

## Experimental investigation and application of soft computing models for predicting flow energy loss in arc-shaped constrictions

Hamidreza Abbaszadeh <sup>a</sup>, Rasoul Daneshfaraz <sup>b,\*</sup>, Veli Sume <sup>c</sup> and John Abraham <sup>d</sup>

<sup>a</sup> Department of Civil Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran

<sup>b</sup> Department of Civil Engineering, Faculty of Engineering, University of Maragheh, Maragheh, Iran

<sup>c</sup> Department of Civil Engineering, Faculty of Engineering and Architecture, Recep Tayyip Erdogan University, Rize, Turkiye

<sup>d</sup> School of Engineering, University of St. Thomas, St. Paul, MN 33901, USA

\*Corresponding author. E-mail: daneshfaraz@maragheh.ac.ir

 HA, 0000-0001-7714-6081; RD, 0000-0003-1012-8342; VS, 0000-0001-8251-2461; JA, 0000-0002-3818-8681

### ABSTRACT

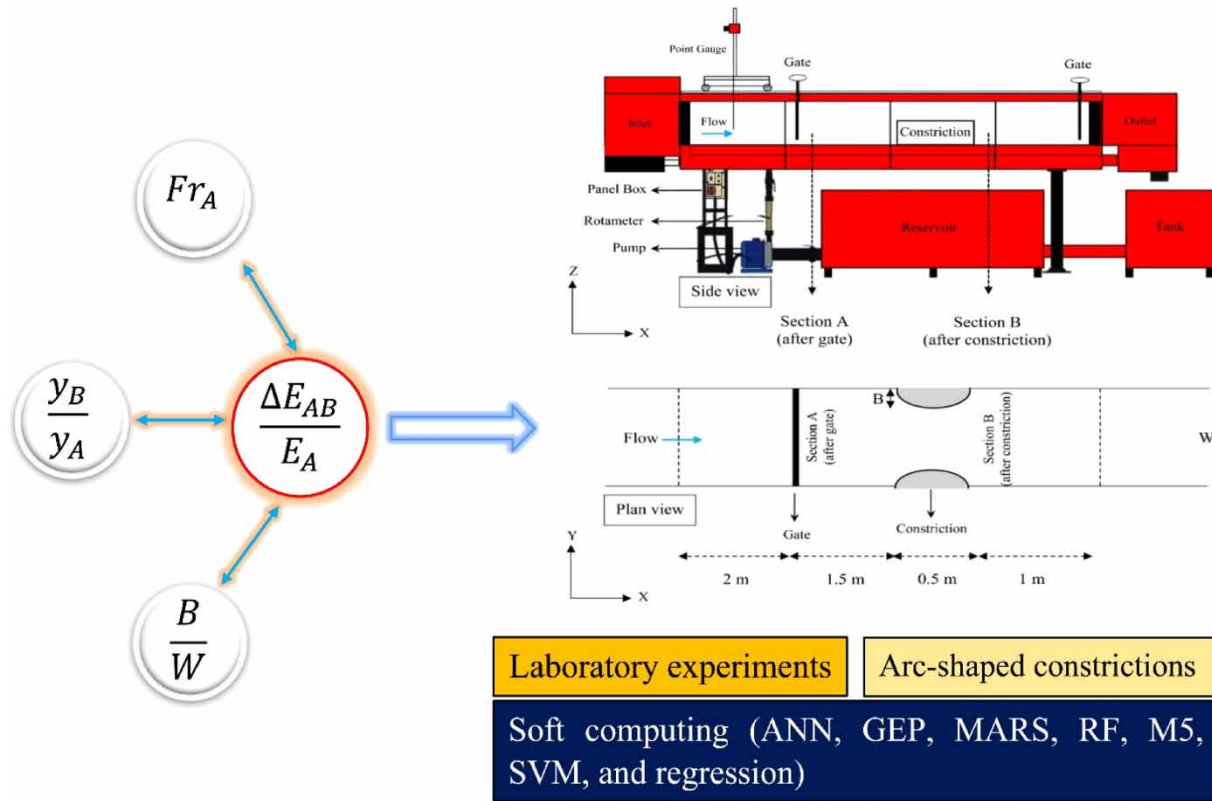
This investigation focuses on flow energy, a crucial parameter in the design of water structures such as channels. The research endeavors to explore the relative energy loss ( $\Delta E_{AB}/E_A$ ) in a constricted flow path of varying widths, employing Support Vector Machine (SVM), Artificial Neural Network (ANN), Gene Expression Programming (GEP), Multiple Adaptive Regression Splines (MARS), M5 and Random Forest (RF) models. Experiments span a Froude number range from 2.85 to 8.85. The experimental findings indicate that the  $\Delta E_{AB}/E_A$  exceeds that observed in a classical hydraulic jump with constriction section. Within the SVM model, the linear kernel emerges as the best predictor of  $\Delta E_{AB}/E_A$ , outperforming polynomial, radial basis function (RBF), and sigmoid kernels. In addition, in the ANN model, the MLP network was more accurate compared to the RBF network. The results indicate that the relationship proposed by the MARS model can play a significant role resulting in high accuracy compared to the non-linear regression relationship in predicting the target parameter. Upon comprehensive evaluation, the ANN method emerges as the most promising among the candidates, yielding superior performance compared to the other models. The testing phase results for the ANN-MLP are noteworthy, with  $R = 0.997$ , average RE% = 0.63%, RMSE = 0.0069, BIAS = -0.0004, DR = 0.999, SI = 0.0098 and KGE = 0.995.

**Key words:** artificial intelligence, constricted arc-shaped, energy loss, nonlinear regression

### HIGHLIGHTS

- This research reinforces the important of investigating the effect of arc-shaped constrictions in the flow path (such as constrictions from bridge piers).
- This investigation improves the design of hydraulic control structures.
- The performance of the ANN, GEP, MARS, M5, RF, SVM, and regression models has been evaluated using quantitative and qualitative indices (KGE, R, RE%, RMSE, BIAS, DR, scatter index (SI)).

GRAPHICAL ABSTRACT



INTRODUCTION

One of the predominant challenges encountered in the downstream segments of many hydraulic structures is the effective management of the surplus kinetic energy within the flow. This need often requires the implementation of energy dissipating structures to avert potential structural damage. These energy dissipating structures are strategically positioned to regulate and diminish velocity, thereby facilitating the dissipation of excess energy downstream. The dissipation process is intricately linked to flow turbulence and disturbances, with a noteworthy portion of energy dissipated through the deliberate constriction of a specific channel area. This study shows the answer to a gap in the study of channel cross-sectional reduction for the purpose of enhancing energy dissipation.

To address elevated levels of energy, hydraulic jumps are often used to reduce kinetic energy. Additionally, impediments strategically positioned within the flow path contribute to energy absorption. In recent years, various studies have been conducted on hydraulic jumps and energy dissipation. Karbasi & Azamathulla (2016) investigated the characteristics of hydraulic jumps on a rough bed using the Gene Expression Programming (GEP) model. The results indicated that the GEP model is capable of predicting the features of a hydraulic jump on a rough bed with an acceptable level of accuracy. A comparison of Artificial Neural Network (ANN) and Support Vector Machine (SVM) models revealed that the performance of these models is slightly superior to the GEP model. Habibzadeh *et al.* (2019) investigated flow characteristics of downstream hydraulic jumps with and without blocks. The range of Froude numbers varied from 3.48 to 6.85. The results of their study indicated that this flow regime of submerged jumps can effectively be used as an energy dissipator within a stilling basin with a length approximately equal to that required for free hydraulic jumps. Ghaderi *et al.* (2020) conducted numerical simulations of free and submerged hydraulic jumps over various roughness shapes in different configurations and under varying Froude numbers using FLOW-3D software. The results indicated that the influence of roughness is more pronounced in reducing the maximum relative velocity in submerged jumps. Additionally, the greatest energy losses occur with triangular roughness elements compared to other models. Nouri *et al.* (2020) investigated the accuracy of M5P, Random Forest (RF) and stochastic M5P models in predicting

the energy loss in cascade spillways. Their results showed that M5P model is more accurate compared to other models. [Rahmanshahi & Shafai Bejestan \(2020\)](#) carried out an experimental study focusing on inclined ramps featuring both smooth and rough surfaces. The investigation took into account various slopes and material sizes. The findings of the study revealed that a higher relative roughness corresponds to more significant energy loss. Moreover, an increase in the ramp's slope was a contributing factor to increased energy loss. In an effort to provide predictive tools, the researchers introduced two mathematical models utilizing a GEP model to estimate energy loss in ramps with both smooth and rough surfaces. [Nasrabadi \*et al.\* \(2021\)](#), utilizing the novel DGMDH technique, focused on predicting the characteristics of submerged hydraulic jumps. The results demonstrated that the DGMDH model, in comparison to the GMDH model, exhibits high accuracy in predicting relative depth, jump length, and relative energy dissipation. Furthermore, they recommended the utilization of this model for estimating the parameters of hydraulic jumps. [Sauda \(2022\)](#) investigated the relative energy loss of a hydraulic jump downstream of multi-gates using ANN model. The results show that ANN model is more accurate than regression; ANN models can be used to predict energy loss in multi-gates. [Heydari \*et al.\* \(2022\)](#) modeled the lengths of hydraulic jumps on rough beds using the Self-Adaptive Extreme Learning Machine (SAELM) machine learning approach. The superior performance of the SAELM model was compared with Multilayer Perceptron Neural Network (MLPNN) and SVM methods. The examination of the model results demonstrated the high effectiveness of the SAELM model. [Mobayen \*et al.\* \(2023\)](#) investigated computational models Multiple Adaptive Regression Splines (MARS) and EPR in estimating energy loss in gabion spillways. The results indicated that the regression equation derived from the EPR model was more complex than the regression equation derived from the MARS model. [Abbaszadeh \*et al.\* \(2023\)](#) experimentally investigated the hydraulic jump parameters in the threshold condition applied in sluice gates. The results revealed that the application of the threshold leads to an increase in energy dissipation and a reduction in the secondary flow depth.

Instances of channel constriction or abrupt reduction in cross-sectional area may arise due to the installation of structures such as bridge piers, causing impediments to the flow ([Chow 1959](#); [Henderson 1966](#)). The presence of bridge piers alters the channel geometry, causing a constriction in the cross-sectional area. This alteration disrupts the smooth flow of water, leading to changes in velocity and pressure distribution. The constriction can result in increased turbulence and flow resistance, affecting the overall hydraulic behavior of the channel.

[Hager & Dupraz \(1985\)](#) experimentally investigated the characteristics of flow in a sudden constriction. They reported a good correlation between their research results and theoretical relationships. [Wu & Molinas \(2001\)](#) examined subcritical flow facing a short constriction along the flow. The relationship proposed for calculating the flow discharge showed good agreement with previous research findings. [Dey & Raikar \(2005\)](#) focused on experimental investigation of scour in a long constriction. Their results indicated that reducing the constriction width leads to an increase in scour depth. In the investigation conducted by [Jan & Chang \(2009\)](#), which focused on hydraulic jumps within a rapidly varied contracted flow, the conclusions from the experimental findings emphasized the substantial influence of bed angle on the relative length of hydraulic jumps. Remarkably, that study suggested that this dependence is distinct from the constriction angle of the side-walls. Furthermore, they presented theoretical relationships for the secondary jump depth, considering factors such as the constriction cross-section and the bed slope. Similarly, [Das \*et al.\* \(2014\)](#) investigated the experimental study of relative energy loss in chutes with various slopes and constrictions. They explored energy loss within rapidly contracting flows. Their study revealed a positive correlation between energy dissipation and the slope of the rapid flow. This underscores the significance of bed slope as a determining factor in the dissipation process, as observed in their empirical results. [Daneshfaraz \*et al.\* \(2022a\)](#) investigated hysteresis in triangular constrictions experimentally. They reported that by increasing the cross-sectional area of the triangular constriction, the relative depth decreases, and with changes to the discharge under the same conditions, the flow depth changes.

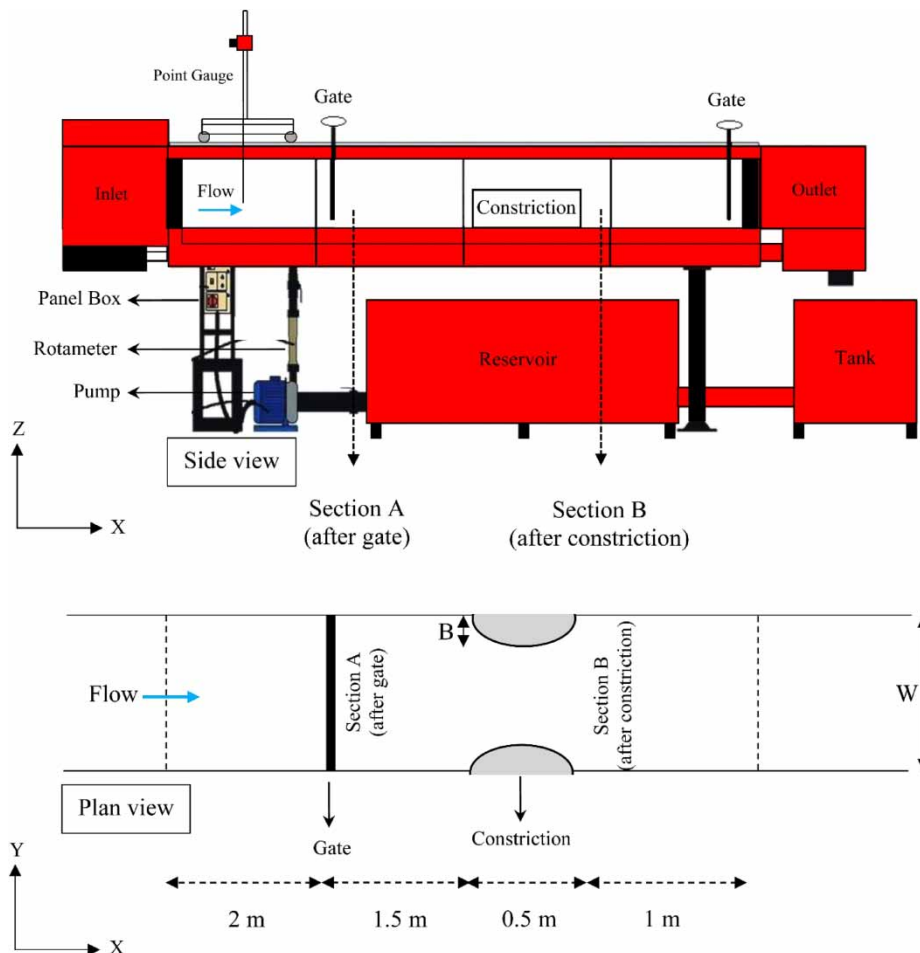
The literature review reveals a gap in the study of channel cross-sectional reduction for the purpose of enhancing energy loss. Consequently, given the significance of this matter, the current research aims to experimentally investigate the impact of arc-shaped constrictions in the flow path. Additionally, the necessity for conducting new research in the field of intelligent modeling and soft computing for assessing essential relative energy loss, which has not been addressed so far, is underscored. To address these objectives, the present study employs artificial intelligence models, including ANN, SVM, RF algorithm, MARS, M5 algorithm, and GEP. The focus of this research is primarily on examining the accuracy of the mentioned intelligent models in arc-shaped constrictions. Furthermore, relationships are presented in the regression-based model (non-linear polynomial regression) alongside the MARS model. Various

statistical indicators such as R, RMSE, average RE, KGE, BIAS, DR, and SI are scrutinized to evaluate the results of the models. In this research, energy loss in arc-shaped constrictions is investigated within the Froude number range of 2.85–8.85.

## METHODS

### Experimental equipment

A 5-m long laboratory flume with a rectangular  $0.3 \text{ m} \times 0.5 \text{ m}$  cross-section and transparent Plexiglas walls and bottom, was used for experiments. All the experiments were carried out in the hydraulic laboratory of the University of Maragheh. The slope of the channel bed was set to zero. To supply the inflow to the flume, a pump with a nominal power of 800 l per minute was used. Rotameters were installed on the flume for reading the inflow discharge with a relative error of 2%. To reduce the turbulence of the inflow water from the reservoir, several calming screens were used. A point depth gauge with an accuracy of  $\pm 1 \text{ ml}$  was used to measure the water depth in the flume. Arc-shaped constrictions were installed at dimensions of 0.5 m in length, 0.05 and 0.075 m in width (0.10 and 0.15 m from both sides) at a distance of 1.5 m from the inlet. Experiments were conducted over a range of Froude numbers from 2.85 to 8.85. Figure 1 shows a schematic view of the flume including a top view that displays the arc-constrictions.



**Figure 1** | Schematic view of experimental flume.

**Relationships and parameters**

Applying conservation of energy, the computation of energy dissipation between sections A and B is conducted for both free and submerged flow conditions, as indicated in Equations (1) and (2) (Fatehi-Nobarian *et al.* 2023).

$$\Delta E_{AB} = E_A - E_B = \left( y_A + \frac{V_A^2}{2g} \right) - \left( y_B + \frac{V_B^2}{2g} \right) \tag{1}$$

$$\Delta E_{AB} = \left( y_{SA} + \frac{V_A^2}{2g} \right) - \left( y_B + \frac{V_B^2}{2g} \right) \tag{2}$$

In the provided equations, the variables are defined as follows:  $y_A$  and  $y_B$  denote the flow depth at sections A and B,  $y_{SA}$  represents the submerged flow depth at section A,  $V_A$  and  $V_B$  are the flow velocities at sections A and B,  $g$  denotes gravitational acceleration,  $E_A$  and  $E_B$  stand for the specific energy at sections A and B, and  $\Delta E_{AB}$  signifies the specific energy difference between the two sections. These parameters collectively contribute to the calculation of energy dissipation between the specified sections under both free and submerged flow conditions.

Determining and selecting the input parameters are important steps in modeling processes that use intelligent methods. In this section, the dimensionless parameters affecting the energy loss in the constriction of the flow path are introduced and different combinations of the parameters are used for modeling (Table 1).

In the current research, the most significant parameters affecting the flow energy loss in the constriction are:

$$f_1(Q, W, B, L, E_A, E_B, y_A, y_B, g, \rho, \mu) = 0 \tag{3}$$

where  $Q$  represents discharge,  $W$  represents the channel width,  $B$  represents the constriction width dimensions,  $L$  represents the constriction length,  $\rho$  represents the specific gravity of fluid and  $\mu$  represents the dynamic viscosity. According to the  $\pi$ -Buckingham theorem and selecting  $y_A$ ,  $\rho$ , and  $g$ , as repeated parameters, dimensionless parameters are obtained.

$$f_2\left(Fr_A, \frac{W}{y_A}, \frac{B}{y_A}, \frac{L}{y_A}, \frac{E_A}{y_A}, \frac{E_B}{y_A}, \frac{y_B}{y_A}, Re_A\right) = 0 \tag{4}$$

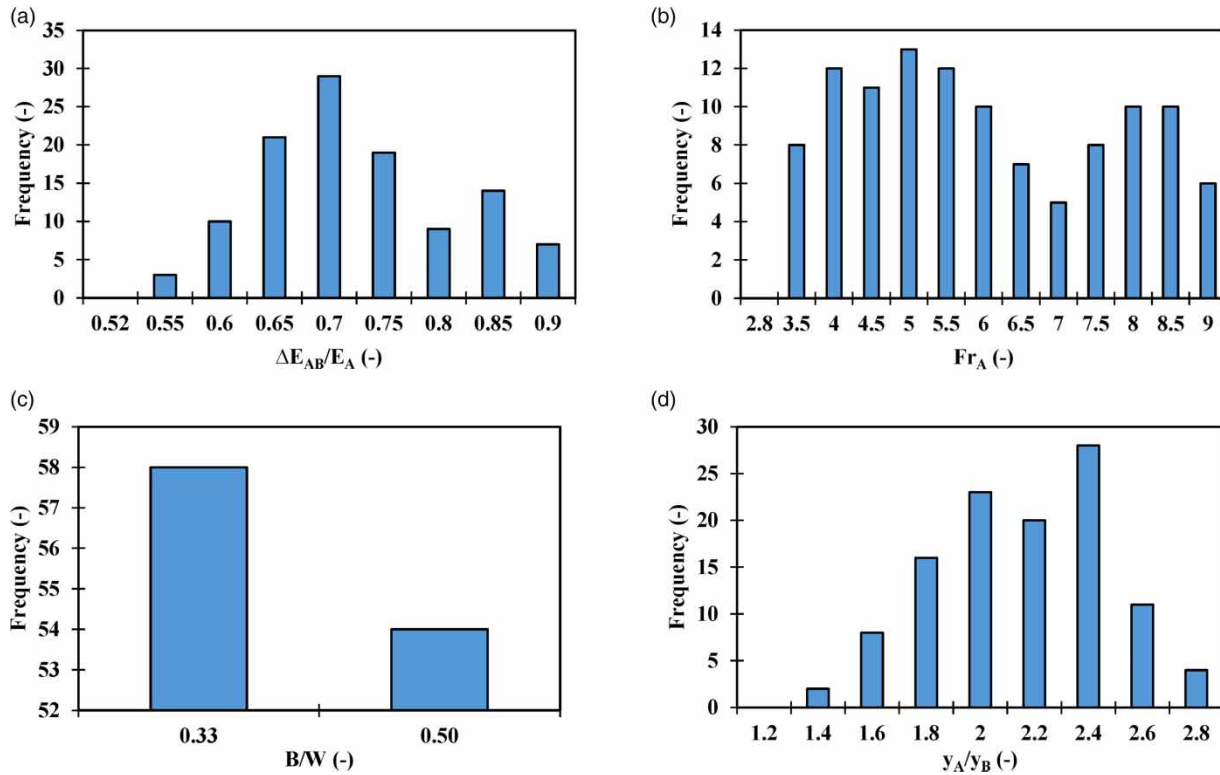
where  $Fr_A$  denotes the Froude number and  $Re_A$  denotes the Reynolds number. In the present research, the flow is turbulent, so the effect of  $Re_A$  is ignored (Nasrabadi *et al.* 2021; Norouzi *et al.* 2023). In addition, some parameters of Equation (3) have certain values and are not part of the research objectives, so they were ignored, too (Rahmanshahi & Shafai Bejestan 2020). White’s theorem provides a useful insight that dimensionless parameters can be obtained through various mathematical operations such as division, multiplication, addition, or subtraction of other dimensionless parameters (White 2016; Daneshfaraz *et al.* 2022b). In the present study, the most significant dimensionless parameters affecting energy dissipation are expressed as follows:

$$\frac{\Delta E_{AB}}{E_A} = f_3\left(Fr_A, \frac{B}{W}, \frac{y_B}{y_A}\right) \tag{5}$$

The ranges of investigated parameters are presented in Table 1. The histogram of parameters  $\Delta E_{AB}/E_A$ ,  $Fr_A$ ,  $B/W$ , and  $y_B/y_A$  is shown in Figure 2.

**Table 1** | The range of parameters of the current research

Parameters	Min.	Max.	Average	Model no.	Parameters
$\Delta E_{AB}/E_A$	0.524	0.890	0.670	1	$Fr_A, B/W$
$Fr_A$	2.854	8.858	5.795	2	$Fr_A, y_B/y_A$
$B/W$	0.333	0.500	0.415	3	$B/W, y_B/y_A$
$y_B/y_A$	1.299	2.750	2.061	4	$Fr_A, B/W, y_B/y_A$



**Figure 2** | The histogram of (a)  $\Delta E_{AB}/E_A$ , (b)  $Fr_A$ , (c)  $B/W$ , and (d)  $y_B/y_A$ .

### Support Vector Machine

SVM is a supervised learning model that is used for classification and prediction (Vapnik 1995). The SVM uses constrained optimization to minimize structural error and ultimately obtain an optimal solution. SVM estimates a function associated with the dependent variable, which in turn depends upon multiple independent input parameters. The relationship between the variables is quantified by an algebraic function with some perturbation (allowable error  $\epsilon$ ) (Norouzi *et al.* 2021).

$$f(x) = W^T \phi(X) + b \tag{6}$$

$$Y = f(x) + noise \tag{7}$$

Here,  $W$  is the coefficient vector,  $T$  is the transpose of  $W$ ,  $b$  is a constant term included in the regression function, and  $\phi$  is the kernel function. The goal is to discover the function  $f(x)$  by training the SVM model with a set of examples (training set). The SVM regression function can be written as:

$$f(x) = \sum_{i=1}^N \bar{a}_i \phi(X_i)^T \phi(X) + b \tag{8}$$

The variable  $a_i$  corresponds to the average Lagrange coefficients. The computation of  $\phi(X)$  can be complex. In the SVM regression model, a kernel function is employed, and its intricacy is contingent on the scale of the training data and the dimensions of the feature vector. Four commonly utilized kernel types in practice are the Linear kernel, Polynomial kernel, Sigmoid kernel, and radial basis function (RBF) kernel. These kernels play a pivotal role in shaping the SVM

regression model and cater to diverse modeling scenarios (Hassanzadeh & Abbaszadeh 2023).

$$K(X_i, X_j) = (X_i, X_j) \quad (9)$$

$$K(X_i, X_j) = (1 + (X_i, X_j))^d \quad (10)$$

$$K(X_i, X_j) = \tanh(-a(X_i, X_j) + C) \quad (11)$$

$$K(X_i, X_j) = \exp\left(-\|X - X_i\|^2 / \sigma^2\right) \quad (12)$$

In the above equations,  $K(X_i, X_j)$  represents the covariance or kernel function, calculated at points  $X_i$  and  $X_j$ . The functions  $a$ ,  $C$ ,  $d$ , and  $\sigma$  denote kernel functions. The term  $d$  represents the degree of the polynomial,  $\sigma$  is the variance and hyper parameter, and  $C$  is a positive integer that acts as a penalty factor when model training errors occur. Here, the values of Capacity, Epsilon and Gamma are 10, 0.1 and 10.

### Artificial Neural Network

An artificial neural network incorporates input layers, hidden neural layers, and output neural layers to process decision making. In the network, the neuron is the foundational structure that modifies an input before sending output information to the later layer. The architecture of the neural connections is mathematical operations that can be nonlinear and the interactions amongst the neural layers results in complex and nonlinear behavior. While each neuron acts individually, the behavior of the network is cohesive. They resemble, in some respects, the neural activity of the human brain but there are differences in their training, behavior, and capacity (Al-Bulushi *et al.* 2012). Here, the values of Min hidden units and Max hidden units are 3 and 21, respectively. The values of Networks to train and Networks to retain were introduced to the Statistica 12 software as 20 and 5, respectively. In Figure 3, the architecture of the artificial neural network model is presented. In the present research, the ANN model with three input neurons, one hidden layer (with 21 neurons), and one output neuron has been employed (Norouzi *et al.* 2020; Ayaz *et al.* 2024).

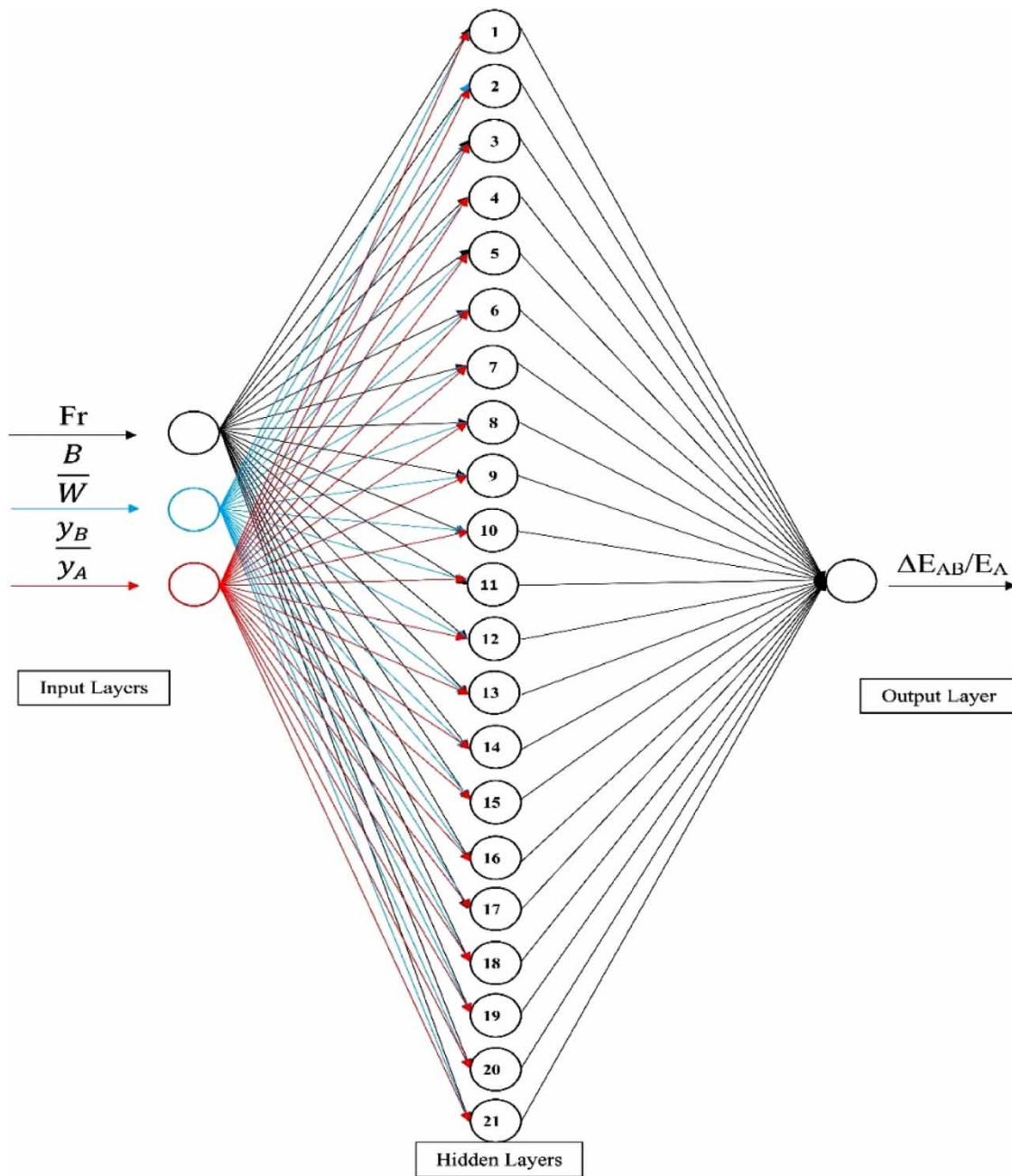
### Random Forest

RF is a supervised learning model that is often done by the bagging method. The bagging method combines learning models to improve their performance. A random forest builds multiple, merged decision trees yield superior predictions (Sun *et al.* 2020). RFs can be used for both classification and regression problems. A random forest, maps input data to outputs in the training or model fitting phase. During training, the model is fed data that is relevant to the problem domain that the model needs to learn to make predictions (Jahed Armaghani *et al.* 2020). The model learns the relationships among the data and the values the user wants to predict. In the RF, number of trees = 500, minimum no. of cases = 5, maximum no. of levels = 10, minimum no. in child node = 5, and max. no. of nodes = 100. The tree graph of the present model is shown in Figure 4.

### Multiple Adaptive Regression Splines

Friedman (1991), initially introduced the MARS model for predicting continuous numerical outputs, and it is a non-parametric local model. The term non-parametric implies that the model structure is unknown before modeling; this model does not use all relevant data at once. Instead, it divides the data into subsets and performs modeling for each of these subsets, referred to as local models. In this model, the capability exists to reveal hidden nonlinear patterns in datasets with a large number of variables. Therefore, by employing this method, it becomes possible to define the estimation function, eliminating the need to combine multiple statistical methods. The foundation of this approach is based on functions called basis functions, where for each explanatory variable, a basis function is defined as Equation (13):

$$g_k(x) = \max(0, x - t_k) \quad (13)$$



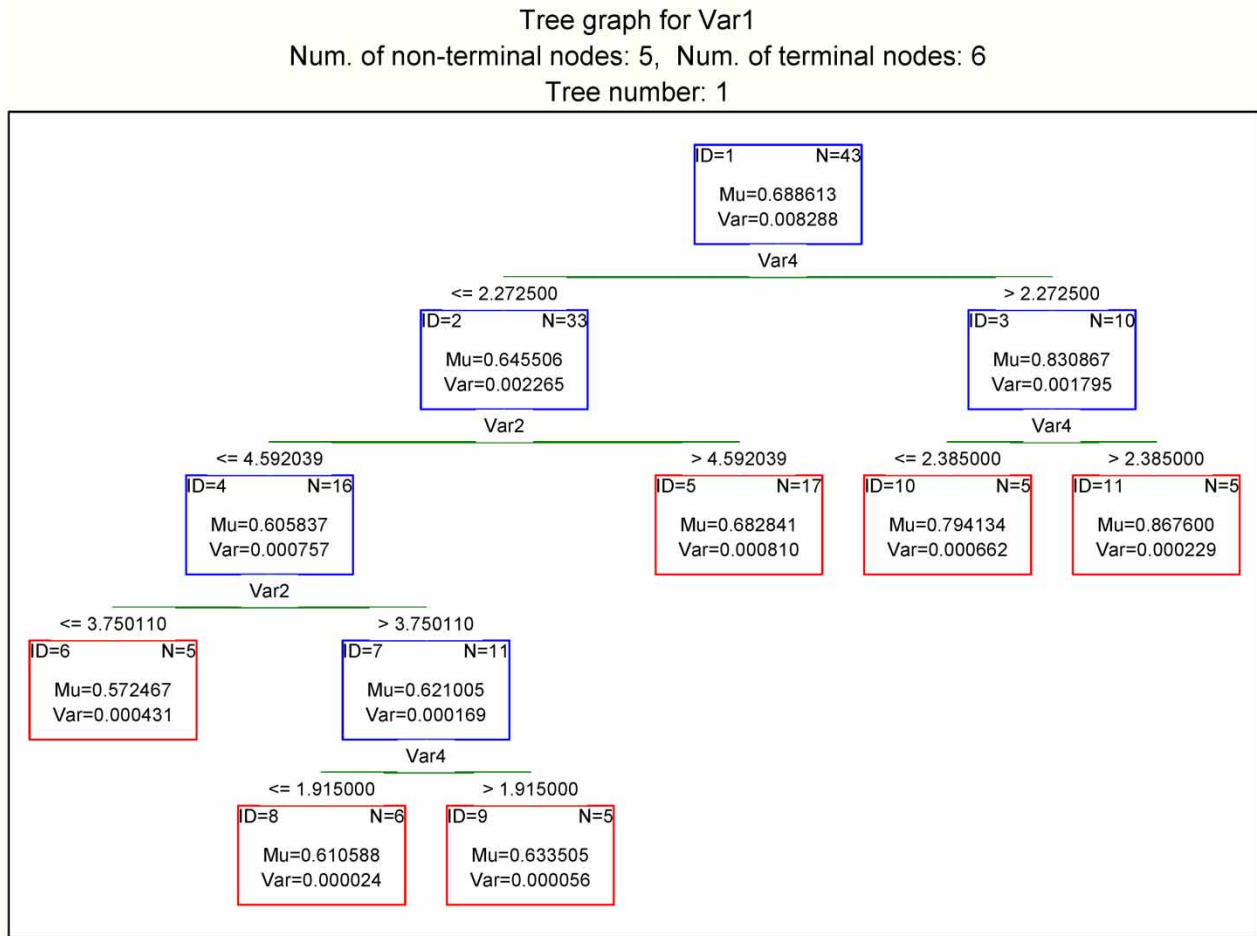
**Figure 3** | ANN model architecture.

Here,  $x$  is the named node, and in practice, it is one of the observations of that explanatory variable. These functions are called spline functions, with  $t$  being the reflected pairs. The general form of the MARS model is defined as follows:

$$Y = \sum_{k=1}^K B_k \times g_k(X) \tag{14}$$

In Equation (14),  $Y$  is the estimated value of the response variable,  $X$  is the vector of explanatory variables,  $B_k$  is the basis function, and  $C_k$  are coefficients determined by minimizing the sum of squared residuals. Each basis function may





**Figure 4** | The tree graph of the model.

take the form of a linear spline function or the product of two or more of them, indicating interaction effects. The Multiple Adaptive Regression Splines model divides the space of explanatory variables into distinct regions with specific nodes, which result in the maximum reduction in the sum of squared errors. The fitting of the MARS model occurs in two stages. In the forward stage, a large number of basis functions with different nodes are successively added to the model, producing a complex and overfit model. In the backward stage or pruning stage, basis functions with less importance and impact on the estimation are removed. For the MARS model, the following settings were defined: Maximum number of Basis Functions: 21, Degree of interaction: 1, and Penalty and Threshold values: 2 and 0.005, respectively.

### M5 algorithm

The M5 algorithm creates binary branches based solely on a single variable. Each node in the algorithm divides its information into two parts based on a condition defined at that node. In the M5 algorithm, the problem space is divided into subdomains, and for each subdomain, a multiple-variable linear regression model is fitted. The algorithm explores possible separations in the multi-variable space and automatically generates models for each of the domains. The standard deviation parameter of the target values is used as the error measurement criterion for branching at each node. Specifically, the feature that leads to the greatest reduction in the standard deviation for each node is chosen as the preferred feature for branching. The reduction in standard deviation, employed as the error function in the M5

algorithm, is defined as Equation (15) (Pal & Deswal 2009):

$$SDR = sd(T) - \sum \left| \frac{T_i}{T} \right| \times sd(T_i) \quad (15)$$

where  $sd$  denotes the standard deviation,  $T$  includes the samples reaching the considered node, and  $T_i$  represents the samples obtained from the division of the considered node based on the selected feature. The M5 algorithm examines all possible scenarios for branching based on a specific feature and ultimately selects a scenario that can reduce the error function more than others. After completing the tree-building algorithm, a multiple-variable linear regression model is fitted to the existing samples in each internal node. In the current research, the M5Rule option of WEKA software was used to model the M5.

### Gene expression programming

This approach is considered part of evolutionary algorithms, all of which are grounded in the principles of Darwinian evolution. These algorithms define an objective function in the form of criteria and then employ a learned function to measure and compare various solution methods (Najafzadeh 2019). In a step-by-step process of refining data structures, they ultimately present a suitable solution method. Gene expression programming is a recent method among these evolutionary algorithms, and due to its sufficient accuracy, it is considered the most conventional and widely used approach. The primary domain of gene expression programming is the same as genetic algorithms, with the distinction that this method uses branches instead of bit strings. Each branch consists of a set of terminals (problem variables) and a set of functions (primary operators) (Mohammed & Sharifi 2020). Table 2 shows the values of the parameters defined for the GEP model in the GenXpro Tools 4 software. The parameters and their rates to estimate the desired parameter if the population is considered up to 10,000 are listed in Table 2.

### Statistical indicators

The following statistical indicators have been used in this study (Saberi-Movahed *et al.* 2020; Daneshfaraz *et al.* 2022b; Agarwal *et al.* 2023; Najafzadeh *et al.* 2023; Nourani *et al.* 2023; Najafzadeh & Mahmoudi-Rad 2024):

$$RE\% = \frac{\Delta E_{AB}/E_{AObs} - \Delta E_{AB}/E_{Acal}}{\Delta E_{AB}/E_{AObs}} \times 100 \quad (16)$$

**Table 2** | Values of parameters used in the GEP model

Parameters	Value
Head size	7
Chromosomes	30
Number of genes	3
Mutation rate	0.044
Inversion rate	0.1
One point recombination rate	0.3
Two point recombination rate	0.3
Gene recombination rate	0.1
IS transposition rate	0.1
RIS transposition rate	0.1
Gene transposition rate	0.1
Fitness function error rate	RMSE
Linking function	+
Generation number	10,000

Here, RE indicates the relative error.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\Delta E_{AB}/E_{A Obs} - \Delta E_{AB}/E_{A Cal})^2}{n}} \tag{17}$$

Here, RMSE indicates the root mean square error;  $n$  indicates the total data.

$$BIAS = \frac{1}{n} \sum_{i=1}^n (\Delta E_{AB}/E_{A Cal} - \Delta E_{AB}/E_{A Obs})_i \tag{18}$$

$$DR = \frac{\sum_{i=1}^n \left( \frac{\Delta E_{AB}/E_{A Cal}}{\Delta E_{AB}/E_{A Obs}} \right)_i}{n} \tag{19}$$

$$SI = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n ((\Delta E_{AB}/E_{A Cal i} - \overline{\Delta E_{AB}/E_{A Cal}}) - (\Delta E_{AB}/E_{A Obs i} - \overline{\Delta E_{AB}/E_{A Obs}}))^2}}{\frac{1}{n} \sum_{i=1}^n \Delta E_{AB}/E_{A Obs i}} \tag{20}$$

where SI indicates the scatter index.

$$KGE = 1 - \sqrt{(R - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2}$$

$$\beta = \frac{\overline{\Delta E_{AB}/E_{A Cal}}}{\overline{\Delta E_{AB}/E_{A Obs}}}, \gamma = \frac{CV_{Cal}}{CV_{Obs}} = \frac{\sigma_{Cal}/\overline{\Delta E_{AB}/E_{A Cal}}}{\sigma_{Obs}/\overline{\Delta E_{AB}/E_{A Obs}}} \tag{21}$$

$$R = \frac{\sum_{i=1}^n (\Delta E_{AB}/E_{A Obs i} - \overline{\Delta E_{AB}/E_{A Obs}})(\Delta E_{AB}/E_{A Cal i} - \overline{\Delta E_{AB}/E_{A Cal}})}{\sqrt{\sum_{i=1}^n (\Delta E_{AB}/E_{A Obs i} - \overline{\Delta E_{AB}/E_{A Obs}})^2 \sum_{i=1}^n (\Delta E_{AB}/E_{A Cal i} - \overline{\Delta E_{AB}/E_{A Cal}})^2}}$$

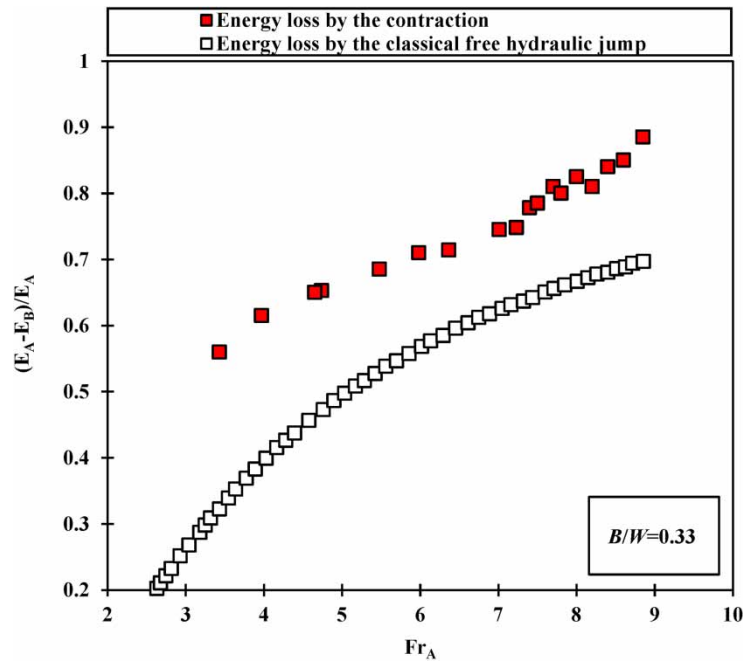


Figure 5 | The changes in the energy loss.

**Table 3** | The experimental results of the present study

No.	$\Delta E_{AB}/E_A$	$Fr_A$	$B/W$	$y_B/y_A$	no.	$\Delta E_{AB}/E_A$	$Fr_A$	$B/W$	$y_B/y_A$
1	0.559	3.430	0.330	1.642	57	0.830	8.686	0.330	2.540
2	0.571	3.514	0.330	1.681	58	0.841	8.858	0.330	2.580
3	0.575	3.651	0.330	1.722	59	0.524	2.889	0.500	1.298
4	0.585	3.714	0.330	1.755	60	0.531	3.054	0.500	1.380
5	0.595	3.852	0.330	1.781	61	0.541	3.145	0.500	1.440
6	0.601	3.942	0.330	1.825	62	0.551	3.265	0.500	1.460
7	0.608	3.958	0.330	1.855	63	0.562	3.314	0.500	1.470
8	0.614	3.968	0.330	1.903	64	0.578	3.385	0.500	1.475
9	0.618	4.452	0.330	1.911	65	0.582	3.407	0.500	1.482
10	0.624	4.158	0.330	1.924	66	0.591	3.565	0.500	1.500
11	0.625	4.142	0.330	1.932	67	0.601	3.612	0.500	1.550
12	0.631	4.280	0.330	1.941	68	0.605	3.785	0.500	1.600
13	0.635	4.369	0.330	1.950	69	0.612	3.854	0.500	1.620
14	0.642	4.457	0.330	1.967	70	0.628	3.985	0.500	1.650
15	0.641	4.551	0.330	1.971	71	0.625	4.025	0.500	1.655
16	0.645	4.624	0.330	1.983	72	0.629	4.112	0.500	1.672
17	0.652	4.765	0.330	1.991	73	0.634	4.265	0.500	1.690
18	0.658	4.737	0.330	2.044	74	0.639	4.245	0.500	1.717
19	0.645	4.865	0.330	2.051	75	0.645	4.614	0.500	1.720
20	0.651	4.952	0.330	2.067	76	0.650	4.453	0.500	1.750
21	0.655	5.015	0.330	2.071	77	0.655	4.548	0.500	1.770
22	0.658	5.114	0.330	2.082	78	0.658	4.632	0.500	1.780
23	0.658	5.254	0.330	2.085	79	0.662	4.785	0.500	1.790
24	0.665	5.345	0.330	2.089	80	0.668	4.856	0.500	1.814
25	0.671	5.425	0.330	2.100	81	0.665	4.844	0.500	1.821
26	0.672	5.458	0.330	2.114	82	0.701	4.915	0.500	1.840
27	0.672	5.476	0.330	2.157	83	0.675	5.021	0.500	1.850
28	0.665	5.585	0.330	2.165	84	0.682	5.145	0.500	1.920
29	0.671	5.625	0.330	2.171	85	0.685	5.225	0.500	1.928
30	0.675	5.745	0.330	2.182	86	0.691	5.336	0.500	1.940
31	0.683	5.836	0.330	2.197	87	0.695	5.445	0.500	1.980
32	0.685	5.914	0.330	2.200	88	0.701	5.585	0.500	1.985
33	0.689	5.981	0.330	2.218	89	0.705	5.685	0.500	1.998
34	0.692	6.415	0.330	2.214	90	0.711	5.758	0.500	2.000
35	0.698	6.245	0.330	2.225	91	0.714	5.811	0.500	2.069
36	0.704	6.365	0.330	2.234	92	0.720	6.150	0.500	2.080
37	0.708	6.565	0.330	2.244	93	0.734	6.180	0.500	2.140
38	0.714	6.776	0.330	2.252	94	0.740	6.254	0.500	2.145
39	0.718	6.834	0.330	2.268	95	0.750	6.361	0.500	2.170
40	0.725	6.956	0.330	2.278	96	0.765	6.895	0.500	2.216
41	0.722	7.010	0.330	2.249	97	0.780	7.219	0.500	2.256
42	0.728	7.156	0.330	2.278	98	0.801	7.385	0.500	2.280

*(Continued.)*

**Table 3** | Continued

No.	$\Delta E_{AB}/E_A$	$Fr_A$	$B/W$	$y_B/y_A$	no.	$\Delta E_{AB}/E_A$	$Fr_A$	$B/W$	$y_B/y_A$
43	0.731	7.230	0.330	2.287	99	0.811	7.454	0.500	2.285
44	0.740	7.365	0.330	2.287	100	0.820	7.558	0.500	2.310
45	0.745	7.454	0.330	2.294	101	0.825	7.632	0.500	2.350
46	0.755	7.587	0.330	2.301	102	0.830	7.758	0.500	2.380
47	0.765	7.654	0.330	2.306	103	0.838	7.865	0.500	2.420
48	0.758	7.712	0.330	2.314	104	0.845	7.948	0.500	2.450
49	0.765	7.836	0.330	2.325	105	0.850	8.025	0.500	2.480
50	0.770	7.958	0.330	2.325	106	0.855	8.114	0.500	2.500
51	0.771	8.025	0.330	2.345	107	0.865	8.225	0.500	2.550
52	0.791	8.145	0.330	2.355	108	0.870	8.365	0.500	2.580
53	0.801	8.226	0.330	2.355	109	0.875	8.445	0.500	2.620
54	0.815	8.354	0.330	2.458	110	0.880	8.526	0.500	2.650
55	0.822	8.425	0.330	2.488	111	0.885	8.652	0.500	2.680
56	0.825	8.523	0.330	2.500	112	0.890	8.858	0.500	2.750

Here, KGE indicates the Kling Gupta Efficiency;  $R$  indicates the correlation coefficient;  $\beta$  indicates the average calculated data relative to the average observed data;  $\gamma$  indicates the standard deviation of the calculated data relative to the standard deviation of the observed data.

If the resulting calculations yield  $0.7 < KGE < 1$ , then the performance is characterized as ‘very good’. If  $0.6 < KGE < 0.7$ , the results are ‘good’. For  $0.4 < KGE < 0.5$  or  $0.5 < KGE < 0.6$ ; respective descriptors ‘acceptable’ and ‘satisfactory’ are used. If, however,  $KGE < 0.4$ , the results are ‘unsatisfactory’ (Gupta *et al.* 2009). If  $DR = 1$ , the soft computing technique shows the most efficient performance,  $DR > 1$  shows over predictions, and  $DR < 1$ , under prediction (Najafzadeh *et al.* 2022). If the BIAS index equal 0 shows the most efficient performance, when  $BIAS > 0$  indicates over prediction and  $BIAS < 0$  indicates under prediction.

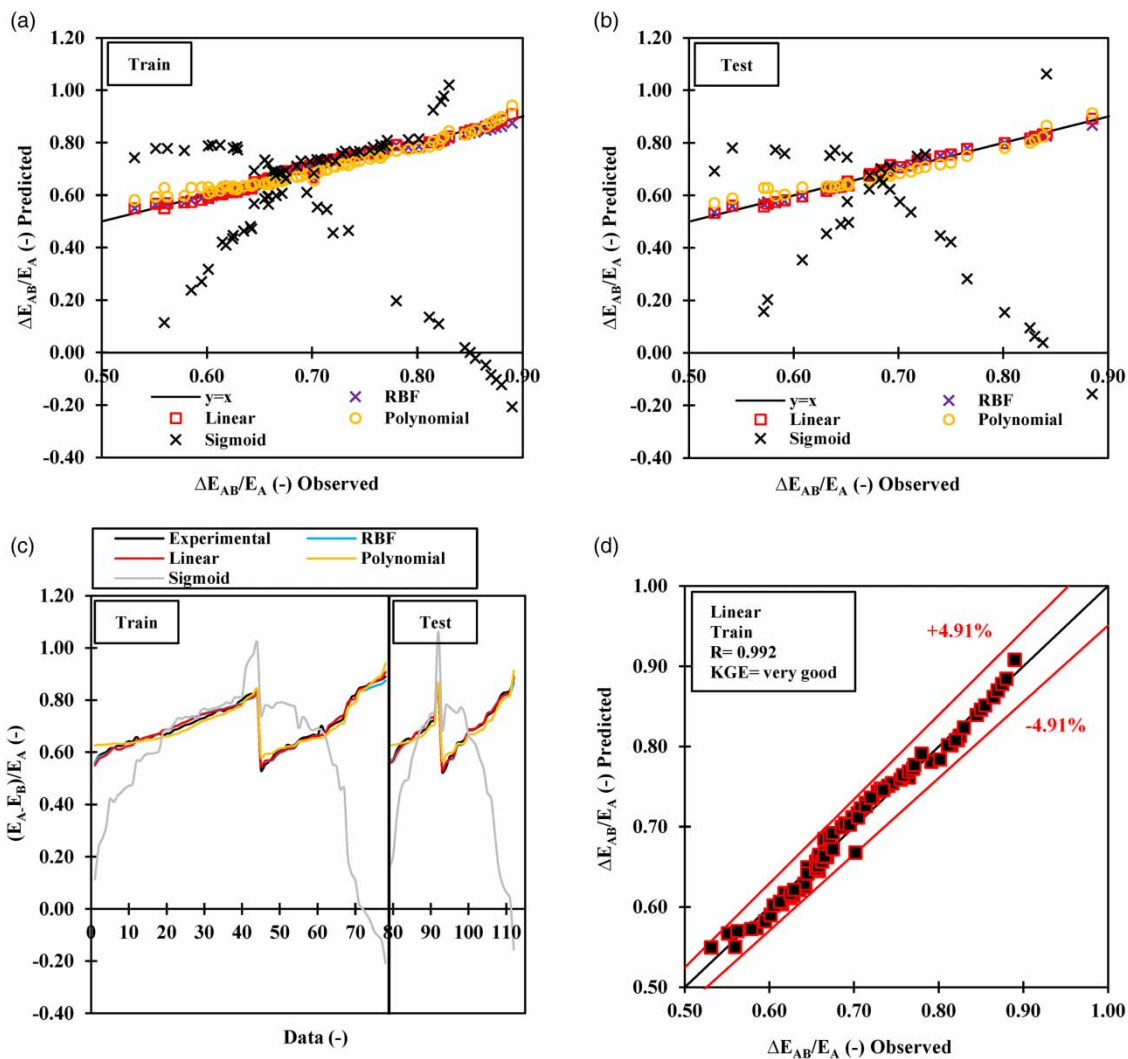
## RESULTS AND DISCUSSION

The increase of the Froude number within each constriction amplifies the energy loss. This phenomenon stems from the diminished flow depth post-gate, inducing heightened flow velocity, thus triggering a hydraulic jump accompanied by increased turbulence and air entrainment. Consequently, this elevates the tailwater depth. The impact of backwater further accentuates energy dissipation by increasing the depth in the constriction region, leading to greater energy loss. Comparatively, the energy loss attributed to constrictions exceeds that of classical free hydraulic jumps, elucidated by turbulent flows preceding the constriction (Figure 5). Unlike free hydraulic jumps, energy loss in arc-shaped constrictions arises from the hydraulic jump, local loss, and the constriction elements. Table 3 encapsulates the experimental findings of this research.

Various dimensionless parameters were considered as inputs for different models, and the relative energy loss was chosen as the output and the target feature. An attempt was made to apply advanced data mining methods to estimate the relative energy loss. In the application of data mining methods to forecast relative energy loss, a partitioning strategy was employed. Seventy percent of the available data was allocated to the training phase, facilitating model development, while the remaining 30% was reserved for the testing phase to assess predictive performance. This division ensures a robust evaluation of the model’s generalization capability on unseen data. According to Table 4, based on model no. 4, among the Linear, Polynomial, RBF, and Sigmoid kernels, the linear kernel was chosen as the superior kernel for the Support Vector Machine model based on its favorable statistical indicator results. The results of Linear and RBF kernels are close to each other. This matter can be either directly chosen or determined through modeling based on the best

**Table 4** | Statistical indicators for the SVM model

Statistical indicator	Kernels			
	Linear	Polynomial	RBF	Sigmoid
R (test)	0.994	0.966	0.992	0.436
KGE (test)	0.980	0.887	0.985	-1.514
RMSE (test)	0.010	0.024	0.014	0.362
Average RE% (test)	1.334	3.047	1.442	34.91
BIAS	-0.0002	-0.0022	-0.0012	-0.168
DR	0.999	0.999	0.998	0.787
SI	0.015	0.035	0.017	0.525



**Figure 6** | Experimental and predicted relative energy loss in the SVM model. (continued).

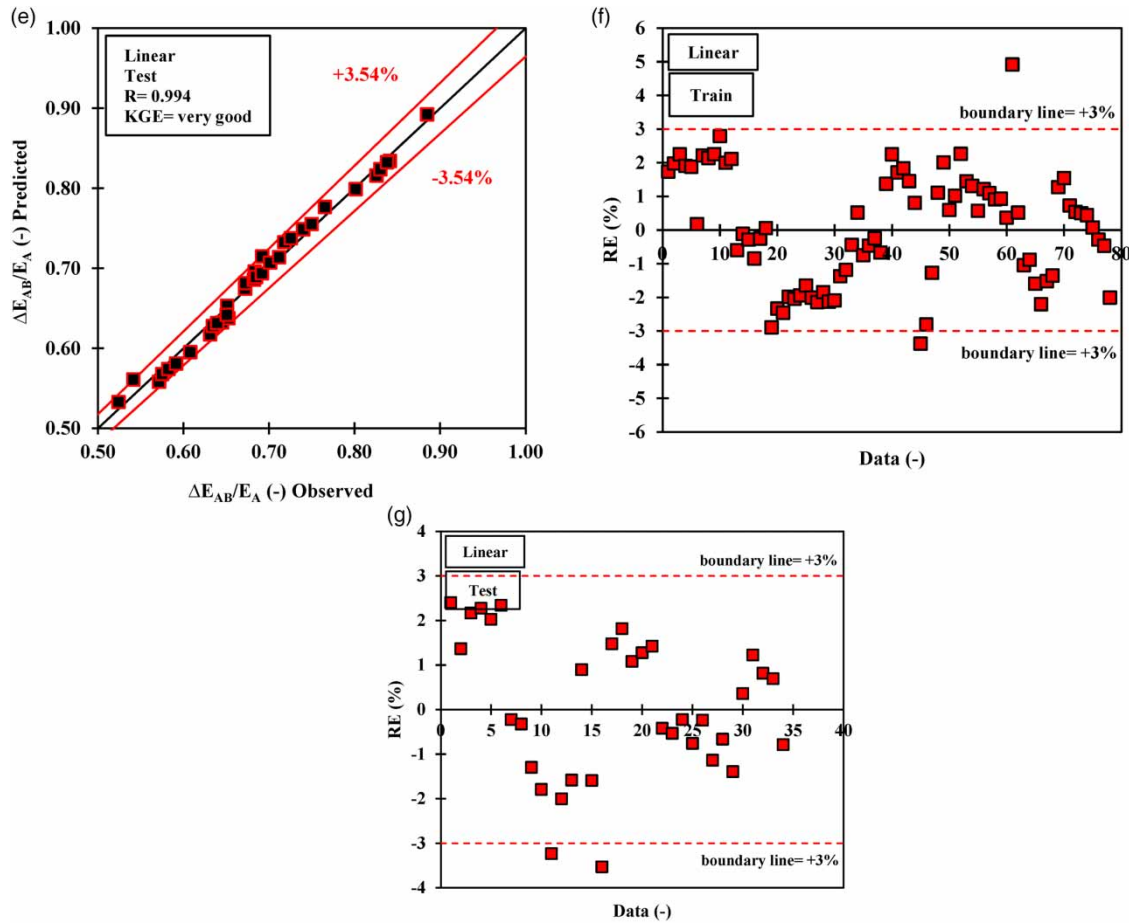


Figure 6 | Continued.

results (Rezaee *et al.* 2023). The selection of the best kernel was made by ensuring that statistical indicators such as R, RMSE, average RE, KGE, BIAS, DR, and SI had satisfactory performance compared to experimental results. The laboratory and predicted values of relative energy loss for various kernels are shown in Figure 6(a) and 6(b). In addition, Figure 6(c) shows the results of the training and testing phases for different datasets. As observed, the linear kernel exhibits high accuracy compared to other kernels and predicts relative energy loss with high precision. As indicated in Figure 6(d) and 6(e), the statistical outcomes for the Linear kernel during the training phase are  $R = 0.992$ ,  $RMSE = 0.0114$ , average  $RE\% = 1.41$ ,  $BIAS = -0.0005$ ,  $DR = 0.999$ ,  $SI = 0.0161$  and  $KGE = 0.982$ . Correspondingly, in the testing phase, these values are  $0.994$ ,  $0.0101$ ,  $1.33\%$ ,  $-0.0002$ ,  $0.999$ ,  $0.015$ , and  $0.980$ , respectively. Figure 6(f) and 6(g) illustrates that, for the superior kernel, a substantial portion of data in both training and testing phases falls within the  $\pm 3\%$  relative error band. This observation underscores high solution accuracy, with over 97% of the data residing within the  $\pm 3\%$  error band during both training and testing phases.

The results of SVM, ANN, RF, MARS, GEP, and M5 models to predict the relative energy loss are presented in Table 5. According to Table 5, it was observed that model no. 4 with three input parameters provides favorable statistical results compared to other models and was selected as the superior model in the processing. Model no. 3 does not have sufficient accuracy to predict the relative energy loss. Also, the comparison of two models no. 1 and 2 shows that replacing the use of parameter  $Fr_A$  significantly improves the accuracy of modeling, which indicates the high impact of parameter  $Fr_A$  in predicting relative energy loss.

As shown in Figure 7, the solution accuracy in the MLP network type increased compared to the RBF network type. The statistical results for R, RMSE, average RE%, BIAS, DR, SI, and KGE for the MLP network type in the training phase are

**Table 5** | The results of statistical indicators, obtained from SVM-Linear, ANN-MLP, RF, GEP, MARS and M5 for different combinations

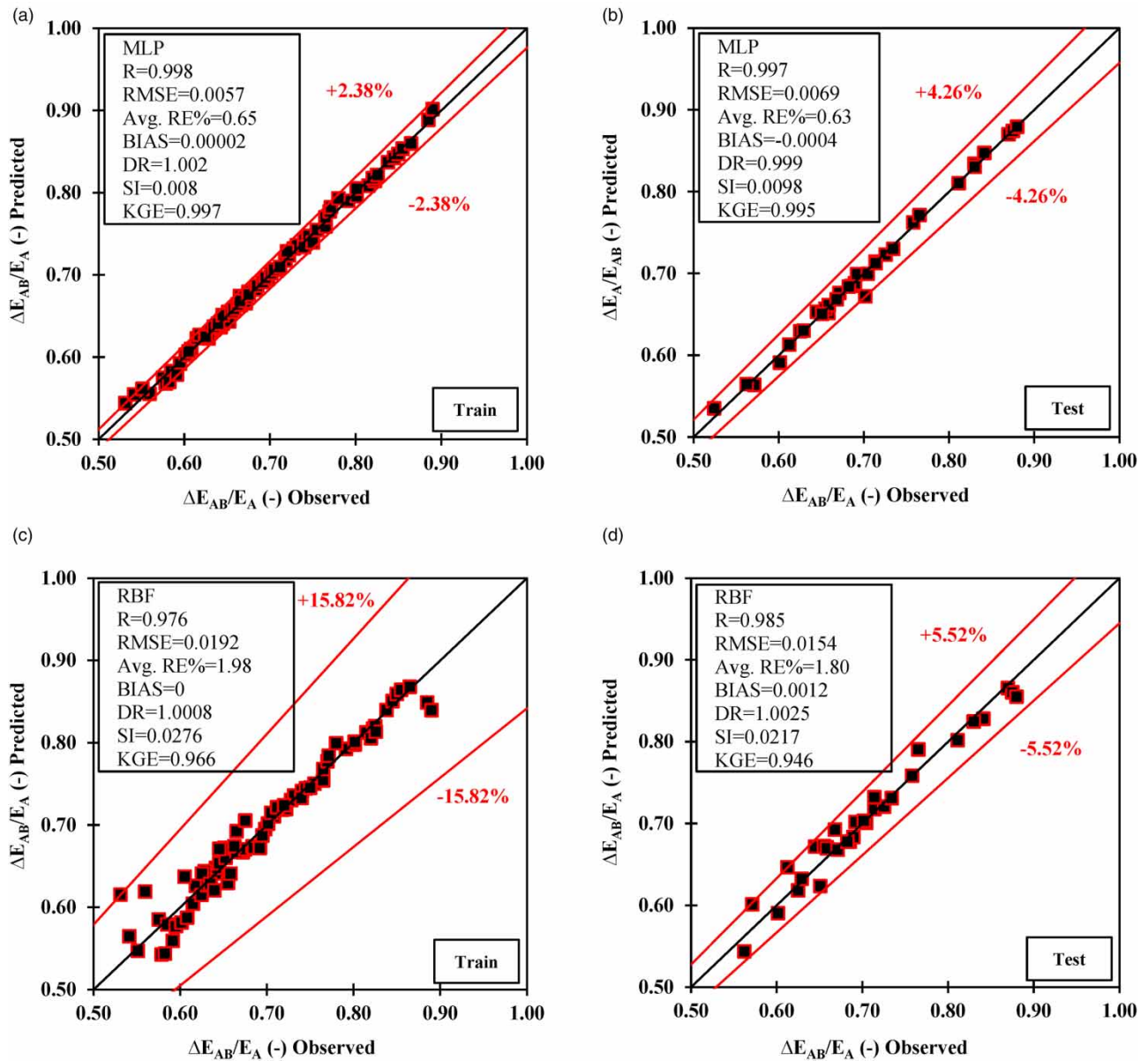
Model no.	method	Train				Test			
		R (-)	Avg. RE (%)	RMSE (-)	KGE (-)	R (-)	Avg. RE (%)	RMSE (-)	KGE (-)
1	SVM	0.795	3.54	0.0982	0.794	0.814	3.52	0.0980	0.815
	ANN	0.802	2.94	0.0854	0.800	0.818	2.57	0.0842	0.824
	RF	0.741	4.14	0.0992	0.755	0.752	4.05	0.0991	0.755
	GEP	0.800	2.98	0.0868	0.810	0.810	2.58	0.0867	0.807
	M5	0.748	4.01	0.0972	0.757	0.757	4.00	0.0945	0.750
	MARS	0.814	2.45	0.0792	0.814	0.825	2.24	0.0795	0.821
2	SVM	0.868	2.55	0.0765	0.868	0.870	2.82	0.0700	0.872
	ANN	0.894	1.48	0.0545	0.884	0.892	1.54	0.0514	0.892
	RF	0.825	2.92	0.0923	0.825	0.826	2.85	0.0852	0.814
	GEP	0.884	1.46	0.0548	0.868	0.892	1.84	0.0714	0.884
	M5	0.840	2.85	0.0892	0.845	0.825	2.70	0.0801	0.825
	MARS	0.892	1.35	0.0546	0.887	0.892	1.58	0.0500	0.898
3	SVM	0.558	4.95	0.1455	0.560	0.545	5.25	0.1580	0.545
	ANN	0.585	5.14	0.1865	0.465	0.582	4.88	0.1784	0.485
	RF	0.485	5.85	0.2550	0.465	0.454	5.92	0.2655	0.445
	GEP	0.575	5.25	0.1885	0.500	0.545	4.84	0.1854	0.484
	M5	0.495	5.72	0.2684	0.471	0.458	5.87	0.2595	0.450
	MARS	0.584	4.80	0.1718	0.484	0.584	4.84	0.1725	0.588
4	SVM	0.992	1.41	0.0114	0.982	0.994	1.33	0.0101	0.980
	ANN	0.998	0.65	0.0057	0.997	0.997	0.63	0.0069	0.995
	RF	0.966	2.58	0.0242	0.867	0.964	3.32	0.0270	0.859
	GEP	0.992	1.28	0.0117	0.984	0.996	0.80	0.0088	0.973
	M5	0.971	2.30	0.0212	0.968	0.978	2.25	0.0190	0.926
	MARS	0.997	0.71	0.0062	0.996	0.997	0.70	0.0074	0.995

0.998, 0.0057, 0.65%, 0.00002, 1.002, 0.008 and 0.997, respectively. These values for the testing phase are 0.997, 0.0069, 0.63%, -0.0004, 0.999, 0.0098, and 0.995, respectively. In the RBF network type, the statistical results in the training phase are 0.976, 0.0192, 1.98%, 0, 1.0008, 0.0276 and 0.966, respectively. For the testing phase, these values are 0.985, 0.0154, 1.80%, 0.0012, 1.0025, 0.0217, and 0.946, respectively. In the MLP and RBF networks, the data are within the ±4.26% and ±5.52% relative error bands in the testing phase. Therefore, based on the above results, the ANN-MLP method was recognized as the superior model in this phase. This phenomenon has been observed in various research studies, such as [Sauda \(2022\)](#) and [Momeneh & Nourani \(2023\)](#), where the ANN model demonstrates a significant level of accuracy in predicting various parameters.

In [Figure 8\(a\)](#) and [8\(c\)](#), the scatter plots of data in the training and testing phases for the RF model are provided. The statistical outcomes for R, RMSE, average RE%, BIAS, DR, SI, and KGE in the training phase are 0.966, 0.0242, 2.58%, -0.0002, 1.0023, 0.0347 and 0.867, respectively. Similarly, for the testing phase, these values are 0.964, 0.0270, 3.32%, 0.0070, 1.0129, 0.0385 and 0.859, indicating the model’s performance in both training and testing phases. In [Figure 8\(b\)](#) and [8\(d\)](#), the overlap between experimental and predicted data for different datasets in the training and testing phases is shown, indicating a noticeable difference between them.

In [Figure 9](#), the scatter plot illustrates the training and testing stages for all diverse cases in the GEP and M5 models. The statistical indices R, RMSE, average RE%, BIAS, DR, SI, and KGE have values of 0.992, 0.0117, 1.28%, -0.0027, 0.966, 0.0168 and 0.984, respectively, during the training phase for GEP model. Similarly, for the testing phase, these indices are 0.996, 0.0088, 0.80%, -0.0004, 0.999, 0.0125 and 0.973, respectively. The GEP model has also yielded favorable results predicting energy loss in hydraulic jumps, as evidenced by the study conducted by [Rahmanshahi & Shafai Bejestan \(2020\)](#). Although their research focused on rough bed conditions, the nature of the investigation, which involves the examination of energy loss, aligns well with the results of the current study. The GEP model,

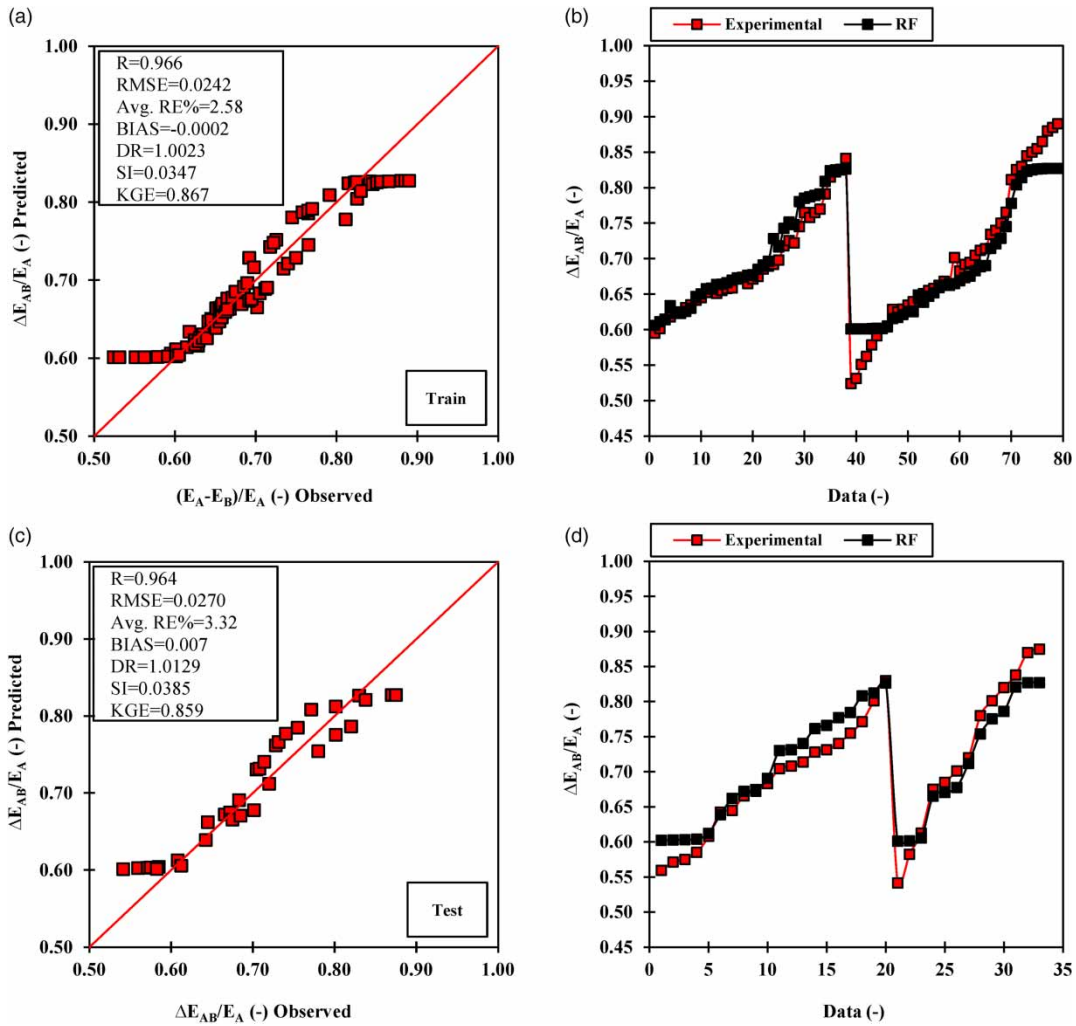




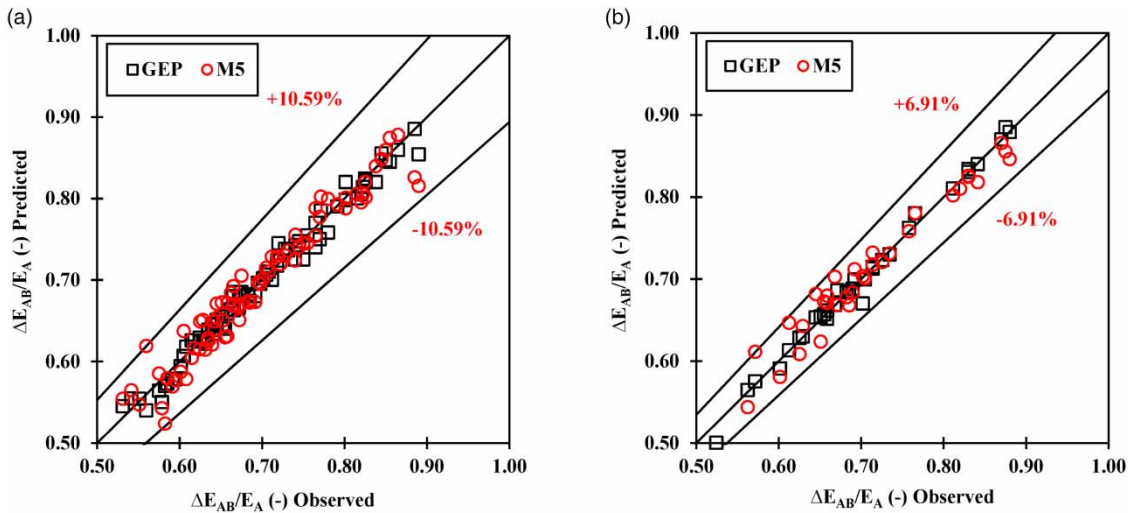
**Figure 7** | Experimental relative energy loss versus predicted relative energy loss in the training and test phase (a, b) in the MLP network, (c, d) in the RBF network in the ANN method.

when compared to the M5 model, exhibits greater precision and closer alignment with experimental results. Specifically, statistical indicators for the M5 model during the training phase ( $R = 0.971$ ,  $RMSE = 0.0212$ , average  $RE\% = 2.30$ ,  $BIAS = -0.0016$ ,  $DR = 0.998$ ,  $SI = 0.0304$ ,  $KGE = 0.968$ ) are noteworthy. These results persist in the testing phase as well, with values of ( $R = 0.978$ ,  $RMSE = 0.0190$ , average  $RE\% = 2.25$ ,  $BIAS = 0.0007$ ,  $DR = 1.0023$ ,  $KGE = 0.926$ ).

The regression relationship for the MARS model was derived from 70% of the data and was validated with the remaining 30%. The regression Equation (22), presented in Table 6, comprises 6 basis functions. Examination of the results indicates that all dimensionless parameters  $Fr_A$ ,  $B/W$ , and  $y_B/y_A$  play a crucial role in predicting the relative energy loss. Statistical indices during the training phase ( $R = 0.997$ ,  $RMSE = 0.0062$ , average  $RE\% = 0.71$ ,  $BIAS = 0$ ,  $DR = 1$ ,  $SI = 0.0090$ ,  $KGE = 0.996$ ) and testing phase ( $R = 0.997$ ,  $RMSE = 0.0074$ , average  $RE\% = 0.70$ ,  $BIAS = -0.0005$ ,  $DR = 0.999$ ,  $SI = 0.0106$ ,  $KGE = 0.995$ ) demonstrate high performance (Figure 10).



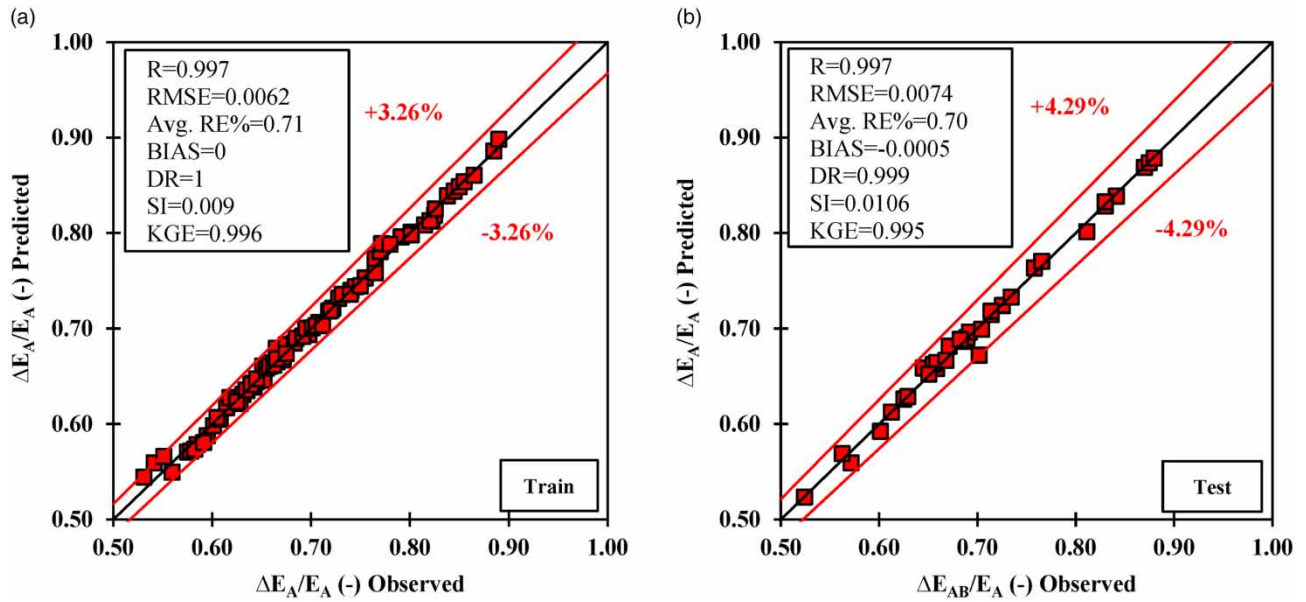
**Figure 8** | Comparison of the experimental relative energy loss and the RF model in phases (a, b) training, (c and d) testing.



**Figure 9** | Comparison of the experimental and predicted relative energy loss of the GEP and M5 models in phases (a) training, (b) testing.

**Table 6** | The equation of the MARS model

Equation (22) of MARS and Basis Function (BF)	Coefficient	Value	Coefficient	Value
$\Delta E_{AB}/E_A = a + b \times BF(1) - c \times BF(2) + d \times BF(3) - e \times BF(4) + f \times BF(5) - g \times BF(6)$	a	$7.80 \times 10^{-1}$	h	$7.45 \times 10^0$
$BF(1) = \max(0, Fr_A - h)$	b	$4.98 \times 10^{-2}$	i	$7.45 \times 10^0$
$BF(2) = \max(0, i - Fr_A)$	c	$1.90 \times 10^{-2}$	j	$3.30 \times 10^{-1}$
$BF(3) = \max(0, B/W - j)$	d	$3.51 \times 10^{-1}$	k	$2.35 \times 10^0$
$BF(4) = \max(0, k - y_B/y_A)$	e	$2.19 \times 10^{-1}$	l	$6.15 \times 10^0$
$BF(5) = \max(0, Fr_A - l)$	f	$3.36 \times 10^{-2}$	m	$4.55 \times 10^0$
$BF(6) = \max(0, Fr_A - m)$	g	$2.37 \times 10^{-2}$	-	-



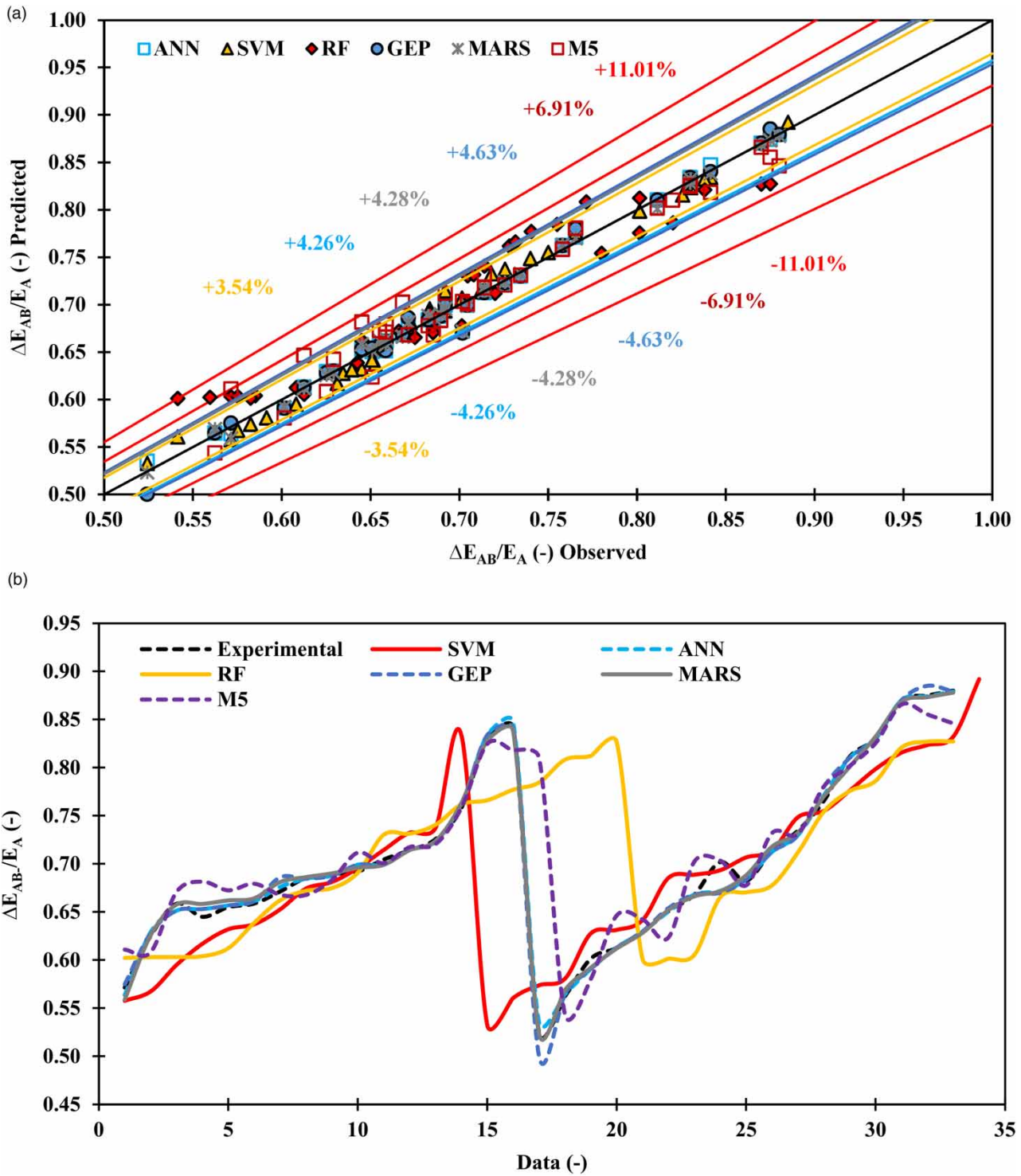
**Figure 10** | Comparison of the experimental relative energy loss and the MARS model in (a) training and (b) testing phases.

In the current study, a nonlinear polynomial regression relationship has been proposed using SPSS for predicting the relative energy loss within the specified research range:

$$\frac{\Delta E_{AB}}{E_A} = 0.002(Fr_A)^{2.098} + 0.290\left(\frac{B}{W}\right)^{0.969} + 0.320\left(\frac{y_B}{y_A}\right)^{0.640} \tag{23}$$

The statistical results for the mentioned relationship are as follows:  $R = 0.988$ , average  $RE\% = 1.108\%$ , and  $RMSE = 0.0098$ . Additionally, the  $KGE$  for this relationship falls within the ‘very good’ range. These statistical indicators suggest that the provided relationship exhibits high accuracy in predicting the amount of energy loss, with over 99% of the data falling within the relative error range of  $\pm 3\%$ . It should be noted that the relationship provided by MARS model is more accurate compared to Equation (23).

In order to select the optimal model from SVM, ANN, RF, GEP, M5, and MARS models, the results of the top performers in each group are depicted in Figure 11. According to Figure 11(a), it can be observed that for the RF model, the values fall within the range of relative error of  $\pm 11\%$ . The corresponding  $RMSE$  and average  $RE\%$  values for this model are 0.0270 and



**Figure 11** | (a) Experimental relative energy loss values versus predicted values and (b) comparison of relative energy loss values for different data in the test phase.

3.32%, respectively. For the SVM-Linear model, the data falls within the relative error range of  $\pm 3.54\%$ , showing favorable results compared to the RF model. The RMSE and average RE% values for this model are 0.0101 and 1.33%, respectively. In the case of the GEP model, the data falls within the relative error range of  $\pm 4.63\%$  with RMSE = 0.008 and average

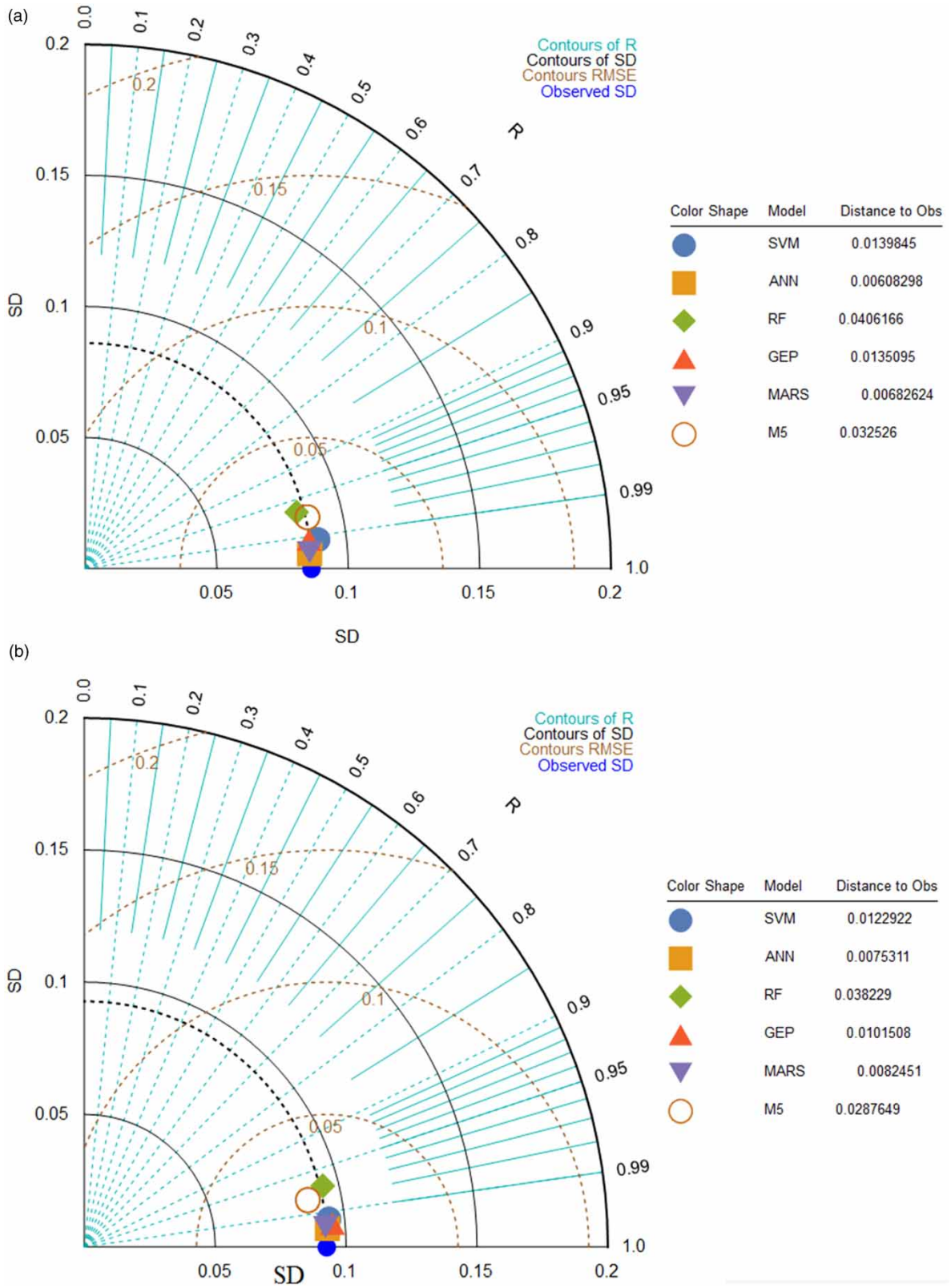


Figure 12 | Taylor and Violin diagrams: (a, c) training and (b, d) testing. (continued.).

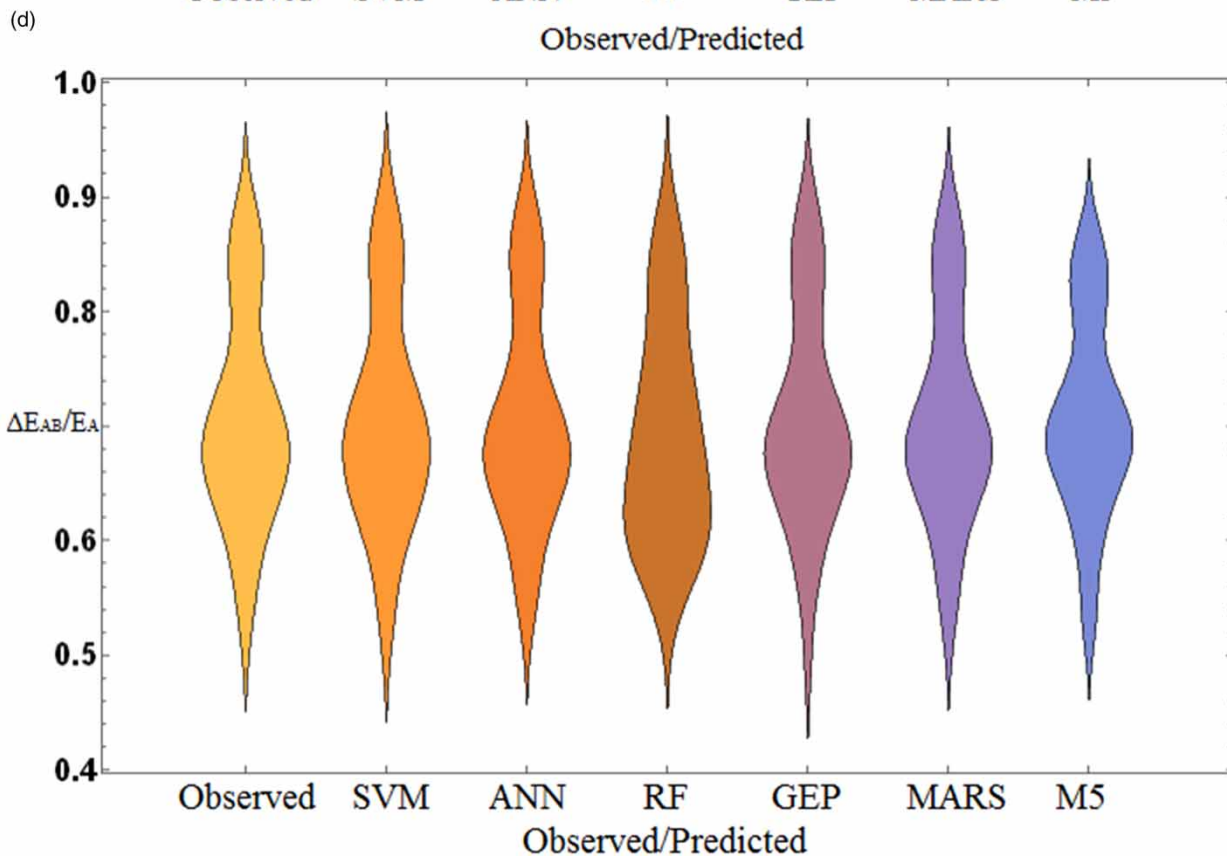
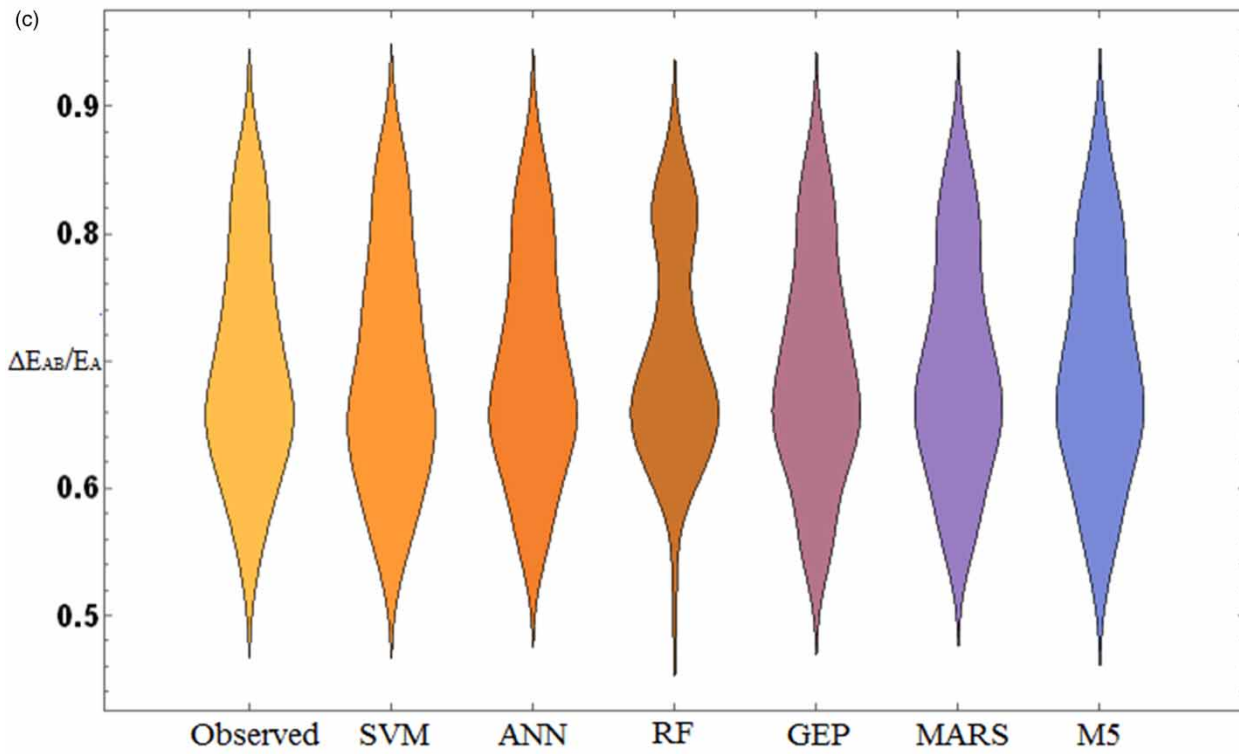


Figure 12 | Continued.

RE% = 0.80%. Similarly, these values for the M5 model are 6.91%, 0.0190, and 2.25%, respectively. The MARS model exhibits satisfactory results, although the ANN-MLP method statistically outperforms the previous models and is closer to experimental results. For the ANN-MLP method, the data falls within the relative error range of  $\pm 4.26\%$ . The corresponding values for this model are 0.0063 and 0.63%. A comparison of the relative energy decline obtained from various models with experimental results indicates better data overlap in the ANN method with experimental results (Figure 11(b)).

As depicted in Figure 12, Taylor diagrams were employed for the decomposition, analysis, and evaluation of models. A notable advantage of the Taylor diagram is its utilization of two statistical indices: the correlation coefficient and the standard deviation (Taylor 2001). A closer proximity of the predicted values to the observed values in terms of correlation coefficient and standard deviation implies a more accurate prediction. The performance of Taylor diagram in Figure 12 illustrates that the ANN model exhibits the highest efficiency and performance. This is evidenced by the predicted standard deviation having a small difference from the observed standard deviation and a high correlation coefficient. According to all evaluation criteria, the examined models demonstrate suitable performance in estimating relative energy loss. Among them, the ANN and MARS models exhibit highest accuracy. The violin plot visually represents numerical data. Occasionally, relying solely on the mean and median proves insufficient for a comprehensive understanding and interpretation of a dataset. Questions may arise regarding whether the majority of sample values cluster around the median, or if most values are concentrated near the maximum and minimum, with no data encompassing the mean. In such instances, a distribution plot is valuable for providing insights. The violin plot amalgamates features of both a box plot and a density plot, effectively highlighting peaks and distributions within the data. In scenarios where samples exhibit multiple peaks, the violin plot adeptly delineates the presence of these peaks, elucidating their coordinates and relative fluctuations.

## CONCLUSIONS

The current research investigates experimental and data mining methods, including SVM, ANN, RF, GEP, MARS, M5 algorithm and regression equation for predicting the relative energy loss in constrictions along the flow path. Experiments were conducted within the Froude number range of 2.85–8.85. 70% of the data were used for the training phase and 30% for the testing phase for all mentioned models. Experimental results indicate that the arc-shaped constriction leads to a relative energy loss in the range of 0.50–0.89. In the SVM model, the examination of various kernels revealed that the linear kernel outperforms polynomial, RBF, and Sigmoid kernels when compared to experimental results. The statistical indicators of the correlation coefficient ( $R$ ), RMSE, mean percentage relative error (average RE%), BIAS, DR, SI, and KGE for the SVM-Linear model in the testing phase are 0.994, 0.0101, 1.33%,  $-0.0002$ , 0.999, 0.0105 and 0.980, respectively. For the ANN method with MLP and RBF networks, the ANN-MLP approach shows more accurate results compared to other network types. Specifically, the statistical indicators for ANN-MLP in the testing phase are  $R = 0.997$ , RMSE = 0.0069, average RE = 0.63%, BIAS = 0.0004, DR = 0.999, SI = 0.0098, and KGE = 0.995. In the RF model, results are comparatively weaker than the other models. ANN-MLP outperforms SVM, GEP, M5, MARS and RF models and is closer to experimental results. It should be noted that the MARS model yields results very close to the ANN Model, and the equation provided by the MARS model can confidently be used. The statistical indicators for the MARS model in the testing phase are  $R = 0.997$ , RMSE = 0.0074, average RE% = 0.71%, BIAS =  $-0.0005$ , DR = 0.999, SI = 0.0106, and KGE = 0.995. The non-linear polynomial regression equation, in comparison to the MARS equation, exhibits relatively lower accuracy but can be used with high confidence. The non-linear regression and MARS relationships are presented in the scope of the present research.

## FUNDING

There is no funding source.

## DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

## CONFLICT OF INTEREST

The authors declare there is no conflict.

## REFERENCES

- Abbaszadeh, H., Norouzi, R., Sume, V., Kuriqi, A., Daneshfaraz, R. & Abraham, J. 2023 Sill role effect on the flow characteristics (experimental and regression model analytical). *Fluids* **8** (8), 235.
- Agarwal, S., Mukherjee, D. & Debbarma, N. 2023 Analysis of extreme annual rainfall in North-Eastern India using machine learning techniques. *AQUA – Water Infrastructure, Ecosystems and Society* **72** (12), 2201–2215.
- Al-Bulushi, N. I., King, P. R., Blunt, M. J. & Kraaijveld, M. 2012 Artificial neural networks workflow and its application in the petroleum industry. *Neural Computing and Applications* **21**, 409–421.
- Ayaz, M., Chourasiya, S. & Danish, M. 2024 Performance analysis of different ANN modelling techniques in discharge prediction of circular side orifice. *Modeling Earth Systems and Environment* **10** (1), 273–283.
- Chow, V. R. 1959 *Open-Channel Hydraulics*. McGraw-Hill, New York.
- Daneshfaraz, R., Sadeghfam, S., Aminvash, E. & Abraham, J. P. 2022a Experimental investigation of multiple supercritical flow states and the effect of hysteresis on the relative residual energy in sudden and gradual contractions. *Iranian Journal of Science and Technology, Transactions of Civil Engineering* **46** (5), 3843–3858.
- Daneshfaraz, R., Norouzi, R., Abbaszadeh, H. & Azamathulla, H. M. 2022b Theoretical and experimental analysis of applicability of sill with different widths on the gate discharge coefficients. *Water Supply* **22** (10), 7767–7781.
- Das, R., Pal, D., Das, S. & Mazumdar, A. 2014 Study of energy dissipation on inclined rectangular contracted chute. *Arabian Journal for Science and Engineering* **39**, 6995–7002.
- Dey, S. & Raikar, R. V. 2005 Scour in long contractions. *Journal of Hydraulic Engineering* **131** (12), 1036–1049.
- Fatehi-Nobarian, B., Nourani, V. & Ng, A. 2023 Application of meta-heuristic methods in the optimization of geometrical sections in trapezoidal channels in jump energy loss. *AQUA – Water Infrastructure, Ecosystems and Society* **72** (8), 1539–1552.
- Friedman, J. H. 1991 Multivariate adaptive regression splines. *The Annals of Statistics* **19** (1), 1–67.
- Ghaderi, A., Dasineh, M., Aristodemo, F. & Ghahramanzadeh, A. 2020 Characteristics of free and submerged hydraulic jumps over different macroroughnesses. *Journal of Hydroinformatics* **22** (6), 1554–1572.
- Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. 2009 Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology* **377** (1–2), 80–91.
- Habibzadeh, A., Rajaratnam, N. & Loewen, M. 2019 Characteristics of the flow field downstream of free and submerged hydraulic jumps. *Proceedings of the Institution of Civil Engineers-Water Management* **172** (4), 180–194.
- Hager, W. H. & Dupraz, P. A. 1985 Discharge characteristics of local, discontinuous contractions. *Journal of Hydraulic Research* **23** (5), 421–433.
- Hassanzadeh, Y. & Abbaszadeh, H. 2023 Investigating discharge coefficient of slide gate-sill combination using expert soft computing models. *Journal of Hydraulic Structures* **9** (1), 63–80.
- Henderson, F. M. 1966 *Open Channel Flow*. Macmillan, New York.
- Heydari, M., Shabanlou, S. & Sanahmadi, B. 2022 Self-adaptive extreme learning machine-based prediction of roller length of hydraulic jump on rough bed. *ISH Journal of Hydraulic Engineering* **28** (2), 152–162.
- Jahed Armaghani, D., Asteris, P. G., Askarian, B., Hasanipanah, M., Tarinejad, R. & Huynh, V. V. 2020 Examining hybrid and single SVM models with different kernels to predict rock brittleness. *Sustainability* **12** (6), 2229.
- Jan, C. D. & Chang, C. J. 2009 Hydraulic jumps in an inclined rectangular chute contraction. *Journal of Hydraulic Engineering* **135** (11), 949–958.
- Karbasi, M. & Azamathulla, H. M. 2016 GEP to predict characteristics of a hydraulic jump over a rough bed. *KSCE Journal of Civil Engineering* **20**, 3006–3011.
- Mobayen, R., Najafzadeh, M. & Farrahi-Moghaddam, K. 2023 Evaluation of regression-based soft computing techniques for estimating energy loss in gabion spillways. *Environment and Water Engineering* **9** (2), 241–255.
- Mohammed, A. Y. & Sharifi, A. 2020 Gene Expression Programming (GEP) to predict coefficient of discharge for oblique side weir. *Applied Water Science* **10** (6), 1–9.
- Momeneh, S. & Nourani, V. 2023 Performance evaluation of artificial neural network model in hybrids with various preprocessors for river streamflow forecasting. *AQUA-Water Infrastructure, Ecosystems and Society* **72** (6), 947–968.
- Najafzadeh, M. 2019 Evaluation of conjugate depths of hydraulic jump in circular pipes using evolutionary computing. *Soft Computing* **23** (24), 13375–13391.
- Najafzadeh, M. & Mahmoudi-Rad, M. 2024 New empirical equations to assess energy efficiency of flow-dissipating vortex dropshaft. *Engineering Applications of Artificial Intelligence* **131**, 107759.
- Najafzadeh, M., Oliveto, G. & Saberi-Movahed, F. 2022 Estimation of scour propagation rates around pipelines while considering simultaneous effects of waves and currents conditions. *Water* **14** (10), 1589.
- Najafzadeh, M., Basirian, S. & Li, Z. 2023 Vulnerability of the rip current phenomenon in marine environments using machine learning models. *Results in Engineering* **21**, 101704.
- Nasrabadi, M., Mehri, Y., Ghassemi, A. & Omid, M. H. 2021 Predicting submerged hydraulic jump characteristics using machine learning methods. *Water Supply* **21** (8), 4180–4194.



- Norouzi, R., Arvanaghi, H., Salmasi, F., Farsadzadeh, D. & Ghorbani, M. A. 2020 A new approach for oblique weir discharge coefficient prediction based on hybrid inclusive multiple model. *Flow Measurement and Instrumentation* **76**, 101810.
- Norouzi, R., Sihag, P., Daneshfaraz, R., Abraham, J. & Hasannia, V. 2021 Predicting relative energy dissipation for vertical drops equipped with a horizontal screen using soft computing techniques. *Water Supply* **21** (8), 4493–4513.
- Norouzi, R., Ebadzadeh, P., Sume, V. & Daneshfaraz, R. 2023 Upstream vortices of a sluice gate: An experimental and numerical study. *AQUA – Water Infrastructure, Ecosystems and Society* **72** (10), 1906–1919.
- Nourani, B., Arvanaghi, H., Pourhosseini, F. A., Javidnia, M. & Abraham, J. 2023 Enhanced support vector machine with particle swarm optimization and genetic algorithm for estimating discharge coefficients of circular-crested oblique weirs. *Iranian Journal of Science and Technology, Transactions of Civil Engineering* **47**, 3185–3198.
- Nouri, M., Sihag, P., Salmasi, F. & Kisi, O. 2020 Energy loss in skimming flow over cascade spillways: Comparison of artificial intelligence-based and regression methods. *Applied Sciences* **10** (19), 6903.
- Pal, M. & Deswal, S. 2009 M5 model tree based modelling of reference evapotranspiration. *Hydrological Processes: An International Journal* **23** (10), 1437–1443.
- Rahmanshahi, M. & Shafai Bejestan, M. 2020 Gene-expression programming approach for development of a mathematical model of energy dissipation on block ramps. *Journal of Irrigation and Drainage Engineering* **146** (2), 04019033.
- Rezaee, A., Bozorg-Haddad, O. & Chu, X. 2023 Comparison of data-driven methods in the prediction of hydro-socioeconomic parameters. *AQUA-Water Infrastructure, Ecosystems and Society* **72** (4), 438–455.
- Saberi-Movahed, F., Najafzadeh, M. & Mehrpooya, A. 2020 Receiving more accurate predictions for longitudinal dispersion coefficients in water pipelines: Training group method of data handling using extreme learning machine conceptions. *Water Resources Management* **34** (2), 529–561.
- Sauida, M. F. 2022 Simulation of relative energy loss downstream of multi-gate regulators using ANN. *Cogent Engineering* **9** (1), 2017388.
- Sun, D., Lonbani, M., Askarian, B., Jahed Armaghani, D., Tarinejad, R., Thai Pham, B. & Huynh, V. V. 2020 Investigating the applications of machine learning techniques to predict the rock brittleness index. *Applied Sciences* **10** (5), 1691.
- Taylor, K. E. 2001 Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* **106** (D7), 7183–7192.
- Vapnik, V. N. 1995 *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.
- White, F. M. 2016 *Fluid Mechanics*, 8th edn. McGraw Hill Education, Secaucus, NJ.
- Wu, B. & Molinas, A. 2001 Choked flows through contractions. *Journal of Hydraulic Engineering* **127** (8), 657–662.

First received 29 December 2023; accepted in revised form 12 February 2024. Available online 23 February 2024