NAU Journal of Civil Engineering (NAUJCVE) October/November/December 2024

Nnamdi Azikiwe University Journal of Civil Engineering (NAUJCVE)

Volume-2,Issue-3,pp-42-51

www.naujcve.com

Research Paper Open Access

Performance Evaluation of Different Machine Learning Models For compressive Strength Prediction of Laterized Portland Cement-Activated Pozzolanic Concrete

[1]Faluyi, Feyidamilola and [2]Arum, Chinwuba

[1]Department of Civil Engineering, Federal University Oye Ekiti, Nigeria.

[2]Department of Civil Engineering, Federal University of Technology, Akure, Nigeria

Corresponding Author: feyidamilola.faluyi@fuoye.edu.ng

Abstract

Activated pozzolans/ geopolymer concrete is no doubt one of the major attempts to make the world liveable and ensure sustainable construction through the reduction of carbon footprint. A correct prediction of the compressive strength is important for the acceptability and use of such concrete within a complex system. In this study, four machine learning models consisting of support vector regression (SVR), artificial neural network (ANN), extreme gradient boosting (XGBoost), and random forest (RF), were employed to predict the compressive strength of hybrid laterized ordinary Portland cement-activated rice husk ash (OPC-ARha) concrete using eight input variables. Data from 192 concrete specimens were used for training and testing in the ratio of 153:39. The R^2 values for SVR, ANN, XGBoost, and RF were 0.7172, 0.9496, 0.9200, and 0.9239 respectively, making ANN the model with the highest absolute fraction of variance (R^2). Also from the result of the root mean squared error (RMSE), ANN proved better than the other models. However, the least mean absolute error (MAE) was obtained using RF. Feature importance of the eight variables used in the strength prediction showed that curing age is the most important factor followed by Portland cement content and water/binder ratio.

Keywords: *Strength prediction, Machine learning, Activated rice husk ash, Laterized concrete, Concrete strength*

Date of Submission: 25-11-2024 Date of acceptance: 02-12-2024

1. INTRODUCTION

The discourse of sustainability in concrete construction is undeniably and imperatively connected to low-energy alternative binders such as Activated pozzolans/ geopolymers of which Activated Rice Husk Ash is one (Sagar and Sivakumar, 2020). Compressive strength prediction is valuable for cost reduction in construction. However, accuracy in predicting the strength of concrete is a fundamental issue when developing a prediction model. Prediction models based on machine learning are progressively becoming popular owing to their growing accuracy and computational capacity (Awoyera, 2020). Machine learning tools can carry out classification and regression and handle the intricate non-linear relationships that characterize concrete variables. According to Ghosh and Ransinchung (2022), such interaction between the dependent and independent variables in concrete is best handled by a non-linear model, corroborated by a high degree of accuracy. Pan *et al*(2022) demonstrated that support vector machines could be effectively employed in the strength prediction of concrete that utilizes recycled aggregate. de-Prado-Gil *et al*(2022) used various machine learning (ML) approaches to assess the split tensile strength of self-compacting concrete and found that ML performed very well in the strength assessment of concrete. According to Golafshani and

Kashani (2022), using a machine learning approach to model the compressive strength of concrete, we can have an insight into the effect of each variable on the properties of such concrete. The low mean absolute percentage error (MAE) of random forest and xgboost showed that the machine learning approach could be employed to accurately and reliably predict the compressive strength of concrete thereby reducing experimental cost (Xu*et al*, 2021; Mai *et al*, 2021). Experimental results from the use of ensemble ML techniques showed that XGBoost can be accurately used to predict the strength of high-performance concrete (Elhishi *et al*, 2023).

B. THEORETICAL ANALYSIS

Machine Learning Models

A. Support Vector Regression

Support vector regression (SVR) is a kind of SVM that uses a supervised learning algorithm for regression jobs. It is used to predict a continuous output value for an input values using a function. The prediction is based on the mapping function. SVR may use both non-linear and linear kernels. A linear kernel is simply between two input vectors, while a non-linear kernel is a more complex function that can capture more intricate patterns in the data. The characteristics of the data and the complexity of the job will determine the kernel to use in SVR. SVR has brilliant generalization competency coupled with high accuracy (Singh, 2023).In an ordinary SVR learning problem, a set of training data (D) is given by Equation (1) to generate a dependency between the input x and output y.

$$D = \{x_i , y_i\}_{i=1}^{N} \qquad (1)$$

From the set of training data D in Equation 1, the objective function of the SVR can be formulated as shown in Equation (2)

$$\text{Min.}\, J_p\,(w, e) = \frac{1}{2} w^T w\phi + c\sum_{i=1}^{N} e_i^2 \qquad (2)$$

Subject to;

$$y_i = w^T\phi(x) + b + e_i, i = 1, \ldots, N \qquad (3)$$

where *ei* is the error variable; c indicates the regularization constant.

In predicting the compressive strength of high-performance concrete (HPC) with radial basis function (RBF) and polynomial kernels, it was concluded that the SVR approach is very effective in the prediction of strength properties of any type of concrete (Imran *et al*, 2022)

B. Artificial Neural Network (ANN)

ANN is an algorithm patterned along the operation of the human brain network of neurons (Burns, 2022) capable of machine learning, classification, regression, and pattern recognition. A lot of research has been on the application of ANN for concrete properties prediction with strength being the main focus. ANN was found to be more accurate than regression in building concrete prediction models (Prayogo *et al*, 2020). Flexibility, speed, and ease of use have made ANN to become one of the most preferable tools in concrete strength prediction (Nagarajan*et al*, 2020). Multi-layer perceptron (MLP) and radial basis function (RBF) are supervised ANN networks for prediction with the former being mostly used in concrete strength prediction. According to Muliauwan *et al,* (2020), the back propagation (BP) algorithm of MLP has been adjudged quite effective (Hashempour *et al*, 2020) and therefore the most commonly used learning and training algorithm of the ANN. The activating process of each neuron is shown in Equations (4) and (5)

$$netk = \Sigma wkjoj \qquad (4)$$

$$yk = f(netk) \qquad (5)$$

Where $netk$ is neurons activation to k, j is the previous layer's neurons set. wkj is the weight of the connection betwixt neurons k and neurons j, oj is neurons $j$output, and $yk$is the representative of logical transfer function and sigmoid calculated output

ANN is robust in the prediction and identification of the effects of various variables on the deformation characteristics of recycled concrete aggregate / recycled glass (RCA/RG) blend (Ghorbani *et al*, 2021). Salimbahrami and Shakeri (2021) also found out that ANN is a better tool for GGBFS concrete strength prediction when compared with multiple linear regression (MLR) using k-fold cross-validation. ANN and SVR are among the most widely utilized methods of ML that have been employed to predict the strength characteristics of concrete (de-Prado-Gil*et al*, 2022, Thang *et al*, 2022).

C. XGBoost

Extreme Gradient Boosting (XGBoost) ML is a scale-able and efficient implementation of the gradient boosting framework by Friedman (2001). According to Li *et al* (2018), XGBoost is a boosting algorithm branch whose main purpose is to combine several weak classifiers to form a strong one, a lifting tree model that integrates many tree models into a strong classifier. XGBoost can be used for both classification and regression. Its use in concrete strength prediction has shown it as a robust and efficient ML tool.Predicted parameters and results can be expressed as Equations(6):

$$L(\phi) = \sum_i l(y_i \hat{y}_i) + \sum_k \Omega(f_k) \tag{6}$$

Where $L(\phi)$ denotes the expression in linear space, i means the number of ith samples, k means the number of kth tree, and $\hat{y}_i$ means the prediction value of ith sample in x_i and Ω is the regularization parameter.

D. Random Forest

Random forest (RF) is an ML procedure grown by a process called bagging, short for bootstrap aggregation. In RF, the bootstrap sample is used to grow trees independently out of the whole data set. According to Shaqadan (2016), the number of trees to be grown and the attributes chosen at each split are the main parameters to tune in implementing RF trees. A good feature of RF is its ability to handle large numbers of variables with comparatively small data sets and its use in providing feature importance, an appraisal of the effect of the individual variable on the overall model (Denil *et al*, 2014)]. For regression purposes, data are divided into homogeneous groups known as nodes, this partition is founded on data of splitting variables. The prediction of the outcome can then be done by tracing the path from the root node based on the split.

III. MATERIALS AND METHODS

A. Materials

The main binder used in the OPC-ARha concrete is Portland limestone cement conforming to BS/EN 197 –1:2000, of grade 32.5R. The precursor (Rha) used for the research was obtained from the calcination of rice husk using the electric furnace at a temperature of 650 °C. The sum of SiO_2, Fe_2O_3, and Al_2O_3 oxides in the precursor is 73.96% which confirms its suitableness as a natural pozzolan as per ASTM C618-19. The alkaline activator is a combination of sodium silicate (Na_2SiO_3) and sodium hydroxide solution (NaOH) with a combined specific gravity of 1.53 obtained from African Fertilizer and Chemicals, Agbara in Ogun state, Nigeria. The coarse aggregate was crushed granite of sizes 4.75 – 19 mm. The main fine aggregate used was natural sharp river sand graded to a minimum particle size of 0.150 mm and passing 4.75 mm sieve conforming to BS 882:1992. The second fine aggregate was laterite which was sieved using a 5.0 mm sieve. The fineness modulus of the laterite is 3.98, enabling its classification as coarse-fine according to ASTM C136/C136M-19.

B. Mix Design

A mix ratio of 1:2:4 (binder**:**fine aggregate**:**coarse aggregate) was used for the control and the laterized OPC-ARha hybrid concrete specimens. As shown in Table 1, the control mix was without laterite and activated RHA, while the other mixtures contained activated RHA and laterite in various proportions. The batching and mixing of materials was by weight in kilogram (kg). Portland limestone cement was replaced with activated RHA at 10%, 20% and 30% while laterite was also used to replace sharp river sand at 10%, 20% and 30%.

Table 1 Mix proportions of OPC- ARHA concrete (kg/m^3)

Group*	Designation*	OPC	RHA	Coarse Agg	Sand	Laterite	Activator	Total Water
C	Control	326.0	0	1304	652.0	0	0	238
	C1	326.0	0	1304	586.8	65.2	0	238
	C2	326.0	0	1304	521.6	130.4	0	238
	C3	326.0	0	1304	456.4	195.6	0	238
R1	R1A	293.4	32.6	1304	652.0	0	14.67	239.4
	R1B	293.4	32.6	1304	586.8	65.2	14.67	239.4
	R1C	293.4	32.6	1304	521.6	130.4	14.67	239.4
	R1D	293.4	32.6	1304	456.4	195.6	14.67	239.4
R2	R2A	260.8	65.2	1304	652.0	0	29.34	248.7
	R2B	260.8	65.2	1304	586.8	65.2	29.34	248.7
	R2C	260.8	65.2	1304	521.6	130.4	29.34	248.7
	R2D	260.8	65.2	1304	456.4	195.6	29.34	248.7
R3	R3A	228.2	97.8	1304	652.0	0	44.01	251.1
	R3B	228.2	97.8	1304	586.8	65.2	44.01	251.1
	R3C	228.2	97.8	1304	521.6	130.4	44.01	251.1
	R3D	228.2	97.8	1304	456.4	195.6	44.01	251.1

* C = concrete with 100% OPC, R1 = concrete with 10% RHA, R2 = concrete with 20% RHA, R3 = concrete with 30%, RHA, A = 0% laterite, B = 10% laterite, C = 20% laterite, D = 30% laterite,

C. Specimen Description

The specimens for the determination of the compressive strength were concrete cubes of size 150mm x 150mm x 150 mm. The concrete samples were cured for 7, 28, 56, and 91 days. Compressive strength was taken as the average of three specimens and in conformity to BS EN 196- 1:2005.

D. Data Description

Eight input variables were used, which comprised ordinary Portland cement (OPC), rice husk Ash (Rha), sand, laterite, slump value, water/ cement ratio, Alkaline activator/water ratio, and age of concrete. The output variable was the compressive strength (CS). Table 2 describes the details of the input variables. Data from a total of 192 concrete specimens were used for training and testing in the ratio of 80:20. Pearson correlation coefficient (R) is widely used to measure the statistical relationship between two variables (Xu *et al*, 2021 while Pearson correlation plot was used to check the relationship among the variables. The result in Figure 1 shows the correlation between the variables. The perfect negative correlation between OPC and Rha, sand and laterite, and the near-perfect relationship between water/cement ratio and alkali activator/water ratio are expected because Rha was used to replace OPC, likewise, laterite was used to replace sand in the concrete mix. Also, the water/cement ratio and alkali activator/water are directly related by water used in the concrete.

Table 2 The Details of The Input Data Used For Analysis

Input Variables	Minimum	Maximum	Range
Ordinary Portland Cement (OPC) (kg/m^3)	228.2	326	97.8
Rice Husk Ash (RHA) (kg/m^3)	0	97.8	97.8
Fine Aggregate (Sand) (kg/m^3)	456.4	652	195.6
Fine Aggregate (Laterite) (kg/m^3)	0	195.6	195.6

Slump Value (mm)	10	80	70
Water / Cement ratio	0.67	1.52	0.85
Alkaline Activator / Water ratio	0.0	0.139	0.139
Age (days)	7days	91days	84 days

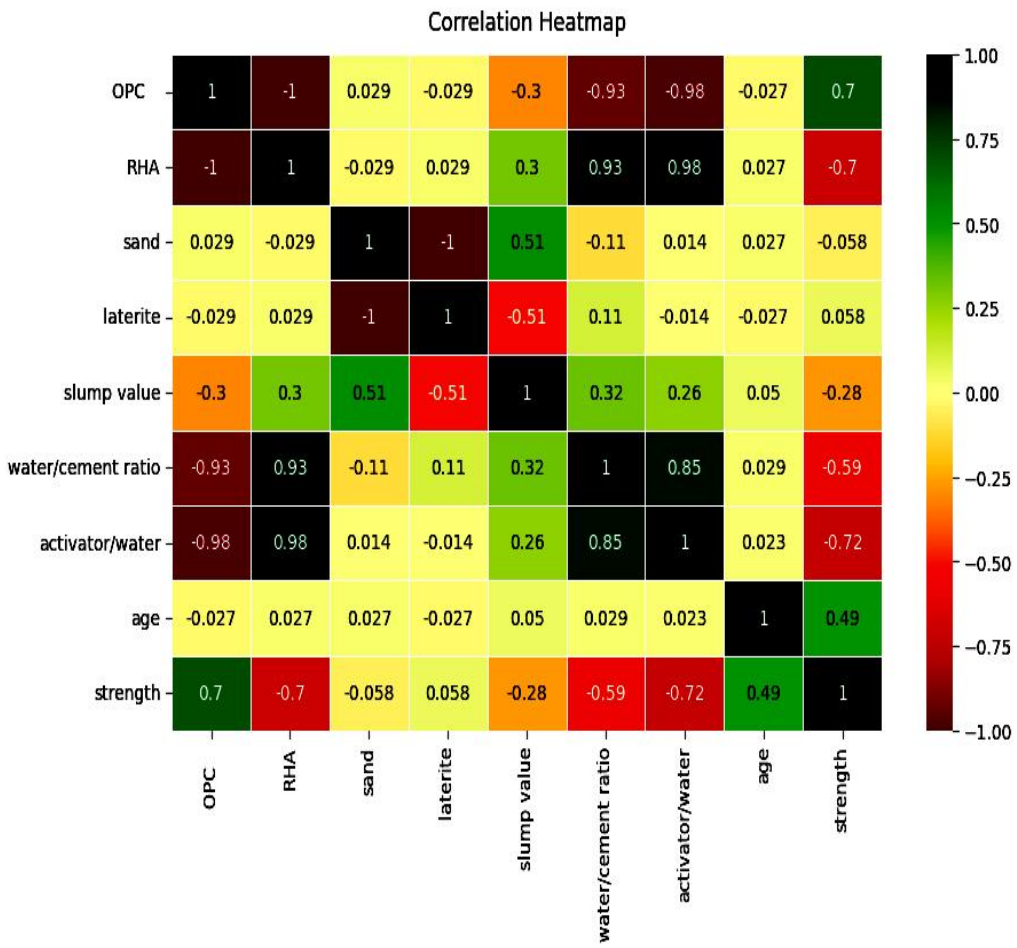


Fig.1. Correlation Coefficients for the Independent and Dependent Variables

E. Parameter Settings for Models

All the ML models were initially given initial parameter values. However, it was observed that hyper-parameter tuning using 'GridSearchCV' which is available in Python 'sklearn' yielded the best tuning/ setting for each of the ML. 'GridSearchCV' works by providing the best setting based on parameters that will produce the highest R^2. Therefore parameters supplied by 'GridSearchCV' were used for the ML models.

IV. RESULTS AND DISCUSSION

A. Variable Importance

Sensitivity analysis helps to understand the influence of each input variable on the output variables. The higher the sensitivity values, the more significant the impact of the input variables is on the output variable. According to Shang *et al*(2022), the input variables have a notable effect on the prediction of the output variables.

To evaluate the importance and impact of each input variable on the prediction variable, SHAP (SHapley Additive exPlanations) was used to check the feature importance (measured as the mean absolute Shapley values) and feature value (impact on model output). The age of the concrete, cement content, and water/cement ratio were found to be the most important factors determining the compressive strength of hybrid

OPC-Activated Rha concrete. The variable that had the least effect on the prediction outcome was the laterite content. This is shown in Figures 2a-2b. Increase in age and OPC content resulted in higher strength, while an increased in water/cement ratio caused a reduction in strength.

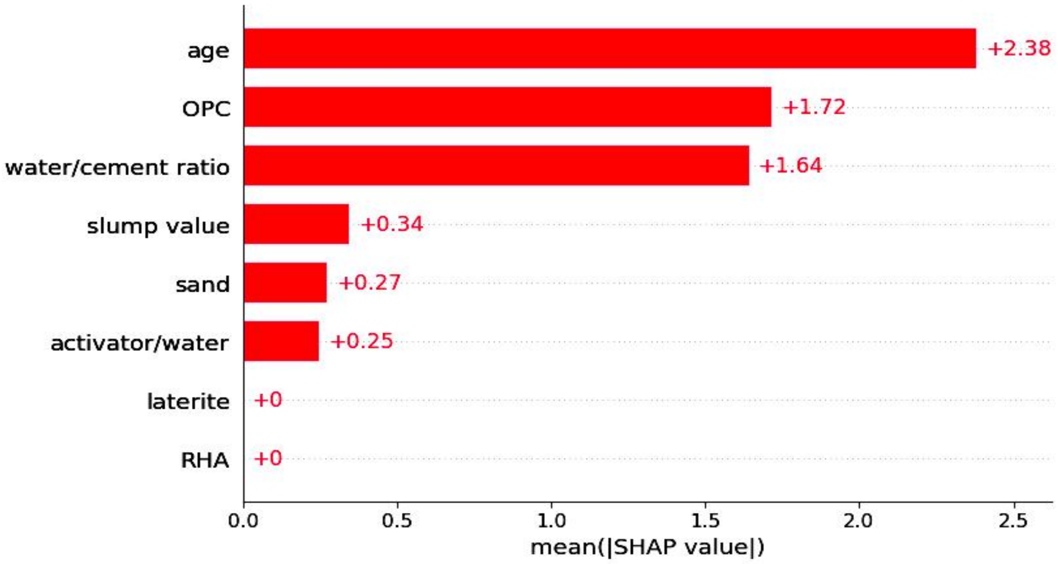


Fig. 2a. SHAP feature importance measured as the mean absolute Shapley value

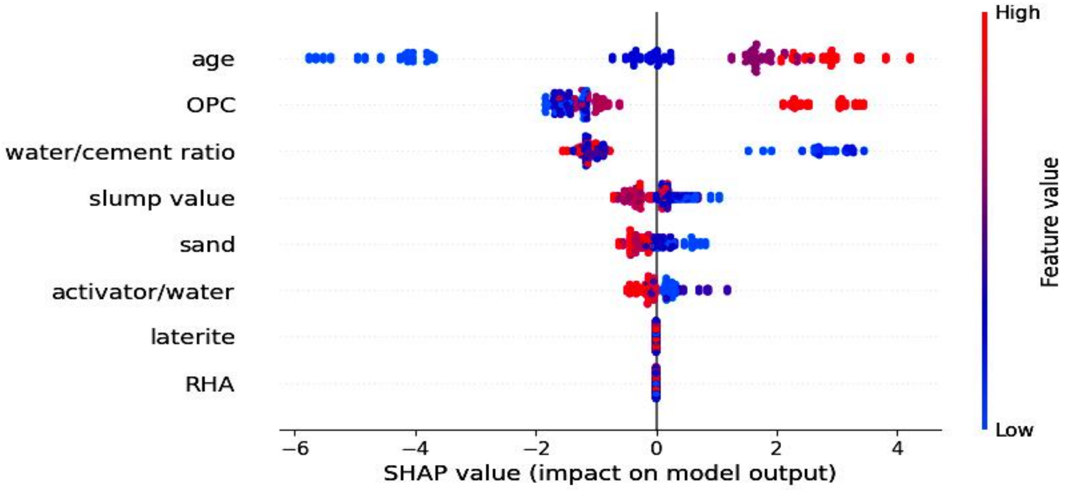


Fig. 2b. SHAP bee swarm feature values

B. Model Evaluation

Model evaluation is very important in data analysis. Three main metrics (Equations 7 to 9) were used for evaluation in these models. They include:

i. R Square/Adjusted R Square,

$$R = \frac{n\Sigma yy' - (\Sigma y)(y')}{\sqrt{n(\Sigma y^2) - (\Sigma y)^2}\sqrt{n(\Sigma y'^2) - (\Sigma y')^2}} \quad (7)$$

ii. Mean Square Error (MSE)/Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y - y')^2} \quad (8)$$

iii. Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y - y'| \quad (9)$$

Where *y* is the measurement, *y'* is the corresponding prediction and *n* is number of data points.

In R^2 the closer the value is to 1, the better the model fits. Mean absolute error (MAE) and Root mean squared error (RMSE) are used to describe the differences between predictive values and the actual values. For a good fit, their value should be close to zero. Compared with MAE, RMSE gives outliers more weight by squaring to amplify deviation. Therefore, RMSE is more sensitive to outliers and reflects the variation of error; MAE is more robust to outliers, and better reflects the real situation of predicted value errors. The metrics were used to compare the performance of the various ML approaches.To better represent the magnitude of the prediction error change, λ is introduced to represent the difference between RMSE and MAE. The smaller the λ, the more stable the prediction result (Li *et al*., 2018).From the values obtained on the four ML models used for the prediction, it can be deduced that SVR, XGBoost, ANN, and RF can be effectively used to predict the compressive strength of hybrid OPC-ARha concrete as seen in Table 3.

R-Square

As observed from Figure 3a among the four ML models, SVR has the least performance with an R-square of 0.7172, while the other three ML models have an R-square of over 0.90. ANN with an R-square of 0.9496 has the best performance.

Fig. 3a: Bar plot of the R-Square values

RMSE and MAE

RF and ANN have the lowest mean absolute error (MAE) and root mean squared error (RMSE) of 0.9565 and 1.2261 respectively as shown in Figure 3b.As expected, SVR has the highest RMSE and MAE values. λ being the difference between RMSE and MAE was introduced to better represent the magnitude of the prediction error and as an index of model stability. From Figure 3c, ANN with the least λ value of 0.2393 is considered more stable and the best performing of the four MLs considered.

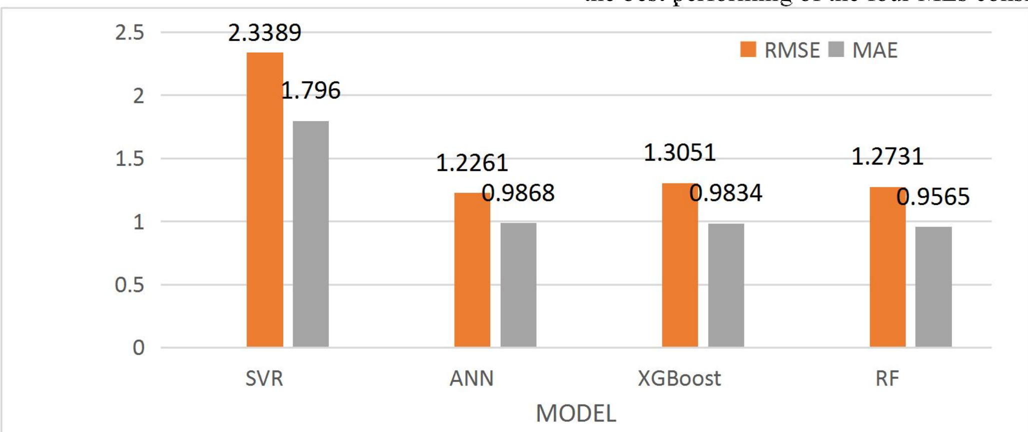


Fig.3b Bar plot of the RMSE and MAE

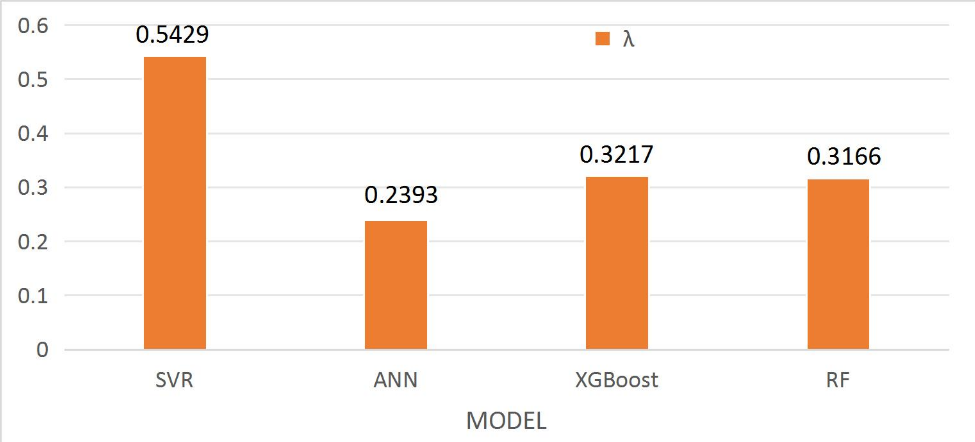
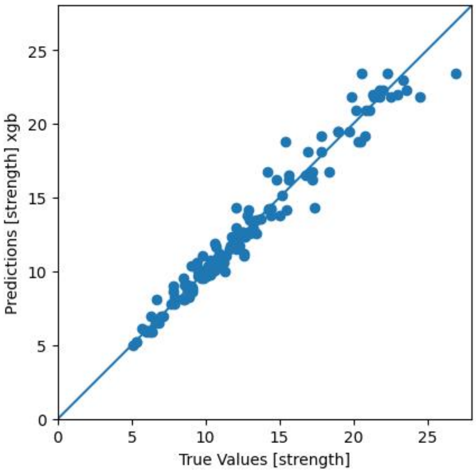


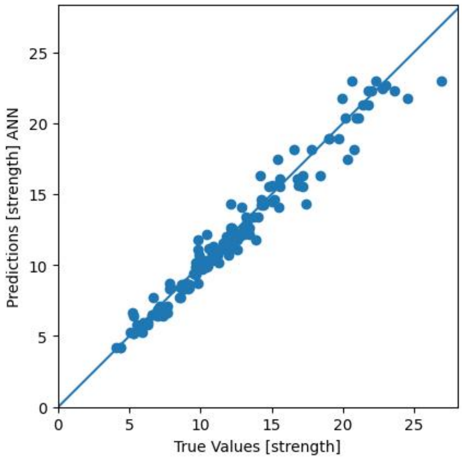
Fig.3c: Bar plot of λ

XGBoost was slightly outperformed by RF while SVR with the highest λ value of 0.5429 was the worst performing of the four MLs considered.

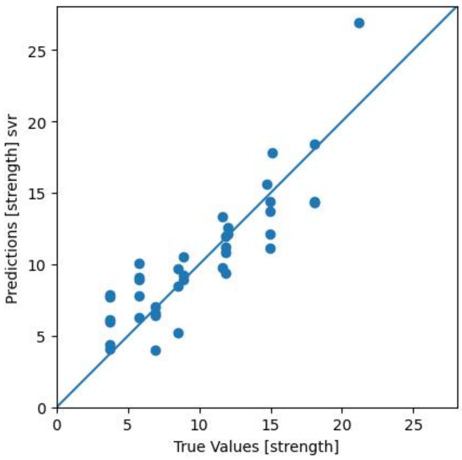
A plot of true values against the predicted values using the test data as shown on Figures 4a-d showed that ANN, XGBoost and RF being more linear in plot performed better than SVR.



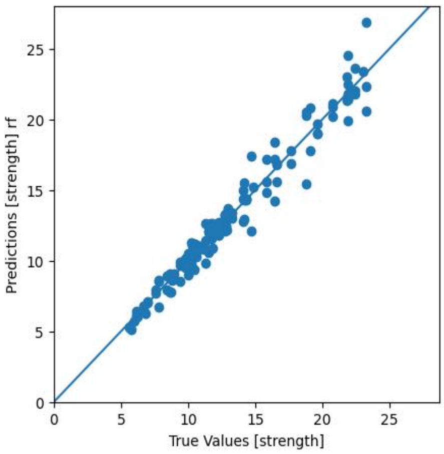
(a)



(b)



(c)



(d)

Figs 4 a-d Plot of true values against model predicted values using testing data

V. CONCLUSION

In this work, four ML models were used to predict the compressive strength of OPC-ARha concrete. The influence and importance of the input variables indicated that age (curing days) has the highest positive impact on the compressive strength of the concrete. The higher the age, the stronger the concrete. This was closely followed by ordinary Portland cement content indicating that when water curing is being employed, Portland cement dictates the compressive strength of hybrid OPC-ARha even though activated Rha is expected to gain from the hydration heat of OPC. Laterite content was found to have the least influence on the matrix's compressive strength. Comparing the four MLs, ANN with an R-square of 0.9496 was found to perform better than RF(0.9239), XGBoost (0.9200), and SVR(0.7172). Although RF has an MAE of 0.9565 which was the least for the four ML employed, overall ANN was adjudged the best of the models.

In conclusion, it has been demonstrated that SVR, ANN, XGBoost, and RF can be successfully used to predict the compressive strength of laterized OPC-ARha concrete and that ANN offers the best prediction of the models considered.

REFERENCES

[1] Awoyera, P.O; Kirgiz, M.S; Viloria, A and Ovallos-Gazabong D. (2020).Estimating Strength Properties of Geopolymer Self-compacting Concrete Using Machine Learning Techniques. Journal of Material Research and Technology, 9(4):9016–9028.

[2] ASTM C618-19. Standard Specification for Coal Fly Ash and Raw or Calcinated Natural Pozzolan for Use in Concrete. ASTM International, PA.

[3] ASTM C136/C136M-19. (2019). Standard Test Method for Sieve Analysis of Coarse and Fine Aggregates.

ASTM International, PA. 2019.

[4] BS EN 196-1. (2005). Method of Testing Cement. British Standards Institution, London. 2005

[5] BS EN 197-1:2000. (2000). Cement, Composition, Specification and Conformity Criteria for Common Cement. British Standards Institution, London. 2000.

[6] BS 882:1992.(1992).Specification for Aggregates from Natural Sources for Concrete. British Standards Institution, London. 1992. W4 4AL.

[7] Burns, E. (2022). An In-depth Guide to Machine Learning in the Enterprise.Available online at: www.techtarget.com. Assessed on August 26, 2022.

[8] Denil, M ; Matheson, D. and Freitas, N.(2014).Narrowing the Gap: Random Forests in Theory and in Practice.Proceedings of the 31 st International Conference on Machine Learning, Beijing, China. 2014.

[9] de-Prado-Gil, J; Palencia, C; Jagadesh, P; Martínez-García, R.A.(2022).Comparison of Machine Learning Tools That Model the Splitting Tensile Strength of Self-compacting Recycled Aggregate Concrete. Materials, 15(12):1-20.

[10] Elhishi, S; Elashry, A.M; El-Metwally, S. (2023). Unboxing Machine Learning Models for Concrete Strength Prediction Using XAI. Scientific Reports, 13:19892.

[11] Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics, 29(5): 1189-1232.

[12] Ghosh, A. and Ransinchung, R. N. (2022). Application of Machine Learning Algorithm to Assess the Efficacy of Varying Industrial Wastes and Curing Methods on Strength Development of Geopolymer Concrete. Construction and Building Materials, 341:127828

[13] Ghorbani, B; Arulrajah, A; Narsilio, G; Horpibulsuk, S and Bo, M.W. (2021).Dynamic Characterization of Recycled Glass-Recycled Concrete Blends Using Experimental Analysis and Artificial Neural Network Modeling. Soil Dynamics and Earthquake Engineering, 2021;142.

[14] Golafshani, E. M. and Kashani, A. (2022).Modeling the Compressive Strength of Concrete Containing Waste Glass Using Multi-Objective Automatic Regression.Neural Computing and Applications, 34(19):17107-17127.

[15] Hashempour, M; Heidari, A and Shahi jounaghani, M.(2020). The Efficiency of Hybrid BNN-DWT for Predicting the

Construction and Demolition Waste Concrete Strength. International Journal of Engineering Transactions B: Applications, 33(8): 1544-1552.

[16] Imran, H; Al-Abdaly, N. M; Shamsa, M. H; Shatnawi, A; Ibrahim, M and Ostrowski, K. A.(2022). Development of Prediction Model to Predict the Compressive Strength of Eco-Friendly Concrete Using Multivariate Polynomial Regression Combined With Stepwise Method. Materials,15: 317.

[17] Mai, H.V.T; Nguyen, T. A; Ly, H. B and Tran, V. Q.(2021). Prediction Compressive Strength of Concrete Containing GGBFS Using Random Forest Model. Advances in Civil Engineering.

[18] Muliauwan, H. N; Prayogo, D; Gaby, G and Harsono, K. (2020).Prediction of Concrete Compressive Strength Using Artificial Intelligence Methods. IOP Conference Series 1625;Journal of Physics.

[19] Pan, X; Xiao, Y; Suhail, S.A; Ahmad, W; Murali, G; Salmi, A and Mohamed, A. (2022). Use of Artificial Intelligence Methods for Predicting the Strength of Recycled Aggregate Concrete and The Influence of Raw Ingredients.Materials, 15: 4194.

[20] Prayogo, D; Santoso, D.I; Wijaya, D; Gunawan, T and Widjaja JA. (2020).Prediction of Concrete Properties Using Ensemble MachineLearning Methods. IOP Conference Series 1625;Journal of Physics.

[21] Li, Y; Zou, C; Berecibar, M; Naninimaury, E; Chan, J.C; Den Bossche, P.V; Van Mierlo, J and Omar N. (2018). Random Forest Regression for Online Capacity Estimation of Lithium-ion Batteries. Applied Energy, 232: 197–210.

[22] Sagar, B. and Sivakumar, M.V.N. (2020).An Experimental and Analytical Study on Alccofine Based High Strength Concrete. International Journal of Engineering, Transactions A: Basics, 33(4):530-538.

[23] Salimbahrami, S.R. and Shakeri, R. (2021). Experimental Investigation and Comparative Machine-Learning Prediction of Compressive Strength of Recycled Aggregate Concrete. Soft Computing, 25: 919–932

[24] Shang M, Li H, Ahmad A, AhmadW, Ostrowski KA, Aslam F, Joyklad P, Majka TM. (2022) . Predicting the Mechanical Properties Of Rca-Based Concrete Using Supervised Machine Learning Algorithms. Materials,15: 647.

[25] Shaqadan, A. (2016). Prediction of Concrete Mix Strength Using Random Forest Model. International Journal of Applied Engineering Research, 11(22): 11024-11029.

[26] Singh, A. (2023).Support Vector Regression Model: A Regression-Based Machine Learning Approach. Available online at:https://medium.com/analytics-vidhya. Accessed on August 02, 2023.

[27] Thang, V.L.; Cung, L. and Nguyen, D.S. (2022). An application of artificial neural network to predict the compressive strength of concrete using fly ash and stone powder waste products in central vietnam. International Journal of Engineering, Transactions B: Applications, 35(05):967-976.

[28] Xu, J; Zhou, L; He, G; Ji, X; Dai, Y and Dang, Y. (2021). Comprehensive Machine Learning-based Model for Predicting Compressive Strength of Ready-mix Concrete. Materials,14(5):1068-86