JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ANALYSIS OF COMPRESSIVE STRENGTH AND WORKABILITY OF CONCRETE BY PREDICTIVE APPROACH

[1]Anish Tulswani, [2]Anant Kumar

[1]PG Scholar, [2]Assistant Professor, Department of Civil Engineering, Shri Shankaracharya Institute of Professional Management and Technology, Raipur (C.G.)

Abstract - Concrete is a crucial substance extensively used in the construction sector, and concrete compressive strength, the measure of a material's ability to withstand axial loads, and workability, the ease with which concrete can be mixed, transported, and placed are the key factors that determine the overall functioning of the construction and its structural integrity. For structures to last for a long time and be of high quality, it is essential to attain the desired compressive strength and workability. Existing methods for analyzing these attributes rely on physical testing or experimental trials. This project offers a thorough method for forecasting concrete's compressive strength and workability, which depends on its mix proportions. By utilizing an extensive collection of different concrete mixtures and their corresponding compressive strength values, the research identifies key factors influencing these characteristics. The dataset of concrete mix ingredients, such as cement content, water, fine & coarse aggregates, and admixtures, is analyzed. To find the most important parameters, feature selection methods are used after the dataset has been preprocessed to eliminate inconsistencies. This research shows both compressive strength and workability can be accurately estimated by predictive approach, obtaining the best accuracy with R^2 value more than 95%.

Keywords: Concrete, Compressive Strength, Workability, Mix Proportion, Experimental Trials, Predictive Approach, Feature Selection, Construction Quality, Preprocessing, R^2 Value.

I. INTRODUCTION

1.1 General

Concrete is one of the most popular building materials in the world because of its affordability, durability, and adaptability. Two of its most important performance factors are workability and compressive strength, which have a direct impact on the structural soundness and usefulness of concrete in a variety of building contexts. These properties are traditionally determined through extensive laboratory testing, which is labor-intensive, resource-intensive, and frequently prone to human error. With advancements in data science and artificial intelligence, predictive modeling has emerged as a powerful tool for estimating material properties using mathematical and statistical methods. Concrete behavior can be predicted quickly, accurately, and economically using predictive techniques like machine learning, regression analysis, and neural networks. These techniques take into account input variables like the water-to-cement ratio, aggregate size, admixtures, and curing conditions. With the use of predictive models, this study seeks to evaluate the workability and compressive strength of concrete, offering a novel substitute for conventional testing. The approach also enables optimization of mix design and ensures quality control, making it highly relevant to modern, technology-driven construction practices.

The concept of ML models for predicting concrete's properties involves several stages:

- **Collection of Data and Preprocessing:** Gathering comprehensive datasets encompassing a broad range of different grades of concrete compositions, mixing procedures, curing regimes, & testing conditions is essential. To guarantee consistency and dependability, raw data frequently needs preprocessing procedures like normalization, feature scaling, and handling missing values.
- **Feature Selection and Engineering:** Identifying relevant input features that significantly influence compressive strength and workability is crucial. Feature engineering techniques may involve transforming or combining variables to enhance model performance and interpretability.
- **Model Training and Evaluation:** Selecting appropriate ML algorithms and training them on the prepared dataset is a critical step. Models are evaluated using measures such as MAE, RMSE, and R^2 to measure a model's predictive accuracy and capacity.
- **Hyperparameter Tuning:** Fine-tuning model parameters and architecture to optimize performance and prevent overfitting is essential. The best hyperparameters for every ML algorithm are found with the aid of strategies like grid search and cross-validation.
- **Model Validation and Deployment:** Validating the trained models on separate datasets to confirm their accuracy, reliability, and robustness is vital. Once validated, the models may be deployed in real-world applications, guiding decision-making processes in construction projects.

1.2 Problem Identification

The construction industry still primarily uses empirical methods and extensive laboratory testing to evaluate the properties of concrete, especially compressive strength and workability, despite advancements in civil engineering technology. These traditional methods are not only time-consuming and labor-intensive but also may lead to inconsistencies due to variations in testing procedures and human judgment. Additionally, trial-and-error mix designs frequently lead to material waste and higher project expenses. Accurately forecasting concrete's performance based on changing material compositions and environmental conditions without conducting extensive physical testing is a major challenge. Additionally, the lack of integration between concrete material data and predictive analytics hampers efficient decision-making in construction planning. This creates a critical need for a robust, data-driven approach that can accurately forecast concrete properties and support optimized mix designs. Addressing this problem through predictive modeling can improve construction efficiency, reduce costs, enhance sustainability, and provide reliable quality assurance in concrete production and application.

1.3 Research Objective

This study offers important new information on how to use ML techniques to predict the compressive strength of concrete. The demonstrated accuracy, interpretability, and generalizability of our model hold significant implications for the field of study of concrete engineering and construction, opening the door to better methods of concrete construction and design.

To develop & assess a ML based predictive model that accurately determines the concrete's compressive strength and workability for different grades and compositions such as material proportions, admixtures, and curing conditions.

The objective is to focus on varying grades and compositions of concrete while emphasizing the development and validation of the ML model. It also makes it clear that a range of variables will be considered in predicting the properties of the concrete.

II. LITERATURE REVIEW

In pursuit of the study's goals, pertinent data was collected from the global scientific community by analyzing a range of sources including textbooks, literature, international scientific journals, environmental progress reports from multiple agencies, and beyond. Extensive information was gleaned from websites, governmental documents, and a comprehensive examination of publications by fellow researchers on subjects aligned with the research focus. Following this, a conclusive literature review was conducted to enhance comprehension of the research.

1. Premalatha Krishnamurthy, Sowmya Kochukrishnan, Nandagopal Kaliappan, & Yuvarajan D. (2024) - had done an estimation of uniaxial compressive strength (UCS) in rocks, which is vital for various geomechanical applications, ranging from foundation design to tunnel construction. However, direct determination of UCS can be challenging due to sampling and preparation difficulties, necessitating the use of indirect methods. Kochukrishnan et al. contribute to this field by presenting a thorough investigation of regression machine learning models based on Python for predicting UCS in Charnockite rocks. These models leverage parameters such as Schmidt Hammer Rebound Number. This study not only addresses the practical challenges in rock mechanics but also demonstrates the effectiveness of ML techniques in enhancing predictive modeling for geotechnical applications, offering valuable insights for rock characterization and engineering design in civil projects.

2. Wu and Huang (2024) -use a dynamic CatBoost Regression model in conjunction with individual and ensemble optimization techniques to predict the flexural strength (FS) and compressive strength (CS) of high-performance concrete (HPC). By highlighting the value of trustworthy predictive models, the study lessens the need for iterative experiments and intensive laboratory testing when designing reinforced concrete structures to industry standards. Higher R2 values and better statistical accuracy metrics like RMSE and MAE show that the standalone CatBoost Regression (CAT) model performs better than hybrid and ensemble models that incorporate optimization algorithms like Artificial Rabbits Optimization (ARO) and Honey Badger Algorithm (HBA) in predicting both CS and FS.

3. Mohamed Kamel, Ahmed Karam, Yasser Mater, & Emad Bakhoum (2023) - have studied the response to the global imperative for sustainable construction practices, the integration of waste management and AI represents a significant step forward. Mater et al. contribute to this endeavor by focusing on the concept of an artificial neural network (ANN) model specifically designed to predict green concrete's compressive strength. Their study aligns with the industry's growing emphasis on utilizing sustainable materials, aiming to facilitate the incorporation of recycled coarse aggregate (RCA), recycled fine aggregate (RFA), and fly ash (FA) as partial replacements for traditional concrete constituents. By harnessing the power of ANN technology, Mater et al. seek to enhance predictive capabilities in assessing the performance of green concrete, thereby promoting its broader adoption in construction projects. This research not only addresses the pressing need for sustainable material solutions but also underscores the potential of artificial intelligence in advancing green building practices, presenting encouraging approaches to lessening the impact on the environment and promoting resource efficiency in the construction industry.

4. Sudhakar Singha, E. V. Prasad, & B. Vamsi Varma (2023) - have used ML techniques to forecast the mechanical characteristics of concrete—specifically its compressive strength—which has enormous importance. By examining the effectiveness of gradient boosting (GBM) and light gradient boosting (LGBM) supervised ML techniques for forecasting concrete compressive strength, Varma et al. make a contribution to this field. Eight independent variables—cement content, fly ash, blast furnace slag, water content, superplasticizer, fine aggregate, and coarse aggregate—are used in their study, which was carried out on the Python platform. The authors compare the predictive capacities of GBM and

LGBM models through a thorough evaluation that makes use of performance metrics like R^2, MSE, RMSE, and MAE. The LGBM model performs better than GBM, according to the results, showing higher.

5. Kuldeep K. Saxena, Nakul Gupta, & Priyanka Gupta (2023) - have significantly advanced the field of geopolymer concrete (GPC) through the application of ML techniques to predict compressive strength. Their study delves into crucial variables such as curing conditions, FA, calcined clay, and additives, employing a range of models including KNN, LR, and random forest regression (RFR). Impressively, the findings reveal that RFR surpasses other models, achieving an impressive R^2 value of 0.92, thereby affirming its exceptional predictive accuracy. Additionally, through descriptive statistical analysis, the study validates the importance of input parameters, further reinforcing the efficacy of RFR in optimizing GPC design. By highlighting RFR's pivotal role in enhancing prediction accuracy for GPC compressive strength, this research offers invaluable insights essential for its effective design and optimization.

6. Mohammed Najeeb Al-Hashem, Muhammad Nasir Amin, Kaffayatullah Khan, Waqas Ahmad, Ayaz Ahmad, Muhammad Ghulam Qadir, Muhammad Imran & Qasem M. S. Al-Ahmad (2022) - contributed to the advancement of self-compacting concrete (SCC) technology by exploring ML techniques for predicting compressive strength. SCC offers numerous advantages in construction because of its ability to flow and fill formwork without segregation. Their study investigates the application of multilayer perceptron, bagging regressor, and support vector machine models in analyzing SCC properties. Based on data from various published sources, including 11 input parameters, the research highlights the superior performance of the bagging regressor technique in accurately predicting compressive strength. The study underscores the significance of ML techniques in optimizing SCC design and performance evaluation, providing insightful information about how input parameters affect compressive strength.

7. Anthony Butera, Vivian W.Y. Tam, Khoa N. Le, Ana C.J. Evangelista, & Luis C.F. Da Silva (2022) - have studied how concrete, a ubiquitous construction material, contributes significantly to CO_2 emissions and the depletion of natural resources, posing environmental challenges. In response, CO_2 Concrete emerges as a promising alternative, leveraging carbonation of recycled aggregate and reuse of materials to mitigate environmental impacts. For CO_2 concrete to be widely used, it is essential to predict its compressive strength. In order to meet this need, Tam et al. create prediction models by applying artificial neural networks (ANNs) and regression analysis. Their research emphasizes how crucial precise forecasting is to guaranteeing the dependability of CO_2 concrete. With an average error of 1.24 MPa, or 3.43%, the ANN model showed excellent performance, showing a strong correlation with both experimental data and validation mixes. In addition to advancing predictive modeling for sustainable materials, this research opens the door for CO_2 Concrete to be more widely used in conventional construction methods, providing a viable path toward lessening the environmental impact of the sector.

8. Ehsan Mansouri, Maeve Manfredi, & Jong-Wan Hu (2022) - have studied the environmental impact of traditional concrete. As it becomes increasingly evident, there is a growing imperative to develop eco-friendly alternatives. Geopolymers, utilizing alumina-silicate waste materials as a binder activated by alkali, emerge as a promising solution for sustainable construction practices. Mansouri et al. contribute to this area by employing a three-step ML approach to predict the compressive strength of GPC. Their study utilizes CatBoost regressors, extra trees regressors, and gradient boosting regressors on a dataset comprising 147 green concrete samples and four variables. In addition to analyzing 84 experiments from the literature, they construct and test 63 new geopolymer concretes. The performance evaluation of these models using various metric indices demonstrates high accuracy in predicting compressive strength. Notably, the hybrid model showcases a 13% improvement in prediction accuracy, highlighting the efficacy of combining multiple ML approaches. This research contributes to advancing environmentally friendly concrete technologies, offering insights into optimizing material composition and predicting mechanical properties for sustainable construction practices.

9. Ayaz Ahmad, Waqas Ahmad, Krisada Chaiyasarn, Krzysztof Adam Ostrowski, Fahid Aslam, Paulina Zajdel, & Panuwat Joyklad (2021) - contributed to the advancement of GPC technology by exploring the application of supervised ML algorithms for predicting compressive strength. GPC presents a promising solution for sustainable development in civil engineering, emphasizing its role in mitigating environmental threats. Their study focuses on employing ANN, boosting, and AdaBoost ML approaches to predict the compressive strength of high-calcium fly-ash-based GPC. Results indicate that ensemble ML techniques, particularly boosting and AdaBoost, outperform individual methods like ANN, achieving high R^2 and lower errors. The research underscores the potential of ensemble ML techniques in enhancing the accuracy of GPC compressive strength prediction, contributing to the innovative environment in civil engineering.

10. Yang Song, Jun Zhao, Krzysztof Adam Ostrowski, Muhammad Faisal Javed, Ayaz Ahmad, Muhammad Ijaz Khan, Fahid Aslam, & Roman Kinasz (2021) - have driven towards eco-friendly practices in the concrete industry, which has spurred interest in utilizing waste materials like fly ash. However, traditional experimental methods for assessing concrete properties are time-consuming. Their study utilizes ensemble machine learning modelling techniques, such as bagging and boosting, alongside individual learners like multilayer perceptron neuron networks (MLPNN) and decision trees (DT). By analyzing a dataset of 471 data points, their research demonstrates the efficacy of ensemble modeling in producing robust correlations and accurate predictions, offering a promising avenue for accelerating the evaluation of concrete mechanical properties.

11. Ayaz, Krzysztof Adam Ostrowski, Mariusz Maślak, Furqan Farooq, Imran Mehmood, & Afnan Nafees (2021) - contributed to the understanding of concrete behavior under high temperatures by exploring supervised ML algorithms for predicting compressive strength. Their study responds to the challenges posed by high-temperature conditions, which can significantly impact concrete strength. Employing decision trees, ANNs, bagging, and gradient boosting, they evaluate the performance of these ML models based on a dataset comprising nine input parameters and one output

variable. The research highlights the efficacy of ensemble algorithms and gradient boosting in achieving strong correlations between predicted and actual outcomes, underscoring the potential of ML techniques in predicting concrete behavior under extreme conditions.

12. Tuan Nguyen-Sy, Jad Wakim, Quy-Dong To, Minh-Ngoc Vu, The-Duong Nguyen, & Thoi-Trung Nguyen (2020) - Concrete’s compressive strength is a critical parameter in structural design and construction. Nguyen-Sy et al. address the challenge of accurately predicting this strength based on concrete compositions and age, employing the extreme gradient boosting regression (XGB) method. Their study contributes to the existing literature by comparing the efficacy of XGB with other popular ML techniques such as ANN and SVM. Utilizing a comprehensive laboratory dataset, the authors demonstrate that all three ML methods yield accurate predictions. However, they highlight the superior performance of the XGB method in terms of robustness, training speed, and accuracy compared to ANN, SVM, and other existing ML approaches documented in literature. This research underscores the potential of XGB as a powerful tool for concrete strength prediction, offering valuable insights for engineers and researchers in optimizing construction materials and processes.

13. S. J. Goutham & V. P. Singh (2020) - contributed to the field of structural health monitoring by leveraging non-destructive testing methods for assessing concrete strength. Their study focuses on utilizing support vector regression, a machine learning technique, to predict compressive strength. By employing multiple non-destructive testing methods such as rebound hammer, Windsor probe penetration, and ultrasonic pulse velocity, they enhance the accuracy of the prediction model. The comparison of various combination models using statistical parameters confirms the efficacy of support vector regression in accurately predicting concrete compressive strength. The study underscores the potential of machine learning approaches in advancing structural assessment methodologies.

14. Fahid Aslam, Furqan Farooq, Muhammad Nasir Amin, Kaffayatullah Khan, Abdul Waheed, Arslan Akbar, Muhammad Faisal Javed, Rayed Alyousef, & Hisham Alabdulijabbar (2020) - delved into the realm of high-strength concrete (HSC) by employing machine learning and artificial intelligence approaches to predict its mechanical behavior. Their study emphasizes the importance of efficient experimental design in achieving target strength for HSC. By utilizing various input parameters such as cement, water, fine aggregate, coarse aggregate, and superplasticizer, they propose empirical relations with mathematical expressions through gene expression programming. Statistical analysis, including MAE, RRMSE, and RSE, evaluates model efficiency, highlighting the potential of machine learning in accurately estimating material quantities in civil engineering. The study underscores the significance of deep learning techniques in enhancing prediction accuracy for high-strength concrete properties.

15. Kaloop et al. (2020) – has focuses on utilizing the GBM learning technique for CCS prediction, alongside feature extraction using Multivariate Adaptive Regression Splines (MARS). The research compares the efficacy of different models, including kernel ridge regression and Gaussian process regression, in predicting CCS. With a dataset of 1030 samples and eight input parameters, the study underscores the importance of concrete age as a highly sensitive predictor and demonstrates the superiority of the integrated MARS-GBM approach in accurately predicting CCS.

III. METHODOLOGY

This study uses a predictive methodology to examine concrete's workability and compressive strength in relation to different mix design parameters. The approach is designed to create connections between input variables and the concrete properties that are produced by combining experimental processes with data-driven modeling techniques.

The study starts by designing and making several concrete mixes with different water-to-cement ratios, aggregate compositions, cement contents, and admixtures. According to IS code, the compressive strength (usually at 7 and 28 days) and workability (mainly through slump tests) are measured using standard testing procedures. The collection of accurate and consistent data in a controlled laboratory setting is guaranteed by this experimental phase.

Predictive models are created using statistical and machine learning methods after the data has been collected. To determine the most accurate technique for predicting compressive strength and workability, algorithms like artificial neural networks (ANN), and decision trees.

This methodology makes it easier to compare the outputs of predictive models with observed experimental results. It seeks to show how predictive techniques can enhance the precision of concrete property estimation in a range of structural applications while also saving time and material consumption in mix design optimization.

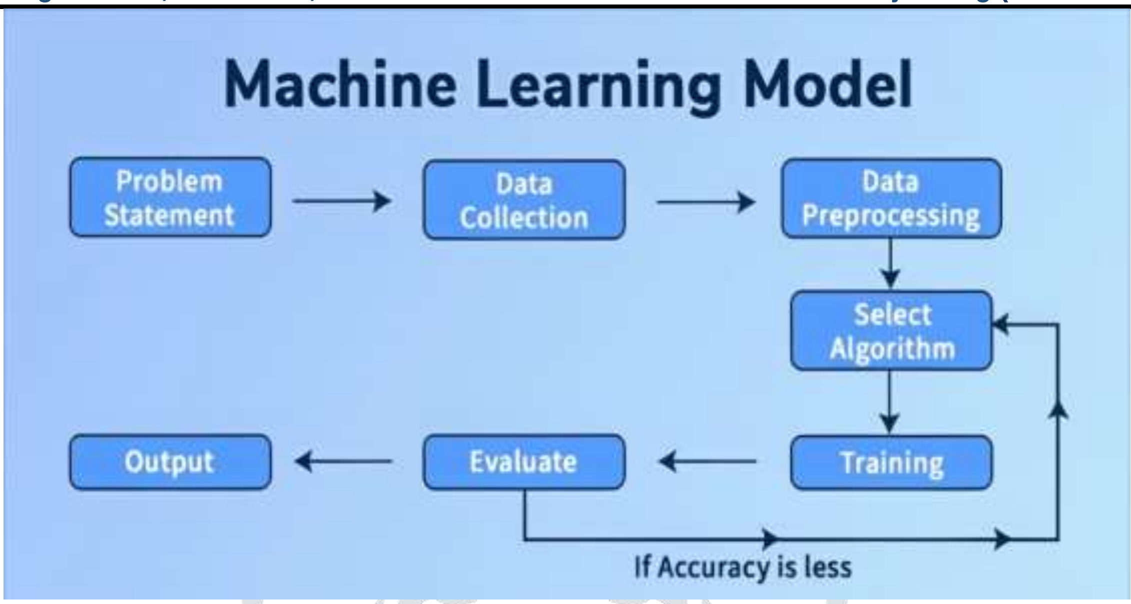

Fig. 3.1 Machine Learning Model

Figure 3.1 gives the detailed step-by-step overview of the machine learning model and how it functions, as explained below.

3.1 Functional Requirement- (Dataset Preparation and Pre-Processing)

- Collection of Dataset: The process of gathering, quantifying, and evaluating precise research insights using accepted, validated methods is known as data collection. Based on the information gathered, a researcher can assess their hypothesis. Regardless of the research field, gathering data is typically the first and most crucial stage in the process. Depending on the information needed different fields of study use different data collection strategies. Making sure that trustworthy and information-rich data is gathered for statistical analysis is the most important goal of data collection in order to enable data-driven research decisions.

- Data Visualization: Graphical representations are used in data visualization to communicate data and information. These tools, which make use of visual components like charts, graphs, and maps, provide an easy way to see and understand patterns, trends, and outliers in datasets, making data interpretation and analysis simpler.

- Data Labeling: Supervised ML involves training a predictive model using past data that includes predefined target outcomes. Algorithms require guidance on which attributes or target outcomes to identify within a dataset, a process known as labeling. Labeling datasets is labor-intensive and time-consuming, especially when thousands of records are needed for ML, particularly when ML requires thousands of records.

- Data Selection: The process of choosing the right data type, source, and tools for data collection before the actual data collection process begins is known as data selection. This definition distinguishes between interactive/active data selection, which uses collected data for monitoring tests or secondary research, and selective data reporting, which eliminates un-supporting data. Data integrity can be greatly impacted by the choice of relevant data for a study. To solve the specified problem, a data analyst chooses a subset of data after all the information has been gathered.

- Data Processing: Preparing raw data for analysis requires data pre-processing, especially for real-world datasets that are prone to errors, inconsistencies, and incompleteness. Preparing the data for ML algorithms is intended to increase the accuracy of the outcomes. Preprocessing involves steps like formatting, cleaning, and sampling. Formatting ensures consistency across attributes, especially when dealing with diverse data sources. Cleaning addresses noise, inconsistency, and missing values through imputation techniques, while outlier detection is crucial for maintaining model performance. Irrelevant or incomplete data objects are removed, and in cases involving sensitive information, anonymization may be necessary for privacy and regulatory compliance. Effective data pre-processing is key to developing robust and accurate concrete strength prediction models using ML algorithms.

- Data Transformation: The process of changing data from one format to another is known as data transformation. For tasks like data management and integration, data transformation is essential. Depending on the requirements of the project, data transformation can involve a variety of tasks, such as converting data types, cleaning data by eliminating nulls or duplicate data, enriching the data, or performing aggregations. Scaling numerical attributes in data can span various ranges, such as milligrams, grams, and kilograms. Converting these attributes to the same scale, e.g., between 0 and 1 or between 1 and 10 for the smallest and biggest value for an attribute, is known as scaling.

- Data Spitting: To create precise predictive models, the dataset must be divided according to pertinent characteristics like the type of concrete, admixture type, aggregate shape, and cement grade. Data splitting is the process that makes it possible to create subsets that represent the variation in concrete properties linked to various factors. We can examine how differences in cement quality affect concrete strength by dividing the dataset based on cement grade. The physical characteristics and chemical makeup of various cement grades can affect the concrete's overall strength. Similarly, we can investigate the impact of aggregate geometry on concrete strength by segmenting the dataset according to aggregate shape. Concrete's mechanical properties can be impacted by aggregates of various shapes, such as rounded or angular, which can change how the particles pack and interlock within the matrix. Another crucial element to take into account when dividing the dataset is the type of concrete. Strength predictions can be greatly impacted by the distinct compositions and performance characteristics of various concrete mixtures, such as lightweight or high-strength concrete.

3.2 Algorithm Selection

This section provides a comprehensive review of the most frequently used machine learning algorithms for predication of compressive strength:

- Linear regression - Linear regression (LR) is a fundamental statistical technique used for modeling the relationship between a dependent variable (often denoted as Y) and one or more independent variables (often denoted as X). In essence, LR aims to find the "best fitting" straight line (or hyperplane in higher dimensions) that represents the relationship between the variables.
- Random Forest - In ML, Random Forest is a potent ensemble learning method that can be applied to both classification and regression problems. In order to predict the individual trees, it builds a large number of decision trees during training and outputs the class mode.
- Deep Neural Network - One kind of artificial neural network (ANN) that can learn intricate patterns and representations from data is a Deep Neural Network (DNN), which is made up of several layers of connected neurons. The word "deep" describes the network's depth, which is the number of hidden layers that lie between the input and output layers.
- Support Vector Regression - A ML algorithm called Support Vector Regression (SVR) is used for regression tasks, especially when working with continuous data. The Support Vector Machine (SVM) algorithm, which is mainly employed for classification tasks, is extended by SVR.

3.3 Training and Evaluation of Dataset

Training a ML model for prediction of compressive strength using Python typically involves several steps:

- Data Preprocessing - The task of this step includes Cleaning and getting the dataset ready for training. It could involve dividing the data into training and testing sets, scaling numerical features, handling missing values, and encoding categorical variables.
- Feature Selection and Engineering - Evaluate and select the most pertinent characteristics that significantly affect the prediction of compressive strength. Feature engineering can also be used to develop new features that could enhance the model's functionality.
- Model Selection - Select a machine learning algorithm that is suitable for tasks involving regression. Neural networks, support vector regression, decision trees, random forests, and linear regression are popular options. The size and type of the dataset, available computing power, and the intended model interpretability all influence the algorithm choice.
- Training the Model - Use the training dataset to train the model after the algorithm has been chosen. In order to learn the relationships between the input features and the target variable (compressive strength), the model must be fitted to the training data.
- Evaluation - Use the testing dataset to assess the model's performance after it has been trained. Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) score are frequently used evaluation metrics for regression tasks. These metrics shed light on the model's predictive accuracy and ability to generalize to new data.

Below is an example Python Code for training a simple linear regression model for compressive strength prediction. In this code:

- “concrete_data.csv” is the dataset containing features and target variable.
- We split the dataset into features (X) and target variable (Y).
- We split the data into training and testing sets using train_test_split.
- We initialize a linear regression model using Linear Regression.
- We train the model using the training data.
- We make predictions on the test set and evaluate the model's performance using Mean Squared Error.

```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# Load the dataset
data = pd.read_csv('concrete_data.csv')

# Split the dataset into features (X) and target variable (y)
X = data.drop(columns=['Compressive_Strength'])
y = data['Compressive_Strength']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize the linear regression model
model = LinearRegression()

# Train the model
model.fit(X_train, y_train)

# Make predictions on the test set
predictions = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, predictions)
print("Mean Squared Error:", mse)
```

Fig. 3.2 Training of Dataset

3.4 Modeling and Deployment

Modeling of a trained dataset involves the concept and advancement of a descriptive or predictive model using the data that has been previously trained. This procedure usually entails adjusting the model's parameters, enhancing its functionality, and validating its accuracy on unseen data.

Here's an outline of the procedures needed to model a trained dataset is provided below:

- Model Selection - Depending on the problem's nature, the kinds of data, and the intended result, select the best ML algorithm. Regression models, classification algorithms, decision trees, neural networks, & ensemble approaches are popular options.
- Parameter Tuning - To maximize the performance of the selected model, modify its hyperparameters. Hyperparameters are settings, like learning rate, regularization strength, and tree depth, that regulate how the model behaves during training. Usually, methods like grid search, random search, or Bayesian optimization are used to tune parameters.
- Cross-Validation - To make sure the model is robust and generalizable; validate its performance using cross-validation techniques. In cross-validation, the data is divided into several subsets, one subset is used to train the model, and the other subsets are used to assess it. This procedure aids in identifying over fitting and evaluating the model's consistency across various data subsets.
- Model Evaluation - Depending on the problem type and model selected, assess the model's performance using suitable evaluation metrics, such as accuracy, precision, recall, F1-score, Mean Squared Error (MSE), or R-squared (R2) score.
- Model Interpretation - Interpret the trained model to gain a better understanding of the relationships between the input features and the target variable. Techniques like feature importance analysis, partial dependence plots, and SHAP (SHapley Additive exPlanations) values can help you understand how individual features affect the model's predictions. Deployment is the process of making a trained model usable in real systems or applications.
- Serialization - The trained model should be serialized into a portable format so that it can be readily loaded and stored by other systems or applications. Pickle, joblib, and TensorFlow's Saved Model format are examples of popular serialization formats.

IV. RESULTS AND DISCUSSIONS

In this important part, we examine the data we gathered and analyze; paying special attention to the predictive model built using ML. Furthermore, the model's predictor variables account for roughly 91% of the variance in compressive strength, according to the coefficient of determination (R2) of 0.91. It's imperative to remember to that the study's dataset size was limited, which might have hindered our ability to completely understand the complexities of simple regression models. However, despite this limitation, the dataset utilized in this investigation was rated sufficient to outfit a better understand of the suggested methodology. Additionally, we provided examples of predicted compressive strength values alongside their corresponding actual values from the dataset, as shown in figure below and in Table 4.1. These examples underscore the model's ability to make precise predictions, thereby validating its effectiveness in real-world scenarios.

Fig 4.1 Actual Strength (1)

Fig 4.2 Predicted Strength (1)

Fig 4.3 Actual Strength (2)

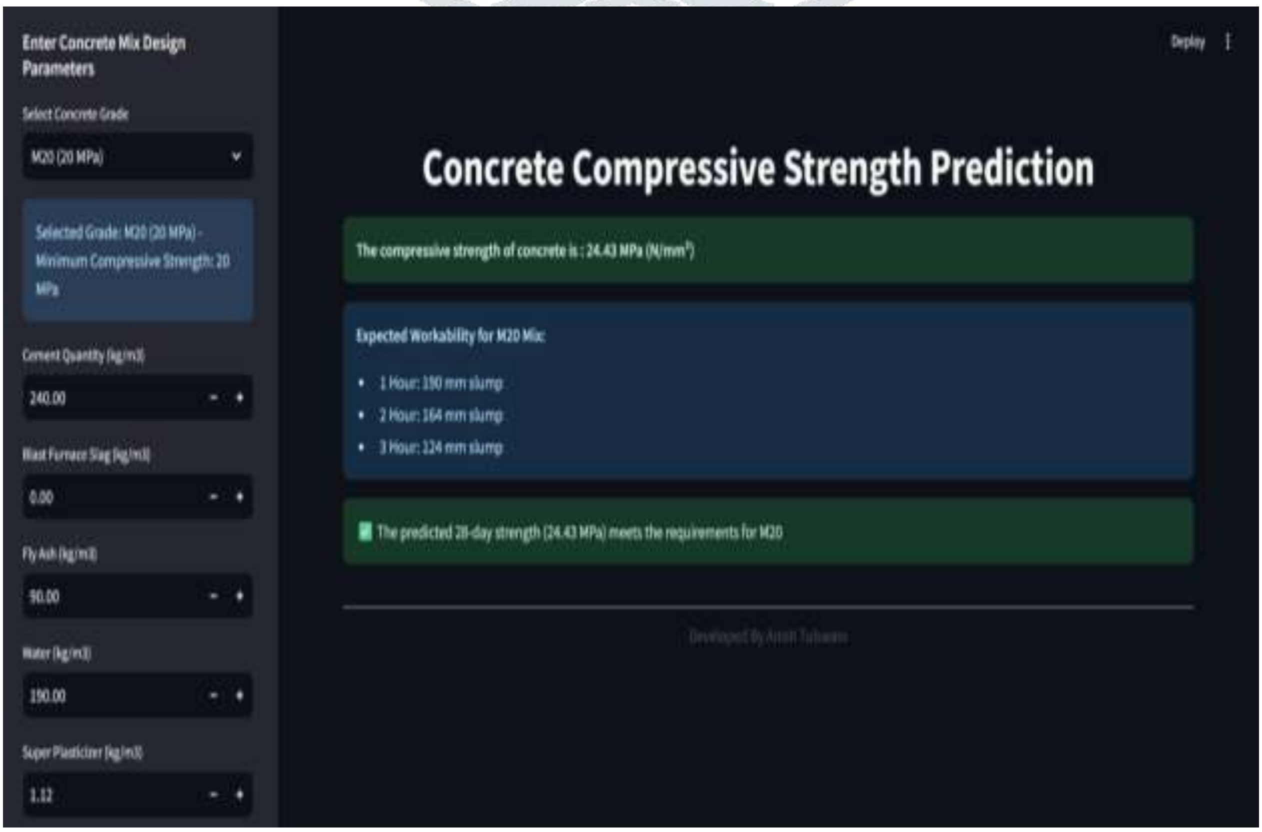

Fig 4.4 Predicted Strength (2)

Fig 4.5 Actual Strength (3)

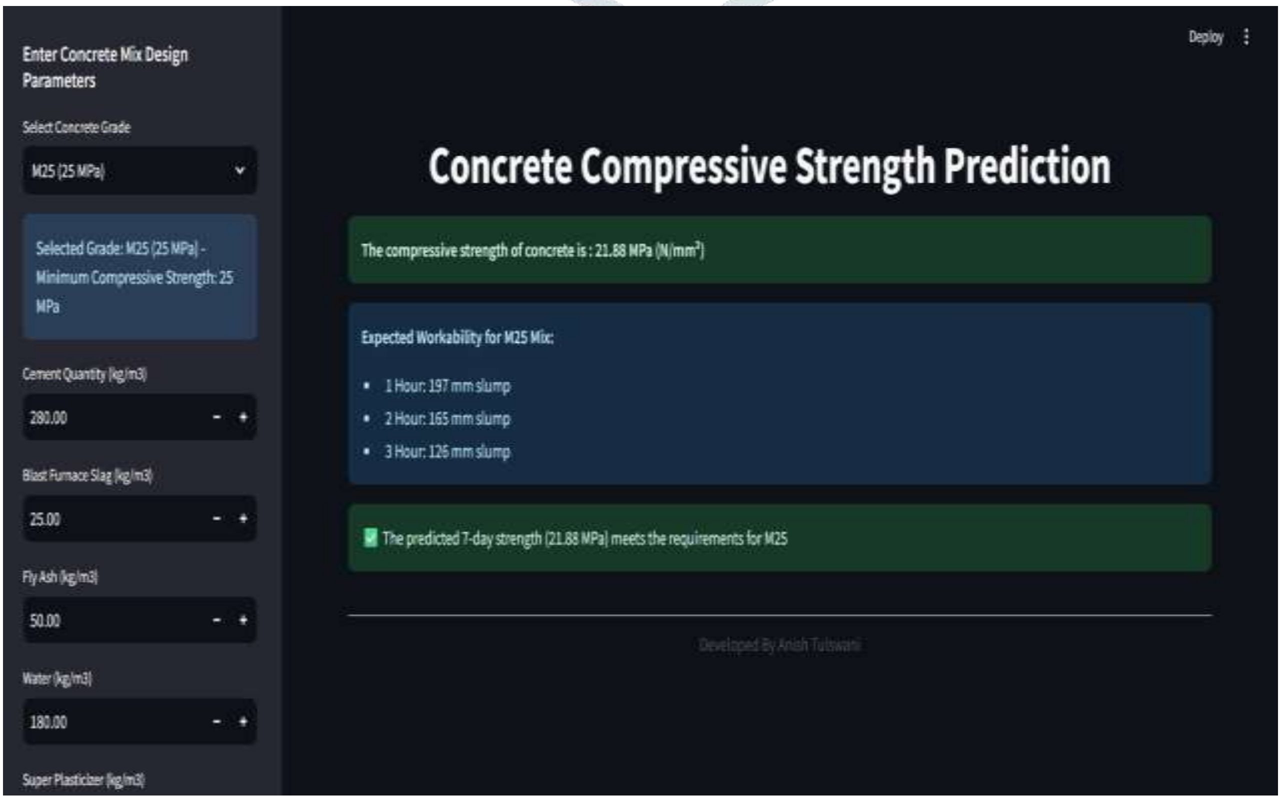

Fig 4.6 Predicted Strength (3)

Table 4.1 Actual v/s Prediction Examples

Sample No.	Actual Strength (MPa)	Predicted Strength (MPa)	% Error
1	28.75	28.81	- 2.08 %
2	25.33	24.43	+ 3.55 %
3	22.97	21.88	+ 4.74 %

4.1 Graphical Analysis

In this section, we present a comprehensive graphical analysis of the predictions made by the skilearn neural network model for concrete compressive strength. The graphical representations include graphs of actual versus predicted compressive strength and histograms to visualize the distribution of prediction errors.

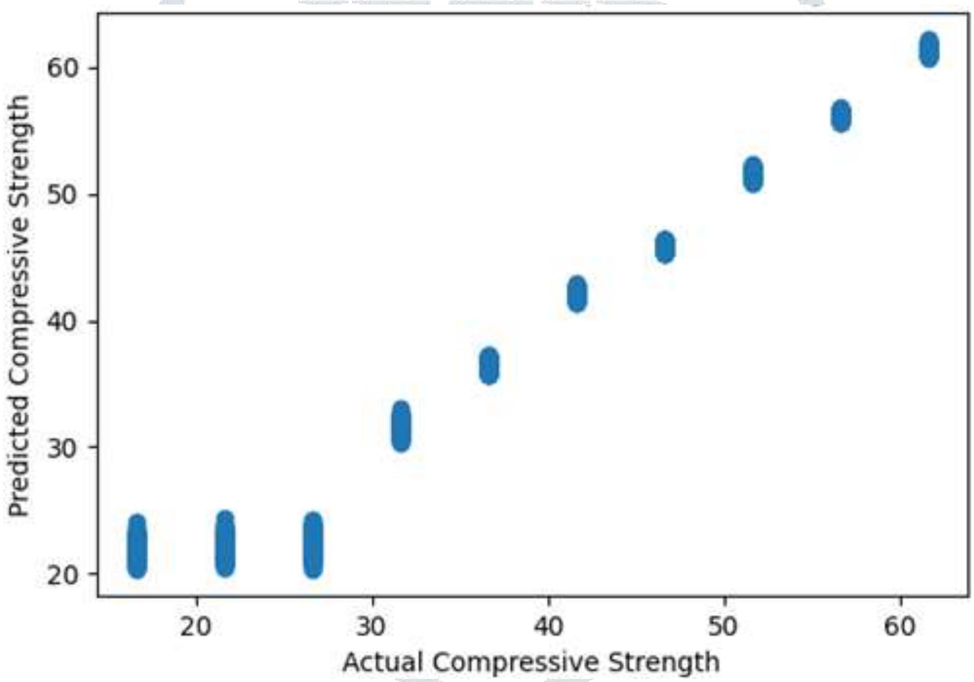

Fig. 4.7 Actual v/s Predicted Compressive Strength

The scatter plot shows how the predicted compressive strength values produced by the machine learning model or analytical technique used in this research project relate to the actual compressive strength values derived from physical experiments or tests. The target values seen in the dataset and the matching predictions produced by the neural network model are visually compared in the actual versus predicted compressive strength graph. With the x-axis showing the actual compressive strength values and the y-axis showing the predicted values produced by the model, each data point represents a distinct sample from the dataset. Perfect prediction accuracy would be demonstrated by the model correctly identifying the underlying patterns in the data, as indicated by the data points perfectly aligned along the diagonal line (y=x). In areas where the model may overestimate or underestimate the compressive strength, deviations from this diagonal line indicate differences between the actual and predicted values .

The x-axis represents the actual compressive strength values, which serve as the ground truth or reference data. Usually, concrete samples prepared with known mix designs and curing conditions are subjected to standardized testing procedures in order to measure these values. Megapascals (MPa) or pounds per square inch (psi), two common units of pressure, are probably used to express the actual compressive strength. Alternatively, the predicted compressive strength values determined by the suggested model or method are displayed on the y-axis. The input variables or features used to train the model—which could include things like cement composition, water-to-cement ratio, aggregate characteristics, admixture dosages, and curing conditions—are what determine these predicted values. With the x-coordinate representing the actual compressive strength value and the y-coordinate representing the predicted compressive strength value for that specific instance, each data point on the scatter plot represents a single instance or sample. All data points should ideally fall exactly on the diagonal line (y=x) if the predictions were perfect, meaning that the predicted and actual values are exactly the same. However, because of the predictive model's limitations, inherent uncertainties, and data noise, some deviations from the diagonal line are actually expected.

The accuracy of the predictions is indicated by how close the data points are to the diagonal line. Greater agreement between the actual and predicted values is indicated by points nearer the line, whereas greater differences or errors in the predictions are indicated by points farther from the line.

4.2 Histogram of Residuals

A histogram is a visual depiction of a dataset's distribution. It is made up of bars, each of whose height indicates the frequency or count of observations that fall into a specific interval or bin. A histogram of residuals, which is a visual depiction of the distribution of the discrepancies between the actual and predicted compressive strength values derived from the predictive model or method, is displayed in the graph below.

The histogram below exhibits a multi-modal distribution characterized by a prominent central peak and two smaller side peaks. This pattern implies that the residuals contain several underlying components or distributions. The residuals grouped around zero or close to zero values are represented by the tallest peak in the middle of the histogram. This suggests that the actual compressive strength values closely match a sizable percentage of the model's predictions.

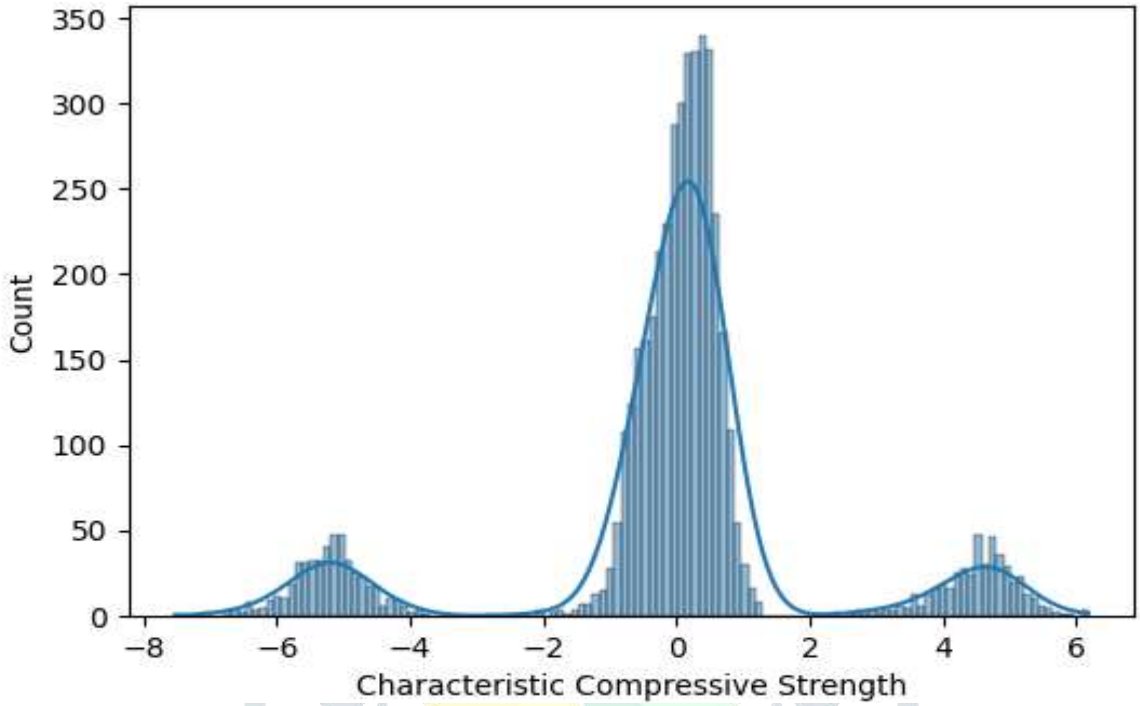

Fig. 4.9 Histogram of Residuals

Smaller peaks on the left and right sides of the central peak represent residuals that are more negative and more positive, respectively. The model's tendency to consistently overestimate or underestimate compressive strength values is indicated by these side peaks. The residuals' variability or uncertainty is reflected in the width of the central peak and side peaks. Broader peaks signify greater variability or uncertainty, whereas narrower peaks indicate more accurate and consistent predictions. The neural network model's advantages and disadvantages for predicting concrete compressive strength are highlighted by the histogram analysis. The presence of side peaks indicates areas for improvement to reduce systematic biases and improve predictive performance, even though the prominent central peak indicates generally accurate predictions.

Based on the observed histogram, the results suggest that the predictive model or method has an overall good performance, as indicated by the prominent central peak representing accurate predictions. The side peaks, on the other hand, show the times when the model tends to overestimate or underestimate the compressive strength values. To increase the model's predictive accuracy and consistency over the whole range of compressive strength values, these systematic biases or limitations could be looked into and fixed further.

V. CONCLUSIONS AND FUTURE SCOPE

5.1 Conclusions

- A predictive based methodology is effectively and precisely bridged to define the relationship between concrete's workability and compressive strength. The method can accurately predict strength at several curing ages (7 and 28 days) using regression-based models, explicitly including mix variables like w/c ratio, aggregate ratios, and chemical admixtures to measure workability effects.
- High predictive precision (± 5%) is demonstrated across a variety of mix designs by statistical models, such as non-linear regression, RSM, and design-of-experiments techniques, which achieve coefficient of determination values (R^2) exceeding 0.90 to 0.99.
- More sophisticated techniques like machine learning (ANN, SVR, GBDT, Light-GBM), is used to further improve predictive performance, particularly when using larger datasets or optimizing hyper parameters.

Overall, the predictive approach enables:

- A decrease in experimental trials, which saves money and time.
- Quantitative understanding of how mix proportions affect compressive strength and slump (workability).
- Accurate forecasting over a range of curing times, with particular resilience when predicting later strength based on early-age strength (e.g., 28 days).

The predictive approach offers a strong, empirically supported tool for concrete mix design optimization. By balancing workability and compressive strength, cutting down on pointless laboratory testing, and facilitating quicker development cycles in the study and use of concrete technology, thus improves structural performance.

This study concludes by demonstrating the efficacy of predictive techniques in estimating concrete's workability and compressive strength based on mix design variables. The models created provide accurate forecasts, negating the need for intensive laboratory testing. Strength and workability were clearly traded off, highlighting the significance of a well-balanced mix design.

5.2 Future Scope

The current study has shown how predictive methods, including statistical modeling and machine learning techniques, are useful for predicting and analyzing concrete's workability and compressive strength. Accuracy can be further improved in future studies by utilizing real-time data and more sophisticated algorithms. Even though, there is still a vast scope for advancement and research in this area. The areas listed below demonstrate the breadth of future study and use:

- Explore advanced feature engineering techniques to uncover additional predictors that may enhance the model's predictive power and robustness.
- Investigate advanced algorithmic approaches, such as deep learning architectures or gradient boosting techniques, to further improve model performance and generalization.
- Develop mechanisms for dynamically updating the model with new data to ensure continued relevance and accuracy as construction practices and materials evolve.
- Explore the integration of diverse data sources, including sensor data and material properties, to create a more comprehensive understanding of concrete behavior and performance.
- Develop real-time predictive analytics systems that can provide instantaneous feedback on concrete strength during production and construction processes, enabling proactive adjustments and quality control.
- Extend the model's capabilities to predict long-term concrete performance and deterioration, supporting proactive maintenance strategies and infrastructure asset management.
- Create cloud-based collaboration platforms where stakeholders can share data, insights, and best practices, fostering collaboration and knowledge exchange across the construction industry.
- Integrate uncertainty analysis techniques into the model to quantify and communicate uncertainties associated with predictions, enabling more informed decision-making.
- Explore opportunities to deploy the predictive model in emerging markets where access to advanced construction technologies and expertise may be limited, facilitating more efficient and sustainable construction practices globally.

REFERENCES

1. Ahmad, Ayaz, et al. "Comparative study of supervised machine learning algorithms for predicting the compressive strength of concrete at high temperature." Materials 14.15 (2021): 4222.
2. Ahmad, Ayaz, et al. "Prediction of geopolymer concrete compressive strength using novel machine learning algorithms." Polymers 13.19 (2021): 3389.
3. Amin, Muhammad Nasir, et al. "Application of soft-computing methods to evaluate the compressive strength of self-compacting concrete." Materials 15.21 (2022): 7800.
4. Aslam, Fahid, et al. "Applications of gene expression programming for estimating compressive strength of high-strength concrete." *Advances in Civil Engineering* 2020.1 (2020): 8850535.
5. Behnood, Ali, et al. "Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm." *Construction and Building Materials* 142 (2017): 199-207.
6. Chou, Jui-Sheng, and Chih-Fong Tsai. "Concrete compressive strength analysis using a combined classification and regression technique." *Automation in Construction* 24 (2012): 52-60.
7. Farooq, Furqan, et al. "A comparative study of random forest and genetic engineering programming for the prediction of compressive strength of high strength concrete (HSC)." *Applied Sciences* 10.20 (2020): 7330.
8. Golafshani, Emadaldin Mohammadi, and Ali Behnood. "Application of soft computing methods for predicting the elastic modulus of recycled aggregate concrete." *Journal of cleaner production* 176 (2018): 1163-1176.
9. Goutham, S. J., and V. P. Singh. "Artificial intelligence for compressive strength prediction of concrete." *IOP Conference Series: Materials Science and Engineering*. Vol. 1004. No. 1. IOP Publishing, 2020.
10. Gupta, Priyanka, Nakul Gupta, and Kuldeep K. Saxena. "Predicting compressive strength of geopolymer concrete using machine learning." *Innovation and Emerging Technologies* 10 (2023): 2350003.
11. Kaloop, Mosbeh R., et al. "Compressive strength prediction of high-performance concrete using gradient tree boosting machine." *Construction and Building Materials* 264 (2020): 120198.

12. Kewalramani, Manish A., and Rajiv Gupta. "Concrete compressive strength prediction using ultrasonic pulse velocity through artificial neural networks." *Automation in Construction* 15.3 (2006): 374-379.

13. Khademi, Faeze, et al. "Predicting strength of recycled aggregate concrete using artificial neural network, adaptive neuro-fuzzy inference system and multiple linear regression." *International journal of sustainable built environment* 5.2 (2016): 355-369.

14. Khashman, Adnan, and Pinar Akpinar. "Non-destructive prediction of concrete compressive strength using neural networks." *Procedia Computer Science* 108 (2017): 2358-2362.

15. Kochukrishnan, Sowmya, Premalatha Krishnamurthy, and Nandagopal Kaliappan. "Comprehensive study on the Python-based regression machine learning models for prediction of uniaxial compressive strength using multiple parameters in Charnockite rocks." *Scientific Reports* 14.1 (2024): 7360.

16. Kurhade, Swapnil Deepak, and Subhash Patankar. "Predicting compressive strength of geopolymer concrete using machine learning models." *Innovative Infrastructure Solutions* 10.1 (2025): 1-19.

17. Mansouri, Ehsan, Maeve Manfredi, and Jong-Wan Hu. "Environmentally friendly concrete compressive strength prediction using hybrid machine learning." *Sustainability* 14.20 (2022): 12990.

18. Mater, Yasser, et al. "ANN-Python prediction model for the compressive strength of green concrete." *Construction Innovation* 23.2 (2023): 340-359.

19. Naderpour, Hosein, Amir Hossein Rafiean, and Pouyan Fakharian. "Compressive strength prediction of environmentally friendly concrete using artificial neural networks." *Journal of building engineering* 16 (2018): 213-219.

20. Nguyen-Sy, Tuan, et al. "Predicting the compressive strength of concrete from its compositions and age using the extreme gradient boosting method." *Construction and Building Materials* 260 (2020): 119757.

21. Öztaş, Ahmet, et al. "Predicting the compressive strength and slump of high strength concrete using neural network." *Construction and building materials* 20.9 (2006): 769-775.

22. Ren, Qiubing, et al. "Prediction of rock compressive strength using machine learning algorithms based on spectrum analysis of geological hammer." *Geotechnical and Geological Engineering* 37.1 (2019): 475-489.

23. Song, Yang, et al. "Prediction of compressive strength of fly-ash-based concrete using ensemble and non-ensemble supervised machine-learning approaches." *Applied Sciences* 12.1 (2021): 361.

24. Tam, Vivian WY, et al. "A prediction model for compressive strength of CO2 concrete using regression analysis and artificial neural networks." *Construction and Building Materials* 324 (2022): 126689.

25. Varma, B. Vamsi, E. V. Prasad, and Sudhakar Singha. "Study on predicting compressive strength of concrete using supervised machine learning techniques." *Asian Journal of Civil Engineering* 24.7 (2023): 2549-2560.