

Estimating Coastal Dyke Leakage Flow Using Support Vector Machine (SVM) and Multivariate Adaptive Regression Spline (MARS) Model

Adel Asakereh*, Farhad Choobi**, Mohammad Bagherzadeh***, Reza Mirzaee****

ARTICLE INFO

RESEARCH PAPER

Article history:

Received:

September 2023

Revised:

January 2024

Accepted:

June 2024

Keywords:

Soft computing,

Seepage discharge,

Coastal dyke,

SVM method,

MARS model

Abstract:

Checking for leakage flow in hydraulic and marine structures during design practice is a crucial step, as uncontrolled leakage can cause irreparable damage. Soft computing methods can be used to easily model, analyze, and control complex systems. This study uses Support Vector Machine (SVM) method to predict leakage discharge of coastal dykes. Five different models are employed to achieve this goal, with parameters including the length of the cutoff blanket, dyke depth, and water head considered. The best support vector machine model is checked using a multivariate adaptive regression spline model (MARS) for prediction. Results show that the model including all parameters predicts settlement discharge with an acceptable accuracy compared to the laboratory model, with a coefficient of determination and root mean square coefficient of 0.949 and 0.058, respectively in the train stage, and 0.93 and 0.06, in the test phase estimates. The dyke depth parameter has the most significant effect on leakage flow, while the water head has a negligible effect on the model. Although the adaptive regression multivariate spline model accurately estimates the annual dyke leakage flow rate, it is marginally less accurate than the support vector machine method.

1. Introduction

Monitoring leakage flow within the body of earth dams, as well as the foundations of hydraulic and marine structures, is crucial for assessing water loss, stability, water pressure distribution, and designing elements including drain thickness and length. Terzaghi [1] pioneered investigating the impact of clay curtains and cutoffs on leakage discharge in earth dams and introduced the concept of critical hydraulic gradient. Bligh [2] defined the creep length as the combined horizontal and vertical distance traversed by a fluid particle from the upstream bed surface. Building upon prior work, Lane [3] formulated his creep theory after analysing over 200 dams worldwide. He postulated that the vertical head drop exceeds the horizontal movement in seepage flow. In a significant contribution, Neuman and Witherspoon [4] explored seepage flow beneath earth dams by employing the finite element method.

They concluded that the finite element analysis, is potentially a powerful approach for comprehensively simulating seepage-related challenges. The USBR [5] proposed that the length of the upstream clay blanket should ideally be at least 10 times the magnitude of the upstream head for effective control of seepage losses. In a study by Ojha et al. [6], they determined the critical hydraulic gradient for boiling. Benmebarek [7] used the FLAC-2D computer code to conduct numerical analyses on seepage-induced failure in sandy foundations. Fontana [8] assessed critical heads to predict hydraulic structure failure due to uplift pressure on a sand foundation, while comparing the outlet hydraulic gradient with the critical gradient. Yousefi et al. [9] investigated the conditions that cause boiling, exploring both vertical and angled shields. Their findings demonstrated that a depth ratio of 0.4 between the shield and foundation is most effective at preventing boiling. Irzooki [10] conducted a numerical investigation on a homogeneous earth dam with a horizontal drain located in the toe. They presented an empirical equation to calculate sediment flow using SEEP/W. Then, using an artificial neural network (ANN), they showed that the length of the horizontal drain

*Corresponding author: Associate Professor, Faculty of Civil Engineering, Semnan University, Semnan, Iran. Email: asakereh@semnan.ac.ir

**MSc student, Faculty of Civil Engineering, Semnan University, Semnan, Iran.

*** PhD student, Department of Civil Engineering, Urmia University, Urmia, Iran.

****PhD student, Faculty of Civil Engineering, Semnan University, Semnan, Iran.

has a more significant effect on the seepage flow than the upstream slope of the dam. Kheiri et al. [11] also conducted a numerical study on the seepage phenomenon under the earth dam using SEEP/W. Their results showed that the best location to construct a cutoff wall is beneath the clay core. Soft computing methods like artificial intelligence (AI) have been increasingly utilized in recent years to solve hydraulic problems, water resources, and more. These methods result from new scientific efforts that enable easy modeling, analysis, and control of complex systems with great success. Among these methods are neural networks (ANN), fuzzy inference system (ANFIS), and support vector machine (SVM) methods [12]. Sharqi et al. [12] estimated the piezometric head of an earth dam using hybrid artificial intelligence models. Their results showed that using nonlinear hybrid models with artificial neural networks increased modeling accuracy. Najafzadeh [13] used genetic programming (GP), the M5 algorithm, and the group method of data handling (GMDH) to predict piezometric pressure and leakage discharge of the Shahid Kazmi Bukan earth dam. The M5 algorithm had the best performance in piezometric pressure estimation, whereas, compared to the GMDH, it performed the best in leakage flow estimation. Bagherzadeh et al. [14] used a numerical model and AI to estimate the energy dissipation of a vertical drop fall with stepped edges. Their results suggest that an artificial neural network (ANN) is the best model for predicting energy dissipation, with three steps being the optimal number for the flow. Mirzaee et al. [15] used various models to estimate the scour depth caused by symmetrical cross jets. Their results showed that the neural network method had the highest accuracy and the tailwater depth was the most influential parameter.

Calculating the leakage flow rate of hydraulic structures, particularly coastal dykes, is essential to effectively manage the salinity of coastal lands and ensure the stability of marine structures. This precision is important not only for managing

coastal salinity but also for maintaining marine construction stability. Soft computing methods are significant in modeling and forecasting desired parameters. The investigation highlights the potential of soft computing approaches, focusing on accurate prediction. The support vector machine (SVM) technique stands out as a sophisticated form of regression, adept at estimating values across diverse scenarios. This study uses SVM to discern leakage discharge of coastal dykes with five models. Additionally, to evaluate the accuracy of the SVM method, a comparative analysis is conducted using the adaptive regression multivariate spline technique. This secondary study aims to estimate the annual leakage flow rate of dykes, focusing on the optimal SVM model. To achieve this goal, the research relies on the laboratory data originally compiled by Sedghi-Asl et al. [16].

2. Material and Methods

The data used in this study to predict leakage flow under coastal dykes was obtained from experiments conducted by Sedghi-Asl et al. in 2010 [16]. The experiments were carried out in a laboratory flume with dimensions of 8 meters by 1 meter and plexiglass walls. Fine sand was used to create the most critical potential situation under erosion, to a depth of 0.5 meters. Piezometers were placed in 4 rows at depths of 10, 20, 30, and 40 cm from the flume bed to measure the piezometric pressure downstream of the wall. The range of considered parameters for the laboratory model is presented in Table 1, and the schematic of the laboratory model is shown in Figure 1.

Table 1: Range of laboratory model parameters

Parameter		Cutoff blanket (L_B)	Water head (h)	Dyke depth (d)
Range (m)	Min.	1	0.05	0.1
	Max.	3.5	0.25	0.5

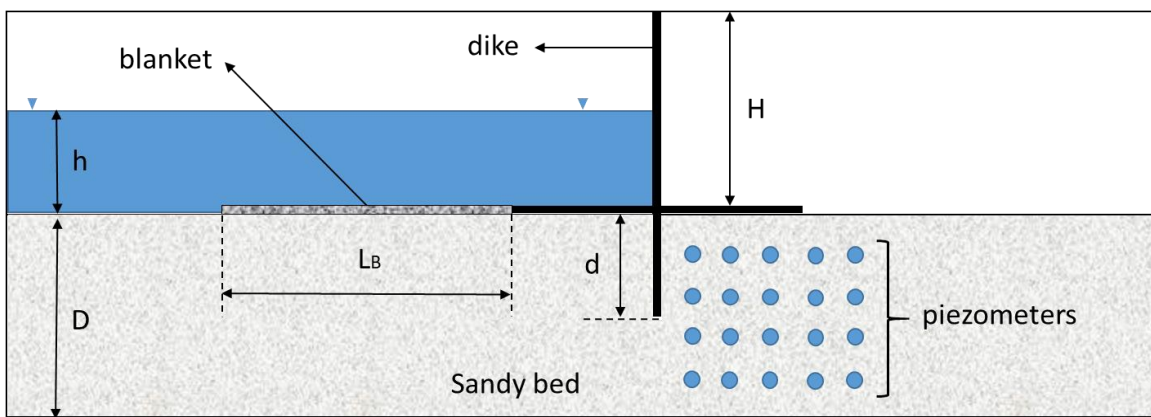


Fig. 1: Scheme of the laboratory model

Using the Buckingham- π theorem, we can define dimensionless laboratory parameters for the model's geometric and hydraulic conditions. These parameters include $\frac{q}{q_0}$, $\frac{L_B}{h_m}$, $\frac{d}{D}$, and $\frac{h}{H}$. Here, q is the leakage flow rate, q_0 is the highest leakage flow rate, L_B is the cutoff blanket length, h_m is the highest water head (0.25 m), d is the sheet pile depth in the foundation, D is the foundation depth (0.5 m), and H is the sheet pile height (0.5 m). Equation (1) presents the leakage flow rate.

$$\frac{q}{q_0} = f\left(\frac{L_B}{h_m}, \frac{d}{D}, \frac{h}{H}\right) \quad (1)$$

The Support Vector Machine (SVM) algorithm is a supervised learning method used for pattern recognition, which was originally introduced by Vapnik [17]. It uses the regression method to solve classification and prediction problems. SVM works by training and testing, linearly classifying the data, and selecting a line with a high confidence margin in the linear division of the data. In other words, the main task of SVM is a linear classification of data, and it is optimal to select a line in the division process that includes a high degree of reliability. The regression SVM model can take the problem to a higher dimensional space by the kernel method. Due to the separation of data into two classes, there is an infinite number of lines in the two-dimensional. Unlike other genetic algorithms like method Roulette Wheel Tree, SVM always has a unique solution and takes less time to achieve the optimal solution. [18- 20].

Multivariate adaptive regression splines (MARS) were first introduced by Fredman [21]. This model is a non-parametric regression method and provides a flexible algorithm to organize the relationship between the input parameters and the target variable and its prediction [22]. The MARS model classifies the inputs in the training period using the divide-and-conquer strategy and forms a stepwise regression relationship for each part [23]. Based on equation 2 in this model, the value of variable Y is calculated based on input x.

$$Y = \sum_{i=1}^n \beta_i BF_i(x) \quad (2)$$

In this relation, BF_i , β_i , and n represent the i-th basis function, the weight of the i-th basis function, and the number of basis functions respectively. The value of the basis functions is calculated using the (3) and (4) relations.

$$BF_i = \max(0, x - c_i) \quad (3)$$

$$BF_i = \max(0, c_i - x) \quad (4)$$

In the above relations, c_i represents the i-th basis node or knot. The MARS model's structure is presented in Figure 2 [15].

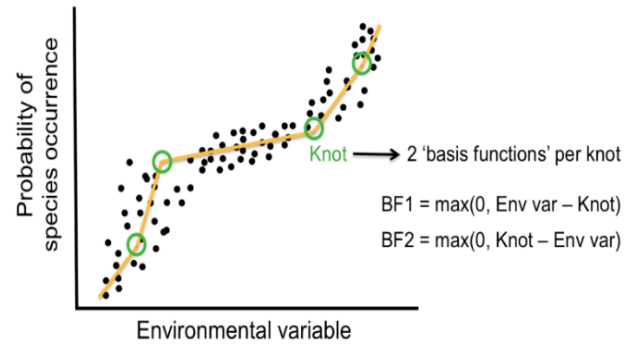


Fig. 2: Examples of nodes in the MARS model [15]

In the current research model, the maximum number of basic functions was set to 21, with a degree of interaction value of 1, a penalty value of 2, and a threshold value of 0.0005 in the implementation settings [15].

In this study, a support vector machine was used to predict leakage flow in coastal dykes based on two evaluation parameters described by relations 5 and 6. The model with the RMSE value approaching zero and the R^2 value approaching one is considered the best [15, 19].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n \left[\left(\frac{q}{q_0} \right)_{Dep} - \left(\frac{q}{q_0} \right)_{Pre} \right]^2}{n}} \quad (5)$$

$$R^2 = \frac{\left(\sum_{i=1}^n \left[\left(\frac{q}{q_0} \right)_{Dep} - \overline{\left(\frac{q}{q_0} \right)_{Dep}} \right] \times \left[\left(\frac{q}{q_0} \right)_{Pre} - \overline{\left(\frac{q}{q_0} \right)_{Pre}} \right] \right)^2}{\sqrt{\sum_{i=1}^n \left[\left(\frac{q}{q_0} \right)_{Dep} - \overline{\left(\frac{q}{q_0} \right)_{Dep}} \right]^2} \times \sqrt{\sum_{i=1}^n \left[\left(\frac{q}{q_0} \right)_{Pre} - \overline{\left(\frac{q}{q_0} \right)_{Pre}} \right]^2}} \quad (6)$$

In these relationships, the Pre index relates to predicted values, the Dep index relates to lab values, and n is the data count.

3. Results and Discussions

In this research, the Support Vector Machine (SVM) method used the Radial Basis Function (RBF) kernel function has been employed to predict leakage discharge in coastal dykes. Various proportions were applied during the Training and Testing phases to estimate leakage flow using the SVM algorithm in a dimensionless mode. After assessing different models and selecting the most suitable one, 75% of the data was allocated to the Training stage, while 25% of the data was assigned to the Testing stage. Subsequently, the results from the different models were compared within the SVM.

The methodology followed the following sequence: initially, relevant parameters influencing leakage flow rate were determined through dimensional analysis. Next, the parameters derived from the dimensional analysis were categorized into five distinct combinations based on dependent and independent parameters, as outlined in Table 2. The input combinations, comprising dimensionless independent parameters, were then input into the SVM network. Following this, the kernel parameter (γ) values were defined in the SVM algorithm through trial and error. Among the models implemented, the optimal approach for calculating leakage flow rate was identified as the best model within the SVM method. In this best model, the SVM approach was evaluated using the multivariate spline estimation model of adaptive regression. Finally, the output results of these two models were presented after comparing the evaluation criteria. Refer to Table 2 for a list of model names and the constituent parameters of each model, which are utilized within the SVM program to model leakage flow. These details are provided under the section titled "Structure of the Models."

Table 2: Input combinations to SVM algorithm

Model	Structure of the Models	Model	Structure of the Models
A	$\frac{q}{q_0} = f(d/D)$	D	$\frac{q}{q_0} = f(L_B/h_m, h/H)$
B	$\frac{q}{q_0} = f(d/D, L_B/h_m)$	E	$\frac{q}{q_0} = f(d/D, L_B/h_m, h/H)$
C	$\frac{q}{q_0} = f(d/D, h/H)$		

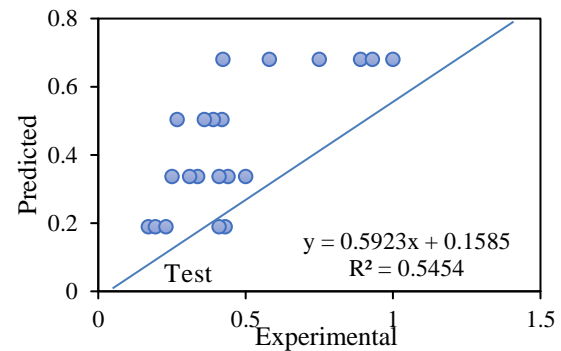
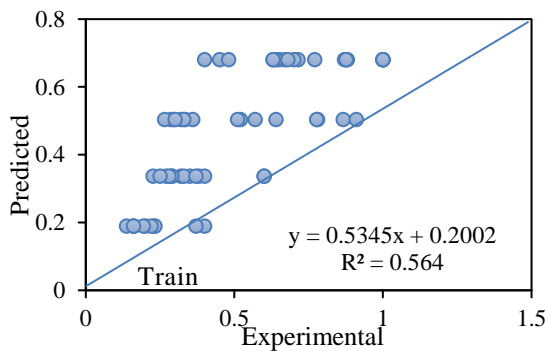
Multiple SVM models were trained and their R^2 and RMSE values were recorded. The results for all combinations were then presented in Table 3, which showed that the issue of then presented in Table 3, which showed that the issue of leakage flow was most effectively addressed when all independent parameters were taken into account. For all

models, the evaluation criteria in training mode consistently outperformed those in test mode, indicating the accuracy of the support vector machine. Model (E) with a correlation coefficient of 0.934 and root mean square error coefficient of 0.06 for the test mode, along with a correlation coefficient of 0.949 and error coefficient of 0.058 for the training mode was selected as the optimal model for this research. Model (D) represented the least accurate calculation of leakage flow rate with RMSE coefficients of 0.22 and 0.211 for test and training modes. These values indicate that calculating the leakage flow rate is not feasible without the parameter (d/D).

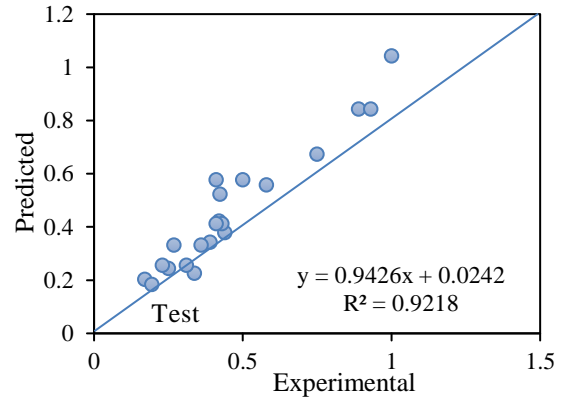
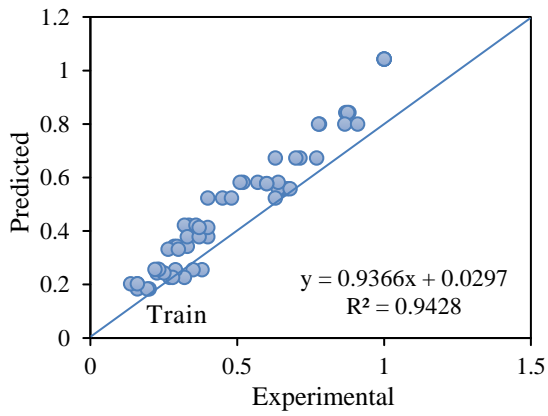
Table 3: Prediction results of different models for leakage flow

Model	Train		Test	
	R^2	RMSE	R^2	RMSE
A	0.564	0.157	0.545	0.161
B	0.943	0.061	0.922	0.066
C	0.606	0.162	0.509	0.171
D	0.312	0.211	0.166	0.220
E	0.949	0.058	0.934	0.060

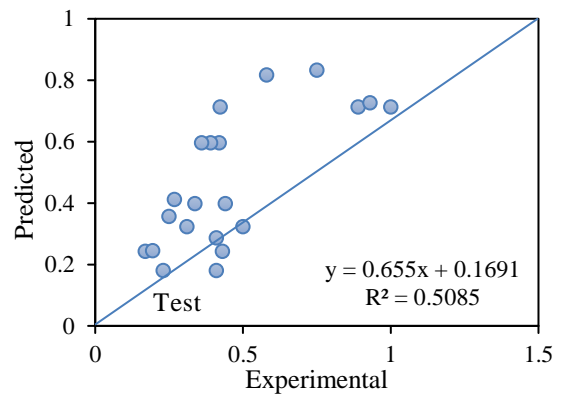
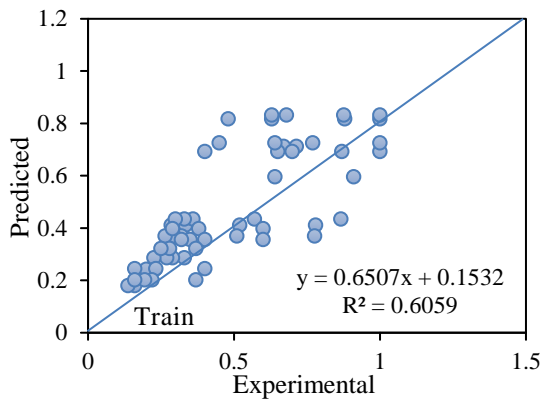
Figure 3 displays the results of all models in both training and testing modes. Upon observing these five models, it can be concluded that the most optimal solution is presented by model (E). In essence, the performance of support vector machine in predicting the leakage flow model, encompassing all relevant parameters, surpasses that of the other models. Furthermore, the marginal discrepancy between the results of this model in the training and test stages underscores the remarkable precision of SVM in the forecasting model (E). A careful examination of Figure 3 reveals a striking correspondence between the values of the laboratory flow rate and predictions for both test and training modes, showcasing an exceptional alignment with the highest correlation coefficient.



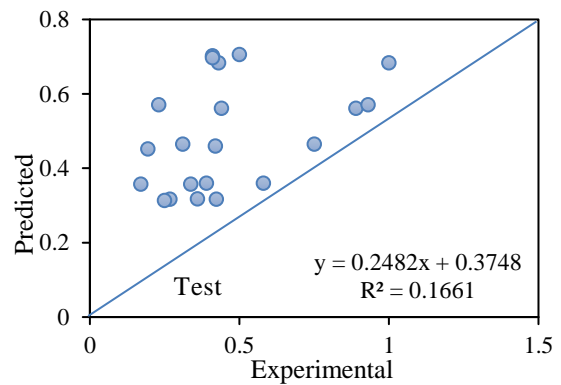
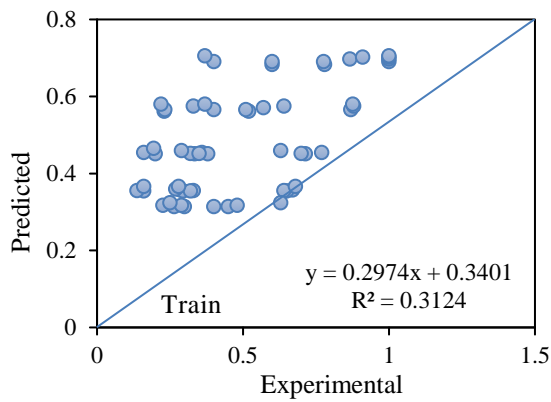
(3a)



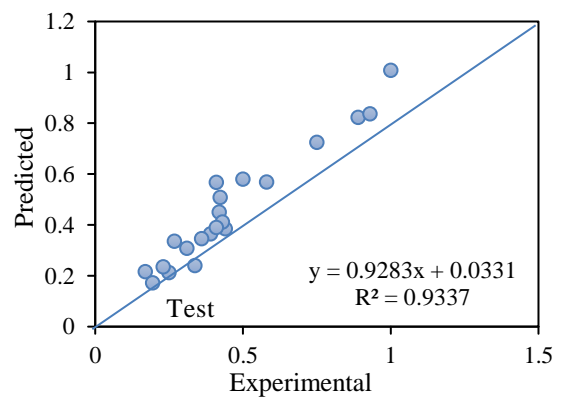
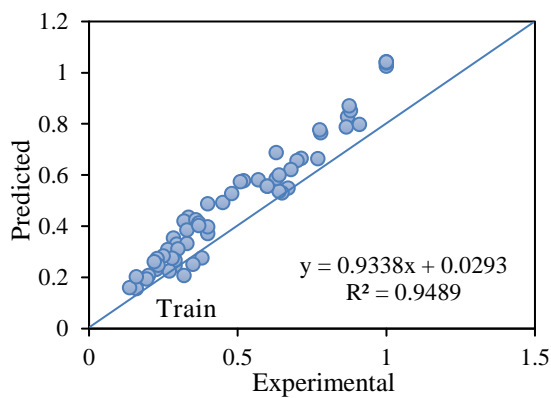
(3b)



(3c)



(3d)



(3e)

Fig. 3: Answer five models for training and testing mode

In numerical studies, sensitivity analysis is used to determine the impact of output variables on input variables in a statistical model. The effect of parameters on the leakage flow rate is determined by removing them one by one in sensitivity analysis. The most and least impactful parameters are also identified through this analysis. This research uses a combination of models to conduct sensitivity analysis by removing an independent parameter in each model. The study found that the worst estimation of leakage flow occurred when the parameter (d/D) was not included in the structural model. The direction of flow (flow lines) increased with the increase of dyke depth leading to the depletion of leakage current energy in the soil. The parameter (L_b/h_m) was recognized as the second most

influential parameter. Model (B) shows that the parameter (h/H) has a minor effect on the leakage flow estimation. First, the best model for predicting leakage flow was identified as model (E) using support vector machine. The same estimation was repeated using the MARS model for comparison. The MARS model had a coefficient of determination of 0.871 and 0.905 for test and training modes respectively, with a root mean square error coefficient of 0.084 and 0.078 for test and training modes. Although the MARS model had a good estimate of the leakage flow rate, the comparison with the SVM model revealed that the SVM model was more reliable at checking the leakage flow rate. Figure 4 shows the comparative graph of leakage flow results for SVM and MARS models.

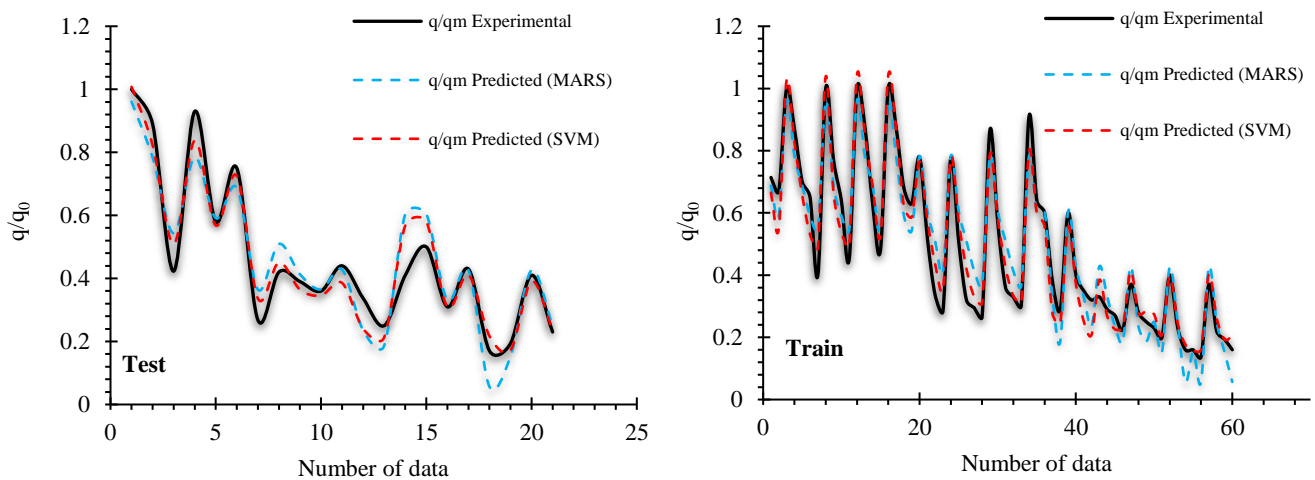


Fig. 4: Model answer for both SVM and MARS

4. Conclusions

To test the ability of the support vector machine method in predicting and estimating leakage flow in coastal dykes, Sedghi-Asl et al.'s [16] laboratory data were used after pre-processing. Of a total of 81 data series, 75% of the data was used for network training, while the remaining 25% was used for testing. Results showed that the SVM method can predict leakage discharge in coastal dykes with acceptable accuracy. Model (E) was identified as the best model in leakage flow estimation with a correlation coefficient of 0.934 a root mean square error coefficient of 0.06 for the test mode, a correlation coefficient of 0.949, and an error coefficient of 0.058 for the training mode. The presence of the parameter (d/D) has the greatest impact on the prediction results. The MARS model was also used to predict the leakage flow rate for the superior model determined by the SVM method, and the comparison of the evaluation criteria of the two methods showed that the accuracy of the SVM method is higher compared to the MARS model method.

References

- [1] Terzaghi, K. (1943). Theoretical soil mechanics, Wiley, New York.
- [2] Bligh, W. G. (1910). Dams, barrages and weirs on porous foundations. Engineering news, 64(26), 708-710.
- [3] Lane, E. W. (1935). Security from under-seepage-masonry dams on earth foundations. Transactions of the American Society of Civil Engineers, 100(1), 1235-1272.
- [4] Neuman, S. P., & Witherspoon, P. A. (1970). Finite element method of analyzing steady seepage with a free surface. Water Resources Research, 6(3), 889-897.
- [5] United States Department of the Interior, Bureau of Reclamation (USBR). (1977). Design of small dams. Washington, DC.
- [6] Ojha, C. S. P., Singh, V. P., & Adrian, D. D. (2003). Determination of critical head in soil piping. Journal of Hydraulic Engineering, 129(7), 511-518.

- [7] Benmebarek, N., Benmebarek, S., & Kastner, R. (2005). Numerical studies of seepage failure of sand within a cofferdam. *Computers and Geotechnics*, 32(4), 264-273.
- [8] Fontana, N. (2008). Experimental analysis of heaving phenomena in sandy soils. *Journal of Hydraulic Engineering*, 134(6), 794-799.
- [9] Yousefi, M., Sedghi-Asl, M., & Parvizi, M. (2016). Seepage and boiling around a sheet pile under different experimental configuration. *Journal of Hydrologic Engineering*, 21(12), 06016015.
- [10] Irzooki, R. H. (2016). Computation of seepage through homogenous earth dams with horizontal toe drain. *Engineering and Technology Journal*, 34(3 part), 430-440.
- [11] Kheiri, G., Javdanian, H., & Shams, G. (2020). A numerical modeling study on the seepage under embankment dams. *Modeling Earth Systems and Environment*, 6(2), 1075-1087.
- [12] Sharghi, E., Nourani, V., & Behfar, N. (2019). Evaluation and application of ensemble AIbased models for estimating piezometric heads in earth fill dams. *Iran-Water Resources Research*, 14(4), 160-169.
- [13] Najafzadeh, M. (2020). Projection of seepage and piezometric pressure in earth dams using soft computational models. *Amirkabir Journal of Civil Engineering*, 52(7), 1783-1796 (In Persian).
- [14] Bagherzadeh, M., Mousavi, F., Manafpour, M., Mirzaee, R., & Hoseini, K. (2022). Numerical simulation and application of soft computing in estimating vertical drop energy dissipation with horizontal serrated edge. *Water Supply*, 22(4), 4676-4689.
- [15] Mirzaee, R., Mohammadi, M., Mousavi, S. F., Bagherzadeh, M., & Hosseini, K. (2023). Application of soft computing techniques to estimate the scouring depth formed by crossing jets. *Water Science and Technology*, 87(8), 1853-1865.
- [16] Sedghi-Asl, M., Rahimi, H., & Khaleghi, H. (2010). Experimental analysis of seepage flow under coastal dikes. *Experimental techniques*, 34(4), 49-54.
- [17] Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.
- [18] Roushangar, K., & Koosheh, A. (2015). Evaluation of GA-SVR method for modeling bed load transport in gravel-bed Rivers. *Journal of Hydrology*, 527, 1142-1152.
- [19] Daneshfaraz, R., Bagherzadeh, M., Esmaeeli, R., Norouzi, R., & Abraham, J. (2021a). Study of the performance of support vector machine for predicting vertical drop hydraulic parameters in the presence of dual horizontal screens. *Water supply*, 21(1), 217-231.
- [20] Bagherzadeh, M., & Mohammadi, M. (2023). Estimation of the Scouring Depth of the Plunge Pool of the Symmetrical Crossing Jets by Support Vector Machine. *Journal of Water and Sustainable Development*, 9(4), 1-12 (In Persian).
- [21] Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, 19(1), 1-67.
- [22] Zhang, W., & Goh, A. T. (2016). Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geoscience Frontiers*, 7(1), 45-52.
- [23] Rezaie-Balf, M., Zahmatkesh, Z., & Kim, S. (2017). Soft computing techniques for rainfall-runoff simulation: local non-parametric paradigm vs. model classification methods. *Water Resources Management*, 31, 3843-3865.



This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license.