Estimation of Wave Overtopping Discharges at Coastal Structures with Combined Slopes using Machine Learning Techniques

Moussa S. Elbisy
Civil Engineering Department, College of Engineering and Islamic Architecture, Umm Al-Qura University, Saudi Arabia
mselbisy@uqu.edu.sa (corresponding author)

Received: 1 March 2024 | Revised: 17 March 2024 | Accepted: 19 March 2024

 DOI: https://doi.org/10.48084/etasr.7175

ABSTRACT

Coastal defense structures are of paramount importance in protecting coastal communities from the adverse impacts of severe weather events and flooding. This study uses machine learning techniques, specifically Decision Tree (DT), Gradient Boosted Tree (GBT), and Support Vector Machine (SVM) models, to estimate wave overtopping discharge at coastal structures with combined slopes employing the recently built EurOtop database. The models were evaluated by deploying statistical metrics and Taylor diagram visualization. The GBT model demonstrated a high level of accuracy in predicting wave-overtopping discharge. Compared to the other models, the scatter index of GBT (0.392) was lower than that of DT (0.512) and SVM (0.823). In terms of the R-index, GBT (0.991) was superior to DT (0.977) and SVM (0.943). The GBT results were also compared with those of previous works. The findings showed that the GBT model significantly decreased the overall error and provided accurate estimations of the wave-overtopping discharge.

Keywords-coastal defense; wave overtopping; prediction; gradient boosted trees; decision trees; support vector machines; safety

I. INTRODUCTION

In recent decades, there has been a gradual development of coastal zones, which have substantial economic importance for countries around the world [1]. Coastal protection infrastructures are designed to mitigate the risks associated with wave-induced overtopping and safeguard people and assets located behind these protective barriers. However, the strain on crucial coastal defenses is expected to escalate due to the confluence of rising sea levels and the projected increase in both frequency and intensity of severe storm surges, all of which are consequences of global climate change [2]. Numerous investigations have examined flood protection strategies aimed at managing future coastal flood risks. However, it is imperative to diligently evaluate the potential wave overtopping of coastal defenses, considering the inherent uncertainties associated with such assessments. Accurate prediction of the propagation of wave overtopping in the inland is essential for the development of coastal defense structures [3].

Wave overtopping is governed by a multitude of parameters. Using analytical methods that rely on simplified representations of the underlying processes may not consistently yield precise predictions due to the intricate and stochastic nature of the overtopping phenomena. Wave overtopping has been extensively investigated in several studies and certain initiatives have been taken, primarily through the adoption of experimental approaches [4]. These investigations have led to the formulation of various empirical prediction equations and the utilization of Artificial Neural Networks (ANNs) [5]. Numerous methods are used to forecast wave overtopping, namely empirical formulas, Machine Learning (ML) techniques, and numerical models. Empirical formulas offer a convenient and straightforward method to obtain an initial approximation of the average wave overtopping discharge. The utilization of numerical modeling often requires significant processing resources that may be cost-prohibitive, thus it is important to acknowledge their limitations. In contrast, ML techniques can provide predictions with minimal delay, while simultaneously achieving a favorable trade-off between precision and efficiency. Additionally, these models facilitate the incorporation of a multitude of control parameters. The utilization of ML techniques is prevalent within the domain of coastal engineering due to their efficacy in knowledge processing, prediction, and forecasting. ML techniques encompass several approaches such as ANNs, tree-based methods, and Support Vector Machines (SVM).

In [6], a unique ANN tool was developed to estimate the wave reflection coefficient (K_r) and the wave transmission coefficients (K_t). The respective studies focused on the estimation of K_r, K_t, and the additional parameter q. In [7], an improved ANN model was presented to analyze different types of coastal structures. This model, published in EurOtop (2018), incorporated input parameters obtained from an expanded database grounded primarily on the CLASH database. In [8], gradient-boosting decision trees were used to predict average wave-overtopping discharges. These models were trained implementing the CLASH database, which contains data on wave overtopping. In [9], XGBoost was put into service to forecast the mean wave overtopping discharge. The specific model was evaluated by being compared to four new physical model datasets. A comprehensive quantitative comparison was conducted, juxtaposing the XGBoost with both previous ML approaches and empirical overtopping equations. The XGBoost model showed a reduction in errors on all the data utilized by a magnitude of up to 5 compared to current overtopping prediction approaches.

In [10], the accuracy of soft computing techniques was assessed using a Multilayer Perceptron Neural Network (MLPNN) and Random Forest Decision Tree (RFDT) for wave-overtopping discharge prediction of vertical coastal structures. The effectiveness of each method was reviewed through the employment of graphs and accuracy measures. The results demonstrated that the RFDT model provided more accurate predictions than MLPNN. In [11], ANN methods were followed, including MPNN, Generalized Regression Neural Network (GRNN), and SVM, to estimate the wave overtopping discharge in rubble mound structures characterized by a straight slope. The results disclosed that GRNN exhibited a high level of prediction accuracy. In [12], a Convolutional Neural Network (CNN) model was presented, applying deep learning techniques to predict wave overtopping in coastal constructions. The validation results displayed that this CNN model was exceptionally precise in calculating wave overtopping discharge based on hydraulic and structural characteristics. The evaluation of the overtopping predictions using a prototype dataset demonstrated that the proposed CNN model exceeded the performance of previous ML models. In [13], a variety of ML methods, such as MPNN, Cascade Correlation Neural Networks (CCNN), GRNN, and SVMs, were deployed to predict wave-overtopping discharge in coastal structures with a linear slope. The results indicated that the GRNN had a notable degree of precision.

This study implements Taylor diagram visualization and multiple evaluation metrics to evaluate the accuracy of several techniques, namely Decision Tree (DT), Gradient Boosted Trees (GBT), and SVM, in estimating wave overtopping of combined-slope coastal structures. The current study aims to develop a comprehensive and precise prediction model that can effectively estimate wave-overtopping discharges for a wide range and types of coastal structures under various wave conditions. This model aims to assist coastal designers in their decision-making processes. Furthermore, the results of this investigation can be exploited for the computation of wave overtopping risk and wave overtopping prediction so that warnings will be issued and emergency evacuations will be facilitated in the face of severe wave events. Also, the aforementioned findings may assist in risk mitigation, and the economic evaluation of coastal protection initiatives.

II. METHODS

A. Data

This study used the recently developed EurOtop database, which encompasses a comprehensive collection of 17,942 tests. The original CLASH database contained around 13,500 systematically organized global tests on wave overtopping discharge (q) [4]. A total of 4,737 tests were employed to investigate rubble mound structures that were devoid of a berm and featured a combined slope. Figure 1 portrays a schematic representation of the rubble mound structure, without a berm, with combined slopes. Table I and Figure 2 present statistical data for the essential parameters.

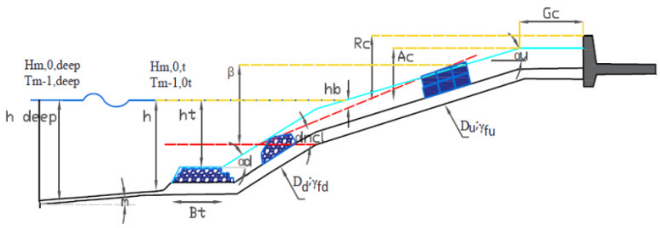

Fig. 1. Cross section with application parameters for ML techniques.

TABLE I. STATISTICAL SUMMARY FOR THE DIMENSIONAL BASIC PARAMETERS OF THE USED DATASET

Parameter	Definition	Number	Units	Minimum	Maximum	Mean	SD
Spreading	Spreading	4737	[-]	0	10	0.3462	1.3999
β	Angle of wave attack	4737	[°]	0	80	3.7205	11.4993
h	Water depth at the structure toe	4737	[m]	0.029	5.01	0.4617	0.3987
Hm0 toe	Significant wave height at the structure toe	4737	[m]	0.017	1.48	0.1271	0.0771
ht	Toe submergence	4737	[m]	0.029	5.01	0.4398	0.4035
Bt	Toe width	4737	[m]	0	2.031	0.0527	0.1330
γf	Roughness factor	4737	[-]	0	1	0.6265	0.3615
D	Average size of the structure elements	4737	[m]	0	1	0.1136	0.2269
cotαd		4737	[-]	0	7	2.3072	1.1931
cotαincl		4737	[-]	0	7	2.3394	1.2029
Rc	Crest height with respect to SWL	4737	[m]	0	2.5	0.1699	0.1519
Ac	(Armor) crest freeboard without crown wall	4737	[m]	0	2.5	0.1614	0.1532
Gc	Crest width or promenade width	4737	[m]	0	1	0.1185	0.1588
q	Wave overtopping discharge	4737	[m^3/s per m]	0.000001	0.0256	0.0008	0.0022

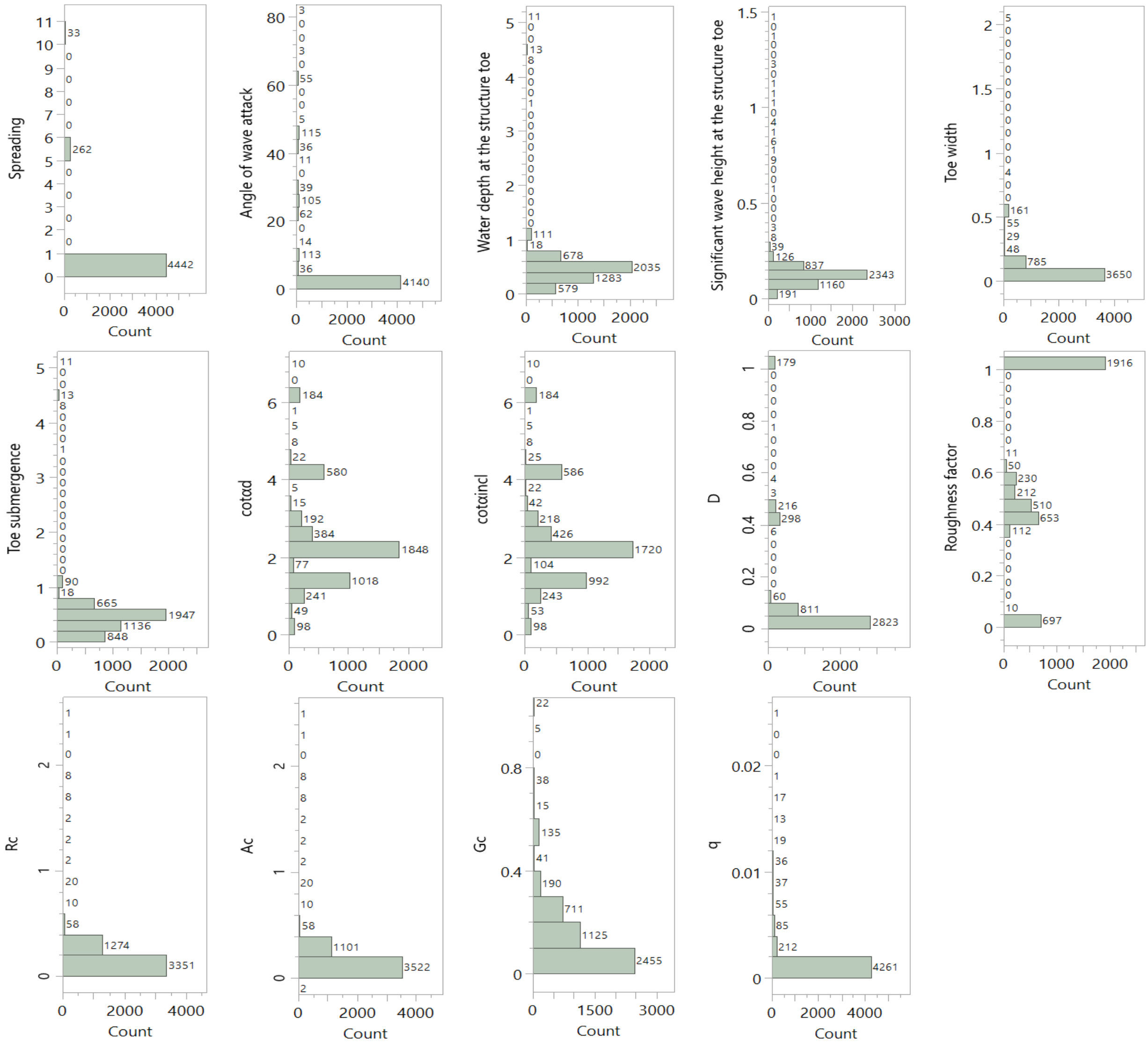

Fig. 2. Parameter spectrum of the assembled database.

B. Methods

ML is widely utilized across a diverse range of applications and disciplines due to their ability to make informed decisions and predictions using recently acquired data [14-17]. ML algorithms undergo repeated training iterations to adapt and identify patterns. Enhanced algorithms frequently yield greater precision in decision-making and predictive results. There are several factors to consider when selecting an algorithm, including the scale of the training data, the precision and interpretability of the output, the duration of the training process, and the number of features involved. ML can be classified into three distinct groups: supervised ML, unsupervised ML, and reinforcement and semi-supervised learning [16]. This study deployed supervised ML due to its ability to accurately predict continuous variables, often known as regression. Specifically, this study assessed the performance of GBT, DT, and SVM.

The SVM technique was initially introduced as a solution for pattern recognition tasks. Over time, it has been further developed to handle non-linear regression issues by incorporating the ε-insensitive loss function [17]. SVMs are well recognized as robust instruments that operate on the notion of structural risk reduction, aiming at lowering the upper-bound risk functionally associated with generalization performance. By employing a non-linear mapping function, the SVM transforms the initial input data into a feature space of larger dimensionality, subsequently applying a basic linear function to it.

DT is a supervised ML model that is commonly used for classification tasks. It has a tree structure with branches that signify potential values associated with each node and nodes that represent characteristics of a categorized group [16]. The tree structure is easily comprehensible and provides a lucid framework to facilitate decision-making. However, there are various drawbacks connected to the former, including the issues of overfitting and errors resulting from bias and variation. An effective strategy to mitigate overfitting is to employ pre-pruning on the decision tree, whereby the growth of the tree is constrained to prevent it from reaching its maximum size [18].

The GBT algorithm combines many DTs to create a powerful predictive model. GBT refers to a collection of regression or classification tree models. This technique engages a collection of prediction models with low individual predictive power, often incorporating a DT algorithm, to generate a more robust and accurate prediction model. GBT operates by sequentially constructing each subsequent tree and taking advantage of the errors made by the previous tree for learning purposes. The iterative procedure of detecting and revising the pattern is subsequently performed until no further pattern can be discerned, the cumulative discrepancy between the observed and predicted values converges to zero, and the projected values converge to the actual values [19].

III. PREPARE RESULTS AND DISCUSSION

The careful consideration of input and output variables is a fundamental part of the development of an ML model. A subset of 13 parameters from the EurOtop database was chosen to develop ML-based models to predict wave overtopping. These selected parameters provide a succinct representation of the overtopping discharge test. Therefore, the fundamental data need to be dimensionless to mitigate significant fluctuations in the raw parameter values, thus improving the precision and dependability of ML models. All factors that define the heights of structures, specifically the vertical measurements of the toe and crest, are dimensionless with respect to the Hm0 toe. The dimensions of structure widths, namely the horizontal measures of the toe and crest, are all expressed in a dimensionless manner relative to the wavelength. The wavelength (*Lm1,0t*) can be calculated by using:

$$Lm1{,}0t = 1.56\, Tm1{,}0t^2 \tag{1}$$

The non-dimensional wave overtopping rate *Sq* is given by:

$$Sq = \frac{q}{\sqrt{g\ Hm0\ toe^{3}}} \tag{2}$$

The final input and output dimensionless parameters utilized in developing machine learning models are: *Spreading, β, h/Lm1,0t, Hm0 toe/Lm1,0t, ht/Hm0 toe, Bt/Lm1,0t, Cotα, cotaincl, γf, D/Hm0 toe, Rc/ Hm0 toe, Ac/ Hm0 toe, Gc/Lm1,0t, Sq*. The ML-based models were evaluated using Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), correlation coefficient (R), and Scatter Index (SI). These statistical indicators are provided as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left|Sq_{meas} - Sq_{pred}\right|^2 \tag{3}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Sq_{meas} - Sq_{pred}\right| \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|Sq_{meas} - Sq_{pred}\right|^2} \tag{5}$$

$$SI = \frac{RMSE}{Sq_{meas}} \tag{6}$$

$$R = \frac{\sum_{i=1}^{n}(Sq_{meas}-\overline{Sq_{meas}})(Sq_{pred}-\overline{Sq_{pred}})}{\sqrt{\sum_{i=1}^{n}(Sq_{meas}-\overline{Sq_{meas}})^2 \sum_{i=1}^{n}(Sq_{pred}-\overline{Sq_{pred}})^2}} \tag{7}$$

where Sq_{meas} and Sq_{pred} are the dimensionless measured and predicted values, n is the number of observations, and $\overline{Sq}_{meas}$ and $\overline{Sq}_{pred}$ are respectively the average of Sq_{meas} and Sq_{pred}.

The results of the ML models were derived utilizing hold-out validation based on the average results obtained for each set of test data. Using DT, the predicted wave overtopping results had MSE = 0.00001, MAE = 0.0013, RMSE = 0.0027, SI = 0.512, and R = 0.977. Figure 3 illustrates the correlation between the observed and estimated wave overtopping values employing the DT, GBT, and SVM models. When engaging the GBT model, the predicated wave overtopping results yielded MSE = 0.00001, MAE = 0.0011, RMSE = 0.002, SI = 0.392, and R = 0.991. These metrics show that the predicted values were relatively close to the corresponding measured values. The SVM model yielded MSE = 0.00002, MAE = 0.0017, RMSE = 0.0043, SI = 0.823, and R = 0.943.

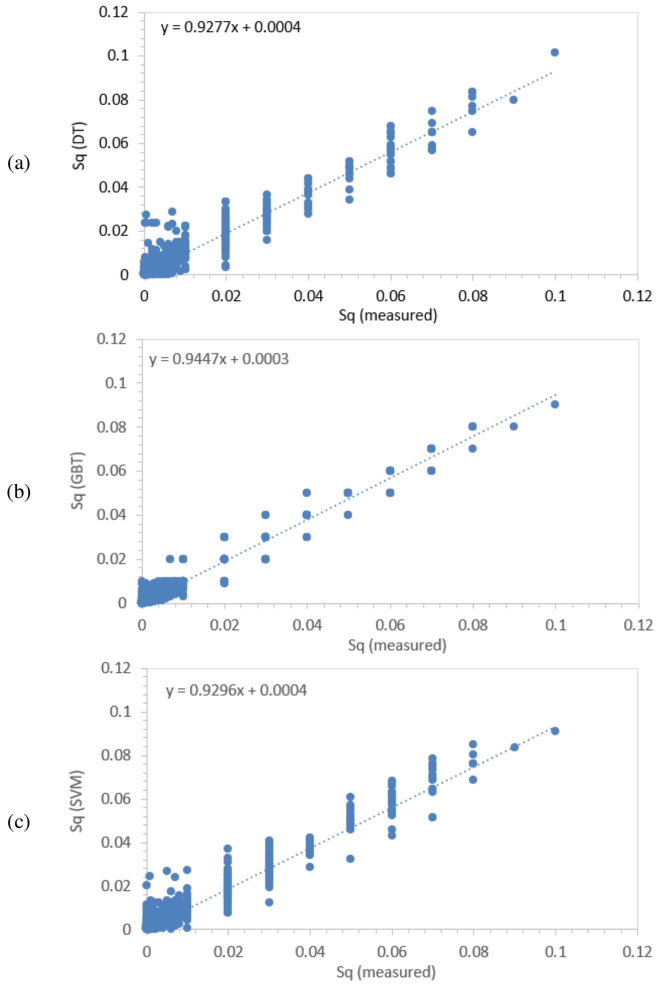

Fig. 3. Comparison between measured and predicted dimensionless overtopping discharge by DT (a), GBT (b), and SVM (c) models.

A. Comparison between DT, GBT, and SVM

Figure 3 displays the correlation between the observed and predicted wave-overtopping values deploying different ML-based models. In general, the predictive accuracy or effectiveness of a wave overtaking model is improved when the discrete points are situated in closer proximity to the trend line. The discrete data points for the GBT model exhibited a notable proximity to the trend line. In contrast, the discrete data points for the DT and SVM models were ranked as the second and third closest to the trend line, respectively.

Taylor diagrams are used to assess the appropriateness of various models, considering their performance in terms of R, RMSE, and standard deviation. Figure 4 depicts a Taylor diagram of the methods followed in this study. The Taylor diagram allows for the comparison and evaluation of anticipated and observed values based on their level of agreement. The horizontal and vertical axes demonstrate the standard deviation, the radial lines denote the R score, and the distance from the reference signifies the RMSE value. The evaluation of a model's accuracy is determined by the degree of proximity between each model and the corresponding actual value (reference). The reliability of a prediction model increases as its proximity to actual data increases. The GBT model manifested superior performance in terms of R score and standard deviation compared to the DT and SVM models. Furthermore, it demonstrated a lower RMSE. This suggests that the GBT model is more accurate in predicting real values. This observation is visually shown in Figure 4.

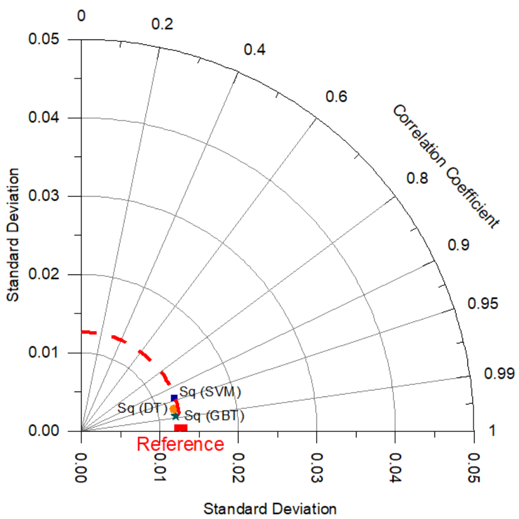

Fig. 4. Taylor diagram visualization of GBT, DT, and SVM models performance.

The GBT model had better predictive performance in terms of MSE, MAE, RMSE, SI, and R values, whereas the SVM model ranked second in terms of performance. The findings indicate that the GBT model outperformed the other models in terms of forecasting wave-overtopping discharge. Furthermore, these findings reveal that the GBT model significantly reduced overall error and effectively predicted wave-overtopping discharge.

B. Comparison of GBT with Previous Models

To further evaluate the GBT model, its performance was compared to the ANN models proposed in [20] and [6], and the results showed that the GBT model exhibited superior performance, as shown in Figure 5.

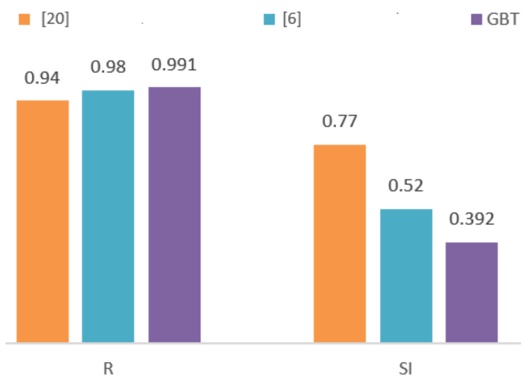

Fig. 5. Comparison between GBT, and models in [20] and [6].

IV. CONCLUSION

This study attempted to test the applicability of DT, GBT, and SVM models in predicting wave-overtopping discharge. These ML models were trained on 4,737 data points from the EurOtop database using 13 predictors out of the 31 parameters that are included in the database. The predicted performances of the models were assessed through the utilization of a Taylor diagram and six statistical indices. GBT demonstrated a high level of accuracy in estimating the discharge of wave-overtopping. The SI value of the GBT model (0.392) was found to be lower than that of DT (0.512) and SVM (0.823). Furthermore, the R-index value of the GBT model (0.991) was higher compared to that of DT (0.977) and SVM (0.943). Juxtaposed with the ANNs presented in [6] and [20], the GBT model significantly reduced prediction errors and had better prediction accuracy. These findings indicate that the GBT model significantly decreased the overall error and provided accurate estimations of the wave-overtopping discharge.

REFERENCES

[1] S. Dong, S. Abolfathi, M. Salauddin, Z. H. Tan, and J. M. Pearson, "Enhancing climate resilience of vertical seawall with retrofitting - A physical modelling study," *Applied Ocean Research*, vol. 103, Oct. 2020, Art. no. 102331, https://doi.org/10.1016/j.apor.2020.102331.

[2] M. Salauddin, J. J. O'Sullivan, S. Abolfathi, Z. Peng, S. Dong, and J. M. Pearson, "New insights in the probability distributions of wave-by-wave overtopping volumes at vertical breakwaters," *Scientific Reports*, vol. 12, no. 1, Sep. 2022, Art. no. 16228, https://doi.org/10.1038/s41598-022-20464-5.

[3] M. B. Mehrabani, H.-P. Chen, and M. W. Stevenson, "Overtopping failure analysis of coastal flood defences affected by climate change," *Journal of Physics: Conference Series*, vol. 628, no. 1, Mar. 2015, Art. no. 012049, https://doi.org/10.1088/1742-6596/628/1/012049.

[4] J. W. van der Meer, H. Verhaeghe, and G. J. Steendam, "The new wave overtopping database for coastal structures," *Coastal Engineering*, vol. 56, no. 2, pp. 108–120, Feb. 2009, https://doi.org/10.1016/j.coastaleng.2008.03.012.

[5] T. Pullen, N. W. H. Allsop, T. Bruce, A. Kortenhaus, H. Schüttrumpf, and J. W. van der Meer, "EurOtop wave overtopping of sea defences and related structures: assessment manual," *Die Küste*, 2007.

[6] S. M. Formentin, B. Zanuttigh, and J. W. van der Meer, "A Neural Network Tool for Predicting Wave Reflection, Overtopping and Transmission," *Coastal Engineering Journal*, vol. 59, no. 1, Mar. 2017, https://doi.org/10.1142/S0578563417500061.

[7] B. Zanuttigh, S. M. Formentin, and J. W. van der Meer, "Prediction of extreme and tolerable wave overtopping discharges through an advanced neural network," *Ocean Engineering*, vol. 127, pp. 7–22, Nov. 2016, https://doi.org/10.1016/j.oceaneng.2016.09.032.

[8] J. P. den Bieman, J. M. Wilms, H. F. P. van den Boogaard, and M. R. A. van Gent, "Prediction of Mean Wave Overtopping Discharge Using Gradient Boosting Decision Trees," *Water*, vol. 12, no. 6, Jun. 2020, Art. no. 1703, https://doi.org/10.3390/w12061703.

[9] J. P. den Bieman, M. R. A. van Gent, and H. F. P. van den Boogaard, "Wave overtopping predictions using an advanced machine learning technique," *Coastal Engineering*, vol. 166, Jun. 2021, Art. no. 103830, https://doi.org/10.1016/j.coastaleng.2020.103830.

[10] M. S. Elbisy, F. A. Osra, and Y. S. Alyafei, "Soft computing techniques for predicting wave overtopping discharges at vertical coastal structures," *GEOMATE Journal*, vol. 23, no. 98, pp. 205–211, Oct. 2022.

[11] A. H. Alshahri and M. S. Elbisy, "Prediction of Wave Overtopping Discharges at Coastal Structures Using Artificial Neural Networks and Support Vector Machine Tecniques," *GEOMATE Journal*, vol. 23, no. 99, pp. 56–62, Nov. 2022.

[12] Y. T. Tsai and C. P. Tsai, "Predictions of Wave Overtopping Using Deep Learning Neural Networks," *Journal of Marine Science and Engineering*, vol. 11, no. 10, Oct. 2023, Art. no. 1925, https://doi.org/10.3390/jmse11101925.

[13] M. S. Elbisy, "Machine learning techniques for estimating wave-overtopping discharges at coastal structures," *Ocean Engineering*, vol. 273, Apr. 2023, Art. no. 113972, https://doi.org/10.1016/j.oceaneng.2023.113972.

[14] IBM Cloud Education, "IBM Cloud Learn Hub," *Machine Learning. July*, 2020.

[15] K. Theofilatos, S. Likothanassis, and A. Karathanasopoulos, "Modeling and Trading the EUR/USD Exchange Rate Using Machine Learning Techniques," *Engineering, Technology & Applied Science Research*, vol. 2, no. 5, pp. 269–272, Oct. 2012, https://doi.org/10.48084/etasr.200.

[16] D. Dhall, R. Kaur, and M. Juneja, "Machine Learning: A Review of the Algorithms and Its Applications," in *Proceedings of ICRIC 2019*, 2020, pp. 47–63, https://doi.org/10.1007/978-3-030-29407-6_5.

[17] V. Vapnik, S. Golowich, and A. Smola, "Support Vector Method for Function Approximation, Regression Estimation and Signal Processing," in *Advances in Neural Information Processing Systems*, 1996, vol. 9.

[18] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, no. 3, pp. 159–190, Nov. 2006, https://doi.org/10.1007/s10462-007-9052-3.

[19] T. Sharma, P. Gupta, V. Nigam, and M. Goel, "Customer Churn Prediction in Telecommunications Using Gradient Boosted Trees," in *International Conference on Innovative Computing and Communications*, 2020, pp. 235–246, https://doi.org/10.1007/978-981-15-0324-5_20.

[20] M. R. A. van Gent, H. F. P. van den Boogaard, B. Pozueta, and J. R. Medina, "Neural network modelling of wave overtopping at coastal structures," *Coastal Engineering*, vol. 54, no. 8, pp. 586–593, Aug. 2007, https://doi.org/10.1016/j.coastaleng.2006.12.001.