





Contents lists available at SCCE

Journal of Soft Computing in Civil Engineering

Journal homepage: [www.jsoftcivil.com](http://www.jsoftcivil.com)



## Predictive Equations for Estimation of the Slump of Concrete Using GEP and MARS Methods

Ismail Husein<sup>1</sup>, Ramaswamy Sivaraman<sup>2</sup>, Sarwar Hasan Mohammad<sup>3</sup>, Forqan Ali Hussein Al-Khafaji<sup>4</sup> , Sokaina Issa Kadhim<sup>5</sup>, Yousof Rezakhani<sup>6\*</sup> 

1. Department of Mathematics, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
  2. Associate Professor, Department of Mathematics, Dwaraka Doss Goverdhan Doss Vaishnav College, Arumbakkam, University of Madras, Chennai, India
  3. Ph.D., Lecture, College of Engineering, Sulaimani Polytechnic University, Sulaimani, Iraq
  4. Department of Media, Al-Mustaqbal University College, 51001, Babylon, Hillah, Iraq
  5. Building and Construction Technical Engineering Department, College of Technical Engineering, The Islamic university, Najaf, Iraq
  6. Department of Civil Engineering, Pardis Branch, Islamic Azad University, Pardis, Iran
- Corresponding author: [yousofrezakhani@pardisiau.ac.ir](mailto:yousofrezakhani@pardisiau.ac.ir); [yousofrezakhani@gmail.com](mailto:yousofrezakhani@gmail.com)

 <https://doi.org/10.22115/SCCE.2023.389726.1619>

### ARTICLE INFO

Article history:  
Received: 13 March 2023  
Revised: 20 June 2023  
Accepted: 17 July 2023

Keywords:  
GEP;  
MARS;  
Slump concrete;  
Data-driven methods.

### ABSTRACT

This paper developed two robust data-driven models, namely gene expression programming (GEP) and multivariate adaptive regression splines (MARS), for the estimation of the slump of concrete (SL). The main feature of the proposed data-driven methods is to provide explicit mathematical equations for estimating SL. The experimental data set contains five input variables, including the water-cement ratio (W/C), water (W), cement (C), river sand (Sa), and Bida Natural Gravel (BNG) used for the estimation of SL. Three common statistical indices, such as the correlation coefficient (R), root mean square error (RMSE), and mean absolute error (MAE), were used to evaluate the accuracy of the derived equations. The statistical indices revealed that the GEP formula (R=0.976, RMSE=19.143, and MAE=15.113) was more accurate than the MARS equation (R=0.962, RMSE=23.748, and MAE=16.795). However, the application of MARS, due to its simple regression equation for estimating SL, is more convenient for practical purposes than the complex formulation of GEP.

How to cite this article: Husein I, Sivaraman R, Mohmmad SH, Al-Khafaji FAH, Kadhim SI, Rezakhani Y. Predictive equations for estimation of the slump of concrete using GEP and MARS methods. J Soft Comput Civ Eng 2024;8(2): 1–18. <https://doi.org/10.22115/scce.2023.389726.1619>

2588-2872/ © 2023 The Authors. Published by Pouyan Press.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



## 1. Introduction

Concrete is an essential construction material that plays a crucial role in the development of infrastructure, providing strength and durability to buildings, bridges, and other structures [1]. Therefore, many studies have been performed on the properties of concrete using different conventional statistical models and data-driven methods due to the material's importance. The ease with which concrete may be blended, poured, compacted, and finished is referred to as its workability. A concrete mixture that is hard to mix and compact will add to the cost of management and result in inadequate strength, durability, and attractiveness. For producing high-quality concrete, the workability of concrete is a critical component that must be studied [2]. The slump test is frequently employed to assess the concrete's mechanical properties. Slump is an essential metric for gauging the consistency of concrete quality, significantly impacting the quality of civil engineering projects [3].

Data-driven methods are a particularly successful and reliable replacement compared to traditional regression analysis for complex systems whose objective is determining relationships between input and output variables [4–9]. Data-driven models are appropriate alternatives to conventional models and are widely used to model concrete properties [10–15]. Many scholars worldwide are interested in using data-driven techniques to evaluate concrete characteristics. The main shortcoming of traditional regression techniques is their inability to provide appropriate estimation results for complex problems [16,17]. The soft computing method exceeds the difficulties and drawbacks of regression analysis and provides astounding and precise findings.

Data-driven models have become increasingly popular in the concrete industry, and many studies have utilized these models to improve various aspects of concrete production. For instance,

Cao et al. [18] employed machine-learning techniques to estimate the porosity of high-performance concrete. Their research revealed that gradient-boosting trees outperformed random forests regarding prediction accuracy. In a study conducted by Golafshani et al. [19], a combination of Particle Swarm Optimization (PSO) and a fuzzy inference system was utilized to model the compressive strength (CS) of eco-friendly concrete. The work of Golafshani et al. [20] introduced robust modeling for approximating the CS of concrete employing an artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS) techniques. The integration of Grey Wolf Optimizer (GWO) enhanced the performance of these models. Badawi et al. [21] proposed the application of ANN to predict the CS and slump of concrete by considering input parameters related to the concrete mix design. The developed ANN model, implemented using the MATLAB neural network toolbox, exhibited a strong correlation with experimental data. Soft computing approaches for estimating the CS and slump of concrete were discussed by Timur Cihan [22], who aimed to identify optimal techniques for normalization, regression, and feature selection to achieve accurate predictions.

Tang et al. [23] introduced a hybrid machine learning model that combined Support Vector Regression (SVR) with Grid Search (GS) to accurately predict the CS of fly ash concrete. Through experimentation with 98 datasets, this hybrid model demonstrated its potential as an effective method for CS prediction, outperforming the stand-alone SVR model. Behnood and

Golafshani [24] employed the M5P algorithm to model the mechanical properties of concrete incorporating waste foundry sand (WFS). Rajakarunakaran et al. [25] proposed the use of machine learning-based regression approaches to estimate the CS of self-compacting concrete (SCC).

Estimating the slump of concrete is a complex task that requires considering the nonlinear relationship between concrete constituents, such as water, cement, and aggregate. These materials have components that interact nonlinearly with the slump of concrete, making estimating the slump a challenging task. Moreover, the slump of concrete is an essential property that influences the workability and performance of concrete, making it a critical factor in construction. To address this challenge, some investigations have used data-driven models for the prediction of the slump of the concrete. For instance, Chine et al. [26] utilized multiple linear regression (MLR) and ANN models to predict the slump of concrete. Their findings suggested that ANN was more accurate than MLR in approximating slump for different grades of concrete.

Similarly, Agrawal and Sharma 2010 [27] combined ANN and genetic algorithms (GA) to estimate the slump of concrete. They showed that their combined model improved the prediction accuracy of the slump compared to the stand-alone ANN model. Rezaie and Sadighi [28] generated a linear regression analysis and an adaptive neuro-fuzzy inference system (ANFIS) to predict the slump of lightweight aggregate concrete. Their results indicated that ANFIS was more accurate than linear regression in predicting the slump of concrete. Islam et al. [29] developed statistical analysis and regression models to estimate the slump of concrete incorporating rice husk ash (RHA) and laboratory results. Furthermore, Onikeku et al. [30] employed both ANN and MLR approaches to predict the slump of concrete containing two blended agro-waste materials and achieved acceptable results. Öztaş et al. [31] demonstrated the effectiveness of ANN in predicting the slump values of high-strength concrete. Singh et al. [32] developed an ANN model for determining the slump of concrete using laboratory tests. Yeh [33] showed that ANN outperformed traditional regression methods in predicting the slump of concrete.

Recently, Yusuf et al. [34] developed various ANN models with different numbers of hidden nodes to predict the slump of concrete. They evaluated the performance of ANN and MLR using statistical measures and created several ANN models for the estimation of the slump. Their results indicated that the ANN model with twenty hidden nodes had the best performance in predicting the slump, and an MLR equation was also obtained to predict the slump.

Recent studies have demonstrated numerous applications of black-box methods, such as ANN, for predicting the characteristics of concrete, particularly in estimating concrete slumps. However, black box models suffer from the limitation of not providing explicit relationships among the variables involved in a complex problem [35–37]. A review of previous studies has revealed limited utilization of the GEP and MARS methods in concrete slump estimation, despite their capability of offering mathematical relationships. Applying GEP and MARS methods has been relatively restricted compared to black box methods. Nonetheless, these methods have showcased their ability to estimate complex parameters accurately. Moreover, the relationships these models provide can be easily applied by engineers in practical applications.

This study employed two powerful data-driven methods, gene expression programming (GEP) and multivariate adaptive regression splines (MARS), to predict the slump of concrete. GEP and MARS are highly effective techniques for simulating complex processes, as they can represent intricate input-output relationships without requiring prior knowledge of the phenomenon. The primary objective of this study is to develop explicit models for predicting  $SL$ . The proposed GEP and MARS models provide mathematical equations that can be used to predict the slump of concrete.

## 2. Material and methods

### 2.1. Data samples

The data samples used for developing data-driven models were obtained from a previous study by Yusuf et al. [34]. Yusuf et al. [34] conducted experimental works for the prediction of  $SL$ . They used Bida natural gravel (BNG) as coarse aggregate in concrete mixes and developed ANN models and MLR equations to predict  $SL$ . The main factors investigated by Yusuf et al. [34] that influenced concrete slump ( $SL$ ) were the water-cement ratio ( $W/C$ ), water ( $W$ ), cement ( $C$ ), sand ( $Sa$ ), and Bida natural gravel ( $BNG$ ). They examined 36 concrete mixes for the measurement of ( $SL$ ). Therefore, the functional relationship form below was considered for modeling  $SL$ .

$$SL = f(W/C, W, C, Sa, BNG) \quad (1)$$

Table 1 provides the main statistical parameters for the prediction of  $SL$ .

**Table 1**

The main statistical parameters for the prediction of  $SL$  [34].

Parameter	Category	Min	Max	Average
$W/C$		0.40	0.60	0.50
$W$ ( $kg/m$ )		129.07	283.72	194.10
$C$ ( $kg/m^3$ )	<b>Inputs</b>	303.02	523.28	390.54
$Sa$ ( $kg/m^3$ )		496.50	1023.16	703.30
$BNG$ ( $kg/m^3$ )		778.77	1262.00	1011.85
$SL$ ( $mm$ )	<b>Output</b>	0	270	67.88

### 2.2. Overview of GEP and modeling $SL$

The gene expression programming (GEP) algorithm is a computational algorithm and a member of the extended genetic algorithms and genetic programming [38]. The results of GEP are computer programs developed by modifying their sizes, forms, and compositions to produce complicated tree structures [39]. The GEP algorithm employs linear chromosomes composed of genes generally structured in a head and a tail.

GEP comprises a combination of five major elements, including the fitness function's definition, the terminal set and mathematical functions definition, the determination of chromosome structures such as the number of genes, determining the linking function and the control parameters, and the stop criterion [40]. The results of GEP are expressed in the form of tree-like structures (namely, sub-expression trees (sub-ETs)). Moreover, GEP includes a unique multi-genic feature that enables the evolution of more complicated programs with numerous sub-programs [41]. There is a collection of fixed-length symbols for each GEP gene, and these symbols can represent any component of the function and terminal sets. The function set may include any user-defined function or the basic mathematical operators (+, −, ×, and /).

Each gene in the GEP comprises mathematical operators, variables, and constants used to encode a mathematical formula. The GEP parameters used to calculate the slump of concrete are listed in Table 2. The outcome of the GEP model is shown in Fig. 1. For modeling *SL*, the input variables, including (*W/C*), (*W*), (*C*), (*Sa*), and (*BNG*) were used. Fig. 1 indicates the Sub-ETs were obtained for estimating *SL* with GEP implementation.

For modeling slump concrete with the GEP approach, the Gene Xpro Tools software was utilized. The setting parameter values of the GEP model and genetic operators, such as mutation, inversion, and transportation, are displayed in Table 2. Ebtehaj and Bonakdari [38] suggested that a population size between 30 and 100 can yield satisfactory results. Therefore, this investigation employed 50 chromosomes, selected by trial and error. The generation number was obtained at 300000, and RMSE fitness functions for GEP model development were selected via trial and error. In addition, the function set was considered +, −, ×, ÷, *Exp*, *Ln*,  $x^2$ ,  $x^3$ ,  $\sqrt{\quad}$ ,  $\sqrt[3]{\quad}$ , *Sin*, *Cos* and *Atan*. It is worth mentioning that the RMSE fitness function has the best performance in similar studies [8,42,43]. Previous studies found favorable outcomes using the additional function as a linking function between sub-ETs.

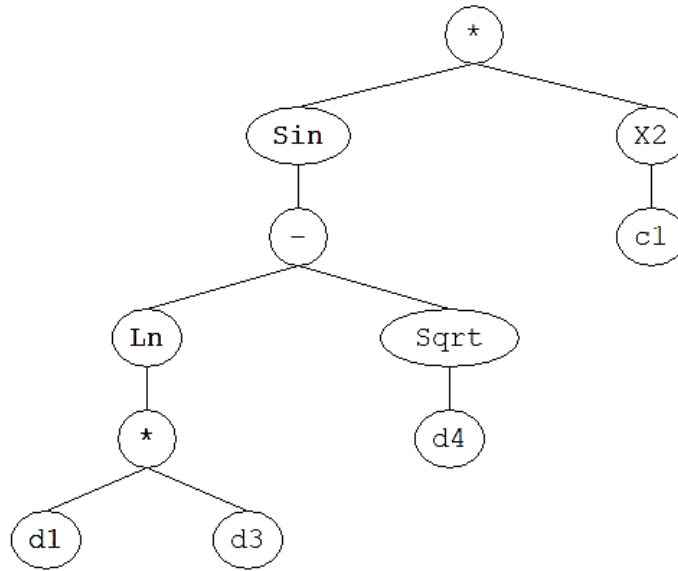
**Table 2**

The setting parameters of the GEP model for the estimation of the slump of concrete.

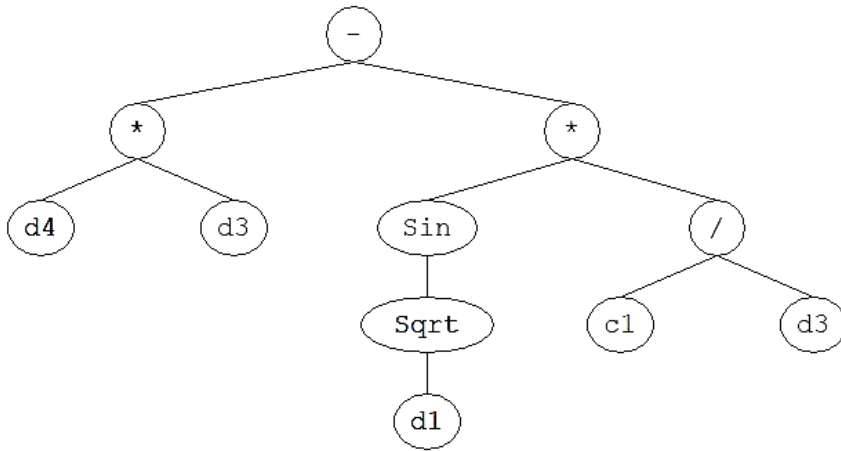
Parameter	Value
Number of genes	3
Mutation rate	0.044
One-point recombination	0.3
Two-point recombination	0.3
IS transposition	0.1
RIS transposition	0.1
Gene recombination	0.1
Gene transposition	0.1

It is worth mentioning the setting parameter of GEP is based on previous studies, and trial and error processes were obtained.

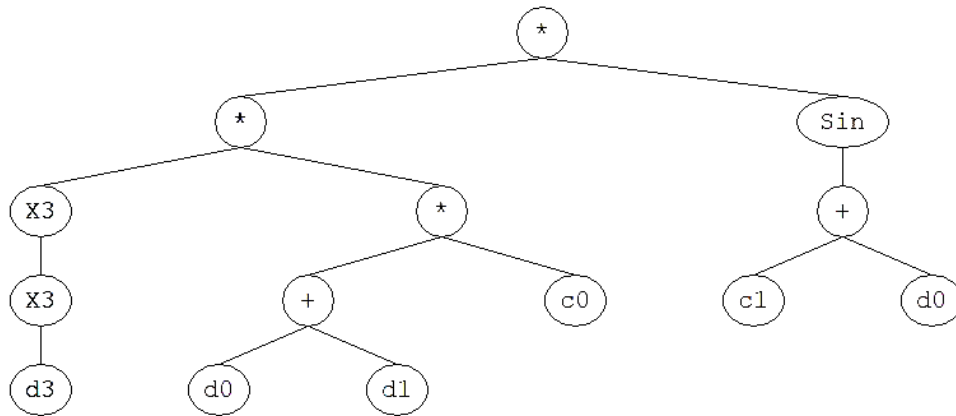
Sub-ET 1



Sub-ET 2



Sub-ET 3



**Fig. 1.** Sub-ETs obtained from GEP for prediction of  $SL$ .

The explicit equations related to Fig. 1 are as follows:

$$\begin{aligned}
\text{Sub - ET 1} &= (\text{Sin}(\ln(d_1 \times d_3) - \sqrt{d_4})) \times c_1^2 \\
\text{Sub - ET 2} &= (d_4 \times d_3) - \text{Sin}(\sqrt{d_1}) \times (c_1/d_3) \\
\text{Sub - ET 3} &= d_3^9 \times [(d_0 + d_1) \times c_0] \times \text{Sin}(c_1 + d_0)
\end{aligned} \tag{2}$$

The value of SL was obtained as follows:

$$SL = \text{Sub - ET 1} + \text{Sub - ET 2} + \text{Sub - ET 3} \tag{3}$$

where the values of  $c_1$  is Sub-ET 1, Sub-ET 2, and Sub-ET 3 are -8.966126, -8.851715, and 6.780976, respectively. In addition, the variables of  $d_0$ ,  $d_1$ ,  $d_3$  and  $d_4$  are *BNG*, *C*, *W/C* and *W*, respectively.

### 2.3. Overview of MARS and modeling SL

Multivariate adaptive regression splines (MARS) is a famous data-driven model commonly used in the civil engineering field with success. MARS is a robust non-parametric regression method that can model complex relationships between dependent and independent variables. The algorithm generates a regression model that predicts the dependent variable based on several independent variables using piecewise regression parts (called basis functions) [44].

The MARS algorithm constructs a piecewise linear regression model by dividing the independent variables into smaller subregions and fitting simple basis functions to each subregion [45]. The algorithm selects the appropriate basis function and independent variables for each subregion and determines the breakpoints or knots that define the boundaries of each subregion.

In the MARS method, the basis function is crucial in capturing the underlying relationships between the input and target variables. MARS utilizes a set of basis functions defined as piecewise linear segments. The algorithm starts with a simple model consisting of a constant term and gradually adds basis functions to capture non-linearities and interactions. At each step, the algorithm assesses the contribution of potential basis functions using a statistical criterion, such as the generalized cross-validation (GCV) score.

The equation of MARS and related basis functions are as follows [46]:

$$SL = \beta_0 + \sum_{i=1}^M \beta_m BF_m(x) \tag{4}$$

$$BF_m(x) = \max(0, c - x) \tag{5}$$

or

$$BF_m(x) = \max(0, x - c) \tag{6}$$

where,  $\beta_0$  is the constant value and,  $\beta_m$  is the corresponding coefficient of BF.  $BF_m(x)$  is the  $m^{\text{th}}$  basis function,  $x$  is the input variable, and  $c$  is the threshold value of the input variable.

The MARS method provided an explicit equation as follows for the prediction of *SL*:

$$SL = 31.6395 - 0.733976 \times BF_1 - 652.467 \times BF_2 - 4.49362 \times BF_3 + 7.2058 \times BF_4 - 0.817229 \times BF_5 \tag{7}$$

where,  $BF_1 = \max(0, 181.66 - W)$ ,  $BF_2 = \max(0, W/C - 0.55)$ ,  $BF_3 = \max(0, W - 209.41)$ ,  $BF_4 = \max(0, W - 192.04)$  and  $BF_5 = \max(0, C - 383.84)$ .

The GCV value of the proposed MARS equation was 1414.98.

### 3. Results and discussions

Common statistical measures, such as root mean squared error (RMSE), correlation coefficient (R), and mean absolute error (MAE), are used to evaluate the accuracy of the proposed algorithms. These statistical measures are as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (8)$$

$$R = \left( \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \right) \quad (9)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (10)$$

where  $x$  and  $\bar{x}$  are observed, and the mean value of  $SL$ . In addition  $y$  and  $\bar{y}$  are predicted and the mean value of  $y$ , respectively.  $n$  is the total number of data. The statistical measurements for the training and testing data set are tabulated in Table 3.

**Table 3**

Statistical values of the GEP and MARS models for estimation of  $SL$  for training and testing datasets.

Approach	RMSE	R	MAE
GEP (Train)	18.553	0.978	14.011
GEP (Test)	19.366	0.975	15.537
MARS (Train)	19.744	0.974	16.127
MARS (Test)	25.118	0.957	17.051

The two data-driven models, GEP and MARS, were used to predict the slump of the concrete. Both models were developed using training data, and their accuracy and performance were evaluated using testing data. As seen in Table 3, the GEP model achieved an RMSE of 18.553, an R of 0.978, and an MAE of 14.011 when developed on the training data. In addition, when evaluated on the testing data, the GEP model achieved an RMSE of 19.366, an R of 0.975, and an MAE of 15.537. On the other hand, the MARS model achieved an RMSE of 19.744, an R of 0.974, and an MAE of 16.127 when developed on the training data. Moreover, when assessed on the testing data, the MARS model achieved an RMSE of 25.118, an R of 0.957, and an MAE of 17.051.

Comparing the results of the proposed models, it can be seen that the GEP model performed better than the MARS model in terms of RMSE and MAE for both training and testing data. The



correlation coefficient (R) for both models is relatively high, indicating a strong linear relationship between the predicted and observed values. It is important to note that while the GEP model performed better overall, the more complex structure of GEP was obtained for the prediction of *SL* compared to the simple equation provided by the MARS model.

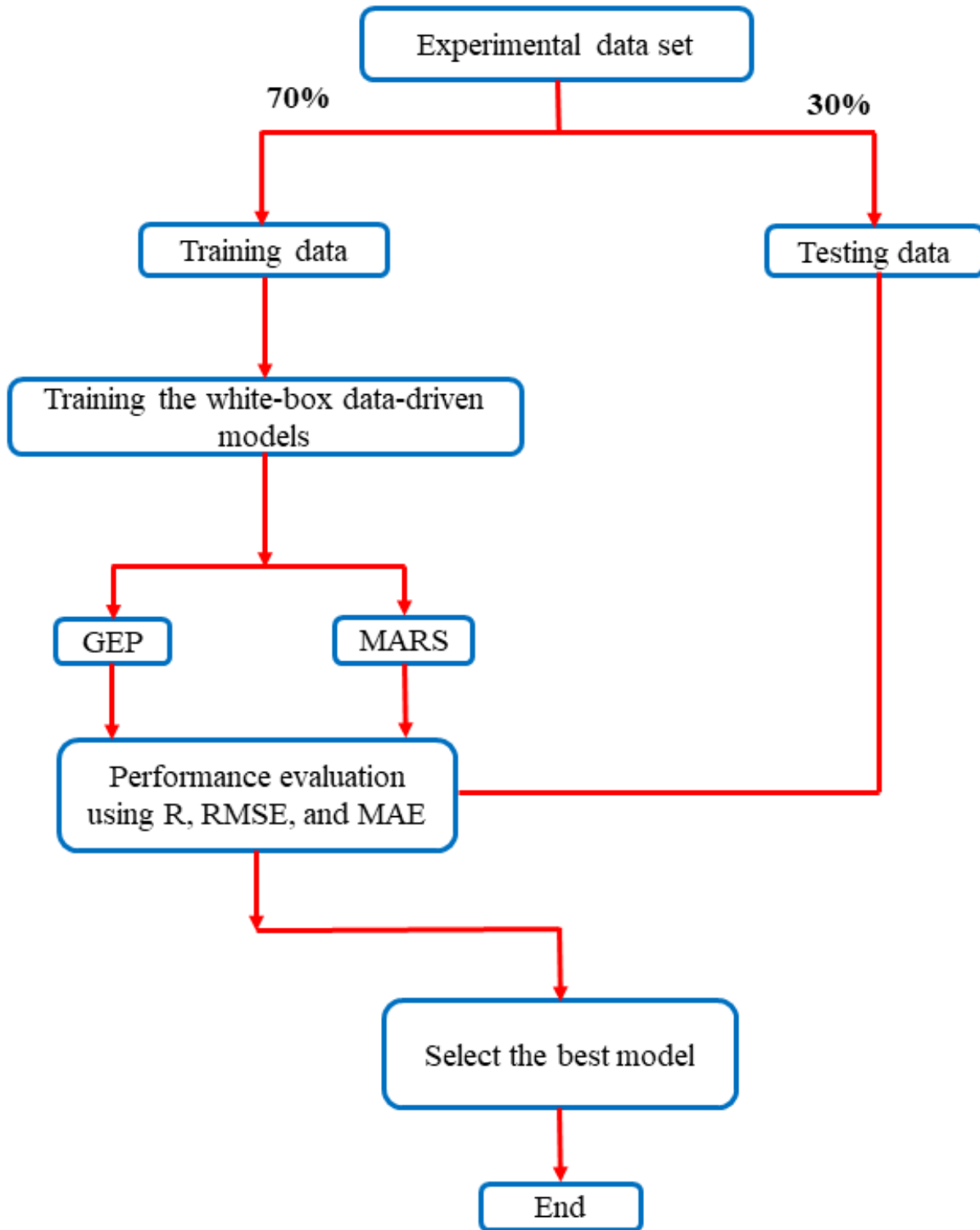
A more in-depth analysis can provide valuable insights into their implications for civil engineering applications when comparing the performance of the GEP and MARS models in terms of error indicators, including RMSE and MAE. RMSE and MAE are commonly used error metrics to assess the performance of data-driven methods. A lower value of RMSE and MAE indicates better agreement between the model predictions and the actual observed values. RMSE represents the average magnitude of the prediction errors made by the model. A lower RMSE indicates that the model's predictions are closer to the actual observed values. MAE measures the average magnitude of the absolute differences between the model predictions and the observed values. It provides a similar interpretation as RMSE but on an absolute scale. Like RMSE, a lower MAE indicates better agreement between the model predictions and the observed values.

In the context of predicting slump concrete, the GEP model demonstrated lower RMSE and MAE values compared to MARS during both the training and testing stages. These results lead to the conclusion that the GEP model outperforms MARS in terms of accuracy for slump concrete prediction. The superior accuracy of GEP over MARS in predicting concrete slumps holds promising practical implications for civil engineering applications. The more accurate predictions obtained from GEP can provide engineers and construction professionals with reliable information regarding the workability and consistency of concrete mixes. This, in turn, enables better planning and optimization of construction processes, leading to improved quality control, cost efficiency, and overall project performance. However, it should be noted that GEP provides a complex equation for predicting the slump of concrete. While the complex equation may offer higher accuracy, it may also pose challenges in terms of interpretation and implementation in real-world scenarios.

On the other hand, although MARS may exhibit lower accuracy compared to GEP in predicting concrete slumps, it still holds practical merits in civil engineering applications. MARS can provide interpretable models, allowing engineers to gain insights into the relationships between variables. Its ability to capture interactions between predictors affecting the slump of concrete. Moreover, MARS provided a more straightforward equation with less complexity than the GEP equation for predicting the slump (SL). This simplicity can be advantageous regarding model interpretation and computational efficiency, especially in cases with limited data or quick exploratory analyses.

In summary, the superior accuracy of GEP in predicting concrete slumps offers significant practical implications for civil engineering applications. However, the complexity of the GEP equation should be considered, as it may affect interpretation and implementation. Despite its lower accuracy, MARS provided interpretable models with simpler equations, making it a viable option for gaining insights into variable relationships and conducting efficient analyses in concrete slump prediction.

The Flowchart of the present study can be summarized in Fig 2.



**Fig. 2.** Flowchart of the present study for the preparation of the slump of concrete.

Moreover, the training and testing data's graphical representations are shown in Figs. 3–6.

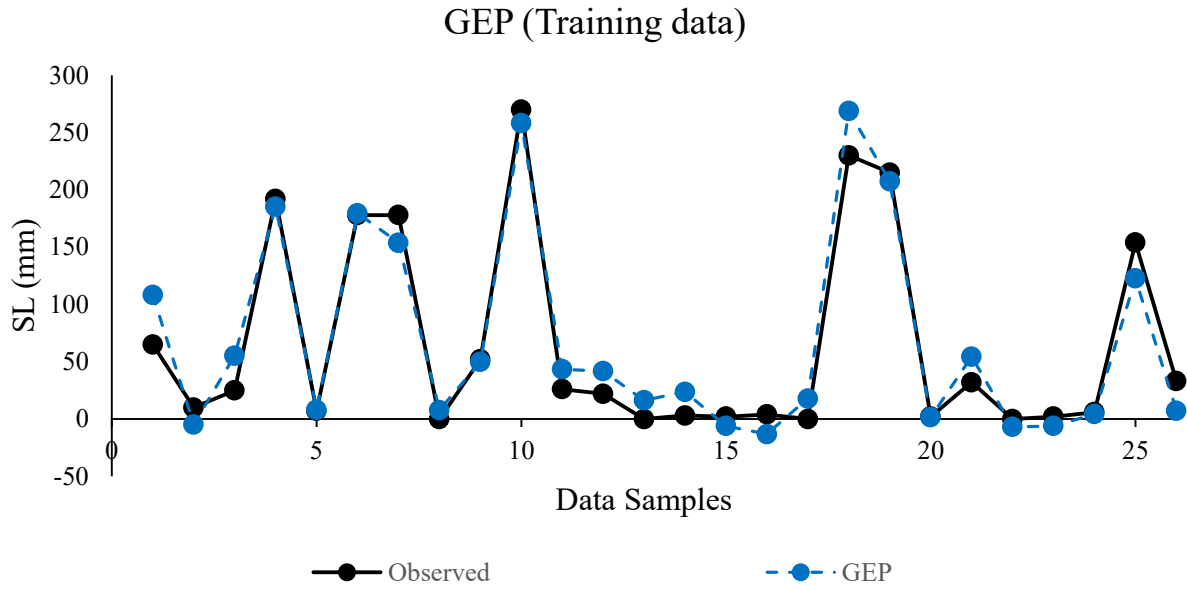


Fig. 3. GEP results for prediction of *SL* using training data.

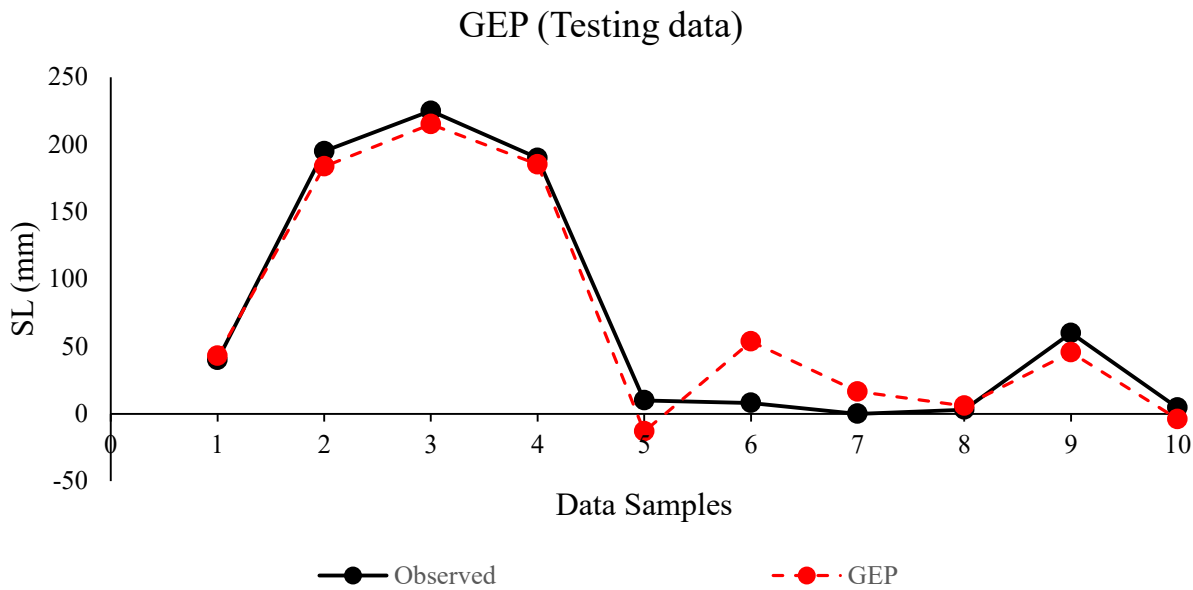
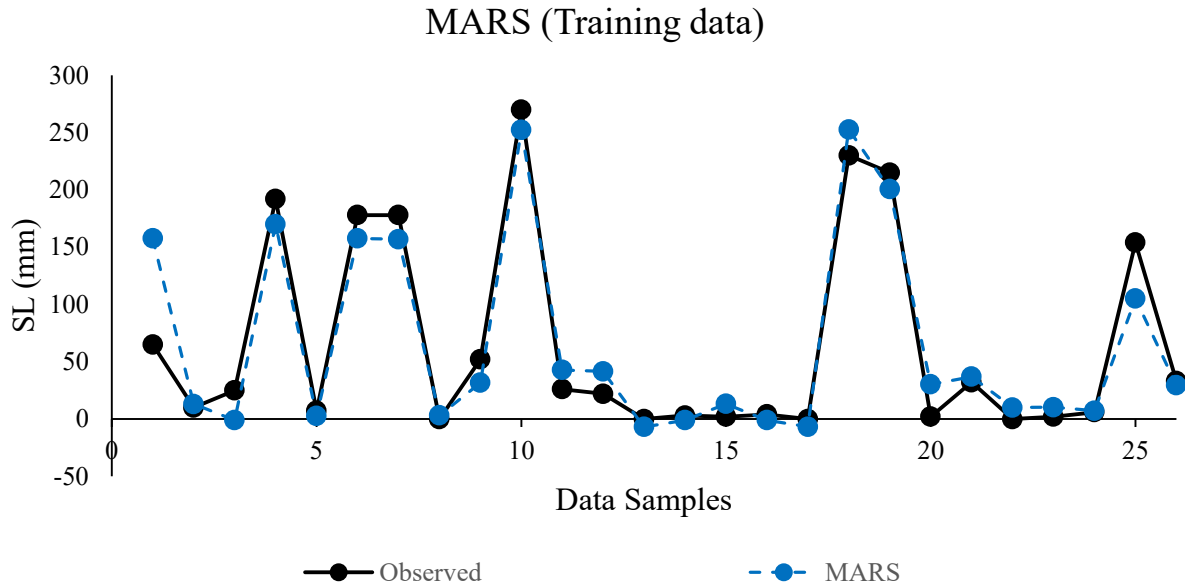
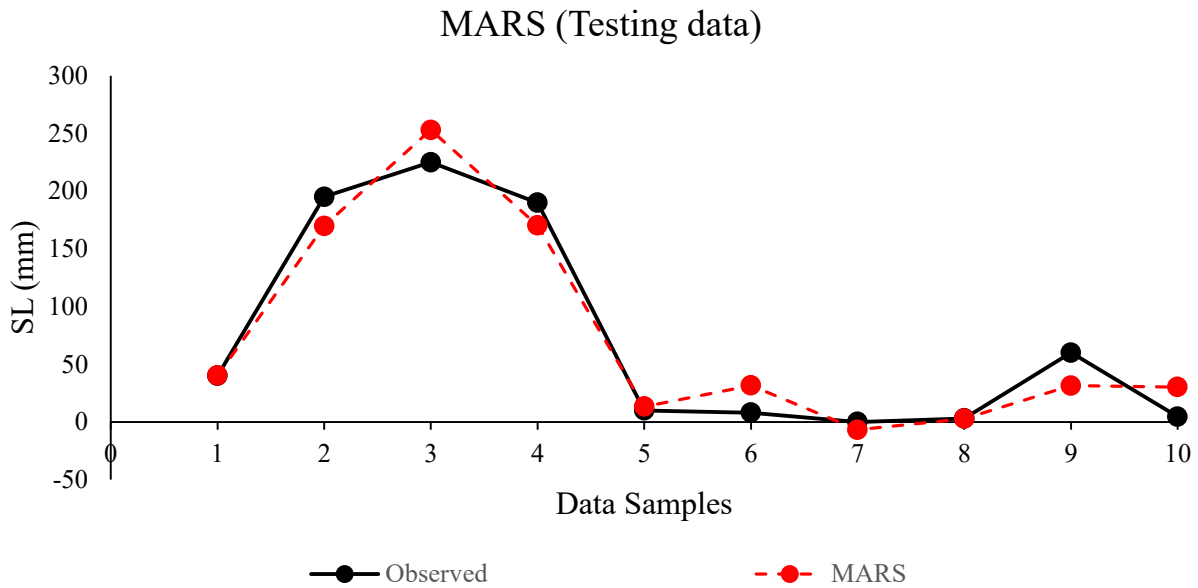


Fig. 4. GEP results for prediction of *SL* using testing data.



**Fig. 5.** MARS results for prediction of  $SL$  using training data.



**Fig. 6.** MARS results for prediction of  $SL$  using testing data.

These figures indicated that the GEP model better captured the complex relationships between independent variables for the prediction of  $SL$ . The testing data results for the GEP model had lower prediction errors and were closer to the observed values than the MARS model. The training data results also showed that the GEP model had better fit and generalization ability, as it closely followed the observed values and had fewer prediction errors. In addition, for more comparison, the study's results were compared for all data sets with the MLR model proposed by the earlier study [34]. Table 4 provides the values of statistical measures for the GEP, MAR, and MLR models for the prediction of  $SL$ .

**Table 4**Statistical values of the proposed models for estimation of *SL* for all data sets.

Approach	RMSE	R	MAE
GEP	19.143	0.976	15.113
MARS	23.748	0.962	16.795
MLR [34]	29.417	0.942	23.309

The results showed that GEP outperformed MARS in terms of accuracy for predicting the slump in concrete. The statistical measures of GEP, including RMSE, R, and MAE, were 19.143, 0.976, and 15.113, respectively. On the other hand, the statistical measures of MARS, including RMSE, R, and MAE, were 23.748, 0.962, and 16.795, respectively. It is important to note that both algorithms performed well in predicting the *SL*, as indicated by the high correlation coefficients (R) and low RMSE and MAE values. However, GEP demonstrated a higher level of accuracy, as reflected in the lower values of the other statistical measures.

In addition, comparing the results with the MLR model with RMSE = 29.417, R = 0.942, and MAE = 23.309 revealed the highest performance of the proposed data-driven models, GEP and MARS, for the prediction of *SL*. It is worth mentioning that compared to the black-box model (i.e., the ANN model), the explicit mathematical expression was proposed for the prediction of *SL*. Furthermore, the values of R=0.98 for the ANN model by Yusuf et al. [34] compared to the R values of GEP (R=0.976) and MARS (R=0.962) indicated the acceptable performance of GEP and MARS as powerful data-driven models for prediction of *SL*.

Therefore, it was concluded that the proposed white-box data-driven models were more accurate than the traditional regression approach. In fact, traditional regression techniques face inherent limitations when accurately estimating complex problems related to concrete properties. These limitations arise due to the linear nature of traditional regression models, which struggle to capture the nonlinear relationships and interactions present in such complex systems. To address these limitations, researchers have turned to data-driven models that offer greater flexibility and adaptability in capturing complex patterns. GEP and MARS are two such data-driven approaches that have shown promise in overcoming the limitations of traditional regression techniques.

GEP is an evolutionary-based, data-driven model that can automatically evolve mathematical expressions to model complex relationships. By incorporating nonlinear functions and interactions, GEP enables a more accurate estimation of concrete properties compared to traditional regression techniques. Similarly, MARS is a flexible data-driven model that can capture nonlinear relationships and interactions using a piecewise regression approach. By adaptively partitioning the data and fitting regression models within each partition, MARS provides a more robust estimation of complex concrete properties.

It is worth mentioning that the mathematical equations presented by the GEP and MARS methods can be beneficial for civil engineers in estimating slump concrete. Civil engineers can use the proposed equations to estimate the slump of concrete without knowing soft computing

methods. The presented mathematical equations do not require any special software for the estimation of the concrete slump. Moreover, Unlike the ANN method, the GEP model is regarded as a white-box, data-driven approach capable of establishing mathematical relationships among the relevant variables in the problem. Furthermore, in comparison to the MLR technique, GEP can generate a complex equation that accurately depicts the relationship between the influencing variables in the problem, facilitating the estimation of the concrete slump with greater precision.

#### 4. Summary and conclusions

This study compared the accuracy and performance of two explicit data-driven algorithms, gene expression programming (GEP) and multivariate adaptive regression splines (MARS), for predicting the slump of concrete (*SL*). Statistical measures such as RMSE, MAE, and R were used to assess the method's accuracy. According to the evaluation metrics, the GEP method makes better predictions than MARS and the regression method. The outcomes demonstrated that the *SL* values predicted by GEP and MARS could accurately estimate the *SL*.

In conclusion, this study has demonstrated that GEP is a more accurate algorithm than MARS for predicting *SL*. The findings of this study may have important implications for the concrete industry, as accurate predictions of *SL* values can help optimize the production process and ensure the quality of the final product. The GEP model uses complex structures and complex equations to estimate concrete slumps. In contrast, the MARS model has used simple regression relationships to estimate concrete slumps.

These findings suggest that GEP can be a more effective algorithm for predicting a slump in concrete. The higher accuracy of GEP can be attributed to its ability to model nonlinear relationships and interactions between variables to predict *SL* values. In contrast, MARS has a less complex formula for predicting *SL* and is more convenient to estimate *SL* than the complicated GEP formula.

Overall, the results of this study provide valuable insights into the application of data-driven methods for predicting *SL* values. The superior performance of GEP over MARS highlights the importance of selecting appropriate algorithms for accurate and reliable predictions. These findings can be useful for researchers and engineers working in the field of concrete technology, as well as in other fields where accurate prediction of complex systems is critical. This study considered two white-box data-driven models for concrete slump estimation, including the GEP and MARS models. To compare the performance of these two models, it is suggested for future work that the obtained results from the present study be compared with other white-box data-driven models that can provide mathematical equations for concrete slump estimation, such as decision trees (DTs) and group method of data handling (GMDH) approaches.

#### Funding

This research received no external funding.

## Conflicts of interest

The authors declare no conflict of interest.

## Authors contribution statement

**Ismail Husein:** Conceptualization, Methodology, Original draft; **Ramaswamy Sivaraman:** Original draft, Formal analysis, Investigation, Methodology; **Sarwar Hasan Mohammad:** Writing – review & editing, Investigation; **Forqan Ali Hussein Al-Khafaji:** Visualization, Software; **Sokaina Issa Kadhim:** Resources, Validation; **Yusof Rezakhani:** Analyzed the key findings, Supervision, Writing – review & editing. All authors have read and agreed to the published version of the manuscript.

## References

- [1] Moein MM, Saradar A, Rahmati K, Rezakhani Y, Ashkan SA, Karakouzian M. Reliability analysis and experimental investigation of impact resistance of concrete reinforced with polyolefin fiber in different shapes, lengths, and doses. *J Build Eng* 2023;69:106262. <https://doi.org/10.1016/j.jobe.2023.106262>.
- [2] Ghazanfari N, Gholami S, Emad A, Shekarchi M. Evaluation of GMDH and MLP Networks for Prediction of Compressive Strength and Workability of Concrete. *Bull La Société R Des Sci Liège* 2017;855–68. <https://doi.org/10.25518/0037-9565.7032>.
- [3] Chen Y, Wu J, Zhang Y, Fu L, Luo Y, Liu Y, et al. Research on Hyperparameter Optimization of Concrete Slump Prediction Model Based on Response Surface Method. *Materials (Basel)* 2022;15:4721. <https://doi.org/10.3390/ma15134721>.
- [4] Ghalandari M, Ziamolki A, Mosavi A, Shamshirband S, Chau K-W, Bornassi S. Aeromechanical optimization of first row compressor test stand blades using a hybrid machine learning model of genetic algorithm, artificial neural networks and design of experiments. *Eng Appl Comput Fluid Mech* 2019;13:892–904. <https://doi.org/10.1080/19942060.2019.1649196>.
- [5] Band SS, Janizadeh S, Chandra Pal S, Saha A, Chakraborty R, Melesse AM, et al. Flash Flood Susceptibility Modeling Using New Approaches of Hybrid and Ensemble Tree-Based Machine Learning Algorithms. *Remote Sens* 2020;12:3568. <https://doi.org/10.3390/rs12213568>.
- [6] Ghasemi M, Samadi M, Soleimani E, Chau K-W. A comparative study of black-box and white-box data-driven methods to predict landfill leachate permeability. *Environ Monit Assess* 2023;195:862. <https://doi.org/10.1007/s10661-023-11462-9>.
- [7] Alipour M, Hashemi Gholpayeghani SMR. Real-time non-uniform EEG sampling. *Biomed Signal Process Control* 2021;70:102961. <https://doi.org/10.1016/j.bspc.2021.102961>.
- [8] Ahmadi S, Kamalian M, Askari F. Literature Review of Estimating the Bearing Capacity of Rough Footings by the Stress Characteristic Lines Method. *Bull Earthq Sci Eng* 2020;7:147–61.
- [9] Fayaz SA, Zaman M, Butt MA. Numerical and Experimental Investigation of Meteorological Data Using Adaptive Linear M5 Model Tree for the Prediction of Rainfall. *Rev Comput Eng Res* 2022;9:1–12. <https://doi.org/10.18488/76.v9i1.2961>.
- [10] Madina B, Gumilyov LN. Determination of the Most Effective Location of Environmental Hardenings in Concrete Cooling Tower Under Far-Source Seismic Using Linear Spectral Dynamic Analysis Results. *J Res Sci Eng Technol* 2020;8:22–4. <https://doi.org/10.24200/jrset.vol8iss1pp22-24>.

- [11] Naderpour H, Mirrashid M. Bio-inspired predictive models for shear strength of reinforced concrete beams having steel stirrups. *Soft Comput* 2020;24:12587–97. <https://doi.org/10.1007/s00500-020-04698-x>.
- [12] Naderpour H, Haji M, Mirrashid M. Shear capacity estimation of FRP-reinforced concrete beams using computational intelligence. *Structures* 2020;28:321–8. <https://doi.org/10.1016/j.istruc.2020.08.076>.
- [13] Naderpour H, Khatami SM, Barros RC. Prediction of Critical Distance Between Two MDOF Systems Subjected to Seismic Excitation in Terms of Artificial Neural Networks. *Period Polytech Civ Eng* 2017. <https://doi.org/10.3311/PPci.9618>.
- [14] Akbarzadeh MR, Ghafourian H, Anvari A, Pourhanasa R, Nehdi ML. Estimating Compressive Strength of Concrete Using Neural Electromagnetic Field Optimization. *Materials (Basel)* 2023;16:4200. <https://doi.org/10.3390/ma16114200>.
- [15] Parsa P, Naderpour H. Shear strength estimation of reinforced concrete walls using support vector regression improved by Teaching–learning–based optimization, Particle Swarm optimization, and Harris Hawks Optimization algorithms. *J Build Eng* 2021;44:102593. <https://doi.org/10.1016/j.jobe.2021.102593>.
- [16] Zhu P, Saadati H, Khayatnezhad M. Application of probability decision system and particle swarm optimization for improving soil moisture content. *Water Supply* 2021;21:4145–52. <https://doi.org/10.2166/ws.2021.169>.
- [17] Samadi M, Jabbar E. Assessment of Regression Trees and Multivariate Adaptive Regression Splines for Prediction of Scour Depth Below the Ski-Jump Bucket Spillway. *J Hydraul* 2012;7:73–9. <https://doi.org/10.30482/jhyd.2012.85350>.
- [18] Cao C. Prediction of concrete porosity using machine learning. *Results Eng* 2023;17:100794. <https://doi.org/10.1016/j.rineng.2022.100794>.
- [19] Alhakeem ZM, Jebur YM, Henedy SN, Imran H, Bernardo LFA, Hussein HM. Prediction of Ecofriendly Concrete Compressive Strength Using Gradient Boosting Regression Tree Combined with GridSearchCV Hyperparameter-Optimization Techniques. *Materials (Basel)* 2022;15:7432. <https://doi.org/10.3390/ma15217432>.
- [20] Golafshani EM, Behnood A, Arashpour M. Predicting the compressive strength of normal and High-Performance Concretes using ANN and ANFIS hybridized with Grey Wolf Optimizer. *Constr Build Mater* 2020;232:117266. <https://doi.org/10.1016/j.conbuildmat.2019.117266>.
- [21] M. H. Badawi Y, Hummaida Ahmed Y. Prediction of Concrete Compressive Strength & Slump using Artificial Neural Networks (ANN). *FES J Eng Sci* 2021;9:84–9. <https://doi.org/10.52981/fjes.v9i2.682>.
- [22] Timur Cihan M. Prediction of Concrete Compressive Strength and Slump by Machine Learning Methods. *Adv Civ Eng* 2019;2019:1–11. <https://doi.org/10.1155/2019/3069046>.
- [23] Tang F, Wu Y, Zhou Y. Hybridizing Grid Search and Support Vector Regression to Predict the Compressive Strength of Fly Ash Concrete. *Adv Civ Eng* 2022;2022:1–12. <https://doi.org/10.1155/2022/3601914>.
- [24] Behnood A, Golafshani EM. Machine learning study of the mechanical properties of concretes containing waste foundry sand. *Constr Build Mater* 2020;243:118152. <https://doi.org/10.1016/j.conbuildmat.2020.118152>.
- [25] Rajakarunakaran SA, Lourdu AR, Muthusamy S, Panchal H, Jawad Alrubaie A, Musa Jaber M, et al. Prediction of strength and analysis in self-compacting concrete using machine learning based regression techniques. *Adv Eng Softw* 2022;173:103267. <https://doi.org/10.1016/j.advengsoft.2022.103267>.



- [26] Chine W-H, Chen L, Hsu H-H, Wang T-S, Chiu C-H. Modeling Slump of Concrete Using the Artificial Neural Networks. 2010 Int. Conf. Artif. Intell. Comput. Intell., IEEE; 2010, p. 236–9. <https://doi.org/10.1109/AICI.2010.287>.
- [27] Agrawal V, Sharma A. Prediction of slump in concrete using artificial neural networks 2010;70:25–32.
- [28] Rezaie M, Sadighi N. Prediction of slump and density of lightweight concretes using ANFIS and linear regression. *Int J Civ Eng Technol* 2017;8:1635–48.
- [29] Islam MN, Zain MFM, Jamil M. PREDICTION OF STRENGTH AND SLUMP OF RICE HUSK ASH INCORPORATED HIGH-PERFORMANCE CONCRETE. *J Civ Eng Manag* 2012;18:310–7. <https://doi.org/10.3846/13923730.2012.698890>.
- [30] Onikeku O, Shitote SM, Mwero J, Adedeji AA, Kanali C. Compressive Strength and Slump Prediction of Two Blended Agro Waste Materials Concretes. *Open Civ Eng J* 2019;13:118–28. <https://doi.org/10.2174/1874149501913010118>.
- [31] Öztaş A, Pala M, Özbay E, Kanca E, Çag˘lar N, Bhatti MA. Predicting the compressive strength and slump of high strength concrete using neural network. *Constr Build Mater* 2006;20:769–75. <https://doi.org/10.1016/j.conbuildmat.2005.01.054>.
- [32] Singh P, Bhardwaj S, Dixit S, Shaw RN, Ghosh A. Development of Prediction Models to Determine Compressive Strength and Workability of Sustainable Concrete with ANN, 2021, p. 753–69. [https://doi.org/10.1007/978-981-16-0749-3\\_59](https://doi.org/10.1007/978-981-16-0749-3_59).
- [33] Yeh I-C. Modeling slump flow of concrete using second-order regressions and artificial neural networks. *Cem Concr Compos* 2007;29:474–80. <https://doi.org/10.1016/j.cemconcomp.2007.02.001>.
- [34] Yusuf A, Abdullahi M, Sadiku S, Aguwa JI, Alhaji B, Folorunso TA. Modelling Slump of Concrete Containing Natural Coarse Aggregate from Bida Environs Using Artificial Neural Network. *J Soft Comput Civ Eng* 2021;5:19–38. <https://doi.org/10.22115/scce.2021.268839.1272>.
- [35] Dwijendra NKA, Sharma S, Asary AR, Majdi A, Muda I, Mutlak DA, et al. Economic Performance of a Hybrid Renewable Energy System with Optimal Design of Resources. *Environ Clim Technol* 2022;26:441–53. <https://doi.org/10.2478/rtuct-2022-0034>.
- [36] Seyedhosseini SM, Esfahani MJ, Ghaffari M. A novel hybrid algorithm based on a harmony search and artificial bee colony for solving a portfolio optimization problem using a mean-semi variance approach. *J Cent South Univ* 2016;23:181–8. <https://doi.org/10.1007/s11771-016-3061-9>.
- [37] Bouchaala F, Ali MY, Matsushima J, Bouzidi Y, Jouini MS, Takougang EM, et al. Estimation of Seismic Wave Attenuation from 3D Seismic Data: A Case Study of OBC Data Acquired in an Offshore Oilfield. *Energies* 2022;15:534. <https://doi.org/10.3390/en15020534>.
- [38] Ebtahaj I, Bonakdari H. No-Deposition Sediment Transport in Sewers Using Gene Expression Programming. *J Soft Comput Civ Eng* 2017;1:29–53. <https://doi.org/10.22115/scce.2017.46845>.
- [39] Samadi M, Sarkardeh H, Jabbari E. Prediction of the dynamic pressure distribution in hydraulic structures using soft computing methods. *Soft Comput* 2021;25:3873–88. <https://doi.org/10.1007/s00500-020-05413-6>.
- [40] Naderpour H, Sharei M, Fakharian P, Heravi MA. Shear Strength Prediction of Reinforced Concrete Shear Wall Using ANN, GMDH-NN and GEP. *J Soft Comput Civ Eng* 2022;6:66–87. <https://doi.org/10.22115/SCCE.2022.283486.1308>.
- [41] Golmohammadi A-M, Tavakkoli-Moghaddam R, Jolai F. Concurrent cell formation and layout design using a genetic algorithm under dynamic conditions. *J Res Sci Eng Technol* 2015;2:5–9. <https://doi.org/10.24200/jrset.vol2iss01pp5-9>.
- [42] Aslanova F. A Comparative Study of the Hardness and Force Analysis Methods Used in Truss Optimization with Metaheuristic Algorithms and Under Dynamic Loading. *J Res Sci Eng Technol* 2020;8:25–33. <https://doi.org/10.24200/jrset.vol8iss1pp25-33>.

- [43] Zahmatkesh S, Rezakhani Y, Arabi A, Hasan M, Ahmad Z, Wang C, et al. An approach to removing COD and BOD based on polycarbonate mixed matrix membranes that contain hydrous manganese oxide and silver nanoparticles: A novel application of artificial neural network based simulation in MATLAB. *Chemosphere* 2022;308:136304. <https://doi.org/10.1016/j.chemosphere.2022.136304>.
- [44] Shafagh Loron R, Samadi M, Shamsai A. Predictive explicit expressions from data-driven models for estimation of scour depth below ski-jump bucket spillways. *Water Supply* 2023;23:304–16. <https://doi.org/10.2166/ws.2022.421>.
- [45] Samadi M, Jabbari E, Azamathulla HM, Mojallal M. Estimation of scour depth below free overfall spillways using multivariate adaptive regression splines and artificial neural networks. *Eng Appl Comput Fluid Mech* 2015;9:291–300. <https://doi.org/10.1080/19942060.2015.1011826>.
- [46] Samadi M, Sarkardeh H, Jabbari E. Explicit data-driven models for prediction of pressure fluctuations occur during turbulent flows on sloping channels. *Stoch Environ Res Risk Assess* 2020;34:691–707. <https://doi.org/10.1007/s00477-020-01794-0>.