Computers and Concrete, *Vol. 24, No. 4* (*2019*) *295-302*
DOI: https://doi.org/10.12989/cac.2019.24.4.295

Prediction of the compressive strength of fly ash geopolymer concrete using gene expression programming

Iyad S. Alkroosh*[1] and Prabir K. Sarker[2a]

[1]*Department of Civil Engineering, College of Engineering, University of Al-Qadisiyah, Baghdad Rd., Al Diwaniyah, Republic of Iraq*
[2]*School of Civil and Mechanical Engineering, Curtin University, Kent Street, Perth, Australia*

(*Received March 21, 2019, Revised July 9, 2019, Accepted August 10, 2019*)

Abstract. Evolutionary algorithms based on conventional statistical methods such as regression and classification have been widely used in data mining applications. This work involves application of gene expression programming (GEP) for predicting compressive strength of fly ash geopolymer concrete, which is gaining increasing interest as an environmentally friendly alternative of Portland cement concrete. Based on 56 test results from the existing literature, a model was obtained relating the compressive strength of fly ash geopolymer concrete with the significantly influencing mix design parameters. The predictions of the model in training and validation were evaluated. The coefficient of determination (R^2), mean (μ) and standard deviation (σ) were 0.89, 1.0 and 0.12 respectively, for the training set, and 0.89, 0.99 and 0.13 respectively, for the validation set. The error of prediction by the model was also evaluated and found to be very low. This indicates that the predictions of GEP model are in close agreement with the experimental results suggesting this as a promising method for compressive strength prediction of fly ash geopolymer concrete.

Keywords: geopolymer concrete; prediction; GEP; compressive strength; training; validation

1. Introduction

Maintaining a healthy and clean environment has become a subject of interest for most nations worldwide. Emissions of carbon dioxide (CO_2) from industrial areas such as cement factories pose a significant threat to the environment. The estimated contribution of cement factories in carbon dioxide emission to atmosphere is between 5%-8% of globally liberated CO_2 (Rubenstein 2012, Andrić *et al.* 2015). About one tonne of CO_2 is released to the atmosphere by the production of tonne of cement (Roy 1999). Uses of industrial by-products such as fly ash as a cement replacement has emerged as an effective strategy to overcome the environmental damage from cement industry and in response to the demand for maintaining a high-quality environment. Fly ash geopolymer concrete provides significant environmental benefit by reducing the carbon emission of concrete production (ACAA 2003). Geopolymer is produced by the reaction of an aluminosilicate source material such as low calcium fly ash with an alkaline liquid such as a mixture of sodium hydroxide (NaOH) and sodium silicate (Na_2SiO_3) (Davidovits 1999, Palomo *et al.* 1999, Barbosa *et al.* 2000). Fly ash from the bituminous and anthracite coals is usually referred as Class F or low-calcium fly ash. It consists of mainly an aluminosilicate glass and has less than 10 percent CaO (Heidrich 2002). The reaction product is an inorganic polymer and has a strong bonding property. This inorganic binder is used to bind the aggregates together in geopolymer concrete instead of traditional Portland cement binder.

Since the reaction product of geopolymer binder is chemically different from the hydration product of Portland cement, the mechanical properties of geopolymer concrete vary from those of Portland cement concrete. Therefore, the correlations developed for predicting physical behaviour of concrete based on Portland cement cannot be directly applied to geopolymer concrete. This urges on the necessity for developing models to predict the mechanical properties of fly ash geopolymer concrete in order for its extensive use by the construction industry. The mechanical properties of concrete such as tensile strength, flexural strength and modulus of elasticity are usually correlated well with its compressive strength. These mechanical properties of concrete can be predicted from the compressive strength by using the relationships given in the concrete design codes and standards. Also, compressive strength is used as the main target parameter in the mix design of concrete. Therefore, it is essential to understand the effect of each mix variable and predict the compressive strength in order to conduct the mix design of geopolymer concrete.

It has been found from a review of the existing literature that compressive strength of low-calcium fly ash geopolymer concrete depends on several factors such as binder content, sodium hydroxide molarity, curing temperature and the duration of curing (Hardjito and Rangan 2005). Palomo (1999) determined the alkaline activator type, curing temperature and curing time as the factors affecting the reaction rate of fly ash based

*Corresponding author, Associate Professor
E-mail: iyad.alkroosh@qu.edu.iq
[a]Associate Professor
E-mail: P.Sarker@curtin.edu.au

Copyright © 2019 Techno-Press, Ltd.

ISSN: 1598-8198 (Print), 1598-818X (Online)

geopolymer. Van Jaarsveld *et al*. (2002, 2003) concluded that the water content, and the curing and calcining condition of kaolin clay affected the properties of geopolymers. They also determined that source materials particularly CaO and water-to-fly ash ratio govern the properties of geopolymers. Barbosa *et al.* (1999) reported the importance of the molar compositions of the oxides present in the mixture and the water content. van Deventer (2000), Swanepoel and Strydom (2002), Xu and van Deventer (2002) investigated the factors that influence the compressive strength of geopolymers and determined that the percentage of CaO and K_2O, the molar ratio of Si to Al in the source material, the type of alkali liquid, the extent of dissolution of Si, and the molar Si-to-Al ratio in the solution significantly influenced the compressive strength.

A comprehensive review of the literature has determined that a reliable model for predicting the compressive strength of fly ash geopolymer concrete has not been obtained so far. The review also found that employing the traditional statistical regression methods for correlating the compressive strength with its factors would not bring a viable solution due to the complexity of the problem. Artificial intelligence techniques can be efficient in dealing with such problem through utilising computers to perform enormously iterated work and develop the problem's solution without any prior assumptions to the form of the relationship between the input and the output.

The commonly known artificial intelligence techniques are artificial neural networks (ANNs) and evolutionary algorithms such as genetic programming (GP). In the field of civil engineering and the concrete area, number of scholars attempted solutions of complex problems via the application of ANNs. Mukherjee and Deshpande (1995) investigated the suitability of ANNs for modelling an initial design of reinforced-concrete rectangular single-span beams and obtained good results. Yeh (2006) explored the use of ANNs for modelling slump of fly ash and slag concrete and concluded that the complex nonlinear relationship between concrete components and concrete slump was predictable. Younis and Pilakoutas (2013) applied ANNs and developed a relationship correlating the properties of recycled aggregate (RA) with the strength of recycled aggregate concrete. Duan *et al.* (2013) used ANNs for predicting the elastic modulus of recycled aggregate concrete and demonstrated that the constructed ANN model could predict well the elastic modulus of concrete made with recycled aggregate derived from different sources. Deshpande *et al.* (2014) utilised back propagation network to predict the 28-day compressive strength of recycled aggregate concrete (RAC). They determined that ANN learn from the examples and grasps the fundamental domain rules governing strength of concrete. Lingama and Karthikeyan (2014) applied ANNs for predicting compressive strength of high-performance concrete containing binary quaternary blends and concluded that ANN is convenient and time saving tool. Behnood *et al.* (2015) used M5' model tree algorithm to predict the elastic modulus of recycled aggregate concrete. They found that the M5' algorithm model has accuracy over 80 percent, which is well above the accuracy of other models. Gonzalez-Taboada *et al.* (2016) proposed expressions for the modulus of elasticity and the splitting tensile strength of structural recycled concretes. They used multivariable regression and genetic programming for the development of the expressions. Gazder *et al.* (2017) modelled the compressive strength of blended cement concrete using ANN. They determined that the use of 2-phase learning algorithm provide better predictions than single-phase traditional ANN models.

However, ANN techniques can provide solutions of problems through black-box models. They provide a solution of the problem under consideration in the form of connecting weights between input and output. GP on the other hand provides a solution of the problem in the form of parse trees which illustrates the interaction between the input variables and the output. Therefore, GP can give better access for understanding the nature of the relationship between the input and the output. The increasing number of applications of this technique in Civil Engineering field confirms the feasibility of its use as a modelling tool. Chen (2003) applied macroevolutionary genetic programming to estimate the compressive strength of concrete. Ozbay (2008) showed that genetic programming was capable in formulation of slump flow diameter, V-funnel flow time, compressive strength, ultrasonic pulse velocity and electrical resistivity of SCCs. Sonebi and Cevik (2009) formulated the fresh and hardened properties of self-compacting concrete containing pulverised fuel ash using genetic programming. The obtained formula showed a good level of accuracy in predicting the slump flow. Gandomi *et al.* (2010), presented a study involved application of genetic programming for deriving a model to predict the compressive strength of carbon fibre-reinforced plastic (CFRP) confined concrete cylinders. They proposed a formula to predict the ultimate compressive strength of concrete cylinders with an acceptable level of accuracy. Kiani et al. (2016) utilised evolutionary approach and proposed a simple model to predict the compressive strength preformed-foam cellular concrete. Castelli *et al.* (2017) developed a model to predict strength of high-performance concrete based on GP and they showed that semantic genetic programming can speed the convergence of the search process.

Gene expression programming (GEP) is an extended version of GP, in which linear chromosomes consisting of multi genes evolve to solve the problem. In recent engineering applications, GEP has been found more efficient than GP in modelling process. Nazari *et al.* (2011) modelled the split tensile strength and percentage of water absorption of concrete containing TiO_2 based on ANNs and GEP. Mousavi *et al.* (2012) demonstrated that gene expression programming was effective in developing a model for prediction of the compressive strength of high-performance concrete. Alkroosh and Ammash (2015) applied GEP for modelling punching shear strength of normal and high strength reinforced concrete flat slabs. The developed model was accurate and predicts shear strength more accurately than traditional methods. Abdollahzadeh *et al.* (2016) developed 20 models for predicting the compressive strength of recycled aggregate concrete containing silica fume using GEP. The models showed good prediction. Saridemir (2016) utilised GEP and obtained a model to predict flexural and splitting tensile strength of

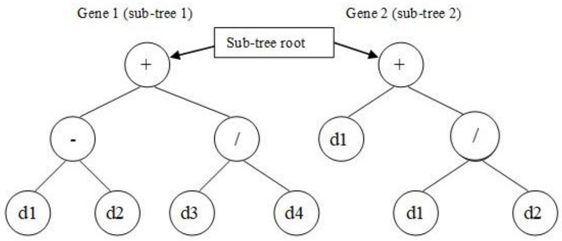

Fig. 1 Expression tree of the written chromosome

concrete containing different proportions of fly ash. This study employed the GEP for developing a model to predict the compressive strength of fly ash based geopolymer concrete, evaluated the accuracy of the developed model and conducted a parametric analysis to investigate the effects of input variables on the compressive strength.

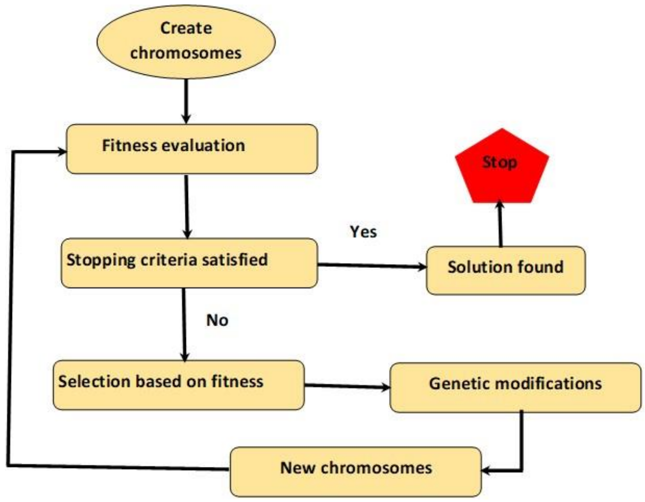

Fig. 2 Model development by GEP algorithm

2. Gene Expression Programming (GEP)

As a problem-solving approach inspiring natural phenotype/genotype system, gene expression programming was invented by Ferreria (2000). It can be described as a hybrid progeny of two parents of evolutionary algorithms namely genetic algorithms (GA) and genetic programming (GP) created to implement a robust problem-solving strategy. The inherent similarity of GEP from GA is that it uses the evolution of linear computer programs (individuals or chromosomes) of fixed length while from GP inherited expression) trees (ETs) which are ramified structures of different sizes and shapes expressing the evolved programs. Chromosome, gene and expression tree are the principal actors of gene expression programming. The chromosome is composed of genes which are encoded in smaller sub-programs. Every gene has a constant length and includes a head that contains functions (e.g., +, -) and terminals (e.g., d1, d2, which are the symbolic representations of the input variables), and a tail composed of terminals only. A typical GEP gene is written as follows: +.+.-./.*d1*./.**d1.d2.d3.d4.d1.d2**, where: “**.**” is the separation mark between the symbols, and **d0**, **d1**, and **d2** are variables known as terminals. The functions and italic symbols represent the gene head, while the black bold symbols represent the tail. This written format is named *K*-expression or *Karva* notation (Ferreira 2002), which can be converted into the ET as shown in Fig. 1. The tree is a spatial illustration demonstrating the interactions among the gene’s components on the map of solution.

In GEP, evolution is performed during each iteration via subjecting the chromosomes to genetic variation introduced by mutation and other genetic operators and recombination. Mutation means randomly selecting any component of the gene's head or tail and replacing it with any other randomly selected component from the function or terminal set. Rrecombination is the process in which two chromosomes are paired and split at the same point to exchange their components downward to the merging point.

As illustrated in Fig. 2, the GEP model development begins with creating an initial population of computer programs chosen randomly from the predefined sets of functions and terminals. The functions can contain basic mathematical operators (e.g., +, −, ×, /) or any other user-defined functions, whereas the terminals may consist of numerical constants, logical constants or variables. Each program (chromosome) is executed and its fitness is evaluated through the fitness function, which measures how well the chromosome is in competition with the rest of population.

Chromosomes are then selected for further development based on their fitness. The ones that have a higher fitness level are given a higher chance of being reselected, whereas the chromosomes with less fitness are deleted or given a slim chance of reselection. The selected programs are then exposed to further developmental operations, which are performed through genetic variations such as mutation and recombination. New offspring of chromosomes with new traits are generated and used to replace the existing population. The chromosomes of the new generation are then subjected to the same developmental process, which is repeated until the stopping criteria are satisfied.

3. Model development

3.1 Data input

The data used for development of the GEP model was obtained from the literature and comprise results of 56 experiments compiled by Hardjito and Rangan (2005). The tested geopolymer concretes were made of low calcium (ASTM Class F) fly ash, an alkaline liquid consisting of a mixture of sodium hydroxide and sodium silicate, and coarse and fine aggregates. Some of the mixtures contained extra water to improve workability, especially when the alkaline liquid content was low. The amount of used fly ash was in a range of 408 kg/m^3 to 476 kg/m^3. The concentration of NaOH ranged from 8M to 16M. The total volume of the aggregates was 75% to 80% of the concrete. The added extra water varied from 0 to 27 kg/m^3. The freshly mixed concretes had adequate workability. The

Table 1 Data division statistics

Variable	Data set	Mean	Standard deviation	Minimum	Maximum	Range
Alkaline/fly ash (by mass)	Training set	0.36	0.01	0.35	0.40	0.04
	Validation set	0.36	0.01	0.35	0.40	0.04
$Na_2SiO_3/NaOH$(by mass)	Training set	2.42	0.35	0.40	2.51	2.11
	Validation set	2.30	0.60	0.40	2.51	2.11
$NaOH$ molarity	Training set	12.42	2.64	8.00	16.00	8.00
	Validation set	11.85	2.88	8.00	14.00	6.00
Added water/fly ash (by mass)	Training set	0.02	0.02	0.00	0.06	0.06
	Validation set	0.02	0.02	0.00	0.05	0.05
Curing temperature (°C)	Training set	62.79	17.70	30.00	90.00	60.00
	Validation set	57.69	16.02	30.00	90.00	60.00
Curing time (hr)	Training set	22.23	7.62	4.00	48.00	44.00
	Validation set	22.46	5.55	4.00	24.00	20.00
Gravel/fly ash (by mass)	Training set	2.93	0.18	2.70	3.20	0.50
	Validation set	2.97	0.21	2.70	3.20	0.50
Sand/fly ash (by mass)	Training set	1.43	0.16	1.20	1.60	0.40
	Validation set	1.42	0.15	1.20	1.60	0.40
Compressive strength (MPa)	Training set	47.02	14.77	8.00	89.00	81.00
	Validation set	45.38	14.00	17.00	63.00	46.00

initial heat curing temperature and time varied from 30°C to 60°C and from 4 hours to 24 hours, respectively. The specimens were stored at room temperature after the initial heat curing and tested at the age of 7 days. The compressive strength tests were carried out on 100×200 mm cylinder specimens in accordance with the Australian Standards (AS 1012.9 2014). The reported compressive strength results were the mean values obtained from at least 3 identical specimens.

3.2 Data division

Obtaining a robust model using artificial intelligence requires dividing the data input into training and validation sets. This step is undertaken in order to ensure that the developed model will generalise when implemented. Artificial intelligence models are extracted using large size of data. When there is noisy data, the model tends to overfit which is the ability of the model to memorise rather than generalise the relationship between input and output. To avoid overfitting, the data are divided into two sets, namely training set and validation set. The training set represents a part of the data used for calibration of the model which involves implementing modelling strategy to predict the unknown relationship between the input and the corresponding output. The validation set represents a part of the data which is not included during training phase used to evaluate the generalisation capability of the model. The ratio of training and validation set is not definite, however researchers (e.g., Master 1993, Ferreria 2000) suggest that 80-90% of the data be assigned to training set and 10-20% to the validation set.

The curtail part of data splitting is separating the data sets into training and validation. Several methods are suggested for data division such as simple random sampling, trial and error method, systematic sampling and convenience sampled (Reitermanova´ 2010). After a careful consideration of the proposed methods of data division, randomly dividing the data into two statistically consistent sets was found as the appropriate method for this study. In this method, the mean and standard deviation of the training set of data need to be as close as possible to those of the validation set of data. This approach was recommended by Master (1993). To achieve this, the method proposed by Shahin *et al.* (2004) was adopted. The total data set consisting of 56 experimental results was divided into statistically consistent training and validation sets. Forty-three experimental results (80%) of the total data were used for training and 13 experiments (20%) of the total data were used for validation.

The extreme values are included in the training set, as artificial intelligence techniques perform better in interpolation rather than extrapolation. The data division statistics are presented in Table 1.

3.3 Identification and selection of input variables

For identification of the input variable, the approach suggested by Dreyfus (2005) was adopted. The author recommends that in process such as physical or chemical the variables that have influence on the output are analysed in detail by experts. Therefore, when selecting the variables that influence the output, experts should be consulted. This approach was adopted and the significant factors that govern the compressive strength of fly ash geopolymer concrete were identified through a comprehensive study of the available relevant literature. It was concluded that the molarity of sodium hydroxide, sodium silicate to sodium hydroxide ratio, added water, curing time, curing temperature and amount of coarse and fine aggregate have strong influence on the compressive strength of the concrete. Therefore, these factors were included as input variables of the GEP model. The measured compressive strength was the dependent targeted output.

Table 2 Optimal model setting

Parameter	Optimal value
Number of chromosomes	25
Number of genes	3
Head size	18
Function set	+, -, ÷, ×, x^2, x^3, $\sqrt[3]{...}$
Fitness function	MAE
Linking function	-
Mutation rate	0.05

In order to obtain a model of good performance, the input parameters were selected with considerations to exclude the parameters of insignificant influence on the performance. Lachtermacher and Fuller (1994) suggested that when developing a model by GEP, reducing the number of input variables will result in evolution of small size chromosomes. Consequently, quicker convergence can take place. Therefore, the authors attempted to keep the number of input variables small and exclude the insignificant ones.

3.4 Modelling using GEP

The development of a GEP model began with determination of the optimal modelling setting such as number of chromosomes, chromosome structure, functional set, fitness function, linking function and rates of genetic operators. Choosing proper values of those parameters expedite access to a robust solution (Ferarria 2002).

In this study, the optimal values of parameters were determined consecutively in several attempts using commercially available software package GeneXproTools 4 (2001). During each attempt, program runs were conducted in which the value of one parameter was varied whereas the values of other parameters were set constant. Each run was stopped after 50,000 generations after which minimal improvement on the fitness of the output was noticed. The mean absolute error (MAE) for training and validation set was evaluated at the end of each run. When MAE of the two sets was small and as close as can be, the corresponding value of the parameter was selected as optimal. Table 2 presents the optimal parameters used for the model development.

Development of the GEP model continued by conducting runs using the optimal parameters provided in Table 2. At the end of each run, a proposed solution of the studied problem was designated. The best solution was selected by analysing the generated solutions and evaluating them based on the selection criteria. The selection criteria used in this study were that the expression should be consistent with the engineering understanding of the compressive strength of fly ash geopolymer concrete; it shows the minimum scatter around the line of equality of predicted and experimental values i.e., coefficient of determination, R^2>0.80 for both training and validation sets; and the mean value of measured to predicted ratios is within 10%. If two or more expressions satisfied the selection criteria, the shorter and simpler one was selected for further improvement in the following step. The model simplification was the last step before accepting the GEP model.

4. Results and models evaluation

The GEP model was given in a form of expression tree consisting of three sub-trees similar to the example shown in Fig. 1. The tree was translated to mathematical expression by Eqs. (1)-(4).

$$f_c = f_1 - f_2 - f_3 \tag{1}$$

$$f_1 = \left[d_5 - d_8\left(d_6 + d_6^3 - d_8^2 + \sqrt{d_5 - d_4} - d_1\right)\right] - (d_6 - d_2) \tag{2}$$

$$f_2 = \frac{d_3\begin{bmatrix} d_5(d_3 + d_4 + d_5)^{0.5} + (d_4 d_6 - d_1) \\ -\left(\frac{9.954 + d_7}{2}\right) - 6d_1 d_6 \end{bmatrix}}{d_8} \tag{3}$$

$$f_3 = \frac{[\{d_0(d_1 + d_7 - 2.528)^{0.5}\} + \pi + d_7]\left(\frac{d_5}{d_0 d_4 + 2.924 d_0}\right)}{d_6^3} + d_6 \tag{4}$$

Where, f_c: compressive strength of fly ash geopolymer concrete in MPa; d_0: ratio of sodium silicate to sodium hydroxide by mass; d_1: sodium hydroxide molarity; d_2: ratio of alkaline liquid to fly ash by mass; d_3: ratio of added water to fly ash by mass; d_4: curing time in hours; d_5: curing temperature in degree Celsius; d_6: curing method taken as 1 for oven curing and 2 for steam curing; d_7: ratio of gravel to fly ash by mass; d_8: ratio of sand to fly ash by mass.

4.1 Evaluation of accuracy of the model

The performance of the GEP model in predicting the compressive strength of fly ash geopolymer concrete was evaluated by calculating coefficient of determination, R^2, mean (μ), standard deviation (σ), mean squared error (MSE), mean absolute error (MAE) and root mean squared error (RMSE) for the training and validation sets. The results of calculation were presented in Table 3.

R^2 was determined to evaluate the correlation of predicted and measured values and how close the points of correlation to the line of best fit. The optimal value of R^2 is unity which means that the predicted value is equal to the measured value. The coefficient of determination was calculated as per Rodgers and Nicewander (1988) from Eq. (5).

$$R = \left[\frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(f_i - \bar{f}\right)}{\sqrt{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 \sum_{i-1}^{n}\left(f_i - \bar{f}\right)^2}}\right]^2 \tag{5}$$

Where, R: coefficient of determination; y_i: experimental value; f_i: predicted value; $\bar{y}$: mean of experimental values; $\bar{f}$: mean of predicted values; n: number of observations.

As shown in Table 3, the value of R^2 is 0.89 for both the training and validation sets. This means that the GEP model

Table 3 Statistical evaluation of the model robustness in training and validation

Statistical measure	Training set	Validation set
R^2	0.89	0.89
μ	1.00	0.99
σ	0.12	0.13
MAE (MPa)	4.07	3.51
MSE $(MPa)^2$	24.42	22.83
RMSE (MPa)	4.94	4.78

is robust in predicting the compressive strength, as eighty nine percent of the variance in measured output is predictable from the input. The propinquity of R^2 for the two sets also indicate that the model has a good generalisation.

The mean of predicted to measured values was calculated to evaluate how on average the predicted values were close to the measured. The ideal value of mean is 1.0 indicating that on average the predicted values are equal to targeted values. If mean value is more than 1.0, on average the model tends to over-predict the targeted output. Conversely, the model under-predicts the targeted output if the mean value is less than 1.0.

The mean was calculated from Eq. (6).

$$\mu = \frac{1}{n}\frac{y_i}{f_i} \tag{6}$$

Where, n: number of observations; y_i: experimental value; f_i: predicted value.

As shown in Table 3, the calculated mean values of training and validation sets were 1.0 and 0.99, respectively. This indicates that the model is robust but slightly tends to under-predict the compressive strength which is conservative for mix designs.

The standard deviation (*SD*) of predicted to measured compressive strength was calculated to inspect how the values of the ratio of predicted to measured compressive strength were spread out from the mean. The closer the value of *SD* to zero indicates that most of the values scatter near to the average and the prediction is good, whereas closer to one value of *SD* means that the values are spread out indicating to poor prediction. The *SD* was calculated from Eq. (7).

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \tag{7}$$

Where, σ: standard deviation; x_i: ratio of measured to predicted strength of case i; $\bar{x}$: mean ratios of measured to predicted strength of all cases; n: number of observations.

As presented in Table 3, the calculated *SD* values were 0.12 and 0.13 for the training and validation set, respectively. This shows minimal number of outliers and provides evidence of strong performance of the model.

The average error was calculated to evaluate accuracy of the GEP model in prediction the compressive strength. The error was calculated from Eqs. (8)-(10).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| \tag{8}$$

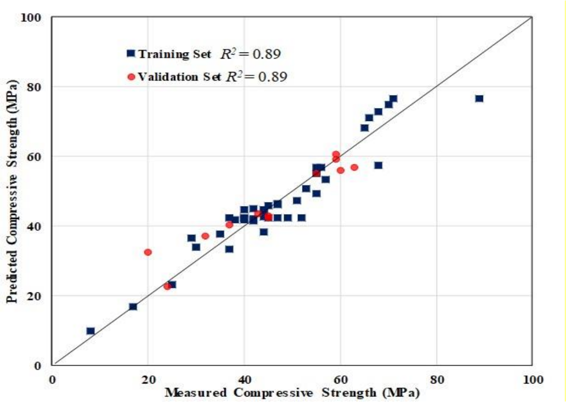

Fig. 3 Performance GEP model in training and validation sets

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2 \tag{9}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2 \tag{10}$$

Where, n: number of observations; f_i: predicted value; y_i: experimental value.

Table 3 shows the calculated errors in predictions of the developed model in training and validation sets. The error is relatively small and the three equations give error in training set slightly different with the error in validation set. This indicate to good prediction and generalization.

The plot of measured versus predicted compressive strength, as depicted in Fig. 3, illustrates a minimum scatter of the points around the line of equality for both training and validation sets. This is consistent with the error measures presented in Table 3 and confirms the robustness of the developed GEP model in predicting compressive strengths of fly ash geopolymer concrete.

4.2 Parametric study

This study was conducted to comprehend the behaviour of the model in relation to the variations of input variables and assess whether the obtained results agree with experimental results. The study involved assuming a hypothetical input which lies within the range of training data and examining the output when one input variable was varied whereas the others were set constant to their mean values. The graphical presentation of parametric study is shown in Fig. 4.

Inspection of the figure shows that compressive strength increases at different rates with the increase of each of the input variables except with amount of water, which inversely affects the compressive strength. The results of the parametric study are consistent with experimental findings. The laboratory test results show that compressive strength increases with the increase of sodium silicate to sodium hydroxide ratio, sodium hydroxide molarity and curing temperature. On the other hand, the compressive strength decreases with the increase of amount of water added. The results of parametric analysis agree well with

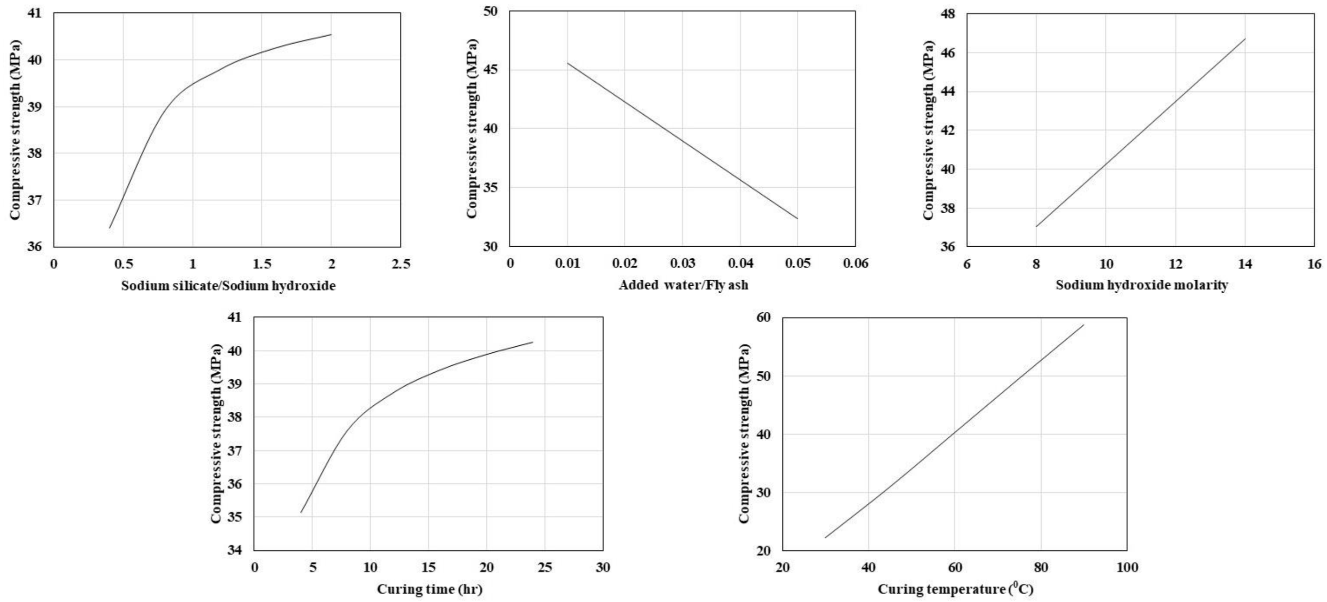

Fig. 4 The effects of input variables variations on compressive strength

the experimental results, hence adding reliability to the developed model.

The significance of each independent input variable in relation to the compressive strength was also investigated in the parametric study. The concentration of sodium hydroxide, curing time and temperature are the factors that have strong influence on compressive strength. This is because the concentration of the alkali, curing temperature and curing time accelerate the reaction of fly ash with the alkali producing stronger geopolymer binder. On the other hand, the added free water creates more voids in the hardened reaction product resulting in weaker geopolymer binder.

5. Conclusions

This study presented an application of GEP to correlate between the compressive strength of fly ash geopolymer concrete and its governing mix design parameters. The adopted algorithm settings, which are number and structure of chromosomes, fitness function, linking function and rates of genetic operators, for model development was effective in bringing satisfactory output in small number of generations. The values of evaluation measures determined by statistical analysis indicate a strong performance of the model. The graphical presentation of predicted versus experimental compressive strength show minimum scatter around the line of equality. All the considered variables have different levels of effect on the compressive strength of fly ash geopolymer concrete. A parametric study using the developed model has shown that concentration of sodium hydroxide, curing temperature and curing time are the most influential factors on the compressive strength and the amount of water affect the strength inversely. It was also found that results of the study were in a good agreement with experimental findings. Therefore, the results confirm that GEP was capable to develop a robust model for predicting the compressive strength of fly ash geopolymer concrete with a good level of accuracy.

References

Abdollahzadeh, G., Jahani, E. and Kashir, Z. (2016), "Predicting of compressive strength of recycled aggregate concrete by genetic programming", *Comput. Concrete*, **18**(2), 155-163. http://dx.doi.org/10.12989/cac.2016.18.2.155.

ACAA (American Coal Ash Association: 74) (2003), Fly Ash Facts for Highway Engineers, Aurora, USA.

Alkroosh, I. and Ammash, H. (2015), "Soft computing for modelling punching shear of reinforced concrete flat slabs", *Ain Shams Eng. J.*, **6**(2), 439-448. https://doi.org/10.1016/j.asej.2014.12.001.

Andrić, I., Jamali-Zghal, N., Santarelli. M. and Le Corre, O. (2015), "Environmental performance assessment of retrofitting existing coal fired power plants to co-firing with biomass: carbon footprint and energy approach", *J. Clean. Prod.*, **103**, 13-27. https://doi.org/10.1016/j.jclepro.2014.08.019.

AS (Australian Standard) (1999), Methods of Testing Concrete-Method 9: Determination of the Compressive Strength of Concrete Specimens.

Barbosa, V.F., MacKenzie, K.J. and Thaumaturgo, C. (1999), "Synthesis and characterisation of sodium polysialate inorganic polymer based on alumina and silica", *Proc. 99 Int. Geopolymer Conf. Saint-Quentin*, France.

Barbosa, V.F., MacKenzie, K.J. and Thaumaturgo, C. (2000), "Synthesis and characterisation of materials based on inorganic polymers of alumina and silica: Sodium polysialate polymers", *Int. J. Inorganic Mater.*, **2**(4), 309-317. https://doi.org/10.1016/S1466-6049(00)00041-6.

Behnood, A., Olek, J. and Glinicki, M. (2015), "Predicting modulus elasticity of recycled aggregate concrete using M5' model tree algorithm", *Constr. Build. Mater.*, **94**, 137-147. https://doi.org/10.1016/j.conbuildmat.2015.06.055.

Castelli, M., Trujillo, L., Gonçalves, I. and Popovič, A. (2017), "An evolutionary system for the prediction of high performance concrete strength based on semantic genetic programming", *Comput. Concrete*, **19**(6), 651-658.

https://doi.org/10.12989/cac.2017.19.6.657.

Chen, L. (2003), "Study of applying macroevolutionary genetic programming to concrete strength estimation", *J. Comput. Civil Eng.*, **17**(4), 290-294. https://doi.org/10.1061/(ASCE)0887-3801(2003)17:4(290).

Davidovits, J. (1999), "Chemistry of geopolymeric systems terminology", *Proc. 99 International Geopolymer Conf. Saint-Quentin*, France.

Deshpande, N., Londhe, S. and Kulkarni, S. (2014), "Modeling compressive strength of recycled aggregate concrete by artificial neural network, model tree and nonlinear regression", *Int. J. Sustain. Build. Environ.*, **3**, 187-198. https://doi.org/10.1016/j.ijsbe.2014.12.002.

Dreyfus, G. (2005), *Neural Networks Methodology and Applications*, Berlin Heidelberg, Springer-Verlag, Germany.

Duan, Z.H., Kou, S.C. and Poon, C.S. (2013a), "Prediction of compressive strength of recycled aggregate concrete using artificial neural networks", *Constr. Build. Mater.*, **40**, 1200-1206. https://doi.org/10.1016/j.conbuildmat.2012.04.063.

Duan, Z.H., Kou, S.C. and Poon, C.S. (2013b), "Using artificial neural networks for predicting the elastic modulus of recycled aggregate concrete", *Constr. Build. Mater.*, **44**, 524-532. https://doi.org/10.1016/j.conbuildmat.2013.02.064.

Gandomi, A.H., Alavi, A.H. and Arjmandi, P. (2010), "Genetic programming and orthogonal leas t squares: a hybrid approach to modelling the compressive strength of CFRP-confined concrete cylinders", *J. Mech. Mater. Struct.*, **5**(5), 735-753. https://doi.org/10.2140/jomms.2010.5.735.

Gazder, U., Al-Amoudi, O., Saad Khan, S. and Maslehuddin, M. (2017), "Predicting compressive strength of blended cement concrete with ANNs", *Comput. Concrete*, **20**(6), 627-634. https://doi.org/10.12989/cac.2017.20.6.627.

Gonzalez-Taboada, I., Gonzalez-Fonteboa, B., Martinez-Abella, F. and PerezOrdonez, J. (2016), "Prediction of the mechanical properties of structural recycled concrete using multivariable regression and genetic programming", *Constr. Build. Mater.*, **106**, 480-499. https://doi.org/10.1016/j.conbuildmat.2015.12.136.

Hardjito, D. and Rangan, B.V. (2005), "Development and properties of low-calcium fly ash-based geopolymer concrete", Research Report GC1, Faculty of Engineering, Curtin University of Technology, Perth, Australia.

Heidrich, C. (2002), "Ash utilisation - An australian perspective", *Proc. Int. Conf. on Geopolymers*, Melbourne, Australia.

Kiani, B., Gandomi, A., Sajedi, S. and Liang R. (2016), "New formulation of compressive strength of preformed-foam cellular concrete: an evolutionary approach", *J. Mater. Civil Eng.*, **28**(10), 04016092. https://doi.org/10.1061/(ASCE)MT.1943-5533.0001602.

Lachtermacher, G. and Fuller, J. (1994), "Back-propagation in hydrological times series forcasting", *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, Eds. Hipel K.W., Panu U. S., Singh V.P., **229**, Kluer Academic Publisher Group, The Netherlands.

Master, T. (1993), *Practical Neural Network Recipes in C++*, Academic Press, San Diego, California.

Mousavi, S.M., Aminian, P. and Gandomi, A.H. (2012), "A new predictive model for compressive strength of HPC using gene expression programming", *Adv. Eng. Softw.*, **45**(1), 105-114. https://doi.org/10.1016/j.advengsoft.2011.09.014.

Mukherjee, A. and Deshpande, J. (1995), "Modelling initial design process using artificial neural networks", *J. Comput. Civil Eng.*, **9**(3), 194-200. https://doi.org/10.1061/(ASCE)0887-3801(1995)9:3(194).

Nazari, A. and Riahi, S. (2011), "Prediction split tensile strength and water permeability of high strength concrete containing TiO_2 nanoparticles by artificial neural network and genetic programming", *Compos. Part B: Eng.*, **42**(3), 473-488. https://doi.org/10.1016/j.compositesb.2010.12.004.

Ozbay, E., Gesoglu, M. and Guneyisi, E. (2008), "Empirical modelling of fresh and hardened properties of self-compacting concretes by genetic programming", *Constr. Build. Mater.*, **22**(8), 1831-1840. https://doi.org/10.1016/j.conbuildmat.2007.04.021.

Palomo, A., Grutzeck, M. and Blanco, M. (1999), "Alkali-activated fly ashes, a cement for the future", *Cement Concrete Res.*, **29**(8), 1323-1329. https://doi.org/10.1016/S0008-8846(98)00243-9.

Reitermanov´a, Z. (2010), "Data splitting", *WDS'10 Proceedings of Contributed Papers*, Part I, 31-36.

Rodgers, J.L. and Nicewander, W.A. (1988), "Thirteen ways to look at correlation coefficient", *Am. Statist.*, **42**(1), 59-66. https://doi.org/10.1080/00031305.1988.10475524.

Roy, D.M. (1999), "Alkali-activated cements, opportunities and challenges", *Cement Concrete Res.*, **29**(2), 249-254. https://doi.org/10.1016/S0008-8846(98)00093-3.

Rubenstein, M. (2012), "Policy shifts toward an energy system transition: The dynamics of advocacy coalitions and New York State's renewable portfolio standard", MS Thesis, New York.

Saridemir, M. (2016), "Empirical modeling of flexural and splitting tensile strengths of concrete containing fly ash by GEP", *Comput. Concrete*, **17**(4), 489-498. https://doi.org/10.12989/cac.2016.17.4.489.

Shahin, M., Maier, H. and Jaksa, M. (2004), "Data division for developing neural networks applied to geotechnical engineering", *J. Comput. Civil Eng.*, **18**(2), 105-114. https://doi.org/10.1061/(ASCE)0887-3801(2004)18:2(105).

Sonebi, M. and Cevik, A. (2009), "Genetic programming based formulation for fresh and hardened properties of self-compacting concrete containing pulverised fuel ash", *Constr. Build. Mater.*, **23**(7), 2614-2622. https://doi.org/10.1016/j.conbuildmat.2009.02.012

Standards Australia (2000), Methods of Testing Concrete. Method 10 Determination of Indirect Tensile Strength of Concrete Cylinders ('Brazil' or splitting test): 8.

Standards Australia (2014), Methods of Testing Concrete-Compressive Strength Tests-Concrete, Mortar and Grout Specimens (AS 1012.9-2014).

Swanepoel, J.C. and Strydom, C.A. (2002), "Utilisation of fly ash in a geopolymeric material", *Appl. Geochem.*, **17**(8), 1143-1148. https://doi.org/10.1016/S0883-2927(02)00005-7.

van Jaarsveld, J.G., van Deventer, J.S. and Lukey, G.C. (2002), "The effect of composition and temperature on the properties of fly ash and Kaolinitebased geopolymers", *Chem. Eng. J.*, **89**(1-3), 63-73. https://doi.org/10.1016/S1385-8947(02)00025-6.

van Jaarsveld, J.G., van Deventer, J.S. and Lukey, G.C. (2003), "The characterisation of source naterials in fly ash-based geopolymers", *Mater. Lett.*, **57**(7), 1272-1280. https://doi.org/10.1016/S0167-577X(02)00971-0.

Xu, H. and van Deventer, J.S. (2000), "The geopolymerisation of Alumino-Silicate ninerals", *Int. J. Min. Pr.*, **59**(3), 247-266. https://doi.org/10.1016/S0301-7516(99)00074-5.

Xu, H. and van Deventer, J.S. (2002), "Geopolymerisation of multiple minerals", *Min. Eng.*, **15**(12), 1131-1139. https://doi.org/10.1016/S0892-6875(02)00255-8.

Yeh, C. (2006), "Exploring concrete slump model using artificial neural networks", *J. Comput. Civil Eng.*, **20**(3), 217-221. https://doi.org/10.1061/(ASCE)0887-3801(2006)20:3(217).

Younis, K.H. and Pilakoutas, K. (2013), "Strength prediction model and methods for improving recycled aggregate concrete", *Constr. Build. Mater.*, **49**(2013), 688- 701. https://doi.org/10.1016/j.conbuildmat.2013.09.003.

CC