Arabian Journal for Science and Engineering
https://doi.org/10.1007/s13369-024-09042-1

RESEARCH ARTICLE-CIVIL ENGINEERING

Efficient machine learning models for estimation of compressive strengths of zeolite and diatomite substituting concrete in sodium chloride solution

Giyasettin Ozcan[1] · Burak Kocak[2] · Eyyup Gulbandilar[3] · Yilmaz Kocak[4]

Received: 19 December 2023 / Accepted: 28 March 2024
© The Author(s) 2024

Abstract
This study implements a set of machine learning algorithms to building material science, which predict the compressive strength of zeolite and diatomite substituting concrete mixes in sodium chloride solution. Particularly, Random Forest, Support Vector Machine, Extreme Gradient Boosting, Light Gradient Boosting, and Categorical Boosting algorithms are exploited and their optimal parameters are tuned. In the training and testing of these models, 28 day, 56 day, and 90 day compressive strength observations of 63 samples of 7 different concrete mixtures substituting Portland cement, zeolite, diatomite, zeolite + diatomite were used. Consequently, compressive strength experimentation results and machine learning predictions were compared through statistical methods such as RMSE, MAPE, and R^2. Results denote that the prediction performance of machine learning is improving with tuned models. Particularly, RMSE, MAPE, R^2 scores of Categorical Boosting are, respectively, 1.15, 1.45%, and 98.03% after parameter tuning design. The results denote that presented machine learning model can provide an advantage in the cost and duration of the compressive strength experiments.

Keywords Zeolite · Diatomite · Compressive strength · Random forest · Gradient boosting · Machine learning

1 Introduction

Conventional concrete is generally not successful in preventing the ingress of moisture and aggressive ions due to its porous structure [1]. For this reason, the use of supplementary cementitious materials in concrete is becoming more common day by day [2–6]. Mineral additives such as zeolite and diatomite are among the important building materials used extensively in cement and concrete technology, due to their economic and ecological advantages as well as contributing to the improvement of properties such as durability and mechanical properties of concrete [7–11]. For this reason, it is necessary to understand mechanical properties of concrete structures under external loads. In this context, compressive strength, which is directly concerned with the safety of structures, is one of the most basic mechanical properties, and necessary in determining the performance of structures. The compressive strength of concrete is affected by many factors such as cement type, mineral and chemical additives, mixing ratios, curing conditions, mixing, transportation, placement and testing methods of concrete [2, 4, 5, 12–14]. To determine the compressive strength value of concrete, cubic or cylindrical samples are produced according to certain mixing ratios, cured under suitable conditions and the results of the compressive strength are determined by breaking them at certain times. Nevertheless, these experiments are costly in terms of both time and economy. For this purpose, different from traditional experimental methods and some artificial intelligence methods are used to predict the compressive strength of concrete with mixing ratios in concrete design.

To reduce time and economic costs one alternative option is machine learning, where studies utilized machine learning algorithms to solve a wide variety of problems in cement and concrete technology including fuzzy logic (FL), artificial neural networks (ANN) and adaptive neuro-fuzzy inference

✉ Giyasettin Ozcan
gozcan@uludag.edu.tr

1 Department of Computer Engineering, Faculty of Engineering, Bursa Uludag University, Bursa, Turkey

2 Department of Computer Engineering, Institute of Postgraduate Education, Duzce University, Duzce, Turkey

3 Department of Computer Engineering, Faculty of Engineering and Architecture, Osmangazi University, Eskisehir, Turkey

4 Department of Civil Engineering, Faculty of Engineering, Duzce University, Duzce, Turkey

Published online: 18 April 2024

system (ANFIS) [1, 15–20]. These studies aimed to determine the target feature's best prediction accurately.

Machine learning, which imitates human behavior, introduces a wide spectrum of models, be due to input data features [21]. In contrast to the standard algorithmic approach, ML models do not explicitly define the instructions that solve a problem. Instead, an algorithm is constructed via statistical analysis of input data [22]. During statistical analysis, the input data are split into two parts. These are defined as response and predictor variables. Hereof, measured observation to be predicted is defined as a response, whereas measurable variables that affect response are defined as predictors. A machine learning algorithm finds functions which input predictor variables and outputs a response variable. For this goal, various machine learning algorithms have been developed. Some of them exploit the statistical properties of input [23], whereas others divide the space via multidimensional geometrical tools such as lines, or hyperplanes [24].

One alternative machine learning algorithm is defined as ensemble learning which utilizes weak learners. The algorithm questions whether weak learners may yield good performance when there's diversity among different learners. Machine learning algorithms such as Random Forest (RF) [25], Support Vector Machines (SVM) [24], and Boosting [26] present efficient ensemble learning.

Random Forest is composed of decision trees where each tree fits to a subset, which is sampled independently through bootstrapping. Since independent sampling, Random Forests reduce overfitting and improve accuracy. Like Random Forest, gradient boosting is a tree-based model and an ensemble of weak learners. They focus on residuals and improve residuals via boosting. In recent years, several gradient boosting algorithms have been presented which presented accurate results [27]. These algorithms are Extreme Gradient Boosting (XGBoost) [28], Light Gradient Boosting (LightGBM) [29] and Categorical Boosting (CatBoost) [30]. The Gradient Boosting algorithms have become popular in literature since their fast and accurate prediction properties [31, 32]. These researchers claim that gradient boosting may handle large and complex data efficiently. Therefore, the algorithms are introducing cost efficiency.

These algorithms, which have gained popularity recently due to these positive features, are used to predict different properties of concrete. To name a few, Zhang et al. [33] studied the application of a beetle antennae search (BAS) algorithm based on random forest (RF) in predicting the compressive strength of lightweight self-compacting concrete. The results show that the BAS-RF model achieves high-prediction accuracy. Abunassar et al. [34] adopted SVM to estimation the compressive strength in self-compacting concrete containing fly ash and silica fume. Feng et al. [35] applied ANN, SVM and AdaBoost for compressive strength prediction for concrete. According to the results obtained from the modeling, they stated that the AdaBoost algorithm was adopted to estimate the compressive strength of concrete. Nguyen and Hoang [36] used a metaheuristic-optimized Extreme Gradient Boosting Machine and Deep Convolutional Neural Network to predict computer vision-based classification of concrete spall severity. According to their results, it is stated that XGBoost provides outstanding classification performance.

As can be seen from the literature survey, it is seen that many artificial intelligence methods are used due to both the complex structure of concrete in terms of its components and economic and ecological reasons. Therefore, in this study, instead of traditional algorithms such as FL, ANN, ANFIS, which have been used extensively to predict concrete compressive strength, SVM, RF, XGBoost, LightGBM and CatBoost algorithms, which have attracted attention recently, were preferred. In addition, it is thought that the prediction of the compressive strengths of concretes with different concrete contents and cured under different conditions will contribute to the literature. In this study, these machine learning algorithms were used to predict the compressive strength of zeolite and/or diatomite substituted concrete mixtures cured in sodium chloride solution. Firstly, a total of 63 concrete specimens of 7 different mixtures were prepared in the laboratory, and their compressive strengths at 28, 56 and 90 day were determined. Then, the models were trained with 42 experimental outcome data and tested with 21 experimental outcome data. Seven parameters including zeolite, diatomite, PC, aggregate, water, hyper plasticiser and curing time (days) were used as predictor values, and compressive strength value was used as response.

According to data features and analysis of recent Machine Learning research in literature, the most convenient Machine Learning Algorithms are investigated. By tuning parameters of the algorithms and data design, efficient Machine Learning models are found and compared. Thus, designed machine learning models are utilized for the prediction of compressive strength of concrete in building material science.

2 Experimental procedure

The main materials of the study were zeolite, diatomite, Portland cement, aggregate, water, and hyperplasticizer. Portland cement, which is defined as CEM I 42.5 R, acquired from Bolu Cement Plant and complies with TS EN 197-1 requirement, is exploited for material binding [37]. Zeolite and diatomite used as supplementary cementitious materials were supplied from Turkish Zeolite and ASU Chemistry and Mining Firms, respectively. The chemical compounds with ARL 8680 S model X-ray spectrometer (XRF) device were determined. The surface area with Toni Technik 6565

Table 1 Chemical, mechanical and physical properties of zeolite, diatomite and PC

		Zeolite	Diatomite	PC
Chemical compounds, wt%	SiO_2	68.85	79.56	18.68
	Al_2O_3	11.71	6.54	4.67
	Fe_2O_3	1.29	2.76	3.53
	CaO	3.97	2.45	64.56
	MgO	1.06	0.79	0.98
	SO_3	0.18	0.48	3.00
	Na_2O	0.29	2.63	0.14
	K_2O	2.19	0.69	0.73
	Free CaO	–	–	1.74
	Loss on ignition	10.00	3.88	3.92
Mechanical properties (compressive strength), MPa	7 day	–	–	29.6
	28 day	–	–	52.8
Physical properties	Blaine, cm^2/g	5740	13,640	4249
	Specific gravity	2.18	2.28	3.17

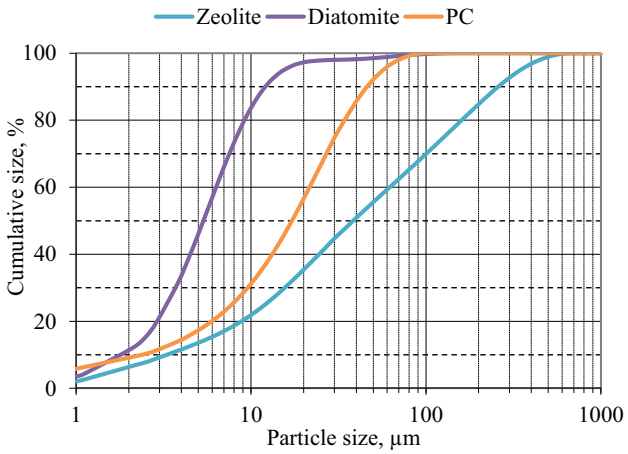


Fig. 1 Particle size distributions of zeolite, diatomite, and PC

model Blaine device and specific gravity with Quantachrome MVP-3 model device were determined. Table 1 presents the chemical, mechanical and physical properties of zeolite, diatomite and PC.

Table 1 denotes that zeolite and diatomite are primarily composed of silicon dioxide (SiO_2), where the mass fractions of $SiO_2 + Al_2O_3 + Fe_2O_3$ are 81.85 and 88.86% for zeolite and diatomite, respectively. This complies with the stated requirements ($SiO_2 + Al_2O_3 + Fe_2O_3 > 70\%$) for a natural pozzolan [38, 39]. The definite weights of the PC, zeolite, and diatomite are 2.18, 2.28 and 3.17 g/cm^3, whereas their specific surfaces (Blaine) are, respectively, 5740, 13,640 and 4249 cm^2/g (Table 1).

Particle size distribution determined with Malvern Hydro 2000 G device of zeolite, diatomite and PC is given in Fig. 1.

The figures of the particle size analysis prove that diatomite has finer particle size. According to their 90% sub-sieve fractions, the particle size of zeolite, diatomite and PC is 280, 13 and 45 µm, respectively.

XRD analyses that disclose the mineralogical structure with Rikagu miniflex model via XRD device using Cu K_α ($\lambda = 1.54$ A°) radiation of zeolite, diatomite and PC were presented in Fig. 2.

As observed from the XRD patterns of zeolite, it has a crystal structure of quartz (SiO_2) and clinoptilolite ($(Na,K)_6[Al_6Si_{30}O_{72}]20H_2O$) (Fig. 2a) [40]. In silica form, diatomite is composed of cristobalite (SiO_2) and has an amorphous structure (Fig. 2b) [41]. PC is in crystal form, and its fundamental structure is composed of alite–tricalcium silicate ($3CaOSiO_2$), belite–dicalcium silicate ($2CaOSiO_2$), tricalcium aluminate ($3CaOAl_2O_3$) and tetracalcium aluminoferrite ($4CaOAl_2O_3Fe_2O_3$) (Fig. 2c) [6].

In the concrete mixture design, crushed sand and crushed stone obtained from the aggregates of Asar Stream in Düzce region were used. Aggregate's physical characteristics were introduced in Table 2.

In order to produce concrete specimens, a new generation hyper plasticizer (fluid 70) is provided by AYDOS Construction Chemicals Factory. Here, the characteristics of the provided hyper plasticizer are disclose as follows: solid matter content 34.32%, intensity 1.184 (20 °C), pH value 7.26 (20 °C). The well water of Doğanlı village in Düzce was utilized as the mixing water, which is free of impurities. Selected water is used for concrete specimen preparation which would retard the normal setting of concrete.

In the study, seven different concrete mixtures manufactured using CEM I type (PC 42.5 R), and partial replacement

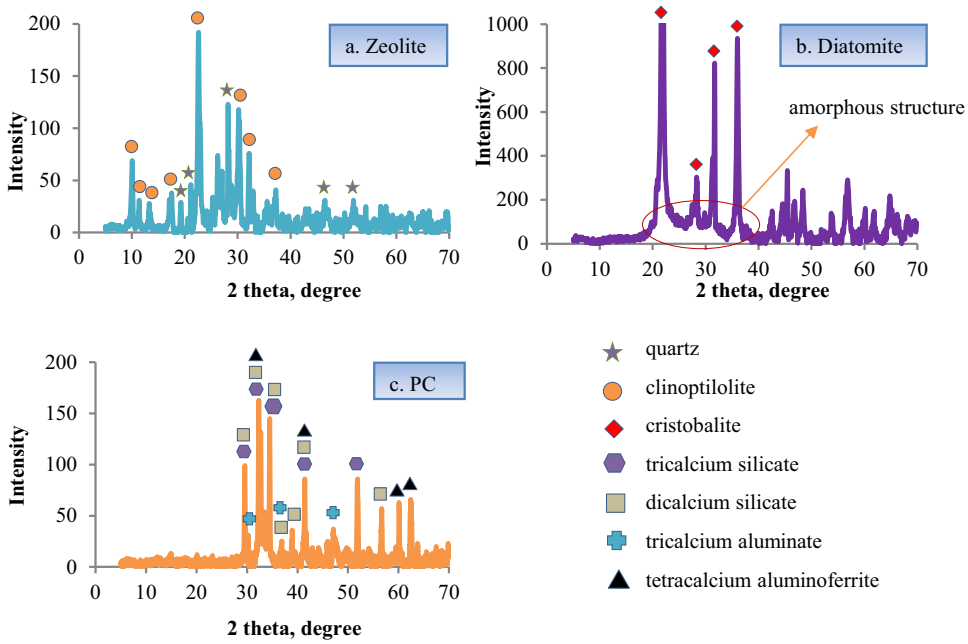


Fig. 2 XRD analysis of zeolite, diatomite and PC

Table 2 Aggregate's physical characteristics

	Aggregate grading		
	0–5, mm	5–19, mm	19–30, mm
Moisture content, %	1.25	1.32	1.41
Water absorption, %	0.61	1.16	1
Saturated and surface-dry weight	2.64	2.65	2.69
Dry weight	2.63	2.62	2.66
Dense Unit Weight, g/cm^3	1.66		
Loose Unit Weight, g/cm^3	1.48		
Organic impurities	Organic matter is harmless (the color of the liquid is from colorless to light yellow)		

of PC by zeolite and/or diatomite were used. Reference concrete mixture, named as R, was made with 100% PC. Zeolite substituted concrete mixtures, named as 10Z and 20Z, were made with 90% PC and 10% zeolite, and 80% PC and 20% zeolite. Diatomite substituted concrete mixtures, named as 10D and 20D, were made with 90% PC and 10% diatomite, and 80% PC and 20% diatomite. Zeolite and diatomite substituted concrete mixtures, named 5Z5D and 10Z10D, were made with 90% PC and 5 + 5% zeolite and diatomite, and 80% PC and 10 + 10% zeolite and diatomite. For concrete mixture design, materials' amounts to be put into the mixture have been determined within the framework of the method stated in TS 802 standard [42]. According to TS EN 12350–2, the consistency of concrete in fresh concrete has been stated for each mixing group separately [43]. While the samples are being prepared for this study, it is aimed to make a mixture design with a slump between 7 and 10 cm. With this

Table 3 Material proportion of each concrete group per 1 m^3

Materials		R, kg	10Z, kg	20Z, kg	10D, kg	20D, kg	5Z5D, kg	10Z10D, kg
Aggregate, mm	0–5	822	843	855	831	822	849	855
	5–19	586	602	611	593	586	606	611
	19–30	428	439	446	433	428	442	445
Total		1836	1884	1912	1857	1836	1897	1911
Zeolite		–	40	80	–	–	20	40
Diatomite		–	–	–	40	80	20	40
PC		400	360	320	360	320	360	320
Water		139.7	139.7	123.2	139.7	123.3	139.7	124.2
Hyper plasticizer		4.800	4.320	4.800	4.32	4.800	4.320	3.840

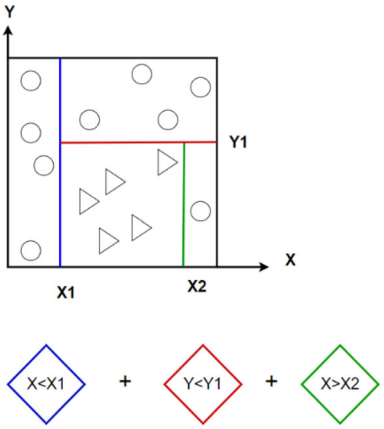


Fig. 3 Simple visualization of decision tree

mix design, at least C30/37 (cylinder/cubic specimens), class concrete according to TS EN 206–1 was tried to be obtained [44]. For these purposes, the material quantities of 1m^3 fresh concrete that will provide these slump quantities and C30/37 class concrete requirements are given in Table 3.

The slump test results of the concretes produced according to this mix design were determined as 7 cm for PC, 10D, 10Z and 20 Z coded concretes, 7.8 cm for 10D10Z coded concrete, 8 cm for 20D coded concrete and 9.3 cm for 5D5Z coded concrete.

The produced concrete specimens with a size of 15 × 15 × 15 cm^3 cubic molds were poured. The specimens were stored in the mold for 24 h. After this period, the concrete specimens were taken out from the mold and cured in tap water at 23 ± 2 °C for 28 day. After the curing, the specimens were subjected to full immersion 5 wt% sodium chloride (NaCl) solution at 23 ± 2 °C for 56 day, and 90 day. Finally, a total of 63 compressive strength results obtained from three concrete samples for each concrete type and hydration day were determined according to the TS EN 12390–3 standard [45].

In the next step of this research, it is aimed to predict the compressive strength of these concrete samples through the compressive strength results obtained and efficient machine learning models.

3 Modeled machine learning algorithms

Machine learning aims to imitate human learning by utilizing data and algorithms. Detailed information about machine learning can be found in literature [21]. Formally, given that there exists N training examples in dataset D as Eqs. (1):

$$D = \{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\} \quad (1)$$

where x_i is the feature vector of ith training sample and y_i is the label. Machine learning seeks an approximation function $f(x)$ which aims to map xi to y_i. For this goal, it is necessary to minimize expected loss function, $L(y, \mathrm{F}(x))$.

In the course of time, novel machine, learning algorithms started to minimize $L(y, \mathrm{F}(x))$ and therefore introduced more accurate prediction performance. Depending on the type of data features, different learning algorithms are presented. Recently, Support Vector Machines [24], Random Forest [25], Generalized Boosted Regression [26], and Gradient Boosting [27] algorithms introduced promising accuracy performance for continuous tabular datasets.

Support Vector Machines are introduced by [24] and consist of hyperplanes, where each hyperplane divides space into two. SVM maps data into high-dimensional space when data are not linearly separable. In terms of continuous data, Support Vector Regression, SVR, predicts the relationship between input features and continuous target. For this goal, SVR aims to minimize error Support Vector Regression use the same principles.

Random Forest (RF), based on multiple decision trees, is an alternative technique, where tree-like model of decisions considers chances of outcomes [25]. Furthermore, RF is an ensemble method and computes the consensus of multiple decision trees, where split strategy of each tree is based on a random subset of features. Such a design of RF reduces variance and improves accuracy. On the other hand, bias increases and interpretability of the model lessens. Simple visualization of decision tree is given in Fig. 3.

Another ensemble technique, Boosting, introduces a different solution by utilizing weak learners. In each iteration, a weak learner f_t takes input object x_i and maps to y_i. Each weak learner, f_t, produces a hypothesis $h(x_i)$ that minimizes the current error by means of determining optimal coefficient. The summation of the weak classifiers denotes machine learning output (Fig. 3).

Advantages of Boosting and Random Forest paves the path for Gradient Boosted Machines [27]. Recently Gradient Boosting Algorithms have been introduced. The Gradient Boosting Algorithms combine the pros of Random Forest and Boosting. For this goal, Gradient Boosting combine weak learners as in Boosting and exploits multiple decision trees [31]. In other words, it builds an additive approximation of F_i (x) (Eqs. (2)). Formally,

$$F_m(x) = F_{m-1}(x) + \rho_m h_m(x) \quad (2)$$

where hypothesis of the m^{th} step is multiplied by weight function, ρ_m. In contrast to Random Forest, Gradient Boosting computes the loss function, $\mathrm{L}(y, \mathrm{F}_m(x))$, value of m^{th} iteration, The Gradient Boosting aims to compensate loss of mth iteration at $(m+1)^{\text{th}}$ step. The algorithm aims to reduce loss after necessary steps as denoted in Fig. 4.

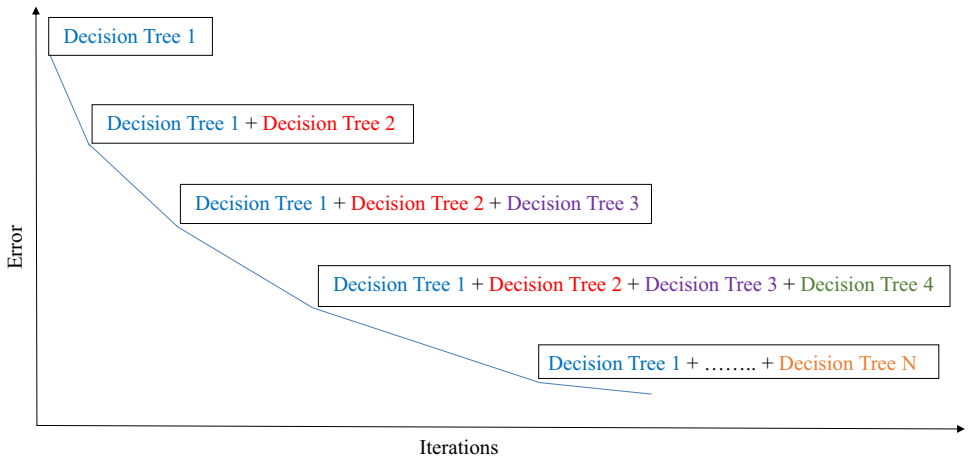


Fig. 4 Reducing error by gradient boosting

Table 4 The input and output quantities in machine learning

Type	Parameter	Unit	Minimum	Maximum	Mean	Std. error	Std. deviation
Input	Zeolite	kg/m^3	0	80	25.714	4.326	28.037
Input	Diatomite	kg/m^3	0	80	25.714	4.326	28.037
Input	PC	kg/m^3	320	400	348.571	4.372	28.334
Input	Aggregate	kg/m^3	1836	1912	1876.143	4.791	31.047
Input	Water	kg/m^3	123.2	139.7	132.786	1.248	8.086
Input	Hyper plasticizer	kg/m^3	3.840	4.800	4.457	0.053	0.340
Input	Curing time	Day	28	90	58.000	3.959	25.658
Output	Compressive strength	MPa	42.8	68.5	54.641	1.281	8.300

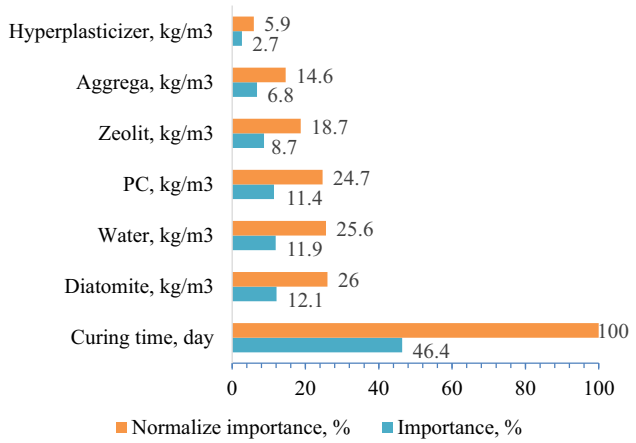


Fig. 5 Degree of importance of the normalized and non-normalized input variables

While Gradient Boosting is recently efficient, the first gradient boosting implementation suffered from overfitting, and the iterative process was regulated poorly.

An alternative gradient boosting algorithm, XGBoost is presented by [28]. The algorithm reduces the decision tree structure and utilizes randomization strategies. It maintains a pruning strategy for better loss function.

Gradient boosting enables "learning to rank" feature, which is very important for search engines. Consequently, gradient boosting studies attracted search engine developers as well.

To compensate for flaws of gradient boosting, Light GBM and the corresponding library are presented by [29]. Like XGBoost, Light GBM introduced tools that enable early stopping, sparse tree optimization, and multiple loss functions. Particularly, Light GBM introduces new features that introduce a new sub-sampling technique that generates tree samplings for the ensembles. Furthermore, the features convert sparse features into a compact form. Gradient boosting utilizes decision trees.

Finally, CatBoost is presented as an alternative gradient boosting method [30]. CatBoost builds balanced trees during extension, where balanced trees yield efficiency during processing numeric, categorical, or text features. To prevent overfitting, CatBoost introduces ordered boosting, where training data are processed on a subset of data. At the same time, residuals are computed on another subset.

Table 5 10-fold Cross-Validation of most efficient (CatBoost) models' iteration results for training and testing

k-fold	Iteration	Train	Test	Total time (ms)	Besttest	Bestiteration
0/10	0	50.1347409	53.0879136	0.290	8.287968198	99
	50	2.6935974	8.5319438	12.9		
	99	1.0549624	8.2879682	27.4		
1/10	0	49.9466672	54.5779848	0.318	8.72175192	99
	50	2.7659658	8.9588626	12.3		
	99	0.9745736	8.7217519	27.3		
2/10	0	50.4969344	49.7301024	0.334	1.8472464	92
	50	3.1287471	4.5034862	13.3		
	99	0.9873375	1.8563985	30.2		
3/10	0	50.9477169	44.5529578	0.268	1.8138280	99
	50	2.7856656	3.1270102	12.5		
	99	1.0026561	1.8138280	27.5		
4/10	0	50.1587291	53.4680173	0.280	1.7976718	99
	50	2.9678633	5.1698082	14		
	99	1.0715407	1.7976718	30.8		
5/10	0	50.0394238	54.3601449	0.285	3.0140160	99
	50	2.9470848	6.5135215	14.1		
	99	1.0197767	3.0140160	30.6		
6/10	0	50.5963860	48.8908573	0.276	1.947215478	99
	50	2.8694671	4.4287598	13.8		
	99	1.0324651	1.9472155	28.2		
7/10	0	50.5910319	49.0392438	0.282	2.3638611	99
	50	2.6090235	3.5660049	14.6		
	99	0.9484628	2.3638611	34.7		
8/10	0	50.5561449	49.0398305	0.284	2.3368041	99
	50	2.9200877	3.3825799	15.4		
	99	0.9408362	2.3368041	29.8		
9/10	0	50.9511071	44.9380104	0.280	1.345387398	99
	50	2.7871325	2.4953553	14.6		
	99	1.0247966	1.3453874	33.7		

The performance of the algorithms also depends on the optimal parameter selection. The learning rate of gradient boosting is one of the prominent parameters. In order to reduce loss, new trees are added during Gradient Boosting. The weight of each tree is controlled by learning rate. Loss function, tree depth, L2 regularization and iterations are also other prominent parameters of gradient boosting.

4 Experimental design and model parameters

In this study, R programming language and necessary machine learning packages were used to estimate the compressive strength of zeolite and/or diatomite substituted concrete mixtures cured in sodium chloride solution. In order to create machine learning models, a total of 63 concrete samples belonging to 7 different mixtures were prepared in the laboratory, and the compressive strengths of these samples were determined. The models of machine learning the models were trained with 42 data from experimental results; the remaining 21 data were exploited for testing. For train and test of these models, seven parameters including zeolite, diatomite, PC, aggregate, water, hyper plasticizer and curing time (day) were employed as input variables, and one parameter including compressive strength was used as output variable (Table 4).

IBM SPSS 22.00 package program was used for statistical evaluations in this study. Basic statistical parameters of the input and output variables such as minimum, mean, maximum and standard deviation were determined (Table 4).

The significance level was presented to reveal the effect of both normalized and non-normalized input variables on the output variable (Fig. 5). While curing time and water are the main factors of all concretes, the contributions of diatomite and zeolit are decent.

According to Fig. 5, it can be stated that curing time (day) provides the most significant contribution to the output variable, and the hyperplasticizer provides the lowest contribution. It can be said that other input variables contribute approximately equally.

To predict compressive strength, five different machine learning are utilized. Particularly, RF, SVM, XGBoost, LightGBM and CatBoost, were implemented, and compressive strength of zeolite and diatomite substituted concrete mixtures cured in sodium chloride solution were predicted. Random Forest is preferred since it handles overfitting via averaging multiple decision trees. Consequently, the algorithm itself is more robust against noise. On the other hand, SVM is able to handle high-dimensional data with small size. Therefore, properties of SVM are compatible with the research. Finally, three Gradient Boosting algorithms are modeled since their recent appealing accuracy performance for multidimensional data. Moreover, Gradient Boosting enables a wide range of parameter tuning options.

Fundamental parameters of the machine learning algorithms are investigated. Optimal parameters are found for each compressive strength prediction algorithm. Tested parameters depend on machine learning algorithms. In terms of Random Forest, a number of trees are tuned. In terms of SVM, default parameters of e1071 package are implemented. Learning rate, maxdepth parameters are tuned for the XGBoost and learning rate is tuned for LightGBM. On the other hand, learning rate, depth, loss function, and metric period parameters are tuned for CatBoost. Unless tuned, Default parameters are used in each algorithm.

Cross-Validation is implemented during model development. The Cross-Validation procedure resamples dataset for evaluation. The procedure also helps to flag overfitting and selection bias. After investigating various models, the most efficient (CatBoost) models' tenfold Cross-Validation is presented to estimate the best models' prediction ability for an independent dataset (Table 5).

Table 5 denotes that Total Computation Time of tenfold cross-validation is reasonable. The results present that low errors exist in both training and validation sets, which implies that model is well adapted to both training and validation sets. Differences in training and validation errors vary in each fold. While some folds have higher errors, the rest of them yield lower errors. Results denote that the best performance for each fold usually occurs at the last iteration. In general, developed model satisfactorily fits the generalization.

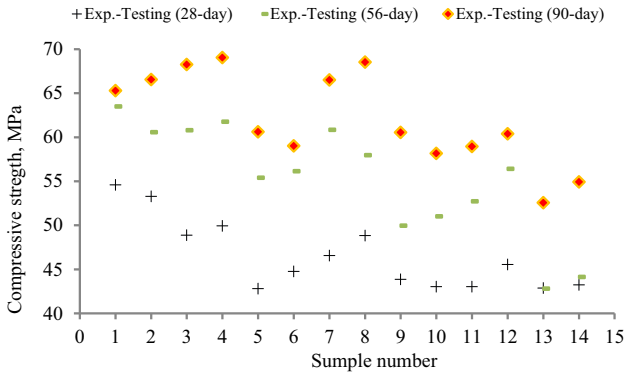


Fig. 6 Compressive strength results for training machine learning models

5 Results and discussion

Performance evaluation of machine learning models has been implemented by three indicators, which are root-mean squared error (RMSE), mean absolute percentage error (MAPE) and *R*-squared (R^2). The indicators are calculated with the following Eqs. (3), (4), and (5), respectively [46, 47].

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - r_i)^2} \tag{3}$$

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{N}\left|\frac{r_i - y_i}{r_i}\right| \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - r_i)^2}{\sum_{i=1}^{N}(r_i - y_m)^2} \tag{5}$$

N in the equations indicates the total number of data sets used, y_i is the estimated value, r_i is the actual value, and y_m is the average of the actual values.

For training machine learning models, 42 of the randomly selected experimental data were utilized, whereas the remaining 21 experimental data were employed for testing. In order to present detailed information, sample numbers and experimental results for training were given in Fig. 6.

Experimental results with predictions of machine learning models were presented in Fig. 7.

Also, sample number, concrete group, input values, experimental results, and predictions of machine learning models were given in Table 6.

Comparison of compressive strength experimental results with prediction results of the machine learning models were given in Fig. 8.

In Fig. 8, the linear least square fit line, corresponding linear equation, and the R^2 values are denoted. The figures denote that model predictions are close to the experimental outputs. Particularly, Fig. 8e shows that the CatBoost model

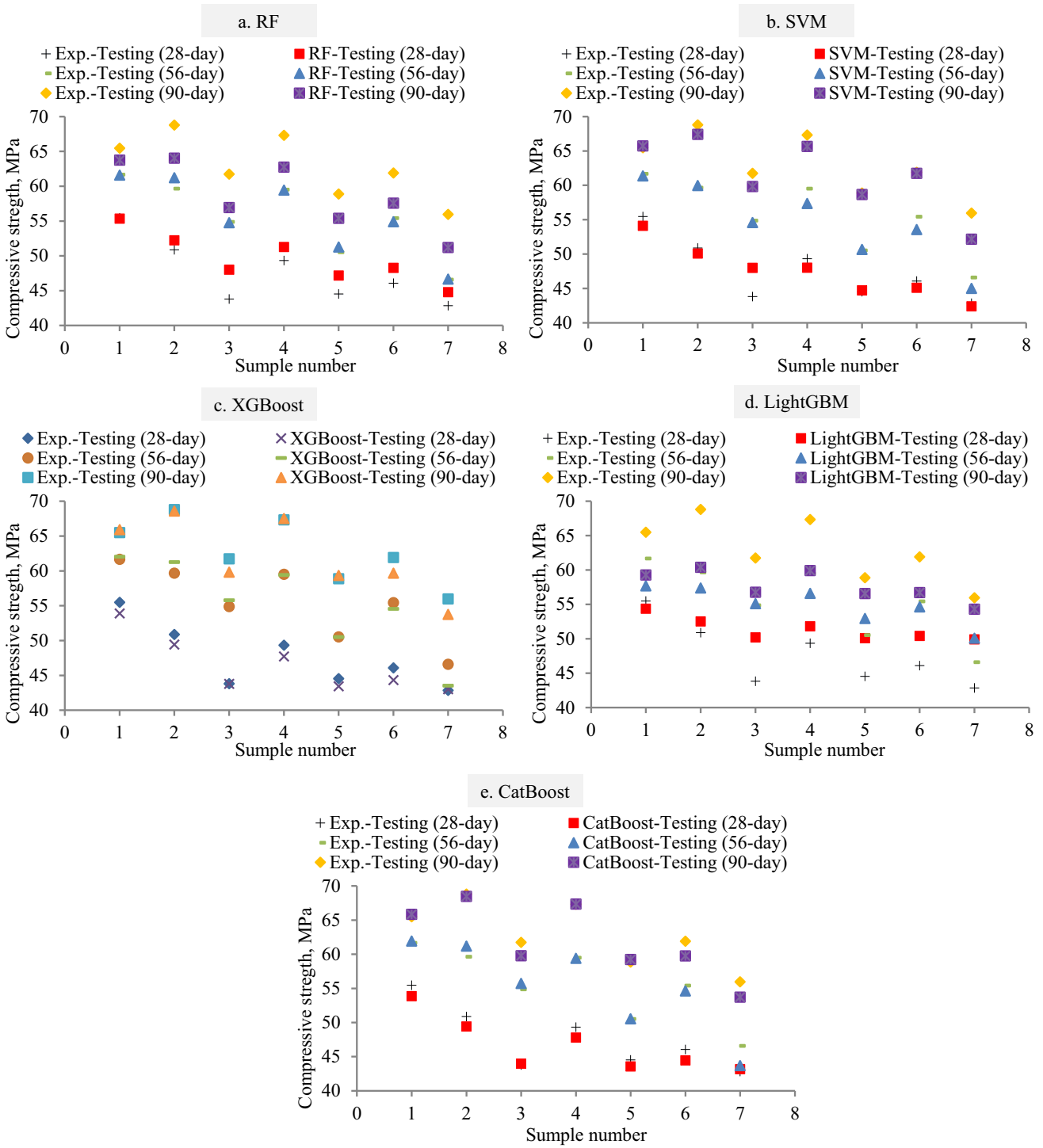


Fig. 7 Comparative evaluation of experimental compressive strength results against predictions of machine learning models

Table 6 Comparison of experimental compressive strength results with predictions of the machine learning models

Concrete group	Data used in the model construction							Compressive strength, MPa					
	Zeolite, kg/m^3	Diatomite, kg/m^3	PC, kg/m^3	Water, kg/m^3	Hyper plasticizer, kg/m^3	Aggregate, kg/m^3	Curing time, day	Exp	RF	SVM	XGBoost	LightGBM	CatBoost
R	0	0	400	139.7	4.800	1836	28	55.5	55.3	54.1	53.9	54.4	53.9
10D	0	40	360	139.7	4.320	1857	28	50.9	52.2	50.1	49.4	52.5	49.4
20D	0	80	320	123.3	4.800	1836	28	43.8	48.0	48.0	43.8	50.2	44.0
10Z	40	0	360	139.7	4.320	1884	28	49.3	51.3	48.0	47.7	51.8	47.8
20Z	80	0	320	123.2	4.800	1912	28	44.5	47.2	44.7	43.4	50.1	43.6
5D5Z	20	20	360	139.7	4.320	1897	28	46.1	48.3	45.1	44.3	50.4	44.5
10D10Z	40	40	320	124.2	3.840	1911	28	42.8	44.8	42.4	43.0	49.9	43.2
R	0	0	400	139.7	4.800	1836	56	61.7	61.6	61.4	62.0	57.7	61.9
10D	0	40	360	139.7	4.320	1857	56	59.7	61.2	60.0	61.3	57.4	61.2
20D	0	80	320	123.3	4.800	1836	56	54.9	54.7	54.6	55.8	55.1	55.7
10Z	40	0	360	139.7	4.320	1884	56	59.5	59.4	57.4	59.4	56.6	59.4
20Z	80	0	320	123.2	4.800	1912	56	50.5	51.3	50.7	50.5	52.9	50.6
5D5Z	20	20	360	139.7	4.320	1897	56	55.4	54.9	53.5	54.5	54.6	54.6
10D10Z	40	40	320	124.2	3.840	1911	56	46.6	46.6	45.0	43.5	50.1	43.7
R	0	0	400	139.7	4.800	1836	90	65.5	63.8	65.7	65.9	59.3	65.9
10D	0	40	360	139.7	4.320	1857	90	68.8	64.1	67.4	68.6	60.4	68.5
20D	0	80	320	123.3	4.800	1836	90	61.7	56.9	59.8	59.8	56.8	59.8
10Z	40	0	360	139.7	4.320	1884	90	67.3	62.8	65.7	67.5	59.9	67.4
20Z	80	0	320	123.2	4.800	1912	90	58.9	55.4	58.7	59.3	56.6	59.2
5D5Z	20	20	360	139.7	4.320	1897	90	61.9	57.6	61.8	59.7	56.7	59.8
10D10Z	40	40	320	124.2	3.840	1911	90	56.0	51.2	52.2	53.7	54.3	53.7

presented a good learning function that maps input variables to response variables accurately. Henceforth, the parameters of the best-performing model (Catboost model) are introduced in Table 7.

As a consequence of the results, it is necessary to analyze outperforming CatBoost more clearly. For this purpose, tenfold cross-validation procedure is implemented with the above model parameters.

The importance of each predictor feature is introduced in Fig. 5. As can be seen from Fig. 5, "Curing Time" and "Diatomite" are the most important predictor features since they are the main components of concrete. While hyperplasticizer and aggrega are the least important predictors.

In Table 8, the performances of the machine learning models are compared by RMSE, MAPE and R^2 statistical parameters.

The RMSE, MAPE and R^2 evaluations of the RF, SVM, XGBoost, LightGBM, and CatBoost models, respectively, have been computed as follows: (2.43–2.96%-91.50%), (1.42–1.69%-96.39%), (1.19–1.49%-98.01%), (3.89–5.31%-97.99%), and (1.15–1.45%-98.03%).

The yield of RMSE evaluations is very close to 0 for each machine learning model. On the other hand, MAPE results are still below 10% in all the models. These values indicate that the models present high accuracy, and therefore, their yield is efficient. Particularly, when MAPE results of the CatBoost model is taken into account, it can be seen that its compressive strength estimations yield an accuracy of 98.55%. Furthermore, it is seen that the R^2 value is very close to one in all models. Evaluation data imply that the machine learning model estimates close values. When all statistical indices are examined together, it is observed that the predictions of the models at the 28, 56, and 90 days are very close to the real values.

To investigate each test result, and analyze the errors, the experimental observations on the curing time (day) and the compressive strength predictions by the SVM, RF, XGBoost, LightGBM, and CatBoost models were compared. Figure 9 and Table 9 present the compressive strength data at 28 day of hydration and the error rates with the percentage difference between the results, respectively.

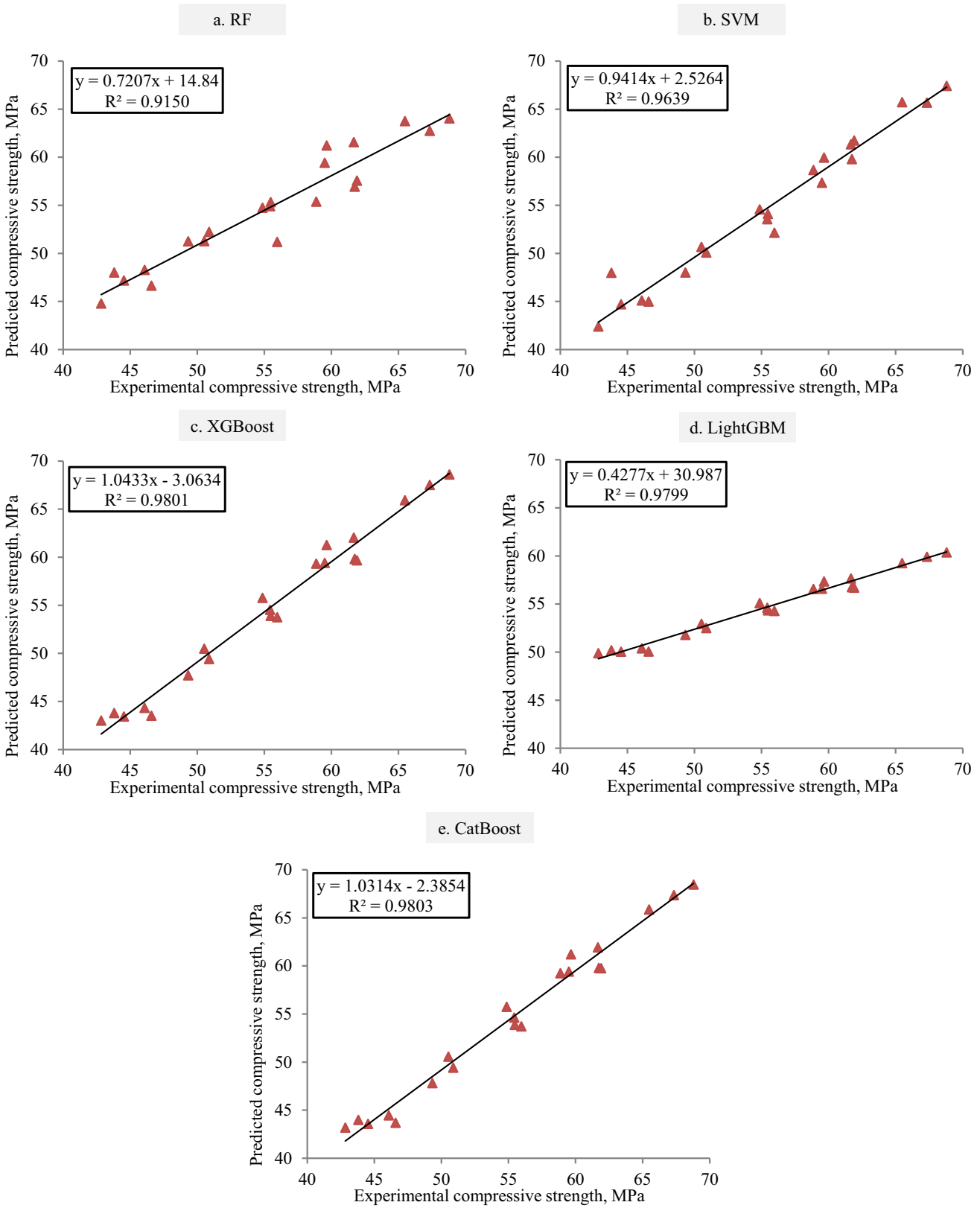


Fig. 8 Comparative evaluation of experimental compressive strength results against predictions of machine learning models fit line

Table 7 Designed Catboost models' parameters

Catboost parameter	Value
Iterations	100
Learning rate	0.1
Depth	6
Loss function	RMSE
Eval metric	RMSE
Od_type	Iter
Metric period	50

Table 8 Statistical analysis of compressive strength predictions of machine learning models

Statistical parameters	RMSE	MAPE, %	R^2, %
RF	2.43	2.96	91.50
SVM	1.42	1.69	96.39
XGBoost	1.19	1.49	98.01
LightGBM	3.89	5.31	97.99
CatBoost	1.15	1.45	98.03

According to the experimental compressive strength results of the concrete at 56-day, the smallest value of 42.83 MPa in the 20D coded concrete, while the highest value with 55.47 MPa in the concrete by R code was introduced in (Fig. 10). The Figure denotes that least gap was in the XGBoost model with 0.05% percent difference and 0.02 MPa error in 20Z coded concrete, and the most gap was in the LightGBM model with 14.15% percent difference and − 7.06 MPa error in 10Z10D coded concrete (Table 9).

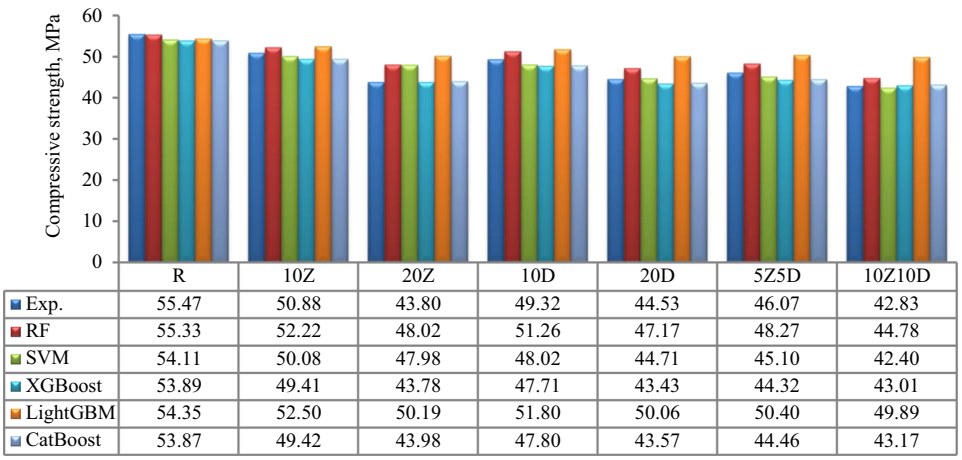
Figure 10 and Table 10 indicate the compressive strength data at 56-day of hydration and the error rates with the percentage difference between the results, respectively.

According to the experimental compressive strength results of the concrete at the 56 day, the smallest value with 46.58 MPa is presented by the 10Z10D coded concrete, where the highest value of 61.66 MPa is introduced by

Table 9 Prediction performance of machine learning models for compressive strength at 28 day

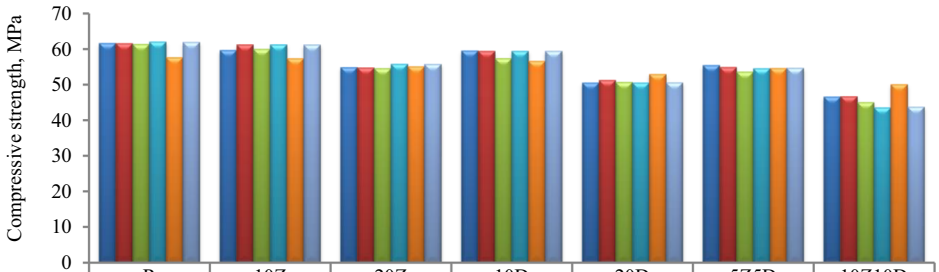
Models		*R*	10Z	20Z	10D	20D	5Z5D	10Z10D
RF	Error, MPa	0.13	− 1.35	− 4.21	− 1.94	− 2.65	− 2.20	− 1.95
	Percentage difference, %	0.24	− 2.58	− 8.78	− 3.78	− 5.61	− 4.56	− 4.36
SVM	Error, MPa	1.36	0.79	− 4.18	1.30	− 0.18	0.97	0.44
	Percentage difference, %	2.51	1.58	− 8.70	2.71	− 0.39	2.14	1.03
XGBoost	Error, MPa	1.57	1.47	0.02	1.61	1.10	1.75	− 0.18
	Percentage difference, %	2.92	2.97	**0.05**	3.37	2.53	3.96	− 0.41
LightGBM	Error, MPa	1.12	− 1.62	− 6.39	− 2.48	− 5.53	− 4.33	− 7.06
	Percentage difference, %	2.05	− 3.09	− 12.73	− 4.78	− 11.04	− 8.59	**− 14.15**
CatBoost	Error, MPa	1.59	1.46	− 0.18	1.52	0.96	1.61	− 0.34
	Percentage difference, %	2.96	2.95	− 0.40	3.17	2.20	3.63	− 0.80

The largest and smallest error values are deliberately shown in bold



	R	10Z	20Z	10D	20D	5Z5D	10Z10D
Exp.	55.47	50.88	43.80	49.32	44.53	46.07	42.83
RF	55.33	52.22	48.02	51.26	47.17	48.27	44.78
SVM	54.11	50.08	47.98	48.02	44.71	45.10	42.40
XGBoost	53.89	49.41	43.78	47.71	43.43	44.32	43.01
LightGBM	54.35	52.50	50.19	51.80	50.06	50.40	49.89
CatBoost	53.87	49.42	43.98	47.80	43.57	44.46	43.17

Fig. 9 Values of actual compressive strength and predicted by the machine learning models at the 28 day



	R	10Z	20Z	10D	20D	5Z5D	10Z10D
Exp.	61.66	59.65	54.86	59.50	50.52	55.42	46.58
RF	61.59	61.23	54.75	59.44	51.27	54.90	46.65
SVM	61.36	59.97	54.57	57.35	50.67	53.55	44.99
XGBoost	62.03	61.26	55.77	59.40	50.48	54.53	43.50
LightGBM	57.67	57.36	55.10	56.59	52.94	54.61	50.06
CatBoost	61.93	61.20	55.74	59.41	50.56	54.62	43.69

Fig. 10 Values of actual compressive strength and predicted by the machine learning models at the 56 day

Table 10 Prediction performance of machine learning models for compressive strength at 56 day

Models		*R*	10Z	20Z	10D	20D	5Z5D	10Z10D
RF	Error, MPa	0.07	− 1.57	0.11	0.06	− 0.75	0.52	− 0.07
	Percentage difference, %	0.12	− 2.57	0.21	0.10	− 1.47	0.95	− 0.15
SVM	Error, MPa	0.30	− 0.32	0.29	2.15	− 0.15	1.87	1.59
	Percentage difference, %	0.49	− 0.53	0.54	3.74	− 0.31	3.50	3.52
XGBoost	Error, MPa	− 0.37	− 1.61	− 0.91	0.10	0.03	0.89	3.07
	Percentage difference, %	− 0.59	− 2.63	− 1.63	0.17	**0.06**	1.64	**7.07**
LightGBM	Error, MPa	4.00	2.29	− 0.24	2.91	− 2.42	0.81	− 3.48
	Percentage difference, %	6.93	4.00	− 0.44	5.14	− 4.58	1.49	− 6.96
CatBoost	Error, MPa	− 0.27	− 1.54	− 0.88	0.09	− 0.04	0.80	2.89
	Percentage difference, %	− 0.43	− 2.52	− 1.57	0.16	− 0.09	1.47	6.61

The largest and smallest error values are deliberately shown in bold

the R coded concrete (Fig. 10). On the other hand, the least change was observed by the XGBoost model with 0.06% difference and 0.03 MPa error in 20D coded concrete, and highest change was observed in the XGBoost model with 7.07% difference, 3.07 MPa error in 10Z10D coded concrete (Table 10).

Figure 11 and Table 11 show the compressive strength data at 56-day of hydration and the error rates with the percentage difference between the results, respectively.

In terms of experimental compressive strength results of the concrete at the 90 day, the smallest value was observed during 55.96 MPa in the 10Z10D coded concrete, where the highest value of 68.8 MPa was encountered by the 10Z coded concrete (Fig. 10). Moreover, the least change was monitored with the LightGBM model with − 0.04% percent difference, and − 0.03 MPa error in 10D coded concrete. Finally, maximal change was monitored by LightGBM model with 13.94% percent difference and 8.42 MPa error in 10Z coded concrete (Table 11).

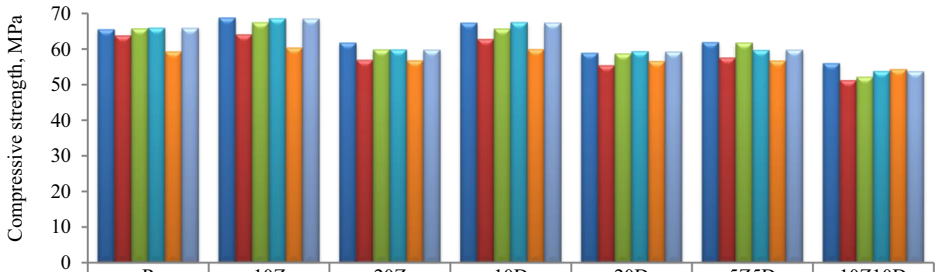
In summary, machine learning models in general and Gradient Boosting models in particular, introduced efficient prediction performance. Especially, CatBoost introduced the most efficient prediction performance and minimum error rates.

The gradient boosting models utilize multiple decision trees as with weak learners. On the other hand, each gradient boosting model differs through decision tree construction methodologies and splitting decision tree construction strategies. In this way, each algorithm aims to ensure low variance and high accuracy.

In terms of this study, the leaf-wise split architecture of CatBoost yields efficient prediction performance for compressive strength of cement data, where complex trees improve prediction performance, where the machine learning model does not cause overfitting.

6 Conclusion and recommendation

In this study, compressive strength results at 28 day, 56 day and 90 day consisting of seven different concrete mixtures with or without zeolite and diatomite in sodium chloride



	R	10Z	20Z	10D	20D	5Z5D	10Z10D
Exp.	65.48	68.80	61.73	67.33	58.87	61.90	55.96
RF	63.76	64.06	56.93	62.76	55.38	57.57	51.19
SVM	65.74	67.43	59.82	65.68	58.67	61.75	52.15
XGBoost	65.92	68.59	59.81	67.51	59.33	59.67	53.75
LightGBM	59.26	60.38	56.76	59.92	56.57	56.70	54.28
CatBoost	65.85	68.46	59.79	67.36	59.24	59.75	53.71

Fig. 11 Values of actual compressive strength and predicted by the machine learning models at the 90 day

Table 11 Prediction performance of machine learning models for compressive strength at 90 day

Models		*R*	10Z	20Z	10D	20D	5Z5D	10Z10D
RF	Error. MPa	1.72	4.74	4.80	4.56	3.48	4.33	4.76
	Percentage difference. %	2.70	7.40	8.43	7.27	6.29	7.52	9.31
SVM	Error. MPa	− 0.26	1.37	1.91	1.65	0.20	0.15	3.80
	Percentage difference. %	− 0.39	2.03	3.19	2.51	0.34	0.24	7.29
XGBoost	Error. MPa	− 0.44	0.21	1.93	− 0.18	− 0.46	2.23	2.21
	Percentage difference. %	− 0.66	0.30	3.22	− 0.27	− 0.78	3.74	4.10
LightGBM	Error. MPa	6.22	8.42	4.97	7.41	2.30	5.20	1.68
	Percentage difference. %	10.50	**13.94**	8.76	12.37	4.06	9.17	3.09
CatBoost	Error. MPa	− 0.37	0.34	1.95	− 0.03	− 0.37	2.15	2.24
	Percentage difference. %	− 0.56	0.49	3.26	**− 0.04**	− 0.63	3.60	4.18

The largest and smallest error values are deliberately shown in bold

(NaCl) solution have been examined developing five different machine learning algorithms, which are SVM, RF, XGBoost, LightGBM, and CatBoost. In developing these five models, 63 experimental data, each consisting of eight features, were used. Among these data, 42 experimental data were randomly selected for training and the remaining 21 data were reserved for testing. According to the findings obtained:

- Our particle size analyses denote that, the particles of diatomite are fine, PC particles are medium, and zeolite particles are coarse.
- In terms of specific gravities, zeolite 2.18, diatomite 2.28, and PC weighs 3.17. Hence, zeolite has the lowest specific gravity.
- As a result of the XRD analyses, PC and zeolite have a crystal structure, and diatomite has an amorphous structure.
- Chemical analyses indicate that zeolite and diatomite are positive in terms of pozzolanic features ($SiO_2 + Al_2O_3 + Fe_2O_3 > 70\%$), where such observation complies with the stated requirements standards for a natural pozzolan.
- The compressive strengths of the concrete samples depend on material type, substitution rate, and concrete age. The speed of strength gain of the diatomite and zeolite substituted concrete samples is slower than the one of reference concrete. However, on the 90th day, the compressive strength equivalent to the reference concrete was obtained at a 10% substituted rate.
- When experiments and RF model is compared, RMS, MAPE and R^2 were found as 2.43, 2.96%, and 91.5%, respectively.
- When experiments and SVM model is compared, RMS, MAPE and R^2 were found as 1.42, 1.69%, and 96.39%, respectively.
- In terms of evaluating our gradient boosting models, results are found as follows: RMS, MAPE and R^2 were found as 1.19, 1.49%, and 98.01% for XGBoost, 3.89, 5.31%, and 97.99% for LightGBM, 1.15, 1.45%, and 98.03% for CatBoost, respectively.

The results conclude that the compressive strength of concrete substituting zeolite, diatomite, and zeolite + diatomite can be predicted with recent machine learning algorithms in a quite short period. In addition, highly accurate prediction results denote that the gradient boosting-based machine learning models are practicable and cheap solutions for concrete mixes. Furthermore, these systems can reduce losses in both elapsed time and financial costs during the preparation of the cement mortars and concretes by exploiting various additives.

Artificial Intelligence in general and Machine learning in particular are still progressing. More accurate chemical behavior predictions can be procured when different types of concrete mixtures with various supplementary cementitious materials are utilized.

Acknowledgements This study is financially supported by Düzce University Presidency of Scientific Research Projects with the project code number 2011.03.HD.009. Furthermore, Düzce Yiğitler Beton supported the research by providing facilities for experiments to be carried out. The authors would like to thank both supporters.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Declarations

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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Khan, M.I.: Predicting properties of high performance concrete containing composite cementitious materials using artificial neural networks. Autom. Constr. **22**, 516–524 (2012). https://doi.org/10.1016/j.autcon.2011.11.011
2. Yang, H.; Liu, J.; Jia, X.; Zhou, Y.; Ji, H.: Influence of NaCl concentrations on the crack-sealing behavior of superabsorbent polymers in cementitious materials. Constr. Build. Mater. **243**, 118228 (2020). https://doi.org/10.1016/j.conbuildmat.2020.118228
3. Gerengi, H.; Kocak, Y.; Jazdzewska, A.; Kurtay, M.: Corrosion behavior of concrete produced with diatomite and zeolite exposed to chlorides. Comput. Concr. **19**(2), 161–169 (2017). https://doi.org/10.12989/cac.2017.19.2.161
4. Park, J.C.; Jung, H.J.: An experimental study on the steel corrosion behavior in NaCl-immersed concrete under hydraulic pressure. Constr. Build. Mater. **345**, 128347 (2022). https://doi.org/10.1016/j.conbuildmat.2022.128347
5. Asrar, N.; Malik, A.U.; Ahmad, S.; Mujahid, F.S.: Corrosion protection performance of microsilica added concretes in NaCl and seawater environments. Constr. Build. Mater. **13**(4), 213–219 (1999). https://doi.org/10.1016/S0950-0618(99)00016-1
6. Pınarcı, İ; Kocak, Y.: Hydration mechanisms and mechanical properties of pumice substituted cementitious binder. Constr. Build. Mater. **335**, 127528 (2022). https://doi.org/10.1016/j.conbuildmat.2022.127528
7. Kurtay, M.; Gerengi, H.; Kocak, Y.; Chidiebere, M.A.; Yildiz, M.: The potency of zeolite and diatomite on the corrosive destruction of reinforcing steel in 1 M HNO_3 environment. Constr. Build. Mater. **236**, 117572 (2020). https://doi.org/10.1016/j.conbuildmat.2019.117572
8. Coelho, T.P.P.; Bezerra, B.P.; Verza, J.R.; Luz, A.P.; Morelli, M.R.: Physico-mechanical properties of metakaolin and diatomite-based geopolymers. Mater. Lett. **349**, 134784 (2023). https://doi.org/10.1016/j.matlet.2023.134784
9. Yang, Y.; Shen, Z.; Wu, W.; Zhang, H.; Ren, Y.; Yang, Q.: Preparation of a novel diatomite-based PCM gypsum board for temperature-humidity control of buildings. Build. Environ. **226**, 109732 (2022). https://doi.org/10.1016/j.buildenv.2022.109732
10. Teja, G.S.; Ravella, D.P.: Studies on self-curing self-compacting concretes containing zeolite admixture. Mater. Today: Proc. **43**, 2355–2360 (2021). https://doi.org/10.1016/j.matpr.2021.01.682
11. Lv, Z.; Jiang, A.; Jin, J.: Influence of ultrafine diatomite on cracking behavior of concrete: an acoustic emission analysis. Constr. Build. Mater. **308**, 124993 (2021). https://doi.org/10.1016/j.conbuildmat.2021.124993
12. Ghorbani, S.; Ghorbani, S.; Elmi, A.; Soleimani, V.; Taji, I.; Mohammadi-Khatami, M.; Tavakkolizadeh, M.; Brito, J.: Simultaneous effect of granite waste dust as partial replacement of cement and magnetized water on the properties of concrete exposed to NaCl and H_2SO_4 solutions. Constr. Build. Mater. **288**, 123064 (2021). https://doi.org/10.1016/j.conbuildmat.2021.123064
13. Aliabdo, A.A.; Abd Elmoaty, M.; Emam, M.A.: Factors affecting the mechanical properties of alkali activated ground granulated blast furnace slag concrete. Constr. Build. Mater. **197**, 339–355 (2019). https://doi.org/10.1016/j.conbuildmat.201811.086
14. Baltazar-Zamora, M.A.; Bastidas, D.M.; Santiago-Hurtado, G.; Mendoza-Rangel, J.M.; Gaona-Tiburcio, C.; Bastidas, J.M.; Almeraya-Calderón, F.: Effect of silica fume and fly ash admixtures on the corrosion behavior of AISI 304 embedded in concrete exposed in 3.5% NaCl solution. Materials **12**(23), 4007 (2019). https://doi.org/10.3390/ma12234007
15. Aruntaş, H.Y.; Bulgu, M.; Baydaş, M.T.; Albayrak, M.: Estimation of physical and mechanical properties of portland composite cement produced by using raw perlite with fuzzy logic model. Arab. J. Sci. Eng. (2024). https://doi.org/10.1007/s13369-023-08577-z
16. Kocak, B.; Pınarcı, İ; Güvenç, U.; Kocak, Y.: Prediction of compressive strengths of pumice-and diatomite-containing cement mortars with artificial intelligence-based applications. Constr. Build. Mater. **385**, 131516 (2023). https://doi.org/10.1016/j.conbuildmat.2023.131516
17. Kazemi, R.; Shadnia, R.; Eskandari-Naddaf, H.; Zhang, L.: The properties of cement-mortar at different cement strength classes: experimental study and multi-objective modeling. Arab. J. Sci. Eng. **47**(10), 13381–13396 (2022). https://doi.org/10.1007/s13369-022-06820-7
18. Fernández, J.; Chiachío, J.; Chiachío, M.; Barros, J.; Corbetta, M.: Physics-guided bayesian neural networks by abc-ss: application to reinforced concrete columns. Eng. Appl. Artif. Intell. **119**, 105790 (2023). https://doi.org/10.1016/j.engappai.2022.105790

19. Nasir, M.; Gazder, U.; Khan, M.U.; Rasul, M.; Maslehuddin, M.; Al-Amoudi, O.S.B.: Prediction of strength of plain and blended cement concretes cured under hot weather using quadratic regression and ANN tools. Arab. J. Sci. Eng. **47**(10), 12697–12709 (2022). https://doi.org/10.1007/s13369-022-06586-y
20. Chen, B.; Wang, L.; Liu, Y.; Wu, X.; Qin, Y.; Xia, L.: Optimization of high-performance concrete mix ratio design using machine learning. Eng. Appl. Artif. Intell. **122**, 106047 (2023). https://doi.org/10.1016/j.engappai.2023.106047
21. Alpaydin, E.: Introduction to Machine Learning. MIT press, Cambridge (2020)
22. Hastie, T.; Tibshirani, R.; Friedman, J.: The elements of statistical learning: data mining, inference, and prediction. Springer, New York (2009)
23. LeCun, Y.; Bengio, Y.; Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015). https://doi.org/10.1038/nature14539
24. Boser, B.E.; Guyon, I.M.; Vapnik, V.N.: A training algorithm for optimal margin classifier. Proceedings of the 5th ACM Workshop (Pennsylvania, USA). pp. 144–152 (1992). https://doi.org/10.1145/130385.130401
25. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
26. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Ann. Stat. **29**(5), 1189–1232 (2001)
27. Freund, Y.; Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. **55**(1), 119–139 (1997). https://doi.org/10.1006/jcss.1997.1504
28. Chen, T.; Guestrin, C.: Xgboost: A scalable tree boosting system, In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (San Francisco, CA, USA). pp. 785–794 (2016). https://doi.org/10.1145/2939672.2939785
29. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.;Ye, Q.; Liu, T.Y.: Lightgbm: A highly efficient gradient boosting decision tree, 31st Conference on Neural Information Processing Systems. Long Beach, CA, USA. (2017)
30. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A. V.; Gulin, A.: CatBoost: unbiased boosting with categorical features, Advances in neural information processing systems, 31 (2018)
31. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G.: A comparative analysis of gradient boosting algorithms. Artif. Intell. Rev. **54**(3), 1937–1967 (2021). https://doi.org/10.1007/s10462-020-09896-5
32. Lee, S.; Nguyen, N.H.; Karamanli, A.; Lee, J.; Vo, T.P.: Super learner machine-learning algorithms for compressive strength prediction of high performance concrete. Struct. Concr. **24**(2), 2208–2228 (2023). https://doi.org/10.1002/suco.202200424
33. Zhang, J.; Ma, G.; Huang, Y.; Aslani, F.; Nener, B.: Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression. Constr. Build. Mater. **210**, 713–719 (2019). https://doi.org/10.1016/j.conbuildmat.2019.03.189
34. Abunassar, N.; Alas, M.; Ali, S.I.A.: Prediction of compressive strength in self-compacting concrete containing fly ash and silica fume using ANN and SVM. Arab. J. Sci. Eng. **48**(4), 5171–5184 (2023). https://doi.org/10.1007/s13369-022-07359-3
35. Feng, D.C.; Liu, Z.T.; Wang, X.D.; Chen, Y.; Chang, J.Q.; Wei, D.F.; Jiang, Z.M.: Machine learning-based compressive strength prediction for concrete: an adaptive boosting approach. Constr. Build. Mater. **230**, 117000 (2020). https://doi.org/10.1016/j.conbuildmat.2019.117000
36. Nguyen, H.; Hoang, N.D.: Computer vision-based classification of concrete spall severity using metaheuristic-optimized extreme gradient boosting machine and deep convolutional neural network. Autom. Constr. **140**, 104371 (2022). https://doi.org/10.1016/j.autcon.2022.104371
37. TS EN 197-1: Cement- Part 1: compositions and conformity criteria for common cements. Turkish Standards, Ankara-Turkey. (2012)
38. TS 25: Natural pozzolan (Trass) for use in cement and concrete - Definitions, requirements and conformity criteria. Turkish Standards, Ankara-Turkey. (2015)
39. ASTM C618: Standard specification for coal fly ash and raw or calcined natural pozzolan for use in concrete. American society for testing and materials. West Conshohocken, PA, USA. (2003)
40. Varela, M.B.; Ramírez, S.M.; Ereña, I.; Gener, M.; Carmona, P.: Characterization and pozzolanicity of zeolitic rocks from two Cuban deposits. Appl. Clay Sci. **33**(2), 149–159 (2006). https://doi.org/10.1016/j.clay.2006.04.006
41. Jia, Y.; Han, W.; Xiong, G.; Yang, W.: A method for diatomite zeolitization through steam-assisted crystallization with in-situ seeding. Mater. Lett. **62**(16), 2400–2403 (2008). https://doi.org/10.1016/j.matlet.2007.12.005
42. TS 802: Design Concrete Mixes. Turkish Standards, Ankara-Turkey. (2009)
43. TS EN 12350-2: Testing fresh concrete - Part 2: Slump test. Turkish Standards, Ankara-Turkey. (2010)
44. TS EN 206-1: Concrete- Part 1: Specification, performance, production and conformity. Turkish Standards, Ankara-Turkey. (2014)
45. TS EN 12390-3: Testing hardened concrete - Part 3: Compressive strength of test specimens. Turkish Standards, Ankara-Turkey. (2019)
46. Sakthivel, P.B.; Ravichandran, A.; Alagumurthi, N.: Modeling and prediction of flexural strength of hybrid mesh and fiber reinforced cement-based composites using artificial neural network (ANN). GEOMATE J. **10**(19), 1623–1635 (2016)
47. Özcan, F.; Atiş, C.D.; Karahan, O.; Uncuoğlu, E.; Tanyildizi, H.: Comparison of artificial neural network and fuzzy logic models for prediction of long-term compressive strength of silica fume concrete. Adv. Eng. Softw. **40**(9), 856–863 (2009). https://doi.org/10.1016/j.advengsoft.2009.01.005