Springer

KSME

Journal of Mechanical Science and Technology 24 (6) (2010) 1273~1278
www.springerlink.com/content/1738-494x
DOI 10.1007/s12206-010-0330-7

Formulation of elastic modulus of concrete using linear genetic programming [†]

Amir Hossein Gandomi[1,*], Amir Hossein Alavi[2], Mohammad Ghasem Sahab[1] and Parvin Arjmandi[1]

[1]*Structural Health Monitoring Research Group, College of Civil Engineering, Tafresh University, Tafresh, Iran*
[2]*College of Civil Engineering, Iran University of Science & Technology, Narmak 16844, Tehran, Iran*

(Manuscript Received September 28, 2009; Revised March 3, 2010; Accepted March 3, 2010)

Abstract

This paper proposes a novel approach for the formulation of elastic modulus of both normal-strength concrete (NSC) and high-strength concrete (HSC) using a variant of genetic programming (GP), namely linear genetic programming (LGP). LGP-based models relate the modulus of elasticity of NSC and HSC to the compressive strength, as similarly presented in several codes of practice. The models are developed based on experimental results collected from the literature. A subsequent parametric analysis is further carried out to evaluate the sensitivity of the elastic modulus to the compressive strength variations. The results demonstrate that the proposed formulas can predict the elastic modulus with an acceptable degree of accuracy. The LGP results are found to be more accurate than those obtained using the buildings codes and various solutions reported in the literature. The LGP-based formulas are quite simple and straightforward and can be used reliably for routine design practice.

Keywords: Tangent elastic modulus; Linear genetic programming; Compressive strength; Normal and high strength concrete; Formulation

1. Introduction

The elastic modulus of concrete is a key factor in structural and material engineering. Designers need the elastic modulus for estimating immediate and time-dependant deformation, determining modular ratio, and evaluating the stiffness of buildings and members. The modulus of elasticity is also important in reinforced and pre-stressed concrete for creep and shrinkage evaluation, as well as in crack control, especially at an early age [1, 2]. The modulus of elasticity is defined in the region in which Hooke's law is obeyed for the material as the ratio of stress over strain [3]. In mechanics, Hooke's law of elasticity is an estimation that states that the amount of strain is linearly related to the stress [4]. This can be determined from the slope of a stress-strain curve created during tensile tests on a sample of concrete. As shown in Fig. 1, in a typical stress-strain diagram of concrete, the first part of the curve is nearly a straight line with some curvature at σ, which is equal to half of the maximum value, σ_u. The initial slope of the stress-strain curve defines the initial or tangent modulus used with the parabolic stress method. The slope of the chord connecting the origin of the coordinate system to $0.5\sigma_u$ determines the secant modulus of elasticity, which is generally used in straight-line stress calculation [5].

Despite its importance, tensile strength (and elastic modulus) is not usually measured in the site for compliance purposes. It is often estimated from the measured compressive strength based on the empirical relationships proposed by various codes of practice. This is mainly to avoid performing laborious and time-consuming direct measurements from load-deformation curve [6].

By extending developments in computational software and hardware, several computer-aided data-mining approaches have been developed. The idea is that a pattern recognition system learns adaptively from experience and extracts various discriminators, each appropriate for its purpose. Artificial neural networks (ANNs) and fuzzy logic (FL) are the well-known pattern recognition methods. ANNs and FL have been utilized for the prediction of tangent elastic modulus of normal-strength concrete (NSC) and high-strength concrete (HSC) [7, 8]. Although ANNs and FL are successful in prediction, their inability to explicitly produce prediction equations limits them to be used by researchers. A new alternative approach, which is based on the data alone to determine the structure and parameters of the model, is known as genetic programming (GP) [9, 10]. GP is a variation of genetic algorithms in which computer programs are evolved to find solutions to problems [10]. The main advantage of GP-based approaches is their ability to generate prediction equations, which can be easily manipulated in practical circumstances. GP and its variants have successfully been applied to civil engineering problems [11]. In recent years, a particular subset

[†] This paper was recommended for publication in revised form by Associate Editor Chang-Wan Kim

[*]Corresponding author. Tel.: +98 912 479 4415, Fax: +98 21 4442 2790
E-mail address: a.h.gandomi@gmail.com

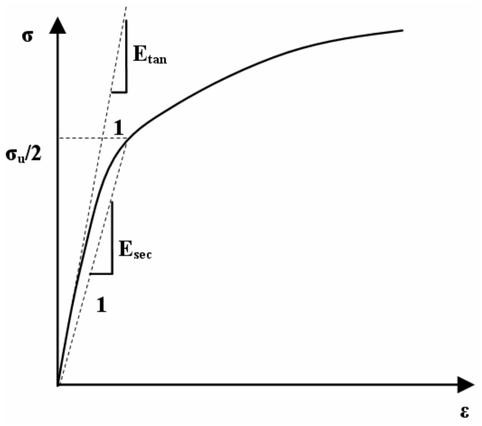


Fig. 1. Typical stress-strain diagram of concrete.

of GP with a linear structure similar to the DNA molecule in biological genomes, called linear genetic programming (LGP) [12], has emerged. LGP is a machine-learning approach that evolves the programs of an imperative language or machine language instead of the traditional GP [9] expressions of a functional programming language. Based on the numerical experiments, the LGP approach can significantly outperform other similar techniques and can be utilized as an efficient alternative to the traditional tree-based GP (TGP) [12, 13]. There have been little scientific efforts directed at applying LGP to civil engineering tasks [14, 15].

In this paper, the LGP technique is utilized to find mathematical relationships between the elastic modulus and the compressive strength of NSC and HSC. A comparative study between the results obtained by LGP and those obtained from the buildings codes [16–21], compatibility-aided models [22, 23], FL model [7], and ANN model [8] was conducted. A reliable database of previously published test results [22, 24–25] was utilized to develop the models.

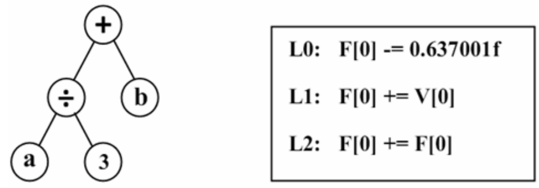
2. Genetic programming

GP is a symbolic optimization technique that creates computer programs for solving problems using the principle of Darwinian natural selection [9]. It was introduced by Koza [9] as an extension of the genetic algorithms. In GP, a random population of individuals (trees) is created to achieve high diversity. The symbolic optimization algorithms present the potential solutions by structural ordering of several symbols.

2.1 Linear genetic programming

LGP is a subset of GP that has recently emerged. Comparing LGP to the traditional TGP, several main differences are observed. Linear genetic programs (LGPs) have graph-based functional structures and evolve in an imperative programming language C/C ++ and machine code, rather than in expressions of a functional programming language like LISP (see Fig. 2). Unlike TGP, structurally non-effective codes coexist with effective codes in LGPs.

Non-effective codes in genetic programs, referred to as introns, represent instructions without any influence on the program behavior. Structural introns act as a protection that reduces the effect of variation on the effective code. The introns allow variations to remain neutral in terms of fitness change [12]. Because of the imperative program structure in LGP, these non-effective instructions can be identified efficiently. This allows the corresponding effective instructions to be extracted from a program during runtime. Since only effective programs are executed, evaluation can be accelerated significantly (see Fig. 3).



(a) (b)

Fig. 2. Comparison of GP program structures: (a) TGP; (b) LGP ([28]).

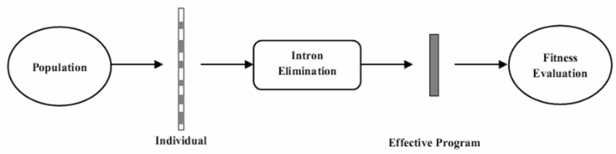


Fig. 3. Elimination of non-effective code in LGP (Only effective programs are executed [12]).

The instructions from imperative languages are restricted to operations that accept a minimum number of constants or memory variables, called registers (r), and assign the result to a destination register, e.g., $r_0 := r_1 + 1$. A part of a linear genetic program in C code is represented as follows [12]:

```
void LGP (double r [5])
{ ...
r [0] = r [5] + 70;
r [5] = r [0] - 50;
if (r [1] > 0)
if (r [5] > 2)
r [4] = r [2] * r [1];
r [2] = r [5] + r [4];
r [0] = sin (r [2]);
}
```

where register r [0] holds the final program output. LGPs can be converted into a functional representation by successive replacements of variables, starting with the last effective instruction [13]. Automatic Induction of Machine Code by Genetic Programming (AIMGP) is a particular form of LGP. In AIMGP, evolved programs are stored as linear strings of native binary machine code, which are directly executed by the processor. The absence of an interpreter and complex memory handling results in a significant speedup in the AIMGP execution compared to TGP. This machine-code-based LGP approach searches for the computer program and the constants simultaneously.

Table 1. Parameter settings for the LGP algorithm.

Parameter	Settings
Population size	100–1,000
Maximum program size	64
Initial program size	40
Crossover rate	0.5, 0.95
Homologous crossover	0.95
Mutation rate	0.9
Block mutation rate	0.3
Instruction mutation rate	0.4
Data mutation rate	0.4
Function set	+, -, ×, /, √
Number of demes	20

The machine-code-based LGP uses the following steps to evolve a computer program that predicts the target output from a data file of inputs and outputs [12]:

I. Initializing a population of randomly generated programs.
II. Running a tournament. In this step, four programs are randomly selected from the population. They are compared based on fitness; two programs are picked as the winners and two as the losers.
III. Transforming the winner programs. The two winner programs are copied and transformed probabilistically as follows:
   - Parts of the winner programs are exchanged with each other to create two new programs (crossover); and/or
   - Each of the tournament winners are changed at random to create two new programs (mutation).
IV. Replacing the loser programs in the tournament with the transformed winner programs. The winners of the tournament remain unchanged.
V. Repeating steps two through four until convergence. A program defines the output of the algorithm that simulates the behavior of the problem to an arbitrary degree of accuracy.

A comprehensive description of the basic parameters used to direct the search for a linear genetic program can be found in Brameier and Banzhaf [12].

3. Development of the LGP models

The main goal of this paper is to obtain explicit formulation of the tangent elastic modulus (E_c) of NSC and HSC in terms of compressive strength (f_c) as follows:

$$E_c = f(f_c) \tag{1}$$

Hence, one parameter was used for the LGP models as the input variable. Using NSC-related and HSC-related databases, two different LGP-based formulas for the elastic modulus of NSC and HSC were obtained. In order to propose a generic model for both NSC and HSC, another LGP model was developed based on all available test results. The various parameters involved in the LGP predictive algorithm are shown in Table 1. The parameter selection affects the model generalization capability of LGP. They were selected based on some previously suggested values [14, 15] and after some trial runs. The parameter settings are shown in Table 1.

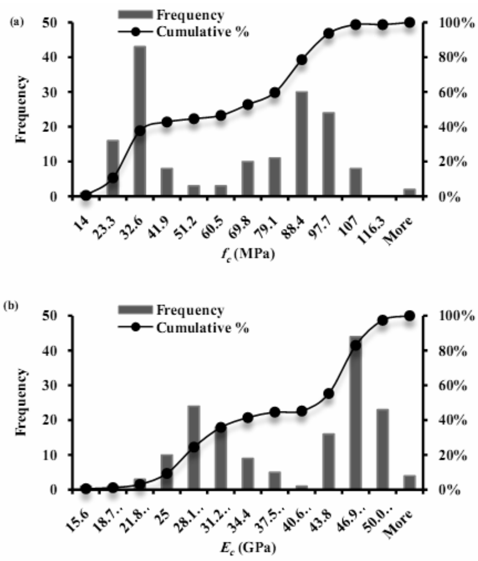


Fig. 4. The histograms of: (a) compressive strength and (b) elastic modulus for all data.

For the LGP-based modeling, a computer software called Discipulus [29] was used, which works based on the AIMGP platform. The best LGP models were chosen based on the following multi-objective strategy:

i. Providing the best fitness value on the training set of data.
ii. Providing the best fitness value on a test set of unsee data.

For this purpose, the following objective function (OBJ) was constructed to judge how well the model-predicted output agrees with the experimentally measured output. The selections of the best LGP models were deduced by the minimization of the following objective function:

$$OBJ = \left(\frac{No._{Train} - No._{Test}}{No._{All}}\right)\frac{MAE_{Train}}{R^2_{Train}} + \frac{2No._{Test}}{No._{All}}\frac{MAE_{Test}}{R^2_{Test}} \tag{2}$$

where $No._{Train}$, $No._{Test}$, and $No._{All}$ are the number of training, testing, and whole of data, respectively; R and MAE are the correlation coefficient and mean absolute error, respectively.

It is well known that the R value alone is not a good prediction accuracy indicator of a model. This is because by shifting the output values of a model equally, the R value will not change. The constructed objective function takes into account the changes of R and MAE together. Higher R values and lower MAE values result in lower OBJ, and consequently, indicate a more precise model. In addition, the above function considers the effects of different data divisions for training and testing data.

An experimental database, including previously published test results [22, 24-27], was utilized to develop the LGP-based

Table 2. The variables used in the model development.

	Range	Mean	S.D.[b]
HSC			
f_c (MPa)	46.4 - 125.6	84.75	13.60
E_c (GPa)	35.2 - 53.2	45.60	2.72
NSC			
f_c (MPa)	14 - 47.7	27.17	6.23
E_c (GPa)	15.6 - 36.8	27.77	4.13

[a]S.D.: Standard deviation

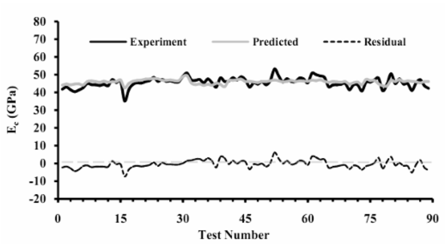


Fig. 5. Predicted versus experimental E_c of HSC using the LGP model.

models. The database contains test results of 89 elastic modulus of HSC and 70 elastic modulus of NSC. To visualize the samples distribution, all of the data are presented by frequency histograms (Fig. 4(a) and (b)). The ranges of samples used in this study are given in Table 2.

For the analysis, the data sets were divided into training and testing subsets. Of the 89 data sets for HSC, 69 values were taken for the training of the LGP algorithm and the remaining 20 values were used for the testing of the generalization capability of the models. For NSC, 57 values were taken for training and the remaining 13 values were used for the testing of the models. Of the total 159 data sets for HSC and NSC, 126 values were taken for the training of the LGP algorithm and the remaining 33 values were used for the testing of the proposed generic model of HSC and NSC.

3.1 Explicit formula for the elastic modulus of HSC

The formulation of the E_c of HSC in terms of f_c, for the best results by the LGP algorithm, is given below:

$$E_{c,LGP} = 2\sqrt{2f_c + 360} \tag{3}$$

In the above equation, E_c and f_c are within GPa and MPa, respectively. A comparison between the LGP-predicted and the experimental E_c of HSC is shown in Fig. 5. The proposed model for the E_c of HSC yields a low OBJ value that is equal to 8.6712.

3.2 Explicit formula for the elastic modulus of NSC

The formulation of the E_c of NSC in terms of f_c, for the best results by the LGP algorithm, is given below:

$$E_{c,LGP} = 2\sqrt{2f_c\sqrt{8\sqrt{f_c} - f_c}} \tag{4}$$

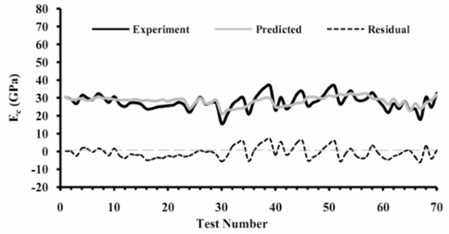


Fig. 6. Predicted versus experimental E_c of NSC using the LGP model.

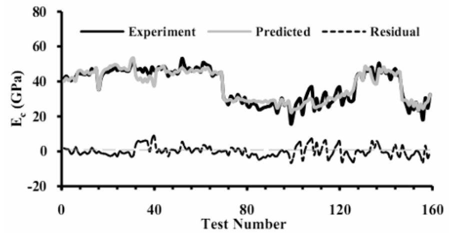


Fig. 7. Predicted versus experimental E_c of HSC and NSC using the LGP model.

In the above equation, E_c and f_c are within GPa and MPa, respectively. A comparison between the LGP-predicted and the experimental E_c of HSC and NSC is shown in Fig. 6. The proposed model for the E_c of NSC yields an OBJ value of 7.4887.

3.3 Explicit formula for the elastic modulus of HSC and NSC

The formulation of the E_c of HSC and NSC in terms of f_c, for the best results by the LGP algorithm, is given below:

$$E_{c,LGP} = 4\sqrt{f_c - 3} + 9 \tag{5}$$

In the above equation, E_c and f_c are within GPa and MPa, respectively. A comparison between the LGP-predicted and the experimental E_c of HSC and NSC is shown in Fig. 7. The proposed model for the E_c of HSC and NSC yields a low OBJ value that is equal to 2.5265.

4. Discussion

As described above, three formulas for the prediction of the E_c of HSC and NSC were obtained through LGP. Performance comparisons of the LGP models – Iranian (NBS 2006) [16], American (ACI 318-95) [17], British (BS 8110) [18], Canadian (CSA A23.3) [19], Norwegian (NS 3473) [20], and Turkish (TS 500) [21] codes – two compatibility-aided models [22, 23], FL model [7], and ANN model [8] for HSC and NSC are presented in Tables 3 and 4. Considering the result of the prediction of the E_c of HSC, it can be seen from Table 3 that the best performance is obtained by the ANN model. With the exception of the ANN and FL models, the proposed LGP model has better performance than the other models in the prediction of the E_c of HSC. In spite of the better performance of the ANN and FL-based models, they still have some fun-

Table 3. Performance statistics of the different models for the E_c of HCS.

Model	OBJ
FL [7]	5.4827
ANN [8]	4.8252
ACI [17]	26.5371
NS [20]	33.2777
CSA [19]	13.9365
Wee et al. [22]	7.5367
Gardner & Zhao [23]	19.6698
LGP (Present Study)	6.2448

Table 4. Performance statistics of the different models for the E_c of NCS.

Model	OBJ
FL [7]	9.0447
ANN [8]	10.4517
ACI [17]	10.6191
NBS [16]	8.6712
BS [18]	7.8557
TS [21]	11.8250
Gardner & Zhao [23]	8.1508
LGP (Present Study)	7.4887

damental disadvantages. Contrary to the LGP models, the ANN and FL models do not give a definite function of calculating the outcome using the input values. Hence, they do not provide a better understanding of the nature of the derived relationship between the different interrelated input and output data. ANN has only final synaptic weights to obtain outcome in a parallel manner. The determination of the fuzzy rules in FL is also a difficult task. The ANN and FL approaches are appropriate to be used as parts of a computer program and are not suitable for practical calculations. It can be seen from Table 4 that the LGP model provides superior performance compared with the other models in the prediction of the E_c of NSC.

In addition, the proposed LGP model for both NSC and HSC (Eq. (5)) yields the best (lowest) OBJ value. The superior performance of the generic model implies the rationale of developing comprehensive models for the E_c of both NSC and HSC, rather than separate models for each of them.

5. Parametric analysis

For further verification of the LGP models, a parametric analysis was performed in this study. The tendency of the elastic modulus predictions to compressive strength (f_c) variations can be determined according to Fig. 10(a)-(c). The results indicate that the elastic modulus of HSC and NSC continuously increases due to increasing f_c. The results of the parametric analysis are the expected cases from a structural engineering viewpoint.

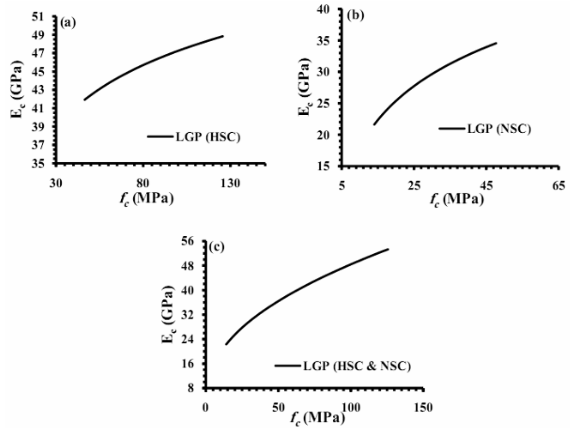


Fig. 8. Parametric analysis of the elastic modulus of HSC and NSC in the LGP models: (a) HSC; (b) NSC; (c) HSC and NSC.

6. Conclusion

In this paper, a linear variant of GP, namely LGP, was utilized to formulate the tangent elastic modulus of HSC and NSC. Three formulas for the elastic modulus were obtained via LGP. A reliable database of previously published elastic modulus test results was used for the training and testing of the prediction models. The LGP-based formulation results were compared with the experimental results, buildings codes, and existing models found in the literature. A subsequent parametric study was also conducted. The major findings obtained in this research are as follows:

(1) The LGP models give reliable estimates of the elastic modulus of HSC and NSC. The proposed formulas outperform the other existing models in nearly all of the cases.

(2) The generic LGP model provides significantly accurate determinations of the elastic modulus of both NSC and HSC.

(3) In addition to the acceptable accuracy, the LGP-based prediction equations are short and simple and can be used for practical engineering purposes.

(4) The robustness of the proposed LGP models was confirmed with the results of the parametric study. As other researchers have mentioned, the results show that the elastic modulus of HSC and NSC continuously increases due to increasing compressive strength.

(5) The LGP models can learn and improve as more data become available, without repeating the development procedures from the beginning.

(6) A major advantage of LGP for determining the elastic modulus lies in its powerful ability to model mechanical behavior without any prior assumptions.

(7) Another distinction of the LGP approach is that the elastic modulus can accurately be estimated without carrying out destructive, sophisticated, and time-consuming laboratory tests.

References

[1] A. Khan et al., Early age compressive stress– strain properties of low, medium and high strength concretes, *ACI Mater. J.,* 92 (6) (1995) 617-24.

[2] H. Mesbah et al., Determination of elastic properties of high-performance concrete at early age, *ACI Mater. J.,* 99 (1) (2002) 37-41.

[3] A. D. McNaught and A. Wilkinson, *Compendium of Chemical Terminology*, (2nd Ed.), IUPAC, Blackwell Scientific Publications, Oxford, (1997).

[4] A. C. Ugural and S. K. Fenster, *Advanced Strength and Applied Elasticity*, (5th Ed.), John Wiley & Sons, (1997).

[5] P. M. Ferguson et al., *Reinforced Concrete Fundamentals*, (5th Ed.), John Wiley & Sons, (1988).

[6] ASTM C 469. *Standard Test Method for Static Modulus of Elasticity and Poisson's Ratio of Concrete in Compression*, Annual Book of ASTM standards, (1994).

[7] F. Demir, A new way of prediction elastic modulus of normal and high strength concrete–fuzzy logic, *Cem. Conc. Res.* 35 (8) (2005) 1531-8.

[8] F. Demir, Prediction of elastic modulus of normal and high strength concrete by artificial neural networks, *Constr. Build. Mater.,* 22 (7) (2008) 1428-1435.

[9] J. R. Koza, *Genetic Programming: On the Programming of Computers by means of Natural Selection*, MIT Press, Cambridge (MA), (1992).

[10] W. Banzhaf et al., *Genetic Programming – An Introduction. On the Automatic Evolution of Computer Programs and its Application*, dpunkt/Morgan Kaufmann, Heidelberg/San Francisco, (1998).

[11] A. H. Alavi et al., Multi Expression Programming: A New Approach to Formulation of Soil Classification, *Eng. Comput.,* 26 (2010) 111-118.

[12] M. Brameier and W. Banzhaf, *Linear Genetic Programming*, Springer Science+Business Media, New York, (2007).

[13] M. Oltean and C. Grosan, A comparison of several linear genetic programming techniques, *Adv. Complex Syst.,* 14 (4) (2003) 1-29.

[14] A. H. Gandomi et al., New formulation for compressive strength of CFRP confined concrete cylinders using linear genetic programming, *Mater. Struct.* (2010) in press.

[15] A. H. Alavi and A. H. Gandomi, Energy-based numerical correlations for soil liquefaction assessment, *Comput. Geotech.* (2010) in press.

[16] NBS, *Analysis and Design of Reinforced Concrete Buildings*, National Building Standard, Part 9, Iran, (2006).

[17] ACI 318-95, *Building Code Requirements for Structural Concrete*, ACI Manual of Concrete Practice Part 3: Use of concrete in Buildings –Design, Specifications, and Related Topics. Detroit, Michigan, (1996).

[18] L. J. Parrott, *A Literature Review of High Strength Concrete Properties*, British Cement Association, ISBN 0 7210 13724, (1988).

[19] CSA A23.3-94, *Design of Concrete Structures*, Canadian Standard Association, Rexdale, Ontario, Canada, (1995).

[20] NS 3473, *Norwegian Council for Building Standardization, Concrete Structures Design Rules*, Stockholm, (1992).

[21] TS 500, *Requirements for Design and Construction of Reinforced Concrete Structures*, Turkish Standardization Institute, Ankara, (2000).

[22] T. H. Wee et al., Stress–strain relationship of high strength concrete in compression, *J. Mater. Civ. Eng.,* 8 (2) (1994) 70-76.

[23] D. Mostofinejad and M. Nozhati, Prediction of the modulus of elasticity of high strength concrete, *Iranian J. Sci. Tech., Trans. B Eng.,* 2005 29 (B3) 85-99.

[24] T. Ozturan, *An Investigation of Concrete Abrasion as Two Phase Material*, Thesis (PhD), Istanbul, Faculty of Civil Engineering, Istanbul Technical University (in Turkish), (1984).

[25] M. Gesoglu et al., Effects of end conditions on compressive strength and static elastic modulus of very high strength concrete, *Cem. Conc. Res.,* 32 (10) (2002) 1545-1550.

[26] M. J. Shannag, High strength concrete containing natural pozzolan and silica fume, *Cem. Conc. Comp.,* 22 (8) (2000) 399-406.

[27] M. Turan, Iren M. Strain stress relationship of concrete, *J. Eng. Arch.,* 12 (1) (1997) 76-81.

[28] A. H. Gandomi et al., Discussion on Genetic programming for retrieving missing information in wave records along the west coast of India, *Appl. Oce. Res.,* 30 (2008) 338-339.

[29] M. Conrads et al., *Discipulus—Fast Genetic Programming Based on AIM Learning Technology*, Register Machine Learning Technologies Inc., Littleton, CO., (2004).

Amir Hossein Gandomi received his B.Sc. and M.Sc. degrees in Civil and Structural Engineering from Iran University of Science & Technology and Tafresh University, respectively. He is currently a lecturer at Tafresh University and serves as a researcher in National Elites Foundation in Iran. His research interests include Structural Health Monitoring, and Artificial Intelligence Application in Civil Engineering.