Research Article

Nashat S. Alghrairi*, Farah N. Aziz, Suraya A. Rashid, Mohd Z. Mohamed and Amer M. Ibrahim

Machine learning-based compressive strength estimation in nanomaterial-modified lightweight concrete

https://doi.org/10.1515/eng-2022-0604
received January 16, 2024; accepted February 24, 2024

Abstract: The development of nanotechnology has led to the creation of materials with unique properties, and in recent years, numerous attempts have been made to include nanoparticles in concrete in an effort to increase its performance and create concrete with improved qualities. Nanomaterials are typically added to lightweight concrete (LWC) with the goal of improving the composite's mechanical, microstructure, freshness, and durability qualities. Compressive strength is the most crucial mechanical characteristic for all varieties of concrete composites. For this reason, it is essential to create accurate models for estimating the compressive strength (CS) of LWC to save time, energy, and money. In addition, it provides useful information for planning the construction schedule and indicates when the formwork should be removed. To predict the CS of LWC mixtures made with or without nanomaterials, nine different models were proposed in this study: the gradient-boosted trees (GBT), random forest, tree ensemble, XGBoosted (XGB), Keras, simple regression, probabilistic neural networks, multilayer perceptron, and linear relationship model. A total of 2,568 samples were gathered and examined. The most significant factors influencing CS during the modeling process were taken into account as input variables, including the amount of nanomaterials, cement, water-to-binder ratio, density, the content of lightweight aggregates, type of nano, fine and coarse aggregate content, and water. The performance of the suggested models was assessed using a variety of statistical measures, including the coefficient of determination (R^2), scatter index, mean absolute error, and root-mean-squared error (RMSE). The findings showed that, in comparison to other models, the GBT model outperformed the others in predicting the compression strength of LWC mixtures enhanced with nanomaterials. The GBT model produced the best results, with the greatest value of R^2 (0.9) and the lowest value of RMSE (5.286). Furthermore, the sensitivity analysis showed that the most important factor influencing the prediction of the CS of LWC enhanced with nanoparticles is the water content.

Keywords: compressive strength, machine learning, regression model, nanomaterials, nano silica, nano metakaolin, lightweight concrete

*** Corresponding author: Nashat S. Alghrairi,** Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia, UPM, 3400 Serdang, Malaysia, e-mail: nashaatalgrara@gmail.com
Farah N. Aziz: Housing Research Centre, Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Malaysia, e-mail: farah@upm.edu.my
Suraya A. Rashid: Institute of Nanoscience and Nanotechnology, Universiti Putra Malaysia, 43400 UPM Serdang, Malaysia, e-mail: suraya_ar@upm.edu.my
Mohd Z. Mohamed: Department of Mechanical Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Malaysia, e-mail: zuhri@upm.edu.my
Amer M. Ibrahim: Department of Scientific Affairs, Diyala University, Baghdad, Iraq, e-mail: amereng76@yahoo.com

1 Introduction

It is possible to improve the mechanical characteristics and performance of lightweight concrete (LWC) by modifying it with nanoparticles. Compressive strength is a crucial property of concrete that is used to evaluate material's load-bearing ability and structural integrity. Experimental testing has always been the primary method for estimating compressive strength (CS) although it can be labor and resource intensive. However, new developments in machine learning (ML) techniques have created new opportunities for precise and effective CS estimation in LWC enhanced with nanoparticles.

To create prediction algorithms, ML-based CS estimation includes training models with data from experimental tests and other pertinent characteristics. The CS of LWC with nanomaterial changes can then be estimated using these algorithms, which take into account a variety of input variables like mix proportions, curing conditions, type and dose of the nanomaterials, and others. With the use of ML algorithms, this method analyzes intricate relationships and patterns in the data to provide precise CS estimates.

There are many benefits to applying ML-based CS estimates in LWC enhanced with nanoparticles. First, it speeds up development by reducing reliance on expensive and time-consuming experimental testing, which conserves resources. Second, by taking into account several factors at once, it offers a more thorough knowledge of how nanomaterials affect CS. This makes it possible to identify the most important variables influencing CS and optimize the mixture design.

Furthermore, complex nonlinear interactions that may be difficult for traditional analytical approaches to capture and handle vast datasets are also capabilities of ML models. Predictions become more reliable as a result of the capacity to estimate CS more precisely and robustly. Furthermore, as new data become available, ML models may be updated and enhanced regularly, which over time will increase their accuracy and flexibility.

It is crucial to remember that estimating CS using ML is not without its difficulties. Important factors to take into account include the training dataset's representativeness and quality, the choice of relevant input variables, and the models' ability to generalize to various contexts. The models must be carefully validated and verified to guarantee their efficacy.

In conclusion, a major development in concrete technology has been made with the introduction of ML-based CS estimation in LWC modified with nanoparticles. It can estimate CS more quickly, accurately, and economically. This might lead to more effective mixture design optimization and the creation of novel LWC materials with improved mechanical properties. The construction industry has considerable potential when it comes to applying ML to research and development of concrete.

The main aim of this study was to provide multiscale models to predict the CS of LWC-containing nanoparticles. To that end, a variety of laboratory work data, approximately 2,568 tested specimens with varying cement content, fine aggregate content, normal coarse aggregate content, lightweight aggregate content, water, water/cement ratio, density, percent of nano, and type of nano were taken into consideration with various analysis approaches aiming:

(i) To ensure that the construction industry can use the models provided without any theoretical.
(ii) To conduct statistical analysis and identify the influence of various parameters on the CS of LWC.
(iii) To quantify and provide a systematic multiscale model to predict the CS of LWC containing nanoparticles.
(iv) To use statistical assessment tools to identify the most reliable model among nine different model strategies (XGB, random forest (RF), linear regression (LR), SL, Keras, multilayer perceptron (MLP), probabilistic neural networks simple regression trees (SRT), and TER) for predicting the CS of LWC.

2 Literature review

When designing and building concrete, one of the most important quality control parameters is the LWC's CS. The widespread use of CS can be ascribed to the ease of its testing process and its strong association with other crucial factors, such as durability and tensile strength. Furthermore, structural engineers make considerable use of CS when designing structures [1]. Because LWC is highly complicated and heterogeneous, designing and producing it generally require trial and error. This is due to the fact that there is a wide range of mixture ingredients that can be used to make concrete, such as cement, water, admixtures, aggregates, and supplemental cementitious materials (SCMs), all of which have a wide range of physical and chemical characteristics [1–3]. The scenario becomes more difficult due to the potential impact of environmental conditions like temperature and humidity on the characteristics of concrete [4,5]. To enable high-throughput mixture design and minimize the time-consuming trial-and-error methods currently employed in the manufacturing of concrete, a predictive model that is efficient and dependable and that can estimate CS based on mixture proportions is required. A strong predictive model would provide more optimized mixtures that satisfy design specifications and raise mixture efficiency [6,7]. The field of ML is growing quickly and has demonstrated significant promise for applications in civil engineering [8,9]. Large datasets may be analyzed by ML algorithms, which can then extract valuable information to help with decision-making, classification, and prediction [10]. Based on a variety of input parameters, such as dosage of cement, sand content, normal aggregate content, lightweight aggregate content, water mix, density, and other SCMs such as nano silica and nano metakaolin (NMK), and the ratio of water to cementitious materials (W/Cm) [9], ML algorithms can be used to predict the CS [7,11], modulus of elasticity [3,12], flexural strength [13,14], pore solution connectivity [9], and other properties of concrete. The ability of ML to process enormous amounts of data considerably more quickly and precisely than conventional approaches is one of the main advantages of utilizing it in concrete analysis. This enables engineers to minimize the amount of experimental testing and improve the mix design, which can save money and time while increasing the efficiency of the project as a whole [6]. Furthermore, ML can provide detailed information about the intricate relationships between various factors and highlight connections that conventional statistical methods could have overlooked [10]. Using ML to model the intricate structure of concrete may prove to be a very successful strategy. The CS of concrete has been predicted using a number of ML models, including gradient-boosted trees (GBT), extreme gradient boosting (XGBoost), RFs, tree ensemble (TE), LR, MLP, SRT, Keras learning network, and artificial neural networks (ANNs, i.e., PNN) [6]. It has been demonstrated that these models are capable of accurately predicting the CS of concrete. To offer more illustrations beyond the already referenced literature, Maherian et al. [15] intended to use support vector machine (SVM) and Gaussian process regression (GPR) ML approaches to create predictive models for the CS of concrete containing nano-silica (NS). Cement concentration, aggregates, NS characteristics, water–binder ratio (W/B), and forecast time were among the input factors. According to the findings, SVM performed better in CS prediction than GPR. According to Zaman et al. [16], curing time and silica concentration are important determinants of the 28-day CS of cement paste with nanocomposites. ML methods, in particular, ensemble approaches combining ANN and RF, efficiently predict this strength, attaining an R^2 value of 0.97. In addition to proposing streamlined formulas for calculating strength, Kioumarsi et al. [17] created an ML-based model for forecasting the CS of concrete containing ground granulated blast furnace slag (GGBFS). DT, RF, SVM, K-nearest neighbors (KNNs), and ANN were among the ML techniques used with a large dataset comprising 625 experimental trials. Performance metrics were used to evaluate the model's correctness, and a sensitivity analysis was conducted to determine how GGBFS affected the CS of concrete. The results of this study were used to propose equations for the model. Shahmansouri et al. [18] conducted research to examine the possibility of using an ANN model to forecast the electrical resistivity and CS of natural zeolitic concrete (NZC). Seven input variables were used in the study: specimen age, W/Cm, cement, natural zeolite, gravel, sand, and superplasticizer contents. The experimental data from 324 NZC specimens were used. The model's correctness was validated by extensive computer studies, providing a dependable and effective substitute for expensive and time-consuming experiments in the optimization of environmentally friendly concrete mixtures. Young et al. [7] examined a large data set (>10,000 observations) of measured CSs from real job site mixtures and their corresponding mixture proportions in a different study. This study focused on the industrial significance of predicting concrete's CS using statistical and ML approaches based on mixture proportions. The models were also used to determine the best concrete combinations that meet target strength requirements while minimizing cost and embodied CO_2 impact. Alshammari [19] optimized concrete compositions for sustainability and strength using artificial intelligence. By using a dataset of 1,150 typical concrete formulas to train an ML model, it was possible to forecast concrete strength at different ages with previously unheard of precision. The model created high-performance concrete with a lower environmental impact by taking carbon embodiment into account. Li et al. [6] addressed the difficulties in formulating desired features as she investigated the transformational potential of ML in practical research. Their review, which addresses the implementation, application, and interpretation of ML algorithms, demonstrates the benefits of ML in concrete research. Surono et al. [20] developed a deep learning and ML model for COVID-19 classification using convolutional neural network (CNN) architectures and ML algorithms. The article compares the accuracy of different CNN models with ML algorithms for classifying CXR lung images, with the MobileNetV2-SVM structure achieving the highest accuracy of 93%. Bayesian optimization was used to optimize the ML algorithms, resulting in improved accuracy. The SVM algorithm outperformed other ML algorithms such as Naive Bayes, KNN, and decision tree in terms of accuracy. The MobileNetV2-SVM model achieved an accuracy of 93% in classifying the CNN-extracted features from the lung images. Precision, recall, and F1-score values for the MobileNetV2-SVM model were 0.83 each. The SVM algorithm outperformed other ML algorithms, such as Naive Bayes, KNN, and decision tree, in terms of accuracy. The research provided valuable insights into the development of an accurate and efficient diagnostic system for COVID-19 using state-of-the-art CNN models and optimized ML algorithms.

NS's effectiveness depends on a number of variables, including its dosage and particle size, the dispersion technique employed, and the mixture's W/Cm, among others [21]. Srinivas [22] claim that because of the latter's coarse capillary pore structure, which made it difficult for NS particles to effectively fill the pores and reinforce the microstructure, the use of NS significantly improved the mechanical properties of concrete mixtures with a W/Cm ratio of 0.36 as opposed to those with a ratio of 0.55. Taher and Ismael [23] studied the incorporation of NS into asphalt mixtures to improve the resistance of the asphalt mixture to rutting. The study showed that adding 4% NS effectively improved the asphalt mixture against corrosion and corrosion. Moreover, the addition of 6%

NS improved Marshall's stability by 33%. In addition, the softening point of asphalt increased and the mixture showed greater resistance to permanent deformation and corrosion at high temperatures. Yasin et al. [22] developed an ML model to improve LWC mix. The researchers used light volcanic tuff with unique properties. ANNs have been successfully used to predict the optimum content of Tuff fine aggregate in LWC, with a range of CSs from 20 to 50 MPa. The ANN model showed a clear agreement between the predicted values and the experimental ones, indicating its accuracy in predicting the properties of LWC. The use of ANNs in concrete mix design can help minimize the need for costly and time-consuming experimental investigations, leading to cost and time savings in construction projects. The ANN model developed in this research can be used to predict new data ranges, indicating its potential for application in different scenarios. The ANN-FF method provided higher accuracy in predicting the CS compared to the ANN-CF method, especially for Tuff percentages up to 50. The developed ANN model can contribute to expanding the applications of Tuff aggregate in LWC production, offering potential benefits in terms of cost savings and increased efficiency.

However, the significance of NS's ability to improve concrete characteristics increases as the *W*/*C* ratio drops and finer capillary pores appear [24]. It is also well known that the mechanical characteristics and workability of concrete are significantly influenced by the size of NS particles. Rao and Maruthi, [25] studied the effects of NS and Metakaolin (MK) on different concrete qualities. When preparing concrete, MK and NS are utilized in part lieu of cement. In the current study, MK is initially used to substitute cement to the tune of 5 and 10% by weight. Subsequent research is conducted by substituting NS at 1, 2, and 3% by weight of cement for MK at 5 and 10%. The modulus of elasticity, flexural strength, split tensile strength, CS, and other characteristics of M25 grade concrete containing MK and NS are assessed for structural applications, and the outcomes are compared with the controlled concrete. The test results show that, in comparison to the controlled concrete, the concrete created using a mixture of 5% MK and 2% NS exhibited greater strength. Therefore, it can be said that concrete made with a mixture of 5% MK and 2% NS is suitable for structural applications. Not only it process by its nucleating impact but also it raises the need for plasticizer with rising NS and NMK dosages, changing the mixture's fresh and hardened properties in the process and adding to the cementitious system's complexity [21].

Based on the published literature in recent years, there is a lack of use of all parameters that affect CS. In addition, some appeared to use a few algorithms conducting statistical analysis, and this in itself constitutes a step in the accuracy of the results. Moreover, not more than 1,440 samples were recorded for predicting compression resistance. As well, a few pieces of literature, which are interested in light concrete-enhanced nanomaterials, were also identified. Even if it is found, the focus is on the use of two nanomaterials or three. This misses the effect of the types of nanotechnologies that can be used (Table 1).

3 Methodology

In ML-based CS estimates for LWC modified with nanomaterials, the specific set of CS parameters chosen is usually

Table 1: Concrete-strength prediction methods using machine-learning algorithms

Author	Algorithm	Data volume	Results
Lai and Serra [26]	ANN	240	Relative error less than 5%
Kewalramani and Gupta [27]	ANN	864	Maximum error rate 25.69%
Naderpour et al. [28]	ANN	139	R = 0.8926, MSE = 0.004447
Morsy et al. [29]	ANN	205	R^2 = 0.919
Wang et al. [30]	GA-SVM	24	Maximum relative error 2.42%
Kurihara and Maruyama [31]	SVM	80	R-0.94
Kurihara and Maruyama [31]	SVM	239	R^2 = 0.87, RMSE = 4.86, MAPE = 9.81%
Wang et al. [30]	BP, RBF	19	Relative error of less than 6%
Maqbool et al. [32]	BP-ANN	30	Absolute error less than 5.0%
Nivethitha and Dharmar [33]	BP	251	Coefficient of variation = 0.112
Wu et al. [34]	RF	56	R^2 = 0.969, RMSE = 0.0149
Ortenzi et al. [35]	RF	1,030	R^2 = 0.902, MAE = 3.761, MAPE = 12.807, RMSE = 5.342
Farooq et al. [36]	RF, GEP	357	R^2 = 0.96 (RF), R^2 = 0.9 (GEP)
Feng et al. [37]	AdaBoost	1,030	R^2 = 0.952, MAPE = 11.39%, RMSE = 4.856

chosen for a number of reasons. 1) Relevance to compressive strength: The parameters chosen ought to affect the concrete's CS, either directly or indirectly. These variables could be the kind and amount of nanomaterials used, the water-to-cement ratio, the curing environment, the characteristics of the aggregate, and other elements that are known to affect CS. The purpose of selecting these characteristics was to identify the critical elements that influence the development of strength in LWC modified with nanomaterials. 2) Availability and accessibility: In both experimental and real-world settings, the parameters used should be readily observable or achievable. Data accessibility is essential for ML model validation and training. It might not be feasible to regularly utilize parameters for CS estimation if they necessitate intricate or costly testing processes. Thus, using measurable and easily accessible characteristics makes the process of gathering data easier. 3) Influence on nanomaterial effects: Particular attention should be paid to how nanomaterials affect CS in the parameters. For example, the kind and dosage of nanomaterials may have a substantial impact on the strength development; hence, the model may accurately reflect their influence by considering these characteristics. Through explicit consideration of the parameters linked to nanoparticles, the ML model is able to efficiently estimate the CS of LWC modified with nanomaterials. 4) Statistical significance: There should be evidence of statistical significance between the selected parameters and CS. The parameters that significantly affect the model's predicted accuracy and have a high link with CS can be found via statistical analysis. By adding statistically significant parameters, the model can be made to concentrate on the most important elements while avoiding the addition of superfluous or unnecessary variables. 5) Model complexity and overfitting: A balance between model complexity and overfitting is the reason for choosing a certain set of parameters. Overfitting, in which the model performs badly on untested data and becomes unduly specialized to the training dataset, can result from having too many parameters. However, leaving out crucial variables could lead to a simplistic model that is unable to adequately represent the nuanced nature of the CS relationship. As a result, great thought is paid to incorporate the most important parameters without adding needless complication.

It is worth noting that some previous studies used different parameters and others used the same parameters, but not in this number. It is crucial to remember that the criteria for CS can be chosen based on the particular goals of the study, the information at hand, and the properties of the LWC and nanomaterials under investigation. The selection of parameters should have a rationale that is in line with the objectives of the research, guarantee the feasibility of gathering data, and improve the precision and dependability of the CS estimation model that is based on ML.

A total of 2,568 datasets were collected in 2023 from the previous work on the CS of LWC. In the literature, there is a range of data regarding LWC with different base source materials, including fly ash, nano silica fume, nano Al_2O_3, nano Fe_2O_3, NMK, nano clay, and so on. But in this article, the authors take those papers that use NMK and NS as base source materials to prepare LWC-containing nanoparticles. The models used nine input parameters to restrict authors from using more datasets in the developed models. The collected datasets were statistically analyzed and split into two groups. The larger group, which included 1,535 (78%) datasets, was used to create the models. The second group consists of 433 (22%) datasets used to test the proposed models. The input dataset consists of the cement content range from 92 to 686 kg/m^3, fine aggregate range from 440 to 1,078 kg/m^3, normal coarse aggregate range from 0 to 1,532 kg/m^3, lightweight weight aggregate range from 0 to 1,273 kg/m^3 water range from 100 to 257 l/m^3, W/C range from 0.23 to 0.66, density range from 710 to 2,656 kg/m^3% of nanoparticles range from 0 to 110%. The former dataset was then used to propose different models to predict the CS of LWC and compared with the actual experimental CS (MPa). After that, the developed models were assessed by some statistical criteria such as coefficient of determination R^2, root-mean-square error (RMSE), absolute mean error AME, scatter index (SI), and objective value (OBJ) to indicate the most reliable and accurate model. Further details of the data collection and modeling work are summarized in the form of a flow chart, as depicted in Figure 1.

4 Statistical evaluation

The statistical study carried out in this section aims to demonstrate whether or not there are significant correlations between the input parameters and the CS of LWC. To illustrate the distribution of each variable with CS, statistical functions such as variance, skewness, kurtosis, and standard deviation were determined. All considered variables, including (i) nanomaterials content, (ii) cement content, (iii) W/B, (iv) water content, (v) density content, (vi) fine aggregate, (vii) normal coarse aggregate content, (viii) lightweight aggregate content, and (ix) type of nano, were plotted and analyzed with CS. A strong negative value for the kurtosis parameter indicates that the distribution tails are shorter than those of a normal distribution; a positive value is represented by longer tails. A large negative value

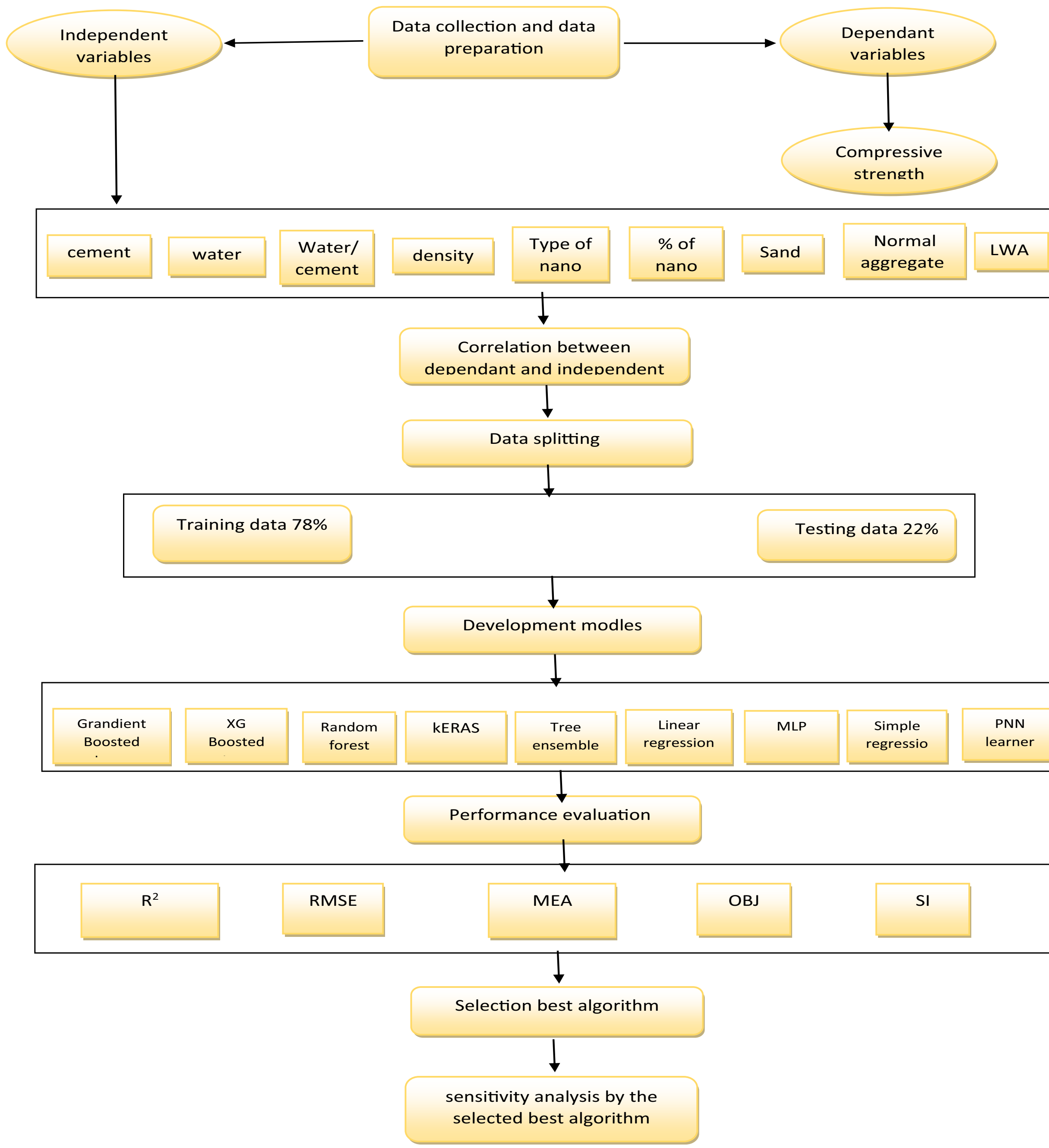

Figure 1: Flow chart of the MLA methods.

for the skewness parameter indicates a long left tail, while a positive value indicates a right tail. Enough details about each variable that is thought to be an input parameter are provided below:

4.1 Cement content

Ordinary Portland Cement Type I, which met ASTM C 150 requirements, was the type of Portland cement used in most LWC combinations. The specific gravity ranged from 3.05 to 3.2, while the specific surface area ranged from 300 to 400 m^2/kg. The amount of cement ranged from 92 to 686 kg/m^3, with a median of 373 kg/m^3, a standard deviation of 108.65 kg/m^3, and a variance of 11804.26 kg/m^3, due to the higher cement and binder content in the LWC. The variables for skewness and kurtosis are, respectively, −0.3423 and −0.2553. Figure 2 shows the relationship between cement concentration and CS with a normal distribution and histogram of LWC mixtures modified with nanomaterials.

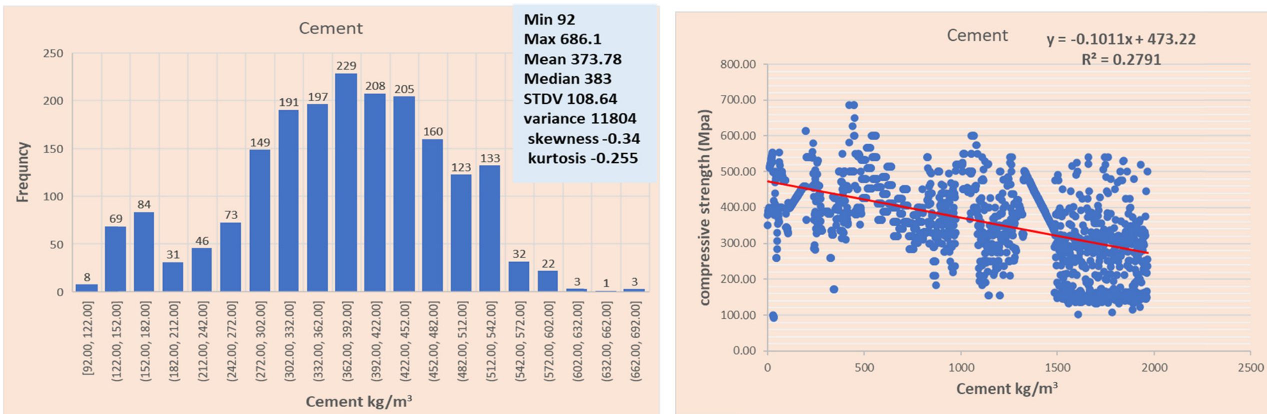

Figure 2: The variation between CS and cement content with normal distribution and Histogram of LWC mixtures modified with nanomaterials.

4.2 Fine aggregate content

River sand with a maximum aggregate size of 4.75 mm and a specific gravity ranging from 2.65 to 2.75 was the fine aggregate employed in the previous research. In addition, its gradation complied with ASTM C 33's requirements. The maximum and minimum fine aggregate content in the LWC mixtures were 440 to 1,072 kg/m^3, with a median of 769 kg/m^3, a standard deviation of 114 kg/m^3, and a variance of 13,071.1 kg/m^3, based on the total of 2,568 collected data from the literature. Skewness and kurtosis, two additional functional parameters for the fine aggregate dosage in the LWC mixtures, are −0.337 and 0.124, respectively. Figure 3 shows the correlation between CS and fine aggregate content using a histogram of LWC mixtures modified with nanomaterials.

4.3 Normal aggregate content

According to the literature, coarse aggregate for the creation of LWC was crushed stone or gravel with a maximum aggregate size of 20 mm. The greatest and minimum coarse aggregate content, based on all 2,568 gathered data from previous investigations, were 0 and 1,532 kg/m^3, with a median of 407 kg/m^3, a standard deviation of 442 kg/m^3, and a variance of 1,961 kg/m^3. The statistical values for skewness and kurtosis are 0.255 and −1.77, respectively. Figure 4 displays the relationship between CS and coarse aggregate content using a histogram of LWC mixtures that have been altered using nanomaterials.

4.4 Density

The density of LWC mixtures modified with nanomaterials ranged from 710 to 2,656 kg/m^3, with a median of 1,881 kg/m^3, based on the overall data gathered from previous investigations. Other variables like variance, standard deviation, skewness, and kurtosis are 95,717, 309.38, −0.268, and −0.668, respectively, according to the statistical analysis. Furthermore, Figure 5 reports the variation in the

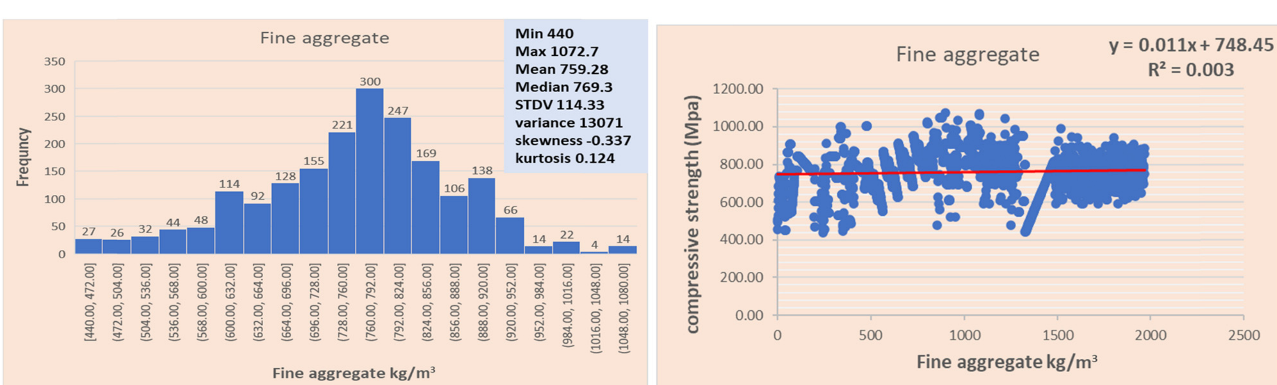

Figure 3: The variation between CS and fine aggregate content with the histogram of LWC mixtures modified with nanomaterials.

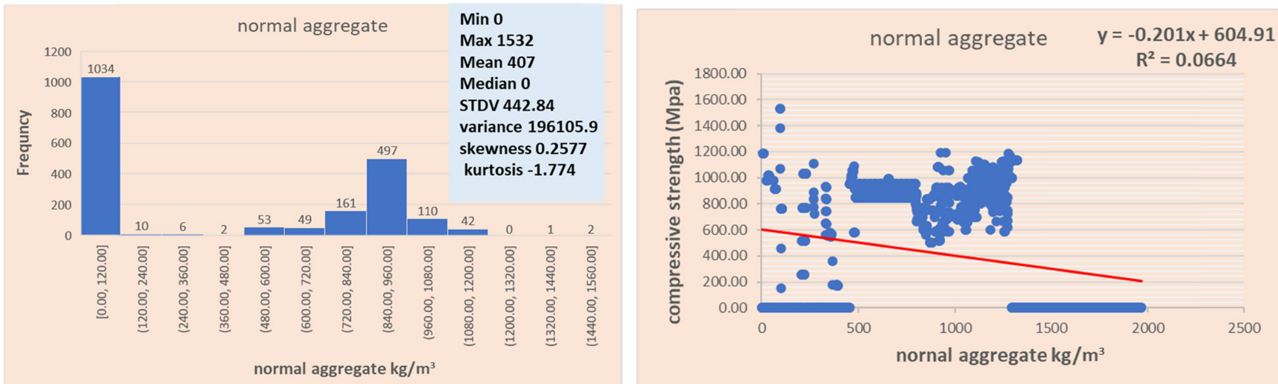

Figure 4: The variation between CS and coarse aggregate content with the histogram of LWC mixtures modified with nanomaterials.

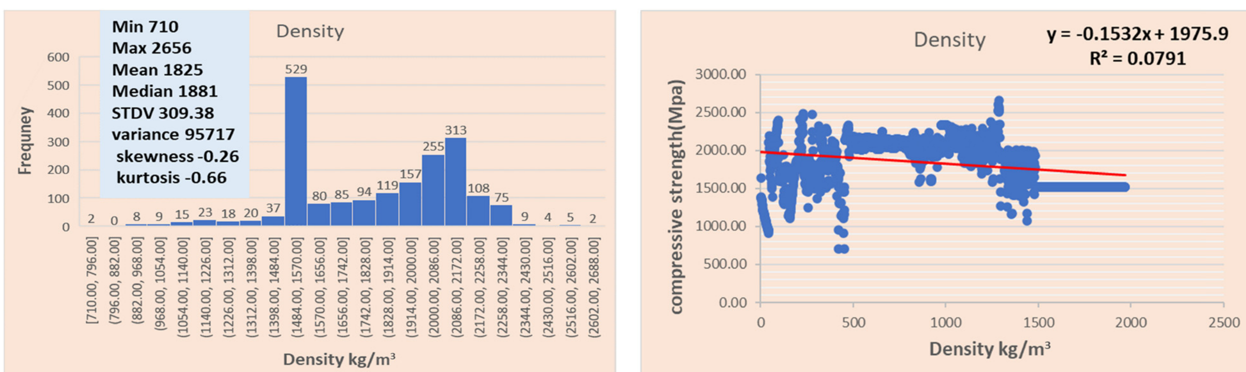

Figure 5: The variation between CS and density with normal distribution and histogram of LWC mixtures modified with nanomaterials.

Histogram and CS and density of LWC mixtures modified with nanomaterials.

4.5 Lightweight aggregate content

In the literature, lightweight aggregate for the creation of LWC was described as being crushed lightweight stone with a maximum aggregate size of 17 mm. The lightweight aggregate content ranged from 0 to 1,276 kg/m^3, with a median of 454 kg/m^3, a standard deviation of 276.7 kg/m^3, and a variance of 71,777 kg/m^3, based on all 2,568 collected data from previous studies. The statistical values for skewness and kurtosis are 0.451 and −0.829, respectively. Figure 6 displays the relationship between CS and lightweight

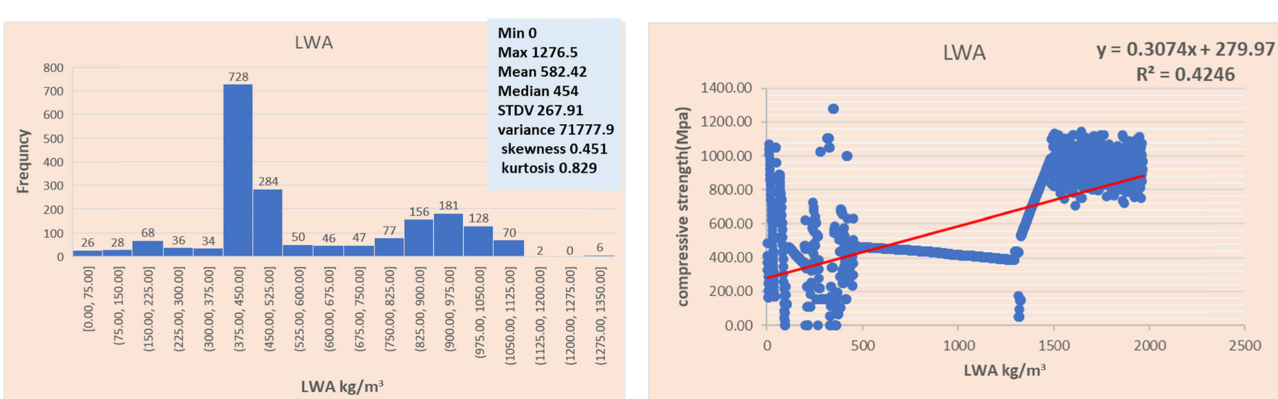

Figure 6: The variation between CS and lightweight aggregate content with the histogram of LWC mixtures modified with nanomaterials.

aggregate content using a histogram of LWC mixtures that have been altered with nanomaterials.

4.6 Water content

From the 2,568 data collected, potable water was used in the process of mixing concrete components. The minimum amount of water used was 100 l/m^3, the largest amount was 275 l/m^3, the mean was 180 l/m^3, the standard deviation was 26.55 l/m^3, and the variance was 704, and The statistical values for skewness and kurtosis are −0.239 and 0.355, respectively. Figure 7 displays the relationship between CS and water content using a histogram of LWC mixtures that have been altered with nanomaterials.

4.7 Water/cement ratio

The *W*/*C* ratio of LWC mixtures modified with nanomaterials ranged from 0.26 to 0.66 with a median of 0.45 based on the overall data collected from previous investigations. Other characteristics including variance, standard deviation, skewness, and kurtosis are 0.006, 0.08, 0.279, and 0.211, respectively, according to the statistical analysis. Furthermore, Figure 8 reports the variation in the histogram and CS and *W*/*C* ratio of LWC mixtures treated with nanomaterials. Figure 4 shows that the CS of LWC modified with nanomaterials was dramatically reduced with an increase in the *W*/*B* ratio.

4.8 % of nanomaterials

The nanomaterials employed in the mix proportions had a particle diameter below 50 nm, a surface-to-volume ratio between 50 and 200 m^2/g, and a purity of greater than 99%, according to the dataset, which included 2,568 data samples from the literature. With a median of 11.36%, the minimum and highest proportion of nanomaterials utilized in the LWC combinations to replace the cement

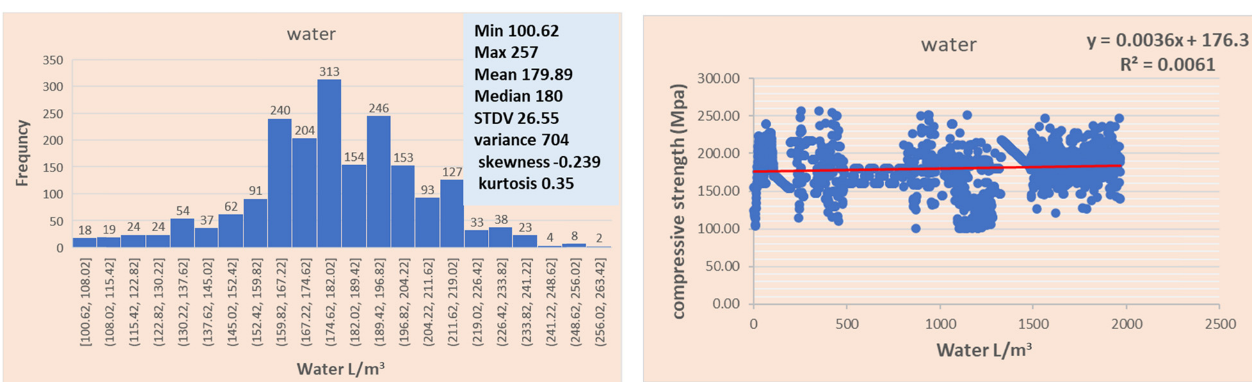

Figure 7: The variation between CS and water content with a histogram of LWC mixtures modified with nanomaterials.

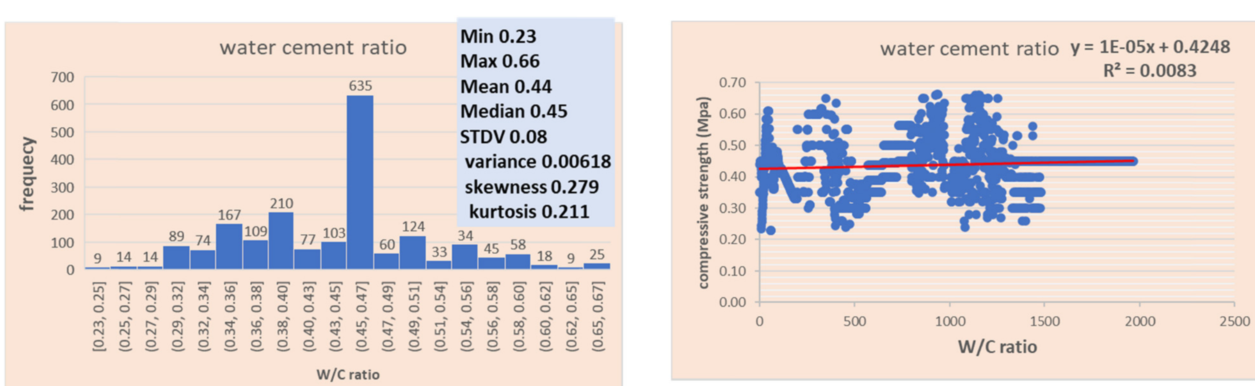

Figure 8: The variation between CS and *W*/*C* ratio with normal distribution and histogram of LWC mixtures modified with nanomaterials.

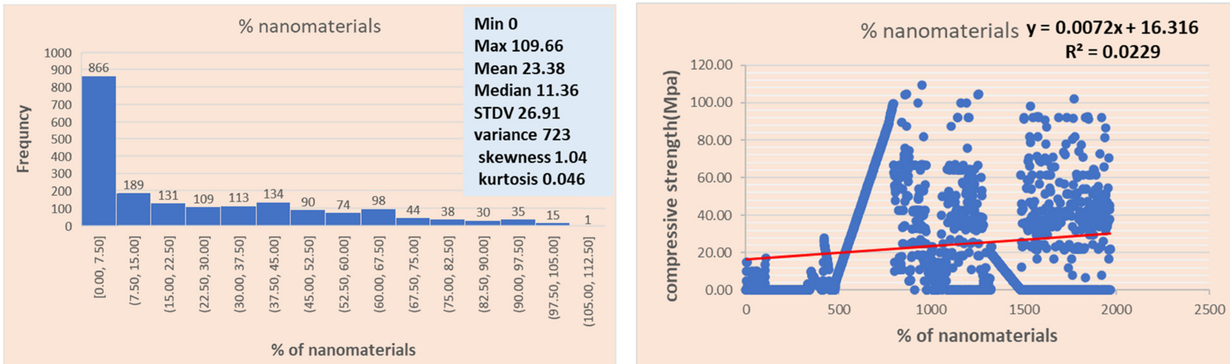

Figure 9: The variation between CS and % of nanomaterials with normal distribution and histogram of LWC mixtures modified with nanomaterials.

weight ranged from 0 to 110%. In addition, the variance, skewness, kurtosis, and standard deviation are, respectively, 26.91, 723.9, 1.041, and 0.046%. In Figure 9, the histogram analysis and the relationship between CS and nanomaterials content are displayed.

5 ML modeling

None of the nine variables under investigation have a straightforward mathematical relationship with CS. Consequently, CS can now be predicted statistically and analytically without a lot of testing or material consumption thanks to the KNIME application. The coefficient of determination, R^2, verified the veracity of the findings [38,39]. A number of algorithms have been employed to assess how factors affect CS [40]. The data were divided into two parts for all algorithms, and the program was trained and tested, obtaining the ideal weights and biases. By finding the coefficient of determination for each algorithm, the accuracy of each algorithm was evaluated, and the highest R^2 value represents the best because it has less error and less dispersion [41]. The study included the use of algorithms with excellent effectiveness. For example, XGBoosted, ANNs, LR, gradient boosted, RF regression, and other algorithms. Part of it will be reviewed.

5.1 XGBoost model

Distributed gradient boosting library is a learning technique used to train models via ML quickly and scale them. It is also known as an ensemble learning technique through which the results of several ineffective models can be combined to increase their accuracy. It was developed to become gradient boosting to be more famous than other algorithms due to its ability to process large groups and its ability to show results in diverse and sophisticated task groups. The XGBoost approach is based on a decision tree [41].

5.2 ANNs MLP

ANN is a type of algorithm in ML. Interconnected artificial nerve cells or nodes are used and are built in layers similar in composition to the human brain. They have been used in engineering disciplines, especially building and construction. It is considered a relatively new and simple computing technology. This technology has proven its ability to predict and address relationships between input and output parameters, and these neurons are responsible for processing data. An ANN contains several layers. An ANN consists of grouping neurons so that they are connected by synapses to form a network. The data is entered, subjected to digital processing, and then the data are exited from the output layer [40].

5.3 RF Model

The RF model can be used to predict CS in various engineering applications. RF is an ML algorithm that is known for its simplicity, versatility, and ability to handle both classification and regression tasks. It has been successfully applied to predict the properties of materials such as concrete. For example, in the study by Vu, an RF model was developed to predict the slump and strength of concrete using mixed mineral admixtures from blast furnace slag and silica fume [42,43]. Similarly, in the study by Ali and

Suthar, an RF-based model outperformed other ML algorithms in predicting the CS of concrete. The RF model achieved high correlation coefficient values and low mean absolute error (MAE) and root-mean-square error values, indicating its accuracy in predicting concrete strength [44]. The advantages of using the RF model to predict CS include its simplicity, versatility, and suitability for both classification and regression tasks [42]. RF models have been successfully applied to a variety of engineering problems, including predicting the slump and strength of concrete. RF models have also been found to outperform other ML algorithms, such as M5P and LR, in terms of prediction accuracy for concrete CS [45]. The RF model has been proven to be the most suitable technique for predicting CS in various concrete mix designs, including geopolymer concrete and glass fiber-reinforced concrete [46]. However, there are also some limitations to using the RF model. It requires a large amount of training data and can be computationally expensive. In addition, the RF model may not perform well if the input parameters are not properly selected or if there are interactions between the input variables that are not captured by the mode. Therefore, the RF model can be a valuable tool for predicting CS in LWC and other engineering materials.

5.4 LR model

The LR model was used to predict CS in the papers by Alabi and Mahachi [47], Imran et al. [48], and Gkountakou and Papadopoulos. Alabi and Mahachi compared the performance of LR with ANN in forecasting the CS of geopolymer recycled concrete based on selected pozzolans. Imran et al. developed a white-box ML model called multivariate polynomial regression to predict the CS of eco-friendly concrete and compared it with LR and SVM. Gkountakou and Papadopoulos studied the prediction of compressive cement strength using fuzzy linear regression and adaptive neuro-fuzzy inference system methods [49]. In this research, LR was used to analyze the collected data (2,568), and it proved its efficiency and ease of obtaining data, but it was not optimal.

5.5 Gradient Boosted Trees Models (GBTs)

GBTs have been used in several studies to predict the CS of concrete. Liu and Sun applied an explainable boosting machine to predict concrete CS and determine the contribution of mix ratio factors on the strength [50]. Dahish and Almutairi constructed a baseline model using a state-of-the-art ML method to predict concrete CS [50]. Dahish and Almutairi proposed a categorical gradient boosting (CatBoost) model with feature importance, feature interaction, partial dependence plot, and shapley additive explanations to explain the mechanism of ML models for predicting CS [50]. Tran and Nguyen developed a hybrid ML model combining GB and Bayesian optimization algorithms to predict the CS of concrete containing ground glass particles. Falah et al. examined the capability of a deep learning neural network approach, along with other AI models, to simulate the CS of environmentally friendly concrete containing recycled aggregate [51].

5.6 Keras network learner

Deep learning is a subfield of artificial intelligence and ML that builds learning models using ANNs as the foundation. In deep learning, “deep” refers to the quantity of layers in a neural network. The “deep learning” set of algorithms was inspired by the composition and functions of the human brain [52]. It uses a large amount of structured and unstructured data to forecast outcomes and effectively instruct machines. The main area of difference between deep learning and ML technologies is in the presentation of the data [12]. This study makes use of Keras, one of the deep learning programming interfaces, which can handle large amounts of data and multiple layers [15]. High-level deep learning application programming interface Keras was designed with human ease of use in mind and is easy to learn. It is written in Python and may be used to construct any type of neural network. Keras is built upon several deep learning frameworks, including TensorFlow and Theano, to name only two. To enable speedy experimentation with creating deep neural networks, it emphasizes being fundamental, modular, and extensible. It also provides comprehensive, expert-level knowledge about deep learning.

5.7 TE regression

TE regression methods have been applied to predict the CS of concrete in several studies. These methods include RF [12], decision tree regressor, and RF regressor [53,54]. The gradient-boosting regression tree has also been used as an ensemble learning framework [54]. In addition, XGBoost [55], light gradient boosting machine (LightGBM) [31], and category boosting (CatBoost) have been employed. These ensemble methods have shown promising results in predicting the CS of concrete, with the LightGBM model achieving the best comprehensive performance. The RF and GB models have also demonstrated strong potential in predicting the CS of self-compacting recycled aggregate concrete. Therefore, TE regression methods have proven to be effective in predicting concrete CS [56].

5.8 Probabilistic neural networks PNN distributed delay activation (DDA)

For assessing the CS of concrete, probabilistic neural networks, also known as DDA PNNs, offer a special set of benefits. Complex nonlinear correlations between the CS of concrete and its characteristics can be captured by PNNs [57]. Compared to linear models, they can produce predictions that are more accurate because they can represent complex patterns and interactions in the data. PNNs are useful for studying a variety of concrete datasets since they can handle both continuous and categorical variables. Unlike some other neural network architectures, PNNs do not require iterative optimization, which results in a quicker training procedure. It is possible to swiftly store the training data in the network during the training phase. PNNs are appropriate for real-time applications because, once trained, they can quickly predict new concrete data. Because PNNs retain training samples directly and use similarity metrics to conduct regression or classification, they are resistant to noisy input. They are appropriate for concrete datasets that can contain measurement mistakes or outliers since they can manage noisy observations and outliers without materially compromising overall performance. PNNs provide class probabilities or predicted strength values along with related uncertainty, resulting in outputs that are easy to understand. It is simpler to explain the outcomes to stakeholders and subject matter experts thanks to the interpretability of the model, which aids in understanding its reasoning and decision-making process. PNNs are capable of transfer learning and good generalization. They can successfully apply the knowledge they have learned to new concrete strength analysis tasks or related domains after being trained on a dataset. When working with comparable concrete datasets or studying various concrete combinations, this capacity saves time and resources [58].

5.9 SRT

Numerous studies employ simple regression (SR) to forecast the CS of concrete. In a different study, an ML model for forecasting concrete CS was trained using the multiple linear regression techniques. The figure below shows how to forecast concrete's CS using ML by using SR techniques [59].

6 Model implementation and analysis of prediction results

To assess the efficacy and identify which models more accurately predict the CS of LWC, a number of statistical performance metrics have been used, such as the coefficient of determination (R^2), RMSE, MAE, Mean Absolute Percentage Error (MAPE), SI. Generally, the suggested artificial neural language models (ML) are put into practice in four stages. The gathered database is divided into training (78%) and testing (22%) datasets at random in the first step. To get rid of the scale effect, the second step is to normalize all inputs to a range of [0, 1]. Using a grid search strategy, the optimal hyperparameter values for training implementation are determined in the third stage. Tenfold cross-validation is a technique used to lessen bias resulting from random sampling of the training set. Finally, the test dataset (22%) is used to assess the model's performance, which is then quantified using the six performance indicators previously discussed. The analysis and evaluation of each model's prediction findings are shown as follows.

7 Comparison between developed models

Except for the R^2 value, which has the best value of one, the best value for all other evaluation metrics parameters is zero. It can be argued that a model performs poorly when SI > 0.3, moderately well when 0.2 < SI < 0.3, well when 0.1 < SI < 0.2, and excellently when SI < 0.1 in relation to the SI parameter [60] (Figures 10 and 11).

As Table 2 shows, the LR model performs the worst (R^2 = 0.49), while the GBT model performs the best regardless

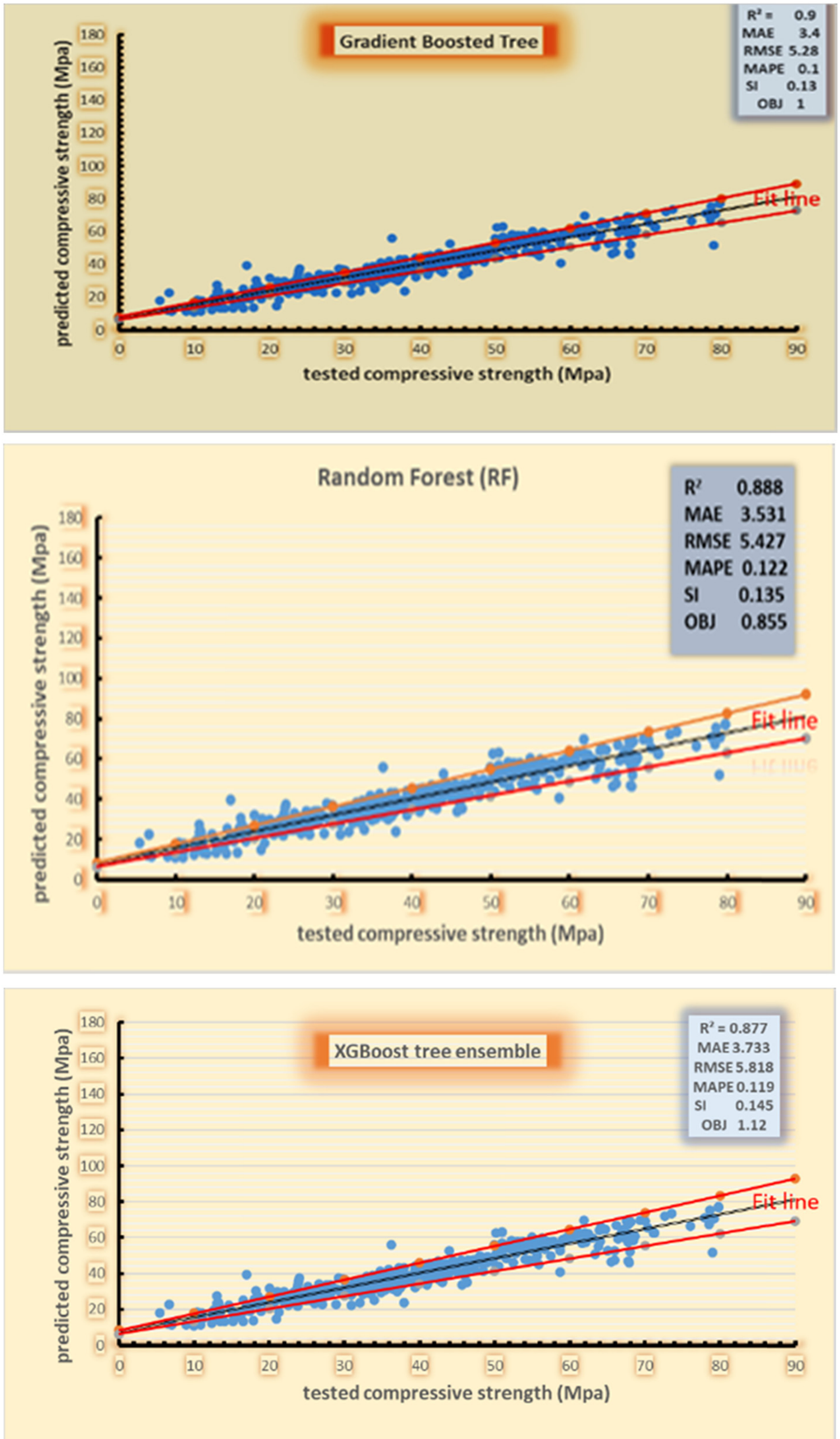

Figure 10: Model implementation and analysis of prediction results.

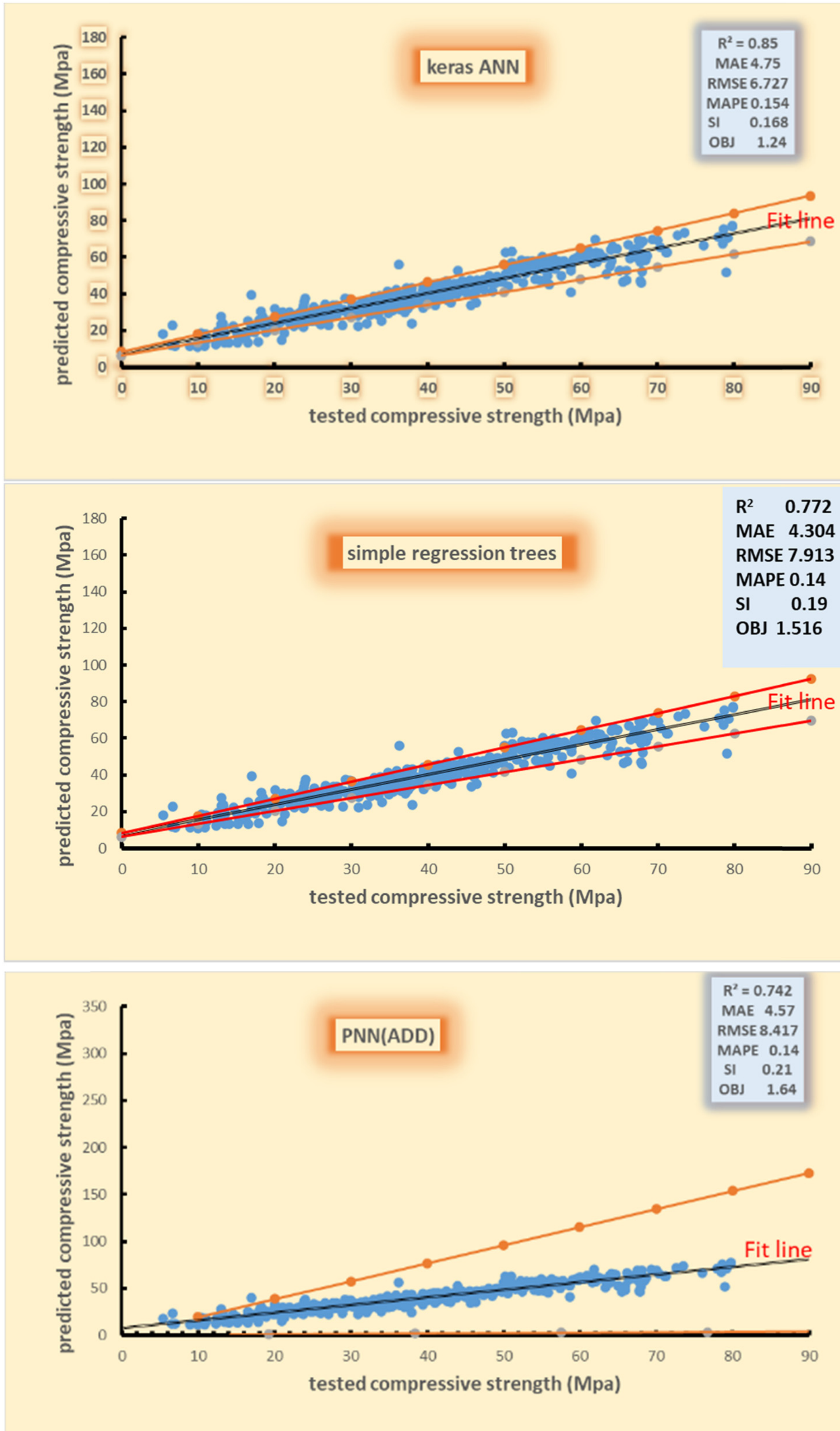

Figure 10: Continued

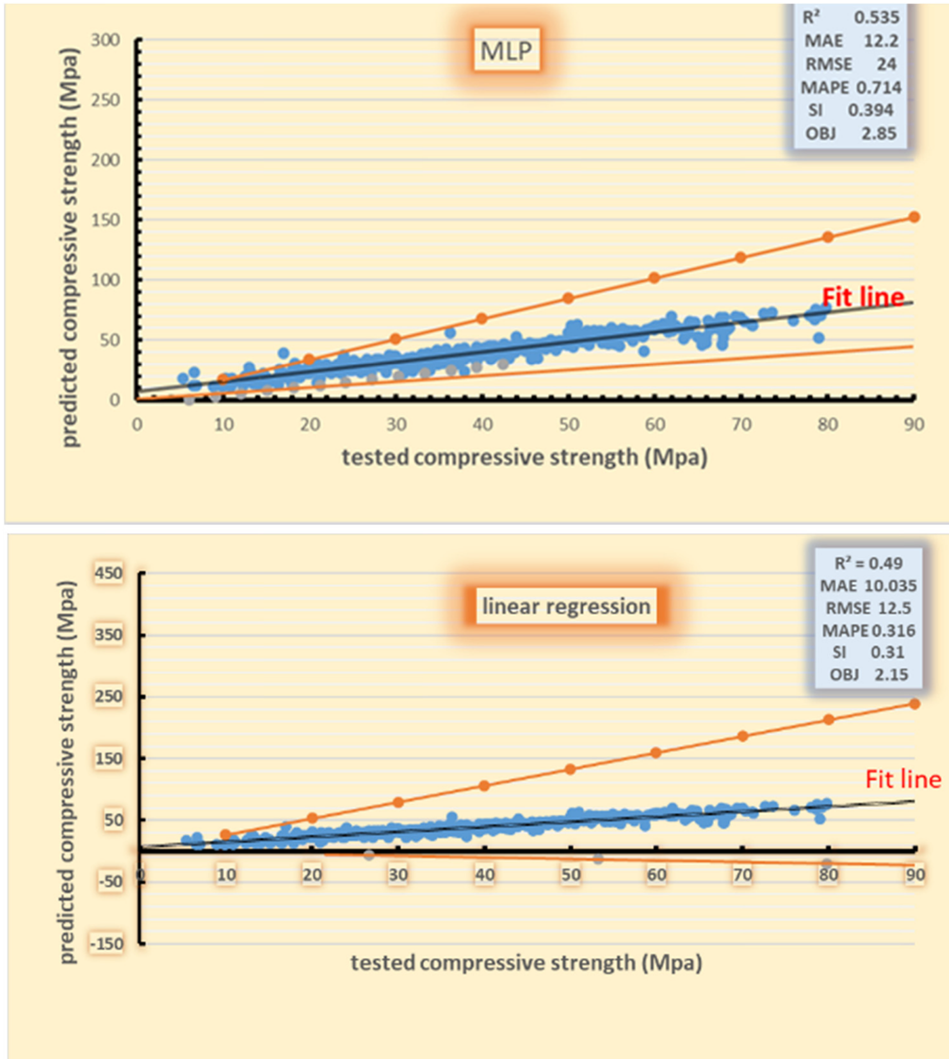

Figure 10: Continued

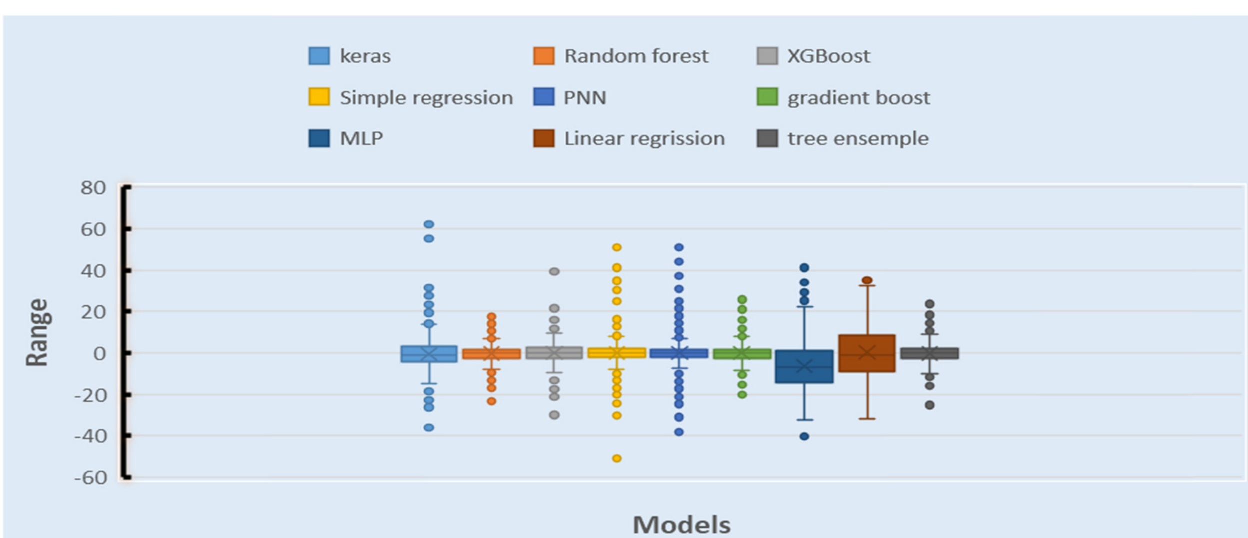

Figure 11: SI for residual errors of the developed models.

Table 2: The analysis and evaluation of each model's prediction findings

Algorithms	R^2	MAE	RMSE	MAPE	SI	OBJ
GBTs	0.9	3.4	5.28	0.1	0.13	1
RF	0.888	3.531	5.427	0.122	0.135	0.855
TE	0.885	3.84	5.62	0.136	0.14	1.1
XGBoost	0.877	3.733	5.818	0.119	0.145	1.12
Keras	0.85	0.475	6.73	0.154	0.168	1.24
SR	0.772	4.304	7.913	0.14	0.19	1.516
PNN	0.742	4.57	8.417	0.14	0.21	1.64
MLP	0.535	12.2	24	0.714	0.394	2.85
Linear regression	0.49	10.035	12.5	0.316	0.31	2.15

of the testing set used (R^2 = 0.9). It is also observed that, in order of performance, XGBoost, RF, and Keras after GBT. Reduced variation in the GBT model and the plotted data are close to the fit line, both of which indicate a smaller projected value error. As mentioned previously, five different statistical tools such as RMSE, MAE, SI, OBJ, and R^2 were implemented to assess the efficiency of the proposed models. Among the nine different models, the GBT model has higher R^2 with lower RMSE and MAE values compared to other models. Also, Table 2 presents the comparison among model estimations of CS of nanomaterial-modified LWC mixes using testing data. Moreover, it can be noticed from the figures that the predicted and measured values of CS are closer to the GBT model, which indicates the superior performance of the GBT model compared to other models. The OBJ values for all proposed models are shown in Figure 12. The values for GBT, RF, TE, XGB, Keras, SR, PNN, MLP, and LR are 1, 1.05, 1.1, 1.12, 1.24, 1.51, 1.64, 2.15, and 2.85, respectively. The OBJ value of the GBT model is lower when compared with another model. This also demonstrates that the GBT model is more efficient regarding the estimation of the CS of LWC mixtures modified with nanomaterials. The scatter interval (SI) values for GBT, RF, TE, XGBoost, Keras, SR, PNN, LR, and RPOP (MLP) are 0.132, 0.136, 0.14, 0.145, 0.168, 0.198, 0.21, 0.312, and 0.395, respectively, as shown in Figure 13. GBT demonstrated excellent performance, as indicated by its SI value, which ranged from 0.1 to 0.2. Like the other performance characteristics, the GBT model has lower SI values than the other models. The SI values are lower than the other models by 3, 6, 10, 27, 50, 59, 136, and 199%, respectively. The figures below show the results of the SI for residual errors of all constructed models.

The SI values for GBT, RF, TE, XGBoost, Keras, SR, PNN, LR, and RPOP are 0.132, 0.136, 0.14, 0.145, 0.168, 0.198, 0.21, 0.312, and 0.395, respectively, as shown in Figure 6. GBT demonstrated excellent performance, as indicated by its SI value, which ranged from 0.1 to 0.2. Like the other performance characteristics, the GBT model has lower SI values than the other models. The SI values are lower than the other models by 3, 6, 10, 27, 50, 59, 136, and 199%, respectively. The figures below show the results of the SI for residual errors of all constructed models.

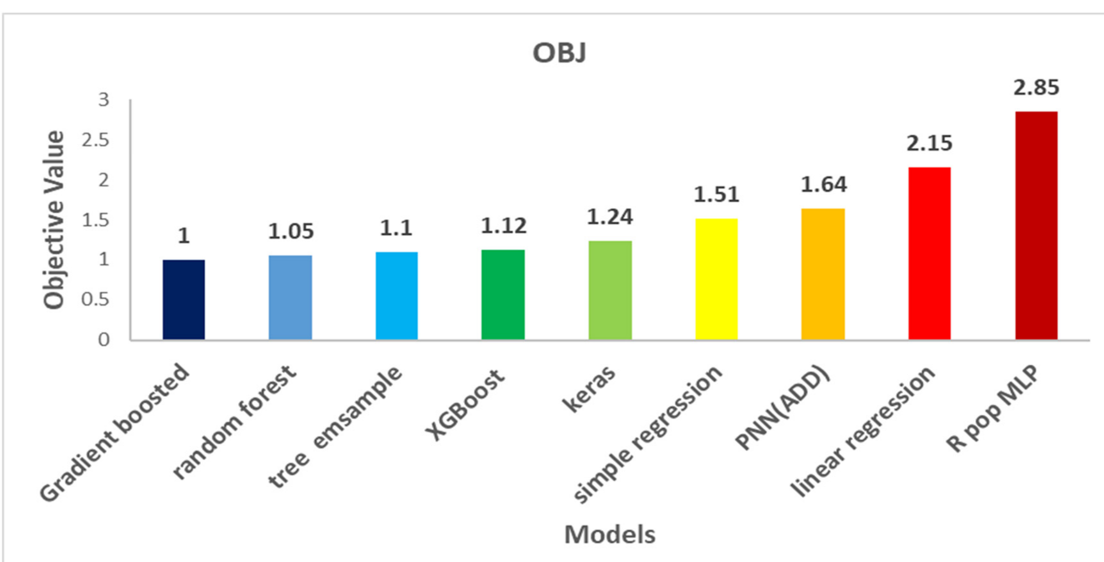

Figure 12: Comparing the OBJ performance parameters of different developed.

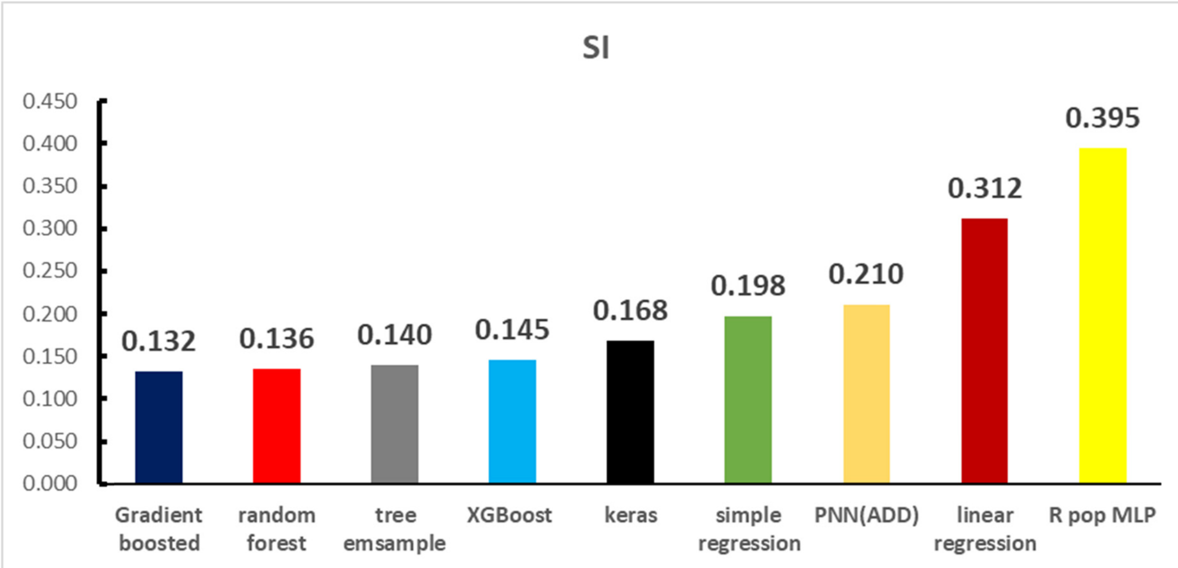

Figure 13: Comparing the SI performance parameters of different develop models.

8 Statistical assessment of design parameters with compressive strength

In this section, statistical analysis has been done to show whether the CS has any meaningful relationships with the input parameters. To achieve this, all nine parameters were plotted and evaluated using the actual CS to determine the extent to which each variable affects the CS. The correlation data between each variable used as an input parameter and CS are shown in the following figures.

8.1 Relationship between cement content and compressive strength

Figure 14 illustrates the nonlinear connection between (cement content) and (CS) based on the dataset, which

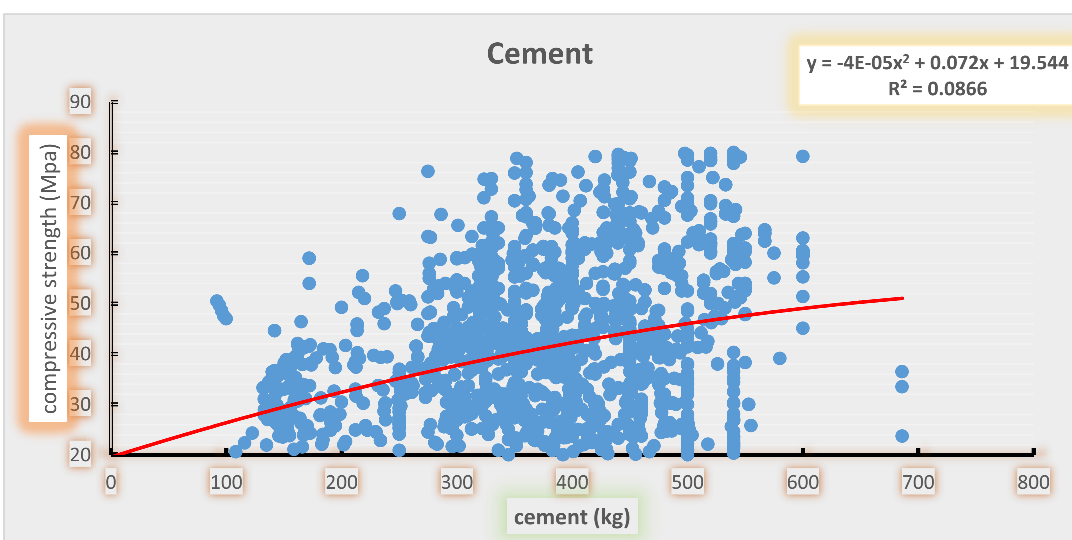

Figure 14: Relationship between cement content and compressive strength.

includes 1,968 samples. The findings show that the (cement content) value is useful and practicable between 250 and 550 kg/m^3, after which it starts to decline since a higher cement dose reduces CS [61]. As a result, the value of R^2 = 0.0866 indicates that the entire collection of data has a negligible relationship to the CS.

8.2 Relationship between sand and compressive strength

According to the collocated data distribution, it was found that the relationship between sand and CS second-degree nonlinear relation as illustrated in Figure 15. It is evident from the preceding data that the effective sand content ranges from 600 to 930 kg/m^3. When the amount is exceeded above the limit, the concrete loses some of its CS; when it is decreased, the concrete becomes less fluid and encourages the settling of coarse material, which further deteriorates the concrete. As a result, the value of R^2 = 0.0336 indicates that the entire collection of data has a negligible relationship to the CS.

8.3 Relationship between normal aggregate and compressive strength

Figure 16 shows that the weight of the impact aggregate is between 500 and 1,200 kg/m^3 and that the relationship

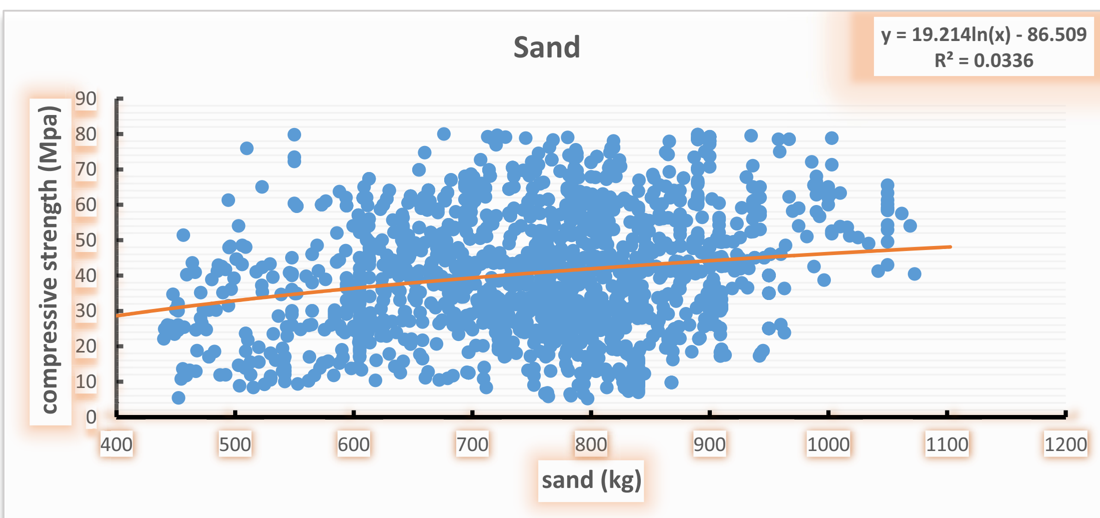

Figure 15: Relationship between sand and compressive strength.

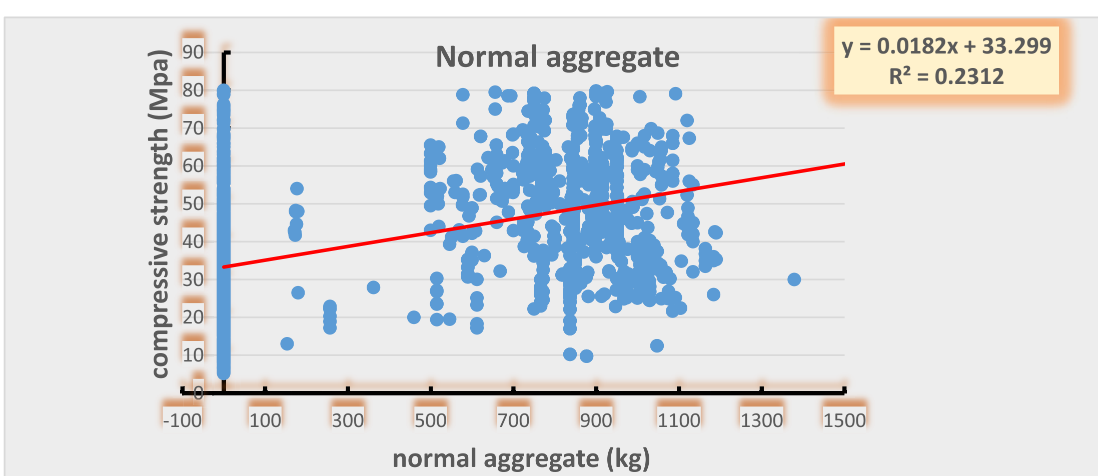

Figure 16: Relationship between normal aggregate and compressive strength.

between normal coarse aggregate and CS is linear. The figure shows that the relationship between ordinary coarse aggregate and CS is linear and that the coefficient of determination R^2 was 0.23. This indicates the presence of an average effect of ordinary coarse aggregate on CS.

8.4 Relationship between lightweight aggregate and compressive strength

The following figure illustrates the significant impact of lightweight coarse aggregate on CS at quantities ranging from 370 to 800 kg/m^3, demonstrating a second-degree nonlinear relationship between the two. The data in the following figure indicate that the coefficient of determination R^2 was generally weak (0.0689), but it is effective and affects the CS when the aggregate is employed in a studied quantity between 370 and 800 kg/m^3 (Figure 17).

8.5 Relationship between water and compressive strength

The distributed data, as shown in the Figure below, indicates that the relationship between the water and CS. The results show that the relationship is nonlinear and second order and that the effect of water is significant when its quantity is between 120 and 180 l/m^3. The results show that the coefficient of determination R^2 is 0.073, meaning that the relationship exists, but it increases if the amount of water is within the optimal limits between 120 and 180 l/m^3 (Figure 18).

8.6 Relationship between water/cement ratio and compressive strength

The distributed data, as shown in Figure 19, indicates that the relationship between the water–cement ratio and the CS is a linear inverse relationship. The data showed that increasing the water–cement ratio causes a decrease in CS, and this is what previous studies have proven. The coefficient of determination R^2 is 0.0684, and the best CS can be obtained when the water-to-cement ratio is 0.35–0.55.

8.7 Relationship between density and compressive strength

Density data showed that the relationship between it and the CS is nonlinear, but rather a second-degree equation and that increasing it leads to an increase in the CS. Density data showed that it is effective when it is between 1,500 and 2,400 kg/m^3. The value of the coefficient of determination R^2 was equal to 0.2 (Figures 20 and 21).

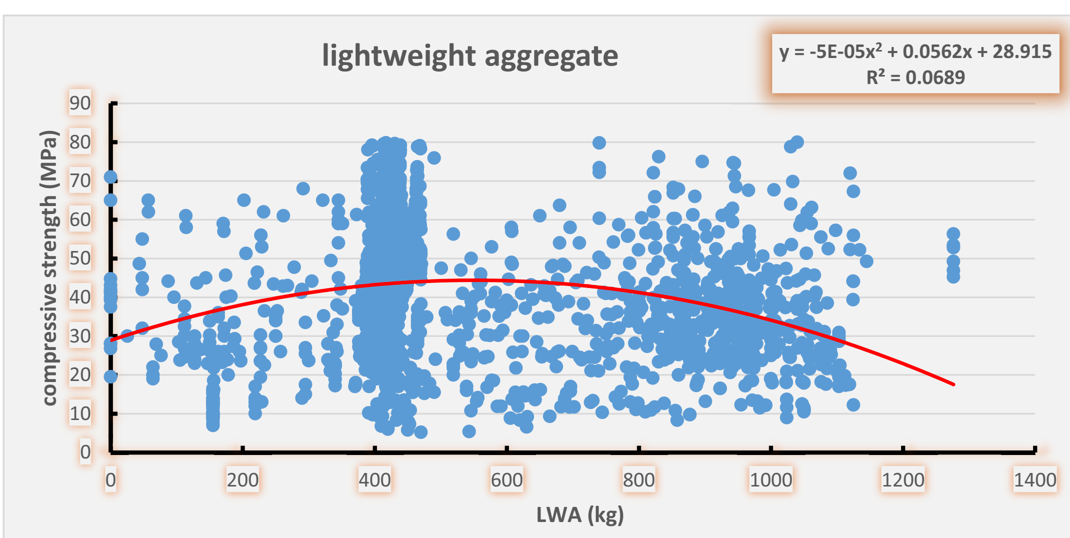

Figure 17: Relationship between lightweight aggregate and compressive strength.

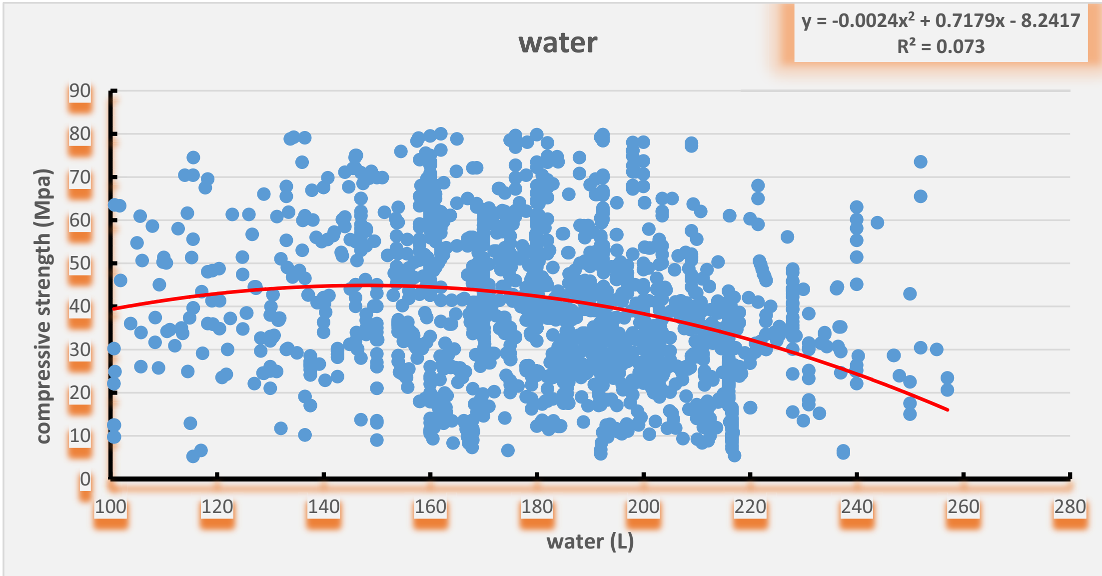

Figure 18: Relationship between water and compressive strength.

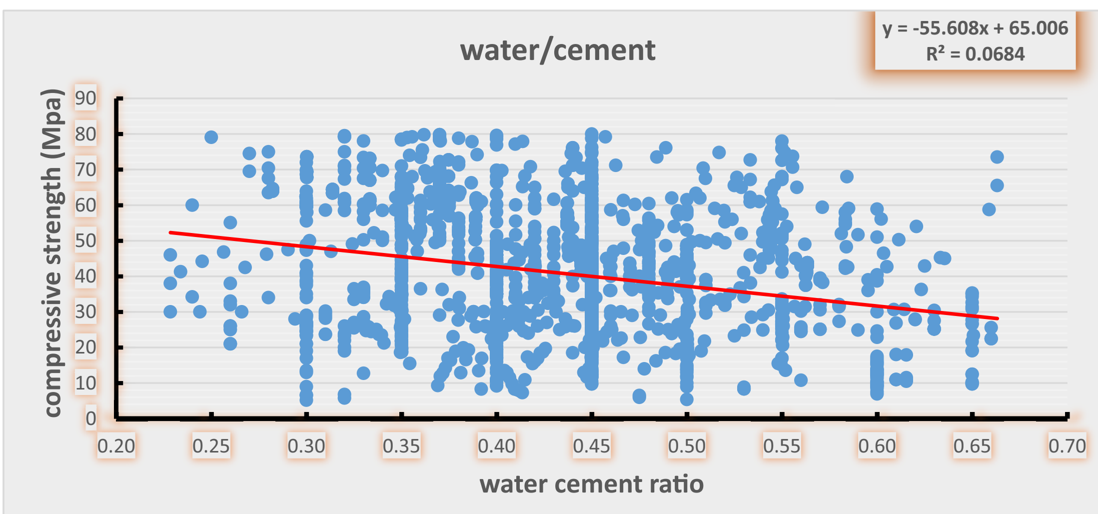

Figure 19: Relationship between water/cement ratio and compressive strength.

9 Relationship between % of nanomaterial and compressive strength

9.1 Validation of the superior ML model

The statistical tests discussed earlier demonstrated that the GBT approach outperformed other ML models in terms of accuracy, error tolerance, and performance. The model's robustness, precision, and effectiveness were then confirmed through testing and validation on a new experimental database of 14 samples and 9 CS-affecting characteristics. This group represents laboratory work done in Iraq's Diyala University laboratories. Light-weight concrete with lightweight porcelainite stone was employed. To meet American regulations ACI211.2, which state that the CS cannot be less than 17 mega-Pascals and the density cannot be greater than 2,000 kg/m^3, it was built with a 17.3 mega-Pascal CS [62]. The mechanical characteristics and CS of LWC were improved by adding two nanomaterials (NS

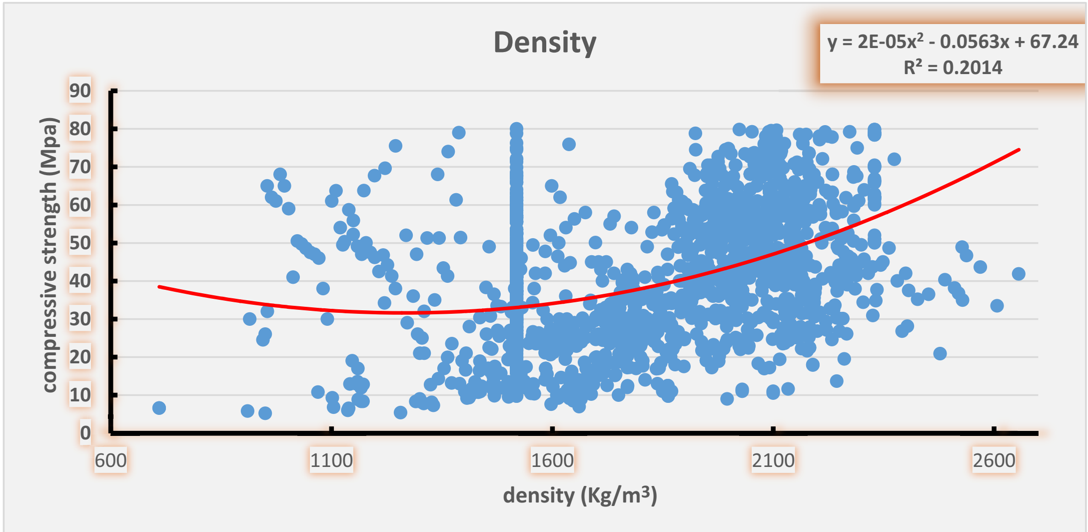

Figure 20: Relationship between density and compressive strength.

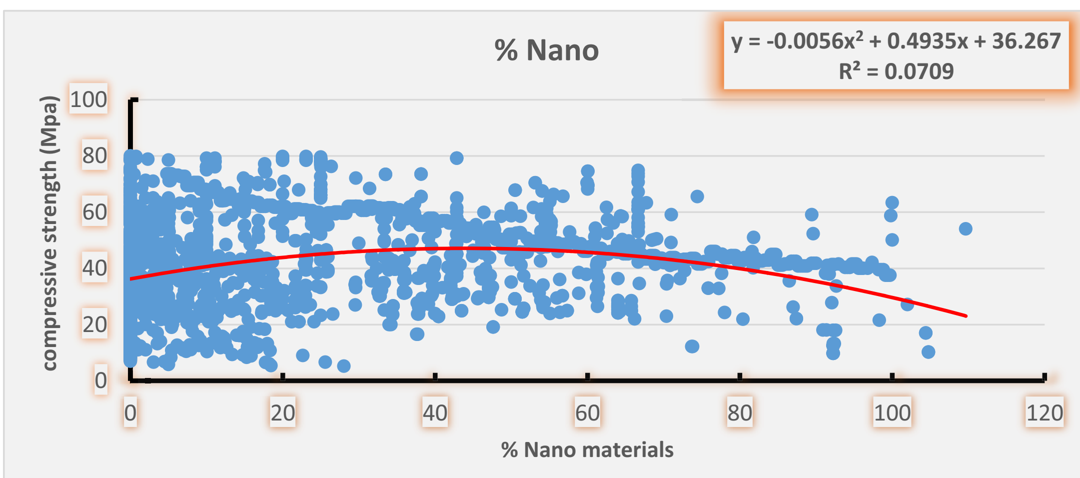

Figure 21: Relationship between % of nanomaterial and compressive strength.

and NMK) in varying amounts to suit all building components. Compressive strength improved significantly for all mixes at 28 days of age, increasing by two times or more, according to the results. This database had not been explicitly utilized in ML methods for the aforementioned model. In the sense that, as opposed to using this specific set of data, the model was trained using different sets of data for testing and training. Fourteen datasets received preprocessing before being fed into the model, as shown in the flowchart of all operations performed on the data from preprocessing and eventually provided to the model in Figure 22.

Based on actual data gathered from laboratory testing, the results demonstrated a high degree of accuracy in CS prediction (R^2 = around 98%). This great precision indicates that the concrete produced in the laboratory meets the intended objective. The NMK and NS dosage ratios utilized in laboratory work are not exactly the same as those used in earlier research, but by applying the GBT algorithm, it is now simple to determine which CS is needed and what proportion is used. This eliminates the need to pour new light concrete into the lab, which adds to the expense, time, and material requirements. This is intended for the purpose of utilizing the artificial intelligence application KNIME. Figure 23 shows the relationship between the experimental actual CS results and the predicted CS.

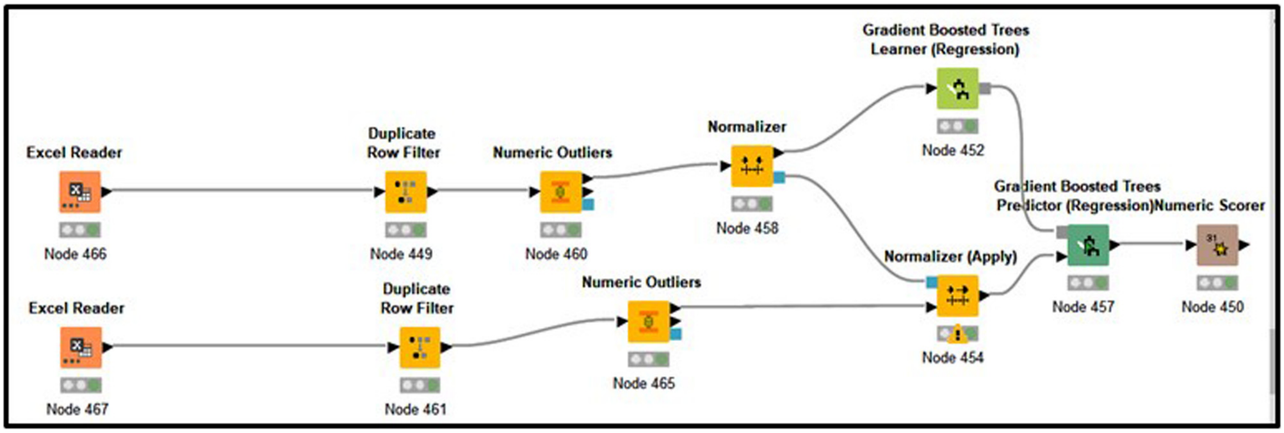

Figure 22: Flowchart of all operations performed on the data from preprocessing and eventually provided to the model.

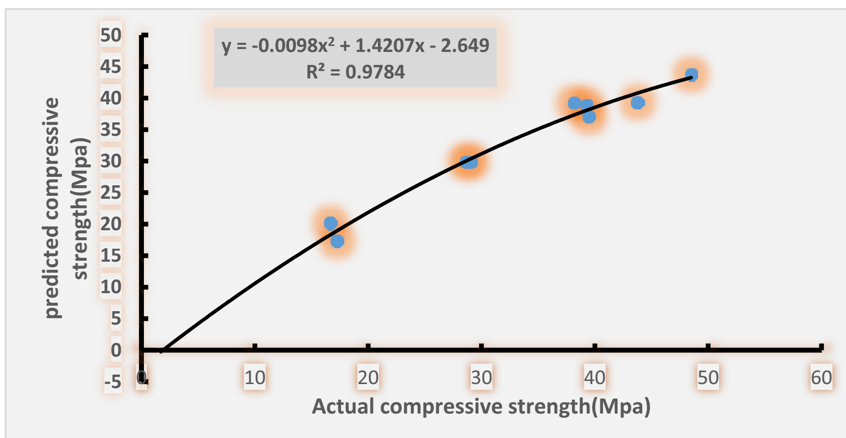

Figure 23: Relationship between the experimental (actual) CS results and the predicted CS.

10 Sensitivity analysis

To find and assess the most impacting variable that influences the CS of LWC mixtures modified with nanomaterials, a sensitivity comparison was performed for the models [50]. The most efficient model, GBT, was chosen for the sensitivity analysis. During the sensitivity analysis, several different training data sets were used, and for each set, a single input variable was extracted at a time. The assessment parameters such as R^2, RMSE, and MAE were calculated for each training (Figure 12). The OBJ values for all developed models (Figure 13) are compared with the SI performance parameters of different developed model datasets independently. The results of the sensitivity analysis are reported in Table 3. It is obvious from the results that water is the most important and influencing variable for the CS prediction of LWC mixtures modified with nanomaterials. In this study, the water for the obtained data was ranged from 100.2 to 257 L/m^3. This can be approved by almost all experimental results collected from previous studies. Figure 24 shows

Table 3: Comparison of developed model results for different ML models

Sr. No	Removed parameter	R^2	RMSE	Ranking
1	None	0.9	**5.286**	None
2	Water	0.769	**8.024**	1
3	Density	0.79	**7.657**	2
4	Sand	0.81	**7.443**	3
5	*W/C*	0.833	**6.862**	4
6	% of nano	0.846	**6.63**	5
7	Lightweight aggregate	0.846	6.525	6
8	Cement	0.862	6.052	6
9	Type of nano	0.87	5.973	7
10	Normal aggregate	0.872	5.934	8

Bold values represent the order of factors from the most influencing the compressive strength to the least

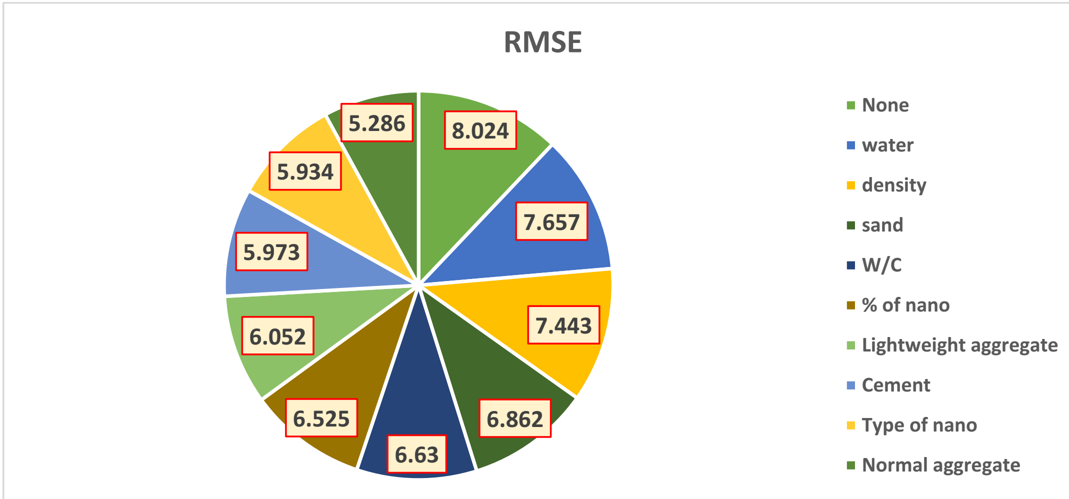

Figure 24: Sensitivity analysis using GBT model.

	Cement	Sand	Normal aggregate	LWC	Water	W/C	Density	Type of nano	% of nano	Compressive
Cement	1									
Sand	-0.17	1								
Normal aggregate	0.20	0.19	1							
LWC	-0.46	-0.16	-0.59	1						
Water	0.10	-0.22	-0.31	0.14	1					
W/C	-0.48	0.18	0.12	-0.04	0.21	1				
Density	0.32	0.24	0.72	-0.50	-0.16	0.01	1			
Type of nano	-0.46	0.26	0.08	0.37	0.04	0.23	-0.05	1		
% of nano	-0.36	0.15	0.35	-0.07	-0.24	0.28	0.15	0.35	1	
Compressive	0.31	0.21	0.53	-0.21	-0.22	-0.19	0.47	0.10	0.13	1

Figure 25: Correlation matrix for input variables and target (output).

the sensitivity analysis using the GBT model. Table 3 and Figure 24 display the results of the essential variables sensitivity analysis (Figure 25).

11 Results and discussion

11.1 Main findings of the present study

Through comparison between the results obtained using the nine algorithms for the same samples collected (2,568), and through the five statistical performance measures (R^2, MAE, MAPE, RMSE, and SI), it was found that the best performance obtained was for GBTs, where R^2 was 0.9, MRSE was 5.28, MAE was 3.4, MAPE was 0.1, and SI was 0.132. These results indicate accuracy and reliability, as when compared to the real results, they are very close to the optimal results line, and the value has little dispersion, as shown in Figure 10, in contrast to the rest of the algorithms, in which the accuracy (R^2) decreased from 0.888 to 0.49.

The superiority of GBTs over other algorithms in accuracy in the process of predicting compressive resistance is due to the accuracy of dealing with different inputs and building models that simulate reality [63].

11.2 Comparison with other studies

Through the samples he collected from previous studies from reliable sources, and as shown in the review section of previous studies and the table, it was found that there is a clear weakness in the number of data he collected and used to predict CS. In addition, the number of algorithms used ranged from three to seven algorithms. Moreover, not using a sufficient number of parameters affects the CS. Also, there was a lack of clarity in the data on the use of nanomaterials to improve CS because most studies used one or two nanomaterials and were analyzed. In this article, approximately 2,568 samples were collected from reliable previous studies, sorted, and purified from missing and duplicate information. In addition, nine parameters affecting CS were identified, and the relationship between them and CS was studied once and between each parameter and another time. In addition, the effect of these parameters on lightweight and ordinary concrete was studied. It is worth noting that this study included an analysis of evidence of concrete containing many types of nanomaterials in multiple proportions to give a clear picture of the effect of nanomaterials on concrete. As for the number of algorithms used, nine algorithms that were known in previous studies in predicting the CS of concrete were chosen, to demonstrate the best, most accurate, and easiest for all users, even if they are not proficient in programming.

11.3 Implication and explanation of findings

Using a large number of data and excluding duplicates and missing ones gave the KNIME statistical program the power to analyze the results with high accuracy. In addition, KNIME is easy to use, has an open-source platform for data analysis, and makes the program's source code freely available, allowing anyone to access, modify, and distribute the system. In addition, choosing the GBT algorithm as the best and most accurate algorithm was consistent with many researchers, and this in itself gave high reliability to the results.

11.4 Strengths and limitations

The strength of this work is summarized in the use of the KNIME program in data analysis, as it has ease of use, access to its updates, accurate analysis, and ease of dealing with many of the important algorithms used in predicting the CS of concrete, without complexity when compared to some codes and software such as Python, Matlab, and others. The sufficient number of data also played an effective role in giving accuracy to the results that were predicted when compared with real data obtained from laboratory work. Increasing the number of algorithms and parameters and including data for most nanomaterials used to improve the CS of concrete gave a good picture for predicting CS.

The limitations of work could be in data quality and availability: To train ML models, a large, representative dataset is needed. Nevertheless, it can be difficult to obtain big and varied information for predicting concrete strength. To train accurate models, high-quality data with precise and trustworthy measurements must be available. Predictions that are skewed or poorly generalized to new contexts can result from incomplete or biased data.

Limited understanding of causality: ML models are very good at identifying patterns and correlations in data. They might not, however, offer an in-depth understanding of the fundamental causal connections between input factors and CS. Models are useful for identifying traits, but they might not reveal the underlying scientific ideas that underpin concrete strength. This restriction makes it more difficult for the models to be applied to novel materials or circumstances that were not included in the training set.

12 Conclusion

Because it depends on the kind, quantity, and homogeneity of its elements as well as on compaction, curing time, environmental factors, and equipment employed, estimating the CS of concrete is a complicated task. The difficulty of the issue is further increased by adding nanoparticles to the concrete mixture. Applying a strong ML model to this intricate issue could yield dependable outcomes and ultimately lower the expenses associated with traditional approaches. This work used the KNIME platform to employ several ML techniques, including GBTs, RF, XGBoost, PNN, RProp MLP, ANN, SR, and LR, on a dataset of 2,568 samples that were taken from 152 peer-reviewed academic publications.

1. Out of all the ML models that were assessed, GBT and RF produced R^2 scores of 0.9 and 0.89, respectively, and were the most accurate in predicting the CS of LWC with NS and NMK. With fewer than 10% prediction error, CS can be predicted in both models. These ratings are considered robust given the difficulty of the challenge and the unpredictability of the input parameters.

2. As anticipated, the LR model showed the lowest dependability due to the nonlinear relationship between the input parameters and CS.
3. After examining the data, it was discovered that the GBTs approach is among the finest models that were employed since it produced the best coefficient of determination (R^2 = 90%), MAE of 3.4, RMSE = 5.28, and MAPE = 0.1, or less error rate. When compared to the other methods, the dispersion coefficient's value was the lowest, with SI = 1.3
4. It was determined that the Water was the predominant variable influencing the CS, followed by the density and sand. Comparing the results from this study with other models in the literature, it becomes evident that the dataset used significantly influences the identification of the most critical factor affecting the CS.
5. The QBT model's link between CS and nanomaterial percentage indicates that a dose of roughly (0–20)% for nanomaterials is the best amount to obtain the required characteristics. The efficacy of nanomaterials appears to decrease beyond this point, presumably as a result of workability and dispersion issues. The dataset obtained from the literature and the outcomes of the experimental work were used to create the machine-learning models in this study. An even bigger data set is needed to create models that are more trustworthy and robust. Because of this, data were gathered for a sample size greater than 2,500, since prior research has shown that the more data entered, the more accurate the conclusions and predictions. The purpose of this study was to aid in data collection and the creation of more reliable machine-learning models for the prediction of concrete mixture CS. To move further, ML modeling approaches should also be used to study how nanoparticles affect the mechanical qualities of LWC, such as CS.
6. The findings, which were supported by earlier research, indicated that the amount of water and the ratio of water to cement had an inverse influence on CS.
7. The results showed that with increasing density, the CS increases, meaning that density is directly proportional to the resistance, and this is what has been proven in previous studies.
8. The best cement dosage, according to the findings, was between 290 and 525 kg/m^3.
9. The results clearly show that the most sensitive and influential variable for the prediction of CS is water, followed by density, sand, and the *W*/*C* ratio, which is indicated in blue. The lightweight aggregate and percentage of nano have a mild impact. The remaining factors, which included cement, nano type, and regular aggregate, all showed negligible effects and some had no effect on the anticipated CS; their respective R^2 and RMSE values were 5.9 and 0.87.

13 Recommendations

Based on the work that has been carried out in this study on the use of different modeling techniques to forecast the CS of LWC modified with nanomaterials, the scope and gaps for further studies have been discussed and highlighted in the following:

(a) Use of these modern techniques and input of all variable parameters that can affect the CS.

(b) Developing empirical models to predict the CS of different types of LWC.

(c) Using these model techniques to propose empirical equations for other mechanical properties of LWC composites like splitting tensile strength, flexural strength, and modulus of elasticity.

(d) Conducting laboratory experiments to validate the developed models.

(e) Take benefits from these models and other intelligence techniques to standardize the mix design of LWC composites just like traditional concrete.

14 Future work

The GBT method is used in this article to forecast the CS of concrete materials. The test variables are separated into nine inputs (such as cement, water, additives, coarse/fine aggregates, etc.) and one output (CS value) after 2,568 sets of concrete compressive tests are gathered. A training set and a testing set are created by further dividing the entire data set. The training set's GBoosted trees create the model, which is subsequently assessed by the testing set. The following can be inferred from the results for future work:

1. Collect more data to introduce new parameters that affect CS.
2. Using the same algorithms to predict some other properties of concrete, such as tensile and bending strength.
3. Trying to collect all published data in a unified database so that all researchers can view and use it.
4. Conducting a comprehensive introduction to the KNIME program for ease of use by researchers and providing the necessary data for this.

Acknowledgments: The authors acknowledge his gratitude to the University of Diyala, especially the Department of

Civil Engineering, for providing the laboratories and equipment to complete experimental work.

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

Data availability statement: Data available under request.

References

[1] Neville AM, Brooks JJ. Civil_Engineering_Materials.Pdf. England: University of Leeds; 1987.
[2] Scrivener KL, Kirkpatrick RJ. Innovation in use and research on cementitious material. Cem Concr Res. 2008;38(2):128–36. doi: 10.1016/j.cemconres.2007.09.025.
[3] Zhang M, Li M, Shen Y, Ren Q, Zhang J. Multiple mechanical properties prediction of hydraulic concrete in the form of combined damming by experimental data mining. Constr Build Mater. 2019;207:661–71. doi: 10.1016/j.conbuildmat.2019.02.169.
[4] Terms RT, Publications NRC. Archive archives des publications du CNRC Effect of temperature on thermal properties of high-strength concrete. J Mater Civ Eng. 2003-03-01 Web, 1561, no. April, 2003;15(2):101–7. doi: 10.1061/(ASCE)0899-1561(2003)15.
[5] Kim JK, Hun Han S, Kyun Park S. Effect of temperature and aging on the mechanical properties of concrete: Part II. Prediction model. Cem Concr Res. 2002;32(7):1095–100. doi: 10.1016/S0008-8846(02)00745-7.
[6] Li Z, Yoon J, Zhang R, Rajabipour F, Srubar III WV, Dabo I, et al. Machine learning in concrete science: applications, challenges, and best practices. npj Comput Mater. 2022;8(1):127. doi: 10.1038/s41524-022-00810-x.
[7] Young BA, Hall A, Pilon L, Gupta P, Sant G. Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods. Cem Concr Res. 2019;115:379–88. doi: 10.1016/j.cemconres.2018.09.006.
[8] Ben Chaabene W, Flah M, Nehdi ML. Machine learning prediction of mechanical properties of concrete: Critical review. Constr Build Mater. 2020;260:119889. doi: 10.1016/j.conbuildmat.2020.119889.
[9] Albert C, Isgor OB, Angst U. Exploring machine learning to predict the pore solution composition of hardened cementitious systems. Cem Concr Res. 2022;162:107001. doi: 10.1016/j.cemconres.2022.107001.
[10] Zhou ZH. Open-environment machine learning. Natl Sci Rev. 2022;9(8):nwac123. doi: 10.1093/nsr/nwac123.
[11] Dai L, Wu X, Zhou M, Ahmad W, Ali M, Sabri MM, et al. Using machine learning algorithms to estimate the compressive property of high strength fiber reinforced concrete. Materials. 2022;15(13):1–19. doi: 10.3390/ma15134450.
[12] Han T, Siddique A, Khayat K, Huang J, Kumar A. An ensemble machine learning approach for prediction and optimization of modulus of elasticity of recycled aggregate concrete. Constr Build Mater. 2020;244:118271. doi: 10.1016/j.conbuildmat.2020.118271.
[13] Behnood A, Golafshani EM. Machine learning study of the mechanical properties of concretes containing waste foundry sand. Constr Build Mater. 2020;243:118152. doi: 10.1016/j.conbuildmat.2020.118152.
[14] Huang Y, Zhang J, Tze Ann F, Ma G. Intelligent mixture design of steel fibre reinforced concrete using a support vector regression and firefly algorithm based multi-objective optimization model. Constr Build Mater. 2020;260:120457. doi: 10.1016/j.conbuildmat.2020.120457.
[15] Maherian MF, Baran S, Bicakci SN, Toreyin BU, Atahan HN. Machine learning-based compressive strength estimation in nano silica-modified concrete. Constr Build Mater. 2023;408:133684. doi: 10.1016/j.conbuildmat.2023.133684.
[16] Zaman A, Alyami M, Shah S, Rehman MF, Hakeem IY, Farooq F. Forecasting the strength of micro/nano silica in cementitious matrix by machine learning approaches. Mater Today Commun. 2023;37:1–21. doi: 10.1016/j.mtcomm.2023.1-21.
[17] Kioumarsi M, Dabiri H, Kandiri A, Farhangi V. Compressive strength of concrete containing furnace blast slag; optimized machine learning-based models. Clean Eng Technol. 2023;13:100604. doi: 10.1016/j.clet.2023.100604.
[18] Shahmansouri AA, Yazdani M, Hosseini M, Akbarzadeh Bengar H, Farrokh Ghatte H. The prediction analysis of compressive strength and electrical resistivity of environmentally friendly concrete incorporating natural zeolite using artificial neural network. Constr Build Mater. 2022;317:no. May 2021 125876. doi: 10.1016/j.conbuildmat.2021.125876.
[19] Alshammari S. The effect of Nano-Silica on porosity and strength. Energies. 2018;6(1):1–8. http://journals.sagepub.com/doi/10.1177/1120700020921110%0Ahttps://doi.org/10.1016/j.reuma.2018.06.001%0Ahttps://doi.org/10.1016/j.arth.2018.03.044%0Ahttps://reader.elsevier.com/reader/sd/pii/S1063458420300078?token=C039B8B13922A2079230DC9AF11A333E295FCD8.
[20] Surono S, Afitian MYF, Setyawan A, Arofah DKE, Thobirin A. Comparison of CNN classification model using machine learning with bayesian optimizer. HighTech Innov J. 2023;4:531–42.
[21] Tawfik TA. Hybrid Effect of Nano Silica and Metakaolin on Mechanical Properties of Cement Mortar. Int J Eng Res. 2019;V8(03):211–5. doi: 10.17577/ijertv8is030145.
[22] Srinivas K. Nanomaterials for concrete technology. Int J Civil Struct Environ Infrastruct Eng Res Dev. 2014;4:79–90.
[23] Taher ZK, Ismael MQ. Rutting prediction of hot mix asphalt mixtures modified by nano silica and subjected to aging process Zainab. Civ Eng J. 2023;9:1–14.
[24] Sikora P, Rucinska T, Stephan D, Chung S-Y, Elrahman MA. Evaluating the effects of nanosilica on the material properties of lightweight and ultra-lightweight concrete using image-based approaches. Constr Build Mater. 2020;264:1–15.
[25] Rao DVP, Maruthi SV. Effect of nano-silica on concrete containing metakaolin. Int J Civ Eng Technol. 2016;7(1):104–12.
[26] Lai S, Serra M. Concrete strength prediction by means of neural network. Constr Build Mater. 1997;11(2):93–8. doi: 10.1016/S0950-0618(97)00007-X.
[27] Kewalramani MA, Gupta R. Concrete compressive strength prediction using ultrasonic pulse velocity through artificial neural networks. Autom Constr. 2006;15(3):374–9. doi: 10.1016/j.autcon.2005.07.003.
[28] Naderpour H, Rafiean AH, Fakharian P. Compressive strength prediction of environmentally friendly concrete using artificial neural networks. J Build Eng. 2018;16:213–9.

[29] Morsy MS, Alsayed SH, Aqel M. Hybrid effect of carbon nanotube and nano-clay on physico-mechanical properties of cement mortar. Constr Build Mater. 2011;145–9.
[30] Wang W-C, Wang S-Y, Lin C-H. Effect of hot environment on strength and heat transfer coefficient of nano-clay concrete paper title. J Mater Sci Chem Eng. 2016;4(7):45–52. doi: 10.4236/msce.2016.47007.
[31] Kurihara R, Maruyama I. Impact of TiO_2 nanoparticles on drying shrinkage of hardened cement paste. J Adv Concr Technol. 2018.
[32] Maqbool M, Awan S, Soroushian P, Ali A, Yousaf M, Awan S. Carbon nano fibre reinforcements in concrete muhammad. Indones J Sci Technol. 2019;4(1):1–16.
[33] Nivethitha D, Dharmar S. Effect of zinc oxide nanoparticle on strength of cement mortar. Int J Sci Technol Eng. 2016;3(5).
[34] Wu X, Zheng S, Feng Z, Chen B, Qin Y, Xu W, et al. Prediction of the frost resistance of high-performance concrete based on RF-REF: A hybrid prediction approach. Constr Build Mater. 2022;333:1–13.
[35] Ortenzi MA, Basilissi L, Farina H, Di Silvestro G. Evaluation of crystallinity and gas barrier properties of films obtained from PLA nanocomposites synthesized via '"in situ"' polymerization of L-lactide with silane-modified nanosilica and montmorillonite. Eur Polym J. 2015.
[36] Farooq F, Nasir Amin M, Khan K, Rehan Sadiq M, Javed MF, Aslam F, et al. A comparative study of random forest and genetic engineering programming for the prediction of compressive strength of high strength concrete (HSC). Appl Sci (Switz). 2020;10(20):1–18. doi: 10.3390/app10207330.
[37] Feng D-C, Liu ZT, Wang XD, Chen Y, Chang JQ, Wei DF, et al. Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. Constr Build Mater J. 2020;230:1–11.
[38] Li J. Assessing the accuracy of predictive models for numerical data: Not r nor r2, why not? Then what? PLoS ONE. 2017;12(8):1–16. doi: 10.1371/journal.pone.0183250.
[39] Onyutha C. From R-squared to coefficient of model accuracy for assessing 'goodness-of-fits. Geosci Model Dev. 2020;2020:1–25. doi: 10.5194/gmd-2020-51.
[40] Minaz Hossain M, Nasir Uddin M, Abu Sayed Hossain M. Prediction of compressive strength ultra-high steel fiber reinforced concrete (UHSFRC) using artificial neural networks (ANNs). Mater Today: Proc. 2023;1–9. doi: 10.1016/j.matpr.2023.02.409.
[41] Abualkibash M. Machine learning in network security using knime analytics. Int J Netw Secur Its Appl. 2019;11:1–14.
[42] Vu VT. Prediction of the slump and strength of high strength concrete using random forest model. J Sci Tech. 2020;6:113–45.
[43] Wang K, Ren J, Yan J, Wu X, Dang F. Research on a concrete compressive strength prediction method based on the random forest and LCSSA-improved BP neural network. J Build Eng. 2023;76:1–15.
[44] Xu B, Tan Y, Sun W, Ma T, Liu H, Wang D. Study on the prediction of the uniaxial compressive strength of rock based on the SSA-XGBoost Model. Sustainability. 2023;1–17.
[45] Ali I, Suthar M. Comparison between Random forest and M5P to predict the compressive strength of concrete modified with solid wastes. Punjab, India: Chandigarh University Mohali; 2023.
[46] Al-Abdaly NM, Al-Taai SR, Imran H, Ibrahim M. Development of prediction model of steel fiber-reinforced concrete compressive strength using random forest algorithm combined with hyperparameter tuning and k-fold cross-validation. Najaf, Iraq: Al-Furat Al-Awsat Technical University Najaf Munazira str.; 2021.
[47] Alabi SA, Mahachi J. Utilizing artificial neural network and multiple linear regression to model the compressive strength of recycled geopolymer concrete. Int J Integr Eng. 2022;14(4):43–56. doi: 10.30880/ijie.2022.14.04.005.
[48] Imran H, Al-Abdaly NM, Shamsa MH, Shatnawi A, Ibrahim M, Ostrowski KA. Development of prediction model to predict the compressive strength of eco-friendly concrete using multivariate polynomial regression combined with stepwise method.pdf. Mater (Basel). 2021;1–15.
[49] Dahish HA, Almutairi AD. Effect of elevated temperatures on the compressive strength of nano-silica and nano-clay modified concretes using response surface methodology. Case Stud Constr Mater. 2023;18(March):e02032. doi: 10.1016/j.cscm.2023.e02032.
[50] Liu G, Sun B. Concrete compressive strength prediction using an explainable boosting machine model. Case Stud Constr Mater. 2023;18:1–17.
[51] Falah MW, Hussein SH, Saad MA, Ali ZH, Tran TH, Ghoniem RM, et al. Compressive strength prediction using coupled deep learning model with extreme gradient boosting algorithm: environmentally friendly concrete incorporating recycled aggregate. Hillah, Iraq: AL-Mustaqbal University College; 2022. pp. 1–22.
[52] Dubey A, Rasool A. Usage of deep learning in recent place its title, and an unchanged title of a short paper should be given in the place foreseen for the paper title. Moreover, it is necessary to write here 'research or technical paper in a full version' applications. Arch Mater Sci Eng. 2022;115:49–57.
[53] Jia JF, Chen XZ, Bai YL, Li YL, Wang ZH. An interpretable ensemble learning method to predict the compressive strength of concrete. Structures. 2022;46:no. September 201–13. doi: 10.1016/j.istruc.2022.10.056.
[54] Ruchitha HK, Sanjana Sharma BS, Sneha, Kiran N, Rao T, Ashwini K. Evaluation of unconfined compressive strength of cement using ensemble learning regression. 2022 IEEE Int Conf Electron Comput Commun Technol Conecct 2022; 2022. p. 1–7. doi: 10.1109/CONECCT55679.2022.9865736.
[55] de-Prado-Gil J, Palencia C, Silva-Monteiro N, Martínez-García R. To predict the compressive strength of self compacting concrete with recycled aggregates utilizing ensemble machine learning models. Case Stud Constr Mater. 2022;16:no. March e01046. doi: 10.1016/j.cscm.2022.e01046.
[56] Zhang Z, Zhu X, Liu D. Model of gradient boosting random forest prediction zhidong. Xi'an, China: Xidian University; 2022.
[57] Roberson MM, Inman KM, Carey AS, Howard IL, Shannon J. Probabilistic neural networks that predict compressive strength of high strength concrete in mass placements using thermal history. Comput Struct. 2022;259:1–13.
[58] Achong PSA, Guntor NAA. Concrete strength prediction using linear regression of machine learning algorithm peggy. RTCEBE Homepage. 2021;2:691–9.
[59] Pham HV, Dinh MN, Luong S, Chung TC, Nguyen QD. Discussion on complexity and accuracy of high-performance concrete's compressive strength deep learning models. Proc. - 2022 RIVF Int Conf Comput Commun Technol RIVF 2022; 2022. p. 647–52. doi: 10.1109/RIVF55975.2022.10013897.
[60] Golafshani EM, Behnood A, Arashpour M. Predicting the compressive strength of normal and high-performance concretes using ANN and ANFIS hybridized with grey wolf optimizer. Constr Build Mater J. 2020;232:1–14.
[61] Li L, Jia P, Dong J, Shi L, Zhang G, Wang Q. Effects of cement dosage and cooling regimes on the compressive strength of concrete after post-fire-curing from 800 C. Constr Build Mater. 2017;142:208–20.
[62] ACI Committee. ACI 211.2-98 supersedes ACI 211.2-91 and became effective. USA: American Concrete Institute; 1998. p. 1–18.
[63] Zhu J, Fang S, Yang Z. Prediction of concrete strength based on random forest and gradient boosting machine. Wuhan, China: Huazhong Univ. of Sci. & Tech; 2023.