A new formulation for strength characteristics of steel slag aggregate concrete using an artificial intelligence-based approach

Paul O. Awoyera¹a, Iman Mansouri²*, Ajith Abraham³b and Amelec Viloria⁴a

¹Department of Civil Engineering, Covenant University, Ota, Nigeria
²Department of Civil Engineering, Birjand University of Technology, 97175-569 Birjand, Iran
³Machine Intelligence Research Labs, Auburn, Washington, 98071, United States
⁴Universidad de la Costa, Barranquilla, Colombia

(Received June 13, 2020, Revised February 25, 2021, Accepted March 4, 2021)

Abstract. Steel slag, an industrial reject from the steel rolling process, has been identified as one of the suitable, environmentally friendly materials for concrete production. Given that the coarse aggregate portion represents about 70% of concrete constituents, other economic approaches have been found in the use of alternative materials such as steel slag in concrete. Unfortunately, a standard framework for its application is still lacking. Therefore, this study proposed functional model equations for the determination of strength properties (compression and splitting tensile) of steel slag aggregate concrete (SSAC), using gene expression programming (GEP). The study, in the experimental phase, utilized steel slag as a partial replacement of crushed rock, in steps 20%, 40%, 60%, 80%, and 100%, respectively. The predictor variables included in the analysis were cement, sand, granite, steel slag, water/cement ratio, and curing regime (age). For the model development, 60-75% of the dataset was used as the training set, while the remaining data was used for testing the model. Empirical results illustrate that steel aggregate could be used up to 100% replacement of conventional aggregate, while also yielding comparable results as the latter. The GEP-based functional relations were tested statistically. The minimum absolute percentage error (MAPE), and root mean square error (RMSE) for compressive strength are 6.9 and 1.4, and 12.52 and 0.91 for the train and test datasets, respectively. With the consistency of both the training and testing datasets, the model has shown a strong capacity to predict the strength properties of SSAC. The results showed that the proposed model equations are reliably suitable for estimating SSAC strength properties. The GEP-based formula is relatively simple and useful for pre-design applications.

Keywords: concrete; steel slag; strength properties; genetic expression programming; experimental data

1. Introduction

Owing to the increasing demand for conventional construction materials, resulting from increasing urbanization and changing human lifestyles, there has been a growing interest, over the years, concerning sustainable development in the built environment. The solution to the scarcity of materials has been found in the use of alternative materials, routinely emanating as industrial or construction rejects. The aforementioned sources are known to pollute the environment and consequently creating various health measures.

Thus, sustainable measures majorly lie in finding lasting solutions to material use and environmental management. On the path of materials, numerous examples have been tested for the possibility of incorporation into fresh cementitious systems, thus ensuring that natural materials sources are preserved. These include materials originating locally (crushed rocks, stone dust, laterite) or those sourced from industrial, construction, and demolition debris (glass, ceramics, steel slags, marble).

Several experimental studies have dwelt on reusing waste materials for the production of new products, with significant results (Cheng et al. 2016, Erdem et al. 2018, Madurwar et al. 2013, Mehta and Ashish 2019, Paris et al. 2016). In the study by Paris et al. (2016), numerous supplementary cementitious materials (SCMs) were evaluated, and a nontraditional use of SCMs were proposed. From all, it is clear that steel slag, a byproduct of the steel rolling process, is one that has been overly explored (Jiang et al. 2018). The prospect of this material is high; a report by Guo et al. (Guo et al. 2018a) has shown that steel slag is strategically reused for various applications in industrial regions like Japan (98.4% rate) Europe (87.0% rate), and United States (84.4% rate). However, it is practically not feasible for steel slag to be used for the fresh production of steel, except only for secondary applications. In some less developed countries in Africa, steel slag is used for filling failed spots on roads or mostly piled up within the premises of the production site (Awoyera et al. 2016), thereby constituting a nuisance to the environment. The application of steel slag covers three categories, in the form of powder...
for cementitious mixture, fine and coarse aggregate in concrete. Steel slag, in powder form of about 5%, has been used to suppress the swelling potential and improved the strength of expansive soil (Wu et al. 2019). Zalnezhad and Hesami (2019) utilized steel slag aggregate and bitumen emulsion to enhance the performance of micro-surfacing mixture. It was reported that steel slag-containing mixtures exhibited a more pronounced behavior in terms of rutting and stripping distresses.

Artificial intelligence applications have been used in a wide range of engineering problems (Abraham et al. 2017, Abraham et al. 2008, Nedjah et al. 2009). Based on experimental studies, several predictive models are generated to facilitate the reuse of proposed products. At the same time, model development helps to prevent repeated experimentation and waste of material resources. The common predictors have been developed based on linear and multiple regression equations, artificial neural networks (ANN), and fuzzy analogy. All the methods have been tried on steel slag aggregate concrete (Awoyera 2018). However, a significant limitation on the methods above is that they only function as a predictor, without giving a correct formulation that determines the performance of the materials. In the last decade, several successful attempts have been made using gene expression programming (GEP), for solving various problems in the field of engineering (Abdollahzadeh et al. 2016, Castelli et al. 2017, Cladera et al. 2014, D’Aniello et al. 2015, Ebrahimzade et al. 2018, Golafshani et al. 2014, Güneyisi et al. 2013, Güneyisi and Nour 2019, Hodhod et al. 2018, Mahdavi Jafari and Khayati 2018, Mansouri et al. 2018, Mansouri et al. 2017, Mansouri and Farzampour 2018, Murad 2020, Nour and Güneyisi 2019, Saridemir 2016, Tsai and Liao 2019). GEP technique offers a cutting-edge solution to problems across the field of engineering. The solutions obtained from the analysis fit to satisfy sustainability considerations in the field of engineering. Therefore, in this study, a new formulation of strength characteristics of steel slag aggregate concrete (SSAC) based on the GEP technique is proposed. The study considers steel slag as a partial replacement for conventional aggregate (crushed rock), where other materials were kept constant. The need to replace the coarse aggregate portion was due to the fact that coarse aggregate constitutes the largest portion of the concrete constituents. So, it is considered an economical approach to introduce steel slag into a concrete mixture. Overall, the outcome of this study is expected to serve as a reference for researchers and other stakeholders in the construction and building sector.

2. State of the art on steel slag aggregate concrete

The production of concrete using steel slag has become popular in the last couple of decades. It has been reported from different quarters that steel slag aggregate has somewhat similar properties as the conventional aggregates (Maslehuddin et al. 2003, Subathra Devi and Gnanavel 2014, Yi et al. 2012). It is also commonly studied for use as supplementary cementitious material (Wang and Suraneni 2019). Although, steel slag tends to swell within a short space of time, however, studies (Akinwumi 2014, Oliveira et al. 2018) have suggested stockpiling steel slag in a cool environment for up to 14 days before use to control its swelling behavior.

In a study by Gupta and Sachdeva (Gupta and Sachdeva 2019), the potential of using Argon Oxygen Decarburization (AOD) steel slag for the production of concrete was investigated. Steel slag contents varying from 10-25% in step of 5% was used a partial replacement of cement for the development of concrete pavement. The study established that AOD slag could suitably replace segments in rigid pavement construction. In a related study, the strength and durability properties of a five years old concrete made with steel slag powder were evaluated (Han and Zhang 2018). The authors reported higher compressive strength, lower porosity, and very low permeability in concrete as steel slag content was increased. Ding et al. (2019) explored the pressure sensitivity of a new smart polymer concrete incorporating steel slag and graphite in an epoxy resin concrete, thus checking the pressure-sensitive properties of the concrete in uniaxial compression. Their study showed that there was an increase in the resistance and strain of the concrete containing graphite and steel slag, with evidence of concordant monotonocity. It has also been reported that steel slag, when used as a partial replacement of conventional aggregate in concrete, is effective for process such as: gamma radiation shielding (Baalamurugan et al. 2019), enhancing static and impact behaviors of concrete at about 20% substitution rate (Guo et al. 2018b), improving water permeability of pervious concrete (Lang et al. 2019), increasing the mechanical and fracture properties of roller-compacted concrete (owing to its sharp edges and surface roughness) (Rooholamini et al. 2019), long-term effect on compressive strength of ultra-high performance concrete (through filling effect) (Zhang et al. 2019).

In terms of microstructure, a study by Liu and Guo (Liu and Guo 2019) revealed the presence of a dense pore structure for hardened composite binders made with steel slag. It was shown that concrete after ten years possessed pore ranging from 3.2 to 20 nm. Also, it was reported that a hardened composite binder containing steel slag exhibited a higher amount of Portlandite (Ca(OH)₂). In another study, the durability (carbonation at 99.9% CO₂ and a pressure of 0.10 MPa) feature of concrete incorporating steel slag was evaluated (Mo et al. 2017). The results of the study showed a significant increase in the compressive strength of the concrete after it had been exposed to CO₂ curing.

Overall, findings have shown that steel slag exhibits appreciable features that support its utilization in cementitious composites. However, its full adoption is limited due to the lack of standard framework on its performance. Therefore, it is an appropriate step to develop a formulation to back the utilization of steel slag in the area of concrete production.

3. Description of the database

The formulation that is derived in this study is based on experimental results obtained by the author (Awoyera et al.
Table 1: Input and output data used for formulation

<table>
<thead>
<tr>
<th>Cement (kg/m³)</th>
<th>Sand (kg/m³)</th>
<th>Granite (kg/m³)</th>
<th>Steel slag (kg/m³)</th>
<th>Water/cement ratio</th>
<th>Age (days)</th>
<th>Compressive strength (MPa)</th>
<th>Split-tensile strength (MPa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>286</td>
<td>572</td>
<td>1144</td>
<td>0</td>
<td>0.5</td>
<td>7</td>
<td>21.02</td>
<td>2.3</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>1144</td>
<td>0</td>
<td>0.5</td>
<td>7</td>
<td>21.5</td>
<td>2.42</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>1144</td>
<td>0</td>
<td>0.5</td>
<td>28</td>
<td>22.3</td>
<td>2.47</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>915.2</td>
<td>228.8</td>
<td>0.5</td>
<td>7</td>
<td>23.23</td>
<td>2.59</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>915.2</td>
<td>228.8</td>
<td>0.5</td>
<td>14</td>
<td>23.82</td>
<td>2.62</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>915.2</td>
<td>228.8</td>
<td>0.5</td>
<td>28</td>
<td>25.2</td>
<td>2.68</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>686.4</td>
<td>457.6</td>
<td>0.5</td>
<td>7</td>
<td>26.48</td>
<td>2.99</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>686.4</td>
<td>457.6</td>
<td>0.5</td>
<td>14</td>
<td>27.57</td>
<td>3.02</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>686.4</td>
<td>457.6</td>
<td>0.5</td>
<td>28</td>
<td>29.6</td>
<td>3.08</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>457.6</td>
<td>686.4</td>
<td>0.5</td>
<td>7</td>
<td>31.1</td>
<td>3.39</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>457.6</td>
<td>686.4</td>
<td>0.5</td>
<td>14</td>
<td>32.19</td>
<td>3.42</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>457.6</td>
<td>686.4</td>
<td>0.5</td>
<td>28</td>
<td>32.7</td>
<td>3.5</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>228.8</td>
<td>915.2</td>
<td>0.5</td>
<td>7</td>
<td>33.15</td>
<td>4.19</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>228.8</td>
<td>915.2</td>
<td>0.5</td>
<td>14</td>
<td>35.34</td>
<td>4.24</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>228.8</td>
<td>915.2</td>
<td>0.5</td>
<td>28</td>
<td>37.2</td>
<td>4.34</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>0</td>
<td>1144</td>
<td>0.5</td>
<td>7</td>
<td>34.67</td>
<td>4.77</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>0</td>
<td>1144</td>
<td>0.5</td>
<td>14</td>
<td>36.91</td>
<td>5.3</td>
</tr>
<tr>
<td>286</td>
<td>572</td>
<td>0</td>
<td>1144</td>
<td>0.5</td>
<td>28</td>
<td>39.6</td>
<td>5.67</td>
</tr>
<tr>
<td>2015, Awoyera et al. 2016). In the studies, strength properties of steel slag aggregate concrete, covering compressive strength of 150 mm cubes and split-tensile strength of 150×300 mm cylinders, have been determined based on a variety of constituent materials and other key factors. Steel slag was used as a partial replacement of crushed rock; the water/cement ratio ranged from 0.5, 0.55 and 0.6, curing regime were 7, 14, and 28 days periods.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Both cement and river sand (fine aggregate) portions were kept constant throughout the experimentation. The database employed is presented in Table 1, showing the materials and factor variations and the corresponding strength outputs. It is noteworthy that the strength results presented were an average of a triplicate test on samples using a compression machine.

### 4. Modeling technique

#### 4.1 The theoretical background of GEP

Gene expression programming (GEP), a programming tool based on a genetic algorithm developed by Ferreira (Ferreira 2001), is well known for its capability in providing solutions to the complex situation of a relatively minimal data set. Predefined solvers are not necessary for the process. The use of GEP has become popular in several civil and structural engineering applications (Cevik and Soneri 2008, Farzampour et al. 2019, Ozbay et al. 2010, Tsi 2013). GEP is capable of predicting the performance of a structural component based on predefined experimental data set. Some application of GEP includes the prediction of strength, durability, and elastic properties of concrete.

The dynamic nature of the GEP makes it produce model output, which is coded in the form of chromosomes, using expression trees, Karva language, and many programming languages, such as VBA, MATLAB, CBB. This also allows GEP to generate empirical expressions for problems where there are no analytical expressions. Overall, the best fit experimental result is achieved by simply deleting or adding up the various factors involved. In GEP expressions, there are one or more elements called genes in the chromosomes, which have a head and tail (Murad 2020). The gene’s tail is comprised of terminal symbols, inform of constants or variables (1, a, b, c). However, the gene’s head is comprised of functions and terminal symbols (1, a, b, √, cos, *, /). The outputs in GEP are influenced by some factors, such as the complexity of the functions, which normally caused an increment in a number of genes, or a high number of chromosomes, which results in increased running time (Gandomi et al. 2014). Generally, modeling in GEP requires selecting a fitness function, and terminals and functions needed for building the chromosomes. This is then followed by the determination of the number of genes, head length, and the number of chromosomes, before finally choosing genetic operators and linking functions.

#### 4.2 GEP modeling

In this study, a popular software, GeneXproTools (GepSoft 2015), was used for the development of the GEP models for the prediction of strength properties of concrete containing steel slag. By randomly selecting gene numbers, head size, and linking functions, various GEP models were developed. However, a model that best fits the experimental data set was finally selected. Table 2 presents the parameters chosen for developing the GEP model.

Generally, the GEP model has a genome, which is made up of linear, symbolic string and chromosomes of fixed length. The genes represented in the expression trees of the model are of various sizes and shapes, and such genes are of fixed length. However, it generally not advisable to excessively increase the number of genes, as this could result in the development of complex GEP models. Mathematical functions and operators are employed in GEP model formulation. The use of mathematical operators does linking of genes. The genes are of the utmost impact on model development. The number of constants in the developed model is largely dependent on the constants present in the genes. The rate at which mutation takes place is backed by the creation of diversity and change of genome or replacements of elements (such as changing function or terminal with another). In the process of replacement of functions, new chromosomes are formed by linking original chromosomes. A combination of genes is carried out using functions such as ×, +, -, /, etc. The experimental data sets were used for model development. In general, GEP uses about 65-75% of the dataset as the training dataset, while the remaining data were used as a validation dataset.

The new formulation of strength characteristics of steel slag aggregate concrete (SSAC) was derived by adopting the experimental data set of the authors. The data sources are presented in section 3. In a specified order, all input data were used to develop a constant sequence of inputs employed in model development. The input nodes include the cement (C), sand (S), gravel (G), steel slag (Sj), water/cement ratio (w/c), and age (A).

The variables used in the model development in GEP includes, training data having both the input and the outputs. An optional set of data having an equal number and sequence of input/output variables was utilized for examining and testing the performance of the model generated. The statistical analyses of the data presented in Table 1 have been highlighted in Table 3. There it is clearly shown that there is a strong agreement between the training and testing data sets, and in essence, it is established that the two sets of data show an identical population. The parameters got the predictors are Cement (C), Sand (S), Granite (G), steel slag (Sj), water/cement ratio (w/c), and age (A).

In this study, gene expression programming was executed using GeneXproTools 4.0 software. Series of mathematical operations was adopted so as to ensure there is accuracy in the model. The GEP data were modified several times in order to achieve an optimum model having the best fitness characteristics. It is also good to state that the GEP model developed contains trigonometric

<table>
<thead>
<tr>
<th>Table 2 Parameters considered for GEP formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function set</td>
</tr>
<tr>
<td>Generation number                            5396101</td>
</tr>
<tr>
<td>Number of chromosomes                        150</td>
</tr>
<tr>
<td>Head size                                    10</td>
</tr>
<tr>
<td>Number of genes                              1</td>
</tr>
<tr>
<td>Linking function                             +</td>
</tr>
<tr>
<td>Mutation rate                                0.044</td>
</tr>
<tr>
<td>Inversion rate                               0.1</td>
</tr>
</tbody>
</table>
operations, in addition to some exponential functions (Majidifard et al. 2019). Eq. (1) was used to determine the correlation coefficient (R) of the model, which also indicates the fitness of the predicted variables as compared to experimental (actual) data. It is known statistically that a satisfactory model would have higher values of R coefficients. Such a model will possess a better capacity to approximate data.

\[
R = \frac{\sum(t_i - \bar{t}) (o_i - \bar{o})}{\sqrt{\sum(t_i - \bar{t})^2} \sqrt{\sum(o_i - \bar{o})^2}}
\]  
(1)

where \(o_{mean}\) and \(t_{mean}\) are the averages of the GEP model output \(o_i\) and target output \(t_i\) values, respectively.

The mathematical formulations for the compressive and split tensile strengths of SSAC are given in Eqs. (2) and (3), respectively.

\[
P = 0.1(C \times S)^{1/6} + \frac{S_g}{9.57(A-7.45)} + 10^{w/c} + 0.5G + 2S_g
\]  
(2)

\[
T = 0.11P
\]  
(3)

The equation terms are represented as follows: The constants in the first line of Table 3 are Cement \((C)\), Sand \((S)\), Granite \((G)\), steel slag \((S_g)\), water/cement ratio \((w/c)\), age \((A)\), compressive strength \((P)\), and split-tensile strength \((T)\).

In Fig. 1, the model expression tree is depicted, and Figs. 2 and 3 present the prediction performance of the proposed model.

From the model architecture shown in Fig. 1 and representation of the factors, it can, thus, be deduced that no strict dependence exists among the input factors, and the overall performance of the model.

Fig. 2 shows the comparison between the experimental and predicted compressive strength of SSAC, for the train and test data sets. The correlation coefficient for the training and testing datasets was calculated as 0.935 and 0.981, respectively.
The suitability in terms of accuracy and reliability of the proposed GEP model was assessed by comparing the performances of the GEP models to the experimental data for strength characteristics of steel slag aggregate concrete (SSAC). As presented in Figs. 4 and 5, there were fluctuations of normalized strengths of SSAC factors versus experimental values for compressive and split tensile strengths, respectively. After this, the study described and discussed the unique characteristics of each case.

5. Results and discussion

In general statistics, the value of $R^2$ should occur between zero to one, from which the value one is an indication that a perfect correlation is exhibited between the predicted and experimental datasets. Whereas, zero value implies that there is no correlation. However, it should be noted that an $R^2$ of 1 is never an indication that the prediction of the dataset is perfect (Sadeghian and Fam 2015). It is only an indication that both the predicted and the experimental datasets can be linearly correlated. Therefore, in this study assumptions and accuracy of the proposed model were not based on $R^2$ values only. Besides, the study adopted other proven statistical indexes, covering mean square error (MSE), mean absolute percentage error (MAPE), and root means square error (RMSE). They were used to establish the performance of the models.

Mathematical expressions for the error values are presented in Eqs. (4)-(6). With this data, the capability of the models is established based on the, lowest values of MAE and MAPE.

\[
MSE = \frac{1}{N} \sum_{i=1}^{N} (o_i - t_i)^2 \tag{2}
\]

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (o_i - t_i)^2} \tag{3}
\]

\[
MAPE = 100 \times \frac{1}{N} \left| \frac{t_i - o_i}{t_i} \right| \tag{4}
\]

where $o_i$ is GEP model output, $t_i$ indicate target output, and $N$ denotes the number of the sample in the database.
Tables 4 and 5 show the stated statistical parameters that were used for the prediction models for the compressive and split tensile strengths of SSAC, respectively. Table 4 shows that the RMSE of the developed GEP model was approximately 1.0 MPa and 0.2 MPa for compressive and split tensile strengths. The low error values in the GEP models confirm that the proposed models are reliable and accurate. This is established in the field of statistical evaluation (Gandomi et al. 2017). Overall, the developed model is well suited for concrete produced using ground granulated blast furnace slag, and its production within the confines of the factors considered for this experimentation. This model is expected to serve as standard framework for application of steel slag for concrete strength prediction.

6. Conclusions

A method is introduced that uses gene-expression programming (GEP) symbolic regression to form a nonlinear combination of steel slag aggregate concrete (SSAC) strengths. Motivated by the difficulty in forecasting compressive and split tensile strengths of steel slag aggregate concrete, we test the ability of GEP to predict the strengths. Input to GEP is 6 independent parameters. The data consists of 54 samples of experimental tests conducted by the authors. The GEP-based model accurately predicts the strengths of SSAC. The proposed model simultaneously takes into account the role of several important factors, including cement, sand, gravel, steel slag, water/cement ratio, and age, representing the behavior of the concrete strength.

The validity of the model was tested for a part of test results beyond the training data domain. The validation phases confirm the efficiency of the model for its general application to the compressive and split tensile strengths of SSAC. The minimum absolute percentage error (MAPE), and root mean square error (RMSE) for compressive strength are 6.9 and 1.4, and 12.52 and 0.91, for the train and test datasets, respectively. With the consistency of both the training and testing datasets, the model has shown a strong capacity to predict the strength properties of SSAC. For the split-tensile strength, minimum absolute percentage error (MAPE), and root mean square error (RMSE) are 12.30 and 0.4, and 4.97 and 0.17, for the train and test datasets, respectively.

Based on the results of the study, it is shown that the GEP models proposed are efficient for practical use of SSAC strength assessment. Moreover, the study revealed that there is no strict dependence among the input factors and the model performance, based on the evaluation of the factor interactions.

References


