International Journal of Research in Engineering, Science and Management
Volume 7, Issue 3, March 2024
https://www.ijresm.com | ISSN (Online): 2581-5792

Predicting the Compressive Strength of Concrete Specimen using Artificial Intelligence

Nihalahmad R. Faras[1*], Pruthviraj V. Sarniak[2], Ranjeet S. Patil[3], Akshay M. Patil[4], Anil R. Jadhav[5], V. T. Gaikwad[6]

[1,2,3,4,5]Under Graduate Student, Department of Civil Engineering, Padmabhooshan Vasantraodada Patil Institute of Technology, Budhgaon, Sangli, India
[6]Associate Professor, Department of Civil Engineering, Padmabhooshan Vasantraodada Patil Institute of Technology, Budhgaon, Sangli, India

***Abstract*: Concrete is a widely used construction material, and accurately predicting its compressive strength is essential to ensure structural integrity and safety of the structure. The compressive strength of concrete is an important property that decides its load-bearing capacity. The compressive strength of concrete can be determined by conducting a compressive strength test using a universal testing machine. However, this test is time-consuming and expensive. Further, it requires at least seven days of curing to get the idea about compressive strength of concrete. Machine learning can be used to predict the compressive strength of concrete. Machine learning is a subset of artificial intelligence that can be used to learn and derive patterns from data and finally make predictions. This project focuses on developing multiple machine learning models to predict the compressive strength of concrete cube specimen and compare their performance based on various metrics. This study uses 1030 data tests from concrete compressive quality tests obtained from University of California, Irvine, to illustrate the utilize of AI forecast models. The obtained results of the recreation appear that these artificial intelligence methods can build predictive models with great precision.**

***Keywords*: Artificial Intelligence, Artificial Neural Network, Concrete compressive strength, Prediction, Support Vector Machine.**

1. Introduction

Concrete is a mixture of mixture of cement, aggregate and water. A proper concrete mixture requires workability for fresh concrete and durability and strength for the hardened stage. Water is needed for the chemical reaction to form a cement paste and offers workability for fresh concrete.

Among many concrete characteristics, compressive strength is usually considered the most valuable hardened property of concrete. It is measured by breaking cylindrical concrete specimens in a compression- testing machine at 28 days of standard curing. The testing procedure and standard size of test specimens are in accordance with American society for Testing and Materials (ASTM). Several factors might affect the compressive strength of the concrete, such as age, ingredients, water to cement ratio, curing conditions, etc. Typically, the compression test result of concrete at 28 days is considered as a standard to determine the quality of concrete.

If the compression test result does not meet the required strength, the mix design needs to be replaced, which might be labor consuming and time consuming. To minimize the risk of a specific concrete mix design falling short of compression strength requirement at the age of 28 days, a method to predict the 28-day strength from its primary ingredients is vital. Traditionally, the experimental method is broadly used to study the properties of materials. In recent years, the application of artificial intelligence- based models such as ANN to predict the concrete mechanical properties has increased significantly. Those models have an ability to learn from the data to establish the non-linear relationship between the inputs and outputs for the complex engineering issues.

Problem statement:

How we can well predict the compressive strength of concrete cube specimen, at the given characteristic strength of previous concrete cube specimen data using the Artificial Intelligence.

A. *Future Scope of Study*

Though the work in the project is sufficient to prove the usefulness of machine learning in predicting the compressive strength of concrete cube specimens, it is possible to use an even larger dataset containing more samples of concrete cube specimens. Larger datasets can improve the accuracy of the model. In our experiment, we utilized 1030 datasets to train the machine learning algorithms. By using larger datasets, we can help in better training of the models and also improve the accuracy of the models since the generalization capability of the machine learning models increases. Also, in this project, we have used three machine learning models. To get more accurate results, we can look for more machine learning models appropriate for this purpose. It is also possible to improve the accuracy of predictions further by using some deep learning algorithms. Finally, it is necessary that the data should be extracted in csv format and more data samples must be collected and used in the training process for the algorithms.

B. *Theoretical Concepts*

Artificial intelligence (AI) is the field of computer science and engineering that works on the creation of intelligent machines and software that can think and act like humans. These machines use algorithms and data to make decisions and

*Corresponding author: nihalfaraz20@gmail.com

perform tasks, which normally require human intelligence, such as learning, problem-solving or decision-making.

It has the potential to bring about many revolutionary changes multiple industries, ranging healthcare to finance. In fact, there are multiple fields undergoing a massive change due to this new technology. Some common applications of AI include natural language processing, machine learning, and automation.

But at the same time, even though AI has the potential to drastically improve the world, ethical concerns and doubts are also being raised about the future of human-machine interactions and their overall impact on human lives. As this technology continues to improve, it is essential to take into consideration the potential implications and hazards of its widespread use.

1) Artificial Intelligence, Machine Learning and Deep Learning

Artificial intelligence is a broader field of computer science and engineering that involves statistics also. Its focus is to create intelligent machines and software, which can assist humans in better decision making. While machine learning is a subset of artificial intelligence, which involves the use of algorithms and data to enable a machine or computer to learn patterns from the data and perpetually improve its performance on a specific task without the need of being explicitly programmed.

Deep learning is a subset of machine learning which utilizes artificial neural networks to model high-level abstractions in data.

In summary, AI is the superset that includes both machine learning and deep learning, with machine learning being a subset of AI and deep learning being a subset of machine learning.

2) Brief History of Artificial Intelligence

The idea of "a machine that thinks" can be traced back to ancient Greece. In the evolution of artificial intelligence, the following milestones are noteworthy:

1. In 1950, Alan Turing published introduced the Turing Test to decide whether a computer can demonstrate the same level of intelligence as a human.
2. In 1956, the term "artificial intelligence" was coined at the first-ever AI conference at Dartmouth College by John McCarthy.
3. In the 1980s, neural networks which use a backpropagation algorithm to train themselves became widely used in AI applications.
4. In 2015, Baidu's Minwa supercomputer employed a special kind of deep neural network called a convolutional neural network to recognize and classify images with a higher rate of accuracy than the average human.
5. In 2016, DeepMind's AlphaGo program, beat Lee Sedol, the world champion Go player.
6. And recently, in 2022, an artwork made by Artificial Intelligence won first place at the Colorado State Fair’s fine arts competition. Midjourney enables users with no art training to develop high-quality artwork. It produces multiple renditions, and allows users to choose the best of the available options. It is an artificial intelligence software that converts text into images.

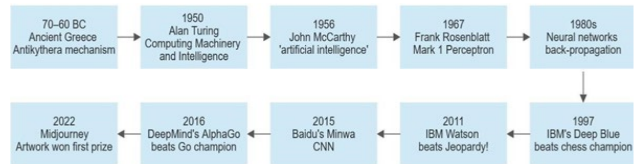

Fig. 1. Overview of History of AI

2. Literature Review

Artificial intelligence-based estimation of ultra-high-strength concrete’s flexural property (Authors: Wang et.al, 2022): This research has used 15 prediction input variables. The AI techniques used include AdaBoost, Bagging, Gradient Boosting and Extreme Gradient Boosting. To compare the results obtained from these models, k-fold cross validation has been used. The data sample is split into ‘k’ number of smaller samples, which is the reason behind the name: K-fold Cross Validation. In order to set up an ensemble learning method, they have selected the base models to be aggregated. Most of the times, a single base learning algorithm is utilized, so that we have homogeneous weak learners that are trained in different ways. The ensemble model obtained can be called as “homogeneous”.

Prediction of concrete compressive strength using Artificial Intelligence methods (Authors: Muilauwn et.al, 2020): This study applies multiple artificial intelligence (AI) methods to find the most accurate input and output relationships within concrete mixtures. The three types of AI methods that will be used in this study are artificial neural networks (ANN), support vector machine (SVM), and linear regression (LR). This study uses 2000 data samples from concrete compressive strength tests obtained from Massachusetts Institute of Technology, to demonstrate the use of AI prediction models. The obtained results of the simulation show that these artificial intelligence methods can build predictive models without conducting any expensive experiments in the laboratory with good accuracy.

Artificial intelligence for compressive strength prediction (Authors: Goutham et.al, 2020): In the study carried out, a prediction model was generated to predict the compressive strength of the concrete using support vector regression which is a machine learning technique. The prediction modelling was made in python language. It has been found that RBH model produces the best prediction output in comparison to other NDT techniques. The values of MAPE, MAD, RMSE and R2 also support the findings.

Research Gap Analysis:

- Studies have relied heavily on small datasets (<200 samples). This leads to overfitting of AI models and poor performance on new models.
- Lack of large and universal datasets.
- Non availability of data sets in desired digital formats.
- Manual extraction - cumbersome and labour intensive.

- No use of transfer learning.

3. Objectives

- To study databases used to predict compressive strength of concrete.
- To understand the change in compressive strength of concrete with change in proportions of the ingredients.
- To develop and study models used to predict compressive strength of concrete.
- To compare accuracy of the developed models.

4. Methodology

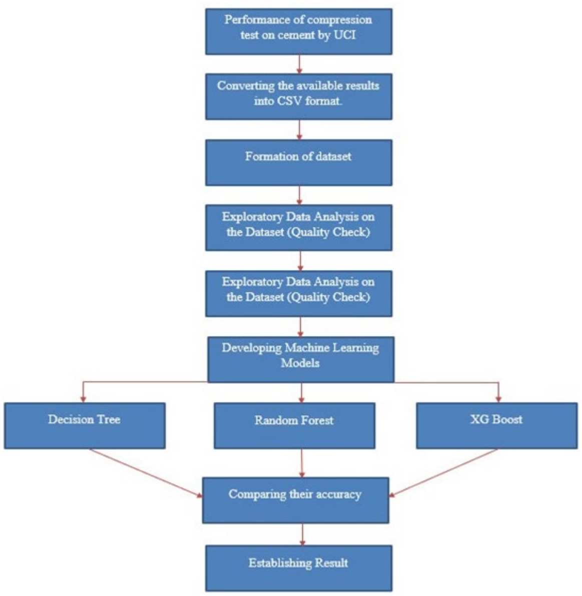

Fig. 2. Workflow diagram

The Procedural methodology is divided into five major sections:

1) Identifying the domain where AI can be applied
2) Studying the datasets
3) Applying algorithms to get the result
4) Getting the results
5) Establishing the accuracy of the models

The process takes the AI algorithms into consideration. The given data is divided into training dataset and testing dataset. Using training dataset, you make an AI algorithm and then use the algorithm on the testing dataset to generate predictions of compressive strength of concrete.

The next step is to evaluate the performance. We have the output from the testing datasets (strength) and we have the output from the algorithm, (predicted strength).

It is very important to split the given dataset into training and test dataset. The purpose of the training dataset is to discover a predictive relationship by using the model.

Once the model is mature, the test dataset is used to get the accuracy of the hypothesis. Since the model has not seen the test dataset during the training phase, it gives us unbiased results. Therefore, it is important to keep the test dataset separate from the training dataset.

There are lots of datasets available on the internet of the samples of concrete cubes which have different components like the cement, blast furnace slag, fly ash, fine aggregates, coarse aggregates, water, superplasticizer and age. People have made these cubes and have waited for number of days and then got it tested in the laboratory to calculate the strength. The data is taken as the input and output is the predicted compressive strength of the concrete.

A. *Exploratory Data Analysis*

1) *Definition*

Exploratory data analysis (EDA) is a method of analyzing data sets to summarize their main characteristics, mostly by using statistical graphics and other data visualization methods.

2) *Need for EDA*

1. To get maximum data insights.
2. Discover underlying structure
3. For extracting important variables from the data.
4. Detect outliers (if any).
5. To check if there are any null values.
6. To understand and test the underlying assumption.

3) *EDA Procedure*

1. Understanding the dataset – counts, mean, standard deviation, minimum values, and maximum value.
2. Checking for null values and getting their count.
3. Detecting and counting the number of outliers by defining interquartile range.
4. Plotting the distribution of columns.

B. *Understanding The Dataset*

cement	blast_furnace	fly_ash	water	superplasticizer	coarse_agg	fine_agg	age	compressive str
540	0	0	162	2.5	1040	676	28	79.99
540	0	0	162	2.5	1055	676	28	61.89
332.5	142.5	0	228	0	932	594	270	40.27
332.5	142.5	0	228	0	932	594	365	41.05
198.6	132.4	0	192	0	978.4	825.5	360	44.3
266	114	0	228	0	932	670	90	47.03
380	95	0	228	0	932	594	365	43.7
380	95	0	228	0	932	594	28	36.45
266	114	0	228	0	932	670	28	45.85
475	0	0	228	0	932	594	28	39.29
198.6	132.4	0	192	0	978.4	825.5	90	38.07
198.6	132.4	0	192	0	978.4	825.5	28	28.02
427.5	47.5	0	228	0	932	594	270	43.01
190	190	0	228	0	932	670	90	42.33

Fig. 3. Overview of dataset

Experimental data were obtained from a machine learning repository held at the University of California, Irvine (UCI) that was collected by Yeh. A total of 1030 concrete samples evaluated by various university research laboratories were used to test the prediction models of each AI method. All tests were carried out on a 15 cm cylindrical concrete specimen using standard procedures.

The different ingredients that were used in the concrete are:

1. Cement
2. Blast Furnace Slag
3. Fly Ash
4. Water
5. Superplasticizer
6. Coarse Aggregate
7. Fine Aggregate
8. Age

Variables	Unit	Min	Mean	Max	Standard deviation
X_1: Cement	kg/m^3	102.0	281.17	540.0	104.51
X_2: Blast-furnace slag	kg/m^3	11.0	107.28	359.4	61.88
X_3: Fly ash	kg/m^3	24.5	83.86	200.1	39.99
X_4: Water	kg/m^3	121.8	181.57	247.0	24.35
X_5: Superplasticizer	kg/m^3	1.7	8.49	32.2	4.04
X_6: Coarse aggregate	kg/m^3	801.0	972.92	1,145.0	77.75
X_7: Fine aggregate	kg/m^3	594.0	773.58	992.6	80.18
X_8: Age of testing	Day	1.0	45.66	365.0	63.17
Y: HPC compressive strength	MPa	2.3	35.82	82.6	16.71

Fig. 4. Describing the dataset

X= Input Parameter (Independent Variable)
Y= Output Parameter (Dependent Variable)

C. *Checking for Null Values*

cement	blast_furnace_slag	fly_ash	water	superplasticizer	coarse_aggregate	fine_aggregate	age	concrete_compressive_strength
False	False	False	False	False	False	False	False	False
False	False	False	False	False	False	False	False	False
False	False	False	False	False	False	False	False	False
False	False	False	False	False	False	False	False	False
False	False	False	False	False	False	False	False	False

Fig. 5. Check for null values

The presence of null value is indicated by a 'True' in the cell and if the cell is not null it is indicated by a false. The count of the null values in each column was checked which came out to be zero indicating the absence of null values in the column.

```
cement                          0
blast_furnace_slag              0
fly_ash                         0
water                           0
superplasticizer                0
coarse_aggregate                0
fine_aggregate                  0
age                             0
concrete_compressive_strength   0
```

Fig. 6. Count of null values

D. *Outlier detection*

Outliers are values within a dataset that vary largely from the other values in the dataset— they are either too larger, or significantly smaller. Outliers indicate variabilities in a measurement, experimental errors, or a novelty.

Outlier Lower limit = Q1 - (Q3-Q1) * 1.5
Outlier Upper Limit = Q3 + (Q3-Q1) * 1.5
Q1= First quartile, Q3 = Third quartile

For outlier detection here, a box plot is used. It is also known as box and whisker plot. A box plot is a chart that displays data from a five-number summary including one of the measures of central tendency. The figure below shows the different parts of box plot.

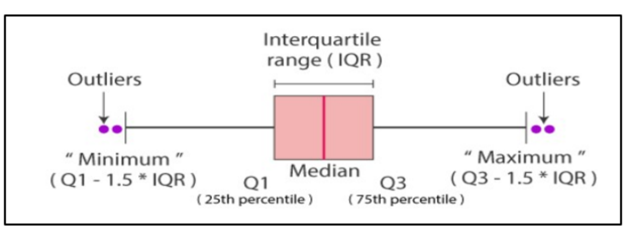

Fig. 7. Different parts of boxplot

To check outliers of the data box plots of each column were plotted. The presence of dots indicates the presence of outliers. Thus, the following observations can be made

- Cement: No Outliers
- Slag: Outliers Present
- Ash: No Outliers
- Water: Outliers Present
- Superplasticizer: Outliers Present
- Coarse Aggregate: No Outliers
- Fine Aggregate: Outliers Present
- Age: Outliers Present
- Strength: Outliers Present

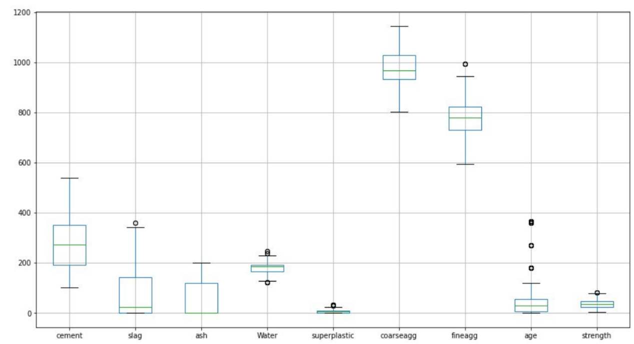

Fig. 8. Box plot

E. *Checking the Distribution of Data*

The distribution of data is found out by plotting the Gaussian curve to understand whether the data is rightly skewed or left skewed.

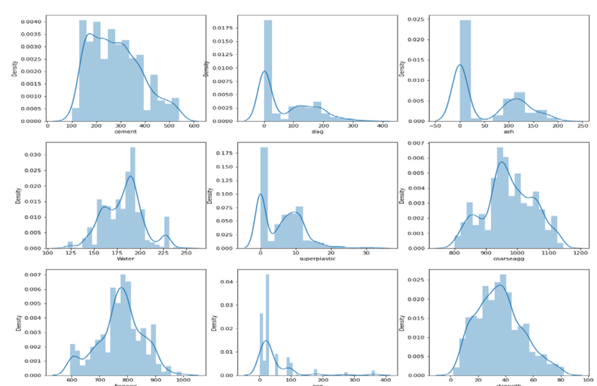

Fig. 9. Distribution plot

- Cement is almost normal
- Slag has 3 Gaussians and is rightly skewed
- Ash has 2 Gaussians and is rightly skewed.
- Water has 3 Gaussians and rightly skewed
- Superplastic: 2 Gaussians, rightly skewed
- Coarse aggregate: 3 Gaussians, almost normal
- Fine aggregate: 2 Gaussians, almost normal
- Age: multiple Gaussians, rightly skewed

F. *Understanding the Inter-Relationships of the Ingredients Using Pair plot*

To understand how the compressive strength changes according to the change in the ingredients a scatter pair- plot is used. A pair plot shows the variance of the attributes with every other attribute.

Positive Co-relation: When the points within the chart are rising, moving from left to right, at that point the diffuse plot appears a positive relationship. It implies the values of one variable are expanding with regard to another.

Negative Co-relation: When the points in the scatter chart fall whereas moving left to right, then it is called a negative correlation. It implies the values of one variable are diminishing with regard to another.

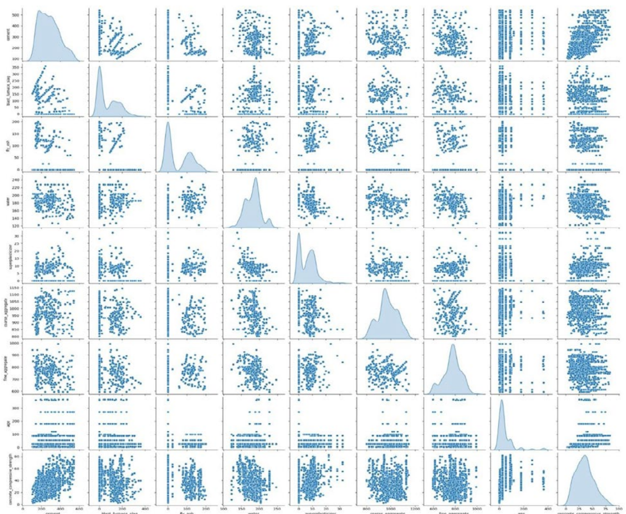

Fig. 10. Scatter pair plot

Closure: Determination of data of the aggregate size can be determined by using Gaussian curve.

G. Co-Relation Matrix

A correlation matrix is used to compare the coefficients of different features (or attributes) within the data. It allows us to see how much (or how little) correlation exists between different variables.

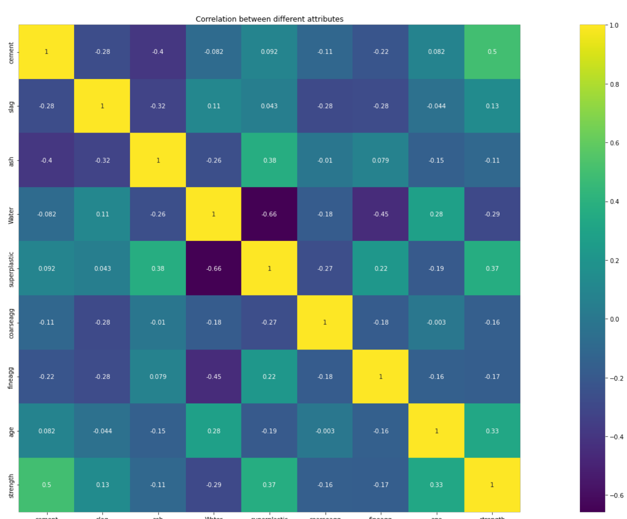

Fig. 11. Co-relation matrix

The closer the value is to 1 (or -1), the better the relationship. The closer the number is to 0, the weaker the relationship.

Closure: A negative coefficient will tell us that the relationship is negative, that is, it gets lower as a value increases. Similarly, a positive coefficient means that when one value increases, the other value also increases.

H. Developing Machine Learning Models

Machine learning models are computer programs that are utilized to recognize patterns in data or make predictions. We utilized 3 models which are as follows:

1) XGBoost Regressor

1. In Extreme Gradient Boosting (XGBoost), new decision trees are added iteratively to the base model.
2. Then, each of the subsequent trees in the ensemble corrects the errors made by the previous tree.
3. The final prediction of this algorithm is the combined output of all the individual predictions made by each of the trees.
4. To enhance its performance, there several hyperparameters that can be tuned to maximize the accuracy.
5. Some of the important hyperparameters are learning rate and the maximum depth.

2) Decision Tree Regressor

1. Decision Tree Regressor uses a tree-like structure, hence the name decision tree.
2. In a decision tree, a question is asked at each node and based on the answer, relevant next node is selected.
3. In any decision tree, the top most node is known as the root node, while nodes without any subsequent nodes are called leaf nodes.
4. Here, the resultant values are predicted based on simple decision rules.
5. These rules are derived from the patterns of the data.
6. Finally, the average of predictions from each of the leaf nodes is calculated as the final value of the target variable.

3) Random Forest Regressor

1. Random Forest Regressor is another tree-based machine learning model.
2. It used ensemble of parallel decision trees (bagging) to predict the value of the target variable.
3. In this algorithm, multiple decision trees are created during the training of the model.
4. Ultimately, the final value is calculated by aggregating the outputs given by each of the trees.
5. Fundamentally, it works on the concept of collective knowledge.

Decision Tree	Random Forest	XG Boost
Single decision tree	Random forest trees are built independently.	XG boost ,a tree is dependent on the previous tree.
Top down method	Parallel combination	Series Combination
Lowest time	Lesser time	More time.
small size data	Medium size data for random forest	Large data set is used
Low computation.	High Computation	Very High Computation

Fig. 12. Comparison of models

I. Training the Model

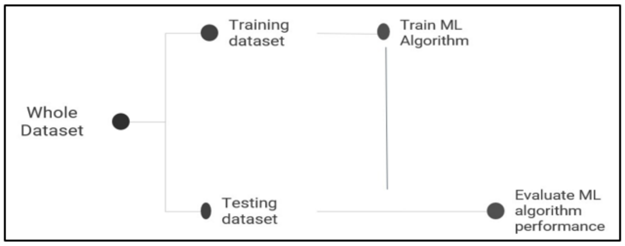

Fig. 13. AI workflow

The given information is partitioned into preparing dataset (70%) and by testing the datasets (30%). Utilizing and preparing dataset leads in the formation of AI calculation and then utilize those on the testing dataset in order to produce the expectations of the compressive quality of the concrete.

The next step over here is in order to access the execution. We have the datasets that we have tested and then we also have the yield that consists of anticipated strength from the calculations.

The next step is to assess the execution. We have the yield from the testing datasets (quality) and we have the yield from the calculation (anticipated strength). It is exceptionally important to part the given dataset into preparing and test dataset.

Table 1
Results of the model

	Actual	Decision tree prediction	Random forest prediction	XG Boost Prediction
75	72.10	62.05	58.502033	61.583641
166	38.70	52.42	46.570800	42.678562
237	54.77	77.30	62.906527	64.068291
153	52.50	33.94	42.532950	43.810886
113	28.99	29.87	28.431100	23.790628
128	66.70	65.20	68.752100	68.238274
86	33.73	33.70	29.913300	32.215401
44	35.23	34.67	32.050300	33.161350
98	27.94	21.07	30.166600	31.844952
225	10.39	13.66	11.881200	12.910582
256	37.23	37.91	37.697500	37.863071
199	40.68	40.87	37.605800	37.959454
289	48.59	49.99	47.323800	42.555294

The table 1 shows the actual compressive strength of the specimen along with the predicted results from the 3 models. The first column shows us that the sampling is done in a randomized manner. To find out which model gives the most accurate predictions, we use an indicator called R square. R square formed the basis of comparison of various models generated. The range of R2 is 0-1 where 1 indicates 100% accuracy in prediction.

Algorithm	R_square
Random Forest	0.885212
Decision Tree Regressor	0.773357
XG Boost	0.904431

Fig. 14. Accuracy of the model

Closure: After the comparisons between the algorithms XG boost is determined as the most efficient machine learning algorithm.

5. Conclusion

This study presents a comparative study between 3 Machine Learning models, namely Decision Tree, Random Forest and XG Boost, in predicting the concrete compressive strength based on 1030 concrete samples. These samples were used to form a database that was used to create a prediction model as well as to test the accuracy of the prediction model formed. Accuracy indicators are used in evaluating the performance of each method, including R, RMSE, MAPE, and MAE. Finally, R square was used as a final indicator from which it was concluded that XG Boost shows the greatest accuracy among the three.

Change in the compressive strength is observed along with change in ingredients. The models are trained with different data sets and then the results are incurred. Out of the three machine learning models selected, XG Boost is determined as the most accurate model.

References

[1] Prayogo D, "Metaheuristic-based machine learning system for prediction of compressive strength based on concrete mixture properties and early-age strength test results", *Civil Engineering Dimension*, vol. 20, no. 1, 21-29, March 2018.
[2] Wang et. al, "Artificial intelligence-based estimation of ultra-high-strength concrete's flexural property", *Case study in Construction Materials*, vol. 17, December 2022, e01243.
[3] Goutham et. al, "Artificial Intelligence for Compressive Strength Prediction of Concrete," IOP Conf. Series: Materials Science and Engineering, ICRADM India, 2020.
[4] Muilauwn et. al, "Prediction of concrete compressive strength using Artificial Intelligence methods," in *Journal of Physics-Conference series*, 2nd ICSI, 1985, pp. 585–590.
[5] http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html?
[6] https://www2.deloitte.com/nl/nl/pages/data-analytics/articles/part-1-artificialintelligence-defined.html
[7] https://www.ibm.com/in-en/cloud/learn/what-is-artificial-intelligence
[8] https://www.mckinsey.com/business-functions/operations/ourinsights/artificialintelligence-construction-technologys-next-frontier
[9] https://scikitlearn.org/stable/modules/generated/sklearn.linear_model.LinearRegre ssion.html
[10] https://scikitlearn.org/stable/modules/generated/sklearn.linear_model.Ridge.html
[11] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Lasso.html
[12] https://scikitlearn.org/stable/modules/generated/sklearn.neural_network.MLPRegr essor.html
[13] https://xgboost.readthedocs.io/en/stable/python/python_api.html
[14] https://scikitlearn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegress or.html
[15] https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.RandomForestR egressor.html
[16] https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.GradientBoosti ngRegressor.html
[17] https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.AdaBoostRegre ssor.html
[18] https://scikitlearn.org/stable/modules/generated/sklearn.neighbors.KNeighborsRegressor.html
[19] https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.BaggingRegressor.html
[20] https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html
[21] https://www.kaggle.com/datasets/elikplim/concrete-compressive-strength-data-set