Advances and Applications in Mathematical Sciences
Volume 21, Issue 7, May 2022, Pages 4151-4161

PREDICTING COMPRESSIVE STRENGTH OF CALCINED CLAY, FLY ASH-BASED GEOPOLYMER COMPOSITE USING SUPERVISED LEARNING ALGORITHM

PRIYANKA GUPTA[1], NAKUL GUPTA[1*], SUDHIR GOYAL[1] and ANUSHREE[2]

[1]Department of Civil Engineering
GLA University, Mathura
Uttar Pradesh-281406, India

[2]Department of Computer
Engineering and Applications
GLA University, Mathura
Uttar Pradesh-281406, India

Abstract

This study employs random forest regression (RFR), decision tree (DT), and support vector regression (SVR) algorithm for the calcined clay (CC) - fly Ash (FA) geopolymer composite compression strength prediction. However, when compared to SVR, and DT, RFR yields a better result. The input variables comprise FA, calcined clay, and coarse aggregate; crushed stone dust as fine aggregate, water, and super plasticizers, alkaline solution, curing time, curing temperature. Models' performance was analyzed using statistical measure like mean absolute error (MAE), root square error (RSE), and root mean square error (RMSE). On one particular data sample, a random forest (RF), decision tree, and SVR were employed and compared. The overall data set comprises 75 data points, of which 80% were utilized for model testing and 20% were used for training models. The results suggest that the RFR is more accurate than the other two models used to predict compressive strength.

2020 Mathematics Subject Classification: 68T01.

Keywords: Fly ash, calcined clay, support vector regression, decision tree, random forest regression.
*Corresponding author; E-mail: nakul030588@gmail.com
Received September 25, 2021; Accepted November 22, 2021

1. Introduction

In the construction industry, waste utilization is an important factor to save the environment along with sustainable development. Due to India's reliance on coal-based electricity generation, there is a problem with FA disposal. [1]. Some 7% of worldwide CO2 release is attributed to the cement industry [2]. Geopolymer is made by combining aluminosilicate materials like FA, metakaolin, and steel slag with an alkaline liquid activator in a geopolymerization process [3]. There are two types of polymeric compounds used in industrial materials: crystalline and non-crystalline (amorphous orglassy structure) [4]. Several studies were performed using machine learning approaches in GPC. In one study RFR was used to estimate the strength due to the compression of manufactured sand concrete [5]. K-nearest neighbors (KNN) and RFR were machine learning tools used to know behavior information of laboratory values and results [6]. RFR is a set algorithm that gives more accuracy compared to gene expression programming (GEP) [7]. GEP and artificial neural networks (ANN) models were applied to a geopolymer, with 12M NaOH solution and water glass, an alkaline solution to FA ratio of 0.33. FA was partially exchanged with silica fume (SF) and GGBS along with the conventional material (river sand) [8].

In the present case, a combination of FA and calcined clay, molarities and curing temperature were used to find compressive strength. Figure 1 is describing parameters of the study. Three machine learning methods i.e., decision tree, RFR, and SVR used to test and trained data. But the results show that RFR is a more accurate method in terms of R2. May recent research applications are developed using machine learning classification models [9-14].

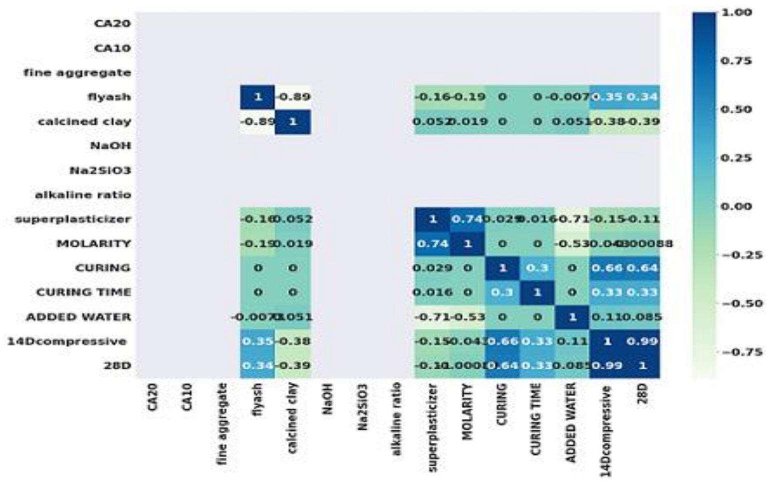

Figure 1. Description of Data used in different models.

2. Methodology of Research

Each machine learning method is detailed briefly in this section.

A. Decision tree

With numerical attributes, decision trees can be understood geometrically as hyperplanes, each orthogonal on the same axis. The Decision Tree (DT) is the model of a tree diagram for classification or regression [15]. If a change of probability at one chance node does not inherently need changes of probabilities in any other chance node, a DT can be separated [16].

B. Random Forest Regression

Dietterich first developed the idea of randomized node optimization in which the randomized approach opts for the decision at each node rather than a deterministic optimization [17]. In comparison to other machine learning algorithms, RF also handles big data sets more efficiently [18]. Figure 2 represents the RF model working.

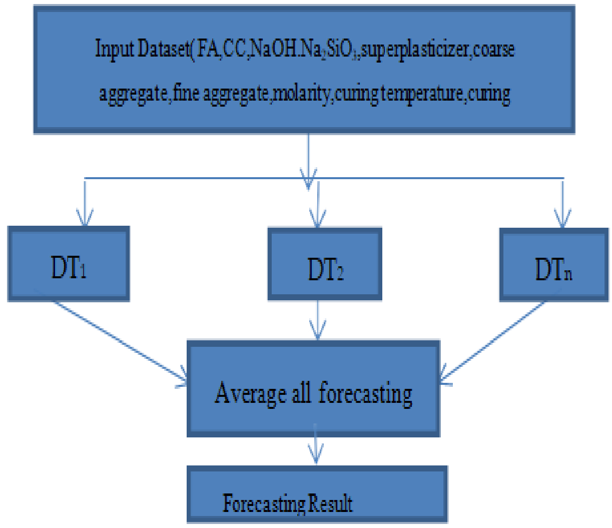

Figure 2. Random Forest Model.

C. Support Vector Regression

Support vector machines are a type of probability distribution that could be employed for classification and regression analysis refers to a statistical theory of learning [19]. The principle of SVR is to locate a hyper-plane to optimize the distance between the data points of two groups. The hyperplane dimension depends on how many characteristics the problem had. They influence both the location and the direction of the hyper-plane directly [20].

3. Model Development

In the current study RFR, SVR, and DT have been used to evaluate the compressive strength of FACC-based geopolymer composites. The following are elucidated fundamental aspects for the development of such models.

A. Data collection, extraction, and pre-processing functionality

Geopolymer Composite is formed using FA and CC as cementitious material with an alkaline solution. The compressive strength of the CCFA geopolymer composite has been studied. Input data used in the study were 8M,10M,12M, 14M and 16M of NaOH, Na2SiO3, ambient temperature, 80oC, and 100oC for curing, curing time of 24 and 48 hours, different % variation of

CC in FA, coarse aggregate passing 20mm and 10mm sieve, fine aggregate. A dataset of 15 attributes (14 data input features and 1 output) and 75 samples had been produced utilizing laboratory work to study this problem methodically. 100mm size cube was used to test the compressive strength of GPC.100mm size cube's compressive strength was tested in an automatic compression testing machine. 80% of data set was trained and 20% tested. For all generated machine learning algorithms, a specific 'random state' was assigned the Python package scikit-learn to be able to contain identical training and testing data. The relevance of the data illustrates how each characteristic helps the output prediction.

B. Performance assessment of the model

Different static measures were used to assess the performance of the implemented machine learning algorithms. The determination coefficient (R2 - value) was determined as well as, the root-mean-square errors (RMSE), the mean absolute error (MAE), Mean square error (MSE) were determined. These are the following metrics [15].

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | xi - x | \tag{1}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_{pred} - y_{ref})^2 \tag{2}$$

$$RMSE = \sqrt{\sum \frac{(y_{pred} - y_{ref})^2}{n}} \tag{3}$$

Where n is the total number of datasets, x and y_{ref} are data set references, x_i and y_{pred} are model values anticipated. The reflective practice shows the link between experimental and predicted results in the resulting value of the model.

4. Results and Discussion

Below is a description of the prediction performance of the (machine learning) ML (RFR, SVR, and DT) model used in this investigation, described in Section 2. A discussion of the prediction accuracy and the usefulness of the tuned algorithms are evaluated.

A. ML models used for Hyper-Parameter adjustment

All models were initially designed to obtain the best precision to prevent overpowering and to estimate strength properties. In the case of the RFR and SVR, DT designed in a random forest, the minimal split quantity necessary for everyone leaves node, the minimum sample number required for each leaf node, the number of characteristics to take into consideration irregular development of the tree, and the highest possible tree depth (max depth) of each tree [21]. In addition to the variables referred to above, the SVR model also has a considerable impact on model performance at the pace of learning indicating the participation of each tree. Identical data were used [22].

B. Machine Learning Models Predictive Performance

Figure 3 shows the parameters of strength due to compression for 28 days considered from laboratory work compared to forecasted values in the different modes of learning. It may be noted that all algorithms have learned to forecast strength due to compression with a like pattern in prediction against the observed output by use of the non-linear relationship. Figures 4 and 5 illustrate the rest of the anticipated output for training and testing data sets (i.e., compressive strength). The remainder of every instance is the gap in the anticipated number of the measured actual output numbers. Residuals varied slightly around 0 for the RFR models used, which indicates acceptable prediction ability.

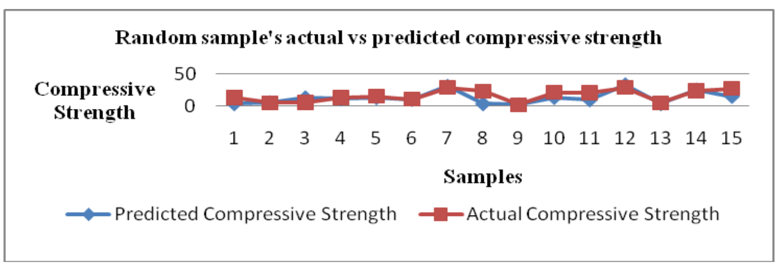

Figure 3. Compressive Strength for predicted and actual values of FA-CC GP.

In addition, the R2 value of the training and testing datasets is illustrated in figure 4 and figure 5 along with MSE, RMSE, and MAE. Out of all three models used, RFR models with a maximum value of R2 as 0.79 in testing and 0.93 in training were obtained. The residuals and the R2 value of the RFR and DT prediction model show an acceptable prediction for the compressive strength of geopolymer composite materials. Different statistical approaches as R2, MAE, RMSE, and MSE were used to better define model performance. A different statistical analysis, presented in section 3, was calculated for the predictive performances of the models. The statistical result of training and testing is shown in table 1. These low values in training imply a greater accuracy in forecasting compressive strength in the created machine learning model. The best predictive output exhibited by the DT model was its higher R2 value (1.0) and the lower RMSE and MAE amount as 0.0 and 0.0. It was shown that for CCFA GPC, RFR and SVR values of R2 were 0.79 and 0.317 respectively, the highest and lowest of the 3 models used. R2 is considered very poor if range less than 0.3, moderate lies between 0.3 to 0.5, acceptable if ranges are between 0.5 to 0.7, and excellent if ranges above 0.7 [23-25].

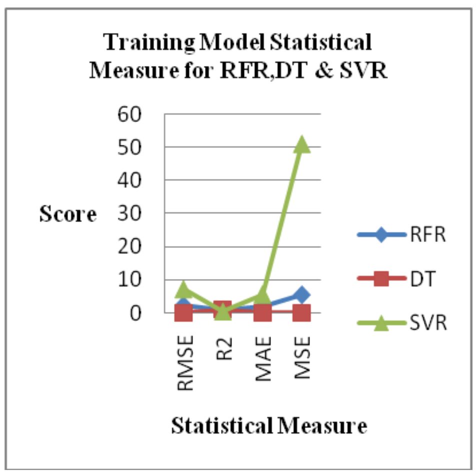

Figure 4. Statistical modeling of RFR, SVR, and DT for training.

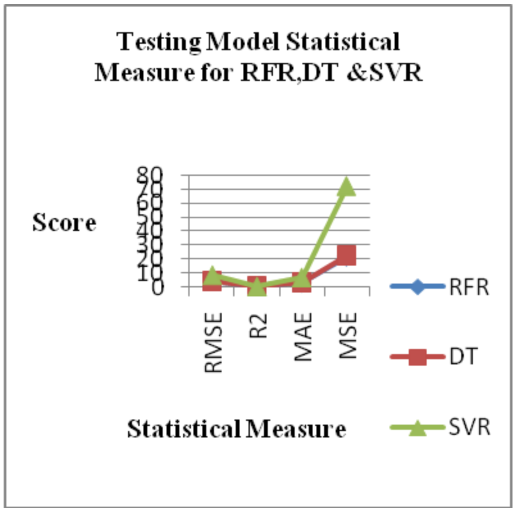

Figure 5. Statistical modeling of RFR, SVR, and DT for testing.

Table 1. Statistical result for models used in training and testing.

Model	Training				Testing			
	MSE	RMSE	R2	MAE	MSE	RMSE	R2	MAE
RFR	5.39	2.323	0.93	1.825	21.933	4.683	0.793	3.393
DT	0.0	0.0	1.0	0.0	22.73	4.768	0.786	3.47
SVR	51.08	7.147	0.408	5.56	72.584	8.519	0.317	6.843

5. Conclusions

This study examines the forecast of compressive strength of the machine learning model developed for CCFA geopolymer composites with actual laboratory values.

1. Out of the 3 models used RFR performed better.

2. The testing phase results of RFR were 0.79, 4.683, and 3.393 for R2, RMSE, and MAE indicating correctness by decreasing the error difference between targeted and anticipated values, according to statistical analytical checks.

3. The findings demonstrate that five variables influenced strength prediction: CCFA content, molarity, curing duration, and curing temperature.

4. As a result, the RFR model shows a significant strategy for forecasting geopolymer concrete compressive strength with temperatures variation, with the potential to develop other important geopolymer concrete characteristics.

5. Understand and forecast the compressive strength of the proposed CCFA geopolymer concrete.

6. Additionally, new laboratory data can be collected in the next studies to enhance the performance results of forecasts.

Abbreviations

Symbol	Full-Form	Symbol	Full-Form	Symbol	Full-Form
RF	Random Forest	KNN	K-Nearest Neighbors	R^2	R square
CCFA	Calcined Clay Flyash	MSE	Mean Squared Error	CO_2	Carbon Dioxide

References

[1] O. A. Abdul Kareem, A. M. Mustafa Al Bakri, H. Kamarudin, I. KhairulNizar, Ala'eddin A. Saif, Effects of elevated temperatures on the thermal behavior and mechanical performance of fly ash geopolymer paste, mortar, and lightweight concrete, Construction and Building Materials 50 (2014), 377-387. doi:10.1016/j.conbuildmat.2013.09.047

[2] V. Malhotra, Role of supplementary cementing materials in reducing greenhouse gas emissions, in 1999, doi:10.1201/9781482272215-23.

[3] P. Duxson, A. Fernández-Jiménez, J. L. Provis, et al., Geopolymer technology: the current state of the art. J. Mater Sci. 42 (2007), 2917-2933 doi.org/10.1007/s10853-006-0637-z.

[4] Davidovits, J. Geopolymers, Geopolymers Inorganic polymeric new materials, Journal of Thermal Analysis 37 (1991), 1633-1656. https://doi.org/10.1007/BF01912193

[5] H. B. Ly and V. Q. Tran, Estimation of compressive strength of concrete containing manufactured sand by random forest, International Journal of Scientific and Technology Research 9 (2020), 564-567.

[6] O. A. Mohamed, M. Ati and O. F. Najm, Predicting compressive strength of sustainable self-consolidating concrete using random forest, Key Engineering Materials 744 (2017), 141-145, doi.org/10.4028/www.scientific.net/KEM.744.141.

[7] Mohsin Ali Khan, Shazim Ali Memon, Furqan Farooq, Muhammad Faisal Javed, Fahid Aslam and Rayed Alyousef, Compressive Strength of Fly-Ash-Based Geopolymer Concrete by Gene Expression Programming and Random Forest, Advances in Civil Engineering 2021, 17 pages, 2021. doi.org/10.1155/2021/6618407

[8] Paul O. Awoyera, Mehmet S. Kirgiz, A. Viloria and D. Ovallos-Gazabon, Estimating strength properties of geopolymer self-compacting concrete using machine learning techniques, Journal of Materials Research and Technology 9 (2020) 9016-9028 doi.org/10.1016/j.jmrt.2020.06.008.

[9] V. Madaan and A. Goyal, Predicting Ayurveda-Based Constituent Balancing in Human Body Using Machine Learning Methods, in IEEE Access 8 (2020), 65060-65070, doi:10.1109/ACCESS.2020.2985717.

[10] R. Kaur, V. Madaan and P. Agrawal, Diagnosis of arthritis using k-nearest neighbor approach, International Conference on Advanced Informatics for Computing Research (2019), 160-171. https://link.springer.com/chapter/10.1007/978-981-15-0108-1_16.

[11] P. Agrawal, V. Madaan, A. Roy, R. Kumari and H. Deore, FOCOMO: Forecasting and monitoring the worldwide spread of COVID-19 using machine learning methods, Journal of Interdisciplinary Mathematics (2021), 1-25.

[12] C. Gupta, P. Agrawal, R. Ahuja, K. Vats, C. Pahuja and T. Ahuja, Pragmatic analysis of classification techniques based on hyperparameter tuning for sentiment analysis, International Semantic Intelligence Conference 2786 (2021), 453-459. http://ceur-ws.org/Vol-2786/Paper54.pdf.

[13] P. K. Verma, P. Agrawal, I. Amorim and R. Prodan, WELFake: word embedding over linguistic features for fake news detection, IEEE Transactions on Computational Social Systems 8(4) (2021), 881-893.

[14] A. Sharma, P. Agrawal, V. Madaan and S. Goyal, Prediction on Diabetes Patient's Hospital Readmission Rates, International Conference on Advanced Informatics for Computing Research, DOI-10, ACM ICPS Digital Library (2019), (https://dl.acm.org/citation.cfm?doid=3339311.3339349).

[15] P. Gupta, N. Gupta, K. K. Saxena and S. Goyal, A novel hybrid soft computing model using stacking with ensemble method for estimation of compressive strength of geopolymer composite, Advances in Materials and Processing Technologies (2021) doi:10.1080/2374068X.2021.1945271.

[16] A. Saltelli, S. Tarantola, F. Campolongo and Marco Ratto, Sensitivity analysis in practice: a guide to assessing scientific models, Wiley Online Library, 2004.

[17] T. G. Dietterich, An experimental comparison of three methods for constructing ensembles of decision trees: bagging, Boosting, and Randomization, Machine Learning 40 (2000), 139-157 doi.org/10.1023/A:1007607513941.

[18] Vladimir Svetnik, Andy Liaw, Christopher Tong, J. Christopher Culberson, Robert P. Sheridan and Bradley P. Feuston, Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling, Journal of Chemical Information and Computer Sciences 43(6) (2003), 1947-1958 doi:10.1021/ci034160g.

[19] K. Cheng and Z. Lu, Adaptive Bayesian support vector regression model for structural reliability analysis, Reliability Engineering and System Safety 206 (2021) 107286.doi/10.1016/j.ress.2020.107286.

[20] Vladimir N. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.

[21] Ly, Hai-Bang, Eric Monteiro, Tien-Thinh Le, Vuong M. Le, Morgan Dal, Gilles Regnier and Binh T. Pham, Prediction and Sensitivity Analysis of Bubble Dissolution Time in 3D Selective Laser Sintering Using Ensemble Decision Trees Materials 12(9) (2019), 1544 doi.org/10.3390/ma12091544.

[22] Zaher Mundher Yaseen, Ravinesh C. Deo, Ameer Hilal, Abbas M. Abd, Laura Cornejo Bueno, Sancho Salcedo-Sanz and Moncef L. Nehdi, Predicting compressive strength of lightweight foamed concrete using extreme learning machine model, Advances in Engineering Software. 115 (2018) 112-125.doi.org/10.1016/j.advengsoft.2017.09.004

[23] D. S. Moore, W. Notz and M. A. Fligner, The Basic Practice of Statistics, W.H. Freeman and Company, 2013.

[24] C. Gupta, A. Jain and N. Joshi, A Novel Approach to feature hierarchy in Aspect Based Sentiment Analysis using OWA operator, In Proceedings of 2nd International Conference on Communication, Computing and Networking Springer, Singapore (2019), 661-667.

[25] S. Taneja, C. Gupta, K. Goyal and D. Gureja, An enhanced k-nearest neighbor algorithm using information gain and clustering, In 2014 Fourth International Conference on Advanced Computing and Communication Technologies IEEE 8 (2014), 325-329.