DOI: 10.6977/IJoSI.202209_7(3).0006
J. P. Mishra, K. Singh and H. Chaudhary / Int. J. Systematic Innovation, 7(3), 64-77 (2022)

Intelligent ocean wave height prediction system using light GBM model

Jai Prakash Mishra[1], Kulwant Singh[2] and Himanshu Chaudhary[3*]
[1,2,3]Department of Electronics & Communication Engineering, Manipal University Jaipur, Jaipur, Rajasthan 303007.
[*] Corresponding author E-mail: himanshu.chaudhary@jaipur.manipal.edu
(Received 4 July 2022; Final version received 29 July 2022; Accepted 23 August 2022)

Abstract

Forecasting the heights of marine waves is an important tool for offshore and coastal engineering and a huge undertaking in marine detection and warning. However, a precise forecast of the Sea Wave Height (SWH) is challenging and outstanding to waves' volatility and fluctuation characteristics. Therefore, our research proposes an Intelligent Ocean Wave Height Prediction system using a light gradient boosting machine learning. Initially, this research extracts the wave speed, peak wave direction, zero up crossing wave period, wave period, and SWH, among the wave-based properties. Then, the retrieved data are fed into the proposed light GBM, which operates well with the high-dimensional data that makes our proposed approach easy to interpret. The proposed method can also be utilized to estimate wave height because light GBM performs with redundant data in time-window-size data and is noise-insensitive. Experimental results reveal that our light GBM significantly improves the accuracy of numerical predictions of ocean wave height. When compared to the baseline, our proposed approach achieves lower error than the Multilayer Perceptron Neural Network (MPNN), Cascade Correlation Neural Network (CCNN), General Regression Neural Network (GRNN), and Radial Basis Function Neural Network (RBFNN), with error rates of 5.01 %, 44.33 %, 6.22 %, and 2.23 %, respectively. As a result, our proposed technique has a lower MAPE of 2.21 % compared to baseline approaches.

Keywords: Sea wave height, light GBM, Machine learning, Forecasting, Wave variables.

1. Introduction

Marine disasters represent a danger to many countries worldwide, resulting in massive deaths and economic damage. The growing growth of a variety of offshore businesses has piqued attention in the effective wave forecast characteristics [Hashim et al., (2016)] [1998]. High-energy ocean waves Significant Wave Height (SWH) can collapse ships and finish marine or seaside infrastructure. It endangers human lives, agriculture output, and the long-term viability of aquaculture goods. Wave height in sea forecasting is a difficult and significant issue in seaside and marine engineering due to waves' difficult and unpredictable nature [Amin, (2013)]. It is essential to estimate marine wave heights promptly and precisely in disaster warning and emergency prevention, as well as the development and renovation of seashore area construction, marine transportation, and environmental protection [Sabatier, (2007)]. As a result, accurate SWH forecasting is crucial since it can help reduce societal and business losses.

Furthermore, Sea Wave Height forecasting can provide certain advantages. Altunkaynak & Wang, (2012) created a new approach for predicting SWH by combining a genetic algorithm with Kalman filtering. These approaches reduced mean relative and mean square errors over ANN and demonstrated its superiority. Improved ship routes based on Sea Wave Height forecasts, for example, could skip stormy sea regions, reducing sailing time and fuel expenditures. Moreover, forecasting sea wave heights can give valuable information for the army and naval planning activities. Sea wave height forecasting systems have been developed for years due to their relevance and practical uses. Empirical and numeric form Sea Wave Height prediction algorithms had tremendous comprehensibility at the beginning, but low performance and generalization capability. Moreover, the authors [Nitsure et al., (2012)] used genetic programming to forecast wave heights based on wind data. Prediction findings with lead periods of up to 12 and 24 hours were satisfactory, with coefficients of correlation between predicted and measured values greater than 0.87.

When predicting ocean wave heights, a variety of methodologies can be used. For meaningful wave height forecasting, experimental, mathematical, and soft computing methodologies have been described. Tidal variation, wind blowing in different directions, the depth and physical qualities of water are all external elements that affect the wave's speed and height. In a numerical model that incorporates wave propagation, these physical processes and interactions are represented as a different equation. To overcome the challenges in the wave propagation model, a powerful computational infrastructure is usually necessary. Prahlada & Deka, (2015) attempted to create a hybrid wavelet and artificial neural network model for SWH prediction beyond multistep lead time by utilizing the advantageous properties of both. The given strategy has been demonstrated to be both effective and practicable. This method requires creating a physical design of the tallness of a sea wave, which necessitates a thorough understanding of basic physical methods, as well as a significant financial investment and time commitment [Yoon et al., (2011)]. When an emergency scenario occurs in the water, faster and more precise forecasting systems must be developed to estimate wave heights rapidly. For example, in [Deo et al., (2001)] the authors suggested a 3-layer feed-forward network model that forecasts the heights of ocean waves in various seas and examines the characteristics that impact ocean wave height prediction. The authors [Zamani et al., (2008), Malekmohamadi et al., (2011] investigated many data-driven designs based on Artificial neural networks (ANNs) and Instance-based learning in-depth (IBL). Experiments revealed that ANNs had a modest edge over the IBL in forecasting severe wave conditions, and ANNs also have a competitive advantage in predicting extreme wave conditions. Mahjoobi & Etemad-Shahidi, (2008) using Support Vector Machines, developed a model for estimating the heights of ocean waves and, discovered that the Support Vector Machine validation loss is lower than the traditional neural network [Mahjoobi & Mosabbeb, (2009)], the effectiveness of arrangement and regression trees in predicting the heights of marine waves was explored. The forecast findings suggest that the decision tree can be utilized as an effective method with a reasonable error range. However, predicting wave heights based on information about storm generation is essentially an indefinite and unpredictable method that is hard to represent using deterministic equations. Machine learning approaches employ statistics to understand better the spatial and temporal relationships buried in historical time series. It makes it a perfect option for a machine learning model. It focuses on detecting a probability plot in a set of input data and then applying the same strategy to forecast the desired attribute. This research paper's primary feature is as follows:

Traditional empirical or numerical-based forecasting models are used in existing research, but they have speed and accuracy constraints. Our research introduces light Gradient Boosting Machine (light GBM), which accurately predicts wave height by overcoming the problems in existing research, and we remove the wave height column from the data frame to avoid loss.

The following is how the rest of the article is put together: Section 2 discusses the associated work. The proposed technique, including its processing phases, is discussed in detail in Section 3. Section 4 explains the outcome of the research. Section 5 concludes with the findings.

2. Related works

Due to the difficulties of data gathering and processing power limits in the previous century, SWH prediction is primarily based on actual or mathematical models. Leading to a shortage of intellectual processes, these strategies have high readability but poor prediction accuracy and poor applicability. Due to the rapid evolution of ML theory, many ML methods, such as Support Vector Regression, Bayesian Network, XG Boost, extreme learning machine, and ANN, have been successfully employed in Sea Wave Height forecasting. In contrast to prior empirically or numerically based "hard computing" methodologies, these approaches were called "soft computing."

Cornejo-Bueno et al. [Cornejo-Bueno et al., (2016)] proposed employing a hybrid clustering evolutionary algorithms, an extreme learning machine technique for marine energy applications in SWH and flow prediction, and received positive results.

Abhigna et al. [Abhigna et al., (2017)] investigated SWH prediction by means of correlation Coefficient (CC) and Mean Square Error (MSE) for Feed Forward and Recurrent Neural Networks trained using Levenberg Marquardt (LM), Conjugate Gradient

(CG), and Bayesian Regularization (BR). A year's worth of data from anchored buoys in the Bay of Bengal was utilized to train the network, and data for the next year was anticipated. Compared to other techniques, the Recurrent Neural Network with Bayesian Regularization fared the best.

Nikoo et al. [Nikoo et al., (2018)] used a fuzzy K-nearest neighbor (FKNN) model to forecast Sea Wave Height, where wind direction changes affect fetch length. FKNN outperformed BN, regression tree induction, and support vector regression in terms of wave height prediction, especially for wave heights greater than 2 m. Wei and Hsieh [Wei & Hsieh, (2018)] used ANN in two settings to examine the feasibility of forecasting waves using data from a neighboring buoy. The study found that the model incorporating information from the neighboring buoy outperformed the existing works. Wang et al. [Wang et al., (2018)] used a Mind Evolutionary Algorithm-Back Propagation neural network hybrid method (MEA-BP). Yang et al. [Yang et al., (2019)] aimed to forecast SWH based on a CS–BP model, taking into account the edges of backpropagation neural networks (BP) and cuckoo search algorithms (CS), and the suggested model has good potential for wave height prediction. In a recent study, Zhang and Dai [Zhang & Dai, (2019)] used the conditional limited Boltzmann machine in the traditional deep belief network to forecast SWH. The measurement criterion demonstrated that the newly suggested technique is quite good at predicting short-term and severe occurrences. Son et al. [Son et al., (2020)] used the bi-directional convolutional Long Short Term Memory technique to estimate real-valued Sea Wave Height from a series of consecutive ocean photos, and they got low error indices.

Fan et al. [Fan et al., (2020)] consumed Long Short Term Memory to predict Sea Wave Height with greater accuracy for various forecasting time horizons and developed a simulating wave's nearshore-LSTM to generate a single-point prediction. A lot of earlier research on SWH prediction has focused on employing superficial machine learning models like BP, SVM, etc. Still, they have failed to leverage the deep correlations between historical data over time fully. SWH has been effectively predicted using LSTM. However, one notable drawback of LSTM is that it requires many parameters for training. As a result, the training procedure is time-intensive and prone to overfitting.

Choi et al. [Choi et al., (2020)] proposed using deep neural network-based algorithms to estimate extreme wave heights in real-time from raw marine images. First, the authors calculated the appropriate wave height level using a single ocean picture. A classification model is built depending on CNN. Second, the authors proposed using a regression model to estimate significant wave heights from many marine images. This technique extracts Spatio-temporal information from time-series pictures using convolutional LSTM. Quach and colleagues [Quach et al., (2020)] Predicted SWH by extracting data from Synthetic Aperture Radar (SAR) images using a CNN. When compared to earlier efforts, their DL-based solution performed much better. This achievement established the viability and efficacy of deep convolutional techniques for the Sea Wave Height forecast. Gao et al. [Gao et al., (2021)] studied unique operational wave forecasting methods in the Bohai Sea, building wave height prediction models for three places. Using training samples of sea level wind and wave heights, this method is developed on a long short-term memory (LSTM) neural network.

As a result, the literature mentioned above studies have various limitations, and to overcome the constraints in existing works, a novel technique is required. As a result, our research provides a novel network that can accurately forecast wave heights, as discussed in the next phase.

3. Intelligent ocean wave height prediction system using light GBM model

Marine disasters cause serious damage to many countries worldwide, resulting in thousands of casualties and significant economic losses. Ships and marine or coastal infrastructure can be destroyed by ocean tides with a high Significant Wave Height. It is critical to anticipate sea wave heights rapidly and exactly in disaster warning and emergency preparedness and the construction and preservation of seaside and offshore infrastructure, marine transport, and ecological security. The existing research employs traditional empirical or numerical-based forecasting models to detect wave height, but it is complex and inaccurate; hence, our research introduces the light GBM, which predicts wave height over a half-hour period using a tree-based machine learning method. To filter out data instances and generate a split value, light GBM employs a design known as Gradient-based One-Side Sampling (GOSS).

We used wave speed, peak wave direction, zero up crossing wave period, wave period, and SWH as inputs. The light GBM model, over other soft computing models, enables us to input many features. All wave variables were used to train the model, including wave period and peak wave direction. Three new features, comprising considerable wave height data, were generated before training, including wavelength, wave number, and sea surface temperature. Before training the model, the current wave height column was removed from the data frame to avoid loss. Fig. 1 depicts the framework of the proposed method.

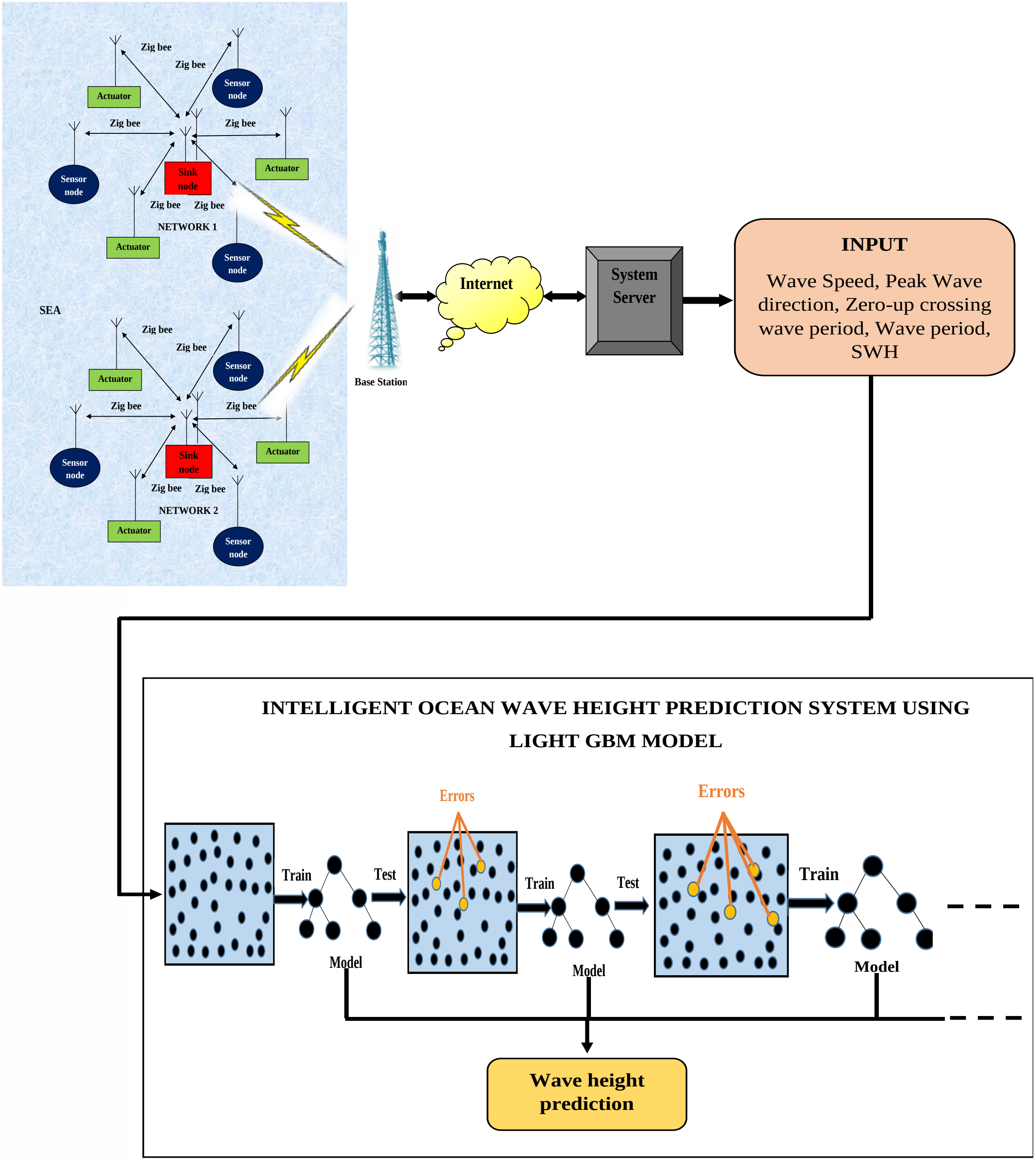

Fig. 1. Structure of the proposed method

3.1 Dataset description

This research collected the data from Coastal Data System – Waves (Mooloolaba) [26], in which wave characteristics were measured and calculated using data acquired by a Mooloolaba-based wave monitoring buoy. It is used to calculate sea levels, direction, and period. The motion (or heave) of a wave monitoring buoy as it drifts up and down with each incoming wave is monitored and processed electronically. The information recorded by the wave monitoring buoys is transmitted via radio signal to a nearby receiving station. The wave data is stored, analyzed, and summarized by a computer connected to the radio receiver. Wave monitoring station data is retrieved each hour, and fresh wave charts are posted to the web every 20 minutes. From 2017 until the present, we acquired 43729 data points, 70 percent of which were utilized for training and 30 percent for testing. Significant wave height (H_s), maximum wave height (H_{max}), zero-up crossing wave period (T_z), peak energy wave period (T_P), peak direction, and Sea surface temperature are the variables in the gathered datasets (SST).

Table 1. Properties of various features used in this dataset [26]

Variables	**Minimum**	**Maximum**	**Average**	**Standard Deviation δ**
Significant wave height (H_s)	0.294	4.257	1.238	0.53
Maximum wave height (H_{max})	0.51	7.906	2.09	0.897
Zero up-crossing wave period (T_z)	3.076	10.92	5.617	0.928
Peak energy wave period (T_P)	2.72	21.12	9.00	2.39
Peak direction	5	358	98.63	24.28
Sea Surface Temperature (SST)	19.8	28.65	23.95	2.23

From table 1, the variables extracted from the Mooloolaba dataset, such as H_s, H_{max}, T_z, T_P, peak direction and SST values are plotted in the following fig. 2.

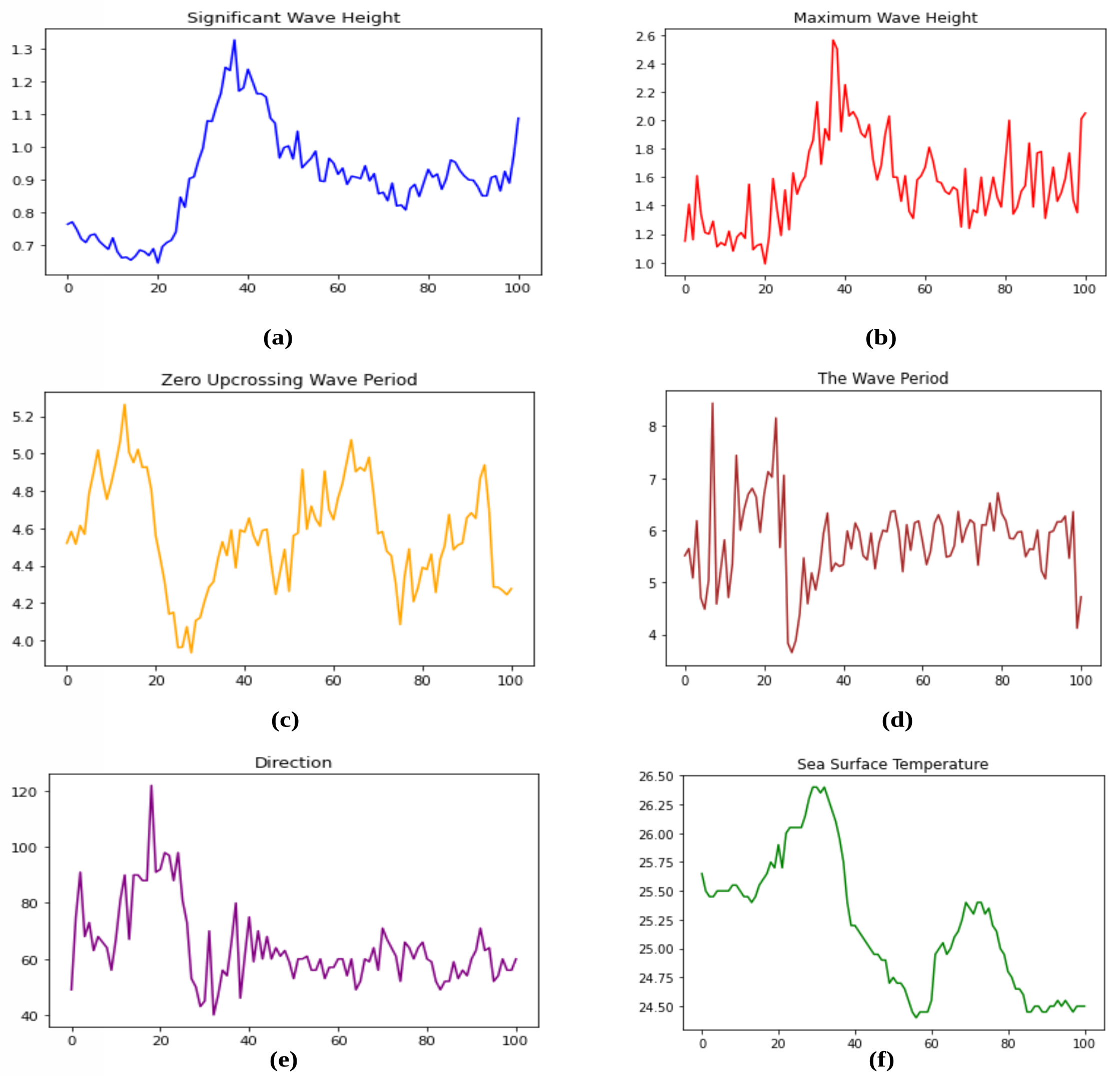

Fig. 2. Variables extracted from the dataset [26]

In fig. 2(a), H_s represents the significant wave height; while analyzing a wave record, the highest one-third of the wave heights in the record is taken into account. Fig. 2(b) shows that the maximum sea level (H_{max}), is the height of the highest single wave. Fig. 2(c) depicts the zero up-crossing wave period (T_z), which is the average of a wave record's zero-up crossing wave periods. Fig. 2(d) illustrates the peak energy wave period (T_P), which is the period of the waves that produce highest energy. Fig. 2(e) shows the peak direction, which is the angle measured in degrees from true north from which the largest waves are flowing.

The Sea Surface Temperature (SST) is measured in degrees Celsius by a wave measurement buoy, as shown in fig. 2(f). Furthermore, the wave speed (C) is calculated based on the following stated formulas:

Wavelength (L) = $gT^2/_{2\pi}$

Wave Number (k) = $2\pi/_{L}$

Angular Frequency (ω) = $2\pi/_{T_P}$

Wave Speed (C) = $\omega/_{k}$ (1)

The other factors are the hour of the day, the day of the month, and the month of the year. As a consequence, in this study, the wave frequency, peak wave direction, zero up crossing wave period, wave period, and SWH were employed as inputs. These values are then passed to the light GBM, which will be discussed in further detail in the following section.

3.2 Proposed light gradient boosting machine (lgbm) approach for wave height prediction

Traditional empirical or numerical-based forecasting models are used in existing research. However, they are slow and inaccurate. Our research uses light GBM, a tree-based machine-learning algorithm, to accurately estimate wave height with minimum training time. light GBM is an ensemble technique that uses Decision Trees (DT), which perform data trapping and improving. Trapping and improving are two common statistical techniques for increased accuracy.

Furthermore, light GBM employs the forward distributing technique. The residue is matched by a negative slope in each iteration of learning a decision tree. First, prepare the training dataset as $D_z = \{(x_{zi}, y_{zi})\}_{i=1}^{n_z}, x_{zi} \in R_z^{m_z+1}, y_{zi} \in R_z$, where n_z is the number of samples. $x_{zi}(i = 1,2 \dots n_z)$ is m_z dimensional input vector and $m_z = N_{zw_z} \times N_{zf_z} + C$, where N_{zw_z} is the size of the time frame, N_{zf_z} is a number of characteristics that have been chosen, C is the wave speed. $y_{zi}(i = 1,2 \dots n_z)$ is one of the dimensional wave height predictions is presented in fig. 3.

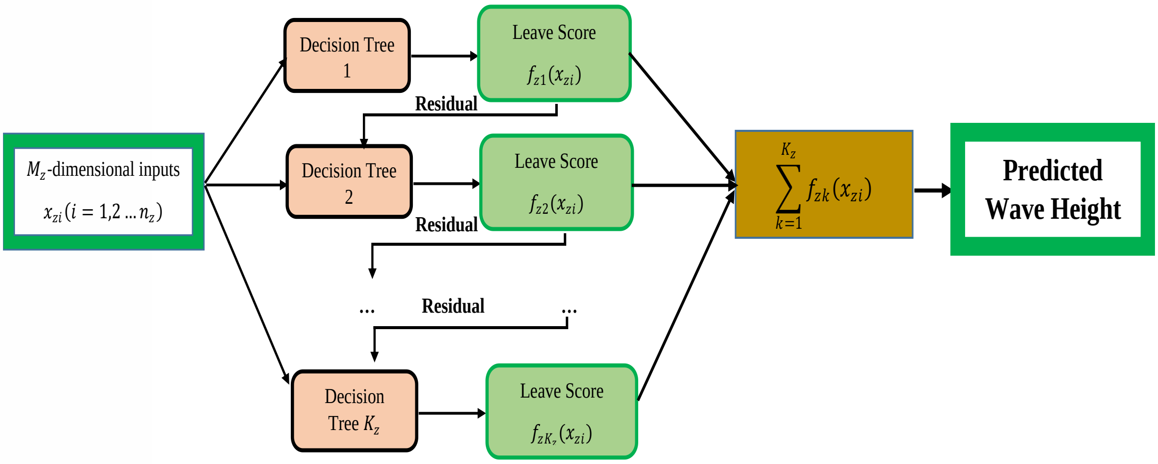

Fig. 3. Light GBM structure for wave height prediction

The height of the waves is calculated by aggregating the prediction of each tree in a group of trees:

$$\widehat{y_{zi}} = \sum_{k=1}^{K_z} f_{zk}(x_{zi}), f_{zk} \in F_z \quad (2)$$

where K_z is the total number of trees, F_z is a space containing all potential tree structures, and f_{zk} is one of the trees with leaf scores. By decreasing the objective, f_{zk} is obtained:

$$f_{zk} = \arg\min_{f_{zk}} \sum_{i=1}^{n_z} L_z(y_{zi}, \widehat{y_{zi}}^{(k)}) + \Omega(f_{zk}) \quad (3)$$

Where L_z is the loss function for training and Ω is the regularisation purpose, generally taken by the equation (4):

$$\Omega(f_{zk}) = \gamma T_z + \frac{1}{2}\lambda \sum_{j=1}^{T_z} \omega_j^2 \quad (4)$$

Where γ is the penalty parameter for the number of leaves T_z, and ω is the leaves' weights. When L_z employs a squared error loss function, the loss becomes:

$$L_z\left(y_z, \widehat{y}_z^{(k-1)} + f_{zk}(x_z)\right) = [y_z - \widehat{y}_z^{(k-1)} - f_{zk}(x_z)]^2 \quad (5)$$

$$= [r_z - f_{zk}(x_z)]^2$$

f_{zk} is obtained using residual fitting r_z. Using a quadratic approximation, at round k, we can define the function to minimize as:

$$f_{zk} \simeq \arg\min_{f_{zk}} \sum_{i=1}^{n_z} [g_{zi}, f_{zk}(x_{zi}) + \frac{1}{2} h_{zi} f_{zk}^2(x_{zi})] + \Omega(f_{zk}) \quad (6)$$

Where $g_{zi} = \partial_{\widehat{y}_{z(k-1)}} L_z\left(y_{zi}, \widehat{y}_z^{(k-1)}\right), h_{zi} = \partial^2_{\widehat{y}_{z(k-1)}} L_z\left(y_{zi}, \widehat{y}_z^{(k-1)}\right)$. As a consequence, minimizing this objective function generates a new tree. Furthermore, the decision tree divides each node based on how it acquires information. Eq. (7) gives the variance gain of splitting feature j at location d_z for a node:

$$V_{zj\backslash O_Z}(d_z) = \frac{1}{n_{O_Z}}\left\{\frac{\left(\sum\{x_{zi}\in O_Z: x_{zij}\le d_z\} g_{zi}\right)^2}{n^j_{zl\backslash O_Z}(d_z)} + \frac{\left(\sum\{x_{zi}\in O_Z: x_{zij}> d_z\} g_{zi}\right)^2}{n^j_{r_z\backslash O_Z}(d_z)}\right\} \quad (7)$$

where O_Z is the number of samples on a fixed decision tree node,

$n_{O_Z} = \sum I_z[x_{zi} \in O_Z]$, $n^j_{zl\backslash O_Z}(d_z) = \sum I_z[x_{zi} \in O_Z: x_{zij} \le d_z]$, $n^j_{r_z\backslash O_Z}(d_z) = \sum I_z[x_{zi} \in O_Z: x_{zij} > d_z]$.

All samples must be scanned to select the optimum partition point to calculate the information gain. When dealing with samples with enormous numbers and dimensions derived from significant wave height, efficacy and accuracy are challenging to achieve. Light GBM employs the Gradient-based One-Side Sampling (GOSS) approach to reduce the quantity of training information generated when a node divides, as shown in Eq. (7). To avoid loss, the current wave height column was deleted from the data frame before training the model.

$$V_{zj}(d_z) = \frac{1}{n_{O_Z}}\left\{\frac{\left(\sum_{x_{zi}\in A_{zl}} g_{zi} + \frac{1-a}{b}\sum_{x_{zi}\in B_{zl}} g_{zi}\right)^2}{n^j_{zl}(d_z)} + \frac{\left(\sum_{x_{zi}\in A_{zr_z}} g_{zi} + \frac{1-a}{b}\sum_{x_{zi}\in B_{zr_z}} g_{zi}\right)^2}{n^j_{r_z}(d_z)}\right\} \quad (8)$$

where A_z is a subset of the top a×100% examples with greater inclines, and B_z is a subset randomly selected from the outstanding set of (1-a) ×100% instances with lesser gradients. $A_{zl} = \{x_{zi} \in A_z: x_{zij} \le d_z\}$, $A_{zr_z} = \{x_{zi} \in A_z: x_{zij} > d_z\}$, $B_{zl} = \{x_{zi} \in B_z: x_{zij} \le d_z\}$, $B_{zr_z} = \{x_{zi} \in B_z: x_{zij} > d_z\}$.

The GOSS approach determines the split point by computing the $V_{zj}(d_z)$ rather than all occurrences, with a narrower sample of instances, reducing computing load and increasing noise signal redundancy. The following figure 4 provides the overall flowchart of the proposed light GBM approach.

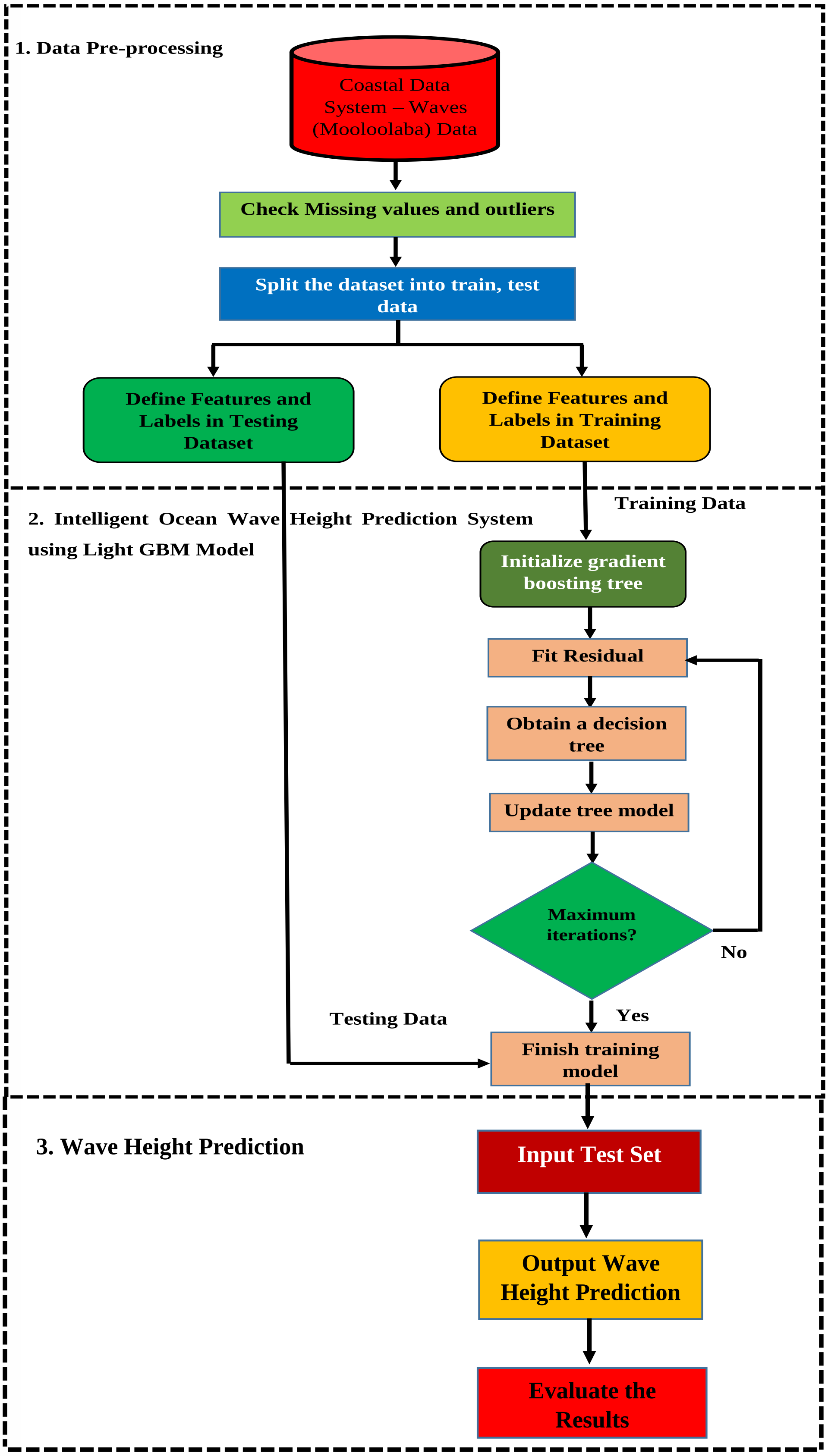

Fig. 4. Flowchart of the proposed approach

As a result, this research attains higher accuracy. Moreover, the following section discusses the performance and comparison results of the proposed method.

4. Result and discussion

This sector details the proposed approach's implementation outcomes, as well as its performance and comparative findings.

Tool : PYTHON 3
OS : Windows 7 (64 bit)
Processor : Intel Premium
RAM :8GB RAM

4.1 Performance parameters

The performance parameters of the proposed Light Gradient Boosting Machine technique are explained in this section.

4.1.1 Training and validation accuracy

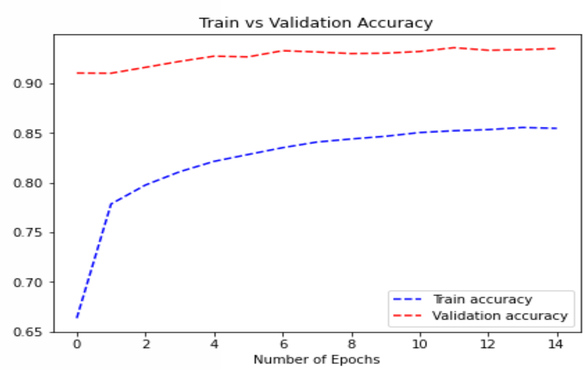

Fig. 5. Accuracy in training and validation

From fig. 5, Accuracy in Training and Validation are 0.83 and 0.926 at epoch 14. As a result, the validation accuracy is greater than the training accuracy by using our proposed light GBM approach, which shows the effectiveness of the proposed approach.

4.1.2 Training and validation loss

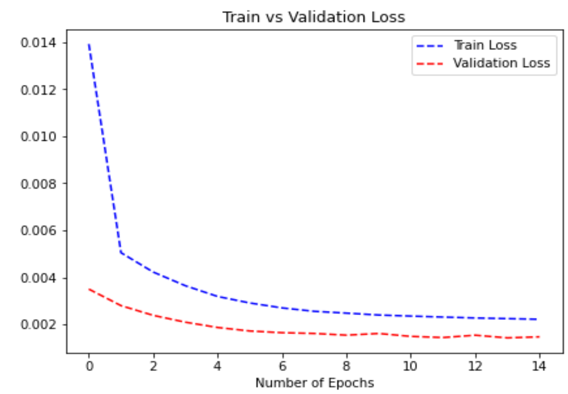

Fig. 6. Loss of training and validation

The training and validation loss is 0.0032 and 0.002 at epoch 14, respectively. From fig. 6, the validation loss is lesser than the training loss by using our proposed light gradient boosting machine approach.

4.1.3 Significant wave height (H_s) *prediction*

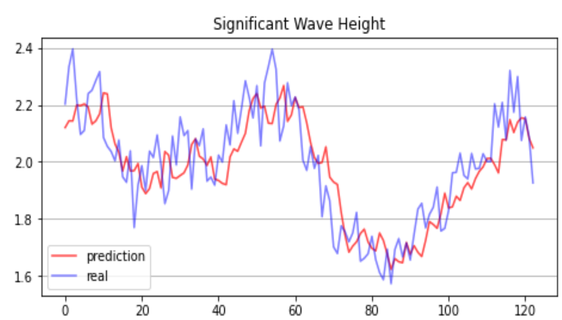

Fig. 7. Substantial wave height (H_s) prediction

Fig. 7 illustrates the major wave height prediction findings. This substantial wave height (H_s) is predicted by using our proposed machine learning light gradient boosting machine approach. The anticipated wave height is substantially comparable to the actual significant wave height, demonstrating the efficacy of the proposed technique.

4.1.4 Root mean square error (RMSE) analysis

To assess the presentation of our method, we use measures such as Root Mean Squared Error (RMSE).

Root Mean Square Error (RMSE)

To obtain the Root Mean square error, compute the residual (the difference among prediction and reality) for every portion of information, the norms of the residual for each piece of data, the average of residuals, and the square root of that mean. RMSE is widely used in supervised training systems since it utilizes and needs real observations at each projected data point. The square error is determined using the following formula:

$$\text{RMSE}=\sqrt{\frac{\sum_{i=1}^{N_z}(x_{zi}-\widehat{x_{zi}})^2}{N_z}} \tag{9}$$

Where, x_{zi} are temporal sequence of genuine observations, $\widehat{x_{zi}}$ is the time series estimation, N_z is the number of missing points.

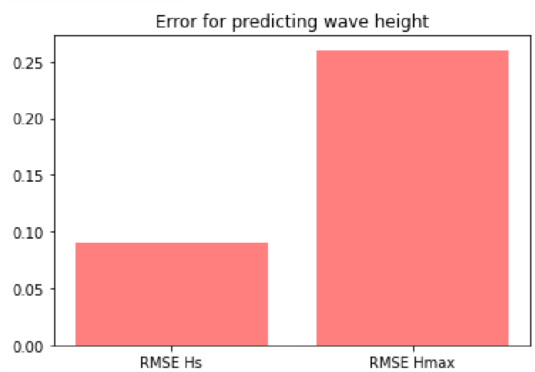

Fig. 8. RMSE for $\boldsymbol{H_s}$ and $\boldsymbol{H_{max}}$

Fig. 8 illustrates the RMSE for the significant wave height and maximum tallness. The obtained RMSE values of H_s, H_{max} are 0.092 and 0.262, respectively, by using our proposed light GBM approach.

4.2 Comparison analysis

This section compares the proposed approach to other current techniques. In terms of NMSE, MSE, MAPE, and R values, the developed model's forecasting performance is compared to that of ANN models such as Multilayer Perceptron's Neural Network (MPNN) [Elbisy & Elbisy, (2021)], Cascade Correlation Neural Network (CCNN) [Elbisy & Elbisy, (2021)] Radial Basis Function Neural Network (RBFNN) [Elbisy & Elbisy, (2021)] and General Regression Neural Network (GRNN) [Elbisy & Elbisy, (2021)].

4.2.1 Comparison of normalized mean squared error

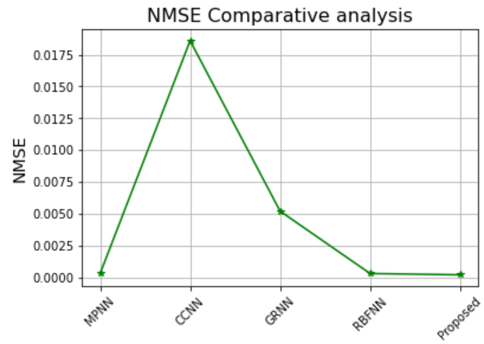

Fig. 9. Comparison of normalized mean squared error

The overall Normalized Mean Squared Error comparison is shown in Fig. 9. The NMSE of the proposed technique improves by using light GBM. Our proposed approach attains lesser error when compared to the baseline as Multilayer Perceptron Neural Network (MPNN) [Elbisy & Elbisy, (2021)], Cascade Correlation Neural Network (CCNN) [Elbisy & Elbisy, (2021)], General Regression Neural Network (GRNN) [Elbisy & Elbisy, (2021)], and Radial Basis Function Neural Network (RBFNN) [Elbisy & Elbisy, (2021)] such as 0.0003, 0.0186, 0.0052, and 0.0003 respectively. As a result, our novel technique has an error of 0.0002, which is less than baseline approaches.

4.2.2 Comparison of mean squared error

The Mean squared error (MSE) is used to calculate the grade of error in statistical models. It is determined by the average squared difference between detected and predicted principles.

$$\text{MSE} = \frac{1}{n_z}\sum_{i=1}^{n_z}(y_{zi}-\widehat{y_{zi}})^2 \tag{10}$$

Where, n_z is the number of information sets, y_{zi} are the observed values, $\widehat{y_{zi}}$ is the forecasted value.

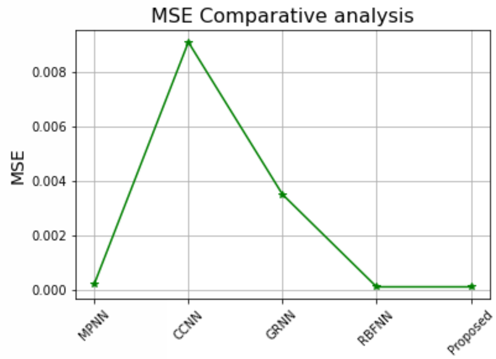

Fig. 10. Comparison of mean squared error (MSE)

The overall MSE comparison is shown in Fig. 10. The MSE of the proposed technique improves by using light GBM. Our proposed approach attains lesser error when compared to the baseline as Multilayer Perceptron Neural Network (MPNN) [Elbisy & Elbisy, (2021)], Cascade Correlation Neural Network (CCNN) [Elbisy & Elbisy, (2021)], General Regression Neural Network (GRNN) [Elbisy & Elbisy, (2021)], and Radial Basis Function Neural Network (RBFNN) [Elbisy & Elbisy, (2021)] such as 0.0002, 0.0091, 0.0035, and 0.0001 respectively. As a result, our novel technique has an MSE of 0.0001, which is less than the baseline approaches.

4.2.3 Comparison of mean absolute percentage error (MAPE)

The Mean absolute percentage error (MAPE) of a predicting system is used to assess its accuracy. It computes the average absolute percentage inaccuracy of each entry in a dataset to determine how close the predicted quantities were to the actual amounts.

$$\text{MAPE}\ (\%) = \frac{1}{n_z}\sum_{t=1}^{n_z}\left|\frac{A_{zt}-F_{zt}}{A_{zt}}\right| \tag{11}$$

Where, n_z is the number of times the summation is iterated, A_{zt} is the real rate, F_{zt} is the forecasted rate.

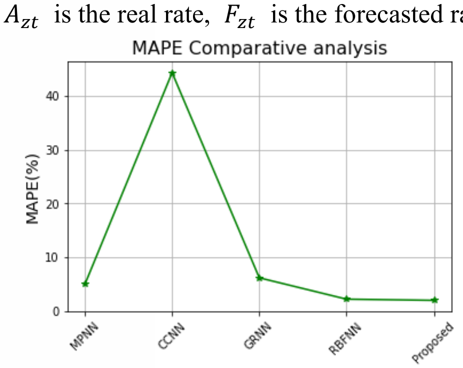

Fig. 11. Comparison of MAPE

The overall Mean Absolute Percentage Error comparison is shown in Fig. 11. The MAPE of the proposed technique improves by using light GBM. Our proposed approach attains lesser error when compared to the baseline as Multilayer Perceptron Neural Network (MPNN) [Elbisy & Elbisy, (2021)], Cascade Correlation Neural Network (CCNN) [Elbisy & Elbisy, (2021)], General Regression Neural Network (GRNN) [Elbisy & Elbisy, (2021)], and Radial Basis Function Neural Network (RBFNN) [Elbisy & Elbisy, (2021)] such as 5.01%, 44.33%, 6.22%, and 2.23%. As a result, our novel technique has a MAPE of 2.21%, which is less than the baseline approaches.

4.2.4 Comparison of correlation coefficient (R)

The correlation coefficient is formulated as follows:

$$R = \frac{\sum_{i=1}^{N_z}(P_{zi}-\overline{P_{zi}})(O_{zi}-\overline{O_{zi}})}{\sqrt{\sum_{i=1}^{N_z}(P_{zi}-\overline{P_{zi}})^2\sum_{i=1}^{N_z}(O_{zi}-\overline{O_{zi}})^2}}$$

Where, $O_{zi}, P_{zi}, N_z, \overline{O_{zi}}$ and $\overline{P_{zi}}$ indicates the observed value, the forecasted rate, the observed number, the detected mean rate, and the forecast mean rate, in that order.

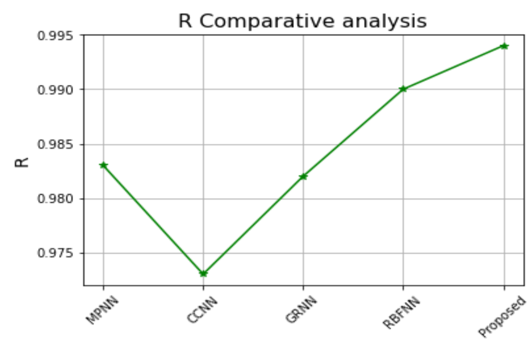

Fig. 12. Comparison of correlation coefficient (R)

The overall R comparison is shown in Fig. 12. The R-value of the proposed technique improves by using light GBM. Our proposed approach attains a higher R-value when compared to the baseline like Multilayer Perceptron Neural Network (MPNN) [Elbisy & Elbisy, (2021)], Cascade Correlation Neural Network (CCNN) [Elbisy & Elbisy, (2021)], General Regression Neural Network (GRNN) [Elbisy & Elbisy, (2021)], and Radial Basis Function Neural Network (RBFNN) [Elbisy & Elbisy, (2021)] such as 0.983, 0.973, 0.982, and 0.99 respectively. As a result, our novel technique has a Correlation Coefficient of 0.994, which is higher than baseline approaches.

As a result, the error such as RMSE, NMSE, MSE, and MAPE values obtained by using our proposed approach as 0.092, 0.0002, 0.0001, and 2.21 respectively, which is outperformed when compared to the existing techniques. Also, the accuracy is high while there are less errors.

5. Conclusion

The LightGBM approach is proposed in this research for predicting sea wave height. The proposed approach is validated using the Mooloolaba Coastal Wave dataset. Processing of data, feature selection, and time window processing all contributed to strong performance on both the training and testing datasets. In comparison to earlier ensemble approaches and neural networks, the proposed method achieves high accuracy and efficiency due to low error, and the prediction becomes increasingly accurate. Finally, the developed method attains less error when compared to findings obtained by existing state-of-the-art approaches. Additionally, we observe that the numerical model's forecast of wave height sometimes lags. The predictions may not come true because our model only considers the significant wave height as an input. In further study, we will incorporate sea surface wind, atmospheric pressure, and other variables as input and influences to examine the effects of the wave type and meteorological components. To improve accuracy, deep learning or other optimization algorithms might be examined in future.

References

Abhigna, P., Jerritta, S., Srinivasan, R. & Rajendran, V. (2017, April). Analysis of feed forward and recurrent neural networks in predicting the significant wave height at the moored buoys in Bay of Bengal, In 2017 International Conference on Communication and Signal Processing (ICCSP),1856-1860. IEEE.

Altunkaynak, A. & Wang, K.H. (2012). Estimation of significant wave height in shallow lakes using the expert system techniques, Expert Systems with Applications, 39(3), 2549-2559.

Amin, A.E. (2013). A novel classification model for cotton yarn quality based on trained neural network using genetic algorithm, Knowledge-Based Systems, 39, 124-132.

Choi, H., Park, M., Son, G., Jeong, J., Park, J., Mo, K. & Kang, P. (2020). Real-time significant wave height estimation from raw ocean images based on 2D and 3D deep neural networks, Ocean Engineering, 201, 107129.

Cornejo-Bueno, L., Nieto-Borge, J.C., García-Díaz, P., Rodríguez, G. & Salcedo-Sanz, S. (2016). Significant wave height and energy flux prediction for marine energy applications: A grouping genetic algorithm–Extreme Learning Machine approach, Renewable Energy, 97, 380-389.

Deo, M.C., Jha, A., Chaphekar, A.S. & Ravikant, K. (2001). Neural networks for wave forecasting, Ocean engineering, 28(7), 889-898.

Elbisy, M.S. & Elbisy, A.M. (2021). Prediction of significant wave height by artificial neural networks and multiple additive regression trees, Ocean Engineering, 230, 109077.

Fan, S., Xiao, N. & Dong, S. (2020). A novel model to predict significant wave height based on long short-term memory network, Ocean Engineering, 205, 107298.

Gao, S., Huang, J., Li, Y., Liu, G., Bi, F. & Bai, Z. (2021). A forecasting model for wave heights based on a long short-term memory neural network, Acta Oceanologica Sinica, 40(1), 62-69.

Hashim, R., Roy, C., Motamedi, S., Shamshirband, S. & Petković, D. (2016). Selection of climatic parameters affecting wave height prediction using an enhanced Takagi-Sugeno-based fuzzy methodology, Renewable and Sustainable Energy Reviews, 60, 246-257.

https://www.data.qld.gov.au/dataset/coastal-data-system-waves-mooloolaba

Mahjoobi, J. & Etemad-Shahidi, A. (2008). An alternative approach for the prediction of significant wave heights based on classification and regression trees, Applied Ocean Research, 30(3), 172-177.

Mahjoobi, J. & Mosabbeb, E.A. (2009). Prediction of significant wave height using regressive support vector machines, Ocean Engineering, 36(5), 339-347.

Malekmohamadi, I., Bazargan-Lari, M.R., Kerachian, R., Nikoo, M.R. & Fallahnia, M. (2011). Evaluating the efficacy of SVMs, BNs, ANNs and ANFIS in wave height prediction. Ocean Engineering, 38(2-3), 487-497.

Nikoo, M.R., Kerachian, R. & Alizadeh, M.R. (2018). A fuzzy KNN-based model for significant wave height prediction in large lakes, Oceanologia, 60(2), 153-168.

Nitsure, S.P., Londhe, S.N. & Khare, K.C. (2012). Wave forecasts using wind information and genetic programming, Ocean Engineering, 54, 61-69.

Prahlada, R. & Deka, P.C. (2015). Forecasting of time series significant wave height using wavelet decomposed neural network, Aquatic Procedia, 4, 540-547.

Quach, B., Glaser, Y., Stopa, J.E., Mouche, A.A. & Sadowski, P. (2020). Deep learning for predicting significant wave height from synthetic aperture radar, IEEE Transactions on Geoscience and Remote Sensing, 59(3), 1859-1867.

Sabatier, F. (2007). US Army Corps of Engineers, Coastal Engineering Manual (CEM), Engineer Manual 1110-2-1100. US Army Corps of Engineers, Washington, DC (6 volumes), Méditerranée. Revue géographique des pays méditerranéens/Journal of Mediterranean geography, (108), 146.

Son, G., Choi, H., Lee, J.H. & Kang, P. (2020). Significant Wave Height Regression from a Raw Ocean Image with Convolutional LSTM and 3D Convolutional Networks, J. Korean Op. Res. Manag. Sci. Soc, 45, 11-24.

Wang, W., Tang, R., Li, C., Liu, P. & Luo, L. (2018). A BP neural network model optimized by mind evolutionary algorithm for predicting the ocean wave heights, Ocean Engineering, 162, 98-107.

Wei, C.C. & Hsieh, C.J. (2018). Using adjacent buoy information to predict wave heights of typhoons offshore of northeastern Taiwan. Water, 10(12), 1800.

WMO, Guide to Wave Analysis & Forecasting, World Meteorological organization, WMO-No.702. (1998).

Yang, S., Xia, T., Zhang, Z., Zheng, C., Li, X., Li, H. & Xu, J. (2019). Prediction of significant wave heights based on CS-BP model in the south China sea, IEEE Access, 7, 147490-147500.

Yoon, H., Jun, S.C., Hyun, Y., Bae, G.O. & Lee, K.K. (2011). A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer, Journal of hydrology, 396(1-2), 128-138.

Zamani, A., Solomatine, D., Azimian, A. & Heemink, A. (2008). Learning from data for wind–wave forecasting, Ocean engineering, 35(10), 953-962.

Zhang, X. & Dai, H. (2019). Significant wave height prediction with the CRBM-DBN model, Journal of Atmospheric and Oceanic Technology, 36(3), 333-351.

Autobiographies:

Jai Prakash Mishra is a Research Scholar at Manipal University Jaipur, Jaipur, India. He received his M.Tech degree in Embedded System in 2014 from Jaipur National University, Jaipur, India. Currently he is working on "IoT Based Sea Wave Turbulence Monitoring System using Multiple MEMS Sensors".

Kulwant Singh is working in the area of MEMS design and fabrication since July 2007. He began his research work as M.Tech trainee in the area of design and fabrication of MEMS pressure sensor at CSIR-CEERI, Pilani, India. After completing his M.Tech. (Electronics Design and Technology) from Tezpur University, Assam, in 2008, he joined as an assistant professor in the Dept. of ECE, SRM University, India. In 2010, he joined as a Ph.D. Scholar at Nanomaterials and Devices Research Laboratory, SNST, NIT Calicut and worked in collaboration with CSIR-CEERI. At present he is working as Associate Professor, in the Department of Electronics and Communication Engineering, Manipal University Jaipur, Rajasthan. His current fields of interest are micro/nano electronics, device design and fabrication and MEMS/NEMS sensors.

Himanshu Chaudhary is working as an Associate Professor, in the Department of Electronics and Communication Engineering, Manipal University Jaipur, Rajasthan since 2016. He received his M.E. degree in Automatic Controls & Robotics in 2000 from the Electric al Engineering Department of Maharaja Sayajirao University, Baroda, Gujarat, India. He has completed his PhD from the Department of Electrical Engineering from IIT Roorkee in 2014 in Robotics. He has around 20 years of teaching and industry experience in Electronics and Communication Engineering. He has published various research papers in SCI Indexed Journals, and he is Certified Peer Reviewer of IEEE, Elsevier and various international journals of repute. His research interests include Soft Computing, Intelligent control, Artificial Intelligence, Fuzzy Control, Automation, Robotic and MEMS Sensors.