International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-10, Issue-5, September 2022
https://doi.org/10.55524/ijircst.2022.10.5.15
Article ID IRPV1057, Pages 107-112
www.ijircst.org

Predicting the Concrete Properties Using Machine Learning-A Step Towards Smart Infrastructure

Sana Shabir Khan[1], Anuj Sachar[2], and Sandeep Singla[3]

[1]M.Tech Scholar, Department of Civil Engineering, RIMT University, Mandigobindgarh, Punjab, India
[2]Assistant Professor, Department of Civil Engineering, RIMT University, Mandigobindgarh, Punjab, India
[3]Professor, Department of Civil Engineering, RIMIT University, Mandigobindgarh, Punjab, India

Correspondence should be addressed to Sana Shabir Khan; sana00632@gmail.com

ABSTRAC- The mechanical properties of concrete mixtures are of a great concern when engineers need to provide the estimation of the concrete strength in addition to forecast the behavior of innovative concrete types. Predicting such mechanical properties like compressive strength, shear strength, tensile strength, and elastic modulus of concrete etc has motivated researchers to pursue reliable models for predicting mechanical strength. Empirical and statistical models, such as linear and nonlinear regression, have been widely used. However, these models require laborious experimental work to develop, and can provide inaccurate results when the relationships between concrete properties and mixture composition and curing conditions are complex. To overcome such drawbacks, several Machine Learning models have been implemented as an alternative approach for predicting the mechanical strength of concrete. The present study reviews ML models for forecasting the mechanical properties of concrete, including artificial neural networks, support vector machine, decision trees, etc. The application of each model and its performance has been discussed and analyzed.

KEYWORDS- Machine Learning (ML), Root mean square error (RMSE), Compressive Strength, Multivalued Linear Regression (MVLR), Artificial Neural Network (ANN), Decision Trees

I. INTRODUCTION

Rapid creation of newer concrete varieties, sparked by the construction industry's ever-increasing demands, has prompted more research into developing new predictive models that can calculate the characteristics of concrete. Predicting concrete's mechanical properties has been a major research project that could help designers meets the criteria of various design codes and standards. Traditional methods for predicting mechanical, rheological, durability, and other qualities of concrete were based on empirical relationships derived from statistical analysis of experimental data, using linear and nonlinear regression models. One of the drawbacks for using those empirical models is the high cost and time taken by trial batches required to create those models. Furthermore, traditional models struggle to deal with complex materials, making them unreliable for calculating the properties of various types of concrete. E.g., Chou et al., noted that due to the complex link between mechanical strength and mixture constituents, several features of high-performance concrete (HPC) cannot be easily predicted[1]. Furthermore, because the effects of new mixture elements are omitted in most existing empirical equations supplied in design codes and standards, their suitability to measure the strength of fresh concrete types with further new properties is questioned.

ML approaches have recently been introduced as a strong candidate for forecasting concrete mechanical strength to compensate for the shortcomings of traditional linear and nonlinear regression models. Such prediction methods can save time and money by eliminating the need for costly and time-consuming trial batches and accompanying experimental effort to obtain the specified concrete strength. There are two major types of machine learning approaches: supervised learning and unsupervised learning [2]. For estimating the mechanical properties of concrete, the former is more usually used. ML models for supervised learning are computer algorithms capable of creating patterns and hypotheses from a given dataset in order to estimate future values.

Despite the fact that different models have been presented to achieve the same purpose, namely the prediction of concrete mechanical strength, their structure and procedure can differ greatly. Artificial neural networks (ANN), support vector machines (SVM), decision trees, and evolutionary algorithms(EA) are the four basic types of machine learning approaches used to estimate concrete strength and each model's performance is assessed using a variety of statistical indicators. These different ML methods are used to forecast concrete's mechanical properties i.e., concrete strength, such as compressive strength, tensile strength, shear strength, and elastic modulus. Those models are usually used to a large dataset that is separated into subsets for training (TR), validation (VAL), and testing (TS). Model training is done with the training set. Validation data allows for an unbiased assessment of the model's fit on training data, as well as the prevention of model over fitting by halting the training process as the error grows. Finally, the model is tested on real-world data to see how well it predicts.

II. MACHINE LEARNING MODELS

Several statistical methods that characterize model fitting have been used to measure the performance of ML algorithms. Some of the statistical metrics for evaluating ML models are Correlation coefficient(R), Coefficient of determination (R2), Mean square error (MSE), Root mean square error (RMSE), Mean absolute error (MAE), Mean absolute percentage Error (MAPE) etc.
These statistical measures are used to evaluate the performance of machine learning approaches and can also be used to compare the efficacy of different algorithms.

A. Artificial Neural Network

Information propagation occurs via linkages that accept data from a processing element (neuron) and send it to subsequent neurons. Each piece of information is given a weight that reflects the importance of input variables to outputs [3]. When a neuron receives information, it uses a combination function to combine it with information from other neurons. After that, the combined data is sent to the following nodes. This iterative process is repeated until the algorithm accurately fits the data, as shown by the error rate convergence, or until the maximum number of iterations is reached [4,5]. The back propagation neural network (BPNN) approach has been widely employed by researchers to train ANN [6]. It is a local search strategy that updates the weights and biases of the ANN using learning algorithms such as gradient descent and Levenberg-Marquardt. The cost function, which commonly expresses the error between actual and expected strength, is minimized using this method. The compressive strength of high-performance concrete (HPC) was predicted using a back propagation neural network (BPNN) [7-9]. Concrete materials and the age of testing were used as input factors in the model. BPNN outperformed regression models in terms of accuracy, according to the results of the performance evaluation. For example, Compressive strength has been predicted in Ground granulated blast furnace slag concrete type using BPNN wherein Cement, blast furnace slag, super plasticizer, aggregates, water and age of samples were the inputs and R^2 as evaluation measure[10]. Similarly, Compressive strength of Silica fume concrete has been predicted using BPNN with inputs as Cement, amount of silica fume replacement, water content, amount of aggregate, plasticizer content, and age of samples and evaluation measures as Mean absolute relative error, MSE[11,12]. The results showed that BPNN was capable of accurately predicting compressive strength, outperforming empirically constructed models.

B. Decision Tree Classifier

The Decision Tree Classifier is a sequential structure, which combines nodes with targeted edges, which organizes a series of questions about predictions (attributes) and their possible answers. The various nodes used are internal nodes with two or more outbound but only one inbound edge, leaf or terminal nodes with one input but without outbound and root area with no input but two or more outbound. The attribute type determines the test status in this separator e.g. binary attributes produce two possible outcomes, but either the word attribute, the ordinal attribute or a continuous element includes the attribute values. A few criteria for selecting attributes or rules of segregation such as Information gain, profit rating, gini index etc., help determine how tuples in a particular area will be categorized.
Based on the training data the decision tree separator produces tree structure as shown in Figure 1 below:

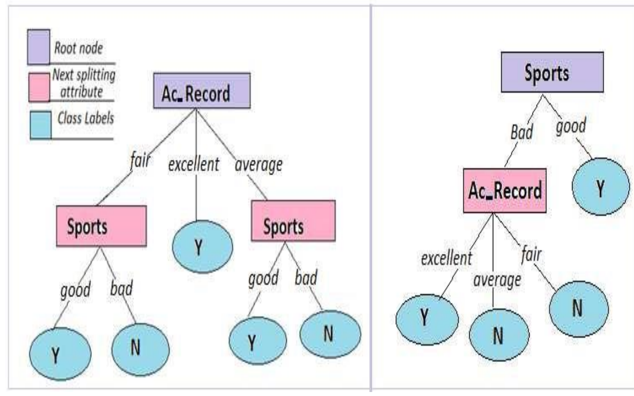

Figure 1: Tree Structure with splitting attribute at Root as (a) "Ac-Record" ; (b) "Sports"

C. Random Forest Classifier

The Random Forest Classifieris an ensemble model that works on this kind of technique wherein the predictions from multiple decision trees (base classifiers). Figure 2 shows the flow of data in Random Forest Classifier.

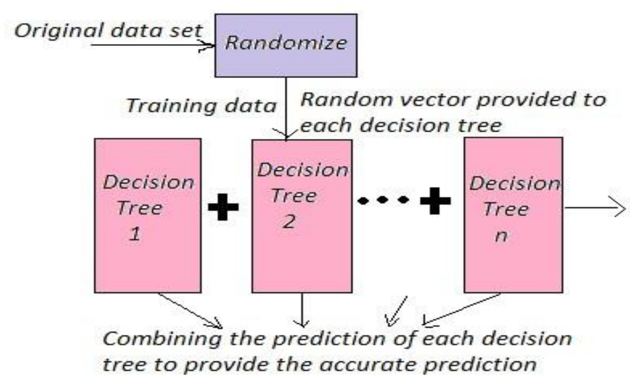

Figure 2: Random Forest Classifier

Random forest (RF) has also been used as a forecasting method in a number of studies. RF combines many decision trees, each of which is created from scratch using the bagging approach [13]. The bagging method, also known as bootstrap aggregation, is a two-step ensemble training method that includes bootstrap and aggregation. By randomly resampling the original set of data, identically distributed and independent datasets are formed in the first phase. The additional datasets are used to train the base predictors independently in the second stage. The aggregation approach is used to acquire results by averaging the predictions of each tree predictor. Several studies have utilized the RF to forecast the mechanical strength of concrete. Han et al. [14], for example, used RF to predict the compressive strength of HPC. Mangalathu et al. [15] used the same approach to predict the shear strength of RC beam-column joints in a previous work. Both research' findings were in good agreement, confirming RF's capacity to make accurate predictions. Zhang et al. [16] did another study in which they used RF to evaluate the uniaxial compressive strength of SCC.

D. Linear Regression

It is the most powerful tool in statistics wherein a model is created in which estimation is made for the value of predictor from a set of values and the error involved in estimation is measured. The focus is on the occurrence of minimum error in making predictions.

In the example of regression line (y=ax+b), which represents the linear form of regression, the prediction is calculated as:

$$\text{Prediction} = a * \text{Prediction} + b$$

We have another form of regression i.e. multivariate linear regressions (MVLR) that is the complicated form of linear regression. It consists of multiple predictors as shown below

$$\text{Prediction } Y = a1(x1) + a2(X2) + a3(X3) + b$$

Another model is of Non linear Regression form that can be of the form:

$$\text{Prediction } Y = a1(X) + a2(X2) + b$$

The logistic form of regression is where the model predicts the value in the form of just YES or NO. Logistic Regression is among widely used statistical technique for creating the predictive model.

III. SELECTION OF MODEL INPUTS

The selection of the most relevant features for training and evaluating the various ML models is critical to simplifying and improving the models' performance. Human expertise and experience, in addition to computational efforts, are required to pick the most appropriate parameters for running ML models. This allows for the precise selection of inputs that have a significant impact on concrete strength while avoiding parameters that have a minor impact, potentially saving calculation time. For forecasting concrete strength, several researches used common features. Binder content, aggregates, and mineral additives like fly ash and blast furnace slag, for example, have all been thoroughly incorporated Aggregates have a significant impact on the mechanical strength of concrete. The strength of concrete materials is influenced by the hardness, granular size distribution, and cleanness of aggregates. Due to the favorable effect of their pozzolanic characteristics and micro filler effect on the compressive strength of concrete, supplementary cementations materials such as fly ash, blast furnace slag, and silica fume are among the most widely integrated materials in concrete [17-21].

IV. EVALUATIONS AND EXPERIMENTATION

For the sake of prediction of concrete properties, several statistical and mathematical models have been implemented; however they are not reliable as the results are less accurate. Taking this into considerations, machine learning (Artificial Intelligent) algorithms are used wherein the predictions of different properties of concrete are being made based on the previous data/history (comprising of compositions / descriptors) that actually determine the properties of material. This shall in turn help in decreasing the necessity of performing physical experiments. This shall inturn save time and cost. The performance is being evaluated on the basis of evaluation metric as RMSE (Root Mean Squared Error) values and the properties to be mechanical properties viz. compressive strength, splitting etc.

RMSE is a function of the difference between the predicted value and the actual value for a particular data point. It is the standard deviation of errors and is a measure of the dispersion of these errors. It is calculated by taking the square root of the average of all the squared residuals or errors

A. Dataset Description

Data has been acquired from UCI repository and is available on Kaggle repository The dataset consists of 1030 instances with 9 attributes i.e. 8 input attributes and 1 output attribute. The input attribute values (Variable values) represent the amount of raw material in kg/m^3 and the output variable i.e. Compressive Strength contains the values represented measured in MPa. Concrete is the most important material in civil engineering. The concrete compressive strength is a highly nonlinear function of age and ingredients. These ingredients include cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, and fine aggregate. Figure 3 below shows the Concrete data set representation in WEKA software.

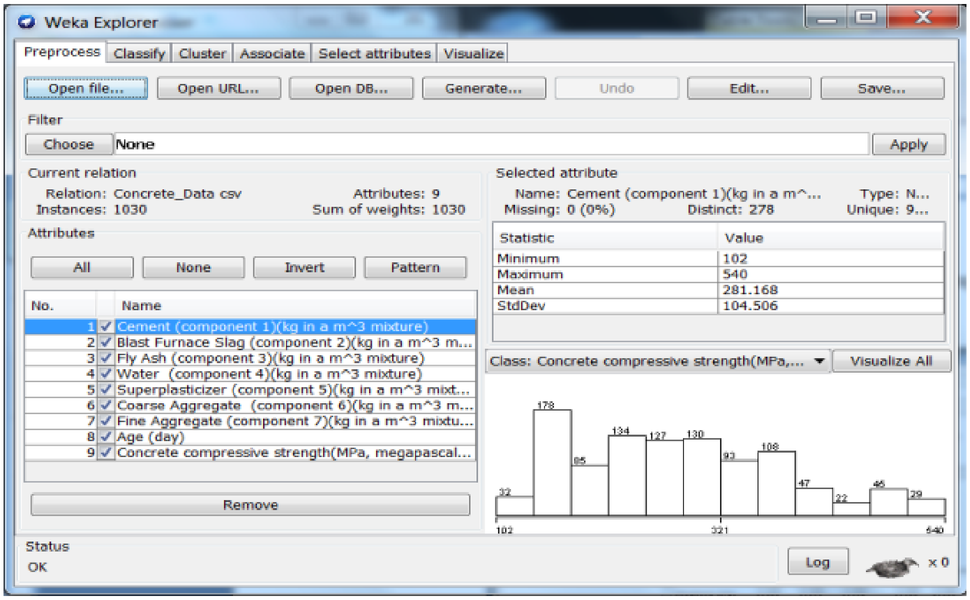

Figure 3: Data description of Concrete Dataset

Using Training sets and Tests sets in WEKA:
No. of instances : 1030
Number of Attributes: 9
@relation 'Concrete_Data csv'
@attribute 'Cement (component 1)(kg in a m^3 mixture)' numeric
@attribute 'Blast Furnace Slag (component 2)(kg in a m^3 mixture)' numeric
@attribute 'Fly Ash (component 3)(kg in a m^3 mixture)' numeric
@attribute 'Water (component 4)(kg in a m^3 mixture)' numeric
@attribute 'Superplasticizer (component 5)(kg in a m^3 mixture)' numeric
@attribute 'Coarse Aggregate (component 6)(kg in a m^3 mixture)' numeric
@attribute 'Fine Aggregate (component 7)(kg in a m^3 mixture)' numeric
@attribute 'Age (day)' numeric
@attribute 'Concrete compressive strength(MPa, megapascals) ' numeric
@data ……

In order to test the efficiency of our learning models we use training and test sets. In the supervised learning we provide the training set to build a learning model and further, we provide the test set so as to check the performance. For this we divide our dataset into two, usually disjoint, subsets. One subset will act as a training set and the other as test set. The training set, which is used to build a predictive model, consists of the predictor attributes as well as the prediction (class label) attribute.

The next step is to split the "concrete.arff" dataset into 40% testing set and 60% training set.

For this we use the WEKA filter – "Randomize "Filter so as to create a random permutation.

Further, another filter "Remove Percentage" is applied two times. First by keeping option "invert Selection" as 'false' and then 'true' so as to keep the 20% of the dataset saved as a test set and rest as the training set , respectively. By following these above steps, we get two datasets:

The "concretetraining.arff" with 50% of the instances in the original datasets

The "concretetTest.arff" with 50% of the instances in the original data

The number of instances in original dataset were 1030. We started by using the training set in the preprocess panel. This is followed by the selection of the particular algorithm we are concerned with. The 10 fold cross validation option is being selected. Next to use our sets in the experiments we choose the training set and move to the "Classify" panel and choose the procedure that we have to use and start the experiment. After that we apply the same procedure on our testing set to check what it predicts on the unseen data. For that, we select "supplied test set" and choose the testing dataset that we created.

B. Decision Tree-Based Technique

Applying J48 Decision Tree based Classifier on training data set as shown in figure and table below:

==Test Set Evaluation==

Table 1: Evaluation of J48 Classifier on test datset

Experiment using J48 Classifier	
Time taken to test model on supplied test set:	0.68 seconds
Correlation Coefficient	0.9175
Mean absolute error	5.1991
Root mean squared error	6.6911

C. *Random Forest Technique*

Applying Random Forest based Classifier on training data set as shown in table and figure below:

==Test Set Evaluation==

Table 2: Evaluation of Random Forest Classifier on test dataset

Experiment using Random Forest Classifier	
Time taken to test model on supplied test set:	0.35 seconds
Correlation Coefficient	0.9507
Mean absolute error	3.6956
Root mean squared error	5.2891

D. *Neural Network Based Technique*

Applying Multi-Layer Perceptron Classifier on training data set as shown in table and figure below:
==Test Set Evaluation==

Table 3: Evaluation of Multilayer Perceptron Classifier on test dataset

Experiment using Multilayer Perceptron Classifier	
Time taken to test model on supplied test set:	0.15 seconds
Correlation Coefficient	0.8713
Mean absolute error	6.7629
Root mean squared error	8.4936

Based on the experimental evaluations made, as in table 1 to 3, random forest has provided better results with MAE (mean absolute error) as 3.6956 and RMSE(Root mean Square Error) as 5.2891 as compared to the RMSE values resulted through Multilayer perceptron i.e. 8.4936 and that of J48 by 6.6911. Random forest, which makes predictions from multiple decision trees (base classifiers), gives the minimum RMSE value with 74.5% correctly classified instances.

E. *Evaluations Based on Previous study*

Evaluation based on the experimental data used on AASHTO Mechanistic-Empirical Pavement Design Guide (MEPDG) to analyze the effects of concrete constituent materials in its main machine and bad concrete structures has been performed [42]. The 10 feature set data includes samples of concrete with various proportions of the mixture including Air Content, Slump, force of pressure, strength, modulus of elasticity, Coefficient of thermal expansion (CTE), and Poisson ratio.
The models like MVLR (Multivariate Linear Regression), Decision tree, Radom forest has been applied for evaluation to check for the best predictor of concrete property like (elasticity, compressive strength, thermal expansion, tensile strength etc). Following table presents the RMSE values against each concrete property evaluated through various models. The data is based on 28 day evaluation and has been shown below in table 4.

Table 4: Evaluation of various models on test dataset for concrete properties

Concrete Properties ⇩	Algorithm Used in evaluation	MVLR	Decision Tree	Random Forest
Compressive Strength		0.67	0.98	0.73
Elasticity modulus		0.88	1.29	0.94
Poissons Ratio		0.94	1.29	0.96
Splitting Tensile Strength		0.85	1.18	0.87
Coefficient Of Thermal expansion		0.94	0.96	0.93

Based on the evaluations made on the data, using the Artificial Intelligent algorithms, Random-forest has given the better results in terms on RMSE value for each of the concrete property i.e. Compressive strength as 0.73 as compared to 0.67 by MVLR and 0.98 by Decision Tree (J48). Further, the elasticity modulus has also been proved better through Decision tree with RMSE value as 0.94 as compared to 0.88 by MVLR and 1.29 by Decision tree. The poisons Ratio, Splitting Tensile strength and coefficient of Thermal strength has been proved better through Random forest with RMSE values 0.96, 0.87 and 0.93 respectively.

V. CONCLUSIONS

The extracted knowledge should be appropriate enough to match the real outcomes. To ensure this, evaluation on the existing Machine Learning models has been performed for forecasting the mechanical properties of concrete including linear regression, artificial neural networks, decision trees, and random forest keeping in view various important evaluation measures like RMSE etc.
Several recent research have been undertaken to forecast the mechanical strength of concrete, examining the advantages and disadvantages of various methodologies. Forecasting the strength of complicated concrete mixtures using ML models (ANN, Linear Regression decision trees, and random forest) are used to address the issues of frequent error and construction time are the four major types of machine learning models studied in this thesis.
Following steps were followed:

- Understanding the Machine learning techniques for concrete data analysis.
- Understanding the features of concrete and their effect on the various properties like compressive strength.
- Study of the work that has been previously done in this research area with focus on all the extents.
- Comparing the results of various Machine Learning techniques taking into consideration various evaluation measure like RMSE.

- Implement and test the selected Machine Learning with special focus on data of constructions.
- Based on the evaluations made on the data, using the Artificial Intelligent algorithms, Random forest has given the better results in terms on RMSE values. In future, the models shall be evaluated based on various realistic datasets to be collected from different construction organizations in future.

REFERENCES

[1] J.S. Chou, C.F. Tsai, A.D. Pham, Y.H. Lu, Machine learning in concrete strength simulations: multi-nation data analytics, Constr. Build. Mater. 73 (2014) 771–780.

[2] K.P. Murphy, Machine Learning: A Probabilistic Perspective, 2012.

[3] M.A. Derousseau, J.R. Kasprzyk, W.V.S. Iii, Cement and concrete research computational design optimization of concrete mixtures: a review, Cem. Concr.Res. 109 (2018) 42–53.

[4] M. Bourdeau, E. Nefzaoui, X. Guo, P. Chatellier, Modeling and forecasting building energy consumption: a review of data-driven techniques, Sustain.Cities Soc. 48 (2019) 101533.

[5] M. Hemmat Esfe, S. Wongwises, A. Naderi, A. Asadi, M.R. Safaei, H. Rostamian,M. Dahari, A. Karimipour, Thermal conductivity of Cu/TiO2–water/EG hybrid nanofluid: experimental data and modeling using artificial neural network and correlation, Int. Commun. Heat Mass Transf. 66 (2015) 100–104.

[6] J. Xu, Y. Chen, T. Xie, X. Zhao, B. Xiong, Z. Chen, Prediction of triaxial behavior of recycled aggregate concrete using multivariable regression and artificial neural network techniques, Constr. Build. Mater. 226 (2019) 534–554.

[7] C. Deepa, K. SathiyaKumari, V.P. Sudha, Prediction of the compressive strength of high performance concrete mix using tree based modeling, Int.J. Comput. Appl. 6 (2010) 18–24.

[8] J. Chou, D. Ph, C. Chiu, D. Ph, M. Farfoura, I. Al-taharwa, Optimizing the prediction accuracy of concrete compressive strength based on a comparison of data-mining, Techniques 25 (2011) 242–253.

[9] S. Chithra, S.R.R.S. Kumar, K. Chinnaraju, F.A. Ashmita, A comparative study on the compressive strength prediction models for High Performance Concrete containing nano silica and copper slag using regression analysis and Artificial Neural Networks, Constr. Build. Mater. 114 (2016) 528–535.

[10] C. Bilim, C.D. Atis_, H. Tanyildizi, O. Karahan, Predicting the compressive strength of ground granulated blast furnace slag concrete using artificial neural network, Adv. Eng. Softw. 40 (2009) 334–340

[11] F. Özcan, C.D. Atis_, O. Karahan, E. Uncuoğlu, H. Tanyildizi, Comparison of artificial neural network and fuzzy logic models for prediction of long-term compressive strength of silica fume concrete, Adv. Eng. Softw. 40 (2009) 856–863.

[12] J. Chou, D. Ph, C. Chiu, D. Ph, M. Farfoura, I. Al-taharwa, Optimizing the prediction accuracy of concrete compressive strength based on a comparison of data-mining, Techniques 25 (2011) 242–253

[13] Q. Han, C. Gui, J. Xu, G. Lacidogna, A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm, Constr. Build. Mater. 226 (2019) 734–742.

[14] S. Chehreh Chelgani, S.S. Matin, S. Makaremi, Modeling of free swelling index based on variable importance measurements of parent coal properties by random forest method, Meas. J. Int. Meas. Confed. 94 (2016) 416–422.

[15] S. Mangalathu, J. Jeon, Classi fi cation of failure mode and prediction of shear strength for reinforced concrete beam-column joints using machine learning techniques, Eng. Struct. 160 (2018) 85–94.

[16] J. Zhang, G. Ma, Y. Huang, J. sun, F. Aslani, B. Nener, Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression, Constr. Build. Mater. 210 (2019) 713–719.

[17] P. Nath, P.K. Sarker, Effect of GGBFS on setting, workability and early strength properties of fly ash geopolymer concrete cured in ambient condition, Constr. Build. Mater. 66 (2014) 163–171.

[18] Y. Ayaz, A.F. Kocamaz, M.B. Karakoç, Modeling of compressive strength an UPV of high-volume mineral-admixtured concrete using rule-based M5 rule and tree model M5P classifiers, Constr. Build. Mater. 94 (2015) 235–240.

[19] Bhanu P. Koya, Sakshi Aneja, Rishi Gupta & Caterina Valeo, Comparative analysis of different machine learning algorithms to predict mechanical properties of concrete, Mechanics of Advanced Materials and Structures, (2021)

[20] Shunbo Zhao, Feijia Hu, Xinxin Ding, Mingshuang Zhao, Changyong Li, Songwei Pei, Dataset of tensile strength development of concrete with manufactured sand, Data in Brief,Volume 11, 2017,Pages 469-472

[21] Khan, Shabia Shabir, and S. M. K. Quadri. "Prediction of angiographic disease status using rule based data mining techniques", Biological Forum-An International Journal. Vol. 8. No. 2. 2016.