

RESEARCH

Open Access



Comparative Analysis of Gradient-Boosting Ensembles for Estimation of Compressive Strength of Quaternary Blend Concrete

Ismail B. Mustapha¹, Muyideen Abdulkareem^{2*} , Taha M. Jassam², Ali H. AlAteah³, Khaled A. Alawi Al-Sodani³, Mohammed M. H. Al-Tholaia³, Hatem Nabus¹, Sophia C. Alih⁴, Zainab Abdulkareem^{5,6} and Abideen Ganiyu⁷

Abstract

Concrete compressive strength is usually determined 28 days after casting via crushing of samples. However, the design strength may not be achieved after this time-consuming and tedious process. While the use of machine learning (ML) and other computational intelligence methods have become increasingly common in recent years, findings from pertinent literatures show that the gradient-boosting ensemble models mostly outperform comparative methods while also allowing interpretable model. Contrary to comparison with other model types that has dominated existing studies, this study centres on a comprehensive comparative analysis of the performance of four widely used gradient-boosting ensemble implementations [namely, gradient-boosting regressor, light gradient-boosting model (LightGBM), extreme gradient boosting (XGBoost), and CatBoost] for estimation of the compressive strength of quaternary blend concrete. Given components of cement, Blast Furnace Slag (GGBS), Fly Ash, water, superplasticizer, coarse aggregate, and fine aggregate in addition to the age of each concrete mixture as input features, the performance of each model based on R^2 , RMSE, MAPE and MAE across varying training–test ratios generally show a decreasing trend in model performance as test partition increases. Overall, the test results showed that CatBoost outperformed the other models with R^2 , RMSE, MAE and MAPE values of 0.9838, 2.0709, 1.5966 and 0.0629, respectively, with further statistical analysis showing the significance of these results. Although the age of each concrete mixture was found to be the most important input feature for all four boosting models, sensitivity analysis of each model shows that the compressive strength of the mixtures does increase significantly after 100 days. Finally, a comparison of the performance with results from different ML-based methods in pertinent literature further shows the superiority of CatBoost over reported the methods.

Keywords Compressive strength, Quaternary blend concrete, Gradient-boosting algorithms, Ensemble machine learning

Journal information: ISSN 1976-0485 / eISSN 2234-1315.

*Correspondence:

Muyideen Abdulkareem

muyikareem@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **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/>.

1 Introduction

Climate change and global warming have accelerated due to increasing emissions of greenhouse gases (GHG). This has led to serious environmental problems, such as drought, flood, heat waves etc. (Pandey & Kumar, 2022). The production of concrete used in the construction industry remains one of the largest sources of GHG, and accounts for about 50% of global emissions (Allujami et al., 2022a, 2022b; Di Filippo et al., 2019). GHG from concrete production is expected to increase as demand for concrete keeps surging due to human development. The production of Portland cement (PC) produces vast amount of CO₂ through a process called calcination of Calcium oxide (CaO). This calcination accounts for around 7% of the global CO₂ emissions to the atmosphere (Benhelal et al., 2019). This emission is expected to increase as the annual consumption of cement would rise from its present 4000 million tonnes to about 6000 million tonnes by the year 2060 (Moreira & Arrieta, 2019). These figures show the need for sustainable and more environmental-friendly materials to replace cement partially or fully, not only to meet the growing demand, but to reduce emissions of CO₂ (Ebid et al., 2022; Mikulčić et al., 2016).

In view of the abovementioned problems, industrial wastes have been used in production of concrete. This approach results in a drastic decrease in PC used in construction as well as prevents environmental degradation caused by disposal of these hazardous industrial waste (Agrawal et al., 2021; Hashim & Tantray, 2021). The use of industrial wastes can reduce about 80% of GHG emissions of normal concrete. The commonly used industrial wastes that act as supplementary cementitious material in concrete include fly ash (FA), ground granulated blast furnace slag (GGBS) and silica fume (SF) (Hammad et al., 2021; Hashmi et al., 2021; Okashah et al., 2020). They have been used as partial replacements for cement when producing improved and more sustainable concrete. This practice is favoured by the availability of large quantity of these industrial wastes as about 300 million tonnes of FA is produced annually with only 25% of this production being used up for concrete production (Dan et al., 2021). Similarly, annual global production of GGBS is around 280 million tonnes with less than 10% of this production being utilised in concrete production (Kamath et al., 2021).

In the production of concrete for structural usage, an in depth and accurate knowledge of the properties are required (Ebid & Deifalla, 2022; Salem & Deifalla, 2022; Song et al., 2021). Compressive strength, being the most important property can be improved by partial replacement of cement with these cementitious industrial wastes in the accurate proportions. The compressive strength

is generally ascertained by testing (crushing) concrete specimens (cubes or cylinders), usually after 28 days of casting (Allujami et al., 2022a, 2022b; Ebid & Deifalla, 2021). However, this method of obtaining the compressive strength of concrete is time consuming, tedious and expensive (Badra et al., 2022; Silva et al., 2020). In addition, the desired strengths are often not attained, thus being less effective (Deifalla & Salem, 2022; Salami et al., 2022). This has led researchers to the use of machine learning (ML) and artificial intelligence (AI) algorithms to obtain the mechanical properties of concrete. The use of AI and ML techniques, such as decision tree (DT), artificial neural network (ANN), support vector machine (SVM), and extreme learning machine (ELM), in estimating (predicting) concrete properties takes into account certain parameters of the concrete (such as concrete mix proportions and concrete age) and its constituents to achieve reliable estimations (Gupta et al., 2006; Mustapha et al., 2022).

Several ML approaches have been proposed over the years for accurate estimation of compressive strength of concrete. For example, Cook et al. (2019) presented a hybrid ML model that combined firefly algorithm (FFA) with random forests (RF) to predict the compressive strength of concrete. A correlation between the input variables and output was developed by training the hybrid (RF-FFA) model with two different categories of data sets. They concluded that the hybrid RF-FFA model performed better than standalone ML models, such as SVM, RF, M5Prime model-tree algorithm and multilayer perceptron-ANN (MLP-ANN). Shariati et al. (2020) presented a novel hybrid ML approach using grey wolf optimizer to predict the compressive strength of concrete with partial replacement of cement. The results were compared to those obtained via an adaptive neuro-fuzzy inference system (ANFIS), extreme learning machine (ELM), ANN, support vector regression (SVR) with radial basis function (RBF) kernel (SVR-RBF), and another SVR with a polynomial function kernel (SVR-Poly).

Dao et al., (2020a, 2020b) applied an optimized conventional ANN to predict the compressive strength of foamed concrete. Dry density was included as an input parameter, while the volume of foam was ignored in their study. The results showed a high correlation R^2 of 0.97 for the models. The authors referred to ANN as a black-box model, since it provides no practical information about the predicted model, and citing the vast hidden neurons as major impediments to developing an empirical relation between input and output parameters. Abellán-García (2020) presented an ANN model with four layers to predict the compressive strength of ultra-high-performance concrete (UHPC). A total of 927 data samples and 18 mixture design variables were used as input. While

impressive results were similarly reported, the proposed approach shares a common shortcoming with other aforementioned approaches in that the knowledge of the contribution of each input feature in the model predictions of the concrete mixtures is lacking. Besides, the results reported in most of these studies are still open to further improvement.

The quest for more accurate estimation of compressive strength of HPC has inspired the use of nature inspired classifiers, such as genetic expression programming (GEP). For instance, Ullah et al. (2022) applied a database of 191 data points to develop a relationship between the mix design parameters and compressive strength of foamed concrete using gene expression programming (GEP). The input variables were cement content, sand content, water to cement ratio, foam volume, while the output parameters were the dry density and compressive strength. The results showed that 95% of the predicted compressive strength had error values that were less than 2%. Recently, Shah et al. (2022) presented a comparative analysis using different ML techniques to predict the compressive strength of sugarcane bagasse ash (SCBA) concrete. The ML techniques included random forest regression (RFR), GEP and SVM. The results were compared to experimental testing. The input variables were water–cement ratio, cement content, SCBA dosage (SCBA%), the quantity of fine aggregate and coarse aggregate. The results showed that the R^2 of all the ML techniques were all above 0.85, and the RRMSE and performance index (PI) were less than 10% and 0.2%, respectively, with GEP producing the most accurate results across the compared methods. While GEP allow generation of simple mathematical equations for built models, it can be computationally expensive. Besides, its performance has long been shown to be similar or lower than other existing genetic programming methods (Oltean & Grosan, 2003). In fact, recent studies on compressive strength estimation such as (Fakharian et al., 2023; Salami et al., 2022; Song et al., 2021) have shown via empirical results that ML methods such as ANN and classifier ensembles outperform GEP across several evaluation metrics.

Boosting methods are a class of ensemble machine learning methods that have found wide application in many real-life domains with impressive results (Babajide Mustapha & Saeed, 2016). They generally enhance learning by merging the predictions of several simple base learners into a composite whole (Tanha et al., 2020). Different implementations of boosting ensemble have also been employed by several researchers for compressive strength estimation. For example, Kaloop et al. (2020) investigated the use of a multivariate adaptive regression splines (MARS) model to extract the optimum inputs

to use for compressive strength design of HPC. The extracted features were fed to a gradient-tree-boosting machine (GBM). While improved results over comparative methods were reported, the authors also found concrete age to be the most influential input parameter. Feng et al. (2020) applied an adaptive boosting algorithm (Ada-boost) to predict the compressive strength of concrete given curing time and mixture contents as input variables. Using tenfold cross validation method for model validation, the authors reported notable improvement in performance over classical methods, such as ANN and SVM. Nguyen-Sy et al. (2020) demonstrated an accurate prediction of the compressive strength of concrete using an extreme gradient-boosting (XGBoost) model. Sensitivity analysis was carried out to optimize the numbers of estimators by varying them from 100 to 1000 while keeping the default values of other hyperparameters constant. An increase in the number of estimators was found to generally lead to increased model accuracy.

In another related study, Cui et al. (2021) proposed a novel XGBoost prediction model based on grey relation analysis (GRA) for the estimation of compressive strength of concrete containing slag and metakaolin. Empirical findings showed that XGBoost outperformed ANN and its genetic algorithm hybridized variant (GA-ANN). Similar study by Nguyen et al. (2021) concluded that XGBoost and gradient-boosting regressor (GBR) models outperformed the likes of SVM and MLP for prediction of compressive strength and tensile strength of HPC.

Apart from XGBoost, there are other gradient-boosting implementations that have found application in concrete property estimation. For instance, Alabdullah et al. (2022) applied LightGBM to estimate the values of Rapid Chloride Penetration Test (RCPT) in a metakaolin-based high strength concrete. Using 201 experimental samples, input variables such as binder content, concrete age, water–binder ratio, metakaolin percentage, and content of fine and coarse aggregates were used to train a LightGBM model which yielded results with R^2 value of 0.9738. Likewise, Mahjoubi et al. (2022) investigated the use of LightGBM in the estimation of the compressive strength of UHPC with similarly high prediction accuracy. In another pertinent study, de-Prado-Gil et al. (2022) applied a CatBoost (CBT) model to predict the compressive strength of a self-compacting concrete. The study was conducted using 381 data samples. Experimental findings show that the cement content had the highest influence on model output.

There has also been a notable growth in the application of deep learning methods for compressive strength estimation in recent years. Jang et al. (Jang et al., 2019) proposed image-based compressive strength estimation

of concrete using three deep neural network (DNN) architectures, namely, ResNet, GoogLeNet, and AlexNet. Images of the surfaces of specially produced specimens were captured with a portable digital microscope and used to train each model for compressive strength estimation. Empirical results show that the DNN models outperformed the fully connected ANNs with ResNet showing the best performance. In addition, a deep learning-based estimation of compressive strength of fiber-reinforced concrete at elevated temperatures was proposed in (Chen et al., 2021). Using the concrete mix, heating profile, and fiber properties as model inputs, three variations of convolutional neural networks (CNN) models were shown to outperform several models that include SVR, ANN and Adaboost. In addition, deep learning models such as CNN have been hybridized with evolutionary algorithms, such as GA for improved performance (Ranjbar et al., 2022). More recently, Hoang (2023) proposed a deep learning-based estimation of the compressive strength of rice husk ash-blended concrete using an asymmetric loss function. Results from this study showed better performance than ANN and multi-variate adaptive regression splines.

The pursuit of accurate estimation of compressive strength of concrete has inspired myriad of research studies over the years, each seeking to achieve this goal via some machine learning methods. However, findings as indicated from the foregoing show that the gradient-boosting ensembles and DNN-based approaches stand out, mostly performing better than popular methods, such as SVR, classical ANN, GEP, KNN and their hybrid variants amongst others. The gradient ensembles methods are particularly the focus of this study, given their high accuracy and interpretability. Besides, a comprehensive comparative study on gradient-boosting algorithms for prediction of compressive strength of quaternary blend concrete remains lacking. Such study has the potentials of guiding field engineers on the choice of computational tools for accurate and reliable estimation of properties when designing concrete.

Thus, this study aims to compare the performance of four gradient-boosting algorithms in estimating the compressive strength of quaternary blend concrete. The algorithms are gradient-boosting regressor (GBR), light gradient-boosting model (LGBM), eXtreme gradient boosting (XGB), and CatBoost (CBT). In the training phase, hyperparameter optimization of each algorithm is first carried out using fivefold cross validation to ensure optimal model performance. Twenty optimal models were built, five for each gradient-boosting algorithm, using different training–test splits to obtain best performing model in terms of mean squared error. The input variable are the proportions of cement, ground

granulated blast furnace slag (GGBS), fly ash (FA), water, superplasticizer, coarse aggregate, fine aggregate, and concrete age. The performance of each of the final model is evaluated using four popularly used statistical measures, namely, root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and coefficient of determination (R^2). A sensitivity analysis is carried out to understand the importance and contribution of the input/predictor variables. Finally, a comparison of the obtained results with results in previous literatures (other methods).

The key contributions of this study are highlighted as follows:

- Prediction of the compressive strength of quaternary blend concrete using CBT.
- A comprehensive comparative analysis of gradient-boosting algorithms (GBR, CBT, XGB and LGBM) for the estimation of quaternary blend concrete.
- An intuitive insight into the importance and contribution of input features for the estimation of quaternary blend concrete.
- Comparison of performance of gradient-boosting algorithms with results from previous studies.

2 Computational Methods

The gradient-boosting ensembles considered in this research are gradient-boosting regressor (GBR), light gradient-boosting model (LGBM), eXtreme gradient boosting (XGB), and CatBoost (CBT). These models have been selected based on their performance in pertinent studies relating to estimation of mechanical properties of concrete. The advantage model interpretability offers makes it especially useful for field engineers, allowing them to understand the impact of input parameters without undergoing tedious and time-consuming laboratory experiments. Each of the selected methods are detailed in what follows.

2.1 Gradient-Boosting Regressor

Gradient-boosted decision trees (GBDT) have been widely used in machine learning. However, gradient-boosting regressor (GBR) (Friedman, 2002) is arguably the earliest well-known implementation of the idea of gradient descent boosting of decision trees that optimizes an arbitrary differentiable loss function via stagewise additive approach in model building. Every iteration of the model building process involves fitting a classification and regression tree (CART) on the negative gradient (i.e., the residual error between the estimated and the target output) of an arbitrary loss function (Friedman, 2002). Gradient boosting of decision trees has been shown to be robust to overfitting while producing highly competitive

results especially while modelling noisy data. In addition, it is also interpretable as it offers relative importance of input features used in model building. The two main hyperparameters for optimal Gradient boosting are the number of boosting stages and the shrinkage parameter, also known as the learning rate (Friedman, 2001).

In general, in GBR, the model is initialized with a constant value γ (A tree with just one leaf node) that minimizes the loss over all the samples as in the following equation:

$$F^0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, \gamma) \quad (1)$$

This is followed by several iterations of negative gradient computation of the loss function L and its subsequent usage to fit a decision tree and addition of a new model to the ensemble as in the following equation:

$$F(x) = F^{m-1}(x) + \nu f^m(x) \quad (2)$$

where ν is the shrinkage parameter used to control overfitting. Although, GBR is used for regression problem in the present study, it is also suitable for classification problems. Extensive details of the theoretical foundation of gradient-boosting regressor can be found in (Friedman, 2001, 2002).

2.2 XGBoost

Another gradient-boosting implementation that is considered in this study is the extreme gradient-boosting (XGBoost) algorithm. XGBoost is an optimized variant of gradient boosting that combines the predictions of several “weak” classification and regression tree (CART) learners to develop a “strong” learner using additive training strategies (Chen et al., 2015). XGBoost is especially known for preventing overfitting efficiently through a simplified objective function that combines the loss and regularization terms. The regularized optimization objective is as in the following equation:

$$Obj = \sum_m^n l(y_m, \hat{y}_m) + \sum_k^K \Omega(f_k) \quad (3)$$

where l is the loss function that measures the difference between the experimental, y_m , and the estimated \hat{y}_m output; Ω is the regularization term given as the following equation:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{i=1}^T w_i^2 \quad (4)$$

where T and w are the number of leaves and the score on each leaf, respectively; γ and λ are constants for controlling the degree of regularization. Although used for

regression problem in this study, XGBoost is suitable for all types of supervised learning problems. See Chen et al. (2015) for detailed background on this algorithm.

2.3 LightGBM

Another novel implementation of gradient-boosted decision tree (GBDT) that has been proposed to address the scalability and efficiency problem of its traditional counterpart is LightGBM (LGBM) (Ke et al., 2017). Unlike the traditional GBDT which entails the time-consuming process of scanning all data samples to estimate the information gain of all possible split points for each tree node, LGBM proposes two novel techniques called gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB). In the GOSS, only samples with large gradients are considered important and used in the estimation of information gain for split point selection. Thus, a significant proportion of data samples are excluded when estimating the information with little or no impact on the accuracy of estimated gain. On the other hand, the EFB technique carries out the NP-hard problem of bundling mutually exclusive features (i.e., they rarely take nonzero values simultaneously) to reduce the number of features with negligible impact on the split point determination accuracy. Although used for regression problem in this study, LGBM is suitable for all supervised learning problems. Further details on LGBM can be found in (Ke et al., 2017).

2.4 CatBoost

Similar to the aforesaid GBDT algorithms, CatBoost (CBT) is also a machine learning algorithm that leverages gradient boosting on decision trees. CBT is a unique GBDT implementation that is known for its categorical feature handling capability (Dorogush et al., 2018). The two main algorithmic advances introduced in CBT are the implementation of ordered boosting which is a permutation-driven alternative to the classic algorithm, and an innovative algorithm for processing categorical features. Both techniques were created to fight a prediction shift caused by a special kind of target leakage present in all currently existing implementations of gradient-boosting algorithms. Likewise, CBT has the advantage of using a new schema for leaf values calculation when selecting tree structures, which greatly alleviates the problem of overfitting. Although used for regression problem in this study, CBT is suitable for all supervised learning problems. Extensive details on CBT can be found in (Dorogush et al., 2018).

3 Methodology

3.1 Data Description

The quaternary concrete data applied in this study are experimental results obtained from (Lichman, 2013). The compressive strength which is the most important property of concrete should be accurately and reliably modelled for a quaternary concrete. Thus, the data has been carefully selected to cover the compressive strength for a wide range of days, ranging from 1 to 365 days. To the best of our knowledge, these data are the largest and most widely used data set for compressive strength estimation. Hence, its popularity makes the results of this experiment comparable to a wide range of previous studies. The variables used as input in the modelling are age (days), portions of cement (Kg/m^3), GGBS (Kg/m^3), FA (Kg/m^3), water (Kg/m^3), super plasticizer (Kg/m^3), fine aggregate (kg/m^3) and coarse aggregate (Kg/m^3). Fig. 1 presents a visual distribution of each feature. The numerical values of the basic statistics of the features of the 1030 data

samples are also presented in Table 1. The statistics of the data set show the mean, standard deviation, minimum value, lower quartile, middle quartile (median), upper quartile, and maximum value to indicate consistency and suitability for use in this study.

In addition, a correlation analysis of all the input variables to the output, the compressive strength, is also presented to understand how changes in each input variable bring about corresponding changes in output. Correlation Coefficient (CC) was used to assess the sensitivity of each component (feature) of the concrete mixture to the compressive strength (MPa) (Mustapha et al., 2022; Salami et al., 2021). From Fig. 2, it can be observed that the input variables (cement, GGBS, water, superplasticizer, coarse aggregate, fine aggregate and age) have varying degrees of correlation with the output. Four of the input variables (cement, GGBS, superplasticizer and age) are positively correlated with the output, whereas the remaining four (fly ash, water, coarse aggregate and

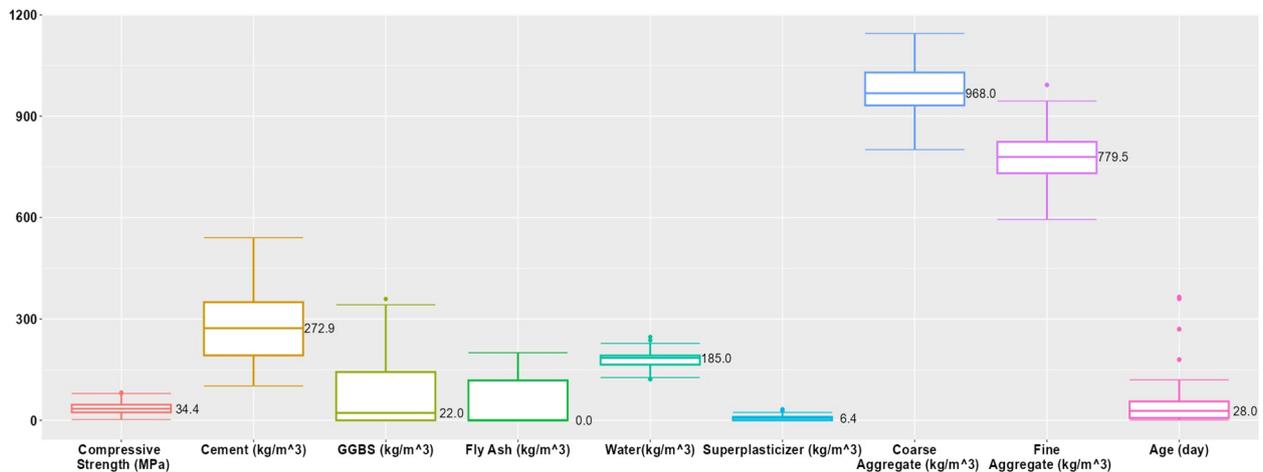


Fig. 1 Boxplots of distribution of compressive strength and input features of data sets

Table 1 Descriptive statistics of variables used in modelling

Stat	Compressive strength (MPa)	Cement (comp. 1) (kg/m^3)	GGBS (comp. 2) (kg/m^3)	Fly Ash (comp. 3) (kg/m^3)	Water (comp. 4) (kg/m^3)	SP (comp. 5) (kg/m^3)	Coarse Aggregate (comp. 6) (kg/m^3)	Fine Aggregate (comp. 7) (kg/m^3)	Age (day)
Count	1030	1030	1030	1030	1030	1030	1030	1030	1030
Mean	35.81796	281.1679	73.89583	54.18835	181.5673	6.20466	972.9189	773.5805	45.66
Std.	16.70574	104.5064	86.27934	63.997	21.35422	5.973841	77.75395	80.17598	63.17
Min.	2.33	102	0	0	121.8	0	801	594	1
25%	23.71	192.375	0	0	164.9	0	932	730.95	7
50%	34.445	272.9	22	0	185	6.4	968	779.5	28
75%	46.135	350	142.95	118.3	192	10.2	1029.4	824	56
Max.	82.6	540	359.4	200.1	247	32.2	1145	992.6	365

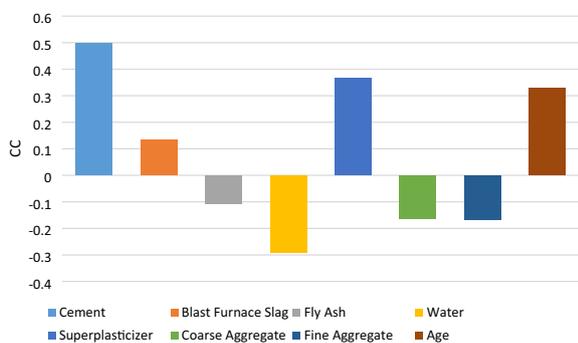


Fig. 2 Correlation of input variables with compressive strength (MPa)

fine aggregate) are inversely correlated. Positive correlation here implies that an increase or decrease in these input variables result in corresponding increment or decrement in the compressive strength, respectively. On the other hand, increase in the inversely correlated variables leads to decrease in the compressive strength of concrete and vice versa.

3.2 Experimental Setup

The steps involved in the experimental setup of this research is depicted in Fig. 3. Following the statistical description of each variable of the data set (see Sect. 3.1) is data normalization. This is a common pre-processing stage in most machine learning pipeline to avoid numerical overflow while keeping the input variables within a uniform range. Due care has been taken to split the data into training and test partitions before data normalization to avoid data leakage (O’Neil & Schutt, 2013). All input variables were normalized, such that the values are within the range of -1 and 1.

Cross validation is often used to assess the generalization capability of models in ML by splitting a given data set into two parts, where a portion is used for model training and the other is used to test how well the trained model is likely to generalise to an unseen data. However, due to varying ratios of training–test splits that have been reported in the literature, the performance of GBR, XGB, LGBM and CBT with optimized hyperparameters are initially examined across five training–test ratios

that include 90:10, 85:15, 80:20, 75:25 and 70:30. The experimental results of this process are presented and discussed in Sect. 5.2. The hyperparameter optimization for each model is carried out using only the training split of the data set to ensure that each model does not have access to the test partition prior to testing as in real-life application of machine learning. Each of the gradient-boosting algorithms considered in this study has a wide range of tuneable hyperparameters for optimal model performance; however, only a few have been selected for optimization. An exhaustive search of every possible combination of values within a specified range for each selected hyperparameters is used to train each model using fivefold cross validation. In other words, the training data are further divided into 5 equal partitions, each of which is, respectively, used to test the performance of a model trained with the remaining four partitions using a combination of hyperparameters at a time. The combination of hyperparameters that produce the best (lowest) average mean squared error over this process is deemed the optimal model parameters that is used to train the model on the entire training set before testing with the test partition that was initially set aside. As reported by Nguyen-Sy et al. (2020), increasing the number of estimators is similarly found to generally result in improved model performance. Hence, a search space of 10 to 1000 estimators is considered in this study.

Table 2 shows the hyperparameter search space and the optimal combination of hyperparameters for the 90:10 training–test split for all models. The model trained with optimal parameters is then evaluated using the evaluation metrics described in Sect. 3.3. All experiments were performed using python programming language. The Scikitlearn (Pedregosa et al., 2011) implementation of gradient-boosting regressor was used for GBR model, whereas the official python implementations of XGB, LGBM and CBT were similarly used for the respective model implementations.

3.3 Evaluation Metrics

To evaluate the performance of the developed machine learning models in this study, widely accepted statistical metrics such as the coefficient of determination (R^2), root

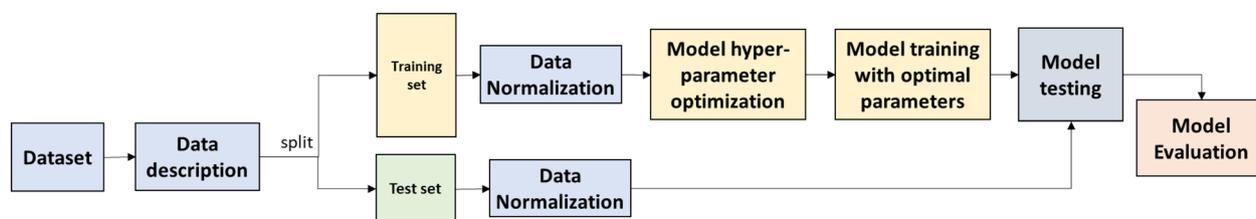


Fig. 3 Research flow

Table 2 Optimal hyperparameters for gradient-boosted models

Hyperparameter	Search space	Optimal parameters			
		GBR	LGBM	XGB	CBT
No of estimators	10–1000	520	780	890	980
Learning rate	0.01–0.5	0.12	0.13	0.12	0.12
Max features	0.5–1	0.7	1	0.7	Default
Subsample	0.7–1	0.8	Default	0.8	Default
Max depth (depth)	1–10	5	5	5	5
Num leaves	0–10	–	10	Default	Default
alpha (L1 Reg)	0.01–2.0	–	0.14	0.6	–
Lambda (L2 Reg)	0.01–2.0	–	Default	Default	1

mean squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) were applied. The mathematical formulation for each of these metrics are presented in Eqs. (5)–(8):

$$R^2 = 1 - \frac{\sum_{m=1}^n (y_m - \hat{y}_m)^2}{\sum_{m=1}^n (y_m - \bar{y})^2} \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{m=1}^n (y_m - \hat{y}_m)^2} \quad (6)$$

$$\text{MAE} = \frac{1}{n} \sum_{m=1}^n |y_m - \hat{y}_m| \quad (7)$$

$$\text{MAPE} = \frac{1}{n} \sum_{m=1}^n \frac{|y_m - \hat{y}_m|}{\max(\varepsilon, |y_m|)} \quad (8)$$

where y_m is the experimental output, \hat{y}_m is the model estimated output, \bar{y} is the mean of the experimental output, $\bar{\hat{y}}$ is the mean of the estimated output, n is the number of samples. MAPE also has ε which stands for an arbitrarily positive small constant to avoid division by zero when y_m is zero. For each of MAPE, RMSE and MAE, the lower the value, the better the model. On the contrary, achieving an R^2 value close to 1 is the goal of the learning algorithm, i.e., the closer the R^2 value to 1, the better. A baseline model which always predict the mean of the experimental output \bar{y} will have an R^2 of value 0, whereas a worse model than the baseline will produce a negative R^2 value.

In addition to the results based on these evaluation metrics, a ranking test using Friedman's test (Friedman, 1940) is also carried out to test the null hypothesis that the means of the results of the gradient-boosting ensemble

methods are the same at significance level of 0.05. If this null hypothesis is rejected, Holms's test (Holm, 1979) is performed as a post-hoc analysis of the pairwise comparison of the performance these methods is carried to establish if one is significantly better. The null hypothesis of the Holms's test is that the mean of the results of a pair of groups is equal. All statistical analysis were carried out on the STAC web platform for statistical analysis (Rodríguez-Fdez et al., 2015).

4 Results and Discussion

4.1 Model Performance Across Varying Training–Test Splits

As hinted in Sect. 3.2, the lack of a globally accepted training–test split ratio inspired a preliminary study on five popular training–test ratios that include 90:10, 85:15, 80:20, 75:25 and 70:30 (e.g., 75:25 implies that 75% of the data set is used for training, while the remaining 25% is used for testing). For each training–test ratio and learning algorithm, hyperparameter optimization is first carried out as described in Sect. 4.2 before model training and testing for estimation of compressive strength. The training and test performance of GBR, XGB, LGBM and CBT for the different training–test splits in terms of RMSE, R^2 , MAPE and MAE is presented in Fig. 4. As expected, the training performance of the model for the different training–test splits is generally better than their respective test performance across the evaluation metrics. However, being the true measure of the performance of the models, the test performances are relatively impressive given the marginal difference between the training and test scores. The general trend from Fig. 4 shows that as the test fraction of the training–test ratio increases, the models' respective performance tends to decrease across the evaluation metrics. Moreover, unlike the remaining training–test ratios, 90:10 consistently produced the best performance across the evaluation metrics for each learning algorithm; corroborating what was reported in (Salami et al., 2021). Hence, the result of the 90:10 training–test ratio for each of GBR, XGB, LGBM and CBT is selected and discussed in detail in the next section. The mean of the training and test scores for each model (with standard deviation) over the different ratios are also presented for each evaluation metric in Table A of the supplementary material.

4.2 Performance Comparison of Best Performing Model for Each Algorithm

Table 3 presents the training and test scores based on the evaluation metrics for compressive strength estimation using the ML methods under study. The best result for each metric is highlighted in bold. In terms of R^2 which measures how well the models approximate the experimental compressive strengths of each concrete mixture,

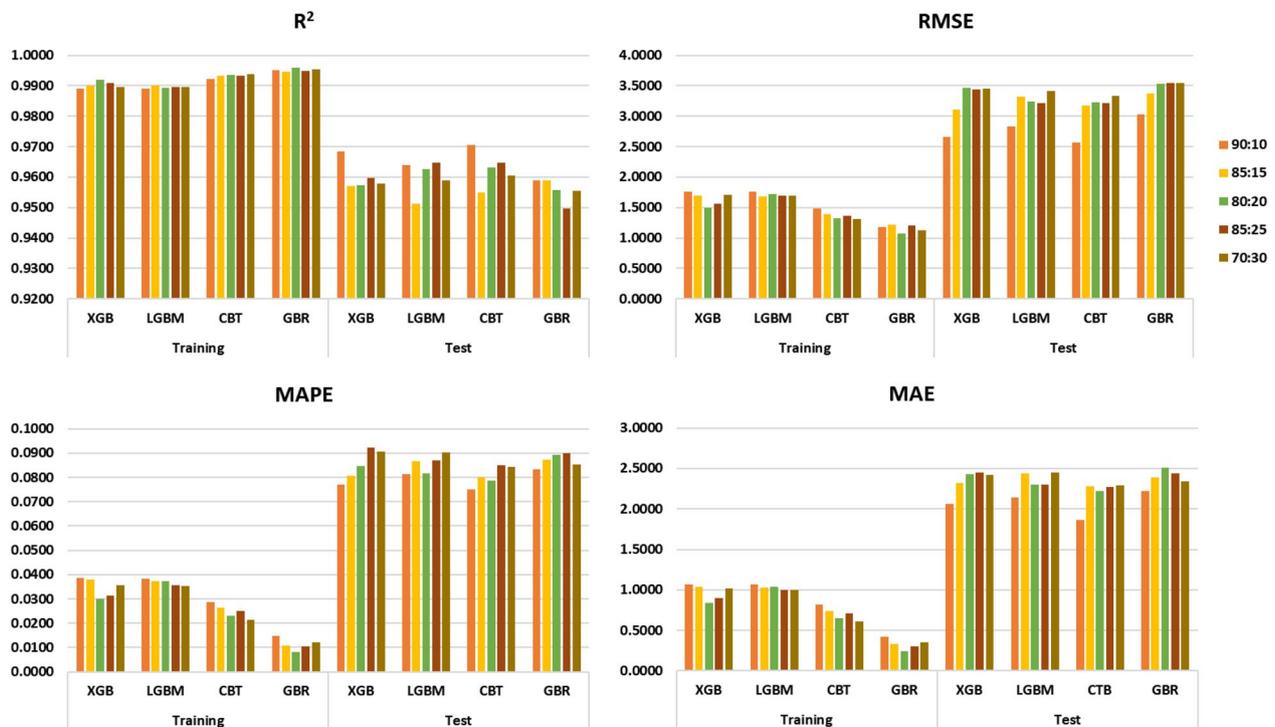


Fig. 4 Training and test performance of ML models with different training–test splits

Table 3 Training and testing performance of the models (↑ Higher is better, ↓ lower is better)

	Training				Test			
	↑ R^2	↓RMSE (MPa)	↓MAE (MPa)	↓MAPE	↑ R^2	↓RMSE (MPa)	↓MAE (MPa)	↓MAPE
XGB	0.9909	1.6016	0.9246	0.0330	0.9764	2.4972	1.9032	0.0744
LGBM	0.9890	1.7578	1.0599	0.0392	0.9745	2.5963	2.0067	0.0788
CBT	0.9930	1.4045	0.7218	0.0256	0.9838	2.0709	1.5966	0.0629
GBR	0.9950	1.1826	0.4259	0.0148	0.9731	2.6642	1.9013	0.0717

the training and test performances of GBR (0.9950 and 0.9731), XGB (0.9909 and 0.9764), LGBM (0.989 and 0.9745) and CBT (0.993 and 0.9838) are, respectively, very impressive given the small generalization gaps of 0.0219, 0.0145, 0.0145, and 0.0092 between the training and test performances of the respective models. This implies that despite fitting the training data to near perfection, the models are still able to generalize their training performance quite well. However, comparatively, the test R^2 score of 0.9838 achieved by CBT is better than 0.9731, 0.9764 and 0.9745 produced by GBR, XGB and LGBM, respectively. This indicates a performance improvement of 1.1%, 0.75% and 0.95% over the trio, respectively.

A comparison of the experimental and estimated compressive strengths by the gradient-boosted ML

models are presented in Fig. 5. Fig. 5 shows the scatter plots of the estimated compressive strengths plotted against experimental ones with the respective line of best fit for the training and test phases of each of GBR, CBT, XGB and LGBM models. The plots intuitively illustrate how correlated the model estimations are to the experimental values. The corresponding R^2 (i.e., coefficient of determination) value on each plot summarises its performance with a single score. In general, the plots show that despite producing a more correlated training estimations of the compressive strength, GBR produced the least correlated estimates in the test phase. The test compressive strength estimations of CBT are most correlated with the experimental values, followed by the XGB then LGBM.

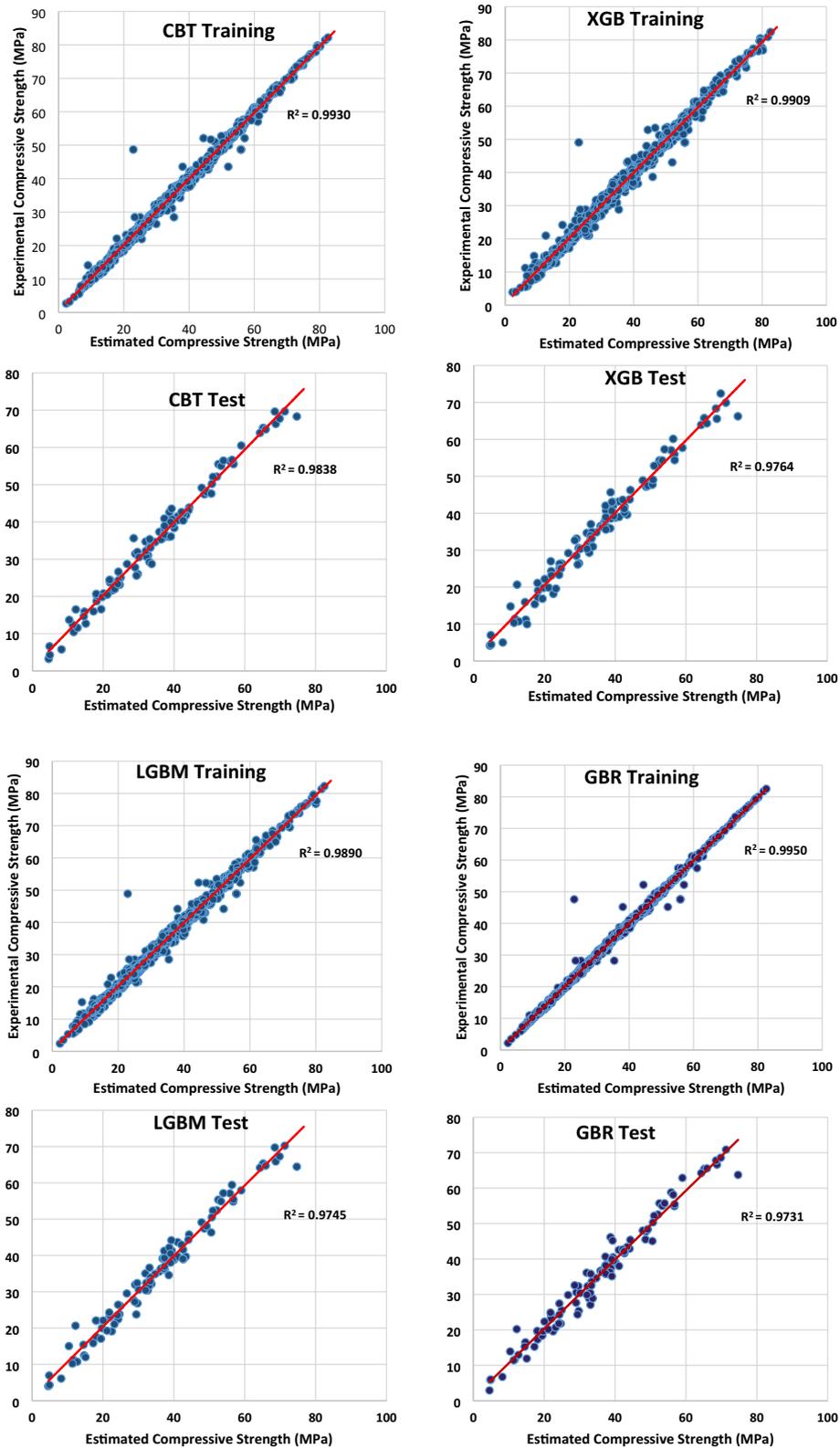


Fig. 5 Comparison of experimental and estimated compressive strength for the training and test phases of each model

Also presented in Table 3 for each model are the respective training and test performances in terms of RMSE, MAE and MAPE. It is worthy of note that unlike R^2 , these statistical evaluation measures seek to approximate the errors between the experimental values and model estimations as described in Sect. 4.3. Based on these metrics, the respective training and test performances of GBR (RMSE=1.1826 MPa and 2.6642 MPa; MAE=0.4259 MPa and 1.9013 MPa; MAPE=0.0148 and 0.0717), XGB (RMSE=1.6016 MPa and 2.4972 MPa; MAE=0.9246 MPa and 1.9032 MPa; MAPE=0.033 and 0.0744), LGBM (RMSE=1.7578 MPa and 2.5963 MPa; MAE=1.0599 MPa and 2.0067 MPa; MAPE=0.0392 and 0.0788) and CBT (RMSE=1.4045 MPa and 2.0709 MPa; MAE=0.7218 MPa and 1.5966 MPa; MAPE=0.0256 and 0.0629) are very impressive given the respective generalization gaps of 1.4816 MPa, 0.8956 MPa, 0.8385 MPa and 0.6664 MPa in terms of RMSE, 1.4754 MPa, 0.9786 MPa, 0.9468 MPa and 0.8748 MPa in terms of MAE as well as 0.0569, 0.0414, 0.0396 and 0.0373 in terms of MAPE for the respective models. Amongst these models, the GBR model produced the largest differences between the training and test scores, hence the least generalization despite fitting the training data best. On the other hand, the CBT model generalizes best while also producing the best test performance across the different metrics. Although, the GBR model fits the training data best, in terms of test performance which is the true measure of model performance, CBT produced a superior performance to GBR, XGB and LGBM across all the error-based evaluation metrics with a performance improvement ranging from 17% to 22%, 16% to 20.4% and 12% to 20% in terms of RMSE, MAE and MAPE, respectively.

Presented in Figs. 6, 7, 8 and 9 are the superimposed line plots of experimental and estimated compressive strengths for the training and test phases (a and b) alongside the corresponding error plots (c and d) for each of the considered gradient-boosting models. The errors for the training and test phases of each model are obtained by subtracting the estimated value of compressive strength for each data sample from its corresponding experimental value in the data sets. Since the aim of the model is to estimate the actual compressive strength as closely as possible, the lesser the deviation of the error plot from zero, the better.

It can be observed from the test error plots (Fig. 6d) that the CBT model shows the least deviation as it only deviates by error more than an |5| at only two occasions (sample indexes 66 and 87) compared to seven, five and three cases in GBR (sample indexes 35, 64, 66, 69, 71, 75 and 86 as in Fig. 9d), XGB (sample indexes 17, 35, 58, 66 and 69, as shown in Fig. 7d) and LGBM (sample indexes 35, 66 and 75 as in Fig. 8d) models, respectively.

It is noteworthy that while all the models, respectively, exceeded |5| error mark on sample index 66, the GBR model notably deviated by |11| on this sample index; making it the least performing model in this regard.

4.3 Average Performance of Models

To further ensure that the performance of the gradient-boosted machine learning algorithms compared in this study is not by chance, the same experiment was repeated 100 times for each of the models using the same set of optimal hyperparameters presented in Table 2. The original data was repeatedly split into training–test partitions for different repetitions of the experiment using different random seeds to ensure that different sets of training and test samples were used each time over the whole process. The mean and standard deviation of the training and test performances of each of GBR, XGB, LGBM and CBT over the 100 repetitions are presented in Fig. 10 for each statistical evaluation measures. As expected, and hinted earlier, the average training performance of each model is generally better than the corresponding average test performance across the evaluation metrics with GBR mostly performing best in this regard followed by CBT.

Similarly, the training performance shows minimal deviation from their respective means compared to the test performance. In terms of the test performance, CBT ($R^2=0.9506$, RMSE=3.6051, MAE=2.2462, MAPE=0.0774) generally produced the best average performance based on all evaluation metrics, whereas GBR ($R^2=0.9444$, RMSE=3.8406, MAE=2.4247, MAPE=0.0836) ranks lowest in all but MAE and MAPE, where it shows comparable or slightly better performance than LGBM ($R^2=0.9467$, RMSE=3.7644, MAE=2.4386, MAPE=0.0862) and XGB ($R^2=0.9468$, RMSE=3.7638, MAE=2.4371, MAPE=0.0854) on average. Although, XGB marginally outperform LGBM on the specific result presented in Table 3, the average performance of XGB and LGBM are mostly similar with XGB slightly performing better over the hundred repetitions. Overall, CBT ranks best on the average, followed by XGB, LGBM, then GBR across all the evaluation measures.

4.4 Statistical Analysis of Results

In addition, a statistical analysis of the obtained results in terms of R^2 and RMSE is presented here. Using the test results from 100 repetitions of experiments from the preceding section, the null hypothesis of the Friedman's test is rejected given p values of 0.00000 (less than significance level of 0.05) for both R^2 and RMSE results, respectively. The Friedman's ranking tests for both R^2 and RMSE rank the gradient boosting ensembles algorithms similarly in descending order as follows, CBT > XGB > LGBM > GBR. While this ranking signifies that CBT

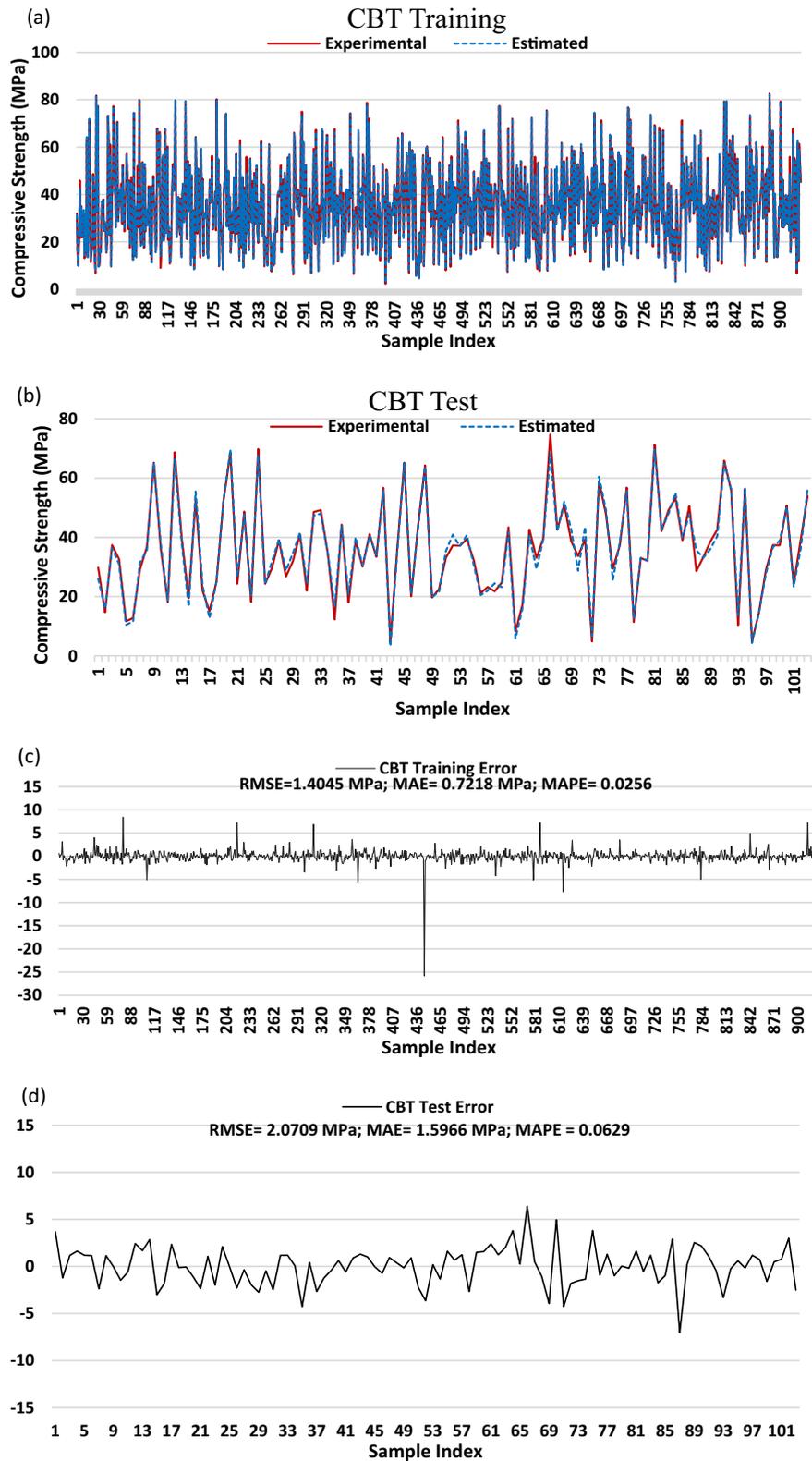


Fig. 6 Superimposed line plots of experimental and estimated compressive strength for **a** training and **b** test phases and corresponding error plots over the **c** training and **d** test data for CatBoost

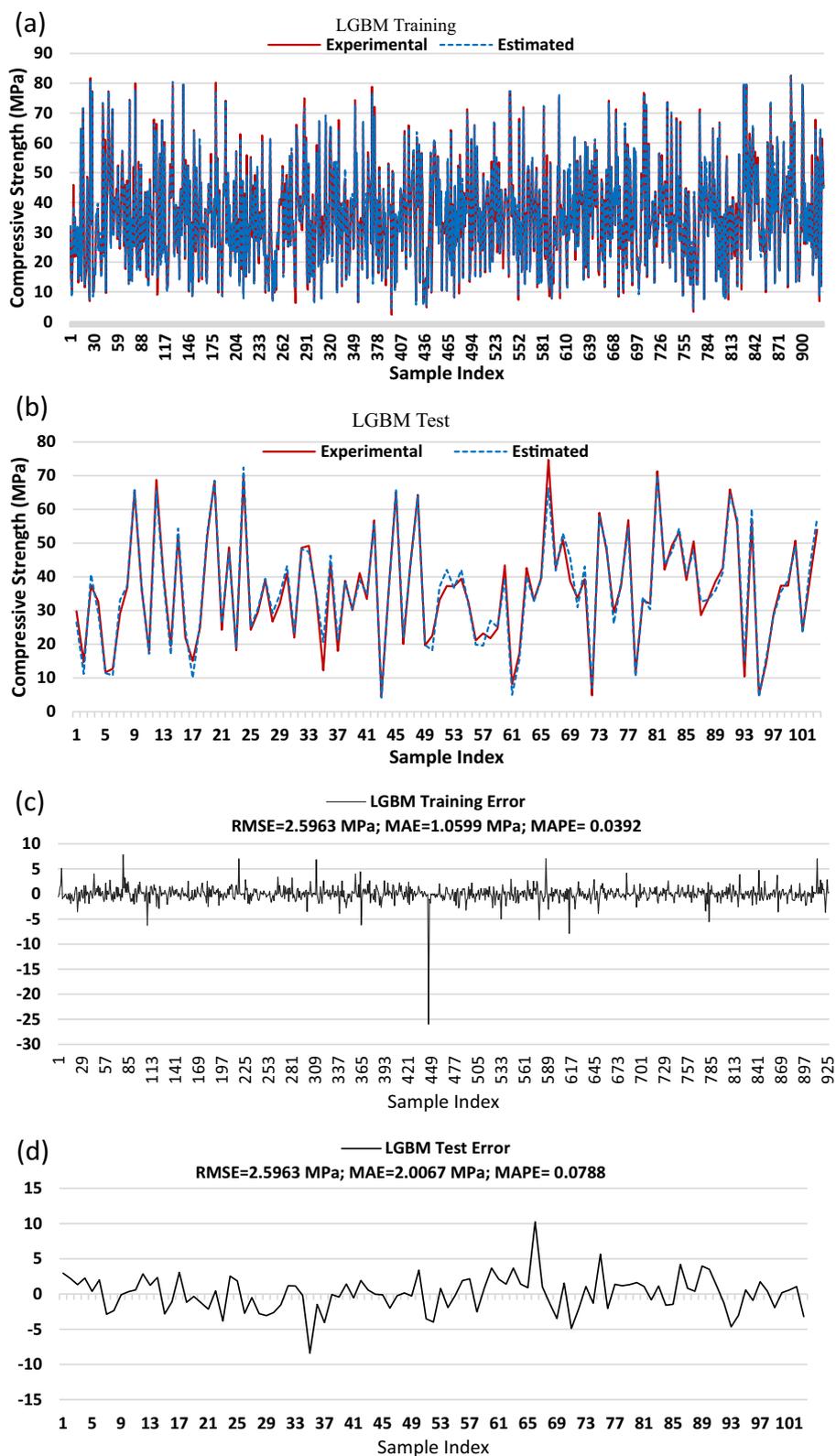


Fig. 7 Superimposed line plots of experimental and estimated compressive strength for **a** training and **b** test phases and corresponding error plots over the **c** training and **d** test data for LightGBM

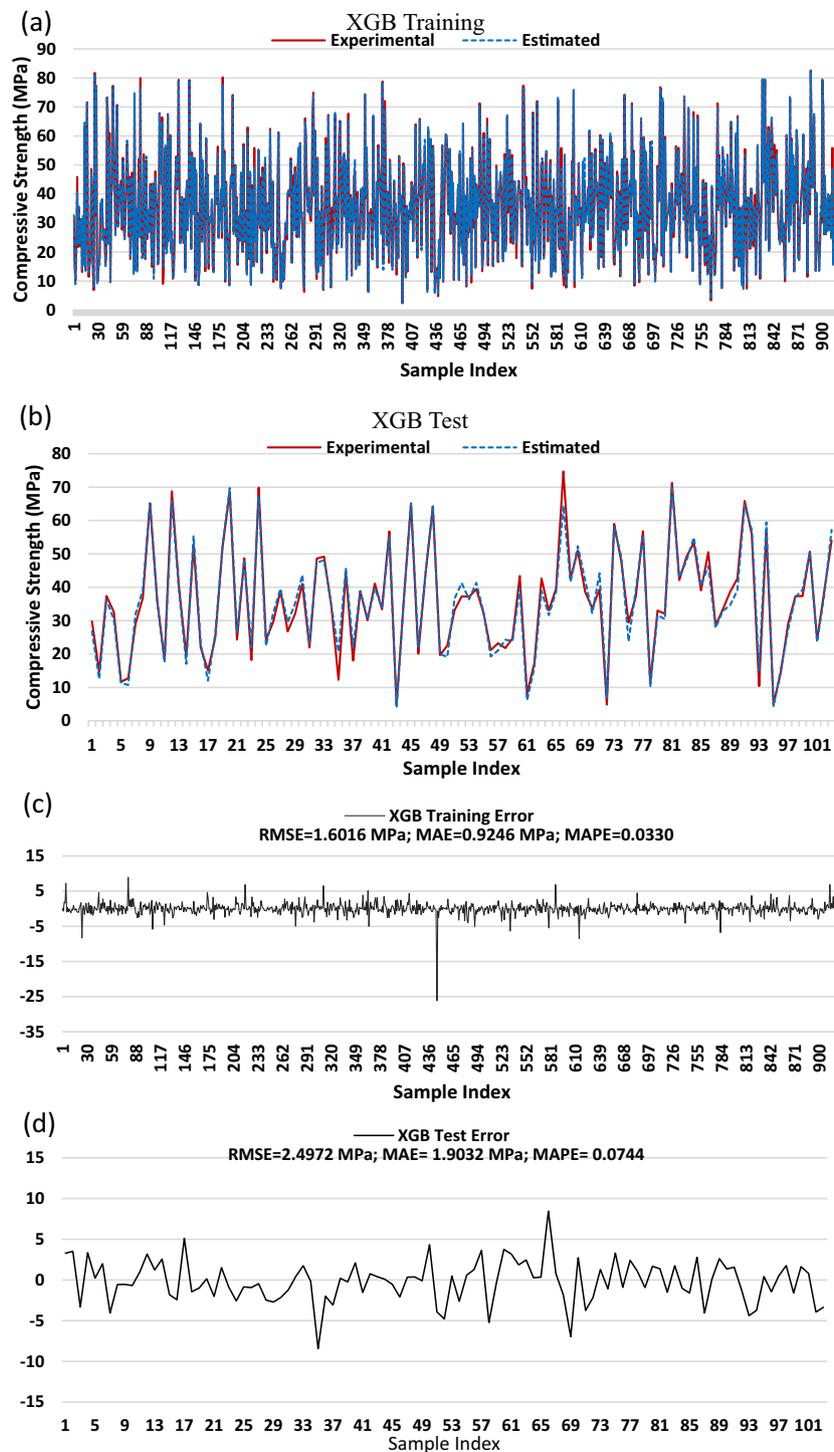


Fig. 8 Superimposed line plots of experimental and estimated compressive strength for **a** training and **b** test phases and corresponding error plots over the **c** training and **d** test data for XGBoost

ranks highest amongst the algorithms, how significant the difference between each pair of algorithms remains unclear. Hence, the need for pairwise post hoc analysis

using Holm’s test. The Holm’s test results for R^2 and RMSE are presented in Table 3. It can be observed that the null hypothesis at significance level of 0.05 is rejected

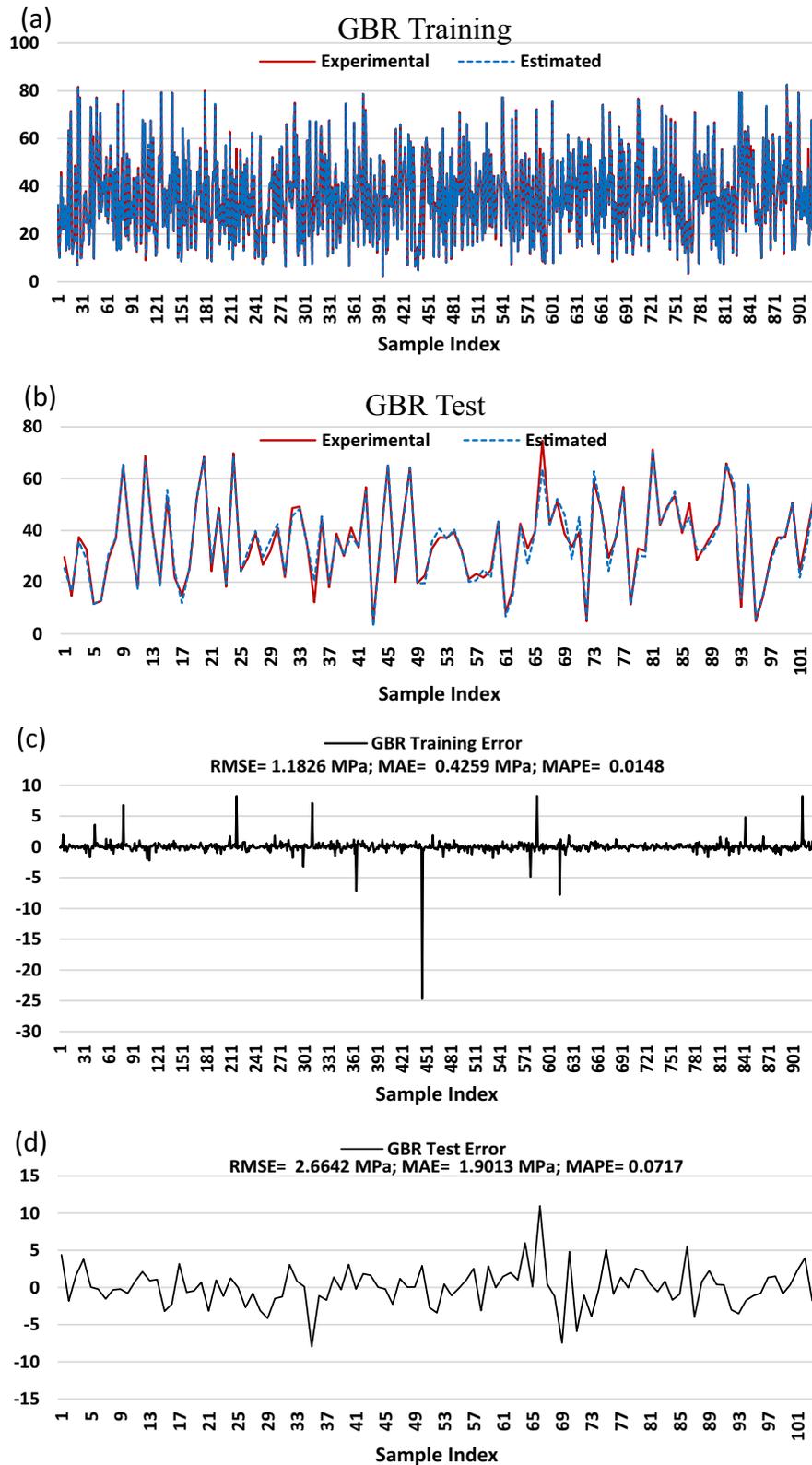


Fig. 9 Superimposed line plots of experimental and estimated compressive strength for **a** training and **b** test phases and corresponding error plots over the **c** training and **d** test data for gradient-boosting regressor

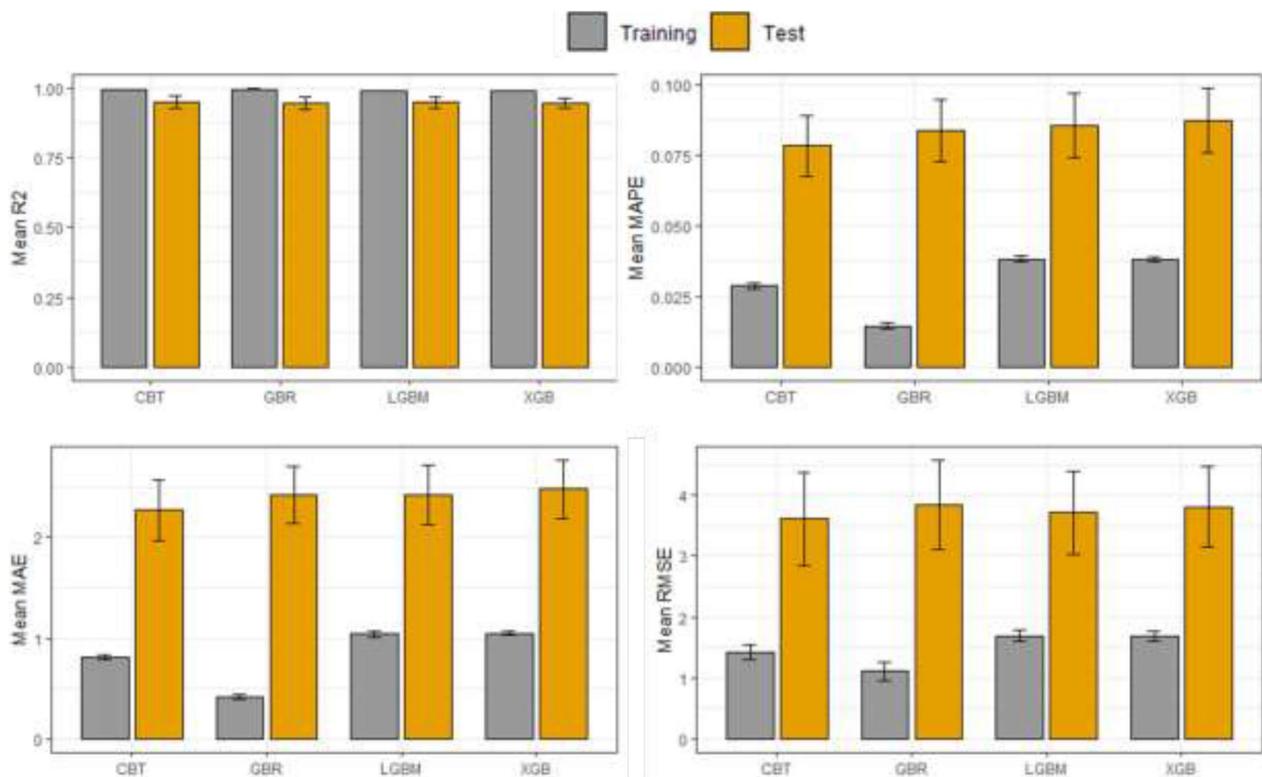


Fig. 10 Mean (\pm Standard Deviation) performance of gradient-boosted models over 100 repetitions of experiments

for all pairwise combination except LGBM vs XGB for both evaluation metrics. This shows that, although XGB ranks higher than LGBM, the difference between them is not statistically significant. Conversely, CBT is significantly better than any other methods (Table 4).

4.5 Feature Importance

Being able to understand or interpret the decision or the cause of the decision a machine learning model makes is integral to improved human understanding of the data, the model and relationship between them. The quest for this has paved way for a whole new active area of research known as interpretable machine learning (Murdoch et al., 2019). Similarly, this section seeks to provide insight into the decision of each of the considered machine learning models in this study relative to the data set. While earlier works on compressive strength estimation have rarely explored this line of research, there has been a notable increase in studies exploring this line of research. Some of which have investigated the importance of input features in the prediction of mechanical properties of pervious concrete using extreme gradient boosting and support vector regression as well as Adaboost (Feng et al., 2020; Güçlüer et al., 2021; Mustapha et al., 2022). In this study, the feature importance function which can be

Table 4 Results of pairwise post-hoc analysis using Holm's test

Comparison	Statistic	Adjusted <i>p</i> value	Null hypothesis
R²			
GBR vs CBT	8.43469	0.00000	Rejected
LGBM vs CBT	5.90936	0.00000	Rejected
XGB vs CBT	5.69224	0.00000	Rejected
XGB vs GBR	2.74245	0.01829	Rejected
LGBM vs GBR	2.52533	0.02312	Rejected
LGBM vs XGB	0.21712	0.82811	Accepted
RMSE			
GBR vs CBT	8.63897	0.00000	Rejected
LGBM vs CBT	6.07266	0.00000	Rejected
XGB vs CBT	5.92649	0.00000	Rejected
XGB vs GBR	2.71248	0.02003	Rejected
LGBM vs GBR	2.56631	0.02056	Rejected
LGBM vs XGB	0.14617	0.88378	Accepted

called on each of the fitted models of the Python implementations of CatBoost, LightGBM, XGBoost and gradient-boosting regressor is used to get the contribution of each input feature to the respective models.

Figs. 11, 12, 13 and 14, respectively, present a ranking of the input features for CBT, LGBM, XGB and GBR in

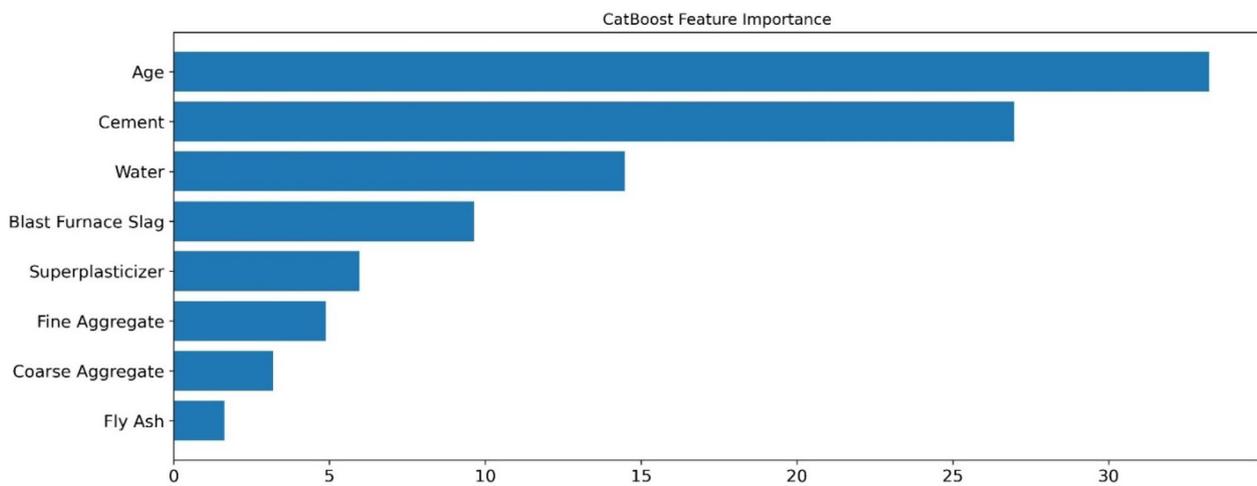


Fig. 11 Feature importance of CatBoost model

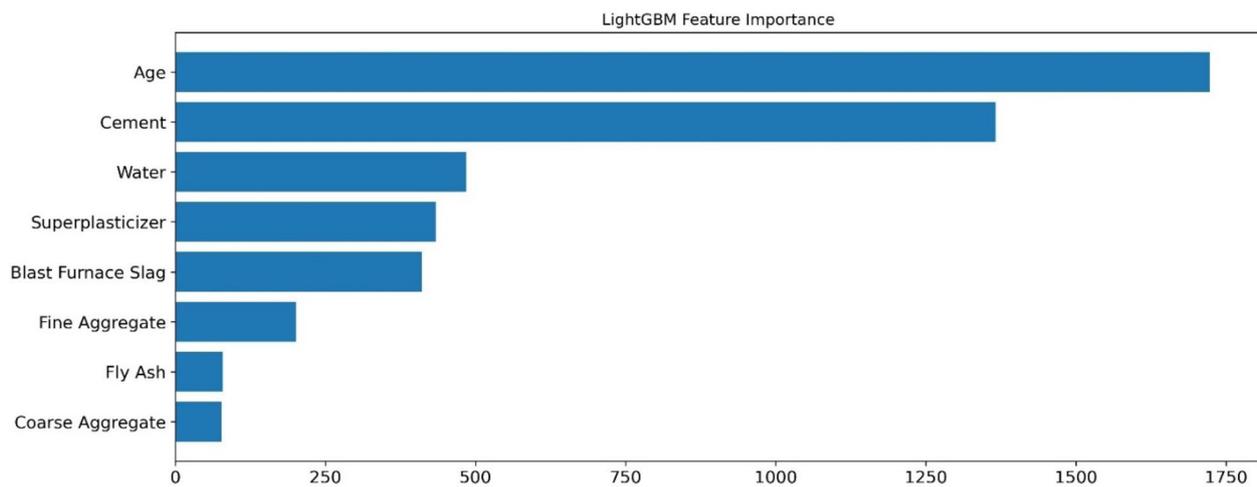


Fig. 12 Feature importance of LightGBM model

descending order of importance. There is consensus amongst all the models that the top three most important feature to the estimation of compressive strength are the Age (in days) of each of the concrete mixtures followed by the quantity of cement (in kg/m^3), then water (in kg/m^3). This confirms what has been reported in earlier studies that the compressive strength of concrete increases with time (Abdulkareem et al., 2019; Sharmila & Dhinakaran, 2016). At the bottom end of the feature importance ranking is coarse aggregate (in kg/m^3) with the least relevance to the predictive performance of XGB and LGBM, whereas the fly ash (in kg/m^3) component of each mixture has the least contribution to the predictive decision of the GBR and CBT models. These findings further corroborate what has been reported in pertinent

works relating the importance of age, cement as well as water quantity in the estimation of compressive strength of concrete (Cakiroglu et al., 2023; Feng et al., 2020; Güçlüer et al., 2021).

4.6 Sensitivity Analysis

A sensitivity analysis of all the input variables employed in estimating the compressive strength is presented here to understand how changes in each input variable bring about corresponding changes in the estimated model outputs. It is noteworthy that while the correlation analysis presented in Fig. 2 can be viewed as a form of sensitivity analysis, it only represents the static relationship between each input variable and the output irrespective of the model. Here, the relationship between the input

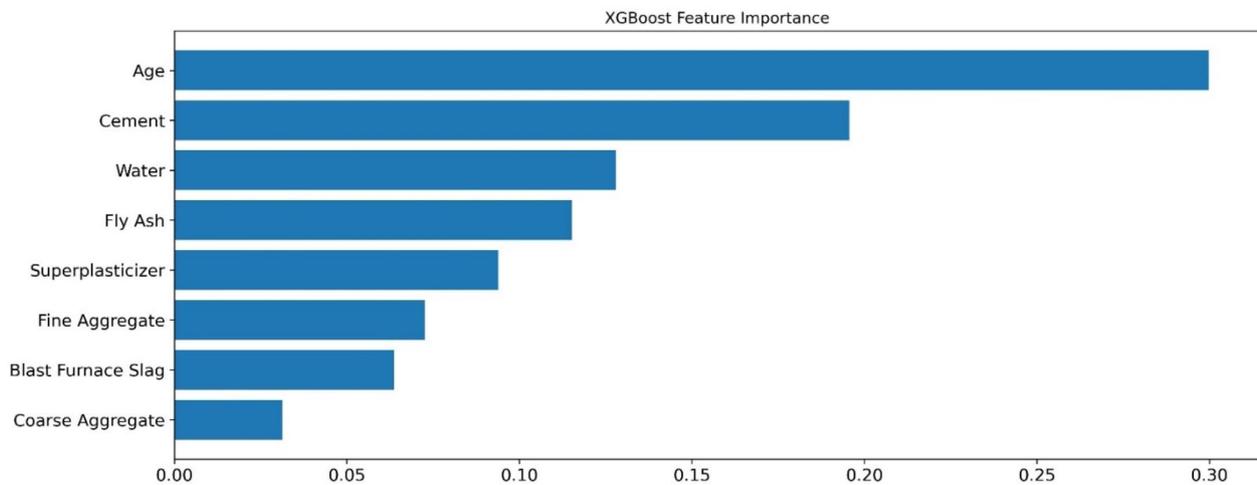


Fig. 13 Feature importance of XGBoost model

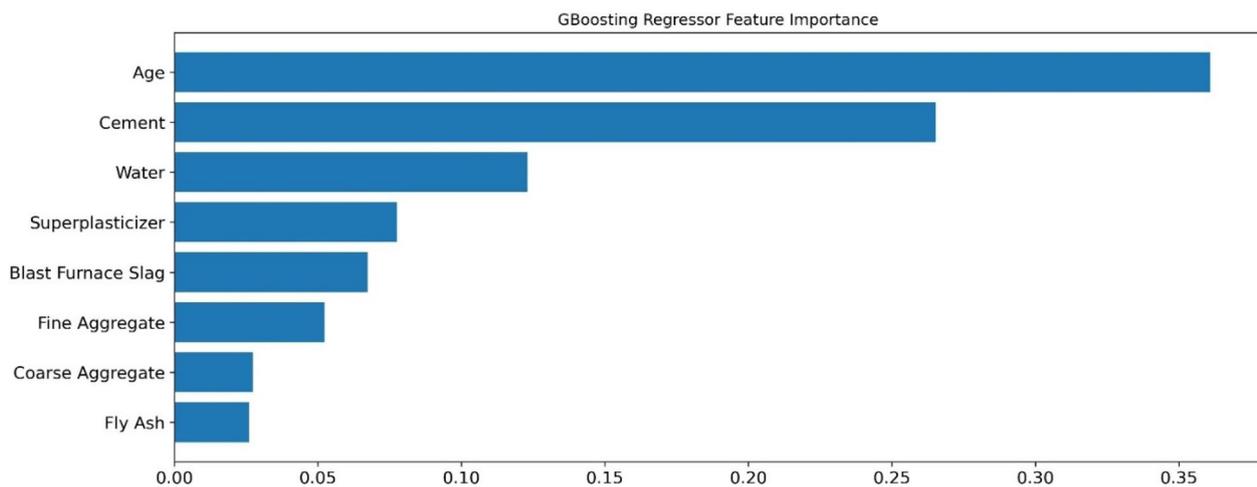


Fig. 14 Feature importance of gradient-boosting regressor model

variables and the estimated output from the perspective of each model is presented. This is achieved by showing the marginal effect each feature has on the predicted outcome of GBR, CBT, LGBM and XGB models with the aid of partial dependence plots (PDP) (Hastie et al., 2009). The PDP is a global method that considers all instances and gives a statement about the global relationship of a feature with the predicted outcome. In the current study, each gradient-boosting ensemble model has been fitted to estimate the compressive strength of concrete mixtures and PDP is used to visualize the relationships each model has learnt as presented in Fig. 15a–d for CBT, GBR, LGBM and XGB, respectively.

It is interesting to note that the relationship between each input feature and the estimated output (compressive

strength) exhibit similar trend across the gradient-boosting models. For instance, the relationship between cement quantity and the estimated compressive strength is linear for all models, with increasing cement quantity yielding corresponding increase in compressive strength across the models. Similar pattern can be observed in relation to the age of the concrete mixtures albeit the compressive strength plateaus after about 100 days, indicating no significant increase in the compressive strength of the mixtures after this period. While the range of training compressive strength values (which is 2.33–82.6 MPa in this study) used for model building in highly influential to model estimations, representative works such as (Abdulkareem et al., 2019; Sharmila & Dhinakaran, 2016) alluded to slower increase in compressive strength of

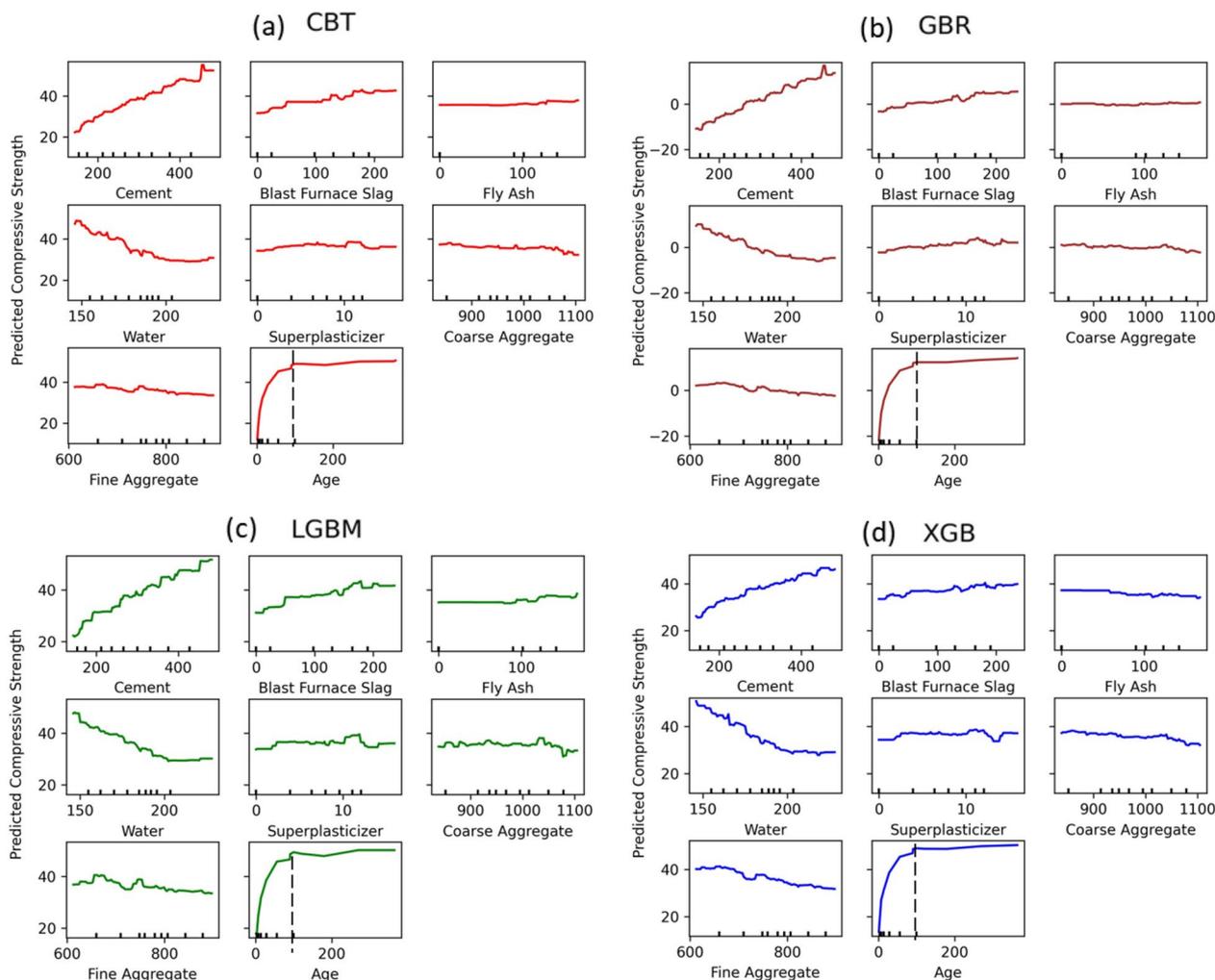


Fig. 15 Partial dependence plots for the **a** CBT, **b** GBR, **c** LGBM and **d** XGB compressive strength estimation models

concrete mixtures after the first 3 months. On the other hand, an inverse relationship exists between the model estimations and water quantity across the models, with increase in water quantity from 150 to 200 kg/m³ resulting in decrease in compressive strength. Interestingly, the estimated compressive strength does not decrease across the models when water quantity increases beyond 200 kg/m³. For other input features, such as fine aggregate and blast furnace slag, the estimated compressive strength slowly and marginally decreases as the former increases, while a marginally decreasing trend can be observed as the latter increases. The intuitive nature of the input–output relationships shown by the models reflect well the models learn from the given data.

4.7 Comparison with Previous Works

Given that compressive strength is one of the most important structural material properties in concrete

research and design, several studies have developed intelligent approaches for its accurate estimation over the past years. A considerable number of these studies have used either part or whole of the Lichman (2013) data set used in this research. Hence, it is considered worthwhile to compare the results obtained herein with the best results that have been reported in pertinent studies. Admittedly, ensuring an objective comparison of performance with previous studies can be challenging, given the differences in statistical evaluation metrics, training–test split ratios (e.g., some may use 90:10 ratio, while others may use 70:30), sample size (e.g., some may use a subset of the data set, while others use the complete 1030 samples) and the general experimental setup. Notwithstanding, the comprehensive nature of the experiments carried out in this study naturally answers some of these concerns. Table 5 presents details of the representative studies grouped by experimental design, algorithm,

Table 5 Comparison with previous studies

Researcher	Algorithm	Data set size	R ²	RMSE	MAE	MAPE
Based on average performance/multiple cross validation						
(Chou et al., 2011)	ANN	1030	0.909	5.030		0.109
	SVM		0.886	5.619		0.128
	Multiple regression (MR)		0.611	10.429		0.317
	Multiple Additive Regression Tree (MART)		0.911	4.949		0.139
	Bagging Regression Tree (BRT)		0.890	5.572		0.142
(Erdal, 2013)	Decision Tree (DT)	1030	0.818			
	Bagging DT		0.879			
	Gradient-boosted DT		0.889			
	Random Sub-spaced DT		0.868			
(Chou et al., 2014)	SVM	1030		5.59	3.75	0.12
	Stacked CART+ SVM+ LR			5.08	3.52	0.12
(Feng et al., 2020)	Adaboost	1030	0.952	4.856	3.205	0.114
(Farooq et al., 2021)	Modified Random Forest Ensemble	1030	0.923	4.6	3.23	
(Chen et al., 2021)	CNN	1030	0.97	3.98	2.68	
	CNN-AP		0.97	4.09	2.92	
	CNN-MP		0.96	4.18	2.89	
This Study	XGB	1030	0.947	3.764	2.437	0.085
	LGBM		0.947	3.764	2.439	0.086
	CBT		0.951	3.605	2.246	0.077
	GBR		0.944	3.841	2.425	0.084
Based on single cross validation						
(Erdal et al., 2013)	ANN	1030	0.909	5.57	4.18	
	Bagged ANN		0.928	4.87	3.60	
	Gradient-boosted ANN		0.927	5.24	4.09	
	Wavelength-bagged ANN		0.94	4.54	3.30	
	Wavelength gradient-boosted ANN		0.953	5.75	4.83	
(Silva et al., 2020)	ANN	1030	0.89	5.9		
	SVM		0.83	7.5		
	Random forest		0.90	5.6		
(Dao et al., 2020a, 2020b)	Gaussian Process Regression (GPR-52)	1030	0.884	5.702	4.058	
	Gaussian Process Regression (GPR-32)		0.888	5.597	3.913	
	GPR using Exponential Kernel (GPR-EXP)		0.888	5.600	3.924	
	GPR using Square Exponential Kernel (GPR-SQEXP)		0.878	5.849	4.242	
	GPR using Rational Quadratic Kernel (GPR-RSQ)		0.880	5.793	4.182	
	Levenberg-Marquardt ANN		0.890	5.447	4.274	
(Feng et al., 2020)	Adaboost	1030	0.982	2.20	1.64	0.0678
(Salami et al., 2021)	LSSVM-CSA	1030	0.954	3.335		
	GP		0.894	4.662		
This Study	XGB	1030	0.976	2.497	1.903	0.074
	LGBM		0.974	2.596	2.007	0.079
	CBT		0.984	2.071	1.597	0.063
	GBR		0.973	2.664	1.901	0.072

data size and performance in terms of R², RMSE, MAE and MAPE where applicable. To ensure a fairly objective comparison, only studies that used the whole data set (i.e., 1030 samples) have been compared. The comparison

has been grouped into two main categories, namely, average performance and cross validation performance, respectively. The mean results of 100 repetitions of the experiments presented in Sect. 4.3 (Fig. 10) is compared

under the average performance category with the best results from studies in which experimental results were conducted using k-fold cross validation and the average performance reported, whereas the best results from studies that evaluate their models based on training–test cross validation are grouped under the cross validation category and compared with results presented in Table 3.

Table 5 presents the comparison of obtained results with the best from previous studies. A general observation from the table is the extensive use of ensemble models and paucity of gradient-boosted models in compressive strength estimation of quaternary blend concrete. In terms of average performance, the best performance found in relevant studies was reported in Feng et al. (Feng et al., 2020), where the proposed Ada-boost model yielded $R^2=0.952$, $RMSE=4.856$ MPa, $MAE=3.205$ MPa and $MAPE=0.114$. Compared to the best average performance obtained in this study, the CBT model produced a better result in all the evaluation metrics (25.76% $RMSE$, 29.92% MAE and 32.46% $MAPE$ improvements, respectively) except in terms of R^2 , where the score of 0.952 reported is marginally better than that average R^2 of 0.951 obtained over 100 repetitions (about 0.1% improvement). It should also be noted the average performances of GBR, XGB and LGBM in terms of $RMSE$, MAE and $MAPE$ are also better than what was reported in (Feng et al., 2020). Likewise, the best cross validation performance found in the literature is $R^2=0.982$, $RMSE=2.20$ MPa, $MAE=1.64$ MPa and $MAPE=0.0678$ reported in Feng et al. (Feng et al., 2020). In comparison with the best results obtained in this study, the R^2 , $RMSE$, MAE and $MAPE$ values of 0.984, 2.071 MPa, 1.597 MPa and 0.063 are better with performance improvement of 0.2%, 5.86%, 2.62% and 0.48%, respectively.

The impressive performance of the gradient-boosting models presented in this study generally reflect the robustness each of each model to different evaluation approaches for compressive strength of quaternary blend concrete estimation. However, it should be noted the performance reported in this study is limited to 1030 concrete mix with age ranging from 1 to 365 days.

5 Conclusion

A comparative analysis of prediction of compressive strength of quaternary blend concrete with gradient-boosted ensembles is presented in this study. Four popular gradient-boosting implementations, namely, gradient-boosting regressor (GBR), light gradient-boosting model (LGBM), extreme gradient boosting (XGB) and CatBoost (CBT) were, respectively, used to build models for compressive strength estimation and results based on an out-of-sample test set as well as average cross

validation are presented. Four popular evaluation metrics were used for performance evaluation with results showing that CBT outperformed other methods across all the metrics with values of 0.9838, 2.0709, 1.5966 and 0.0629 as the R^2 , $RMSE$, MAE and $MAPE$ values, respectively. An analysis of the most important features to model performance also shows that the age, quantity of cement and water in the concrete mixture have highest contributions to the compressive strength estimation of each model. In addition, a sensitivity analysis of the model prediction with varying values of input features confirms the importance of these features, notably showing no significant increase in compressive strength estimations after the first 100 days. Moreover, a comparison of results with findings from previous studies also shows the superiority of CBT and the other gradient-boosting models in estimating compressive strength. CBT not only outperform the models on single evaluation with an out of sample test but also in terms of average performance. It is hoped that these findings will further increase the awareness of the predictive capabilities of CBT amongst and thus, increase its use alongside the growing computational tools at their disposal.

This study, though comprehensive, is not without limitations. In relation to the data set, though, a fairly large representative one in concrete properties estimation, we acknowledge that machine learning models are only as good as their training data. Hence, the findings reported are based on the range of values reported in Sect. 3.1. Besides, the data set is not representative of all types of concrete mixtures, such as the rubberized recycled aggregate concretes and heat-treated concretes (Cakiroglu et al., 2023; Chen et al., 2021). These are viable areas for future investigation.

In addition, the relentless quest for improved accuracy of concrete properties and specifically compressive strength estimation has led to innovative learning methods, such as advanced deep learning algorithms with specialised loss functions Hoang (2023) and metaheuristic optimized DNN (Ranjbar et al., 2022) as well as ensemble of ensemble models (Lee et al., 2023). While these methods have potential shortcomings that relates to computational cost and overfitting, future works will explore feature selection, using only top-ranking features that contribute most to each model performance as shown in the feature importance and sensitivity analysis.

Abbreviations

ANFIS	Adaptive neuro-fuzzy inference system
ANN	Artificial neural network
CaO	Calcium oxide
CBT	Catboost
CO ₂	Carbo dioxide
DT	Decision tree

ELM	Extreme learning machine
FA	Fly ash
FFA	Firefly algorithm
GBDT	Gradient-boosted decision trees
GBM	Gradient-tree-boosting machine
GBR	Gradient-boosting regressor
GEP	Gene expression programming
GGBS	Ground granulated blast furnace slag
GHG	Greenhouse gas
GRA	Grey relation analysis
HPC	High-performance concrete
LightGBM	Light Gradient-Boosting Machine
MAPE	Mean absolute percentage error
MARS	Multivariate adaptive regression splines
ML	Machine learning
MLP	Multilayer perceptron
MSE	Mean squared error
PC	Portland cement
R^2	Coefficient of determination
RBF	Radial basis function
RF	Random forests
RFG	Random forest regression
RMSE	Root mean squared error
RRMSE	Relative root mean squared error
SCBA	Sugarcane bagasse ash
SF	Silica fume
SVM	Support vector machine
SVR	Support vector regression
UHPC	Ultra-high-performance concrete
XGBoost	Extreme gradient boosting

Acknowledgements

The authors thank UCSI University for financing this research via Research Excellence and Innovation Grant (REIG-FETBE-2020/041).

Author contributions

IBM: Data curation, experimental analysis. MA: Supervision, methodology. TMJ: Data processing, data curation. AHA: Data processing, Reviewing and Editing the manuscript. KAAA: Data processing, Writing- Original draft preparation. MMHA: Data processing, Writing- Original draft preparation. HN: Data curation, experimental analysis. SCA: Supervision, Final draft. ZA: Experimental testing. AG: Conceptualization of idea, Methodology. All authors read and approved the final manuscript.

Funding

The authors thank UCSI University for financing this research via Research Excellence and Innovation Grant (REIG-FETBE-2020/041).

Availability of data and materials

All data applied in this study are available on demand.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

The authors have given permission for publication.

Competing interests

The authors declare that they have no conflict of interest.

Author details

¹Computer Science Department, School of Computing, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia. ²Faculty of Engineering and Built Environment, UCSI University, 56000 Kuala Lumpur, Malaysia. ³Department of Civil Engineering, University of Hafr Al Batin, Hafr Al Batin, Saudi Arabia. ⁴Institute of Noise and Vibration, School of Civil Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia. ⁵Computer and Information Sciences Department, University of Strathclyde, Glasgow G1 1XQ, UK. ⁶Department

of Telecommunication Science, University of Ilorin, Ilorin, Nigeria. ⁷Department of Civil Engineering, British University of Bahrain, Saar, Bahrain.

Received: 22 July 2023 Accepted: 27 November 2023

Published online: 02 April 2024

References

- Abdulkareem, M., Ayeronfe, F., Abd Majid, M. Z., Sam, A. R. M., & Kim, J.-H.J. (2019). Evaluation of effects of multi-varied atmospheric curing conditions on compressive strength of bacterial (*Bacillus subtilis*) cement mortar. *Construction and Building Materials*, 218, 1–7.
- Abellán-García, J. (2020). Four-layer perceptron approach for strength prediction of UHPC. *Construction and Building Materials*, 256, 119465.
- Agrawal, Y., Gupta, T., Siddique, S., & Sharma, R. K. (2021). Potential of dolomite industrial waste as construction material: A review. *Innovative Infrastructure Solutions*, 6(4), 1–15.
- Alabdullah, A. A., Iqbal, M., Zahid, M., Khan, K., Amin, M. N., & Jalal, F. E. (2022). Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis. *Construction and Building Materials*, 345, 128296.
- Allujami, H. M., Abdulkareem, M., Jassam, T. M., Al-Mansob, R. A., Ibrahim, A., Ng, J. L., & Yam, H. C. (2022a). Mechanical Properties of Concrete Containing Recycle Concrete Aggregates and Multi-Walled Carbon Nanotubes Under Static and Dynamic Stresses. *Case Studies in Construction Materials*, 17, e01651.
- Allujami, H. M., Abdulkareem, M., Jassam, T. M., Al-Mansob, R. A., Ng, J. L., & Ibrahim, A. (2022b). Nanomaterials in recycled aggregates concrete applications: Mechanical properties and durability. A review. *Cogent Engineering*, 9(1), 2122885.
- Babajide Mustapha, I., & Saeed, F. (2016). Bioactive molecule prediction using extreme gradient boosting. *Molecules*, 21(8), 983.
- Badra, N., Haggag, S. A., Deifalla, A., & Salem, N. M. (2022). Development of machine learning models for reliable prediction of the punching shear strength of FRP-reinforced concrete slabs without shear reinforcements. *Measurement*, 201, 111723.
- Benhelal, E., Shamsaei, E., & Rashid, M. I. (2019). Novel modifications in a conventional clinker making process for sustainable cement production. *Journal of Cleaner Production*, 221, 389–397.
- Cakiroglu, C., Shahjalal, M., Islam, K., Mahmood, S. F., Billah, A. M., & Nehdi, M. L. (2023). Explainable ensemble learning data-driven modeling of mechanical properties of fiber-reinforced rubberized recycled aggregate concrete. *Journal of Building Engineering*, 76, 107279.
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., & Cho, H. (2015). "Xgboost: extreme gradient boosting." R package version 0.4–2 1(4): 1–4.
- Chen, H., Yang, J., & Chen, X. (2021). A convolution-based deep learning approach for estimating compressive strength of fiber reinforced concrete at elevated temperatures. *Construction and Building Materials*, 313, 125437.
- Chou, J.-S., Chiu, C.-K., Farfoura, M., & Al-Taharwa, I. (2011). Optimizing the prediction accuracy of concrete compressive strength based on a comparison of data-mining techniques. *Journal of Computing in Civil Engineering*, 25(3), 242–253.
- Chou, J.-S., Tsai, C.-F., Pham, A.-D., & Lu, Y.-H. (2014). Machine learning in concrete strength simulations: Multi-nation data analytics. *Construction and Building Materials*, 73, 771–780.
- Cook, R., Lapeyre, J., Ma, H., & Kumar, A. (2019). Prediction of compressive strength of concrete: critical comparison of performance of a hybrid machine learning model with standalone models. *Journal of Materials in Civil Engineering*. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0002902](https://doi.org/10.1061/(ASCE)MT.1943-5533.0002902)
- Cui, L., Chen, P., Wang, L., Li, J., & Ling, H. (2021). Application of extreme gradient boosting based on grey relation analysis for prediction of compressive strength of concrete. *Advances in Civil Engineering*. <https://doi.org/10.1155/2021/887839>
- Dan, A. K., Bhattacharjee, D., Ghosh, S., Behera, S. S., Bindhani, B. K., Das, D., & Parhi, P. K. (2021). *Prospective utilization of coal fly ash for making advanced materials* (pp. 511–531). Springer.
- Dao, D. V., Adeli, H., Ly, H.-B., Le, L. M., Le, V. M., Le, T.-T., & Pham, B. T. (2020a). A sensitivity and robustness analysis of GPR and ANN for high-performance

- concrete compressive strength prediction using a Monte Carlo simulation. *Sustainability*, 12(3), 830.
- Dao, D. V., Ly, H.-B., Vu, H.-L. T., Le, T.-T., & Pham, B. T. (2020b). Investigation and optimization of the C-ANN structure in predicting the compressive strength of foamed concrete. *Materials*, 13(5), 1072.
- Deifalla, A., & Salem, N. M. (2022). A machine learning model for torsion strength of externally bonded FRP-reinforced concrete beams. *Polymers*, 14(9), 1824.
- de-Prado-Gil, J., C. Palencia, P. Jagadesh & Martínez-García R. (2022). A comparison of machine learning tools that model the splitting tensile strength of self-compacting recycled aggregate concrete. *Materials* 15(12): 4164.
- Di Filippo, J., Karpman, J., & DeShazo, J. (2019). The impacts of policies to reduce CO2 emissions within the concrete supply chain. *Cement and Concrete Composites*, 101, 67–82.
- Dorogush, A. V., V. Ershov and A. Gulin (2018). "CatBoost: gradient boosting with categorical features support." arXiv preprint [arXiv:1810.11363](https://arxiv.org/abs/1810.11363).
- Ebid, A., & Deifalla, A. (2022). Using artificial intelligence techniques to predict punching shear capacity of lightweight concrete slabs. *Materials*, 15(8), 2732.
- Ebid, A. M., & Deifalla, A. (2021). Prediction of shear strength of FRP reinforced beams with and without stirrups using (GP) technique. *Ain Shams Engineering Journal*, 12(3), 2493–2510.
- Ebid, A. M., Deifalla, A. F., & Mahdi, H. A. (2022). Evaluating shear strength of light-weight and normal-weight concretes through artificial intelligence. *Sustainability*, 14(21), 14010.
- Erdal, H. I. (2013). Two-level and hybrid ensembles of decision trees for high performance concrete compressive strength prediction. *Engineering Applications of Artificial Intelligence*, 26(7), 1689–1697.
- Erdal, H. I., Karakurt, O., & Namli, E. (2013). High performance concrete compressive strength forecasting using ensemble models based on discrete wavelet transform. *Engineering Applications of Artificial Intelligence*, 26(4), 1246–1254.
- Fakharian, P., Eidgahee, D. R., Akbari, M., Jahangir, H., & Taeb, A. A. (2023). *Compressive strength prediction of hollow concrete masonry blocks using artificial intelligence algorithms*. Elsevier.
- Farooq, F., Ahmed, W., Akbar, A., Aslam, F., & Alyousef, R. (2021). Predictive modeling for sustainable high-performance concrete from industrial wastes: A comparison and optimization of models using ensemble learners. *Journal of Cleaner Production*, 292, 126032.
- Feng, D.-C., Liu, Z.-T., Wang, X.-D., Chen, Y., Chang, J.-Q., Wei, D.-F., & Jiang, Z.-M. (2020). Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Construction and Building Materials*, 230, 117000.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29, 1189–1232.
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367–378.
- Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics*, 11(1), 86–92.
- Güçlüer, K., Özbeyaz, A., Göymen, S., & Günaydin, O. (2021). A comparative investigation using machine learning methods for concrete compressive strength estimation. *Materials Today Communications*, 27, 102278.
- Gupta, R., Kewalramani, M. A., & Goel, A. (2006). Prediction of concrete strength using neural-expert system. *Journal of Materials in Civil Engineering*, 18(3), 462–466.
- Hammad, N., El-Nemr, A., & Hasan, H.E.-D. (2021). The performance of fiber GGBS based alkali-activated concrete. *Journal of Building Engineering*, 42, 102464.
- Hashim, M., & Tantray, M. (2021). Developing and optimizing foam concrete using industrial waste materials. *Innovative Infrastructure Solutions*, 6(4), 1–10.
- Hashmi, A. F., Shariq, M., & Baqi, A. (2021). An investigation into age-dependent strength, elastic modulus and deflection of low calcium fly ash concrete for sustainable construction. *Construction and Building Materials*, 283, 122772.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.
- Hoang, N.-D. (2023). Compressive strength estimation of rice husk ash-blended concrete using deep neural network regression with an asymmetric loss function. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, 47(3), 1547–1565.
- Holm, S. (1979). "A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 65–70.
- Jang, Y., Ahn, Y., & Kim, H. Y. (2019). Estimating compressive strength of concrete using deep convolutional neural networks with digital microscope images. *Journal of Computing in Civil Engineering*, 33(3), 04019018.
- Kalooop, M. R., Kumar, D., Samui, P., Hu, J. W., & Kim, D. (2020). Compressive strength prediction of high-performance concrete using gradient tree boosting machine. *Construction and Building Materials*, 264, 120198.
- Kamath, M., Prashant, S., & Kumar, M. (2021). Micro-characterisation of alkali activated paste with fly ash-GGBS-metakaolin binder system with ambient setting characteristics. *Construction and Building Materials*, 277, 122323.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154.
- Lee, S., Nguyen, N. H., Karamanli, A., Lee, J., & Vo, T. P. (2023). Super learner machine-learning algorithms for compressive strength prediction of high performance concrete. *Structural Concrete*, 24(2), 2208–2228.
- Lichman, M. (2013). UCI machine learning repository, Irvine.
- Mahjoubi, S., Meng, W., & Bao, Y. (2022). Auto-tune learning framework for prediction of flowability, mechanical properties, and porosity of ultra-high-performance concrete (UHPC). *Applied Soft Computing*, 115, 108182.
- Mikulčić, H., Klemeš, J. J., Vujanović, M., Urbanec, K., & Duić, N. (2016). Reducing greenhouse gasses emissions by fostering the deployment of alternative raw materials and energy sources in the cleaner cement manufacturing process. *Journal of Cleaner Production*, 136, 119–132.
- Moreira, L., & Arrieta, F. (2019). Thermal and economic assessment of organic Rankine cycles for waste heat recovery in cement plants. *Renewable and Sustainable Energy Reviews*, 114, 109315.
- Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116(44), 22071–22080.
- Mustapha, I. B., Z. Abdulkareem, M. Abdulkareem and A. Ganiyu (2022). "Predictive Modeling of Physical and Mechanical Properties of Pervious Concrete using XGBoost." arXiv preprint [arXiv:2206.00003](https://arxiv.org/abs/2206.00003).
- Nguyen, H., Vu, T., Vo, T. P., & Thai, H.-T. (2021). Efficient machine learning models for prediction of concrete strengths. *Construction and Building Materials*, 266, 120950.
- Nguyen-Sy, T., Wakim, J., To, Q.-D., Vu, M.-N., Nguyen, T.-D., & Nguyen, T.-T. (2020). Predicting the compressive strength of concrete from its compositions and age using the extreme gradient boosting method. *Construction and Building Materials*, 260, 119757.
- Okashah, A. M., Abdulkareem, M., Ali, A. Z., Ayeronfe, F., & Majid, M. Z. (2020). Application of automobile used engine oils and silica fume to improve concrete properties for eco-friendly construction. *Environmental and Climate Technologies*, 24(1), 123–142.
- Oltean, M., & Grosan, C. (2003). A comparison of several linear genetic programming techniques. *Complex Systems*, 14(4), 285–314.
- O'Neil, C., & Schutt, R. (2013). *Doing data science: Straight talk from the frontline*. O'Reilly Media, Inc.
- Pandey, A., & Kumar, B. (2022). Utilization of agricultural and industrial waste as replacement of cement in pavement quality concrete: a review. *Environmental Science and Pollution Research*. <https://doi.org/10.1007/s11356-021-18189-5>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Ranjbar, I., Toufigh, V., & Boroushaki, M. (2022). A combination of deep learning and genetic algorithm for predicting the compressive strength of high-performance concrete. *Structural Concrete*, 23(4), 2405–2418.
- Rodríguez-Fdez, I., A. Canosa, M. Mucientes and A. Bugarín (2015). STAC: a web platform for the comparison of algorithms using statistical tests. 2015 IEEE international conference on fuzzy systems (FUZZ-IEEE), IEEE.
- Salami, B. A., Iqbal, M., Abdullaheem, A., Jalal, F. E., Alimi, W., Jamal, A., Tafsiroj-jaman, T., Liu, Y., & Bardhan, A. (2022). Estimating compressive strength of lightweight foamed concrete using neural, genetic and ensemble machine learning approaches. *Cement and Concrete Composites*, 133, 104721.
- Salami, B. A., Olayiwola, T., Oyehan, T. A., & Raji, I. A. (2021). Data-driven model for ternary-blend concrete compressive strength prediction using

- machine learning approach. *Construction and Building Materials*, 301, 124152.
- Salem, N. M., & Deifalla, A. (2022). Evaluation of the strength of slab-column connections with FRPs using machine learning algorithms. *Polymers*, 14(8), 1517.
- Shah, M. I., Javed, M. F., Aslam, F., & Alabduljabbar, H. (2022). Machine learning modeling integrating experimental analysis for predicting the properties of sugarcane bagasse ash concrete. *Construction and Building Materials*, 314, 125634.
- Shariati, M., Mafipour, M. S., Ghahremani, B., Azarhomayun, F., Ahmadi, M., Trung, N. T., & Shariati, A. (2020). A novel hybrid extreme learning machine–grey wolf optimizer (ELM-GWO) model to predict compressive strength of concrete with partial replacements for cement. *Engineering with Computers*. <https://doi.org/10.1007/s00366-020-01081-0>
- Sharmila, P., & Dhinakaran, G. (2016). Compressive strength, porosity and sorptivity of ultra fine slag based high strength concrete. *Construction and Building Materials*, 120, 48–53.
- Silva, P. F., Moita, G. F., & Arruda, V. F. (2020). Machine learning techniques to predict the compressive strength of concrete. *Revista Internacional De Métodos Numéricos Para Cálculo y Diseño En Ingeniería*. <https://doi.org/10.23967/j.rimni.2020.09.008>
- Song, H., Ahmad, A., Farooq, F., Ostrowski, K. A., Maślak, M., Czarnecki, S., & Aslam, F. (2021). Predicting the compressive strength of concrete with fly ash admixture using machine learning algorithms. *Construction and Building Materials*, 308, 125021.
- Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N., & Asadpour, M. (2020). Boosting methods for multi-class imbalanced data classification: An experimental review. *Journal of Big Data*, 7(1), 1–47.
- Ullah, H. S., Khushnood, R. A., Ahmad, J., & Farooq, F. (2022). Predictive modelling of sustainable lightweight foamed concrete using machine learning novel approach. *Journal of Building Engineering*, 56, 104746.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ismail B. Mustapha BSc, MSc, Research Assistant at the Universiti Teknologi Malaysia, Malaysia.

Muyideen Abdulkareem BSc, MSc, PhD, CEng, Assistant Professor of Civil Engineering at the UCSI University, Malaysia.

Taha M. Jassam BSc, MSc, Lecturer of Civil Engineering at the UCSI University, Malaysia.

Ali H. AlAteah BSc, MSc, PhD, Assistant Professor of Civil Engineering at the University of Hafr Al Batin, Saudi Arabia.

Khaled A. Alawi Al-Sodani BSc, MSc, PhD, Assistant Professor of Civil Engineering at the University of Hafr Al Batin, Saudi Arabia.

Mohammed M. H. Al-Tholaia BSc, MSc, PhD, Assistant Professor of Civil Engineering at the University of Hafr Al Batin, Saudi Arabia.

Hatem Nabus BSc, MSc, Research Assistant, at the Universiti Teknologi Malaysia, Malaysia.

Sophia C. Alih BSc, MSc, PhD, Associate Professor of Civil Engineering at the Universiti Teknologi Malaysia, Malaysia.

Zainab Abdulkareem BSc, MSc, Research Assistant at the University of Strathclyde and the University of Ilorin, Glasgow & Nigeria.

Abideen Ganiyu BSc, MSc, PhD, CEng, Assistant Professor of Civil Engineering at the University of Bahrain, Bahrain.