Hindawi
Advances in Civil Engineering
Volume 2023, Article ID 8272566, 13 pages
https://doi.org/10.1155/2023/8272566

Research Article

Evaluation of Hybrid Soft Computing Model's Performance in Estimating Wave Height

Tzu-Chia Chen,[1] **Zryan Najat Rashid**,[2] **Biju Theruvil Sayed**,[3] **Arif Sari**,[4] **Ahmed Kateb Jumaah Al-Nussairi**,[5] **Majid Samiee-Zenoozian**,[6] **and Mehrdad Shokatian-Beiragh**[6]

[1]*College of Management and Design, Ming Chi University of Technology, New Taipei City, Taiwan, China*
[2]*Technical College of Informatics, Sulaimani Polytechnic University, Sulaymaniyah, Iraq*
[3]*Department of Computer Science, Dhofar University, Salalah, Oman*
[4]*Department of Management Information Systems, Girne American University, Kyrenia, North Cyprus, via Mersin 10, Turkey*
[5]*Al-Manara College for Medical Sciences, Maysan, Iraq*
[6]*Department of Water Resources Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran*

Correspondence should be addressed to Mehrdad Shokatian-Beiragh; m.shokatian.s@gmail.com

Received 7 May 2022; Revised 14 September 2022; Accepted 21 March 2023; Published 18 April 2023

Academic Editor: Mohammad Najafzadeh

In coastal and port engineering, wind-generated waves have always been a crucial, fundamental, and important topic. As a result, various methods for estimating wave parameters, including field measurement and numerical methods, have been proposed over time. This study evaluates the wave height at Sri-Lanka Hambantota Port using soft computing models such as Artificial Neural Networks (ANNs) and the M5 model tree (M5MT). In order to overcome its nonstationarity, the primary wave height time series were divided into subtime series using the wavelet transform. The collected subtime series were then utilized as input data for ANN and M5MT in order to determine the wave height. For the sake of the model performance, the daily wind and wave data from the Acoustic Wave and Current (AWAC) sensor for Hambantota Port in 2020 and Sanmen Bay in 2017 were used in this study. The training state utilizes 80% of the available data, while the test state uses 20%. The Root Mean Square Error (RMSE) of the ANN, M5, WANN, and Wavelet-M5 models in the Hambantota Port for the test stage are 0.12, 0.11, 0.04, and 0.06, respectively. While in Sanmen Bay, the RMSE of the ANN, M5, WANN, and Wavelet-M5 models for the test stage are 0.14, 0.16, 0.06, and 0.08, respectively. According to the findings of this study, the accuracy of WANN and Wavelet-M5 hybrid models in evaluating wave height is superior to that of classic ANN and M5MT, and it is recommended that WANN and Wavelet-M5 hybrid models be used to estimate wave height.

1. Introduction

Water waves are the most obvious, almost permanent phenomena on the surface of any water basin, such as wetlands, lakes, rivers, reservoirs behind dams, bays, seas, and oceans. They are usually defined as the surface oscillation of the fluid surface [1]. Wave study is the first step for any study and activity in order to identify the factors affecting the behavior and conditions in the sea [2]. In coastal areas, waves play an important role in determining the geometry and shape of beaches. The height of the sea waves, while creating the first feeling about the occurrence of the wave, is the most important parameter in all issues raised in coastal engineering studies. In designing marine structures such as platforms, breakwaters, and jetties, the main parameter in determining their various components' stability and design is the wave height in the region [3, 4]. When waves approach coastal areas, they are deformed due to various phenomena such as shallow, scattering, refraction, and reflection, which are important in various aspects such as management, protection, and exploitation of the coast, environment, fisheries, navigation, and construction of structures [5–7]. The study of sea wave's offshore and onshore structures develops basic knowledge in the field of coastal engineering and the physics of the sea and waves. In coastal areas, determining the pattern of waves and coastal currents is the most important; features are proposed to identify the factors affecting the marine environment, coastal areas, and coastal structures [8]. The beach's geometry, shape, sedimentation, erosion, and many other physical and dynamic phenomena are directly affected by waves and currents. Wind waves are the most important waves observed at sea and have the greatest impact on human activities in the marine environment; therefore, when it comes to forecasting waves for engineering purposes, mainly wind waves are considered [9]. Although field measurements are the most accurate way to obtain the wave parameters of any region, the field measurement method alone will not be able to respond when determining waves in a wide area [10]. Today, using numerical models as an efficient tool for simulation and then studying complex natural processes open the way for many technical and engineering issues, including the state of the sea. Soft computing methods such as model tree (MT), gene expression programming (GEP), multivariate adaptive regression spline (MARS), adaptive-neuro fuzzy inference system (ANFIS), and Bayesian Network (BN) have proven successful applications for modeling various ocean engineering problems [11–16]. In addition, many studies demonstrated the combination of properties of different soft computing methods with evolutionary algorithms causing an improvement in the power prediction of phenomena in solving ocean environment problems [17–21].

Due to the random and irregular nature of the sea, estimating the height of the waves is associated with inherent uncertainty. Uncertainty in estimating the wave height and the consequent forces acting on the structure causes uncertainty in the design of the members of the marine structures. Also, the coefficients used to determine the drag and inertia forces are always uncertain. Given the capabilities of mathematical models with the help of numerical simulation, using these methods in predicting wave properties at sea is appropriate. Since the forecast wave parameters are essential for the design of coastal structures and for naval operations, different methods such as semiempirical methods such as Coastal Engineering Manual (CEM) and Sverdruv Munk Bretschneider (SMB) and numerical models such as MIKE21, Wavewatch III, and SWAN were used [22, 23]. Soft computing methods such as Artificial Neural Networks (ANNs), fuzzy inference systems, decision trees, and genetic algorithms are also used.

There are two approaches to modeling sea parameters in general, namely, conceptual (white-box) and systemic (black-box). White box models are based on governing mathematical equations and physical parameters of the phenomenon. The purpose of these models is to rely on scientific research on how the main components of each sea parameters cycle work to fully understand the mechanism and how the components work together. Hence, understanding and interpreting white-box models are more straightforward than black-box models. In black-box models, it is difficult to present equations and mathematical relations in them, and the physical parameters affecting them cannot be easily estimated. Black-box models estimate the desired output by receiving input and performing a series of mathematical operations. Black box models have parameters and coefficients that are estimated according to observational input and output data [24]. Therefore, black-box models depend on input and output data in terms of quantity and quality of data.

In this study, an attempt was made to develop an efficient wave evaluation model based on the innovation hybrid models. This study evaluates wave height at Sri Lanka Hambantota Port and China Sanmen Bay using ANNs (surrogate of the nonlinear model) and the M5 model tree (M5MT, surrogate of the multivariate linear regression model). For this purpose, the wind and wave data were gathered in Hambantota Port, Sri-Lanka 2020 and Sanmen Bay, China 2017. Primary wave height time series were divided using the wavelet transform to overcome nonstationarity.

2. Study Area and Data Processing Methods

Sri Lanka's Hambantota International Port is a deep water port in the country's south and directly faces the North Indian Ocean (Figure 1). After the Port of Colombo, it is Sri Lanka's second-largest port. In its plan for the Hambantota Port, the Sri Lankan government thought it would deliver commercial benefits and logistical feasibility. The dominant wave directions range from 157.5° to 225°. There are approximately 95% of waves concentrated between the South and Southern South West ($H_s > 2.2$ m), and the predominant wave direction is southward, with about 60% occurrences [25].

Sanmen County is a coastal county in the eastern part of China's Zhejiang Province. There are approximately 400,000 people living in the county, which has a total land area of 1,072 km^2. Sanmen Bay is a semienclosed bay, and the easterly direction of the waves is the most common one in this region ($H_s > 1.5$ m). [26]. An Acoustic Wave and Current (AWAC) meter was used to measure the waves that were used in this study.

The smoothed wave spectrum is used to figure out wave spectral parameters like the zeroth order spectral moment (m_0), the maximum spectral energy density $S\ (f_p)$, and the mean wave periods (T_{01} and T_{02}). The following are some definitions of the wave parameters used in the study:

$$
\begin{aligned}
T_{01} &= \frac{m_0}{m_1}, \\
T_{02} &= \sqrt{\frac{m_0}{m_2}}, \\
m_n &= \int_0^{\infty} f^n S(f)\mathrm{d}f; n = 0, 1, 2, \cdots,
\end{aligned}
\tag{1}
$$

where $S(f)$ is the spectral energy density at frequency f and m_{n} is the n^{th} order spectral moment.

The JONSWAP spectra formulation was widely advocated for describing wind-generated waves with durations below 20 seconds equation (3). As a result, the spectral density of the input JONSWAP spectrum is minimal at low frequencies (0.03 Hz).

$$E(f) = \frac{\alpha g^2}{f^5} \exp\left[-\frac{5}{4}\left(\frac{f_p}{f}\right)^4\right] \gamma^{\exp\left[-\frac{\left(f - f_p\right)^2}{2\sigma^2 f_p^2}\right]}, \tag{2}$$

$f_p = 22[g^2/(U_{10}F)]^{1/3}$, where F is wind fetch length, U_{10} is wind speed at 10 meter above the water surface level, g is gravity acceleration, σ is the shape parameter, and γ is the bandwidth parameter.

The significant wave height, H_s, calculated using in this study, is obtained from equation (4).

$$H_S = 4\sqrt{m_0}, m_0 = \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} \eta^2 dt, \tag{3}$$

$t_2 - t_1$ is the time domain and η is the free-surface elevation.

A filter-pass module is utilized to achieve H_s. Using the discrete Fourier transform, this module may apply a band-pass filter to the surface elevation at various frequency ranges and time steps. This is done to avoid the generation of time-series data about surface elevations at a large number of places.

$$\begin{aligned} H_x(X,Y) &= 4\sqrt{m_x(X,Y)}, \\ m_x(X,Y) &= \int_f^{f+}\left[\left(\int_{t_1}^{t_2} \eta(X,Y)\cos(2\pi f \cdot t)dt\right)^2 + \left(\int_{t_1}^{t_2} \eta(X,Y)\sin(2\pi f \cdot t)dt\right)^2\right]\frac{\Delta T}{2} df. \end{aligned} \tag{4}$$

3. Materials and Methods

This study uses ANN and M5MT models and their combination with wavelet transform decomposition. Therefore, a review of the theoretical foundations of these methods seems necessary. The framework of this study is presented in Figure 2.

Normalization of data is the first stage in designing a forecast using machine learning. It can facilitate the training process [27]. This data fall between 0 and 1. Normalization data are presented as follows:

$$Z = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \tag{5}$$

where Z is the normalized data value, x_i is the data before normalization, and $x_{\min}$ and $x_{\max}$ are the prenormalization minimum and maximum data values, respectively.

3.1. Artificial Neural Network (ANN). Each ANN model is typically made up of three layers (Figure 3). The input layer is responsible for introducing network input parameters, the output layer is responsible for network output parameters, and the hidden layer (layers between the input and output layers) is responsible for information processing [28–31].

The main control parameters of artificial neural network methods are between neurons, which are called connection resistors called weights. Each neuron receives the weighted outputs ($W_{j,i}$, x_i) of the neurons of the previous layer, and together they produce a net input to the neuron j (net$_j$) according to the following equation:

$$\text{net}_j = \sum W_{j.i} x_i + b_j. \tag{6}$$

The multilayer perceptron (MLP) neural networks are a type of a progressive neural network in which each neuron in one layer is connected to the neurons in the next layer. MLP learning, like multilayer networks, employs a variety of learning algorithms, the most common of which is the error propagation algorithm. An algorithm was used in the current study. Matlab tool was used to simulate ANN structures and determines the best structure.

The ANN architecture is critical to the network's understanding of variable relationships. The problem always dictates a portion of the neural network architecture [33]. According to the problem, the number of network inputs equals eight, and the number of output layer neurons equals one. To obtain this, various architectures were created, trained, and tested. Finally, in a two-layer network with five neurons in the first layer (hidden layer) and one neuron in the second layer (output layer), the transfer function Tangent Sigmoid for the first layer is introduced as the best network architecture in this prediction. The network architecture and the linear transfer functions (purelin) and (tansig) for the second layer are shown in Figure 3.

3.2. M5 Model Tree. The decision tree in data mining is a model used to represent classifiers and regressions. This tree consists of a number of nodes and branches [34]. The leaves represent the classes in the decision tree that performs the classification operation. In each of the other nodes (nonleaf nodes), a decision is made according to one or more specific attributes. The decision tree is a popular data mining technique because of its simplicity and comprehensibility; in other words, the decision tree alone describes everything and does not need an expert to interpret the output [35]. In fact, it is a graphical method, and because of its interpretation, it may be easier to classify than other techniques. Obviously, having too many nodes in a tree can make it difficult to graphically display the decision tree. The first step in creating a tree model is to use a branch criterion performed by one of

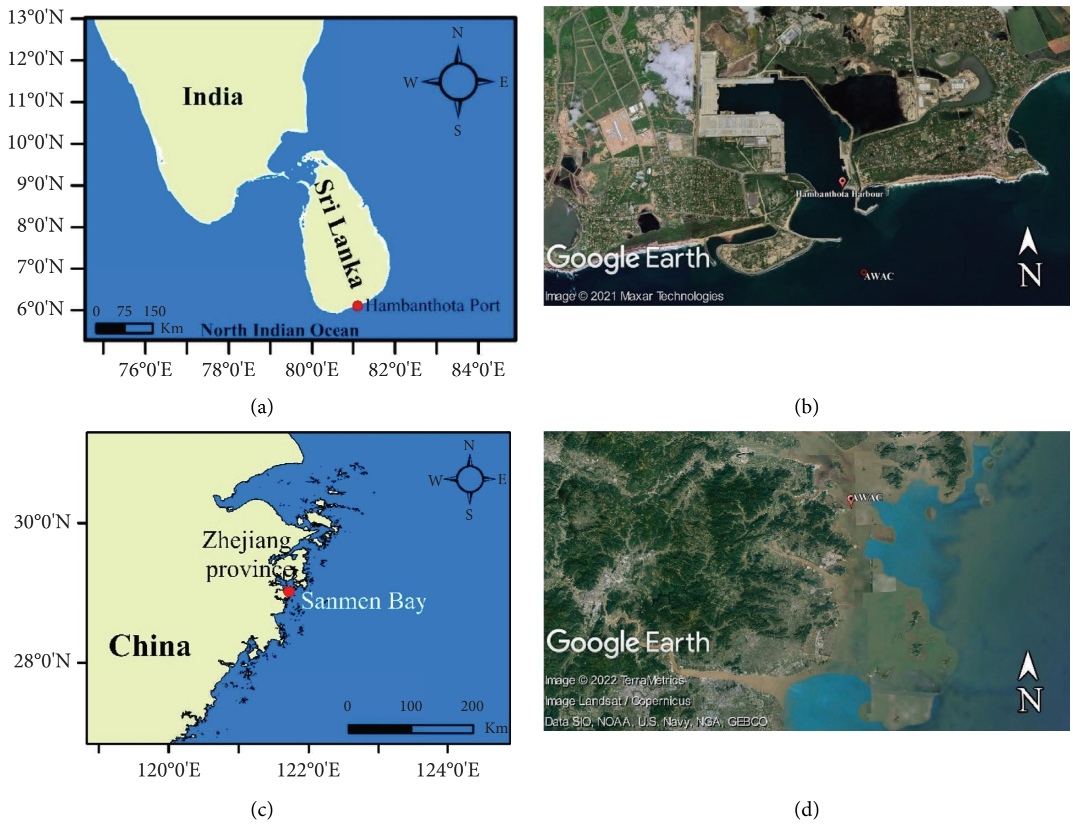

Figure 1: Study area. (a and b) Hambantota Port, Sri Lanka; (c and d) Sanmen Bay, China.

the predictor variables. The branching criterion for the M5 algorithm is based on the standard deviation function of the values of each class obtained in each node. This method is the basis of classification methods called entropy [36]. Entropy can be interpreted as a measure of the turbulence of a system. The branch criterion expresses the amount of error in that node, and the model calculates the minimum expected error as a result of testing each attribute in that node. Model error is generally measured by measuring the accuracy of predicting target values of unseen items [37]. The equation for calculating the standard deviation reduction (SDR) is as follows:

$$\mathrm{SDR} = \mathrm{Sd}\left(T_i\right) - \sum_{i=1}^{N} \frac{\left|T_i\right|}{\left|T\right|} Sd\left(T_i\right), \tag{7}$$

where T is a set of samples that enter each node, Sd indicates standard deviation, and N displays the data number. Because of the branching process, the data in the child nodes have a lower standard deviation than the data in the mother node and are thus purer. M5 chooses the trait that maximizes the expected reduction after maximizing all possible branches.

The formed tree in the M5MT needs to have its branches trimmed so that the overfitting issue can be resolved. This is accomplished by switching out a subtree for a leaf in the tree. Therefore, the second step in the process of designing a tree model is to perform a pruning operation on the mature tree and then replace the subtrees with linear regression functions. This technique for the generation of tree models divides the space of input parameters into areas that contain smaller subspaces and then fit a linear regression model in each of those areas.

3.3. Wavelet Transform. Wavelet transform is one of the efficient mathematical transformations in the field of signal processing. Mathematical transformations are used to obtain additional information from a signal that is not available from the signal itself. Wavelet analysis like Fourier analysis, which is one of the most popular mathematical transformations, deals with the expansion of functions, but this expansion is based on wavelets [38, 39]. The wavelet is a characteristic function of a hypothesis with a mean of zero and, unlike trigonometric polynomials, is studied locally in space. In this way, a closer relationship between some functions and their coefficients is possible, and more numerical stability is provided in the reconstruction and calculations [2]. Any application that is based on fast Fourier transformation can be formulated using wavelets to obtain more spatial or temporal information [40]. A wavelet function is a function that has two important properties, namely, fluctuating and short-lived. In other words, $\psi\,(x)$ is a wavelet function if and only if its Fourier transform $\psi\,(\omega)$ satisfies the following condition:

$$\int_{-\infty}^{+\infty} \frac{\left|\psi(\omega)\right|}{\left|\omega\right|^{2}} d\omega < +\infty. \tag{8}$$

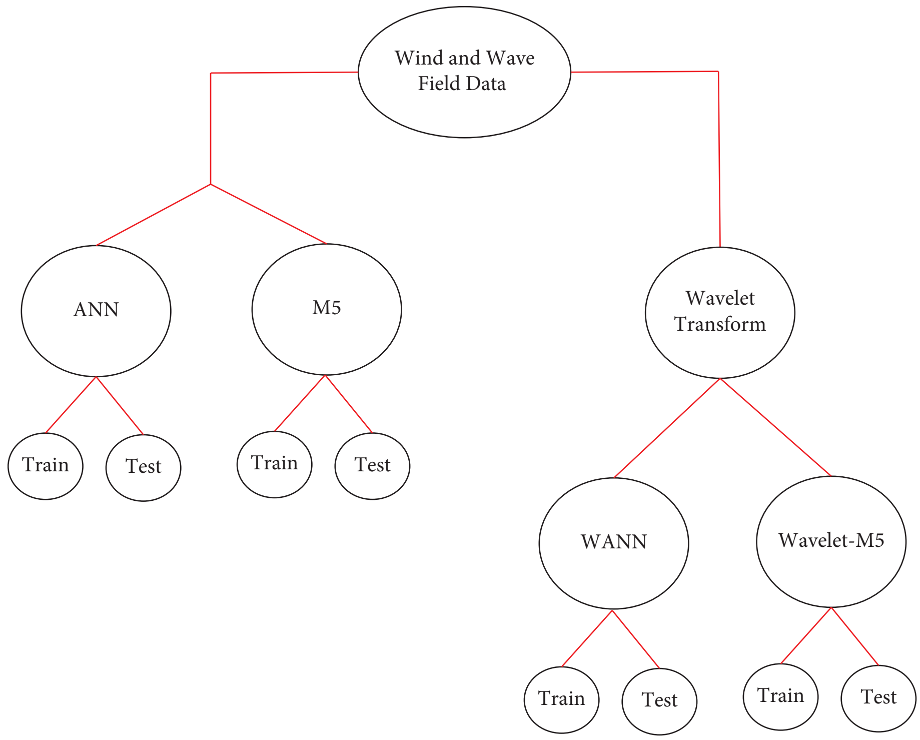

FIGURE 2: Data analysis framework.

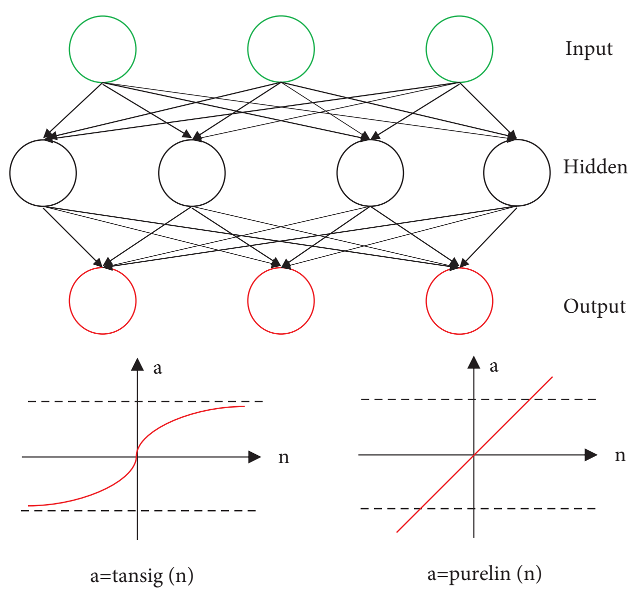

FIGURE 3: ANN structure schemes.

This condition is known as the wavelet acceptance condition $\psi\ (x)$. The previous relation can be considered equivalent to the following formula that must be satisfied:

$$\psi(0) = \int_{-\infty}^{+\infty} \psi(x)\mathrm{d}x = 0. \tag{9}$$

This property of a function with a mean of zero is not very restrictive, and many functions can be called wavelet functions based on it. $\psi\ (x)$ is a mother wavelet function in which the functions used in the analysis are scaled and shifted along with the analyzed signal by two mathematical operations of shifting and scaling. Finally, the wavelet coefficients at any point in the signal (b) and for any value on the scale (a) can be calculated by the following equation:

$$\psi_{a.b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right). \tag{10}$$

The operation of scaling, as a mathematical operator, expands or compresses the signal for the assumed function f(t); if ($s < 1$), the expanded state is f(st), and if ($s > 1$), the compressed state is the function f(t). As shown in equation (3), in the definition of the wavelet transform, the term of scale (a) is in the denominator, and therefore, if it is ($a < 1$), the signal is compressed, and if it is ($a > 1$), the signal is expanded. Also, in the previous equation, parameter (b) is modeled as a function of delay or precedence [41]. Finally, the continuous wavelet transform (CWT) can be written as follows:

$$\mathrm{CWT}_{(a.b)} = Wf_{(a.b)} = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x)\psi\left(\frac{x-b}{a}\right)\mathrm{d}x = \int_{-\infty}^{+\infty} f(x)\psi_{a.b}(x)\mathrm{d}x. \tag{11}$$

3.4. Hybrid Wavelet-Artificial Neural Network Model (WANN). The WANN model has a structure that is composed of three layers. The first layer of the network comprises wavelet neurons, and their input is a subseries obtained by applying a wavelet transform to a time series of wave height evaluations. In order to determine the weight coefficients of wind speed at the height of 10 meters in the network structure, the WANN model uses the neural network to perform the necessary calculations [42]. The time series of the wave height assessment is initially segmented into subseries using a variety of scales according to the structure of this model. For instance, time series can be segmented into one long-term scale and several short-term scales (in order to monitor transient properties and fluctuations).

3.5. Hybrid Model Wavelet-M5. The Wavelet-M5 hybrid model that has been proposed has a total of four stages [40]. In the first step, information pertinent to the study area's wave height evaluation is compiled from the data collected [43]. The preprocessing of data is the second step and is necessary because there is a possibility that the estimated height of the waves will change depending on the spatial and temporal distributions of the data. Effective preprocessing has the potential to make data-driven methods more productive. One of the potential approaches to preprocessing the data is the use of wavelet analysis. Clustering the data is done in the third step of the process, not only so that the data can be organized into similar groups but also so that the structure of the model can be optimized. Clustering the data serves the following two purposes: first, it helps organize the collected information into meaningful categories; second, it enables the third step of the process, which is to optimize the structure of the model. When determining the nature of the connection between independent and dependent variables, we use the M5MT, which is an application of the tree classification method. The repetitive patterns that are present in the data are identified and extracted during the final stage of the model that has been proposed. This is done in order to finally provide tree regression models for each of the subgroups [8, 44].

4. Results and Discussion

The wave characteristics at any given time are determined by the current wind speed and previous wind speeds. As a result, the height of the waves may be affected by the wind speed 10 hours earlier. The following equation is used to simulate and estimate the height of waves in Hambantota Port.

$$H_s = f\left(U_t, U_{t-1}, U_{t-2}, U_{t-3}, U_{t-4}, U_{t-5}, U_{t-6}, U_{t-7}\right), \tag{12}$$

where t is the time in hour, U is wind speed at 10 meter above the sea level in the Buoy location, and H_s is the observed significant wave height in the Buoy location. In order to evaluate and develop the models, wind data and Acoustic Waves and Current (AWAC) statistics of the Hambantota Port, in 2020 have been used. For this purpose, 80% of the data has been used to train soft computational models, and the rest of the data have been used to evaluate and validate the performance of trained models.

To evaluate the performance of models, statistical measures are utilized. For verification and quantitative evaluation of the performance of the presented models, statistical indicators such as Nash–Sutcliffe Model Efficiency Coefficient (NSE), Mean Average Error (MAE), Root Mean Square Error (RMSE), and correlation coefficient (R) have been used. In the mentioned relations of N number of observational data, X_i and Y_i indicate observational and predicted parameter, respectively. $\overline{X}$ and $\overline{Y}$ are average observational and predicted values, respectively. The performance of models was evaluated using the error indices.

$$\begin{aligned}
\mathrm{NSE} &= 1 - \frac{\sum_{i=1}^{N}\left(X_i - Y_i\right)^2}{\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^2},\\
\mathrm{MAE} &= \frac{1}{N}\sum_{i=1}^{N}\left|X_i - Y_i\right|,\\
\mathrm{RMSE} &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_i - Y_i\right)^2},\\
R &= \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum\left(X_i - \overline{X}\right)^2 \sum\left(Y_i - \overline{Y}\right)^2}}.
\end{aligned} \tag{13}$$

4.1. Development of ANN Model. A simple perceptron network with sigmoid transfer function was used to develop the model of artificial neural networks. Determining and selecting the optimal middle layers and the number of

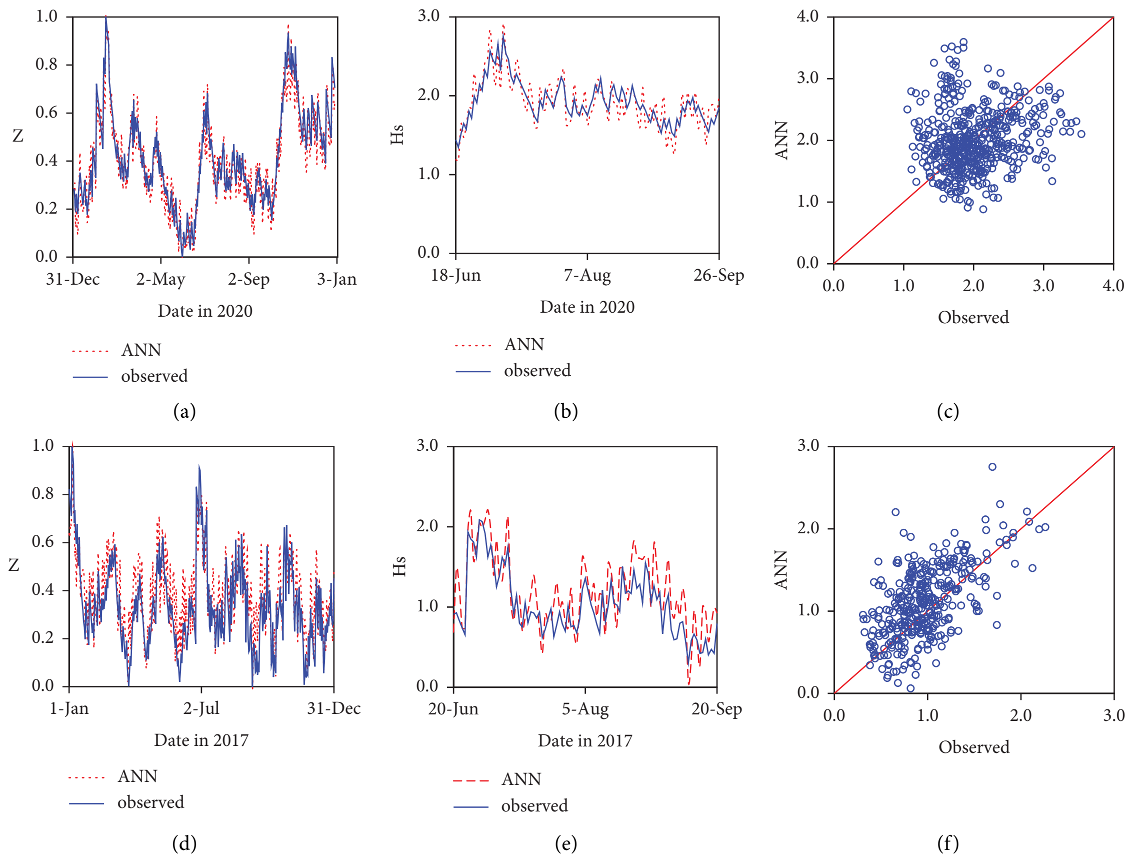

FIGURE 4: Times series of test and train normalized significant wave height and the corresponding scatter diagram by the classic ANN model (a–c) in the Hambantota Port (2020) and (d–f) in the Sanmen Bay (2017).

TABLE 1: Results of different modeling for the Hambantota Port at daily scales.

Input	Output	Case study	Stage	Model	Efficiency criteria				*P*-value
					NSE	MAE	RMSE	R	
U_t	H_{st}	Hambantota Port	Train	ANN	0.74	0.24	0.06	0.37	0.017
			Test		0.66	0.27	0.12	0.34	0.021
		Sanmen Bay	Train		0.70	0.26	0.08	0.34	0.024
			Test		0.63	0.31	0.14	0.30	0.037
		Hambantota Port	Train	WANN	0.93	0.18	0.01	0.43	<0.001
			Test		0.90	0.21	0.04	0.40	<0.001
		Sanmen Bay	Train		0.91	0.20	0.03	0.42	<0.001
			Test		0.87	0.22	0.06	0.38	0.001
		Hambantota Port	Train	M5	0.72	0.23	0.11	0.39	0.011
			Test		0.64	0.28	0.13	0.36	0.033
		Sanmen Bay	Train		0.69	0.26	0.07	0.35	0.021
			Test		0.60	0.30	0.16	0.30	0.051
		Hambantota Port	Train	Wavelet-M5	0.94	0.17	0.03	0.46	<0.001
			Test		0.89	0.20	0.06	0.43	0.002
		Sanmen Bay	Train		0.88	0.18	0.04	0.42	<0.001
			Test		0.84	0.22	0.08	0.38	0.002

neurons in this layer in the ANN model have always been a contentious issue. However, research has shown that the use of a middle layer can be useful for modeling complex and nonlinear problems. The middle layer was obtained to be 18 through trial and error. It is important to note that a low number of training repetitions can result in incomplete training, whereas a high number of repetitions can result in network retention or disruption during the training phase. As a result, the optimal number of repetitions should be considered so that the model's quality is acceptable for both training and testing. According to previous research, this is between 150 and 200 [40]. As a result, an ANN model with

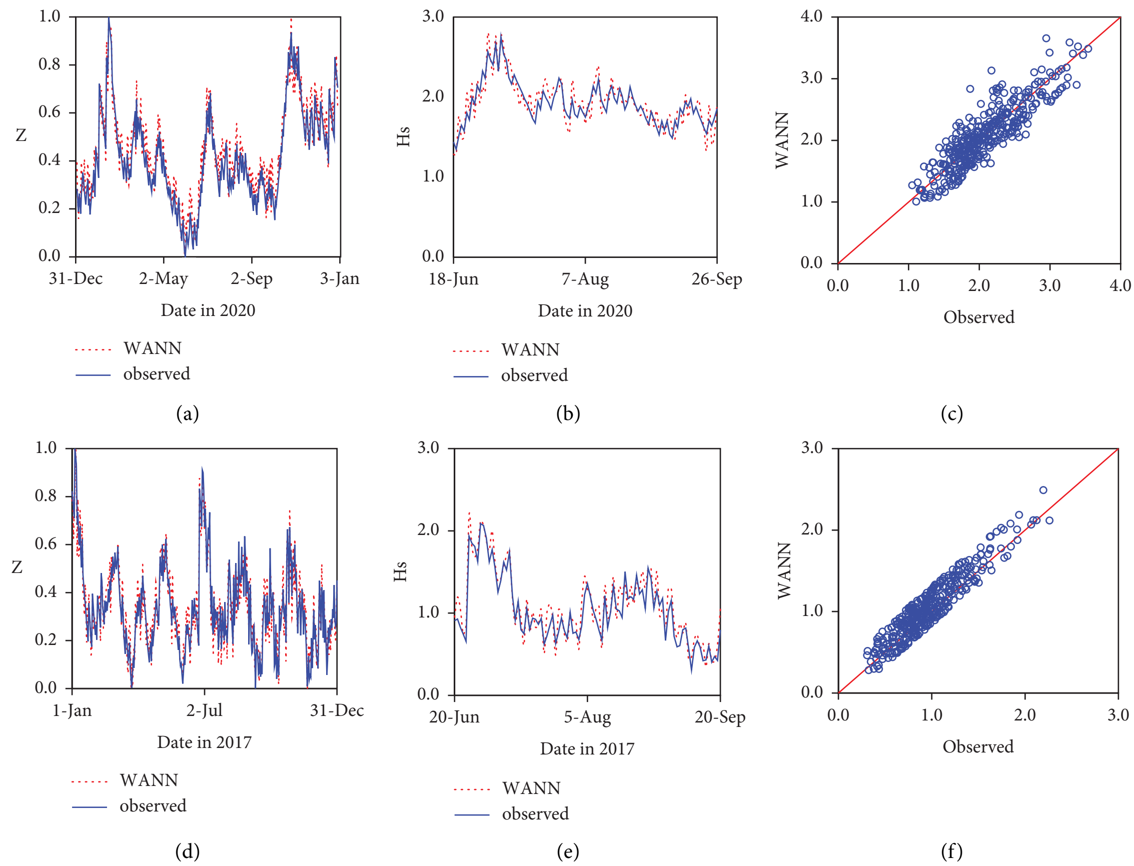

Figure 5: Times series of test and train normalized significant wave height and the corresponding scatter diagram by the WANN model (a–c) in the Hambantota Port (2020) and (d–f) in the Sanmen Bay (2017).

an optimal arrangement of $8 \times 17 \times 1$ was obtained to estimate wave height. To estimate the number of neurons in the hidden layer, Nielson's relationship is defined as follows:

$$N^H \leq 2N^I + 1, \tag{17}$$

where N^H is the number of neurons in the hidden layer and N^I is the number of input parameters.

In the Hambantota Port, the straightforward ANN model did not perform particularly well when attempting to predict the significant wave height. It is possible that the inability of the simple ANN model to deal with the instability of the input time series is the single most important factor contributing to this result. Meanwhile, in the Sanmen Bay as well as Hambantota Port, the ANN model's performance was not desirable. Figure 4 presents the findings of a comparison between the significant wave heights that were observed and those that were simulated using ANN for the train and test states. The scatter plot for the observed and simulated results in the train and test states is depicted in Figures 4(c) and 4(f) for Hambantota Port and Sanmen Bay, respectively. The values for the efficiency criteria are presented in Table 1, which compares the train state to the test state. The traditional ANN model did not perform very well in terms of predicting significant wave height in both case studies.

4.2. Development Wavelet-Neural Network Model (WANN). In this study, the outcomes of the WANN model and the ANN model were compared with one another. Decomposing significant wave height data into subseries using a wavelet neural network (WANN) model is a technique that can be used to improve the accuracy of ANN models. In a manner analogous to that of ANN modeling, the WANN models were created by applying various ANN architectures to various input combinations. The results of different decomposition levels for an input are listed in Table 1, along with the best performance indices. To obtain the best possible outcomes, various levels of decomposition, ranging from level 2 to level 5, were scrutinized. Applying the wavelet transform should result in an increase in the accuracy of the model in comparison to the traditional ANN model and should also result in an improvement in the model's efficiency, as shown in Table 1. A comparison of the simulated and observed time series of the significant wave height for the test and train states is presented in Figure 5, which is based on the WANN model. Figures 5(c) and 5(f) represent a scatter plot of observed and simulated results in the train and test states in the HambantotaPort and Sanmen Bay, respectively. As can be seen, the WANN model outperforms the traditional ANN model in terms of performance in the both case studies. As a result, the WANN hybrid model is significantly more appropriate for use in the investigation of

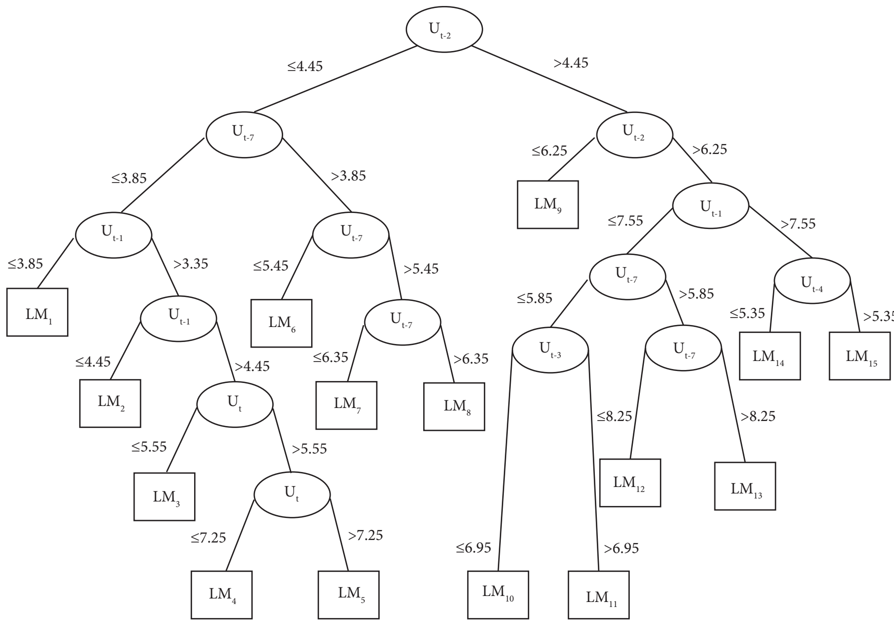

FIGURE 6: M5MT structure to evaluate wave height.

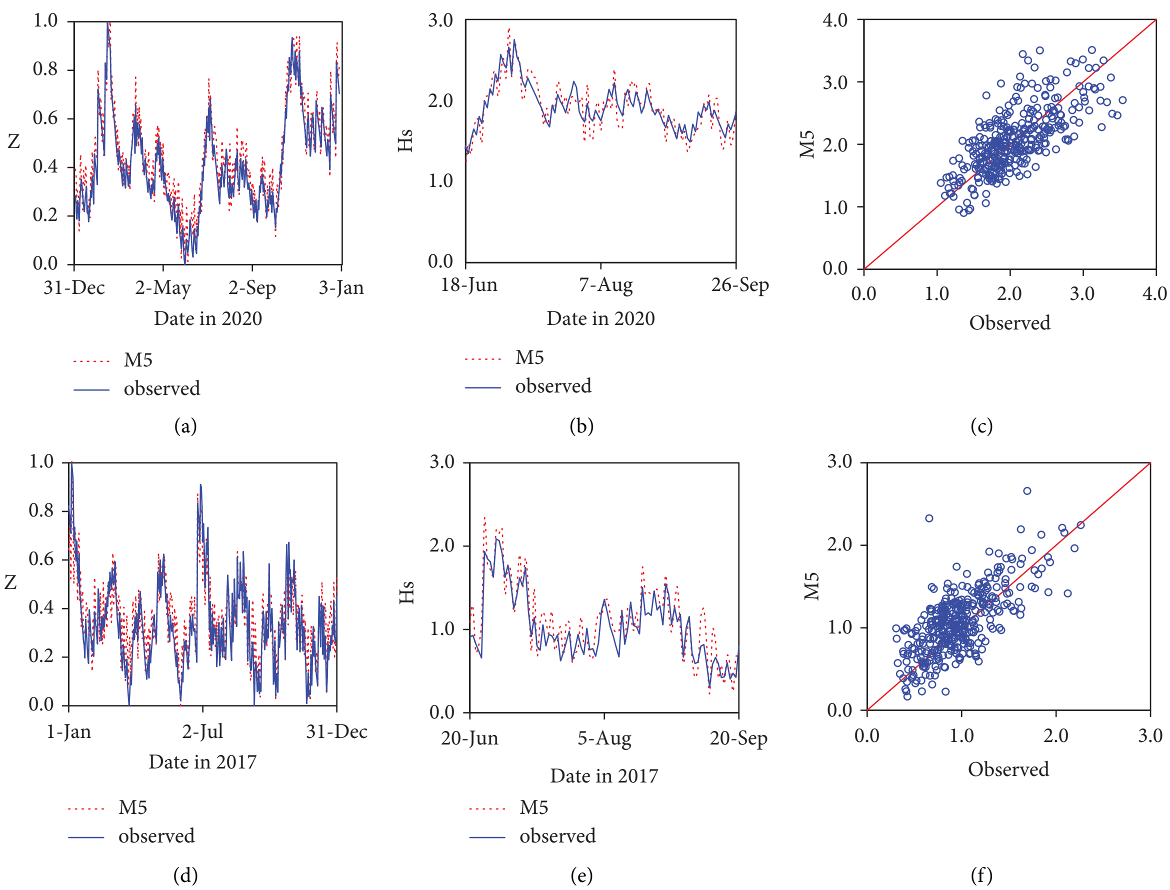

FIGURE 7: Times series of test and train normalized significant wave height and the corresponding scatter diagram by the M5MT model (a–c) in the Hambantota Port (2020) and (d–f) in the Sanmen Bay (2017).

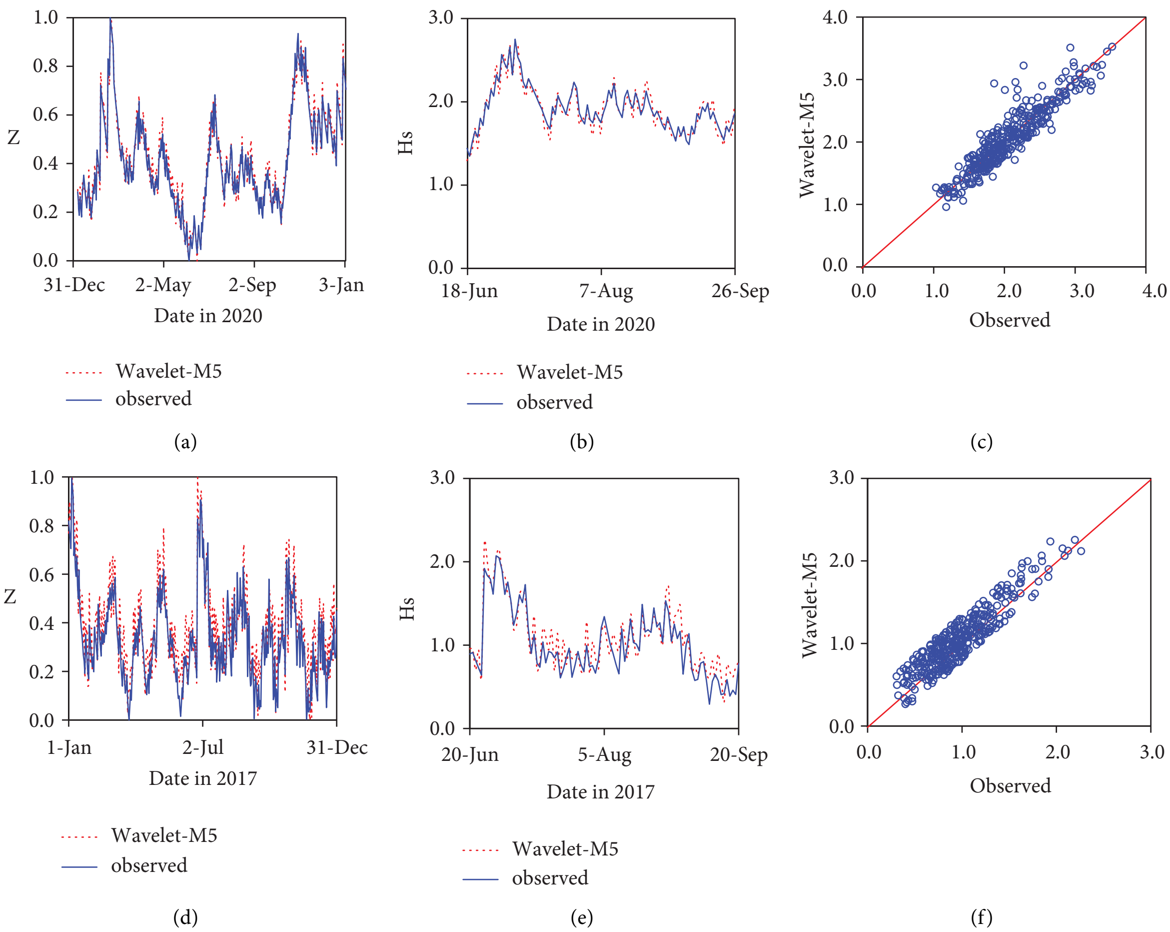

Figure 8: Times series of test and train normalized significant wave height and the corresponding scatter diagram by the wavelet- M5 model (a–c) in the Hambantota Port (2020) and (d–f) in the Sanmen Bay (2017).

the process of wave height. It is essential to utilize wavelet transformation in order to boost the quantity of data that are input. Although the model improved the accuracy of prediction, in addition, the number of calculations increased significantly, which increased the amount of time required to perform the calculations.

4.3. Development M5 Model Tree. It is possible to partition a difficult modeling issue into a number of manageable subtasks, and the solution is to combine the answers to all of these problems. In order to obtain an estimate of the height of the waves in Hambantota Port, the tree structure depicted in Figure 6 was obtained through the utilization of the M5MT. In the M5MT, regression relations in the final leaves are used to estimate the wave height. This is done so that the target parameter can be estimated based on the input variables that have been introduced to the model. It can be seen in the tree structure that the M5MT offers up for inspection. At the bottom of the tree, the final leaf, which contains fifteen rectangles, is obtained. As a result, fifteen regression relationships are presented in order to estimate the height of the waves in Hambantota Port. Moving from the root node at the top of the tree to the final leaves at the bottom of the tree is sufficient, according to the M5MT, in order to fulfill the requirements of each rule. The numbers are displayed on the various branches, each representing a boundary between the various relationships shown. In order to categorize models that are nonlinear by their very nature, the M5MT divides the nonlinear model into classes that are capable of being modeled by a straightforward linear regression. Table 1 presents the findings of the M5MT analysis of the efficiency of tree performance. Figure 7 is a comparison of simulated and observed time series for the significant wave height for the test and train states. This comparison is based on the M5MT. Figures 7(c) and 7(f) illustrate a scatter plot comparing the observed and simulated results obtained in the train and test states in the Hambantota Port and Sanmen Bay, respectively. The results are summarized in Table 1, and they demonstrate that the performance of the M5MT is comparable to that of a traditional ANN model. The M5MT did not perform particularly well due to the stochastic nature of the sea state process. This was similar to the problem the traditional ANN model had.

4.4. Development Wavelet M5 Model Tree (Wavelet-M5). Wavelet decomposition was used to transform the wave height time series into subsignals to manage the involved trend in the main series, which is analogous to the WANN

Table 2: Comparison of the hybrid models with ANN and M5MT (according to efficiency criteria average).

Model	Case study	Train (%)	Test (%)
WANN vs. ANN	Hambantota Port	39	41
	Sanmen Bay	36	38
Wavelet-M5 vs. M5	Hambantota Port	36	37
	Sanmen Bay	30	33
WANN vs. M5	Hambantota Port	35	37
	Sanmen Bay	32	35
Wavelet-M5 vs. ANN	Hambantota Port	33	34
	Sanmen Bay	30	32

model. After that, M5MT was constructed with the inputs from each of the subsignals. Table 1 contains the detailed findings regarding the Wavelet-M5 performance regarding its effectiveness. In Figure 8, a comparison of simulated and observed time series for significant wave height is presented for the test and train states based on the wavelet-M5 model. This comparison is made for both of these states when they were in the train state. A scatter plot of the observed and simulated results in the train and test states in the Hambantota Port and Sanmen Bay is depicted in Figures 8(c) and 8(f), respectively. As seen in Table 1, the application of wavelet transform improved both the model's accuracy and overall efficiency, which occurred as a direct consequence of the unfavorable results produced by the M5MT. As a result, the Wavelet-M5 hybrid model has a performance that is significantly superior to that of the M5 model in both case studies. The fact that the M5MT does not perform any sort of data preprocessing is one of the most straightforward conclusions that can be drawn from the table. As a result, the M5NT on its own cannot be considered a tool for processing data.

4.5. Comparison of the Models. The results of the WANN and wavelet- M5 models have less scatter compared to the results of other models, and the data are closer to the optimal line. This can be seen in the scatter plots, which are shown in Figures 4, 5, 7, and 8. While the results of the traditional ANN model and the M5MT are significantly scattered from the optimal line, the results of the M5MT are more consistent. Nevertheless, the results of the comparison indicate that hybrid models are preferable to the straightforward computational model in terms of the desired outcomes. As a result, in the Hambantota Port, the NSE values for ANN and M5MT are 0.74 and 0.72, respectively, for the train state and 0.66 and 0.64, respectively, for the test state. On the other hand, it was discovered that the WANN and wavelet-M5 models are more accurate than the ANN and M5MT. As a result, the NSE for the WANN and wavelet- M5 models is 0.93 and 0.94 for the train state, and it is 0.90 and 0.89 for the test state, respectively.

In the Sanmen Bay, the NSE values for ANN and M5MT are 0.70 and 0.69, respectively, for the train state and 0.63 and 0.60, respectively, for the test state. On the other hand, it was discovered that the WANN and wavelet-M5 models are more accurate than the ANN and M5MT. As a result, the NSE for the WANN and wavelet-M5 models is 0.91 and 0.84 for the train state, and it is 0.87 and 0.84 for the test state, respectively. As a result, the performance of the proposed hybrid model is desirable in both case studies, and it is comparable to the quality of the WANN hybrid model, which is known as the model that is considered to be the optimal model.

According to Table 2, hybrid models' performance is better compared to simple models' performance. As a result, the performance of the hybrid wavelet- M5 model is improved by 37 percent compared to the M5MT performance. In addition, the WANN demonstrated a performance that was approximately 41% superior to that of the traditional ANN model.

According to Table 2, hybrid models' performance is better compared to simple models' performance. As a result, the performance of the hybrid wavelet-M5 model is improved by 37 and 33 percent compared to the M5MT performance in Hambantota Port and Sanmen Bay, respectively. Furthermore, in Hambantota Port and Sanmen Bay, the WANN performed approximately 41 and 38 percent better than the classic ANN model, respectively.

5. Conclusion

Machine learning is used in a variety of sciences, including coastal engineering and management. Surveying the wave's height nearshore and ports is critical for achieving sustainable development. The performance of soft computing models, including the classic ANN and M5MT versus hybrid models of WANN and Wavelet-M5 models is evaluated in the Hambantota Port, Sri Lanka, and Sanmen Bay, China. In this study, the wind and wave daily data from the AWAC sensor for Hambantota Port in 2020 and Sanmen Bay 2017 were used. Statistical indicators and scatter plots were used to compare the performance of these models. Wave evaluation was used to examine the characteristics of each model. Approximately 80% of the data was used to train and evaluate soft computational models, with the remainder used to validate how well the models performed after training in both case studies. The results show that hybrid methods incorporating wavelet decomposition improve simulation accuracy. Furthermore, the obtained results in both case studies demonstrated that the wavelet-M5 and WANN models outperformed the individual ANN and M5 models. Meanwhile, in the ANN method, problem information and knowledge are stored in a large set of coefficients and weights of connections between neurons, making it difficult to determine the relationship between input variables and the target parameter. Because the values of the statistical indicators of the models are so similar, it is suggested that the WANN and Wavelet-M5 hybrid models should be used to evaluate the wave height in the study area.

When the model is presented with more than two-time series inputs and when wavelet transformations are performed, the number of inputs significantly increases, which is regarded as a limitation. It should also be noted that this results in an increased number of inputs. This approach relies on data, which is a limitation. If field data are not available, we need simulation results to further realize the wave characteristic.

Data Availability

The data supporting the findings of this study are available upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] M. Zanganeh, "Improvement of the ANFIS-based wave predictor models by the particle swarm optimization," *Journal of Ocean Engineering and Science*, vol. 5, no. 1, pp. 84–99, 2020.

[2] W. Huang and S. Dong, "Improved short-term prediction of significant wave height by decomposing deterministic and stochastic components," *Renewable Energy*, vol. 177, pp. 743–758, 2021.

[3] W. Yong, J. Zhou, D. Jahed Armaghani et al., "A new hybrid simulated annealing-based genetic programming technique to predict the ultimate bearing capacity of piles," *Engineering with Computers*, vol. 37, no. 3, pp. 2111–2127, 2021.

[4] A. Mojtahedi, M. S. Beiragh, I. Farajpour, and M. Mohammadian, "Investigation on hydrodynamic performance of an environmentally friendly pile breakwater," *Ocean Engineering*, vol. 217, Article ID 107942, 2020.

[5] R. Tur and S. Yontem, "A comparison of soft computing methods for the prediction of wave height parameters," *Knowledge-Based Engineering and Sciences*, vol. 2, no. 1, pp. 31–46, 2021.

[6] R. Tür, "Maximum wave height hindcasting using ensemble linear-nonlinear models," *Theoretical and Applied Climatology*, vol. 141, no. 3–4, pp. 1151–1163, 2020.

[7] S. Yang, T. Xia, Z. Zhang et al., "Prediction of significant wave heights based on CS-BP model in the South China sea," *IEEE Access*, vol. 7, pp. 147490–147500, 2019.

[8] A. K. Ball, R. Das, S. S. Roy, D. R. Kisku, and N. C. Murmu, "Modeling of EHD inkjet printing performance using soft computing-based approaches," *Soft Computing*, vol. 24, no. 1, pp. 571–589, 2020.

[9] A. Sharafati, M. Haghbin, D. Motta, and Z. M. Yaseen, "The application of soft computing models and empirical formulations for hydraulic structure scouring depth simulation: a comprehensive review, assessment and possible future research direction," *Archives of Computational Methods in Engineering*, vol. 28, no. 2, pp. 423–447, 2021.

[10] F. Sayyahi, S. Farzin, and H. Karami, "Forecasting daily and monthly reference evapotranspiration in the aidoghmoush basin using multilayer perceptron coupled with water wave optimization," *Complexity*, vol. 2021, Article ID 6683759, 12 pages, 2021.

[11] M. Najafzadeh, J. Shiri, and M. Rezaie-Balf, "New expression-based models to estimate scour depth at clear water conditions in rectangular channels," *Marine Georesources and Geotechnology*, vol. 36, no. 2, pp. 227–235, 2018.

[12] M. Najafzadeh and G. Oliveto, "Scour propagation rates around offshore pipelines exposed to currents by applying data-driven models," *Water (Switzerland)*, vol. 14, no. 3, p. 493, 2022.

[13] M. Najafzadeh and F. Saberi-Movahed, "GMDH-GEP to predict free span expansion rates below pipelines under waves," *Marine Georesources and Geotechnology*, vol. 37, no. 3, pp. 375–392, 2019.

[14] M. Najafzadeh, A. Etemad-Shahidi, and S. Y. Lim, "Scour prediction in long contractions using ANFIS and SVM," *Ocean Engineering*, vol. 111, pp. 128–135, 2016.

[15] M. Banan-Dallalian, M. Shokatian-Beiragh, A. Golshani, and A. Abdi, "Use of a Bayesian Network for storm-induced flood risk assessment and effectiveness of ecosystem-based risk reduction measures in coastal areas (Port of Sur, Sultanate of Oman)," *Ocean Engineering*, vol. 270, Article ID 113662, 2023.

[16] H. A. Yonesi, A. Parsaie, A. Arshia, and Z. Shamsi, "Discharge modeling in compound channels with non-prismatic floodplains using GMDH and MARS models," *Water Supply*, vol. 22, no. 4, pp. 4400–4421, 2022.

[17] M. Najafzadeh, "Neuro-fuzzy GMDH systems based evolutionary algorithms to predict scour pile groups in clear water conditions," *Ocean Engineering*, vol. 99, pp. 85–94, 2015.

[18] M. Najafzadeh, F. Saberi-Movahed, and S. Sarkamaryan, "NF-GMDH-Based self-organized systems to predict bridge pier scour depth under debris flow effects," *Marine Georesources and Geotechnology*, vol. 36, no. 5, pp. 589–602, 2018.

[19] M. Najafzadeh, G. A. Barani, and M. R. Hessami-Kermani, "Evaluation of GMDH networks for prediction of local scour depth at bridge abutments in coarse sediments with thinly armored beds," *Ocean Engineering*, vol. 104, pp. 387–396, 2015.

[20] A. Parsaie, A. H. Haghiabi, S. D. Latif, and R. P. Tripathi, "Predictive modelling of piezometric head and seepage discharge in earth dam using soft computational models," *Environmental Science and Pollution Research*, vol. 28, no. 43, pp. 60842–60856, 2021.

[21] A. Parsaie and A. H. Haghiabi, "Mathematical expression for discharge coefficient of Weir-Gate using soft computing techniques," *Journal of Applied Water Engineering and Research*, vol. 9, no. 3, pp. 175–183, 2020.

[22] M. Banan-Dallalian, M. Shokatian-Beiragh, A. Golshani, A. Mojtahedi, and M. A. Lotfollahi-Yaghin, "Study of the effect of an environmentally friendly flood risk reduction approach on the oman coastlines during the gonu tropical cyclone (case study: the coastline of sur)," in *Proceedings of the 2nd International Conference on Oceanography for West Asia (RCOWA 2020)*, p. 11, Tehran, Iran, September 2020.

[23] M. Banan-Dallalian, M. Shokatian-Beiragh, A. Golshani, A. Mojtahedi, M. A. Lotfollahi-Yaghin, and S. Akib, "Study of the effect of an environmentally friendly flood risk reduction approach on the Oman coastlines during the gonu tropical cyclone (case study: the coastline of sur)," *Eng*, vol. 2, no. 2, pp. 141–155, 2021.

[24] M. Samadi, H. Sarkardeh, and E. Jabbari, "Prediction of the dynamic pressure distribution in hydraulic structures using

soft computing methods," *Soft Computing*, vol. 25, no. 5, pp. 3873–3888, 2021.

[25] M. Umair, M. A. Hashmani, and M. H. B. Hasan, "Survey of sea wave parameters classification and prediction using machine learning models," in *Proceedings of the 2019 1st International Conference on Artificial Intelligence and Data Sciences*, pp. 1–6, AiDAS, Ipoh, Malaysia, September 2019.

[26] Y. Zhou, Q. Ye, W. Shi, B. Yang, Z. Song, and D. Yan, "Wave characteristics in the nearshore waters of Sanmen bay," *Applied Ocean Research*, vol. 101, Article ID 102236, 2020.

[27] S. Gracia, J. Olivito, J. Resano, B. Martin-del-Brio, M. de Alfonso, and E. Álvarez, "Improving accuracy on wave height estimation through machine learning techniques," *Ocean Engineering*, vol. 236, Article ID 108699, 2021.

[28] M. Parvan, A. R. Ghiasi, T. Y. Rezaii, and A. Farzamnia, "Transfer learning based motor imagery classification using convolutional neural networks," in *Proceedings of the ICEE 2019 - 27th Iranian Conference on Electrical Engineering*, pp. 1825–1828, Yazd, Iran, April 2019.

[29] M. Samadi, M. H. Afshar, E. Jabbari, and H. Sarkardeh, "Prediction of current-induced scour depth around pile groups using MARS, CART, and ANN approaches," *Marine Georesources and Geotechnology*, vol. 39, no. 5, pp. 577–588, 2021.

[30] M. R. Nikoo, R. Kerachian, and M. R. Alizadeh, "A fuzzy KNN-based model for significant wave height prediction in large lakes," *Oceanologia*, vol. 60, no. 2, pp. 153–168, 2018.

[31] V. Nourani, A. Molajou, H. Najafi, and A. Danandeh Mehr, "Emotional ANN (eann): a new generation of neural networks for hydrological modeling in IoT," *Transactions on Computational Science and Computational Intelligence*, Springer, Cham, Switzerland, 2019.

[32] G. Kuntoji, M. Rao, and S. Rao, "Prediction of wave transmission over submerged reef of tandem breakwater using PSO-SVM and PSO-ANN techniques," *ISH Journal of Hydraulic Engineering*, vol. 26, no. 3, pp. 283–290, 2020.

[33] A. Molajou, V. Nourani, A. Afshar, M. Khosravi, and A. Brysiewicz, "Optimal design and feature selection by genetic algorithm for emotional artificial neural network (EANN) in rainfall-runoff modeling," *Water Resources Management*, vol. 35, no. 8, pp. 2369–2384, 2021.

[34] E. Androulakis and G. Galanis, "A two-step hybrid system towards optimized wave height forecasts," *Stochastic Environmental Research and Risk Assessment*, vol. 36, no. 3, pp. 753–766, 2022.

[35] S. Kundapura, V. H. Arkal, and J. L. S. Pinho, "Below the data range prediction of soft computing wave reflection of semicircular breakwater," *Journal of Marine Science and Application*, vol. 18, no. 2, pp. 167–175, 2019.

[36] M. Koopialipoor, D. Jahed Armaghani, A. Hedayat, A. Marto, and B. Gordan, "Applying various hybrid intelligent systems to evaluate and predict slope stability under static and dynamic conditions," *Soft Computing*, vol. 23, no. 14, pp. 5913–5929, 2019.

[37] M. R. Kaloop, D. Kumar, F. Zarzoura, B. Roy, and J. W. Hu, "A wavelet - particle swarm optimization - extreme learning machine hybrid modeling for significant wave height prediction," *Ocean Engineering*, vol. 213, Article ID 107777, 2020.

[38] S. Akbarifard and F. Radmanesh, "Predicting sea wave height using Symbiotic Organisms Search (SOS) algorithm," *Ocean Engineering*, vol. 167, pp. 348–356, 2018.

[39] A. H. Zaji, H. Bonakdari, and S. Shamshirband, "Standard equations for predicting the discharge coefficient of a modified high-performance side weir," *Scientia Iranica*, vol. 8, no. 2, p. 1069, 2017.

[40] V. Nourani, A. Davanlou Tajbakhsh, A. Molajou, and H. Gokcekus, "Hybrid wavelet-M5 model tree for rainfall-runoff modeling," *Journal of Hydrologic Engineering*, vol. 24, no. 5, Article ID 04019012, 2019.

[41] V. Nourani, A. Molajou, A. D. Tajbakhsh, and H. Najafi, "A wavelet based data mining technique for suspended sediment load modeling," *Water Resources Management*, vol. 33, no. 5, pp. 1769–1784, 2019.

[42] M. Gangappa, C. Kiran Mai, and P. Sammulal, "Classification of land cover images using modified water wave Optimization-based hybrid classifier," *International Journal of Computers and Applications*, vol. 43, no. 10, pp. 1–11, 2019.

[43] I. Ebtehaj, H. Bonakdari, S. Shamshirband, and K. Mohammadi, "A combined support vector machine-wavelet transform model for prediction of sediment transport in sewer," *Flow Measurement and Instrumentation*, vol. 47, pp. 19–27, 2016.

[44] H. Karami, S. Karimi, H. Bonakdari, and S. Shamshirband, "Predicting discharge coefficient of triangular labyrinth weir using extreme learning machine," *Artificial Neural Network and Genetic Programming*, vol. 29, pp. 983–989, 2018.