

Article

Strength Estimation and Feature Interaction of Carbon Nanotubes-Modified Concrete Using Artificial Intelligence-Based Boosting Ensembles

Fei Zhu ^{1,2}, Xiangping Wu ^{3,*}, Yijun Lu ⁴ and Jiandong Huang ^{4,*}¹ School of Fine Arts, Suzhou Vocational University, Suzhou 215104, China; zhufeiedu@163.com² School of Civil Engineering, China University of Mining and Technology, Xuzhou 221116, China³ Department of Gem Design Engineering, KAYA University, Gimhae 50830, Republic of Korea⁴ School of Civil Engineering, Guangzhou University, Guangzhou 510006, China; yijun.lu@e.gzhu.edu.cn

* Correspondence: xiangping_wu@163.com (X.W.); jiandong.huang@hotmail.com (J.H.)

Abstract: The standard approach for testing ordinary concrete compressive strength (CS) is to cast samples and test them after different curing times. However, testing adds cost and time to projects, and, therefore, construction sites experience delays. Because carbon nanotubes (CNTs) vary in length, composition, diameter, and dispersion, experiment and formula fitting alone cannot reliably predict the strength of CNTs-based composites. For empirical equations or traditional statistical approaches to properly forecast complex materials' mechanical characteristics, various significant parameters, databases, and nonlinear relationships between variables must be considered. Machine learning (ML) tools are the most advanced for accurate predictions of material behaviour. This study employed gradient boosting, light gradient boosting machine, and extreme gradient boosting techniques to forecast the CS of CNTs-modified concrete. Also, in order to explore the influence and interaction of various features, an interaction analysis was conducted. In terms of R^2 , gradient boosting, light gradient boosting machine, and extreme gradient boosting models proved their accuracy. Extreme gradient boosting had the highest R^2 of 0.97, followed by light gradient boosting machine and gradient boosting with scores of 0.94 and 0.93, respectively. This type of research may help both academics and industry forecast material properties and influential elements, thereby reducing lab test requirements.

Keywords: carbon nanotubes; compressive strength; prediction models; interaction analysis; machine learning



Citation: Zhu, F.; Wu, X.; Lu, Y.; Huang, J. Strength Estimation and Feature Interaction of Carbon Nanotubes-Modified Concrete Using Artificial Intelligence-Based Boosting Ensembles. *Buildings* **2024**, *14*, 134. <https://doi.org/10.3390/buildings14010134>

Academic Editor: Grzegorz Ludwik Golewski

Received: 18 December 2023

Revised: 26 December 2023

Accepted: 29 December 2023

Published: 4 January 2024



Copyright: © 2024 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 order to get the best possible results, concrete technology has recently included nanoparticles [1]. The usage of these materials can improve the durability of many building components [2]. As a result of developments in nanotechnology, new materials are available for use in enhancing the mechanical characteristics and toughness of concrete, cement pastes, and other construction components [3]. The mechanical properties of nanomaterial-modified concrete have recently been investigated in experimental experiments. Researchers have used a variety of nanoscale materials, including carbon nanotubes (CNTs), micro silica, nanoscale aluminium, and micro-clay [4–6]. Concrete is one of the most widely used building materials in the world for construction purposes [7,8]. Crushed stone, freshwater, sand, and gravel are the non-renewable raw resources used the most by the building industry [9]. In addition, this industry consumes over 1.6 billion metric tonnes per year of Portland cement (PC) [10]. PC, a crucial ingredient in concrete, consumes considerable energy and is limited in supply. Additionally, a substantial amount of CO₂ is released into the environment by the cement industry [11]. Thus, the latter is responsible for around 7% of global CO₂ emissions [12]. Modern construction trends, such as the building

of modern bridges, high-rise buildings, and huge water accumulation systems, are also driving the rising demand for concrete [13,14]. Nonetheless, the appearance of nanoscale voids and cracks is a major disadvantage that lowers concrete's performance and reduces its lifespan [15]. Consequently, incorporating nanoparticles into a cementitious matrix increases mechanical strength and makes the material very resistant to cracking [16,17]. Researchers all around the globe are interested in nanotechnology because of its promising applications in a wide range of industries [18]. Particles of diameters between 1 and 100 nm are considered to be in this category [19]. Nanoparticles are employed in cementitious concrete to improve the material's conventional strengths.

CNTs' structural parameters, including their diameter characteristics, specific number of walls, cross-link density, and chirality, have significant effects on the cementitious matrix. Different properties of CNTs, such as fracture linkage, varied nucleation site, permeability, and structural change, allow them to enhance cementitious composites in various ways. To obtain optimal strengthening, it is necessary to understand how a given ingredient will affect the composite's qualities and what factors will determine that effect before using that substance as a supplement. A number of analyses have investigated how CNTs affect the mechanical properties of cementitious composites. Ghaderi [20] employed ternary materials in the formulation of concrete, including waste glass powder, basalt fibres, and CNTs, which are useful for making environmentally friendly buildings. Models for CNT concrete were provided by Lushnikova and Zaoui [21], and the effect of nanotube density on cementation nanocomposite was studied. The mechanical properties and piezoelectric material properties of concrete buildings have been studied recently [22,23]. Cost-effective monitoring of concrete structures may be possible with the use of electrochemical cement nanocomposites incapacitated with CNTs. The use of CNTs in concrete might lead to the development of an exciting new construction material. The aggregation propensity of nanoparticles has a similar profound impact on the shape performance of cementitious composites [24].

Producing a significant quantity of samples and evaluating them after varied lengths of curing time is the standard method for determining the compressive strength (CS) of ordinary concrete [25]. Until the results of the test are collected, normally after 28 days, construction activities on a building site should be halted. As a result, construction sites experience delays, and the testing processes adds extra cost and effort to the project. Testing specimens for distinct ingredient combinations with variable quantities at various ages is neither practicable nor sensible because of being time-consuming, costly, and reliant on real investigations [26–28]. It is difficult to accurately predict the mechanical characteristics of cementitious composites through experiment and formula fitting because of the varied features of CNTs, such as length, content, dispersion, and diameter. In addition, there are several important factors, databases, and nonlinear interactions between variables for empirical equations or typical statistical methodologies that need to be considered to reliably predict the mechanical properties of complex materials. Machine learning (ML) approaches have been used in recent studies to predict the properties of building materials, and the resulting models have shown promising results, outperforming the empirical formulae [29]. One of the most beneficial approaches to forecasting material performance, ML, has been successfully used for cement-based materials. ML approaches for predicting and analysing CNTs/cement-based composites are still in their infancy.

The primary objective of this work was to utilize ML techniques to estimate the CS of CNTs-modified concrete. Additionally, this study intends to assess the significance of input parameters by analysing a dataset obtained through laboratory testing. Three ML techniques, including gradient boosting (GB), light gradient boosting machine (LGBM), and extreme gradient boosting (XGB), were employed to achieve the study's aims. The objectives of this study are as follows: (i) to develop ML models using established Python codes in order to estimate the CS of the CNTs-modified concrete; (ii) to validate and compare the model results using statistical performance indicators and Taylor diagrams; (iii) to investigate the influence and interaction of input parameters using interaction

analysis. The evaluation and assessment of the correctness of each model's estimation was conducted through the utilization of statistical tests, which involved comparing the estimated results with the test results and calculating the coefficient of determination (R^2) and errors. Experimental-based research for the CS evaluation and mix design optimization of building materials requires significant effort, expenses, and time. This proposed research may aid both academics and industry in predicting the material attributes and influential factors, thus eliminating repeated test trials in the laboratory.

2. Research Methods

The research approach that was utilized in this investigation for the purpose of estimating the compressive strength of concrete containing CNTs involved making use of CNT data that was already available from previous literature sources. The information that was gathered was put through processing and analysis using ML methods so that a predictive model could be developed. Figure 1 is an illustration of the methodical approach that the research took, illustrating the successive processes that were carried out for this study. This methodology sought to improve our understanding of the influential factors affecting concrete strength when CNTs were incorporated by leveraging the power of both historical data and advanced techniques for machine learning. The end goal was to contribute to a more informed and efficient concrete composite design.

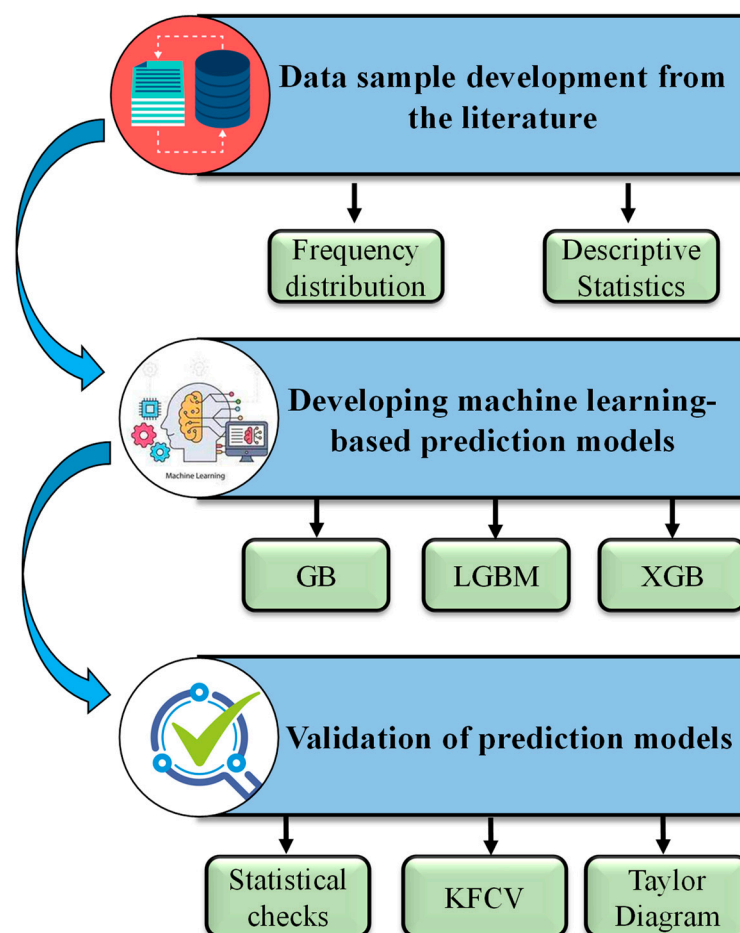


Figure 1. Flow chart of the study.

2.1. Data Description

The purpose of this study was to make a prediction regarding the compressive strength of concrete that contained CNTs. In order for ML algorithms to create the predicted output variable, they require a number of different input variables. Data were taken from

previously published works and included in this review for the purpose of predicting the CS of concrete containing CNTs [15]. In order to accurately forecast the compressive strength of concrete, six characteristics were used as inputs. These parameters were as follows: cement (kg/m^3), curing (days), fine aggregate (FA) (kg/m^3), coarse aggregate (CA) (kg/m^3), w/b, and nanomaterial CNTs (%). The parameters of aggregates, such as their fineness modulus, maximum diameters, and saturated surface dry (SSD), were either unchanging or varied only very slightly. As a result, these parameters were not included in this particular research. Since all of the data points acquired for this study adhered to the American Society for Testing and Materials (ASTM) criteria, it was presumed that the preparation of concrete was the same in each and every instance. In addition to this, the output of the model is substantially impacted by the quantity of data points as well as the numerous input parameters. This study made use of 282 different data points in order to make a prediction regarding the CNTs-based compressive strength of concrete. The dataset was split into training (70% of the data) and testing (30% of the data) for the modelling phase. This allocation was taken considering pertinent literature in the field [30–32]. All of the algorithms needed to run the models were crafted in the programming language Python, and the software that was employed was the Spyder (version 5.4.3) accessed through the Anaconda Navigator. Figure 2 gives a visual depiction of the relative frequency distribution of all parameters used in creating the models. Furthermore, for the database, Pearson’s correlation matrix was produced, as illustrated in Figure 3. A correlation matrix presents the correlation coefficients between pairs of variables in the form of a symmetrical matrix. The intensity and direction of the linear relationship between two variables are quantified by correlation coefficients. The correlation matrix is a frequently employed instrument in the fields of statistics and multivariate analysis for examining the connections between numerous variables [33]. Potential results comprise a perfect negative correlation (denoted by -1), a perfect positive correlation (represented by $+1$), and the absence of any correlation (represented by 0). A positive correlation denotes that when one variable increases, the other variable also increases proportionally; conversely, a negative correlation suggests that when one variable increases, it generally results in a decrease in the other [34]. The most significant input was the cement quantity, which exhibited a strong positive correlation (as indicated by a correlation coefficient of 0.59) with the output (CS). The correlation between the output and CT was also positive, while the impact of w/b, CA, FA, and CNTs was shown to be negative. Statistics used for describing the data are listed in Table 1.

Table 1. Statistical description of the CNTs dataset.

Name	Abbreviation	Total Data	Mean	Standard Deviation	Sum	Minimum	Kurtosis	Skewness	Range	Median	Maximum
Curing time (d)	CT	282	45.30	32.69	12,776	1	3.2	1.5	179.0	28	180
Cement (kg/m^3)	C	282	397.82	45.40	112,186.6	250	-0.3	-0.1	225.0	400	475
Water to binder ratio	w/b	282	0.50	0.08	140.62	0.4	5.1	1.9	0.5	0.49	0.87
Coarse aggregate (kg/m^3)	CA	282	1031.25	163.87	290,813.5	498	1.7	-0.9	968.8	1068.75	1466.8
Fine aggregate (kg/m^3)	FA	282	638.64	163.44	180,096.4	175.5	3.2	1.2	1109.5	608.375	1285
Carbon nanotubes (%)	CNT	282	0.51	1.89	145.21	0	18.4	4.4	10.0	0	10
Compressive strength (MPa)	CS	282	45.14	10.57	12,730.26	14.7	0.6	-0.9	52.0	46.5	66.7

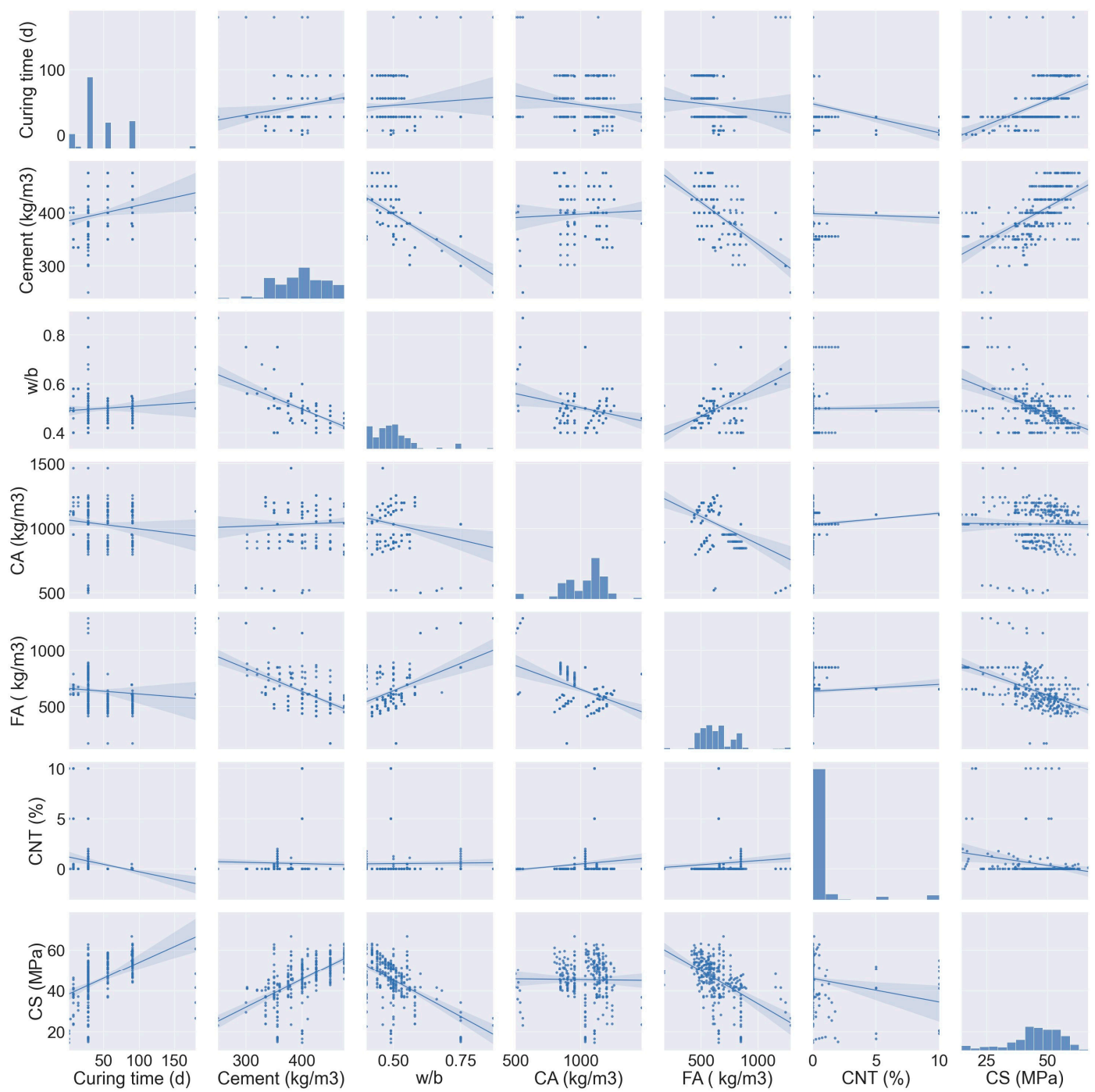


Figure 2. Frequency description of the CNTs dataset.

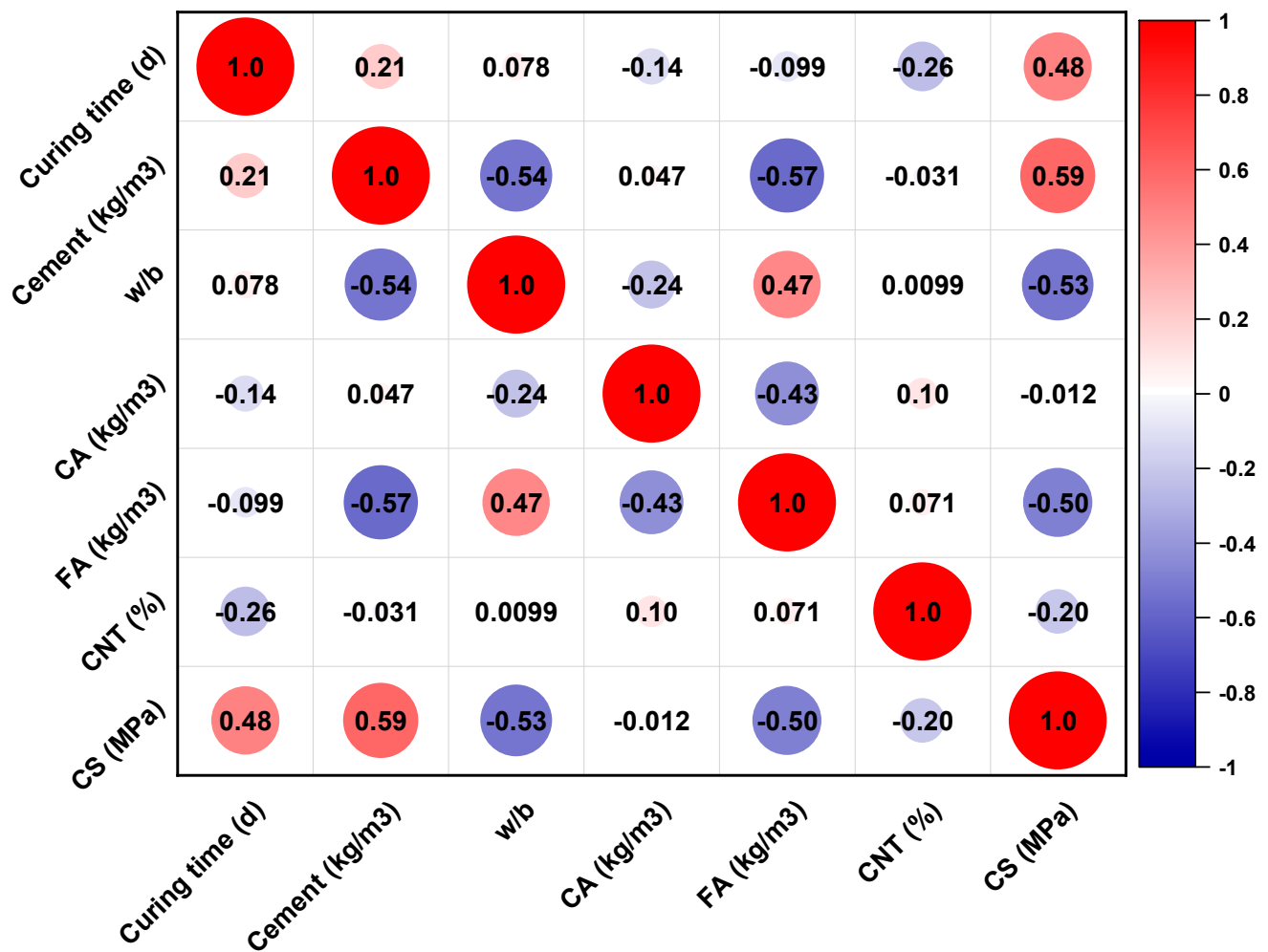


Figure 3. Correlation coefficient of the CNTs dataset.

2.2. Machine Learning Algorithms

2.2.1. Gradient Boosting

In 1999, Friedman [35] introduced GB to the scientific community as an ensemble method for regression and classification. The GB method can only be used for regression. Figure 4 exhibits how the GB approach conducts a comparison of each iteration of the arbitrarily chosen training set to the initial model. Randomly subsampling the data used for training is one way to assist in minimizing overfitting while also increasing the execution accuracy of GB. This can also help speed up the execution of GB. The smaller the percentage of data used for training, the more rapid the regression will be since the model needs to adapt to the new minor data with each and every iteration. The GB method needs tuning parameters, such as n-trees and shrinkage rate. The n-trees parameter represents the count of trees to be produced and it should not be set too low. In addition, the shrinkage rate, which is typically known as the rate of learning, being utilized to all trees in the creation, should not be set too high [36]. Tuning is required because the GB algorithm is used to generate trees.

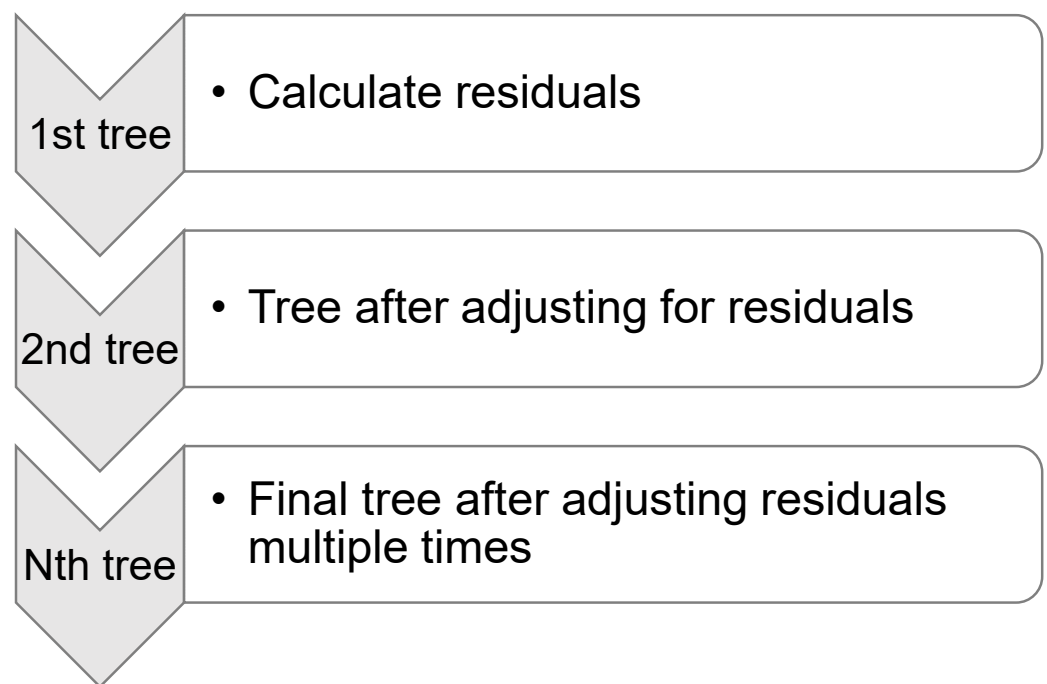


Figure 4. The gradient boosting method is depicted in a schematic form.

2.2.2. Light Gradient Boosting Machine

LGBM is a relatively new gradient learning framework that utilizes decision trees and the concept of boosting [37]. Histogram-based techniques are used to accelerate training, decrease memory use, and use a leaf-wise growth approach with depth limitations, setting it apart from the XGB model. Histogram algorithms work by dividing the space occupied by the eigenvalues of a continuous floating-point variable into a number of discrete bins, say k . The memory requirements of the histogram approach can be lowered by a factor of eight by not having to keep track of intermediate results that have already been sorted, and by saving the value after the discretization of features, which is often small enough to store with an 8-bit integer. The model's precision is unaffected by the approximate partitioning. Overfitting may be successfully avoided due to the regularization impact of the coarser segmentation points. Decision trees utilize a level-wise growth approach, which is an inefficient method since it treats the leaves of the same layer. This wastes a lot of memory. By comparing all of the leaves and selecting the ones with the largest branching gain, the leaf-by-leaf approach is more efficient. As a result, the blade's vertical orientation allows for more error reduction and greater accuracy with the same number of iterations as the horizontal orientation. Overfitting and deeper decision trees are two potential drawbacks of leaf orientation. LGBM prevents inefficient overfitting by placing a depth restriction on the leaf's upper surface. Figure 5 is a simplified representation of the level- and leaf-specific development methods that trees employ [38].

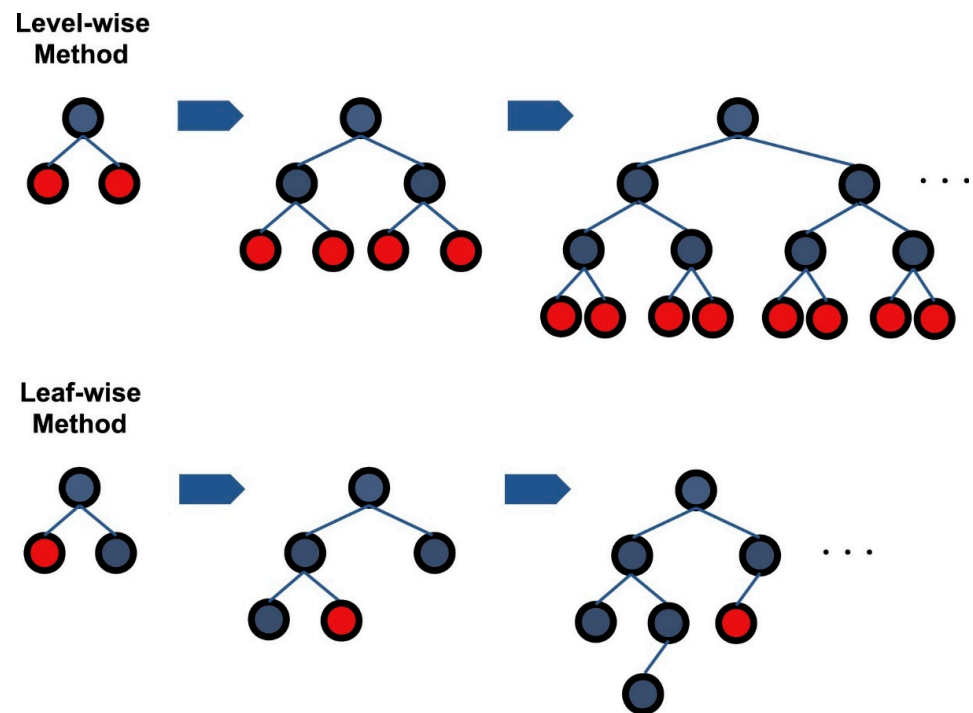


Figure 5. An illustration of the LGBM method [38].

2.2.3. Extreme Gradient Boosting

The XGB approach for machine learning is based on a decision tree, and it uses the gradient descent method to decrease the amount of data that is shed as a result of the addition of new models. According to Shehadeh et al. [39], the decision-tree-based algorithms that are used for machine learning on medium- to small-scale tabular data are deemed to be the most effective among the comparable machine-learning algorithms. XGB is nothing more than a haphazard assortment of CARTs. The CARTs are divided, and the places at which they are most effectively divided are determined to be the minimal objective function. According to Nguyen et al. [40], XGB may be utilized to solve regression problems. XGB leaf node weight computations are guided by the following desired function:

$$Obj = \sum_{j=1}^T \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T$$

where Obj is the objective function, w_j is the leaf j weight vector, G_j is the sum of the derivatives that are partial of the samples at the j th leaf node (defined at the first order), λ is the coefficient for L2 regularisation, H_j is the cumulative sum of the partial derivations of the data points at leaf node j of the second order, and γT regulates the tree's degree of intricacy.

2.3. Model Assessment and Validation Methods

A mathematical technique known as k-fold cross-validation (KFCV) is used to evaluate the effectiveness of the models that are applied to data in order to prevent overfitting and bias in the training set. It then takes the remaining set ($k10-1$) and uses it to train the model [41]. This method partitions the whole data set into $k10$ subsets of data, with one set serving as the testing set ($k1$) out of ten. It was hypothesized by Kohavi in 1995 [42] that KFCV would produce a precise variance and be more appropriate for optimum calculation time. The $K10$ subset is used in this study to perform an analysis and validation of all models, as shown in Figure 6.

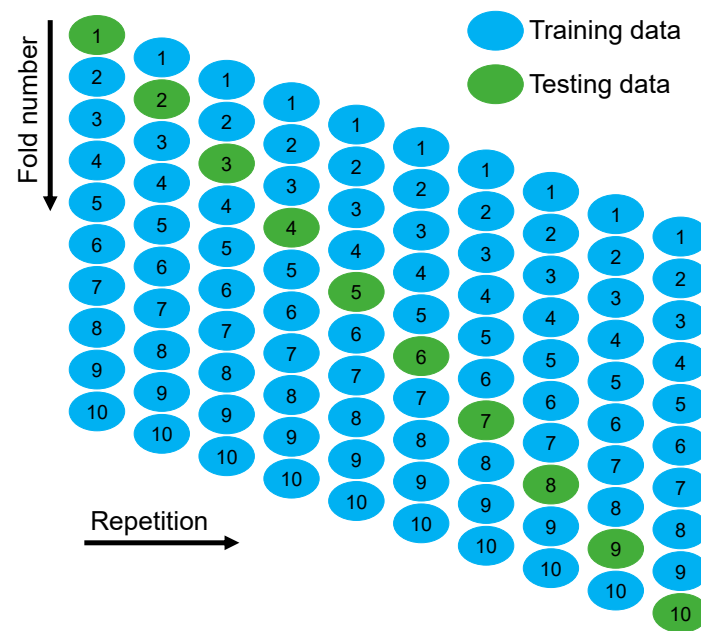


Figure 6. KFCV involves a test and a training set.

Moreover, statistical errors including the mean absolute error root (MAE), root mean square logarithmic error (RMSLE), root mean square error (RMSE), and R^2 , were utilized when assessing a model's performance on a training or testing set. R^2 , also known as the determination coefficient, is a statistic that may be used to measure how well a model can predict future outcomes [43,44]. The advancements that have been made in AI modelling techniques have made it possible to provide more accurate forecasts of the mechanical characteristics of concrete. The GB, LGBM, and XGB models are statistically evaluated and contrasted with one another in this research by means of the determination of error criteria. There are a large number of data points that might perhaps shed light on the inaccuracy of the model. In order to determine whether or not the model is reliable and valid, the coefficient of determination can be utilized. The findings that are produced by models with R^2 values that are above 0.50 are disappointing, whereas the results that are produced by models with R^2 values that lie within the range of 0.65 and 0.75 are encouraging. The value of R^2 may be found by utilizing Equation (1). The MAE uses the same unit system for both its input and its output. A model with an MAE that falls within a certain range has the potential to produce mistakes of a significant nature on occasion. Equation (2) is what we use to compute MAE. The RMSE is the measure of how accurate estimates and measurements are. Error squared is determined by adding up all of the individual error squares. The new method gives more consideration to extreme circumstances than the older computations did, which results in squared differences that are larger in some cases but lower in others. The RMSE may be estimated using Equation (3). The model's ability to reliably anticipate incoming data increases in proportion to the reduction in the value that represents the RMSE. The root-mean-square error, or RMSE, is a useful metric for contrasting models of differing degrees of complexity. Incorporating a larger logarithmic error into the RMSLE enables it to compute the proportional difference between the outcome that was predicted and the one that was actually seen. Because the log conversion presents the intended distribution in a very straightforward manner, it is helpful for analysing results that are right-skewed. The RMSLE can be computed with Equation (4).

$$R^2 = 1 - \frac{\sum_{j=1}^m (p_j - t_j)^2}{\sum_{j=1}^m (t_j - \bar{t})^2} \quad (1)$$

$$\text{MAE} = \frac{\sum_{j=1}^m |t_j - p_j|}{n} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^m (t_j - p_j)^2}{n}} \quad (3)$$

$$\text{RMSLE} = \sqrt{\frac{1}{n} \sum_{j=1}^m (\log(t_j + 1) - \log(p_j + 1))^2} \quad (4)$$

In these equations, t_j represents the original experimental data used to create the model, p_j represents the anticipated result, \bar{t}_j represents the target average value, \bar{p}_i represents the projected mean value, and m is the total number of instances considered.

3. Results and Discussions

3.1. Gradient Boosting Model

The GB model has a remarkable predictive performance, displaying an R^2 value of 0.93, which indicates its capacity to explain a significant percentage of the variation included within the dataset. This demonstrates the model's ability to forecast the future accurately. The linear regression fit of R^2 is also shown in Figure 7. As can be seen in Figure 8, a comprehensive examination of prediction errors has been carried out in order to evaluate the accuracy of the model in terms of providing estimates of actual values. The findings indicate that the bulk of the predictions are astonishingly near to the actual values, with 47.1% of them having an error of less than 1.5 MPa. This is demonstrated by the fact that there is a remarkable correlation between the two. In addition, 35.2% of the model's predictions fall within the range of 1.5 to 4 MPa, which demonstrates the model's reliability in providing correct forecasts throughout a wider spectrum. The fact that just 17.7% of the forecasts are off by more than 4 MPa from the values that really occurred demonstrates the robustness of the model in its ability to deal with difficult circumstances. When looking at the error distribution, we can see that the biggest forecast error is 8.2 MPa, while the smallest error is 0.1 MPa. This shows that the model is able to handle a diverse set of data points. The fact that the model has an error rate of 2.19 MPa on average underlines the robust prediction skills it possesses over a wide range of data sets.

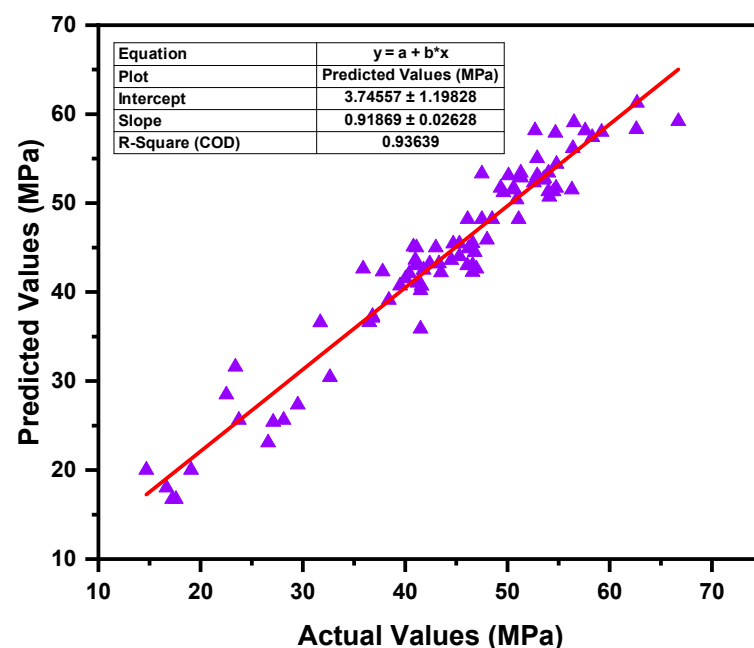


Figure 7. Actual vs. predicted values for the GB model.

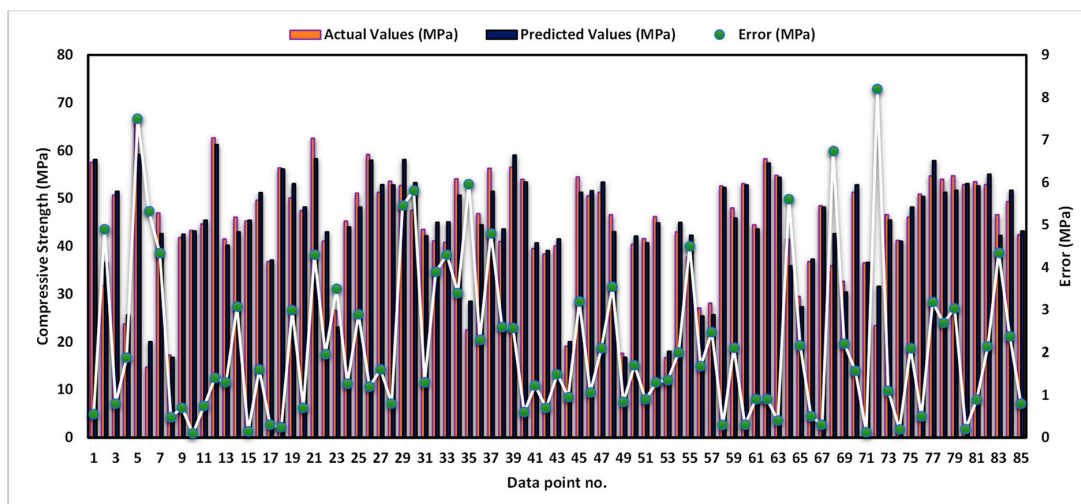


Figure 8. Error distribution for the GB model.

3.2. Light Gradient Boosting Machine Model

Figure 9 illustrates that the LGBM model has an impressively high R^2 value of 0.94, which suggests that it has an exceptional level of predictive accuracy. This score is a reflection of how well the model is able to explain a sizeable percentage of the variability present in the data. In addition, a comprehensive error analysis has been carried out in order to assess how well the model performs in terms of its ability to forecast real values, as can be seen in Figure 10. According to the findings, the bulk of the forecasts are quite close to the actual values, with 52.9% of them having an error that is lower than 1.5 MPa. In addition, 29.4% of forecasts fall within the range of 1.5 to 4 MPa, which demonstrates the model's constant accuracy over a larger range of values. The fact that just 17.7% of forecasts are off by more than 4 MPa from the actual values demonstrates how resilient the model is when it comes to dealing with severe circumstances. The dependability of the model is shown by the error distribution, which reveals that the model's biggest forecast error measures 7.33 MPa, while the model's lowest mistake measures just 0.05 MPa. The model has high prediction skills for a variety of data points, as seen by its moderate error of 1.99 MPa, which is calculated as an average over all of the data points.

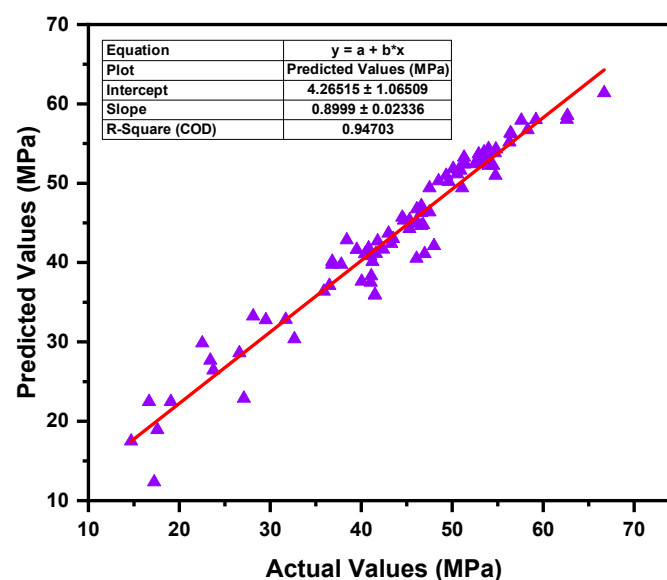


Figure 9. Actual vs. predicted values for the LGBM model.

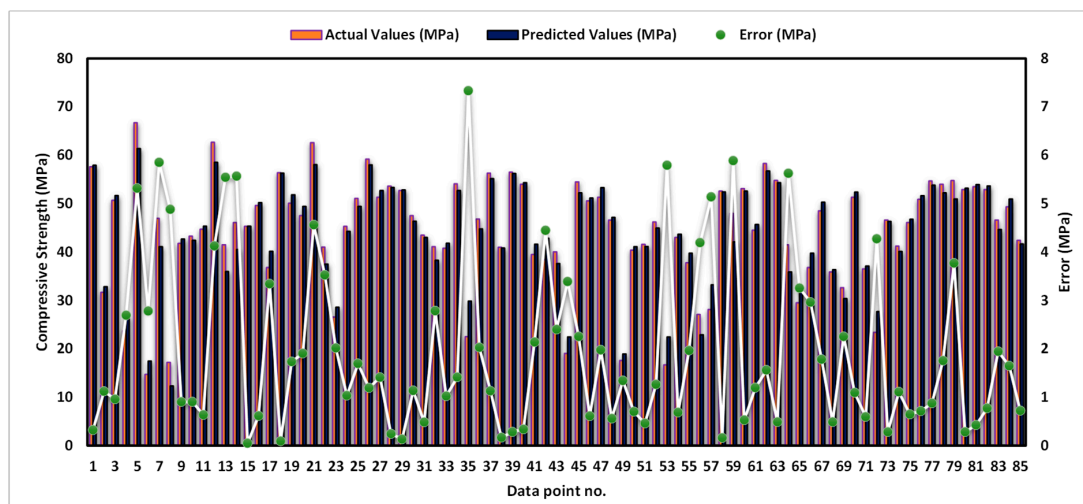


Figure 10. Error distribution for the LGBM model.

3.3. Extreme Gradient Boosting Model

The XGB regression model demonstrates exceptional predictive capability, as seen by its remarkable R^2 value of 0.97. This indicates that the model is very successful in explaining approximately 97% of the variation in the data, showing its resilience in capturing underlying patterns, as illustrated in Figure 11. In addition, as can be shown in Figure 12, an in-depth investigation of the error distribution demonstrates that the model has an excellent level of accuracy. Its precision in forecasting values within a tight margin is demonstrated by the fact that 61.1% of the forecasts have errors that are less than 1.5 MPa. This is a major fraction of the predictions. In addition, 36.5% of forecasts come within the range of 1.5 MPa to 4 MPa, which demonstrates its adaptability in managing a wider range of value ranges. Even when errors are present that are more than 4 MPa, the model still maintains an acceptable degree of performance, with just 2.4% of predictions falling into this category. The reliability of the model's forecasts is shown by the fact that its biggest mistake was calculated to be 5.5 MPa, while its lowest error was calculated to be a meagre 0.005 MPa. The model has been shown to have an inaccuracy of 1.44 MPa on average, which further demonstrates the outstanding precision and dependability with which it can estimate target values.

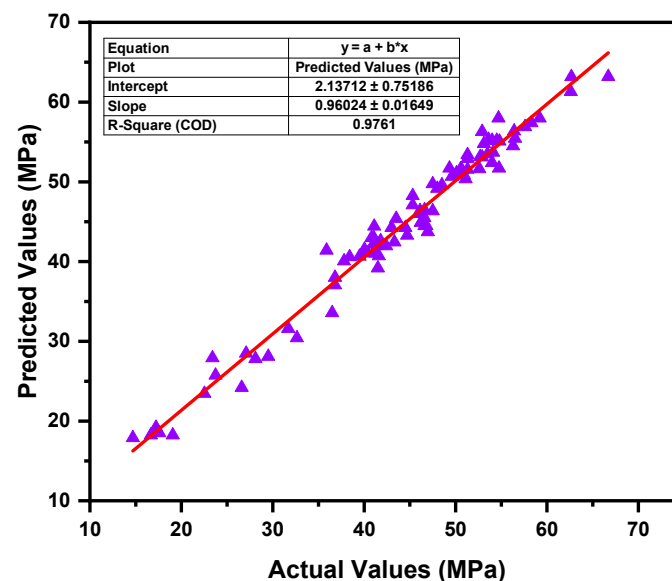


Figure 11. Actual vs. predicted values for the XGB model.

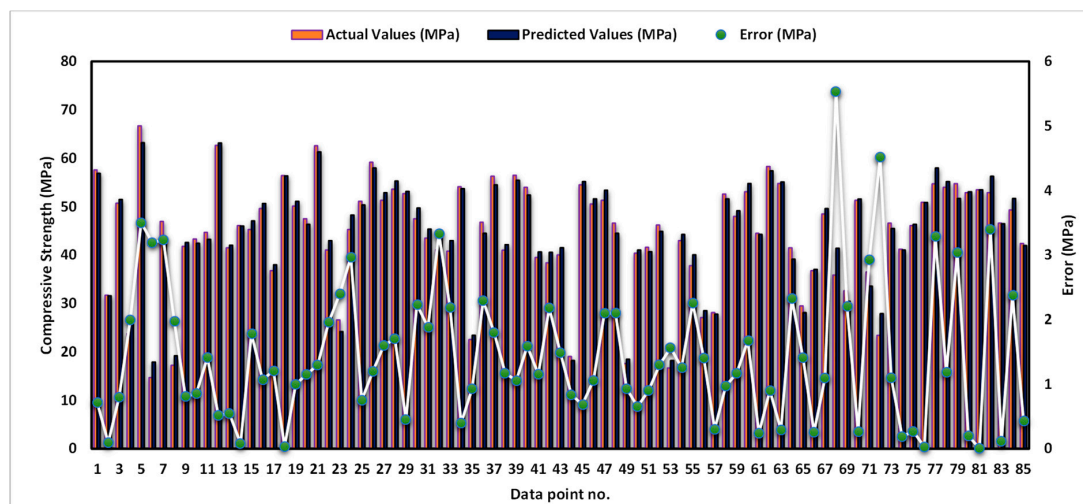


Figure 12. Error distribution for the XGB model.

3.4. K-Fold Cross-Validation Outcomes

KFCV was used to conduct an in-depth analysis of the performance of three ML models, namely GB, LGBM, and XGB. As can be seen in Table 2 and Figure 13, the findings demonstrated outstanding performance across the board for all of the assessment criteria. XGB came out on top as the best performer, attaining the greatest R^2 value of 0.981 on a constant basis, which indicates that it has a remarkable capacity to explain data variation. In addition, it had the lowest MAE of 3.499 MPa, which indicates that it is extremely accurate, and the lowest RMSE of 2.605 MPa, which demonstrates that it can make accurate predictions. In addition, XGB showed the lowest RMSLE values, with 0.071 MPa being the lowest possible value. This demonstrates how well it can handle data on a variety of scales. Despite the fact that both GB and LGBM performed wonderfully, with impressive R^2 values of 0.935 and 0.940, respectively, as well as competitive MAE, RMSE, and RMSLE scores, XGB continually beat both, making it the best option for this specific CNTs dataset.

Table 2. KFCV of all models.

K-Fold	GB Model				LGBM Model				XGB Model			
	R^2	MAE (MPa)	RMSE (MPa)	RMSLE (MPa)	R^2	MAE (MPa)	RMSE (MPa)	RMSLE (MPa)	R^2	MAE (MPa)	RMSE (MPa)	RMSLE (MPa)
1	0.82	3.65	2.08	0.036	0.71	2.47	2.57	0.068	0.88	3.33	2.52	0.033
2	0.92	4.02	4.60	0.072	0.91	3.95	3.89	0.017	0.94	3.25	3.78	0.012
3	0.83	2.43	2.82	0.042	0.89	3.28	3.21	0.014	0.87	3.01	1.04	0.035
4	0.89	3.29	3.80	0.044	0.87	3.19	3.02	0.022	0.96	3.50	3.71	0.029
5	0.91	2.27	2.96	0.078	0.94	2.73	1.16	0.071	0.95	2.49	2.52	0.071
6	0.83	3.09	3.23	0.017	0.93	2.14	3.96	0.027	0.90	3.36	2.13	0.015
7	0.68	3.86	3.95	0.036	0.83	3.02	3.68	0.069	0.98	1.17	3.75	0.034
8	0.85	1.53	2.46	0.047	0.86	2.31	3.11	0.078	0.88	1.31	1.65	0.012
9	0.88	3.17	3.53	0.092	0.86	3.03	1.74	0.036	0.94	2.03	2.38	0.060
10	0.94	1.36	3.84	0.057	0.85	1.74	4.42	0.090	0.93	1.41	2.57	0.009
Max	0.94	4.02	4.60	0.092	0.94	3.95	4.42	0.090	0.98	3.50	3.78	0.071
Min	0.68	1.36	2.08	0.017	0.71	1.74	1.16	0.014	0.87	1.17	1.04	0.009
Mean	0.85	2.87	3.33	0.052	0.87	2.79	3.08	0.049	0.92	2.48	2.60	0.031

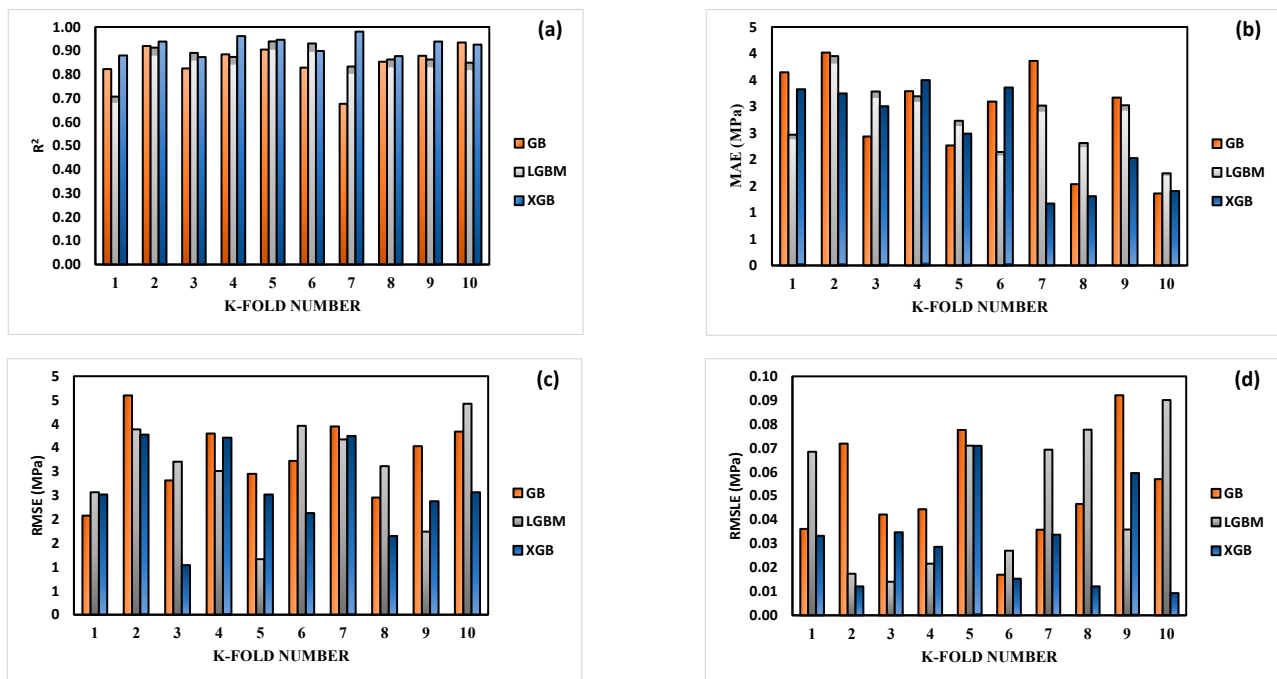


Figure 13. Statistics pertaining to KFCV for each model: (a) R^2 , (b) MAE, (c) RMSE, and (d) RMSLE.

3.5. Statistical Performance Indicators

The statistical analysis conducted on the three independent ML models, GB, LGBM, and XGB, yielded valuable performance data. The XGB model beat the others with a significantly lower MAE of 1.445 MPa, suggesting its higher accuracy in predicting target values. In addition, the XGBoost model showed a substantially reduced mean absolute percentage error (MAPE) at 3.80%, which indicates that it is able to deliver more exact predictions while taking into consideration the magnitude of the data. In addition, the XGBoost model had the lowest RMSE of all of the models tested, coming in at 1.798 MPa. This demonstrates how good the model is in reducing prediction errors across the board. In conclusion, the RMSLE provided more evidence that the model is superior by demonstrating that it attained the lowest value out of the three models (0.041 MPa), indicating that it is able to make predictions that are scaled and accurate on a consistent basis. Table 3 provides an in-depth analysis of the statistical examinations that were performed. In light of all of these findings, it is clear that the XGB model has far more potential for accurate prediction than either GB or LGBM.

Table 3. Statistical checks for the GB, LGBM, and XGB models.

Technique	GB	LGBM	XGB
MAE (MPa)	2.195	2.0	1.445
MAPE	5.70%	5.70%	3.80%
RMSE (MPa)	2.863	2.844	1.798
RMSLE (MPa)	0.065	0.064	0.041

3.6. Taylor Diagrams

A Taylor diagram is a graphical tool that is used to graphically compare and analyse the performance of several models or datasets, notably in the context of data analysis and predictive modelling. It is also known as a “Taylor plot”. It makes it possible to evaluate different models or datasets on the basis of their correlation, RMSE, and standard deviation, which provides insights into the correctness and dependability of the models and datasets. In the field of machine learning, this graphic is especially helpful for model assessment

and selection. Therefore, Taylor diagrams were used to graphically evaluate the results of three different machine learning algorithms (the GB model, the LGBM model, and the XGB model) in the assessment of prediction models for compressive strength in CNTs-based concrete. Figure 14 displays the forecasts made by the models: GB had a value of 9.572, LGBM had a value of 9.576, and XGB was the clear winner with a prediction of 9.579. In particular, XGB's predictions for compressive strength in CNTs-based concrete were very close to the actual values, demonstrating its better accuracy. This result demonstrates why XGB has the potential to be the go-to model for concrete strength prediction, where accuracy and precision are of the utmost importance.

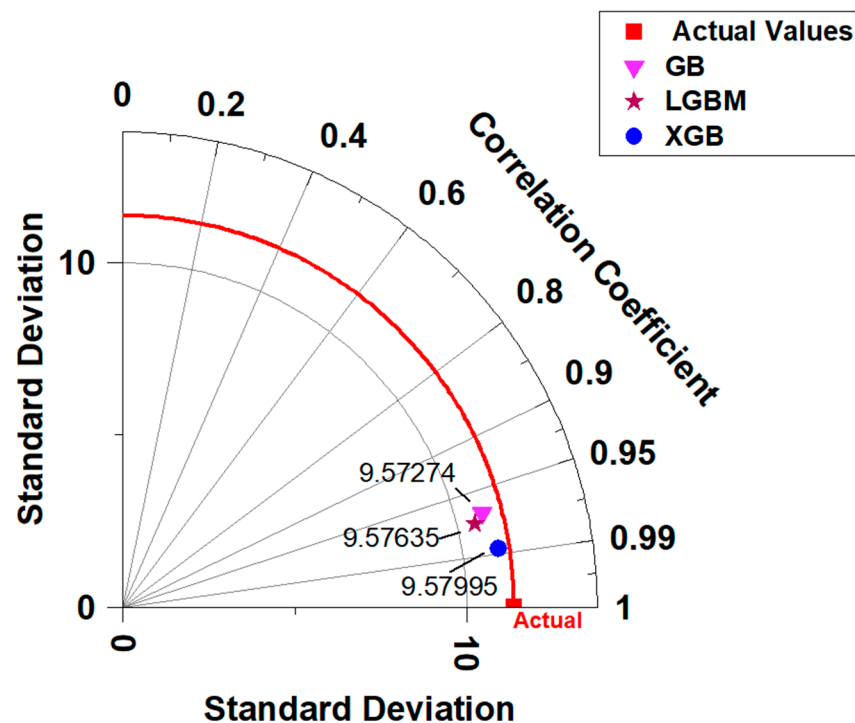


Figure 14. Taylor diagram.

3.7. Interaction Analysis Outcomes

This section investigates the relationships between input factors, specifically raw materials, and the resulting outputs, i.e., CS. Scatter plots were produced in order to illustrate the correlation between different inputs and the CS of CNTs-based concrete, as seen in Figure 15. The bar graphs positioned beside the scatter plot depict the frequency of input or output events. As depicted in Figure 15a, CT exhibited increasing strength with rising CT, which suggests that at later ages of CNTs-based concrete, its strength improves. Similar observations were also noted in prior experimental studies [45]. The interaction of cement with CS was found to be similar to the CT. In order to achieve higher strength, the cement quantity in the mix needs to be kept higher, as displayed in Figure 15b. The impact pattern of w/b, as shown in Figure 15c, demonstrates that lower w/b is favourable for attaining the desirable strength. The interaction of CA and FA, as shown in Figure 15d,e, showed that with rising quantities of CA and FA, strength declines. Figure 15f shows the interaction of CNTs, which implies that a lower percentage of CNTs is feasible for achieving the maximum CS. The optimal percentage of CNTs in the mix might be around 1%. The results of the present study were derived from the input kinds and dataset size employed in the interaction analysis. There is potential for attaining more accurate correlations by expanding the range of input variables inside the database, which is suggested for future research.

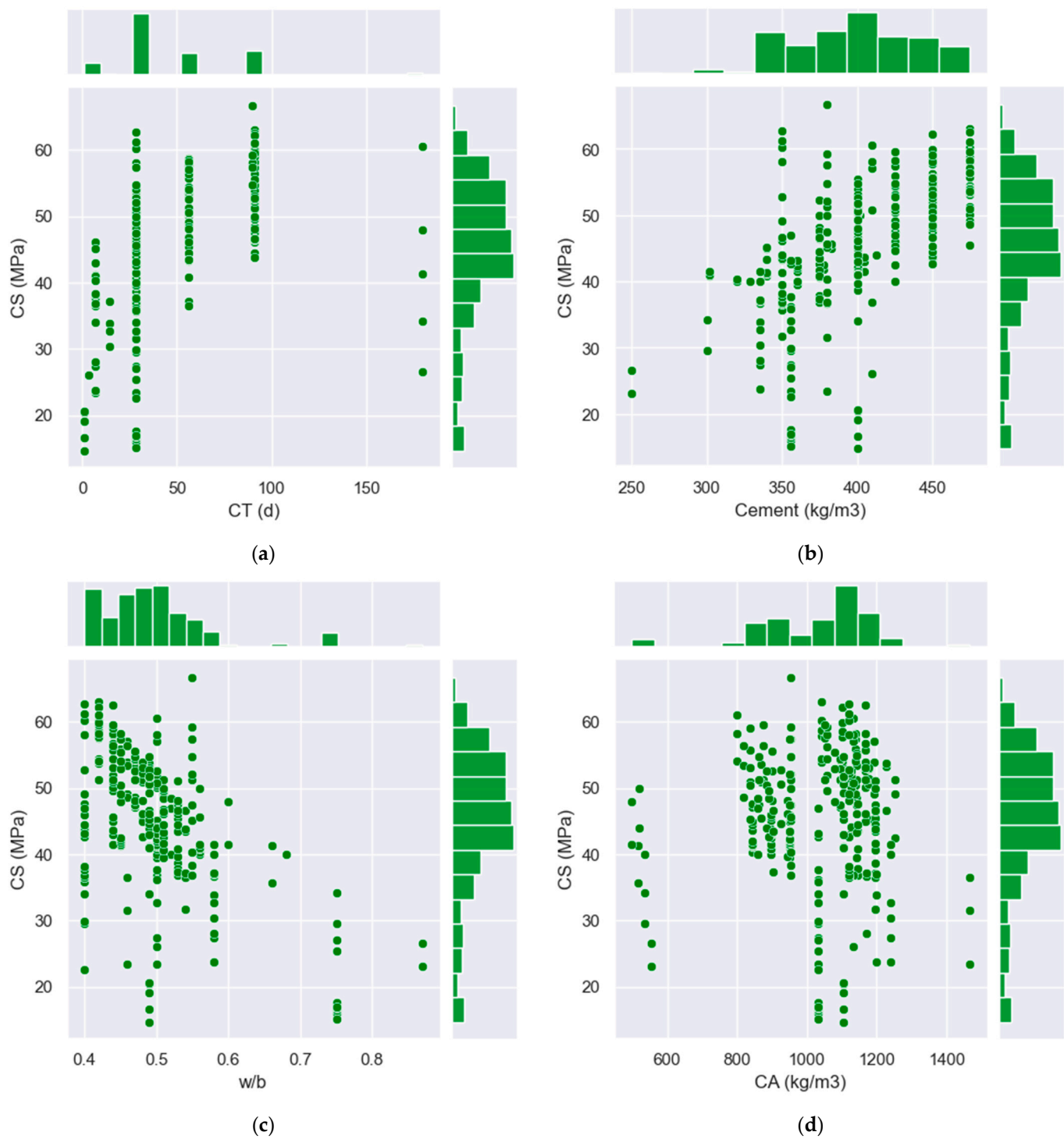


Figure 15. Cont.

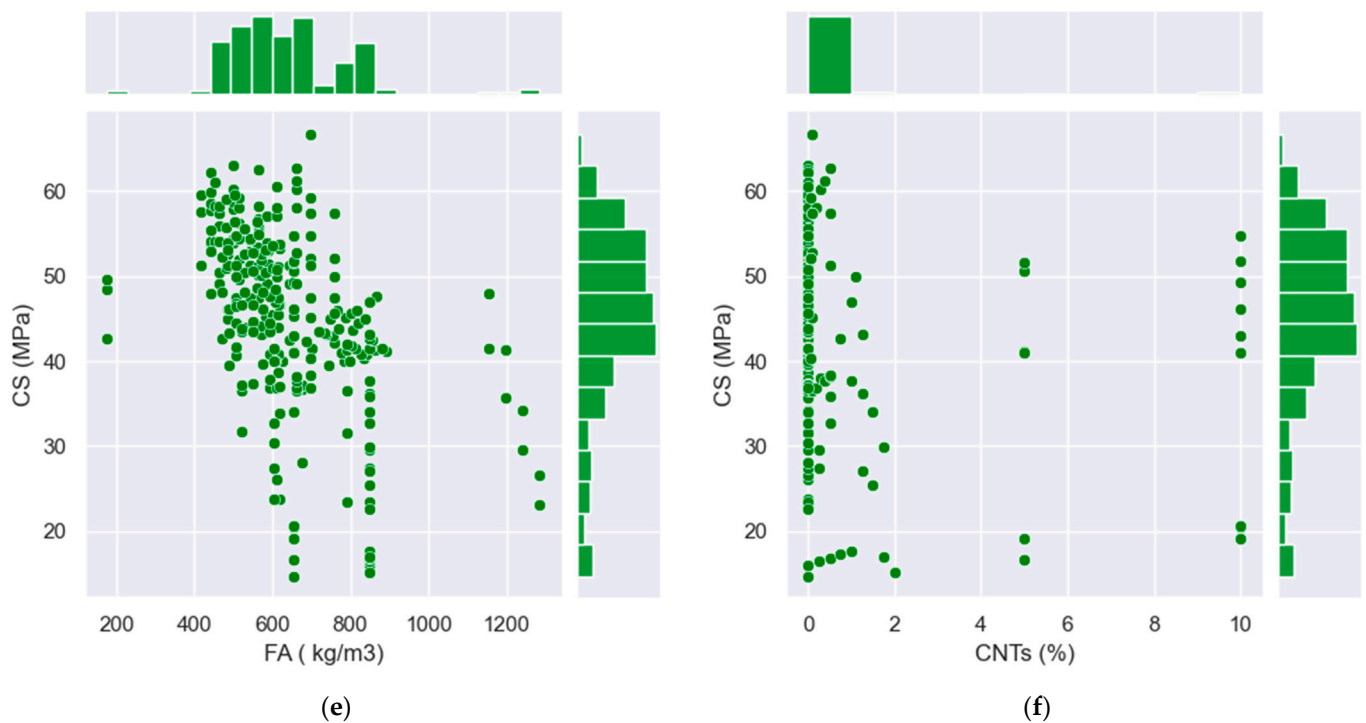


Figure 15. Interaction plots for the input features.

4. Discussions

This research focused on developing and comparing ML models such as GB, LGBM, and XGB for estimating the CS of CNTs-based concrete. The developed models exhibited satisfactory prediction performance. An important aspect of ML models derived from this work is their restriction to a predefined set of eight input parameters. Consequently, the forecasts will be tailored only to CNTs-based concrete. The accuracy of the models' strength forecasts is ensured by their utilization of identical testing procedures and unit measurements. The meticulously derived equations generated by the simulations significantly enhance comprehension of the mix design and the impact of each input parameter. Expanding the parameters of the composite analysis beyond the given eight inputs might potentially invalidate the usefulness of the predicted models. If these models are provided with data that do not align with their original design, they may not function as intended. The models may produce erroneous predictions if the input parameters' units are adjusted or inconsistent. The efficacy of the models relies on maintaining uniform unit sizes. ML models have several applications in the construction sector, including predicting material strength, ensuring quality, assessing risks, forecasting maintenance needs, and improving energy efficiency. Nevertheless, these models have several constraints, including the requirement for human involvement, flawed data, and unreliable models. Subsequent studies may concentrate on overcoming these constraints and enhancing ML solutions through various means, such as leveraging the Internet of Things (IoT), developing hybrid models, implementing explainable AI techniques, taking sustainability factors into account, and adapting data generation and distribution to industry requirements. The previous research done on construction materials using similar methods has been documented in Table 4. These technological advancements have the capacity to improve effectiveness, clarity, comprehensibility, and informed decision-making in the construction industry. All of these factors would lead to a reduction in project delays and an improvement in safety and sustainability. The results of this study might potentially enhance the adoption of CNTs-based concrete in the construction sector, hence fostering the implementation of sustainable building practices.

Table 4. Results comparison with past ML-based studies.

Ref.	Material Studied	Properties Predicted	ML Techniques Employed	Optimum ML Model Noted	R ² of the Best ML Model
Current study	CNTs-based concrete	CS	GB, LGBM, and XGB	XGB	0.97
[32]	Eggshell-based cement mortar	Reduction in CS after acid attack	GB, AdaBoost, and XGB	XGB	0.94
[46]	Steel fibre reinforced concrete	Flexural strength	GB, random forest, and XGB	XGB	0.94
[47]	Geopolymer concrete	CS	Support vector machine, GB, and XGB	XGB	0.98
[48]	Nano-silica modified concrete	Ultrasonic pulse velocity	GB, AdaBoost, and XGB	XGB	0.90

5. Conclusions

This study aimed to use ML algorithms to predict the CS of CNTs-based concrete. A comprehensive dataset of six inputs and 282 points was employed with CS as an output. The GB, LGBM, and XGB models were utilized in the process of result forecasting. Python code was utilized in Spyder (version 5.4.3) to build models. The models' predictability performance was evaluated using statistical checks (R², MAE, RMSLE, MAPE, and RMSE). This study led to the following conclusions:

- In terms of the coefficient of determination (R²), each of the three ML models, GB, LGBM, and XGB, presented convincing evidence of their accuracy. XGB surpassed the other models with the greatest R² of 0.97, while GB and LGBM also attained R² of 0.93 and 0.94, respectively.
- When compared to LGBM and GB, XGB had a low error distribution, and its average error was just 1.44 MPa. This indicated that XGB's forecasts were more accurate and had lower variability than other models' predictions.
- During KFCV, the performance of XGB was consistently superior to that of GB and LGBM since it produced lower values for MAE, RMSE, and RMSLE. The excellence of XGB's predictions is further shown through measurements. Similar results are also shown using statistical checks.
- The findings were verified using the Taylor diagram, which demonstrated that the values predicted by XGB were closer to the actual values than those predicted by GB and LGBM. This also demonstrated that XGB was accurate in predicting the CS of CNTs-based concrete.
- The interaction pattern indicated that at higher values of CT and cement, strength improves. To attain maximum strength, w/b, CA, and FA need to be kept low, while CNTs' percentage needs to be kept around 1%.

This study determined the suitability of ensemble ML methods in predicting the CS of CNTs-modified concrete. Additionally, interaction analysis was conducted to examine the influence and interaction of various features on CS. Using ML tools might eliminate repeated test trials by accurately predicting the desired outcomes. Interaction analysis examines the influence of various features, which might be used in formulating the mix proportions of CNTs-modified concrete.

Author Contributions: Conceptualization, F.Z. and J.H.; methodology, X.W. and Y.L.; software, F.Z.; validation, X.W. and J.H.; formal analysis, Y.L.; investigation, Y.L.; resources, J.H.; data curation, F.Z.; writing—original draft preparation, F.Z.; writing—review and editing, X.W., Y.L. and J.H.; visualization, Y.L.; supervision, X.W. and J.H.; project administration, X.W.; funding acquisition, X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

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

References

- Hassan, A.; Galal, S.; Hassan, A.; Salman, A. Utilization of carbon nanotubes and steel fibers to improve the mechanical properties of concrete pavement. *Beni-Suef Univ. J. Basic Appl. Sci.* **2022**, *11*, 121. [\[CrossRef\]](#)
- Faraj, R.H.; Mohammed, A.A.; Omer, K.M. Self-compacting concrete composites modified with nanoparticles: A comprehensive review, analysis and modeling. *J. Build. Eng.* **2022**, *50*, 104170. [\[CrossRef\]](#)
- Paruthi, S.; Husain, A.; Alam, P.; Husain Khan, A.; Abul Hasan, M.; Magbool, H.M. A review on material mix proportion and strength influence parameters of geopolymer concrete: Application of ANN model for GPC strength prediction. *Constr. Build. Mater.* **2022**, *356*, 129253. [\[CrossRef\]](#)
- Hawreen, A.; Bogas, J.A. Creep, shrinkage and mechanical properties of concrete reinforced with different types of carbon nanotubes. *Constr. Build. Mater.* **2019**, *198*, 70–81. [\[CrossRef\]](#)
- Mohsen, M.O.; Al Ansari, M.S.; Taha, R.; Al Nuaimi, N.; Taqa, A.A. Carbon nanotube effect on the ductility, flexural strength, and permeability of concrete. *J. Nanomater.* **2019**, *2019*, 6490984. [\[CrossRef\]](#)
- Shekari, A.H.; Razzaghi, M.S. Influence of Nano Particles on Durability and Mechanical Properties of High Performance Concrete. *Procedia Eng.* **2011**, *14*, 3036–3041. [\[CrossRef\]](#)
- Lao, J.-C.; Xu, L.-Y.; Huang, B.-T.; Zhu, J.-X.; Khan, M.; Dai, J.-G. Utilization of sodium carbonate activator in strain-hardening ultra-high-performance geopolymer concrete (SH-UHPGC). *Front. Mater.* **2023**, *10*, 1142237. [\[CrossRef\]](#)
- Riaz Ahmad, M.; Khan, M.; Wang, A.; Zhang, Z.; Dai, J.-G. Alkali-activated materials partially activated using flue gas residues: An insight into reaction products. *Constr. Build. Mater.* **2023**, *371*, 130760. [\[CrossRef\]](#)
- Sandanayake, M.; Gunasekara, C.; Law, D.; Zhang, G.; Setunge, S.; Wanijuru, D. Sustainable criterion selection framework for green building materials—An optimisation based study of fly-ash Geopolymer concrete. *Sustain. Mater. Technol.* **2020**, *25*, e00178. [\[CrossRef\]](#)
- Abdalla, L.B.; Ghafor, K.; Mohammed, A. Testing and modeling the young age compressive strength for high workability concrete modified with PCE polymers. *Results Mater.* **2019**, *1*, 100004. [\[CrossRef\]](#)
- Lou, Y.; Khan, K.; Amin, M.N.; Ahmad, W.; Deifalla, A.F.; Ahmad, A. Performance characteristics of cementitious composites modified with silica fume: A systematic review. *Case Stud. Constr. Mater.* **2023**, *18*, e01753. [\[CrossRef\]](#)
- Yang, H.; Liu, L.; Yang, W.; Liu, H.; Ahmad, W.; Ahmad, A.; Aslam, F.; Joyklad, P. A comprehensive overview of geopolymer composites: A bibliometric analysis and literature review. *Case Stud. Constr. Mater.* **2022**, *16*, e00830. [\[CrossRef\]](#)
- Pan, Y.; Tannert, T.; Kaushik, K.; Xiong, H.; Ventura, C.E. Seismic performance of a proposed wood-concrete hybrid system for high-rise buildings. *Eng. Struct.* **2021**, *238*, 112194. [\[CrossRef\]](#)
- Wang, Z.; Pan, W.; Zhang, Z. High-rise modular buildings with innovative precast concrete shear walls as a lateral force resisting system. *Structures* **2020**, *26*, 39–53. [\[CrossRef\]](#)
- Jiao, H.; Wang, Y.; Li, L.; Arif, K.; Farooq, F.; Alaskar, A. A novel approach in forecasting compressive strength of concrete with carbon nanotubes as nanomaterials. *Mater. Today Commun.* **2023**, *35*, 106335. [\[CrossRef\]](#)
- De Maio, U.; Fantuzzi, N.; Greco, F.; Leonetti, L.; Pranno, A. Failure Analysis of Ultra High-Performance Fiber-Reinforced Concrete Structures Enhanced with Nanomaterials by Using a Diffuse Cohesive Interface Approach. *Nanomaterials* **2020**, *10*, 1792. [\[CrossRef\]](#)
- Vitharana, M.G.; Paul, S.C.; Kong, S.Y.; Babafemi, A.J.; Miah, M.J.; Panda, B. A study on strength and corrosion protection of cement mortar with the inclusion of nanomaterials. *Sustain. Mater. Technol.* **2020**, *25*, e00192. [\[CrossRef\]](#)
- Huseien, G.F.; Shah, K.W.; Sam, A.R.M. Sustainability of nanomaterials based self-healing concrete: An all-inclusive insight. *J. Build. Eng.* **2019**, *23*, 155–171. [\[CrossRef\]](#)
- Wang, Y.; Zeng, D.; Ueda, T.; Fan, Y.; Li, C.; Li, J. Beneficial effect of nanomaterials on the interfacial transition zone (ITZ) of non-dispersible underwater concrete. *Constr. Build. Mater.* **2021**, *293*, 123472. [\[CrossRef\]](#)
- Mohammadyan-Yasouj, S.E.; Ghaderi, A. Experimental investigation of waste glass powder, basalt fibre, and carbon nanotube on the mechanical properties of concrete. *Constr. Build. Mater.* **2020**, *252*, 119115. [\[CrossRef\]](#)
- Lushnikova, A.; Zaoui, A. Influence of single-walled carbon nanotubes structure and density on the ductility of cement paste. *Constr. Build. Mater.* **2018**, *172*, 86–97. [\[CrossRef\]](#)
- D'Alessandro, A.; Rallini, M.; Ubertini, F.; Materazzi, A.L.; Kenny, J.M. Investigations on scalable fabrication procedures for self-sensing carbon nanotube cement-matrix composites for SHM applications. *Cem. Concr. Compos.* **2016**, *65*, 200–213. [\[CrossRef\]](#)
- Parvaneh, V.; Khiabani, S.H. Mechanical and piezoresistive properties of self-sensing smart concretes reinforced by carbon nanotubes. *Mech. Adv. Mater. Struct.* **2019**, *26*, 993–1000. [\[CrossRef\]](#)

24. Farooq, F.; Akbar, A.; Khushnood, R.A.; Muhammad, W.L.; Rehman, S.K.; Javed, M.F. Experimental Investigation of Hybrid Carbon Nanotubes and Graphite Nanoplatelets on Rheology, Shrinkage, Mechanical, and Microstructure of SCCM. *Materials* **2020**, *13*, 230. [\[CrossRef\]](#)
25. Lin, W.-T. Effects of sand/aggregate ratio on strength, durability, and microstructure of self-compacting concrete. *Constr. Build. Mater.* **2020**, *242*, 118046. [\[CrossRef\]](#)
26. Adel, H.; Ilchi Ghazaan, M.; Habibnejad Korayem, A. Chapter 9—Machine learning applications for developing sustainable construction materials. In *Artificial Intelligence and Data Science in Environmental Sensing*; Asadnia, M., Razmjou, A., Beheshti, A., Eds.; Academic Press: Cambridge, MA, USA, 2022; pp. 179–210.
27. Huang, J.; Zhang, J.; Li, X.; Qiao, Y.; Zhang, R.; Kumar, G.S. Investigating the effects of ensemble and weight optimization approaches on neural networks' performance to estimate the dynamic modulus of asphalt concrete. *Road Mater. Pavement Des.* **2023**, *24*, 1939–1959. [\[CrossRef\]](#)
28. Huang, J.; Zhou, M.; Zhang, J.; Ren, J.; Vatin, N.I.; Sabri, M.M.S. Development of a new stacking model to evaluate the strength parameters of concrete samples in laboratory. *Iran. J. Sci. Technol. Trans. Civ. Eng.* **2022**, *46*, 4355–4370. [\[CrossRef\]](#)
29. Ben Chaabene, W.; Flah, M.; Nehdi, M.L. Machine learning prediction of mechanical properties of concrete: Critical review. *Constr. Build. Mater.* **2020**, *260*, 119889. [\[CrossRef\]](#)
30. Chen, Z.; Amin, M.N.; Iftikhar, B.; Ahmad, W.; Althoey, F.; Alsharari, F. Predictive modelling for the acid resistance of cement-based composites modified with eggshell and glass waste for sustainable and resilient building materials. *J. Build. Eng.* **2023**, *76*, 107325. [\[CrossRef\]](#)
31. Khan, K.; Ahmad, W.; Amin, M.N.; Rafiq, M.I.; Abu Arab, A.M.; Alabdullah, I.A.; Alabduljabbar, H.; Mohamed, A. Evaluating the effectiveness of waste glass powder for the compressive strength improvement of cement mortar using experimental and machine learning methods. *Heliyon* **2023**, *9*, e16288. [\[CrossRef\]](#)
32. Wang, N.; Xia, Z.; Amin, M.N.; Ahmad, W.; Khan, K.; Althoey, F.; Alabduljabbar, H. Sustainable strategy of eggshell waste usage in cementitious composites: An integral testing and computational study for compressive behavior in aggressive environment. *Constr. Build. Mater.* **2023**, *386*, 131536. [\[CrossRef\]](#)
33. Taffese, W.Z.; Espinosa-Leal, L. A machine learning method for predicting the chloride migration coefficient of concrete. *Constr. Build. Mater.* **2022**, *348*, 128566. [\[CrossRef\]](#)
34. Zheng, X.; Xie, Y.; Yang, X.; Amin, M.N.; Nazar, S.; Khan, S.A.; Althoey, F.; Deifalla, A.F. A data-driven approach to predict the compressive strength of alkali-activated materials and correlation of influencing parameters using SHapley Additive exPlanations (SHAP) analysis. *J. Mater. Res. Technol.* **2023**, *25*, 4074–4093. [\[CrossRef\]](#)
35. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [\[CrossRef\]](#)
36. Dahiya, N.; Saini, B.; Chalak, H.D. 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**. [\[CrossRef\]](#)
37. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems 30 (NIPS 2017), Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS2017), Long Beach, CA, USA, 4–9 December 2017*; Neural Information Processing Systems Foundation, Inc.: La Jolla, CA, USA, 2017.
38. Fan, J.; Ma, X.; Wu, L.; Zhang, F.; Yu, X.; Zeng, W. Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agric. Water Manag.* **2019**, *225*, 105758. [\[CrossRef\]](#)
39. Shehadeh, A.; Alshboul, O.; Al Mamlook, R.E.; Hamedat, O. Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression. *Autom. Constr.* **2021**, *129*, 103827. [\[CrossRef\]](#)
40. Nguyen, H.; Vu, T.; Vo, T.P.; Thai, H.-T. Efficient machine learning models for prediction of concrete strengths. *Constr. Build. Mater.* **2021**, *266*, 120950. [\[CrossRef\]](#)
41. Saud, S.; Jamil, B.; Upadhyay, Y.; Irshad, K. Performance improvement of empirical models for estimation of global solar radiation in India: A k-fold cross-validation approach. *Sustain. Energy Technol. Assess.* **2020**, *40*, 100768. [\[CrossRef\]](#)
42. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 20–26 August 1995*; pp. 1137–1145.
43. Wang, H.-L.; Yin, Z.-Y. Unconfined compressive strength of bio-cemented sand: State-of-the-art review and MEP-MC-based model development. *J. Clean. Prod.* **2021**, *315*, 128205. [\[CrossRef\]](#)
44. Mosavi, A.; Edalatfar, M. A hybrid neuro-fuzzy algorithm for prediction of reference evapotranspiration. In *Recent Advances in Technology Research and Education, Proceedings of the International Conference on Global Research and Education, Kaunas, Lithuania, 24–27 September 2018*; Springer: Cham, Switzerland, 2019; pp. 235–243.
45. Siahkouhi, M.; Razaqpur, G.; Hoult, N.A.; Hajmohammadian Baghban, M.; Jing, G. Utilization of carbon nanotubes (CNTs) in concrete for structural health monitoring (SHM) purposes: A review. *Constr. Build. Mater.* **2021**, *309*, 125137. [\[CrossRef\]](#)
46. Zheng, D.; Wu, R.; Sufian, M.; Kahla, N.B.; Atig, M.; Deifalla, A.F.; Accouche, O.; Azab, M. Flexural Strength Prediction of Steel Fiber-Reinforced Concrete Using Artificial Intelligence. *Materials* **2022**, *15*, 5194. [\[CrossRef\]](#) [\[PubMed\]](#)

47. Khan, K.; Ahmad, W.; Amin, M.N.; Ahmad, A.; Nazar, S.; Al-Faiad, M.A. Assessment of Artificial Intelligence Strategies to Estimate the Strength of Geopolymer Composites and Influence of Input Parameters. *Polymers* **2022**, *14*, 2509. [[CrossRef](#)]
48. Khan, K.; Amin, M.N.; Sahar, U.U.; Ahmad, W.; Shah, K.; Mohamed, A. Machine learning techniques to evaluate the ultrasonic pulse velocity of hybrid fiber-reinforced concrete modified with nano-silica. *Front. Mater.* **2022**, *9*, 1098304. [[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.