Concrete Compressive Strength Prediction by Ensemble Machine Learning Approach

Jyoti Thapa*

1 Former Master Research Scholar, Structural Engineering, School of Engineering, Pokhara University, Pokhara, Nepal

(Manuscript Received 31/03/2024; Revised 15/05/2024; Accepted 16/05/2024)

Abstract

The prediction of concrete compressive strength is a crucial aspect of ensuring the structural integrity and durability of construction projects. In recent years, machine learning approaches have improved upon the limitations of empirical formulas and laboratory testing methods for predicting concrete compressive strength. This study utilizes ensemble machine-learning techniques, such as Bagging, XGBoost, and Stacking models, to enhance the accuracy of concrete compressive strength prediction models. A five-fold cross-validation technique was applied to mitigate the problems of underfitting or overfitting in the regression model. Furthermore, various statistical indices were employed to compare the forecasting performance of these ensemble techniques. The prediction performance of this research revealed that XGBoost achieved the highest R-squared value of 93%, followed by Stacking and Bagging regression models at 92%. Consequently, this research underscores the potential of ensemble techniques as valuable tools in the domain of civil engineering, paving the way for more reliable and efficient construction practices.

Keywords: Concrete compressive strength; Ensemble machine learning; Bagging; XGBoost; Stacking

1. Introduction

Concrete is one of the most used and important materials in civil engineering construction industries. Concrete is made with a combination of different ingredients namely cement, sand, and coarse aggregate. Additionally, water is used to create a bond between these ingredients. Moreover, the properties and workability of concrete are enhanced by using admixture. Thus, the properties and proportion of these parameters significantly influence concrete strength. The compressive strength of concrete fundamentally shows its inherent properties. Furthermore, the compressive strength of concrete stands as a key factor determining the structural integrity and durability of various infrastructural projects [1-11]. Thus, the compressive strength of concrete is a critical parameter that influences the design, construction, and maintenance of various infrastructural projects, ranging from buildings to bridges. As a result, the prediction of concrete compressive strength is essential in the civil engineering construction field.

The concrete compressive strength is often determined based on empirical formulas or laboratory testing [4, 5]. These laboratory methods are time-consuming, resource-intensive, and prone to inaccuracies. To mitigate these limitations accurate prediction methods are crucial. Also, accurate prediction of concrete compressive strength not only ensures the safety and reliability of structures but also optimizes material usage and construction costs. Thus, machine learning can be an appropriate solution to address these limitations. In the last decades, many researchers have worked on the application of robust machine learning techniques in different civil engineering sectors as regression and classification tasks [4, 12]. Additionally, researchers have focused on exploring various techniques to enhance the prediction accuracy of concrete compressive strength by using machine learning (ML) techniques as well [5, 9, 13-17]. Among these techniques, ensemble learning methods have garnered significant attention due to their ability to combine multiple models to achieve superior predictive performance. However, most of the researchers applied single machine-learning techniques to predict concrete compressive strength. Hence, the main aim of this research study is to establish an ensemble learning technique to predict concrete compressive strength.

2. Ensemble Machine Learning Technique

The ensemble machine learning technique is a powerful technique that can improve the predictive performance of models by leveraging the strengths of multiple models. This technique can apply to both
regression and classification problems. Fundamentally this technique consists of three types such as Bagging, Boosting, and Stacking ensemble techniques. These ensemble techniques can reduce bias and variation, increase model variety, and improve the interpretability of the final forecast of regression models [18].

The short form of bootstrap aggregation is indicated by bagging. In this technique, the multiple models are trained independently, and prediction outcomes are combined by averaging for regression and majority vote for classification. Hence, this technique helps to reduce overfitting, increase stability, and enhance the generalization capability of the model [18-20]. In addition, Boosting is another ensemble technique that combines the prediction performance of weak learners in sequential order. This technique initially trains base learners in the entire dataset. After the first model, initially, equal weights are assigned to each instance and adjusted after each iteration. Adjustments are sequentially emphasized with misclassified instances in previous models. Finally, after training all weak learners, their predictions are combined, often through a weighted sum for regression or a weighted voting scheme for classification [18, 20]. Furthermore, the stacking ensemble technique fundamentally consists of two main layers namely base models and meta-models [18, 20, 21]. In this study, different machine learning models such as support vector machine (SVM), random forest (RF), decision tree (DT), k-nearest neighbor (KNN), bagging, boosting, logistic regression, etc. are used as base models, where predictions are made on the same training dataset. The predictions of base models are used as input for the meta-model. After that, the final ensemble prediction is established by combining the best prediction outcomes from different base models. Thus, this stacking technique shows better prediction performance capability, and high flexibility in model diversity acceptance in comparison with a single model.

3. Prediction Model Development

This research involved gathering 776 sets of data on concrete compressive strength from various past studies [9-11, 13-16, 22-31]. The concrete compressive strength (CS) was predicted by taking different input variables namely cement (C), sand (S), coarse aggregate (CA), water (W), and curing time (T). Additionally, fly ash (FA) and superplasticizer were also selected as input variables. These input parameters significantly influence the prediction performance of the target variable. Table 1 outlines the statistical breakdown of these datasets. After the selection of these input and target parameters, firstly all dataset was precisely preprocessed with data cleaning, correlation analysis, data distribution, and rescaling. Afterward, the preprocessed dataset was split into training and testing sets of 80% and 20%, respectively. After that, the selected model was tarin with corresponding optimal hyperparameters. The optimal hyperparameters of all models were established by using the grid search method with a 5-fold cross-validation (CV) method. After that, the train performance of regression models was validated by using a testing set. The outcomes were evaluated by using different statistical indicators presented in section 3.6. If the outcomes are not in satisfactory range, then tune selected hyperparameters. In contrast, if prediction outcomes are in satisfactory range, then compare these results and establish a final prediction regression model. Figure 1 presents the detailed research methodology that is used in this research work.
Table 1: Statistical indices of selected features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Mean</th>
<th>Std</th>
<th>Min</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>300.2</td>
<td>98.27</td>
<td>134.7</td>
<td>218.9</td>
<td>290.1</td>
<td>372</td>
<td>540</td>
</tr>
<tr>
<td>W</td>
<td>182.1</td>
<td>17.96</td>
<td>140</td>
<td>168.2</td>
<td>186</td>
<td>192.9</td>
<td>238.6</td>
</tr>
<tr>
<td>S</td>
<td>765.7</td>
<td>148.1</td>
<td>0</td>
<td>734.9</td>
<td>780.5</td>
<td>830</td>
<td>1820</td>
</tr>
<tr>
<td>CA</td>
<td>1026.9</td>
<td>142.4</td>
<td>410</td>
<td>966</td>
<td>1028.4</td>
<td>1086.8</td>
<td>1385.2</td>
</tr>
<tr>
<td>FA</td>
<td>69.6</td>
<td>63.54</td>
<td>0</td>
<td>0</td>
<td>97</td>
<td>124.8</td>
<td>200.1</td>
</tr>
<tr>
<td>SP</td>
<td>4.6</td>
<td>5.33</td>
<td>0</td>
<td>0</td>
<td>3.7</td>
<td>9.4</td>
<td>28.2</td>
</tr>
<tr>
<td>T</td>
<td>31.4</td>
<td>27</td>
<td>1</td>
<td>14</td>
<td>28</td>
<td>28</td>
<td>100</td>
</tr>
<tr>
<td>CS</td>
<td>30.9</td>
<td>13.4</td>
<td>6.27</td>
<td>20.4</td>
<td>29.8</td>
<td>39.8</td>
<td>79.9</td>
</tr>
</tbody>
</table>

3.1 Data Correlation Analysis

In this study, the correlation between selected input features and target variables was evaluated by using the Pearson correlation coefficients method. In this analysis, the range of correlation lies between -1 to 1. The high negative value is revealed with -1 value and vice versa. Figure 2 shows that compressive strength is positively correlated with cement (C), superplasticizer (SP), and curing time (T) with values of 0.29, 0.14, and 0.54, respectively. That means compressive strength is increased with an increase in the quantity of these ingredients.

On the other hand, this compressive strength is negatively correlated with water content (W), sand (S), coarse aggregate (CA), and fly ash (FA) with values of 0.33, 0.14, 0.09, and 0.01, respectively. Thus, the compressive strength of concrete will decrease with use of these ingredients. Eventually, this correlation analysis underscores the importance of these parameters in accurately forecasting concrete compressive strength. Thus, all selected parameters were considered for further analysis.

3.2 Data Distribution

Figure 3 presents the visualization and distribution of the selected dataset in a combined form of a box-whisker plot and violin plot. The Figure shows the statistical ranges of the dataset with its central tendency and occurrence of outliers. It seems some features namely water (W), sand (S), coarse aggregate (CA), curing time (T), and compressive strength (CS) have some outliers. In the machine learning regression model, the removal of outliers increases the performance of the model, in contrast, this removal may increase uncertainty in the real construction field.

Figure 2: Correlation between input parameters and target variable

Figure 3: Data distribution in box-whisker and violin plot

Thus, the presence of outliers is selected in this research analysis. In addition, the Figure reveals the probability of occurrence by the density curve. It seems the compressive strength of the selected
dataset mostly occurred in a range of 20 MPa to 40 MPa. In contrast, compressive strength with a value above 70 MPa occurs less. In summary, highly dense values are indicated by a wider section, and less occurred values are indicated with a narrow range in the violin plot. Hence, data distribution illustrates the presence of data conditions.

### 3.3 Data Normalization

The selected data sets in this research study contain different scales in input parameters presented in Table 1. Thus, rescaling is essential to keep the uniform scale in all parameters. In addition, the presence of outliers is considered for further regression analysis. In this perspective, min-max normalization was applied for rescaling and to maintain a uniform scale in selected parameters. The following equation (1) was used for rescaling.

\[
X_n = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} 
\]

Where \(X\) represents the actual value of selected parameters, \(X_n\) represents rescaled values. Additionally, \(X_{\text{max}}\) and \(X_{\text{min}}\) correspond to the maximum and minimum values of each input feature, respectively.

### 3.4 Hyper-Parameters Tuning

In this research, firstly the range of hyperparameters was selected by using the grid search method. After that, a 5-fold cross-validation technique was adopted to check performance and tune the optimal hyperparameters. Table 2 demonstrates the summary of optimal hyperparameters used in this research.

<table>
<thead>
<tr>
<th>Model</th>
<th>Optimal Hyperparameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>'C': 50.0, 'gamma': 'scale', 'kernel': 'rbf'</td>
</tr>
<tr>
<td>DT</td>
<td>criterion='mse', max_depth=20, max_features='sqrt', min_samples_leaf=1, min_samples_split=2, splitter='random'</td>
</tr>
<tr>
<td>KNN</td>
<td>'n_neighbors': 5, 'p': 1, 'weights': 'distance'</td>
</tr>
<tr>
<td>RF</td>
<td>random_state=42, max_depth=20, max_features='sqrt', min_samples_leaf=1, min_samples_split=2, n_estimators=100</td>
</tr>
<tr>
<td>XGBoost</td>
<td>colsample_bytree=1, gamma=1, learning_rate=0.1, max_depth=5, n_estimators=200, subsample=0.8</td>
</tr>
<tr>
<td>Bagging</td>
<td>bootstrap=True, bootstrap_features=False, max_features=1.0, max_samples=1.0, n_estimators=50</td>
</tr>
</tbody>
</table>

For the stacking ensemble classifier, its hyperparameters were established based on the optimization results of the base models it comprises. These optimal hyperparameters were then employed to configure each regression model before training the model. Additionally, apart from the optimized hyperparameters, the remaining initialization hyperparameters for each base and meta-regression model were set to their default values as defined in the Scikit-learn libraries.

### 3.5 Base And Meta-Model Selection

In the stacking ensemble technique, base models were selected namely decision tree (DT), support vector machine (SVM), random forest (RF), k-nearest neighbor (KNN), XGBoost, and Bagging. The optimal hyperparameters of all these base regression models were established by using grid search method with a 5-fold cross-validation (CV) method. Subsequently, prediction outcomes of these base models were utilized for further analysis via optimal meta-model. Equally, these base models were utilized as meta-models, and outcomes of these meta-models were compared, and the final optimal meta-model was established. After that, the prediction outcomes of base models were used as features to train the optimal meta-model. Figure 4 illustrates the performance of each selected meta-model. Output results indicated that the random forest (RF) shows the best results as compared to others. The model shows the value of \(R^2\) and mean squared error as 0.88, and 22.06, respectively. These results are good as compared to other meta-models. Thus, this model was used as a meta-model in this study for further analysis of the stacking ensemble regression model.

![Meta-model results comparison](image)

### 3.6 Performance Evaluation Statistical Indicators

In machine learning regression analysis, the output performance of regression analysis has been evaluated by using different statistical indicators. These indicators help researchers, data scientists, and analysts understand how well their models are performing on unseen data. Thus, these statistical indicators
are crucial for assessing the effectiveness and accuracy of predictive models. In this research, the following equations (2) to (9) were applied to evaluate the effectiveness of model hyperparameters, the performance of the regression model, and in comparing different regression models. Eventually, these metrics demonstrate the effectiveness of the stacking ensemble technique over other base models.

R-squared \( (R^2) \)

\[
R^2 = 1 - \frac{\sum \text{of squared regression (SSR)}}{\sum \text{of squared total (SST)}}
\]

Mean Absolute Error (MAE)

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i^a - y_i^p|
\]

Mean Squared Error (MSE)

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i^a - y_i^p)^2
\]

Root Mean Squared Error (RMSE)

\[
RMSE = \left( \frac{1}{n} \sum_{i=1}^{n} (y_i^a - y_i^p)^2 \right)^{1/2}
\]

Relative Root Mean Squared Error (RRMSE)

\[
RRMSE = \sqrt{ \frac{1}{n} \sum_{i=1}^{n} \left( \frac{y_i^a - y_i^p}{y_i^a} \right)^2 }
\]

Mean Absolute Percentage Error (MAPE)

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i^a - y_i^p}{y_i^a} \right| \times 100\%
\]

Mean Relative Error (MRE)

\[
MRE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i^a - y_i^p}{y_i^a} \right|
\]

Variance Accounted For (VAF)

\[
VAF = 1 - \frac{\text{var}(y_i^p)}{\text{var}(y_i^a)} \times 100\%
\]

4. Results and Discussion

In this study, Bagging, XGBoost, and Stacking ensemble learning models were analyzed with optimal hyperparameters. These hyperparameters were established by the Grid search method with fivefold cross-validation techniques. In the case of the Stacking ensemble regression model, the final meta-model was established with the performance of different optimized base models. Thus, in this study, the RF model was used as a meta-model. The outcomes of each ensemble model are described in the following sections.

4.1 Bagging Ensemble Model

The Bagging ensemble model was analyzed with optimal hyperparameters presented in Table 2. The output results are presented in Figure 5 and Figure 6. In Figure 5, predicted results are indicated by blue lines, and actual values of concrete compressive strengths are represented by black lines. Predicted results of the Bagging model exhibit good fitting with actual values.
4.2 Boosting Ensemble Model

Boosting is a robust ensemble technique widely used for regression and classification problems. In this study, the XGBoost ensemble model was analyzed with optimal hyperparameters, which are presented in Table 2. In addition, the output results of this model are presented in Figure 7 and Figure 8. In Figure 7, predicted results are indicated by a pink line, and actual values of concrete compressive strengths are represented by a black line. These results reveal that this ensemble model exhibits good prediction capabilities for forecasting the concrete compressive strength.

Figure 7: XGBoost regression model forecasted outcomes comparison with actual concrete compressive strength

Additionally, this model reveals a good correlation between the actual values and outcomes of forecasted results, which is presented in Figure 8. This model shows an R-square value of 0.93, which indicates that this model has good forecasting capabilities. Moreover, other statistical indices are presented in Table 3.

Figure 8: Correlation between XGBoost forecasted outcomes with actual compressive strength

4.3 Stacking Ensemble Model

In this study, the Stacking ensemble model was analyzed with different base models and RF meta-models with their optimal hyperparameters. These hyperparameters are presented in Table 2. In addition, the output results of this stacking ensemble regression model are presented in Figure 9 and Figure 10. In Figure 9, predicted results are indicated by the green line, and actual values of concrete compressive strengths are represented by the black line. These results reveal that this Stacking ensemble model exhibits good prediction capabilities for forecasting the concrete compressive strength.

Figure 9: Stacking regression model forecasted outcomes comparison with actual compressive strength

Additionally, this model reveals a good correlation between the actual values and outcomes of forecasted results, which is presented in Figure 10. This model shows R-square value of 0.92, which indicates that this model has good forecasting capabilities. Moreover, other statistical indices are presented in Table 3.

Figure 10: Correlation between Stacking regression model forecasted outcomes with actual compressive strength
4.4 Results Comparison

The prediction output results were comprehensively demonstrated in comparison with the actual results in the above section. These comparisons depicted that ensemble machine learning techniques have good fitting to predict the compressive strength of concrete. The detailed statistical parameters of each ensemble regression model are presented in Table 3. In this Table, the R² values of ensemble regression models range from 0.92 to 0.93, which means ensemble models have good prediction performance capabilities. In addition, in this table, it was observed that the R² and VAF values of XGBoost reveal good prediction results as compared to the Stacking and Bagging ensemble regression model. Furthermore, the Table reveals that the prediction errors in XGBoost are also less as compared to other regression models.

Table 3: Regression model forecasted outcomes comparisons with their statistical indices

<table>
<thead>
<tr>
<th>Regression model/Indices</th>
<th>Bagging</th>
<th>XGBoost</th>
<th>Stacking</th>
</tr>
</thead>
<tbody>
<tr>
<td>R²</td>
<td>0.92</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>VAF (%)</td>
<td>91.86</td>
<td>93.25</td>
<td>91.99</td>
</tr>
<tr>
<td>MAE</td>
<td>2.52</td>
<td>2.31</td>
<td>2.57</td>
</tr>
<tr>
<td>MSE</td>
<td>13.61</td>
<td>11.29</td>
<td>13.41</td>
</tr>
<tr>
<td>RMSE</td>
<td>3.69</td>
<td>3.36</td>
<td>3.66</td>
</tr>
<tr>
<td>RRMSE</td>
<td>0.11</td>
<td>0.10</td>
<td>0.11</td>
</tr>
<tr>
<td>MAPE</td>
<td>8.46</td>
<td>7.44</td>
<td>8.49</td>
</tr>
<tr>
<td>MRE</td>
<td>-0.15</td>
<td>0.06</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

5. Conclusions

In this study, different ensemble techniques namely Bagging, XGBoost, and Stacking were selected to predict the compressive strength of concrete. For this purpose, 776 number of datasets were collected from previously published research papers. The output results of these ensemble techniques show good predictive capability. In addition, prediction results establish a good correlation with the testing dataset. The comparison results of R² and VAF depict that the XGBoost has the highest values of 0.93, and 93.25 %, followed by Stacking with values of 0.92, and 91.99%, and Bagging with values of 0.92, and 91.86 %, respectively. In case of errors, the XGBoost has a minimum error in comparison with other Sacking and Bagging models. Eventually, it concludes that the XGBoost ensemble exhibits good prediction results as compared other two ensemble models.

Therefore, this research presented that ensemble techniques significantly improve the prediction performance as compared to the individual non-ensemble regression models. In comparison with all ensemble techniques, the XGBoost shows a good prediction capability to predict concrete compressive strength. Thus, this technique might be applicable to forecast concrete strength in real construction field and further research work as well.

Acknowledgment

I would like to express my gratitude to the researchers whose data sets have been invaluable to the completion of this research work. Their accurate collection and documentation have significantly contributed to the depth of this research work. Without their dedication to advancing knowledge in this field, this study would not have been possible.

References


