



# Machine Learning as a Tool for Estimating Discharge Coefficient for Rectangular Weir with Multiple Slots

Mona A. Hagraas

**Abstract:** *In terms of higher water demand due to increases of agriculture lands which depend on irrigation channels, there is a need to increase the efficiency of weirs already constructed on those channels. The most reasonable solution is to modify the existing weirs instead of replacing them by new ones to avoid costly solutions. In this study, a potential of using different Machine Learning regression algorithms has been investigated to estimate the coefficient of discharge  $C_d$  for rectangular weir with multiple circular slots. Using experimental data set, three Machine Learning regression models; Decision Tree Regressor (DTR), Artificial Neural Networks (ANN) and Support Vector Machine Regressor (SVR) were developed and compared to find most suitable algorithm. Based on the simulation results of the three developed models, it was found that the ANNs algorithm is the superior one which can be used to estimate discharge coefficient  $C_d$  for rectangular weirs with multiple circular slots. It gives the highest matching between measured and predicted values with correlation coefficient ( $R^2$ ) value of 0.759, minimum MAE with value 0.001 and minimum MSE with value 0.022. Finally, an equation using ANNs is presented to estimate the discharge coefficient.*

**Keywords:** Discharge Coefficient, Machine Learning, Rectangular Weir with slots

## I. INTRODUCTION

Weirs are the most common water structures. They may be used to measure the water flow rate and divide water between irrigation canals, etc. There are many types of weirs. According to geometry, they can be classified to rectangular, triangular and Cipolletti weirs. According to shape of crest, they can be classified to Sharp-crested, Broad-crested, Narrow-crested, and Ogee weirs.

In terms of higher water demand due to increases of planted lands which depend on irrigation channels, there is a need to increase the efficiency of constructed weirs on those channels. The most reasonable solution is to modify the existing weirs instead of replacing them by new ones to avoid costly solutions. Modification can be such as; lowering or Widening weir crest or adding slots in weir body. Researches in hydraulic flow characteristics of weir with bottom slots

started long time ago. Wolters et al. [1] suggested set of rating curves of discharge for a structure composed of weir and pipe. Abdel Halim et al., [2] deduced experimentally a relation for estimating the water flow rate over weirs with orifices. Unfortunately, this relation cannot be used as a general case as it was on the basis of used model dimensions. Negm [3],[4],[5] studied experimentally the composite flow over sharp-edged and beneath a gate with different shapes under different flow conditions. He presented a general dimensionless equation to predict the discharge of the composite flow. Elazizy [6] developed experimentally a general equation to estimate the discharge for a weir with a slot. On the basis of energy principal and dimensional analysis theory, Hassan et al. [7] developed multiple regression formulas to compute the discharge over clear over fall weir with a lower circular opening. El-Belasy [8] developed formulae for rehabilitated weirs using SOBEK model based on field measurements from El-Fayoum district. These formulae can be used in designing of required modification for the existing weir in order to increase discharge. Weirs with one opening and more than one opening were tested. Neveen and Ehab [9] tested experimentally three cases of openings arrangements including one, two and three openings. They developed an equation to estimate the discharge coefficient for different suggested combinations. Recently, Artificial Intelligence (AI) models have been strongly employed in modeling of many problems in water engineering field such as investigation of flow characteristics over different types of weirs. Bilhan et al. [10] could predict coefficient of discharge for side weirs using two different types of Neural Networks techniques. He concluded that performance of ANNs models is better than the common regression models. Furthermore, Dursun et al. [11] and Khoshbin et al. [12] employed successfully the Adaptive Neuro-Fuzzy Inference System (ANFIS) to estimate the coefficient of discharge for semi-elliptical labyrinth side weirs and rectangular sharp-crested side respectively.

Parsaie and Haghiabi [13] concluded that the developed model using the SVM is a bit more accurate compared to; Multilayer Perceptron Neural Network (MLPNN), Radial Basis Neural Network (RBFNN) to estimate the discharge coefficient ( $C_d$ ) of triangular labyrinth weir. Recently, Ehsan et al., [14] predicted the triangular weirs coefficient of discharge using the Extreme Learning Machine (ELM). For rectangular side weirs built on trapezoidal cross section, Azimi et al.,

Revised Manuscript Received on April 04, 2020.

\* Correspondence Author

**Mona Abdelhamid Elsayed Hagraas\***, Irrigation and Hydraulics Department, Faculty of Engineering, Ain Shams University, Cairo, Egypt. Email: mona\_hagraas@eng.asu.edu.eg

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

[15] used Support Vector Machines (SVMs) algorithm to estimate their coefficients of discharge.

The main objective of the current research is to develop quick and easy reliable model to simulate combined flow for rectangular weir with multiple circular slots and to estimate its coefficient of discharge. For these purposes three Machine Learning Regression Models based on three different algorithms; Decision Tree Regressor (DTR), Artificial Neural Networks (ANN) and Support Vector Machine Regressor (SVR) were developed and compared to find the most suitable model.

## II. THEORETICAL APPROACH

### A. Governing Equation

Fig. 1 illustrates the sketch of composite structure to show flow over the weir and through openings. Discharge over a rectangular weir can be determined by Equation (1). Flow via circular slot is considered as flow through orifice which can be determined using Equation (2) based on Chanson [16] and [17].

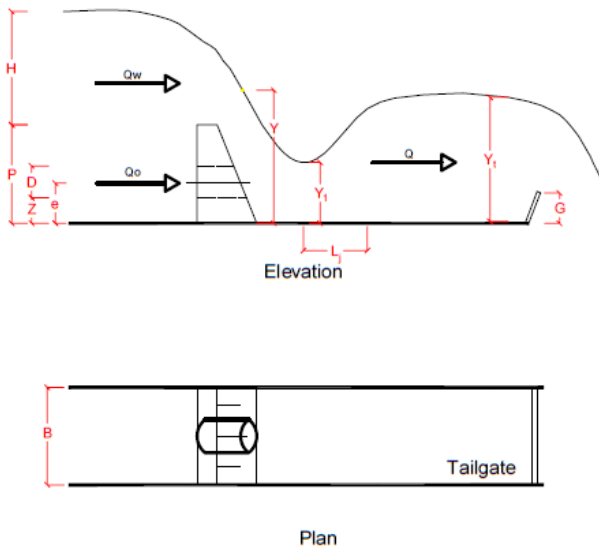


Fig. 1. Sketch of the experimental setup

$$Q = \frac{2}{3} B \sqrt{2g} H^{1.5} \quad (1)$$

$$Q_o = N \left( \frac{\pi D^2}{4} \right) \sqrt{2g} (H + P - y)^{1.5} \quad (2)$$

Where  $Q$  is the theoretical discharge over the weir [ $L^3/T$ ],  $B$  is the weir length [ $L$ ],  $g$  is the gravity acceleration [ $L/T^2$ ] and  $H$  is the head over the weir crest [ $L$ ].

$N$  is number of slots,  $Q_o$  is the theoretical discharge through slots [ $L^3/T$ ],  $D$  is diameter of the circular slot [ $L$ ],  $y$  is the head of water downstream the circular slot [ $L$ ],  $P$  is the weir height [ $L$ ],  $h_f$  is the friction loss through the circular slot [ $L$ ], but it can be neglected due to the small length of the slot. Total theoretical discharge ( $Q_{th}$ ) is calculated as shows in Equation (3)

$$Q_{th} = \sqrt{2g} \left( \frac{2}{3} B H^{1.5} + N \left( \frac{\pi D^2}{4} \right) (H + P - y)^{1.5} \right) \quad (3)$$

If measured discharge is ( $Q$ ), the overall coefficient of discharge  $C_d$  can be estimated using equation (4)

$$C_d = \frac{Q}{Q_{th}} \quad (4)$$

### B. Dimensional Analysis

Dimensional analysis is employed to come up with relationship between the combined structure geometry, number of slots and total discharge coefficient. Equation (5) shows expected variable governing the studied problem.

$$f(Q, C_d, H, P, B, \sqrt{A}, Z, y_1, y_t, L_j, g, \mu, \rho) = 0 \quad (5)$$

Where  $f$  is the functional symbol,  $Q$  is the total measured discharge [ $L^3/T$ ],  $C_d$  is the overall coefficient of discharge [dimensionless],  $Z$  is the vertical distance between the bed and the beginning of the slot [ $L$ ],  $y_1$  is the initial depth of the jump [ $L$ ],  $y_t$  is the tail water depth [ $L$ ],  $L_j$  is length of the submerged hydraulic jump [ $L$ ],  $g$  is the acceleration due to gravity [ $L/T^2$ ],  $\mu$  is the water dynamic viscosity [ $ML^{-1}T^{-1}$ ], and  $\rho$  is the water density [ $M/L^3$ ]. By applying Buckingham's theorem, the following function is achieved:

$$\frac{Q}{P^2 \sqrt{gP}} = f \left( C_d, \frac{H}{P}, \frac{\sqrt{A}}{Z}, F_n, \frac{L_j}{y_t} \right) \quad (6)$$

Where  $F_n$  is the Froude number

### C. Materials and Methods

In this study, three models were developed using three Machine Learning regression algorithms: Support Vector Machine Regressor (SVR), Decision Tree Regressor (DTR), Artificial Neural Networks (ANNs) to estimate the discharge coefficient for rectangular weir with multiple circular slots. A sub codes have been written using the tool of Machine Learning in Python as shown in the workflow, Fig. 2

### D. Data Set Description

In this study, available published experimental data by Neveen and Ehab [9] was used. These experiments were carried out within the Irrigation and Hydraulics Laboratory of the Faculty of Engineering, Ain Shams University. They used 12 models. All the models have a top width of 2 cm, height of 6 cm and constant downstream slope 1:2. Models were divided into three groups each has constant total slots area with different number of slots: one, two and three slots. Experiments were run under subcritical flow condition with weir crest always submerged. Laboratory measured variables which used in this study are: upstream head ( $H$ ), actual discharge ( $Q$ ), slot diameter ( $D$ ) and slot height ( $Z$ ) as shown in Fig. 1

### E. Data Set Pre-processing

The data set used to build the three models has three input variables and one target variable. Numbers of circular slots ( $N$ ), the ratio of relative head ( $H/P$ ) and  $\frac{\sqrt{A}}{Z}$  are categorized as input variables. Coefficient of discharge ( $C_d$ ) is categorized as the target variable.

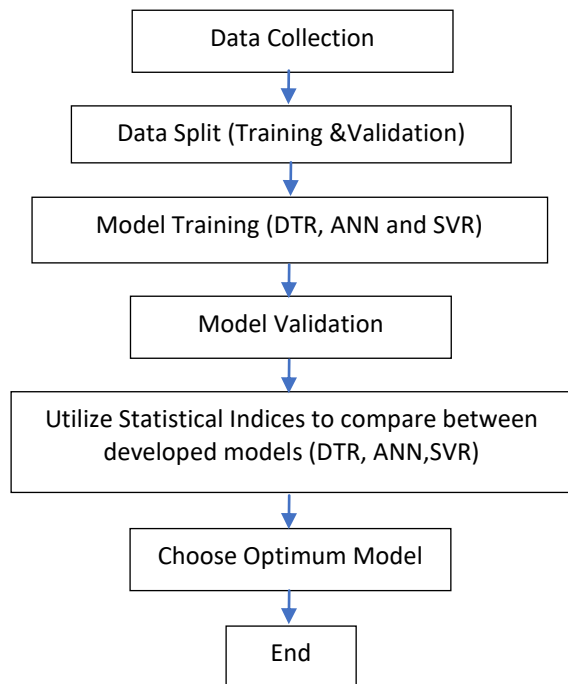


Fig. 2. Methodology and Python workflow

### III. IMPLEMENTATION

#### A. Decision Tree Regressor

Decision Tree Regressor (DTR) is one among the foremost used supervised learning algorithms for regression problems. A decision tree reaches an estimate by asking set of consecutive questions to the input data, each question tightens the possible values till the model reach accepted level of confidence to obtain one prediction. The model determines the arrangement of these question and their content. Steps to build the proposed DTR model are illustrated using flowchart shown in Fig. 3.

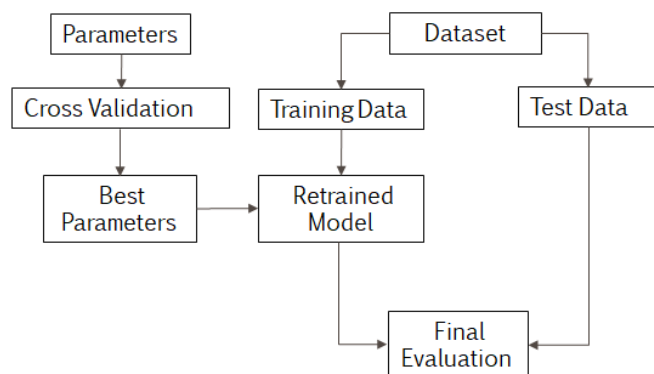


Fig. 3. DTR Flowchart

These steps have been repeated 7 times with 7 different combinations of training/validation data. 20% of the data was selected randomly as validation data using Shuffle Split function. This process is called Hyper-parameter optimization. A python sub-code was created to run optimization automatically. Fig. 4 shows  $C_d$  predicted values versus  $C_d$  measured values using DTR

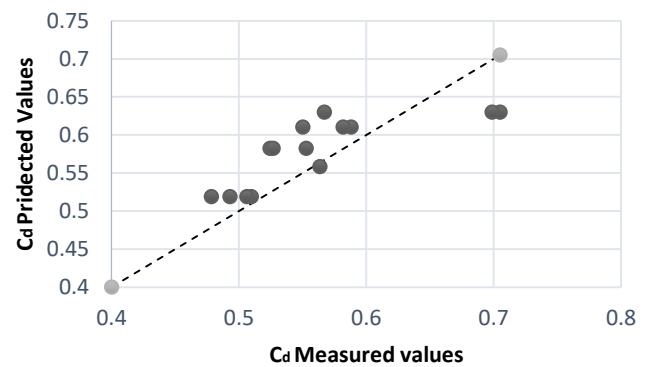


Fig. 4. DTR Model Results:  $C_d$  measured values versus  $C_d$  predicted values

#### B. Artificial Neural Network

Artificial Neural Networks (ANNs) are used efficiently nowadays as computing tools for solving complex non-linear regression problems. ANN typical structure consists of artificial linked nodes (neurons) in multiple layers. Input layer which receives input data. Output layer which produces predicted output results. In between these two layers, there are zero or more hidden layers. The optimum number of neuros and hidden layers are decided by try and error taking into consideration the time taken for training and avoiding overfitting.

In the current study, Artificial Neural Network algorithm was employed to develop a model, using python programming language, to estimate coefficient of discharge as a function of  $(\frac{H}{P}, \frac{\sqrt{A}}{Z}, N)$ . To achieve the best prediction value of the output parameter, different network structures were tried and evaluated.

Fig. 5 shows the structure of the developed ANN model. It consists of one hidden layer to avoid overfitting. The optimum number of nodes have been determined after many trials; starting by 32 nodes and ended by 7 nodes which gives the least error.

After training the network model, weights and biases are fixed. These fixed weight and bias values are shown in Equation (9). Each input value gets multiplied with the weight and adds with bias value. The total sum is the input at each hidden node and pass through an activation function as defined in Equation (7), and further the output from hidden node get multiplied with the weight and adds with the bias value and total sum pass through the activation again as shown in Equation (8).

Back propagation Neural Network (BNN) technique was employed to train proposed network such that the result outputs are nearer to the desired values. Training in BNN is divided into two phases. Phase one is a forward phase; where data is propagated from the input to the output layer. Phase Two is a backward stage where an error, defined as the difference between the measured value and the actual value in the output layer, is propagated backwards in order to modify both of weightings and bias values. In the forward phase, the weighted sum of input components,  $u_j$ , is calculated as:

$$U_j = \sum_{i=1}^n W_{ij} X_i + bias_j \quad (7)$$

where  $w_{ij}$  is the weight between the  $j_{th}$  and the  $i_{th}$  neurons in the preceding layer,  $x_i$  is the output of the  $i_{th}$  neuron in the preceding layer, and  $bias_j$  is the weight between the  $j_{th}$  neuron and the bias neuron in the preceding layer. The output of the  $j_{th}$  neuron in any layer,  $y_j$ , is given in the form:

$$y_j = f(U_j) \quad (8)$$

Where,  $f$  refers to the activation function.

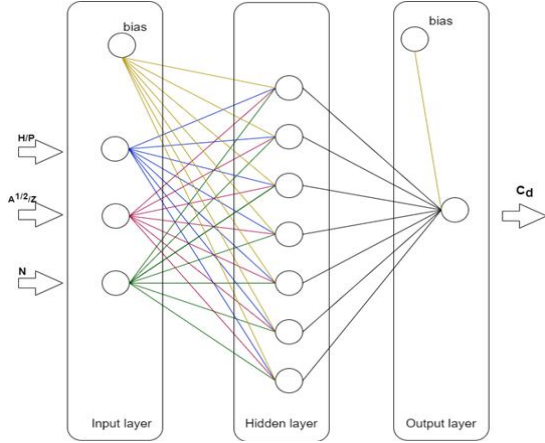


Fig. 5. ANN Structure

In current study, two activation functions; linear and ReLu (rectified linear unit) were applied for testing. It was found that linear activation function gives better results. Mean Square Error (MSE) is used in the current study to compare the performances of regression models. Fig. 6 shows  $C_d$  predicted values versus  $C_d$  measured values using ANN.

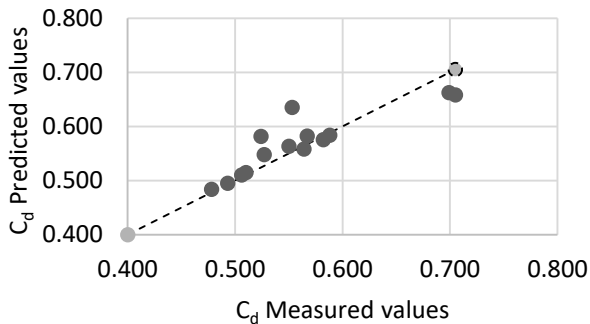


Fig. 6. ANN Model Results:  $C_d$  measured values versus  $C_d$  predicted values

Equation (9) provides trained ANN model for estimating the coefficient of discharge  $C_d$  for rectangular weir with multiple circular slots.

$$C_d = A * [B * C] \quad (9)$$

Where

$$A = [-0.4049 \quad -0.3506 \quad 0.7459 \quad 0.6655 \quad 0.2856 \quad -0.5046 \quad -0.1172 \quad 0.1903]$$

$$B = \begin{bmatrix} 0.2448 & 0.525 & 0.2803 & -0.1930 \\ -0.52 & -0.5213 & -0.1532 & -0.1871 \\ -0.3574 & 0.1707 & -0.5833 & 0.1905 \\ -0.4565 & -0.2263 & 0.3085 & 0.1908 \\ 0.1926 & 0.2519 & 0.1109 & 0.1862 \\ -0.5864 & 0.0228 & -0.3488 & -0.1875 \\ 0.6746 & -0.05235 & -0.4485 & -0.1835 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$C = \begin{bmatrix} H/P \\ \sqrt{A}/Z \\ N \\ 1 \end{bmatrix}$$

### C. Support Vector Machine Regressor

Support Vector Machines (SVM) has been presented by Vapnik [18] as a type of supervised machine learning techniques. Recently it finds broad utilization in the field of computing, environmental, and hydrology applications. SVM depends on the basis of the Statistical Learning Theory and Structural Risk Minimization, which reduces the upper limit generalization error instead off local training error which is typical in other Machine Learning methodologies. SVM algorithms depend on a set of mathematical functions which are known as the kernel. The function of kernel role is to simplify the learning process. It transforms data as input to required form.

The performance of SVM model depends upon SVM parameters: Kernel specific parameter ( $\gamma$ ), and Regularization Cost parameter ( $C$ ) which constrain model's allowance to error Behnam et al., [19].

In this study, different types of kernel functions; *Linear*, *Polynomial*, and *Gaussian Radial Basis* (RBF) Functions; were applied to choose the best fit function. Statistical indices such as; highest correlation coefficient, smallest Mean Absolute Error (MAE), and Mean-Squared Error (MSE); were employed after many trials, as shown in Table (I) to obtain the optimum values for  $C$  and  $\gamma$ . Finally, Linear Kernel function with  $C=1$  and  $\gamma=0.0001$  has been chosen to provide the best results. Fig. 7 shows  $C_d$  predicted values versus  $C_d$  measured values

Table-I: Summary of SVR Model parameters and grid values

| Model Parameters | Grid Values                 |
|------------------|-----------------------------|
| kernel           | <i>RBF, Linear, Poly</i>    |
| degree           | 3, 5, 7, 9, 11              |
| $\gamma$         | 1, 0.1, 0.01, 0.001, 0.0001 |
| $C$              | 0.1, 1, 10, .01             |

### IV. MODELS EVALUATION

In order to accurately evaluate the three developed models, statistical metrics including Correlation Coefficient ( $R^2$  Score), Mean Absolute Error (MAE) which is Absolute of the difference in distance between the measured value and the predicted value, and Mean Squared Error (MSE) have been used.

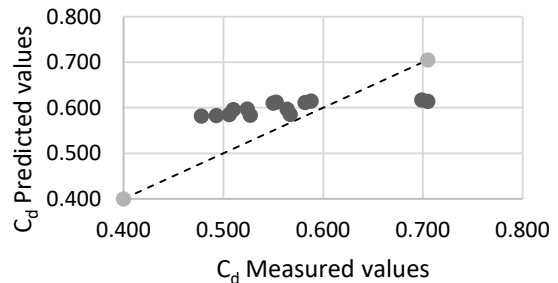


Fig. 7. SVR Model Results:  $C_d$  measured values versus  $C_d$  predicted values

$$MAE = \frac{1}{n} \sum_{i=1}^n |C_{d(predicted)i} - C_{d(measured)i}| \quad (10)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (C_{d(predicted)i} - C_{d(measured)i})^2 \quad (11)$$



$$R^2 = \frac{(n \sum_{i=1}^n C_d(\text{predicted})_i C_d(\text{measured})_i - \sum_{i=1}^n C_d(\text{predicted})_i \sum_{i=1}^n C_d(\text{measured})_i)^2}{(n \sum_{i=1}^n (C_d(\text{predicted})_i)^2 - (\sum_{i=1}^n C_d(\text{predicted})_i)^2)(n \sum_{i=1}^n (C_d(\text{measured})_i)^2 - (\sum_{i=1}^n C_d(\text{measured})_i)^2)} \quad (12)$$

Table II shows comparison for statistical metrics for developed models.

**Table-II: Statistical indices for predicted discharge coefficient by DTR, ANN and SVR**

| Metrics              | DTR   | ANN   | SVR   |
|----------------------|-------|-------|-------|
| <b>R<sup>2</sup></b> | 0.519 | 0.759 | 0.092 |
| <b>MSE</b>           | 0.040 | 0.022 | 0.064 |
| <b>MAE</b>           | 0.002 | 0.001 | 0.005 |

## V. CONCLUSION

In this study, a potential of using three Machine Learning regression algorithms; Decision Tree Regressor (DTR), Artificial Neural Networks (ANN) and Support Vector Machine Regressor (SVR) has been investigated to estimate the coefficient of discharge  $C_d$  for rectangular weir with multiple circular slots. Simulation results of the three developed models have been compared. A major conclusion is that the ANNs algorithm is the superior model to predict the coefficient of discharge  $C_d$  for rectangular weir with multiple circular slots. It gives the highest matching between measured and predicted values with correlation coefficient ( $R^2$ ) value of 0.759, minimum MAE with value 0.001 and minimum MSE with value 0.022. Finally, an equation using ANNs is introduced to estimate the discharge coefficient as a function of  $(\frac{H}{P}, \frac{\sqrt{A}}{Z}, N)$  as input variables.

## REFERENCES

- Wolters, W., Ghobrial, N. S. and Eisa, M., "Calibration of weirs in The Fayoum, Egypt", Irrigation & Drainage Systems, 19871. pp. 173-185.
- Abdel Halim N. A, Sherif MM, El-Zaher, A. S., "On the Fayoum weirs with orifices", J Eng Appl Sci, Faculty of Eng, Cairo Univ, Egypt 1991;38(5):893-904.
- Negm A. M., "Discharge prediction model for simultaneous underflow-overflow". In: Proc 6th int symp on flow modeling and turbulence measurements, Florida, USA, 8-10 September, Balkema Publishers, 1996. p.665-70.
- Negm A. M., "Characteristics of simultaneous overflow-submerged underflow (unequal contractions)". Eng Bull Faculty of Engineering, Ain Shams Univ, Cairo, Egypt 2000; 35(1):137-54.
- Negm A. M. "Experimental investigation on simultaneous flow through combined box culverts and over contracted broad-crested weirs". In: Proc 2nd int conf for advanced trends in engineering, 7-9 April, Faculty of Engineering, Al-Minia Univ, Egypt, 2002.
- Elazizy, I. M., "Hydraulic characteristics of weir with a slot", Engineering Research Journal, Helwan University, Faculty of Eng., Mataria, Cairo, Egypt; 2005. (99), C 149-C 163 .
- Hassan, M. I, Abozeid Gamal, Shehata Salah M., "Hydraulics of clear and submerged overfall weirs with bottom circular-openings". Ain Shams Univ, Eng Bull, Cairo, Egypt 2010;1:115-9.
- El-Belasy, A. M., "Developing Formulae for combined weir and orifice (case study: EL-Fayoum weirs)". Alexandria Engineering Journal, 2013. 52, 763-768
- Neveen Y. Saad, Ehab M. Fattouh, "Hydraulic characteristics of flow over weirs with circular openings". Ain Shams Engineering Journal, 2017. 8, 515-522
- Bilhan O., Emiroglu M.E., Kisi O., "Application of two different neural network techniques to lateral outflow over rectangular side weirs located on a straight channel". Advances in Engineering Software 41, 2010. 831-837.

- Dursun O.F., Kaya N., Firat M., "Estimating discharge coefficient of semi-elliptical side weir using ANFIS". Journal of Hydrology, 2012. V 426-427, Pp. 55-62
- Khoshbin F., Bonakdari H., Ashraf Talesh S.H., Ebtehaj I., Zaji A.H., Azimi H., "Adaptive neuro-fuzzy inference system multi-objective optimization using the genetic algorithm/singular value decomposition method for modelling the discharge coefficient in rectangular sharp-crested side weirs". Engineering Optimization, 2016; 48, pp. 933-948.
- Parsaie A., Haghiabi A.H., "Support Vector Machine to predict the discharge coefficient of Sharp crested w-planform weirs". AUT Journal of Civil Engineering, 2017 (1),pp. 195-204.
- Ehsan et al., "Modeling discharge coefficient of triangular plan form weirs using extreme learning machine". Journal of Applied Research in Water and Wastewater, (2019). Under press. <https://doi.org/10.22126/ARWW.2019.1094>
- Azimi et al., "Design of radial basis function-based support vector regression in predicting the discharge coefficient of a side weir in a trapezoidal channel". Applied Water Science, (2019) 9:78 <https://doi.org/10.1007/s13201-019-0961-5>
- Chanson H., "The hydraulics of open channel flow". London Arnold, 1999
- Chanson H., "The hydraulics of open channel flow". Elsevier Butterworth-Heinemann, 2004
- Vapnik V. N., "The nature of statistical learning theory". Springer, New York, 1995
- Behnam Balouchi , Gholamreza Rakhshandehroo, "Using Physical and Soft Computing Models to Evaluate Discharge Coefficient for Combined Weir-Gate Structures Under Free Flow Conditions". Iran J Sci Technol Trans Civ Eng, (2018) 42:427-438

## AUTHORS PROFILE



**Dr. Mona A. Hagras** is Associate Professor at Irrigation & Hydraulics Department, Faculty of Engineering, Ain Shams University, Cairo, Egypt. She received a PhD and a MSc from the University of Ain Shams. She supervised several theses and published several articles in national and international cited journals. She is a peer reviewer at many international scientific journals. Her areas of interest include Water Resources Management, Hydrology, Hydraulics, and Groundwater modeling.