SCIENTIFIC PAPER
Published: December 30, 2024
https://doi.org/10.13167/2024.29.11

Predictive modelling of sustainable concrete compressive strength using advanced machine learning algorithms

Tejas Joshi[1], Pulkit Mathur[1], Parita Oza[1], Smita Agrawal[1] and Husen Narmawala[1]

[1] Nirma University, School of Engineering, Civil Engineering, 382481, Ahmedabad, India

Corresponding author:
Tejas Joshi
tejas.joshi@nirmauni.ac.in

Received:
June 16, 2024

Revised:
August 19, 2024

Accepted:
August 27, 2024

Published:
December 30, 2024

Citation:
Joshi, T. et al. Predictive modelling of sustainable concrete compressive strength using advanced machine learning algorithms. *Advances in Civil and Architectural Engineering,* 2024, 15 (29), pp. 168-192. https://doi.org/10.13167/2024.29.11

ADVANCES IN CIVIL AND ARCHITECTURAL ENGINEERING (ISSN 2975-3848)

Faculty of Civil Engineering and Architecture Osijek
Josip Juraj Strossmayer University of Osijek
Vladimira Preloga 3
31000 Osijek
CROATIA

Abstract:
Considerable efforts have been made to increase the compressive strength of concrete by incorporating industrial by-products such as recycled aggregates and manufactured sand as partial substitutes for natural materials. However, predicting the compressive strength of concrete remains a challenge due to the influence of various factors, such as the type and proportion of aggregates, the water-cement ratio, and the age of the concrete. This research focuses on the development of machine learning (ML) models to predict concrete's compressive strength (CS) at 7 and 28 days. Fifteen input parameters—cement, natural and recycled fine and coarse aggregates, fly ash, manufactured Sand (M-Sand), water, admixture, w/c ratio, and age—were identified as critical factors influencing CS. A data set of 1030 samples from the literature was used, supplemented by additional experiments with recycled aggregates and manufactured sand. The models were trained on 70 % of the data, and the remaining 30% was used for testing. The results show that ML algorithms are highly effective in predicting CS, with the random forest algorithm achieving the highest accuracy (R^2 = 0,95; error = 3,74). In addition, a novel WebApp has been developed to leverage these models, allowing users to input parameters and quickly obtain CS predictions for concrete mix designs. The user-friendly interface of the WebApp makes it an easily accessible tool for professionals and researchers in concrete engineering. In this study, the potential of ML, in particular the random forest algorithm, is emphasised as a reliable and cost-effective method for predicting concrete CS, providing a valuable alternative to conventional experimental approaches.

Keywords:
machine learning; concrete compressive strength; random forest algorithm; regression analysis; web application

1 Introduction

Concrete, known for its exceptional compressive strength (CS), is a fundamental material in civil engineering and essential for the construction of various infrastructure projects. Its composite nature consisting of cement, ground granulated blast furnace slag (GGBS), fly ash (FA), aggregates, water, and admixtures makes it versatile and adaptable to the specific requirements of different projects [1-3]. Concrete reaches its maximum strength after a curing time of 28 d, making this time frame a standard benchmark for assessing its CS at different stages [4, 5]. Because of its adaptability, concrete is used for a variety of structures, each of which requires a specific grade to fulfil certain performance criteria. Conventionally, the determination of CS requires a labour-intensive process of producing, curing, and testing numerous concrete samples in a laboratory setting. However, with the advent of machine learning (ML) this process has been revolutionised. ML techniques using large data sets can autonomously identify the relationships between the constituent materials of concrete and their CS, enabling accurate predictions [6]. In addition to its construction, ML has proven its versatility in various domains, including medicine, email filtering, and computer vision. When applied to concrete strength prediction, ML not only improves prediction accuracy, but also has the potential to significantly reduce resource consumption [7]. To overcome the challenges associated with conventional methods, researchers have increasingly turned to ML techniques, which offer a more efficient and accurate approach for predicting the CS of concrete based on its constituent properties [1-3]. The application of ML to predict concrete properties has attracted considerable attention in recent years, and numerous studies have investigated its potential to streamline and improve construction practise. For example, Koya et al. [8] applied ML techniques to predict various concrete properties using the Wisconsin concrete mixture database. In their study, multiple ML models were used to estimate six mechanical properties of concrete. It was found that the support vector machine (SVM) model outperformed the other models, as indicated by higher coefficients of determination (R^2) and lower root mean squared error (RMSE) values. This study highlights the potential of ML to reduce unnecessary experimentation and optimise concrete mix designs.

Similarly, Ahmad et al. [9] investigated the use of ML modelling to predict the compressive strength of concrete exposed to high temperatures. They analysed data from 207 samples and used bagging, gradient boosting, decision trees, and artificial neural networks (ANN) to develop prediction models. Their results showed a strong correlation between the predicted and actual compressive strengths, with the ensemble methods performing better, especially under high-temperature conditions. This underscores the robustness of the ML models in adapting to varying environmental conditions and material properties. El-Gamal et al. [10] investigated the behaviour of near-surface-mounted reinforced concrete beams strengthened with glass- and carbon-fibre-reinforced polymers. In the study, a significant increase in the load-carrying capacity of all strengthened beams was observed compared to the reference beams, albeit with reduced ductility. To further advance the field, Kumar et al. [11] used ensemble ML techniques and deep neural networks to improve the accuracy of CS prediction. They investigated various algorithms, such as gradient boosting and random forest, which significantly outperformed conventional methods in terms of prediction accuracy. They also emphasised the importance of optimising the model training procedures to improve the generalisation capability of the ML models for different concrete types and mix designs. Jahangir et al. [12] introduced an innovative signal processing technique, the contourlet transform, to assess the damage severity in prestressed concrete slabs. Their study concluded that this technique can accurately identify the damage locations and severity, providing valuable insights for structural assessments. In another study, Fakharian et al. [13] applied artificial intelligence (AI) techniques, including neural networks and gene expression programming, to estimate the CS of hollow concrete block masonry prisms. Their results showed that ANNs provided the most accurate predictions, outperforming other models based on error metrics such as RMSE and mean absolute error (MAE). This paper further confirms the effectiveness of AI and ML in predicting the mechanical properties of concrete, even in

specialised applications such as masonry prisms. Chen et al. [14] applied multiple AI algorithms to predict the compressive strength of circular steel tubes filled with recycled aggregate concrete and achieved high accuracy with all models, with ANNs providing the highest level of accuracy. Bansal et al. [15] conducted a comparative analysis of ML algorithms for predicting the compressive strength of recycled aggregate concrete and demonstrated the effectiveness of ML models in accurately predicting the concrete strength. These studies collectively highlight the growing potential of ML and AI to revolutionise the prediction of concrete properties in a variety of applications.

Song et al. [16] investigated the prediction of concrete CS using machine learning (ML) techniques with a particular focus on the inclusion of FA as an admixture. In their study, an ANN model was used that considers inputs such as the cement binder, FA, coarse and fine aggregates, water, superplasticiser, and curing age. The model was validated using k-fold cross-validation and RMSE, which showed a strong correlation between the predicted and actual CS values. They concluded that ML models effectively capture the complex interactions between the different concrete components and leads to a more accurate strength predictions. The versatility of ML in civil engineering extends beyond just predicting the CS. Asteris et al. [17] investigated the use of ML to optimise the material composition of cement mortar and demonstrated the potential of these techniques to improve construction quality. Their study showcased how ML algorithms can determine optimal mix proportions and thus improve the overall performance of concrete structures. Similarly, Güçlüer et al. [18] conducted a comparative analysis of different ML models for predicting CS and highlighted the effectiveness of these techniques in optimising mix designs and ensuring the structural integrity of civil engineering projects. In addition to these studies, Pan et al. [19] conducted a comprehensive analysis of ML algorithms for predicting the CS of recycled aggregate concrete. Their research demonstrated the ability of ML models to accurately predict the strength of concrete incorporating recycled materials, thereby promoting sustainability in construction practise. This aligns with the broader trend of using waste materials such as FA and slag in concrete production to reduce the carbon footprint of construction activities [2-4]. Yaseen et al. [20] proposed an improved extreme learning machine (ELM) algorithm for more accurate prediction of concrete properties. Their study highlighted the advantages of the ELM over conventional neural network models, especially in terms of computational efficiency and prediction accuracy. This paper contributes to the ongoing development of advanced ML algorithms tailored to the specific needs of the concrete industry. Torre et. al. [21] also used ANN models to predict the CS of high-performance concrete (HPC) and emphasised the role of ML in improving the durability and sustainability of concrete structures. Their study found that ANN models can accurately predict the CS of HPC even when incorporating supplementary cementitious materials such as FA and slag, which are known to affect the long-term performance of concrete. Building on these advancements, recent studies have expanded the application of ML to CS prediction with innovative approaches. For example, Yu et al. [22] developed a model using the cat swarm optimisation (CSO) algorithm to predict the CS of HPC based on inputs such as aggregate type, cement content, and curing age. This approach significantly reduces the excessive number of trial mixes conventionally required, thereby saving time and resources. Similarly, Feng et al. [23] leveraged an adaptive boosting (AdaBoost) algorithm to construct a robust model from a large data set. By combining weak learners into a stronger predictor, the AdaBoost model effectively correlated input parameters such as aggregate composition and curing time with the CS of different concrete mixes. This method proved to be highly reliable in meeting the targeted strength requirements, further supporting the role of ML in concrete design. Complementing these advances, Tenza-Abril et al. [24] used ultrasonic pulse velocity together with an ANN to predict the CS of segregated lightweight concrete (SLC). The ANN model, trained on an extensive data set, provided rapid and accurate estimates of CS, proving its usefulness in assessing the quality and structural integrity of lightweight aggregate concrete (LWAC).

An overview of the different ML algorithms that have been used in the literature over the last five years is listed in Table 1 [25]. These studies collectively illustrate the growing influence of

ML in revolutionising concrete design and analysis by providing a more efficient and accurate means of predicting concrete properties for a wide range of applications.

Table 1. Literature survey of ML models adopted for predicting CS

ML Algorithm Name	Data set	Year	Ref
Random Forest	131	2019	[26]
Intelligent rule-based enhanced multiclass support vector machine and fuzzy rules	114	2019	[27]
Adaptive neuro-fuzzy inference system	7	2020	[28]
Multivariate	21	2020	[29]
Genetic expression programming	357	2020	[30]
Data envelopment Analysis	114	2021	[31]
Genetic expression programming, Decision Tree, Bagging	270	2021	[32]
Extreme Gradient Boosting (XGBoost), Random Forest, and Support Vector Machine (SVR)	1030	2022	[33]
Firefly Algorithm (FA) and Random Forest (RF)	225	2022	[34]
Genetic expression programming (GEP) and Artificial Neural Network (ANN)	310	2022	[35]
Ensemble and non-ensemble supervised ML approach	471	2022	[16]
Gradient boosting regression tree (GBRT)	1030	2023	[36]

The integration of advanced ML techniques into CS prediction significantly improves accuracy and efficiency and supports the wider use of sustainable materials in concrete production. This advancement not only streamlines concrete testing processes, but also aligns with efforts to reduce the environmental impact of concrete by optimising its strength and durability. The studies reviewed have demonstrated the potential of ML models to revolutionise civil engineering by improving the precision and sustainability of concrete production. The application of ML in civil engineering is expected to drive further innovations in the design, construction, and maintenance of concrete structures as it continues to evolve. This study focused on predicting the compressive strength of concrete using various machine learning algorithms and fifteen input parameters. Although the study demonstrates the efficacy of several ML techniques, including linear regression, decision tree regressor, and random forest regressor, there are potential research gaps. These gaps include examining additional ML algorithms and advanced techniques, considering additional input parameters, evaluating the quality of experimental data, optimising model training procedures, and validating model generalisations across diverse scenarios. This paper examines the evolving landscape of ML techniques in civil engineering, in particular their application in concrete strength prediction. Fig. 1 illustrates the range of ML algorithms used for this purpose, to provide a deeper understanding of material properties and improve prediction accuracy.

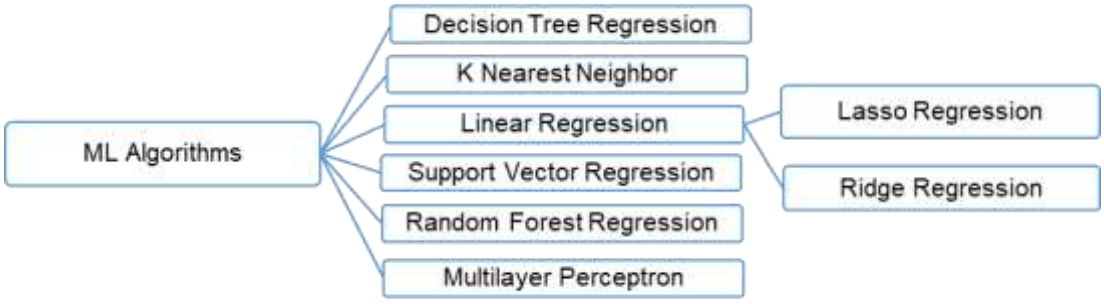


Figure 1. ML algorithms used to predict the compressive strength of concrete

This study examines the role of machine learning in estimating the compressive strength of concrete, focusing on recent progress in enhancing prediction accuracy and minimizing the need for extensive experimental procedures. By studying how concrete and supplementary materials interact to influence its strength, this approach provides solutions to the limitations of traditional testing. The following sections discuss the materials and methods applied, the key results, and the broader impact of using machine learning to advance concrete technology. Altogether, this research underscores the potential for machine learning to improve concrete mix designs and foster more sustainable building practices

2 Research objective

The main objective of this study was to assess the effectiveness of various ML techniques in predicting CS of concrete, focusing on the complex relationships between specific input attributes and the output. To optimise the learning process, two different training-to-testing ratios are proposed and applied to a concrete data set from the UCI repository. Furthermore, in this study, concrete was produced from different materials and different proportions of recycled aggregates, so that the effects of these materials on concrete strength could be analysed. This approach emphasises the sustainability of construction practises. The aim of this study was to identify the most accurate and reliable methods for predicting concrete CS by using both non-ensemble and ensemble ML models. In addition, sensitivity and parametric analyses were performed in this study to understand the relationships between the input parameters, offering a faster and more cost-effective alternative to the conventional experimental methods for predicting CS. Table 2 provides a comprehensive overview of the attribute information in the data set, which consists of eight input variables and one output variable.

Table 2. Brief description of the data set

Name	Data type	Measurement	Description
Cement (C)	Quantitative	kg/m^3 mixture	Input Variable
Blast Furnace Slag (GGBS)			
Fly Ash (FA)			
Water (W)			
Superplasticizer (S)			
Coarse Aggregate (CA)			
Fine Aggregate (FA)			
Age		Day	
Concrete Compressive Strength (CCS)		MPa	Output Variable

3 Background on ML algorithms

The application of AI techniques to predict the CS of concrete offers a promising approach for to address the inherent uncertainties and unpredictability of concrete constituents and their effects. By leveraging AI and ML algorithms, models can learn and adapt to complex fluctuations resulting from variations in material properties. This enables accurate predictions and a deeper understanding of the relationships between the input attributes and CS outcomes. The comparative analysis of multiple ML techniques presented in this study provides valuable insights into their effectiveness in predicting concrete strength and facilitates informed decision-making in construction projects. The proposed learning schemes optimised the training-to-test ratios, thereby improving the learning process for a concrete data set from the UCI repository [37]. This optimisation contributed to improved model performance and reliability. This study also considers the production of concrete using various materials, including different proportions of recycled aggregates, thus emphasising sustainability. By examining the effects of these materials on concrete strength, this study expands the

knowledge base of construction companies and promotes the adoption of innovative technologies in the concrete industry. Over the past few decades, ML models have played a critical role in predicting and categorising various attributes by leveraging knowledge gained from training data. With these techniques, accurate models can be developed without understanding the underlying mechanisms of the processes involved. In this section, various ML algorithms used to predict the CS of concrete using a data set from existing sources are examined.

3.1 Linear regression

The linear regression (LR) model is a fundamental machine learning technique used to predict a single output variable (y) based on one or more input variables (x). It is based on the assumption that there is a linear relationship between the input variables and output. In this model, a linear equation was created by fitting the coefficients $w = w_1, w_2, w_3, ..., w_n$ to the input variables: the goal was to minimise the sum of the squared differences (residuals) between the observed and predicted values, resulting in an optimal linear model. The general form of the linear regression equation is as follows:

$$y = wx + b \quad (1)$$

where y is the predicted output, w is the coefficient of the input variable, and b is the bias term. This bias allows the regression line to shift, similar to how a constant functions in a linear equation. Remarkably, linear regression does not require an activation function as the linearity of the model is sufficient to establish the relationship between the inputs and outputs. This simple but powerful method forms the basis for more complex machine learning algorithms, making it an important part of the toolkit of data scientists and engineers. By understanding the linear relationships in a data set, linear regression models can provide valuable insights and serve as a basis for advanced predictive modelling techniques. There are two types of LR, which are described below.

3.1.1 Lasso Regression L1

'LASSO' stands for 'Least Absolute Shrinkage and Selection Operator' and is a regularisation technique that is frequently used in machine learning to improve the performance of predictive models. It is particularly effective in models with many parameters or where multicollinearity is present among the input variables. The LASSO procedure applies a penalty to the absolute values of the coefficients in the model, effectively shrinking some of them towards zero. This process not only reduces model complexity, but also aids in feature selection by identifying the most important variables. As a result, LASSO is particularly useful in scenarios where a simpler and more interpretable model is desired without compromising prediction accuracy. The strength of LASSO lies in its ability to cope with the overfitting that is common in models with numerous parameters. By bringing the coefficient values closer to the mean and possibly setting some to zero, LASSO helps create a more robust model that generalises well to unseen data. This makes it an invaluable tool for refining models and ensuring that they perform optimally, especially when handling data sets that exhibit multicollinearity or when the number of predictors exceeds the number of observations.

3.1.2 Ridge Regression L2

Ridge regression is a regularisation technique that is often used to deal with multicollinearity in regression models. Multicollinearity occurs when two or more independent variables in a model are highly correlated, meaning that one independent variable can be linearly predicted by another. This correlation can cause instability in the regression model, and thus to large variances in the estimated coefficients, which, in turn, results in predictions that deviate significantly from the true target values. Ridge regression mitigates this problem by adding a penalty term to the loss function that is proportional to the square of the magnitude of the coefficients. This penalty helps shrink the coefficients, reduce their variance, and make the model more stable and reliable, especially in cases where multicollinearity is present. Unlike

LASSO, which can shrink some coefficients to exactly zero, ridge regression usually retains all variables, but reduces their effects, resulting in a more balanced and less overfitted model. This technique is particularly useful in scenarios where the model contains many correlated predictors. By controlling the complexity of the model, ridge regression improves its generalisability to new data, making it a powerful tool for predictive modelling in high-dimensional data sets.

3.2 Decision tree regression

Decision tree regression is a widely used supervised learning algorithm known for its ability to effectively handle both classification and regression tasks. While linear regression is often considered one of the simplest and most practical models, decision trees offer a more nuanced approach that is particularly useful for decision-based problems. A decision tree consists of three types of nodes: root, interior, and leaf. The root node represents the entire data set and serves as a starting point for splitting the data into more specific groups. Interior nodes capture the characteristics of the data set as tree branches according to the decision criteria. These branches guide the data points through a series of binary questions and lead them to leaf nodes that represent outcomes or predictions.

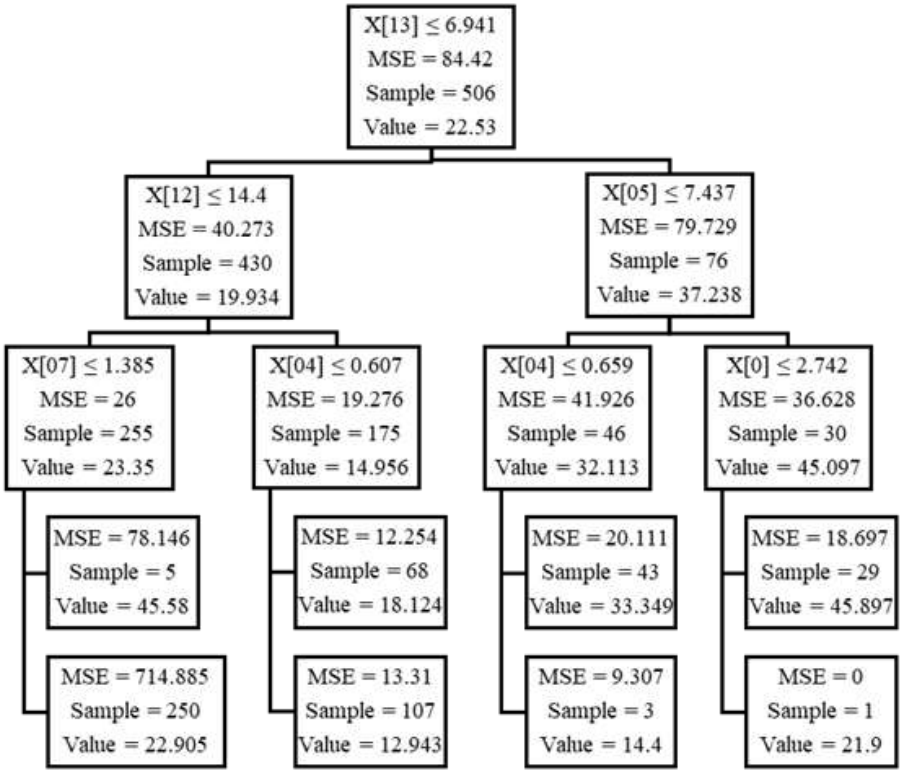


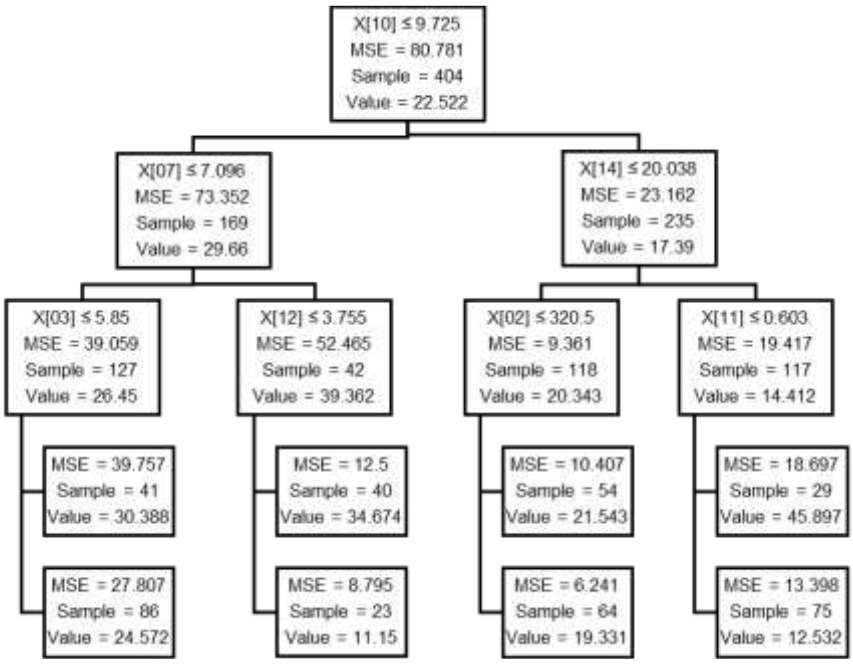
Figure 2. Decision tree algorithm / node hierarchy for the decision tree

The predicted value for each data point is the average value of the dependent variable within the corresponding leaf node. Decision trees are particularly advantageous owing to their simplicity and interpretability. They require only minimal preprocessing of the data and are resilient to non-linear relationships within the data. In addition, decision trees are not dependent on hyperparameters, so they are easy to implement without extensive tuning. However, decision trees can be prone to overfitting, where the model becomes too closely fitted to the training data, capturing noise, and decreasing performance on new and unseen data. This issue is often addressed using a random forest regressor. This is an ensemble technique that builds multiple decision trees and combines their predictions to improve generalisation and reduce overfitting. The robustness of random forests results from the diversity of trees in the ensemble, each trained on different subsets of data and features, which

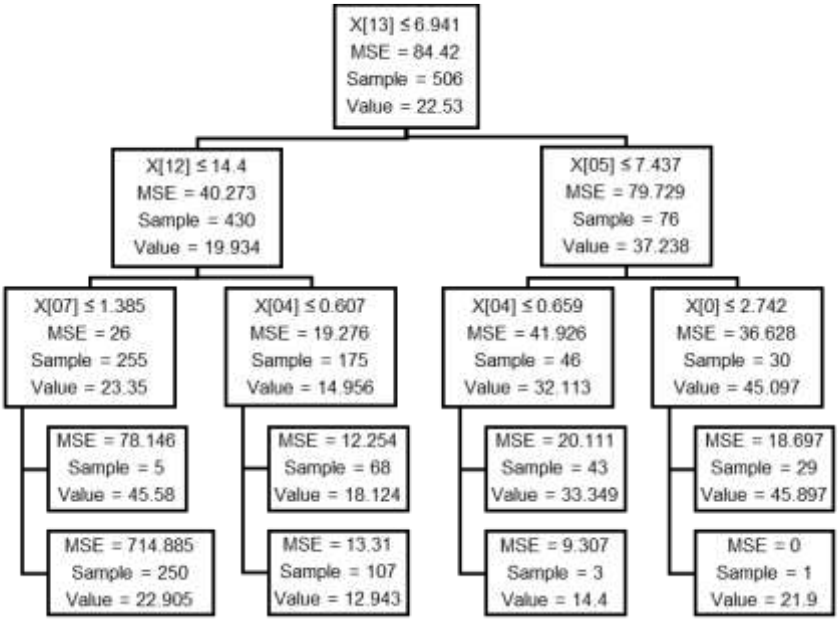
helps mitigate the risk of overfitting. When constructing a decision tree, a hierarchy of nodes based on training data is created. As the tree is built, the data at each node are minimised. The performance of a decision tree is influenced by various factors including the number of splits, maximum number of trees, features selected, and number of nodes. This methodology makes decision trees powerful machine learning tools, especially for problems where the relationships between variables are complex and non-linear. Their ability to handle different types of data and their interpretability make them a popular choice for various applications. The node hierarchy for the decision tree constructed using the sample data is shown in Fig. 2. In this process, a few representative rows and feature samples are selected and added to the decision tree. The data set was then used to train the tree, and training continued until the error was minimised, leading to a final prediction. The performance of a decision tree is influenced by several factors, including the number of splits, maximum number of trees, features selected, and number of nodes.

3.3 Random forest regression

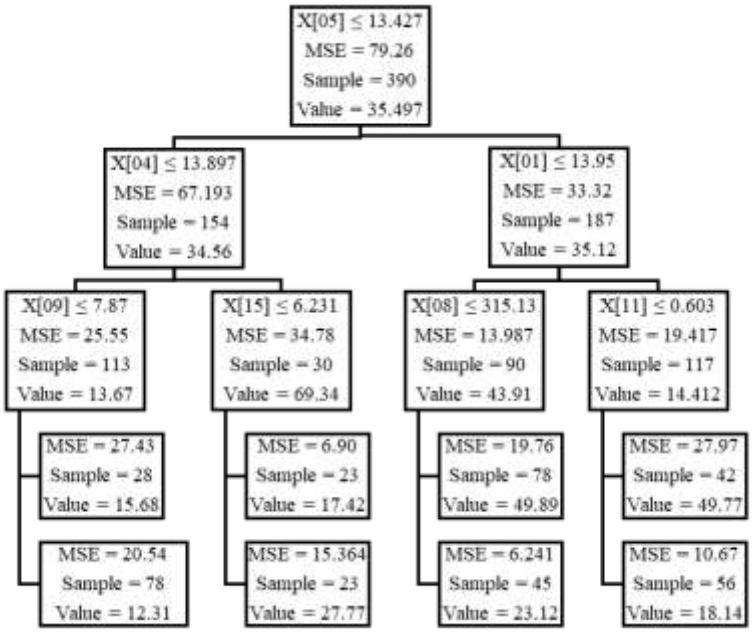
Random forest is an advanced ensemble-learning method that extends the concept of bagged decision trees. This approach uses a collection of decision trees, each trained on a different subset of features, to improve the diversity and robustness of the model. A key feature of random forest is the use of bootstrapping, a sampling technique in which each decision tree is trained on a random selection of data points with replacement. This process, known as bootstrap aggregation or bagging, involves the creation of multiple decision trees using different bootstrapped subsets of training data. The final prediction of the random forest model was determined obtained by averaging the predictions of all the individual trees. This ensemble approach helps reduce the variance that a single decision tree may exhibit, which can often lead to overfitting. Random forest also incorporates a unique feature selection mechanism during training. Each time a tree is split, only a random subset of features is considered rather than evaluating all available features. This method reduces the correlation between the individual trees in the forest and helps prevent overfitting. Specifically, if n is the total number of features, only sqrt{n} features are randomly selected for each split.



(a) Decision tree 1



(b) Decision tree 2



(c) Decision tree 3: Average of all predictions = final prediction

Figure 3. Group of decision trees formed using a different set of input features

This randomness in the feature selection ensures that the model examines a variety of predictors and does not over-rely on a small number of strong predictors. The random forest algorithm is particularly well-suited for handling large data sets with many dimensions. It combines the outputs of multiple decision trees to produce a single robust prediction, making it both efficient and resilient. Fig. 3 illustrates a random forest model created from the input variables and shows how a few representative rows and feature samples were used to train individual trees. Unlike a single decision tree, which may overfit the data by growing too deep, the random forest approach averages the predictions from multiple trees, thereby improving

accuracy and reducing errors. Overall, random forests improve the performance of decision trees by introducing randomness into feature selection and combining multiple trees to create a more stable and accurate predictive model.

3.4 Support vector machine

SVM are prominent machine learning techniques used for both classification and regression tasks. When applied to regression problems, SVM takes the form of support vector regression (SVR). SVR operates by identifying a hyperplane, which in the case of regression, is a straight line that best fits the data. The goal of SVR is to find a hyperplane that keeps the predicted values within a certain threshold of the actual values, rather than directly minimising the error between the predictions and true values. This threshold is defined by a margin that determines how close the data points, known as support vectors, are to the hyperplane. The key idea behind SVR is to find the hyperplane with the maximum margin while keeping the prediction error within a predefined range. The support vectors are the data points closest to the hyperplane on either side. They are crucial for determining the position of the hyperplane. The SVR uses kernel functions to solve complex regression problems with non-linear relationships and high-dimensional data. Kernels transform data into a higher-dimensional space where linear separation is more feasible, thus simplifying the regression task. Common kernel functions include polynomials, radial basis functions (RBF), and sigmoid kernels, each of which provide different methods for mapping the data into the desired space. Overall, SVR is valued for its ability to handle complex and non-linear relationships by leveraging kernel functions and focusing on fitting the hyperplane within a specified margin, making it a powerful tool in machine learning for regression problems.

3.5 Multi-layer perceptron (MLP)

Linear models are often inadequate when it comes to learning complex non-linear relationships between input data and target variables. To address this limitation, MLPs and neural networks offer more sophisticated approaches. An MLP consists of three main types of layers: input, hidden, and output. Each layer contains a specific number of nodes or neurones, with non-linear activation functions applied to the neurones in all layers except the input layer. The architecture of an MLP enables the modelling of complex patterns in data that cannot be captured using linear models. The input layer processes the incoming data, whereas the output layer produces the final predictions or classifications. The hidden layers situated between the input and output layers serve as the core computational components of the network. These hidden layers enable the MLP to learn and model intricate relationships in the data through non-linear activation functions. Training an MLP involves a process called backpropagation, a supervised learning technique. Backpropagation adjusts the weights of the connections between the neurones to minimise the error between the predicted and actual outputs. This iterative learning process ensures that the performance of the network improves over time. Unlike a linear perceptron, which consists of a single layer, an MLP features multiple layers and non-linear activation functions. This structure allows MLPs to discriminate between data points that are not linearly separable, making them highly effective in a variety of complex tasks. Fig. 4 illustrates the typical architecture of an MLP system. One of the challenges in designing an MLP is determining the optimal number of neurones for the hidden layers. Too many neurones can lead to overfitting, wherein the model excessively learns the training data and fails to generalise it to new data. Conversely, too few neurones may result in underfitting, where the model cannot capture the underlying relationships in the data. A common heuristic for selecting the number of neurones is to multiply the number of input variables by a factor such as three; however, this approach may need to be adjusted depending on the specific problem and data set. Overall, MLPs are powerful tools for modelling complex data as they are able to approximate any continuous function. This makes them suitable for problems that linear models cannot solve effectively.

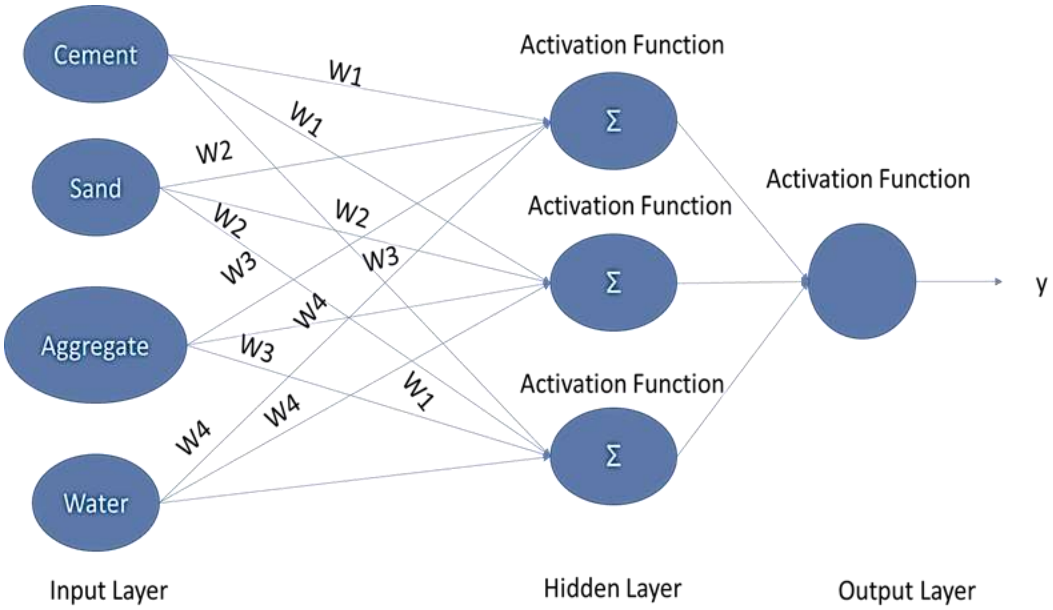


Figure 4. MLP architecture

3.6 K-nearest neighbour

The k-nearest neighbour (KNN) algorithm is a non-parametric technique that is widely used for both classification and regression tasks. The fundamental premise relies on the proximity of data points. During classification, KNN assigns a new data point to the class that is most prevalent among its k-nearest neighbours. The value of k determines the quantity of nearby points to be taken into account. A class with many nearest neighbours determines the class label for a new data point. However, this approach can be affected by outliers or imbalances in the data set, which may skew the classification results. For regression problems, the KNN predicts the value of a new data point by averaging the values of its k-nearest neighbours. This method helps to make predictions based on similar data points, thus providing a smooth approximation of the output values. Despite its simplicity, the KNN is highly effective. However, the parameter k and the distance metric used to measure the proximity of data points must be carefully considered is. In addition, KNN can be computationally intensive, especially for large data sets, as it requires calculating the distances between a new data point and all existing data points in the training set.

4 Experimental program

The existing literature predominantly focuses on ordinary concrete. To fill this gap, various mix design calculations were conducted using recycled aggregates and manufactured sand to produce more sustainable concrete. This approach was intended to expand the data set and eventually lead to the production of concrete samples containing these innovative materials. The materials used during the experimental program are summarised in Fig. 5.

Firstly, a series of experiments were conducted to determine the physical properties of the materials, including specific gravity (SG), water absorption (WA), and moisture content (MS). These properties are essential for the formulation of an accurate mix design. Subsequently, the mixes with different proportions of recycled aggregates and supplementary cementitious materials (SCMs) were calculated as per the guidelines of IS 10262:2019 [38]. The concrete was cast into 150 ×150 × 150 mm cubes and cured in water for 7 and 28 d. The compressive strength (CS) was then tested using a compression testing machine (CTM). Detailed information on the mix designs for the different percentage replacements is listed in Tables 3–6. Table 3 lists the mix design parameters and CS) values for the different concrete formulations. The table includes details on the quantities of cement, sand, coarse aggregates

(CA), recycled coarse aggregates (RCA), water, and water-to-cement (w/c) ratios, along with the CS measurements at 7 and 28 d of curing. The data illustrates the effects of different aggregate types and proportions on the CS of the concrete.

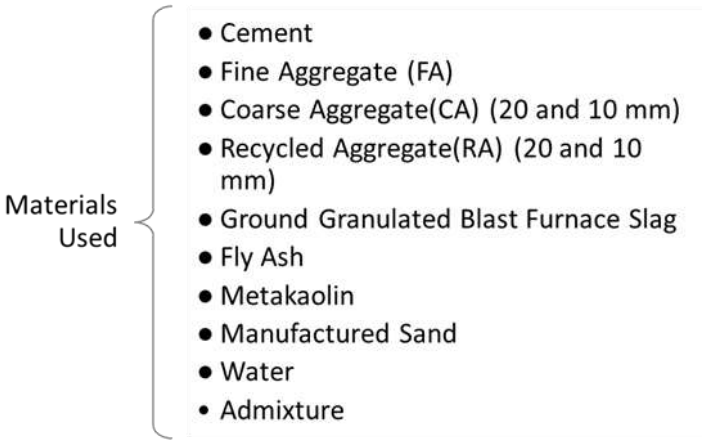


Figure 5. Materials used

Table 3. Mix proportions for concrete using recycled aggregates

Sr. No.	Cement	Sand	CA_20	CA_10	RCA_20	RCA_10	Water	w/c ratio	Age (day)	CS
1	400	578,82	711,46	474,31	0,00	0,00	182,0	0,450	7	24,75
									28	35,80
2	400	578,82	640,32	426,88	46,73	30,60	184,8	0,460	7	25,00
									28	35,00
3	400	578,82	569,17	379,44	83,09	54,41	187,6	0,470	7	27,40
									28	30,70
4	400	578,82	498,02	332,01	155,79	102,01	190,4	0,476	7	28,40
									28	32,40
5	400	578,82	426,88	284,58	207,72	136,02	193,2	0,483	7	26,90
									28	33,10
6	400	578,82	355,73	237,15	259,65	170,03	196,0	0,490	7	26,37
									28	32,45

Table 4 lists the mix design parameters and compressive strength (CS) values for concrete incorporating recycled fine aggregates (RFA). The table shows the amounts of cement, sand, RFA, CA, water, and w/c ratios used, along with the CS measurements at 7 and 28 d. The data highlights the influence of varying RFA proportions on the compressive strength of the concrete.

Table 5 lists the mix design parameters and CS for concrete with manufactured sand. These include the quantities of cement, sand, manufactured sand, CA, water, and water-to-cement (w/c) ratio, along with CS measurements at 7 and 28 d. The data illustrates the effects of varying the proportion of manufactured sand on the compressive strength of the concrete.

Table 6 lists the results of mix design and CS for concrete with additional SCMs such as GGBS, Metakaolin (MK), and FA. The table contains various combinations of cement, aggregates, and SCM, along with the corresponding CS values after 7 and 28 d. It illustrates the effects of different SCMs on the compressive strength of the concrete, highlighting the effects of GGBS, MK, and FA on the final strength results.

Table 4. Mix proportions for concrete using recycled sand

Sr. No.	Sand	RFA	CA_20	CA_10	Water	w/c	Cement	Age (day)	CS
1	400	520,94	54,1	711,46	474,31	182,8	0,457	7	22,20
								28	33,70
2	400	463,06	109,0	711,46	474,31	183,2	0,458	7	23,00
								28	34,50
3	400	405,17	164,0	711,46	474,31	183,6	0,459	7	21,60
								28	33,40
4	400	347,3	218,0	711,46	474,31	184,0	0,460	7	24,20
								28	30,40
5	400	289,41	273,0	711,46	474,31	184,8	0,462	7	23,60
								28	32,33

Table 5. Mix proportions for concrete with manufactured sand

Sr. No.	Sand	Manufactured Sand	CA_20	CA_10	Water	w/c	Cement	Age (day)	CS
1	400	520,94	56,57	711,46	474,31	182,9	0,457	7	22,20
								28	33,70
2	400	463,06	113,14	711,46	474,31	183,3	0,458	7	23,00
								28	34,50
3	400	405,17	169,71	711,46	474,31	183,7	0,459	7	21,60
								28	33,40
4	400	347,3	226,28	711,46	474,31	184,6	0,460	7	24,20
								28	30,40
5	400	289,41	282,86	711,46	474,31	185,4	0,462	7	23,60
								28	32,33

Table 6. Mix proportions for concrete with recycled aggregates and SCMs

Sr. No.	Cement	Sand	CA_20	CA_10	RCA_20	RCA_10	GGBS	MK	FA	Water	w/c	Age (day)	CS
1	362,60	578,82	711,5	474,31	46,73	30,06	155,4	0,0	0,0	175,13	0,483	7	26,4
												28	36,6
2	387,37	578,82	711,5	474,31	46,73	30,06	0,0	77,7	0,0	187,16	0,465	7	26,1
												28	33,0
3	395,37	578,82	711,5	474,31	46,73	30,06	0,0	0,0	51,8	180,68	0,430	7	28,3
												28	36,7

5 Methodology, interpretation, correlation results, and method performance measurements

5.1 Methodology

The data set used in this study was compiled from fresh experimental data generated by the author in the laboratory, the UCI repository [37], and other relevant research literature. The data set comprises 1746 records, each containing 15 input features and one output label, which is the CS. The features and their respective ranges are listed in Table 6. Before applying the data set to different models, it is important to establish a correlation between the different components to improve the understanding and performance of the model. The minimum and

maximum values of the input parameters are listed in Table 7. This table highlights the diversity and scope of data used in the analysis and facilitates a comprehensive understanding of the relationships between the different concrete components and their influence on CS. Identifying these relationships is crucial for optimising mix designs and improving the predictive accuracy of the model.

Table 7. Minimum and maximum values for all parameters

Sr No.	Measurement	Range of values
Input parameters		Range
Cement content (C)	kg/m^3	102-540
Natural fine aggregate (NFA)		0-992,6
Recycled fine aggregate (RFA)		0-800
Natural coarse aggregate (NCA) (20 mm)		0-1295
Natural coarse aggregate (NCA) (10 mm)		0-516,45
Recycled coarse aggregate (RCA) (20 mm)		0-1149
Recycled coarse aggregate (RCA) (10 mm)		0-503,6
Ground Granulated Blast Furnace Slag		0-359,4
Metakaolin		0-84
Fly Ash		0-215
Manufactured Sand		0-725,5
Water		121,8-271
Admixture	lit/m^3	0-126,18
W/C ratio	-	0,26-1,88
Age	days	1-365
Output parameters		Range
Compressive Strength	N/mm^2	2,91-115,20

5.2 Interpretation and correlation of the results

As shown in Fig. 6, CS shows the weakest correlation with the water content and water/cement (w/c) ratio, whereas it shows the strongest correlation with age. This result aligns with our expectations. In addition, there is a notably strong correlation between cement content, GGBS, and CS, which was another expected result. The relationship between the use of superplasticiser and CS is weakly positive, whereas the relationship with water content is weakly negative. Superplasticisers allow for a reduction in water content without affecting the workability of fresh concrete, which in turn increases the CS and improves the cement-to-water ratio.

An inverse relationship between cement and GGBS was observed as the addition of additional cementitious materials reduced the cement content. Recycled aggregates exhibited a strong correlation with water content owing to their high absorption capacity, a trend that was similarly observed, albeit to a lesser extent, for recycled sand. To ensure that the model interprets the data correctly, feature scaling is required as the values of the different attributes span a wide range, making it difficult for the model to treat all features on the same scale. During normalisation, the attribute values are rescaled and usually brought into the range between zero and one. This is essential for the model to learn effectively from the data. A machine learning model learns in two phases: training and testing. During the training phase, the model identifies the relationships between different attributes and the influence of the input features on the output. In the testing phase, the ability of the model to apply the learnt understanding to a new data set is evaluated. In this study, a training-to-testing ratio of 70:30 was chosen as a larger training data set generally leads to better model performance.

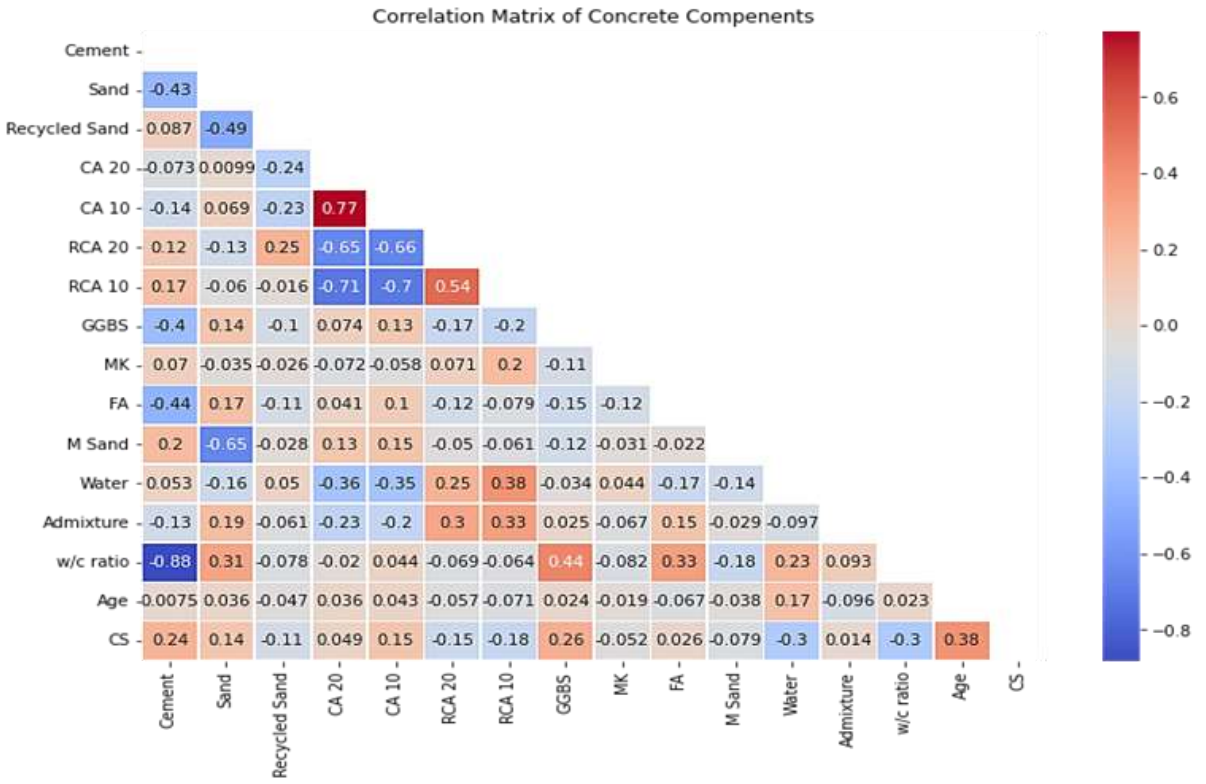


Figure 6. Correlation matrix of the concrete components

5.3 Method performance measurements

The accuracy of the model was evaluated using key statistical metrics, specifically the MAE and the R^2 score, as defined by Equations 2 and 3. The MAE measures the average magnitude of error in a set of predictions and gives a clear indication of how close the predicted values are to the actual values. The R^2 score quantifies the proportion of variance in the dependent variable that is predictable from the independent variables and thus provides information about the overall fit of the model. Using these metrics, the performance of the model was rigorously evaluated to ensure that it not only minimised error but also accurately captured the underlying relationships within the data set. This approach allows for a more robust interpretation of the model's predictive capabilities and ensures that it generalises well to new data.

$$Mean\ absolute\ error\ (MAE) = \frac{1}{N}\sum_{i=1}^{n}|EXP_i - PRE_i| \tag{2}$$

$$Coefficient\ of\ determination\ (R^2) = \frac{\sum_{i=1}^{n}(EXPi - expi) \times (PREi - prei)}{\sum_{i=1}^{n}(EXPi - expi)^2 \times \sum_{i=1}^{n}(PREi - prei)^2} \tag{3}$$

where EXP_i is the experimental value of CS; PRE_i is the predicted value of CS; exp_i mean of the experimental value of CS; pre_i mean of the predicted value of CS.

6 Results and discussion

The performance of the different models in predicting the CS of concrete was assessed using two important statistical measures: MAE and R^2. These metrics were applied to both the training and test data sets to ensure a comprehensive assessment of the accuracy and generalisation capabilities of each model.

6.1 Training data set performance

Table 8 lists the MAE and R^2 values for all models when applied to the training data set. The decision tree regressor (DTR) showed the highest accuracy during training, with a near-perfect R^2 score of 0,99 and a minimal MAE of 0,21 MPa. This indicates that the predicted values agree very well with the experimental measurements. However, the high performance of DTR may also indicate a potential overfitting that could affect its generalisation to new data. The random forest regressor (RFR) also performed exceptionally well, achieving an R^2 of 0,98 and an MAE of 1,69 MPa. A high R^2 value of 0,98 achieved by RFR indicates that approximately 98 % of the variance in the response variable can be explained by the predictors, highlighting the model's strong ability to capture patterns and correlations in the training data. The linear regression (LR) and ridge regression (RR) models produced almost identical results, with moderate R^2 values and MAE scores, reflecting their limited predictive abilities compared to more complex models.

Table 8. Results of the error and fit measures on the training data set

Sr. No.	Model	MAE (MPa)	R^2	Literature			
				[38]	[39]	[40]	[41]
1	LR	8,98	0,64	-	0,98	0,82	0,512
2	LAR	15,22	0,09	-	-	0,81	0,511
3	RR	9,09	0,64	-	-	0,82	0,511
4	DTR	0,21	0,99	0,69	1,0	0,91	0,928
5	RFR	1,69	0,98	0,75	0,99	0,94	0,926
6	SVM	9,27	0,63	0,73	-	-	-
7	MLP	8,31	0,67	-	-	0,80	-
8	KNN	5,76	0,85	0,78	0,91	-	0,944

When comparing the predictive performance of different models in the experimental work with those reported in the literature, several observations can be made. For the LR model, the experimental work yielded an MAE of 8,98 MPa and R^2 of 0,64. These results agree well with those in the literature, which also show a moderate R^2 range of 0,82-0,98 across different studies. The linear autoregressive (LAR) model in the experimental work showed a higher MAE of 15,22 MPa with a low R^2 of 0,09, indicating poor predictive accuracy compared with the literature, where R^2 values as high as 0,81-0,98 are reported. For RR, the experimental results showed an MAE of 9,09 MPa and an R^2 of 0,64; which is consistent with the literature, where R^2 values typically range from 0,2-0,98. The DTR performed exceptionally well in the experimental work, with an MAE of only 0,21 MPa and an R^2 of 0,99. This performance is confirmed by the literature results, which also report high accuracy of the DTR models, with R^2 values close to 1.0 and MAE as low as 0.69 MPa. Similarly, the RFR in the experimental work achieved an MAE of 1.69 MPa and an R^2 of 0.98, closely matching the literature values of R^2 between 0,94 and 0,99. In contrast, the SVM model displayed lower accuracy in the experimental work, with an MAE of 9.27 MPa and R^2 of 0,63, which is comparable to the mixed results reported in the literature, where R^2 values range from 0,73-0,91. The MLP in the experimental work had an MAE of 8,31 MPa and an R^2 of 0,67, which is consistent with the variable performance of MLP models, with R^2 values of approximately 0.80. Finally, the KNN model showed a solid performance in the experimental work, with an MAE of 5,76 MPa and an R^2 of 0,85; which compares favourably with the R^2 values of up to 0,944 reported in the literature. Overall, the results of the experimental work are in general agreement with those from the literature, with some variations likely due to differences in data sets and modelling approaches.

6.2 Test data set performance

Table 9 lists the error and fit measures for all models on the test data set. As expected, the RFR outperformed the DTR in the testing phase, achieving an R^2 of 0,93 and an MAE of 3,74 MPa, thereby confirming its robustness and ability to generalise well to new data. The DTR, which performed perfectly on the training set, showed reduced accuracy on the test set with an R^2 of 0,64 and an MAE of 5,30 MPa, which is a further indication that it may have been overfitted during training. The high R^2 value (0,93 for the RFR on the test data indicates that the model can explain approximately 93% of the variance in the response variable in new instances, showcasing its predictive power and reliability for real-world applications. The KNN model also maintained a solid performance with an R^2 of 0,73 and an MAE of 7,89 MPa, demonstrating a good balance between accuracy and generalisation.

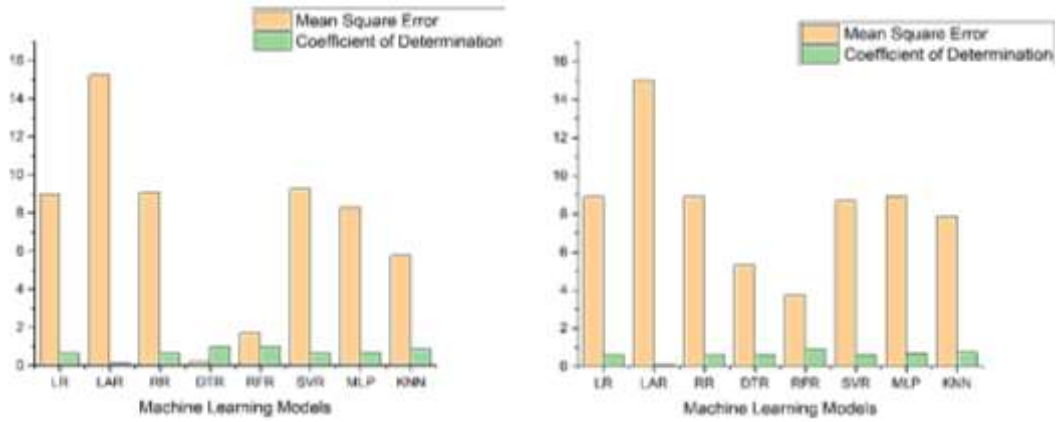
Table 9. Results of the error and fit measures on the test data set

Sr. No.	Model	MAE (MPa)	R^2	Literature	
				[42]	[43]
1	LR	8,89	0,62	0,86	0,562
2	LAR	15,00	0,08	-	0,562
3	RR	8,92	0,62	-	0,561
4	DTR	5,30	0,64	0,97	0,786
5	RFR	3,74	0,93	0,99	0,827
6	SVM	8,68	0,64	-	-
7	MLP	8,96	0,72	-	-
8	KNN	7,89	0,73	0,86	0,797

By comparing the predictive performance of the different models in the experimental work with those reported in the literature, several insights emerged. For the LR model, the experimental work achieved an MAE of 8,89 MPa and R^2 of 0,62. These results are somewhat consistent with the literature, which reported R^2 values of 0,86 but shows a slightly lower MAE of 0,562 MPa, indicating that the experimental work had a slightly lower predictive accuracy. The LAR model in the experimental work exhibited a higher MAE of 15,0 MPa with a very low R^2 of 0,08; indicating poor model performance. This is in contrast to the literature, where R^2 values were not reported, but a similar MAE value of 0,562 MPa was observed, implying that the application of the model or data set could be significantly different. For RR, the experimental results gave an MAE of 8,92 MPa and an R^2 of 0,62, which is consistent with the R^2 values in the literature of approximately 0,561, but there is a lack of literature data for direct comparison of MAE. The DTR showed an MAE of 5,30 MPa and an R^2 of 0,64 in the experimental work. This performance is modest compared to that in the literature, where an R^2 of 0,97 is reported with a significantly lower MAE of 0,786 MPa, suggesting that the model in the experimental work could benefit from further optimisation. The RFR in the experimental work performed well, with an MAE of 3,74 MPa and an R^2 of 0,93. These results are consistent with those from the literature, which show an R^2 of 0,99 and a similar MAE of 0,827 MPa, indicating strong predictive accuracy for this model. For the SVM model, the experimental work gave an MAE of 8,68 MPa and R^2 of 0,64. There is no direct comparison of these values in the literature, so it is challenging to assess the relative performance, although the experimental work suggests moderate accuracy. The MLP in the experimental study achieved an MAE of 8,96 MPa and R^2 of 0,72. Without a direct literature comparison, the performance of the model appears to be moderate. Finally, the KNN model in the experimental work had an MAE of 7,89 MPa and an R^2 of 0,73. These results agree fairly well with the literature, which gives an R^2 of 0.86 and a slightly higher MAE of 0,797 MPa, indicating reasonable agreement between the experimental work and the literature results. Overall, the results of the experimental work show general

agreement with those in the literature, with some discrepancies that could be attributed to differences in data sets, model configurations, or experimental conditions.

The variations of MSE and R^2 for different ML models are shown in Fig. 7. These plots provide compelling evidence that the RFR model outperforms the other models and has the highest R^2 values of 0,98 and 0,93 in the training and testing phases, respectively. The choice to evaluate ML models using MSE and R^2 is justified by their widely recognised importance in regression analysis. MSE quantifies the average squared difference between the predicted and actual values and thus provides insights into model accuracy, whereas R^2 measures the proportion of variance explained by the predictors and thus reflects the model's ability to capture the underlying data relationships.



(a) Training data set (b) Test data set

Figure 7. Variation of the mean square error and the coefficient of determination values

Based on the evidence provided by the plots and performance metrics, the RFR model proved to be the best choice among the ML models evaluated. Its exceptional R^2 values in both the training and testing phases confirmed its strong performance and emphasised its potential for successful application in real-world scenarios. The ability of the model to effectively learn from data, capture complex relationships, and generalise well to new instances makes it a promising choice for accurate predictions and reliable inferences in predicting concrete CS. A closer look at the scatterplots in Fig. 8 and Fig. 9, reveals that the data points that overlap the best-fit line are of considerable importance, especially in models other than the Lasso regression. This overlap suggests a strong correlation between the predicted and corresponding experimental values, indicating that these models have successfully learned the underlying patterns in the data and were able to generalise their predictions to new, unseen instances.

In contrast, the Lasso regression model is an exception. The absence of overlapping data points along the best-fit line in this model illustrates its limited ability to learn effectively from the data set. This shortcoming prevents the Lasso regression from establishing a strong correlation between the predicted and experimental values, which ultimately impairs its ability to accurately generalise to new data points. The visible lack of overlap in the Lasso regression clearly demonstrates its inadequate learning capabilities and limitations in making accurate predictions. These results emphasise the importance of carefully considering the characteristics of a data set when selecting an appropriate regression model. When the relationships between predictors and response variables are complex or non-linear, it is critical to explore alternative regression models that can better capture these patterns. Analysis of the scatter plots highlights the need for models that can effectively learn and generalise to ensure accurate and reliable predictions in real-world applications.

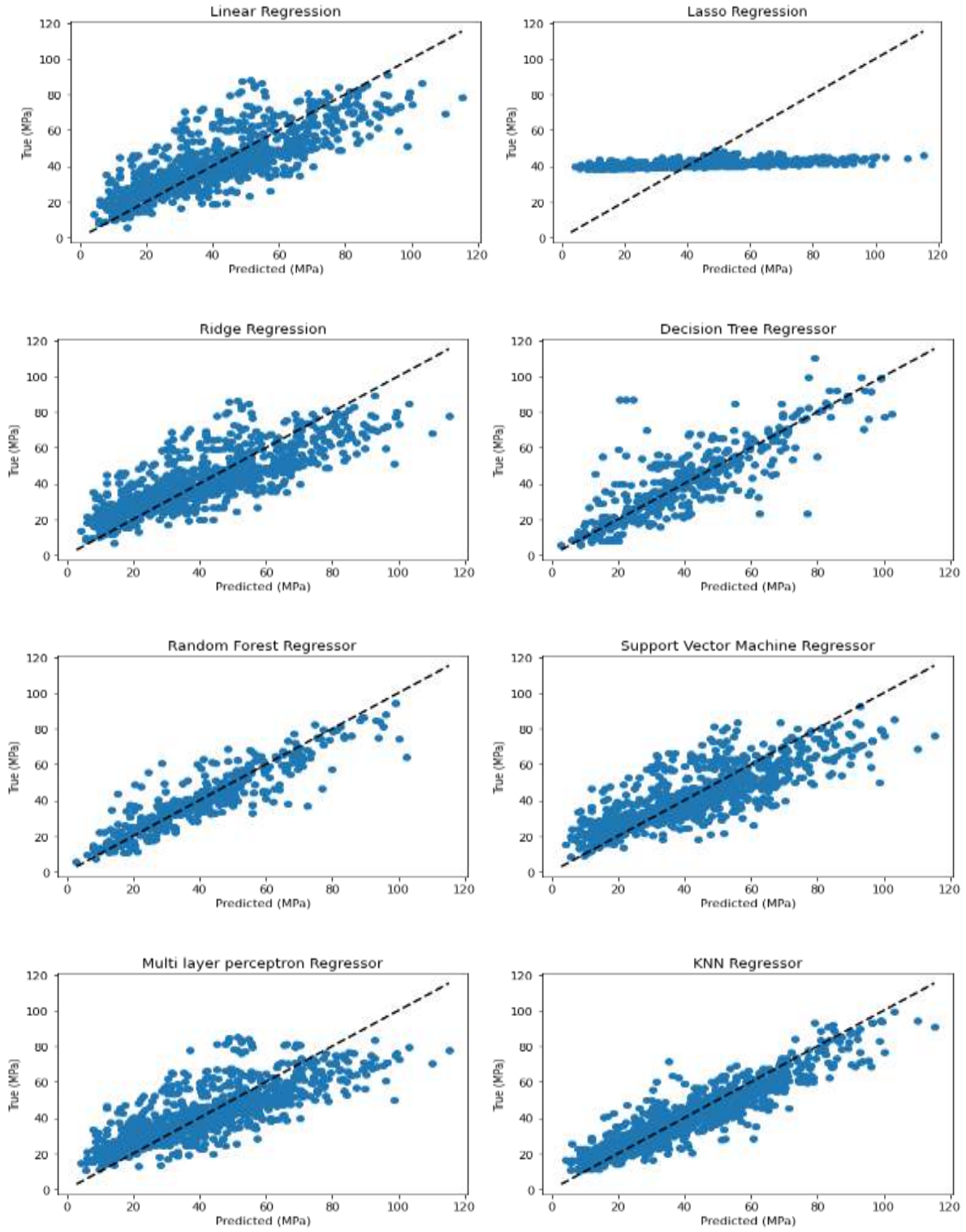


Figure 8. Relationship between the actual and predicted values of CS of all models for the training set

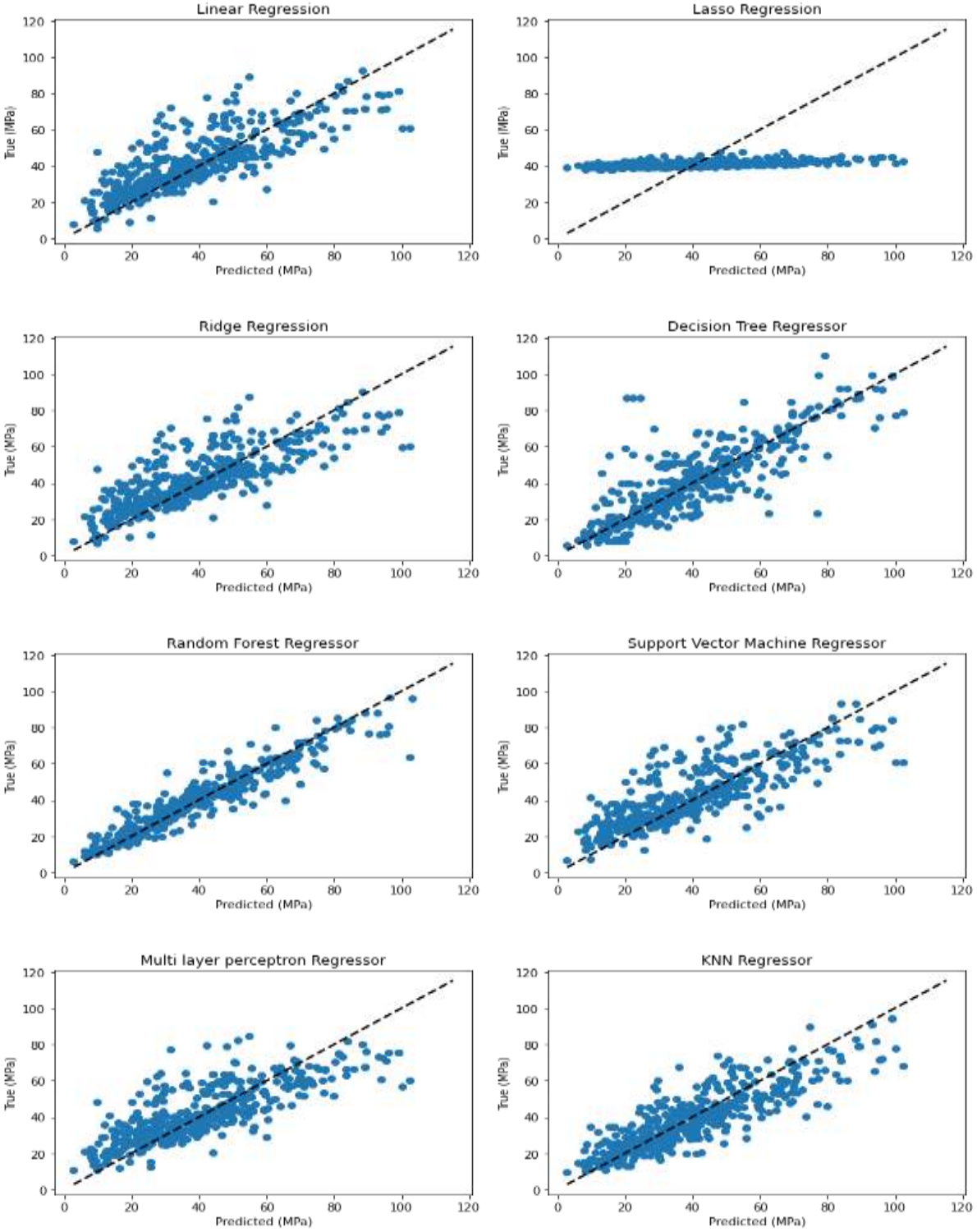


Figure 9. Relationship between the actual and forecasted values of CS of all models for the test set

7 Web application development

After developing several predictive models, a WebApp was created to provide users with an interactive user interface (UI) for predicting concrete mix designs within a defined range of values. The WebApp allows users to enter values that correspond to their intended mix design

and predict the resulting concrete strength. As is evident in Fig. 10, the primary WebApp home page allows users to enter specific components of their mix designs.

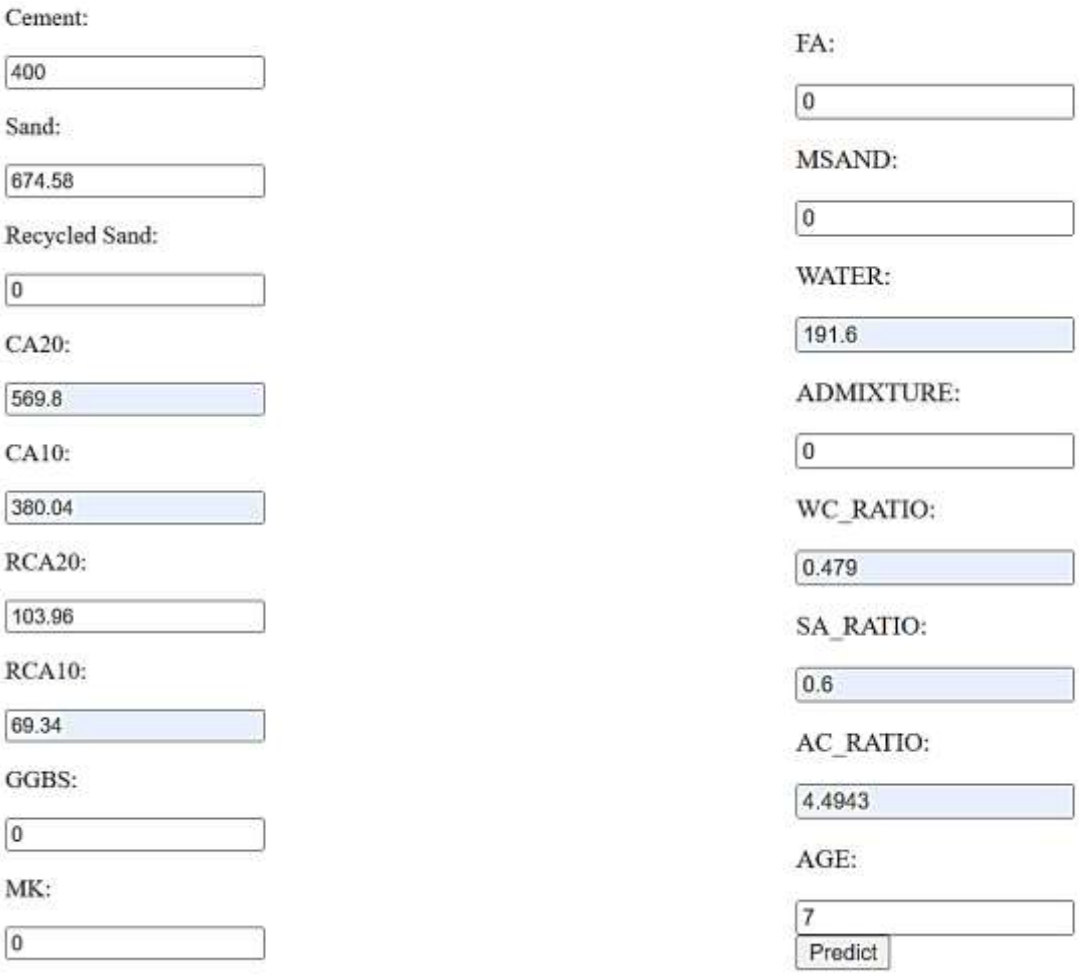


Figure 10. Inputting values for the calculated mix design

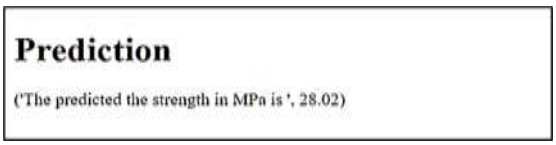


Figure 11. Predicted result for the entered values

Upon entering the necessary input values, users can click on the 'Predict' button, as illustrated in Fig. 11, to receive the predicted CS of the concrete. The WebApp leverages the developed models to make accurate predictions and then presents the predicted CS values to the user. The purpose of this tool was to provide users with a convenient and user-friendly platform to obtain strength predictions for their desired concrete mix designs. By inputting the relevant mix components and using the underlying prediction models, users can quickly gain valuable insights to help them in make informed decisions for construction projects. The intuitive design of the WebApps improves both accessibility and usability, making them indispensable resources for professionals and researchers in the field of concrete engineering. This tool not only streamlines the prediction process, but also allows users to examine different mix designs with ease to ensure that their projects meet the desired performance criteria.

8 Conclusion

In this study, several ML models were developed to predict the CS of concrete. The models were trained using a data set of 1746 entries obtained from laboratory experiments, the UCI repository, and various research studies. The results presented in this paper provide a comprehensive evaluation of the performance and limitations of different ML algorithms for the prediction of concrete CS.

The decision tree regression model showed excellent performance during the training phase, achieving an MAE of 0,21 and an R^2 value of 0,99. However, its inability to accurately predict unseen data during the testing phase showed a tendency to overfit, probably due to its sensitivity to complex variations in the input data. In contrast, the Lasso regression model exhibited showed robustness to changes in the input data, but at the cost of higher prediction errors. Alternatively, the random forest regressor proved to be the most reliable model and outperformed the other algorithms in both the training and testing phases. It had a minimal error and a high correlation coefficient, indicating its strong predictive ability for concrete CS. These results underscore the importance of selecting appropriate ML algorithms for concrete strength prediction and highlight the potential for further research in this area.

Future work could investigate the integration of additional algorithms and advanced scaling techniques to improve the generalisation performance of these models. In addition, the successful development of a web application for the prediction of concrete CS further demonstrates the practical applicability of this research work. This tool not only confirms the effectiveness of predictive models, but also provides a valuable resource for professionals in the field of concrete engineering. To summarise, this study provided important insights into the application of ML techniques to predict concrete strength. It highlighted the strengths and weaknesses of different algorithms, paved the way for future advances, and contributed to the ongoing development of more accurate and reliable prediction tools for concrete technology.

Acknowledgments

The authors are grateful to the Department of Civil Engineering, IT-NU for providing the necessary support to conduct the experiments.

References

[1] Aitcin, P.-C.; Mindess, S. *Sustainability of concrete*. 1st Edition, London: CRC Press, 2011. https://doi.org/10.1201/9781482266696

[2] Mehta, P. K. *Concrete. Structures, Properties and Materials*. United States: Prentice-Hall publication, 1986.

[3] Neville, A. M. *Properties of Concrete*. 5th Edition, New York: John Wiley & Sons, 2012.

[4] Chopra, P.; Sharma, R. K.; Kumar, M. Prediction of compressive strength of concrete using artificial neural network and genetic programming. *Advances in Materials Science and Engineering*, 2016 (1), 7648467. https://doi.org/10.1155/2016/7648467

[5] Noorzaei, J.; Hakim, S. J. S.; Jaafar, M. S.; Thanoon, W. A. M. Development of artificial neural networks for predicting concrete compressive strength. *International Journal of engineering and Technology*, 2007, 4 (2), pp. 141-153.

[6] Akpinar, P.; Khashman, A. Intelligent classification system for concrete compressive strength. *Procedia Computer Science*, 2007, 120, pp. 712-718. https://doi.org/10.1016/j.procs.2017.11.300

[7] Zhang, X.; Dai, C.; Li, W.; Chen, Y. Prediction of compressive strength of recycled aggregate concrete using machine learning and Bayesian optimization methods. *Frontiers in Earth Science*, 2023, 11, 1112105. https://doi.org/10.3389/feart.2023.1112105

[8] Koya, B. P.; Aneja, S.; Gupta, R.; Valeo, C. Comparative analysis of different machine learning algorithms to predict mechanical properties of concrete. *Mechanics of*

Advanced Materials and Structures, 2022, 29 (25), pp. 4032-4043. https://doi.org/10.1080/15376494.2021.1917021

[9] Ahmad, A. et al. Comparative Study of Supervised Machine Learning Algorithms for Predicting the Compressive Strength of Concrete at High Temperature. *Materials*, 2021, 14 (15), 4222. https://doi.org/10.3390/ma14154222

[10] El-Gamal, S. E.; Al-Nuaimi, A.; Al-Saidy, A.; Al-Lawati, A. Efficiency of near surface mounted technique using fiber reinforced polymers for the flexural strengthening of RC beams. *Construction and Building Materials*, 2016, 118, pp. 52-62. https://doi.org/10.1016/j.conbuildmat.2016.04.152

[11] Kumar, R.; Rai, B.; Samui, P. A comparative study of prediction of compressive strength of ultra-high performance concrete using soft computing technique. *Structural Concrete*, 2023, 24 (4), pp. 5538-5555. https://doi.org/10.1002/suco.202200850

[12] Jahangir, H.; Khatibinia, M.; Kavousi, M. Application of contourlet transform in damage localization and severity assessment of prestressed concrete slabs. *Journal of Soft Computing in Civil Engineering*, 2021, 5 (2), pp. 39-67. https://doi.org/10.22115/SCCE.2021.282138.1301

[13] Fakharian, P. et al. Compressive strength prediction of hollow concrete masonry blocks using artificial intelligence algorithms. *Structures*, 2023, 47, pp. 1790-1802. https://doi.org/10.1016/j.istruc.2022.12.007

[14] Chen, L. et al. Axial compressive strength predictive models for recycled aggregate concrete filled circular steel tube columns using ANN, GEP, and MLR. *Journal of Building Engineering*, 2023, 77, 107439. https://doi.org/10.1016/j.jobe.2023.107439

[15] Bansal, T.; Talakokula, V.; Saravanan, T. J. Comparative study of machine learning methods to predict compressive strength of high-performance concrete and model validation on experimental data. *Asian Journal of Civil Engineering*, 2024, 25 (2), pp. 1195-1206. http://dx.doi.org/10.1007/s42107-023-00836-6

[16] Song, Y. et al. Prediction of compressive strength of fly-ash-based concrete using ensemble and non-ensemble supervised machine-learning approaches. *Applied Sciences*, 2021, 12 (1), 361. https://doi.org/10.3390/app12010361

[17] Asteris, P. G. et al. Prediction of cement-based mortars compressive strength using machine learning techniques. *Neural Computing and Applications*, 2021, 33 (19), pp. 13089-13121. https://doi.org/10.1007/s00521-021-06004-8

[18] Güçlüer, K.; Özbeyaz, A.; Göymen, S.; Günaydın, O. A comparative investigation using machine learning methods for concrete compressive strength estimation. *Materials Today Communications*, 2021, 27, 102278. https://doi.org/10.1016/j.mtcomm.2021.102278

[19] Pan, L.; Wang, Y.; Li, K.; Guo, X. Predicting compressive strength of green concrete using hybrid artificial neural network with genetic algorithm. *Structural Concrete*, 2023, 24 (2), pp. 1980-1996. https://doi.org/10.1002/suco.202200034

[20] Yaseen, Z. M. et al. Predicting compressive strength of lightweight foamed concrete using extreme learning machine model. *Advances in Engineering Software*, 2018, 115, pp. 112-125. https://doi.org/10.1016/j.advengsoft.2017.09.004

[21] Torre, A. et al. Prediction of compression strength of high performance concrete using artificial neural networks. *Journal of Physics: Conference Series*, 2015, 582 (1), 012010. https://doi.org/10.1088/1742-6596/582/1/012010

[22] Yu, Y.; Li, W.; Li, J.; Nguyen, T. N. A novel optimised self-learning method for compressive strength prediction of high performance concrete. *Construction and Building Materials*, 2018, 184, pp. 229-247. https://doi.org/10.1016/j.conbuildmat.2018.06.219

[23] Feng, D. C. et al. Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Construction and Building Materials*, 2020, 230, 117000. https://doi.org/10.1016/j.conbuildmat.2019.117000

[24] Tenza-Abril, A. J.; Villacampa, Y.; Solak, A. M.; Baeza-Brotons, F. Prediction and sensitivity analysis of compressive strength in segregated lightweight concrete based

on artificial neural network using ultrasonic pulse velocity. *Construction and building materials*, 2018, 189, pp. 1173-1183. https://doi.org/10.1016/j.conbuildmat.2018.09.096

[25] Paudel, S.; Pudasaini, A.; Shrestha, R. K.; Kharel, E. Compressive strength of concrete material using machine learning techniques. *Cleaner Engineering and Technology*, 2023, 15, 100661. https://doi.org/10.1016/j.clet.2023.100661

[26] Zhang, J. et al. Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression. *Construction and Building Materials*, 2019, 210, pp. 713-719. https://doi.org/10.1016/j.conbuildmat.2019.03.189

[27] Selvaraj, S.; Sivaraman, S. Prediction model for optimized self-compacting concrete with fly ash using response surface method based on fuzzy classification. *Neural Computing & Applications*, 2019, 31 (5), pp. 1365-1373. https://doi.org/10.1007/s00521-018-3575-1

[28] Al-Mughanam, T.; Aldhyani, T. H.; Alsubari, B.; Al-Yaari, M. Modeling of compressive strength of sustainable self-compacting concrete incorporating treated palm oil fuel ash using artificial neural network. *Sustainability*, 2020, 12 (22), 9322. https://doi.org/10.3390/su12229322

[29] Bušić, R.; Benšić, M.; Miličević, I.; Strukar, K. Prediction models for the mechanical properties of self-compacting concrete with recycled rubber and silica fume. *Materials*, 2020, 13 (8), 1821. https://doi.org/10.3390/ma13081821

[30] Aslam, F. et al. Applications of gene expression programming for estimating compressive strength of high-strength concrete. *Advances in Civil Engineering*, 2020, 2020 (1), 8850535. https://doi.org/10.1155/2020/8850535

[31] Balf, F. R.; Kordkheili, H. M.; Kordkheili, A. M. A new method for predicting the ingredients of self-compacting concrete (SCC) including fly ash (FA) using data envelopment analysis (DEA). *Arabian Journal for Science and Engineering*, 2021, 46 (5), pp. 4439-4460. https://doi.org/10.1007/s13369-020-04927-3

[32] Ahmad, A. et al. Prediction of compressive strength of fly ash based concrete using individual and ensemble algorithm. *Materials*, 2021, 14 (4), 794. https://doi.org/10.3390/ma14040794

[33] Wang, M. et al. Research on prediction of compressive strength of fly ash and slag mixed concrete based on machine learning. *Plos One*, 2022, 17 (12), e0279293. https://doi.org/10.1371/journal.pone.0279293

[34] Huang, J. et al. Predicting the compressive strength of the cement-fly ash–slag ternary concrete using the firefly algorithm (fa) and random forest (rf) hybrid machine-learning method. *Materials*, 2022, 15 (12), 4193. https://doi.org/10.3390/ma15124193

[35] Al-Hashem, M. N. et al. Predicting the compressive strength of concrete containing fly ash and rice husk ash using ANN and GEP models. *Materials*, 2022, 15 (21), 7713. https://doi.org/10.3390/ma15217713

[36] Li, D. et al. Machine learning-based method for predicting compressive strength of concrete. *Processes*, 2023, 11 (2), 390. https://doi.org/10.3390/pr11020390

[37] Yeh, I. Modeling of strength of high-performance concrete using artificial neural networks. *Cement and Concrete Research*, 1998, 28, pp. 1797-1808. https://doi.org/10.1016/S0008-8846%2898%2900165-3

[38] Bureau of Indian Standards. IS 10262:2019. *Concrete Mix Proportioning- Guidelines, Indian Standards.* New Delhi: BIS; 2019.

[39] Dabiri, H. et al. Compressive strength of concrete with recycled aggregate; a machine learning-based evaluation. *Cleaner Materials*, 2022, 3, 100044. https://doi.org/10.1016/j.clema.2022.100044

[40] Albaijan, I. et al. Optimal machine learning-based method for gauging compressive strength of nanosilica-reinforced concrete. *Engineering Fracture Mechanics*, 2023, 291, 109560. https://doi.org/10.1016/j.engfracmech.2023.109560

[41] Pallapothu, S. N. R. G.; Pancharathi, R. K.; Midathada, L. Integrating particle packing approach with ML techniques to optimise the compressive strength of RCA based

concrete mixes. *Journal of Building Engineering*, 2024, 94, 109994. https://doi.org/10.1016/j.jobe.2024.109994

[42] Rajakarunakaran, S. A. Prediction of strength and analysis in self-compacting concrete using machine learning based regression techniques. *Advances in Engineering Software*, 2022, 173, 103267. https://doi.org/10.1016/j.advengsoft.2022.103267

[43] Pal, A.; Ahmed, K. S.; Hossain, F. Z.; Alam, M. S. Machine learning models for predicting compressive strength of fiber-reinforced concrete containing waste rubber and recycled aggregate. *Journal of Cleaner Production*, 2023, 423, 138673. https://doi.org/10.1016/j.jclepro.2023.138673