Proceedings of the 40th IAHR World Congress. Rivers – Connecting Mountains and Coasts
21 – 25 August 2023, Vienna, Austria
Edited by Helmut Habersack, Michael Tritthart and Lisa Waldenberger

doi: 10.3850/978-90-833476-1-5_iahr40wc-p0115-cd

Prediction of Wave Overtopping Rates at Sloping Structures Using Artificial Intelligence

M. A. Habib [(1)], S. Abolfathi [(2)] J.J. O'Sullivan [(1)] and M. Salauddin[(1*)]

[(1)] Dooge Centre for Water Resources Research, School of Civil Engineering, University College Dublin, Dublin 4, Ireland,
[(2)] School of Engineering, University of Warwick, CV4 7AL, Coventry, United Kingdom,
*e-mail md.salauddin@ucd.ie

Abstract

The prediction of wave overtopping at coastal defenses is critical to ensure the flood resilience of people and properties in low-lying nearshore coastal areas. With the effects of anthropogenic climate change, the frequency of wave overtopping is expected to increase, along with sea level rise and more frequent damaging storm surges. Established approaches for the prediction of wave overtopping have traditionally relied on physical and numerical modelling and empirical methods. The ubiquity of computational resources has led to the emergence of Artificial Intelligence techniques, such as Machine Learning (ML) algorithms, as a promising approach for predicting wave overtopping. This study investigates the application of four ML models based on Random Forest (RF), Gradient Boosted Decision Trees (GBDT), Support Vector Machines Regression (SVMR) and Artificial Neural Networks (ANN) approach for predicting wave overtopping at sloping breakwaters. Data from the EurOtop II manual, a comprehensive dataset of physical and numerical wave overtopping tests undertaken on a variety of coastal structure geometries, including sloping breakwaters (the focus of this study), underpinned the developed models.. To optimize the data for redundancy, feature transformation and advanced feature selection methods were employed. Hyperparameter tuning was performed to extract the best features for the predictive models. The performance of the developed ML-based models was examined in terms of the coefficient of determination, r^2, and the Pearson correlation coefficient, R, for the measured and predicted overtopping values. The range of r^2 values across the four models varied between 0.69 to 0.87, with Pearson correlations varying between 0.87 and 0.93. The results show that the GBDT model outperformed the other ML models tested in this study.

Keywords: Artificial Intelligence; Machine Learning; Overtopping; ANN; GBDT; Random Forest; SVMR; Sloping Structures.

1. INTRODUCTION

Overtopping occurs at coastal structures when wave heights surpass the height of the structure, causing water to flow over into the hinterland areas. Storm surges are the primary cause of wave height exceedance. Assessment of overtopping at coastal defense structures is vital for the functionality of the structures and for mitigating wave hazards. However, with the prediction of more frequent and intense storms, coupled with sea level rise and high tides (Salauddin et al., 2022; IPCC, 2021; Dong et al., 2021a,b), overtopping is likely to become a more common occurrence and critical issue in low-lying coastal areas. Thus, developing fast and reliable methods for estimating overtopping is essential for effective climate change mitigation and adaptation planning.

There are three popular methods of estimating overtopping namely empirical, numerical, and physical. In empirical methods, equations with arbitrary coefficients are applied to estimate overtopping quantity (e.g., O'sullivan et al., 2020; Dong et al., 2018; Salauddin and Pearson 2018; Goda 2009). In particular, EurOtop (2018) has a comprehensive collection of empirical equations that are a state-of-the-art guidance to estimate overtopping for coastal engineers across the world. Although empirical methods are fast and straightforward, they may not fully capture the underlying complexities of wave-structure interaction. Numerical methods are more capable of simulating wave propagation and interactions with the coastal structure using principles of computational fluid dynamics (e.g., Dang et al., 2023; Mata and van Gent, 2023; Ravindra et al., 2022; Chen et al., 2021; Yeganeh-Bakhtiary et al., 2020). The numerical methods can provide better insights into the overtopping processes but are computationally demanding in nature. Physical methods rely on experimental models involving the observation and measurement of overtopping at laboratory prototypes of coastal defense structures. (e.g., Liu et al., 2022; Salauddin et al., 2021; Dong et al., 2020; Salauddin and Pearson 2020).

The advancement in computational power and data science has enabled the application of Machine Learning (ML) algorithms for estimation of overtopping quantity. A review of the applications of ML-based models (Habib et al., 2022a,b) underpinned the necessity of a common methodological framework for ML applications. The study also identified Artificial Neural Networks (ANN) and Decision Trees as more popular methods for overtopping estimation across different geometries of coastal defense structures. The ability of ML algorithms to effectively identify complex and non-linear patterns in large datasets with high accuracy has made them a promising alternative for overtopping estimation. In fact, ANN has been recognized by EurOtop (2018) as a reliable method for this task. ANN is also in use for overtopping estimation for a significant amount of time in the recent past (see for example Verhaeghe et al., 2008; van Gent 2007; and Victor et al., 2012). The application of Decision Trees (DT) and Support Vector Machines Regression (SVMR) is also evident in recent literature (Elbisy 2023; den Bieman et al., 2021; Chen et al., 2021; Hosseinzadeh et al., 2021; den Bieman et al., 2020) and in all cases were able to predict overtopping with reasonable accuracy.
In this study, we examine the predictive performance of four ML algorithms, namely RF, GBDT, SVMR and ANN to estimate overtopping discharge at sloping structures. Among the four algorithms, ANN is a popular choice in overtopping prediction while the other three algorithms are relatively recent in practice. The data was sourced from the EurOtop (2018) manual which is an extensive collection of data from physical experiments for overtopping at sloping structures. As per the EurOtop (2018) manual, design and overtopping estimation methods for different geometrical configurations are independent and hence, it is possible to investigate the application of ML in sloping structures only. To ensure unbiasedness and eliminate redundancy in the data, data optimization techniques such as scalar transformation and feature selection methods were implemented. Notably, feature selection methods have replaced permutation methods for selecting the best features in Machine Learning (ML) prediction tasks. The dataset was then split into train-test set in the ration of 70%-30% and hyperparameter tuning was incorporated to ensure each of the algorithms fit well into the training data. The training data included validation sets prior to introducing the algorithms to an unseen test set. To determine the accuracy of the algorithms, predicted overtopping rates from the test set were compared with the actual values.

2. MATERIALS AND METHODS

The four ML algorithms used in this study, together with data sourcing, data optimization, and feature selection approaches are discussed in this section.

2.1 Decision Trees

Two of the four ML algorithms used in this study, GBDT and RF are characteristically Decision Trees (DT). Decision Trees are a type of supervised machine learning algorithm that is trained to predict an output variable (known as the dependent or target variable) from a set of independent variables (known as features). DTs can perform both regression and classification tasks. In regression tasks, a continuous or numerical output variable is prediction; for example: overtopping discharge. In classification tasks, the output variable is discrete, and the algorithm is trained to predict a class label (Yeganeh-Bakhtiary et al., 2022).

In regression-based DTs, the training data is recursively partitioned into rectangular regions and the output variable is predicted from the mean or median of these individual regions until a stopping criteria is reached. For example, let us consider a training dataset, X = $\{(x_1, y_1), (x_2, y_2) \ldots, (x_n, y_n)\}$, where x_i is an input feature vector for the i^{th} training data and y_i is the corresponding output variable. The DT algorithm then partitions *X* into a set of rectangular regions R_1, R_2, R_3.... R_n. Each region then has an associated prediction value P which is the mean or median of the output variables in that region. The DT is then finally built based on the input features that distinctly splits the rectangular regions and produces the smallest variance in the output variable. DTs are often favorably applied due to the algorithm's ability to perform prediction tasks independently from the noise and non-linearity of input data (Pedregosa et al., 2011; Kotu and Deshpande, 2015).

In a RF algorithm, a set of DTs is built from a random set of input features and training data. The purpose of an RF is to reduce overfitting and improve generalization by avoiding overexposure to any set of training data. Each DT is allowed to make individual predictions, and the final prediction is made by averaging the predictions of all the DTs. This approach is also known as the bagging technique. In addition to reducing overfitting, another advantage of an RF is that it can handle both categorical and numerical data. Another approach to improve the prediction performance of Decision Trees is the boosting technique. GBDT is an example of this kind. The boosting technique uses a loss function that measures the Mean Squared Error (MSE) between the predicted and actual values. The aim of the boosting algorithm is to minimize this loss function by adjusting numerical coefficients to input data during the training step. The process of minimizing the loss function is called gradient descending. This optimization of the loss function allows for faster and more accurate predictions from the DTs (Sutton, 2005).

2.2 Support Vector Machines

Support Vector Machine is also a supervised ML algorithm that can perform classification and prediction tasks. In a Support Vector Machine Regression (SVMR) algorithm, the input features are mapped to a high-dimensional space by applying a kernel function (Noori et al., 2022). In the next step, the algorithm builds a hyperplane to differentiate the input data into two distinct sets, namely the positive and negative regions. The data points that are hence placed in proximity to the hyperplane are known as support vectors. The objective of the SVMR is to maximize the distance between the hyperplane and support vectors, also known as the margin. The SVMR then applies a loss function to penalize the predictions that are placed outside the margin or on the wrong side of the hyperplane. The loss function of a SVMR is computed by Eqn. 1.

$$\mathrm{L}\,(\mathrm{y}, \mathrm{f}(\mathrm{x})) = \max\,(0, |\mathrm{y} - \mathrm{f}(\mathrm{x})| - \varepsilon) \quad [1]$$

where, y is the actual value and f(x) is the predicted value of the dependent variable and ε is the value of the margin. The loss function essentially penalizes predicted variables outside the margin and the margin can be fine-tuned using a constant known as the regularization parameter. The overall objective of the SVMR is to maximize the margin while simultaneously minimizing the loss function.

2.3 Artificial Neural Network

The concept of Artificial Neural Network (ANN) dates as far as 1943 when it was conceived through the work of McCulloh and Pitts (1943). As mentioned earlier, ANN is a widely practiced ML method in overtopping estimation which is evident from the studies of Formentin et al., (2017), Zanuttigh et al., (2016), van Gent (2007), etc. The authors in these studies opined that ANN is a reliable algorithm to perform prediction tasks in large overtopping datasets where the relationship between the independent and dependent variables is non-linear and/or unknown. However, uncertainty quantification of ANN -based models is important to ensure quality and robustness of the model's prediction. In this study, a feed forward and back propagation ANN algorithm is applied to the overtopping dataset. The feed forward and back propagation enables the ANN to minimize the loss function to the extent of a given threshold (Babaee et al., 2021). The number of hidden layers in an Artificial Neural Network (ANN) depends on the type and complexity of data in the input layer. For this study, we adopted two hidden layers, each consisting of 500 neurons.

2.4 Evaluation Scores

The accuracy of prediction by the four algorithms was expressed in terms of two statistical scores, namely the Coefficient of Determination r^2 and Pearson Correlation Coefficient *R*. The algorithms predicted overtopping discharge, 'q', per unit width of the structure. The overtopping discharge was then converted to a dimensionless quantity ($\frac{q\_predicted}{\sqrt{9.81*(H_{m0,t})^3}}$), where 'q_predicted' is the predicted overtopping discharge in m^3/s, 9.81 is the gravitational acceleration and $H_{m0,t}$ is the water depth at the toe of the structure. Similarly, the actual overtopping discharge from the EurOtop (2018) dataset was also converted to a dimensionless quantity of ($\frac{q\_\mathrm{actual}}{\sqrt{9.81*(H_{m0.t})^3}}$), where 'q_actual' is the actual overtopping discharge in m^3/s and other quantities and symbols have the same meaning as mentioned before. It is particularly important to use dimensionless quantities in scientific research as this measure ensures the physical process is reflected in the results and the results are independent from scale effects. Dimensionless quantities often contribute towards understanding the physics of different phenomena (Xiu et al., 2022).

The Coefficient of Determination (r^2) is expressed by Eqn. 2. The r^2 score accounts for the proportion of total variability in the dependent variable that is interpreted by the independent variables. The r2 score ranges from a minimum of 0 to a maximum of 1. Higher r^2 scores indicate good fit of the prediction model.

$$r^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \bar{y})^2} \quad [2]$$

where, $y_i, \hat{y}_i \; and \; \bar{y}$ represents the observed values, predicted values and mean of all observed values, respectively.

The second statistical score used for assessing the predicted values is the Pearson's *R* coefficient. The Pearson *R* depicts the linear relationship between two sets of variables represented with scores between -1 and 1; the former representing perfect negative linear relationship and the latter representing a perfect positive linear relationship. Ideally, the predicted and actual values in a prediction model should exhibit positive linear relationship. The Pearson *R* is computed using Eqn. 3.

$$R = \frac{\sum(x_i - \hat{x})(y_i - \hat{y})}{\sqrt{\sum(x_i - \hat{x})^2 \sum(y_i - \hat{y})^2}} \quad [3]$$

2.5 Hyperparameter Tuning

Hyperparameter tuning involves fine-tuning certain parameters of a machine learning algorithm that cannot be learned during training. It is a crucial step in customizing a model for a specific dataset, and can significantly improve model performance, accuracy, and generalization, resulting in a more robust prediction process. Hyperparameter tuning also enables easy interpretation of the model, and ensures optimal use of computational resources (Donnelly et al., 2023). In this study, hyperparameter tuning was performed on all ML algorithms using the open-source Python programming language-based library, scikit-learn (Pedregosa et al., 2011).

Table 1 summarizes the different hyperparameters, their typical values and the best possible value for the overtopping dataset. For the SVMR algorithm, the term C signifies a regularization parameter and the kernel or the main function of the algorithm could take the form of a linear, polynomial or rbf (radial basis function).

The hyperparameters for the RF and GBDT are similar as they are characteristically DTs. Among the important hyperparameters in a RF and GBDT algorithm can be mentioned as 'n_estimators'- the number of DTs in RF/GBDT while 'max_depth' and 'min_samples_split' perform the role of reducing overfitting. In a GBDT algorithm, the 'learning_rate' is a key hyperparameter as it determines the convergence rate of the loss function.

Table 1. Summary of Hyperparameter Values

Algorithm	Hyperparameters and Typical Values	Best Values
SVMR	C': 5, 15; 'kernel': linear, rbf, poly; 'gamma': auto, scale	C': 15; 'kernel': rbf, 'gamma': auto
RF	n_estimators': 800, 900, 2000 'min_samples_split': 2, 5, 10 'min_samples_leaf': 1, 2, 4 'max_features': auto, sqrt 'max_depth': 10 to 110 'bootstrap': True , False	n_estimators': 800, 'min_samples_split': 5, 'min_samples_leaf': 1, 'max_features': 'auto', 'max_depth': 110, 'bootstrap': True
GBDT	max_depth': 3, 5, 6, 10, 15, 20; 'learning_rate': 0.01, 0.1, 0.2, 0.3; 'subsample': 0.5, 1.0, 0.1; 'colsample_bytree': 0.4, 1.0, 0.1; 'colsample_bylevel': 0.4, 1.0, 0.1; 'n_estimators': 100, 500, 1000, 1500	max_depth': 10; 'learning rate': 0.2; 'subsample': 0.5; 'colsample_bytree': 0.7; 'colsample_bylevel': 0.4; 'n_estimators': 1000
ANN	activation': 'relu', 'sigmoid'; 'batch_size': 32, 64, 128, 256 epochs': 10, 15, 20, 'alpha': 0.0001, 0.0005	activation': 'relu', 'batch_size': 64, 'epochs': 10, 'alpha': 0.0005

Unlike the other algorithms used in this study, the scope of hypertuning in ANN is limited to some extent (Huang et al., 2012). Among the hyperparameters for ANN listed in Table 1, 'alpha' is the learning rate which signifies the rate at which weights are applied to the input data. A higher learning rate can result in a faster rate of attaining convergence in the loss function while at the same time there is a risk of overshooting the appropriate model loss. On the other hand, if the learning rate is relatively low, this may cause a slower convergence, however, makes the model more stable and accurate eventually. The 'activation' function and the 'epochs' were the other tunable hyperparameters of ANN.

The typical values for hyperparameters were adopted as per the suggestion of scikit learn library (Pedregosa et al., 2011) and other open-source data science platforms.

2.6 Data Optimization

The overtopping dataset used in this study consisted of overtopping parameters in several types of units and scales. This many different units and scales can potentially reduce the accuracy of predictions of the ML

algorithms (Pedregosa et al., 2011). Two main approaches were adopted in this study to ensure uniformity in the scale of data and to reduce redundancy of data, namely scalar transformation and feature selection respectively. The functioning of ANN is susceptible to missing data and hence the missing data were interpolated using the k-Nearest Neighbors (kNN) imputation method. In kNN, the missing values are estimated based on the median or average of the data that is nearest to the missing value. The methodology of kNN is further described in the work of Hastie et al., (2009).

The scalar transformation and feature selection were performed using scikit-learn's libraries (Pedregosa et al., 2011). In scalar transformation, the numerical data is scaled to unit variance. The data is transformed to normal distribution with zero mean and unit variance. The role of feature selection in data science is to remove redundancy and irrelevant features by computing statistical significance of independent variables with respect to the dependent variable **(**Liu et al., 2012). Feature selection was applied to overtopping dataset in the work of den Bieman et al, (2021), as a permutation analysis of the overtopping parameters.

A sequential Forward Selection (SFS) was implemented to filter the relevant overtopping parameters for this study. The SFS method relies on a greedy search algorithm that is initiated with an empty feature set and continues to add features or independent variables in an iterative manner. The SFS algorithm then trains the features and performs prediction of the target variable. The performance of prediction is estimated in terms of the r^2 score and the iterative process of adding features continues until a stopping criterion is satisfied. The algorithm finally produces the number and list of features that had significant statistical significance in the prediction task.

As depicted in Figure 1, the dataset consisted of 30 features and the results of the Sequential Feature Selection (SFS) showed a decline in prediction performance after the inclusion of the 15th feature. Hence, it can be concluded that a dataset containing 15 features is optimal for overtopping prediction in this study.

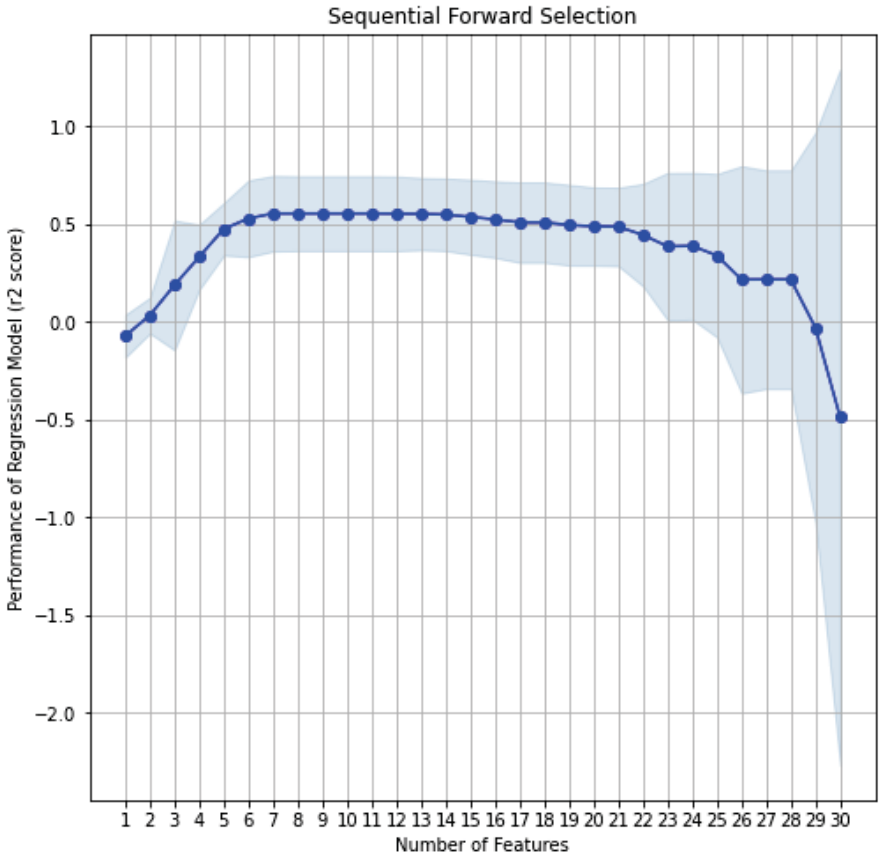

Figure 1. Regression Model Performance for Sequential Forward Feature Selection

2.7 Data Preparation

The EurOtop (2018) is an extensive collection of physical overtopping tests for a range of geometrical configurations of coastal structures. The dataset comprises of approximately 17500 sets of overtopping data along with parametric values of overtopping parameters. Therefore, it was necessary to refine the dataset so that it contained data for simple sloped impermeable sloping walls only. Hence, the refining parameters were set as follows: (the terms are elaborated in the glossary table)

- $H_{m0,t} \geq 0.5$; to include entries for small-scale records

- $\gamma_f = 1$; to include impermeable sloped walls
- $1.33 \leq \cot \alpha \leq 2$; to include walls with mild slope
- $B = 0$; to include sloping walls with no Berm
- $\cot \alpha_u = \cot \alpha_d$; to include simple sloped walls
- $1 \leq RF$ and $CF \leq 3$; to exclude data with lowest reliability and highest complexity in geometry

The refined data consisted of 1079 entries of overtopping data which was approximately 6% of the total number of entries of the EurOtop (2018) dataset. It is to be noted that the minimum number of data points required for overtopping estimation using machine learning depends on several factors, such as the complexity of the problem, the quality of the data, and the choice of the algorithm. However, in existing literature (e.g., Hosseinzadeh et al., 2021; Etemad-Shahidi et al., 2016), several hundred data points were used to train a reliable machine learning models for overtopping estimation. Elbisy (2023) used approximately 2,400 dataset for ANN and SVM models used in the study.

Following the feature selection strategy discussed in the previous section, 15 overtopping parameters/independent variables were selected for final analysis which can be listed as $H_{m0,d}$, $T_{p,d}$, h, $H_{m0,t}$, $T_{m,toe}$, h_t, $\cot \alpha_d$, $\cot \alpha_u$, Rc, B, h_b, $\tan \alpha$, Gc, RF and CF (terms explained in glossary). The dependent or the predictor variable was '*q*' which is the overtopping discharge per unit length of the structure, measured in m^3/s.

3. RESULTS AND DISCUSSION

A common methodology of data preprocessing (scalar transformation and missing data imputation), feature selection and hyperparameter tuning was applied to an overtopping dataset containing 1079 observations of measured overtopping discharge at simple sloped breakwaters. The filtered and then treated data were applied to 4 ML algorithms to predict overtopping discharge following a train-test split of 70%-30%. The algorithms were executed separately in a computer having a Central Processing Unit of 8 cores, a Random Access Memory of 16 Gigabytes and 6 Gigabytes of dedicated graphics memory. The predicted overtopping discharge rates were converted to dimensionless quantities, 'dimensionless 'q' (measured) ($=\frac{q\_measured}{\sqrt{9.81*(H_{m0,t})^3}}$ and 'dimensionless 'q' (predicted) ($=\frac{q\_predicted}{\sqrt{9.81*(H_{m0,t})^3}}$) to compare the accuracy of prediction from the four ML algorithms. The performance of the ML algorithms in terms of predicted vs measured quantities is illustrated in Figure 3. It can be summarized from Figure 3, that the ML algorithms were able to predict overtopping discharge with reasonable accuracy. Particularly, prediction quantities from small overtopping rates were more accurate which is evident from the position of the data points within the 95% Confidence Interval (CI) zone. For all the algorithms, larger overtopping quantities showed dispersion from the regression line.

The scatter depicted in Figure 3 can further be explained using the statistical scores shown in Table 2. Although the GBDT algorithm outperformed the other algorithms in terms of r^2 score, the SVMR algorithm showed similar performance but with a significantly longer time. The RF algorithm appeared the fastest algorithm in terms of computational time, however, with reduced prediction accuracy in comparison to the GBDT and SVMR. The ANN underperformed with respect to the other algorithms in the context of both accuracy and computational time. It is also evident from this study, that DT based algorithms such as GBDT and RF are computationally efficient than kernel-based algorithms such as SVMR and ANN. The computational efficiency of RF and GBDT is coupled with prediction accuracy at the same time.

Table 2. Performance Attributes of the Algorithms

Algorithm	Coefficient of Determination r^2	Pearson's *R* Coefficient	Computational Time (s)
GBDT	0.87	0.93	32
SVMR	0.86	0.93	141
RF	0.79	0.89	22
ANN	0.69	0.87	468

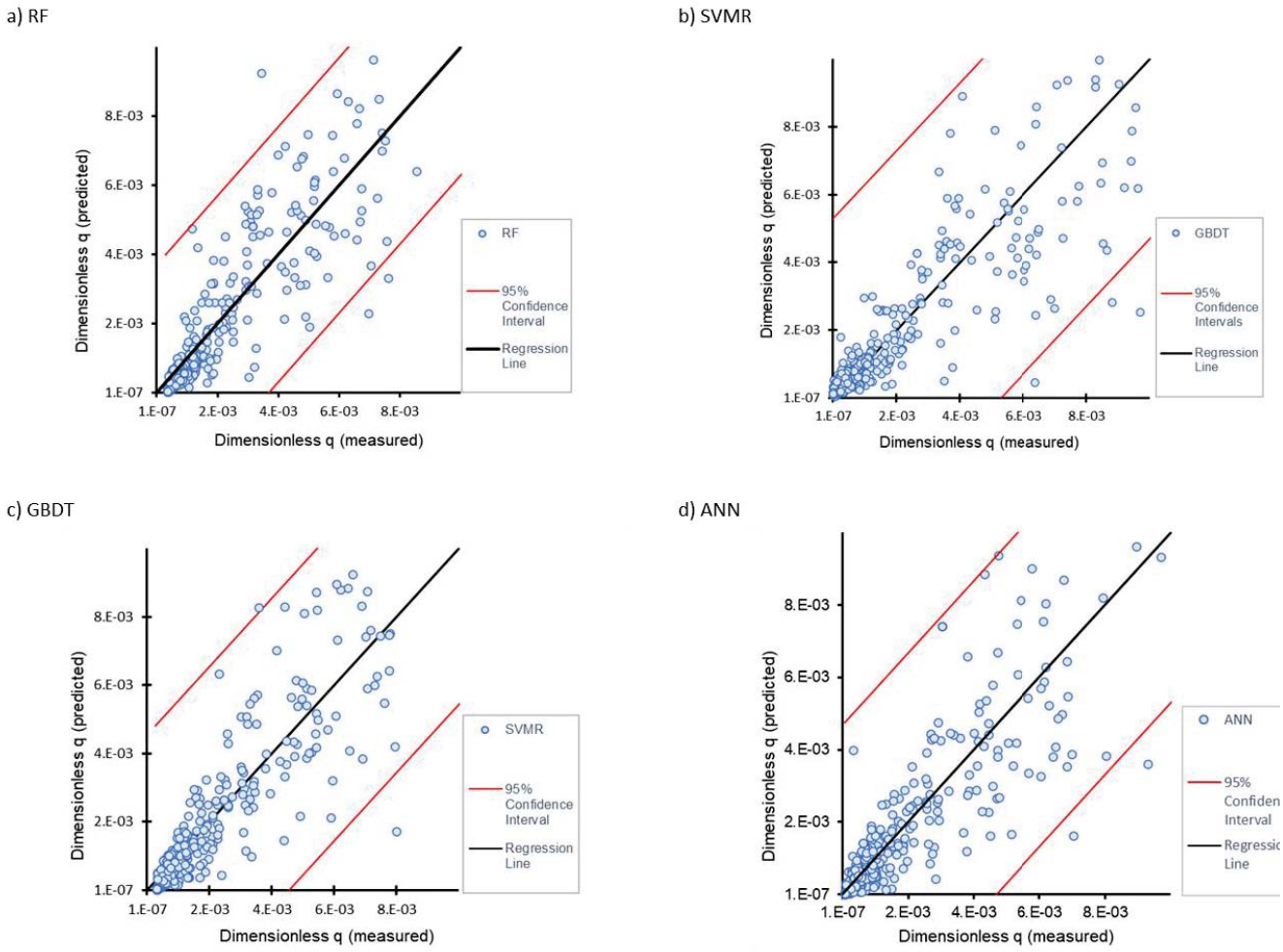

Figure 2. Comparison of predicted dimensionless '*q*' (=$\frac{q\_predicted}{\sqrt{9.81*(H_{m0,t})^3}}$) and actual dimensionless '*q*' (=$\frac{q\_measured}{\sqrt{9.81*(H_{m0,t})^3}}$)

4. CONCLUSION

This study investigates the performance of four algorithms, namely GBDT, RF, SVMR and ANN, in predicting overtopping discharge from the freely available EurOtop (2018) dataset for simple sloped breakwaters. The dataset was imputed for missing values and treated for redundancy by kNN imputation and a feature selection method respectively, before exposing it to the ML algorithms. These steps ensured that the applied dataset was statistically significant in the context of the independent variables and the dependent variable. The train-test ratio was set to 70%-30% and hyperparameter tuning was performed to curtail the individual algorithms according to the given dataset. In addition to hyperparameter tuning, Cross Validation was also ensured that implemented validation of the training data before the algorithms were allowed to predict on the test set. The predicted and actual overtopping discharge quantities were converted to dimensionless quantities to ensure fair comparison.

The accuracy of predictions from the four algorithms was reported using statistical scores, the r^2 and Pearson *R* value. The GBDT algorithm marginally outperformed the SVMR algorithm in terms of the r^2 score, 0.87 and 0.86, respectively, and emerged as the most accurate algorithm in the prediction task. The RF algorithm was, however, computationally more efficient than the GBDT algorithm with the former completing the prediction task 10 seconds earlier than the latter. Overall, the DT based algorithms (RF and GBDT) performed better than the kernel-based algorithms with respect to both prediction accuracy and statistical scores. The performance of ANN did not match with those of the other algorithms, with lower statistical scores and significantly higher completion time.

The common methodological framework adopted in this study for all the algorithms ensured a fair comparison of the performance of the algorithms. However, the main limitation of this study can be cited as the presence of missing values in the dataset.

5. ACKNOWLEDGMENT

With thanks, the authors receive the EurOtop (2018) database from Professor Jentsje W. Van der Meer, Professor Barbara Zanuttigh and Dr Sara Mizar Formentin.

6. GLOSSARY

$H_{m0\,d}$	Significant wave height at deep water (m)
$T_{p\,d}$	Wave period determined at deep water (s)
$H_{m0,\,t}$	Significant wave height at toe of structure (m)
h	Water depth at toe of structure (m)
$T_{m\,toe}$	Avg. wave period determined at toe (s)
h_t	Water depth at toe of structure (m)
$\cot\alpha_d$	cot of angle between structure slope downward berm and horizontal (-)
$\cot\alpha_u$	cot of angle between structure slope upward berm and horizontal (-)
R_c	crest freeboard of structure (m)
B	Berm width, measured horizontally (m)
h_b	water depth on berm (negative means berm is above SWL) (m)
tan α	tan of angle of structure slope (-)
G_c	Width of promenade (m)
RF	Reliability-Factor of test
CF	Complexity-Factor of structure section

Source: EurOtop (2018)

7. REFERENCES

Babaee, M., Maroufpoor, S., Jalali, M., Zarei, M., & Elbeltagi, A. (2021). Artificial intelligence approach to estimating rice yield*. *Irrigation and Drainage*, *70*(4), 732–742. https://doi.org/10.1002/ird.2566.

Chen, S., Chen, H., Peng, C., Wang, Y., & Hu, Y. (2022). Analysis of Factors Influencing Wave Overtopping Discharge from Breakwater Based on an MIV-BP Estimation Model. *Water*, *14*(19), 2967. https://doi.org/10.3390/w14192967

Chen, W., Warmink, J.J., van Gent, M.R.A., & Hulscher, S.J.M.H. (2021). Numerical modelling of wave overtopping at dikes using OpenFOAM®. *Coastal Engineering*, *166*, 103890. https://doi.org/10.1016/j.coastaleng.2021.103890

Dang, B.L., Nguyen-Xuan, H., & Abdel Wahab, M. (2023). An effective approach for VARANS-VOF modelling interactions of wave and perforated breakwater using gradient boosting decision tree algorithm. *Ocean Engineering*, *268*, 113398. https://doi.org/10.1016/j.oceaneng.2022.113398

den Bieman, J.P., van Gent, M.R.A., & van den Boogaard, H.F.P. (2021). Wave overtopping predictions using an advanced machine learning technique. *Coastal Engineering*, *166*. https://doi.org/10.1016/j.coastaleng.2020.103830.

den Bieman, J.P., Wilms, J.M., Van den Boogaard, H.F.P., & Van Gent, M.R.A. (2020). Prediction of Mean Wave Overtopping Discharge Using Gradient Boosting Decision Trees. Water, 12(6), 1703. https://doi.org/10.3390/w12061703.

Dong, S., Salauddin, M., Abolfathi, S., Tan, Z. H., & Pearson, J.M. (2018). The Influence of Geometrical Shape Changes on Wave Overtopping: a Laboratory and SPH Numerical Study. Coasts, Marine Structures and Breakwaters. DOI: https://doi.org/10.1680/cmsb.63174.1217

Dong, S., Abolfathi, S., Salauddin, M., Tan, Z. H., & Pearson, J. M. (2020). Enhancing climate resilience of vertical seawall with retrofitting - A physical modelling study. *Applied Ocean Research*, *103*. https://doi.org/10.1016/j.apor.2020.102331

Dong, S., Abolfathi, S., Salauddin, M., & Pearson, J. M. (2021a). Spatial distribution of wave-by-wave overtopping behind vertical seawall with recurve retrofitting. *Ocean Engineering*, *238*, 109674. https://doi.org/10.1016/j.oceaneng.2021.109674

Dong, S., Salauddin, M., Abolfathi, S., & Pearson, J.M. (2021b). Wave Impact Loads on Vertical Seawalls: Effects of the Geometrical Properties of Recurve Retrofitting. Water 13, no. 20: 2849. https://doi.org/10.3390/w13202849

Donnelly, J., Daneshkhah, A., and Abolfathi, S. (2023). Physics-Informed Neural Networks for Statistical Emulation of Hydrodynamical Numerical Models, EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-5445, https://doi.org/10.5194/egusphere-egu23-5445.

Elbisy, M.S. (2023). Machine learning techniques for estimating wave-overtopping discharges at coastal structures. *Ocean Engineering*, *273*, 113972. https://doi.org/10.1016/j.oceaneng.2023.113972

Etemad-Shahidi, A., Shaeri, S., & Jafari, E. (2016). Prediction of wave overtopping at vertical structures. *Coastal Engineering*, *109*, 42–52. https://doi.org/10.1016/j.coastaleng.2015.12.001

EurOtop. (2018). Manual on wave overtopping of sea defences and related structures. An overtopping manual largely based on European research, but for worldwide application. Van der Meer, J.W., Allsop, N.W.H., Bruce, T., De Rouck, J., Kortenhaus, A., Pullen, T., Schüttrumpf, H., Troch, P. and Zanuttigh, B., www.overtopping-manual.com.

Formentin S.M., Zanuttigh B. & Van der Meer J.W. (2017). A Neural Network tool for predicting wave reflection, overtopping and transmission, Coastal Engineering Journal, 59, No. 2 (2017), 1750006, 31 pp.

Goda, Y. (2009). Derivation of unified wave overtopping formulas for seawalls with smooth, impermeable surfaces based on selected CLASH datasets. *Coastal Engineering*, *56*(4). https://doi.org/10.1016/j.coastaleng.2008.09.007

Habib, M.A., O'Sullivan, J. & Salauddin, M. (2022a). Prediction of wave overtopping characteristics at coastal flood defences using machine learning algorithms: A systematic review. IOP Conf. Ser.: Earth Environ. Sci.1072 012003. DOI: 10.1088/1755-1315/1072/1/012003

Habib, M.A., O'Sullivan, J., and Salauddin, M. (2022b). Comparison of machine learning algorithms in predicting wave overtopping discharges at vertical breakwaters, EGU General Assembly 2022, Vienna, Austria, 23–27 May 2022, EGU22-329, https://doi.org/10.5194/egusphere-egu22-329, 2022.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction *(Vol. 2). Springer*

Hosseinzadeh, S., Etemad-Shahidi, A., & Koosheh, A. (2021). Prediction of mean wave overtopping at simple sloped breakwaters using kernel-based methods. *Journal of Hydroinformatics*, *23*(5), 1030–1049. https://doi.org/10.2166/hydro.2021.046.

Huang, G.B.; Zhou, H.; Ding, X.; Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man Cyber Part B 42, 513–529.

Irías Mata, M., & van Gent, M.R.A. (2023). Numerical modelling of wave overtopping discharges at rubble mound breakwaters using OpenFOAM®. *Coastal Engineering*, *181*, 104274. https://doi.org/10.1016/j.coastaleng.2022.104274

Kotu, V., & Deshpande, B. (2015). Classification. In *Predictive Analytics and Data Mining* (pp. 63–163). Elsevier. https://doi.org/10.1016/B978-0-12-801460-8.00004-5.

Liu H. & Motoda H. (2012). Feature Selection for Knowledge Discovery and Data Mining, *Springer Science & Business Media.*

Liu, N., Salauddin, M., Pearson, J.M., & Abolfathi, S. (2023). The impact of eco-retrofitting on the spatial distribution of wave overtopping – a physical modelling study. IOP Conference Series: Earth and Environmental Science, Volume 1072, 9th International Conference on Coastal and Ocean Engineering (ICCOE 2022) 08/04/2022 - 10/04/2022 Online. DOI: 10.1088/1755-1315/1072/1/012005

Masson-Delmotte, V.P., Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou IPCC, (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.. Cambridge University Press. In Press.

McCulloch, W.S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133. https://doi.org/10.1007/BF02478259

Noori, R., Ghiasi, B., Salehi, S., Esmaeili Bidhendi, M., Raeisi, A., Partani, S., Meysami, R., Mahdian, M., Hosseinzadeh, M., & Abolfathi, S. (2022). An Efficient Data Driven-Based Model for Prediction of the Total Sediment Load in Rivers. *Hydrology*, *9*(2), 36. https://doi.org/10.3390/hydrology9020036

O'Sullivan, J.J., Salauddin, M., Abolfathi, S., and Pearson, J.M. (2020). Effectiveness of eco-retrofits in reducing wave overtopping on seawalls. Coastal Engineering Proceedings (36v). DOI: https://doi.org/10.9753/icce.v36v.structures.13.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). *Scikit-learn: Machine Learning in Python*. https://doi.org/https://doi.org/10.48550/arXiv.1201.0490.

Ravindar, R., Sriram, V., & Salauddin, M. (2022). Numerical modelling of breaking wave impact loads on a vertical seawall retrofitted with different geometrical configurations of recurve parapets. Journal of Water and Climate Change, 13(10), 3644-3674. DOI: https://doi.org/10.2166/wcc.2022.211

Salauddin, M. and Pearson, J.M. (2018). A laboratory study on wave overtopping at vertical seawalls with a shingle foreshore. *Coastal Engineering Proceedings*, 1(36), waves.56. DOI: https://doi.org/10.9753/icce.v36.waves.56

Salauddin, M., & Pearson, J.M. (2020). Laboratory investigation of overtopping at a sloping structure with permeable shingle foreshore. *Ocean Engineering*, *197*. https://doi.org/10.1016/j.oceaneng.2019.106866

Salauddin, M., O'Sullivan, J. J., Abolfathi, S., & Pearson, J.M. (2021). Eco-Engineering of Seawalls—An Opportunity for Enhanced Climate Resilience From Increased Topographic Complexity. *Frontiers in Marine Science*, *8*. https://doi.org/10.3389/fmars.2021.674630

Salauddin, M., O'Sullivan, J.J., Abolfathi, S., Dong, S., & Pearson, J.M. (2020). Distribution of individual wave overtopping volumes on a sloping structure with a permeable foreshore. Coastal Engineering Proceedings, 36, 54, DOI: https://doi.org/10.9753/icce.v36v.papers.54

Salauddin, M., O'Sullivan, J.J., Abolfathi, S. et al., (2022). New insights in the probability distributions of wave-by-wave overtopping volumes at vertical breakwaters. Scientific Reports 12, 16228 (2022). https://doi.org/10.1038/s41598-022-20464-5

Sutton, C. D. (2005). *Classification and Regression Trees, Bagging, and Boosting* (pp. 303–329). https://doi.org/10.1016/S0169-7161(04)24011-1.

Temmerman, S., Meire, P., Bouma, T.J., Herman, P.M.J., Ysebaert, T., & De Vriend, H.J. (2013). Ecosystem-based coastal defence in the face of global change. *Nature*, *504*(7478).

van Gent, M.R.A., van den Boogaard, H.F.P., Pozueta, B., & Medina, J.R. (2007). Neural network modelling of wave overtopping at coastal structures. *Coastal Engineering*, *54*(8). https://doi.org/10.1016/j.coastaleng.2006.12.001

Verhaeghe, H., De Rouck, J., & van der Meer, J. (2008). Combined classifier–quantifier model: A 2-phases neural model for prediction of wave overtopping at coastal structures. *Coastal Engineering*, *55*(5). https://doi.org/10.1016/j.coastaleng.2007.12.002

Victor, L., van der Meer, J.W., & Troch, P. (2012). Probability distribution of individual wave overtopping volumes for smooth impermeable steep slopes with low crest freeboards. *Coastal Engineering*, *64*. https://doi.org/10.1016/j.coastaleng.2012.01.003.

Yeganeh-Bakhtiary, A., EyvazOghli, H., Shabakhty, N., Kamranzad, B. & Abolfathi, S. (2022). Machine Learning as a Downscaling Approach for Prediction of Wind Characteristics under Future Climate Change Scenarios. *Complexity*, vol. 2022, Article ID 8451812, 13 pages. https://doi.org/10.1155/2022/8451812

Yeganeh-Bakhtiary, A., Houshangi, H., & Abolfathi, S. (2020). Lagrangian two-phase flow modeling of scour in front of vertical breakwater. *Coastal Engineering Journal*, *62*(2), 252–266. https://doi.org/10.1080/21664250.2020.1747140

Zanuttigh B., Formentin S.M., and Van der Meer J.W. (2016). Prediction of extreme and tolerable wave overtopping discharges through an advanced neural network, Ocean Engineering, 127, 7-22

Zhao, T., Liu, S., Xu, J., He, H., Wang, D., Horton, R., & Liu, G. (2022). Comparative analysis of seven machine learning algorithms and five empirical models to estimate soil thermal conductivity. *Agricultural and Forest Meteorology*, *323*, 109080. https://doi.org/10.1016/j.agrformet.2022.109080.