HARRAN ÜNİVERSİTESİ MÜHENDİSLİK DERGİSİ

HARRAN UNIVERSITY JOURNAL of ENGINEERING

e-ISSN: 2528-8733 (ONLINE)

Practical Design of Stepped Spillways Using Machine Learning Methods and Fuzzy Inference System

Makine Öğrenme Yöntemleri ve Bulanık Çıkarım Sistemi Kullanılarak Basamaklı Dolusavakların Pratik Tasarımı

Yazar(lar) (Author(s)): *Sadık ALASHAN[1], Sedat GOLGİYAZ[2], Erdinç İKİNCİOĞULLARI[3], Eyyüp Ensar YALÇIN[4]*

[1] ORCID ID: 0000-0003-1769-4590
[2] ORCID ID: 0000-0003-0305-9713
[3] ORCID ID: 0000-0003-2518-980X
[4] ORCID ID: 0000-0001-9446-2991

Bu makaleye şu şekilde atıfta bulunabilirsiniz (To cite to this article): Alashan S., Golgiyaz S., İkincioğulları E., Yalçın E. E., "Practical Design of Stepped Spillways Using Machine Learning Methods and Fuzzy Inference System", ***Harran Üniversitesi Mühendislik Dergisi***, 10(1): 36-50, (2025).

DOI: 10.46578/humder.1638527

HRU Muh Der, 10(1): 36-50 (2025)

Harran Üniversitesi Mühendislik Dergisi

Harran University Journal of Engineering

https://dergipark.org.tr/tr/pub/humder

Araştırma Makalesi (Research Article)

Practical Design of Stepped Spillways Using Machine Learning Methods and Fuzzy Inference System

Sadık ALASHAN[1], Sedat GOLGİYAZ[2], Erdinç İKİNCİOĞULLARI[1*], Eyyüp Ensar YALÇIN[1]

[1]*İnşaat Mühendisliği Bölümü, Mühendislik ve Mimarlık Fakültesi, Bingöl Üniversitesi, Bingöl, Türkiye*

[2]*Bilgisayar Mühendisliği Bölümü, Mühendislik ve Mimarlık Fakültesi, Bingöl Üniversitesi, Bingöl, Türkiye*

Article Information

Received: 12/02/2025
Revised: 04/03/2025
Accepted: 10/03/2025
Published: 31/03/2025

Citation

Alashan S., Golgiyaz S., İkincioğulları E., Yalçın E. E., “Practical Design of Stepped Spillways Using Machine Learning Methods and Fuzzy Inference System”, Harran University Journal of Engineering, 10(1): 36-50, (2025).

Abstract

Energy-dissipating pools or flip bucket structures reduce the energy of downstream flow in conventional spillways. Recently, stepped spillways have been widely used to dissipate the flow of energy downstream. Flows on the stepped spillways are complex and advanced techniques such as Fuzzy Logic (FL), Adaptive Neuro-Fuzzy Inference System (ANFIS), Artificial Neural Network (ANN), Genetic Programming (GP), Deep Learning, and Tree-Based models are required to calculate energy dissipation ratios. Fuzzy Logic has the advantage of considering physical processes when examining problems using rule bases. In this study, energy dissipation over stepped spillways is calculated using machine learning methods and the Fuzzy Inference System in Python programming language. Experimental data by different researchers are used to model stepped spillways. Two new parameters, such as an approach channel and step-top geometric ratios, are used in addition to the literature to obtain energy dissipation ratios on stepped spillways. Artificial Neural Network Regressor (ANN) from machine learning methods gives minimum percentages and absolute errors (-0.117% and 1.398) and maximum R^2 values (0.976) for the testing dataset. Although the accuracy of the ANN method changes with hidden layer sizes and ratios between training and testing data, the Fuzzy Logic (FL) is independent to training data. The FL method represents good results with low Mean Percentages Error (MPE) and Mean Absolute Errors (MAE) (-1.688% and 2.000) and an R^2 value (0.951), and the produced Python function using the fuzzy inference system can be applied easily to different flow conditions and stepped spillways.

Keywords: *Stepped spillways, energy dissipation, skfuzzy, fuzzy logic, artificial neural network regressor, machine learning methods.*

Makine Öğrenme Yöntemleri ve Bulanık Çıkarım Sistemi Kullanılarak Basamaklı Dolusavakların Pratik Tasarımı

Makale Bilgisi

Başvuru: 12/02/2025
Düzeltme: 04/03/2025
Kabul: 10/03/2025
Yayınlanma: 31/03/2025

Alıntı

Alashan S., Golgiyaz S., İkincioğulları E., Yalçın E. E., “Practical Design of Stepped Spillways Using Machine Learning Methods and Fuzzy Inference System”, Harran Üniversitesi Mühendislik Dergisi, 10(1): 36-50, (2025).

Öz

Enerji sönümleyici havuzlar veya sıçratma eşiği yapıları, klasik dolusavakların mansabındaki akım enerjisini sönümlemektedir. Son yıllarda, akım enerjisini sönümlemek amacıyla basamaklı dolusavaklar yaygın olarak kullanılmaktadır. Basamaklı dolusavaklardaki akım karmaşık olup enerji sönümleme oranlarını hesaplamak için ileri tekniklere sahip Bulanık Mantık (BM), Yapay Sinir Ağları Uyarlamalı - Bulanık Çıkarım Sistemi (YSA- BÇS), Yapay Sinir Ağları (YSA), Genetik Programlama (GP), Derin Öğrenme ve Ağaç Tabanlı modeller kullanılmaktadır. BM, kural tabanları kullanarak problemleri incelerken fiziksel süreçleri göz önünde bulundurabilme avantajına sahiptir. Bu çalışmada, Python programlama dilinde makine öğrenmesi yöntemleri ve Bulanık Çıkarım Sistemi kullanılarak basamaklı dolusavakların enerji sönümleme oranları hesaplanmıştır. Basamaklı dolusavakları modellemek için farklı araştırmacıların deneysel verileri kullanılmıştır. Basamaklı dolusavaklarda enerji sönümleme oranlarını tahmin edebilmek için literatürde kullanılan parametrelere ek olarak yaklaşım kanalı ve basamak üstü geometrik oranları gibi iki yeni parametre kullanılmıştır. Makine öğrenmesi yöntemlerinden YSA ile test veri seti için minimum yüzde ve mutlak hata (% -0.117 ve 1.398) ve maksimum R^2 değerleri (0.976) elde edilmiştir. YSA yönteminin doğruluğu gizli katman boyutları ve eğitim-test veri oranlarıyla değişse de Bulanık Mantık (BM) yönteminin sonuçları eğitim verilerinden bağımsızdır. BM yönteminde düşük ortalama yüzdelik ve mutlak hatalar (% -1.688 ve 2.000) ve 0.951 R^2 değeri ile iyi sonuçlar elde edilmiştir ve BM yöntemi kullanılarak üretilen Python fonksiyonu farklı akış koşullarına ve basamaklı dolusavaklara kolayca uygulanabilme imkânına sahiptir.

Anahtar Kelimeler: *Basamaklı dolusavak, Enerji sönümleme, skfuzzy, Bulanık mantık, yapay sinir ağı regresörü, makine öğrenmesi yöntemleri.*

*Corresponding author, e-mail: eikinciogullari@bingol.edu.tr
DOI: 10.46578/humder.1638527

1. INTRODUCTION

Stepped spillways have been used for over 3500 years because it is easy to design and build them [1]. The Akarnian stepped spillway, claimed to be the world's earliest example of a stepped spillway, was built about 1300 BC, according to the literature [1]. Since the beginning of the 20th century, stepped spillways have begun to be constructed with more demand to dissipate the flow energy [2].

The flow energy is dissipated downstream of traditional spillways using energy-dissipating pools or flip bucket structures. Stepped spillways have recently become popular for dissipating the flow energy downstream of small dams [3]. The flow energy can be dissipated on the stepped channel about 70–80% more energy than other spillways [4]. Therefore, it is emphasized that the stepped spillways are more cost-effective than the traditional ones since the size of the energy-dissipating structures installed downstream is decreased by 30–40% [3,5]. The high energy dissipation on the stepped spillway also decreases the flow momentum. Hence, the probability of cavitation decreases, and the cavitation index rises [6]. Stepped spillways are also frequently utilized in water treatment facilities to improve the water's oxygen quantity [7]. Stepped spillways are frequently constructed in roller compacted concrete (RCC) dams with a unit flow rate of up to q=10–15 m^3/sm. Optimizing energy dissipation and extending the structure's active life are the main reasons of constructing stepped spillways [3]. Additionally, stepped spillways reduce ground cavitation in loose ground regions to their high energy dissipation capability. This high energy dissipation capability enables making necessary energy reduction in lower structure lengths. This situation is an advantage in terms of economics, as well as a selection criterion for construction of dams with location problems. Floods affecting spillway design have occurred more frequently in recent years with higher discharges due to the effects of climate change. This necessitates the dissipation of higher discharges over shorter distances and increases the demand for stepped spillways.

As a result of the widespread use of stepped spillways, many researchers have conducted intensive experimental studies on the subject [8–20]. Felder et al. [14] used three models besides the traditional stepped spillway model to assess the aeration rates and energy dissipation capabilities. The studies were carried out for discharges between 0.02 and 0.155 m^3/s. The channel width is 0.52 m, the width of the steps is 0.20 m, and the height of the steps is 0.10 m. For pooled models, thresholds with dimensions of 0.031 m high and 0.015 m wide were put at the ends of the stairs. The results show that the pooled stepped spillway models dissipate energy more slowly than the traditional stepped spillways. Felder et al. [13] used three stepped spillway models: classical, pooled, and combined. In an experimental set with an 8.9° channel angle, a 0.50 m channel width, a 0.318 m step width, and a 0.05 m step height, the researchers conducted 21 experiments with discharges between 0.02 and 0.117 m^3/s. They discovered that the combination model had the highest energy dissipation rate while the conventional stepped spillways had the lowest.
The numerical studies have also intensified with the development of computer technology [21–32]. Meanwhile, researchers in various fields of science and engineering have recently been interested in artificial intelligence (AI) approaches, including Fuzzy Logic, ANFIS, ANN, GP, and Deep Learning models, because these methods can correlate large and complicated datasets without previously knowing how they are related [33,34].

Salmasi and Ozger [35] examined the applicability and accuracy of the ANFIS technique in predicting the correct values of energy dissipation of the skimming flow regime across stepped spillways. The recommended ANFIS model's training and testing determination coefficients were 0.974 and 0.966, respectively. They emphasized that the formulation of accounting for the energy dissipation over stepped spillways using this technique is more effective than regression equations.

Roushangar et al. [36] used the neural computing approach capability to evaluate the energy dissipation performance of the stepped spillway under the nappe flow regime. They utilized ANFIS and Feed-Forward Neural Network (FFNN) techniques to determine the most influential parameters. The findings show that neural computing-based approaches consistently predict energy dissipation over stepped spillways under a nappe flow regime. Also, the ANFIS model outperformed the FFNN model in terms of overall performance. Nevertheless, the essential depth, height, and number of steps were shown to be the variables that influence energy dissipation the most.

Mojtahedi et al. [37] conducted a comprehensive study about the energy dissipation performance of stepped chutes using experimental, numerical, and FL methods. They stated that the prediction accuracy of the FL data was evaluated using the mean absolute percentage error criterion (%7.8). The FL model can be used as an efficient tool for the design and operational control of stepped spillways.

In this study, energy dissipation ratios on stepped spillways are calculated using a FL system with the skfuzzy toolbox, as well as Machine Learning (ML) and Linear Regression (LR) methods in the Python programming language. Experimental results from different researchers are used as the study datasets. The datasets are divided into two sub-series, training and testing datasets, for the ML and LR methods; however, FL does not require training datasets. The FL method uses a rule-based approach instead of training datasets. A function (endistepway) with low error ratios (mean percentage error, -1.688%, and mean absolute error, 2.000) and a high determination coefficient (0.951) has been developed for implementing the FL method in stepped spillways. The function enables easy calculation of energy dissipation ratios in stepped spillways. Also, it employs two new parameters - the approach channel and top step geometric ratio - that have never been used in the literature, along with the step height ratio and spillway angle parameters.

2. MATERIAL and METHOD

2.1 Geometrical Model and Data

Three laboratory test results [13, 14, 38] of the flat-stepped spillway were used to compare the results of the ML, LR and FL methods. The sketch of the flat stepped model is shown in Fig. 1. In here, H is the height of the spillway, L is the length of the spillway, L_a is the length of the approach channel, D_c is the critical flow depth, D is the flow depth on the last step, L_s is the step length, H_s is the step height, and W_s is the channel width.

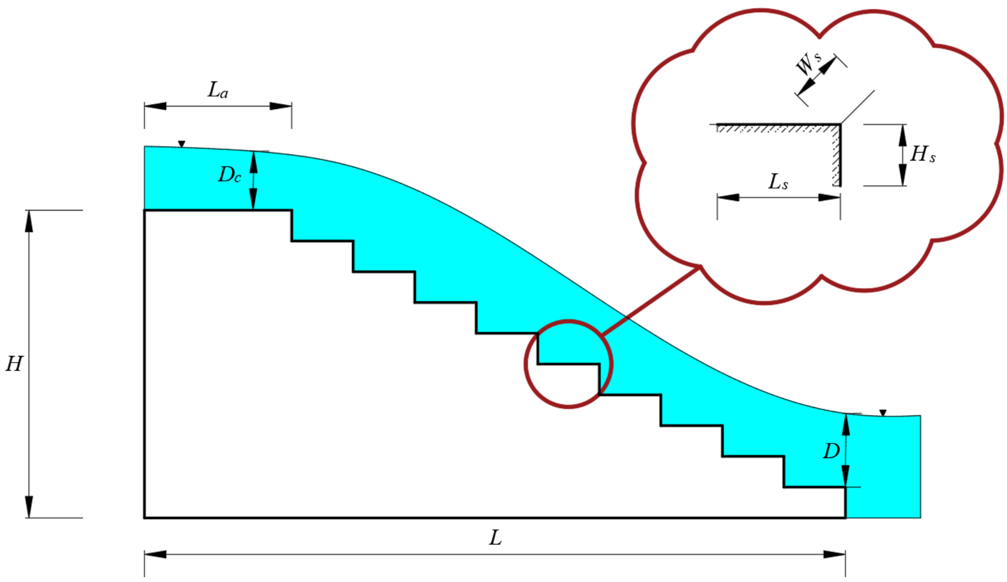

Figure 1. A sketch of the longitudinal section of stepped spillways

2.2 Dimensional Analysis for Stepped Spillways

Considering the specific design of the spillway and the extensive research carried out by a number of experts, the following parameters have been identified as the most important factors influencing the flow over stepped spillways:

The fluid's characteristics encompass essential parameters such as dynamic viscosity (μ), mass density (ρ), and gravitational acceleration (g).

Table 1. The Values of Discharges and Desing Parameters of Stepped Spillways

	Discharge (Q) (m³/s)	**Unit discharge (q) (m³ s⁻¹ m⁻¹)**	**W_s (m)**	**H (m)**	**L_a (m)**	**Chute angle (α) (°)**	**Step number**
Felder et al. [13]	0.030	0.058	0.52	1.00	1.01	26.6	10
	0.038	0.073	0.52	1.00	1.01	26.6	10
	0.049	0.094	0.52	1.00	1.01	26.6	10
	0.063	0.121	0.52	1.00	1.01	26.6	10
	0.075	0.144	0.52	1.00	1.01	26.6	10
	0.090	0.173	0.52	1.00	1.01	26.6	10
	0.097	0.187	0.52	1.00	1.01	26.6	10
	0.113	0.217	0.52	1.00	1.01	26.6	10
Felder et al. [14]	0.018	0.036	0.5	1.05	0.318	8.9	21
	0.027	0.054	0.5	1.05	0.318	8.9	21
	0.039	0.078	0.5	1.05	0.318	8.9	21
	0.049	0.098	0.5	1.05	0.318	8.9	21
	0.061	0.122	0.5	1.05	0.318	8.9	21
	0.076	0.152	0.5	1.05	0.318	8.9	21
	0.091	0.182	0.5	1.05	0.318	8.9	21
	0.105	0.210	0.5	1.05	0.318	8.9	21
	0.117	0.234	0.5	1.05	0.318	8.9	21
Irzooki et al. [38]	0.00183	0.00600	0.305	0.25	0.05	26.6	5
	0.00279	0.00914	0.305	0.25	0.05	26.6	5
	0.00607	0.01992	0.305	0.25	0.05	26.6	5
	0.00642	0.02105	0.305	0.25	0.05	26.6	5
	0.00737	0.02416	0.305	0.25	0.05	26.6	5
	0.00202	0.00661	0.305	0.25	0.05	45	5
	0.00210	0.00688	0.305	0.25	0.05	45	5
	0.00568	0.01862	0.305	0.25	0.05	45	5
	0.00665	0.02180	0.305	0.25	0.05	45	5
	0.00808	0.02650	0.305	0.25	0.05	45	5
	0.00065	0.00214	0.305	0.25	0.025	26.6	10
	0.00104	0.00342	0.305	0.25	0.025	26.6	10
	0.00320	0.01050	0.305	0.25	0.025	26.6	10
	0.00365	0.01198	0.305	0.25	0.025	26.6	10
	0.00750	0.02459	0.305	0.25	0.025	26.6	10
	0.00039	0.00128	0.305	0.25	0.025	45	10
	0.00084	0.00276	0.305	0.25	0.025	45	10
	0.00347	0.01137	0.305	0.25	0.025	45	10
	0.00439	0.01440	0.305	0.25	0.025	45	10
	0.00759	0.02489	0.305	0.25	0.025	45	10

The hydraulic characteristics of the flow include critical parameters such as the depth of the flow (D) and the velocity of the flow (V).

The shape and geometry characteristics of the spillway include key parameters such as step height (H_s), step length (L_s), length of the approach channel (L_a), number of steps (N_s), channel width (W_s) and spillway height (H).

Thus, the dimensional analysis for the stepped spillway is expressed as a function of the parameters $f(H, W_s, N_s, H_s, L_s, L_a, V, D, g, \rho, \mu) = 0$. Within the framework of the π-Buckingham theory, the variables ρ, V and D_c were chosen as repetitive variables. From this analysis the Eq. 1 was derived:

$$\frac{\Delta E}{E_0} = f\left(\frac{H}{D_c}, \frac{W_s}{D_c}, \frac{H_s}{D_c}, \frac{L_s}{D_c}, \frac{L_a}{D_c}, Fr, Re, N_s\right) \tag{1}$$

According to the laboratory studies [13, 14, 38], the difference in the total energy (E_L) was calculated as Eq. (2). In here, E_0 and E_1 are the upstream and downstream energies on the stepped spillway, respectively. Then, the energy dissipation ratio (ΔE) was calculated as shown in Eq. (3). The parameters of the laboratory studies used in this study are given in Table 1.

$$E_L = E_0 - E_1 \tag{2}$$

$$\Delta E = \frac{E_L}{E_0} \tag{3}$$

2.3 Random Forest Regressor

Random Forest (RF) regressor is a popular machine learning algorithm widely employed in regression tasks due to its robustness and effectiveness [39]. RF is an ensemble method formed by multiple decision trees. Each decision tree is constructed using randomly selected features and trained on a subset of the dataset. The creation of these subsets and the way each tree is built introduces randomness to the algorithm. RF aggregates the predictions of each tree to make a more generalizable and robust prediction. This method is resistant to overfitting and often outperforms a single decision tree. Therefore, RF yields effective results even when applied to datasets with complex feature interactions and high dimensionality.

2.4 XGB Regressor

XGBoost (XGB) regressor stands as a prominent algorithm in the realm of gradient boosting, and it is celebrated for its exceptional performance in regression tasks [40]. XGB operates by sequentially adding weak learners, usually decision trees, to form a robust ensemble model. It achieves this by minimizing the gradient of the loss function at each iteration, thus correcting the errors made by the preceding trees. Additionally, XGB employs a weighted sum aggregation method during the ensemble construction, effectively balancing the contribution of each tree while minimizing the overall model error. This meticulous approach allows XGB to capture intricate relationships within the data and produce highly accurate predictions even in the presence of noise or outliers.

2.5 K-Nearest Neighbors Regressor

K-Nearest Neighbors (KNN) regressor stands as a fundamental and intuitive algorithm in the domain of machine learning, particularly in regression tasks [41]. Unlike parametric models, the KNN regressor makes predictions by computing the distance between the query instance and each data point in the feature space. Typically, the Euclidean or Manhattan distance metrics are utilized for this purpose. Subsequently, KNN identifies the k nearest neighbors to the query instance based on these distances. The prediction for the query instance is then calculated as the average (or weighted average) of the target values of these k neighbors. Notably, KNN does not assume any underlying functional form of data, making it particularly useful when the relationship between features and target variables is complex or nonlinear. However, it is essential to choose an appropriate value for the hyper parameter k, as selecting a too small or too large value may lead to suboptimal predictions. Despite its simplicity and interpretability, KNN may suffer from computational inefficiency when dealing with large datasets. Overall, KNN serves as a versatile tool in regression tasks, offering a balance between simplicity and performance.

2.6 Light Gradient Boosting Machine Regressor

Light Gradient Boosting Machine (LGBM or LightGBM) is an algorithm that has gathered significant attention in the machine learning community in recent years. As one of the gradient boosting methods, LGBM stands out for its ability to provide fast training and high performance, especially on large datasets. LGBM is based on the Gradient Boosting Framework, which is a tree-based ensemble method [42]. However, the one of standout features of LGBM is its ability to work faster and more efficiently compared to other GBM implementations. This capability significantly reduces training times for large datasets, making it an ideal choice for large-scale machine learning projects. Another key feature that enhances the performance of LGBM is its ability to automatically capture interactions between features. This allows the model to learn more complex relationships and improve prediction accuracy. Additionally, LGBM offers better memory usage and scalability, enabling it to effectively handle large datasets. Therefore, LGBM is considered a powerful tool for solving regression problems in large datasets.

2.7 Artificial Neural Network Regressor

Artificial Neural Network (ANN) regressor is a powerful algorithm widely used in the field of ML. ANN are designed based on biological neural networks and form the foundation of deep learning techniques used to learn complex relationships [43]. The model is trained on an ANN which consisting of multiple layers, with each layer forming a network of interconnected nodes. Each node produces an output by weighted combinations of input features and then transforms this output using a function called an activation function. The training process is carried out to minimize the error between the actual output and the predicted output of the model. This is typically done using a loss function, and optimization techniques such as backpropagation algorithm are used to update the weights of the network. The ability of ANN to adapt to various data types and problem domains makes it a versatile tool in regression tasks. With its capacity to learn intricate functions and relationships, the ANN excels in capturing nonlinear patterns within datasets, offering valuable insights for predictive modeling.

2.8 Linear Regressor

Linear Regressor (LR) is a regression method that primarily predicts the output variable by utilizing a linear combination of input features. This model typically employs the least squares method to calculate regression coefficients. Regression coefficients represent the weighted sum of variables for each feature, indicating their contributions to changes in the output variable. Despite being a simple model, LR is robust in explaining and interpreting relationships among variables in the dataset, making it particularly preferred for data analysis and exploratory studies. However, this model requires the assumption of linearity in relationships within the dataset and has limitations in modeling complex, nonlinear relationships. In such cases, more complex models or feature engineering techniques may be employed. Nonetheless, LRs advantages, including high interpretability, fast training, and low computational cost, make it a preferred choice in many problem scenarios.

2.9 Fuzzy Logic Model

FL enables the examination of a complex phenomenon by using IF and THEN logical rule base (Fig. 2) [44]. Also, verbal interpretations in FL are used to separate fuzzy sets into subsets such as “Low”, “Medium”, and “High”, thus make easier to explain a nonlinear complex phenomenon. Spillways discharge floodwater from rivers downstream through very steep chute channels. This means that very large potential energy is converted into kinetic energy, resulting in complex air-water mixed flows with very high velocities in nappe, transition, and skimming flow conditions. In FL models, the complex flow is examined by input and output variables. Discharges, step heights, widths, lengths, numbers, and approach channel and spillway lengths are the main factors that form the flow in stepped spillways. According to dimension analysis, dimensionless variables (ratios) can be obtained to examine this complex flow situation and generalize solutions. The parameters are investigated to calculate the energy dissipation ratios(E_L/E_0) in stepped spillways include the approach channel ratio (L_a/L; approach channel length to spillway length), step-top geometric ratio (L_S/W_S; step length to step width), spillway angle ($H/n{\cdot}L_S$; spillway height to cumulative step length) and step height ratio (D_c/H_s; critical depth to step height).

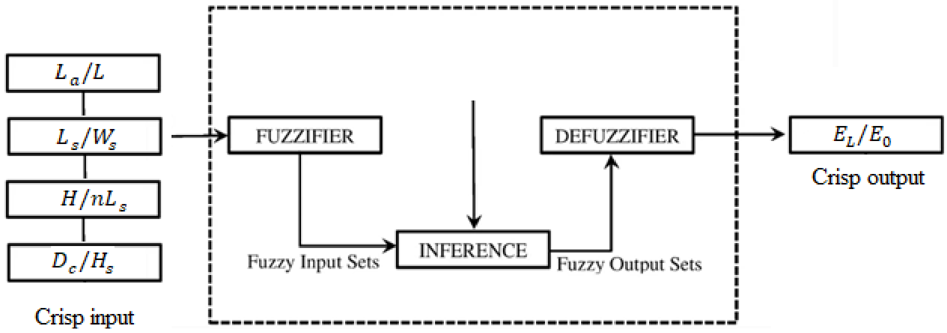

Figure 2. A schematic representation of Fuzzy Inference System [45].

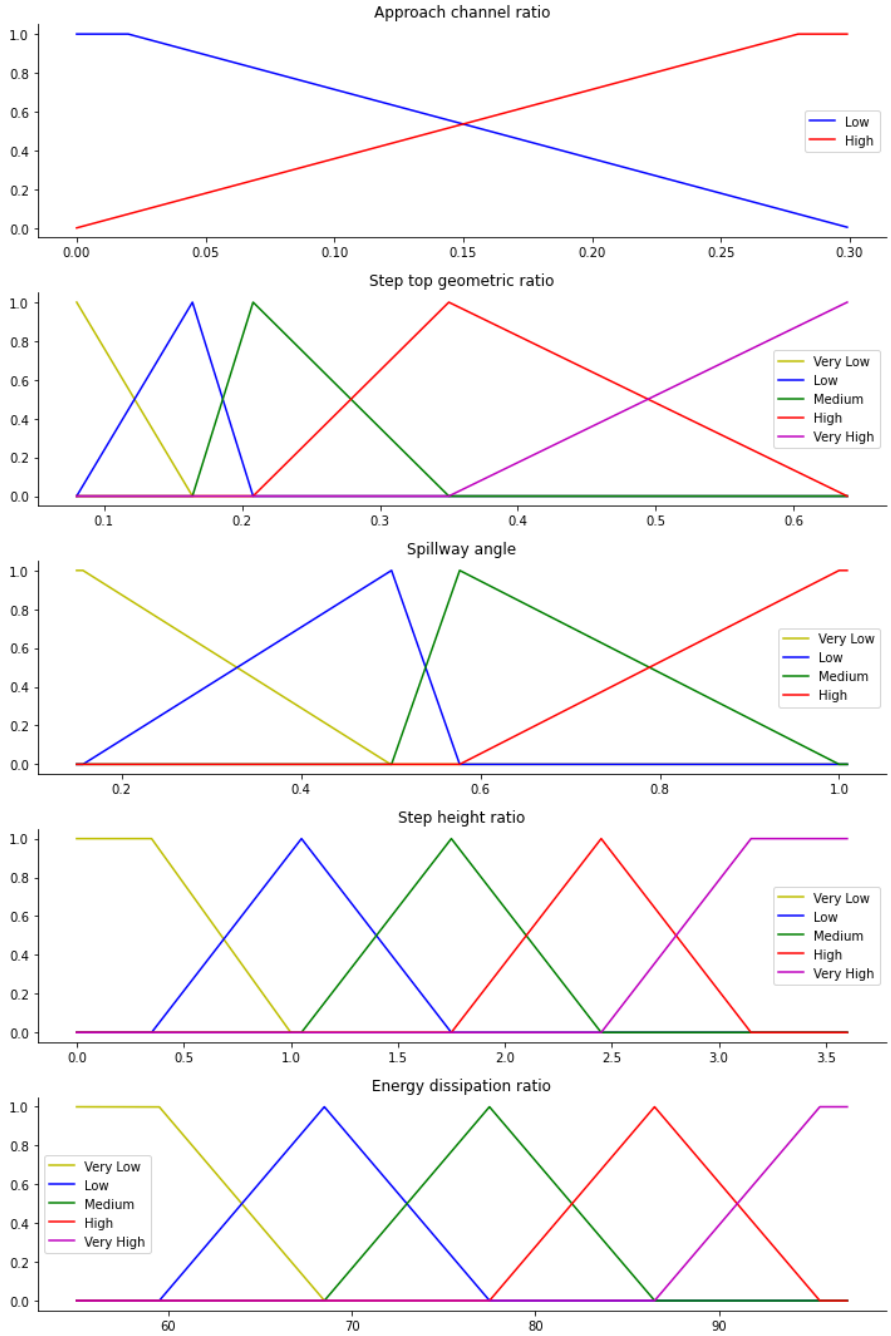

Figure 3. Membership functions of the parameters for design of stepped spillways.

The most important parameter frequently used in the literature to calculate energy dissipation is the step height ratio (D_c/H_s). The step height ratio to critical depth (step height ratio), step length to step width (step top geometric ratio), and energy dissipation values are categorized into five subsets: very low, low, medium, high, and very high (Figure 3). Additionally, the ratio of the spillway height to cumulative step length (spillway angle) is classified into four subsets: very low, low, medium, and high. Lastly, the ratio of the approach channel to spillway length (approach channel ratio) is divided into only two subsets: low and high.

Spyder, a scientific Python development environment, is used with the "skfuzzy" FL toolbox to model energy dissipation on stepped spillways in fuzzy inference system (Appendix). Membership functions for related parameters are given in Figure 3. Trapezoidal and triangular membership functions are used to represent subsets and relations among them. Limit values and membership functions (triangular and trapeozidal) for these subsets are seen in Figure 3.

A fuzzy rule base determines the relationship between 4 input parameters and 1 output parameter. The rule base has 17 rules, of which 7 trigger the very low energy dissipation subset, 6 the medium energy dissipation subset, 2 the low energy dissipation subset, and 2 the very low energy dissipation subset.

3. RESULTS and DISCUSSIONS

In this study, energy dissipation ratios have been predicted with the ratios of the approach channel to spillway length ($\boldsymbol{L_a/L}$), step length to step width ($\boldsymbol{L_S/W_S}$), spillway height to cumulative step length ($\boldsymbol{H/n{\cdot}L_S}$), and critical depth to step height ($\boldsymbol{D_c/H_s}$). Different datasets areas obtained from empirical studies conducted by different researchers in literature.

RF, XGB, KNN, LGBM, ANN, and LR machine learning methods are used to search optimum parameters which can represent best models. The splitting rates between the training and test datasets significantly change the model performances (Figure 4). MAE, MPE, and R^2 values are obtained to test the models' accuracies using 20% testing and 80% training datasets (Figures 5 and 6).

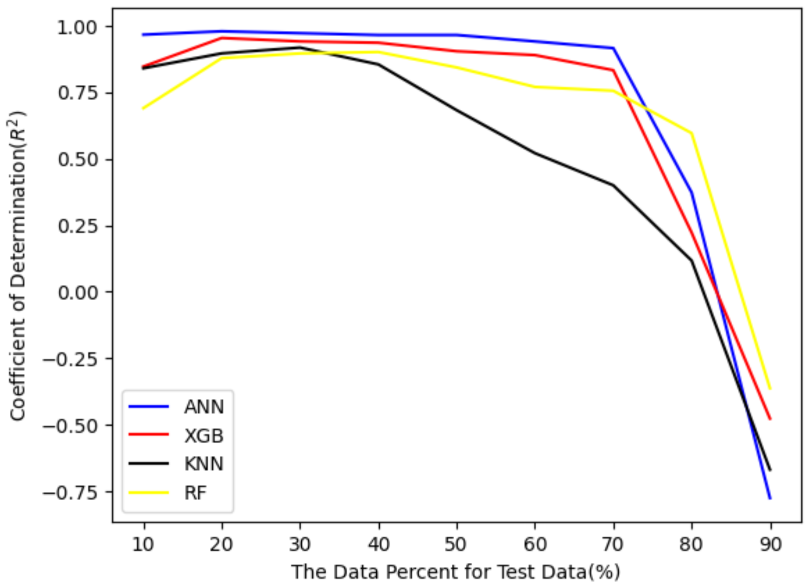

Figure 4. Model performances for different test data rates.

For the decision tree LGBM model, optimum parameters have been calculated as n_estimators = 55, boosting_type = gbdt, learning_rate = 0.35, num_leaves = 23, feature_fraction = 0.8, bagging_fraction = 0.85, max_depth = 2, and min_data_in_leaf = 1. The training and testing dataset give determination coefficients 0.978, and 0.897 respectively. Also, the method has a MAE of 0.87 in the training and 2.91 in the testing dataset. MPE are -0.039% for the training dataset and -2.562% for the testing dataset. The training dataset has higher error rates than the testing dataset, but the error rates remain below 5%. Although the determination coefficient ($\boldsymbol{R^2}$ of testing data (0.897) is good it is significantly lower than that of training dataset This indicates that the method has a low degree of overfitting. The models with high learning capacity and low complexity represent more successful results.

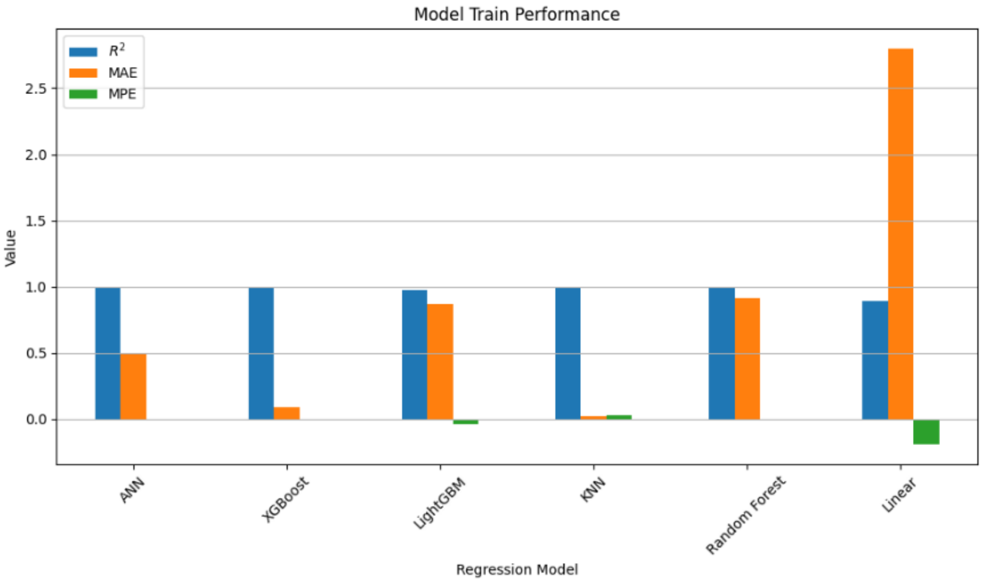

Figure 5. The training data MAE, MPE, and R^2 values obtained from the regression models.

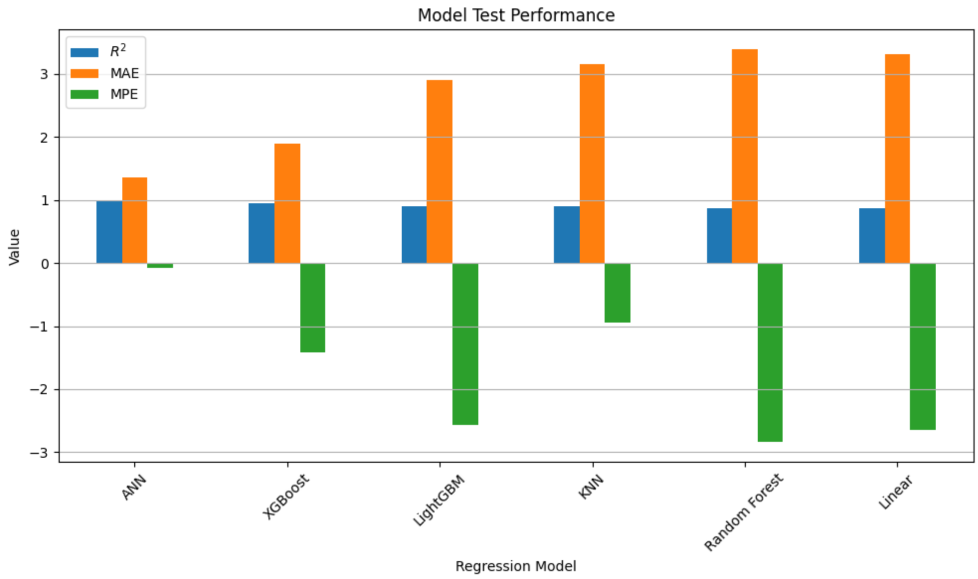

Figure 6. The testing data MAE, MPE, and R^2 values obtained from the regression models.

For the KNN model, the lowest neighbors number gives the best results in searching parameters. For example, n_neighbors=1, 2, and 3 provide 0.999, 0.98, and 0.92 determination coefficients in the training dataset. Similarly, the determination coefficients in the testing dataset are 0.895, 0.78, and 0.65 for the n_neighbors=1, 2, and 3 respectively. The MAE and MPE values in the training dataset are 0.024 and 0.029%, whereas in the testing dataset, these values are 3.164 and -0.943% respectively. As seen in the LGBM, the KNN model has a good R^2 value of 0.895 in the testing dataset, but it (0.999) is very high in the training dataset.

For the RF model, the maximum depth parameter (max_depth) has been obtained as 7 in optimum parameter searching. The training and testing dataset have 0.987 and 0.864 determination coefficients for this value. The MAE values are 0.933 and 3.500 for the training and testing datasets, respectively. As for the MPE, it is -0.182% for the training dataset and 3.048% for the testing dataset. As seen in LGBM and KNN, the error of the method increases significantly in the testing dataset, however these errors remain below 5%. The R^2 values of the testing dataset (0.999) is lower significantly than that of the training dataset (0.895).

For the ANN model, the optimum parameters are detected as hidden_layer_sizes = 75, activation = 'logistic', solver= 'adam', max_iter= 19000, and n_iter_no_change = 500. As the hidden layer sizes increase, the performance of the model also increases (Figure 7). The model's performance increases with its complexity as different from the LGBM model. The training and testing R^2 values are 0.99 and 0.98, respectively. The method has 0.646 MAE, -0.015% MPE in the training dataset and 1.398 MAE, -0.117% MPE in the testing dataset. The errors of the method do not show a significant difference compared to the used ML models. The model is robust, considering that the training and testing performances are similar.

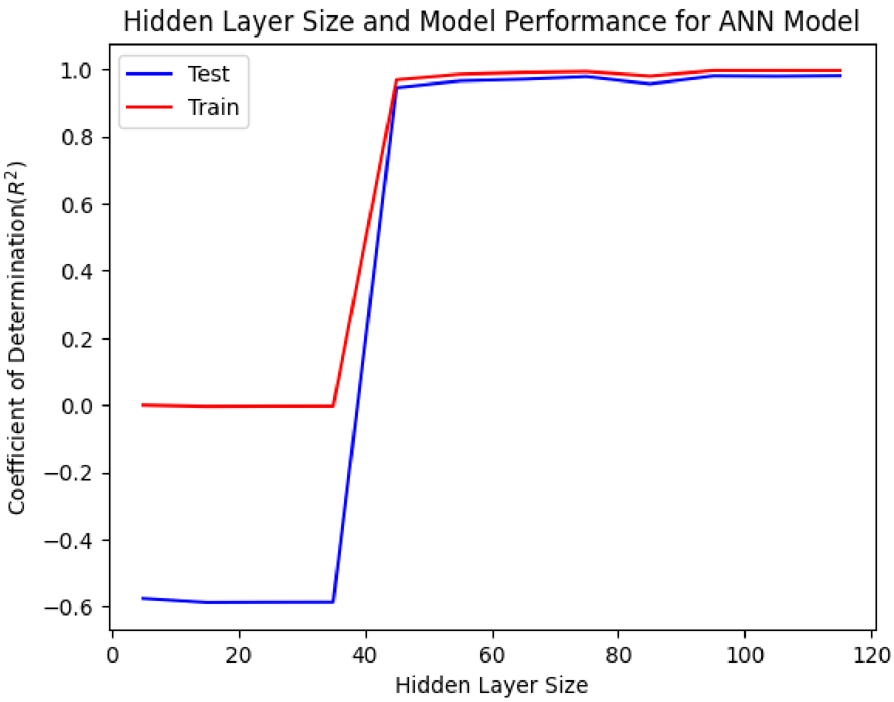

Figure 7. Variations of the performance of the ANN model with hidden layer sizes.

The XGB model gives the best results in low depths as the LGBM model. Optimum parameters for this model have been estimated as max_depth = 2 and n_estimators = 315. The XGB model has high performance with 0.999 and 0.953 R^2 for the training and testing dataset performances. The model is robust for high estimator numbers in the aspect of R^2 (Figure 8). The MAE and MPE is calculated as 0.094 and -0.001% in the training dataset and 1.903 and -1.41% for the testing dataset. The errors in the testing dataset are significantly higher than those of the training dataset, however, it remains below 5%. The R^2 value of the testing dataset is importantly lower than that of the training dataset; however, they are at good levels (0.999 and 0.953).

For the LR model, the training and testing performances are similar. The R^2 of the training and testing datasets are 0.89 and 0.86. Although the model's accuracy is lower than the other models, similar results for the training and testing performances show that the model is robust. The training dataset has a 2.797 MAE and -0.190% MPE, while the testing dataset has values of 3.315 MAE -2.65% MPE. The method shows a significant difference only in the MPE values, while the MPE and R^2 values are approaching each other in the training and testing datasets.

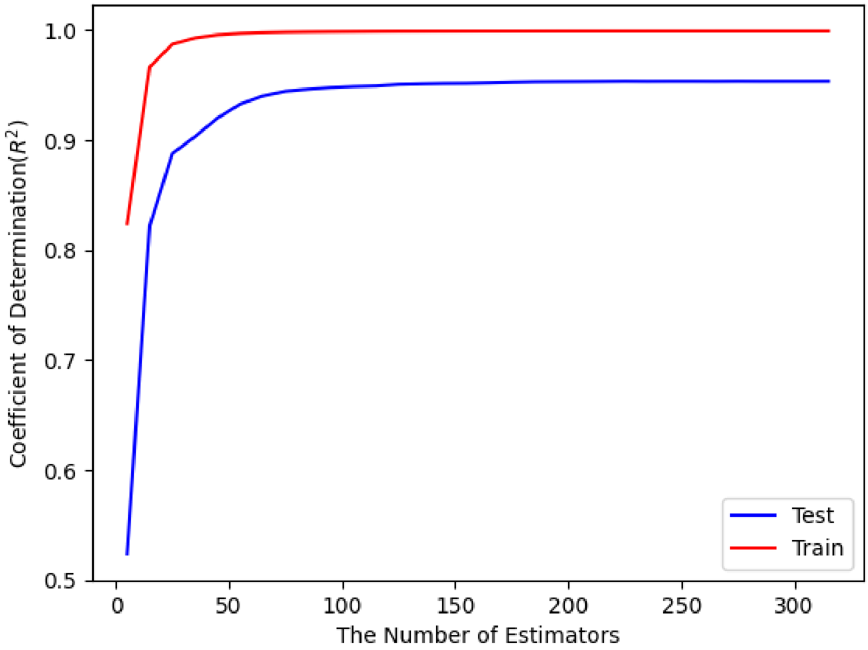

Figure 8. Variations of the number of estimators and model performance for XGB model.

Additionally, the XGB model has been executed to identify important levels in the input parameters. The step height ratio parameter, defined as the critical depth to step height, (D_c/H_s) is identified as the most effective parameter with importance percentage of 60.962% (Figure 9). The next most important parameter, step-top geometric ratio (L_s/W_s), defined as the step length to step width, has an importance percentage of 26.062%. Additionally, the approach channel ratio parameter (L_a/L), defined as the approach channel length to spillway length, and spillway angle parameter ($H/n{\cdot}L_S$), defined as the spillway height to cumulative step length, have importance percentages of 8.389% and 4.586%, respectively.

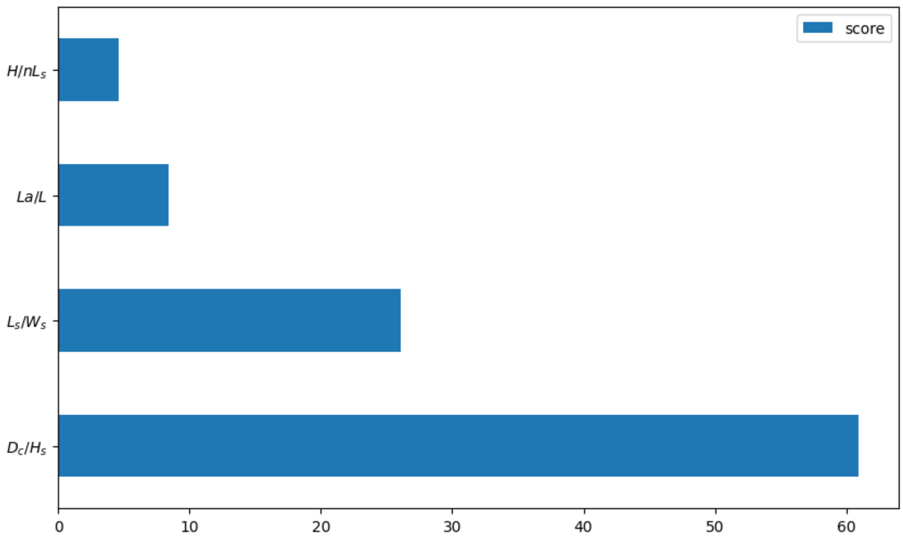

Figure 9. The percentage of the input parameters obtained from the XGB model.

Moreover, an FL model was formed without using the training dataset. FL is not a black box, so the dataset enters fuzzy logic after the programming is finished and is free from training and testing series. Mamdani Inference System [46] provides a relationship between inputs and output parameters. FL yields fuzzy output sets and defuzzification provides crisp values, and the centroid method is used, which considers the center of the output set [47]. MAE and MPE are used to determine FLs efficiency on energy dissipation of stepped spillways. MAE and MPE values are calculated as 2.000 and -1.688%. The results can be accepted as a good aspect of statistics. Also, observed and calculated energy dissipation values are plotted on graphs to visualize results (Fig. 10).

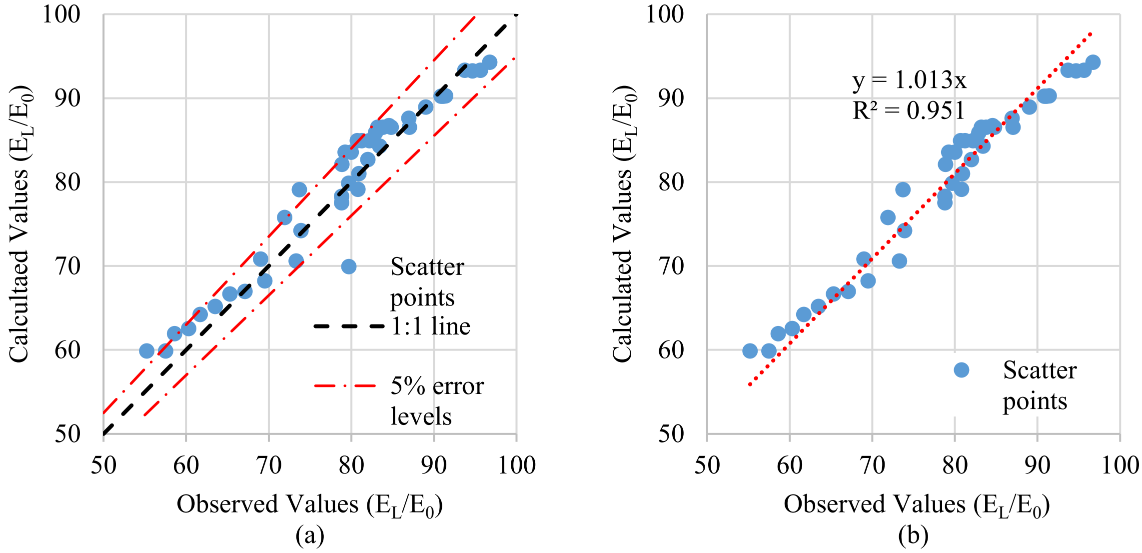

Figure 10. Membership functions of the parameters for design of stepped spillways.

In Fig. 10a, most of the scatter points, which give observed (calculated) energy values on the horizontal (vertical) axis, are in 5% error levels (red lines). This result shows that the error rate is less than some CFD results in the literature [8, 25, 48]. The determination coefficient between observed and calculated energy dissipation values is approximately 0.951 in Fig. 10b.

4. CONCLUSION

Machine Learning methods, including RF, XGB, KNN, LGBM, ANN, and LR regression models yield R^2 ranging from 0.976 (ANN) and 0.864 (LR) for the testing datasets. The FL model only uses testing datasets and has a higher R^2 than the LGBM (0.897), KNN (0.896), RF (0.864), and LR (0.864) models in the testing dataset. As for the XGB model, it has similar MAE (1.903), MPE (-1.41%) and $\boldsymbol{R^2}$ (0.954) values to the FL model, which has MAE, MPE and $\boldsymbol{R^2}$ values of 2.000, -1.688%, and 0.951, respectively. The ANN produces the best results with $\boldsymbol{R^2}$=0.976, MAE=1.398, and MPE=-0.117. However, the errors vary significantly depending on hidden layers, and the ratio of training to testing datasets. Moreover, like other ML models, it is a black-box model and is not based on physical rules. Furthermore, the two new parameters introduced in this study, defined as the approach channel and step-top geometric ratios have importance percentages of 26.062% and 8.389%, respectively.

Unlike ML models that rely on training and testing processes to evaluate complex flow events in stepped spillways, the fuzzy inference system may gain a physical basis using rule bases associated with examined problems. The accuracy of ML methods varies depending on the ratio of data sets divided into training and test data sets. The Fuzzy Inference System (FIS) is independent from the dataset and constructed with fuzzy rules. This represents its effectiveness and priority. The FIS model can be used safely and allows us to calculate the energy dissipation on stepped spillways quickly. The function (endistepway) produced in this study can calculate energy dissipation ratios with low MPE (-1.69) and MAE (2.00) ratios. It does not need advanced computers like computational fluid dynamics programs and has short processing times. It also has open-source codes and can be easily accessed and safely used to calculate the energy dissipation on stepped spillways. This situation enables engineers to model stepped spillways efficiently from economic and technical aspects. Unfortunately, floods continue to increase significantly with the negative effects of climate change [49]. Efficiently designed spillways play an important role in discharging floods from dam reservoirs, which cause great economic and loss of life.

ACKNOWLEDGMENTS

The first version of this study was published as a preprint.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest in this study.

AUTHOR STATEMENT

All authors contributed to the study conception and design. Data collection and analysis were performed by **Sadık Alashan** and **Eyyüp Ensar Yalçın**. The machine learning part was performed by **Sedat Golgiyaz**. The first draft of the manuscript was written by **Erdinç İkincioğulları** and all authors commented on previous versions of the manuscript. All authors read and approved of the final manuscript.

DATA AVAILABILITY

The datasets used during the current study have been provided from literature studies (Felder, Fromm, et al., 2012; Felder, Guenther, et al., 2012; Irzooki, Rasheed Mohammed, & Ameen, 2016)

REFERENCES

[1] Chanson, H. (2000). Forum article. Hydraulics of Stepped Spillways: Current Status. *Jl Hyd Engrg*, ASCE,126.

[2] Chanson H. (2004). Drag reduction in skimming flow on stepped spillways by aeration. J *Hydraul Res*, 42:316–22. https://doi.org/10.1080/00221686.2004.9728397.

[3] Berkün, M. (2007) *Su Yapıları*. Birsen Yayınevi [in Turkish].

[4] Boes, RM., Chanson, H., Matos, J., Ohtsu, I., Yasuda, Y., Takahasi, M. (2000). Characteristics of Skimming Flow over Stepped Spillways. *J Hydraul Eng*, 126:860–73. https://doi.org/10.1061/(asce)0733-9429(2000)126:11(860).

[5] Boes, RM., Hager, WH. (2003). Hydraulic Design of Stepped Spillways. *J Hydraul Eng*, 129:671–9. https://doi.org/10.1061/(asce)0733-9429(2003)129:9(671).

[6] Chanson, H. (1993). Stepped spillway flows and air entrainment. *Can J Civ Eng*, 20:422–35. https://doi.org/10.1139/l93-057.

[7] Chanson, H. (1998). R*eview of studies on stepped channel flows.* Work Flow Charact around Hydraul Struct River Environ, 25.

[8] Bai, ZL., Peng, Y., Zhang, JM. (2017). Three-Dimensional Turbulence Simulation of Flow in a V-Shaped Stepped Spillway. *J Hydraul Eng*, 143:06017011. https://doi.org/10.1061/(asce)hy.1943-7900.0001328.

[9] Felder, S., Chanson, H. (2014). Effects of Step Pool Porosity upon Flow Aeration and Energy Dissipation on Pooled Stepped Spillways. *J Hydraul Eng*, 140:04014002. https://doi.org/10.1061/(asce)hy.1943-7900.0000858.

[10] Zare, HK., Doering, JC. (2012). Effect of rounding edges of stepped spillways on the flow characteristics. *Can J Civ Eng*, 39:140–53. https://doi.org/10.1139/L11-121.

[11] Zhang, G, Chanson, H. (2018). Effects of Step and Cavity Shapes on Aeration and Energy Dissipation Performances of Stepped Chutes. *J Hydraul Eng*, 144:04018060. https://doi.org/10.1061/(asce)hy.1943-7900.0001505.

[12] Zhou, Y., Wu, J., Ma, F., Hu, J. (2020). Uniform flow and energy dissipation of hydraulic-jump-stepped spillways. *Water Sci Technol Water Supply,* 20:1546–53. https://doi.org/10.2166/ws.2020.056.

[13] Felder, S., Fromm, C., Chanson, H. (2012) *The Universıty of Queensland Report Ch86/12 Air Entrainment And Energy Dissipation On A 8.9° Slope Stepped Spillway with Flat and Pooled Steps.* School of Civil Engineering.

[14] Felder, S., Guenther, P., Chanson, H. (2012). T*he University of Queensland Report CH87/12 Air-water flow properties and energy dissipation on stepped spillways: a physical study of several*

pooled stepped configurations. School of Civil Engineering.
[15] Mero, S., Mitchell, S. (2017). Investigation of energy dissipation and flow regime over various forms of stepped spillways. *Water Environ J*, 31:127–37. https://doi.org/10.1111/wej.12224.
[16] Peyras, L., Royet, P., Degoutte, G. (1992). Flow and Energy Dissipation over Stepped Gabion Weirs. *J Hydraul Eng*, 118:707–17. https://doi.org/10.1061/(asce)0733-9429(1992)118:5(707).
[17] Rice, CE., Kadavy, KC. (1996). Model Study of a Roller Compacted Concrete Stepped Spillway. *J Hydraul Eng*, 122:292–7. https://doi.org/10.1061/(asce)0733-9429(1996)122:6(292).
[18] Torabi, H., Parsaie, A., Yonesi, H., Mozafari, E. (2018). Energy Dissipation on Rough Stepped Spillways. *Iran J Sci Technol - Trans Civ Eng*, 42:325–30. https://doi.org/10.1007/s40996-018-0092-5.
[19] Wu, J., Qian, S., Wang, Y., Zhou, Y. (2020). Residual Energy on Ski-Jump-Step and Stepped Spillways with Various Step Configurations. *J Hydraul Eng*, 146:06020002. https://doi.org/10.1061/(asce)hy.1943-7900.0001710.
[20] Zare, HK., Doering, JC. (2012). Inception Point of Air Entrainment and Training Wall Characteristics of Baffles and Sills on Stepped Spillways. *J Hydraul Eng*, 138:1119–24. https://doi.org/10.1061/(asce)hy.1943-7900.0000630.
[21] Ashoor, A., Riazi, A. (2019). Stepped Spillways and Energy Dissipation: A Non-Uniform Step Length Approach. *Appl Sci,* 9:5071. https://doi.org/10.3390/app9235071.
[22] De Carvalho, RF., Amador, AT. (2009). Physical and numerical investigation of the skimming flow over a stepped spillway. *Adv Water Resour Hydraul Eng - Proc 16th IAHR-APD Congr 3rd Symp IAHR-ISHS*, 1767–72. https://doi.org/10.1007/978-3-540-89465-0_304/COVER.
[23] Tabbara, M., Chatila, J., Awwad, R. (2005). Computational simulation of flow over stepped spillways. *Comput Struct*, 83:2215–24. https://doi.org/10.1016/j.compstruc.2005.04.005.
[24] Yalcin, EE., Ikinciogullari, E., Kaya, N. (2023). Comparison of Turbulence Methods for a Stepped Spillway Using Computational Fluid Dynamics. *Iran J Sci Technol - Trans Civ Eng*, 47:3895–911. https://doi.org/10.1007/S40996-023-01127-5/FIGURES/14.
[25] Ghaderi, A., Abbasi, S., Abraham, J., Azamathulla, HM. (2020). Efficiency of Trapezoidal Labyrinth Shaped stepped spillways. *Flow Meas Instrum*, 72:101711. https://doi.org/10.1016/j.flowmeasinst.2020.101711.
[26] Ghaderi, A., Abbasi, S., Di Francesco, S. (2021). Numerical Study on the Hydraulic Properties of Flow over Different Pooled Stepped Spillways. Water, 13:710. https://doi.org/10.3390/w13050710.
[27] Ikinciogullari, E. (2023). Stepped spillway design for energy dissipation. *Water Supply*, 23:749–63. https://doi.org/10.2166/WS.2023.016.
[28] Ikinciogullari, E. (2023). A novel design for stepped spillway using staggered labyrinth trapezoidal steps. Flow Meas Instrum, 93:102439. https://doi.org/10.1016/J.FLOWMEASINST.2023.102439.
[29] Ikinciogullari, E. (2023). Energy dissipation performance of the trapezoidal stepped spillway. *J Eng Res*,11(2A). https://doi.org/10.36909/JER.13649.
[30] Ikinciogullari, E. (2024). Energy dissipation performance of labyrinth and harmonic stepped spillways. *J Hydroinformatics*. https://doi.org/10.2166/HYDRO.2024.221.
[31] Mohammad, M., Tabari, R., Tavakoli, S. (2016). Effects of Stepped Spillway Geometry on Flow Pattern and Energy Dissipation. *Arab J Sci Eng,* 41:1215–24. https://doi.org/10.1007/s13369-015-1874-8.
[32] Reeve, DE., Zuhaira, AA., Karunarathna, H. (2019). Computational investigation of hydraulic performance variation with geometry in gabion stepped spillways. *Water Sci Eng*, 12:62–72. https://doi.org/10.1016/j.wse.2019.04.002.
[33] Korkut, O., Demirel, S., Taşar, B., Kaya, YZ. (2023). Adana İli Referans Evapotranspirasyon Miktarının Bulanık Smrgt, Anfis ve Çoklu Doğrusal Regresyon Kullanılarak Tahmini. *Osmaniye Korkut Ata Üniversitesi Fen Bilim Enstitüsü Derg*, 6:106–20. https://doi.org/10.47495/OKUFBED.1079066.
[34] Üneş, F., Taşar, B., Demirci, M., Zelenakova, M., Kaya, YZ., Varçin, H. (2021). Daily Suspended Sediment Prediction Using Seasonal Time Series and Artificial Intelligence Techniques. *Rocz Ochr Sr*, 23:117–37. https://doi.org/10.54740/ROS.2021.008.
[35] Salmasi, F., Özger, M. (2014). Neuro-Fuzzy Approach for Estimating Energy Dissipation in Skimming Flow over Stepped Spillways. *Arab J Sci Eng*, 39:6099–108. https://doi.org/10.1007/S13369-014-1240-2/METRICS.

[36] Roushangar, K., Akhgar, S., Salmasi, F., Shiri, J. (2016). Neural networks- and neuro-fuzzy-based determination of influential parameters on energy dissipation over stepped spillways under nappe flow regime. https://doi.org/10.1080/09715010.2016.1235472.
[37] Mojtahedi, A., Soori, N., Mohammadian, M. (2020). Energy dissipation evaluation for stepped spillway using a fuzzy inference system. *SN Appl Sci*, 2:1–13. https://doi.org/10.1007/S42452-020-03258-0/
[38] Irzooki, RH., Rasheed Mohammed, J., Ameen, AS. (2016). Computational Fluid Dynamics Modeling of Flow over Stepped Spillway. *J Eng Sci*, 23:1–11.
[39] Geurts, P., Ernst, D., Wehenkel, L. (2006). Extremely randomized trees. *Mach Learn*. https://doi.org/10.1007/s10994-006-6226-1.
[40] Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proc. *ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* https://doi.org/10.1145/2939672.2939785.
[41] Kaya, H., Guler, E., Kırmacı, V. (2023). Prediction of temperature separation of a nitrogen-driven vortex tube with linear, kNN, SVM, and RF regression models. *Neural Comput Appl.* https://doi.org/10.1007/s00521-022-08030-6.
[42] Bakır, R., Orak, C., Yüksel, A. (2024). Optimizing hydrogen evolution prediction: A unified approach using random forests, lightGBM, and Bagging Regressor ensemble model. *Int J Hydrogen Energy*, 67:101–10. https://doi.org/https://doi.org/10.1016/j.ijhydene.2024.04.173.
[43] Golgiyaz, S., Cellek, MS., Daskin, M., Talu, MF., Onat, C. (2023). Flame stability measurement through image moments and texture analysis. *Meas Sci Technol.* https://doi.org/10.1088/1361-6501/acb001.
[44] Zadeh L. A. (1967). Fuzzy sets. Inf Control, 8:38–53.
[45] Singh, S., Singh, S., Banga, VK. (2020). Design of fuzzy logic system framework using evolutionary techniques. *Soft Comput,* 24:4455–68. https://doi.org/10.1007/s00500-019-04207-9.
[46] Mamdani, EH., Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *Int J Man Mach Stud.* https://doi.org/10.1016/S0020-7373(75)80002-2.
[47] Şen Z. (2009). Fuzzy logic and hydrological modeling.
[48] Li, S., Zhang, J., Xu, W. (2018). Numerical investigation of air–water flow properties over steep flat and pooled stepped spillways. *J Hydraul Res*, 56:1–14. https://doi.org/10.1080/00221686.2017.1286393.
[49] Alashan, S. (2020) Innovative Trend Analysis Methodology in Logarithmic Axis. *Konjes* 2020;8:573–85. https://doi.org/doi:10.36306/konjes.668212.