PREDICTING COMPRESSIVE STRENGTH OF RECYCLED AGGREGATE CONCRETE USING M5’ MODEL

Ali BEHNOOD1*, Jan OLEK1**, Michal A. GLINICKI2
1Lyles School of Civil Engineering, Purdue University, 550 Stadium Mall, West Lafayette, IN 47907-2051, USA, e-mail: *abehnood@purdue.edu,**olek@purdue.edu
2Institute of Fundamental Technological Research, Polish Academy of Sciences
Pawinskiego 5B, 01-106 Warszawa, Poland, e-mail: mglinic@ippt.pan.pl

ABSTRACT

Construction industry demands large quantity of recycled materials for sustainable development. The use of recycled aggregate (RA) as a replacement for natural aggregate (NA) represents a sensible approach from technical, environmental, and economic points of view. Due to the substantial differences in the properties of RA and NA, predicting the performance of recycled aggregate concrete has been a concern in many design applications. In this study, M5’ model tree algorithm was used to develop a new model to predict the compressive strength of recycled aggregate concrete. Compared to other soft computing methods, the model tree algorithms offer the following advantages: (a) greater transparency with respect to development of model equations and (b) relative ease of development and implementation. To develop the model tree, 270 data sets were collected from international published literature. The results show that the developed model tree algorithm can well predict the compressive strength of recycled aggregate concrete.

Keywords

M5’ model tree, modulus of elasticity, recycled aggregate, concrete

INTRODUCTION

Predicting mechanical properties of construction materials such as concrete is an important task in design applications. Compressive strength of concrete (CSC) is the major and (as far as the existing design practices are concerned) probably the most important of its mechanical properties. A reliable and accurate model for prediction of CSC can quickly generate the needed design data and thus save time and costs associated with development of alternate designs of the structural components. Conventional methods of predicting the CSC are usually based on statistical analysis and multiple regression (MR) methods [1,2]. Nowadays, machine learning (ML) algorithms have been successfully applied to predict the CSC. Some of these algorithms include artificial neural network (ANN) [3–5], classification and regression trees (CART) [6–8], support vector machine (SVM) algorithms [6,7], and adaptive network-based fuzzy inference systems (ANFIS) [9,10].

Chou et al. [11] compared the performance of ANN, SVM, MR, multiple additive regression tree (MART) and bagging regression tree (BRG) methods to predict the CSC. The results indicated that the ANN and SVM provided the best prediction power for future unseen data. In another research, Chou et al. [6] used advanced machine learning techniques to predict concrete compressive strength. Specifically, those authors used test data from several countries to evaluate the performance of ANN, SVM, CART, linear regression (LR),
multilayer perception (MLP) and ensemble techniques with respect to their strength prediction capabilities. Based on the results of that study, the best individual learning methods were found to be SVM and MLP. However, the prediction accuracy of the ensemble technique was superior to that of single learning models.

Generally, three main sources of aggregates can be used in construction projects: natural, artificial and recycled [12,13]. For sustainable development, many attempts have been made to use the aggregate produced from recycled concrete (from now on referred to as recycled concrete aggregate (RCA) as coarse aggregate in construction projects. The use of RCA in the mixture changes the mechanical properties of concrete and can decrease its compressive strength by as much as 25% [14,15]. The adverse effects of the RCA on compressive strength of concrete can be attributed to the excessive content of the attached mortar, presence of various contaminants, higher water absorption, and lower density of RCA. Due to differences in properties between concretes made with recycled aggregate and concretes made with natural aggregate, the models developed for prediction of the compressive strength of the latter will generally not be reliable and accurate to predict the compressive strength of the former. It should also be mentioned that all the aforementioned studies mainly targeted the compressive strength of concrete made with natural aggregates.

More recently, Duan et al. [16] studied the applicability of artificial neural networks (ANNs) to model the compressive strength of concrete made with recycled aggregate concrete (from now on referred to as recycled aggregate concrete (RAC)). Although ANN technique provides a good alternative to statistical regression and numerical methods, it is not as transparent as regression-based methods and formulas as it does not reveal mathematical expressions for relationships between the input and output variables of the system. In addition, in the ANN approach a trial and error method is required to find the network parameters such as number of hidden layers and neurons, which is time consuming. In some cases, the training time can be as high as 100 times that required for other methods [11].

In the present study, the M5′ algorithm [17], one of the algorithms of the model tree method (MT), was employed for prediction of the compressive strength of RAC. The M5′ is a new machine learning (ML) algorithm that provides understandable mathematical formulas and allows users to have more insight with respect to engineering applications [18]. In this paper, a total of 270 data sets extracted from the published literature were used for the model development. Generalized relationships were obtained to correlate the compressive strength of RAC with mixture proportions and characteristics of individual components.

MATERIALS AND METHODS

Model Trees (MTs)

This paper aims to develop a model for prediction of the compressive strength of recycled aggregate concrete (CS_RAC) using previously mentioned M5′ algorithm. M5′ algorithm is a reconstruction of M5 algorithm that includes four steps [17]. These steps include: (a) splitting the input space and branching of data to grow a complete tree; (b) developing a regression function in each node for pruning and prediction, (c) pruning the tree to avoid the overfitting problem; and (d) smoothing the tree to compensate for the sharp discontinuities caused by the division. In M5′ algorithm, standard deviation reduction (SDR) factor, which is the maximum expected reduction in output errors after branching, is used to construct the regression tree. SDR is calculated by a formula shown in Eq. (1):

\[
SDR = sd(T) - \sum_{T_i} \frac{T_i}{|T|} \times sd(T_i)
\]
Predicting compressive strength of recycled concrete aggregate using M5’ model

where $T$ is the set of the data points that reach the node, $T_i$ is the data point that results from splitting the node and falls into one sub-space according to the chosen splitting parameter and $sd$ is the standard deviation [17]. Standard deviation is considered to represent an error measure for data points that fall into a subspace [19]. A multiple linear regression function is constructed at each inner node after growing the tree. Then, the tree is pruned from the leaves if SDR for linear model in the root of sub-tree is smaller or equal to the expected error for the sub-tree. After pruning, the smoothing process is used to compensate for the potential discontinuities between the pruned leaves and adjacent linear models. This process involves combining the estimated value with the predicted one as shown in Eq. (2):

$$P' = \frac{np + kq}{n + k}$$

where $P'$ is the prediction passed up to the next higher node, $p$ is the prediction passed to the current node from below, $q$ is the value predicted by the model at the current node, $n$ is the number of training instances that reach the node below, and $k$ is smoothing constant. Smoothing process substantially increases the accuracy of predictions [17].

**Data sets**

This study used a collection of data sets (assembled from 270 distinctive data records obtained from published literature [20,14,21–41]) to develop a model for predicting the $CS_{RAC}$. All parameters affecting the compressive strength of RAC need to be identified in order to obtain the most accurate and reliable model. Considering the nature of concrete mixtures, several parameters can be expected to be included in the model. Therefore, an engineering judgment is required to select the most significant factors. The most important factors affecting the compressive strength of RAC, including both quantitative and qualitative indexes, are described in the following sections.

**Input parameters: Mixture proportions**

Mechanical properties of RAC (e.g., compressive strength) are highly dependent on the amounts (and relative ratios) of its different constituents. These constituents include water, cement (binder), admixtures, coarse aggregate, and fine aggregate. In order to make the wide range of possible factors narrower and thus to make the model development statistically feasible, the effect of replacing fine aggregate with RCA as well as the influence of supplementary cementitious materials in the mix were excluded from the list of variables considered in this paper. Although the use of fine recycled concrete aggregate in RAC mixes has been found by some researchers [42] to improve the mechanical properties of the mixes, it is generally associated with negative effects [43–45]. With regard to mixture proportions, water-cement ratio (w/c), coarse aggregate-cement ratio (CA/C), fine aggregate-total aggregate ratio (FA/TA) and volume replacement of natural aggregate (NA) by RCA (r) were selected as input parameters for mix proportions. The volume fraction of coarse RA in RAC was calculated using Eq. (3) [16]:

$$r = \left( \frac{M_{RA}/SD_{RA}}{M_{RA}/SD_{RA} + M_{NA}/SD_{NA}} \right) \times 100$$

where $r$ (%) is the volume fraction of coarse RA in RAC; $M_{RA}$ and $M_{NA}$ (kg/m$^3$) are the amounts of coarse RA and NA used in RAC, respectively.
Input parameters: Component characteristics
The type of cement and its strength play a vital role in development of the ultimate value of CS<sub>RAC</sub>. Since the data set available for use in this study only contained the normal hardening cement, the type of cement is not considered here.

The properties of aggregates used in the concrete mixture will also affect the CS<sub>RAC</sub>. Günaydın and Doğan indicated that the processing conditions of the RCA will affect the properties of RAC [46]. In another study, Poon et al. reported that the level of moisture present in aggregates constitutes another significant factor which can change the properties RAC [47]. In addition, parameters such as the source of the RCA, curing conditions of the concrete and impurities present in RCA will also affect the properties of the mixture [36,48–50]. To account for the potential variability in the moisture levels of aggregates, this study uses water absorption (W<sub>a</sub>) and saturated surface dry specific gravity (SG<sub>SSD</sub>) of the combined (i.e. recycled and natural) coarse aggregates as input parameters for the models (see Eqs. (4) and (5) below):

\[
W_a = \frac{W_{aRA} \times r + W_{aNA} \times (100 - r)}{100} \tag{4}
\]

\[
SG_{SSD} = \frac{SG_{RA} \times r + SG_{NA} \times (100 - r)}{100} \tag{5}
\]

where \( r \) is as defined previously (Eq. (3)); \( W_{aRA}, W_{aNA}, W_a \) (%) are water absorption of coarse recycled aggregate, coarse natural aggregate, and total (i.e. combined) coarse aggregate, respectively; \( SG_{RA}, SG_{NA}, \) and \( SG_{SSD} \) are the saturated surface dry specific gravities of the coarse recycled aggregate, coarse natural aggregate and total coarse aggregates, respectively.

Output parameter
The 28 day cube compressive predicted by the M5’ model was the only output variable studied in this paper. This value was compared with the measured values of the compressive strength associated with the data records used in the development of the models to establish the accuracy of the model. It should also be mentioned that in cases where only cylinder-derived compressive strength was available in the dataset, the UNESCO [51] conversion factors, as reported by Elwell and Fu [52], were used to convert the cylinder compressive strength to 10 cm cube compressive strength.

MODELING AND RESULTS
As mentioned previously in the text, it will be statistically impossible to consider all individual variables and conditions during development of the model. It was therefore assumed that the data records extracted from the published literature represent average conditions (i.e. typical conditions that may be encountered at the construction projects). Consequently, it is proposed to use function shown in Eq. (6) to relate the CS<sub>RAC</sub> to the input parameters:

\[
CS_{RAC} = f_1(r, w/c, FA/TA, CA/C, W_a, SG_{SSD}) \tag{6}
\]

where all the terms are as previously described.

By assuming that \( f_1 \) is a power function, the general expression of the compressive strength of RAC can be written as
Predicting compressive strength of recycled concrete aggregate using M5' model

\[ \text{CSRAC} = a(r)^b \left( \frac{W}{C} \right)^c \left( \frac{FA}{TA} \right)^d \left( \frac{CA}{C} \right)^e (W_a)^f (SG_{SSD})^g \]  

(7)

where \( a, b, c, d, e, f, \) and \( g \) are constants of the equation and may attain different values in different expressions; all other terms are as previously defined.

Since model trees are only capable of producing linear relationships and multiple dependencies are non-linear in Eq. (7), the logarithmic transformation was applied to this type of regression.

Taking the logarithms of the variables in Eq. (7) results in the following linear formula:

\[
\log \text{CSRAC} = \log a + b \log(r) + c \log \left( \frac{W}{C} \right) + d \log \left( \frac{FA}{TA} \right) + e \log \left( \frac{CA}{C} \right) + f \log(W_a) + g(SG_{SSD})
\]

(8)

Eq. (8) can be transformed back to a form that predicts the CSRAC (i.e. Eq. (7)) by taking antilogarithms.

The “train-and-test” technique, being one of the most common approaches used in machine learning methods [19], was applied to develop the model. This technique starts with dividing the data set into two subsets (i.e., the train subset and the test subset). Training of the model is then performed using the train data subset. After training procedure is completed, the developed model is tested and verified by using the test data subset. In this study, 220 data sets were randomly selected and used to train the model and the remaining 50 data sets were used to test the model. The statistics of the parameters used for training the model are listed in Table 1. The developed model tree is presented in Fig. 1.

As can be seen in Fig.1, the generated model looks like an inverted tree in which the roots are located on the top while the branches and leaves are at the bottom. Prediction of CSRAC is achieved by introducing a new data point at the root of the tree and allowing it to find its way down to the leaves by passing through the nodes until reaching a linear model (LM) level. Table 2 shows the coefficients generated for each of the developed linear models (LMs) by the MT (M5') algorithm.

<table>
<thead>
<tr>
<th>Statistical Indicator</th>
<th>CSRAC (MPa)</th>
<th>(w/c)</th>
<th>W_a</th>
<th>SG</th>
<th>(FA/TA)</th>
<th>(CA/C)</th>
<th>(r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>84.87</td>
<td>0.86</td>
<td>28</td>
<td>2.75</td>
<td>0.58</td>
<td>7.40</td>
<td>1.00</td>
</tr>
<tr>
<td>Minimum</td>
<td>10.00</td>
<td>0.30</td>
<td>0.30</td>
<td>2.40</td>
<td>0.31</td>
<td>1.40</td>
<td>0.00</td>
</tr>
<tr>
<td>Mean</td>
<td>40.40</td>
<td>0.59</td>
<td>4.12</td>
<td>2.50</td>
<td>0.42</td>
<td>3.64</td>
<td>0.55</td>
</tr>
<tr>
<td>Standard Error</td>
<td>1.00</td>
<td>0.01</td>
<td>0.20</td>
<td>0.02</td>
<td>0.01</td>
<td>0.07</td>
<td>0.02</td>
</tr>
<tr>
<td>Sample Variance</td>
<td>269.42</td>
<td>0.02</td>
<td>11.26</td>
<td>0.01</td>
<td>0.01</td>
<td>1.26</td>
<td>0.16</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-0.30</td>
<td>-0.73</td>
<td>19.54</td>
<td>17.82</td>
<td>1.66</td>
<td>-0.37</td>
<td>-1.62</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.31</td>
<td>0.32</td>
<td>3.47</td>
<td>-4.32</td>
<td>1.16</td>
<td>0.34</td>
<td>-0.01</td>
</tr>
</tbody>
</table>
Fig. 1. The “model tree” generated by M5’ algorithm used to estimate the compressive strength of RAC.

Table 2. Predicted coefficients for various conditions shown in Figure 1

<table>
<thead>
<tr>
<th>Linear Model</th>
<th>Coefficient</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
<td>e</td>
<td>f</td>
<td>g</td>
</tr>
<tr>
<td>LM1</td>
<td>60.2698</td>
<td>-0.0027</td>
<td>-0.9546</td>
<td>0.8381</td>
<td>0.3117</td>
<td>-0.1013</td>
<td>-0.2742</td>
</tr>
<tr>
<td>LM2</td>
<td>28.9201</td>
<td>-0.0042</td>
<td>-0.6363</td>
<td>-0.0674</td>
<td>-0.0735</td>
<td>-0.0180</td>
<td>0.1149</td>
</tr>
<tr>
<td>LM3</td>
<td>28.7276</td>
<td>-0.0042</td>
<td>-0.5842</td>
<td>-0.0674</td>
<td>-0.0281</td>
<td>-0.0286</td>
<td>0.1149</td>
</tr>
<tr>
<td>LM4</td>
<td>22.3409</td>
<td>0.0065</td>
<td>-0.4564</td>
<td>-0.0674</td>
<td>0.0967</td>
<td>-0.0180</td>
<td>0.3985</td>
</tr>
<tr>
<td>LM5</td>
<td>38.3001</td>
<td>-0.0039</td>
<td>-0.2770</td>
<td>0.1535</td>
<td>0.0884</td>
<td>-0.0180</td>
<td>0.0652</td>
</tr>
<tr>
<td>LM6</td>
<td>39.4912</td>
<td>-0.0039</td>
<td>-0.2855</td>
<td>0.1850</td>
<td>0.0884</td>
<td>-0.0180</td>
<td>0.0652</td>
</tr>
<tr>
<td>LM7</td>
<td>35.8179</td>
<td>-0.0039</td>
<td>-0.2502</td>
<td>0.0044</td>
<td>0.0944</td>
<td>-0.0243</td>
<td>0.0652</td>
</tr>
<tr>
<td>LM8</td>
<td>45.7615</td>
<td>-0.0013</td>
<td>-0.1974</td>
<td>0.1451</td>
<td>0.0475</td>
<td>0.0139</td>
<td>0.0918</td>
</tr>
<tr>
<td>LM9</td>
<td>70.3558</td>
<td>0.0233</td>
<td>-0.2405</td>
<td>-0.0155</td>
<td>-0.1375</td>
<td>-0.2106</td>
<td>-0.3460</td>
</tr>
<tr>
<td>LM10</td>
<td>74.5933</td>
<td>0.0000</td>
<td>1.4031</td>
<td>0.4869</td>
<td>0.0186</td>
<td>-0.0767</td>
<td>-0.2487</td>
</tr>
<tr>
<td>LM11</td>
<td>0.0017</td>
<td>0.0075</td>
<td>-0.2870</td>
<td>-8.6682</td>
<td>1.9664</td>
<td>-0.0466</td>
<td>-1.2474</td>
</tr>
</tbody>
</table>

Error measurements, including discrepancy ratio (DR), mean absolute error (MAE) and root mean square (RMS) were used to evaluate the performance of the developed models. These parameters are defined as follows:

\[ DR = \log \frac{CS_{RAC_P}}{CS_{RAC_m}} \]  
\[ MAE = \frac{1}{N} \sum_{i=1}^{N} |DR_i| \]
Predicting compressive strength of recycled concrete aggregate using M5’ model

$(11)$

\[
RMS = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (DR_i)^2}
\]

where $CS_{RAC_m}$ and $CS_{RAC_p}$ are, respectively, measured and predicted cube compressive strength values of RAC and $N$ is the total number of data records.

A DR equal to zero represents an exact match between the predicted and measured compressive strength. Positive and negative signs for DR indicate, respectively, either the overestimation or underestimation. In statistics, MAE is a quantity used to measure how close predictions are to the measured values. The RMS, also known as quadratic mean, is a statistical measure of the magnitude of a varying quantity. This value is an average of the absolute error. This quantity is especially useful when variables are positive and negative. The values of MAE and RMS, when compared to zero, will also give indication of the performance of the model. An accurate model has MAE and RMS values close to zero. In this study, the accuracy is defined as the percentage of DR values that fall between -5 MPa and 5 MPa (i.e., it is assumed that 5 MPa error in prediction of the $CS_{RAC}$ is ignorable). Fig. 2 shows the histogram of DR values for the M5’ model. As can be seen, a high percentage of DR values are in the range between -5 and 5, which is an indication of the accuracy of the model.

The slope of regression line (SRL), correlation coefficient (CC) and coefficient of determination or ($R^2$) are also some other tools that can be used to evaluate the performance of a model. Correlations coefficient and coefficient of determination are defined as follows:

\[(12)\]

\[
CC = \frac{n \sum_i CS_{RAC_{pi}} CS_{RAC_{mi}}}{\sqrt{n \sum_i (CS_{RAC_{pi}})^2} \sqrt{n \sum_i (CS_{RAC_{mi}})^2}}
\]

\[(13)\]

\[
R^2 = \frac{\sum_i (CS_{RAC_{pi}} - \frac{1}{n} \sum_i CS_{RAC_{pi}}^2)}{\sum_i (CS_{RAC_{mi}} - \frac{1}{n} \sum_i CS_{RAC_{mi}}^2)}
\]

where $N$ is the total number of data records, $CS_{RAC_{pi}}$ is the predicted value of $CS_{RAC}$ for data record $i$ and $CS_{RAC_{mi}}$ is the measured value of $CS_{RAC}$ for data record $i$. 
The closer the value of the slope of the regression line for predicted versus the measured data to unity, the more accurate the model. Moreover, higher values for CC and $R^2$ also provide indication of a more precise and reliable model. Table 3 shows the mean values of the DR, ME, RMS, CC, $R^2$, and standard deviation of the training and testing data sets. The performance of both the training and testing sets of MT when compared with the actual values can be seen in Fig. 3. Error measurements indicate that M5’ model tree is able to predict the CS$_{RAC}$ with high accuracy and reliability.

*Table 3. MAE, RMS, CC, $R^2$ and Standard Deviation of MT (M5’) model*

<table>
<thead>
<tr>
<th>Data group</th>
<th>MAE (Target = 0)</th>
<th>RMS (Target = 1)</th>
<th>CC (Target = ±1)</th>
<th>$R^2$ (Target = 1)</th>
<th>SLR</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>0.034</td>
<td>0.047</td>
<td>0.97</td>
<td>0.94</td>
<td>0.868</td>
<td>0.075</td>
</tr>
<tr>
<td>Testing set</td>
<td>0.076</td>
<td>0.121</td>
<td>0.95</td>
<td>0.89</td>
<td>0.841</td>
<td>0.119</td>
</tr>
</tbody>
</table>

*Fig. 3. Performance of (a) training set, (b) testing set*

Generally, the MT algorithm has some advantages over other soft computing methods such as e.g. ANN. First, MT does not require much trial and error to obtain the best model. Moreover, piecewise regression analysis provides better understanding of the physics of the phenomenon compared to simple least square regression model, which gives only one equation. However, one of the main limitations of MT algorithm is that it only produces linear relationships. In addition, with regard to more complex cases, the transformation of input parameters may not be that simple and may not necessarily lead to a few simple linear formulas [18].

**CONCLUSIONS**

In this study, the M5’ model tree (MT) algorithm was used to develop a series of regression models that can predict the compressive strength of recycled aggregate concrete. The model was developed by using 220 data sets. Each data set consisted of mixture characteristics which were used as model input parameters (i.e. w/c, CA/C, FA/TA, SGSSD, and $W_3$) and measured compressive strength. That measured compressive strength was compared to the predicted compressive strength of a 10 cm RAC cube (i.e. model output). The performance of the developed model was evaluated by statistically comparing the predicted values of the compressive strength with the measured ones using various error measurement parameters. The results of these comparisons indicate that the model tree (MT) technique is able to
produce good predictions of compressive strength of RAC based on the known values of mixture proportions and selected characteristics of aggregates.

REFERENCES

12. Ameri, M., Behnoood, A., Laboratory studies to investigate the properties of CIR mixes containing steel slag as a substitute for virgin aggregates. Construction and Building Materials, 26, 2012, 475–480
29. Schoppe, B.M., Shrinkage and modulus of elasticity in concrete with recycled aggregates. MSc thesis, California Polytechnic State University, 2011, pp 131
30. Performance related approach to use of recycled aggregates. Project report to waste & resources action programme, 2007, pp 77
52. Elwell, J.D., Fu, G., Compression Testing of Concrete: Cylinders vs. Cubes. FHWA report, New York State Department of Transportation, 1995, pp 28