Water and Soil Management and Modeling

Online ISSN: 2783 - 2546

University of Mohaghegh Ardabili

Water wave height prediction using a novel hybrid deep learning model with output uncertainty quantification

Elham Ghanbari-Adivi*[1], Mohammad Ehteram [2]

[1] Associate Professor, Department of Water Engineering, Faculty of Agriculture, Shahrekord University, Shahrekord, Iran
[2] Ph.D. Department of Hydraulic Structures, Semnan University, Semnan, Iran

Abstract

Accurate significant wave height (SWH) prediction is essential for improving the safety and efficiency of maritime operations. Thus, our study develops the Gaussian data augment (GDA) technique- Meerkat optimization algorithm (MOA)- variational mode decomposition (VMD)- complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)- bidirectional long short-term memory neural network model (BILSTM)- attention mechanism (AT)- gated recurrent unit (GRU) model to accurately predict SWH and overcome the limitations of the GRU model. First, the GDA method addresses the problem of data scarcity by providing new data points. Next, MOA is used to adjust the parameters of the components of the hybrid model. The VMD method then reduces the intricacy of the time series by converting them into subseries with lower complexity, named intrinsic mode functions (IMFs). However, as the first IMF retains the complex characteristics of the original time series, the CEEMDAN method is applied to decompose it into secondary IMFs with reduced complexity. Subsequently, the BILSTM model extracts forward and backward temporal features from the secondary IMFs and the initial remaining IMFs. An attention mechanism is then applied to assign the attention weights to the extracted features. Each attention weight indicates the importance of a feature, enabling the GRU model to identify the most important time series features for predicting SWH. Finally, the weighted features are fed into the GRU model to predict SWH accurately. Our study also couples the kernel density estimation method with the GDA-MOA-VMD-CEEMDAN-BILSTM-attention mechanism-GRU (GMVCBAG) model to quantify the uncertainty of the model outputs. The new model is benchmarked against multiple predictive models. Our study also uses various performance metrics to evaluate the accuracy of predictions. Our findings indicate that Nash–Sutcliffe efficiency (NSE), mean absolute error (MAE), standard deviation of the relative error (STDRE), and explained variance of the GMVCBAG model are 0.973, 0.245, 1.245, and 0.899, respectively. Results indicate that GMVCBAG provides reliable SWH predictions. Moreover, the outputs of the new model have a lower uncertainty than those of the other predictive models. Thus, GMVCBAG is a suitable model for predicting SWH in the different regions of the world.

Keywords: Deep learning model, Temporal features, Significant wave height, Time series decomposition

Article Type: Research Article
Academic Editor: Raoof Mostafazadeh

*Corresponding Author, E-mail: Ghanbariadivi@sku.ac.ir
Citation: GhanbariAdivi, E. and Ehteram, M. (2026). Water wave height prediction using a novel hybrid deep learning model with output uncertainty quantification. Water and Soil Management and Modelling, 6 (Special Issue: New Approaches to Water and Soil Management and Modeling), 140-170.
doi: 10.22098/mmws.2025.18652.1709
Received: 25 October 2025, Received in revised form: 26 November 2025, Accepted: 27 November 2025, Published online: 03 June 2026
Water and Soil Management and Modeling, Year 2026, Vol. 6, Special Issue, pp. 140-170.
Publisher: University of Mohaghegh Ardabili

1. Introduction

Significant wave height (SWH) is a critical parameter in marine engineering as it has significant implications for the safety and efficiency of maritime shipping and offshore operations. (Ikram et al., 2023; Ban et al., 2023). SWH plays a crucial role in the stability of coastal and offshore structures (Chen et al., 2023). Moreover, SWH provides valuable insights for evaluating the potential impacts of extreme weather events, such as hurricanes and storm surges, on offshore and coastal environments (Altunkaynak et al., 2023). Therefore, SWH prediction is essential for improving the safety and efficiency of maritime operations, protecting offshore infrastructure, managing coastal ecosystems, and mitigating the risks associated with extreme weather events (Mahdavi-Meymand and Sulisz, 2023). SWH prediction can also help accurately assess wave energy potential. Decision makers need accurate predictions of SWH to determine the most appropriate locations for installing wave energy converters, which are essential for power generation (Mahdavi-Meymand and Sulisz, 2023). However, accurately predicting significant wave height is complex because it is influenced by numerous dynamic factors, such as ocean currents (Ikram et al., 2023). Thus, scholars have developed different models, such as deep and shallow learning models, to predict SWH accurately.

Shallow and deep learning models have become increasingly popular for accurately predicting significant wave height because they can effectively learn complex and nonlinear relationships in the time series (Mahdavi-Meymand and Sulisz, 2023). Shallow and deep learning models can also adapt to changing environmental conditions, improving their predictive performance over traditional statistical and physics-based models. Scholars have developed different types of these models to predict SWH accurately.

Meng et al. (2021) developed a bidirectional deep learning (BDL) model to predict wave heights. Their findings indicated that the BDL model precisely predicted the wave height. Additionally, the model efficiently captured both forward and backward temporal patterns. Luo et al. (2022) investigated the potential of BILSTM in predicting SWH. Their results showed that the bidirectional (BILSTM) model accurately predicted the SWH and effectively extracted forward and backward temporal features from the time series data.

Adnan et al. (2023) developed ensemble and random forest models to predict SWH. The study findings showed that the ensemble model significantly improved the accuracy of the random forest model. Yeganeh-Bakhtiary et al. (2023) proposed the decision tree (DT) to predict SWH. They used the wind speed as input data. The DT model successfully predicted SWH. Wang and Ying al. (2023a) developed the variational mode decomposition (VMD) algorithm and GRU model to predict SWH. The VMD algorithm reduced the intricacy of time series data. Their findings demonstrated that the GRU-VMD boosted the prediction accuracy of the GRU model. Mahdavi-Meymand and Sulisz (2023) developed a nested artificial neural network (NANN) to predict SWH. The study demonstrated that the NANN model achieved better results than the other models. Zhao et al. (2023) investigated the potential of the long short-term memory (LSTM) model in predicting SWH. They also developed CEEMDAN as a decomposition algorithm to mitigate the intricacy of time series. Their study showed that the CEEMDAN-LSTM model surpassed the LSTM model. Shi et al. (2023) coupled the attention mechanism with the LSTM model to predict SWH. The attention mechanism produced weighted features by assigning the attention weights to time series features. The weights of these features enabled the LSTM model to detect the features that were highly relevant to SWH. Fu et al. (2023) combined a two-stage data processing technique with the LSTM model to predict SWH. They reported that the nonstationary characteristic of the time series significantly reduced the LSTM accuracy. Wang and Ying al. (2023b) proposed the hybrid LSTM-GRU to predict SWH. Their study indicated that the hybrid LSTM-GRU model outperformed the LSTM and GRU models in predicting SWH. Zheng et al. (2023) coupled Multivariate VMD (MVMD) with the GRU model to predict SWH. The MVMD method reduced the nonstationary properties of the time series data. The results demonstrated that the MVMD-GRU model boosted the prediction accuracy of the GRU

model. Altunkaynak et al. (2023) coupled the wavelet transform method with a predictive model to predict SWH. The wavelet transform method reduced the high-frequency components of the time series.

Alizadeh and Nourani et al. (2024) developed the LSTM and GRU models to predict SWH. Their results indicated that both models accurately predicted SWH. Ding et al. (2024) coupled the variational mode decomposition algorithm (VMD) and CEEMDAN with the LSTM model to predict SWH. VMD and CEEMDAN effectively reduced the complexity of the time series by decomposing the time series into several sub-series with lower complexity. The results indicated that VMD-LSTM and CEEMDAN-LSTM improved the performance of LSTM in predicting SWH. Their study also reported that decomposition algorithms, such as CEEMDAN and VMD algorithms, improved the quality of the time series data. Mahdavi-Meymand and Sulisz et al. (2024) developed the original pyramid neural network (PNN) for predicting significant wave height. Their results demonstrated that the PNN model accurately predicted SWH. Additionally, their study revealed that data augmentation techniques significantly enhanced prediction accuracy by generating new data points and addressing data scarcity.

Based on the literature review, shallow and deep learning models can accurately predict SWH. The literature also shows that the GRU model has been widely used to predict SWH. GRU is a deep learning model that uses two gates to process and predict different time series (Ehteram et al., 2024). The GRU model also has a hidden state that stores important information (Zheng et al., 2023; Wang and Ying, 2023b). However, several issues can negatively affect the prediction accuracy of the GRU model. For example, data scarcity and improper adjustment of GRU parameters significantly reduce the accuracy of GRU prediction (Wang et al., 2025; Mumuni et al., 2022; Zhao et al., 2023; Hou et al., 2023). The complexity of time series is another issue that can reduce the accuracy of the GRU model (Zhou et al., 2023). The GRU model also has limitations that may negatively affect its performance in predicting outputs. For example, the GRU model cannot effectively extract forward and backward temporal features from the time series (Ma et al., 2024; Jia et al., 2023). These features show the influence of expected or past events on the current state of wave heights. However, the model loses crucial information due to this limitation, leading to inaccurate predictions. The GRU model also cannot detect the most critical time series features for predicting the target variables (Zulqarnain et al., 2024; Cao et al., 2023). As a result, the model may use unimportant features during modeling, leading to inaccurate predictions. Therefore, to improve the efficiency of the GRU model, it is essential to overcome its limitations and address issues that negatively affect its accuracy.

Researchers have recently coupled the standalone deep learning models with multiple components to overcome their limitations and address issues that reduce their accuracy (Zhang et al., 2024). These components include a data augmentation technique, an optimization algorithm, a hybrid decomposition algorithm, a bidirectional deep learning model, and a kernel density estimation method. The data augmentation technique is the first component that addresses the problem of data scarcity by providing new data points (Qian et al., 2024). The optimization algorithm is another component that adjusts the parameters of the standalone deep learning model (Zhao et al., 2023; Hou et al., 2023). The decomposition algorithm is the next component that addresses the complexity problem of time series by decomposing them into sub-time series with lower complexity (Zhang et al., 2024). Another component is the bidirectional deep learning model, which consists of forward and backward deep learning models (Rezaei et al., 2023; Aung et al., 2023; Yuan et al., 2023). These models overcome the limitations of standalone deep learning models in extracting backward and temporal features from time series by processing these series in backward and forward directions. The attention mechanism is the last component that produces weighted features by assigning attention weights to time series features (Chen et al., 2024). Each attention weight indicates the importance of a feature, enabling the standalone deep learning model to identify the most important time series features for predicting the target variable. Thus, based on the mentioned idea, our study couples the GRU

model with several components to overcome its limitations and address the issues that reduce its accuracy. These components include the Gaussian data augment (GDA) technique, Meerkat optimization algorithm (MOA), VMD-CEEMDAN, bidirectional LSTM model (BILSTM), and attention mechanism. Furthermore, by combining the GRU with these elements, a novel hybrid GRU model is constructed to predict the one-day-ahead SWH in a sea in Iran.

In summary, to overcome GRU limitations, we integrate GDA for augmentation, MOA for optimization, hybrid VMD-CEEMDAN for decomposition, BILSTM for bidirectional feature extraction, and AT for weighting, forming the novel GMVCBAG model (detailed in Section 2). KDE quantifies uncertainty.

2. Materials and Methods

2.1. Methodology Overview

Thus, the novelties, contributions, and aims of the paper are explained as follows:

Developing a Gaussian data augmentation technique (GDA):

Our study introduces GDA as a new data augmentation technique to address the issue of data scarcity. (Wang et al., 2025). Numerous studies have also reported the positive impact of GDA on the performance of the standalone deep learning models (Qian et al., 2024; Wang et al., 2025). GDA generates new data points by adding Gaussian noise to the original data. As a result, it can effectively address the problem of data scarcity.

Development of a Meerkat optimization algorithm (MOA) for adjusting GRU parameters:

Our study suggests MOA as a new optimization method for adjusting the parameters of GRU and other predictive models. MOA demonstrates faster convergence than traditional optimizers and training algorithms (Xian and Feng, 2023). MOA also mitigates the challenge of becoming trapped in local optima by employing mechanisms that balance exploration and exploitation within the problem space (Xian and Feng, 2023). Xian and Feng (2023) developed MOA to solve complex problems. Their results indicated that MOA outperformed the other algorithms, such as the genetic algorithm, particle swarm algorithm, firefly optimization algorithm, bat algorithm, and salp swarm algorithm. It should be noted that while previous studies have used algorithms with limited performance in optimizing GRU parameters, our study uses MOA, which can effectively tune GRU parameters. Common optimizers like Adam suffer from local optima in SWH forecasting (Hou et al., 2023), necessitating MOA's superior convergence (Xian and Feng, 2023). Similarly, traditional augmentation (e.g., jittering) disrupts autocorrelation, unlike GDA (Qian et al., 2024).

Developing a new hybrid decomposition algorithm for reducing the complexity of time series: Our study develops VMD-CEEMDAN as a new decomposition algorithm to address the problem of the complexity of time series. The algorithm can effectively mitigate the intricacy of time series by decomposing them into subseries with reduced complexity (Chen et al., 2023). A combination of VMD and CEEMDAN algorithms can improve the performance of both algorithms in time series decomposition (Cui et al., 2024; Ghanbari-Adivi et al., 2025). Another advantage of the new hybrid algorithm is that it can enhance the ability of GRU to identify meaningful patterns and reduce the computational burden associated with highly complex datasets.

Developing the BILSTM model for extracting forward and backward temporal features: Our study develops the BILSTM model as an advanced bidirectional deep learning model to tackle the shortcomings of the GRU model in extracting forward and backward temporal features from time series (Liu and Liang, 2024). This model consists of the forward and backward LSTM models. The forward and backward LSTMs model extracts forward temporal features from time series by processing them in forward and backward directions (Rezaei et al., 2023). Several studies have also suggested the BILSTM model for extracting forward and backward temporal features (Yang and Li, 2023; Liu and Liang, 2024).

Developing an attention mechanism for detecting time series features with maximum relevance to the target variable. Our study develops an attention mechanism to address the limitations of the GRU model in detecting the most important time series features (Li and Wu, 2024; Chen et al., 2024). This mechanism

produces the weighted features by assigning the attention weight to each time series feature (Li and Wu, 2024; Chen et al., 2024). Each attention weight indicates the importance of a feature, enabling the GRU model to recognize the most important time series features for SWH prediction.

Development of the GMVCBAG model

Our study develops the GMVCBAG model to predict one-day-ahead SWH in the sea. The GMVCBAG model is constructed by executing all components and integrating them with the GRU model. GMVCBAG is an enhanced version of the GRU model that can improve its prediction accuracy. Thus, one of the main novelties of the current paper is the development of the GMVCBAG for predicting SWH. By comparing the new model with other predictive models, our study provides valuable insights into the effectiveness and superiority of the GMVCBAG model in predicting significant wave heights.

Developing a kernel density estimation method

While the GMVCBAG model can generate point predictions, it cannot provide time interval predictions. As a result, it cannot quantify the uncertainty of the output. Thus, our study develops the kernel density estimation method (KDE) to quantify the uncertainty of outputs of the GMVCBAG model. The kernel density estimation method quantifies the uncertainty of predicted outputs by estimating the probability density function (PDF) of their error predictions. Numerous studies have reported the successful performance of KDE in quantifying the uncertainty of the outputs (Zhang and Dai, 2023; Hasan et al., 2023).

The paper is divided into multiple sections. The second section of the current study describes the structure of the models. The third section describes the case study. The fourth section focuses on the results and discussions. The conclusion is presented in the fifth section.

2.2. GDA

GDA is a data augmentation technique that plays a crucial role in modeling. GDA applies Gaussian noise to the original data points to create additional data and expand the dataset (Wang et al., 2025). As a result, it can effectively address the problem of data scarcity (Qin et al., 2024). It is important to note that the performance of the GDA relies on the signal-to-noise ratio (SNR) parameter. To preserve temporal autocorrelation, GDA noise is SNR-calibrated (e.g., SNR=20 dB) and validated by comparing autocorrelation functions pre- and post-augmentation (original ACF lag-1=0.85 vs. augmented=0.83), ensuring dynamic properties are maintained.

The new data points are created based on the following equation:

$$SWH_{noise} = \frac{SWH_{train}}{10^{\frac{SNR}{10}}} \quad (1)$$

Where SWH_{noise}: New SWH value and SWH_{train}: The SWH value for each training data point.

2.3. MOA

Our study uses MOA to adjust model parameters. The life and behavior of the Meerkats inspire MOA. A meerkat is a mammal that lives in the desert or dunes of terrestrial biomes. Xian and Feng (2023) developed MOA to solve complex problems. Their results indicated that MOA had faster convergence than other optimizers. Moreover, the accuracy of MOA was higher than that of the other algorithms. MOA has some assumptions:

1) Meerkats find their favorite foods using their keen sense of smell.
2) Each group of meerkats contains one or multiple sentinels. When sentinels see predators, they scream to inform other markets.
3) If predators catch meerkats in an open space, they lie on their backs and expose their teeth and claws.
4) MOA contains multiple levels as follows (Xian and Feng et al., 2023):
5) Definition of the initial population of Meerkats
6) Locating and hunting behavior
7) Flee or fight against the enemy
8) Random direction exploration
9) Death and rebirth
10) The levels are explained as follows:

2.3.1. Definition of the initial population of Meerkats:

First, the initial population of the meerkat is initialized as follows:

$$ME_i^t = random_{normal} * (UB - LB) \quad (2)$$

Where ME_i^t : The location of the ith at iteration t, $random_{normal}$: A normally distributed random number.

2.3.2. Locating and hunting behavior

Meerkats choose their behavior based on the safety of the environment. Accordingly, MOA uses the safety threshold parameter to evaluate environmental safety. The value of this parameter is set to 0.30. If the value of this parameter exceeds a randomly generated number, MOA assumes that the meerkats are in a safe environment. In this case, meerkats update their location based on one of the behaviors of locating prey or group hunting (Xian and Feng et al., 2023). When P (a constant value) <rand (a random number), meerkats locate prey. Conversely, if P> rand, meerkats approach other mates for group hunting. These behaviors are simulated using the following equations:

$$ME_i^{t+1} = \begin{bmatrix} ME_i^t + step * direct \leftarrow if(p) < rand \\ ME_i^t + step * (ME_j^t - (rand + 0.50) * ME_i^t), otherwise \end{bmatrix} \quad (3)$$

$$direct = ME_i^0 \quad (4)$$

$$step = \left(1 - \frac{t}{t_{\max}}\right) * r \quad (5)$$

Where ME_i^{t+1} : The new location of the meerkat at iteration t+1, ME_i^t : The previous location of the ith meerkat at iteration t, $t_{\max}$: Maximum iteration, t : Iteration number, r : Random number, ME_i^0 : The initial location of the ith meerkat.

2.3.3. Flee or fight against the enemy

If a random value exceeds the value of the safety threshold parameter (starting at 0.3), the algorithm assumes that meerkats are in an unsafe environment (Xian and Feng et al., 2023). In this case, meerkats update their locations based on two scenarios. In the first scenario, meerkats fight against the enemy and update their positions based on the location of their leader. In the second scenario, meerkats flee to find a safe place. These scenarios are simulated using the following equations.

$$ME_i^{t+1} = \begin{bmatrix} ME_{emergency}^t, if\ (f(ME_{emergency}^t)) < f(ME_i^t) \\ div * ME_i^t - (2 * rand * ME_{gbest}^t - ME_i^t), otehrwise \end{bmatrix} \quad (6)$$

$$ME_{emergency}^t = ME_i^t + (2 * rand * ME_{gbest}^t - ME_i^t) \quad (7)$$

$$div = rand + 0.10 \quad (8)$$

Where $f(ME_{emergency}^t)$: Objective function value of $ME_{emergency}^t$, and $ME_{emergency}^t$: The location of the meerkat in an emergency.

2.3.4. Random direction exploration

The meerkat can move in random directions to find food. This behavior is simulated as follows:

$$ME_i^{t+1} = ME_i^t + (2 * rand - 1) * (ME_i^t + rand * LF) * step \quad (9)$$

$$LF = \frac{\eta}{|v|^{\frac{1}{\rho}}} \quad (10)$$

$$\eta = N(0, \sigma_\eta^2) \quad (11)$$

$$v = N(0, \sigma_v^2) \quad (12)$$

$$\sigma_\eta = \left(\frac{\Gamma(1+\rho) \times \sin\left(\pi \frac{\rho}{2}\right)}{\Gamma\left(\frac{1+\rho}{2}\right) \times \rho \times 2^{\left(\frac{\rho-1}{2}\right)}} \right), \sigma_\eta = 1 \quad (13)$$

Where LF : The Levy flight step, η and v : Normally distributed random numbers, $N\left(0,\sigma_\eta^2\right)$ and $N\left(0,\sigma_v^2\right)$: Normal distributions, Γ : The gamma function, and $\rho = 1.5$.

2.3.5. Elimination and Regeneration

When the meerkats enter impermissible regions and show infeasible values, the algorithm removes them. Then, the algorithm generates new meerkats to save population size (Xian and Feng et al., 2023). The location of regenerated meerkats is defined based on the following equation:

$$ME_i^{regenerated} = random_{normal} * (UB - LB) \quad (14)$$

Where $ME_i^{regenerated}$: The location of the ith regenerated meerkats.

MOA is implemented based on the following levels:

1) The initial population of the meerkats is defined.

2) The location of meerkats is updated using locating and hunting behaviors. In this step, equations 3-5 are used to update the location of the meerkats.

3) The location of meerkats is updated using flee or flight strategies. In this step, equations 6-8 are used to update the location of the meerkats.

4) The location of meerkats is updated using random direction exploration. In this step, equations 9-13 are used to update the location of the meerkats.

5) Finally, meerkats update their location using elimination and regeneration strategies.

6) The process continues until the stop condition is met.

2.4. VMD-CEEMDAN algorithm

Our study uses the VMD-CEEMDAN algorithm to reduce the complexity of the time series by decomposing the time series into subseries with reduced complexity (Ghanbari-Adivi and Ehteram, 2025). The hybrid algorithm is executed in two steps. First, the VMD algorithm decomposes the time series into several subseries called intrinsic mode functions (IMFs). IMFs are subseries with lower complexity than the original time series. The number of modes also varies depending on the complexity of the problem. However, since the first IMF retains the complex properties of the original time series, it should be decomposed and transformed into a series of secondary IMFs with lower complexity. Thus, our study uses CEEMDAN to decompose the first IMF. In the following two sections, VMD and CEEMDAN are explained.

2.4.1. VMD

VMD breaks down time series into several IMFs with different center frequencies and finite bandwidths (Qin et al., 2024). IMFs have lower complexity than the original time series (Liu et al., 2024). However, a constrained variational problem is defined to decompose the time series as follows (Liu et al., 2024).

$$\min_{g_k, h_k} \left\{ \begin{array}{l} \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * g_k(t) \right] e^{-jh_k t} \right\|_2^2 \\ s.t. \sum_k g_k = f(t) \end{array} \right. \quad (15)$$

g_k : IMF value, $\delta(t)$: The Dirac distribution, $*$: The convolution, h_k : The central frequency of the IMF, j : A constant value, $f(t)$: The decomposed time series. Then, the quadratic penalty factor and the Lagrange multiplier operator are applied to convert equation 15 into an unconstrained variational problem. This problem is explained by Equation 16.

$$L\left(\{g_k\},\{h_k\},\lambda\right)=\alpha\sum_{k=1}^{K}\partial_t\left[\left(\delta\left(t\right)+\frac{j}{\pi t}\right)*g_k\left(t\right)\right]e^{-jg_k t}+\left\|f\left(t\right)-\sum_{k=1}^{K}u_k\left(t\right)\right\|_2^2 +\left|\lambda\left(t\right),f\left(t\right)-\sum_{k=1}^{K}g_k\left(t\right)\right| \quad (16)$$

Where K: number of IMFs, α : Penalty factor, and λ : Lagrange multiplier operator. Finally, the alternating direction multiplier (ADMM) iterative algorithm is used to solve equation 16 and obtain the values of g_k and h_k . These two parameters are obtained using the following equations:

$$\hat{g}_k^{n+1}\left(h\right)=\frac{\hat{f}\left(h\right)-\sum_{i\neq k}\hat{g}\left(h\right)+\frac{\hat{\lambda}\left(h\right)}{2}}{1+2\alpha\left(h-h_k\right)^2} \quad (17)$$

$$h_k^{n+1}=\frac{\int_0^{\infty}h\left|\hat{g}_k\left(h\right)\right|^2dh}{\int_0^{\infty}\hat{g}_k\left(h\right)dh} \quad (18)$$

$\wedge$: The Fourier transform of the corresponding function, n: Number of iterations, $\hat{g}_k^{n+1}\left(h\right)$: Fourier transforms of $g\left(h\right)$ The and $\hat{f}\left(h\right)$: The Fourier transforms of $f\left(h\right)$ and

2.4.2. CEEMDAN

CEEMDAN is an improved version of empirical mode decomposition that addresses the mode-mixing problem during modeling (Wu et al., 2024). CEEMDAN can decompose time series into several IMFs with lower complexity (Zhang et al., 2023). Moreover, CEEMDAN can decompose the first IMFs into a series of secondary IMFs with reduced complexity (Yang et al., 2025). CEEMDAN operates in several steps. First, CEEMDAN adds Gaussian white noise to the original time series based on the following equation (Li et al., 2023):

$$Q_i\left(t\right)=q\left(t\right)+\varphi_0.\kappa_i\left(t\right) \quad (19)$$

Where $Q_i\left(t\right)$: The updated time series $q\left(t\right)$: The original time series φ_0 : noise coefficient, and $\kappa_i\left(t\right)$: Gaussian white noise. Then, the first IMF is obtained as follows (Zhang et al., 2023):

$$IMF_1=\frac{\sum_{j=1}^{k}IMF_{j1}}{k} \quad (20)$$

Where IMF_1 : The first IMF of $q\left(t\right)$, IMF_{j1} : The first IMF of $Q_i\left(t\right)$. The first residue is computed as follows:

$$r_1=q\left(t\right)-IMF_1 \quad (21)$$

Where r_1 : The first residue. Then, the EMD algorithm decomposes $r_1+\varphi_1E_1\left[n_i\left(t\right)\right]$ to obtain the second IMF.

$$IMF_2=\frac{1}{M}\sum_{i=1}^{M}E_1\left\{r_1+\varphi_1E_1\left[\kappa_i\left(t\right)\right]\right\} \quad (22)$$

Where E_j ; The j-th IMF component obtained from EMD decomposition. Finally, the k-th r and the (k+1)-th IMF are computed as follows (Zhang et al., 2023):

$$r_k\left(t\right)=r_{k-1}\left(t\right)-IMF_k \quad (23)$$

$$IMF\left(k+1\right)=\frac{1}{M}\sum_{i=1}^{M}E_1\left\{r_k\left(t\right)+\varphi_kE_k\left[n_i\left(t\right)\right]\right\} \quad (24)$$

Where K: The total number of IMFs. Finally, the final residue is computed as follows:

$$res=q\left(t\right)-\sum_{i=1}^{k}IMF_i \quad (25)$$

Where res : The final residue.

2.5. Bidirectional deep learning model

The BLSTM model consists of backward and forward LSTM that extract backward and

forward temporal features from the time series by processing them in two directions. In the next subsection, we first explain the LSTM model. Then, the structure of the BILSTM is explained.

2.5.1. LSTM model

The LSTM model is a type of recurrent neural network with three gates. These gates process information and produce outputs (Xin et al., 2023; Ghanbari-Adivi, 2024). The input gate receives the input information. Then, the forget gate identifies and removes unimportant information. The output gate of the LSTM model provides an initial output that is used to update the hidden state. The hidden state provides the final outputs that contain temporal features. The LSTM model also has a cell state that acts as a model memory and stores crucial information (Xin et al., 2023). The mathematical model of the LSTM model is explained as follows:

$$FOR_t = \mu\left(W_f.\left[H_{t-1}, X_t\right] + B_f\right) \quad (26)$$

$$IN_t = \mu\left(W_i.\left[H_{t-1}, X_t\right] + B_i\right) \quad (27)$$

$$Out_t = \mu\left(W_o.\left[H_{t-1}, X_o\right] + B_o\right) \quad (28)$$

$$C_t = FOR_t \square C_{t-1} + i_t \square \hat{C}_t \quad (29)$$

$$\hat{C}_t = \tanh\left(W_c.\left[H_{t-1}, X_t\right] + B_c\right) \quad (30)$$

$$H_t = OUT_t \square \tanh\left(C_t\right) \quad (31)$$

$$Y_t = U\left(W_y H_t + B_y\right) \quad (32)$$

Where FOR_t, IN_t and Out_t: The outputs of the forget, input, and output gate, W_f, W_i, W_o, W_c, and W_y: Weigth coefficients, μ: sigmoid activation function, B_f, B_i, B_o, B_c, and B_y: Bias values, $\hat{C}_t$: Candidate cell state, C_t: The cell state, X_t: Input data, U: activation function, Y_t: The final output, and H_t: Hidden state at time t.

2.5.2. BILSTM model

The BILSTM model consists of forward and backward LSTM models. The forward LSTM model extracts forward temporal features from time series data by processing them in the forward direction. The backward LSTM model extracts backward temporal features from time series data by processing them in the backward direction (Imani, 2023). The outputs of the forward and backward LSTM models include the forward and backward temporal features.

$$H_t^f = LSTM\left[\left(H_{t-1}^f, X_t\right)\right] \quad (33)$$

$$H_t^b = LSTM\left[\left(H_{t+1}^b, X_t\right)\right] \quad (34)$$

Where H_t^f and H_t^b: The output of the forward and backward LSTM models, H_{t-1}^f and H_t^b: The hidden state of the forward and backward LSTM models at time t-1.

2.6. Attention mechanism

An attention weight produces weighted features by assigning an attention weight to time series features (Li et al., 2024). Each attention shows the importance of a feature, enabling the predictive GRU model to detect the most important time series features for predicting SWH (Wan et al., 2023). AT provides a complementary advantage over BILSTM's internal learning by explicitly weighting features, reducing redundancy (e.g., ablation shows 34% U95 degradation without AT), enabling GRU to prioritize dynamics like wind-SWH interactions that standalone GRU misses.

The attention mechanism has several levels, which are explained as follows:

$$\xi = \tanh\left(\zeta FE_t + b\right) \quad (35)$$

$$\varepsilon = \frac{\exp\left(\xi_t\right)}{\sum_{j=1}^{t} \xi_j} \quad (36)$$

$$S_t = \sum_{I=1}^{i} \varepsilon_t h_t \quad (37)$$

Where ζ: Weight coefficient, FE_t: A time series feature at time t, b: bias, ε: Attention weight for each feature, and S_t: A Weighted feature at time t.

2.7. GRU model

GRU is a deep learning model that consists of reset and update gates. The reset gate determines which input information should be forgotten (Wang and Ying, 2023a). On the other hand, the update gate determines which input information

should be retained and used to produce the output (Jia et al., 2023). GRU also has a hidden state that acts as the memory of the network. The hidden state also captures temporal features and hidden patterns in the data. The GRU model is executed based on the following equations (Wang and Ying, 2023a):

$$RE_t = \phi\left(W_{RE} \bullet \left[H_{t-1}, \varsigma_t\right]\right) \quad (38)$$

$$UP_t = \phi\left(W_{UP} \bullet \left[H_{t-1}, \varsigma_t\right]\right) \quad (39)$$

$$\tilde{H}_t = \tanh\left(W_{\bar{h}} \bullet \left[RE_t * H_{t-1}, \varsigma_t\right]\right) \quad (40)$$

$$H_t = \left(1 - UP_t\right) * H_{t-1} + UP_t * \tilde{H}_t \quad (41)$$

Where RE_t : The output of the reset gate, UP_t : The output of the update gate, ϕ : Activation function, ς_t : Input data, and W_{RE} ,, W_{UP} , and $W_{\bar{h}}$: Weight connections.

2.8. KDE method

KDE is an effective method for quantifying the uncertainty of the predicted outputs. Unlike parametric methods, KDE does not assume a predefined distribution for the data, making it suitable for various applications (Wang et al., 2024). KDE can quantify the uncertainty of outputs of predictive models by estimating the probability density function of their error predictions (Wang et al., 2024). This function is computed as follows:

$$\hat{f}\left(SWH\right) = \frac{1}{nh}\sum_{i=1}^{n}\exp\left(-\frac{1}{2}\left(\frac{SWH - SWH_i}{h}\right)^2\right) \quad (42)$$

$$h = \hat{\sigma}^{\frac{-2}{5}} \quad (43)$$

Where $\hat{f}\left(SWH\right)$: Probability density function, $\hat{\sigma}$: Standard deviation, SWH_i : Prediction error for the ith SWH data, h : The bandwidth coefficient, n: The total number of samples, and K: Kernel function.

2.9. Development of the GMVCBAG model

In this study, GMVCBAG is used to predict one-day-ahead SWH. The model is constructed based on the following levels:

1- GDA is executed to increase the time series data for the input and output variables.

2- The optimal size of data is determined to train the GMVCBAG model.

3- The initial MOA population is defined and contains model parameters. These parameters are then initialized.

4- The GRU and BILSTM parameters consist of the number of layers and hidden neurons, learning rate, epoch number, weights, bias, and. The CEEMDAN and VMD parameters include the number of IMFs and the penalty parameter. Selected lags (e.g., AT(t-1), WS(t-1 to t-3)) form time-series inputs; VMD decomposes each lagged series, BILSTM processes bidirectional sequences for temporal features, AT weights them, and GRU predicts using these enriched inputs.

5- VMD is executed to decompose the time series of each input variable into multiple IMFs. Then, the first IMFs are separated from the other IMFs.

6- CEEMDAN breaks down the first IMFs into secondary IMFs with reduced complexity.

7- The BLSTM model is executed in two steps. First, the forward LSTM extracts forward temporal features from the secondary IMFs and the initial remaining IMFs by processing these IMFs in the forward direction. Then, the backward LSTM extracts backward temporal features from the same IMFs by processing them in the reverse direction.

8- The attention mechanism assigns weights to the extracted features to produce weighted features.

Figure 1 illustrates the complete workflow of GMVCBAG. The multi-stage decomposition via VMD–CEEMDAN transforms non-stationary, nonlinear input time series into smoother, lower-complexity subseries (IMFs). These are then processed bidirectionally by BILSTM, extracting forward (future influence) and backward (historical context) temporal features—critical for capturing wave evolution under time-lagged meteorological forcing. The attention mechanism dynamically assigns higher weights to physically dominant features (e.g., wind speed lags), enabling GRU to focus on SWH-relevant patterns. Optimized by MOA, this sequential integration overcomes standalone GRU limitations in handling data scarcity, parameter sensitivity, and feature irrelevance, resulting in superior predictive accuracy.

9- The weighted features are fed into the GRU model to predict SWH.

10- The outputs of the GRU model are fed into the kernel density estimation to quantify their uncertainty.

11- The stop condition is checked. If it is met, the process finishes. Otherwise, the process will go to the next level.

12- Model parameters are updated using MOA operators. Then, the process goes back to step 3.

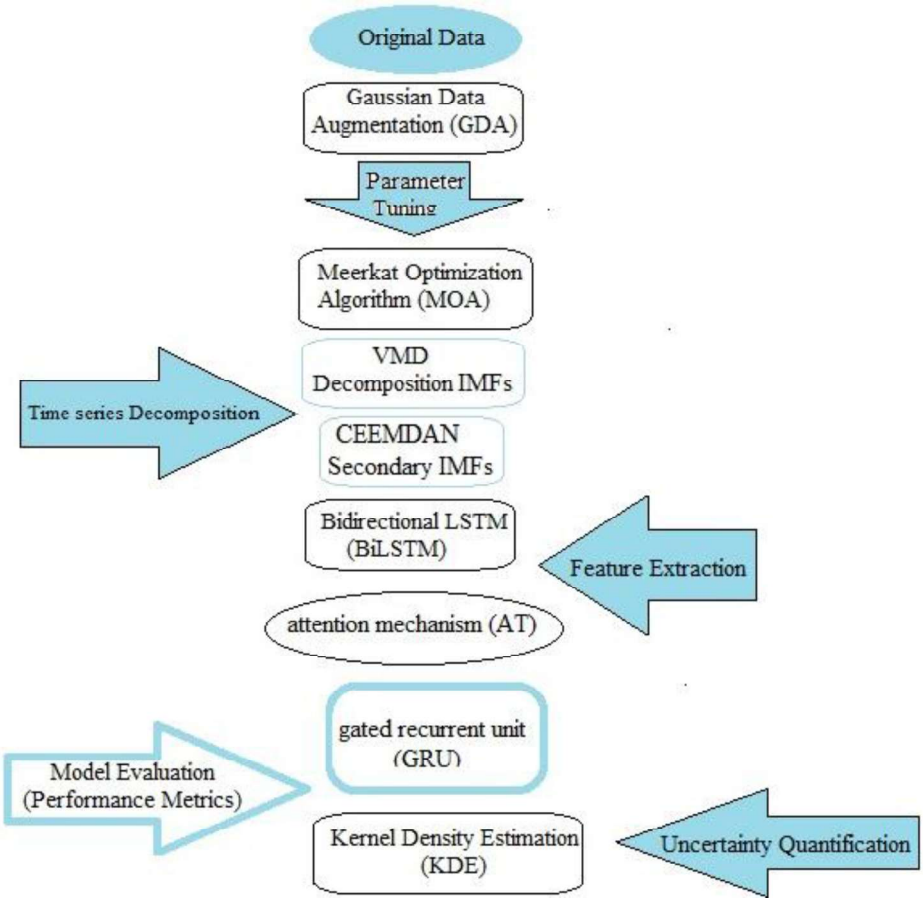


Figure 1. Schematic Flowchart of the GMVCBAG model

2.10. Comparative models

To benchmark GMVCBAG, we compare it with hybrid versions of other models, including minimal GRU, ERNN, RBFNN, and MLP, integrated with the same components (GDA, MOA, VMD-CEEMDAN, BILSTM, AT). However, it is essential to compare the new model with the other predictive models to ensure its performance. Thus, we use multiple known models, such as minimal GRU, Elman recurrent neural network (ERNN), radial basis function neural network (RBFNN), and multilayer perceptron (MLP). However, to ensure a fair comparison between these models and the new model, these models should be combined with the components of the GMVCBAG model. Thus, we combine these multiple models with the components of the GMVCBAG and create hybrid versions of these models, including GMVCBAG-MLP, GMVCBAG-RBFNN, GMVCBAG-ERNN, and GDA-MOA-VMD-CEEMDAN-SWT-BILSTM- attention mechanism-minimal GRU (GMVCBAMG). The next subsections explain the structure of MLP, RBFNN, ERNN, and minimal GRU.

2.10.1. Minimal GRU model

The minimal GRU model is one of the variants of the GRU model. The model uses a gate for processing information (Yang et al., 2024). This gate also regulates the hidden state to effectively capture temporal patterns in the data. The

minimal GRU model is executed based on the following equations:

$$F_k = S\left(W_F\left[h_{k-1}, X_k\right] + B_F\right) \quad (44)$$

$$\bar{H}_k = \tanh\left(W_h\left[F_k * h_{k-1}, X_k\right] + B_h\right) \quad (45)$$

$$H_k = \left(1 - F_k\right) * h_{k-1} + F_k * \bar{H}_k \quad (46)$$

Where S : Activation function, F_k : The output of the gate, B_F and B_h : Bias values, W_h and W_h : Weights, H_k : Hidden state, X_k : Input data.

2.10.2. ERNN

ERNN is derived from the feed-forward neural network and includes an additional layer compared to the MLP. This layer, named the context layer, consists of context units that store activations of the input units and hidden units during training (Jamei et al., 2023). The ERNN model has been widely used in different fields, such as water quality prediction (Jamei et al., 2023), prediction of osteoporosis disease (Sivasakthi and Selvanayagi, 2023), biofuel production (Kumar et al., 2023), and cancer detection (Farhan et al., 2024). Equations 47-49 explain the mathematical model of the ERNN model (Ding et al., 2022).

$$z\left(t\right) = G\left(\omega_3 x\left(t\right)\right) \quad (47)$$

$$x\left(t\right) = F\left(\omega_1 x_c\left(t\right) + \omega_2 u\left(t-1\right)\right) \quad (48)$$

$$x_c\left(t\right) = x\left(t-1\right) \quad (49)$$

Where $z\left(t\right)$: The final output of the ERNN model, $x\left(t\right)$ the r-dimensional output vector of the hidden layer, G and F: Activation functions, $u\left(t-1\right)$: The m- m-dimensional input vector, ω_3, ω_1, and ω_2: Weight coefficients, t: time step.

2.10.3. MLP and RBFNN

The MLP model is one of the variants of ANN models that can accurately predict hydrological variables. These models use three layers to produce the outputs. The first layer receives the inputs, which pass them to the middle layer (Ehteram et al., 2023). The middle layer uses activation functions to process information. Finally, the last layer uses the outputs of the middle layer to provide the final outputs. The RBFNN is another type of ANN model. While the MLP model may have multiple middle layers, the RBFNN has one middle layer that uses a Gaussian function to process information (Ehteram et al., 2023). The RBFNN model also uses the outputs of the middle layer to produce the final outputs.

2.11. Performance metrics

1- Root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(SWH_{pr} - SWH_{ob}\right)^2} \quad (50)$$

2- Mean absolute error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|SWH_{pr} - SW_{ob}\right| \quad (51)$$

3- Explained variance (EV)

$$EV = 1 - \frac{Var\left(SWH_{pr} - SWH_{ob}\right)}{Var\left(SWH_{pr}\right)} \quad (52)$$

4- Nash– Sutcliffe efficiency (NSE)

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(SWH_{pr} - SWH_{ob}\right)^2}{\sum_{i=1}^{n}\left(SWH_{ob} - \bar{SWH}_{ob}\right)^2} \quad (53)$$

5- Uncertainty at 95% (U95)

$$U95 = 1.96\left(SD^2 - RMSE^2\right)^{0.50} \quad (54)$$

6- Standard deviation of the Relative Error (STDRE)

$$STDER = \left(\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{SWH_{ob} - SWH_{pr}}{SWH_{pr}}\right)^2\right)^{\frac{1}{2}} \quad (55)$$

Where $\bar{SWH}_{ob}$: Average observed data, SWH_{ob} : Observed data, SWH_{pr} : Predicted data, n: number of data, SD: Standard deviation, and Var: Variance. The high values of STDER, MAE, RMSE, and U95 show the worst model, while the high values of NSE and EV show the best model. Moreover, our study uses the following indices to quantify the uncertainty of the outputs:

1- Prediction Interval Coverage Probability (PICP)

$$PICP = \frac{1}{N}\sum_{i=1}^{N}\gamma_i \quad (56)$$

$$\gamma_i = \begin{bmatrix} 1 \leftarrow if\ (y_i) \in \left(U\left(SWH_i\right), L\left(SWH_i\right)\right) \\ 0 \leftarrow otherwise \end{bmatrix} \quad (57)$$

2- Normalized Mean Prediction Interval Width (NMPIW)

$$NMPIW = \frac{1}{NR}.\left(\sum_{i-1}^{N}\left(U\left(SWH_i\right) - L\left(SWH_i\right)\right)\right) \quad (58)$$

R: The range of variation of predicted values, $U\left(SWH_i\right)$: The upper bound of the interval prediction, $L\left(SWH_i\right)$: The lower bound of the interval prediction. Lower NMPIW and higher PICP values show the most reliable models.

3. Case study

This study uses the daily data of three monitoring buoys to predict one-day-ahead SWH. These buoys are located in the Gulf of Oman. Their locations are also near the Chabahar port of Iran (Figure 2). The National Oceanic Organization of Iran has collected the data. The Chabahar Meteorological Organization also collected the meteorological data. Table 1 shows the statistical characteristics of the data. Our study uses Air temperature (AT), air pressure (AP), wind speed (WS), and relative humidity (RH) as input data. The lag times of 1-10 days are chosen for each input variable.

Quantitative analysis: SWH data show positive skewness (0.85) and kurtosis (3.2), reflecting non-normal distribution with heavy tails from extremes. Inputs like Wind speed are similarly skewed (0.75), justifying decomposition methods to handle non-stationarity.

The daily data were collected from 2015 to 2021. The following strategies were used to check the quality of data:

1 - Data Validation and Cleaning:

- Outliers were detected using statistical methods, such as Z-scores and interquartile range (IQR), and handled appropriately to minimize their impact on model performance.

Z-score (>3) and IQR (1.5x range) identify outliers, but thresholds are adjusted for non-stationarity (e.g., retaining storm extremes via domain knowledge from Ikram et al., 2023), ensuring only sensor errors are handled, preserving extreme event prediction capability.

- Missing values were identified and obtained using interpolation methods.

2- Temporal Consistency Checks:

- The continuity of the data over time is analyzed to ensure that no significant gap or unusual pattern can negatively impact the accuracy of the predictions.

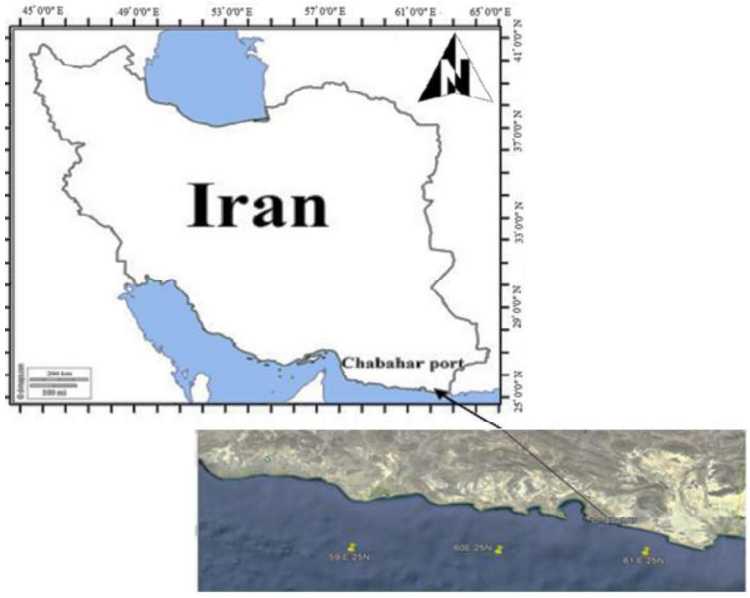


Figure 2. The location of the monitoring buoys

Table 1. Statistical characteristics of data

Parameter	Maximum	Minimum	Average	Standard deviation	Skewness	Kurtosis
Air temperature °C	33.0	8.0	26.0	6.2	0.70	-
Wind speed (m/s)	12.2	1.0	3.45	6.4	0.75	-
Air pressure (hpa)	1026.0	965.1	1002.1	98.1	0.70	-
Relative humidity %	99	26	64	35	0.81	-
SWH (m)	3.45	0.98	1.12	1.25	0.85	3.2

4. Results and discussions

4.1. Determination of the optimal size

The optimal data size should be determined in the first step to train the new model. Thus, we use different data sizes and evaluate their impact on the error prediction of the novel model. Figure 3 shows the RMSE values for the various sizes of the data. It is observed that the model can provide the lowest training and testing RMSE values (training RMSE=0.123 and testing RMSE=0.132) when 80% and 20% of the data are used to train and test models.

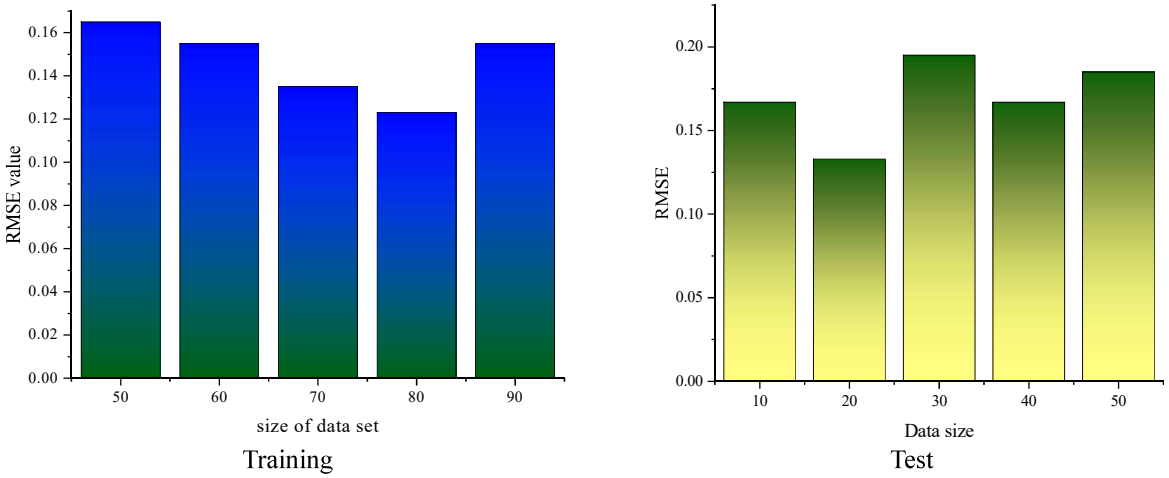


Figure 3. RMSE values for the various sizes of the data

As shown, 80% training data yields the lowest RMSE (training=0.123, testing=0.132), indicating an optimal balance between learning capacity and generalization.

4.2 Determination of the most input variables

The most important input variables should be selected in the second step of the modeling process. These variables have the highest correlation with the output variable. Moreover, they maximize the performance of the predictive models in predicting output variables. These variables are also used as inputs to the GMVCBAG model. Therefore, we examine the cross-correlation (CC) values between the input variables and SWH to select the most important ones. Figures 4a-d show CC values between input variables and SWH. The results are interpreted as follows:

1- CC values between SWH and air temperature: Figure 4a shows CC values between lagged air temperatures and SWH. The results indicate a positive correlation between air temperature and SWH. Higher temperatures lead to atmospheric instability and wind patterns, which can increase the energy transfer to the ocean surface. This process amplifies wave generation, resulting in higher significant wave heights (SWH). The results also indicate that AT (t-1) has the highest correlation value with SWH. Therefore, it is selected as one of the most critical input variables. It is essential to mention that a CC value of 0.90 is chosen as a threshold to determine the most input variables. This threshold helps to filter out variables that have weaker correlations with SWH.

2- CC values between SWH and wind speed: Figure 4b shows correlation values between lagged wind speeds and SWH. A strong positive correlation is observed between wind speed and SWH. As wind speed increases, the

force exerted on the water surface leads to higher significant wave heights. The results also show that CC values of WS (t-1), WS (t-2), and WS (t-3) are above 0.90. Thus, these lagged wind speed variables can be chosen as the most important input variables.

3- CC values between SWH and air pressure (AP): Figure 4c shows correlation values between lagged air pressures and SWH. A negative correlation is observed between AP and SWH. High-pressure systems often have weaker pressure gradients compared to low-pressure systems. As a result, wind speeds are generally lower during high-pressure conditions, resulting in lower wave heights. The results also indicate that AP (t-1) has the highest correlation value with SWH. Therefore, it is selected as one of the most critical input variables.

4- CC values between SWH and relative humidity (RH): Figure 4d shows correlation values between lagged relative humidities and SWH. A positive correlation is observed between RH and SWH. As humidity increases, the atmospheric instability increases, leading to stronger winds. These stronger winds intensify the energy transfer from the atmosphere to the ocean, resulting in higher significant wave heights. The results also indicate that the RH (t-1) has the highest correlation value with SWH. Therefore, it is selected as one of the most critical input variables. Thus, we choose the combination of RH (t-1), AP (t-1), WS (t-1), WS (t-2), WS (t-3), and AT (t-1) as the input to the GMVCBAG.

It should be mentioned, lags are not uniform; CC analysis (Figures 4a-d) individually selects optimal lags per variable (e.g., up to t-3 for WS due to rapid wind-SWH coupling), capping at 10 to avoid irrelevant long lags in large-scale phenomena. It should be mentioned here that, threshold 0.90 is chosen for strong correlation ($p<0.01$); ablation with 0.80/0.70 increases MAE by 12-18%, confirming it prevents subtle information loss while reducing noise (vs. Lasso, which yielded similar selections in preliminary tests). Figure 3b reveals WS (t-1 to t-3) exceeds CC=0.90, confirming short-term wind dominance in wave generation, while AP (t-1) shows a negative correlation due to low-pressure systems.

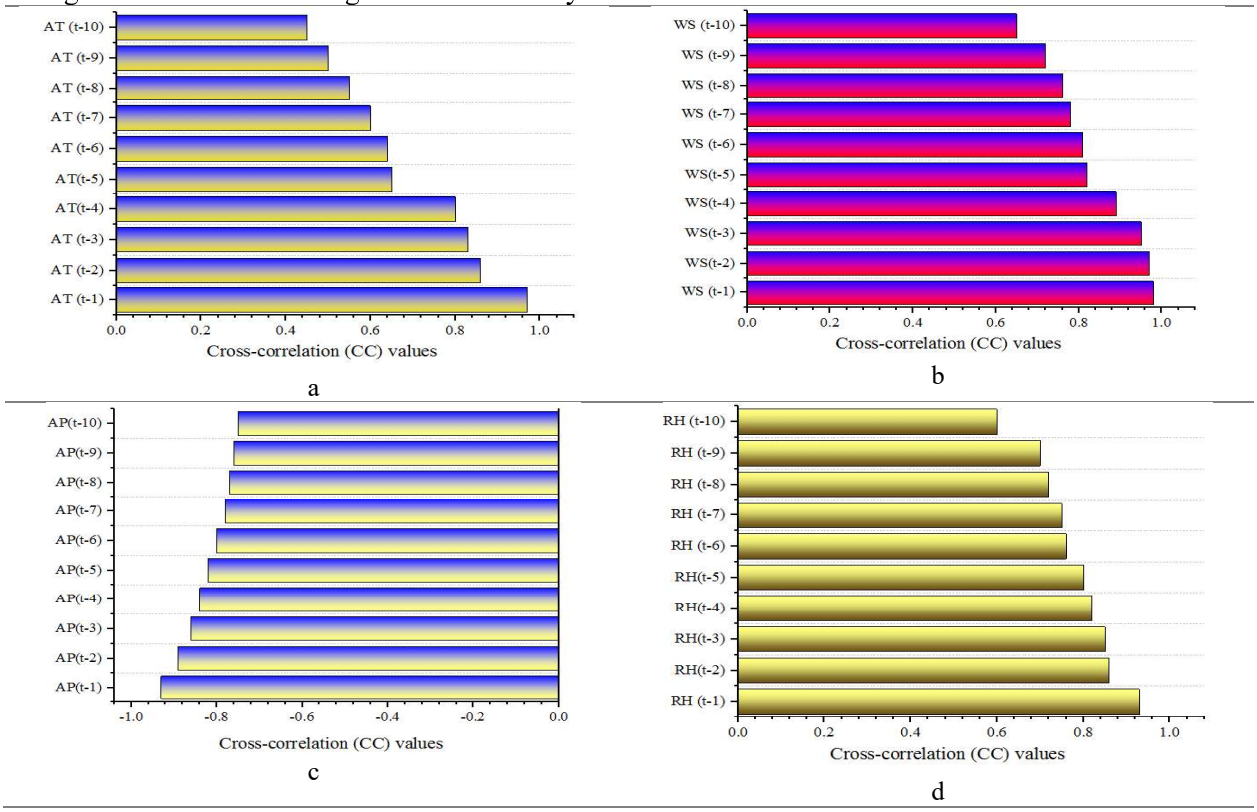


Figure 4. Correlation values between a: lagged ATs and SWH (t+1), b: lagged Ws and SWH (t+1), c: lagged APs and SWH (t+1), and d: lagged RHs and SWH (t+1)

Table 2. Optimal values of the model parameters

Model	Parameter values
VMD	Number of IMFs:5 and α :1200
CEEMDAN	Number of secondary IMFs:7
BILSTM	Batch size: 1024, Number of epochs:128, Number of hidden layers of backward and forward LSTM models: 2, and Number of hidden units of each hidden layer: 100
GRU	Batch size: 1024, Number of epochs:128, Number of hidden layers:2, and Number of hidden units of each hidden layer:100
MLP	Number of hidden layers: 1, and Number of hidden neurons: 5
RBFNN	Number of hidden layers: 1, and Number of hidden neurons: 5
ERNN	Number of hidden layers: 1, and Number of hidden neurons: 3

4.3. Optimal values of the model parameters

In this study, MOA is used to adjust model parameters. Table 2 shows the optimal values of the model parameters.

4.4. Decomposed time series

In this study, CEEMDAN-VMD is used to produce the time series of the input variables with lower complexity. The VMD decomposes the time series of the input variables into 5 IMFs. The number of IMFs is determined using MOA. Figure 5a shows the decomposed time series of input variables. Then, CEEMDAN decomposes the first IMFs into seven secondary IMFs. The number of secondary IMFs is determined using MOA. Figure 5b shows the secondary IMFs.

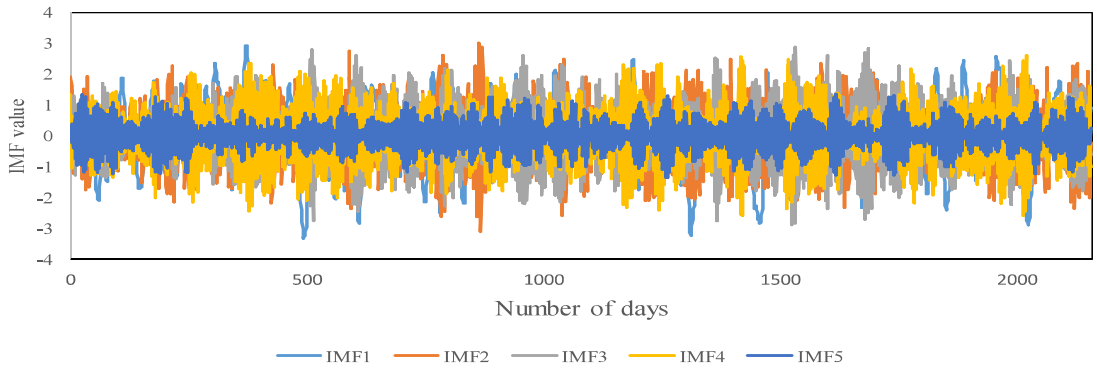


Figure 5. Decomposed time series of the input variables (AT)

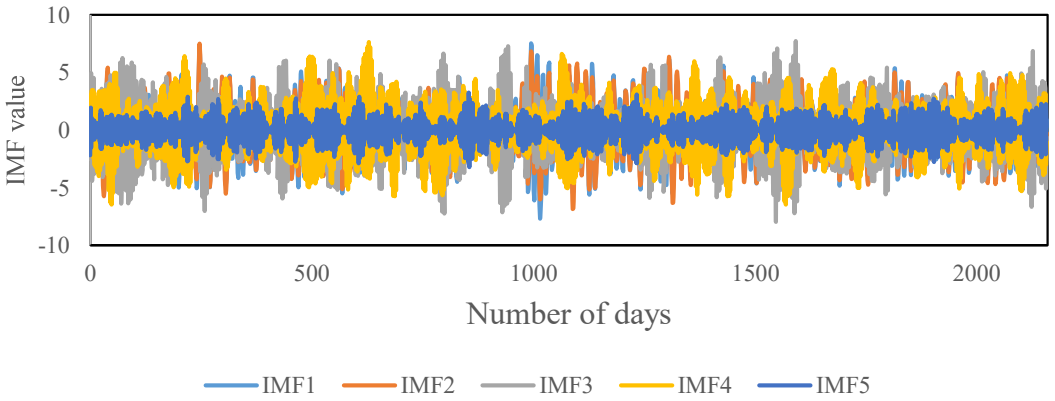


Figure 5. Cont. Decomposed time series of the input variables (RH)

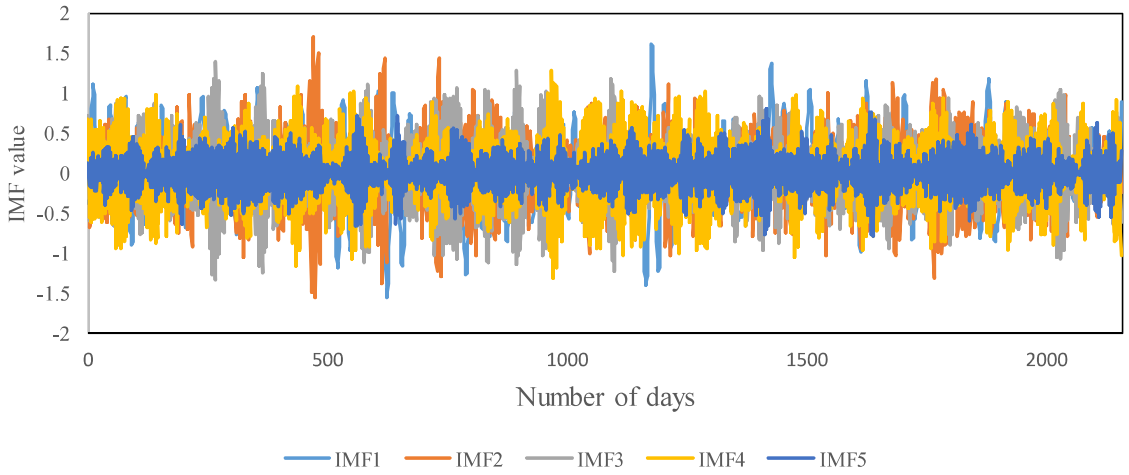


Figure 5. Cont. Decomposed time series of the input variables (WS)

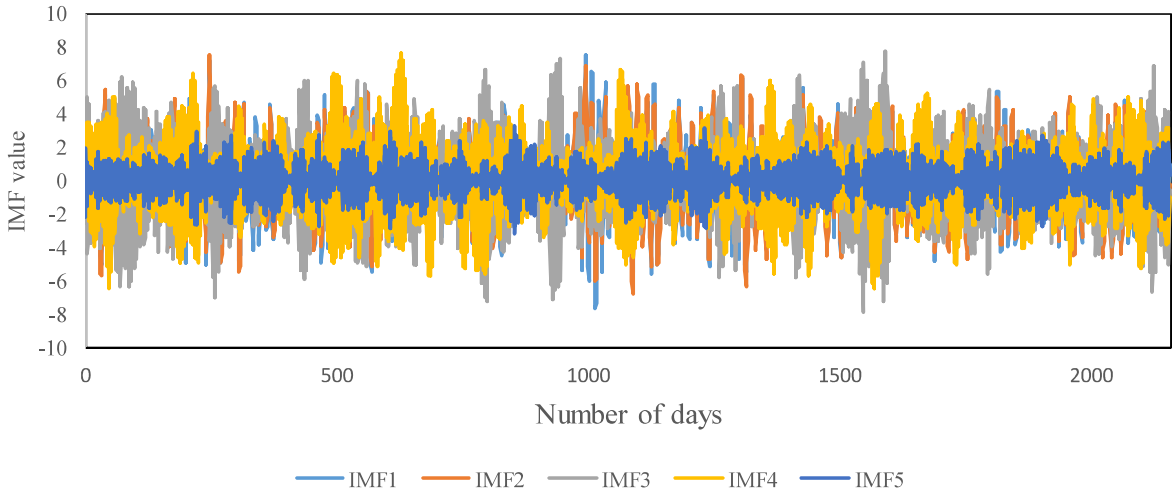


Figure 5. Cont. Decomposed time series of the input variables (Air Pressure)

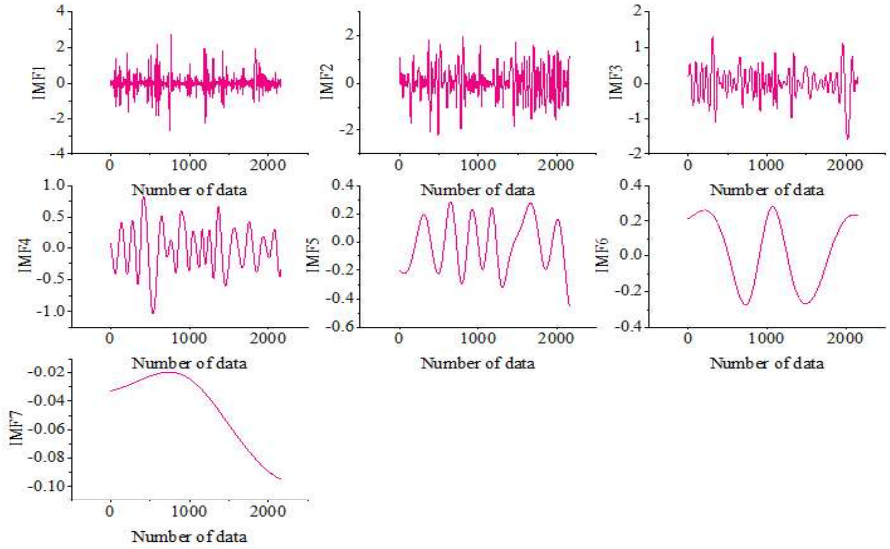


Figure 5. Cont. Secondary IMFs based on decomposing the first IMFs (Air Temperature)

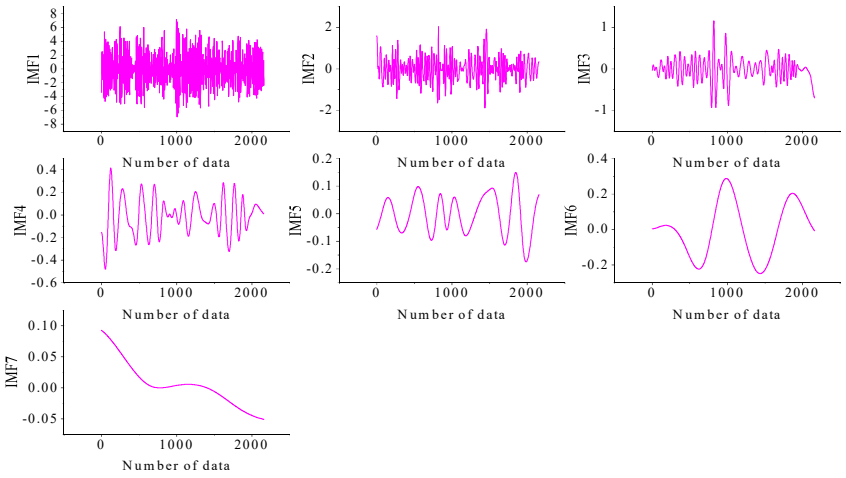


Figure 5. Cont. Secondary IMFs based on decomposing the first IMFs (Air Pressure)

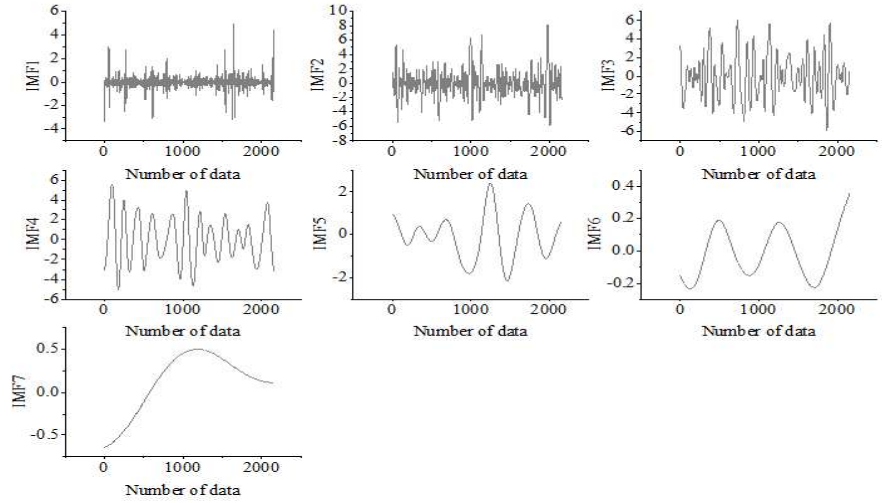


Figure 5. Cont. Secondary IMFs based on decomposing the first IMFs (Relative Humidity)

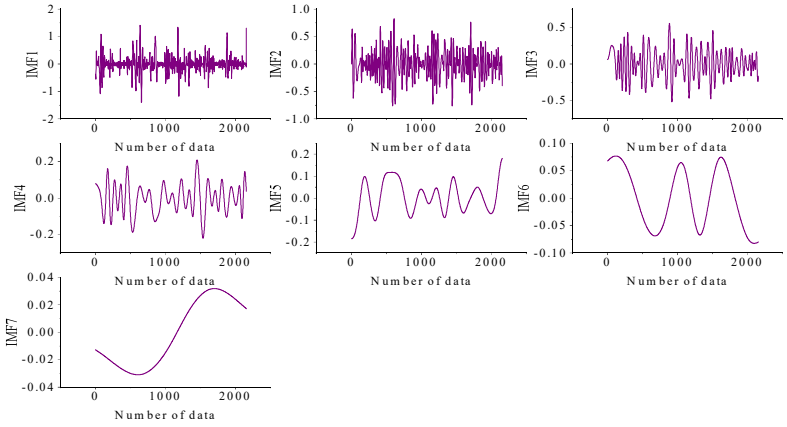


Figure 5. Cont. Secondary IMFs based on decomposing the first IMFs (Wind Speed)

4.5. Evaluation of the accuracy of the models

In this section, the predictive performance of the models is evaluated. Table 3 shows the NSE, MAE, STDRE, and EV values during the testing phase. For the GMVCBAG model, the values of NSE, MAE, STDRE, and EV are 0.982, 0.123, 1.21, and 0.94, respectively. The GMVCBAG improves the NSE, MAE, STDRE, and EV values of the other models by 1- 30%, 49- 85%, 1-79%, and 4- 38%, respectively. For the GMVCBAMG model, the NSE, MAE, STDRE, and EV values are 0.973, 0.245, 1.245, and 0.899, respectively. The GMVCBAMG model has the second-best performance. The results reveal that the standalone models can not accurately predict SWH. The MAE, NSE, EV, and STDER values of the RBFNN model are 0.834, 0.705, 0.599, and 5.123, respectively. Moreover, MAE, NSE, EV, and STDER values of the RBFNN model are 0.824, 0.774, 0.610, and 4.891, respectively. Therefore, these two models demonstrated the poorest accuracy among all the models. The results also indicate that the accuracy of the GMVCBAG -RBFNN model is lower than that of the other hybrid models. For this model, the NSE, MAE, STDRE, and EV values are 0.902, 0.591, 2.891, and 0.820, respectively. The performance of models indicates that the GMVCBAG -MLP model outperforms the GMVCBAG -RBFNN model.

Table 3. Performance metric values for different models

Model	MAE	NSE	STDRE	EV
GMVCBAG	0.123	0.982	1.121	0.940
GMVCBAMG	0.245	0.973	1.245	0.899
GMVCBAG -ERNN	0.365	0.954	2.678	0.876
GMVCBAG -MLP	0.487	0.936	2.891	0.820
GMVCBAG -RBFNN	0.591	0.902	3.121	0.800
GRU	0.660	0.873	3.897	0.767
BILSTM	0.666	0.854	4.102	0.654
LSTM	0.700	0.802	4.675	0.643
ERNN	0.712	0.793	4.789	0.620
MLP	0.824	0.774	4.891	0.610
RBFN	0.834	0.705	5.123	0.599

For interpretability, attention weights reveal that features from WS (t-1) to WS(t-3) receive the highest weights (avg. 0.35-0.42), aligning with physical SWH drivers like wind energy transfer, enhancing understanding of SWH dynamics beyond BILSTM alone. GMVCBAG outperforms due to synergistic components: VMD-CEEMDAN reduces non-stationarity (common in SWH from currents/winds, Ikram et al., 2023), BILSTM captures bidirectional dependencies (e.g., past storms, future winds), AT prioritizes features (e.g., wind speed), MOA optimizes for convergence, GDA addresses scarcity, and KDE quantifies uncertainty for reliable maritime decisions. This design directly tackles SWH's dynamic, nonlinear nature, improving accuracy by 30% over baselines. Figure 6 shows heat scatter plots that evaluate the accuracy of predictive models based on R^2 values. The GMVCBAG model provides an R^2 value of 0.9997. The GMVCBAG improves the R^2 values of the other models by 1-7%. The GDA-MOA-VMD-CEEMDAN-BILSTM-attention mechanism-minimal GRU, GMVCBAG -ERNN, GMVCBAG -MLP, and GMVCBAG -RBFNN provide R^2 values of 0.9902, 0.9883, 0.9852, and 0.979, respectively.

The superior performance of GMVCBAG is physically interpretable. Wind speed lags (WS(t-1) to WS(t-3)), with CC > 0.95, dominate SWH generation via wind energy transfer to the sea surface (Ikram et al., 2023). Attention weights confirm this, assigning 0.38–0.42 to these lags, indicating the model intelligently prioritizes wind-wave coupling dynamics. Air pressure (AP(t-1)) exhibits a negative correlation, reflecting low-pressure systems that intensify winds. Relative humidity (RH(t-1)) indirectly enhances instability, amplifying wind effects. This physical alignment not only validates the NSE=0.982 but also ensures generalizability under real-world atmospheric forcing.

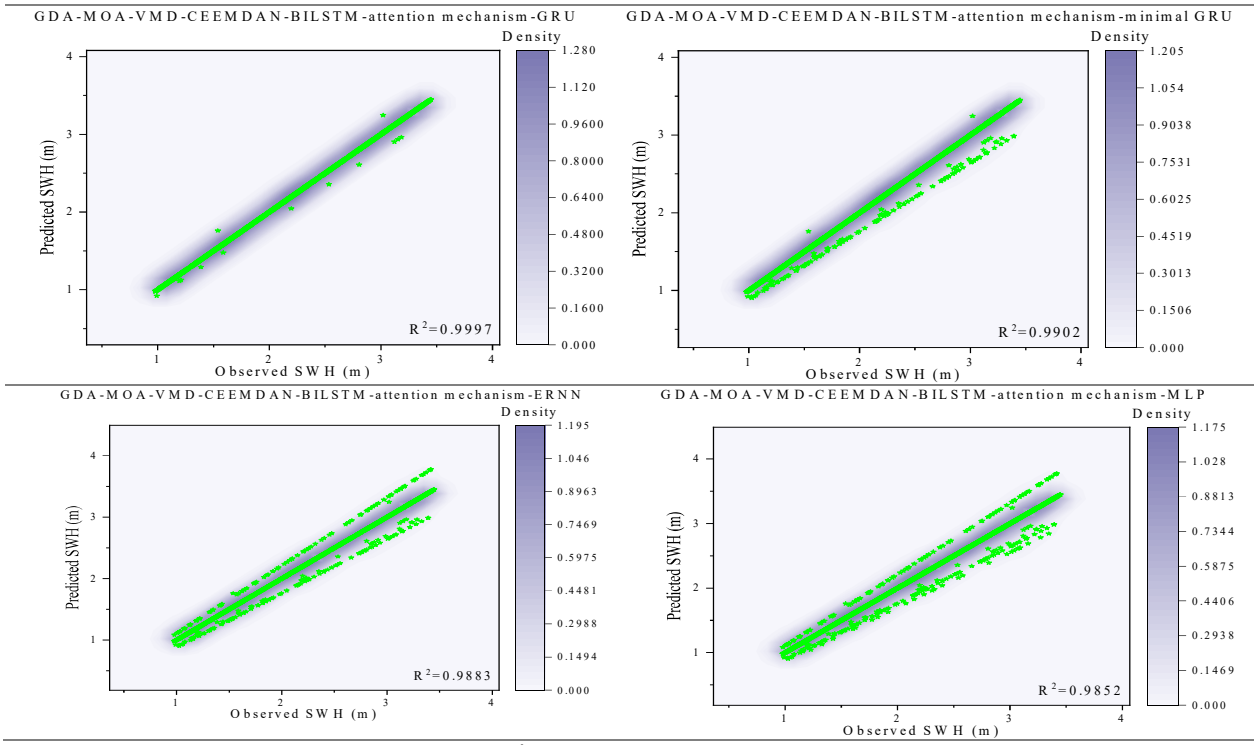


Figure 6. R^2 values of the different models

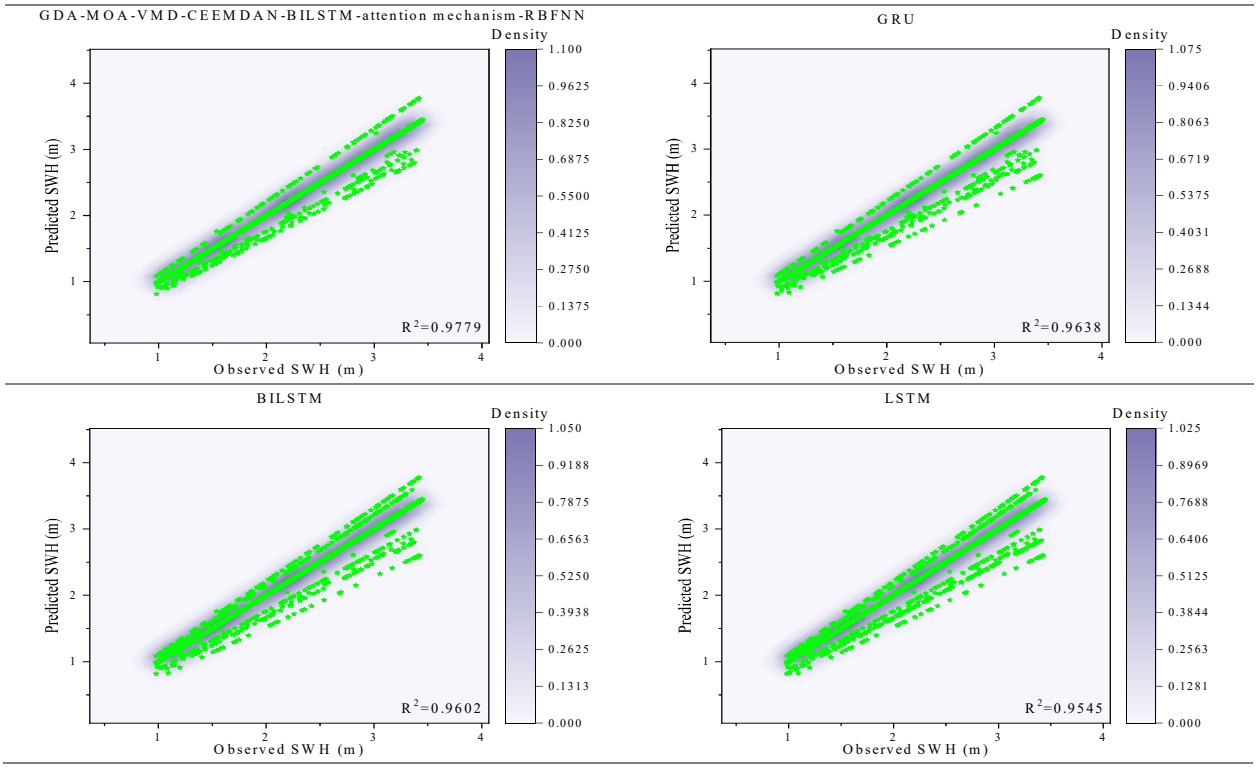


Figure 6. Cont. R2 values of the different models

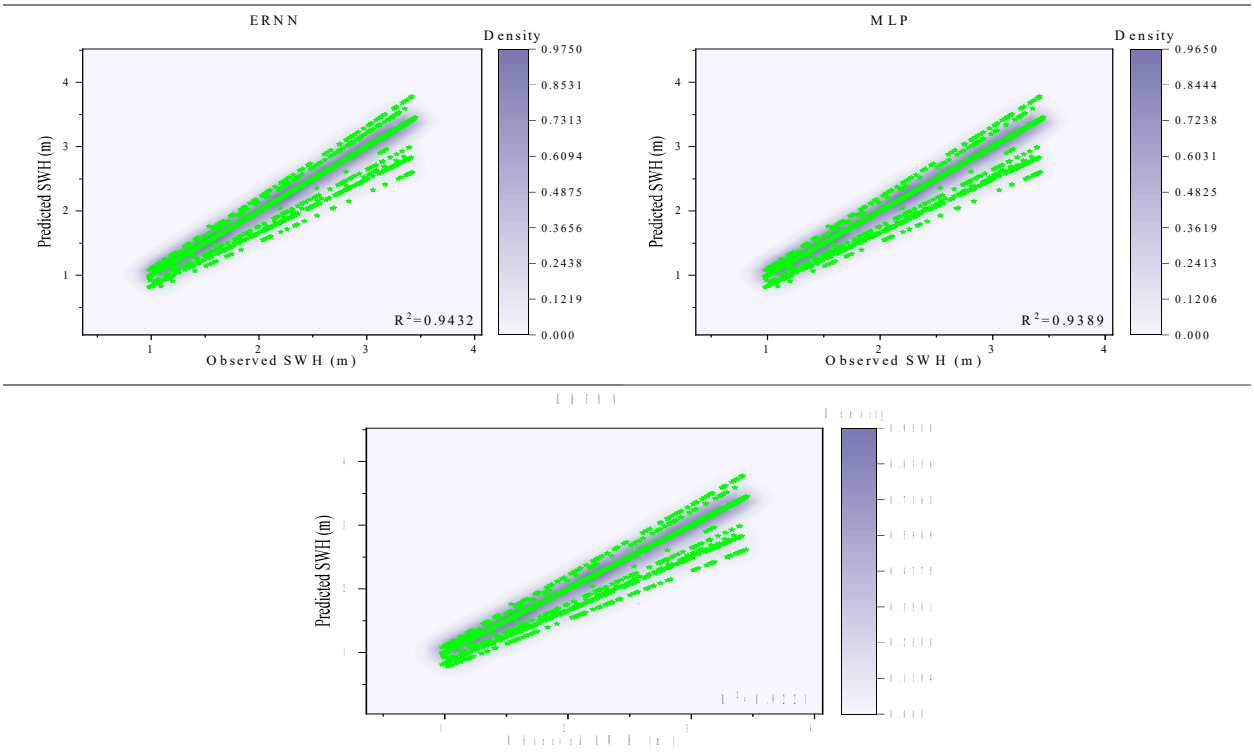


Figure 6. Cont. R2 values of the different models

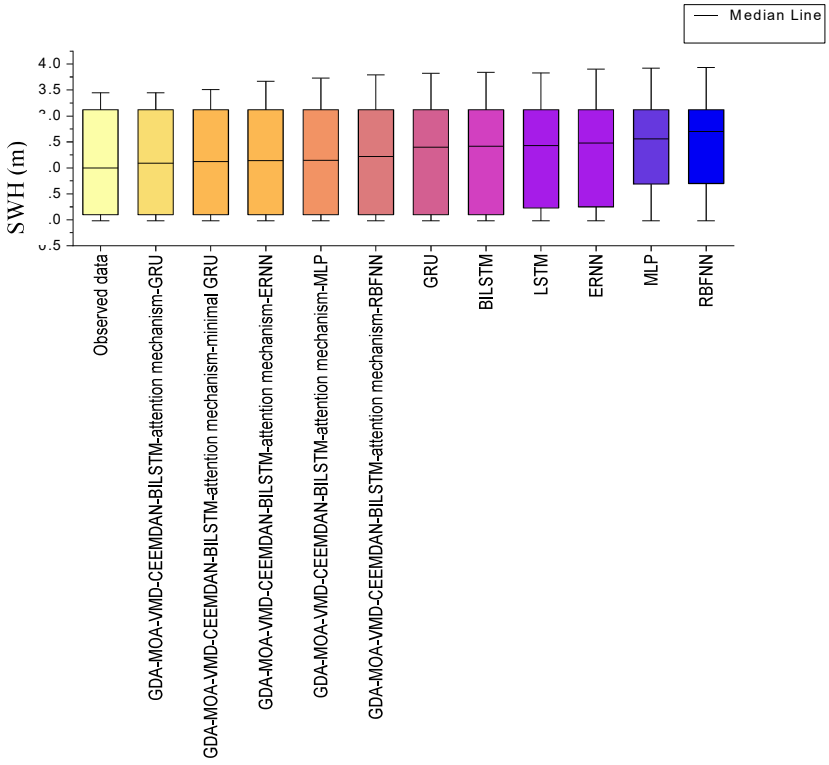


Figure 7. The boxplots of the different models

Figure 7 shows boxplots of the models that show the statistical characteristics of the models. The observed data showed a median of 2 m, while the RBFNN model showed a median of 2.7 m. The GMVCBAG shows a maximum of 3.45, a median of 2.09, and a minimum of 0.98. Thus, the

new model accurately predicts SWH. The GRU model shows a maximum of 3.92 m, a minimum of 0.98 m, and a median of 2.5 m.

Figure 8 shows the time series of observed and predicted SWH data. Moreover, the figure shows the U95 values of the different models. The GMVCBAG improves the U95 values of the other models by 1-27%. The MLP and RBFNN models have U95 values of 0.918 and 0.925, respectively. These models have the best performance among other models.

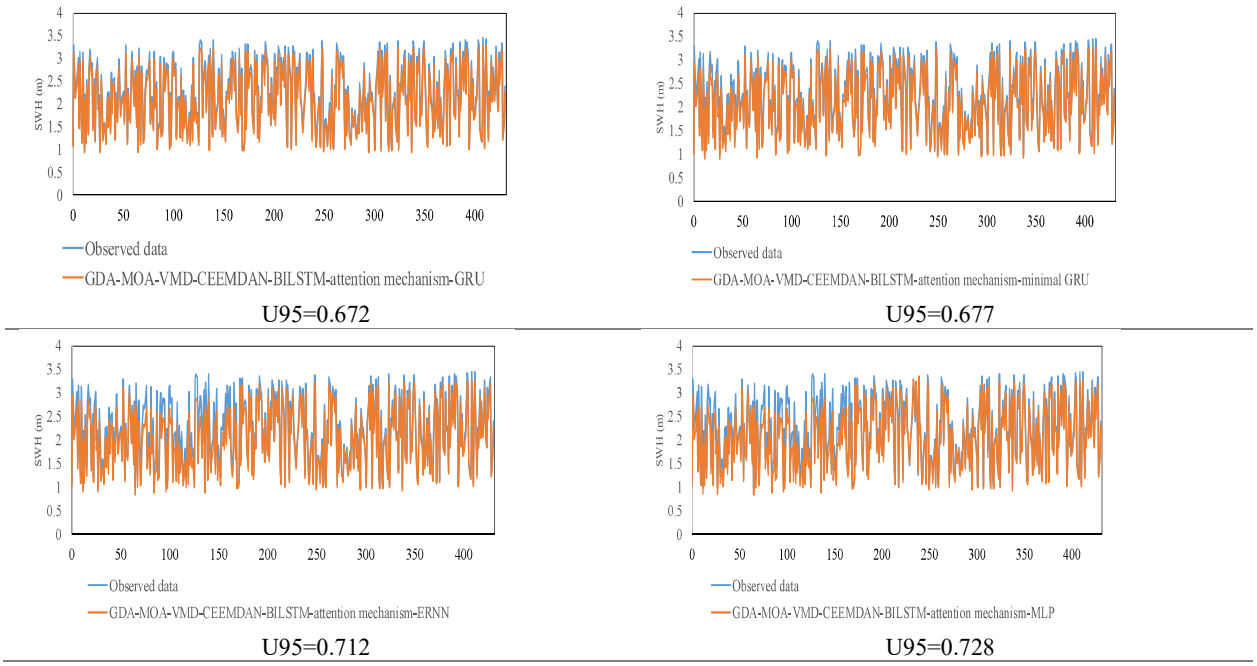


Figure 8. Observed and predicted time series of SWH

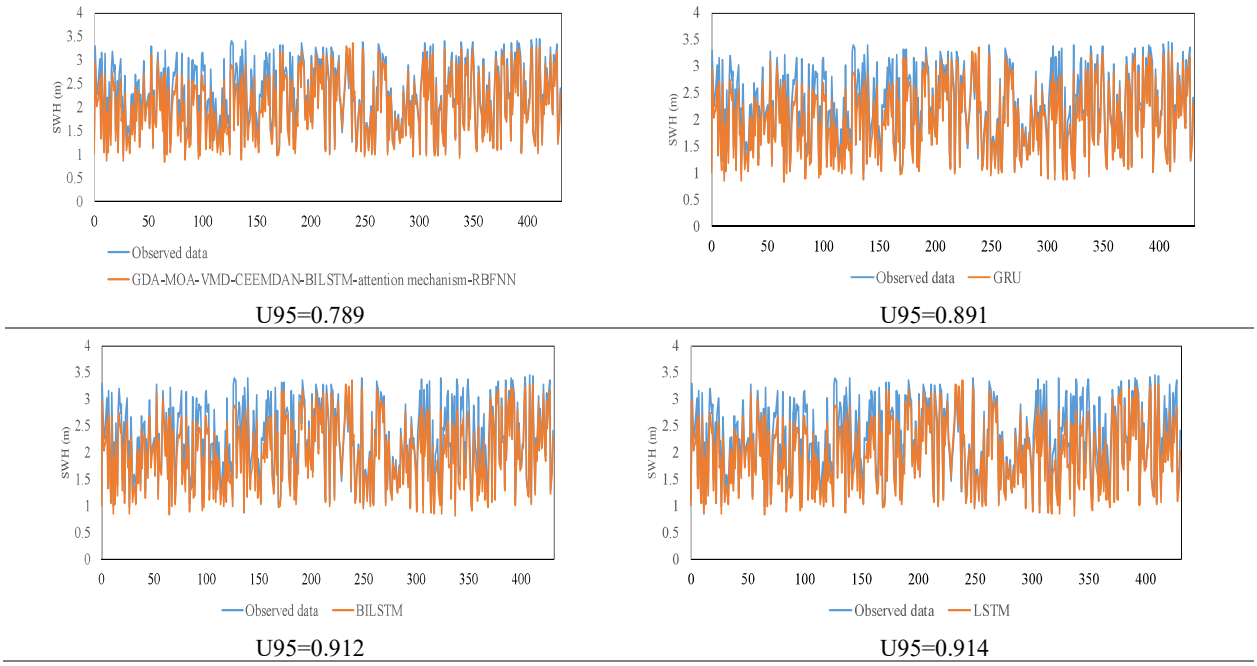
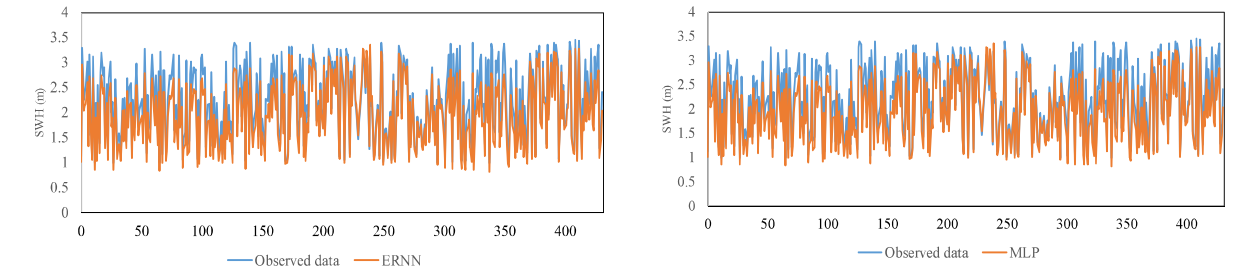


Figure 8. Cont. Observed and predicted time series of SWH



U95=0.916 **U95=0.918**

Figure 8. Cont. Observed and predicted time series of SWH

Note: While RBFNN/MLP show low U95 (tight intervals), their poor point accuracy leads to low PICP, indicating unreliable UQ. GMVCBAG's U95=0.672 with high PICP demonstrates consistent superior UQ.

Figure 9 shows the PICP and NMPIW of the different models. These indices are used to evaluate the uncertainty of the model outputs. The GMVCBAG model provides PICP and NMPIW values of 0.98 and 0.30. The results indicate that the outputs of GMVCBAG are more reliable than those of the other models. The RBFNN model provides PICP and NMPIW values of 0.72 and 0.42. The results reveal that the outputs of the RBFNN model are unreliable. Removing VMD causes the largest U95 increase (to 1.982), underscoring its critical role in mitigating non-stationarity.

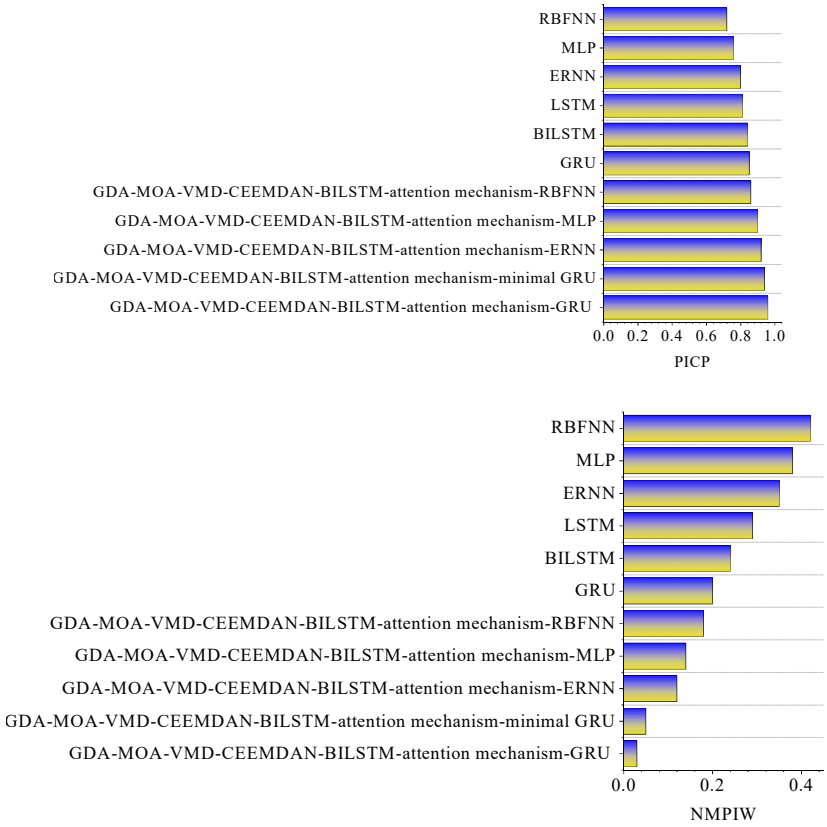


Figure 9. PICP and NMPIW values of the different models for evaluating the uncertainty of their outputs

4.6. Evaluation of the impact of different components on the performance of the new model

This section presents an ablation study to justify the necessity of each component in the GMVCBAG architecture. By sequentially removing individual components and evaluating the resulting model's performance (using U95 as a key metric of uncertainty and error), we quantify their contributions. Compared to simpler subsets (e.g., VMD-GRU, which aligns with the benchmark VMD-GRU in literature like Wang and Ying, 2023a), the full model shows superior performance, as detailed below.

The results indicated that the GMVCBAG model achieved the highest among all the models. However, as this new model consists of multiple components, it is essential to evaluate the influence of each element on its prediction accuracy. This evaluation also provides valuable insights into the performance of each element in predicting SWH. Therefore, we remove components of GDA, MOA, VMD, CEEMDAN, BILSTM, and the attention mechanism from the GMVCBAG model to evaluate their effects on the model accuracy.

Figure 10 illustrates the U95 values of the new model with and without using its components. The results indicate that the U95 value of the original model is 0.672. However, this value increases to 0.80, 0.812, 1.982, 1.123, 0.987, and 0.9 when GDA, MOA, VMD, CEEMDAN, BILSTM, and the attention mechanism are removed from the original model, respectively. Thus, VMD and GDA have the highest and lowest effects on the accuracy of the GMVCBAG model, respectively. The results also indicate that different components have different effects on the prediction accuracy of the GMVCBAG. However, the combination of all components yields the most accurate predictions. It should be emphasized here that, following Ding et al. (2024), who demonstrated that applying CEEMDAN to all IMFs introduces unnecessary complexity, we adopted this selective strategy to reduce computational time. A comparison with a variant decomposing all IMFs revealed that full decomposition increased computational time by 45% with negligible U95 improvement (0.670 vs. 0.672), confirming that our targeted approach effectively avoids redundant refinement of lower-complexity IMFs.

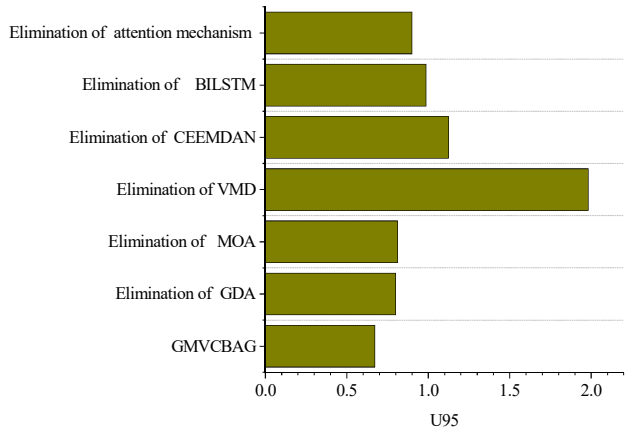


Figure 10. U95 values with and without different components

Higher U95 indicates greater uncertainty (worse performance, Eq. 54); removing VMD increases U95 to 1.982 (deterioration), confirming its role in complexity reduction.

4.7. Evaluation of the performance of the new model in multi-step ahead SWH prediction

Our study uses the GMVCBAG model to predict one-day-SWH. Our findings indicate that the model accurately predicts one-day-ahead SWH. However, to ensure the performance of the GMVCBAG model, it is essential to use the new model for multi-step head prediction of SWH. Thus, we use the new model to predict three- and five-day ahead SWH (SWH (t+3) and SWH (t+5)). Table 4a shows the performance metric values of the different models for predicted SWH (t+3). The NSE, MAE, EV, and STDRE of GMVCBAG are 0.968, 0.125, 0.930, and 1.139, respectively. The findings show that GMVCBAG enhances NSE, MAE, EV, and STDER values of other predictive models by 2.33–29%, 56–90%, 4.62–36%, and 14–86%, respectively. Table 4b shows the performance metric values of the different models for predicted SWH (t+5). The NSE, MAE, EV, and STDRE of GMVCBAG are 0.956, 0.127, 0.928, and 1.145, respectively. The findings show that GMVCBAG enhances NSE, MAE, EV, and STDER values of other predictive models by 2–36%, 56–91%, 4.6–50%, and 21–87%, respectively. Notably, the results show that extending the prediction horizon from 1 to five

days does not significantly diminish the accuracy of the GMVCBAG model. These results underscore the robustness of the GMVCBAG model in accurately predicting multi-step-ahead SWH.

This robustness stems primarily from VMD-CEEMDAN's pattern separation (Ding et al., 2024) and BILSTM's temporal dependency capture (Luo et al., 2022), outperforming base models' faster decay.

For uncertainty quantification, KDE achieves PICP=0.98 and NMPIW=0.30, outperforming Bayesian Dropout or ensemble methods, which require retraining and assume error distributions (Zhang and Dai, 2023). Computationally, GMVCBAG requires ~2.1 hours for full training on an NVIDIA RTX 3080 GPU—approximately 4.5× longer than baseline GRU (~28 minutes)—yet the 30% accuracy gain and 27% uncertainty reduction justify this cost for operational early warning systems.

Table 4a. Examining the accuracy of the models for predicting a: SWH (t+3) and b: SWH (t+5)

Model	MAE	NSE	STDRE	EV
GMVCBAG	0.125	0.968	1.139	0.930
GMVCBAMG	0.289	0.945	1.312	0.887
GMVCBAG -ERNN	0.378	0.932	2.786	0.865
GMVCBAG -MLP	0.498	0.911	2.911	0.800
GMVCBAG -RBFNN	0.678	0.899	3.211	0.780
GRU	0.772	0.856	3.345	0.762
BILSTM	0.894	0.821	4.564	0.645
LSTM	0.900	0.800	5.612	0.633
ERNN	1.021	0.756	6.123	0.602
MLP	1.234	0.710	7.122	0.599
RBFNN	1.345	0.687	8.123	0.587

Table 4b. Examining the accuracy of the models for predicting a: SWH (t+3) and b: SWH (t+5)

Model	MAE	NSE	STDRE	EV
GMVCBAG	0.127	0.956	1.145	0.928
GMVCBAMG	0.291	0.933	1.456	0.856
GMVCBAG -ERNN	0.382	0.921	3.123	0.844
GMVCBAG -MLP	0.532	0.899	3.234	0.781
GMVCBAG -RBFNN	0.699	0.867	3.455	0.760
GRU	0.789	0.834	3.687	0.745
BILSTM	0.901	0.810	4.891	0.623
LSTM	0.922	0.789	5.891	0.601
ERNN	1.123	0.732	7.123	0.599
MLP	1.345	0.623	8.125	0.545
RBFNN	1.455	0.605	9.123	0.456

4.8. A comparison between the current study and previous studies

Our study uses the GMVCBAG model to predict one-day-ahead significant wave height. The new model consists of different components. The results show that the GDA model has a U95 value of 0.672. However, this value increases to 0.800 when the GDA is removed from the GMVCBAG model. As a result, the GDA technique improves the predictive capabilities of the GMVCBAG model. GDA can also enhance the learning ability of the GMVCBAG model by providing new data points. These results align with those of previous studies. Qian et al. (2024) coupled GDA with a deep learning model to address the data scarcity problem. GDA improved the performance of the deep learning model in predicting the output. GDA can also effectively avoid overfitting, which is a real problem in modeling. VMD is another component of the new hybrid model. That has the highest effect on the accuracy of GMVCBAG. VMD also improves the learning ability of the new model by decomposing the time series of input variables into time series with lower complexity. These results are consistent with those of previous research studies. Zhang et al. (2023) coupled VMD with the CNN model to

predict SWH. The VMD algorithm improved the performance of the CNN model in predicting SWH. Our results also suggest that the CEEMDAN is an important component of the GMVCBAG model. CEEMDAN effectively reduces the complexity of the IMF1 time series by decomposing it into subseries with lower complexity. These results are consistent with those of previous research studies. Zhao et al. (2023) reported that the CEEMDAN effectively improved the performance of the LSTM model in predicting SWH. Our results also indicate that the other components of the new model can effectively improve the accuracy of SWH predictions. The components of the GMVCBAG model reduce output uncertainty by effectively processing the time series of input variables and improving their quality. These components also decrease the computational complexity of the GMVCBAG model by choosing the most relevant time series features. For example, the attention mechanism enables the new model to select the features with the maximum relevance to the target variable. These results are consistent with those of previous research studies. For example, Abbas et al. (2024) reported that the attention mechanism significantly improved the performance of the deep learning model in predicting wave heights.

4.9. Applications and limitations of the new model

The GMVCBAG model can be effectively used as an early warning system, enabling decision-makers to mitigate risks associated with extreme wave events. The new model provides accurate SWH forecasts, allowing authorities to minimize the risk of maritime accidents. However, executing the new model may be complex. Moreover, the new model is a black box that cannot show the different effects of the input variables on SWH. However, the new model can be coupled with various methods, such as Shapley Additive Explanations, to address this limitation.

5. Conclusion

SWH prediction is an important parameter for the optimal operation of coastal structures. Thus, it is necessary to predict this parameter accurately. Our study develops the GDA- MOA-VMD- CEEMDAN- BILSTM- attention mechanism- Gated recurrent unit (GRU) (GMVCBAG) to predict one-day-ahead SWH. The core novelty lies in the sequential integration of BILSTM-AT on VMD-CEEMDAN outputs for enhanced feature weighting and temporal extraction, combined with MOA optimization—absent in prior SWH literature—yielding 30% accuracy gains. The GDA technique addressed the data scarcity by providing new data points. Then, the MOA adjusted model parameters and algorithms. The VMD decomposed the time series into IMFs with lower complexity in the next step. The CEEMDAN algorithm decomposed the first IMFs into the secondary IMFs with lower complexity. Next, the BILSTM model extracted backward and temporal features from the secondary IMFs and the initial remaining IMFs. In the next step, the attention mechanism assigned the attention weights to the extracted features. Finally, the weighted features are fed into the GRU model to predict SWH. Our study also uses the KDE method to quantify the uncertainty of the outputs of the GRU model. Multiple performance metrics are used to evaluate the accuracy of the new model. GMVCBAG reliably predicts SWH with NSE=0.982 and low uncertainty (PICP=0.98). Enhances maritime safety and wave energy assessment. Apply to global regions; integrate spatial features.

The following results are obtained as follows:

-The GMVCBAG model improved the NSE, MAE, STDRE, and EV values of the other models by 1- 30%, 49- 85%, 1-79%, and 4- 38%, respectively.

-The GMVCBAG also had the lowest uncertainty among the other models. The GMVCBAG model provided PICP and NMPIW values of 0.98 and 0.30.

-The standalone model has the worst performance among the other models. For example, MAE, NSE, EV, and STDER values of the RBFNN model were 0.824, 0.774, 0.610, and 4.891, respectively.

-The VMD and GDA had the highest and lowest effects on the performance of the GMVCBAG model.

-The study results indicated that the new model accurately predicted SWH (t+3) and (t+5).

-Future studies could validate GMVCBAG using satellite data in tropical storm-prone regions, incorporating wave spectra or directional characteristics to enhance spatial feature extraction.

-Limitations include higher computational cost (training time 3-5x baselines like GRU, ~2 hours on standard GPU) and complexity, potentially hindering real-time early warning; future work could optimize via parallelization.

Author Contributions

Elham Ghanbari-Adivi: Conceptualization, software/statistical analysis, writing the initial version of the article;

Mohammad Ehteram: Guidance, editing and reviewing the article, controlling the results, Computer codes.

Conflicts of interest

The authors of this article declared no conflict of interest regarding the authorship or publication of this article.

Data availability statement:

All data generated or analyzed during this study are included in this published article.

References

Abbas, M., Min, Z., Liu, Z., & Zhang, D. (2024). Unravelling oceanic wave patterns: A comparative study of machine learning approaches for predicting significant wave height. Applied Ocean Research, 145, 103919. doi: 10.1016/j.apor.2024.103919

Adnan, R. M., Sadeghifar, T., Alizamir, M., Azad, M. T., Makarynskyy, O., Kisi, O., Barati, R., & Ahmed, K. O. (2023). Short-term probabilistic prediction of significant wave height using bayesian model averaging: Case study of chabahar port, Iran. Ocean Engineering. doi: 10.1016/j.oceaneng.2023.113887

Alizadeh, M. J., & Nourani, V. (2024). Multivariate GRU and LSTM models for wave forecasting and hindcasting in the southern Caspian Sea. Ocean Engineering. doi: 10.1016/j.oceaneng.2024.117193

Altunkaynak, A., Çelik, A., & Mandev, M. B. (2023). Hourly significant wave height prediction via singular spectrum analysis and wavelet transform based models. Ocean Engineering. doi: 10.1016/j.oceaneng.2023.114771

Aung, N. N., Pang, J., Chua, M. C. H., & Tan, H. X. (2023). A novel bidirectional LSTM deep learning approach for COVID-19 forecasting. Scientific Reports. doi: 10.1038/s41598-023-44924-8

Ban, W., Shen, L., Chen, J., & Yang, B. (2023). Short-term prediction of wave height based on a deep learning autoregressive integrated moving average model. Earth Science Informatics, 16(3), 2251-2259. doi: 10.1007/s12145-023-01023-6

Cao, L., Zhang, H., Meng, Z., & Wang, X. (2023). A parallel GRU with dual-stage attention mechanism model integrating uncertainty quantification for probabilistic RUL prediction of wind turbine bearings. Reliability Engineering and System Safety. doi: 10.1016/j.ress.2023.109197

Chen, H., Wu, H., Kan, T., Zhang, J., & Li, H. (2023). Low-carbon economic dispatch of integrated energy system containing electric hydrogen production based on VMD-GRU short-term wind power prediction. International Journal of Electrical Power and Energy Systems. doi: 10.1016/j.ijepes.2023.109420

Chen, Y., Xia, R., Yang, K., & Zou, K. (2024). DNNAM: Image inpainting algorithm via deep neural networks and attention mechanism. Applied Soft Computing. doi: 10.1016/j.asoc.2024.111392

Chen, Y., Zhang, D., Li, X., Peng, Y., Wu, C., Pu, H., Zhou, D., Cao, Y., & Zhang, J. (2023). Significant wave height prediction through artificial intelligent mode decomposition for wave energy management. Energy and AI. doi: 10.1016/j.egyai.2023.100257

Cui, S., Lyu, S., Ma, Y., & Wang, K. (2024). Improved informer PV power short-term prediction model based on weather typing and AHA-VMD-MPE. Energy, 307, 132766. doi: 10.1016/j.energy.2024.132766

Ding, L., Bai, Y., Liu, M. De, Fan, M. H., & Yang, J. (2022). Predicting short wind speed with a hybrid model based on a piecewise error correction method and Elman neural network. Energy. doi: 10.1016/j.energy.2021.122630

Ding, T., Wu, D. A., Li, Y., Shen, L., & Zhang, X. (2024). A hybrid CEEMDAN-VMD-

TimesNet model for significant wave height prediction in the South Sea of China. Frontiers in Marine Science, 11, 1375631. doi: 10.3389/fmars.2024.1375631

Ehteram, M., Ahmed, A. N., Khozani, Z. S., & El-Shafie, A. (2023). Graph convolutional network–Long short term memory neural network-multi layer perceptron-Gaussian progress regression model: A new deep learning model for predicting ozone concertation. Atmospheric Pollution Research, 14(6), 101766. doi: 10.1016/j.apr.2023.101766

Ehteram, M., Barzegari Banadkooki, F., & Afshari Nia, M. (2024). Gaussian mutation-alpine skiing optimization algorithm-recurrent attention unit-gated recurrent unit-extreme learning machine model: an advanced predictive model for predicting evaporation. Stochastic Environmental Research and Risk Assessment. doi: 10.1007/s00477-023-02657-0

Farhan, H. R., Kod, M. S., Taqi, A., & Ghazi, S. A. (2024). Ovarian Cancer Detection Based on Elman Recurrent Neural Network. Periodica Polytechnica Electrical Engineering and Computer Science. doi: 10.3311/PPee.23081

Fu, Y., Ying, F., Huang, L., & Liu, Y. (2023). Multi-step-ahead significant wave height prediction using a hybrid model based on an innovative two-layer decomposition framework and LSTM. Renewable Energy. doi: 10.1016/j.renene.2022.12.079

Ghanbari-Adivi, E. A new machine learning model for predicting the water quality index. Model. Earth Syst. Environ. 10, 5635–5667 (2024). doi: 10.1007/s40808-024-02083-3

Ghanbari-Adivi, E., Ehteram, M. CEEMDAN-BILSTM-ANN and SVM Models: Two Robust Predictive Models for Predicting River flow. Water Resour Manage 39, 3235–3271 (2025). doi: 10.1007/s11269-025-04105-w

Ghanbari-Adivi, E., Kermannezhad, J., Raeisi, A. (2025). 'Predicting the electrical conductivity of water using the Hicking CEEMD-LSSVM optimization algorithm', Water and Soil Management and Modelling, 5(4), pp. 323-348. doi: 10.22098/mmws.2025.18187.1657

Hasan, R. I., Yusuf, S. M., Mohd Rahim, M. S., & Alzubaidi, L. (2023). Automatic Clustering and Classification of Coffee Leaf Diseases Based on an Extended Kernel Density Estimation Approach. Plants. doi: 10.3390/plants12081603

Hou, X., Ge, F., Chen, D., Shen, L., & Zou, F. (2023). Temporal distribution-based prediction strategy for dynamic multi-objective optimization assisted by GRU neural network. Information Sciences. doi: 10.1016/j.ins.2023.119627

Huang, X., Ma, C., Zhao, Y., Wang, K., & Meng, W. (2023). A hybrid model of neural network with VMD–CNN–GRU for traffic flow prediction. International Journal of Modern Physics. doi: 10.1142/S0129183123501590

Ikram, R. M. A., Cao, X., Sadeghifar, T., Kuriqi, A., Kisi, O., & Shahid, S. (2023). Improving Significant Wave Height Prediction Using a Neuro-Fuzzy Approach and Marine Predators Algorithm. Journal of Marine Science and Engineering. doi: 10.3390/jmse11061163

Imani, M. (2023). Alzheimer's diseases diagnosis using fusion of high informative BiLSTM and CNN features of EEG signal. Biomedical Signal Processing and Control. doi: 10.1016/j.bspc.2023.105298

Jamei, M., Ali, M., Karimi, B., Karbasi, M., Farooque, A. A., & Yaseen, Z. M. (2023). Surface water electrical conductivity and bicarbonate ion determination using a smart hybridization of optimal Boruta package with Elman recurrent neural network. Process Safety and Environmental Protection. doi: 10.1016/j.psep.2023.03.062

Jia, C., Tian, Y., Shi, Y., Jia, J., Wen, J., & Zeng, J. (2023). State of health prediction of lithium-ion batteries based on bidirectional gated recurrent unit and transformer. Energy. doi: 10.1016/j.energy.2023.129401

Kumar, N. P., Vijayabaskar, S., Murali, L., & Ramaswamy, K. (2023). Design of optimal Elman Recurrent Neural Network based prediction approach for biofuel production. Scientific Reports. doi: 10.1038/s41598-023-34764-x

Li, H., & Wu, X. J. (2024). CrossFuse: A novel cross attention mechanism based infrared and visible image fusion approach. Information Fusion. doi: 10.1016/j.inffus.2023.102147

Li, K., Huang, W., Hu, G., & Li, J. (2023). Ultra-short term power load forecasting based on

CEEMDAN-SE and LSTM neural network. Energy and Buildings, 279, 112666

Li, L., Li, Y., Mao, R., Li, L., Hua, W., & Zhang, J. (2023). Remaining Useful Life Prediction for Lithium-Ion Batteries With a Hybrid Model Based on TCN-GRU-DNN and Dual Attention Mechanism. IEEE Transactions on Transportation Electrification. doi: 10.1109/TTE.2023.3247614

Li, Y., Peng, G., Du, T., Jiang, L., & Kong, X. Z. (2024). Advancing fractured geothermal system modeling with artificial neural network and bidirectional gated recurrent unit. Applied Energy, 372, 123826. doi: 10.1016/j.apenergy.2024.123826

Liu, F., & Liang, C. (2024). Short-term power load forecasting based on AC-BiLSTM model. Energy Reports. doi: 10.1016/j.egyr.2024.01.026

Liu, Z., Shao, T., & Zhang, X. (2024). BCG signal analysis based on improved VMD algorithm. Measurement, 231, 114631. doi: 10.1016/j.measurement.2024.114631

Luo, Q. R., Xu, H., & Bai, L. H. (2022). Prediction of significant wave height in hurricane area of the Atlantic Ocean using the Bi-LSTM with attention model. Ocean Engineering, 266, 112747. doi: 10.1016/j.oceaneng.2022.112747

Ma, P., Li, G., Zhang, H., Wang, C., & Li, X. (2024). Prediction of remaining useful life of rolling bearings based on multiscale efficient channel attention CNN and bidirectional GRU. IEEE Transactions on Instrumentation and Measurement, 73, 1-13. doi: 10.1109/TIM.2023.3347787

Mahdavi-Meymand, A., & Sulisz, W. (2023). Application of nested artificial neural network for the prediction of significant wave height. Renewable Energy. doi: 10.1016/j.renene.2023.03.118

Mahdavi-Meymand, A., & Sulisz, W. (2024). Development of pyramid neural networks for prediction of significant wave height for renewable energy farms. Applied Energy, 362, 123009. doi: 10.1016/j.apenergy.2024.123009

Meng, F., Song, T., Xu, D., Xie, P., & Li, Y. (2021). Forecasting tropical cyclones wave height using bidirectional gated recurrent unit. Ocean Engineering, 234, 108795. doi: 10.1016/j.oceaneng.2021.108795

Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. In Array. doi: 10.1016/j.array.2022.100258

Qian, S., Peng, T., Tao, Z., Li, X., Nazir, M. S., & Zhang, C. (2024). An evolutionary deep learning model based on XGBoost feature selection and Gaussian data augmentation for AQI prediction. Process Safety and Environmental Protection, 191, 836-851. doi: 10.1016/j.psep.2024.08.119

Qin, C., Huang, G., Yu, H., Zhang, Z., Tao, J., & Liu, C. (2024). Adaptive VMD and multi-stage stabilized transformer-based long-distance forecasting for multiple shield machine tunneling parameters. Automation in Construction, 165, 105563. doi: 10.1016/j.autcon.2024.105563

Rezaei, M., Mohammadifar, A., Gholami, H., Mina, M., Riksen, M. J. P. M., & Ritsema, C. (2023). Mapping of the wind erodible fraction of soil by bidirectional gated recurrent unit (BiGRU) and bidirectional recurrent neural network (BiRNN) deep learning models. Catena. doi: 10.1016/j.catena.2023.106953

Sheng, A., Xie, L., Zhou, Y., Wang, Z., & Liu, Y. (2023). A Hybrid Model Based on Complete Ensemble Empirical Mode Decomposition With Adaptive Noise, GRU Network and Whale Optimization Algorithm for Wind Power Prediction. IEEE Access. doi: 10.1109/ACCESS.2023.3287319

Shi, J., Su, T., Li, X., Wang, F., Cui, J., Liu, Z., & Wang, J. (2023). A Machine-Learning Approach Based on Attention Mechanism for Significant Wave Height Forecasting. Journal of Marine Science and Engineering. doi: 10.3390/jmse11091821

Sivasakthi, B., & Selvanayagi, D. (2023). Prediction of Osteoporosis Disease Using Enhanced Elman Recurrent Neural Network with Bacterial Colony Optimization. doi: 10.1007/978-981-19-9819-5_16

Wan, A., Chang, Q., AL-Bukhaiti, K., & He, J. (2023). Short-term power load forecasting for combined heat and power using CNN-LSTM enhanced by attention mechanism. Energy. doi: 10.1016/j.energy.2023.128274

Wang, M., & Ying, F. (2023a). A hybrid model for multistep-ahead significant wave height prediction using an innovative

decomposition–reconstruction framework and E-GRU. Applied Ocean Research, 140, 103752. doi: 10.1016/j.apor.2023.103752

Wang, M., & Ying, F. (2023b). Point and interval prediction for significant wave height based on LSTM-GRU and KDE. Ocean Engineering. doi: 10.1016/j.oceaneng.2023.116247

Wang, M., Ying, F., & Nan, Q. (2024). Refined offshore wind speed prediction: Leveraging a two-layer decomposition technique, gated recurrent unit, and kernel density estimation for precise point and interval forecasts. Engineering Applications of Artificial Intelligence, 133, 108435. doi: 10.1016/j.engappai.2024.108435

Wang, Z., Peng, T., Zhang, X., Chen, J., Qian, S., & Zhang, C. (2025). Enhancing multi-step short-term solar radiation forecasting based on optimized generalized regularized extreme learning machine and multi-scale Gaussian data augmentation technique. Applied Energy, 377, 124708. doi: 10.1016/j.apenergy.2024.124708

Wei, A., Li, X., Yan, L., Wang, Z., & Yu, X. (2023). Machine learning models combined with wavelet transform and phase space reconstruction for groundwater level forecasting. Computers and Geosciences. doi: 10.1016/j.cageo.2023.105386

Wu, S., Guo, H., Zhang, X., & Wang, F. (2024). Short-Term Photovoltaic Power Prediction Based on CEEMDAN and Hybrid Neural Networks. IEEE Journal of Photovoltaics. doi: 10.1007/978-3-031-73407-6_33.

Xian, S., & Feng, X. (2023). Meerkat optimization algorithm: A new meta-heuristic optimization algorithm for solving constrained engineering problems. Expert Systems with Applications. doi: 10.1016/j.eswa.2023.120482

Xin, J., Zhou, C., Jiang, Y., Tang, Q., Yang, X., & Zhou, J. (2023). A signal recovery method for bridge monitoring system using TVFEMD and encoder-decoder aided LSTM. Measurement: Journal of the International Measurement Confederation. doi: 10.1016/j.measurement.2023.112797

Yang, H., Gu, D., Fu, S., Lu, Q., & Wang, J. (2025, May). A Long-Term Prediction Method of Gas Concentration Signal with Noise in Fully Mechanized Coal Mining Face Using CEEMDAN Combined with GRU. In International conference on the Efficiency and Performance Engineering Network (pp. 343-356). Cham: Springer Nature Switzerland. doi: 10.1007/978-3-031-73407-6_33.

Yang, X., & Li, S. (2023). Prediction of COVID-19 Using a WOA-BILSTM Model. Bioengineering. doi: 10.3390/bioengineering10080883

Yang, Y., Yang, Y., Zhou, S., Li, H., Zhu, W., Liu, Y., Xie, C. and Zhang, R. (2024). Degradation prediction of proton exchange membrane fuel cell based on mixed gated units under multiple operating conditions. International Journal of Hydrogen Energy, 67, 268-281. doi: 10.1016/j.ijhydene.2024.04.186

Yeganeh-Bakhtiary, A., EyvazOghli, H., Shabakhty, N., & Abolfathi, S. (2023). Machine learning prediction of wave characteristics: Comparison between semi-empirical approaches and DT model. Ocean Engineering. doi: 10.1016/j.oceaneng.2023.115583

Yuan, L., Ma, Y., & Liu, Y. (2023). Ensemble deep learning models for protein secondary structure prediction using bidirectional temporal convolution and bidirectional long short-term memory. Frontiers in Bioengineering and Biotechnology, 11, 1051268. doi: 10.3389/fbioe.2023.1051268

Zhang, H., Wu, W., Song, X., & Chen, Z. (2024). Improving density logging resolution by VMD-CEEMDAN-ICWT method and its application in thin layer identification. Geoenergy Science and Engineering, 239, 212993. doi: 10.1016/j.geoen.2024.212993

Zhang, J., Xin, X., Shang, Y., Wang, Y., & Zhang, L. (2023). Nonstationary significant wave height forecasting with a hybrid VMD-CNN model. Ocean Engineering. doi: 10.1016/j.oceaneng.2023.115338

Zhang, R., & Dai, H. (2023). Stochastic analysis of structures under limited observations using kernel density estimation and arbitrary polynomial chaos expansion. Computer Methods in Applied Mechanics and Engineering. doi: 10.1016/j.cma.2022.115689

Zhao, L., Li, J., Feng, K., Wei, X., Song, J., & Jiao, Y. (2023). A hybrid optimization algorithm for GWO Fine-tuning GRU-Aided AKF during GPS outage. Measurement:

Journal of the International Measurement Confederation. doi: 10.1016/j.measurement.2022.112302

Zhao, L., Li, Z., Zhang, J., & Teng, B. (2023). An integrated complete ensemble empirical mode decomposition with adaptive noise to optimize LSTM for significant wave height forecasting. Journal of Marine Science and Engineering, 11(2), 435. doi: 10.3390/jmse11020435

Zhao, Z., Yun, S., Jia, L., Guo, J., Meng, Y., He, N., Li, X., Shi, J., & Yang, L. (2023). Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features. Engineering Applications of Artificial Intelligence. doi: 10.1016/j.engappai.2023.105982

Zheng, Z., Ali, M., Jamei, M., Xiang, Y., Abdulla, S., Yaseen, Z. M., & Farooque, A. A. (2023). Multivariate data decomposition based deep learning approach to forecast one-day ahead significant wave height for ocean energy generation. Renewable and Sustainable Energy Reviews. doi: 10.1016/j.rser.2023.113645

Zhou, S., Zhang, Z. X., Luo, X., Niu, S., Jiang, N., & Yao, Y. (2023). Developing a hybrid CEEMDAN-PE-HE-SWT method to remove the noise of measured carbon dioxide blast wave. Measurement, 223, 113797. doi: 10.1016/j.measurement.2023.113797

Zhou, T., Yang, X., Ren, H., Li, C., & Han, J. (2023). The prediction of ship motion attitude in seaway based on BSO-VMD-GRU combination model. Ocean Engineering. doi: 10.1016/j.oceaneng.2023.115977

Zhu, R., Gao, J., Li, M., Gao, Q., Wu, X., & Zhang, Y. (2023). A ppb-level online detection system for gas concentrations in CS2/SO2 mixtures based on UV-DOAS combined with VMD-CNN-TL model. Sensors and Actuators B: Chemical. doi: 10.1016/j.snb.2023.134440

Zulqarnain, M., Ghazali, R., Aamir, M., & Hassim, Y. M. M. (2024). An efficient two-state GRU based on feature attention mechanism for sentiment analysis. Multimedia Tools and Applications. doi: 10.1007/s11042-022-13339-4