



Taeyoon Kim<sup>†</sup>, Soonchul Kwon \*<sup>1</sup> and Yongju Kwon<sup>†</sup>

Department of Civil and Environmental Engineering, Pusan National University, 2, Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan 46241, Korea; rlaxodbs3@pusan.ac.kr (T.K.); zenga11@pusan.ac.kr (Y.K.) \* Correspondence: sckwon@pusan.ac.kr; Tel.: +82-515107640

+ These authors equally contributed to this work.

Abstract: The adoption of low-crested and submerged structures (LCS) reduces the wave behind a structure, depending on the changes in the freeboard, and induces stable waves in the offshore. We aimed to estimate the wave transmission coefficient behind LCS structures to determine the feasible characteristics of wave mitigation. In addition, various empirical formulas based on regression analysis were proposed to quantitatively predict wave attenuation characteristics for field applications. However, inherent variability of wave attenuation causes the limitation of linear statistical approaches, such as linear regression analysis. Herein, to develop an optimization model for the hydrodynamic behavior of the LCS, we performed a comprehensive analysis of 10 types of machine learning models, which were compared and reviewed on the prediction accuracy with existing empirical formulas. We found that, among the 10 models, the gradient boosting model showed the highest prediction accuracy with MSE of  $1.0 \times 10^{-3}$ , an index of agreement of 0.996, a scatter index of 0.065, and a correlation coefficient of 0.983, which indicates a performance improvement over the existing empirical formulas. In addition, based on a variable importance analysis using explainable artificial intelligence, we determined the significant importance of the input variable for the relative freeboard ( $R_C/H_0$ ) and the relative freeboard to water depth ratio ( $R_C/h$ ), which confirms that the relative freeboard was the most dominant factor for influencing wave attenuation in the hydraulic behavior around the LCS. Thus, we concluded that the performance prediction method using a machine learning model can be applied to various predictive studies in the field of coastal engineering, deviating from existing empirical-based research.

**Keywords:** low-crested structure; machine learning; gradient boosting; wave attenuation; explainable artificial intelligence

# 1. Introduction

Artificial structures for wave mitigation, such as breakwaters, headlands, detached breakwaters, and submerged breakwaters, are utilized to control coastal erosion problems by reducing incident wave energy and reducing sediment transports. Recently, shoreline deformation from beach erosion and scouring by coastal development has been rapidly increasing, along with sea level rise and external force of storm wave increases due to climate change [1]. Among these, coastal erosion and sedimentation caused by morphological change can lead to changes in the natural environment and ecosystem of coastal areas [2,3]. These problems directly/indirectly affect various factors involved in local economic activities in related field such as fishery and tourisms.

Low-crested submerged structures (LCS), such as detached breakwaters and artificial reefs, reduce the wave behind a structure according to the change in the freeboard at the still water level, thereby protecting onshore environment by inducing the wave [4]. Since the geometrical specifications of such LCS should be set under conditions to obtain the target wave transmission coefficient, the calculation or prediction of the transmission coefficient



Citation: Kim, T.; Kwon, S.; Kwon, Y. Prediction of Wave Transmission Characteristics of Low-Crested Structures with Comprehensive Analysis of Machine Learning. *Sensors* 2021, 21, 8192. https:// doi.org/10.3390/s21248192

Academic Editor: Juan M. Corchado

Received: 5 November 2021 Accepted: 5 December 2021 Published: 8 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).



of a structure is an important factor in designing the structure. Various studies have been conducted to estimate the wave transmission coefficient of low-crested structures, and have proposed formulas for calculating the wave transmission coefficient [5–7]. However, most of the previously proposed formulas of wave transmission coefficients are estimated with assumptions and a regression analysis of mathematical experimental data, which inhibits the clear explanation of the natural phenomena. Existing models are limited to calculating wave transmission coefficients from a limited range of input parameters, resulting in less usefulness and applicability with low accuracy and reliability [8].

To cope with the limits, researchers have been conducting various studies to understand wave motion and wave characteristics around structures by applying numerical modeling [9,10] that are often used to model physical phenomena when analyzing hydrodynamic processes according to structural applications.

In order to simulate wave motion effectively and accurately, the non-hydrostatic effect must be considered. Ai et al. [11] simulated 3D free surface flow to suggest a new fully non-hydrostatic model. They verified the capabilities and numerical stability of the model through various test cases of surface wave motions. The proposed model presented the results of accurately and effectively resolving the motion of shortwaves with shoaling nonlinearity, dispersion, refraction, and diffraction. Ning et al. [12] used a fully nonlinear Boussinesq wave model to analyze runup characteristics according to solitary wave propagation conditions of fringing reefs. Furthermore, they presented results that reasonably reproduced the runup/rundown process for both non-breaking and breaking solitary waves through verification with laboratory experiment results. Hur et al. [13] analyzed hydrodynamic characteristics around permeable submerged breakwater using a 3D numerical scheme. They presented the detailed analysis results for distribution of wave height and wave breaking and average flow around the structure, considering the wave-structuresandy seabed interactions in the 3D wave field. Recently, OlaFlow, an open source licensed open foam-based toolbox, has been developed. OlaFlow is composed of the Reynolds Average Navier Stokes equation and the conservation of mass equation, and the free water surface is tracked using Volume Of Fraction (VOF). Olaflow is widely used to analyze the hydrodynamic characteristics around structures in the field of coastal engineering [14].

Recently, a machine learning model has drawn attention to determine and to predict statistical structures from input/output data using a computer model. Machine learning is an inductive method as a field of artificial intelligence, which finds rules through learning using data and results, rather than a traditional program method that derives results through rules and data. It can easily solve complex engineering problems, and provide regression analysis of nonlinear relationships [15]. Compared to other traditional regression methods, it employs a specific algorithm that can learn from the input data itself and provides very accurate results on the output [16]. In machine-learning-based prediction models, various deep learning algorithms are employed, such as neural networks, decision trees, support vector machines (SVMs), and gradient boosting. In recent years, research using machine learning algorithms have been continuously increasing in the field of coastal engineering, in particular, for problems related to modeling behavior around coastal structures [17,18]. Kim et al. [19] used a neural network to estimate the stability of a breakwater, and presented the improved structure stability compared with the existing empirical formula. Based on the experimental data of Van Der Meer et al. [20], Koc et al. [21] proposed a model to predict the stability of a breakwater using a genetic algorithm.

Herein, to improve the feasible application of machine learning models in the field of coastal conservation engineering, we investigated the application of various machine learning models to predict the wave transmission coefficient of LCS. In addition, we proposed a machine learning pipeline model that selects a machine learning model suitable for data characteristics and performs an overall analysis of the model. Finally, we evaluated the applicability of the machine learning model by analyzing the accuracy and errors associated with the formulas for calculating the wave transmission coefficient of the existing LCS.

#### 2. Methodology

# 2.1. Machine Learning Model

#### 2.1.1. Linear Regression Model

The linear regression model has the advantage that the parameters are linear and can be easily interpreted and analyzed quickly. Linear regression models were developed over 100 years ago and have been widely used over the past decades. However, a very restrictive shape results in low accuracy for data with nonlinear relationships. Linear regression creates a regression model using one or more characteristics and finds parameters w and b that minimize the mean squared error (MSE) between the experimental value (y) and predicted value ( $\hat{y}$ ) (Equations (1) and (2)).

$$\hat{\mathbf{y}} = \mathbf{w}[0] + \mathbf{x}[0] \times \mathbf{x}[1] + \dots + \mathbf{w}[p] \times \mathbf{x}[p] + \mathbf{b}$$
 (1)

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$
(2)

#### 2.1.2. Lasso Regression

In the existing linear regression method, overfitting with poor predictive performance may occur when new data are provided. To solve this problem, a lasso regression was developed using L1 regulation to forcibly constrain the model (Equation (3)).

$$E = MSE + penalty = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \alpha \sum_{j=1}^{m} |w_j|$$
(3)

Here, m is the number of weights, and  $\alpha$  is a penalty parameter that determines w and b that minimize the sum of the MSE and penalty terms.

### 2.1.3. Ridge Regression

Ridge regression is a model with an added L2 constraint to solve the overfitting problem of the linear regression model. The model not only fits the data of the learning algorithm, but also keeps the weights of the model as small as possible (Equation (4)).

$$E = MSE + penalty = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \alpha \sum_{j=1}^{m} |w_j|^2$$
(4)

The weights become zero in lasso regression, whereas in ridge regression, the weights become close to zero but not zero. The difference is that if some of the input variables are important, lasso regression will have a higher accuracy, and if the importance of the input variables is similar overall, the ridge model will have a higher accuracy.

## 2.1.4. SVM

The SVM was introduced by Boser et al. [22], inspired by the concept of statistical learning theory. The SVM is a method of finding a hyperplane composed of support vectors that can classify vectors of linearly different classes with the maximum margin for the distance between them [23]. The machine learning algorithm reflects data that cannot be classified linearly in a low-dimensional space in a high-dimensional space using a kernel function, and classifies it using a hyperplane. Representative types of kernel functions include polynomial, sigmoid, and radial basis function (RBF). In this study, the Gaussian RBF kernel was used and applied to the model [24].

#### 2.1.5. Gaussian Process Regression (GPR)

The GPR model is a probabilistic model based on nonparametric kernels. The Gaussian regression analysis model can be performed when the wave attenuation coefficient, which is the dependent variable, has a Gaussian shape [25]. Specifically, if a specific wave

attenuation coefficient ( $K_t^*$ ) is assumed as a random variable that includes an error, the expected wave attenuation coefficient with the error removed can be expressed as a covariance function between the mean and the error ( $K_t^* = K_t + \varepsilon$ ). Assuming that this error covariance can be interpreted as a kernel function, Bayesian analysis model can predict wave attenuation characteristics [26].

# 2.1.6. Ensemble Method

The ensemble method was created to improve the performance of the classification and regression tree (CART). The method creates a more accurate prediction model by creating several classifiers and combining their predictions. In other words, the method derives a highly accurate prediction model by combining several weak classifier models, and not using a single strong model. Ensemble models can be broadly divided into bagging and boosting models. Bagging method reduces variance by using average or voting methods for the results predicted by various models, and boosting method synthesizes weak classifiers into strong classifiers. In this study, we performed a predictive study using boosting and random forest (RF) ensemble methods.

## (1) Random Forest (RF)

RF is a method that is employed to improve defects such as the variance and the performance fluctuation range of the decision tree being large. RF combines the concept and properties of bagging with randomized node optimization to overcome the shortcomings of existing decision trees and improve the generalization performance. In line with bagging method, the process of extracting bootstrap samples and creating a decision tree for each bootstrap is similar, but instead of selecting the optimal partition within all predictors for each node, the RF randomly extracts predictors and creates optimal partitions within the extracted variables [27]. In other words, RF creates several learners of low importance since it determines slightly different training data through bootstrap to give maximum randomness, and simultaneously combines the randomization of predictors. Important hyperparameters of RF include max\_features, whether to use bootstrap, and n\_estimator. The max\_features parameter represents the maximum number of features to be used in each node, bootstrap is the option to allow duplication in data sampling conditions for each nodel [28].

# (2) Boosting method

Boosting is a method that is used to create strong classifiers from a few weak classifiers, and is a model created by further boosting the weights on the data at the boundary. AdaBoost is the most common and widely used ensemble learning algorithm, and is specifically one of the boosting families of ensemble learning. The main feature of AdaBoost is that after generating a weak classifier using initial training data, the distribution of the training data is adjusted according to the prediction performance dependents on the weak classifier training. The weight of the training sample with low prediction accuracy was increased by using the information received from the classifier in the previous stage. In other words, the method improves learning accuracy by adaptively changing the weights of samples with a low prediction accuracy in the previous classifier. The method combines these weak classifiers with low prediction performance to create a strong classifier with slightly better performance. Gradient boosting method, which is applied in this study, also sequentially adds multiple models in the same way as the Adaboost model [29]. The biggest difference between the two algorithms is the recognition of weak classifiers. While AdaBoost recognizes values that are more difficult to classify by weighting them, Gradient Boost uses a loss function to classify errors. In other words, the loss function is an indicator that can evaluate the performance of the model in learning specific data, and the model result can be interpreted differently depending on which loss function is used.

## 2.2. Analysis of Machine Learning Model

# 2.2.1. Performance Measurement

The correlation coefficient, which indicates the correlation between the predicted output value and the measured value of the model, is an important factor for evaluating the predictive performance of a machine learning model. To analyze the predictive performance of the model, in this study, we measured the performance of the model using the mean square error (MSE), index of agreement (I), scatter index (SI), and  $\mathbb{R}^2$ , which represent the correlation coefficient. For each dataset, the correlation coefficient between the experimental and predicted values is as shown in Equations (5)–(8).

$$MSE = \frac{\sum_{i=1}^{n} (x_i - y_i)^2}{n}$$
(5)

$$I = 1 - \frac{\sum_{i=1}^{n} (y_i - x_i)^2}{\sum_{i=1}^{n} (|y_i - \bar{x}| + |x_i - \bar{x}|)^2}$$
(6)

$$SI = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_{i} - y_{i})^{2}}}{\bar{x}}$$
(7)

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (x_{i} - y_{i})^{2}}{\sum_{i=1}^{n} (x_{i} - \bar{y})^{2}}$$
(8)

Here,  $x_i$  and  $y_i$  are the experimental and predicted values, respectively,  $\bar{x}$  and  $\bar{y}$  are the mean values of the experimental and predicted values, respectively, and n is the sample number. Statistically, the closer  $R^2$  and I are to 1 and the smaller the MSE and SI, the higher is the reliability.

# 2.2.2. Analysis Method of Feature Importance

# (1) eXplainable Artificial Intelligence (XAI)

XAI was developed to help users understand the overall characteristics of how an AI system works and correctly interprets the final result. XAI is a surrogate model that makes possible to explain the process of calculating results for the correlation between input variables and dependent variables by determining the major factors that affect the prediction of a machine learning model. The interpretation of such a machine learning model is an important analysis method for deriving a suitable learning model according to various conditions or to increase the prediction stability of the model through quantitative analysis of predicted values through input variables.

The analysis method of the artificial intelligence system analyzes the characteristics of input variables to interpret the model learning and prediction process, and is divided into global and local interpretations. Global interpretation is a method of interpreting the overall analysis process and results of a model, and local interpretation is a method of interpreting model predictions for a single observation or part of a data set, and interpreting the results derived from the model for one specific input data.

## (2) Shapley Additive exPlanations (SHAP)

The Shapley value is the mean value of the marginal contribution for all possible sets to understand the importance of one characteristic based on game theory (theorizing about what decisions or actions each other takes in situations where multiple themes influence each other). Lundberg et al. [30] developed the SHAP machine analysis model, which achieves the highest accuracy, with a solid theoretical background among the machine learning analysis models that have been released to date, and the Shapely value is given in Equation (9).

$$\varnothing_{i} = \sum_{S \in N/(i)} \frac{|S|!(n-|S|-1)!}{n!} (f(S \cup \{i\} - f(S))$$
(9)

Here, S is a subset of the features used in the model,  $\{i\}$  is the vector of feature values of observations for explanation, n is the number of characteristics. f(S) is the value obtained by subtracting the predicted value from one observation from the average predicted value obtained from the data for a combination of feature values.

SHAP values are obtained by using the conditional expected value function of the machine learning model for Shapley values. The Shapley values for all input features are obtained, and the SHAP values can be interpreted locally and globally with the SHAP mean for each feature for each observation. The input feature importance can be expressed by visualizing the input feature based on the dataset through the average or sum of the absolute values of the SHAP values. In the case of the partial dependence plot provided by SHAP, the value of the input characteristic of each instance and the corresponding SHAP values are expressed as dots for all instances, and the average of the predicted values is calculated by changing the specific characteristic value of each instance. In this study, we analyzed the characteristics of a machine learning model built using the SHAP model.

# 2.3. Empirical Formula of Wave Transmission Coefficient

The wave transmission coefficient represents the ratio of the incident wave height before passing through the LCS and the average wave height after passing through the LCS (Figure 1).





Existing theoretical and empirical equations for the LCS are based on the experimental results of the hydraulic model, and many researchers have proposed empirical equations to predict the wave transmission coefficient using experimental data [31–33].

The suggested equation for the wave transmission coefficient by D'Angremond et al. [32] is as follows (Equations (10) and (11)):

$$K_{t} = -0.4 \frac{R_{c}}{H_{i}} + 0.64 \left(\frac{B}{H_{i}}\right)^{-0.31} \left(1 - e^{-0.5\xi}\right), \qquad \frac{B}{H_{i}} < 8$$
(10)

$$K_{t} = -0.35 \frac{R_{c}}{H_{i}} + 0.51 \left(\frac{B}{H_{i}}\right)^{-0.65} \left(1 - e^{-0.41\xi}\right), \quad \frac{B}{H_{i}} > 12$$
(11)

Here,  $R_c$  is the crest freeboard,  $H_i$  is the incident wave height, B is the crown width  $\xi$  is the surf similarity coefficient for breakwater ( $\xi = \tan \alpha / \sqrt{H_i/L_0}$ ). However, in the aforementioned equation, the effective range of the wave transmission coefficient is limited to 0.075–0.8. Van der Meer [31] suggested a wave transmission coefficient equation based on the breakwater coefficient to improve the accuracy of the wave transmission coefficient (Equations (12) and (13)).

$$K_t = -0.3 \frac{R_c}{H_i} + 0.75 \left(1 - e^{-0.5\xi}\right), \quad \xi < 3$$
(12)

$$K_{t} = -0.3 \frac{R_{c}}{H_{i}} + 0.75 \left(\frac{B}{H_{i}}\right)^{-0.31} \left(1 - e^{-0.5\xi}\right), \quad \xi > 3$$
(13)

Bleck and Oumeraci [33] proposed the exponential decay equation of the wave transmission coefficient of the LCS, according to the relative freeboard (Equation (14)).

$$K_t = 1 - 0.83e^{0.72 \frac{\kappa_c}{H_i}} \tag{14}$$

In previous studies, factors such as relative freeboard  $\left(\frac{R_c}{H_i}\right)$ , relative crest width  $\left(\frac{B}{H_i}\right)$ , front slope of structure (tan  $\alpha$ ), and wave steepness ( $H_i/L_0$ ), are classified as the factors related to wave decay around the LCS, and the empirical formula for this is presented. Figure 1 shows cross-section of LCS structure. Herein, the crest freeboard ( $R_C = h_c - h$ ) indicates the differences between structure depth and water depth, which has a positive value in emerged state and negative value in submergence state in still-water level.

In this study, we compared and reviewed the results of calculating the wave transmission coefficient using the existing empirical formula (Equations (10)–(14)), and the prediction results using a machine learning model.

## 2.4. Model Design Condition and Method

## 2.4.1. Machine Learning Automatic Pipeline Model

In this study, we applied 10 machine learning models, namely linear regression, kernel ridge (KR), ridge, lasso, GPR, SVM, RF, artificial neural network (ANN), gradient boosting regressor (GBR), and AdaBoost, to compare and review the performance dependencies on the characteristics of each model. To determine the optimal conditions of the automatic pipeline model dependent on input data characteristics, we adjusted hyperparameters using Grid-searchCV, and constructed automatic models for 10 machine learning models using the scikit-learn pipeline. The optimal machine learning model, selected through the automatic model, was analyzed to determine the importance of variables affecting the wave control of LCS using the machine learning analysis package SHAP.

#### 2.4.2. Machine Learning Model Configuration and Input Conditions

The 260 items of input data were obtained and applied in this study, with reference to the results of hydraulic model experiments with existing LCS by Seelig [34], Daemrich and Kahle [35], van der Meer [20], and Daemen [36]. Data on the wave transmission coefficient were obtained from DELOS database for permeable structures. The 260 data consist of:

- 81 data on rubble mound emerged/submerged breakwater [Seelig];
- 95 data on tetrapod submerged breakwater [Daemrich and Kahle];
- 31 data on rubble mound emerged/submerged breakwater [Van der Meer];
- 53 data on rubble mound emerged/submerged breakwater [Daemen].

In the data applied to the model, wave attenuation characteristics behind the structure were analyzed using random wave, and the data in range of 0.021 to 0.231 m were applied to wave height, and 0.91–3.66 s to wave period (Table A1). Through various studies, research results on the wave attenuation mechanisms of LCS and various factors that have a dominant influence on wave attenuation are presented. Van der Meer [31] proposed a wave

transmission coefficient equation using  $R_c/H_0$ ,  $B/H_0$  and  $\xi$ . Shin et al. [37] suggested the wave transmission coefficient equation using  $B/L_0$ ,  $R_c/h$ , and Gadomi et al. [38] performed research on porosity and  $h_c/h$ . Therefore, in this study, we used seven dimensionless numbers (X = {X<sub>1</sub>, X<sub>2</sub>,..., X<sub>7</sub>}) as input variables (Table 1) based on previous studies. Here,  $R_c/H_0$  is the relative freeboard,  $B/H_0$  is the relative crest width,  $\xi$  is the surf similarity parameter,  $B/L_0$  is the ratio of the crest width to the wavelength,  $R_c/h$  is the relative freeboard to water depth ratio,  $Dn_{50}/h_c$  is the ratio of the nominal diameter to the crest height, and  $h_c/h$  is the relative structure height. Herein,  $D_{n50}$  means nominal diameter, which is the ratio of median mass of unit ( $M_{50}$ ) and mass density of the rock ( $\rho_r$ ) ( $D_{n50} = (M_{50}/\rho_r)$ . Therefore,  $Dn_{50}/h_c$  parameter is a factor related to the effect of voids as the ratio of the structure height to the nominal diameter. Surf similarity parameter ( $\xi = \tan \alpha/\sqrt{H_i/L_0}$ ) represents the ratio of the front slope (tan  $\alpha$ ) and wave slope ( $H_i/L_0$ ), which is an important parameter in relation to wave breaking. The front slope of the structure applied in this study was in the range of 1:1.38–1:4, and various slope conditions were considered.

Table 1. Definitions and ranges of scaled model parameter.

	Parameter	Definition	Average	Max	Min
X <sub>1</sub>	$R_C/H_0$	Relative freeboard	-0.494	4.0	-8.696
X2	$B/H_0$	Relative crest width	4.525	43.478	0.889
$X_3$	ξ	surf similarity parameter	4.145	10.541	1.181
$X_4$	$B/L_0$	Ratio of the crest width to wave length	0.09	0.424	0.012
$X_5$	R <sub>C</sub> /h	the relative freeboard to water depth ratio	-0.065	0.734	-0.56
X <sub>6</sub>	$Dn_{50}/h_c$	Ratio of the nominal diameter to structure height	0.166	0.336	0.065
$X_7$	h <sub>c</sub> /h	The relative structure height	0.935	1.734	0.44
Y	K <sub>t</sub>	Transmission coefficient	0.482	0.922	0.049

Figure 2 depicts the statistical distributions of the input and output variables. To reflect the same feature scale, the input variable was converted to a range of 0 to 1 using max-min normalization.



Figure 2. Cont.



**Figure 2.** Results and discussion. (a):  $R_C/H_{0;}$  (b):  $B/H_{0;}$  (c):  $\xi$ ; (d):  $B/L_{0;}$  (e):  $R_C/h$ ; (f):  $Dn_{50}/h_c$ ; (g):  $h_c/h$ ; (h):  $K_t$ .

#### 3. Results and Discussion

3.1. Comparison of Machine Learning Model and Model Selection

Recently, research on the development and application of various machine learning techniques has been performed in the field of computer science. The performance of these machine learning models differs according to the characteristics of the input variables. Therefore, to model the fluid mechanical behaviors of the LCS, we analyzed the performance of the model by using 10 linear and nonlinear regression models. Figure 3 presents the performance results of the machine learning model of the artificial coral reef data derived from the machine learning pipeline model. Among the 10 machine learning models, GBR showed the highest model performance with an  $R^2 = 0.983$ , and the linear regression method showed the lowest performance with an  $R^2 = 0.814$ . Table 2 shows the model performance results based on the application of the 10 machine learning methods. The ensemble method (Adaboost, GBR, and RF), including the ANN method, showed the highest model accuracy (under  $1.3 \times 10^{-3}$  MSE) and model performance (>0.979 of R<sup>2</sup>). Among them, GBR yielded the highest model prediction accuracy. This indicates that the boosting method reinforces the weak classifier. In addition, the linear regression models (linear, ridge, and lasso) showed low accuracy, indicating that the application of the linear model cannot reflect the nonlinear characteristics of the data. Furthermore, it is believed that the linear model shows low model prediction performance when the model has nonlinearity between the input variable and the dependent variable. However, the performance of the model could be increased by regulating the  $L_1$  and  $L_2$  weights. We presented the wave transmission prediction results for the LCS using a machine learning model, as shown in Figure A1, which shows the distribution of the experimental values and prediction values of the test set. In terms of designing coastal structure, the estimation of the wave transmission coefficient with high accuracy is most important. As a result of comparing machine learning models, the GBR model shows the highest accuracy in terms of predicting the wave transmission coefficient for LCS structure. Therefore, we performed an analysis by applying the GBR model, which showed the highest accuracy in predicting the hydraulic characteristics around the LCS.



Figure 3. Performance analysis of machine learning models.

		Performanc	e Measures		
ML Regressor	MSE	Ι	SI	<b>R</b> <sup>2</sup>	
M1. AdaBoost	0.0013	0.995	0.073	0.979	
M2. Gradient boost	0.0010	0.996	0.065	0.983	
M3. ANN	0.0013	0.995	0.078	0.979	
M4. RF	0.0013	0.995	0.073	0.979	
M5. SVM	0.0022	0.991	0.101	0.965	
M6. GPR	0.0038	0.984	0.132	0.941	
M7. Lasso	0.0049	0.981	0.150	0.924	
M8. Ridge	0.0050	0.980	0.152	0.922	
M9. KŘ	0.0028	0.989	0.113	0.957	
M10. Linear	0.0121	0.944	0.235	0.814	
					ľ

Table 2. Proposed machine learning regressors and the resulting attenuation coefficient.

3.2. Model Performance Analysis

3.2.1. Results of Splitting a Dataset

To determine the most accurate parameter of the GBR model, we divided the collected data into training data and test data. Traditionally in machine learning, when the number of data is small, the training data and test data are divided by 7:3; however, recently, when the number of data is large, the dataset can be divided by 9:1. We divided the data set into 7:3, 8:2, and 9:1 conditions to perform sensitivity analysis on model accuracy. Table 3 shows the model performance results according to data splitting condition, and the highest  $R^2$  can be obtained under the conditions of 9:1 and 8:2. As the training set ratio increases, the number of training data increases, which allows the GBR model to produce a strong learner. However, since overfitting of the model and generalization of the model may be difficult due to insufficient data in the test set under 9:1 condition, we built the model by applying the 8:2 condition.

Splitting Ratio	MSE	Ι	SI	R <sup>2</sup>
7:3	$1.0  imes 10^{-3}$	0.996	0.065	0.983
8:2	$0.8 imes10^{-3}$	0.997	0.058	0.988
9:1	$0.8 imes10^{-3}$	0.997	0.051	0.988

Table 3. Performance of machine learning model according to splitting a dataset.

Figure 4 shows the prediction results of the training data and test data considering seven input variables using gradient boosting; the horizontal axis represents the experimental values, and the vertical axis represents the distribution of predicted values. As for the results, I was 0.999, SI was 0.032, and R<sup>2</sup> was 0.999 for the training data set, and the MSE was  $0.8 \times 10^{-3}$ , I was 0.997, SI was 0.058, and R<sup>2</sup> was 0.988 for the test data set, indicating the excellent prediction performance for the wave transmission coefficient. As a result, it is deemed that the performance prediction method using such a machine learning model can be applied to various predictive studies in the field of coastal engineering, deviating from existing empirical-based research.



**Figure 4.** Distribution of experimental and predicted values according to the GBR model application. (a): Test set. (b): Training set.

# 3.2.2. 10-Fold Validation Analysis

To verify the model performance of the GBR model, we utilized 10-fold cross-validation. This method was developed to minimize the bias associated with random sampling of the training set. The entire data sample was divided into 10 parts: nine were used for training, and one was used for model validation. The process of cross-validation was performed ten consecutively. The 10-fold cross-validation method ensured the generalization and reliability of the model performance. Figure 5 shows the model performance results obtained using the 10-fold cross-validation method. Figure 5a shows the results of R<sup>2</sup> according to each fold and shows slight fluctuations; however, the minimum and maximum values were 0.958 and 0.987, respectively. Figure 5b shows the minimum value of  $0.97 \times 10^{-3}$ , and the maximum value of  $2.70 \times 10^{-3}$  for MSE, showing that all errors are minimal, and a high level of accuracy is maintained.



Figure 5. Results for 10-fold cross validation. (a) R<sup>2</sup>; (b) MSE; (c) MAE; (d) MAPE.

Table 4 shows the model performance and statistical information using the 10-fold cross-validation method. The mean  $R^2$  is 0.973, and the std is 0.009, demonstrating that the results have a small deviation. In addition, the MAE and MAPE are 0.027 and 0.080, respectively, indicating small prediction errors.

E-14-		Performance Measures					
Folds	MSE	<b>R</b> <sup>2</sup>	MAPE	MAE			
Fold 1	$1.4 imes10^{-3}$	0.971	0.073	0.029			
Fold 2	$1.5 imes10^{-3}$	0.969	0.060	0.027			
Fold 3	$1.0 imes10^{-3}$	0.987	0.076	0.024			
Fold 4	$1.4 imes10^{-3}$	0.979	0.080	0.024			
Fold 5	$1.1 imes 10^{-3}$	0.982	0.056	0.023			
Fold 6	$2.1 imes10^{-3}$	0.965	0.114	0.030			
Fold 7	$1.5 imes10^{-3}$	0.979	0.082	0.026			
Fold 8	$2.7 imes10^{-3}$	0.958	0.070	0.034			
Fold 9	$1.6 imes10^{-3}$	0.961	0.108	0.033			
Fold 10	$1.3 imes10^{-3}$	0.978	0.085	0.027			
Mean	$1.6 imes10^{-3}$	0.973	0.080	0.027			
STD	$0.4 imes10^{-3}$	$9.0 imes10^{-3}$	$18.7 imes10^{-3}$	$3.8 imes10^{-3}$			

Table 4. Performance of 10-fold cross validation.

3.2.3. Evaluation of GBR Model Using Another Data Set

To verify the generalization of developed model additionally, we applied new data set to the model. For the verification for the GBR model, we used 41 data sets from Delft Hydraulics [39] and Allsop [40] as data.

- 20 data on rubble mound LCS (submerged) [Delft Hydraulics].
- 21 data on rubble mound LCS (emerged) [Allsop].

Figure 6 Shows the prediction results for new data set using GBR model. For the new data set, the relationship between the predicted and tested values converges closely to the ideal y = x line, which shows the good prediction results of the proposed model. Quantitative performance measurement showed high accuracy with  $R^2 = 0.93$ , MSE =  $2.3 \times 10^{-3}$ , I = 0.98, SI =  $5.2 \times 10^{-3}$ , which suggests that the GBR model actually has high accuracy in predicting the wave transmission coefficient of LCS even for completely new data sets.



Figure 6. Results for the new data set.

# 3.2.4. Feature Importance Analysis

Around the coastal structures, the wave attenuation effect is not an action that is independent of the input variables, but rather a complex interaction dependent on the variables. Thus, the relative importance of each variable in the model unit to the total observations should also be analyzed. Since the importance of the input variable is a measure of how much the variable affects the dependent variable, analysis for the correlation between the input and dependent variables is important. Therefore, we analyzed the importance of variables that affect the wave attenuation of the LCS. Figure 7 shows the importance of input variables that affect the dependent variable (wave transmission coefficient) when applying the 260 hydraulic model experiment results to the GBR model. Figure 7a shows the variable importance, and the x-axis represents the average of the absolute values of the Shapley values of the input variables throughout the data. In short, this means that the average influence of the input variable on the dependent variable, and the larger the x-axis value, the greater the influence on wave attenuation.

As a result of the variable importance analysis, the SHAP value of relative freeboard  $(R_c/H_0)$  was 0.116, which verifies that  $R_c/H_0$  was the most dominant parameter for wave control and wave energy reduction in hydrodynamics behavior around the LCS. Next, the relative freeboard to water depth ratio  $(R_c/h)$  and relative structure height  $(h_c/h)$  were 0.062, 0.042, respectively, and the results show that the input variable related to the freeboard has a dominant influence over 80% of the total wave height attenuation behind the structure. The freeboard should be prioritized for wave control in the design of the structure, as the attenuation effect of wave energy passing over the LCS along with wave breaking increases as the freeboard increases. The SHAP values of the ratio of the crest width to wavelength  $(B/L_0)$  and relative crest width  $(B/H_0)$  were 0.023 and 0.017, respectively, indicating that the input variable related to the crest width has an effect of

more than 8.6% of the total on the wave attenuation. Figure 7b shows a summary plot combining the feature importance and feature effects of the input variables. Here, they are arranged in order of importance, so that the one with the highest feature importance is placed at the top. The stronger the red shading of the corresponding feature value, the more positive is the influence on the wave transmission coefficient (K<sub>t</sub>), and the stronger the blue shading, the more negative is the influence. As a result, as  $R_c/H_0$ ,  $R_c/h$ ,  $h_c/h$ ,  $B/L_0$ , and  $B/H_0$  increased, the wave transmission coefficient decreased, and as the surf similarity coefficient ( $\xi$ ) increased, the wave transmission coefficient tended to increase. The results showed that this sensitivity trend was in line with engineering practice and physical background.



Figure 7. Analysis of feature importance. (a): SHAP feature importance. (b) Summary plot (feature effects).

#### 3.2.5. Influence of Input Variable Number

In this study, we analyzed the model accuracy by applying an input variable consisting of seven dimensionless numbers (X = { $X_1:R_c/H_0, X_2:B/H_0, X_3:\xi, X_4:B/L_0, X_5:R_c/h, X_6:Dn_{50}/h_c$ , and  $X_7:h_c/h$ }). If it is possible to build a model with high accuracy by excluding insignificant input variables and constructing a model with only important input variables, it is also possible to reduce computational complexity and to derive good results in terms of time efficiency. Accordingly, we analyzed the effects on model performance when some input variables or data were not reflected through various combinations. Table 5 presents the model performance results based on the eight combinations of input variables, and Figure 8 presents the results of the predicted values and experimental values for the eight combinations.

	Table 5.	Performance	measures fo	or analysis	s of differe	ent input v	variable c	ombinations.
--	----------	-------------	-------------	-------------	--------------	-------------	------------	--------------

	Performance Measures						
Combinations	MSET	I		SI		R <sup>2</sup>	
		Test	Train	Test	Train	Test	Train
1: X <sub>1</sub> , X <sub>2</sub> , X <sub>3</sub> , X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub> , X <sub>7</sub>	$0.8 imes10^{-3}$	0.997	0.999	0.058	0.032	0.988	0.999
2: X <sub>2</sub> , X <sub>3</sub> , X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub> , X <sub>7</sub>	$1.3 imes10^{-3}$	0.995	0.999	0.073	0.040	0.981	0.998
3: X <sub>1</sub> , X <sub>4</sub> , X <sub>5</sub> , X <sub>6</sub> , X <sub>7</sub>	$1.1  imes 10^{-3}$	0.996	0.999	0.066	0.070	0.984	0.995
4. X <sub>1</sub> , X <sub>2</sub> , X <sub>3</sub> , X <sub>4</sub> , X <sub>6</sub>	$1.6 imes10^{-3}$	0.994	0.999	0.082	0.041	0.977	0.998
5. X <sub>1</sub> , X <sub>4</sub> , X <sub>5</sub> , X <sub>7</sub>	$1.3 imes10^{-3}$	0.995	0.999	0.072	0.071	0.982	0.995
6. X <sub>2</sub> , X <sub>3</sub> , X <sub>4</sub> , X <sub>6</sub>	$22.5  imes 10^{-3}$	0.894	0.967	0.303	0.351	0.675	0.845
7. X <sub>2</sub> , X <sub>3</sub> , X <sub>6</sub>	$22.9  imes 10^{-3}$	0.889	0.962	0.307	0.370	0.668	0.820
8. X <sub>1</sub> , X <sub>5</sub> , X <sub>7</sub>	$3.7 imes10^{-3}$	0.986	0.987	0.123	0.230	0.947	0.948



**Figure 8.** Predicted and experimental values for the eight combinations. (a): Combination 1; (b): Combination 2; (c): Combination 3; (d): Combination 4; (e): Combination 5; (f): Combination 6; (g): Combination 7; (h): Combination 8.

Combination 1 showed the model performance results when applying pristine seven dimensionless input variables, indicating the highest accuracy with an MSE of  $0.8 \times 10^{-3}$ , and  $R^2$  of 0.988. In contrast, combination 7, which applied four input variables (X<sub>2</sub>:B/H<sub>0</sub>,  $X_3$ :  $\xi$ ,  $X_6$ :Dn<sub>50</sub>/h<sub>c</sub>) showed the lowest accuracy with an MSE of 22.9  $\times$  10<sup>-3</sup>, R<sup>2</sup> of 0.668. In addition, combination 8, which applied three input variables  $(X_1:R_c/H_0, X_5:R_c/h, X_7:h_c/h)$ , showed relatively high accuracy with an MSE of  $3.7 \times 10^{-3}$ , R<sup>2</sup> of 0.947, despite the application of a small number of input variables. It is worth noting that the accuracy of the model does not simply increase with an increase in the number of input variables, as can be seen from the combination 1-8 results. In addition, in combination 2, the relative freeboard  $(X_1:R_c/H_0)$ , which was classified as the most important factor in the sensitivity analysis, was not taken into account; however, combination 2 had a relatively high accuracy with an MSE of  $1.3 \times 10^{-3}$  and an R<sup>2</sup> of 0.981. Even if the relative freeboard (X<sub>1</sub>:R<sub>c</sub>/H<sub>0</sub>) is ignored, it is judged that combination 2 showed high accuracy by considering the factors  $(X_5:R_c/h, X_7:h_c/h)$  related to the freeboard. However, combinations 6–7, which ignored the factors related to the crest height ( $X_1:R_c/H_0$ ,  $X_5:R_c/h$ ,  $X_7:h_c/h$ ), showed low accuracy. In summary, the factors related to the freeboard  $(X_1:R_c/H_0, X_5:R_c/h, X_7:h_c/h)$  are the most important input variables to consider when obtaining predictions with high accuracy.

# 3.3. Comparison of Wave Transmission Coefficient Using Empirical Formulas and Machine Learning Models

Figure 9a,b show the results of substituting all experimental data into the empirical formula for the wave transmission coefficient of LCS at the low-crest submerged breakwater suggested by Van der Meer [31] and D'Angremond [32]. However, the results out of the effective range  $(0.075 < K_t < 0.8)$  suggested by the empirical formula were excluded. As a result of the prediction of the wave transmission coefficient using the Van der Meer empirical equation, the MSE was 0.009, and the determination coefficient ( $R^2$ ) was 0.81, indicating that the overall result of the empirical formula was overestimated, with respect to the experimental value (Figure 9a). As a result of the prediction of the wave transmission coefficient using the D'Angremond empirical formula, the MSE was 0.006 and the correlation coefficient (R<sup>2</sup>) was 0.84, showing fewer errors than the Van der Meer empirical formula and high prediction accuracy (Figure 9b). However, in the case of empirical formulas proposed by Van der Meer [31] and D'Angremond [32], applicable formulas are classified according to the surf similarity parameter ( $\xi = \tan \alpha / \sqrt{H_i/L_0}$ ) and the relative crest width  $(B/H_0)$ , and the effective range of the wave transmission coefficient is limited to  $0.075 < K_t < 0.8$ , which has a disadvantage in that uncertainty increases for other ranges. Figure 9c shows the results of substituting all experimental data into the empirical formula for the wave transmission coefficient of a low-crest submerged breakwater proposed by Bleck and Oumeraci [33]. In the case of the prediction results for the wave transmission coefficient using the empirical formula, the MSE was 0.017, and the determination coefficient  $(\mathbb{R}^2)$  was 0.71, indicating that the result of the empirical formula showed a low similarity to the experimental value overall. In addition, for the experimental value under 0.32, the transmission coefficient was 0.170, with a low prediction accuracy. The wave transmission coefficient of the LCS should consider the influence of various factors, such as crown depth, crown width, and porosity; however, the experimental formula of Bleck and Oumeraci [33] only considered the relative freeboard ( $R_C/H_0$ ), which led to a lower prediction accuracy than the other empirical formulas. Table 6 shows the results of comparing the statistical indicators of the existing empirical formula and the GBR model. Overall, all statistical indicators showed that the results of the boosting model showed higher prediction accuracy than that of the existing empirical formula. Furthermore, unlike the existing empirical formula, the boosting model does not need to set the effective range of the wave transmission coefficient, and does not require a separate formula dependent on the input variable (Figure 9d). In summary, a very accurate wave transmission coefficient can be predicted by inputting the seven input variables required by the machine learning model. This is because machine-learning models can interpret the non-linear relationships between independent and dependent variables. In the case of the empirical formula, analysis is



possible only in the effective range of the wave transmission coefficient, whereas when the GBR model is applied, it shows good predictive performance in all ranges.

**Figure 9.** Comparison of predictive performance of empirical formulas and machine learning model. (a): Comparison of wave transmission formulas (Van der Meer) with experimental data. (b) Comparison of wave transmission formulas (D'Angremond) with experimental data. (c) Comparison of wave transmission formulas (Bleck and Oumeraci) with experimental data. (d) Comparison of GBR model with experimental data.

Table 6. Statistical parameters of the results.

Methods	MSE	Ι	SI	R <sup>2</sup>
Van der Meer (2005)	$9.0 imes10^{-3}$	0.974	0.221	0.810
D'Angremond (1996)	$6.0 imes10^{-3}$	0.983	0.161	0.840
Bleck and Oumeraci (2001)	$17.0  imes 10^{-3}$	0.960	0.238	0.710
Gradient boosting	$0.8 imes10^{-3}$	0.997	0.058	0.988

## 4. Conclusions

In this study, we investigated the hydrodynamic performance modeling of a lowcrested structure using 10 machine learning models, including linear and non-linear models. To construct the model, we used 260 hydraulic model test data for training (80%) and prediction (20%). To predict the wave transmission coefficient behind the structure, we applied seven dimensionless parameters ( $R_C/H_0$ ,  $B/H_0$ ,  $\xi$ ,  $B/L_0$ ,  $R_C/h$ ,  $Dn_{50}/h_c$ , and  $h_c/h$ ). In addition, we evaluated the correlation between the input variable and dependent variable by analyzing the main factors that affect the prediction of machine learning models using XAI. The wave transmission coefficient for the linear model (M8, M9, and M10) among the machine learning models showed low prediction accuracy; however, the ensemble technique, the GBR model (M2) in particular, showed the highest accuracy to predict the wave transmission coefficient of a structure with a given input variable. To validate the machine learning models, we performed a 10-fold cross-validation, which indicates that the resulting  $R^2$  was 0.973, and the mean MAPE was 2.7%, confirming a significantly low prediction error. This small degree of error proves the generalization of the model reasonably. Based on the sensitivity analysis, we confirmed that the input variable for the relative freeboard ( $R_C/H_0$ ), and the relative freeboard to water depth ratio  $(R_C/h)$  show that the importance of independent variables is significant. As a result, freeboard was found to be the most dominant factor influencing wave attenuation in the hydraulic behaviors around the LCS. In addition, we comprehensively analyzed the results of the empirical formulas and machine learning models. In the wave transmission prediction of the trained gradient boosting model, the MSE was 0.8  $\times$  10<sup>-3</sup>, I was 0.997, SI was 0.058, and R<sup>2</sup> was 0.988, which indicates high prediction accuracy and improved wave transmission coeffcient prediction performance, compared with existing empirical results. Since the prediction using machine learning can perform analysis non-linearly, the wave transmission coefficient of a LCS can be predicted precisely and efficiently, in contrast to the regression method adopted by the exiting empirical formula. It is determined that the constructed machine learning automated pipeline model can be utilized for not only wave attenuation studies on LCS, but also various applications in coastal engineering.

**Author Contributions:** Conceptualization, T.K. and Y.K.; development, T.K. and S.K.; writing, T.K., S.K. and Y.K.; data analysis T.K. and Y.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2020R1A6A3A13069154), as well as through a grant (20011068) of the Regional Customized Disaster-Safety R&D Program funded by the Ministry of Interior and Safety (MOIS, Korea).

Conflicts of Interest: The authors declared no conflict of interest.

# Appendix A

Table A1. The indication of the range of the parameters.

Parameter	Definition	Average	Max	Min
H <sub>0</sub> (m)	Wave height	0.123	0.231	0.021
T <sub>0</sub> (s)	Wave period	2.017	3.660	0.910
$R_{C}(m)$	Crest freeboard	-1.081	0.196	-0.420
B (m)	Crest width	0.429	1.000	0.200
Dn <sub>50</sub> (m)	Nominal diameter	0.110	0.161	0.028
tan α	Slope of structure	0.507	0.667	0.250







Figure A1. Cont.



**Figure A1.** Measured versus predicted values for the various models. (a) Linear; (b) Ridge; (c) Lasso; (d) KR; (e) GPR; (f): SVM; (g) RF; (h) ANN; (i) AdaBoost; (j) Gradient boost.

# References

- 1. Mimura, N.; Kawaguchi, E. Responses of Coastal Topography to Sea-Level Rise. In Proceedings of the 25th International Conference on Coastal Engineering, Orlando, FL, USA; 1997; pp. 1349–1360.
- Moghaddam, E.I.; Hakimzadeh, H.; Allahdadi, M.N.; Hamedi, A.; Nasrollahi, A. Wave-induced currents in the northern gulf of Oman: A numerical study for Ramin prot along the Iranian Coast. *Am. J. Fluid Dyn.* 2018, *8*, 30–39.
- 3. Hseih, T.C.; Ding, Y.; Yeh, K.C.; Jhong, R.K. Investigation of Morphological Changes in the Tamsui River Estuary Using an Integrated Coastal and dEstuarine Processes Model. *Water* **2020**, *12*, 1084. [CrossRef]
- 4. Pilarczyk, K.W. Design of low-crested submerged structures—An overview. In Proceedings of the 6th International Conference on Coastal and Port Engineering in Developing Countries, Pianc-Copedec, Colombo, Sri-Lanka, 15–19 September 2003; pp. 1–16.
- Seabrook, S.R.; Hall, K.R. Wave transmission at submerged rubble mound breakwaters. In Proceedings of the 26th International Conference on Coastal Engineering, Copenhagen, Denmark, 22–26 June 1998; pp. 2000–2013.
- Tanaka, N. Wave deformation and beach stabilization capacity of wide-crested submerged breakwaters. In Proceedings of the 23rd National Conference on Coastal Engineering, Fukuoka, Japan, 25–26 November 1976; pp. 152–157. (In Japanese)
- 7. Buccino, M.; Calabrese, M. Conceptual approach for prediction of wave transmission at Low-Crested Breakwaters. J. Waterw. Port Coast. Ocean. Eng. 2007, 133, 213–224. [CrossRef]

- 8. Ahmadian, A.S.; Simons, R.R. Estimaion of nearshore wave transmission for submerged breakwaters using a data-driven predictive model. *Neural Comput. Appl.* **2016**, *29*, 705–719. [CrossRef]
- 9. Losada, I.J.; Losada, M.A.; Martin, F.L. Harmonic generation past a submerged porous step. *Coast. Eng.* **1997**, *31*, 281–304. [CrossRef]
- 10. Martinelli, L.; Zanuttigh, B.; Lamberti, A. Hydrodynamic and morphodynamic response of isolated and multiple low crested structures: Experiments and simulations. *Coast. Eng.* **2006**, *53*, 363–379. [CrossRef]
- 11. Ai, C.; Jin, S.; Lv, B. A new fully non-hydrostatic 3D free surface flow model for water wave motions. *Int. J. Numer. Meth. Fluids* **2011**, *66*, 1354–1370. [CrossRef]
- 12. Ning, Y.; Liu, W.; Sun, Z.; Zhao, X.; Zhang, Y. Parametric study of solitary wave propagation and runup over fringing reefs based on a Boussinesq wave model. *J. Mar. Sci. Technol.* **2018**, *24*, 512–525. [CrossRef]
- 13. Hur, D.S.; Lee, W.D.; Cho, W.C. Three-dimensional flow characteristics around permeable submerged breakwaters with open inlet. *Ocean Eng.* **2012**, *44*, 100–116. [CrossRef]
- 14. Higuera, P.; Lara, J.L.; Losada, I.J. Realistic wave generation and active wave absorption for Navier-Stokes models: Application to OpenFOAM. *Coast. Eng.* 2013, *71*, 102–118. [CrossRef]
- 15. Mitchell, T.M. Machine Learning; McGraw-Hill, Inc.: New York, NY, USA, 1997.
- 16. Salehi, H.; Burgueño, R. Emerging artificial intelligence methods in structural engineering. *Eng. Struct.* **2018**, *171*, 170–189. [CrossRef]
- 17. Kundapura, S.; Arkal, V.H.; Pinho, J.L. Below the data range prediction of soft computing wave reflection of semicircular breakwater. J. Mar. Sci. Appl. 2019, 18, 167–175. [CrossRef]
- 18. Kuntoji, G.; Manu, R.; Subba, R. Prediction of wave transmission over submerged reef of tandem breakwater using PSO-SVM and PSO-ANN techniques. *ISH J. Hydraul. Eng.* **2018**, *26*, 283–290. [CrossRef]
- 19. Kim, D.H.; Park, W.S. Neural network for design and reliability analysis of rubble mound breakwaters. *Ocean Eng.* **2005**, *32*, 1332–1349. [CrossRef]
- 20. Van der Meer, J.W. Rock Slopes and Gravel Beaches under Wave Attack. Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, 1990.
- 21. Koc, M.L.; Balas, C.E.; Koc, D.I. Stability assessment of rubble-mound breakwaters using genetic programming. *Ocean Eng.* **2016**, *111*, 8–12. [CrossRef]
- 22. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A Training Algorithm for Optimal Margin Classifiers. In *5th Annual ACM Workshop on COLT*; Haussler, D., Ed.; ACM Press: Pittsburg, PA, USA, 1992.
- 23. Steve, R.G. *Support Vector Machines for Classification and Regression;* Technical Report; University of Southampton: Southampton, UK, 1998.
- 24. Bruges, C.J.C. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Discov.* **1998**, 2, 121–167. [CrossRef]
- Rasmussen, C.E. Gaussian processes in machine learning. In Advanced Lectures on Machine Learning; Bousquet, O., von Luxburg, U., Ratsch, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 63–71.
- 26. Rasmussen, C.E.; Williams, C.K.I. Gaussian Processes for Machine Learning; The MIT Press: Cambridge, UK, 2010; 266p.
- 27. Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]
- 28. Ali, J.; Khan, R.; Ahmad, N.; Maqsood, I. Random Forests and Decision Trees. Int. J. Comput. Sci. Issues 2012, 9, 272–278.
- 29. Zhou, W.; Eckler, S.; Barszczyk, A.; Waese-Perlman, A.; Wang, Y.; Gu, X.; Feng, Z.-P.; Peng, Y.; Lee, K. Waist circumference prediction for epidemiological research using gradient boosted trees. *BMC Med. Res. Methodol.* **2021**, *21*, 47. [CrossRef] [PubMed]
- Lundberg, S.A.; Lee, S.I. Unified Approach to Interpreting Model Prediction. In Proceedings of the 31st Conference on Neural Information Processing System(NIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 1–10.
- 31. Van der Meer, J.W.; Briganti, R.; Zanuttigh, B.; Wang, B. Wave transmission and reflection at low-crested structures: Design formulae, oblique wave attack and spectral change. *Coast. Eng.* **2005**, *52*, 915–929. [CrossRef]
- 32. D'Angremond, K.; Van der Meer, J.W.; De Jong, R.J. Wave transmission at low-crested structures. In Proceedings of the 25th Coastal Engineering Conference, Orlando, FL, USA, 2–6 September 1996; pp. 2418–2426.
- Bleck, M.; Oumeraci, H. Hydraulic Performance of Artificial Reefs: Global and Local Description. In Proceedings of the 28th International Conference on Coastal Engineering, Cardiff, UK, 7–12 July 2002; pp. 1778–1790.
- Seelig, W.N. Two Dimensional Tests of Wave Transmission and Reflection Characteristics of Laboratory Breakwaters; Technical Report 80-1; Coastal Engineering Research Center: Fort Belvoir, VA, USA; Army Corps of Engineers Waterways Experiment Station: Vicksburg, MS, USA, 1980; p. 187.
- 35. Daemrich, K.; Kahle, W. Schutzwirkung von Unterwasser Wellen Brechern unter dem Einfluss Unregelmassiger Seegangswellen; Technical Report, Report Heft 61; Franzius-Instituts fur Wasserbau und Kusteningenieurswesen: Hannover, Germany, 1985.
- 36. Daemen, I.F.R. Wave Transmission at Low-Crested Structures. Master's Thesis, Delft University of Technology, Delft, The Netherlands, 1991.
- 37. Shin, S.W.; Bae, I.R.; Lee, J.I. Physical Modelling of the Wave Transmission over a Tetrapod Armored Artificial Reef. *J. Coast. Res.* **2019**, *91*, 126–130. [CrossRef]

- 38. Gandomi, M.; Dolatshahi Pirooz, M.; Varjavand, I.; Nikoo, M.R. Permeable Breakwaters Performance Modeling: A Comparative Study of Machine Learning Techniques. *Remote Sens.* **2020**, *12*, 1856. [CrossRef]
- 39. Hydraulics, D. *AmWaj Island Development, Bahrain: Physical Modelling of Submerged Breakwaters*; Report H4087; Delft Hydraulics: Delft, The Netherlands, 2002.
- 40. Allsop, N.W.H. Low-Crested Breakwaters, Studies in Random Waves. In Proceedings of the Coastal Structures'83, Arlington, VA, USA, 9–11 March 1983; pp. 94–107.