The Study of the Performance of Data- Driven Models to Predict the Scour Depth Caused by the Aerated Vertical Jet

Babak Lashkar-Ara[1*], Saman Baharvand[2] and Leila Najafi[3]

[1*]- Corresponding Author, Associate Professor, Civil Engineering Department, Jundi-Shapur University of Technology, Dezful, Iran (*Lashkarara@jsu.ac.ir*).
2- PhD Candidate, Department of Civil Engineering, University of Texas at Arlington, P.O. Box 19019, Arlington, TX 76019, USA.
3- Instructor in Civil Engineering Department Jundi-Shapur University of Technology, Dezful, Iran .

Abstract

High flow discharges coming from the hydraulic structures usually carry a high-velocity jet of flow, which could have different short- and long-term impacts on the river mechanics and the habitat conditions. Scouring is one of the major effects of the incoming flow jet, which, once aerated, has a dynamic behavior and structure. Plunge pools are hydraulic structures to prevent the severe damages of the scouring phenomena. In the present study, due to the high complexity of constructing a physical model, the effect of air entrainment on scoured hole's depth is assessed using the Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) methods. Each soft computing model's performance on the scouring is compared to a Nonlinear Regression Method's result using different statistical measures (RMSE, ME, MAE). The prediction accuracy of ANN, ANFIS, and nonlinear regression using RMSE was calculated as 0.0137, 0.011, and 0.0262, respectively. This study presents a novel achievement in measuring and predicting the scoured hole's depth as one of the most critical phenomena in hydro-environmental science.

Introduction

Predicting scour holes created downstream of the hydraulic structures is a critical issue regarding the probable impacts of the river's morphology. It may cause some changes inside the habitat condition as well. Besides, the emerged scour holes should be in a way to minimize the failure and subverting probability. The effects of different parameters and the existence of nonlinear relations among the scour's parameters raise the complexities of estimating the downstream changes. A typical way to explore the relationship between the scour hole dimension parameters is to recognize the dimensionless parameters and then determine the governing mathematical mapping of the effective parameters.

Extensive research projects were conducted to determine local scour around hydraulic structures, and in most of the studies, the results were presented in an empirical form. Rouse (1940) reviewed the scour hole dimension from the initiation to the hole's equilibrium form. Borman and Bormann and Julien (1991) presented equations for scouring a hole by investigating the jet's characteristics. The dynamics of shear stress flow in the scouring hole were examined by Robinson et al. (2000). The scouring might be affected by the amount of entrained air into the incoming jet based on their findings. Therefore, Equation 1 for the scouring depth was proposed considering air aeration inside the incoming jet (Mason, 1989; Mason & Arumugam, 1985).

$$T = 3.39 q^{0.06} (1+\beta)^{0.3} h^{0.16} / (g^{0.3} D^{0.06}) \quad (1)$$

where T is the scouring depth, q is the specific discharge, h is the tailwater depth, g is the acceleration of gravity, D is the mean diameter of bed materials, and β is the fraction of air inside the incoming water, which was proposed by Ervine (1976). In Equation 1, β is a factor for the definition of scouring process replaced instead of the jet's height (H). Bohrer et al. (1998) investigated the behavior of the jet's velocity and assessed the decreasing trend of the velocity profiles. Two jet types were evaluated with and without air entrance inside the water jet. The results showed that the air-entraining process would decrease the jet's mean velocity. Canepa and Hager (2003) studied the air and water interaction and its effects on scouring. They showed a reverse relationship between the amount of air and the depth of scour hole. Xu et al. (2004) studied the air-entraining impact on scouring caused by the falling jets. They proposed a formula addressing the relationship between the amount of concentrated air and the relative scouring depth. Under the same bed material and tailwater depth, the scour hole profile basically depends on the scour depth and is not significantly affected by air concentration. Jet air entrainment can affect the shape of the scour hole and reduce the depth of the scour.

The experimental research results are used in the present study to predict the scour depth under different hydraulic conditions using artificial neural networks and nonlinear regression-based approaches. The methodology and the results of each part of the study are discussed in the further sections.

Material and Method

Based on the scope of this study, two machine learning approaches, including ANN, ANFIS, along with a nonlinear based regression method, were applied to the experimentally result from the previous studies. Each model's construction and theory are discussed in the following sections.

The Adaptive Neuro-Fuzzy Interface System

ANFIS model is a general investigation tool that approximates continuous real functions on compact sets. This model was developed in 1993 by Jang and used by different hydro-environmental researchers to investigate various aspects and hydraulic parameters, such as the prediction of effective hydrodynamic parameters of the hydraulic jump (Baharvand et al., 2020; Roushangar et al., 2018), the bedload transport (Azamathulla et al., 2009a; Riahi-Madvar & Seifi, 2018), the prediction of meteorological variables (Hassanzadeh et al., 2020), the prediction of flow characteristics over different type of spillways (Azamathulla et al., 2009b; Yildiz et al., 2020), etc. The connection between the ANFIS model nodes is of "directional links" type, in which the node's function introduces the variable or constant parameters to each node (Jang et al., 1997). The architecture of the developed ANFIS model is shown in Figure (1). The example below presents the two if-then fuzzy models' rules, known as the Takagi and Sugeno rule system, used inside the developed prediction model.

Rule 1: If x is A_1, y is B_1, and z is C_1, then
$$f_1 = p_1\,x + q_1\,y + r_1\,z + s_1$$
Rule 2: If x is A_2, y is B_2, and z is C_2, then
$$f_2 = p_2\,x + q_2\,y + r_2\,z + s_2$$

where f_1 and f_2 are the output functions of rules 1 and 2, respectively. Figure 1 shows the structure of the developed ANFIS model.
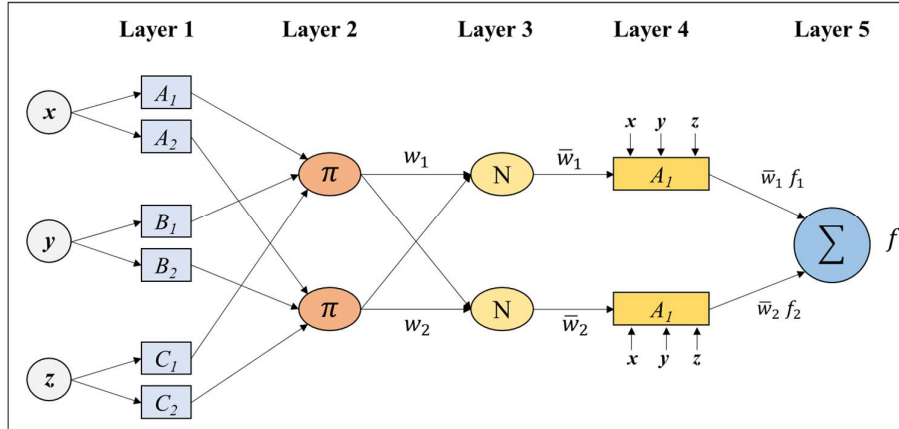
Fig. 1- The architecture of the developed ANFIS model to estimate the scour depth parameters

In Figure (1), $O_{l,i} = \phi A_i(x)$ indicates the node function for each square node i for layer 1, and $i = 1, 2$, x, is the ith input node; A_i is a linguistic label (i.e., "small" or "big") for this node function. When input x satisfies quantifier A_i, $O_{1,i}$, is the membership function (MF) of a fuzzy set A (e.g., A_1, A_2, B_1, B_2, C_1, C_2) and indicates the degree of the produced set. $\phi A_i(x)$ is mainly a Gaussian function ranging between 0 and 1 as the minimum and maximum levels, respectively.

$$\phi A_i(x) = \exp\left(-\left(\frac{x - a_i}{b_i}\right)^2\right) \quad (2)$$

where a_i, b_i are the parameter sets. The circle nodes of layer 2 have index Π, which indicates the multiplication of the inputs. For instance,
$$w_i = \phi A_i(x)\, \phi B_i(y)\, \phi C_i(z)\ , i = 1, 2.$$
The output of each node shows the impression level of a rule. Circle nodes of layer 3 take the label N. Following this, the ratio of the level of an impression of the ith rule on the sum of all rules' levels is computed by the ith node using the Equation (3).

$$w_i = \frac{w_i}{w_1 + w_2}\ ,\quad i = 1, 2 \quad (3)$$

The node function of the square nodes in layer four is calculated using Equation (4).

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i\left(p_i\, x + q_i\, y + r_i\, z + s_i\right) \quad (4)$$

where $\overline{w}_i$ represents the layer three output, and the parameter series is $\{p_i, q_i, r_i, s_i\}$. These layer parameters are called consequent parameters. The single circle nodes of layer five summarizes all incoming signals and returns the result as the final output, as shown in Equation (5).

$$O_{5,i} = \sum_{i=1} \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (5)$$

The output of the ANFIS model is generated using either linear or fixed value functions. Detailed information for the ANFIS model and its function is available in Jang's (1993) study.

Nonlinear Regression

Nonlinear Regression analysis is mainly used when two or more variables might be connected systematically by a nonlinear relationship. The complexity of the nonlinear approaches compared to the linear models is evident because the model's primary function is constructed based on different approximations inside each iteration. Different well-known mathematic methods such as the Gauss-Newton method and the Levenberg-Marquardt method are developed, which can be used as the nonlinear models' primary function.

In nonlinear models, one of the parameters would be the dependent

parameter. Simultaneously, the rest of the variables are used as the independent parameters by the assigned function to predict the dependent term. Equation 6 indicates the present study's primary equation concerning the effective parameters on the scour hole depth produced by the dimensional analysis method.

$$d_{st} = a\, p_t^b \times q_t^c \times r_t^d \times s_t^e \qquad (6)$$

where a, b, c, d, and e are constant coefficients. Also, p_t, q_t, r_t, and s_t are representative of the independent variables.

Model Performance

Four statistical measures are used to examine the fitted function's accuracy by ANN, ANFIS, and MLR models to the testing data. These measures are Main Absolute Error (MAE), Root Mean Square Error (RMSE), determination coefficient (R^2), and slope of the best fit line (m). Statistical functions used for evaluating the models' performance are expressed through Equations (7) to (9).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|O_i - P_i\right| \qquad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(O_i - P_i)^2}{N}} \qquad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O_i})^2} \qquad (9)$$

where O is the observed parameter, $\overline{O}$ is the average of observed parameters, P is the model's predicted parameter, and N is the total data. The regression line gradient between predicted and observed results (m) is also calculated for evaluating the model's

performance. The model's highest accuracy occurred when m=1, which shows that the model predicted the values entirely correct, and the performance of the model would be a hundred percent.

Experimental Data

The experiments are conducted in a pool with 2m length, 1m width, and 1m depth. Figure (2) shows the designed experimental setup for the present study. The discharge is transported to the nozzle by a circular pipe with a 4-inch diameter. For simulating a vertical jet, a roller for the nozzle is fixed at 90 degrees. The tailwater depths are controlled in three scenarios, 0.325m, 0.385m, and 0.435m. The equilibrium time of scouring is determined for 5 hours. The equilibrium time is the appropriate time needed for collecting the bed elevation profile. The bed elevation variation is detected by a Leica Disto-d8 laser meter. Figure 2 shows the nozzle location, scour hole, and governing parameters of the study.

The most critical parameters that affect d_s are the maximum depth of scouring and L_s, the length of scour downstream of the impinging jet, the discharge of spillway Q_w, the incoming air discharge Q_a, the total head H, the velocity of air-water mixed flow (V), the width of the pool (B), the tailwater depth (h), the nozzle diameter (d_n), the mean sediment size (d_{50}), the gravitational acceleration (g), the fluid viscosity (μ), the density of water and (ρ_w), and sediment's density (ρ_s), respectively. Identifying effective parameters is essential for determining the relation between effective parameters of scour hole dimensions. To find the best function addressing the scour depth, d_s and L_s are assumed dependent variables and $Q_W, Q_a, V, H, d_n, B, h, D_{50}, g, \mu, \rho_s, \rho_w$ are considered the study's independent variables. For determining a simple dimensionless equation, the air concentration is identified as Equation 10.

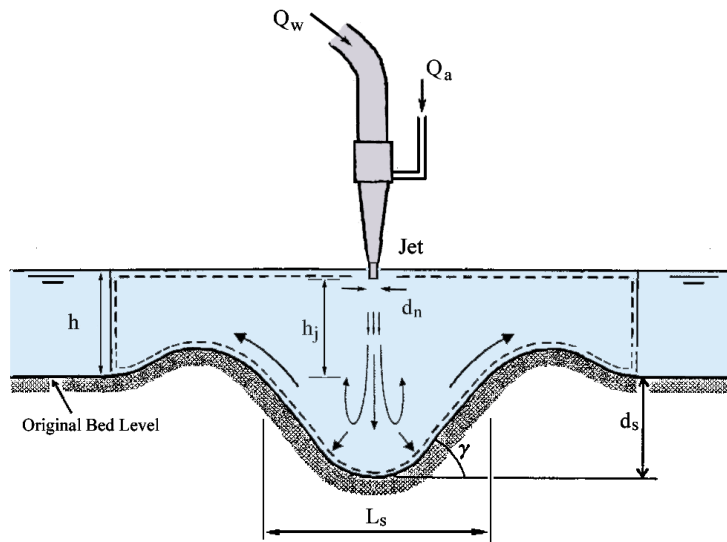$$C_a = \left(\frac{Q_a}{Q_a + Q_w}\right) \times 100 \qquad (10)$$

Fig. 2- Sketch of the pre-aerated jet and its effect in the plunge pool on the scour hole dimension

Table 1- Effective parameters range in the present study

Parameter	Symbol	Unit	Limit
Tailwater depth	H	m	0.26-0.45
Nozzle diameter	d_n	mm	20.5-36.5
Air discharge	C_a	%	0-30.1
Flow velocity (air and water mixture)	V	m/sec	1.46-2.41
Depth of scouring	d_s	m	0.019-0.227
Length of scour downstream of the jet	L_s	m	0.21-0.88

Using π Buckingham theory and by considering ρ, g, and h, as primary variables, the dimensionless equation governing the scour hole dimensions' behavior in a plunge pool can be written as Equations 11 and 12:

$$\frac{d_s}{h} = f(Fr, d_n/h, C_a) \tag{11}$$

$$\frac{L_s}{h} = f(Fr, d_n/h, C_a) \tag{12}$$

An accurate experimental data set is needed to investigate different effective scenarios to optimize the scour hole dimensions downstream a spillway. Table (!) shows the ranges of each parameter used in this study.

In the training stage of the soft computing and MLR approaches, 80 percent of the randomly experimentally driven data is assigned to the training dataset. The rest of the dataset (20) is used to assess each model's accuracy. The result of each model's precision is discussed in the following section.

Results and Discussion

The ANN model is designed and trained 80% of the effective dimensionless parameters. The model's optimal architecture containing the input layer, hidden layers, and their neuron threshold function, iteration cycle, and the output section are determined using MATLAB 7.12. Input parameters are considered as Fr, d_n/h, and C_a $(Fr, d_n/h, C_a)$, and output parameters are d_s and L_s. Different network patterns are used to predict the best function addressing the scour depth. Table (@) shows a summary of the created network architecture and its performance using the statistical measures.

The result of ANN prediction in preparation and test stages are summarized in Table (3).

Figures (3a) and (4a) show the ANN model's performance in estimating d_s/h and L_s/h in the training stage with a correlation coefficient of more than 0.99. The "m" value is almost 1 in both models (Table 3), representing the model's high ability to predict the training stage's scour depth. The fitted model's result on the test dataset is shown in Figures (3b) and (4b).

Table 2- Comparison of the network architecture for dimensionless length and width of scouring

Output parameter	Input parameter	Network architecture	Training		test		Performance function
			R^2	RMSE	R^2	RMSE	
d_s/h	$Fr,d_n/h,C_a$	3-3-1	0.9921	0.0189	0.9915	0.0219	Sigmoid
d_s/h	$Fr,d_n/h,C_a$	3-4-1	0.9922	0.0187	0.9922	0.021	Sigmoid
d_s/h	$Fr,d_n/h,C_a$	3-5-1	0.9922	0.0188	0.9913	0.0222	Sigmoid
d_s/h	$Fr,d_n/h,C_a$	3-6-1	0.9916	0.0195	0.9902	0.0236	Sigmoid
d_s/h	$Fr,d_n/h,C_a$	3-3-3-1	0.9899	0.0213	0.9899	0.0237	Sigmoid
d_s/h	$Fr,d_n/h,C_a$	3-4-4-1	0.9907	0.0205	0.9911	0.0224	Sigmoid
L_s/h	$Fr,d_n/h,C_a$	3-3-1	0.9924	0.0439	0.9909	0.0552	Sigmoid
L_s/h	$Fr,d_n/h,C_a$	3-4-1	0.9917	0.0461	0.9898	0.0579	Sigmoid
L_s/h	$Fr,d_n/h,C_a$	3-5-1	0.9879	0.0555	0.9866	0.0652	Sigmoid
L_s/h	$Fr,d_n/h,C_a$	3-6-1	0.9918	0.0455	0.9903	0.0564	Sigmoid
L_s/h	$Fr,d_n/h,C_a$	3-3-3-1	0.9921	0.0446	0.9907	0.0563	Sigmoid
L_s/h	$Fr,d_n/h,C_a$	3-3-4-1	0.9899	0.0503	0.9884	0.0614	Sigmoid
L_s/h	$Fr,d_n/h,C_a$	3-4-4-1	0.992	0.0449	0.9909	0.0558	Sigmoid

Table 3- Statistical measures of the ANN model

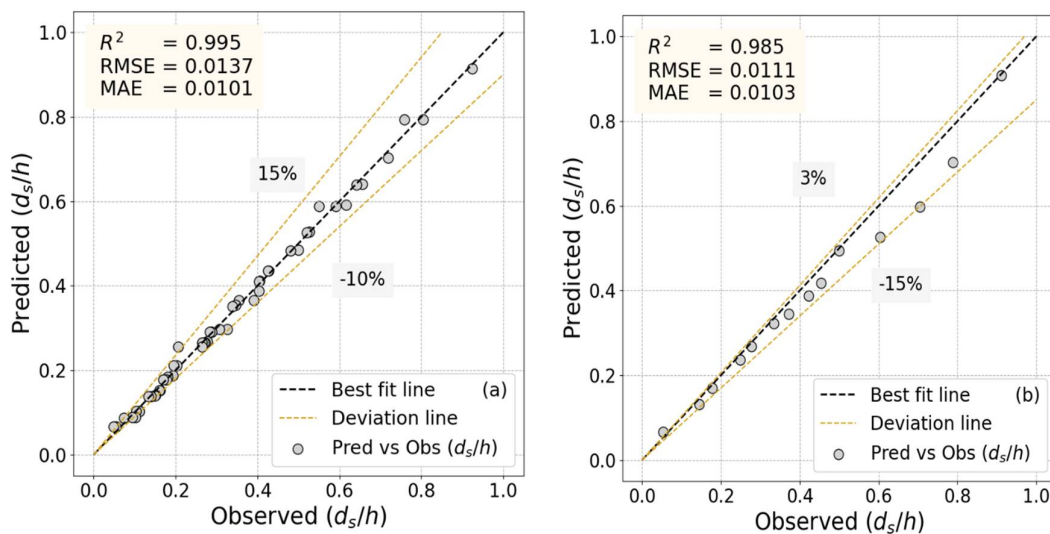| Parameter	Train				Test			
	MAE	RMSE	R^2	m	MAE	RMSE	R^2	m
ds/h	0.0101	0.0137	0.9951	0.9985	0.0103	0.0111	0.985	0.9268
Ls/h	0.0252	0.346	0.999	0.9943	0.0439	0.0518	0.9732	1.0301

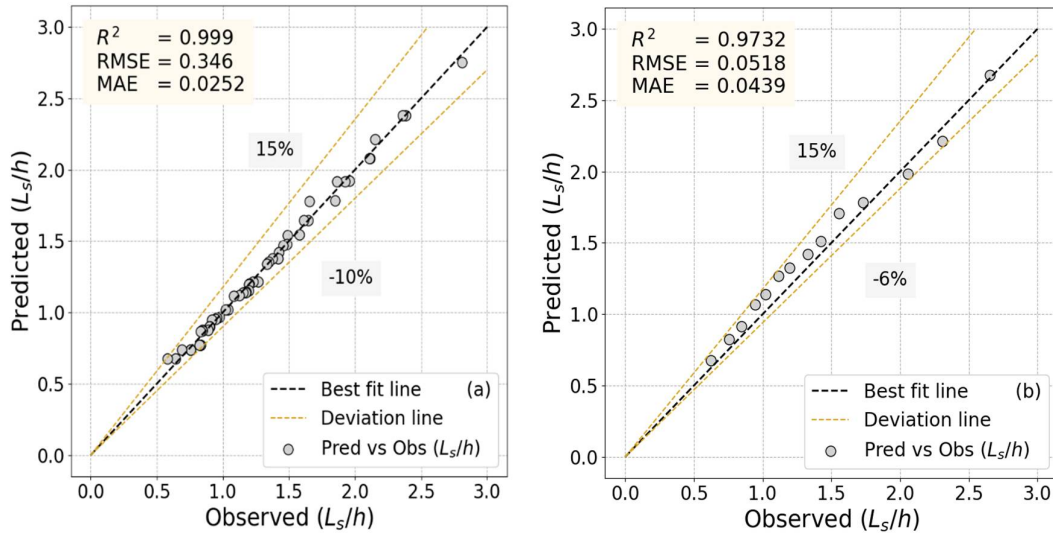**Fig. 3- ANN model performance in estimating d_s/h, a) Train b) Test**

Fig. 4- ANN model performance in estimating L_s/h, a) Train b) Test

**Table 4- Statistical measures of the ANFIS model in estimating dimensionless
length and width of scouring**

Parameter	Train				Test			
	MAE	RMSE	R^2	m	MAE	RMSE	R^2	m
ds/h	0.0078	0.011	0.9973	0.9995	0.0102	0.0139	0.9975	1.0146
Ls/h	0.0191	0.0287	0.9968	1.0023	0.0461	0.0641	0.9929	1.0259
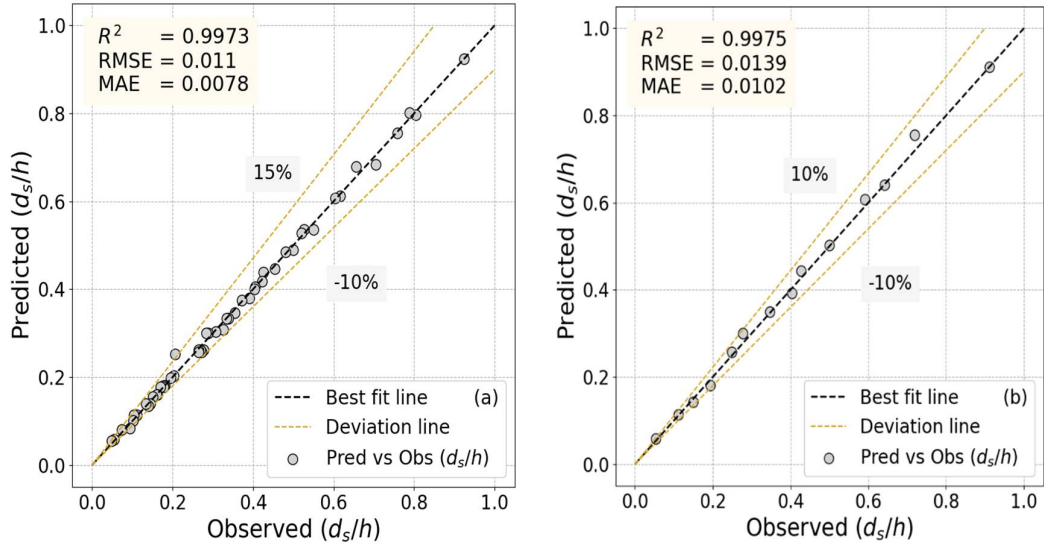
Fig. 5- ANFIS model performance in estimating d_s/h, a) Train b) Test

The developed ANFIS model used in the present study estimates dependent variables in a highly acceptable range. The performance of the ANFIS model is shown in Table (4). Figures (5) and 6 show the ANFIS model's performance in estimating the length and width of scouring. The correlation coefficient values of the test stage show the significant ability of the model to predict the length and width of scouring (Figures 5b and 6b).
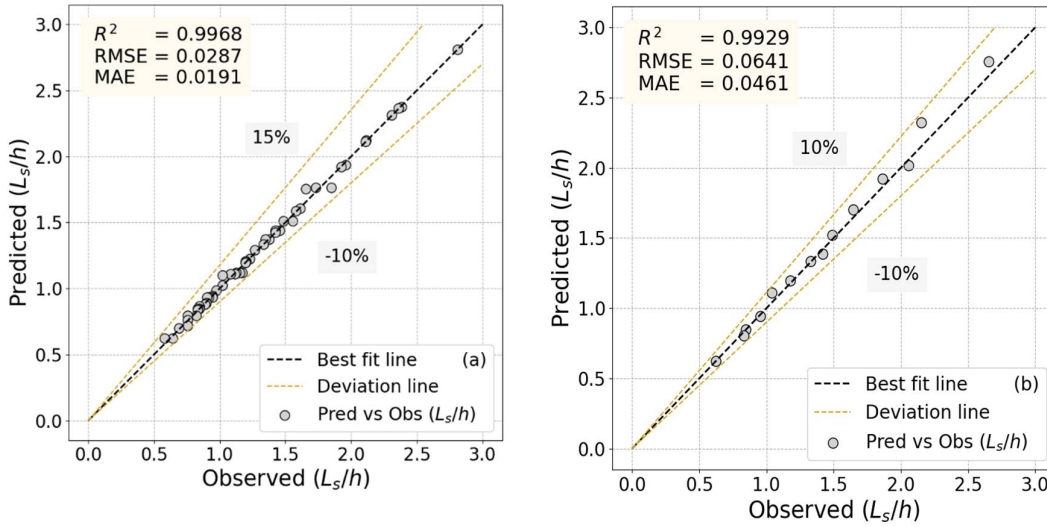
Fig. 6- ANFIS model performance in estimating L_s/h, a) Train b) Test

Table 5-Statistical analysis of nonlinear regression relation for estimating scour hole

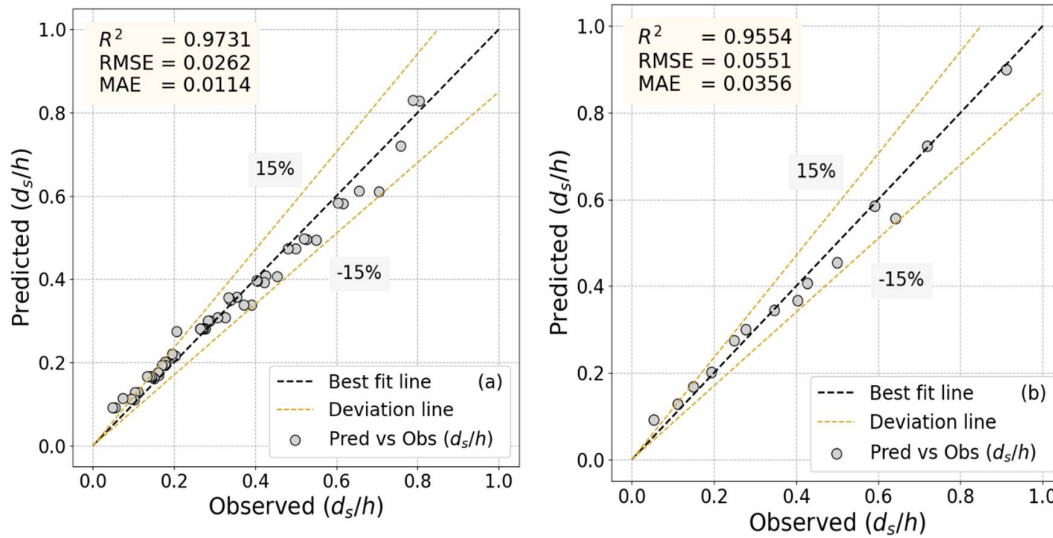| Parameter	Train				Test			
	MAE	RMSE	R^2	m	MAE	RMSE	R^2	m
ds/h	0.0114	0.0262	0.9731	0.993	0.0356	0.0551	0.9554	1.0232
Ls/h	0.0143	0.0334	0.9926	0.999	0.0441	0.0545	0.9922	1.0116

Fig. 7- Nonlinear regression performance in estimating d_s/h, a) Train b) Test

The nonlinear regression-based approach was produced by SPSS software and applied to the same 80 percent of the experimental dataset used for the ANN and ANFIS models. Equations 11 and 12 show the resulted formula addressing the scour depth and length, respectively.

$$d_s/h = 0.7919(Fr)^{-0.5274}(d_n/h)^{2.908}(1-C_a)^{0.8862}$$
(11)

$$L_s/h = 1.465(Fr)^{0.0057}(d_n/h)^{1.011}(1-C_a)^{-1.232}$$
(12)

The performance of Equations (9) and (10) for predicting scour hole dimension are summarized in Table (5).

Figures (7) and (8) show the results of nonlinear regression prediction. For 80% of the training data set, the correlation coefficient of the scour hole parameters d_s/h and L_s/h are 0.97 and 0.99, and for the testing stage of the proposed equations, the coefficients are 0.95 and 0.99, respectively. The "m" value of fitted line is so close to 45 degree for determining equations. For depth and scouring length testing stages, this value is 0.023 and 0.011 higher than the optimal value (m=1), respectively (Table 5).
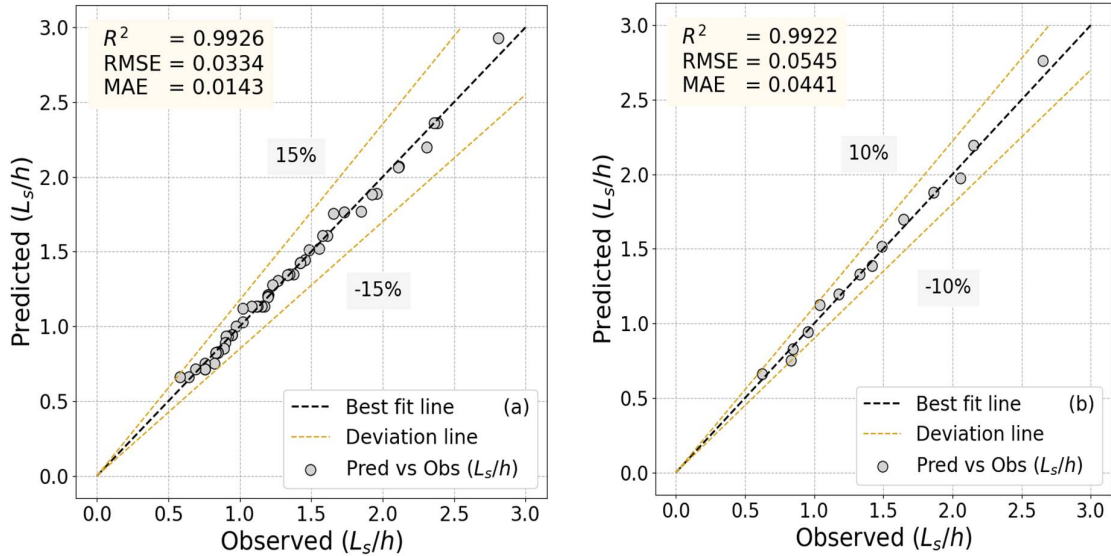

Fig. 8- Nonlinear regression performance in estimating L_s/h, a) Train b) Test
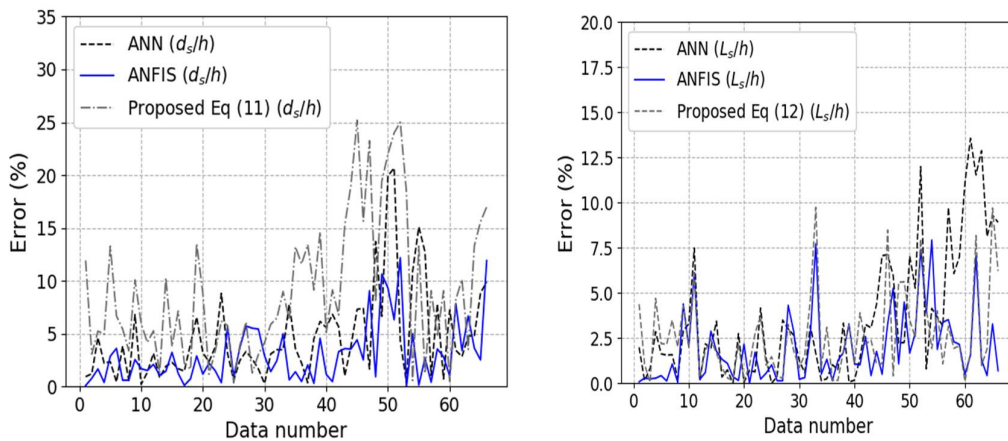

Fig. 9- Error percentage of the data-driven methods for prediction of scour hole dimensions a) depth b) length

The result's comparison shows that the root main square error of ANN, ANFIS, and Nonlinear regression method to predict maximum scour hole downstream of the aerated vertical jet is 0.0137, 0.011, and 0.0262, respectively.

For a better comparison of the developed models' performance in predicting the scour dimensionless hole parameters, the error of each model's predicted values from the real data is shown in Figure (9). The ANFIS model's error fluctuates from 0% to 12.7% in predicting the depth of scour hole Figure (9a), while Figure (9b) shows the error range between 0 to 7.8% for ANFIS model scour hole's length estimation. For the entire data of the study, there is an acceptable average error fluctuation. This study brings novel achievements for the engineering community and indicates the ANFIS model's superiority against ANN and MLR methods. However, the statistical measures prove that the proposed simplified nonlinear equation has acceptable accuracy for the designers and investigators in future studies.

Conclusion

In the present study, the aerated vertical water jet's effect on the created scour hole dimensions was discussed. First, the experimental model was designed, and the effective parameters were identified and used to produce the effective dimensionless parameters of the study. The effective parameters on the scour hole's depth and length were identified as Fr, d_n/h, C_a. Second, two ANN and ANFIS models were trained to predict the scour hole dimensions under different hydrodynamic conditions. In addition to the mentioned data driven approaches, a nonlinear regression-based model was fitted to the training dataset in order to generate a dimensionless formula to define the scour hole's geometrical characteristics. Different statistical measures were used to identify the performance of each three mentioned soft computing techniques. The root main square error for ANN, ANFIS, and nonlinear regression method to predict the scour hole's depth is 0.0137, 0.011, and 0.0262, respectively. Comparing the physical model data and the soft computing models revealed that the ANN and ANFIS had RMSE of 0.346, 0.0287, respectively, to predict the scour hole's length. However, the nonlinear regression approach had the RMSE of 0.0334 for estimating the length of the hole. Hence, the ANFIS model showed the best performance in predicting the scour dimensions. Two empirical equations for the scour geometry were derived as a result of the nonlinear regression model with acceptable accuracy to be used for future studies.

References

1- Azamathulla, H.M., Chang, C.K., Ghani, A.A., Ariffin, J., Zakaria, N.A. and Hasan, Z.A., 2009a. An ANFIS-based approach for predicting the bed load for moderately sized rivers. *Journal of Hydro-Environment Research*, *3*(1), pp.35-44.

2- Azamathulla, H.M., Ghani, A.A. and Zakaria, N.A., 2009b, December. ANFIS-based approach to predicting scour location of spillway. In *Proceedings of the Institution of Civil Engineers-Water Management* (Vol. 162, No. 6, pp. 399-407). Thomas Telford Ltd.

3- Baharvand, S., Jozaghi, A., Fatahi-Alkouhi, R., Karimzadeh, S., Nasiri, R. and Lashkar-Ara, B., 2020. Comparative Study on the Machine Learning and Regression-Based Approaches to Predict the Hydraulic Jump Sequent Depth Ratio. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, pp.1-14.

4- Bohrer, J.G., Abt, S.R. and Wittler, R.J., 1998. Predicting plunge pool velocity decay of free falling, rectangular jet. *Journal of hydraulic engineering*, *124*(10), pp.1043-1048.

5- Bormann, N.E. and Julien, P.Y., 1991. Scour downstream of grade-control structures. *Journal of hydraulic engineering*, *117*(5), pp.579-594.

6- Canepa, S. and Hager, W.H., 2003. Effect of jet air content on plunge pool scour. *Journal of Hydraulic Engineering*, *129*(5), pp.358-365.

Ervine, D. A. (1976). ''The entrainment of air in water.'' *Int. Water Power Dam Constr.,* 28(12), 27–30.

7- Hassanzadeh, Y., Ghazvinian, M., Abdi, A., Baharvand, S. and Jozaghi, A., 2020. Prediction of short and long-term droughts using artificial neural networks and hydro-meteorological variables. *arXiv preprint arXiv:2006.02581*.

8- Jang, J.S., 1993. ANFIS: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*, *23*(3), pp.665-685.

9- Jang, J.S.R., Sun, C.T. and Mizutani, E., 1997. Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]. *IEEE Transactions on automatic control*, *42*(10), pp.1482-1484.

10- Mason, P.J., 1989. Effects of air entrainment on plunge pool scour. *Journal of Hydraulic Engineering*, *115*(3), pp.385-399.

11- Mason, P.J. and Arumugam, K., 1985. Free jet scour below dams and flip buckets. *Journal of Hydraulic Engineering*, *111*(2), pp.220-235.

12- Riahi-Madvar, H. and Seifi, A., 2018. Uncertainty analysis in bed load transport prediction of gravel bed rivers by ANN and ANFIS. *Arabian Journal of Geosciences*, *11*(21), pp.1-20.

13- Robinson, K.M., Cook, K.R. and Hanson, G.J., 2000. Velocity field measurements at an overfall. *Transactions of the ASAE*, *43*(3), p.665.

14- Rouse, H., 1940. Criteria for similarity in the transportation of sediment. *University of Iowa Studies in Engineering*, *20*, pp.33-49.

15- Roushangar, K., Ghasempour, R. and Valizadeh, R., 2018. Effect of channel boundary conditions in predicting hydraulic jump characteristics using an ANFIS-based approach. *Journal of Applied Fluid Mechanics*, *11*(3), pp.555-565.

16- Xu, W., Deng, J., Qu, J., Liu, S. and Wang, W., 2004. Experimental investigation on influence of aeration on plane jet scour. *Journal of Hydraulic Engineering*, *130*(2), pp.160-164.

17- Yildiz, A., Yarar, A., Kumcu, S.Y. and Marti, A.I., 2020. Numerical and ANFIS modeling of flow over an ogee-crested spillway. *Applied Water Science*, *10*(4), pp.1-10.