Civil Engineering and Architecture 11(4): 2012-2032, 2023 http://www.hrpub.org
DOI: 10.13189/cea.2023.110425

The Role of Hybrid Machine Learning for Predicting Strength Behavior of Sustainable Concrete

Bader Aldeen Ayman Almahameed[1,*], Habibur Rahman Sobuz[2]

[1]Maan University College, Al-Balqa Applied University, Jordan
[2]Department of Building Engineering and Construction Management, Khulna University of Engineering and Technology, Khulna-9203, Bangladesh

Received December 4, 2022; Revised February 16, 2023; Accepted March 12, 2023

Cite This Paper in the Following Citation Styles

(a): *[1] Bader Aldeen Ayman Almahameed, Habibur Rahman Sobuz , "The Role of Hybrid Machine Learning for Predicting Strength Behavior of Sustainable Concrete," Civil Engineering and Architecture, Vol. 11, No. 4, pp. 2012 - 2032, 2023. DOI: 10.13189/cea.2023.110425.*

(b): *Bader Aldeen Ayman Almahameed, Habibur Rahman Sobuz (2023). The Role of Hybrid Machine Learning for Predicting Strength Behavior of Sustainable Concrete. Civil Engineering and Architecture, 11(4), 2012 - 2032. DOI: 10.13189/cea.2023.110425.*

Abstract Researchers are actively seeking accurate models for predicting forecasting mechanical strength in response to the proliferation of novel mixtures of concrete and applications. Both linear and nonlinear regression, two types of empirical and statistical models, have seen extensive use. Sustainable concrete is made by introducing supplemental cement elements into concrete mixing, and it finds widespread use in sound attenuation, roofing, thermal insulation, varied tunneling, and geotechnical engineering. The effectiveness of this technology depends on its capacity to provide consistent products with predictable outcomes. In this article, we train and test our ML approaches and modeling using an experimental database comprised of relevant data obtained from numerous prior investigations. Through a new combination of the random forests (RF) model and the Bagging algorithm, this work introduces a hybrid ML model (RF-B) for forecasting the compressive strength of concrete. Bagging is an ensemble approach that aggregates the predictions of numerous models that were each fit to a separate subset of a training dataset. As a second example, Support Vector Regression (SVR) was created to help in finding the activities of parameters in connection to one another in order to forecast the robustness of machine learning models. Multivariate analysis is also another way of reading the data accumulated with a determination coefficient of roughly 0.6. The decision tree regression showed two iterations and R^2 values are 0.7453 and 0.7737 respectively. The cement percentage, density for oven dry conditions, w/c ratio, and additive usage are all used as input factors in the predictive models. Machine learning has many potential benefits for the construction industry, including cost savings, time savings, and less labor intensity. The statistical and graphical representation of contributors and countries in this study can facilitate the development of collaborative projects and the trading of novel ideas and approaches among scholars.

Keywords Sustainable Concrete, Random Forests, Bagging Algorithm, Machine Learning, Support Vector Regression, Multivariate Analysis

1. Introduction

Sustainable development factors heavily into the latest demands are placed on the building sector, particularly in the form of environmental restrictions and the protection of natural resources. Restoration efforts are a major contributor to the worldwide rise in the generation of building debris and demolition waste. The annual building trash output in Tehran is predicted to reach 20 million tons. The concept of digitization in the construction industry extends well beyond simply purchasing the most up-to-date computers, software, servers, or networks, despite the fact that these are also essential components of technical growth. The construction industry has the potential to become one of the most lucrative sectors of the economy if state-of-the-art digital technologies are incorporated into traditional building practices. These technologies include artificial intelligence (AI), big data, machine learning, and the internet of things (IoT) [1]. In recent years, researchers have utilized a variety of approaches to predicting and analyzing the many different qualities that recycled aggregate concrete possesses. The use of artificial neural networks (ANNs) and other methods that are founded on the discipline of machine learning is becoming increasingly widespread. ANN methods are considerably utilized, despite the fact that they have the potential to be useful, in order to forecast how well recycled coarse aggregate (RCA) and concretes in general will perform. The strength of compressive and splitting tensile of RAC with silica fume were predicted by Topcu and Saridemir [2]. Using 168 data sets, Duan et al. [3] suggested an ANN model with 14 input features. In order to determine the connection between the qualities of RCA and the corresponding compressive strength, Chopra et al. [4] ran a regression analysis on 20 data sets.

When it comes to the construction of things, concrete is an absolute necessity. It is significant to produce nicely designed concrete as a durable construction material because transport properties have become one of the critical problems in the process of constructing reinforced concrete structures with long service lives and developing construction technologies for a variety of environmental and economic reasons in recent years. As a result, the manufacturing of nicely designed concrete as a sustainable construction material is important. In spite of this, the creation of concrete necessitates the use of enormous quantities of natural resources such gravel, sand, water, and cement. In addition, the manufacturing of cement consumes around 3 billion tons of raw materials per year [5, 6] and about 2.5% of total worldwide carbon emissions from sources of industry levels [7, 8]. There is a pressing need to find innovative approaches to achieving sustainable development in light of the alarming rates at which natural resources are being depleted, industrial waste is being generated, and the environment is being polluted. Mineral admixtures including powdered granulated blast furnace slag, fly ash, and silica fume used as Supplementary Cementitious Materials (SCM) are among the most effective strategies to lessen the impact on the environment. Because of the pozzolanic reaction, mortars and concretes are made with the help of additives to have better compressive strength, pore structure, and permeability [8, 9]. The lesser cement required that leads to lessen the amount of CO_2 generated in the production of cement, therefore this strategy has the ability to cut expenditures, preserve energy, and minimize waste [7, 8, 10]. Concrete's compressive strength stands out as a key determinant of the material's overall quality. Compressive strength of RAC has been proven to be strongly reliant on recycled aggregates (RA) strength, according to studies. Consequently, RAC constructed of lower-strength RA has a lower strength than RAC made of higher-strength RA; the degree to which the strength is reduced depends on a wide range of variables, including the kind of concrete, the w/c ratio, the moisture percentage, the replacement ratio, etc. [11-14]. Figure 1 shows various models of machine learning that can be used in this study.

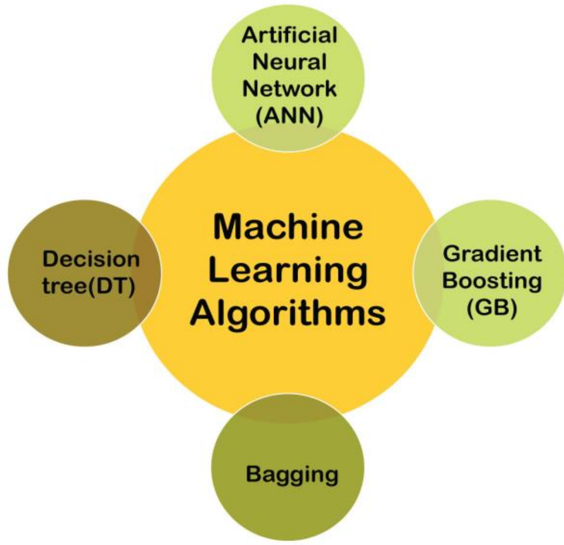


Figure 1. Various models of Machine Learning

This variation in RA and RAC behavior would necessitate extensive testing to learn more about their capabilities. Although useful, these many tests take a lot of time, money, and resources. Therefore, data algorithmic models considered by measurable data can be a useful replacement for this intensive testing when estimating the strength of compressive for concrete. Figure 1 shows the ML algorithms that can be improvised as per requirements.

The objective of this study is to construct a novel hybrid machine learning model to forecast the compressive strength of using all types of replacements that will be the consequence for producing sustainable concrete and carbon free environment. For this purpose, a computer program was developed by using ML and the novelty of using RF-B where bagging algorithm is incorporated. Considering all the SCM data whether it's cementitious one or the coarse aggregate replacement with the recycled one, the compressive strength for different curing periods is also showed and compared by multivariate analysis.

Table 1. Detailed information of the property of materials

Materials	Property
Ordinary Portland Cement	P.O. 42.5 (R), 42.5 MPa (28 days),80%≤clinker+gypsum<95%
River Sand	Fineness modulus: 230, apparent density: 2511 kg/m^3
Gravel	Size: 4.75-16.00 mm, apparent density: 2610 kg/m^3
Rubber powders (NR)	Size: 20 mesh, apparent density: 985 kg/m^3
Rubber powders (NBR)	Size: 20 mesh, apparent density: 1289 kg/m^3
Waste rubber blocks (NR, NBR, and Q)	Dimensions:80mmx40mmx20mm
Sodium Silicate	Modulus: 3.1, Baume degrees: 40°C
Calcium chloride solution	Saturated Solutions

2. Concrete Constituents

The smallest aggregate size was no more than 19 mm in nominal size. Numerous laboratory tests were conducted on gravels, and the results showed that their water absorption and specific gravity in the saturated surface dry (SSD) state were, respectively, 2.37 and 2.57. The experiment's sand is of 4.75 mm which is the maximum size, absorbed 2% of its weight in water, and had a specific gravity of 2.51 under the SSD condition. Both the sand equivalent and the fineness modulus (FM) were determined to be 3.09 according on the standards set forth by ASTM D2419 [15] and ASTM C33 [16], respectively. The powder binder properties and other materials for the rubberized concrete in this study are displayed in Table 1. Powdered cement (C), silica fume (SF), and superplasticizer (SP) make up this substance (SP). The experiment made use of type II cement, which is often employed for a variety of construction projects. Ten percent of the cement volume was replaced with Silica Fume (SF) as per ASTM C1240 standards guideline for using silica Fume [17].

2.1. Recycled Aggregates

Consequences of attached cement paste, recycled aggregate (RA) typically have inferior inherent characteristics compared to natural aggregate (NA). Recycled concrete accounts for 20-30% of the total volume and typically retains the same properties as the parent concrete. Compared to its natural counterpart, recycled aggregate is weaker, more porous, more angular, contains impurities, and is more easily damaged by mechanical and chemical processes. Some of the benefits of employing recycled aggregate include the following.

2.2. Aggregates

Coarse aggregate properties, such as particle size at a maximum level, saturated surface dry (SSD), specific gravity (SG), and water absorption (WA), significantly impact RAC strength. The coarse aggregates employed have widely varying characteristics because of the wide variety of possible sources and crushing procedures that yield them.

Aggregate from a clean, dry river was utilized to make the concrete. The gravel had a nominal maximum size of 16 millimeters, a water absorption value of 1.3%, and a relative density of 2.70 grams per cubic centimeter when dried to a saturated surface. The sand had a relative density of 2.67 g/cm^3 under SSD conditions and an absorption value of 1.8% when wet. This displays the typical range for the grading of the blended aggregate. The table 1 demonstrates that the aggregate grade is acceptable for making concrete.

2.3. Superplasticizer

The chemical admixture is used as additive to make the concrete more workable and mixable in terms of chemically hibernated with the other constituents. The addition of these components significantly improves the mixture's workability, allowing for a lesser w/c ratio or even a smaller amount of cement. To draw conclusions about the performance of superplasticizers, tests like fresh tests and strength based were evaluated using natural (river) and crushed limestone aggregate, respectively. The amount of fine aggregate in a mixture also appears to affect the efficiency of superplasticizers.

2.4. Mineral Admixtures

Concrete's workability, resistance to heat cracking, alkali-aggregate expansion, and sulfate attack can all be improved by adding mineral admixtures (fly ash, silica fume (SF), and slags) in larger quantities than is customarily the case. In addition to silica (SiO_2), alumina (Al_2O_3), and calcium oxide (CaO), there are a few other elements that make up fly ash. The properties of Ground Granulated Blast Furnace Slag (GGBFS) range from cementitious to pozzolanic. Slag cannot be hydrated without an activator. Silica fume consists primarily of amorphous silica particles of extremely small size. When elemental silica or other silicon-based compounds are manufactured in electric arc furnaces, this byproduct is also created.

3. Methodology

The subfields of computer science known as artificial intelligence and machine learning are intrinsically linked. These are the two cutting-edge methods now in use for developing smart machines. Although these are both examples of related technologies and are commonly used interchangeably, they are technically distinct concepts. By analyzing past data, machine learning allows computers to infer future outcomes or make judgments with limited human input. In order for a machine learning model to produce reliable results or predictions, it requires access to vast amounts of structured and semi-structured data. That concrete's strength is typically checked to determine if it is of sufficient quality. We conduct a typical crushing test on a concrete cylinder to measure its Compressive Strength and use that number as a proxy for the overall quality of the concrete. One of the most important aspects of achieving the appropriate longevity is the concrete's strength. Given that 28 days are a long time, what are our next steps? With the help of data science, we can estimate how much of each raw material will be needed to achieve a certain compressive strength, saving a great deal of time and energy. The process of exploratory data analysis (EDA) is crucial to the development of any machine learning project, as it provides a concise summary of the dataset’s most salient features. Using EDA, we can learn about key characteristics by simply glancing at graphical representations of the data.

Hybrid techniques aim to improve model performance and process by combining multiple algorithms. Researches have taken an increased interest in hybrid techniques due to their ability to combine the benefits of multiple models. Since then, many studies [4, 18-23] have looked into models like the Adaptive Neuro-Fuzzy Inference System (ANFIS). By fusing the strengths of ANN and FL, ANFIS models can serve as universal approximators. Despite the fact that the rules of FL are supposed to supply expert knowledge, ANN is used to improve the membership capacity with the goal of lowering the mistake rate in the output [16, 19]. Fuzzy if-then rules based on FL are utilized internally by the algorithm to generate the required input-output pairs. For instance, strength of compressive predictions of polymer considering earth characteristics, concrete including BFS, and concrete containing fly ash were all made using ANFIS [19, 23]. Outperforming another model, the consequences showed that ANFIS's prediction capacity is robust. With the help of ANFIS, one is able to make educated guesses about the shear strength of RC and High Strength Concrete (HSC) beams [24, 25]. The model's accurate forecasts trumped those from authoritative design regulations like the American Concrete Institute and the Canadian Standards Association.

Metaheuristic algorithms were incorporated as another method for optimizing ANN [26, 27]. To find the optimal weights and thresholds for Back Propagation Neural Networks (BPNN), Yuan et al. [20] used a GA model. The idea of natural evolution and selection served as inspiration for Genetic Algorithm (GA), a metaheuristic algorithm. Since it can find a good compromise between global and local optimum solutions, it might be used to fine-tune BPNN. Predicting the compressive strength of slag and fly ash-based concrete was done with a hybrid GA-ANN model. The results of a comparison between GA-ANN and BPNN showed that the former had superior performance. To find the optimal ANN architecture, Behnood et al. [27] created a multi-objective grey wolves’ optimization (MOGWO) approach. Grey wolf pack dynamics and hunting strategies served as inspiration for the MOGWO optimization method. Predicting concrete's compressive strength using silica fume using a hybrid MOGWO-ANN yielded promising results in an experiment. The compressive strength of the concrete was shown to be highly sensitive to the maximum aggregate size, according to the sensitivity analysis. In order to estimate the compressive and tensile strength of HPC, Bui et al. used an ANN model, with its weights and biases tuned by a Modified Firefly Algorithm (MFA). Based on the flitting behavior of tropical fireflies, the firefly algorithm (FA) is a metaheuristics technique [28]. The research described that the MFA-ANN model is considered to make precise forecasts in a relatively short amount of time.

Lee [29] noted that many different approaches have been proposed over the years to forecast the concrete compressive strength; most of them are grounded in the concept of concrete's maturity. Independent research has demonstrated that the w/c ratio isn't the only factor in determining how much strength concrete gains over time; other elements matter too. These other elements are curing period, ratios of materials, proper mixing procedure as sustainable or high strength concrete needs special care with mixing etc. However, a few exceptions have been observed from this rule even though experimental data has proved its practical applicability within large boundaries. Codes and standards currently use empirical equations based on tests of concrete without added cementitious ingredients to estimate compressive strength. It is recommended that research be conducted to determine whether or not these connections hold true for concrete that contains SCMs such as blast furnace slag, silica fume, fly ash, and so on. Increasing our understanding of concrete's nature and how to best optimize the concrete mixture requires a deeper dive into the composition vs. strength relationship [29, 30].

Researchers and practitioners in the engineering and scientific communities have increasingly turned to a novel modeling approach based on fuzzy logic (FL) or neural networks over the past two decades [31]. A type of massively parallel architecture, network neuralties (NNs) are able to perform complex tasks by coordinating a large number of relatively simple computational nodes. Simply described, a neural network's processing elements are a collection of simple computational elements organized in layers, much like neurons in the brain [32]. The theory of fuzzy control has wide-ranging applications, including both linear and nonlinear systems. The complex mathematical models of a regulated body are unnecessary. Using simply engineering knowledge, a straightforward mechanism of control can be established. As a result, it excels in controlling complex structural systems. Models constructed with FL and NNs allow for the estimation of compressive strength. Numerical tests reviewing the impacts of each variable on the mix proportions may be performed quickly and easily with the help of these models [33-37]. The schematic flowchart for achieving goals of an ANN model is shown in the figure 2.

3.1. Machine Learning Methods

3.1.1. Support Vector Machine

The SVM is a supervised learning method that produces mapping functions between input and output using just labeled training data. The input data category determines whether the mapping function is a regression or classification function. In classification, nonlinear kernel functions are often used to map data inputs to a high-dimensional feature space. Thus, the input data become progressively disentangled with relation to the precise input space. Thereafter, maximum-margin hyper planes are generated. Then, the model was built using just the data at the class boundaries. Support Vector Regression, in its turn, creates a model that ignores any training data that is sufficiently similar to the model prediction. The SVMs are assumed to be "kernel procedures" as well.

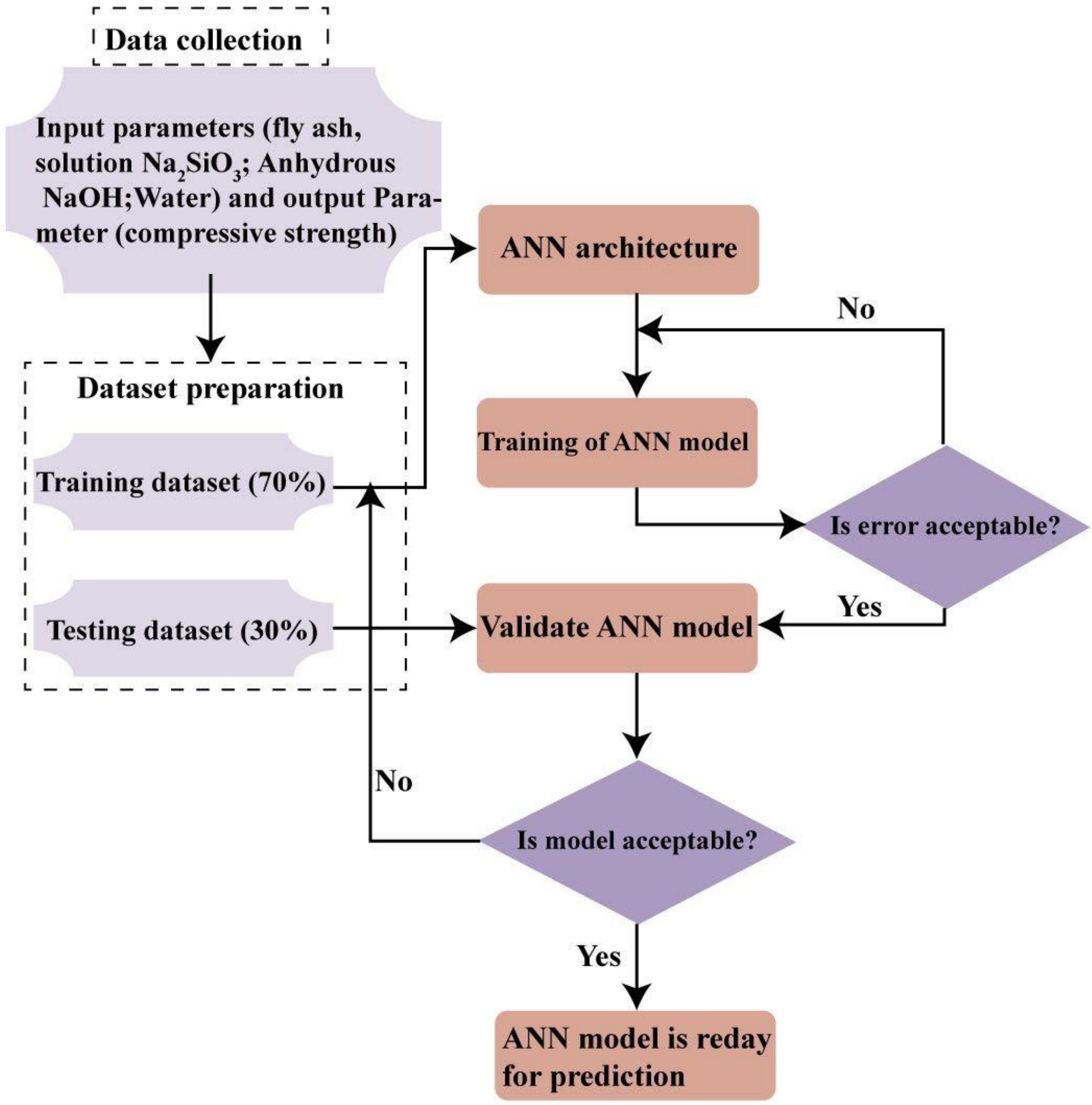


Figure 2. Schematic flowchart for achieving goals of an ANN model

Vapnik and his co-workers at AT&T Bell Laboratories originally unveiled the support vector machine (SVM) in 1995 [38]. Although Pal and Mather and Dibike et al. both employed the SVM extensively in civil engineering applications, it is rarely used to forecast High Performance Concrete (HPC) compressive strength [39,40]. In this study, an epsilon support vector regression model was used to construct an HPC input-output diagram [41], a version of SVM for function estimation.

In SVM regression, the input is first mapped using fixed (i.e., nonlinear) mapping into a dimensional feature space.

$$f(\chi, \omega)=\sum_{j=1}^{m} \omega_j\, g_j(\chi)+b$$

In mathematical notation, the linear model in the feature space, f x, can be written as follows: where, g_j (x), j = 1, m = set of nonlinear transformations; and b = "bias" term. Data are frequently obtained in the preprocessing stage under the assumption that they have a zero mean, hence the bias term is skipped. The loss function, L [y; f (x, ω)], assesses the quality of an estimation.

3.1.2. Random Forests

Combining numerous base learners (often a fast learner, such Decision Tree) into an "ensemble" can increase the accuracy of predictions [42]. In the realm of Tree based ensemble models, Random Forest (RF) is a common favorite. The training set Sn is partitioned arbitrarily into many subgroups for use in training the RF algorithm [43]. A regression tree is developed for each subset. The bagging method combines all of the regression trees to decrease prediction variance and increase prediction accuracy [44]. To be more precise, n samples are drawn at random from the space Sn, with a selection probability of 1/n. These n samples have been chosen at random style.

3.1.3. M5 Model Tree

The M5 Model tree [45] has been used as a data algorithmic model in the present investigation to evaluate the similarities and differences between the ELM and the MARS. To explore correlations in the predictor-predictor and matrix, this hierarchical model considers leaf nodes and is grounded in a binary decision framework [46]. Two procedures [47] are used to create an M5 Model tree. We begin by creating a decision tree of input-output data patterns and attributes by dividing the data into subsets. The next step is to construct a model of the tree.

Think of a training matrix with N samples and d dimensions, where each dimension represents an attribute of the predictand that was used in the training process. The goal of an M5 Tree model [48] is to establish a correlation between the value that is targeted considering the positive cases and the predictor. Model trees (MTs) have significant advantages over other soft computational approaches, including being more accurate than regression trees, more transparent and intelligible than Artificial Neural Networks (ANNs), easy to train, and robust when coping with missing data. M5' Model Tree is a relatively new soft computing approach that has had limited use in structural engineering thus far. Using this technique, Kaveh et al. recently established M5's efficiency relative to several other algorithms for predicting the main ground-motion parameters. The M5' algorithm is described in detail below. In 1992, Quinlan created the M5 model, which was later refined by Wang and Witten in 1997 into the M5' system. When compared to regression trees and ANNs, M5 model trees perform better and are easier to understand. It works well with high-dimensional data and a large number of properties. The three primary components of the method are tree construction, tree pruning, and smoothing. The first tree structure is established by using the splitting criterion. This dividing criterion measures the node's error as the standard deviation of the class values that make it there, and it predicts how much that error will decrease if each attribute is tested there. The attribute with the highest predicted error reduction is then chosen.

Improvising the conquer and divide technique, we build a model represented as tree in which N points are linked to a leaf, representing a test criterion, and then we partition the tree into subsets that correspond to the results of the test. To build subsets for this recursive approach, we use a criterion that accounts for both the spread of class values and the enhancement of accuracy. Random forests graph showed in figure 3 [48,49]. Here the accuracy of this study measured is 0.926858 and 95.0 confidence interval considering lower and upper limits are 76.2% and 99.9% respectively.

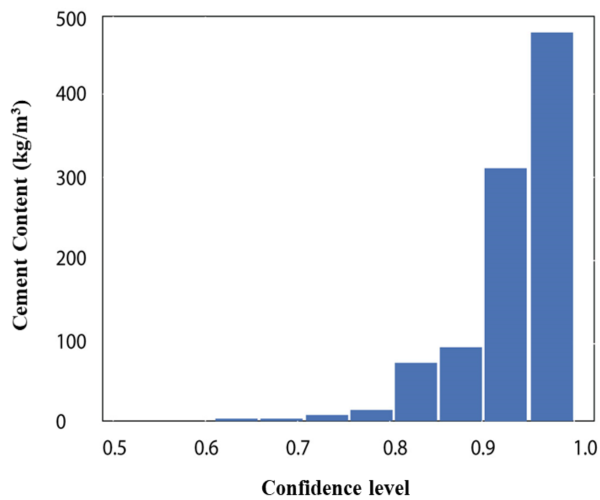


Figure 3. Random Forest Regression

3.1.4. Bagging Regression Trees

Bootstrap aggregating, often known as "bagging", is a "bootstrap" ensemble method that employs multiple classification and regression techniques to reduce prediction variance and produce more accurate results [50]. Bagging is based on the idea that one can train a single learning algorithm on a redistributed training set to produce many independent regression or classification models. In order to produce a training set for each regression model, N instances are drawn at random considering the training set with original replacement. The final training set may include many or few of the initial occurrences. Once many models of regression have been developed, the final forecast is obtained by averaging the values of their respective predictions. Breiman describes a more advanced kind of bagging [51]. Bagging, or bootstrap aggregating, is the practice of arranging bagged samples in a way that strengthens and refines the efficiency of ML algorithms used in tasks like regression and classification. It is commonly employed to lessen discrepancies between observed and anticipated outcomes. Though it is not limited to decision trees, bagging has been widely used in many different types of analyses [52]. It is also a specific example of the method known as "model averaging." During the training phase, Bagging adds additional data to the mix such that each piece of the dataset has an equal probability of appearing. It is impossible to increase predictive power while simultaneously modifying the training set. As a result of optimizing the value of 20 different sub-models that make up the decision tree with bagging, we can get a definitive answer to our output question. One of the most effective methods for developing predictive models, gradient boosting is widely recognized and regarded as a robust tool. It is an ML ensemble algorithm typically used to solve classification and regression issues. In most cases, the decision tree is used as the basis for this forecasted model, but the technique can be applied to any ensemble of projected models. As a gradient boosting tree, the resulting algorithm is viewed as a poor learner when the decision tree offers the outcome. The topic of ranking learning also benefits from the use of gradient boosting. The processing of data from high-energy physics also makes use of it.

3.1.5. Multiple Linear Regression Model (MLR)

It is common practice to use regression models for this purpose, as they allow one to estimate the degree of correlation between input and output variables and to determine the nature of their relationship. While least squares are the most common technique for fitting linear regressions, there are other methods available, such as ridge regression's penalized version of the least square's loss function or the reduction of the "lack of fit" in alternative norms. There are the two most common forms of linear analysis namely Simple linear regression (SLR) and multiple linear regression (MLR). The SLR and MLR methods are used to estimate and predict linear correlations between predictors and criteria variables, respectively. Remember that the MLR is the most popular type of linear regression analysis and that each value of the independent variable has a corresponding value of the dependent variable. Typically, MLR calculates an estimation of the correlation considering single response (dependent variable) and two or more response predictors (Independent variable). It's important to note that the MLR seeks to find a straight line that best predicts each and every one of the data points that includes both the target and the output variables demonstrated by some authors considering the basic structure of an MLR model [12, 50].

3.2. Models Development and Construction

Some of the machine learning regression models (RF-B, M5 Tree, and SVR) considered for predicting foamed concrete compressive strength were created using MATLAB sub-routines. Compressive strength was chosen as the dependent variable (y) and the input matrix (x) was defined as the predictor variables (cement content, coarse and fine aggregate, superplasticizer, water, and mineral admixture data). The entire set of experimental input data was considered into three sections, with 70% being randomly selected and classified into the training set, and 30% into the testing set. A study of pairs of significantly correlated concrete strength values and their respective predictors was conducted to verify the efficacy of the heuristic models.

4. Dataset of the materials as input

By adding a specific volume of supplementary cementitious materials, standard sustainable concrete data in this study comprises of a cement paste (cement and water) or mortar (cement, water, and sand) with a uniform pore structure. Sustainable concrete's compressive strength and other properties can be enhanced with mineral admixtures like silica fume and fly ash. Compressive strength can be improved by using water reducers and superplasticizers to lower the binder-to-water ratio. Sustainable concrete's tensile strength, flexural performance, and toughness can all be enhanced by mixing in materials like glass, steel, carbon, and synthetic fibers; however, the effect on compressive strength is less quantifiable [2].

Research papers covering a range of sustainable concrete densities (600-1,800 kg/m^3), foam addition volumes (from 0 to 100%), and mixture components were culled to create the database utilized in the creation of the models presented here. The data was extracted from the existing literature to build both the RF-B and the corresponding (comparative) model. Data intelligent models were constructed and tested using a total of 370 data points, split evenly between training and validation per testing sets. The projected compressive strength was calculated using cementitious material content, mineral admixture, superplasticizer, water/binder ratio, and SCM volume as inputs. Table 2 shows the input data for hybrid and multivariate analysis.

The table 2 considers various data for concrete mix and these data are secondary in terms of different mixture situations. When the data shows zero, that material is not used in that mixture. Cement is used and supplementary cementitious material is also used for sustainable concrete. Also, recycled aggregate is replaced with coarse aggregate. The mineral and chemical admixture are used as per standards and the ratios with the other constituents.

Table 2. Input Data for hybrid and multivariate analysis

Cement +SCM (kg/m^3)	Mineral Admixture (SF or FA) (kg/m^3)	FA (kg/m^3)	CA + Recycled Aggregate (kg/m^3)	Water (kg/m^3)	Superplasticizer (kg/m^3)
360	40	736	1155	160	1.48
437.69	42	61.13	1052.13	182.92	8.58
250	0	417	1681	123	0
350	0	373	1507	172	0
500	0	700	1110	135	14
500	0	700	1110	135	18
470	52.5	700	1110	135	18
442	78	689	1125	166	0
500	0	630	1260	150	10
425	75	630	1260	150	12.5
400	100	630	1260	150	15
308	0	630	1260	150	15
310	31	940	976	186	7.7
511	51	709	1122	153	20.4
350	0	750	1023	150	0
475	0	750	1065	150	25
390	0	585	1209	195	0
572	0	1345	0	286	0
786	0	1286	0	236	25.9
572	0	1345	0	286	0
441	28	653	1115	164	2.9
465	30	615	1168	149	3.1
450	45	615	1168	149	3.7
500	0	0	820	385	6
400	0	0	1038	308	4.8
470	45	700	1110	135	20
326	58	659	1124	135	3
391	69	689	1172	179	3.5
308	0	933	968	186	6
441	0	615	1115	149	1.9
350	150	750	1023	150	0
475	0	750	1065	150	25
465	30	615	1168	149	3.1
476	0	500	1121.51	193.54	3.34
149	0	0	267.77	31.93	3.52
501	113	1332	1678	307	1.49
487	161	1330	16890	314	1.54
540	106	1256	1697	265	28
528	120	1250	1725	265	24
96	265	678	1090	134	2.5
336	144	560	980	150	1.9
360	114	640	1050	124	2.4
432	48	720	950	119	2.4
635	30	567	742.95	133.35	0

5. Results and Discussion

5.1. Machine Learning Approaches

The many algorithms that considered estimating concrete's compressive strength at elevated temperatures are described in this section. In order to forecast the strength property of concrete, both ensemble and individual algorithms were used. The models were executed using decision tree, bagging, and gradient boosting. Python software was used for all three of the utilized machine learning strategies, and python code was used for all of them. Figure 2 depicts the used algorithms.

Decision trees are a supervised machine learning technique used for both issue distribution and classification. A decision tree's structure resembles that of a flowchart, complete with nodes, branches, and a root. Each leaf node indicates the class tag and each internal node displays a test on an attribute. The branching structure from stem to bud is a visual representation of the categorization scheme. Decision tree nodes come in three main shapes: square, round, and triangle. In its most basic form, it is a method for gaining insight and making sense of anything.

The algorithm used in ANNs is organized like a network of neurons in the homo sapiens brain. When modeling the human brain, an ANN can be thought of as a network of interconnected units or nodes (termed artificial neurons). These neural networks pick up new processing skills by watching others. They are stored within the net's data structure and include an "input" and "result" that are already known, allowing for the generation of probability-weighted associations between the input and the result. Predicting the hardened properties of concrete using an ANN is a hot topic considered in civil engineering right now. This is because of how well it predicts the true properties of strength.

5.2. Multivariate Adaptive Regression

This new approach, called MARS, was created and developed by Friedman, the non-parametric multivariate flexible regression approach. Piecewise regression is a local regression technique that employs a basis function to model the nonlinear association between predictor variables. In addition, when a target is the specified predictand variable, it provides more leeway for analyzing the nonlinear relationship between predictor and predictand. In order to get beyond the restrictions imposed by CART, researchers turned to MARS, an expanded version of CART. In order to fit the spline function, regions in the produced space of predictor variables are overlapped during training. The final, global model is a weighted average of the individual regional ones.

The correlation matrix for input variables is shown in Fig. 1. The majority of the correlations are low (less than 0.5), indicating that multicollinearity problems will not arise from the use of these variables [52]. Stronger correlations have been found between RCA density and RCA water absorption rates. The density of RCA has been shown to be linearly related to its water-absorbing capacity in a prior study, with a determination coefficient of roughly 0.6. Low-density particles may be less likely to absorb water since they are more porous. Since both RCA and NCA serve as coarse aggregates in RCA mixtures, it is not surprising that I5 is closely correlated with I6 (NCA content). Input and output variables in the dataset are shown with distribution bar and histogram figure 4-11.

Figure 4 considered the Cement + SCM distribution with bar and histogram. The cement and SCM mixed concrete showed an improvement in strength rather than only cement mixed concrete. Sustainable concrete generally shows 60-80 MPa and higher properties than the normal strength concrete. Thus from the histogram it shows an upper value average of 60-80 MPa in strength parameter.

Figure 5 shows Mineral admixture distribution with bar and histogram. Prediction of strength considering less mineral admixture will result in higher in strength. Similarity with the Figure 7 where coarse aggregate and recycled aggregate were used to get a sustainable concrete higher in strength properties.

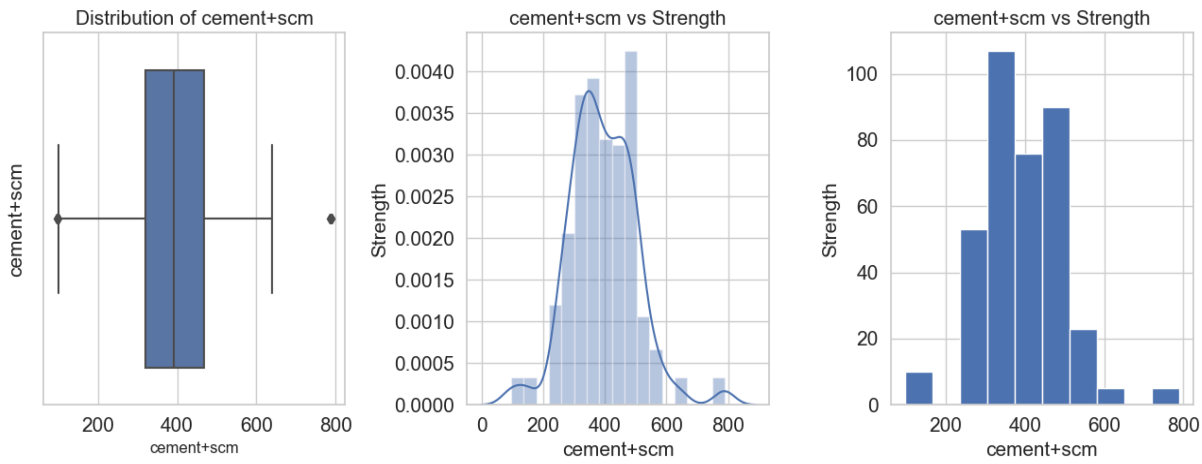


Figure 4. Cement + SCM distribution with bar and histogram

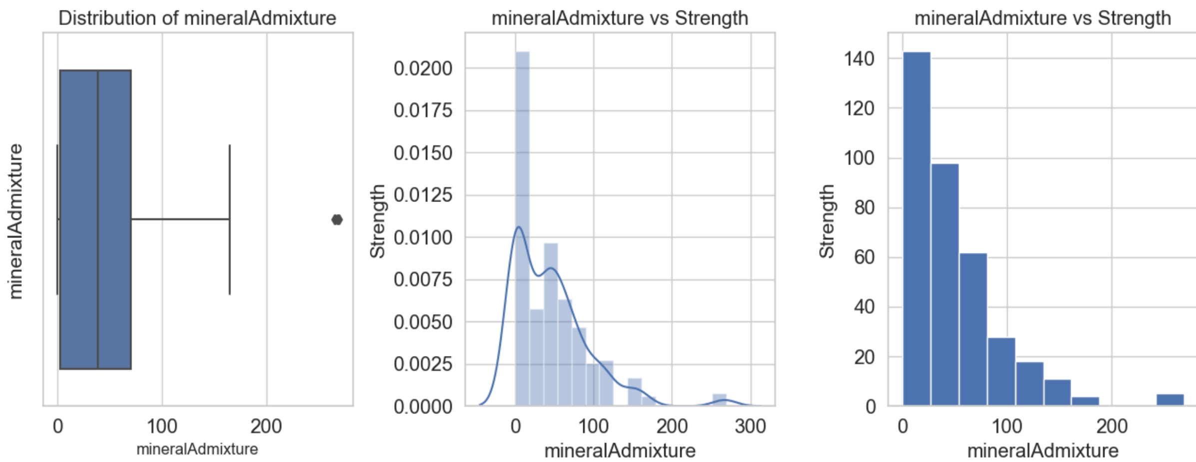


Figure 5. Mineral admixture distribution with bar and histogram

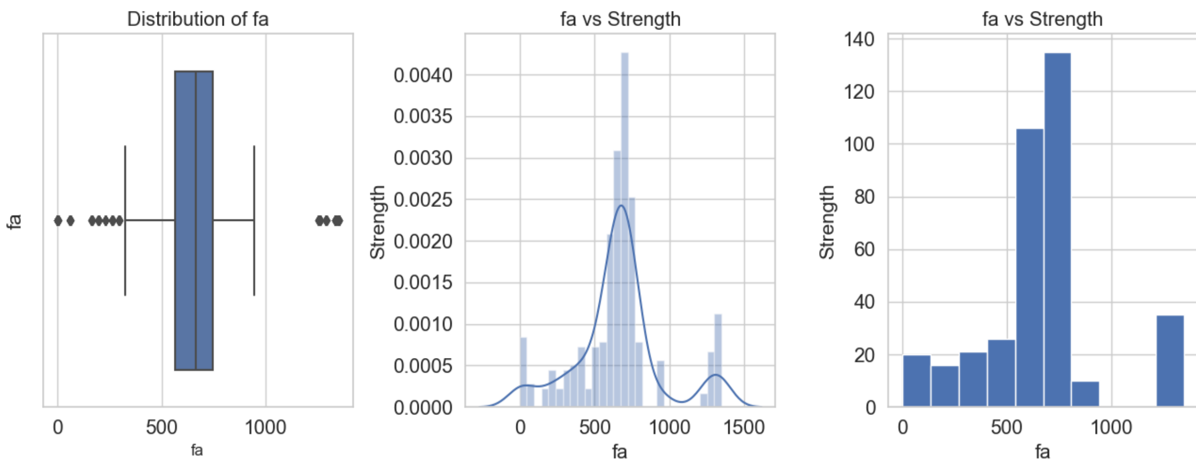


Figure 6. fa distribution with bar and histogram

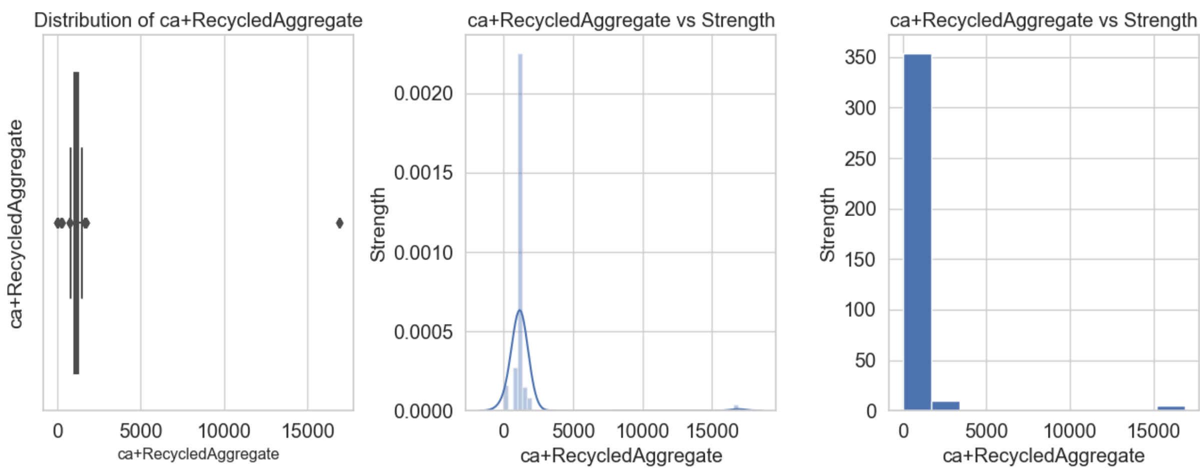


Figure 7. Ca+ recycle aggregate distribution with bar and histogram

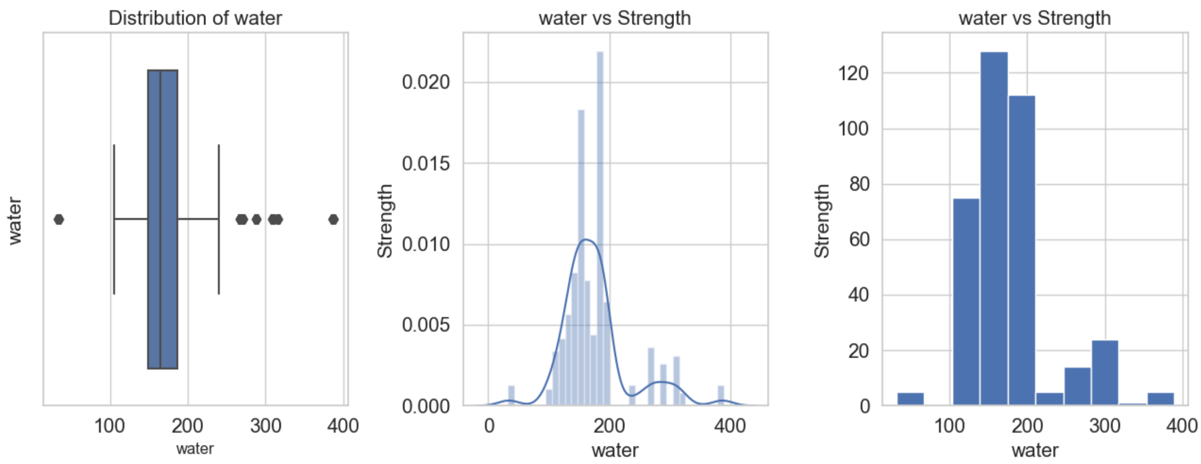


Figure 8. Water distribution with bar and histogram

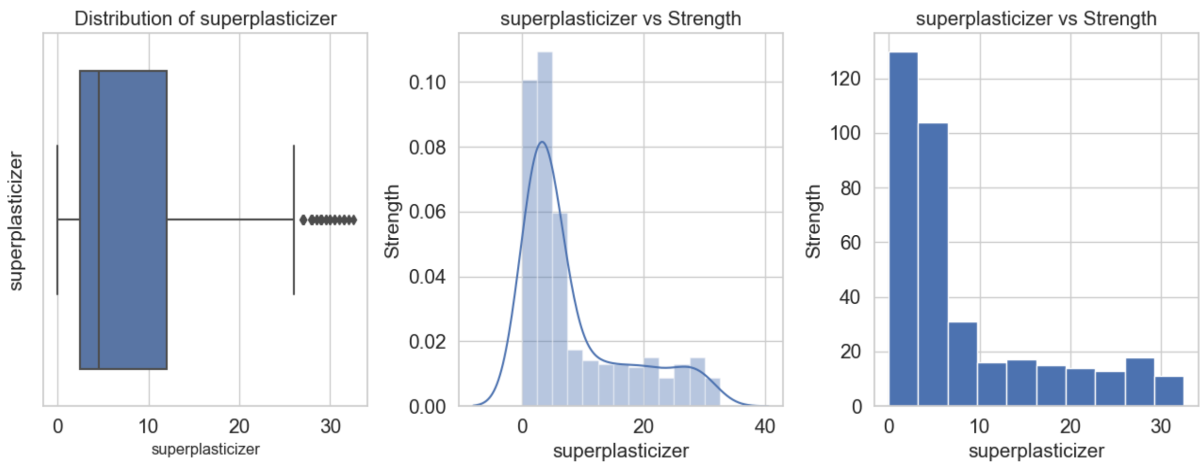


Figure 9. Superplasticizer distribution with bar and histogram

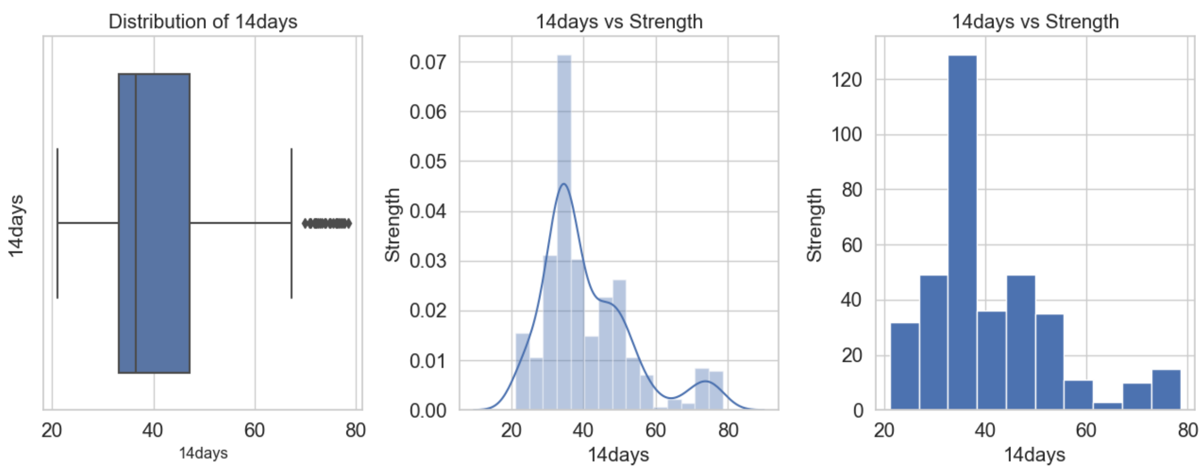
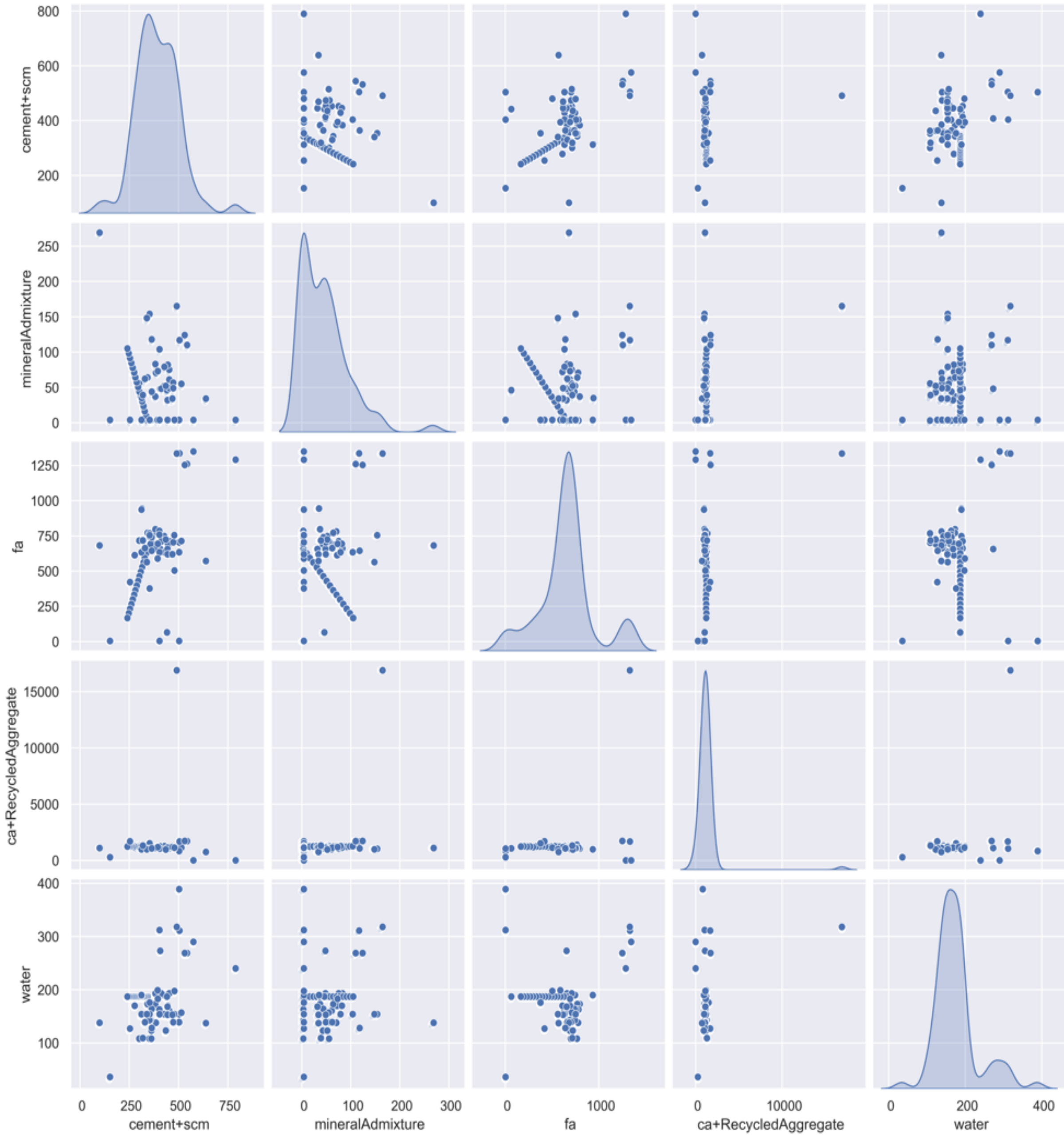
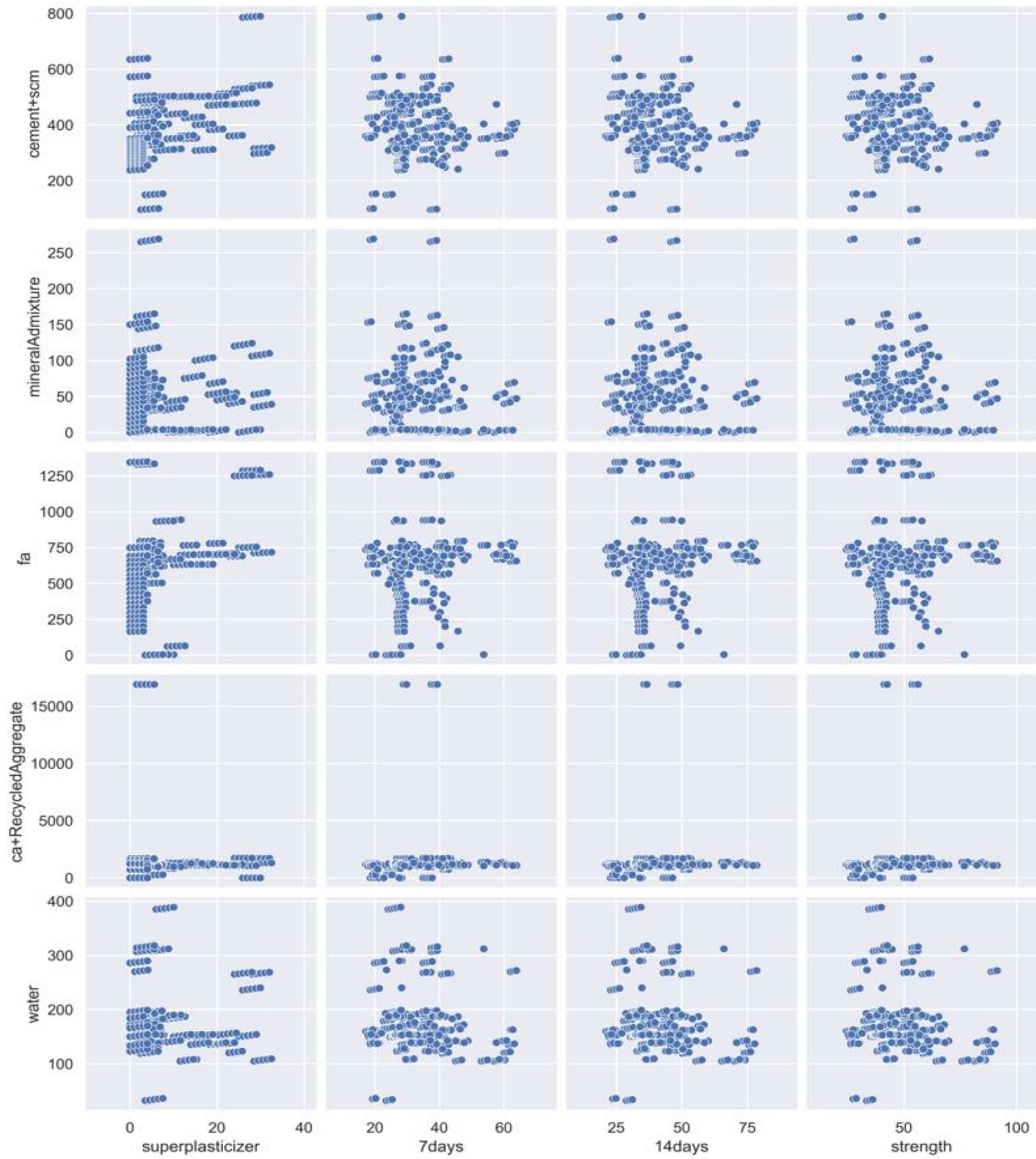


Figure 10. 14 days distribution with bar and histogram

Figures 8-10 illustrate various parameters like water distribution with bar and histogram, superplasticizer distribution with bar and histogram, 14 days curing period distribution with bar and histogram respectively. The distribution of water should be aligned with the cement ratio and must be 100-200 kg/m^3 for sustainable concrete. Also, the curing period and usage of superplasticizer need to be controlled to get a good prediction of result in strength.

All the data accumulated is predicted in the machine learning process by the multivariate analysis method. Models showed the scattered data with different materials mixing data and all are compared with each other. This method describes how one material will react with each other and with the considered strength. The predicted strength should be in the range of 80-60 MPa and after reviewing the scattered graph it shows the maximum values are higher than 50 MPa. So, the multivariate analysis taking materials data as input and strength prediction as output value is aligned with the aim of this research to understand the clarity of the topic. Thus, Figure 11 illustrates Multivariate analysis considering input variables and output as compressive strength.




cement+scm
mineralAdmixture
fa
ca+RecycledAggregate
water
superplasticizer
7days
14days
strength


superplasticizer
7days
14days
strength
cement+scm
mineralAdmixture
fa
ca+RecycledAggregate
water

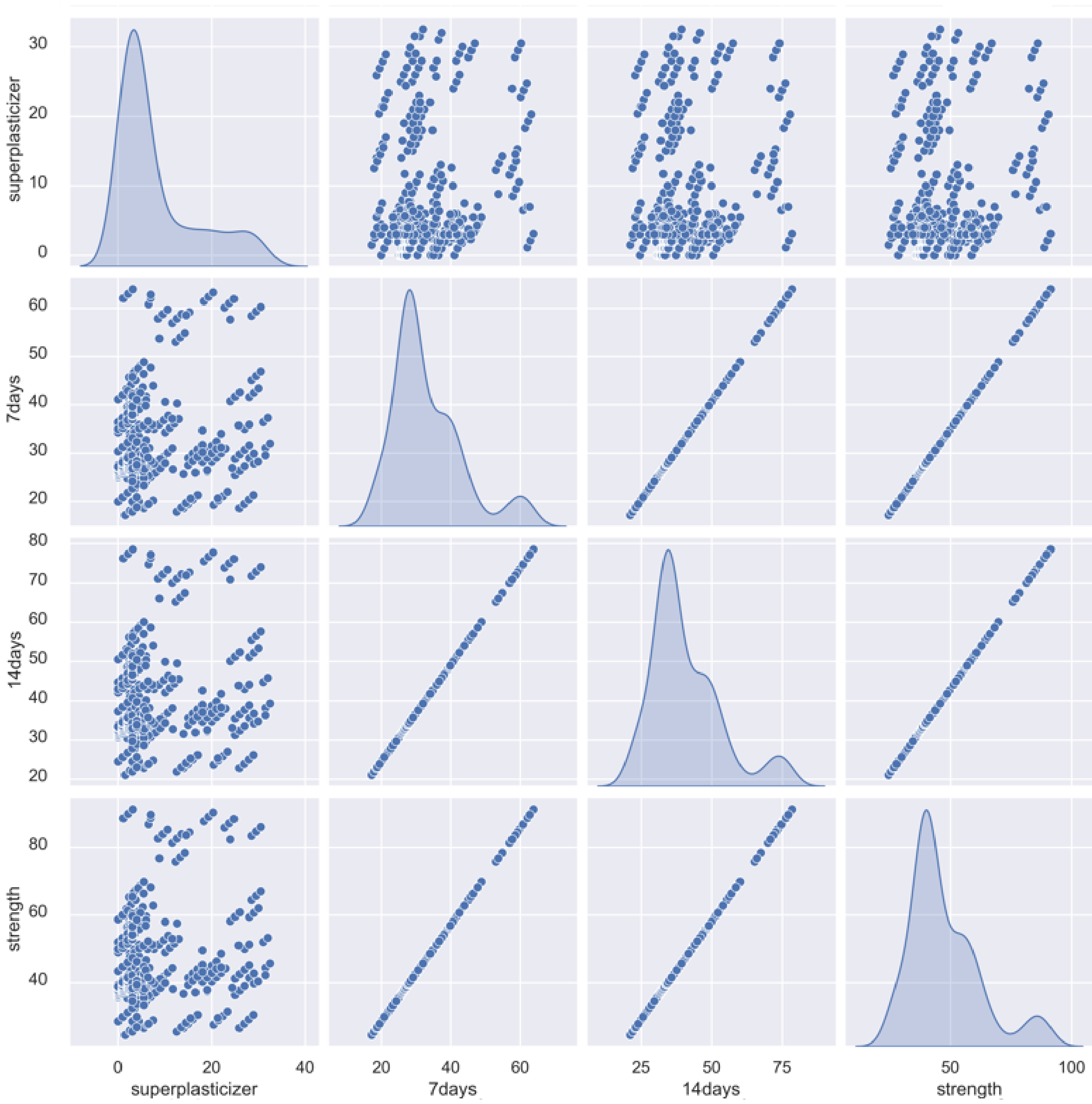


Figure 11. Multivariate analysis considering input variables and output as compressive strength

5.3. Correlation between features

Some of the tangible variables may be intertwined with one another. All potential variables' correlation coefficients have been calculated, and the results are shown in Figure 12. It may be inefficient and difficult to understand the behavior of the input variables on the responsive conditions if the correlation coefficient between them has a high positive or negative value. None of the input variables are significantly related to one another, as can be shown. The value of the correlation is 0.88 which is satisfactorily in terms of input values and ML created by the modelling as well as analysis. Thus, a clear idea of sensitivity analysis of the compressive strength based on experimental database can be summed as development of hybrid model of RF-B. The bagging analysis will really make this type of works readable to the non-users too.

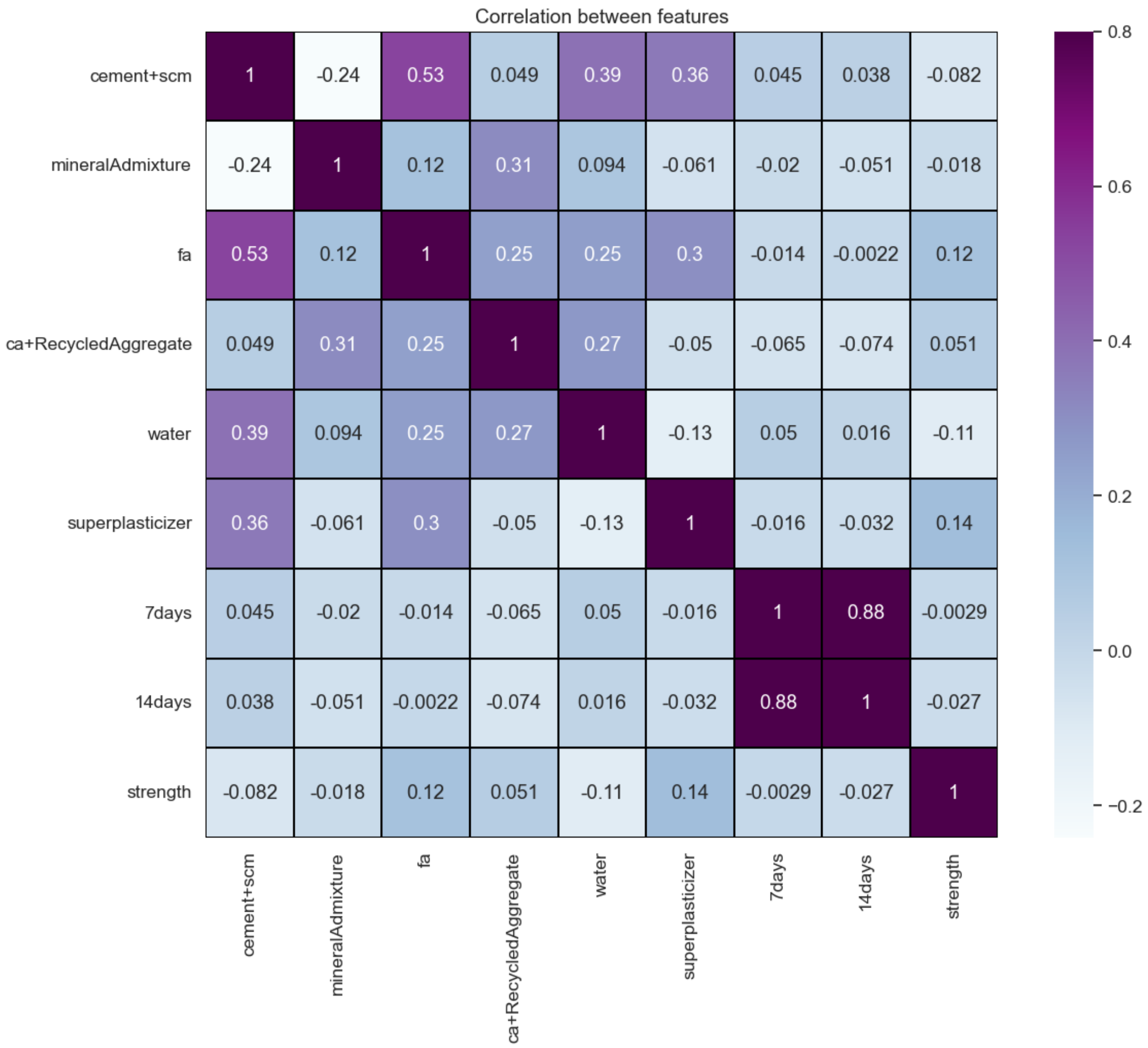


Figure 12. Correlation between various input and predicted strength

5.4. Decision Tree Regression

Decision trees (DT) are able to be constructed at a much faster rate than other types of classification methods. It's possible that these trees can be easily transformed into SQL queries for efficient database access. The accuracy of decision-tree classifiers is on par with or perhaps slightly higher than that of alternative methods of classification. Depending on the amount of data, the amount of memory available on the computer resource, and the method's flexibility, the decision-tree algorithm can be applied serially or in parallel. In the field of remote sensing, DT classifiers have not seen the same level of adoption as analytical or neural/connectionist methods. Decision trees have many clear advantages, including their ability to store data measured on different scales, their independence from assumptions about the summary statistics of the data in each class, and their robustness, in particular, and their ability to deal with non-linear interactions between characteristics and classes. The main benefit is that they may pick the most biased part and make it easier to understand. They are also simple to organize and decipher. Not only that, but they may be used with any kind of data, whether continuous or discrete. There is a more detailed schematic representation of the DT model in Figures 13 and 14. Two Iterations are done to understand the data variation considering the problems around compressive strength. The guarantee of this dataset can be verified by these two iterations because they look quite identical after being analyzed.

First iteration considered the R^2 value as 0.7453 and it improved in the 2nd iteration which shows an accuracy level where $R^2 = 0.7737$. The iteration-1 equation is: $y = 0.3513x^2 + 0.5521x + 0.1107$ and compared with the equation from 2nd iteration which is $y = 0.6716x - 0.0063$.

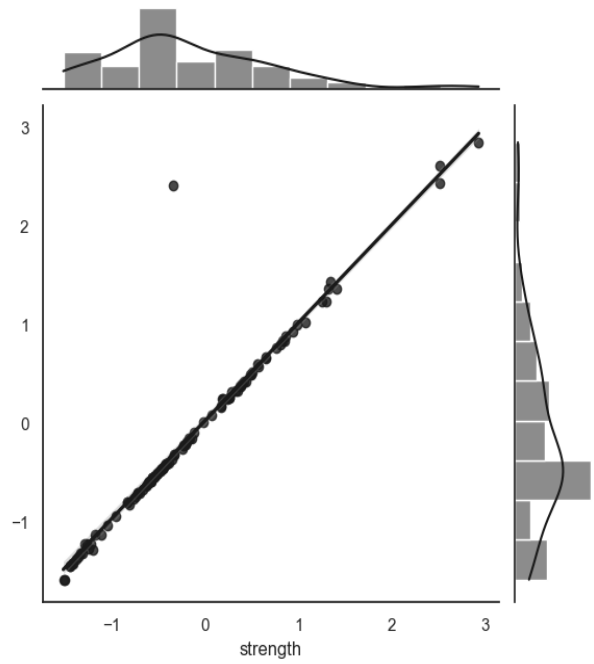


Figure 13. Decision Tree Regression Iteration-1

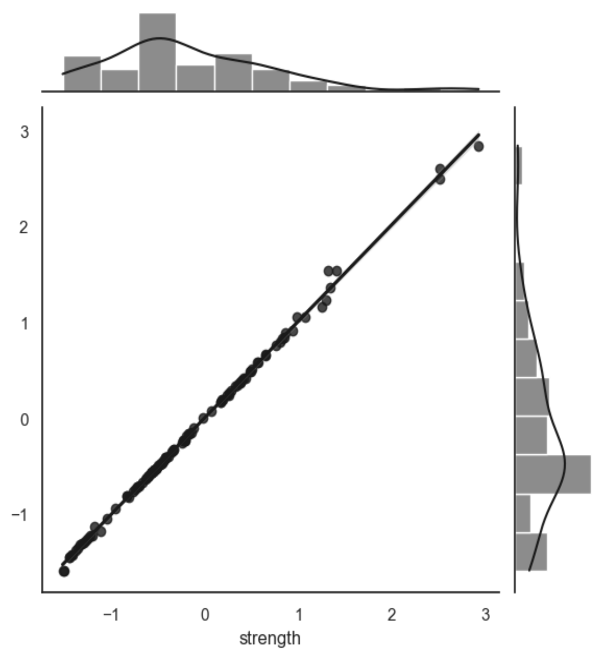


Figure 14. Decision Tree Regression Iteration-2

5.5. Boxplot by K Means Clustering

Another way of checking the data is this boxplot by K means clustering. In this study, we compare how well these ML models are able to predict outcomes. RMSE, MAE, and R^2 values for all models are shown in Fig. 3. Improved model performance is indicated by smaller RMSE and MAE values and larger R^2 values. When discussing performance, RMSE is typically used as the primary statistic for clarity's sake. Overfitting in the models may be discussed with greater ease because results from both testing and validation are presented. Boxplots are used to display the spread and variation of these performance metrics. Through a process of nested five-fold cross-validation, five testing values and twenty-five training values are used to identify the error sets for each model. In what follows, we will examine the methodological merits of each paradigm. Each model's performance is analyzed, and the underlying methodological and architectural factors that contributed to that performance are explored.

The Clustered graphs shown in figure 15 are very much identical for the 7 days and 14 days data. Also, the predicted strength data is aligned with the materials effect to produce sustainable concrete. The superplasticizer usage looked into different and with varied conditions.

5.6. Error Rate Analysis Considering KNN Regression

The error rate analysis is done considering the KNN regression process. KNN stands for K-Nearest Neighbor and this particular regression analysis is done to point out the mean error in one side and K value on the other axis. Example based algorithm that is considered used for both classification and regression is k-nearest neighbor (KNN) [45]. By identifying the k nearest data points to the lab testing object and then using the characteristics of those points to objectify the new classifications, the KNN method can quickly and accurately classify new data. To accomplish this, we calculate a distance between every example in the considered training dataset and the test facilities. The iteration values are considered for the input one and thus the mean error always remained 1.00 for most of the time with respect to K value. Figure 16 shows a clear aspect of the error considered in terms of regression analysis.

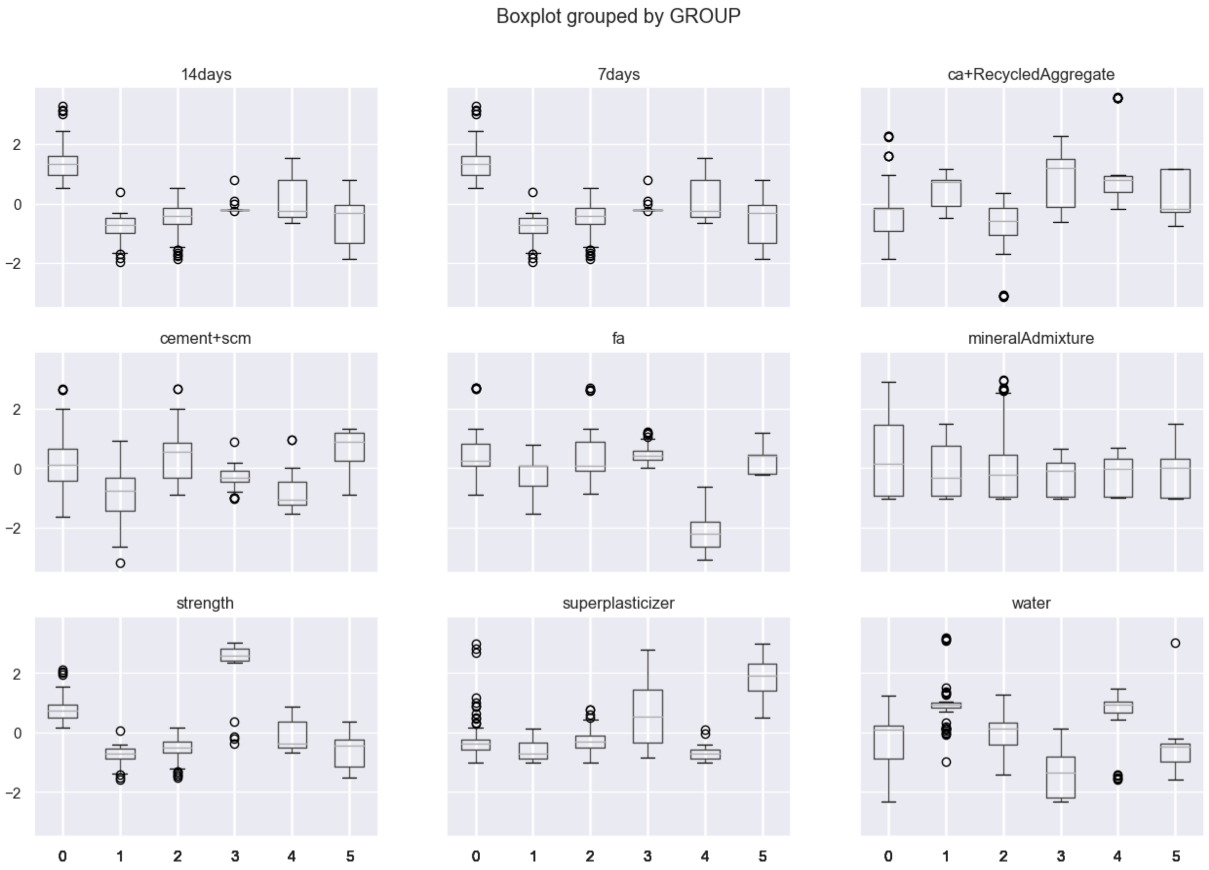


Figure 15. Boxplot by K Means Clustering

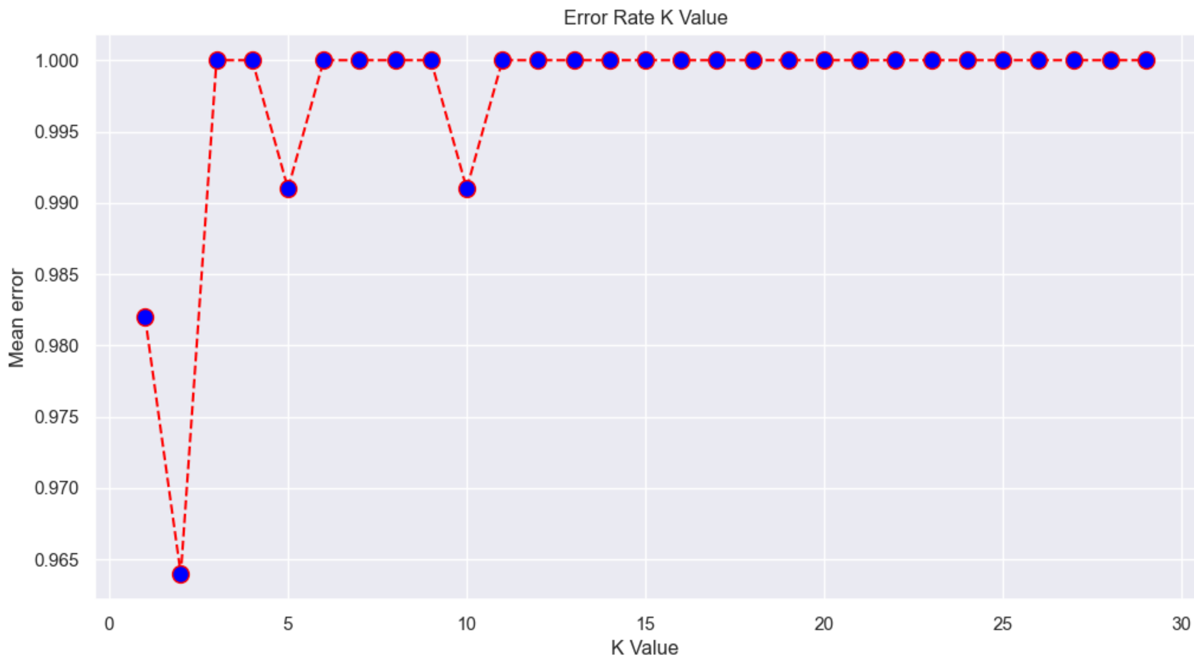


Figure 16. Error rate of KNN Regression

6. Conclusions

This investigation was conducted to predict the strength of compressive for sustainable concrete incorporated with various supplementary cementitious materials (SCMs), recycled aggregate, mineral admixtures etc. to understand the behavior of the hardened properties of concrete particularly compressive strength. Also, the predicted strength showed promising results derived from the hybrid ML model which refers to random forest with bagging algorithm (RF-B). The highlights of this study are:

- The incorporation of various SCMs really has the impact on the environmental perspective and also to cover up the carbon emission at a minimum extent.
- Some soft computing techniques like ANN, SVR and RF-B which are hybridized quite easily to utilize constructing a novel prediction model. These models were used to predict the strength variation in different usage consideration.
- By compiling relevant experimental data found in the public literature, a database of experiments was built. Multiple statistical performance metrics were studied to learn more about the predictability and accuracy/intelligence of a technique to predict concrete's compressive strength; this method could also be used to predict other engineering properties of cementitious materials.
- K-fold cross validation was done in different way considering KNN analysis also used to verify the high precision of the bagging and GB regressors.
- This research highlights the usefulness of supervised ML methods in the construction industry. The mechanical properties of concrete can be predicted with reasonable accuracy using these methods, saving valuable time and money over traditional laboratory experiments. It was also found that, in comparison to using separate algorithms, using an ensemble of machine learning algorithms produces more accurate predictions.
- Laboratory experiments can also be used to assess the models' efficacy by revealing the magnitude of discrepancies between the observed and projected outcomes.
- More than 20 sub-models can be trained on the same data to reduce variance, and optimization will yield the highest R^2 value possible using ensemble techniques. While some methods (like AdaBoost Regressor) can forecast outcomes and be used in comparisons, it is challenging to recommend or comment about any methodology directly based on a small number of trials.

REFERENCES

[1] Gledson, B. J. and Greenwood, D. "The adoption of 4D BIM in the UK construction industry: an innovation diffusion approach," *Engineering, Construction and Architectural Management*, 2017. https://doi.org/10.1108/ECAM-03-2016-0066

[2] Topçu, İ. B. and Sarıdemir, M. "Prediction of mechanical properties of recycled aggregate concretes containing silica fume using artificial neural networks and fuzzy logic," *Computational Materials Science,* vol. 42, no. 1, pp. 74-82, 2008. https://doi.org/10.1016/j.commatsci.2007.06.011

[3] Duan, Z.-H. Kou, S.-C. and Poon, C.-S. "Prediction of compressive strength of recycled aggregate concrete using artificial neural networks," *Construction and Building Materials,* vol. 40, pp. 1200-1206, 2013. https://doi.org/10.1016/j.conbuildmat.2012.04.063

[4] Naderpour, H. A. Rafiean, H. and Fakharian, P. "Compressive strength prediction of environmentally friendly concrete using artificial neural networks," *Journal of Building Engineering,* vol. 16, pp. 213-219, 2018. https://doi.org/10.1016/j.jobe.2018.01.007

[5] Mehta, P. K. "Greening of the concrete industry for sustainable development," *Concrete international,* vol. 24, no. 7, pp. 23-28, 2002.

[6] Yüksel, İ. Bilir, T. and Özkan, Ö. "Durability of concrete incorporating non-ground blast furnace slag and bottom ash as fine aggregate," *Building and Environment,* vol. 42, no. 7, pp. 2651-2659, 2007. https://doi.org/10.1016/j.buildenv.2006.07.003

[7] Aldea, C.-M. Young, Wang, F. K. and Shah, S. P. "Effects of curing conditions on properties of concrete using slag replacement," *Cement and concrete research,* vol. 30, no. 3, pp. 465-472, 2000.

[8] Çakır, Ö. and Aköz, F. "Effect of curing conditions on the mortars with and without GGBFS," *Construction and Building Materials,* vol. 22, no. 3, pp. 308-314, 2008. https://doi.org/10.1016/j.conbuildmat.2006.08.013

[9] Bagel, L. "Strength and pore structure of ternary blended cement mortars containing blast furnace slag and silica fume," *Cement and Concrete Research,* vol. 28, no. 7, pp. 1011-1022, 1998.

[10] Escalante-Garcıa, J. and Sharp, J. "The microstructure and mechanical properties of blended cements hydrated at various temperatures," *Cement and Concrete Research,* vol. 31, no. 5, pp. 695-702, 2001. https://doi.org/10.1016/S0008-8846(01)00471-9

[11] Ajdukiewicz, A. and Kliszczewicz, A. "Influence of recycled aggregates on mechanical properties of HS/HPC," *Cement and concrete composites,* vol. 24, no. 2, pp. 269-279, 2002.

[12] Khademi, F. and Behfarnia, K. "Evaluation of concrete compressive strength using artificial neural network and multiple linear regression models," 2016.

[13] Ryu, J. "An experimental study on the effect of recycled aggregate on concrete properties," *Magazine of concrete research,* vol. 54, no. 1, pp. 7-12, 2002. https://doi.org/10.1680/macr.2002.54.1.7

[14] Tu,T.-Y. Chen, Y.-Y. and Hwang, C.-L. "Properties of HPC with recycled aggregates," *Cement and concrete research,* vol. 36, no. 5, pp. 943-950, 2006.

https://doi.org/10.1016/j.cemconres.2005.11.022

[15] ASTM-D2419-14 (2014). Standard Test Method For Sand Equivalent Value Of Soils And Fine Aggregate. ASTM International. 2015 West Conshohocken, PA.

[16] ASTM-C33/C33M-18 (2018). Standard Specification for Concrete Aggregates. ASTM International. West Conshohocken, PA.

[17] ASTM-C1240-03 (2003). Standard Specification for Use of Silica Fume as a Mineral Admixture in Hydraulic-Cement Concrete, Mortar, and Grout. ASTM International. West Conshohocken, PA.

[18] Ahmadi-Nedushan, B. "Prediction of elastic modulus of normal and high strength concrete using ANFIS and optimal nonlinear regression models," *Construction and Building Materials,* vol. 36, pp. 665-673, 2012.

[19] Dao, D. V. Ly, H.-B. Trinh, S. H. Le, T.-T. and Pham, B. T. "Artificial intelligence approaches for prediction of compressive strength of geopolymer concrete," *Materials,* vol. 12, no. 6, pp. 983, 2019. https://doi.org/10.3390/ma12060983

[20] Khademi, F. Akbari, M. S. Jamal, M. and Nikoo, M. "Multiple linear regression, artificial neural network, and fuzzy logic prediction of 28 days compressive strength of concrete," *Frontiers of Structural and Civil Engineering,* vol. 11, no. 1, pp. 90-99, 2017. https://doi.org/10.1007/s11709-016-0363-9

[21] Khademi, F. Jamal, S. M. Deshpande, N. and Londhe, S. "Predicting strength of recycled aggregate concrete using artificial neural network, adaptive neuro-fuzzy inference system and multiple linear regression," *International Journal of Sustainable Built Environment,* vol. 5, no. 2, pp. 355-369, 2016. https://doi.org/10.1016/j.ijsbe.2016.09.003

[22] Mohammadhassani, M. Nezamabadi-Pour, H. Suhatril, M. and Shariati, M. "An evolutionary fuzzy modelling approach and comparison of different methods for shear strength prediction of high-strength concrete beams without stirrups," *Smart Struct. Syst,* vol. 14, no. 5, pp. 785-809, 2014. DOI: 10.12989/sss.2014.14.5.785

[23] Yuan, Z. Wang, L.-N. and Ji, X. "Prediction of concrete compressive strength: Research on hybrid models genetic based algorithms and ANFIS," *Advances in Engineering Software,* vol. 67, pp. 156-163, 2014. https://doi.org/10.1016/j.advengsoft.2013.09.004

[24] Jaafari, A. Panahi, M. Panahi, B. Panahi, T. Shahabi, H. Bui, D. T. F. Rezaie, and Lee, S. "Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility," Catena, vol.175, pp.430-445, 2019. https://doi.org/10.1016/j.catena.2018.12.033

[25] Mansour, M. Y. Dicleli, M. Lee, J.-Y. and Zhang, J. "Predicting the shear strength of reinforced concrete beams using artificial neural networks," *Engineering Structures,* vol. 26, no. 6, pp. 781-799, 2004. https://doi.org/10.1016/j.engstruct.2004.01.011

[26] Bui, D.-K. Nguyen, T. Chou, J.-S. Nguyen-Xuan, H. and Ngo, T. D. "A modified firefly algorithm-artificial neural network expert system for predicting compressive and tensile strength of high-performance concrete," *Construction and Building Materials,* vol. 180, pp. 320-333, 2018. https://doi.org/10.1016/j.conbuildmat.2018.05.201

[27] Behnood, A. and Golafshani, E. M. "Predicting the compressive strength of silica fume concrete using hybrid artificial neural network with multi-objective grey wolves," *Journal of Cleaner Production,* vol. 202, pp.54-64, 2018. https://doi.org/10.1016/j.jclepro.2018.08.065

[28] Yang, X.-S. *Nature-inspired metaheuristic algorithms*: Luniver press, 2010.

[29] Lee, S.-C. "Prediction of concrete strength using artificial neural networks," *Engineering structures,* vol. 25, no. 7, pp. 849-857, 2003. https://doi.org/10.1016/S0141-0296(03)00004-X

[30] Yeh, I.-C. "Modeling of strength of high-performance concrete using artificial neural networks," *Cement and Concrete research,* vol. 28, no. 12, pp. 1797-1808, 1998. https://doi.org/10.1016/S0008-8846(98)00165-3

[31] Metaxiotis, K., Psarras, J., & Samouilidis, E. (2003). Integrating fuzzy logic into decision suppport systems: current research and future prospects. Information management & computer security, 11(2), 53-59.

[32] Pala, M. Özbay, E. Öztaş, A. and Yuce, M. I. "Appraisal of long-term effects of fly ash and silica fume on compressive strength of concrete by neural networks," *Construction and Building Materials,* vol. 21, no. 2, pp. 384-394, 2007. https://doi.org/10.1016/j.conbuildmat.2005.08.009

[33] Akkurt, S. Ozdemir, S. Tayfur, G. and Akyol, B. "The use of GA–ANNs in the modelling of compressive strength of cement mortar," *Cement and concrete research,* vol. 33, no. 7, pp. 973-979, 2003.

[34] Demir, F. "A new way of prediction elastic modulus of normal and high strength concrete—fuzzy logic," *Cement and Concrete Research,* vol. 35, no. 8, pp. 1531-1538,2005. https://doi.org/10.1016/j.cemconres.2005.01.001

[35] Lai, S. and Serra, M. "Concrete strength prediction by means of neural network," *Construction and Building Materials,* vol. 11, no. 2, pp. 93-98, 1997. https://doi.org/10.1016/S0950-0618(97)00007-X

[36] Öztaş, A., Pala, M., Özbay, E., Kanca, E., Çağlar, N., and Bhatti, M. A. "Predicting the compressive strength and slump of high strength concrete using neural network," *Construction and building materials,* vol. 20, no. 9, pp.769-775, 2006. https://doi.org/10.1016/j.conbuildmat.2005.01.054

[37] Sebastiá, M. Olmo, I. F. and Irabien, A. "Neural network prediction of unconfined compressive strength of coal fly ash–cement mixtures," *Cement and Concrete Research,* vol. 33, no. 8, pp. 1137-1146, 2003. https://doi.org/10.1016/S0008-8846(03)00019.

[38] Gualtieri, J.A., Chettri, S.R., Cromp, R.F., and Johnson, L.F. 1999. Support vector machine classifiers as applied to AVIRIS data. In Proceedings of the 8th Jet Propulsion Laboratory Airborne Earth Science Workshop. Pasadena, Calif., 8–11 February 1999.

[39] Dibike, Y. B. Velickov, S. Solomatine, D. and Abbott, M. B. "Model induction with support vector machines: introduction and applications," *Journal of Computing in Civil Engineering,* vol. 15, no. 3, pp. 208-216, 2001.

https://doi.org/10.1061/(ASCE)0887-3801(2001)15:3(208)

[40] Pal, M. "Ensemble of support vector machines for land cover classification," *International journal of remote sensing*, vol. 29, no. 10, pp. 3043-3049, 2008. https://doi.org/10.1080/01431160802007624

[41] Smola, A. J. and Schölkopf, B. "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no.3, pp.199-222, 2004. https://doi.org/10.1023/B:STCO.0000035301.49549.88

[42] Dietterich, T.G. (2000). Ensemble Methods in Machine Learning. In: Multiple Classifier Systems. MCS 2000. Lecture Notes in Computer Science, vol 1857. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45014-9_1

[43] Liaw, A. and Wiener, M. "Classification and regression by randomForest," *R news*, vol. 2, no. 3, pp. 18-22, 2002.

[44] Schapire, R. E. "The boosting approach to machine learning: An overview," *Nonlinear estimation and classification*, pp. 149-171, 2003.

[45] Quinlan, J.R. (1992) Learning with Continuous Classes. *Proceedings of Australian Joint Conference on Artificial Intelligence*, Hobart 16-18 November 1992, 343-348.

[46] Deo, R. C. Kisi, O. and Singh, V. P. "Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model," *Atmospheric Research*, vol. 184, pp. 149-175, 2017. https://doi.org/10.1016/j.atmosres.2016.10.004

[47] Rahimikhoob, A. Asadi, M. and Mashal, M. "A comparison between conventional and M5 model tree methods for converting pan evaporation to reference evapotranspiration for semi-arid region," *Water resources management*, vol. 27, no. 14, pp. 4815-4826, 2013. https://doi.org/10.1007/s11269-013-0440-y

[48] Bhattacharya, B. and Solomatine, D. P. "Neural networks and M5 model trees in modelling water level–discharge relationship," *Neurocomputing*, vol. 63, pp. 381-396, 2005. https://doi.org/10.1016/j.neucom.2004.04.016Get

[49] Kisi, O. "Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree," *Journal of Hydrology*, vol. 528, pp. 312-320, 2015. https://doi.org/10.1016/j.jhydrol.2015.06.052

[50] Breiman, L. "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123-140, 1996. https://doi.org/10.1007/BF00058655

[51] Chen, P.-L. Tsai, C.-T. Chen, Y.-N. Chou, K.-C. Li, C.-L. Tsai, C.-H. Wu, K.-W. Chou, Y.-C. Li, C.-Y. and Lin, W.-S. "A linear ensemble of individual and blended models for music rating prediction." *KDD Cup*, pp. 21-60, 2012.

[52] Farrar, D. E. and Glauber, R. R. "Multicollinearity in regression analysis: the problem revisited," *The Review of Economic and Statistics*, pp. 92-107, 1967. https://doi.org/10.2307/1937887