M.K., M.H.-N., I.N.G., D.R. and S.L.; methodology and software, M.K.; validation, M.K., M.H.-N. and I.N.G.; formal analysis, M.K.; investigation, M.K., M.H.-N. and S.L.; resources, M.H.-N., I.N.G. and D.R.; data curation, M.K. and S.L.; writing—original draft preparation, M.K., M.H.-N., I.N.G. and S.L.; writing—review and editing, M.K. and M.H.-N.; funding acquisition, I.N.G. and D.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors are grateful for financial support of University North within the project “Trajnost cementnih kompozita”.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Thomas, B.S. Green concrete partially comprised of rice husk ash as a supplementary cementitious material—A comprehensive review. *Renew. Sustain. Energy Rev.* **2018**, *82*, 3913–3923. [CrossRef]
2. Sheheryar, M.; Rehan, R.; Nehdi, M.L. Estimating CO₂ emission savings from ultrahigh performance concrete: A system dynamics approach. *Materials* **2021**, *14*, 995. [CrossRef]
3. Suhendro, B. Toward green concrete for better sustainable environment. *Procedia Eng.* **2014**, *95*, 305–320. [CrossRef]
4. United States Department of Agriculture. Grain: World Markets and Trade Report. 2023. Available online: <https://fas.usda.gov/data/grain-world-markets-and-trade> (accessed on 14 December 2023).
5. Data Bridge Market Research Market Analysis Study. 2022. Available online: <https://www.databridgemarketresearch.com/reports/global-rice-husk-ash-market> (accessed on 1 June 2023).
6. Aprianti, E. A huge number of artificial waste material can be supplementary cementitious material (SCM) for concrete production—A review part, II. *J. Clean. Prod.* **2017**, *142*, 4178–4194. [CrossRef]
7. Boindala, S.P.; Ramagiri, K.K.; Alex, A.; Kar, A. Step-Wise Multiple Linear Regression Model Development for Shrinkage Strain Prediction of Alkali Activated Binder Concrete. In *Proceedings of SECON'19; Lecture Notes in Civil Engineering*; Dasgupta, K., Sajith, A., Unni Kartha, G., Joseph, A., Kavitha, P., Praseeda, K., Eds.; Springer: Cham, Germany, 2020; Volume 46. [CrossRef]
8. Ramagiri, K.K.; Boindala, S.P.; Zaid, M.; Kar, A. Random Forest-Based Algorithms for Prediction of Compressive Strength of Ambient-Cured AAB Concrete—A Comparison Study. In *Proceedings of SECON'21; Lecture Notes in Civil Engineering*; Marano, G.C., Ray Chaudhuri, S., Unni Kartha, G., Kavitha, P.E., Prasad, R., Achison, R.J., Eds.; Springer: Cham, Germany, 2020; Volume 171. [CrossRef]
9. Zhang, M.H.; Lastra, R.; Malhotra, V.M. Rice-husk ash paste and concrete: Some aspects of hydration and the microstructure of the interfacial zone between the aggregate and paste. *Cem. Concr. Res.* **1996**, *26*, 963–977. [CrossRef]
10. Sata, V.; Jaturapitakkul, C.; Kiattikomol, K. Influence of pozzolan from various by-product materials on mechanical properties of high-strength concrete. *Constr. Build. Mater.* **2007**, *21*, 1589–1598. [CrossRef]
11. Antiohos, S.; Tapali, J.; Zervaki, M.; Sousa-Coutinho, J.; Tsimas, S.; Papadakis, V. Low embodied energy cement containing untreated RHA: A strength development and durability study. *Constr. Build. Mater.* **2013**, *49*, 455–463. [CrossRef]

12. Prasittisopin, L.; Trejo, D. Hydration and phase formation of blended cementitious systems incorporating chemically transformed rice husk ash. *Cem. Concr. Compos.* **2015**, *59*, 100–106. [CrossRef]
13. Mehta, P.; Pirtz, D. Use of rice hull ash to reduce temperature in high-strength mass concrete. *J. Proc.* **1978**, *75*, 60–63. [CrossRef]
14. Kishore, R.; Bhikshma, V.; Prakash, P.J. Study on strength characteristics of high strength rice husk ash concrete. *Procedia Eng.* **2011**, *14*, 2666–2672. [CrossRef]
15. Ganesan, K.; Rajagopal, K.; Thangavel, K. Rice husk ash blended cement: Assessment of optimal level of replacement for strength and permeability properties of concrete. *Constr. Build. Mater.* **2008**, *22*, 1675–1683. [CrossRef]
16. Giaccio, G.; de Sensale, G.R.; Zerbino, R. Failure mechanism of normal and high-strength concrete with rice-husk ash. *Cem. Concr. Compos.* **2007**, *29*, 566–574. [CrossRef]
17. de Sensale, G.R. Strength development of concrete with rice-husk ash. *Cem. Concr. Compos.* **2006**, *28*, 158–160. [CrossRef]
18. Sam, J. Compressive strength of concrete using fly ash and rice husk ash: A review. *Civ. Eng. J.* **2020**, *6*, 1400–1410. [CrossRef]
19. Hwang, C.L.; Chandra, S. The use of rice husk ash in concrete. In *Waste Materials Used in Concrete Manufacturing*; Chandra, S., Ed.; William Andrew Publishing: Norwich, NY, USA, 1996; pp. 184–234, ISBN 9780815513933. [CrossRef]
20. Habeeb, G.A.; Mahmud, H.B. Study on properties of rice husk ash and its use as cement replacement material. *Mater. Res.* **2010**, *13*, 185–190. [CrossRef]
21. Fapohunda, C.; Akinbile, B.; Shittu, A. Structure and properties of mortar and concrete with rice husk ash as partial replacement of ordinary Portland cement—A review. *Int. J. Sustain. Built Env.* **2017**, *6*, 675–692. [CrossRef]
22. Nehdi, M.; Duquette, J.; Damatty, E.A. Performance of rice husk ash produced using a new technology as a mineral admixture in concrete. *Cem. Concr. Res.* **2003**, *33*, 1203–1210. [CrossRef]
23. Shatat, M.R. Hydration behavior and mechanical properties of blended cement containing various amounts of rice husk ash in presence of metakaolin. *Arab. J. Chem.* **2016**, *9*, S1869–S1874. [CrossRef]
24. Ahmed, A.E.; Adam, F. Indium incorporated silica from rice husk and its catalytic activity. *Microporous Mesoporous Mater.* **2007**, *103*, 284–295. [CrossRef]
25. Badorul, H.A.B.; Ramadhansyah, P.C.; Hamidi, A.A. Malaysian rice husk ash—Improving the durability and corrosion resistance of concrete: Pre-review. *EACEF-Int. Conf. Civ. Eng.* **2011**, *1*, 607–612. Available online: <https://proceeding.eacef.com/ojs/index.php/EACEF/article/view/428> (accessed on 5 November 2023).
26. Habeeb, G.A.; Fayyadh, M.M. The effect of RHA average particle size on the mechanical properties and drying shrinkage. *Aust. J. Basic. Appl. Sci.* **2009**, *3*, 1616–1622. Available online: <https://www.ajbasweb.com/old/ajbas/2009/1616-1622.pdf> (accessed on 4 November 2023).
27. Anwar, M.; Miyagawa, T.; Gaweesh, M. Using rice husk ash as a cement replacement material in concrete. *Waste Manag. Ser.* **2000**, *1*, 671–684. [CrossRef]
28. Nagrale, S.D.; Hemant, H.; Modak, P.R. Utilization of rice husk ash. *Int. J. Eng. Res. Appl.* **2012**, *2*, 1–5. Available online: https://www.ijera.com/papers/Vol2_issue4/A24001005.pdf (accessed on 5 November 2023).
29. Chindaprasirt, P.; Kanchanda, P.; Sathonsaowaphak, A.; Cao, H.T. Sulfate resistance of blended cements containing fly ash and rice husk ash. *Constr. Build. Mater.* **2007**, *21*, 1356–1361. [CrossRef]
30. Khan, K.; Ullah, M.F.; Shahzada, K.; Amin, M.N.; Bibi, T.; Wahab, N.; Aljaafari, A. Effective use of micro-silica extracted from rice husk ash for the production of high-performance and sustainable cement mortar. *Constr. Build. Mater.* **2020**, *258*, 119589. [CrossRef]
31. Iqtidar, A.; Bahadur Khan, N.; Kashif-ur-Rehman, S.; Faisal Javed, M.; Aslam, F.; Alyousef, R.; Alabduljabbar, H.; Mosavi, A. Prediction of Compressive Strength of Rice Husk Ash Concrete through Different Machine Learning Processes. *Crystals* **2021**, *11*, 352. [CrossRef]
32. Amin, M.N.; Iftikhar, B.; Khan, K.; Javed, M.F.; AbuArab, A.M.; Rehman, M.F. Prediction model for rice husk ash concrete using AI approach: Boosting and bagging algorithms. *Structures* **2023**, *50*, 745–757. [CrossRef]
33. Amlashi, A.T.; Golafshani, E.M.; Ebrahimi, S.A.; Behnood, A. Estimation of the compressive strength of green concretes containing rice husk ash: A comparison of different machine learning approaches. *Eur. J. Environ. Civ. Eng.* **2023**, *27*, 961–983. [CrossRef]
34. Bassi, A.; Manchanda, A.; Singh, R.; Patel, M. A comparative study of machine learning algorithms for the prediction of compressive strength of rice husk ash-based concrete. *Nat. Hazards* **2023**, *118*, 209–238. [CrossRef]
35. Li, C.; Mei, X.; Dias, D.; Cui, Z.; Zhou, J. Compressive Strength Prediction of Rice Husk Ash Concrete Using a Hybrid Artificial Neural Network Model. *Materials* **2023**, *16*, 3135. [CrossRef]
36. Amin, M.N.; Ahmad, W.; Khan, K.; Deifalla, A.F. Optimizing compressive strength prediction models for rice husk ash concrete with evolutionary machine intelligence techniques. *Case Stud. Constr. Mater.* **2023**, *18*, e02102. [CrossRef]
37. Kovačević, M.; Lozančić, S.; Nyarko, E.K.; Hadzima-Nyarko, M. Modeling of Compressive Strength of Self-Compacting Rubberized Concrete Using Machine Learning. *Materials* **2021**, *14*, 4346. [CrossRef] [PubMed]
38. Hastie, T.; Tibsiran, R.; Friedman, J. *The Elements of Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2009.
39. Breiman, L.; Friedman, H.; Olsen, R.; Stone, C.J. *Classification and Regression Trees*; Chapman and Hall/CRC: Wadsworth, OH, USA, 1984.
40. Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
41. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

42. Kovačević, M.; Ivanišević, N.; Petronijević, P.; Despotović, V. Construction cost estimation of reinforced and prestressed concrete bridges using machine learning. *Građevinar* **2021**, *73*, 727. [\[CrossRef\]](#)
43. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [\[CrossRef\]](#)
44. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [\[CrossRef\]](#)
45. Friedman, J.H.; Meulman, J.J. Multiple additive regression trees with application in epidemiology. *Stat. Med.* **2003**, *22*, 1365–1381. [\[CrossRef\]](#)
46. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.
47. Kecman, V. *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*; MIT Press: Cambridge, MA, USA, 2001.
48. Smola, A.J.; Sholkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [\[CrossRef\]](#)
49. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [\[CrossRef\]](#)
50. LIBSVM—A Library for Support Vector Machines. Available online: <https://www.csie.ntu.edu.tw/~cjlin/libsvm/> (accessed on 21 February 2021).
51. Rasmussen, C.E.; Williams, C.K. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2006.
52. Elwell, D.J.; Fu, G. *Compression Testing of Concrete: Cylinders vs. Cubes*; FHWA/NY/SR-95/119; New York State Department of Transportation: New York, NY, USA, 1995.
53. Ikpong, A.A.; Okpala, D.C. Strength characteristics of medium workability ordinary Portland cement-rice husk ash concrete. *Build. Environ.* **1992**, *27*, 105–111. [\[CrossRef\]](#)
54. Ismail, M.S.; Waliuddin, A.M. Effect of rice husk ash on high strength concrete. *Constr. Build. Mater.* **1996**, *10*, 521–526. [\[CrossRef\]](#)
55. Zhang, M.H.; Malhotra, V.M. High-performance concrete incorporating rice husk ash as a supplementary cementing material. *ACI Mater. J.* **1996**, *93*, 629–636. [\[CrossRef\]](#)
56. Feng, Q.G.; Lin, Q.Y.; Yu, Q.J.; Zhao, S.Y.; Yang, L.F.; Sugita, S. Concrete with highly active rice husk ash. *J. Wuhan. Univ. Technol.-Mater. Sci. Ed.* **2004**, *79*, 74–77. [\[CrossRef\]](#)
57. Bui, D.D.; Hu, J.; Stroeven, P. Particle size effect on the strength of rice husk ash blended gap-graded Portland cement concrete. *Cem. Concr. Compos.* **2005**, *27*, 357–366. [\[CrossRef\]](#)
58. Kartini, K.; Mahmud, H.; Hamidah, M. Strength properties of grade 30 rice husk ash. In Proceedings of the 31st Conference on Our World in Concrete & Structure, Singapore, 16–17 September 2006.
59. Sakr, K. Effects of silica fume and rice husk ash on the properties of heavy weight concrete. *J. Mater. Civ. Eng.* **2006**, *18*, 367–376. [\[CrossRef\]](#)
60. Saraswathy, V.; Song, H.W. Corrosion performance of rice husk ash blended concrete. *Constr. Build. Mater.* **2007**, *2*, 1779–1784. [\[CrossRef\]](#)
61. Ganesan, K. Studies on the Performance of Rice Husk Ash and Bagasse Ash Blended Concretes for Durability Properties. Ph.D. Thesis, Anna University, Chennai, India, 2007. Available online: <http://hdl.handle.net/10603/28700> (accessed on 5 November 2023).
62. Mahmud, H.B.; Malik, M.F.A.; Kahar, R.A.; Zain, M.F.M.; Raman, S.N. Mechanical properties and durability of normal and water reduced high strength grade 60 concrete containing rice husk ash. *J. Adv. Concr. Technol.* **2009**, *7*, 21–30. [\[CrossRef\]](#)
63. Mahmud, H.B.; Hamid, N.A.A.; Chin, K.Y. Production of high strength concrete incorporating an agricultural waste—Rice husk ash. In Proceedings of the 2nd International Conference on Chemical, Biological and Environmental Engineering, Cairo, Egypt, 2–4 November 2010; pp. 106–109. [\[CrossRef\]](#)
64. Madandoust, R.; Ranjbar, M.M.; Moghadam, H.A.; Mousavi, S.Y. Mechanical properties and durability assessment of rice husk ash concrete. *Biosyst. Eng.* **2011**, *110*, 144–152. [\[CrossRef\]](#)
65. Chao-Lung, H.; Anh-Tuan, B.L.; Chun-Tsun, C. Effect of rice husk ash on the strength and durability characteristics of concrete. *Constr. Build. Mater.* **2011**, *25*, 3768–3772. [\[CrossRef\]](#)
66. Muthadhi, A. Studies on production of reactive rice husk ash and performance of RHA concrete. In *Indian ETD Repository @ INFUBNET*; Pondicherry University: Puducherry, India, 2012.
67. Ramasamy, V. Compressive strength and durability properties of Rice Husk Ash concrete. *KSCE J. Civ. Eng.* **2012**, *16*, 93–102. [\[CrossRef\]](#)
68. Islam, M.N.; Mohd Zain, M.F.; Jamil, M. Prediction of strength and slump of rice husk ash incorporated high-performance concrete. *J. Civ. Eng. Manag.* **2012**, *18*, 310–317. [\[CrossRef\]](#)
69. Krishna, N.K.; Sandeep, S.; Mini, K.M. Study on concrete with partial replacement of cement by rice husk ash. *IOP Conf. Ser. Mater. Sci. Eng.* **2016**, *149*, 12109. [\[CrossRef\]](#)
70. Singh, P. To study strength characteristics of concrete with rice husk ash. *Indian. J. Sci. Technol.* **2016**, *9*, 1–5. [\[CrossRef\]](#)
71. Tandon, A.; Jawalkar, C.S. Improving strength of concrete through partial usage of rice husk ash. *Int. Res. J. Eng. Technol.* **2017**, *4*, 51–54. Available online: <https://www.irjet.net/archives/V4/i7/IRJET-V4I709.pdf> (accessed on 5 November 2023).
72. Siddika, A.; Mamun, M.A.A.; Ali, M.H. Study on concrete with rice husk ash. *Innov. Infrastruct. Solut.* **2018**, *3*, 18. [\[CrossRef\]](#)
73. He, Z.; Chang, J.; Liu, C.; Du, S.; Huang, M.A.N.; Chen, D. Compressive strengths of concrete containing rice husk ash without processing. *Rev. Romana Mater./Rom. J. Mater.* **2018**, *48*, 499–506. Available online: <https://solacolu.chim.upb.ro/p499-506.pdf> (accessed on 5 November 2023).

74. Singh, R.R.; Singh, D. Effect of rice husk ash on compressive strength of concrete. *Int. J. Struct. Civ. Eng. Res.* **2019**, *8*, 223–226. [[CrossRef](#)]
75. Kumar, P.C.; Rao, P.M. A study on reuse of Rice Husk Ash in Concrete. *Pollut. Res.* **2010**, *29*, 157–163. [[CrossRef](#)]
76. Nisar, N.; Bhat, J.A. Experimental investigation of Rice Husk Ash on compressive strength, carbonation and corrosion resistance of reinforced concrete. *Aust. J. Civ. Eng.* **2021**, *79*, 155–163. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.