# CHAPTER 7 PREDICTING THE 28-DAYS COMPRESSIVE STRENGTH OF OPC AND PPC PREPARED CONCRETE THROUGH HYBRID GA-XGB MODEL

## 7.1 GENERAL

This research concentrates on proposing a computer-based technique for predicting the 28-days compressive strength (CS) of concrete made from OPC and PPC cement, which is a major issue for quality control and design engineers. For these purposes, 1062 datasets were collected from institute laboratory, 524 were belong to the concrete prepared from OPC cement and rest 538 were from PPC cement. This study utilizes eXtreme Gradient Boosting (XGB) algorithm for developing a generalize model due to the advantage of handling the large dataset and outliers. Moreover, Genetic Algorithm (GA) was adopted for tuning the hyper-parameters of XGB algorithm. The K-Fold CV approach was espoused in identifying the dataset to be used for training the model. Furthermore, the robustness of the developed models was identified through K-Fold and literature published datasets. Finally, a user-friendly, labor-free, and economically viable computer software for predicting the CS of concrete was developed to aid field and design engineers. The overall performance of the developed models demonstrates that the study could be useful to field engineers, particularly for projects with limited equipment, inadequate funding, and time limits. However, further modifications to the model or computer-based tool could be made by incorporating the dataset from other countries laboratories as well.

## 7.2 APPLIED SOFT COMPUTING APPROCHES

### 7.2.1 Genetic Algorithm (GA)

GA was first introduced by [288] and further developed by [289] have gained significant popularity in solving optimization problems. GA is based on the concept of survival of the fittest, the mechanism that drives biological evolution. Initially, it starts with a set of solutions represented by chromosomes, called population. Solutions from one population are taken to form a new population with the possibility that the new population will be better than the old one. Further, solutions are selected according to their fitness to form new solutions i.e., offspring. The above process is repeated, through the change in the values of several hyper-parameters i.e., number of generations, population size, mutation and crossover probability until some condition is satisfied. Algorithmically, the basic genetic algorithm (GAs) is outlined as below:

Step 1. Generate random population of chromosome (close to a suitable solution for that problem)

Step 2. Evaluate the fitness value for each of the chromosome

Step 3. Generate new population by repeating the following steps

i. Choose two parental chromosomes depending on the fitness value.

ii. Apply the cross-over probability (cross-over the parents to form new offspring)

iii. Mutate new offspring with the mutation probability.

iv. Incorporate the new offspring in the new population.

Step 4. Use newly generated population for further run of the algorithm.

Step 5. If stopping criteria is satisfied, stop; else go to step 2.

### 7.2.2 eXtreme Gradient Boosting (XGBoost)

XGBoost, developed by [290] , is an advancement of gradient boosting [291] algorithm. It can be used efficiently for both regression and classification problems [292], [293].

Because of working the boosted tree in parallel, the algorithm is able to tackle complicated problems with the highest precision and speed, and it quickly became the algorithm of choice for machine learning researchers and developers.

The gradient boosting algorithm uses various complement functions to estimate the results using Equation 7.1 .

$$\overline{y_i} = y_i^0 + \eta \sum_{j=1}^{m} f_j(S_i) \qquad \text{Equation 7.1}$$

Where, $\overline{y_i}$ denotes the predicted output for the $i^{th}$ data with the parameter vector $S_i$; $m$ denotes the number of estimators corresponding to independent tree structures for each $f_j$; $y_i^0$ is the primary hypothesis i.e., mean of the original parameters in the TR dataset; η is the learning rate.

According to Equation 7.1 in the $j^{th}$ stage, the $j^{th}$ estimator is connected to the model and the prediction of the$j^{th}$ $y_i^{-j}$is calculated from the estimated output $y_i^{-(j-1)}$ in the next step, and the established $f_j$ of the $j^{th}$ complementary estimator is shown in Equation 7.2

$$y_i^{-j} = y_i^{-(j-1)} + \eta f_j \qquad \text{Equation 7.2}$$

where, $f_k$ represents the leaves weight that is established by reducing the objective function of the $j^{th}$ tree and is given by Equation 7.3.

$$f_{obj.} = \gamma N + \sum_{k=1}^{N} \left[ g_k \omega_k + \frac{1}{2}(h_k + \lambda)\omega_k^2 \right] \qquad \text{Equation 7.3}$$

where, N denotes the leaf nodes quantity, λ denotes constant coefficient, γ indicates the complexity parameter, $\omega_k^2$ indicates the leaf weight from 1 to N, $g_k$ and $h_k$ are the summation parameters for the entire dataset associated with k leaf of the initial and previous loss function gradient, respectively.

In order to build the $j^{th}$ tree, a leaf is distributed into several leaves. Such a system is implied by using the gain parameters which is expressed through Equation 7.4.

$$G = \frac{1}{2}\left[\frac{S_L^2}{T_L + \lambda} + \frac{S_R^2}{T_R + \lambda} + \frac{(S_L + S_R)^2}{T_L + T_R + \lambda}\right] \qquad \text{Equation 7.4}$$

Where, G denotes the gain parameters, $S_L$ and $T_L$ denote the left leaf, respectively. Similarly, $S_R$ and $T_R$ denote the subsequent division of the right leaf, respectively. When the gain parameter is approximated to zero, the division criteria are generally assumed. γ and λ are regularization parameters that are indirectly dependent on the gain parameters. A larger regularization parameter can significantly reduce the gain parameter which leads to avoid the leaf convolution phenomenon. However, this will reduce the performance of the model to adapt to the training data.

## 7.3 DATA COLLECTION, ANALYSIS AND PREPARATION

Over almost five years, a final set of 1062 datasets of concrete were collected from laboratory experiments. Out of 1062 datasets, 524 samples were prepared from Ordinary Portland Cement (OPC) and 538 were prepared from Portland Pozzolana Cement (PPC). A sample dataset of concrete prepared from OPC and PPC is summarized in Table 7.5 and Table 7.6, respectively. All the compressive strength tests were conducted on 15x15x15 $cm^3$ cube as per Bureau of Indian Standards (BIS) specifications. The collected test data comprises of water (W), cement (C), sand (or fine aggregate referred to as FA), coarse aggregate (CA: consist of 10mm and 20mm sizes), superplasticizer (SP) and 28-days compressive strength (28-days CS).

The parameters along with their descriptive statistic values are tabulated in Table 7.1 and Table 7.2 for concrete prepared from OPC and PPC cement, respectively. Furthermore, the insight information about the dataset was extracted in the form of histogram plot as shown in Figure 7.1.

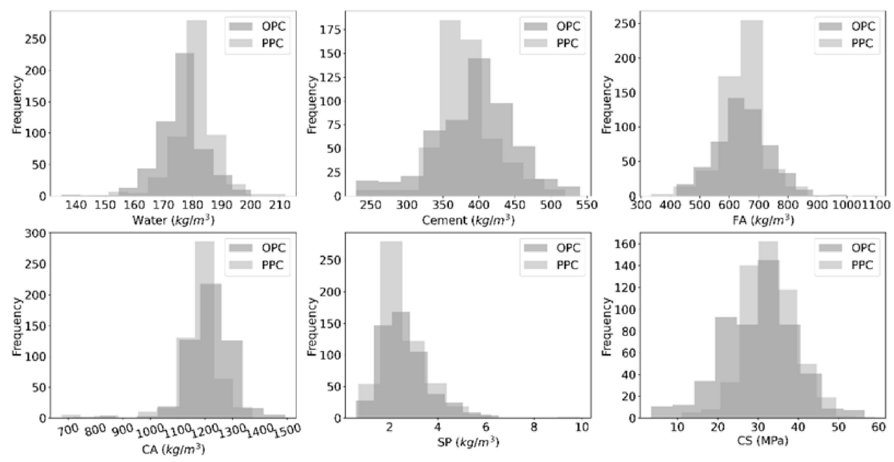


Figure 7.1 Combined histogram plot for input and output parameters of OPC and PPC concrete

Table 7.1 Statistical summary of concrete dataset prepared from OPC

| | Water ($kg/m^3$) | Cement ($kg/m^3$) | FA ($kg/m^3$) | CA ($kg/m^3$) | SP ($kg/m^3$) | 28-Days CS (MPa) |
|---|---|---|---|---|---|---|
| Min | 135.00 | 230.00 | 421.00 | 718.00 | 0.62 | 3.50 |
| Max | 200.10 | 540.00 | 1000.00 | 1492.00 | 8.00 | 56.50 |
| Average | 176.44 | 392.28 | 647.79 | 1216.43 | 2.64 | 30.41 |
| Median | 175.75 | 395.00 | 646.50 | 1216.00 | 2.46 | 30.50 |
| Mode | 180.00 | 400.00 | 640.00 | 1200.00 | 1.80 | 29.50 |
| S.D. | 8.18 | 57.37 | 89.76 | 83.22 | 1.00 | 9.07 |
| Variance | 66.95 | 3290.77 | 8056.50 | 6926.38 | 0.99 | 82.26 |

Table 7.2 Statistical summary of concrete dataset prepared from PPC.

| | Water ($kg/m^3$) | Cement ($kg/m^3$) | FA ($kg/m^3$) | CA ($kg/m^3$) | SP ($kg/m^3$) | 28-Days CS (MPa) |
|---|---|---|---|---|---|---|
| Min | 144.40 | 230.00 | 335.00 | 676.00 | 0.68 | 11.00 |
| Max | 211.75 | 520.00 | 1095.00 | 1373.00 | 9.91 | 59.00 |
| Average | 180.56 | 382.02 | 642.85 | 1176.39 | 2.52 | 32.86 |
| Median | 180.00 | 375.00 | 648.00 | 1184.00 | 2.31 | 31.50 |
| Mode | 180.00 | 360.00 | 648.00 | 1200.00 | 1.80 | 29.50 |
| S.D. | 7.18 | 42.53 | 72.42 | 85.37 | 1.02 | 6.09 |
| Variance | 51.49 | 1809.02 | 5244.06 | 7287.67 | 1.04 | 37.12 |

The major of association between the parameters was identified through Pearson's correlation analysis, which is shown in the form of a matrix plot in Figure 7.2 and Figure 7.3 for OPC and PPC concrete, respectively. It is observed from Figure 7.2 and Figure 7.3 that the water content has weak negative and CA have acceptable positive correlation with CS. Furthermore, the CS of OPC and PPC prepared concrete increases significantly with the increase in cement content and decrease in fine aggregate.

An inappropriately proportioned dataset may distort some of the essential properties of the training dataset, which may be responsible for over-fitting and under-fitting of the model, and subsequently the accuracy of any predictive models. Hence, in order to determine the dataset to be used for training the model, K-Fold data division was used in this study. Initially, the complete dataset was divided into training (TR) and testing (TS) sets. Using four number of folds, 75% were considered for training purposes and the rest 25% were kept for testing the model. Therefore, among a total of 524 datasets of OPC concrete, 393 were taken for TR set and rest 131 for TS set. Similarly, from 538 datasets of PPC concrete, 404 and 134 were taken for TR and TS set.

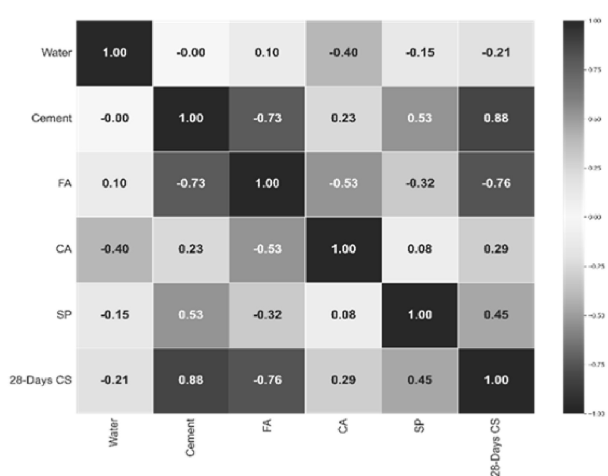


Figure 7.2 Correlation plot of 28-days CS with other input parameters of OPC concrete

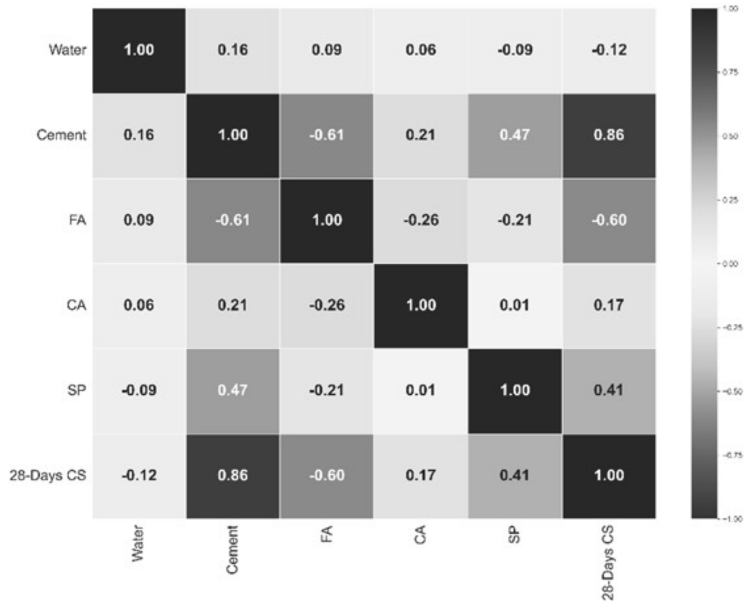


Figure 7.3 Correlation plot of 28-days CS with other input parameters of PPC concrete.

## 7.4 XGB Hyper-Parameters Tuning Through GA

Each machine learning (ML) algorithms consist of different hyper-parameters. The selection of precise hyper-parameters can make a significant impact on the predictive ability of the model. Doing it manually by all possible value is time-consuming and impractical because the number of possible combination is large. Therefore, this study utilized genetic algorithm (GA) for tuning the hyper-parameters of XGB algorithm. For GA, the population size was set to 50, mutation and crossover probability was 0.10 and 0.06, respectively, the number of generation was taken as 1000 and tournament size was 5. The termination criteria of GA were set to number of generation reaching to maximum value. The step-by-step methodology for developing the hybrid GA-XGB model is shown in Figure 7.4.

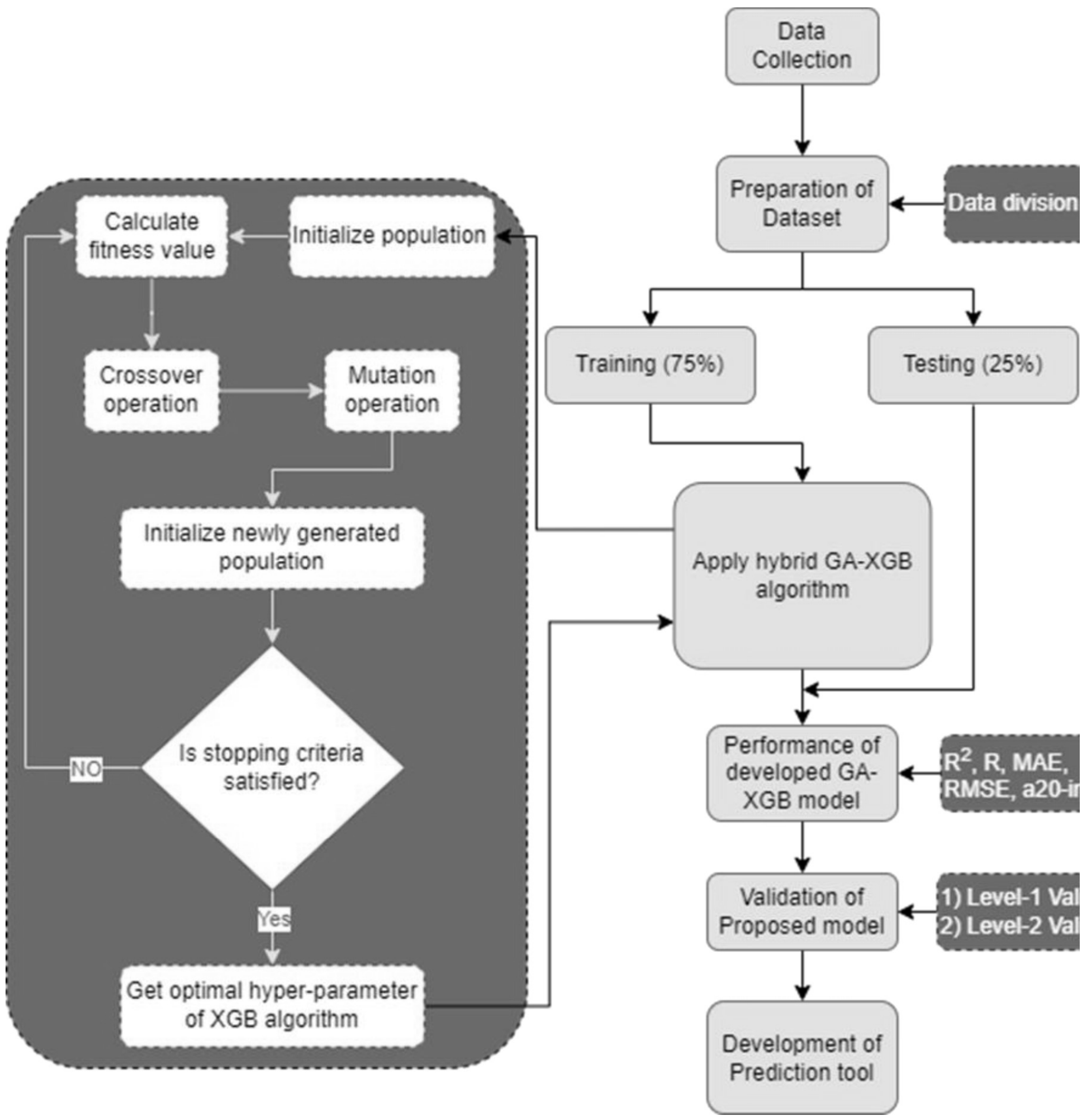


Figure 7.4 Flow diagram for developing the hybrid GA-XGB model.

Using the TR dataset, numerous hyper-parameters of XGB algorithm (as shown in Table 7.3) were optimized through GA. The default values were taken for other hyper parameters. The search ranges of hyper parameters were specified based on few preliminary trials, like, some arbitrary unreasonable values can be bypassed. The tuning ranges along with the optimal values of these hyper parameters of XGB model developed for OPC and PPC prepared concrete are summarized in Table 7.3.

Table 7.3 Optimal combination of XGB algorithm's hyperparameters for OPC and PPC model.

| Hyperparameters | Tuning range | Optimal parameters setting (OPC) | Optimal parameters setting (PPC) |
|---|---|---|---|
| Base score | 0.1 - 0.9 | 0.126105485 | 0.525056074 |
| Subsample | 0.1 - 0.9 | 0.795089736 | 0.616393056 |
| Maximum depth | 2 - 10 | 3 | 3 |
| N estimators | 10 - 100 | 75 | 38 |
| Learning rate | 0.01 - 0.2 | 0.1520120153 | 0.1695540273 |

## 7.5 PERFORMANCE EVALUATION PARAMETERS

Statistical performance of models was identified through five parameters such as coefficient of determination ($R^2$), coefficient of correlation (R), mean absoluter error (MAE), root mean square error (RMSE) and a20-index. These parameters are formulated through Equation 7.5 to Equation 7.9, respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i(a) - y_i(p))^2}{\sum_{i=1}^{N}\left(y_i(a) - \overline{y_i(a)}\right)^2}$$ Equation 7.5

$$R = \frac{\sum_{i=1}^{N}\left(\left(y_i(a) - \overline{y_i(a)}\right)\left(y_i(p) - \overline{y_i(p)}\right)\right)}{\sqrt{\left(y_i(a) - \overline{y_i(a)}\right)^2\left(y_i(p) - \overline{y_i(p)}\right)^2}}$$ Equation 7.6

$$MAE = \left[\frac{1}{N}\sum_{i=1}^{N}|y_i(a) - y_i(p)|\right]$$ Equation 7.7

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i(a) - y_i(p))^2}$$ Equation 7.8

$$a20 - index = \frac{n20}{N} \times 100$$ Equation 7.9

where, $y_i(a)$ = actual value (laboratory obtained value); $y_i(p)$ = predicted value (value obtained through the developed model); $\overline{y_i(a)}$= mean of actual value; $\overline{y_i(p)}$= mean of predicted value; $N$= number of observations.

## 7.6 RESULTS AND DISCUSSION

This section is particularly focused on the overall performance of developed model in terms of statistical performance measurement parameters and visual interpretation through scatter plot, violin plot, radar plot and regression error characteristics (REC) curve. Furthermore, two levels of validation were performed to identify the model's robustness. Lastly, a simple and effective user-friendly prediction tool was generated for the easiness to the field and laboratory engineers.

### 7.6.1 Statistical Performance of GA-XGB Model

Table 7.4 presents the performance of GA-$XGB_{OPC}$ and GA-$XGB_{PPC}$ model in TR and TS dataset, where GA-$XGB_{OPC}$ and GA-$XGB_{PPC}$ are referred to as GA-XGB model for concrete prepared from OPC and PPC cement, respectively. The $R^2$ value obtained for GA-$XGB_{OPC}$ model in TR and TS dataset are 0.961 and 0.902, respectively. This means that the developed GA-XGBOPC model is able to explain 96.1% and 90.2% variability in TR and TS dataset, respectively, of 28-days CS of concrete through W, C, FA, CA and SP as input parameters. Similarly, the variability explained by GA-XGBPPC model in TR and TS dataset are 90.8% and 83.1%, respectively. The R value obtained for both models in TR and TS dataset are > 0.90. The error measured in terms of MAE and RMSE are also least. According to [294]–[296] , if R value is ≥ 0.80 and MAE are least then the developed model is thought to be efficient in predicting the respective parameters. Furthermore, the predictive ability estimated in terms of a20-index also demonstrate that GA-XGBOPC and GA-XGBPPC models can predict more than 90% observation within ±20% variations.

Table 7.4 Statistical performance of GA-XGB model for OPC and PPC concrete

| | | $R^2$ | R | MAE | RMSE | a20-index |
|---|---|---|---|---|---|---|
| GA-$XGB_{OPC}$ | TR | 0.961 | 0.98 | 1.361 | 1.754 | 0.9822 |
| | TS | 0.902 | 0.95 | 2.155 | 2.923 | 0.9160 |
| GA-$XGB_{PPC}$ | TR | 0.908 | 0.95 | 1.248 | 1.740 | 0.9851 |
| | TS | 0.831 | 0.91 | 1.815 | 2.888 | 0.9328 |

### 7.6.2 Visual Interpretation of GA-XGB Model

Figure 7.5 presents the actual versus predicted 28-days CS of concrete prepared from OPC and PPC cement. The center line signifies the 1:1 (or 45º) line whereas the lower and upper line denotes the lower and upper bound which is taken as ±20%. It is observed from Figure 7.5 that the maximum number of data are distributed very close to the line of equality (1:1 line) or within the predefined lower and upper bound of percent variations. Furthermore, the developed models follow a specific and almost similar trend in TR and TS dataset. All of the aforementioned evidence indicates that the GA-XGBOPC and GA-XGBPPC models are fitted adequately.

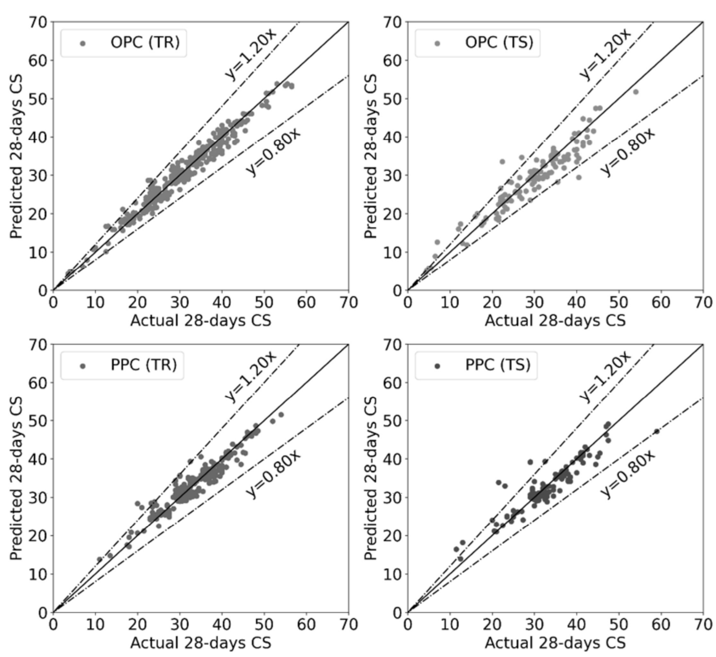


Figure 7.5 Scatter plot for GA-XGB model in TR and TS dataset of OPC and PPC concrete.

The full distribution of error obtained for GA-XGBOPC and GA-XGBPPC models in the TR and TS dataset is presented in the form of a violin plot. In general, the violin plot is similar to the box plot but with a combination of rotated density plot on each side. The left and right-side portions of violin plot are mutually mirror image. Figure 7.6 presents the violin plot for the GA-XGB models in the TR and TS dataset of OPC and PPC prepared concrete. The section of the plot that is shorter and wider corresponds to a greater probability density or frequency in the data set, whereas the section that is longer and narrower reflects a lower frequency. The lower and top nib of plot signifies the range of the error values. As seen from Figure 7.6 that the error obtained for GA-XGBOPC and GA-XGBPPC models in TR and TS datasets varies from almost +40% to -85%. Furthermore, the more wideness and poor thickness of the section in the mid of the plot demonstrate that the 28-days CS for maximum number of datasets can be predicted within ±20% variations. This can also be confirmed from Figure 7.7 showing an error bar plot for GA-XGB models in the TR and TS dataset of OPC and PPC prepared concrete. It is clearly observed from Figure 7 that more than 90% observations can be predicted within ±20% variations.

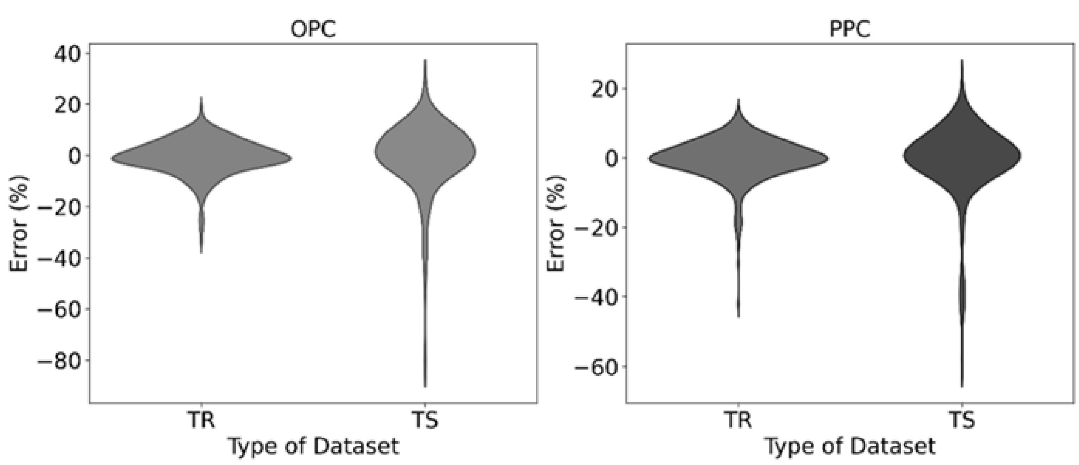


Figure 7.6 Violin plot for GA-XGB model in TR and TS dataset of OPC and PPC concrete.

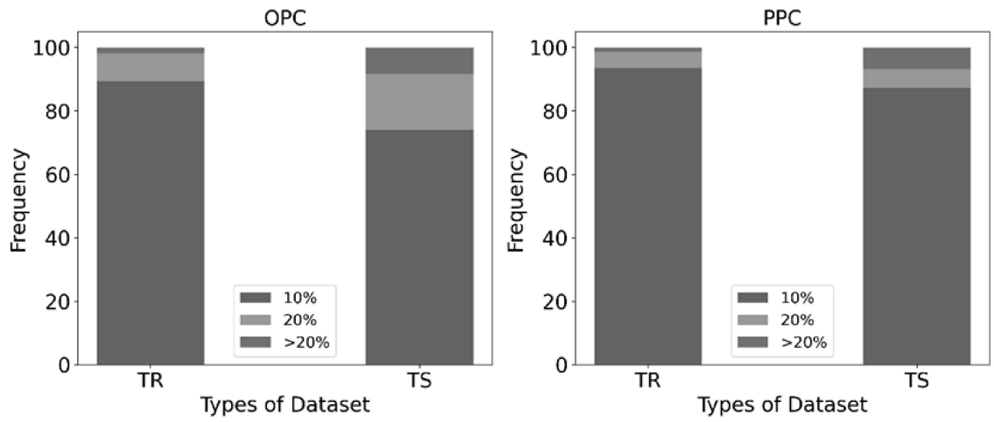


Figure 7.7 Error bar plot for GA-XGB model in TR and TS dataset of OPC and PPC concrete

Regression Error Characteristic (REC) curves are equivalents to the receiver operating characteristics (ROC) curves as in classification problems. The X-axis and Y-axis of REC curve plot signify the error tolerance and accuracy (percentage of point predicted within the tolerance), respectively [297], [298]. An ideal model's curve should pass through the upper left corner and therefore, should have an area under the curve (AUC) value is 1. This means that the model can perfectly discriminate between all the positive and the negative class points. In general, an AUC of 0.5 suggests no discrimination, 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is deemed to be excellent, and more than 0.9 is considered outstanding [298]. The REC curve signifies the amount of error either in the form of squared residual or absolute deviation. Figure 7.8 presents the REC curve plot for GA-XGBOPC and GA-XGBPPC models in the TR and TS datasets. As seen from Figure 7.8 that the REC curve in all cases exists in the upper left corner. Additionally, the AUC value obtained for both models is greater than 0.9 which means that the developed models outperform very well and are specified to be reliable for predicting the 28-days CS of concrete.

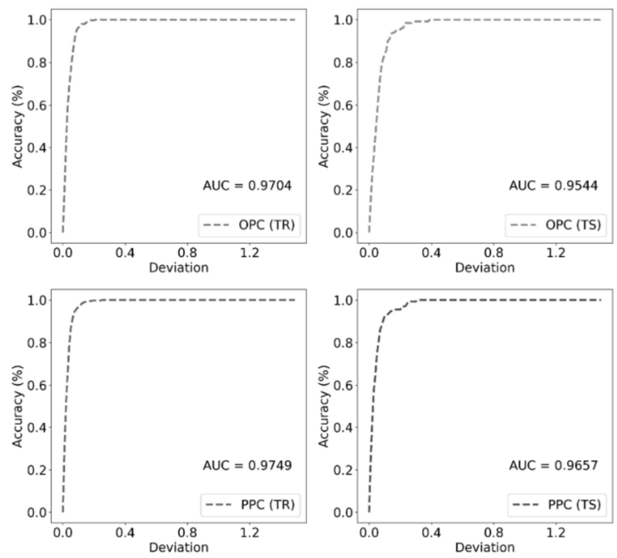


Figure 7.8 REC plot for GA-XGB model in TR and TS dataset of OPC and PPC concrete.

### 7.6.3 Validation of Proposed GA-XGB Model

The developed GA-XGB model exhibits good accuracy in both TR and TS dataset. Nevertheless, the developed model can't be considered reliable until its generalization capability is identified through some computational approaches or using some dataset that has not seen by model earlier. It is always desirable to identify the accuracy of predictive model for the dataset obtained through different environments/conditions/set-ups. In the present study, two types of validation levels were adopted: level-1 validation, through K-Fold cross-validation approach; level-2 validation, the experimental dataset taken from the literature.

#### *7.6.3.1 1st Level Validation*

K-Fold cross-validation (CV) approach was adopted for level-1 validation. Five number of folds were utilized to evaluate the predictive capability of developed GA-XGBOPC and GA-XGBPPC models.

Figure 7.9 and Figure 7.10 depicts the validation results of GA-XGBOPC and GA-XGBPPC models, respectively. The R value obtained corresponding to each of the fold in GA-XGBOPC model is more than 0.83 and for GA-XGBPPC is more than 0.75. The average R value obtained for GA-XGBOPC and GA-XGBPPC model is 0.88 and 0.81, respectively. Hence, it is perceived that the GA-XGBOPC and GA-XGBPPC models are substantially robust in predicting the compressive strength.

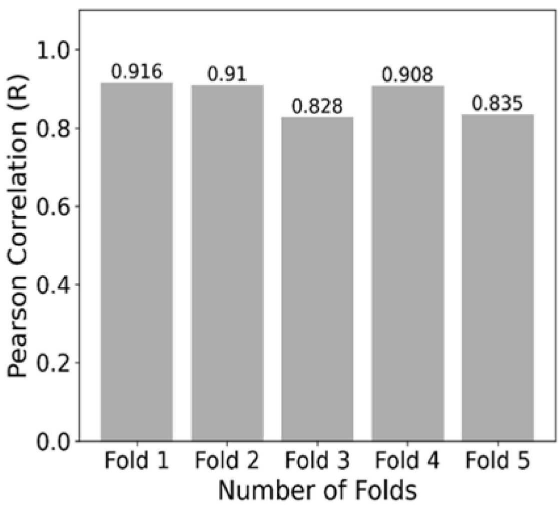


Figure 7.9 Robustness of GA-XGBOPC model through level-1 validation.

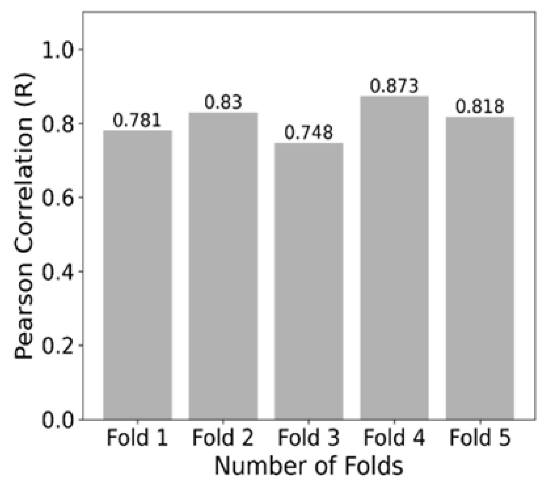


Figure 7.10 Robustness of GA-XGBPPC model through level-1 validation.

#### *7.6.3.2 2nd Level Validation*

For level-2 validation, the literature datasets were used for analysis. This method is considered to be the utmost reliable and significant as it establishes the performance in terms of real-life implementation of the developed model on the field dataset which has not seen by the model earlier. For this analysis, the only datasets were taken from [299] study which exhibit the value of water, cement, FA, CA, SP and 28-days CS parameters within the range of present study dataset. Table 7.7 summarizes the value of different parameters along with the predicted 28 days CS value. Figure 7.11, showing the error radar plot, signifies that all the literature collected dataset can be predicted within ±34%

variations through present study developed model. Moreover, despite being the literature collected dataset from different environmental and geological conditions, the developed model is able to predict almost 80% observations within ±20% variations (please see Figure 7.12). Therefore, the all above evidence obtained from level-1 and level-2 validation states that the developed GA-XGBOPC and GA-XGBPPC model are more robust or generalized in predicting the 28-days CS strength of concrete.


Figure 7.11 Radar plot for GA-XGBOPC model through level-2 validation.

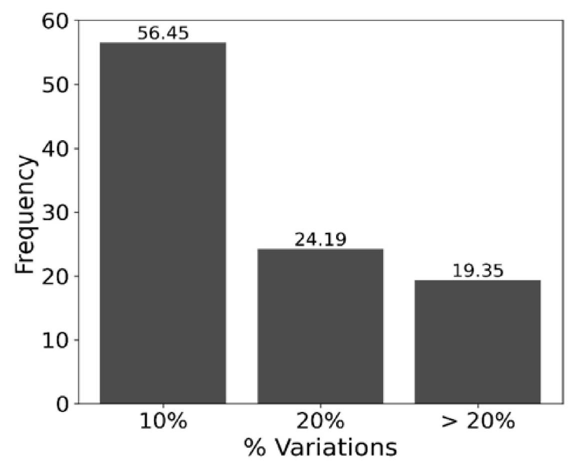


Figure 7.12 Error frequency plot for GA-XGBOPC model through level-2 validation.

Table 7.5 Sample dataset of OPC prepared concrete.

| S. No. | Water ($kg/m^3$) | Cement ($kg/m^3$) | FA ($kg/m^3$) | CA ($kg/m^3$) | SP ($kg/m^3$) | 28-Days CS (MPa) |
|---|---|---|---|---|---|---|
| 1 | 192.6 | 428.0 | 575 | 1207 | 2.14 | 32.5 |
| 2 | 191.5 | 383.0 | 607 | 1218 | 1.92 | 26.5 |
| 3 | 160.0 | 500.0 | 445 | 1302 | 8.00 | 44.5 |
| 4 | 175.5 | 540.0 | 496 | 1280 | 6.48 | 54.0 |
| 5 | 175.0 | 500.0 | 491 | 1300 | 6.25 | 50.5 |
| 6 | 172.5 | 460.0 | 549 | 1288 | 5.98 | 39.5 |
| 7 | 175.5 | 540.0 | 496 | 1280 | 5.94 | 50.5 |
| 8 | 180.0 | 400.0 | 740 | 1200 | 5.60 | 31.0 |
| 9 | 172.5 | 460.0 | 549 | 1288 | 5.52 | 41.0 |
| 10 | 175.0 | 500.0 | 491 | 1300 | 5.50 | 51.0 |
| - | - | - | - | - | - | - |
| - | - | - | - | - | - | - |
| - | - | - | - | - | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| 515 | 200.1 | 435.0 | 637 | 1140 | 1.09 | 30.0 |
| 516 | 175.0 | 350.0 | 640 | 1280 | 1.05 | 24.0 |
| 517 | 175.0 | 350.0 | 640 | 1280 | 1.05 | 30.0 |
| 518 | 175.0 | 350.0 | 640 | 1280 | 1.05 | 24.0 |
| 519 | 175.0 | 350.0 | 640 | 1280 | 1.05 | 30.0 |
| 520 | 191.3 | 510.0 | 450 | 1296 | 1.02 | 40.5 |
| 521 | 200.0 | 400.0 | 665 | 1144 | 1.00 | 22.5 |
| 522 | 185.0 | 370.0 | 602 | 1265 | 0.93 | 22.0 |
| 523 | 197.5 | 395.0 | 645 | 1174 | 0.79 | 24.5 |
| 524 | 196.8 | 410.0 | 616 | 1183 | 0.62 | 25.5 |

Table 7.6 Sample dataset of PPC prepared concrete.

| S. No. | Water ($kg/m^3$) | Cement ($kg/m^3$) | FA ($kg/m^3$) | CA ($kg/m^3$) | SP ($kg/m^3$) | 28-Days CS (MPa) |
|---|---|---|---|---|---|---|
| 1 | 187.0 | 340.0 | 667 | 1182 | 0.68 | 25.5 |
| 2 | 186.0 | 351.0 | 610 | 1229 | 0.70 | 25.0 |
| 3 | 187.5 | 375.0 | 646 | 1196 | 0.75 | 31.5 |
| 4 | 192.5 | 385.0 | 635 | 1186 | 0.77 | 30.5 |
| 5 | 188.4 | 392.5 | 617 | 1185 | 0.79 | 31.5 |
| 6 | 190.6 | 393.0 | 570 | 1204 | 0.79 | 29.0 |
| 7 | 192.0 | 400.0 | 591 | 1173 | 0.80 | 29.5 |
| 8 | 193.0 | 402.0 | 565 | 1190 | 0.80 | 29.5 |
| 9 | 188.6 | 410.0 | 611 | 1173 | 0.82 | 32.5 |
| 10 | 180.0 | 360.0 | 620 | 1257 | 1.08 | 31.0 |
| - | - | - | - | - | - | - |
| - | - | - | - | - | - | - |
| - | - | - | - | - | - | - |
| 529 | 177.4 | 486.0 | 536 | 1240 | 5.35 | 47.0 |
| 530 | 144.4 | 486.0 | 536 | 1240 | 5.35 | 47.0 |
| 531 | 175.5 | 390.0 | 609 | 1226 | 5.85 | 36.0 |
| 532 | 187.0 | 425.0 | 715 | 1150 | 6.16 | 35.0 |
| 533 | 169.0 | 520.0 | 452 | 1262 | 6.76 | 48.0 |
| 534 | 180.0 | 500.0 | 452 | 1280 | 9.50 | 48.5 |
| 535 | 177.5 | 355.0 | 696 | 1160 | 9.91 | 31.5 |
| 536 | 186.9 | 356.0 | 638 | 1196 | 1.78 | 23.5 |
| 537 | 194.0 | 400.0 | 588 | 1176 | 2.00 | 32.5 |
| 538 | 187.0 | 440.0 | 584 | 1168 | 2.20 | 35.5 |

Table 7.7 Dataset collected from [299] study

| S. No. | Water (Kg) | Cement (kg) | FA (kg) | CA (kg) | SP (kg) | Actual 28 days CS | Predicted 28 days CS | % Error |
|---|---|---|---|---|---|---|---|---|
| 1 | 174 | 289 | 933 | 900 | 0.86 | 21.2 | 20.93 | 1.29 |
| 2 | 184 | 308 | 927 | 860 | 0.62 | 23.8 | 22.28 | 6.37 |
| 3 | 183 | 302 | 926 | 860 | 0.91 | 23.8 | 19.28 | 18.99 |
| 4 | 184 | 309 | 821 | 975 | 0.92 | 22.6 | 22.28 | 1.40 |
| 5 | 164 | 328 | 886 | 942 | 1.64 | 31.2 | 23.63 | 24.26 |
| 6 | 164 | 328 | 886 | 942 | 1.64 | 34.9 | 23.63 | 32.29 |
| 7 | 176 | 368 | 805 | 988 | 0.77 | 26.9 | 26.34 | 2.08 |
| 8 | 178 | 360 | 858 | 914 | 1.08 | 30.7 | 25.14 | 18.12 |
| 9 | 165 | 371 | 810 | 1000 | 1.85 | 27.6 | 30.32 | 9.84 |
| 10 | 164 | 367 | 847 | 940 | 1.84 | 28.5 | 30.61 | 7.39 |
| 11 | 169 | 348 | 882 | 902 | 1.74 | 32 | 27.74 | 13.31 |
| 12 | 181 | 370 | 866 | 887 | 1.11 | 30.6 | 24.67 | 19.37 |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . |
| 62 | 171 | 390 | 768 | 969 | 1.17 | 30.6 | 33.02 | 7.91 |

## 7.7 PREDICTION TOOL FOR 28-DAYS CS

Graphical User Interface (GUI) design technique, a system of representing the features visually in the form of computer software or tool, was used. The designed interface was labelled as "Concrete Compressive Strength Prediction". Figure 7.13 presents the structure of the designed prediction tool. In this tool, initially, user have to select the type of cement to be used in concrete. Later, by entering the values of water, cement, FA, CA and SP, the 28-days concrete strength can directly be obtained after pressing the Run button. The designed tool shows 'Successfully Run' message when there is no error in providing the input parameters value. The reset button could be used for initializing the model and input parameters value. The designed interface is not only beneficial for the researchers but much user friendly for the site engineers. The only limitation of using this tool is that the values of input parameters should be in the range given in Table 7.1 and Table 7.2 for OPC and PPC cement, respectively.

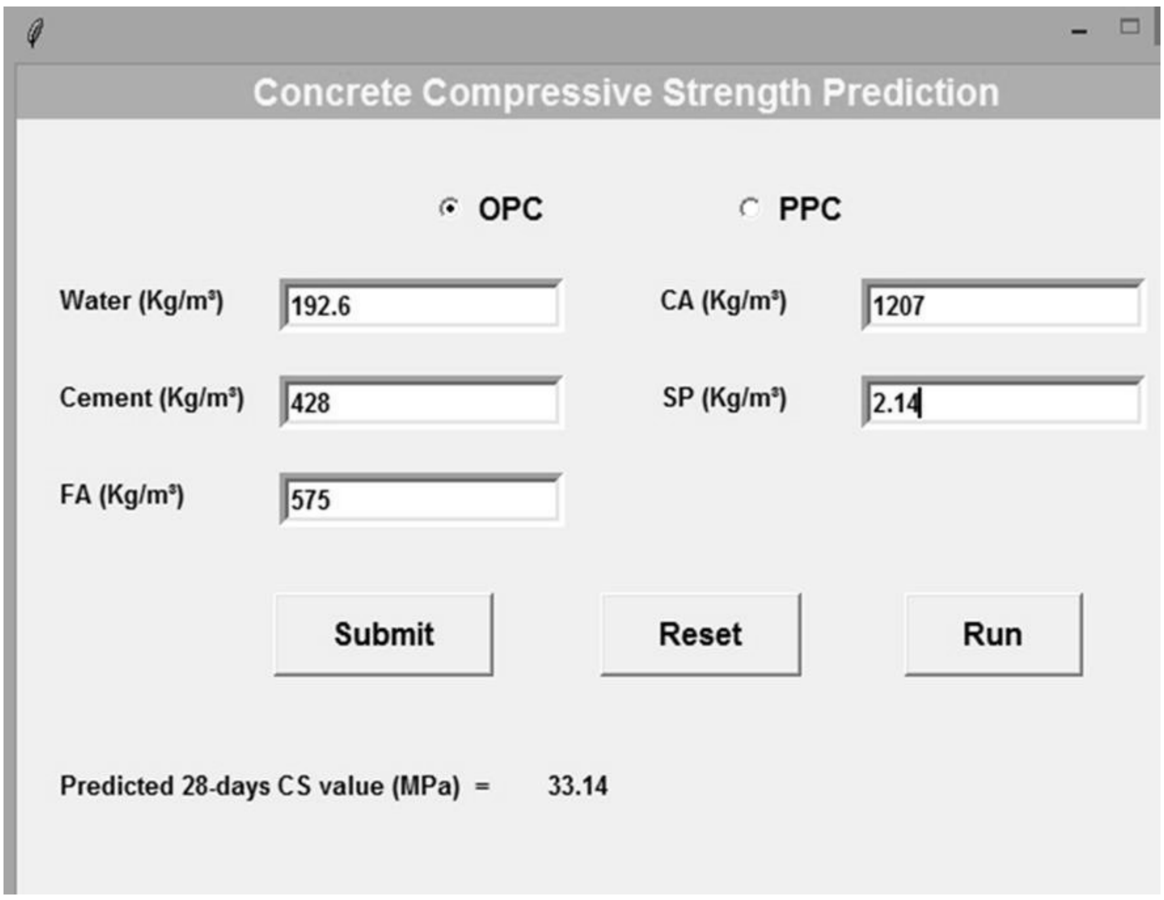


Figure 7.13 Designed interface for predicting the 28-days CS of OPC and PPC prepared concrete.

## 7.8 CONCLUSION

This study attempted to develop a robust computer-based prediction tool for 28-days compressive strength of concrete prepared from OPC and PPC cement. The analysis made use of a 1062 comprehensive dataset compiled from laboratory trials conducted over the past five years. Out of 1062 datasets, 524 were belong to the concrete prepared from OPC cement and rest 538 were from PPC cement. This study utilizes eXtreme Gradient Boosting (XGB) algorithm for developing a generalize model. Furthermore, Genetic Algorithm (GA) was adopted for tuning the hyper-parameters of XGB algorithm. The performance of developed hybrid GA-XGBOPC and GA-XGBPPC model were identified through statistical performance indices such as $R^2$, R, MAE, RMSE and a20-index. The R value obtained for both models in TR and TS dataset are almost ≥ 0.90. The MAE value obtained for GA-XGBOPC and GA-XGBPPC model were 2.155 and 1.815, respectively. Similarly, the RMSE value obtained for GA-XGBOPC and GA-XGBPPC models were 2.923 and 2.888, respectively. Moreover, the developed GA-XGBOPC and GA-XGBPPC models were found to predict more than 90% observation within ±20% variations. Based on the scatter plot and AUC value obtained in REC curve analysis, the developed models were considered to be trustworthy in predicting the 28-days CS of concrete. In addition to the good results obtained in TR and TS phases, the excellency of results achieved in 1st level and 2nd level validation itself authenticate the generalization capability of the established GA-XGBOPC and GA-XGBPPC models. Lastly, the developed user interface tool was found considerably suitable for the future convenience of the site engineers in predicting the 28-days CS of concrete.