

Recent Advances in Data Mining Approaches in Civil Engineering for Water-related Engineering Problems

Lead Guest Editor: Mohammad Najafzadeh

Guest Editors: Amir Molajou and Ata Akbari Asanjan





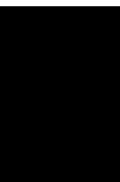
**Recent Advances in Data Mining Approaches
in Civil Engineering for Water-related
Engineering Problems**

Advances in Civil Engineering

**Recent Advances in Data Mining
Approaches in Civil Engineering for
Water-related Engineering Problems**

Lead Guest Editor: Mohammad Najafzadeh

Guest Editors: Amir Molajou and Ata Akbari
Asanjan



Copyright © 2023 Hindawi Limited. All rights reserved.

This is a special issue published in "Advances in Civil Engineering." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.






Chief Editor

Cumaraswamy Vipulanandan, USA














Associate Editors

Chiara Bedon , Italy
Constantin Chaliotis , Greece
Ghassan Chehab , Lebanon
Ottavia Corbi, Italy
Mohamed ElGawady , USA
Husnain Haider , Saudi Arabia
Jian Ji , China
Jiang Jin , China
Shazim A. Memon , Kazakhstan
Hossein Moayedi , Vietnam
Sanjay Nimbalkar, Australia
Giuseppe Oliveto , Italy
Alessandro Palmeri , United Kingdom
Arnaud Perrot , France
Hugo Rodrigues , Portugal
Victor Yepes , Spain
Xianbo Zhao , Australia

Academic Editors

José A.F.O. Correia, Portugal
Glenda Abate, Italy
Khalid Abdel-Rahman , Germany
Ali Mardani Aghabaglou, Turkey
José Aguiar , Portugal
Afaq Ahmad , Pakistan
Muhammad Riaz Ahmad , Hong Kong
Hashim M.N. Al-Madani , Bahrain
Luigi Aldieri , Italy
Angelo Aloisio , Italy
Maria Cruz Alonso, Spain
Filipe Amarante dos Santos , Portugal
Serji N. Amirkhanean, USA
Eleftherios K. Anastasiou , Greece
Panagiotis Ch. Anastasopoulos , USA
Mohamed Moafak Arbili , Iraq
Farhad Aslani , Australia
Siva Avudaiappan , Chile
Ozgur BASKAN , Turkey
Adewumi Babafemi, Nigeria
Morteza Bagherpour, Turkey
Qingsheng Bai , Germany
Nicola Baldo , Italy
Daniele Baraldi , Italy

Eva Barreira , Portugal
Emilio Bastidas-Arteaga , France
Rita Bento, Portugal
Rafael Bergillos , Spain
Han-bing Bian , China
Xia Bian , China
Huseyin Bilgin , Albania
Giovanni Biondi , Italy
Hugo C. Biscaia , Portugal
Rahul Biswas , India
Edén Bojórquez , Mexico
Giosuè Boscato , Italy
Melina Bosco , Italy
Jorge Branco , Portugal
Bruno Briseghella , China
Brian M. Broderick, Ireland
Emanuele Brunesi , Italy
Quoc-Bao Bui , Vietnam
Tan-Trung Bui , France
Nicola Buratti, Italy
Gaochuang Cai, France
Gladis Camarini , Brazil
Alberto Campisano , Italy
Qi Cao, China
Qixin Cao, China
Iacopo Carnacina , Italy
Alessio Cascardi, Italy
Paolo Castaldo , Italy
Nicola Cavalagli , Italy
Liborio Cavaleri , Italy
Anush Chandrappa , United Kingdom
Wen-Shao Chang , United Kingdom
Muhammad Tariq Amin Chaudhary, Kuwait
Po-Han Chen , Taiwan
Qian Chen , China
Wei Tong Chen , Taiwan
Qixiu Cheng, Hong Kong
Zhanbo Cheng, United Kingdom
Nicholas Chileshe, Australia
Prinya Chindaprasirt , Thailand
Corrado Chisari , United Kingdom
Se Jin Choi , Republic of Korea
Heap-Yih Chong , Australia
S.H. Chu , USA
Ting-Xiang Chu , China

Zhaofei Chu , China
Wonseok Chung , Republic of Korea
Donato Ciampa , Italy
Gian Paolo Cimellaro, Italy
Francesco Colangelo, Italy
Romulus Costache , Romania
Liviu-Adrian Cotfas , Romania
Antonio Maria D'Altri, Italy
Bruno Dal Lago , Italy
Amos Darko , Hong Kong
Arka Jyoti Das , India
Dario De Domenico , Italy
Gianmarco De Felice , Italy
Stefano De Miranda , Italy
Maria T. De Risi , Italy
Tayfun Dede, Turkey
Sadik O. Degertekin , Turkey
Camelia Delcea , Romania
Cristoforo Demartino, China
Giuseppe Di Filippo , Italy
Luigi Di Sarno, Italy
Fabio Di Trapani , Italy
Aboelkasim Diab , Egypt
Thi My Dung Do, Vietnam
Giulio Dondi , Italy
Jiangfeng Dong , China
Chao Dou , China
Mario D'Aniello , Italy
Jingtao Du , China
Ahmed Elghazouli, United Kingdom
Francesco Fabbrocino , Italy
Flora Faleschini , Italy
Dingqiang Fan, Hong Kong
Xueping Fan, China
Qian Fang , China
Salar Farahmand-Tabar , Iran
Ilenia Farina, Italy
Roberto Fedele, Italy
Guang-Liang Feng , China
Luigi Fenu , Italy
Tiago Ferreira , Portugal
Marco Filippo Ferrotto, Italy
Antonio Formisano , Italy
Guoyang Fu, Australia
Stefano Galassi , Italy

Junfeng Gao , China
Meng Gao , China
Giovanni Garcea , Italy
Enrique García-Macías, Spain
Emilio García-Taengua , United Kingdom
DongDong Ge , USA
Khaled Ghaedi, Malaysia
Khaled Ghaedi , Malaysia
Gian Felice Giaccu, Italy
Agathoklis Giaralis , United Kingdom
Ravindran Gobinath, India
Rodrigo Gonçalves, Portugal
Peilin Gong , China
Belén González-Fonteboa , Spain
Salvatore Grasso , Italy
Fan Gu, USA
Erhan Güneyisi , Turkey
Esra Mete Güneyisi, Turkey
Pingye Guo , China
Ankit Gupta , India
Federico Gusella , Italy
Kemal Hacıfendioglu, Turkey
Jianyong Han , China
Song Han , China
Asad Hanif , Macau
Hadi Hasanzadehshooiili , Canada
Mostafa Fahmi Hassanein, Egypt
Amir Ahmad Hedayat , Iran
Khandaker Hossain , Canada
Zahid Hossain , USA
Chao Hou, China
Biao Hu, China
Jiang Hu , China
Xiaodong Hu, China
Lei Huang , China
Cun Hui , China
Bon-Gang Hwang, Singapore
Jijo James , India
Abbas Fadhil Jasim , Iraq
Ahad Javanmardi , China
Krishnan Prabhakan Jaya, India
Dong-Sheng Jeng , Australia
Han-Yong Jeon, Republic of Korea
Pengjiao Jia, China
Shaohua Jiang , China

MOUSTAFA KASSEM , Malaysia
Mosbeh Kaloop , Egypt
Shankar Karuppannan , Ethiopia
John Kechagias , Greece
Mohammad Khajehzadeh , Iran
Afzal Husain Khan , Saudi Arabia
Mehran Khan , Hong Kong
Manoj Khandelwal, Australia
Jin Kook Kim , Republic of Korea
Woosuk Kim , Republic of Korea
Vaclav Koci , Czech Republic
Loke Kok Foong, Vietnam
Hailing Kong , China
Leonidas Alexandros Kouris , Greece
Kyriakos Kourousis , Ireland
Moacir Kripka , Brazil
Anupam Kumar, The Netherlands
Emma La Malfa Ribolla, Czech Republic
Ali Lakirouhani , Iran
Angus C. C. Lam, China
Thanh Quang Khai Lam , Vietnam
Luciano Lamberti, Italy
Andreas Lampropoulos , United Kingdom
Raffaele Landolfo, Italy
Massimo Latour , Italy
Bang Yeon Lee , Republic of Korea
Eul-Bum Lee , Republic of Korea
Zhen Lei , Canada
Leonardo Leonetti , Italy
Chun-Qing Li , Australia
Dongsheng Li , China
Gen Li, China
Jiale Li , China
Minghui Li, China
Qingchao Li , China
Shuang Yang Li , China
Sunwei Li , Hong Kong
Yajun Li , China
Shun Liang , China
Francesco Liguori , Italy
Jae-Han Lim , Republic of Korea
Jia-Rui Lin , China
Kun Lin , China
Shibin Lin, China

Tzu-Kang Lin , Taiwan
Yu-Cheng Lin , Taiwan
Hexu Liu, USA
Jian Lin Liu , China
Xiaoli Liu , China
Xuemei Liu , Australia
Zaobao Liu , China
Zhuang-Zhuang Liu, China
Diego Lopez-Garcia , Chile
Cristiano Loss , Canada
Lyan-Ywan Lu , Taiwan
Jin Luo , USA
Yanbin Luo , China
Jianjun Ma , China
Junwei Ma , China
Tian-Shou Ma, China
Zhongguo John Ma , USA
Maria Macchiaroli, Italy
Domenico Magisano, Italy
Reza Mahinroosta, Australia
Yann Malecot , France
Prabhat Kumar Mandal , India
John Mander, USA
Iman Mansouri, Iran
André Dias Martins, Portugal
Domagoj Matesan , Croatia
Jose Matos, Portugal
Vasant Matsagar , India
Claudio Mazzotti , Italy
Ahmed Mebarki , France
Gang Mei , China
Kasim Mermerdas, Turkey
Giovanni Minafò , Italy
Masoomah Mirrashid , Iran
Abbas Mohajerani , Australia
Fadzli Mohamed Nazri , Malaysia
Fabrizio Mollaioli , Italy
Rosario Montuori , Italy
H. Naderpour , Iran
Hassan Nasir , Pakistan
Hossein Nassiraei , Iran
Satheeskumar Navaratnam , Australia
Ignacio J. Navarro , Spain
Ashish Kumar Nayak , India
Behzad Nematollahi , Australia

Chayut Ngamkhanong , Thailand
Trung Ngo, Australia
Tengfei Nian, China
Mehdi Nikoo , Canada
Youjun Ning , China
Olugbenga Timo Oladinrin , United Kingdom
Oladimeji Benedict Olalusi, South Africa
Timothy O. Olawumi , Hong Kong
Alejandro Orfila , Spain
Maurizio Orlando , Italy
Siti Aminah Osman, Malaysia
Walid Oueslati , Tunisia
SUVASH PAUL , Bangladesh
John-Paris Pantouvakis , Greece
Fabrizio Paolacci , Italy
Giuseppina Pappalardo , Italy
Fulvio Parisi , Italy
Dimitrios G. Pavlou , Norway
Daniele Pellegrini , Italy
Gatheeshgar Perampalam , United Kingdom
Daniele Perrone , Italy
Giuseppe Piccardo , Italy
Vagelis Plevris , Qatar
Andrea Pranno , Italy
Adolfo Preciado , Mexico
Chongchong Qi , China
Yu Qian, USA
Ying Qin , China
Giuseppe Quaranta , Italy
Krishanu ROY , New Zealand
Vlastimir Radonjanin, Serbia
Carlo Rainieri , Italy
Rahul V. Ralegaonkar, India
Raizal Saifulnaz Muhammad Rashid, Malaysia
Alessandro Rasulo , Italy
Chonghong Ren , China
Qing-Xin Ren, China
Dimitris Rizos , USA
Geoffrey W. Rodgers , New Zealand
Pier Paolo Rossi, Italy
Nicola Ruggieri , Italy
JUNLONG SHANG, Singapore


Nikhil Saboo, India
Anna Saetta, Italy
Juan Sagaseta , United Kingdom
Timo Saksala, Finland
Mostafa Salari, Canada
Ginevra Salerno , Italy
Evangelos J. Sapountzakis , Greece
Vassilis Sarhosis , United Kingdom
Navaratnarajah Sathiparan , Sri Lanka
Fabrizio Scozzese , Italy
Halil Sezen , USA
Payam Shafigh , Malaysia
M. Shahria Alam, Canada
Yi Shan, China
Hussein Sharaf, Iraq
Mostafa Sharifzadeh, Australia
Sanjay Kumar Shukla, Australia
Amir Si Larbi , France
Okan Sirin , Qatar
Piotr Smarzewski , Poland
Francesca Sollecito , Italy
Rui Song , China
Tian-Yi Song, Australia
Flavio Stochino , Italy
Mayank Sukhija , USA
Piti Sukontasukkul , Thailand
Jianping Sun, Singapore
Xiao Sun , China
T. Tafsirojjan , Australia
Fujiao Tang , China
Patrick W.C. Tang , Australia
Zhi Cheng Tang , China
Weerachart Tangchirapat , Thailand
Xiabin Tao, China
Piergiorgio Tataranni , Italy
Elisabete Teixeira , Portugal
Jorge Iván Tobón , Colombia
Jing-Zhong Tong, China
Francesco Trentadue , Italy
Antonello Troncone, Italy
Majbah Uddin , USA
Tariq Umar , United Kingdom
Muahmmad Usman, United Kingdom
Muhammad Usman , Pakistan
Mucteba Uysal , Turkey

Ilaria Venanzi , Italy
Castorina S. Vieira , Portugal
Valeria Vignali , Italy
Claudia Vitone , Italy
Liwei WEN , China
Chunfeng Wan , China
Hua-Ping Wan, China
Roman Wan-Wendner , Austria
Chaohui Wang , China
Hao Wang , USA
Shiming Wang , China
Wayne Yu Wang , United Kingdom
Wen-Da Wang, China
Xing Wang , China
Xiuling Wang , China
Zhenjun Wang , China
Xin-Jiang Wei , China
Tao Wen , China
Weiping Wen , China
Lei Weng , China
Chao Wu , United Kingdom
Jiangyu Wu, China
Wangjie Wu , China
Wenbing Wu , China
Zhixing Xiao, China
Gang Xu, China
Jian Xu , China
Panpan , China
Rongchao Xu , China
HE YONGLIANG, China
Michael Yam, Hong Kong
Hailu Yang , China
Xu-Xu Yang , China
Hui Yao , China
Xinyu Ye , China
Zhoujing Ye, China
Gürol Yildirim , Turkey
Dawei Yin , China
Doo-Yeol Yoo , Republic of Korea
Zhanping You , USA
Afshar A. Yousefi , Iran
Xinbao Yu , USA
Dongdong Yuan , China
Geun Y. Yun , Republic of Korea

Hyun-Do Yun , Republic of Korea
Cemal YİĞİT , Turkey
Paolo Zampieri, Italy
Giulio Zani , Italy
Mariano Angelo Zanini , Italy
Zhixiong Zeng , Hong Kong
Mustafa Zeybek, Turkey
Henglong Zhang , China
Jiupeng Zhang, China
Tingting Zhang , China
Zengping Zhang, China
Zetian Zhang , China
Zhigang Zhang , China
Zhipeng Zhao , Japan
Jun Zhao , China
Annan Zhou , Australia
Jia-wen Zhou , China
Hai-Tao Zhu , China
Peng Zhu , China
QuanJie Zhu , China
Wenjun Zhu , China
Marco Zucca, Italy
Haoran Zuo, Australia
Junqing Zuo , China
Robert Černý , Czech Republic
Süleyman İpek , Turkey

Contents

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

Tzu-Chia Chen , Zryan Najat Rashid , Biju Theruvil Sayed , Arif Sari, Ahmed Kateb Jumaah Al-Nussairi, Majid Samiee-Zenoozian, and Mehrdad Shokatian-Beiragh 

Research Article (13 pages), Article ID 8272566, Volume 2023 (2023)





Prediction of the Coefficient of Pressure Fluctuations during the Hydraulic Jump Using ELM, GMDH, and M5MT

Tzu-Chia Chen , Biju Theruvil Sayed , Maria Jade Catalan Opulencia , Raed H. C. Alfilh , Maki Mahdi Abdulhasan , and Sayed Hashmat Sadat 

Research Article (14 pages), Article ID 2495631, Volume 2022 (2022)

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

¹College of Management and Design, Ming Chi University of Technology, New Taipei City, Taiwan, China

²Technical College of Informatics, Sulaimani Polytechnic University, Sulaymaniyah, Iraq

³Department of Computer Science, Dhofar University, Salalah, Oman

⁴Department of Management Information Systems, Girne American University, Kyrenia, North Cyprus, via Mersin 10, Turkey

⁵Al-Manara College for Medical Sciences, Maysan, Iraq

⁶Department 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

Copyright © 2023 Tzu-Chia Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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 on-shore 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 Sverdrup 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 $1,072$ km². 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:

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

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).

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

$f_p = 22[g^2/(U_{10} F)]^{1/3}$, 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).

$$H_s = 4\sqrt{m_0}, m_0 = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \eta^2 dt, \quad (3)$$

$t_2 - t_1$ 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.

$$H_x(X, Y) = 4\sqrt{m_x(X, Y)},$$

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

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:

$$Z = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad (5)$$

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:

$$net_j = \sum W_{j,i} x_i + b_j. \quad (6)$$

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

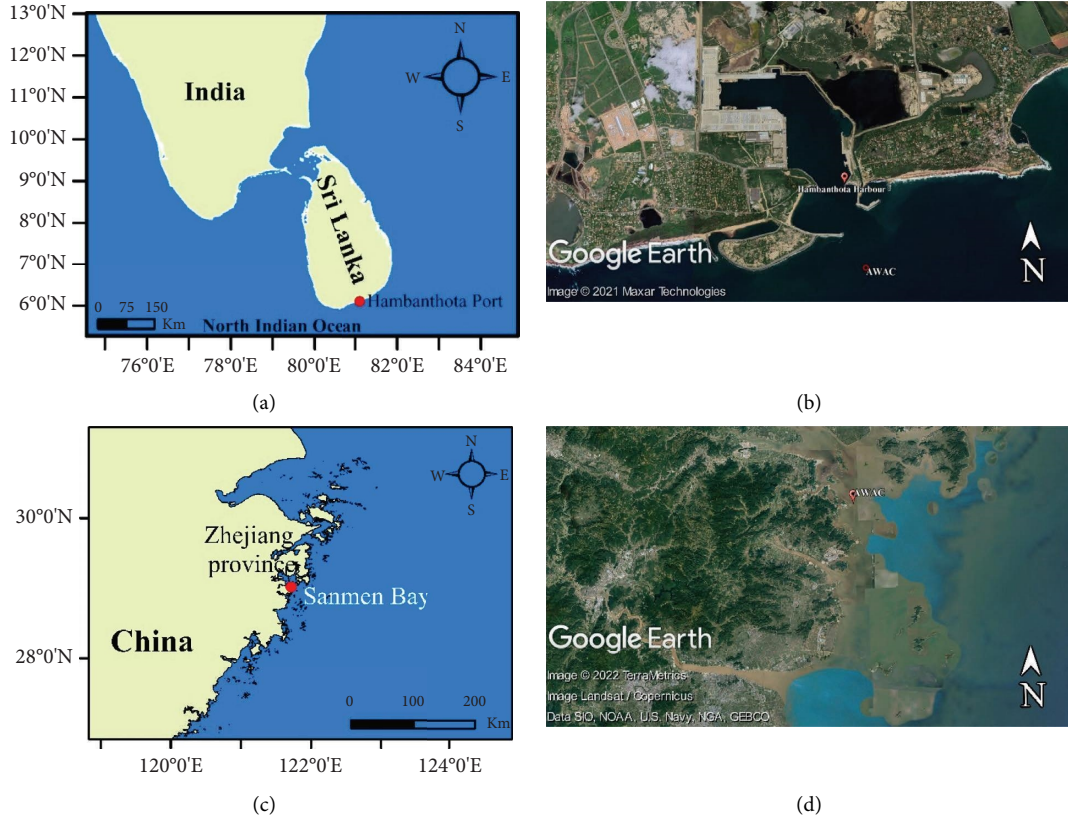


FIGURE 1: Study area. (a and b) Hambantota Port, Sri Lanka; (c and d) Sanmen Bay, China.

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:

$$\text{SDR} = \text{Sd}(T) - \sum_{i=1}^N \frac{|T_i|}{|T|} \text{Sd}(T_i), \quad (7)$$

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:

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

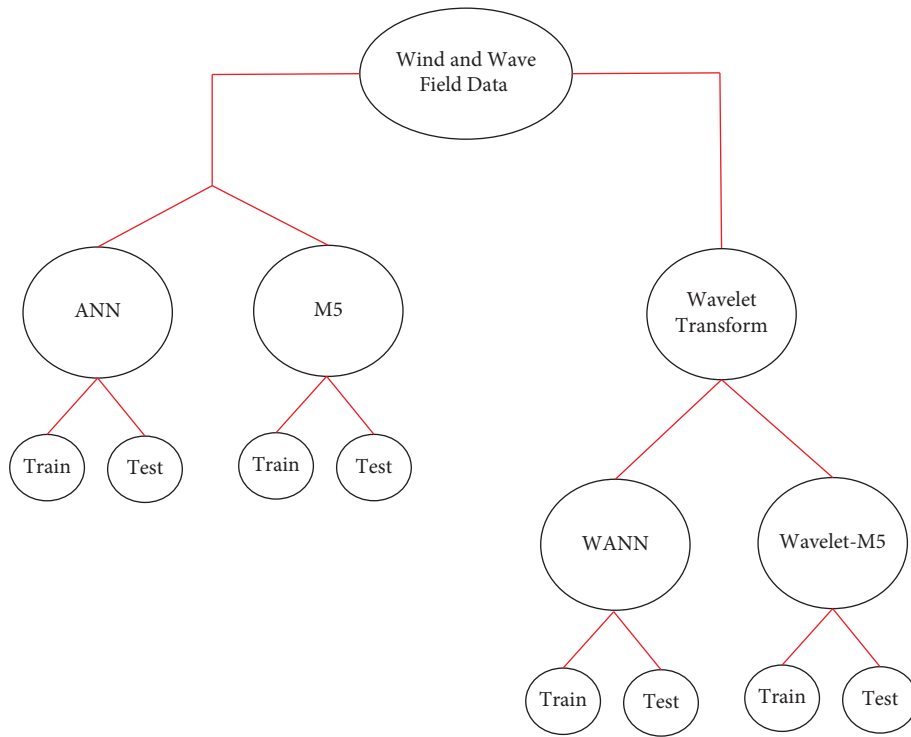


FIGURE 2: Data analysis framework.

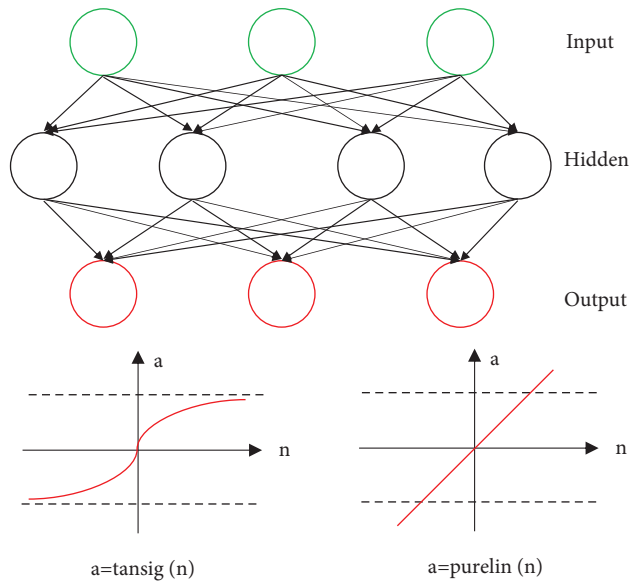


FIGURE 3: ANN structure schemes.

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:

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

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 $(a < 1)$,

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

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

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:

speed 10 hours earlier. The following equation is used to simulate and estimate the height of waves in Hambantota Port.

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

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. \bar{X} and \bar{Y} are average observational and predicted values, respectively. The performance of models was evaluated using the error indices.

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

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

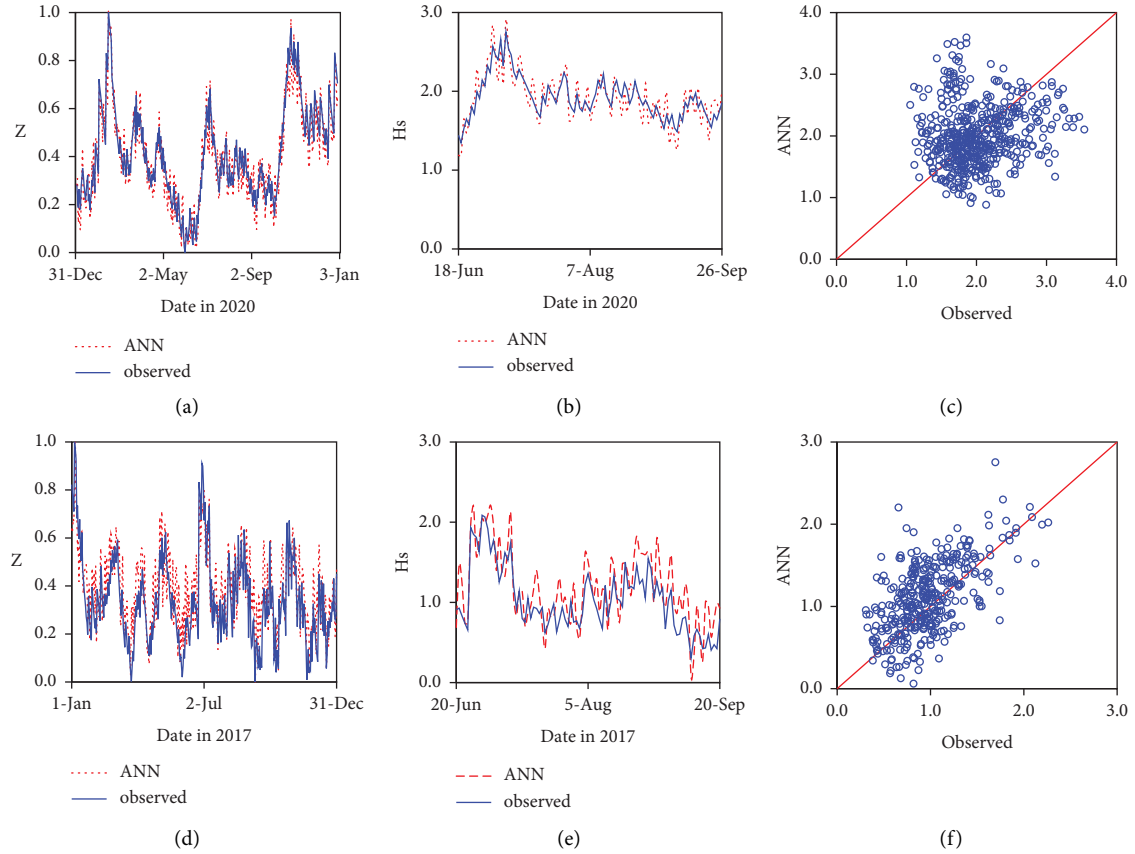


FIGURE 4: Times series of test and train normalized significant wave height and the corresponding scatter diagram by the classic ANN model (a–c) in the Hambantota Port (2020) and (d–f) in the Sanmen Bay (2017).

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

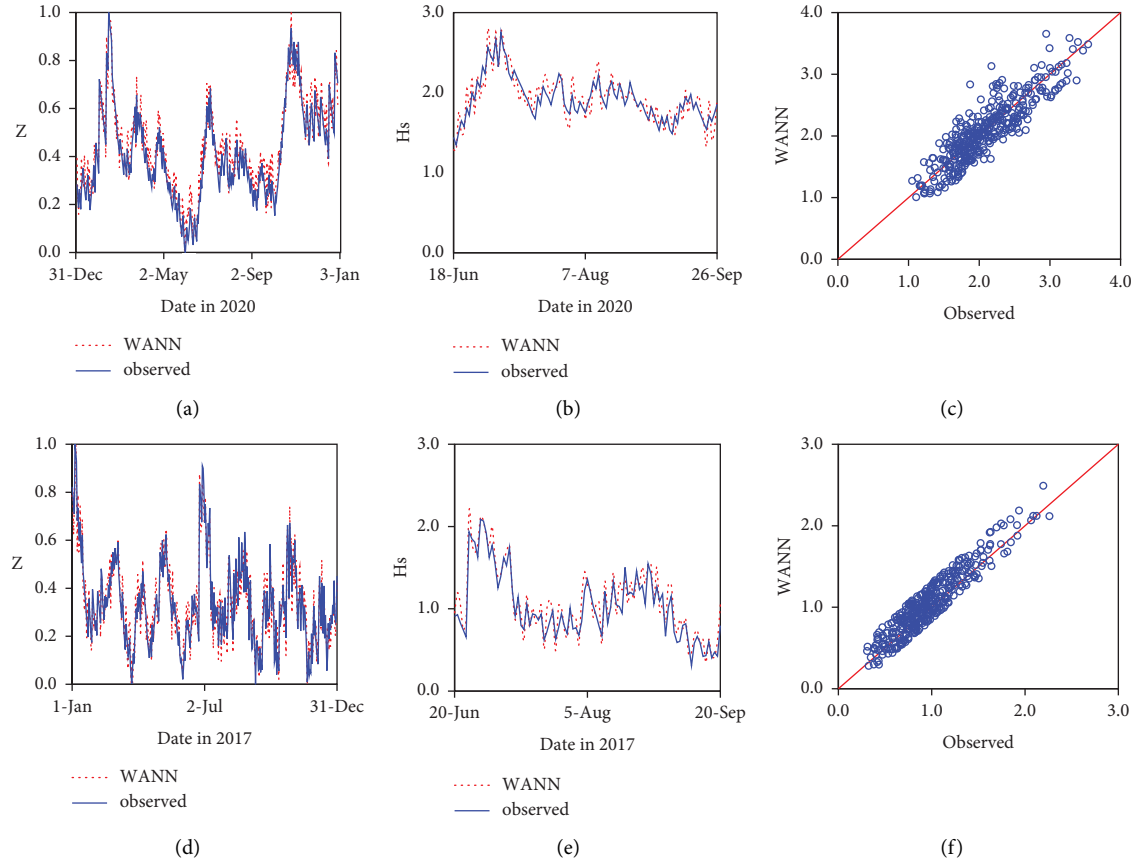


FIGURE 5: Times series of test and train normalized significant wave height and the corresponding scatter diagram by the WANN model (a–c) in the Hambantota Port (2020) and (d–f) in the Sanmen Bay (2017).

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:

$$N^H \leq 2N^I + 1, \quad (17)$$

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 Hambantota Port 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

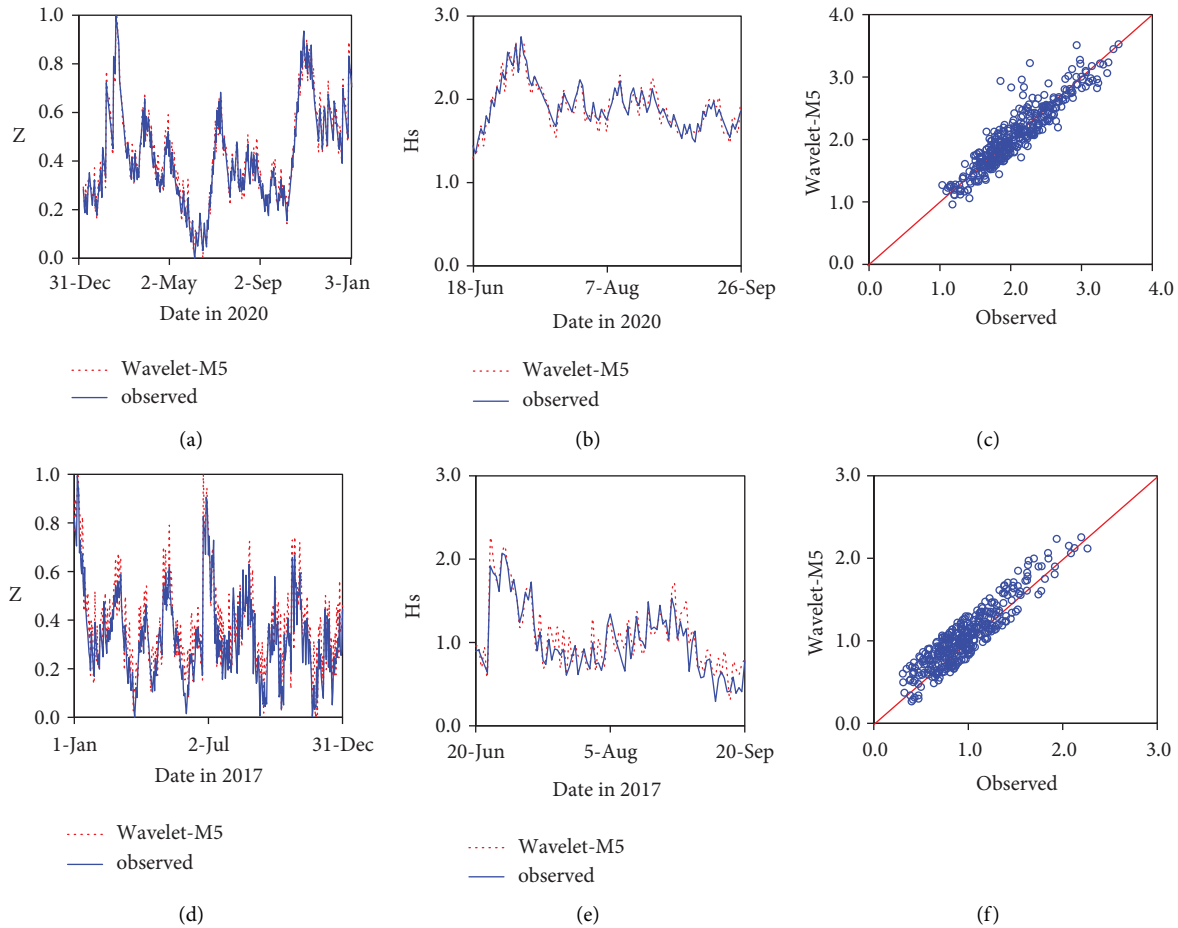


FIGURE 8: Times series of test and train normalized significant wave height and the corresponding scatter diagram by the wavelet- M5 model (a–c) in the Hambantota Port (2020) and (d–f) in the Sanmen Bay (2017).

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 aidoghmouth 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. Haghbi, 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. Haghbi, "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. Farzamnna, “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 semi-circular 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. Shamsirband, “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. Sammual, “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. Shamsirband, 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. Shamsirband, “Predicting discharge coefficient of triangular labyrinth weir using extreme learning machine,” *Artificial Neural Network and Genetic Programming*, vol. 29, pp. 983–989, 2018.

Research Article

Prediction of the Coefficient of Pressure Fluctuations during the Hydraulic Jump Using ELM, GMDH, and M5MT

Tzu-Chia Chen ¹, Biju Theruvil Sayed ², Maria Jade Catalan Opulencia ³,
Raed H. C. Alfilh ⁴, Maki Mahdi Abdulhasan ⁵ and Sayed Hashmat Sadat ⁶

¹Department of Industrial Engineering and Management, Ming Chi University of Technology, New Taipei City, Taiwan

²Department of Computer Science Dhofar University, Salalah, Oman

³College of Business Administration Ajman University, Ajman, UAE

⁴Refrigeration and Air-conditioning Technical Engineering Department, College of Technical Engineering Islamic University, Najaf, Iraq

⁵Al-Nisour University College, Baghdad, Iraq

⁶Kabul University, Kabul, Afghanistan

Correspondence should be addressed to Sayed Hashmat Sadat; s.h.sadat@ku.edu.af

Received 23 June 2022; Accepted 28 August 2022; Published 23 September 2022

Academic Editor: Mohammad Najafzadeh

Copyright © 2022 Tzu-Chia Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pressure fluctuations are a critical phenomenon that can endanger the safety and stability of hydraulic structures, especially stilling basins. Hence, the accurate estimation of the dimensionless coefficient of pressure fluctuations (C_p') is critical for hydraulic engineers. This study proposed predictive soft computing models to estimate C_p' on sloping channels. Therefore, three robust soft computing methods, including extreme learning machine (ELM), group method data of handling (GMDH), and M5 model tree (M5MT), were used to estimate C_p' . The results revealed that ELM was more accurate than GMDH and M5MT methods when comparing statistical indices, including correlation coefficient (CC), root mean square error (RMSE), mean absolute error (MAE), scatter index (SI), index agreement (I_a), and BIAS values. The performance of ELM was found to be more accurate (CC = 0.9183, RMSE = 0.0067, MAE = 0.0051, SI = 11.88%, I_a = 0.9569) when compared with the results of GMDH (CC = 0.8818, RMSE = 0.0078, MAE = 0.0058, SI = 13.89%, I_a = 0.9361) and M5MT (CC = 0.6883, RMSE = 0.0120, MAE = 0.0090, SI = 21.28%, I_a = 0.7905) in the testing stage. In addition, the BIAS values revealed that ELM slightly overestimated the values of C_p' especially at the peak point compared with GMDH and M5MT results. Overall, the suggested soft computing techniques worked well for predicting pressure fluctuation changes in the hydraulic jump.

1. Introduction

Stilling basins are the most widely used dissipation hydraulic structures of large dams. Energy dissipation in stilling basins is related to hydraulic jumps with high turbulent flow. Hydraulic jump is the common phenomenon for flow energy dissipation in the stilling basins. This phenomenon transforms the supercritical flow into a subcritical flow at a short distance. This is also accompanied by large-scale turbulence, surface waves, air entrainment into the flow, an increase in flow depth, and considerable energy dissipation in the water flow [1, 2].

The turbulent flow within stilling basins is related to the movement of large-scale vortices and severe pressure fluctuations, which may cause significant damage in stilling basins through cavitation, erosion, lifting force, and material fatigue [3]. The importance of pressure fluctuations in hydraulic jumps was revealed after the destruction of the Karnafoli and Malpasos dams in Bangladesh and Mexico, respectively [4]. Therefore, pressure fluctuations in hydraulic structures have included a considerable volume of hydraulic engineers' investigations. Hydraulic models are standard tools for measurements of pressure fluctuations. Experimental studies were

carried out for the analysis of characteristics of pressure fluctuations at the flip buckets and stilling basins.

Soft computing methods are alternative tools compared with traditional regression approaches and analytical solutions for modeling complicated problems and systems [5]. Moreover, soft computing techniques are much more cost-effective and have less run time than experimental studies. It is noteworthy that soft computing models are newer tools than experimental studies. The capability of different soft computing approaches has been proven for modeling C_p' in the hydraulic jump [6–8]. These models were developed and assessed using experimental data. The suggested soft computing methods were more accurate than conventional regression schemes for predicting C_p' during the hydraulic jump. Moreover, Samadi et al. [9] modeled dynamic pressure distribution in flip bucket spillways using a variety of soft computing approaches and provided simple, high-precision mathematical expressions for estimating it.

According to the literature review's assessment, there were few investigations into the performance of soft computing techniques for estimating pressure fluctuations in hydraulic structures. In addition, the applications of ELM, GMDH, and M5MT methods have not yet been reported in estimating pressure fluctuations. However, the results mentioned above show that soft computing methods are a desirable alternative to model pressure fluctuations in hydraulic structures.

The ELM algorithm is an advanced method that is constructed on a single hidden layer feed-forward neural network (SFLN) [10]. The ELM improves training time and accuracy by transforming training data into fixed-length batches and only updating the weight without retraining the trained samples. The ELM approach has been successfully reported for longitudinal dispersion coefficients in water pipelines [11], compressive strength concrete estimation [12], and predicting total dissolved gas [13].

The group method of data handling is a set of induction techniques that can be used to make mathematical modeling of multi-parametric datasets. GMDH algorithm uses an inductive approach to sort and order more complex polynomial models, with an external criterion selecting the best result. The application of GMDH has been widely and successfully carried out in scour depth estimation and water-related engineering problems [14–16].

M5MT is a common decision tree method that uses piecewise multiple linear regression equations to approximate nonlinear functions. The M5MT was applied to predict drought events [17], scour depth [15, 18], alga growth in reservoirs [19], the capacity of shallow foundations [20], and groundwater modeling [21].

The major goal of this study is to investigate the usefulness and capability of ELM, GMDH, and M5MT for predicting C_p' is taking place during the hydraulic jump on sloping channels. The author's knowledge indicated that ELM, GMDH, and M5MT applications had not been investigated yet for prediction C_p' . Therefore, the current

research in the first study evaluates the ability of ELM, GMDH, and M5MT to predict C_p' occurring in hydraulic jumps on sloping channels.

2. Materials and Methods

The data description and soft computing methods, including ELM, GMDH, and M5MT, are briefly explained in the following subsections.

2.1. Data Description. Data were collected from reliable and published experimental results of Gunal [22]. During the hydraulic jump phenomenon in the laboratory, pressure fluctuations occurred on the sloping channel. Gunal's [22] experiments were conducted in a 91 cm wide and 320 cm long rectangular flume. The inclined angles of the sloping channel were set to 10, 20, and 30 degrees. The cross-sectional representation of the hydraulic jump downstream of the sloping channel is displayed in Figure 1.

Güven [7] determined the relationship between $\sqrt{\overline{P'^2}}$ as root mean square value of the pressure fluctuations with characteristics of the flow and geometry of channels as follows:

$$\sqrt{\overline{P'^2}} = f(y_t, y_1, V_1, x, g, \rho_w, \theta), \quad (1)$$

where $\sqrt{\overline{P'^2}}$ is calculated from the following equation:

$$\sqrt{\overline{P'^2}} = \frac{1}{\sqrt{N}} \sqrt{\sum_{i=1}^N [P(x, y, n\Delta t) - \overline{P}(x, y)]^2}, \quad (2)$$

where $\overline{P}(x, y)$ is the mean pressure, $P(x, y, n\Delta t)$ is the value instantaneous of pressure, N is the number of data collected in a discrete-time series, and Δt is the sampling time step.

In (1), y_t is the depth of tail-water downstream of the stilling basin; y_1 is the gate opening of sluice gate; V_1 is the velocity issuing from the gate; x is the distance between the horizontal and inclined parts' intersection; g is the acceleration of gravity; ρ_w is water's mass density; and θ is the angle of the sloping section of the flume. Finally, the dimensional analysis provided the dimensionless functional form of (1) as follows [7]:

$$C_p' = f\left(Fr_1, \frac{x}{y_1}, \frac{y_t}{y_1}, \theta\right), \quad (3)$$

where $C_p' = \sqrt{\overline{P'^2}}/V_1^2/2g = RMS/V_1^2/2g$ and $Fr_1 = V_1/\sqrt{gy_1}$ are the dimensionless coefficient of pressure fluctuations and inflow Froude number, respectively. C_p' is an important factor used to describe pressure fluctuations.

2.2. Data Preparation for the Development of the Soft Computing Approaches. Overall, 112 values were obtained for C_p' under different flow conditions by recording pressure data. The statistical indices in the testing and training data sets should be similar; otherwise, we will be unable to test our models under certain hydrodynamic conditions that are

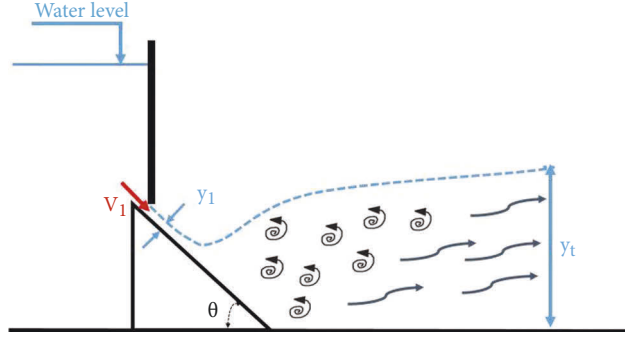


FIGURE 1: The cross-sectional representation of a hydraulic jump downstream of a sloping channel.

designated for the train set [23]. Therefore 80% of all data is randomly chosen for training, and the rest is used to assess the created models. The major statistical parameters of the testing and training data set are shown in Table 1. As seen in Table 1, the statistical parameters for testing and training subsets were close.

2.3. Extreme Learning Machine (ELM) and Model Development. Although the standard backpropagation neural network has several advantages, it also has some limitations, including tuning the parameter settings by selecting the number of hidden layers, momentum coefficient, and learning rate values. It suffers from the iterative learning process used to determine the weights, which takes an extended period. In addition, there is a possibility of a local minimum. By contrast, ELM provides it with the advantages of rapid convergence, fewer parameters to tweak, and a high degree of generalization [24].

The ELM comprises a single hidden layer that contains L nodes. The following equation is valid for N arbitrarily separate samples (x_i, t_i) , with an activation function $g(x)$, randomly distributed weights (w_i) , randomly distributed biases (b_i) , and output weights (β_i) .

$$\sum_{i=1}^L \beta_i g(w_i x_j + b_i) = O_j, j = 1, \dots, N, \quad (4)$$

where $\beta = [\beta_1, \dots, \beta_L]^T$ is a vector containing the output weights of the hidden layer of L nodes and the output node. It is possible to approximate N samples with zero error using typical SLFN with L hidden nodes and activation function $g(x)$. This means that

$$\sum_{j=1}^N \|O_j - t_j\| = 0. \quad (5)$$

In other words, there are β_i , w_i , and b_i such that

$$\sum_{i=1}^L \beta_i g(w_i x_j + b_i) = t_j, j = 1, \dots, N. \quad (6)$$

Let

$$H = \begin{bmatrix} g(w_1 x_1 + b_1) & \dots & g(w_L x_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(w_1 x_N + b_1) & \dots & g(w_L x_N + b_L) \end{bmatrix}, \quad (7)$$

$$\beta = [\beta_1^T \dots \beta_L^T]^T.$$

And

$$T = [t_1^T \dots t_L^T]^T. \quad (8)$$

Then,

$$H\beta = T. \quad (9)$$

H is referred to as the neural network's hidden layer output matrix. Huang et al. [24] prove $L \leq N$ hidden neurons as necessary for infinitely differentiable activation functions. H remains unchanged once the biases and weights of hidden nodes are established, and only β needs to be estimated. The least-squares solution with a minimum norm for (9) is as follows:

$$\min \|H(w_1, w_L, b_1, b_L)\beta - T\|. \quad (10)$$

H is square if the neurons in the hidden layer equal the training set (N), and β may be calculated by inverting H . Nevertheless, to obtain higher generalization, the number of hidden nodes is modified, and it may be less than N . Then, the Moore–Penrose generalized inverse of H must be utilized in this case.

$$\hat{\beta} = H^\dagger T, \quad (11)$$

where H^\dagger can be used to validate the algorithm because it is the Moore–Penrose generalized inverse of H . The ELM model that has been constructed has the following simple general form for the prediction of C_p' :

$$C_p' = \left[\frac{1}{(1 + \exp(-\text{In} W \times \text{In} V + \text{BHI}))} \right]^T \times \text{Out} W, \quad (12)$$

where $\text{In} V$ is the input variables, $\text{In} W$ is the input weight matrix, BHI is the hidden neuron vector's bias, and $\text{Out} W$ is the output weight vector. The following relations were obtained for the ELM model to calculate C_p' .

TABLE 1: The main statistical parameters of training and testing data set for developing the proposed soft computing methods.

Dataset	Variables	Parameter	Min	Max	Avg	St. dev.
Training	Input	Fr_1	4.944	8.662	6.712	1.473
		x/y_1	0.750	31.875	11.993	7.761
		y_t/y_1	6.750	11.111	8.992	1.721
	Output	θ	10	20	20.333	8.670
		C_p'	0.027	0.096	0.056	0.017
Testing	Input	Fr_1	4.944	8.662	6.325	1.357
		x/y_1	0.750	28.333	12.550	8.731
		y_t/y_1	6.750	11.111	8.509	1.613
	Output	θ	10	30	18.636	8.888
		C_p'	0.028	0.090	0.057	0.018

The variables of Fr_1 , x/y_1 , y_t/y_1 , and θ are considered as input variables of soft computing methods for the prediction of C_p' .

$$\begin{aligned}
 \text{InV} &= \begin{bmatrix} Fr_1 \\ \frac{y_t}{y_1} \\ \frac{x}{y_1} \\ \theta \end{bmatrix}, & \text{OutW} &= \begin{bmatrix} -0.05 \\ 0.03 \\ -0.01 \\ -0.09 \\ 0.02 \\ -0.02 \\ -0.06 \\ -0.01 \\ 0.04 \\ 0.01 \end{bmatrix}. & (13) \\
 \text{InW} &= \begin{bmatrix} -0.41 & -0.17 & 0.88 & -0.61 \\ -0.56 & -0.98 & -0.92 & 0.02 \\ 0.96 & -0.24 & -0.65 & -0.94 \\ -0.81 & -0.91 & -0.66 & -0.14 \\ 0.44 & 0.68 & 0.55 & 0.16 \\ 0.61 & -0.88 & -0.93 & -0.86 \\ -0.31 & -0.65 & -0.02 & 0.36 \\ 0.92 & 0.98 & 0.49 & 0.74 \\ 0.57 & -0.32 & 0.53 & -0.90 \\ -0.07 & 0.06 & -0.39 & -0.86 \end{bmatrix}, \\
 \text{BHI} &= \begin{bmatrix} -0.68 \\ 0.29 \\ -0.99 \\ 0.55 \\ 0.30 \\ -0.84 \\ 0.72 \\ -0.24 \\ -0.97 \\ -0.96 \end{bmatrix},
 \end{aligned}$$

2.4. Group Method of Data Handling (GMDH) and Model Development. The GMDH polynomial neural network introduced by Ivakhnenko [25] is a feed-forward neural network. This algorithm is a self-organizing system that is used to search progressively for the optimal solution to complicated nonlinear problems. This algorithm approximates the relationship between input and output based on quadratic polynomials. Therefore, GMDH creates new neurons in each layer by connecting pairs of neurons with quadratic polynomials. The goal of a mathematical model problem of GMDH is to discover a function (\hat{f}) that can be used to approximate an original function (f) in order to estimate the model's output (\hat{y}) for an assumed input vector X including n input variables [26]. For this, given n data instances of multi-input single-output data pairs, the following results are obtained:

$$y_i = f(X) = f(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}); (i = 1, 2, \dots, M). \quad (14)$$

A mathematical formulation describes the general equation between input-output variables. The goal now is to

design a GMDH network in such a way that the square of the deviation output and the estimated output is as little as possible, which means:

$$\sum_{i=1}^M [\hat{f}(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}) - y_i]^2 \longrightarrow \min. \quad (15)$$

The mathematical expression between input and output variables can be described by a sophisticated discrete variant of the Volterra function called the Kolmogorov–Gabor polynomial. This series is presented in the following forms.

$$yw_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n w_{ijk} x_i x_j x_k + \dots \quad (16)$$

It is employed in this study to calculate the GMDH network's quadratic polynomial, which can be represented as [27]:

$$\hat{y} = G(x_i, x_j) = w_0 + w_1 x_i + w_2 x_j + w_3 x_i x_j + w_4 x_i^2 + w_5 x_j^2. \quad (17)$$

The mathematical expression (18) shows how neurons in a GMDH network are linked together to create the equation between input-output variables. Using the least-squares regression method, the weighting coefficient values of (17) are determined. This means that the deviation output, y , and the one that is calculated, \hat{y} , for each pair of x_i and x_j input variables are as minimal as possible. Thus, the weighting coefficients values of the quadratic function G_i are determined in order to optimize the fit of the output throughout the entire number of sample data pairs, that is [27].

$$E = \frac{\sum_{i=1}^M (y_i - G_i)^2}{M} \longrightarrow \min. \quad (18)$$

In its ordinary form, the GMDH method considers all possible combinations of two input variables from a total of n input variables in order to generate the regression polynomial in the form of (17) that fits the dependent data ($y_i, i = 1, 2, \dots, M$) the best in a least-squares scheme. As a consequence, the initial layer of the GMDH network's architecture will be selected with $C_n^2 = n(n-1)/2$ input neurons for the creation of the quadratic polynomial based on observations $\{(y_i, x_{ip}, x_{iq}); (i = 1, 2, \dots, M)\}$ for varied $p, q \in \{1, 2, \dots, n\}$. To put it another way, it is currently possible to generate M data triples

$\{(y_i, x_{ip}, x_{iq}); (i = 1, 2, \dots, M)\}$ from observations by utilizing such $p, q \in \{1, 2, \dots, n\}$ in the following way:

$$\begin{bmatrix} x_{1p} & x_{1q} & y_1 \\ x_{2p} & x_{2q} & y_2 \\ \vdots & \vdots & \vdots \\ x_{mp} & x_{mq} & y_m \end{bmatrix}. \quad (19)$$

By replacing the quadratic sub-expression in the shape of (17) for each row of M data triples, the matrix equation shown below is simple to construct.

$$AW = Y, \quad (20)$$

where W and Y are the vectors of unknown quadratic polynomial weighting coefficients in (17) and a vector containing the observed values of outputs, respectively.

$$\begin{aligned} W &= \{w_0, w_1, w_2, w_3, w_4, w_5\}^T, \\ Y &= \{y_1, y_2, y_3, \dots, y_M\}^T. \end{aligned} \quad (21)$$

The superscript T denotes the matrix's transposition. It is self-evident that:

$$A = \begin{bmatrix} 1 & x_{1p} & x_{1q} & x_{1p}x_{1q} & x_{1p}^2 & x_{1q}^2 \\ 1 & x_{2p} & x_{2q} & x_{2p}x_{2q} & x_{2p}^2 & x_{2q}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{mp} & x_{mq} & x_{mp}x_{mq} & x_{mp}^2 & x_{mq}^2 \end{bmatrix}. \quad (22)$$

To solve regression analysis problems, we employ the least-squares approach.

$$W = (A^T A)^{-1} A^T Y. \quad (23)$$

This returns the vector with the best quadratic (17) weighting coefficients for the entire set of M data triples. It is worth noting that this technique is replicated for each neuron in the subsequent hidden layer, depending on the network's connectivity structure. Using the GMDH algorithm, the following GMDH network was obtained for the prediction C_P' (see Figure 2).

With respect to Figure 2, the following equations were obtained for the prediction C_P' using the GMDH approach.

The first layer output:

$$\begin{aligned} (C_P')_1 &= 0.1101 - 0.0110 \left(\frac{y_t}{y_1} \right) + 3.6904(\theta) + 6.2904 \times 10^{-4} \left(\frac{y_t}{y_1} \right)^2 + 8.3932 \times 10^{-5} (\theta)^2 - 3.0321 \times 10^{-4} \left(\frac{y_t}{y_1} \right) (\theta), \\ (C_P')_3 &= 0.1253 - 0.0180 (Fr_1) - 5.5504 \times 10^{-4} (\theta) + 0.0013 (Fr_1)^2 + 9.1284 \times 10^{-5} (\theta)^2 - 3.02542 \times 10^{-4} (Fr_1) (\theta), \\ (C_P')_4 &= 0.0979 - 0.0079 (Fr_1) + 0.0014 \left(\frac{x}{y_1} \right) + 8.1446 \times 10^{-5} (Fr_1)^2 - 6.6223 \times 10^{-5} \left(\frac{x}{y_1} \right)^2 + 5.4575 \times 10^{-5} (Fr_1) \left(\frac{x}{y_1} \right), \\ (C_P')_6 &= 0.0109 + 0.01502 \left(\frac{y_t}{y_1} \right) + 0.0011 \left(\frac{x}{y_1} \right) - 0.0011 \left(\frac{y_t}{y_1} \right)^2 - 6.2790 \times 10^{-5} \left(\frac{x}{y_1} \right)^2 + 5.8831 \times 10^{-5} \left(\frac{y_t}{y_1} \right) \left(\frac{x}{y_1} \right). \end{aligned} \quad (24)$$

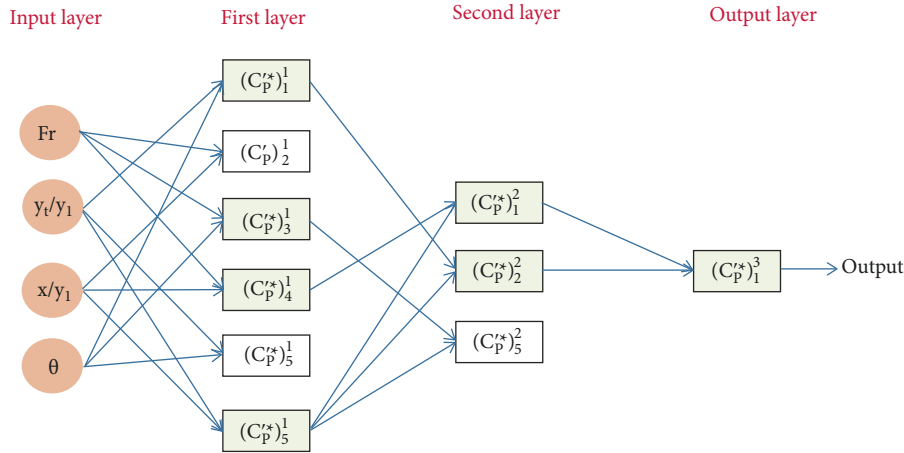


FIGURE 2: The GMDH network for prediction C'_p .

The second layer output:

$$\begin{aligned}
 (C'_p)^2_1 &= 0.1382 - 2.6925(C'_p)^1_4 - 1.7543(C'_p)^1_6 - 1.3848 \times 10^3 \left((C'_p)^1_4 \right)^2 - 1.3189 \times 10^3 \left((C'_p)^1_6 \right)^2 + 2.7568 \times 10^3 \left((C'_p)^1_4 \right) \left((C'_p)^1_6 \right), \\
 (C'_p)^2_2 &= 0.1671 - 0.7202(C'_p)^1_1 - 4.8455(C'_p)^1_6 + 31.2869 \left((C'_p)^1_1 \right)^2 + 70.9947 \left((C'_p)^1_6 \right)^2 - 41.3749 \left((C'_p)^1_1 \right) \left((C'_p)^1_6 \right), \\
 (C'_p)^2_3 &= 0.1845 - 1.1091(C'_p)^1_3 - 5.1238(C'_p)^1_6 + 36.6985 \left((C'_p)^1_3 \right)^2 + 75.5852 \left((C'_p)^1_6 \right)^2 - 45.3275 \left((C'_p)^1_3 \right) \left((C'_p)^1_6 \right).
 \end{aligned} \tag{25}$$

The output layer:

$$(C'_p)^3_1 = 0.0337 - 0.4779(C'_p)^2_1 + 0.1849(C'_p)^2_2 + 6.7083 \left((C'_p)^2_1 \right)^2 + 5.1880 \left((C'_p)^2_2 \right)^2 - 0.2282 \left((C'_p)^2_1 \right) \left((C'_p)^2_2 \right). \tag{26}$$

Finally, the value of C'_p is calculated from the above formulations of GMDH. As seen with replacing the input variables with the GMDH formulation, the $(C'_p)^3_1$ approximate the outcome GMDH network for prediction of C'_p .

2.5. M5 Model Tree (M5MT) and Model Development. M5MT is one of the most widely used decision tree techniques in data-driven modeling. In the M5MT, the entire input domain is recursively partitioned into subdomains, with each subdomain being predicted using a multiple linear regression model. The graphical M5MT is constructed from a root node, the number of binary branches, a group of inner nodes (splits), and a number of terminal nodes (leaves) [28]. For this reason, the resulting tree model has a clear decision structure and is understandable for everyone.

The constructing, pruning, and smoothing of the tree are the three main components of the M5MT algorithm. The splitting criterion is used to construct the primary tree. The

expected reduction in error resulting from evaluating each attribute at the node is calculated using this splitting criterion. A measure of the error at a node is defined as the standard deviation of the class values that reach the node. After that, the attribute with the highest anticipated error reduction is chosen. The following formula is used to get the standard deviation reduction (SDR) for M5MT:

$$\text{SDR} = s d(T) - \sum_i \frac{|T_i|}{|T|} \times s d(T_i), \tag{27}$$

where T denotes the set of instances that reach the node, T_i is the result of separating the node according to the attribute chosen, and $s d$ denotes the standard deviation [29]. To stop splitting, either only a few examples remain, or their class values are less than 5% of the initial instance set's standard deviation. It is possible to encounter an over-fitting problem based on training data during the creation tree process. Pruning is a technique that has been employed in trying to alleviate this difficulty in the past. It decreases the size of the

model tree by deleting splits that don't meaningfully enhance prediction.

For the test data, the pruning algorithm uses an estimate of the predicted error at each node. To begin, the absolute deviation between the observed and estimated output values for each of the training cases entering the node is averaged. Due to the fact that the trees were constructed specifically for this dataset, the average will underrate the predicted error for new instances. This is compensated for by multiplying the output value by the factor $(n + \nu)/(n - \nu)$, where n denotes the size of the training instances received at the node [30]. In addition, ν is the number of model attributes that signify the output value at that node. The leaf node can be omitted if the estimated error is lower at the parent. As a result, this multiplication is done to make sure that new data, rather than training data, don't get underestimated by the multiplication.

Quinlan [31] describes a smoothing approach that employs the leaf model to determine the estimated value during the smoothing phase. The value is then smoothed by merging it with the linear model's estimated value for each node along the path back to the root. This requires the following calculation:

$$p' = \frac{np + kq}{n + k}, \quad (28)$$

p' is a prediction that exceeds the upper node; p is a prediction from below that is passed to the current node; q is the model's predicted value at the node; n represents the number of training instances that have made it to the preceding node, and Wang and Witten [28] constant is denoted by k . Using the M5MT created, the graphical tree, is shown in Figure 3.

The decision rules concerned with Figure 3 are as follows:

$$LM(4) = C_p' = -0.0041 \times Fr_1 + 0.0792. \quad (29)$$

As seen, four linear rules were obtained for the prediction C_p' . Concerning the values of Fr_1 , the appropriate rule was selected then C_p' was calculated.

3. Results and Discussion

Statistical indices such as correlation coefficient (CC), root mean square error (RMSE), mean absolute error (MAE), index of agreement (Ia), scatter index (SI), and BIAS index was employed to assess the qualitative evaluation of the developed suggested models.

$$\begin{aligned} CC &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \\ RMSE &= \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}}, \\ MAE &= \frac{1}{N} \sum_{i=1}^N |x_i - y_i|, \\ Ia &= 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (|x_i - \bar{x}| + |y_i - \bar{y}|)^2}, \\ SI &= \frac{RMSE}{\bar{x}} \times 100, \quad BIAS = \bar{y} - \bar{x}, \end{aligned} \quad (30)$$

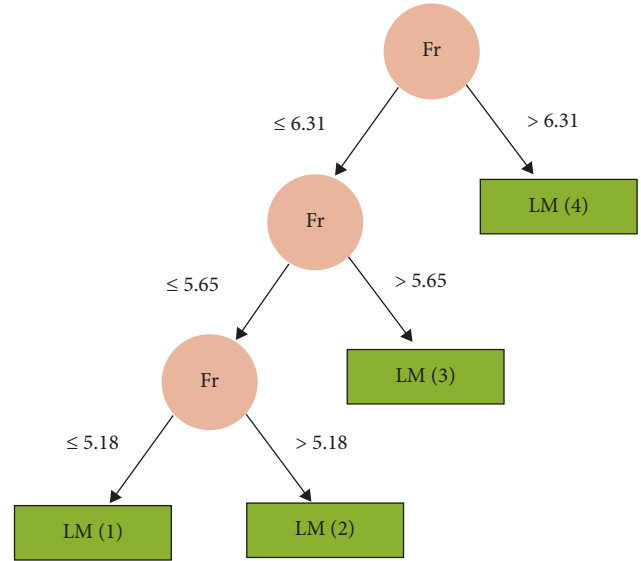


FIGURE 3: M5MT created the regression tree to predict C_p' .

where x_i and y_i denote measured and predicted values, \bar{x} and \bar{y} represent the average of measured and predicted values, and N denotes the number of the dataset. In Table 2, the statistical index values for the proposed soft computing methods in the training and testing stages for the prediction of C_p' .

Table 2 shows that the ELM model performed C_p' prediction with the lowest errors (RMSE, MAE, and SI) and higher coefficient correlation (CC) and index agreement (Ia) than the GMDH and M5MT models in both the training and testing stages. As observed in Table 2, the ELM method with $CC = 0.9183$, $RMSE = 0.0067$, and $MAE = 0.0051$ has the best prediction accuracy compared with GMDH and M5MT. On the other hand, it can be deduced from Table 2 that GMDH ($CC = 0.8818$, $RMSE = 0.0078$, and $MAE = 0.0058$) and M5MT ($CC = 0.6883$, $RMSE = 0.0120$, and $MAE = 0.0090$) have the second and third level of accuracy for the prediction C_p' in the testing stage. Furthermore, the scatter index ($SI = 11.88\%$) for the ELM method is smaller than the SI for GMDH ($SI = 13.89\%$) and M5MT ($SI = 21.28\%$). It is noteworthy that the BIAS values of ELM for training and testing data are 0.0008 and 0.0020, respectively, indicating a slightly overestimated C_p' . However, the BIAS values of GMDH in training and testing were -0.0005 and 0.0016 , and for M5MT, they were -0.0003 and 0.0014 for the training and testing stages, respectively. Therefore, it can be concluded that, overall, ELM was more conservative for prediction C_p' than GMDH and M5MT.

By comparing two soft computing approaches, including the ELM (as the best model) and the M5MT (as the worst model) in the testing stage, it can be shown that the ELM produced significantly lower errors in prediction C_p' than the M5MT. The values of RMSE and MAE obtained via ELM showed that they decreased by about 79.10% and 76.47% compared with the M5MT in the testing stage, respectively.

Besides, the CC and Ia values for ELM increased by 33.42% and 21.05% compared with M5MT. The error values

TABLE 2: The statistical indices for ELM, GMDH, and M5MT for the prediction of C_p' .

Approach	Data set	CC	RMSE	MAE	Ia	SI (%)	BIAS
ELM	Training	0.9770	0.0046	0.0040	0.9841	8.00	0.0008
ELM	Testing	0.9183	0.0067	0.0051	0.9569	11.88	0.0020
GMDH	Training	0.9135	0.0074	0.0066	0.9512	12.85	-0.0005
GMDH	Testing	0.8818	0.0078	0.0058	0.9361	13.89	0.0016
M5MT	Training	0.8309	0.0108	0.0090	0.8467	18.81	-0.0003
M5MT	Testing	0.6883	0.0120	0.0090	0.7905	21.28	0.0014

of RMSE and MAE from GMDH in the testing stage showed that these values improved by about 53.85% and 55.17% compared with those obtained from M5MT, respectively. Moreover, the CC and Ia values of GMDH increased by about 28.11% and 18.42% compared with M5MT. Based on error measures, it is clear that the ELM model was better than the GMDH and M5MT in terms of accuracy.

The mathematical expressions for the prediction C_p' were formulated from the proposed soft computing models. It seems the M5MT result was more transparent and easier to use than ELM and GMDH. M5MT provided four multivariate linear models for the prediction of C_p' . Although the derived rules from M5 had less accuracy than ELM and GMDH, these rules are more straightforward for predicting C_p' . On the other hand, GMDH presented complex mathematical formulations for the prediction of C_p' . The GMDH used two hidden layers of neurons for the prediction of C_p' . The GMDH utilized the quadratic polynomial of input variables and the combination of the best neurons in each layer to predict C_p' . ELM generated the coefficients matrixes for the prediction of C_p' . It seems the ELM mathematical shapes were simple compared with the GMDH equations.

M5MT method provided four simple equations for the prediction of C_p' based on dividing the domain of the problem. It is worth mentioning that M5MT selected the appropriate rule concerning only one input variable (i.e., Froude number) and used two variables, including Fr_1 and x/y_1 in their rules. However, the ELM and GMDH used all the independent variables to generate the predictive expressions for estimating C_p' values. Although M5MT has led to the creation of simple rules for predicting C_p' , these simple formulas cause M5MT to be less accurate than ELM and GMDH expressions. M5MT divided the domain of input variables into four subdomains and represented four regression equations to estimate C_p' . In fact, separating the domain of the problem into the local subdomains and combining their results caused an improvement in the accuracy compared with the single equation.

Figures 4–9 compare observed values of C_p' versus estimated values obtained from ELM, GMDH, and M5MT during the training and testing stages. As observed in scatter plots, the predicted C_p' values from ELM were more concentrated around the ideal line (the 45-degree line).

In addition, the variation of C_p' values obtained by soft computing methods versus observed C_p' demonstrated the capability of estimating C_p' by the proposed approaches. These figures graphically confirmed the higher ELM accuracy than GMDH and M5MT.

These figures indicated that ELM performed better in the training and testing stages than GMDH and M5MT. As shown, the ELM reasonably estimated C_p' in the training and testing stages (bias = 0.0008 and 0.0020). The remarkable point is that in Figures 4 and 5, the peak values of C_p' by ELM were estimated well. In contrast, GMDH and M5MT slightly underestimated the values of C_p' , especially at the peak C_p' values. From the comparison between Figures 4–9, it can be deduced that ELM is more skillful and accurate than GMDH and M5MT in the prediction of C_p' .

The present study results were compared with earlier research conducted by Samadi et al. [8]. They used three data-driven algorithms, such as MARS, GEP, and CART, for the prediction of C_p' . Samadi et al. [8] indicated that CART results for the prediction of C_p' have three and four non-terminal and terminal nodes, respectively. The CART tree structure used only the Fr_1 variable with threshold values of 6.31, 5.18, and 5.65 for the CART tree structure. The present study used another common decision tree technique, namely M5MT, to predict C_p' .

It is worth noting that M5MT and CART are two common decision tree algorithms used for regression problems. In the present study, M5MT, among the four input variables included, Fr_1 , x/y_1 , y_t/y_1 , and θ , only selected the Fr_1 variable for prediction of C_p' . As illustrated in Figure 3, M5MT's tree structure uses three nonterminal and four terminal nodes. In addition, the splitting values for M5MT were 6.31, 5.18, and 5.65. A comparison of the tree structures of the decision tree (i.e., M5MT) and the proposed CART presented by Samadi et al. [8] revealed that the two models have similar structures. They employed a splitting variable (Fr_1) and the same splitting values for the generation of regression trees. As a result, the graphical structures of these two decision tree algorithms for selecting the independent variable (Fr_1) and creating four if-then rules are similar. However, M5MT provided linear regression functions while CART presented constant values in their terminal nodes. This was the main distinctive characteristic of the differences between M5MT and CART concerning the nature of CART and M5MT. It should be noted that due to the nature of the M5MT algorithm, the Fr_1 parameter was an attribute that caused error reduction for the prediction of C_p' in the regression tree obtained by M5MT. In addition, the present study's findings about selecting the Fr_1 parameter is completely consistent with the results of the CART tree provided by Samadi et al. [8].

Also, the statistical measures indicated the results of M5MT and CART were more or less the same. M5MT with RMSE = 0.0120 and MAE = 0.0090, compared with CART

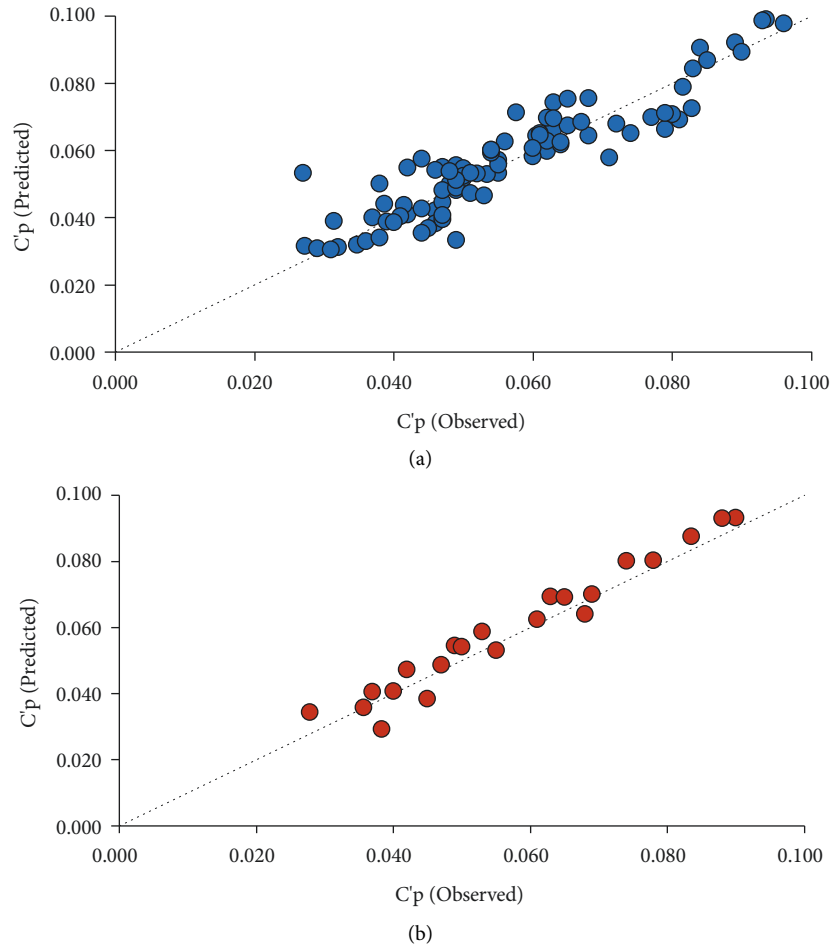


FIGURE 4: Scatter plots of measured and estimated values of C_p' for ELM during (a) training stage and (b) testing stage.

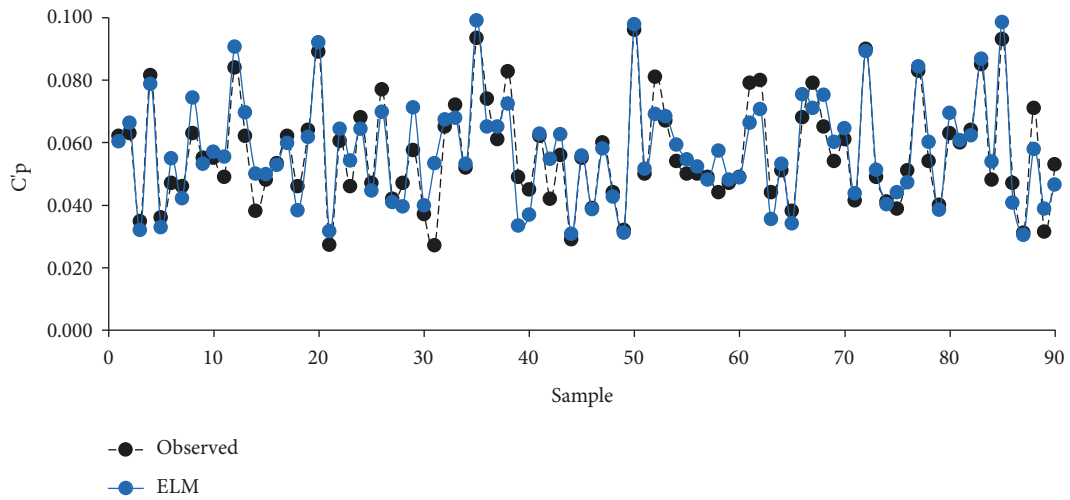


FIGURE 5: Continued.

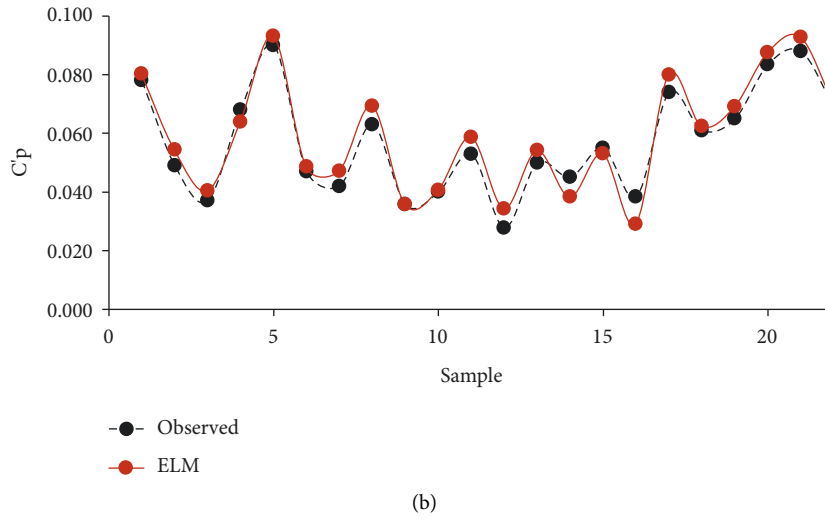


FIGURE 5: Variation of C_p' with ELM during the training stage (a) and testing stage (b).

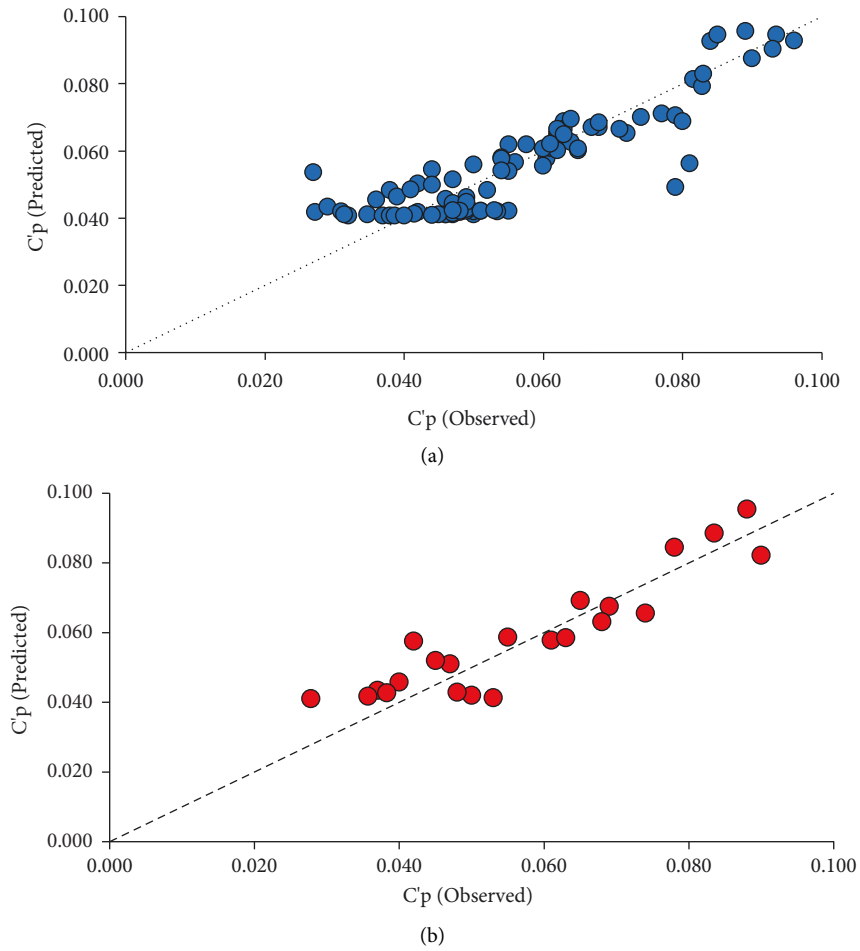


FIGURE 6: Scatter plots of measured and estimated values of C_p' for GMDH during (a) training stage and (b) testing stage.

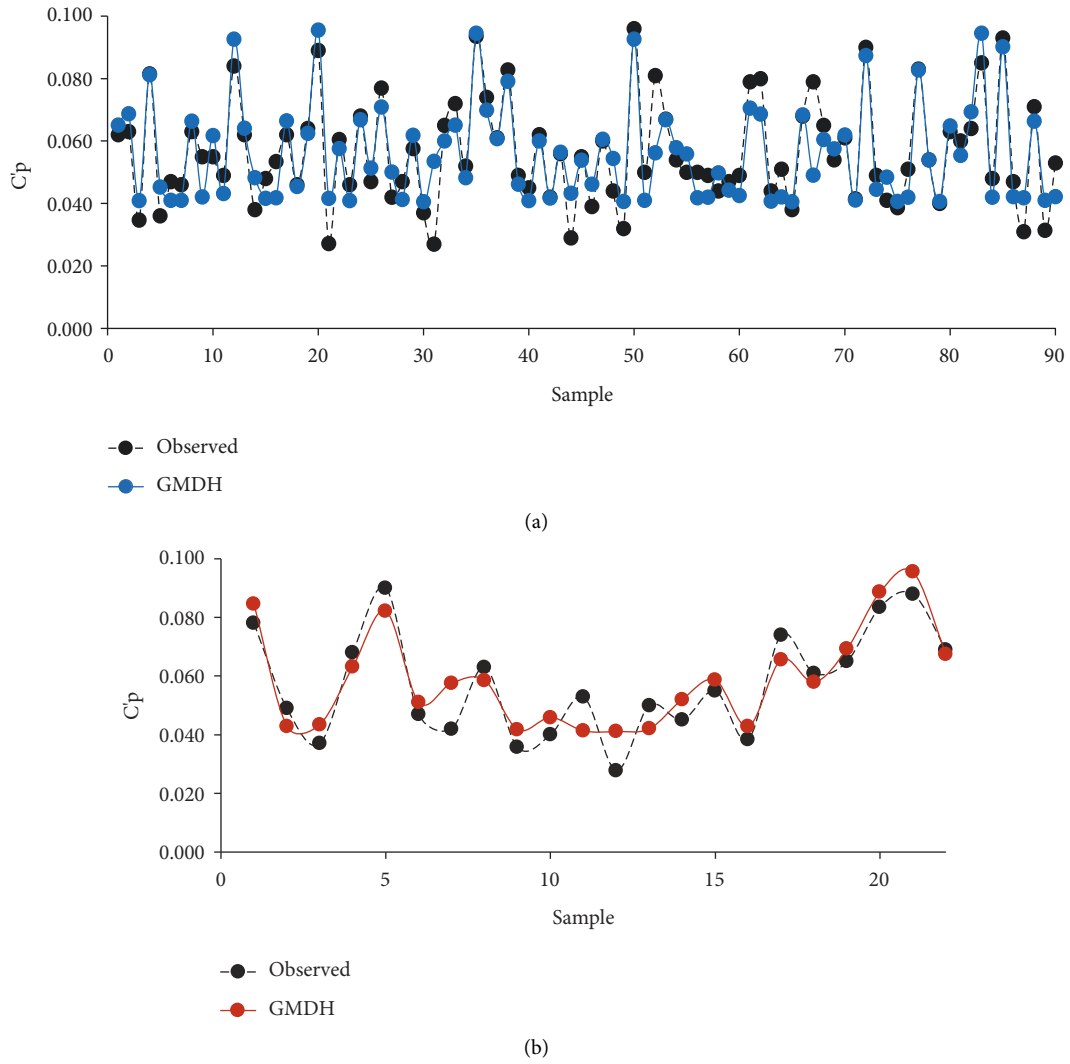


FIGURE 7: Variation of C_p with GMDH during the training stage (a) and testing stage (b).

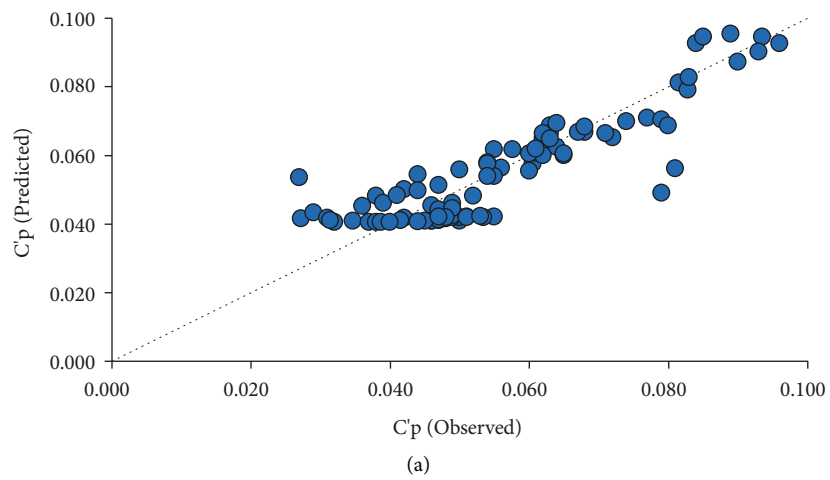


FIGURE 8: Continued.

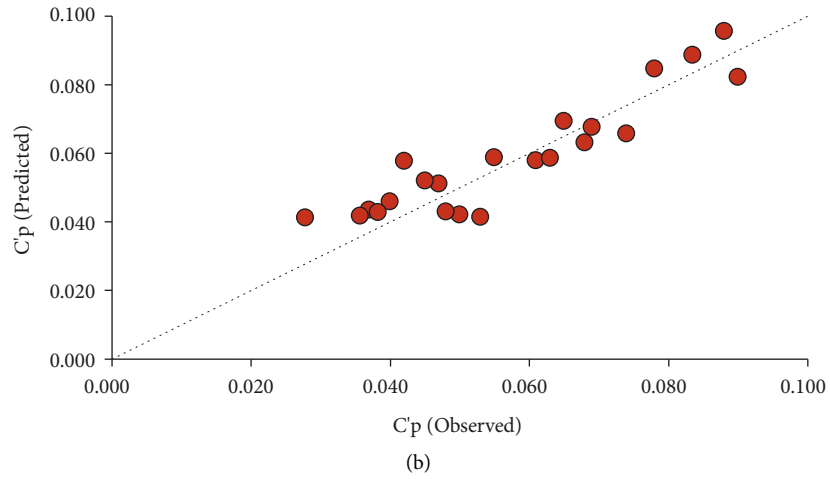
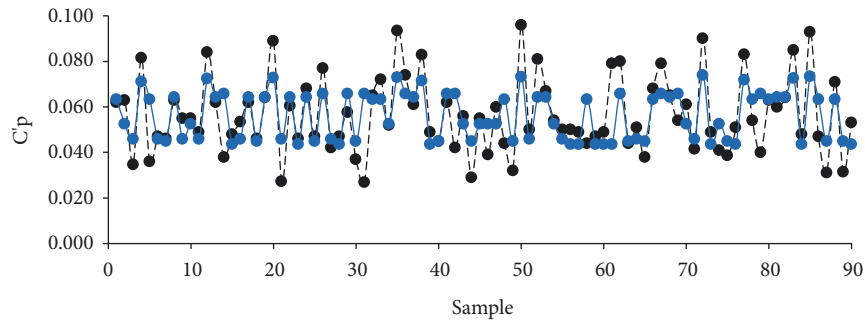
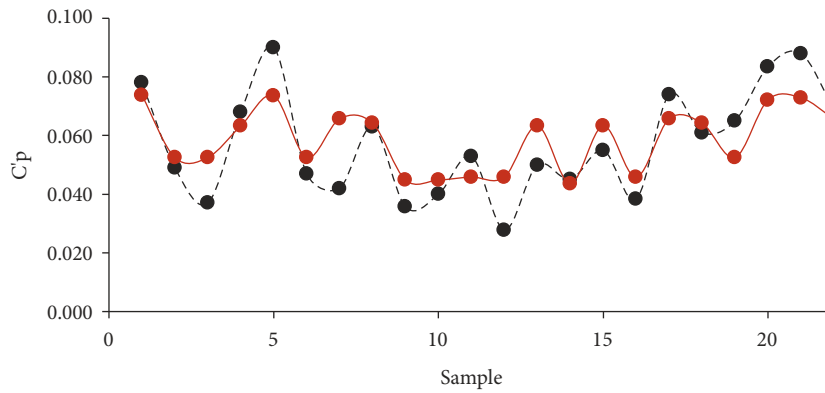


FIGURE 8: Scatter plots of measured and estimated values of C_p' for M5MT during (a) training stage and (b) testing stage.



—●— Observed
—●— M5MT



—●— Observed
—●— M5MT

FIGURE 9: Variation of C_p' with M5MT during the training stage (a) and testing stage (b).

[12] with RMSE = 0.012, MAE = 0.009, was similar results. However, M5 generated multivariate linear regression equations while CART generated the constant values for the prediction of C_p' . The linear equations by M5MT caused the flexibility and generalizability of M5 to improve the prediction of C_p' compared with the constant values of C_p' that was yielded by CART. Compared with GMDH in the present study with GEP [12], both algorithms provided nonlinear mathematical expressions for the prediction of C_p' . The accuracy of GMDH with RMSE = 0.0078 and MAE = 0.0058 is slightly better than GEP results with RMSE = 0.008 and MAE = 0.006. Moreover, compared with the formulas provided by GEP and GMDH, GEP requires more difficult calculations related to trigonometric and algebraic functional sets appearing in the computing process, including exp, ln, sin, and atan functions. In contrast, GMDH used polynomial quadratic equations. From this perspective as well, it seems the computational effort was less than GEP. Finally, the comparison results of ELM with MARS indicated that both algorithms had almost identical results.

4. Summary and Conclusions

Stilling basins are used widely as hydraulic dissipation structures in large dams. An accurate estimation of C_p' within stilling basins is a critical issue for hydraulic engineering for the design of stilling basins. This study employed three soft computing methods, including ELM, GMDH, and M5MT, to estimate the C_p' that occurred during the hydraulic jump in the sloping channels.

Different soft computing models were developed to estimate C_p' according to the dimensionless parameters and experimental data. ELM showed the lowest error values of RMSE, MAE, and SI in the training and testing stages. In addition, ELM has the highest correlation coefficient and Ia values than those obtained from GMDH and M5MT. Therefore, the proposed soft computing models provided sufficiently accurate results due to this problem's complexity.

ELM formulated the matrices of coefficients for the prediction of C_p' . In addition, GMDH provided mathematical quadratic polynomials and combined input variables for the prediction of C_p' . M5MT generated four simple rules for the estimation of C_p' . It seems the application of M5MT was the most straightforward method for the prediction of C_p' . It is noteworthy that complexity degree equations were derived from ELM and GMDH for the prediction C_p' have been more than M5MT rules.

In summary, the ELM method provided a weight matrix for predicting C_p' . GMDH generated second-order polynomial equations for the prediction of C_p' . M5MT developed piecewise multiple linear regression equations for the calculation of C_p' . GMDH and M5MT methods generated explicit and clear mathematical expressions to estimate C_p' . In addition, M5MT divided the problem domain into subdomains and fitted local linear models to compute C_p' . It seems that in terms of the degree of complexity of the developed models in estimating C_p' , the ELM model has the highest complexity, followed by GMDH and M5MT models. However, in contrast to the complexity degree of the

proposed models, their computational accuracy has increased for the estimation of C_p' so that the ELM model has the highest accuracy, followed by GMDH and M5MT.

Finally, the comparisons of the results with previous research revealed that the proposed applications of soft computing methods have good performance for prediction C_p' .

Further works can be considered to use pressure field data for modeling with soft computing methods. It is also recommended that hybrid data-driven models with evolutionary algorithms be used instead of stand-alone data-driven models to figure out the coefficient of pressure fluctuations.

Data Availability

The data references are described in the text of the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] W. H. Hager, "Energy Dissipators and Hydraulic Jump," *Kluwer Academic Publ., Water Science and Technology Library*, vol. 8, 1992.
- [2] H. Chanson (ED.), *Energy Dissipation in Hydraulic Structures*, CRC Press, FL, USA pp. 178, 2015, ISBN: 9780367575731.
- [3] V. Fiorotto and A. Rinaldo, "Turbulent pressure fluctuations under hydraulic jumps," *Journal of Hydraulic Research*, vol. 30, no. 4, pp. 499–520, 1992a.
- [4] V. Fiorotto and A. Rinaldo, "Fluctuating uplift and lining design in spillway stilling basins," *Journal of Hydraulic Engineering*, vol. 118, no. 4, pp. 578–596, 1992b.
- [5] A. H. Zaji, H. Bonakdari, S. R. Khodashenas, and S. Shamsirband, "Firefly optimization algorithm effect on support vector regression prediction improvement of a modified labyrinth side weir's discharge coefficient," *Applied Mathematics and Computation*, vol. 274, pp. 14–19, 2016.
- [6] A. Guven, M. Günal, and A. Çevik, "Prediction of pressure fluctuations on sloping stilling basins," *Canadian Journal of Civil Engineering*, vol. 33, no. 11, pp. 1379–1388, 2006.
- [7] A. Guven, "A predictive model for pressure fluctuations on sloping channels using support vector machine," *International Journal for Numerical Methods in Fluids*, vol. 66, no. 11, pp. 1371–1382, 2011.
- [8] M. Samadi, H. Sarkardeh, and E. Jabbari, "Explicit data-driven models for prediction of pressure fluctuations occur during turbulent flows on sloping channels," *Stochastic Environmental Research and Risk Assessment*, vol. 34, no. 5, pp. 691–707, 2020.
- [9] 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.
- [10] Q. Chen, C. Wang, and L. Song, "Prediction of low-temperature rheological properties of SBS modified asphalt," *Advances in Civil Engineering*, vol. 2020, Article ID 8864766, 8 pages, 2020.
- [11] F. Saberi-Movahed, M. Najafzadeh, and A. Mehrpooya, "Receiving more accurate predictions for longitudinal dispersion coefficients in water pipelines: training group method

- of data handling using extreme learning machine conceptions,” *Water Resources Management*, vol. 34, no. 2, pp. 529–561, 2020.
- [12] M. M. Hameed, M. A. Abed, N. Al-Ansari, and M. K. Alomar, “Predicting compressive strength of concrete containing industrial waste materials: novel and hybrid machine learning model,” *Advances in Civil Engineering*, vol. 2022, Article ID 5586737, 19 pages, 2022.
- [13] M. K. AlOmar, M. M. Hameed, N. Al-Ansari, and M. A. AlSaadi, “Data-driven model for the prediction of total dissolved gas: robust artificial intelligence approach,” *Advances in Civil Engineering*, vol. 2020, Article ID 6618842, 20 pages, 2020.
- [14] M. Najafzadeh, G. A. Barani, and H. M. Azamathulla, “GMDH to predict scour depth around a pier in cohesive soils,” *Applied Ocean Research*, vol. 40, pp. 35–41, 2013.
- [15] M. Najafzadeh and G. Oliveto, “More reliable predictions of clear-water scour depth at pile groups by robust artificial intelligence techniques while preserving physical consistency,” *Soft Computing*, vol. 25, no. 7, pp. 5723–5746, 2021.
- [16] M. Najafzadeh, M. Rezaie-Balf, and A. Tafarajnoruz, “Prediction of riprap stone size under overtopping flow using data-driven models,” *International Journal of River Basin Management*, vol. 16, no. 4, pp. 505–512, 2018.
- [17] A. Barzkar, M. Najafzadeh, and F. Homaei, “Evaluation of drought events in various climatic conditions using data-driven models and a reliability-based probabilistic model,” *Natural Hazards*, vol. 110, no. 3, pp. 1931–1952, 2022.
- [18] M. Samadi, E. Jabbari, and H. M. Azamathulla, “Assessment of M5’ model tree and classification and regression trees for prediction of scour depth below free overfall spillways,” *Neural Computing & Applications*, vol. 24, no. 2, pp. 357–366, 2014.
- [19] N. C. Jung, I. Popescu, P. Kelderman, D. P. Solomatine, and R. K. Price, “Application of model trees and other machine learning techniques for algal growth prediction in Yongdam reservoir, Republic of Korea,” *Journal of Hydroinformatics*, vol. 12, no. 3, pp. 262–274, 2010.
- [20] R. Khorrami, A. Derakhshani, and H. Moayedi, “New explicit formulation for ultimate bearing capacity of shallow foundations on granular soil using M5’ model tree,” *Measurement*, vol. 163, Article ID 108032, 2020.
- [21] M. Najafzadeh, F. Homaei, and S. Mohamadi, “Reliability evaluation of groundwater quality index using data-driven models,” *Environmental Science and Pollution Research*, vol. 29, no. 6, pp. 8174–8190, 2022.
- [22] M. Gunal, *Numerical and Experimental Investigation of Hydraulic Jumps*, Ph.D. Thesis, University of Manchester Institute of Science and Technology, Manchester, 1996.
- [23] B. Bhattacharya, R. K. Price, and D. P. Solomatine, “Machine learning approach to modeling sediment transport,” *Journal of Hydraulic Engineering*, vol. 133, no. 4, pp. 440–450, 2007.
- [24] G. B. Huang, Q. Y. Zhu, and C. K. Siew, “Extreme learning machine: theory and applications,” *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
- [25] A. Ivakhnenko, “The group method of data handling in long-range forecasting,” *Technological Forecasting and Social Change*, vol. 12, no. 2–3, pp. 213–227, 1978.
- [26] M. Najafzadeh and F. Saberi-Movahed, “GMDH-GEP to predict free span expansion rates below pipelines under waves,” *Marine Georesources & Geotechnology*, vol. 37, no. 3, pp. 375–392, 2019.
- [27] 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.
- [28] Y. Wang and I. H. Witten, “Induction of model trees for predicting continuous classes,” University of Economics, Faculty of Informatics and Statistics, Prague, 1997.
- [29] N. C. Jung, “Eco-hydraulic modelling of eutrophication for reservoir management,” CRC Press, FL, USA, 2010, ISBN: 9780415573825.
- [30] I. H. Witten and E. Frank, *Data Mining-Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publisher, San Francisco, CA, 2005.
- [31] J. R. Quinlan, *Learning with continuous classes*, A. Adams and L. Sterling, Eds., pp. 343–348, World Scientific, Singapore, 1992.