International Journal of GEOMATE, Oct., 2022, Vol.23, Issue 98, pp.205-211
ISSN: 2186-2982 (P), 2186-2990 (O), Japan, DOI: https://doi.org/10.21660/2022.98.3580
Geotechnique, Construction Materials and Environment

SOFT COMPUTING TECHNIQUES FOR PREDICTING WAVE OVERTOPPING DISCHARGES AT VERTICAL COASTAL STRUCTURES

Moussa S. Elbisy*[1], Faisal A. Osra[1], and Yasser S. Alyafei[1]

[1]Civil Engineering Dept., College of Engineering and Islamic Architecture, Umm Al_Qura University, Makkah, Saudi Arabia.

*Corresponding Author, Received: 02 Aug. 2022, Revised: 09 Sept. 2022, Accepted: 20 Oct. 2022

ABSTRACT: Coastal zone management, including coastal protection, is one of the main concerns of humanity. Coastal zone management includes protection against extreme waves and floods and affects social and economic life. An important consideration in the construction of coastal structures is the estimation of wave overtopping. In fact, wave overtopping has been responsible for several coastal structure failures. This study sought to evaluate the accuracy of soft computing techniques using Multilayer Perceptron Neural, MLPNN, Network and Random Forest Decision Tree, RFDT, for wave overtopping discharge prediction of vertical coastal structures. The new EurOtop database, which was recently created by gathering a lot of data from various hydraulic testing, was also used to create soft computing techniques for the precise prediction of wave overtopping discharge. Each wave overtopping prediction method's effectiveness has been quantitatively assessed using charts and statistically assessed using accuracy measures. The evaluation's findings demonstrated that the RFDT model provided more accurate predictions than the MLPNN technique.

Keywords: Coastal Management, Wave overtopping, Prediction, Soft computing, Coastal structures Safety.

1. INTRODUCTION

In coastal zones, assessing the risk of wave overtopping of marine structures is critical for assessing the danger of both the structures collapsing and any flooding of the protected areas. For coastal structure development and safety considerations, a reliable estimate of wave overtopping rate is essential. Extreme overtopping events can cause water to flow over the crest at rapid speeds, endangering infrastructure and people. These incidents are extremely dangerous because they have resulted in the washing into the sea of people, vehicles, and even trains.

Approximately 10% of the population of England lives within areas potentially at risk from flooding or coastal erosion, whilst approximately 12% of the agricultural land in England is also located in these areas. The value of assets at risk of floods and coastal erosion is estimated to be£214 billion, with average yearly damages of £2.8 billion (without defenses) [1]. In England, without any flood and coastal defenses, the average annual economic damage from flooding and coastal erosion would be over £2.6 billion per year. In July 2020, the amount invested in the new 6-year investment program in the UK had doubled to £5.2 billion for 2021 to 2027 [2].

As a consequence, a precise assessment of the wave overtopping rate in terms of several hydraulic and structural design elements is required. Several hydraulic experiments have been conducted, resulting in the development of design diagrams or empirical formulations. A variety of approaches for predicting the quantity of allowed wave overtopping of certain structures under specified wave conditions and structural attributes are available. These approaches include empirical equations, soft computing methods, and numerical models. Empirical equations are the simplest and most reliable method for predicting wave overtopping. These formulas give a reasonably quick and simple way to acquire a preliminary estimate of the predicted mean wave overtopping discharge. Numerical techniques of prediction have the benefit of being able to be customized for any type of structure and provide far more precise information on the overtopping flow. However, it requires a large amount of computing time and money [3]. Soft computing methods allow the impacts of a large number of governing parameters to be included, resulting in adaptable tools capable of representing complex structural geometries and changeable wave conditions. Several effective examples of soft computing methods have been created and employed in coastal engineering in a variety of applications during the last several years [4]. Soft computing methods include artificial neural networks (ANNs), tree-based methods, and support vector machines.

Wedge *et al*. (2005) used two types of neural networks, which are multi-layer perceptron networks (MLPNN) and radial basis function networks (RBFNN). They observed that the neural network approach has a significantly lower processing cost and provides generic prediction across a variety of structures and sea states. Radial basis function networks also perform significantly better than multi-layer perceptron networks and the curve-fitting (parametric regression) regime, and their accuracy is comparable to that of custom numerical simulations [5]. Van Gent *et al*. (2007) presented an ANN model for various types of coastal structures as a part of the CLASH project. The model for the prediction of wave overtopping discharges was MLPNN with three layers, and the number of hidden neurons that were used was 20 [6]. Verhaeghe *et al*. (2008) developed a two-phase neural model to improve prediction accuracy. The quantifier is refined using the bootstrap technique. They discovered the optimal neural quantifier architecture (13 input parameters, 25 neurons in the hidden layer, and 1 output parameter). The results demonstrate that the combined classifier–quantifier result is demonstrably superior to the quantifier result alone [7]. Bieman et al. (2020) applied gradient-boosting decision trees to the prediction of mean wave overtopping discharges.

The model was trained using the CLASH wave overtopping database. By decreasing the error on the prediction of the CLASH database by a factor of 2.8, the model has been proven to outperform the current neural network model provided by Van Gent et al. (2007) [8]. The purpose of this study is to assess the accuracy of soft computing techniques utilizing the ANN approach through the use of a Multilayer Perceptron Neural Network. Also, utilizing the tree-based methods through the use of the Random Forest model for wave overtopping discharge prediction of vertical coastal structures. This study is expected to provide valuable up-to-date information for estimating wave overtopping risk, forecasting wave overtopping for warning and emergency evacuation of people in the event of extreme waves, risk minimization, and economic assessment of coastal protection projects.

2. MATERIALS AND METHODS

2.1 Data

The data set used in this research is the new EurOtop database which is based on the European CLASH project. The new enlarged database contains over 17,000 tests, approximately 13,500 of which are for wave overtopping only [9]. The data used is the data that is convened for vertical walls only and assigned for the training of the soft computing methods, which were 2279 tests. Fig. 1 illustrates the physical meaning of the parameters impacting vertical coastal structures. The 13 parameters were analyzed through statistical measurements to summarize the characteristics of a data set, as shown in Table 1.

2.2 Methods

The methods used in this study are Multilayer Perceptron Neural Network (MLPNN) and Random Forest Decision Tree (RFDT). The MLPNN is considered one of the most popular methods of the ANNs and has three layers: an input layer, an output layer, and a hidden layer. Three basic elements were recognized when modeling artificial neurons. To begin, there is a collection of synapses, or connecting links, each of which has its own weight or strength. The synaptic weight *wkj* is multiplied by a signal *xj* at the input of synapse *j* coupled to neuron *k*. Second, there's an adder for summing the input signals, weighted by the respective synaptic strengths of the neuron. Finally, there is an activation function for restricting the magnitude of a neuron's output [10]. A simple model of a neuronal process is shown in Fig. 2. Random forests are soft computing methods that use a variety of techniques. In this method, there are successive series of tree classifiers, each tree casts a unit vote for the most popular class, and the results are merged to produce the final sort of result [11].

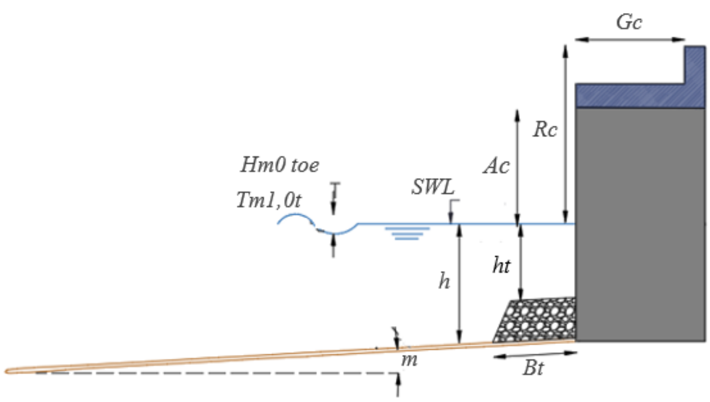

Fig. 1 Cross section with application parameters for soft computing methods.

Table 1 Summary of statistical for the dimensional basic parameters of the used dataset.

Parameter	N	Units	Minimum	Maximum	Range	Mean	SD
m	2279	[1: *m*]	10.000	1050.000	1040.000	641.704	459.962
β	2279	[°]	0.000	0.000	0.000	0.000	0.000
h	2279	[*m*]	0.016	1.280	1.264	0.507	0.291
Hm0 toe	2279	[*m*]	0.024	0.603	0.579	0.121	0.081
Tm1,0t	2279	[*s*]	0.290	7.500	7.210	1.638	0.662
ht	2279	[*m*]	0.000	1.280	1.280	0.409	0.280
Bt	2279	[*m*]	0.000	1.132	1.132	0.129	0.193
γf	2279	[-]	0.495	1.000	0.505	0.990	0.041
D	2279	[*m*]	0.000	0.038	0.038	0.002	0.007
Rc	2279	[*m*]	0.000	1.460	1.460	0.197	0.196
Ac	2279	[*m*]	-0.217	1.460	1.677	0.178	0.200
Gc	2279	[*m*]	0.000	1.376	1.376	0.015	0.068
q	2279	[m^3/s per *m*]	0.00E+00	1.46E-02	1.46E-02	6.07E-04	1.52E-03

Fig. 3 depicts a schematic diagram of the random forest algorithm's information flow. This approach has the ability to handle big data sets with increased dimensionality. It is one of the dimensionality reduction strategies since it can handle hundreds of input variables and discover the most significant variables.

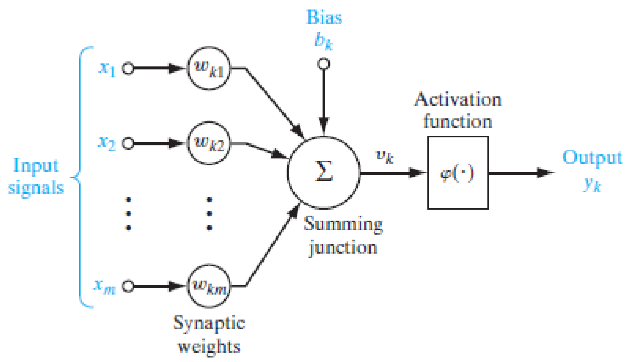

Fig. 2 The model of an artificial neuron.

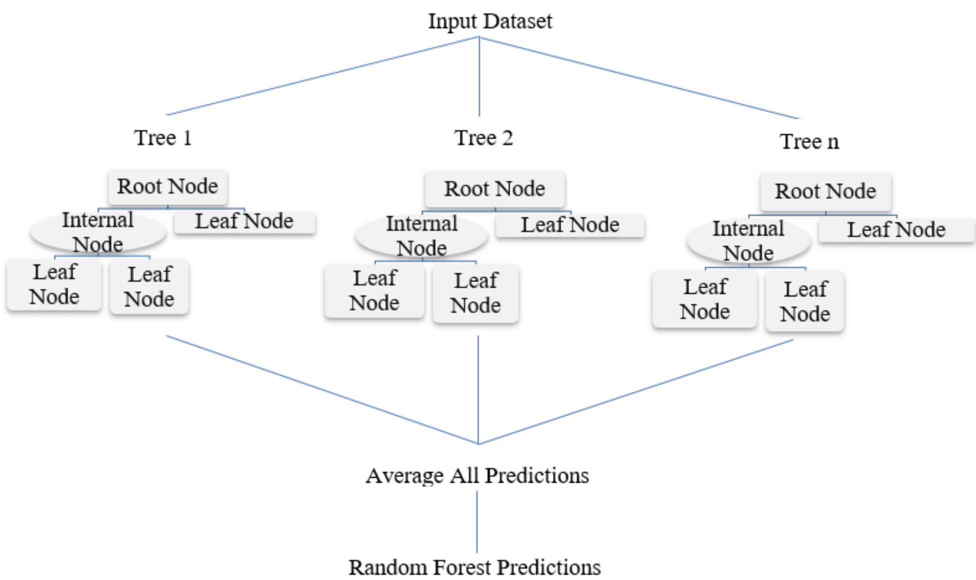

Fig. 3 Conceptual diagram of the random forest model.

3. RESULTS AND DISCUSSION

The new EurOtop database recommends not using basic parameters as input to the models since some of the data is from small-scale models while other data is from full-scale prototypes. Consequently, the basic data should be dimensionless to prevent the large variation in raw parameter values. The reason for making dimensionless parameters is to make the soft computing models more accurate and reliable. The parameters characterizing structural heights (vertical measurements of toe and crest) are all dimensionless with the *Hm0 toe*. While the parameters characterizing structure widths (horizontal measurements of toe and crest) are all dimensionless with the wavelength.

The wavelength (*Lm1,0t*) can be calculated by using the following equation:

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

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

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

The final input and output dimensionless parameters used in developing soft computing models are (*m, β, h/Lm1,0t, Hm0 toe/Lm1,0t, ht/Hm0 toe, Bt/Lm1,0t, γf, D/Hm0 toe, Rc/Hm0 toe, Ac/Hm0 toe, Gc/Lm1,0t, Sq*).

The performance of MLPNN and RFDT models were assessed based on standard deviation (*SD*), mean square error (*MSE*), root mean square error (*RMSE*), mean absolute error (*MAE*), mean absolute percentage error (*MAPE*), correlation coefficient (*R*), discrepancy ratio (*DR*), coefficient of performance (*COP*), average absolute error (*AAE*), scatter index (*SI*), root mean square percentage error (*RMSPE*) and Willmott index (*WI*). The equations for these statistical indicators are as follows:

$$SD = \sqrt{\frac{\sum_{i=1}^{n}\left(Sq_{pred}-\overline{Sq_{pred}}\right)^2}{n}} \tag{3}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left|Sq_{meas} - Sq_{pred}\right|^2 \tag{4}$$

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Sq_{meas} - Sq_{pred}\right| \tag{6}$$

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

$$DR = log\frac{Sq_{pred}}{Sq_{meas}} \tag{8}$$

$$COP = \frac{\sum_{i=1}^{n}Sq_{pred}}{\sum_{i=1}^{n}Sq_{meas}} \tag{9}$$

$$AAE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left|Sq_{meas}-Sq_{pred}\right|}{Sq_{meas}}\right) \tag{10}$$

$$SI = \frac{RMSE}{\overline{Sq_{meas}}} \tag{11}$$

$$RMSPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left|Sq_{meas}-Sq_{pred}\right|}{Sq_{meas}}\right)^2 \tag{12}$$

$$WI = 1-\frac{\sum_{i=1}^{n}\left|Sq_{meas}-Sq_{pred}\right|^2}{\sum_{i=1}^{n}\left(\left|Sq_{meas}-\overline{Sq_{meas}}\right|+\left|Sq_{pred}-\overline{Sq_{meas}}\right|\right)^2} \tag{13}$$

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

The MLPNN configuration parameters used to build the training are listed in Table 2. Table 3 describes the architecture of MLPNN. Table 4 also displays the MLPNN model's training statistics.

The MLPNN outcomes seemed inappropriate to predict the overtopping rates as illustrated in Fig. 4; the points were clustered and less spread, as proven by statistical indicators. The *MSE*, *RMSE*, *MAE*, and *MAPE* were 8.50E-05, 0.009, 0.006, and 0.57, respectively. *R* was little at 0.31, but *COP* was good at around 1.04. *AAE*, *SI*, and *RMSPE* were weighty high exactly at 268.66, 2.16, and 211590.38 %, respectively. The *WI* was just under half that at 0.44. Fig. 5 approved that MLPNN was unable to predict the overtopping rates higher than 0.001 $m^3/s/m$, the model gave a conservative value in these ranges.

Table 2 MLPNN configuration parameters.

Model parameters	Parameters used
Type of analysis	Regression
Validation method	Cross validation
Number of cross validation folds	10
Number of layers	3 (1 hidden)
Hidden layer 1 neurons	Search from 2 to 20, step:1
Hidden layer activation function	Logistic

Table 3 Architecture of MLPNN model.

Layer	Neurons	Activation	Min. Weight	Max. Weight
Input	10	Pass thru	---------	---------
Hidden 1	11	Logistic	-8.782e-001	8.714e-001
Output	1	Linear	-2.774e-001	3.005e-001

Table 4 MLPNN training statistics.

Process	Time	Evaluations	Error
Conjugate gradient	00:00:01.3	932,732	9.18e-005

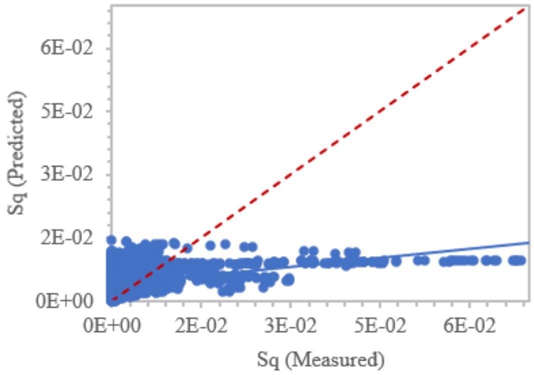

Fig. 4 Comparison between the measured and predicted dimensionless overtopping discharge by MLPNN model.

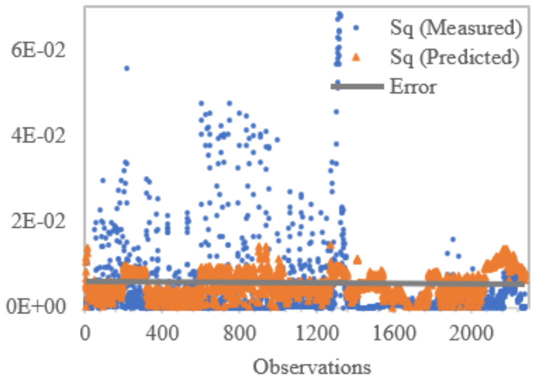

Fig. 5 Error between the measured and predicted dimensionless overtopping discharge by MLPNN model.

The model's parameters and features of the RFDT model are listed in Table 5.

Table 5 RFDT model parameters.

Model parameters	Parameters used
Type of analysis	Regression
Validation method	Hold-out validation
Splitting the dataset	60 % of the dataset is used for training and 40 % of the dataset is used for testing the model.
Sampling type	Automatic
Local random seed	1992
number of trees	140
Maximal depth of tree	7
Criterion on which attributes will be selected for splitting	Least square
Minimal gain	0.01
Minimal leaf size	2
Minimal size for split	4

The results of RFDT model are depicted in Fig. 6; the prediction points are distributed about the optimum line with some dots which were altogether overestimated. From the statistical indexes, *MSE*, *RMSE*, *MAE*, and *MAPE* were 8.97E-06, 0.003, 0.002, and 0.18, respectively. *R* was considered high at 0.94, and *COF* was 1.04. *AAE*, *SI*, and *RMSPE* were pretty great at 56.57, 0.74, and 37830.85 %, respectively. *WI* was found at 0.96. Fig. 7 showed that the predicted points were roughly comparable to the observed points with the minimum absolute error.

3.1 Comparison Between the MLPNN model and RFDT model.

In this section, the prediction abilities of the MLPNN and RFDT methods were compared. Table 6 summarizes the accuracy metrics of the models. As seen from the table, the MLPNN led to a comparatively high percentage of uncertainty with an *RMSE* of 0.009, and the corresponding *MAPE* was about half percent for the models. A modest correlation has been demonstrated between the measured and predicted values of MLPNN model, as *R* is equal 0.31 as shown in Fig. 8. *AAE*, *SI*, and *RMSPE* achieved poor results, underscoring the inaccuracy of the model for predicting wave overtopping. Furthermore, a lower *WI* (0.44) indicated a significantly lower match between the measured and predicted values.

RFDT model exhibited much better performance when compared to MLPNN model. The model had lower error values, since *RMSE* and *MAPE* were 0.003 and 0.18, respectively. The correlation coefficient is close to one as illustrated in Fig. 8, which means higher performance.

The *AAE* and *RMSPE* were lower by around 80 percent from the NN model, whereas *SI* was also lower by about 65 %. This proved that RFDT method has a better prediction estimate than the other approach investigated in the study. The *WI* was close to unity at 0.96 as shown in Fig. 8 , indicating that the data was efficiently fit in this model.

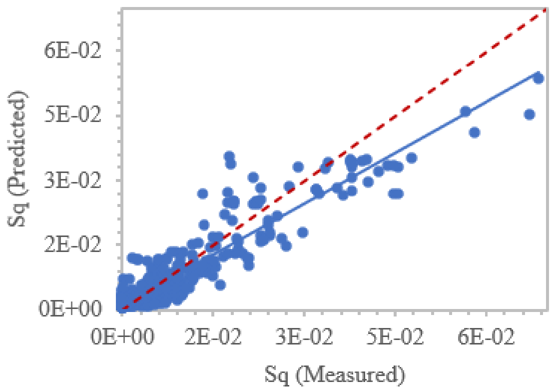

Fig. 6 Comparison between the measured and predicted dimensionless overtopping discharge by RFDT model.

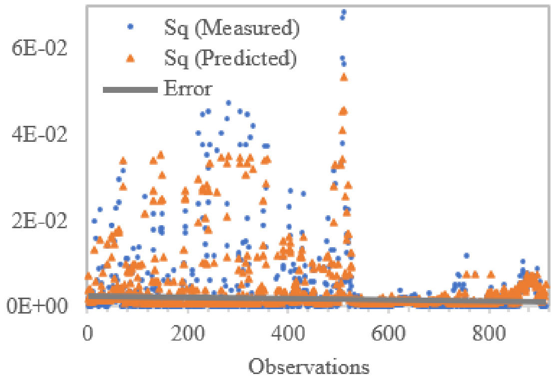

Fig. 7 Error between the measured and predicted dimensionless overtopping discharge by RFDT model.

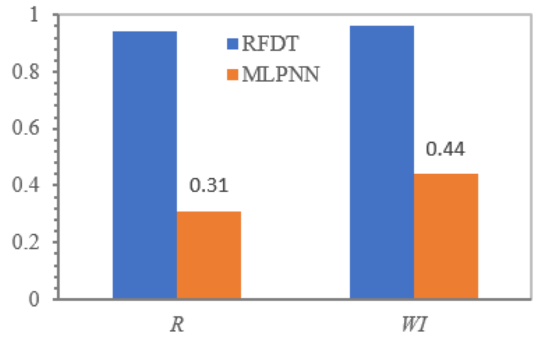

Fig. 8 Comparison between *R* and *WI* values of soft computing methods.

Table 6 Performance comparison among soft computing methods for vertical costal structures.

Model	*SD*	*MSE*	*RMSE*	*MAE*	*MAPE*	*R*	$-1 \leq DR \leq 1$	*COP*	*AAE*	*SI*	*RMSPE %*	*WI*
MLPNN	0.0044	8.50E-05	0.009	0.006	0.57	0.31	69.52	1.04	268.66	2.16	211590.38	0.44
RFDT	0.0072	8.97E-06	0.003	0.002	0.18	0.94	79.81	1.04	56.57	0.74	37830.85	0.96

Generally, from Fig. 9, the MLPNN model tends to massively underestimate the overtopping rates as it was far from the actual and predicted dash line, while the RFDT model was closer to the actual and predicted dash line, which indicates satisfactory performance of this model.

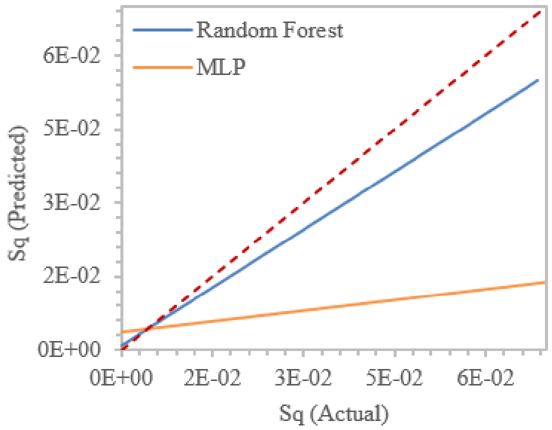

Fig. 9 A scatter plot of the measured and predicted values of wave overtopping discharge for the MLPNN and REDT methods.

The discrepancy ratios of each method for the data in Table 6 are shown in Fig. 10. *DR* is a method to judge the closeness of predictions to measurements qualitatively by plotting the histogram of *DR* values. The more the *DR* values get to zero, the more exact the match between measured and predicted overtopping rates will be.

It is clear from Fig. 10 that 65 % of *DR* values for RFDT model were close to zero. On the other side, only 41 % of matching between the measured and predicted values in MLPNN. This means that using the RFDT meathod to predict wave overtopping discharges will give reasonable and reliable estimations higher than using MLPNN method.

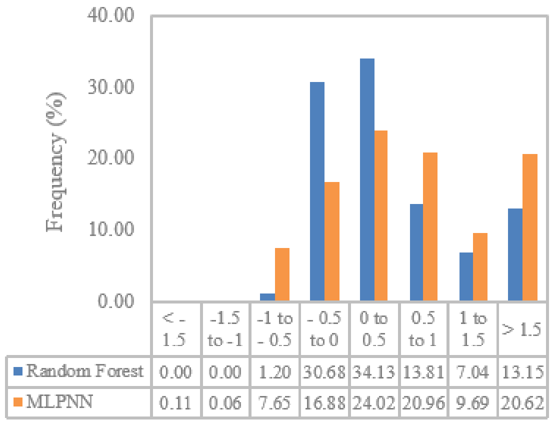

Fig. 10 Comparison of discrepancy ratio values.

4. CONCLUSION

Estimating wave overtopping discharges, which is traditionally a key criterion in the design of coastal structures, is one of the most important concerns in coastal areas. Engineers use these calculations to calculate the maximum amount of water that can exceed a vertical wall for planning, design, and risk assessment of coastal structures like vertical walls.This study developed an approach for soft computing and performance measures to predict wave overtopping discharge at verticle coastal structures using Multilayer Perceptron Neural, and Random Forest Decision Tree, models. By using eleven performance measures, the various models were compared to achieve a comprehensive comparison of the predictive methods used.

The EurOtop database was used to train and validate the MLP and REDT models. A total of 2279 data points were trained for vertical coastal structures, covering a wide variety of parameters. The results of each method's prediction were examined qualitatively using charts and statistically using accuracy measures to test the accuracy of the methods. According to the results of soft computing methods, RFDT has demonstrated its competence in predicting overtopping discharge. The method provided better estimates than the MLPNN, as it resulted in lower errors between predicted and measured values. All of the diagrams and accuracy evaluations demonstrated that the RFDT approach had the best prediction performance.

5. ACKNOWLEDGMENTS

The authors would like to express his high appreciation to the referees of the paper for their critical review as well as their valuable comments that improved the paper to its present form.

6. REFERENCES

[1] Maritime H., and Hazard F., Assessment of Economic Value of National Assets At Risk From Flooding and Coastal Erosion, 2000, pp. 1–10.

[2] Environment Agency, Flood and Coastal Erosion Risk Management Report: 1 April 2019 to 31 March 2 020, updated 29 April 2021, 2021, pp. 1–43.

[3] Pillai K., Etemad-Shahidi A., and Lemckert C., Wave overtopping at berm breakwaters: Review and sensitivity analysis of prediction models, Coast. Eng., Vol. 120, 2017, pp. 1–21.

[4] Formentin S.M., Zanuttigh B., and Van Der Meer J.W., A Neural Network Tool for Predicting Wave Reflection, Overtopping and Transmission, Coast. Eng. J., Vol. 59, 2017.

[5] Wedge D., Ingram D., McLean D., Mingham C., and Bandar Z., A Global-Local Artificial Neural Network with application to wave overtopping prediction, Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Vol. 3697 LNCS, 2005, pp. 109–114.

[6] Van Gent M.R.A., van den Boogaard H.F.P., Pozueta B., and Medina J.R., Neural network modelling of wave overtopping at coastal structures, Coast. Eng., Vol. 54, 2007, pp. 586–593.

[7] Verhaeghe H., De Rouck J., and van der Meer J., Combined classifier-quantifier model: A 2-phases neural model for prediction of wave overtopping at coastal structures, Coast. Eng., Vol. 55, 2008, pp. 357–374.

[8] Den Bieman J.P., Wilms J.M., van den Boogaard H.F.P., and van Gent M.R.A., Prediction of mean wave overtopping discharge using gradient boosting decision trees, Water (Switzerland), Vol. 12, 2020, pp. 1–13.

[9] Van der Meer J., W. A., N.W.H., Bruce T., De Rouck J., Kortenhaus A., *et al.*, EurOtop. Manual on wave overtopping of sea defences and related structures. An overtopping manual largely based on European research, but for worldwide application, 2018, pp. 320.

[10] Haykin S., Neural Networks and Learning Machines, edit Prentice Hall. New Jersey, USA. L, 2008, pp.1-885.

[11] Liu Y., Wang Y., and Zhang J., New Machine Learning Algorithm : Random Forest, 2012, pp. 246–252.