# Statistical analysis of the performance of the soft computing based prediction model for shrinkage of concrete including mineral admixtures

Kasım Mermerdaş
Harran University, Faculty of Engineering
Department of civil engineering
Şanlıurfa, Turkey
kasim.mermerdas@harran.edu.tr

Mohamed Moafak Arbili
Hasan Kalyoncu University, Faculty of Engineering
Department of civil engineering
Gaziantep, Turkey
mohamed.moafak@yahoo.com

Farman Khalil Ghaffoori
Hasan Kalyoncu University, Faculty of Engineering
Department of civil engineering
Gaziantep, Turkey
farman_gh@yahoo.com

***Abstract*— The study covers a numerical study to derive a soft computing based mathematical model for prediction of the shrinkage of concretes having different mix compositions. The model will be derived by means of gene expression programming (GEP). In order to generate the model, a comprehensive data set compiled from technical literature will be used. The prediction parameter will be assigned as water-to-binder ratio (w/b), silica fume content (SF) in kg/m$^3$, fly ash content (FA) in kg/m$^3$, cement (C) content in kg/m$^3$, aggregate/binder ratio, compressive strength (fc) in MPa, type of shrinkage for shrinkage (drying or autogenous shrinkage), and drying time in days. The GEP model obtained in this study will be compared to an available formula which has been presented by first two authors of this paper. The available model was generated by artificial neural networks (NN). The comparison of the models will be carried out through some statistical parameters such as correlation coefficient (R), mean absolute percent error (MAPE), means square error (MSE), and root mean square error (RMSE). Moreover, in order to assess performances of the proposed models an advanced statistical analysis method called "Wilcoxon rank sum test" was applied. The test indicates whether the obtained prediction and experimental datasets are statistically equivalent or not at a specified level of significance.**

***Keywords-* *Shrinkage of concrete, soft computing, genetic programming, artificial neural networks, statistical analysis***

## I. INTRODUCTION

Concrete is the most widely used construction material all over the world. However, shrinkage is of concern when it relates to durability of concrete structure. Excessive shrinkage may cause concrete cracking, even structural failure. Thus, cracking may lead to increased corrosion rate of steel reinforcement in concrete structure. In the view of global sustainable development [1]. Therefore, researchers start to make use of blending of two or three SCMs to optimize durability and cost for the benefit of engineers, owners, contractors and material suppliers. The industrial by-products used as SCMs, such as fly ash and silica fume, have become more efficient admixtures to diminish the shrinkage effects and increase the durability of concrete, and usage of SCMs could substantially reduce the final cost of concrete mixtures since these materials are quite cheaper in comparison to Portland cement [2].

The problems encountered in the field of engineering are generally unstructured and imprecise influenced by intuitions and past experiences of a designer [3]. Complexity to mathematically model real world problems has compelled the human civilization to search for nature inspired computing tools. The evolution of such computing tools revolves around the information processing characteristics of biological systems. In contrast to conventional computing, these tools are rather "soft" as they lack the exactness and therefore placed under the umbrella of a multidisciplinary field called soft computing. Soft Computing is an emerging collection of methodologies, which aim to exploit tolerance for imprecision, uncertainty and partial truth to achieve robustness, tractability and total low cost [4].

Soft computing techniques have a self-adapting characteristic paving a way for development of automated design systems. A synergistic partnership exploiting the strengths of these individual techniques can be harnessed for developing hybrid-computing tools [4].

For example, some applications of soft computing are invited in the following fields on

a) Structural Engineering: Vanluchene and Sun [5] presented an introduction to neural network by using back-propagation algorithm to solve three different structural

TABLE I THE EXPERIMENTAL DATA COLLECTED FROM THE TECHNICAL LITERATURE FOR MODELING STUDY

| **Data source (586)** | **Input** | | | | | | | | **Output** |
|---|---|---|---|---|---|---|---|---|---|
| | **X1** | **X2** | **X3** | **X4** | **X5** | **X6** | **X7** | **X8** | **Y** |
| | w/b | SF | FA | cement | Agg/b. | $f_c$ Mpc @28 days | Type Shrinkage | Dry Time | Shrinkage |
| **Zhang et al [10]** | 0.27-0.35 | 0-50 | 0 | 446-498 | 3.38-3.70 | 57.33-86.94 | 0-1 | 1-98 | 34-282 |
| **Wongkeo et al [11]** | 0.49 | 0-42 | 0-269 | 269-538 | 2.64-2.75 | 29.05-69.05 | 1 | 7-91 | 93-1100 |
| **Yoo et al [12]** | 0.30 | 0-88 | 0-175 | 408-583 | 2.68-2.57 | 54.8-69.8 | 0 | 1-49 | 39-400 |
| **Khatib et al [13]** | 0.36 | 0 | 0-400 | 100-500 | 3.25-3.5 | 11-72.58 | 1 | 2-56 | 5-432 |
| **Khatri and Sirivivatnanon [14]** | 0.34-0.36 | 0-46 | 0-100 | 282-425 | 4.15-4.30 | 65-94.99 | 1 | 7-400 | 267-895 |

engineering problems related to pattern recognition, decision making and problems that have numerically complex solutions.

b) Concrete Strength Modeling: in the study of Özcan et al [6], compressive strength prediction was done by using ANN and Fuzzy logic.

c) Geotechnical Engineering: Shahin et al. [7] used neural networks for predicting settlement of shallow foundations on cohesion less soils. The predictive ability of ANN is compared with three of the most commonly used traditional methods.

d) Earthquake Engineering: Lee and Han [8] developed efficient neural network models for generation of artificial earthquakes and response spectra.

The purpose of this study is to perform an analytical study to obtain soft computing based mathematical model for prediction of autogenous and drying shrinkage of concretes incorporating silica fume (SF) and fly ash (FA). The model was produced by means of genetic algorithm based soft computing method called gene expression programming (GEP). The GEP model was compared with the available formulation which has been suggested by Mermerdaş and Arbili [9]. The comparison was made on the statistical bases. The statistical parameters utilized were mean absolute percent error (MAPE), mean square error (MSE), root mean square error (RMSE), and the correlation coefficient (R). Moreover, another statistical method called "Wilcoxon Rank sum test" was also benefited to indicate whether the means of the data populations (the experimental and the predicted ones by GEP and NN) are same or not.

## II. DESCRIPTION OF THE DATABASE USED FOR MODELING

The proposed GEP formulation of shrinkage (S) was derived using a set of 586 experimental data available in the technical literature [10-14] for training and testing the proposed model.

Table 1 summarizes the selected experimental data. In detail, the generated models for shrinkage following input parameters: w/b (water/binder), SF (silica fume) content in kg/m$^3$, FA (fly ash) content in kg/m$^3$, C (cement) content in kg/m3, aggregate/binder ratio, fc (compressive strength) in MPa, type of shrinkage (for drying shrinkage 1, for autogenous shrinkage 0) and dry time in days.

All data samples were put in an order to establish a consistent sequence of the inputs to be used for derivation of the models. Thus, eight inputs parameters were utilized for development of prediction models. The data set was randomly divided into two parts to obtain training and testing databases. GeneXProtools5.0 software was used for derivation of the GEP based mathematical model.

## III. PROPOSED GEP MODEL

The prediction model derived from GEP is presented in Eq. 1. The GEP parameters used for derivation of the mathematical models are given in Table 2. In order to provide an accurate model, various mathematical operations were used.

$$S = \cos\left[e^{\sqrt[3]{d_0}}\right] - 4.88385 +$$

$$+\cos\left[\sin\left(\tan\left(\ln 1.733399 + 10^{1.733399} - d_6 \times d_5 + 1.733399\right)\right)\right] +$$

$$d_0\left[\frac{\left(\frac{\tan(d_5 \times d_4)}{d_5 + d_6}\right)}{\sqrt{d_0}}\right] + \sqrt{e^{d_5 - (-3.03775)} \times (d_6)} +$$

$$\ln d_7 + \left[\tan^{-1}\left(\sqrt[4]{(\cos d_3 \times d_2 + d_5)^3}\right)\right]^3 +$$

$$10^{\left[\sin\left(\tan^{-1}(d_7 - \ln d_7) + d_5\right)\right]} - d_5 +$$

$$\sqrt[32]{\tan^{-1} d_2 \times (-5.326263 - d_4)} \times d_1 + d_7$$

$$+\cot(-5.326263 + d_4) + \sqrt[3]{d_0} \tag{1}$$

Where d0 = w/b (water/binder); d1 = SF (silica fume); d2 = FA (fly ash); d3= C (cement); d4 = (aggregate/binder); d5= fc (compressive strength); d6 = (type of shrinkage); d7= (drying time)

TABLE II. GEP PARAMETERS USED FOR PROPOSED MODELS

.

| | Parameters | *S* for shrinkage |
|---|---|---|
| P1 | Function Set | +, -, *, /, √, ^, ln, exp, sin, tan, inverse, Pow |
| P2 | Number of generation | 99521 |
| P3 | Choromosomes | 30 |
| P4 | Head size | 10 |
| P5 | Linking function | Addition |
| P6 | Number of genes | 10 |
| P7 | Mutation rate | 0.044 |
| P8 | Inversion rate | 0.1 |
| P9 | One-point recombination rate | 0.3 |
| P10 | Two-point recombination rate | 0.3 |
| P11 | Gene recombination rate | 0.1 |
| P12 | Gene transposition rate | 0.1 |

The performance of the proposed GEP prediction model in Eq. 1 is graphically demonstrated in Fig. 1 for training and in Fig 2 for testing data sets. It seems that there is a far trend in the variation of the data between predicted and experimental data. Correlation coefficients equal to 0.863 and 0.789 were calculated for training and testing databases, respectively, thus indicating reasonably strong correlation between actual and predicted values. Moreover, close values of the correlation coefficients may be considered as an evidence for the consistency and good fitness of the proposed model.

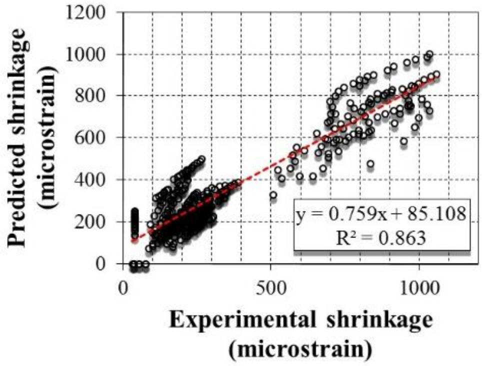


Figure 1 Predicted shrinkage values from GEP vs. experimental data for training

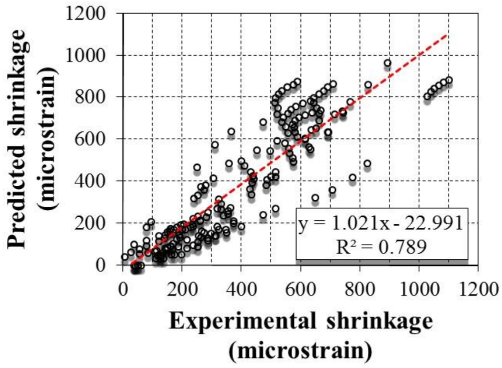


Figure 2 Predicted shrinkage values from GEP vs. experimental data for testing

In order to compare the prediction of the proposed models with experimental shrinkage, Figures 3-7 were plotted. Figure 3 includes the experimental and predicted autogenous shrinkage values, while the other figures contain drying shrinkage values.

Observing figure 3 it can be seen that prediction performance of GEP for autogenous shrinkage values between 0-100 microstrain is totally misleading. The GEP model yielded both invalid (0 microstrain) and extremely overestimated values. However, NN model performed well in this interval. Moreover, for the higher autogenous shrinkage values (> 100 microstrain), NN model demonstrated almost prefect estimation performance while GEP model mostly gave underestimated results.

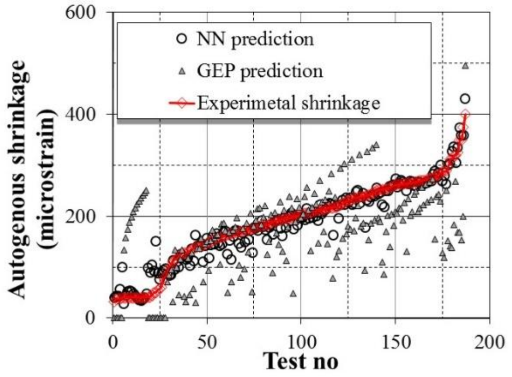


Figure 3 Comparison of experimental autogenous shrinkage values with those predicted by NN and GEP

For drying shrinkage values, GEP model had overestimated results between experimental values of 0-300 microstrain (Figure 4). However, as the experimental drying shrinkage values increased the tendency of GEP estimation decreased. Especially for the drying shrinkage values of 900-1200 microstrain all of the GEP values were below the experimental findings (Figure 7). On the other hand, NN model achieved more precise and accurate prediction performance in all of the intervals.

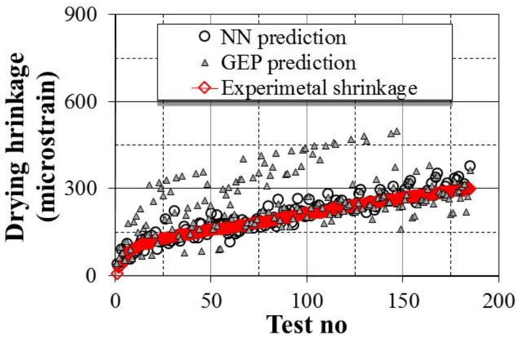


Figure 4 Comparison of experimental drying shrinkage values between 0-300 microstrain with those predicted by NN and GEP

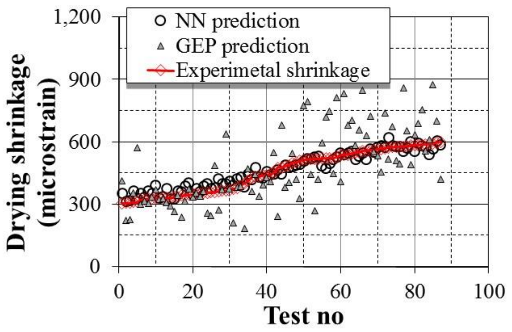


Figure 5 Comparison of experimental drying shrinkage values between 300-600 microstrain with those predicted by NN and GEP

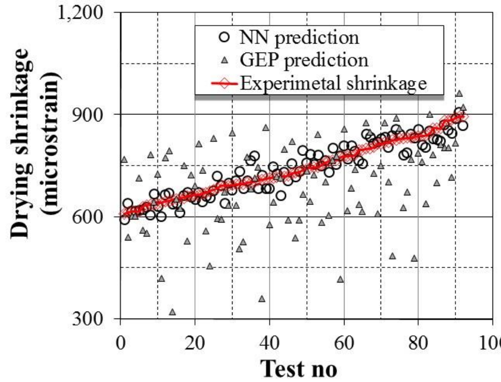


Figure 6 Comparison of experimental drying shrinkage values between 600-900 microstrain with those predicted by NN and GEP

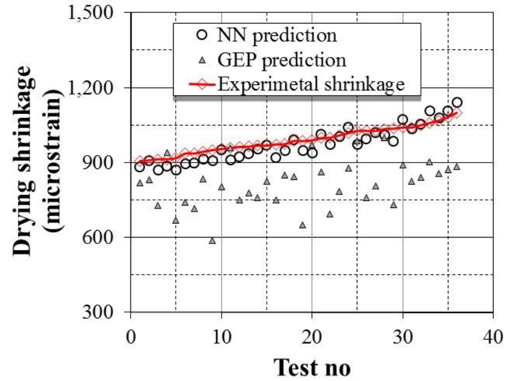


Figure 7 Comparison of experimental drying shrinkage values between 900-1200 microstrain with those predicted by NN and GEP

## IV. DATA ANALYSIS AND PERFORMANCE EVALUATION OF THE PREDICTION MODELS

In order to make a better comparison in the prediction performances of the suggested empirical relations developed in this study, the following statistical parameters were calculated and given in Table 3

The correlation coefficient, R, (Eq 2):

$$R = \frac{\sum (m_i - m')(p_i - p')}{\sqrt{(m_i - m')^2 (p_i - p')^2}} \tag{2}$$

Mean absolute percent error, MAPE, (Eq 2):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{m_i - p_i}{m_i}\right| \times 100 \tag{3}$$

Mean square error, MSE, (Eq 3):

$$MSE = \frac{\sum_{i=1}^{n}(m_i - p_i)^2}{n} \tag{4}$$

Root mean square error, RMSE, (Eq 4):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(m_i - p_i)^2}{n}} \tag{5}$$

where m’ and p’ are mean values of measured (mi) and predicted (pi) values, respectively

For further statistical assessment of the results, "Wilcoxon rank sum test" was applied and *P* value for each group determined to decide on the acceptance of the model. The Wilcoxon rank-sum test is a nonparametric alternative to the two sample *t*-test which is based solely on the order in which the observations from the two samples fall. Since the equality between means of target shrinkage ($S_1$) values and the shrinkage values proposed by the models ($S_2$) were investigated, two-tailed test with 5% level of significance ($\alpha$=0.05) was considered.

*P* values for each group are calculated. If *P* value is greater than $\alpha$ then there is no sufficient evidence to reject null hypothesis ($H_0$). That means the predicted model can be a useful tool for prediction of strength of concrete. The hypothesis;

$$\mathrm{H}_0 : S_1 = S_2$$
$$\mathrm{H}_1 : S_1 \neq S_2$$

If $P > \alpha$, then accept $H_0$ hypothesis and reject $H_1$ hypothesis (The shrinkage values of predicted and target values are the same. The estimation model is valid). If $P < \alpha$, then reject $H_0$ hypothesis and accept $H_1$ hypothesis (The shrinkage values of predicted and target values are not the same; The estimation model is invalid).

The test results are given in Table 4. As can be seen from Table 4, the *P* values determined are greater than the critical $\alpha$ value (0.05) for the model. According to the statistical analysis by Wilcoxon rank sum test GEP model and the NN model were proved to be applicable. It can also be observed from Table 3 that the proposed GEP formulation for shrinkage strain prediction is able to follow a very close trend to the experimental values.

TABLE III. STATISTICAL PARAMETERS FOR COMPARING THE PREDICTION MODELS

| | **The prediction model** | |
|---|---|---|
| **Statistical Parameter** | GEP | NN |
| MAPE | 40.1 | 9.6 |
| MSE | 13289.4 | 584.2 |
| RMSE | 67.76 | 11.42 |
| R | 0.903 | 0.996 |

Table 3 indicates that previously proposed NN model is more accurate than GEP model. However, when observing Table 4 results the mean shrinkage values of the predicted data population by GEP is equal to that of experimental values. Therefore, it can be concluded that the proposed GEP model can statistically be accepted as a prediction tool with 95% level of the confidence. The divergence of the numerical values of the GEP model presented in Table 3 is due to some extreme values obtained from the proposed model. These extreme values are the invalid "0" values given by the model This problem may be overcome by pre-processing the data by means of some other techniques like particle swarm optimization. However, the advantage of using GEP model is to obtain a quite simple mathematical expression when compared to NN model. Moreover, in GEP model the data is entered to the model as they are. Whereas, for NN model the data must be normalized before entering to the model.

Although the beneficiation of GEP model is more simple than NN model it still seems to be complex for hand calculation. Therefore, through using a suitable computational tool (like spreadsheet modeling) the proposed formulation can be benefited for researches on shrinkage of the concretes containing mineral admixtures.

TABLE IV. WILCOXON RANK SUM TEST RESULT

| | P-value | Acceptance of the model |
|---|---|---|
| GEP | 0.581>0.05* | YES |
| NN | 0.846>0.05* | YES |

*0.05 indicates 5% statistical significance or 95% level of confidence

## V CONCLUSIONS

The following conclusions can be drawn

- Numerical modeling of shrinkage of concrete containing mineral admixtures was conducted gene expression programming (GEP). To this aim, available experimental data presented in the existing literature were used to derive the model. In order to evaluate their efficiency and advantages, the performance of the proposed model was compared to that provided by the collected and an available model derived from neural networks.
- A comparison with the existing NN modeling for the collected data referred that the NN model provide better prediction results than the GEP model. The errors obtained from GEP model were higher than NN model's prediction results. However, GEP model uses plain data, while NN requires the input data to be processed before entering the prediction model.
- The model for shrinkage estimation of the concretes produced with mineral admixtures can efficiently be constructed using GEP. Although the constructed GEP model showed lower performance than the available NN model, the statistical analyses proved that the accuracy of the proposed models is good enough to be utilized for prediction purposes.

## REFERENCES

[1] M. Tia et al. "Modulus of elasticity, creep and shrinkage of concrete" Gainesville: Department of Civil & Coastal Engineering, College of Engineering, 2005

[2] E.H. Yang, Y. Yang, V.C. Li, "Use of high volumes of fly ash to improve ECC mechanical properties and material greenness." ACI Materials Journal, 2007, pp. 620-628.

[3] V. Chandwani, V. Agrawal, R.Nagar “Applications of Soft Computing in Civil Engineering: A Review” International Journal of Computer Applications, 2013, pp. 13 – 20

[4] D.K. Chaturvedi, “Soft Computing Techniques and its applications in Electrical Engineering” Studies in Computational Intelligence, 2008, 103, 1-10.

[5] R.D. Vanluchene, R. Sun, “Neural Networks in Structural Engineering”, Computer-Aided Civil and Infrastructure Engineering, 1990, pp. 207-215.

[6] F. Özcan, C. Atis, O. Karahan, E. Uncuoglu, H. Tanyildizi, “Comparison of artificial neural network and fuzzy logic models for prediction of long term compressive strength of silica fume concrete” Advances in Engineering Software, 2009, pp. 856–863

[7] M.A. Shahin, H.R. Maier, M.B. Jaksa, “Predicting settlement of Shallow Foundations using Neural Networks”, Journal of Geotechnical and Geoenvironmental Engineering, 2002, pp. 785-793.

[8] S.C. Lee, S.W. Han, “Neural-network-based models for generating artificial earthquakes and response spectra” Computers and Structures, 2002, pp. 1627–1638.

[9] K. Mermerdaş, M.M. Arbili “Explicit formulation of drying and autogenous shrinkage of concretes with binary and ternary blends of silica fume and fly ash” Construction and Building Materials, 2015, pp. 371-379

[10] M.H. Zhang, C.T. Tam, M.P. Leow, “Effect of Water-to-Cementitious Materials Ratio and Silica Fume on the Autogenous Shrinkage of Concrete” Cement and Concrete Research, 2003, pp. 1687–1694

[11] W. Wongkeo, P. Thongsanitgarn, A. Chaipanich, “Compressive Strength of Binary and Ternary Blended Cement Mortars Containing Fly Ash and Silica Fume Under Autoclaved Curing” TIChE International Conference, Hatyai, Songkhla THAILAND, 2012.

[12] S. Yoo, S. Kwon, S. Jung “Analysis technique for autogenous shrinkage in high performance concrete with mineral and chemical admixtures” Construction and Building Materials, 2012, pp. 1–10

[13] J. Khatib, “Performance of self-compacting concrete containing fly ash” Construction and Building Materials, 2008, pp. 1963–1971

[14] R. Khatri and V. Sirivivatnanon “Effect of different supplementary cementitious materials on mechanical properties of high performance concrete” Cement and Concrete Research, 1995, pp. 209-220.