Predicting Energy Loss Over Vegetated Dikes Utilizing Machine Learning Techniques

Nadir Murtaza[1*], Ghufran Ahmed Pasha[1]

1. *Civil Engineering Department, University of Engineering and Technology Taxila, 47050, Pakistan*

* **Corresponding Author:** Email: nadirkhattak1122@yahoo.com

Abstract

Vegetated dike has a significant role in energy loss of flooding, however, it is a challenging task to accurately predict the energy loss. Therefore, the present research work attempted to estimate the energy loss of flood over a vegetated dike utilizing machine learning techniques (ML) including random forest (RF) and extreme gradient boosting with particle swarm optimization (XGBoost-PSO). Dataset of various parameters like Froude number (Fr), velocity reduction (Vo/V), ho/h, (ho: initial water depth, h: water in a flume with vegetated dike), ho/B, (B: channel width), and energy loss (E) was calculated from the experiment performed in a controlled laboratory setting. Moreover, SHAP analysis was performed to investigate the impact of critical parameters on energy loss. The result of the findings demonstrates the superior performance of the XGBoost-PSO due to a higher R-value of 0.99 and a lower MSE value of 0.0345. The SHAP analysis result indicates that in the case of the RF model parameter ho/h has a significant impact on energy loss while ho/B in the case of the XGBoost-PSO model. The findings of the present research provide a precise estimation of the energy loss while designing a vegetated dike in a flood-prone region.

Keywords: Machine learning; energy loss; random forest; particle swarm optimization; SHAP analysis

1. Introduction

Flooding is defined as the occurrence of water on land and areas that tend to be dry most of the time, usually due to precipitation [1]. Among all types of flooding, both flash floods and river floods are widespread in Pakistan. The mountainous and sub-mountainous region of the country is more vulnerable to very heavy flash floods because of steep slopes and varied climates. Specifically, the hill torrents that collect water drained from the hilly and foothills including water derive a good part of their volume, and torrents specialized in flash floods despite these having averaged low annual rainfall.

Past works have investigated several flow characteristics under subcritical and supercritical situations by altering total discharge but fixing the channel bed [2-6]. The findings of the undertaken studies have been mostly based on assessing the energy efficiency of the defense systems [7-10]. Ahmed and Ghumman's [9] research shows that single-vegetation and hybrid defense systems (dyke-moat-vegetation) in subcritical flow conditions are associated with 32 percent and 46 percent energy reduction respectively, while Pasha and Tanaka [7] conclude that dense vegetation reduces even higher energy than intermediate or sparse vegetation under supercritical conditions though flow discharge was regulated by change in the channel bed. Furthermore, various studies have been conducted on hydraulic jumps through defense systems in all regimes of flow [11]. A similar undulated hydraulic jump was reported by Retsinis and Papanicolaou [11] in an open channel with rapidly varied flow while Pasha and Tanaka [7] observed undulated hydraulic jumps under steady subcritical flow conditions in agreement with Ahmed and Ghumman [9]. These studies in combination suggest defense systems for floods by reducing energy dissipation because of hydraulic jump, and modification in the drag coefficient under both the subcritical and supercritical state by varying discharge and keeping the channel bed fixed.

There is clear evidence for the fact that vegetation, being one of the major characteristics of natural rivers, controls flow resistance. On the same note, vegetation offers several ecological benefits but turns out to be a factor that complicates the flow by increasing flow resistance while at the same time affecting sediment transport. In regards to engineering concerns of flood, river segments generally go through vegetation management [12]. As vegetated surfaces and granular sediments are associated with different roughness, dissimilar measurements are required to model flow accurately [13]. There is a trend to determine the flow resistance as applied to vegetated parts of river systems. This study found vegetation significant in river conservation and rehabilitation since it is involved in the physical processes that determine rivers' stability and the biology shaping riverine ecosystems and water flow [14]. Some act on the short-term mobility of sediments and the accumulation of sediment, vertical and lateral flow, and flow resistance, thanks to interactions between the sediment applicants [14]. Floodplain flow processes are frequently convoluted due to the dynamic between flow and vegetation and may go over the erosion threshold, according to Stone and Shen [15]. Water velocity, particularly in floods, is also influenced by vegetation, by increasing water depth and decreasing the ability of the river to discharge floods making flood events worse [16]. Moreover, vegetation in a fixed-bed flume causes total channel resistance thus worsening water depth and improving flow velocity [14].

A clear gap is noted in the previous research in which the focus was to explore the flood energy loss in a controlled laboratory setting under diverse flow conditions. However, the present research work utilized a dataset of controlled laboratory utilizing vegetated dikes to explore the significance of the machine learning techniques in predicting energy loss. The dataset of energy loss was measured under the condition of varying the distance between a house model located downstream side of the vegetated dike under different flow conditions. The dataset of a controlled laboratory setting was utilized in various machine learning techniques such as RF, and XGBoost-PSO as an input parameter while energy loss was considered as a predicted/output factor. Furthermore, the impact of the critical parameter was assessed using a SHAP analysis.

2. Methodology

2.1 Data Collection and Machine Learning Technique

Experiments were performed in a controlled laboratory environment for collecting a dataset of various parameters including flow velocity and depth without and with placing vegetated dike in a channel. The flow velocity and water depth were measured utilizing an electromagnetic velocity meter and rail-mounted point gauge. Based on flow svelocity and depth various non-dimensional parameters like Froude number ($Fr = V/(gho)^{0.5}$, g: gravitational acceleration), ho/h, Vo/V, and ho/B were calculated. Moreover, the energy loss was estimated from the controlled laboratory environment using $E = \frac{E_1 - E_2}{E_1}$, where E1 and E2 are the specific energy on the upstream and downstream sides of the vegetated dike and a similar approach was adopted in previous research [17-20]. Therefore, parameters such as Fr, ho/h, ho/B, and Vo/V were considered as input while E was an output parameter in random forest and extreme boosting gradient with particle swarm optimization. After collecting a dataset of selected parameters machine learning techniques such as RF and XGBoost-PSO were utilized to predict the energy loss over the vegetated dike. The Random Forest is an example of an ensemble learning technique that employs a set of decision trees for a better prediction of results. Every distinct tree is only trained on a randomly selected portion of the data and the overall result is reached from averaging all trees' outputs. The RF gives an energy loss value by calculating Fr, ho/h, ho/B, and Vo/V. This setup enhances stability and reduces the incidence of fitting in elaborate work conditions of flow and depth in the laboratory. The XGBoost-PSO model splits eXtreme Gradient Boosting (XGBoost), this is an effective boosting method, which categorizes sequential decision trees, with Particle Swarm Optimization (PSO) to optimize hyperparameters. The XGBoost uses parameters such as Fr, ho/h, ho/B, and Vo/V in predicting energy loss, while PSO improves the accuracy of the model and computation time for the model parameters. Furthermore, a SHAP analysis was performed based on the RF and XGBoost-PSO models to highlight the importance of the various parameters on the energy loss over the vegetated dike. Table 1 summarizes the range of various input parameters utilized in the machine learning techniques.

Table 1: Range of various input and output parameters utilized in the current research. Where SD is the standard deviation and E is the energy loss

Parameters	ho/B	Fr	Vo/V	ho/h	E
Maximum	0.31	0.409	3.37	0.34	17.99
Minimum	0.2	0.203	0.247	0.278	11.546
SD	0.038	0.08	1.371	0.017	1.872
Average	0.259	0.324	2.19	0.311	14.883

3. Result and Discussion

3.1 Performance Indicator (RF & XGBoost-PSO)

Figure 2a-b depicts the result of the performance indicators and predicted values of the RF and XGBoost-PSO models. When comparing the performance of two machine learning techniques, namely RF and XGBoost with PSO, the correlation coefficient, and the mean square error were used. The results of using the XGBoost-PSO framework as the predictive model showed satisfac-

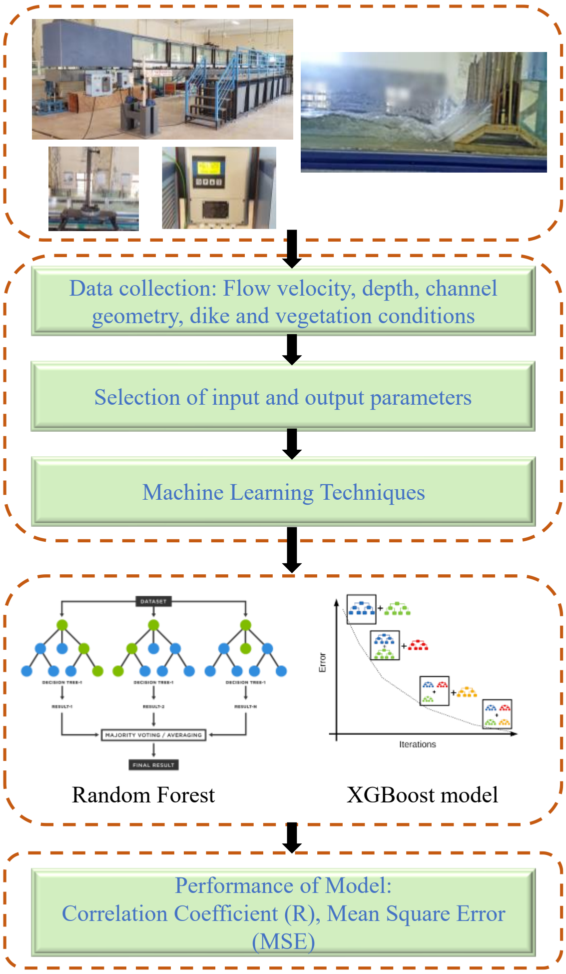


Fig. 1: Methodology framework of the current research

tory metrics as the value of R was 0.99 and MSE was 0.0345 which means; a very good and strong positive correlation between n predicted and actual values and margin of error respectively. Nonetheless, the RF model had a higher R-value of 0.978 and MSE of 0.071, and a similar result was reported in a study conducted by Khan et al [21], but those values were slightly lower than the indicators of superordination of XGBoost-PSO. Hence, the improved performance of the proposed hybrid model, XGBoost-PSO, can be linked to the integration of the components. XGBoost came into the picture because of its strong capability to handle nonlinear relationships and minimum error by boosting [22]. To include additional complexity of the Particle Swarm Optimization (PSO), which helped to fine-tune these hyperparameters to improve the convergence rate and the accuracy of the model. This results in a better generalization of the data set and a closer approximation to the original results as shown below. Since a lower MSE was obtained for XGBoost-PSO, it means that XGBoost-PSO is more precise in its predictions as well as minimizing prediction errors than RF. Therefore, the proposed approach, the XGBoost-PSO model, is more acceptable in this work for this application because of the better accuracy and lesser error compared to the traditional PSO model.

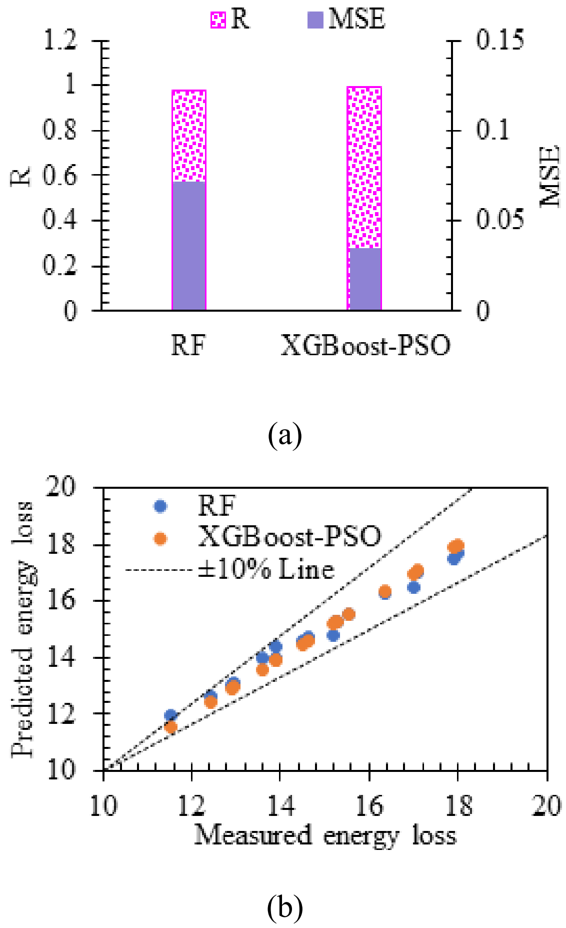


Fig. 2: Performance indicators and predicted values of different machine learning techniques (a) performance indicator of the random forest and extreme gradient boosting with particle swarm optimization (b) predicted and observed values of the energy loss.

3.2 SHAP Analysis

Figure 3a-b illustrates the impact of various parameters on the energy loss over vegetated dike using SHAP analysis. The SHAP analysis conducted for the Random Forest model highlights the influence of specific parameters on energy loss through a vegetated dike. The results indicate that parameters such as the Froude number (Fr), relative depth ratio (ho/h), relative width ratio (ho/B), and velocity ratio (Vo/V) significantly impact the energy loss. The Froude number (Fr), a dimensionless parameter representing the flow regime, plays a crucial role in determining energy loss. Higher Froude numbers indicate supercritical flow conditions, where kinetic energy dominates. This increases turbulence around the vegetated dike, leading to greater energy dissipation. Consequently, Fr has a notable impact on energy loss, as higher values intensify interactions between the water flow and vegetation. The relative depth ratio (ho/h) also affects energy loss. This ratio represents the water depth over the dike compared to the downstream depth. A higher ho/h ratio implies a deeper overflow over the dike, which can lead to increased flow resistance due to submerged vegetation, amplifying energy dissipation. Similarly, the relative width ratio (ho/B), which compares the water depth over the dike to the dike's width, influences energy loss. A higher ho/B ratio indicates that the water depth is more significant relative to the dike width, enhancing the interaction between flow and vegetation, which raises energy dissipation through drag and turbulence. Finally, the velocity ratio (Vo/V), representing the overflow velocity relative to the main flow velocity, affects energy loss. Higher Vo/V ratios suggest a more intense flow over the dike, increasing energy dissipation as the vegetation absorbs and disrupts the flow.

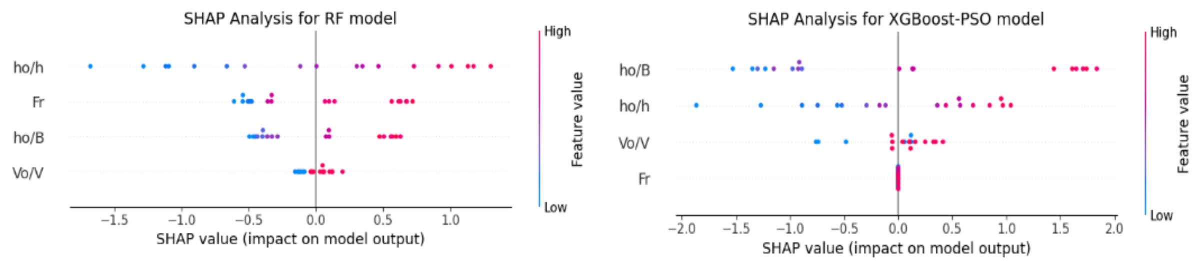
The parameters, Froude number (Fr), relative depth ratio (ho/h), relative width ratio (ho/B), and velocity ratio (Vo/V)—have a significant impact on energy loss through a vegetated dike as they dictate flow behavior and interaction with vegetation. Higher values of these parameters increase turbulence, drag, and flow resistance. This intensifies energy dissipation as water interacts more dynamically with the vegetative surface, making these parameters crucial for optimizing flood mitigation designs.

4. Conclusion

In the current research, a controlled laboratory setting was performed utilizing a vegetated dike under various flow conditions. An experimental dataset was collected under subcritical flow conditions through a vegetated dike to predict energy loss using machine learning techniques such as random forest (RF) and extreme gradient boosting with particle swarm optimization (XGBoost-PSO). Therefore, the conclusion of the current research is following.

The findings of the current research conclude a super performance and prediction capacity of the XGBoost-PSO technique in comparison to the RF model due to a higher correlation coefficient value of 0.99 and a lower mean square error value of 0.0345. Moreover, the SHAP analysis performed in the current research shows the significant impact of the initial water depth ratio on the channel width. The study concluded that by increasing the ratio of ho/B, energy loss also increases.

The current recommended that future research should focus on the investigation of various hydraulic forces through vegetated-dike and for the prediction of hydraulic forces advanced machine learning techniques should be utilized.



(a) (b)

Fig. 3: Result of SHAP analysis for different machine learning techniques (a) random forest model (b) XGBoost-PSO

5. References

[1] Lavell, A., Oppenheimer, M., Diop, C., Hess, J., Lempert, R., Li, J., & Myeong, S. (2012). Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of working groups I and II of the intergovernmental panel on climate change (IPCC), 3, 25-64.

[2] Murtaza N, Pasha GA, Khan ZU, Alotaibi S, Akbar Z, Khedher KM. Impact of dyke and vegetation on fluid force and moment reduction under sub and supercritical flow conditions. Physics of Fluids. 2024 Oct 1;36(10). https://doi.org/10.1063/5.0237696

[3] Iqbal, S., Siddique, M., Hamza, A. et al. Computational analysis of fluid dynamics in open channel with the vegetated spur dike.

Innov. Infrastruct. Solut. 9, 345 (2024). https://doi.org/10.1007/s41062-024-01636-w

[4] Murtaza, N., Pasha, G.A., Tanaka, N. et al. Analysis of Hydraulic Jump and Energy Dissipation in Flow Through Emergent Vegetation Under Varying Froude Numbers. Iran J Sci Technol Trans Civ Eng (2024). https://doi.org/10.1007/s40996-024-01571-x

[5] Murtaza, N., Pasha, G., Murtaza, A., Ahmed, A., Raza, H., & Khalid, M. (2024a). Sloping Perspectives: Investigating Flow Hydrodynamics in Vegetated Open-Channels Under Varying Bed Slope. Technical Journal, 3(ICACEE), 626-638. Retrieved from https://tj.uettaxila.edu.pk/index.php/technical-journal/article/view/1873

[6] Murtaza, N., Pasha, G.A., Tanaka, N. et al. (2025), Optimal configuration of a hybrid defense system for hill torrents: an experimental investigation for effective management. Nat Hazards. https://doi.org/10.1007/s11069-025-07309-w.

[7] Pasha, G. A., & Tanaka, N. (2017). Undular hydraulic jump formation and energy loss in a flow through emergent vegetation of varying thickness and density. Ocean Engineering, 141, 308–325. https://doi.org/10.1016/j.oceaneng.2017.06.049

[8] Pasha, G.A., Tanaka, N., (2020). Characteristics of a hydraulic jump formed on upstream vegetation of varying density and thickness. Journal of Earthquake and Tsunami. 14 (3), 2050012 (1-28). https://doi.org/10.1142/S1793431120500128

[9] Ahmed, A., & Ghumman, A. R. (2019). Experimental Investigation of Flood Energy Dissipation by Single and Hybrid Defense System. Water, 11(10), 1971.

[10] Ahmed, A., Valyrakis, M., Razzaq Ghumman, A., Pasha, G. A., & Farooq, R. (2021). Experimental Investigation of Flood Energy Dissipation through Embankment Followed by Emergent Vegetation. Periodica Polytechnica Civil Engineering.

[11] Retsinis, Eugene, and Panayiotis Papanicolaou. 2020. "Numerical and Experimental Study of Classical Hydraulic Jump" Water 12, no. 6: 1766. https://doi.org/10.3390/w12061766

[12] Aberle, Jochen; Järvelä, Juha (2013). Flow resistance of emergent rigid and flexible floodplain vegetation. Journal of Hydraulic Research, 51(1), 33–45. https://doi.org/10.1080/00221686.2012.754795

[13] Huthoff, F. (2012). Theory for flow resistance caused by submerged roughness elements. Journal of hydraulic research, 50(1), 10-17. https://doi.org/10.1080/00221686.2011.636635

[14] Huai, Wx., Li, S., Katul, G.G. et al. (2021), Flow dynamics and sediment transport in vegetated rivers: A review. J Hydrodyn 33, 400–420. https://doi.org/10.1007/s42241-021-0043-7

[15] Stone, B. M., & Shen, H. T. (2002). Hydraulic Resistance of Flow in Channels with Cylindrical Roughness. Journal of Hydraulic Engineering, 128, 500-506. https://doi.org/10.1061/(ASCE)0733-9429(2002)128:5(500).

[16] Wang, C., Zheng, Ss., Wang, Pf. et al. (2015), Interactions between vegetation, water flow and sediment transport: A review. Journal of Hydrodynamic 27, 24–37. https://doi.org/10.1016/S1001-6058(15)60453-X.

[17] N. Murtaza, G.A. Pasha, H. Hamidifar, U. Ghani, A. Ahmed, Enhancing flood resilience: Comparative analysis of single and hybrid defense systems for vulnerable buildings, International Journal of Disaster Risk Reduction, Volume 116, 2025, 105078, https://doi.org/10.1016/j.ijdrr.2024.105078.

[18] GA Pasha, M Asghar, N Murtaza, AR Ghumman. (2024), Impact of floating debris on houses during floods and vegetation-based mitigation, Proceedings of the ICE Engineers-Water. https://doi.org/10.1680/jwama.23.00055

[19] Murtaza, N., Pasha, G.A., Khan, Z.U., Alotaibi, S., Akbar, Z. and Khedher, K.M., 2024. Impact of dyke and vegetation on fluid force and moment reduction under sub and supercritical flow conditions. Physics of Fluids, 36(10). https://doi.org/10.1063/5.0237696

[20] Murtaza, N., Pasha, G.A., Khan, Z.U. et al. (2025), Assessing Dyke and Moat systems for hydrodynamic reduction in super-critical flow: a laboratory and ANN approach. Innov. Infrastruct. Solut. 10, 14. https://doi.org/10.1007/s41062-024-01836-4

[21] Khan, Z. U., Khan, D., Murtaza, N., Pasha, G. A., Alotaibi, S., Rezzoug, A., Benzougagh, B., & Khedher, K. M. (2024). Advanced Prediction Models for Scouring Around Bridge Abutments: A Comparative Study of Empirical and AI Techniques. Water, 16(21), 3082. https://doi.org/10.3390/w16213082

[22] Murtaza, N., Khan, D., Rezzoug, A., Khan, Z.U., Benzougagh, B. and Khedher, K.M., 2025. Scour depth prediction around bridge abutments: A comprehensive review of artificial intelligence and hybrid models. Physics of Fluids, 37(2). https://doi.org/10.1063/5.0244974