Results in Engineering 24 (2024) 103573

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

Research paper

Data-driven models for significant wave height forecasting: Comparative analysis of machine learning techniques

Ahmet Durap [a,b,c,*]

[a] Engineering and Natural Sciences Faculty, Istanbul Medipol University, 34181, Beykoz, İstanbul, Turkey
[b] Division of Coastal Sciences and Engineering, Civil Engineering Department, Civil Engineering Faculty, Istanbul Technical University, Maslak 34469, Istanbul, Turkey
[c] Division of Civil Engineering, The University of Queensland, Brisbane, 4072, Australia

ARTICLE INFO

Keywords:
Comparative machine learning techniques
Significant wave height (SWH) prediction
Dynamic coastlines
Wave characteristics
Stradbroke Island
Queensland

ABSTRACT

Accurate prediction of significant wave height (SWH) is critical for coastal safety, marine operations, and disaster management. Traditional numerical models for wave prediction are computationally intensive and often lack accuracy, prompting a shift towards data-driven methods. This study explores the efficacy of machine learning (ML) models in forecasting SWH in the coastline of North Stradbroke Island, Queensland, Australia. Using a dataset spanning 2010 to 2022, the study employs wave characteristics such as maximum wave height (Hmax), wave periods (Tz, Tp), peak direction, and sea surface temperature (SST) as predictors. Three ML models—Linear Regression, Decision Tree, and Random Forest—were trained and evaluated using performance metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2). The Random Forest model demonstrated the best predictive accuracy with the lowest MSE (0.0074) and highest R^2 (0.958), outperforming both Linear Regression and Decision Tree models. This improvement in prediction accuracy supports the model's application for coastal management, ensuring better forecasting of wave conditions. The proposed approach shows significant advantages, including better handling of non-linearities and reduced computational costs compared to conventional numerical methods, such as SWAN and WAM, making it highly applicable for real-time wave forecasting, particularly for regions with complex coastal dynamics such as North Stradbroke Island.

1. Introduction

Extreme marine phenomena impact human life, impede coastal city development, and threaten human well-being [1].Hurricanes, in particular, are extremely dangerous marine weather phenomena that endanger the security and welfare of people living in coastal cities [2]. The passage of hurricanes entails intense winds and rainfall. Simultaneously, the enormous waves generated by hurricanes can cause ships to capsize, damage sea platforms, and result in significant disasters for sea transportation, construction, and fisheries [3]. Accurate wave prediction can mitigate specific disasters and decrease economic losses in marine transportation, offshore oil and gas exploration, marine scientific research, water operations, and military activities [4].

The marine meteorological environment is highly intricate, characterized by a combination of numerous waves originating from various directions, each with different amplitudes and periods [5–7]. A single wave is often not indicative. As a result, significant wave height (SWH) is used as the main ocean wave measurement metric [8,9]. Hence, precise analysis and forecasting of ocean waves can furnish essential marine meteorological reference parameters for the normal progression of maritime activities and prevent the detrimental impacts of extreme marine meteorological events on human production and life (" [10,11])

SWH is a crucial parameter in coastal engineering, marine operations, and disaster management, influencing the design and resilience of coastal infrastructure. SWH plays a vital role in various applications, such as determining the resilience of coastal infrastructure to wave attacks [12], assessing wave overtopping risks [13], predicting wave runup [14], scouring [15], pipe soil interactions [16], wave damping caused by breakwater [17], waves resonance and understanding coastal flow slides (=dilative slope failure) mechanisms. Accurately predicting SWH is essential for mitigating risks associated with coastal erosion, coastal vulnerability [18,19] port design, and marine energy generation [20]. Furthermore, SWH forecasting is key for developing sustainable wave energy systems and safeguarding offshore installations.

SWH prediction is indispensable for a wide range of activities. It supports coastal safety by enabling the forecasting of severe weather

* Corresponding author.
E-mail address: ahmetdurap@gmail.com.

https://doi.org/10.1016/j.rineng.2024.103573
Received 19 August 2024; Received in revised form 27 November 2024; Accepted 27 November 2024
Available online 1 December 2024

events, enhances maritime operations through improved fuel efficiency and navigation, and is essential for aquaculture, military operations, and the developing wave energy sector [21–24]. Accurate SWH and wave period modeling is critical for optimizing wave energy extraction [25].

The growth of offshore industries, particularly marine energy and oil and gas, has highlighted the need for refined wave data. Numerous studies have been conducted to improve the accuracy and efficiency of wave characterization methods to support the design, operation, and management of offshore assets [26,27]. Over the years, different methods have been developed, ranging from traditional numerical models to advanced deep learning techniques.

The Seasat mission, launched in 1978, was a groundbreaking satellite dedicated to studying the Earth's oceans [28]. One of its primary instruments was a radar altimeter, which emitted radar pulses towards the ocean surface and measured the time it took for the signal to return. This technique allowed scientists to determine the height of the ocean surface relative to a reference ellipsoid [28–30].

To extract valuable information about ocean waves from the altimeter data, scientists employed a process called waveform retracking. This technique involved modeling the waveform and identifying key features, such as the leading edge and peak amplitude. By analyzing these features, researchers could estimate the significant wave height (SWH), a measure of the average height of the highest one-third of the waves.

While Seasat's radar altimeter provided valuable insights into ocean wave conditions, it faced certain limitations. For instance, Seasat altimeter required significant calibration and validation efforts to ensure the accuracy of altitude, wave height, and wind speed measurements [29,30]. Various atmospheric and geophysical effects influenced the measurement accuracies, necessitating correction models. A significant wave height-dependent bias was identified in Seasat-measured surface heights, which could not be neglected for high-accuracy applications. The altimeter's ability to resolve short wavelength geoid features was limited, with high coherence only for wavelengths longer than 70 km [31]. Oceanographic effects were more significant than instrument noise in limiting geoidal resolution thereby affecting its ability to measure low wave heights and dynamic height changes accurately [32].

The advent of radar altimetry with the 1985 Geosat mission marked a significant milestone in global oceanography. These satellite-borne instruments transmit radar pulses to the sea surface and record the returning signals. To derive crucial parameters like SWH, the received waveforms undergo a rigorous analysis process known as waveform retracking. This technique involves modeling the waveform and extracting relevant information [33].

Earlier studies have indicated that altimeter-derived SWH measurements are on par with buoy measurements in open ocean areas. However, validation studies show that altimeter SWH data from Sentinel-3 and Jason-3 are generally accurate in coastal regions, but errors increase near the shore due to tidal currents and water depth variations [34]. Sentinel-6 and SARAL/AltiKa altimeters show high accuracy in open ocean but face challenges in coastal areas due to environmental interferences [33,35].

As can be seen that early methods relied heavily on complex numerical models, which were resource-intensive and computationally expensive [36].

Numerical models, including SWAN [37,38], WAM, Wave Watch III [39–41], and JONSWAP [42], have been employed to study the growth and decay of wind-wave energy spectra. These models are used to predict wave characteristics in various ocean conditions, considering factors such as wind fetch, duration, and spatial variability. Numerical models for wave prediction are computationally intensive and have been criticized for their insufficient accuracy [43,44]. Moreover, the requirement for detailed local bathymetry data makes it difficult [44] to apply these models consistently across diverse oceanic regions [43]. Given the challenges posed by numerical methods in wave parameter prediction, researchers have shifted their focus to data-driven modeling. These approaches harness the power of data to uncover hidden relationships between variables, offering a more efficient and potentially accurate solution [45]. Data-driven modeling has become a popular approach for predicting wave parameters in coastal and ocean engineering. These models utilize meteorological and sensor data to estimate spatially variable wave characteristics, including significant wave height and period. The field has experienced rapid advancements in the application of data-driven methods for wave forecasting over the past decade [46].

Past buoy data is commonly used to train forecasting models for specific locations. [47] made a foundational contribution by applying an ANN to predict wave period and height using wind data at three Indian sites, demonstrating the feasibility of data-driven approaches for wave forecasting.

Several studies, including those by [48–51] have utilized ANNs for wave forecasting. [52]compared SVM with different kernel functions to ANNs and found SVM with RBF kernel to be more effective for wave height prediction. [53] evaluated multiple soft computing techniques, including ANNs, SVMs, Bayesian networks, and ANFIS, for wave height forecasting. [54]proposed a genetic programming approach as an alternative to numerical models for predicting peak wave height. [55] compared MRAN and GAP-RBF with SVR and ELM for SWH prediction across multiple locations. [56]assessed the performance of MTEANN against SVR and ELM for wave energy flux forecasting at various time horizons. [57]further developed the MTEANN model. Ali et al. proposed a novel MLR-CWLS model for short-term wave height prediction. Furthermore, LSTM neural networks have been utilized to wave height forecasting [58–61].

Wavelet-based models are widely used in various fields for signal processing, time series forecasting, and other predictive tasks. However, these models often face challenges in accurately predicting real-world data due to several inherent limitations and improper implementations [62].

The performance of wavelet-based models can be unstable due to the empirical selection of wavelet layers, which affects the accuracy of predictions. Optimizing wavelet layers can improve prediction accuracy

The evolution of SWH prediction has transitioned from resource-intensive numerical models to advanced machine learning techniques. Modern approaches, including LR, LSTM, Decision Tree, Random Forest, and hybrid models, have significantly improved the accuracy and efficiency of SWH predictions. These advancements are crucial for various marine and coastal applications, ensuring better preparedness and safety.

The primary objective of this study is to investigate the effectiveness of machine learning models in predicting significant wave height (SWH) of North Stradbroke Island, Queensland, Australia. This study introduces a novel approach by incorporating a comparative analysis of three distinct machine learning models—Linear Regression, Decision Tree, and Random Forest—using a comprehensive dataset spanning from 2010 to 2022. This comparative analysis has not been extensively explored in previous research, particularly in the context of regions with complex coastal dynamics like North Stradbroke Island. Moreover, the novel part of this study is to investigate the impact of feature variables (maximum wave height (Hmax), wave periods (T_z,T_p), peak direction, and sea surface temperature (SST) on target variable (significant wave height, H_s), with a specific focus on these three ML models alongside their learning curves of the models to provide a deeper understanding of their performance and reliability. This study introduces a novel approach by incorporating feature and target variables, which has not been exclusively explored in previous research. The findings from this study are expected to provide new insights into SWH, which could have significant implications for coastal management, marine operations, and disaster management. Specifically, this study aims to answer the following research questions: Specifically, this study aims to answer the following research questions:

How do different machine learning models (Linear Regression, Decision Tree, and Random Forest) compare in terms of accuracy and reliability in predicting significant wave height?

What are the key predictors (e.g., maximum wave height, wave periods, sea surface temperature) that most significantly influence the accuracy of these models?

How can the integration of multiple wave characteristics enhance the predictive performance of machine learning models?

By addressing these questions, this study seeks to fill a critical knowledge gap in the field of coastal engineering and marine science. The novelty of this research lies in its comprehensive evaluation of machine learning models in a specific coastal region with unique wave dynamics. Previous studies have often focused on individual models or generalized approaches without considering the specific characteristics of the study area. This research, therefore, provides a more tailored and context-specific analysis, which is crucial for developing accurate and reliable wave height prediction models.

The paper is structured to first introduce the importance of accurate wave height prediction and discuss the limitations of existing methods. It then outlines the study area and data, detailing the ML models and algorithms used, along with the data preprocessing, model training, and evaluation processes. Following this, the results of the ML models are presented, including a comparison of their performance. The discussion highlights the implications of these findings for improving SWH prediction. The paper concludes with a summary of the key outcomes and acknowledges the limitations of the study.

2. Study area and data

2.1. Sites selection and description

The value of data is getting more and more indispensable for scientific researchers because of its critical role in driving evidence-based decision-making, enabling accurate predictions, validating hypotheses, and fostering innovation across diverse fields through advanced analytical and computational tools [63]. The region of interest for wave height prediction encompasses North Stradbroke Island, located in Moreton Bay, Queensland, Australia. The island is situated approximately 30 km southeast of Brisbane (Fig. 1).

The economy and industry of coastal regions, including North Stradbroke Island, are heavily influenced by marine and coastal activities. Accurate SWH predictions can significantly benefit the local economy by improving the safety and efficiency of maritime operations, which are vital for trade and transportation. Enhanced wave forecasts can aid in the planning and execution of offshore projects, such as oil and gas exploration and marine renewable energy installations, thereby boosting industrial growth.

Furthermore, the tourism industry, which is a major economic driver for many coastal regions, can benefit from reliable wave height predictions. Ensuring the safety of water sports enthusiasts and beachgoers can enhance the attractiveness of coastal destinations. Additionally, accurate wave forecasts can help in the management and preservation of coastal ecosystems, which are crucial for maintaining biodiversity and supporting fisheries.

By addressing these economic and industrial impacts, this study underscores the broader significance of improving SWH prediction models. The findings from this research are expected to contribute to safer and more efficient coastal and marine operations, ultimately supporting the sustainable development of coastal regions.

This region was chosen for its critical importance in wave height prediction, attributed to the following key factors:

i. The island boasts a diverse coastline featuring various beaches, points, and bays [64]. This complexity creates unique wave refraction and diffraction patterns, influencing the distribution and intensity of waves across different areas of the island (Edward [65]).
ii. North Stradbroke Island's eastern coastline is directly exposed to the vast expanse of the Pacific Ocean. This exposure makes the island susceptible to receiving swells generated by distant storms and weather systems, leading to larger waves and a greater potential for hazardous conditions [66].
iii. The ocean floor surrounding North Stradbroke Island exhibits varying depths and contours [67]. These bathymetric features interact with incoming swells, modifying wave heights and directions as they approach the island. Shallow areas can cause

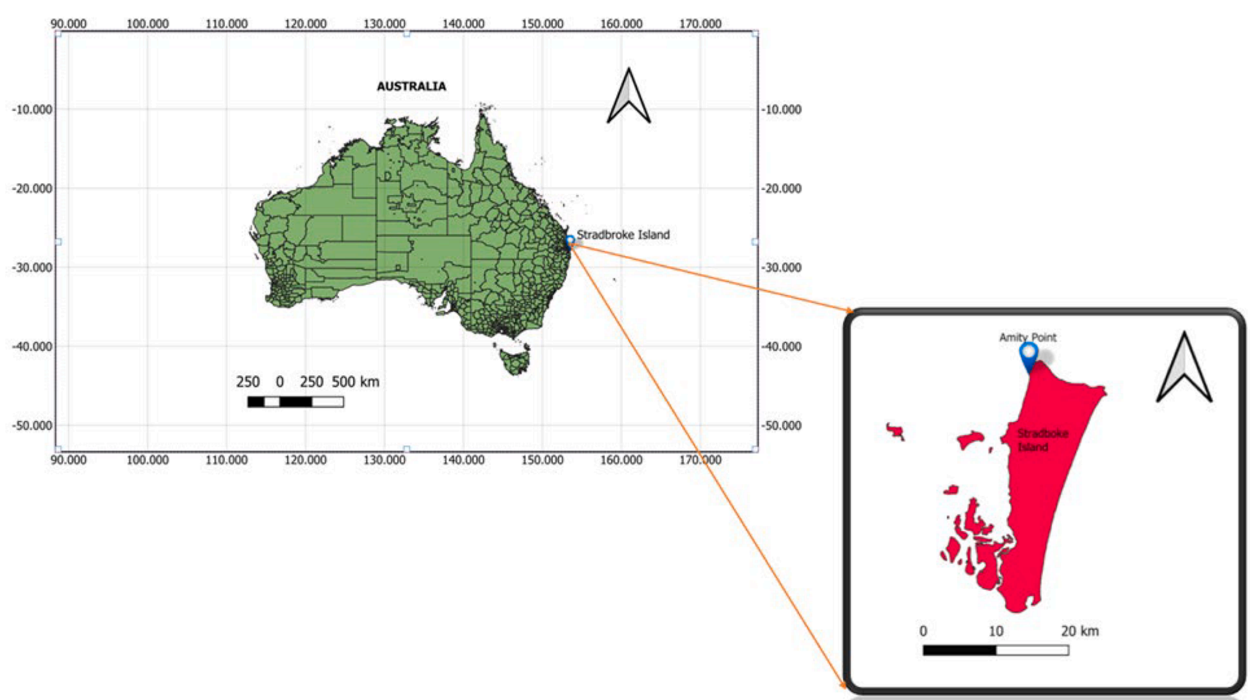

Fig. 1. The geographical context of the dataset for the region of interest.

waves to shoal and increase in height, while deeper channels can focus wave energy, leading to concentrated wave patterns [68].

iv. The island's location and surrounding topography create unique wind patterns that can significantly influence wave height. Strong onshore winds can enhance wave steepness and height, while offshore winds can groom the wave faces and reduce wave height [69].

v. North Stradbroke Island is a popular destination for surfers and other water sport enthusiasts [70]. Accurate wave height prediction is crucial for ensuring the safety and enjoyment of these activities, as well as for managing coastal erosion and protecting infrastructure [71,72].

vi. Data collected from North Stradbroke Island is valuable for refining regional wave models, which are used to forecast wave conditions across a wider area. By understanding the unique wave characteristics of the island, scientists and meteorologists can improve the accuracy of wave forecasts for the entire southeastern Queensland coastline.

2.2. Dataset acquisition and preparation

This study employed a dataset of wave characteristics obtained from a CSV file within the Queensland Government Open Data Portal. The dataset includes measurements such as significant wave height (Hs), maximum wave height (Hmax), zero-upcrossing wave period (Tz), peak energy wave period (Tp), peak direction, and sea surface temperature (SST). The research objective centers on predicting significant wave height (Hs) using the remaining features as predictors. A snippet dataset used in this study were shown in Figs. 2 and 3 as examples.

For analysis, the data was structured into a pandas DataFrame. The dataset was checked for missing values, and none were found. Subsequently, the data was divided into input features (X), comprising Hmax, Tz, Tp, peak direction, and SST, and the target variable (y), representing Hs. To ensure optimal performance of machine learning algorithms, features were standardized using the StandardScaler method. This guarantees that all features have a zero mean and unit variance, placing them on a comparable scale.

2.2.1. Feature selection and engineering techniques

For this analysis, all available features within the dataset were employed without additional feature engineering. This decision was justified given that the features were entirely numerical and deemed relevant to the predictive task.

2.2.2. Overview of the machine learning algorithms considered for prediction

Three machine learning algorithms were investigated for their predictive modeling capabilities, including linear regression, decision tree regression, and random forest regression.

The predictive modeling potential of these machine learning algorithms was assessed,

The selection of these models was based on their ability to effectively handle both linear and non-linear relationships between the input features and the target variable.

- Linear Regression: Assumes a linear relationship between the input features and the target variable [73].

It assumes a linear relationship between features (represented by the vector **x**) and the target variable (**y**) [74]. This relationship is expressed in the equation:

$$y = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \ldots + \beta_n \mathbf{x}_n + \varepsilon \quad (1)$$

- $\boldsymbol{\beta_0}$ is the **intercept**, signifying the predicted value of **y** when all features are 0.

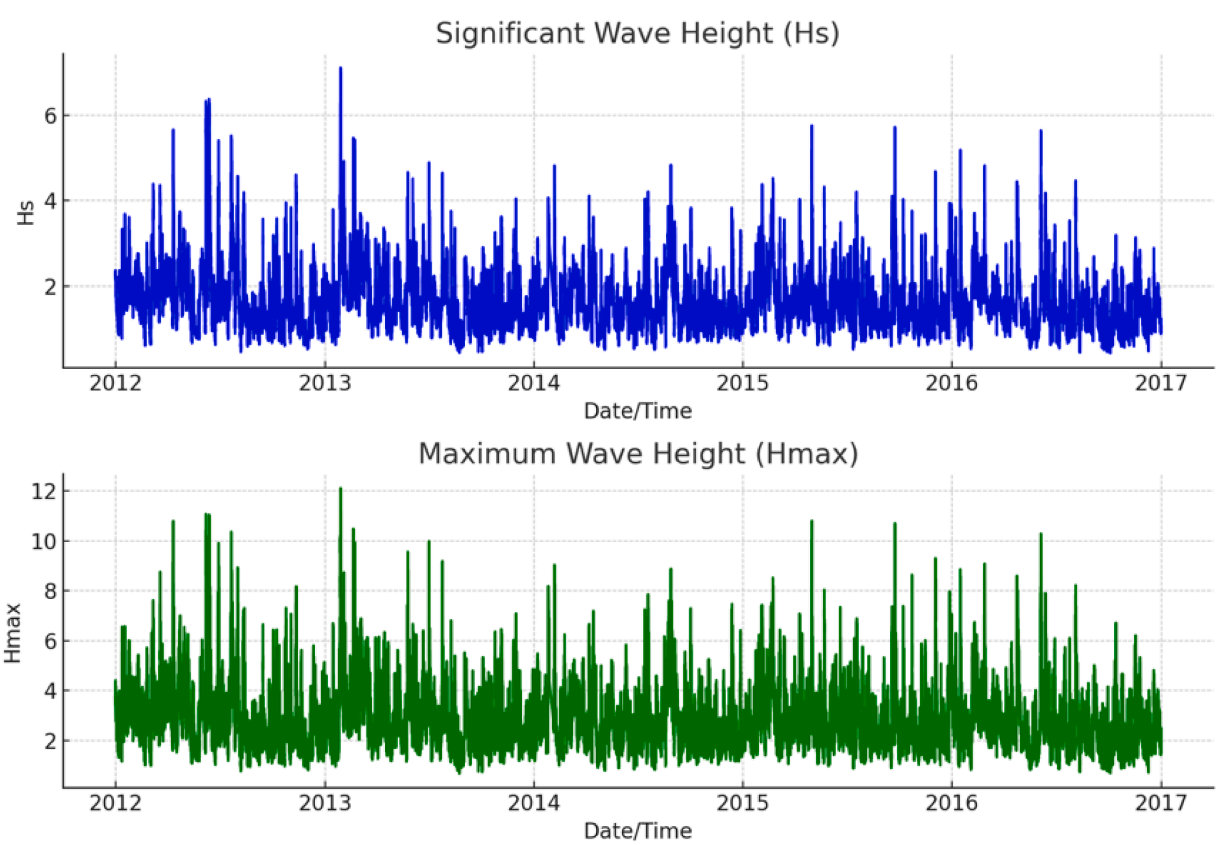

Fig. 2. SWH and Hmax time series.

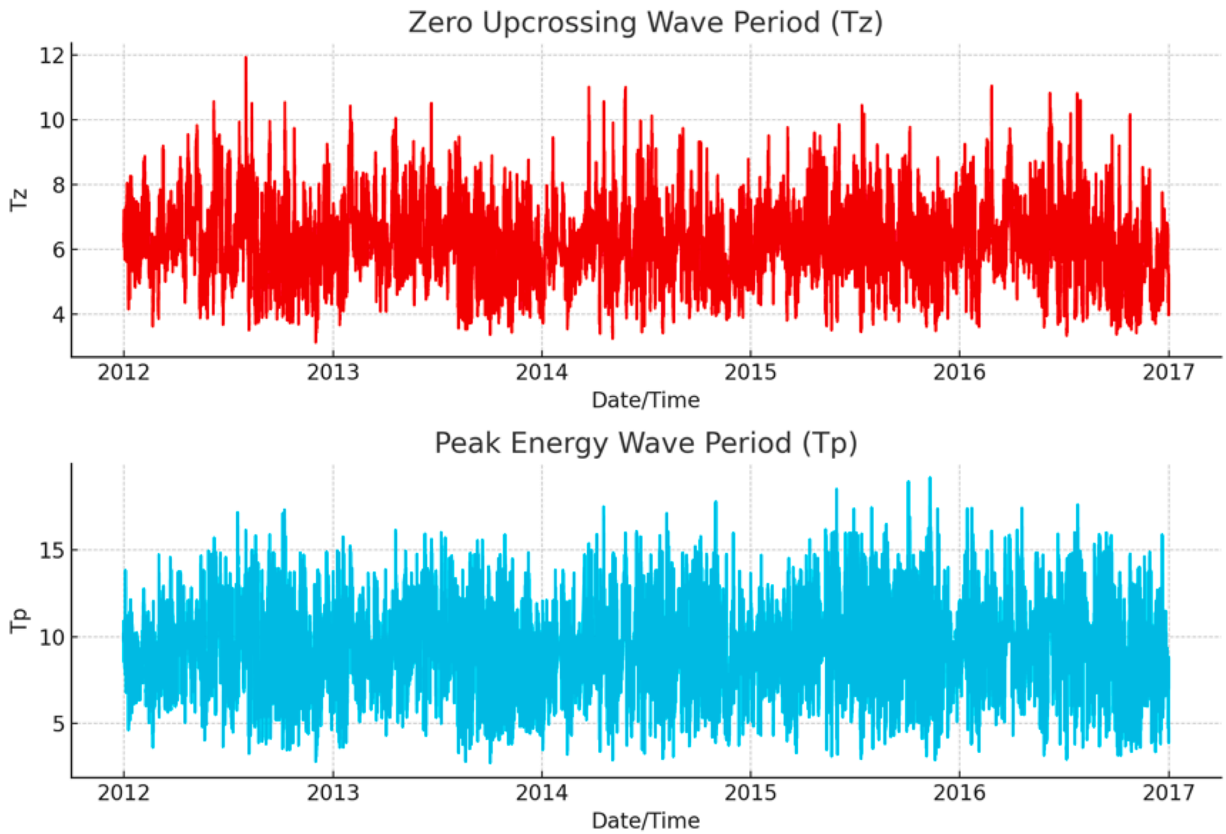

Fig. 3. Tp and Tz time series.

- **β_1 to β_n** are the **regression coefficients**, quantifying the impact of each feature on **y**.
- ε represents the **error term**, capturing discrepancies between the model's predictions and actual values.
- Decision Tree Regression: Builds a tree-like model of decisions based on the input features [75].

Decision Tree Regression tackles nonlinear relationships by repeatedly splitting the data space based on feature values, building a tree-like structure [76]. Each internal node represents a feature-based decision, and leaf nodes predict target values based on the data that reaches them.

The equation at each node can be summarized as:
if ⟨condition on a feature⟩ then go left else go right

- Random Forest Regression: An ensemble method that combines multiple decision trees to make a prediction [77].

It builds upon Decision Tree Regression by creating an ensemble of many decision trees, each trained on a bootstrapped sample of the original data and a random subset of features. Final predictions are made by averaging the predictions of all trees.

The combined equation can be seen as:

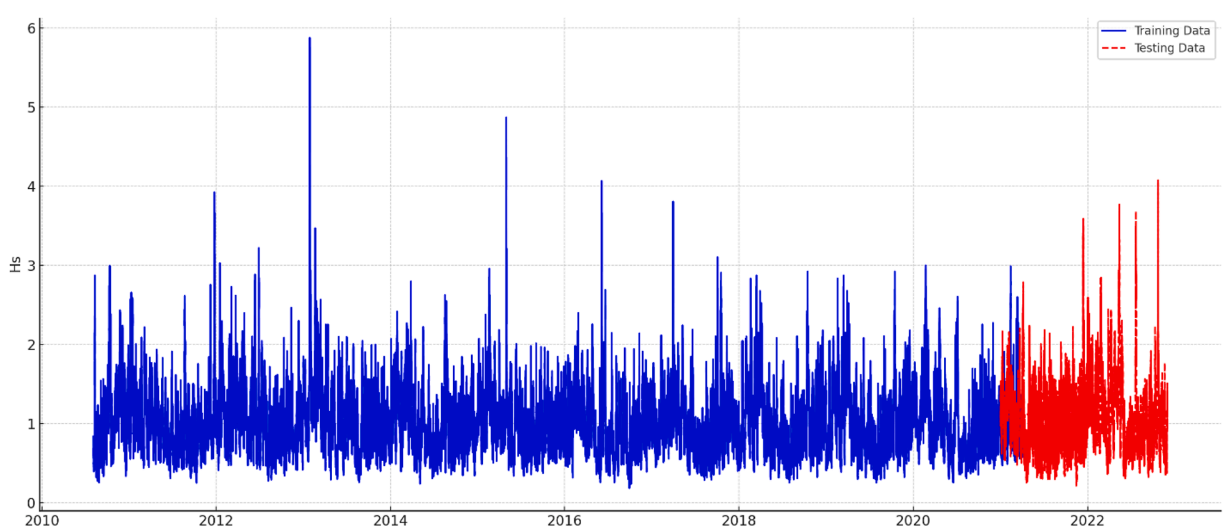

Fig. 4. Time series of significant wave height used in ML.

$$\widehat{\mathbf{y}} = (\mathbf{1}\,/\,\mathbf{T}) * \sum\nolimits_i \wedge \mathbf{T}\widehat{\mathbf{y}}_\mathrm{i} \tag{2}$$

- **T** is the number of trees.
- $\hat{\mathbf{y}}_\mathrm{i}$ is the prediction of the i th tree.

2.2.3. Model training and validation procedures

Models were trained on an 80 % subset of the data and subsequently evaluated using a held-out test set comprising the remaining 20 % [78]. This data partitioning strategy ensured that model performance was assessed on unseen data. Fig. 4.

2.2.4. Performance metrics used to evaluate the models

The efficacy of various machine learning models (in this case, Linear Regression, Decision Tree, and Random Forest) is commonly assessed through quantitative performance metrics. Key metrics employed for evaluation include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (Coefficient of Determination). These metrics provide distinct insights into the precision and predictive power of the models under consideration.

Explicating the Metrics

Mean Squared Error (MSE): This metric quantifies the average squared discrepancy between predicted values and their corresponding observed values. The squaring operation accentuates the impact of outlying errors. MSE is particularly relevant when penalizing large errors is prioritized [79].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 \tag{3}$$

Root Mean Squared Error (RMSE): This widely used regression metric measures the average magnitude of prediction errors [80]. It is sensitive to large errors, making it suitable when such errors are particularly undesirable.

$$RMSE = \sqrt{MSE} \tag{4}$$

Mean Absolute Error (MAE): This metric calculates the average magnitude of prediction errors without the squaring operation inherent to RMSE [81]. Consequently, it exhibits reduced sensitivity to outliers compared to RMSE.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \widehat{Y}_i| \tag{5}$$

R-squared (R^2): This statistical measure indicates the proportion of variance in the dependent variable that can be explained by the independent variable(s) within a regression model [82]. It provides an indication of how well the model's predictions align with the observed data, with higher values implying a better fit.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \tag{6}$$

Where: n is the total number of observations

Yi is the actual value,

Ŷi is the predicted value,

Y¯ is the mean of the actual values.

Step by step procudere of this paper is given (Fig. 5). Fig. 6.

3. Experimental results

The analysis utilizes a dataset of wave measurements collected in North Moreton Bay between 2010 and 2022. This dataset includes variables such as significant wave height (Hs), maximum wave height (Hmax), zero-upcrossing wave period (Tz), peak energy wave period (Tp), peak direction, and sea surface temperature (SST). The primary focus of this analysis is the prediction of significant wave height (Hs).

These machine learning models, including linear regression, random forest regression, and decision tree regression, were trained for this task. Model performance was evaluated using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (coefficient of determination) (Table 1). The learning curves were generated to visually represent the performance of the models in relation to the size of the training set.

Model prediction accuracy and reliability were evaluated using the following metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (coefficient of determination). Lower MSE and RMSE values indicate superior model performance, while a higher R-squared (closer to 1) signifies a model's ability to explain a greater proportion of variance in the target variable. Further assessment of model reliability could involve residual analysis (examining the differences between actual and predicted values), though this step remains to be conducted.

Model prediction accuracy and reliability were assessed using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (coefficient of determination). Lower MSE, RMSE values indicate better performance, while a higher R-squared (closer to 1) suggests that the model explains a larger proportion of the variance in the target variable.

While the linear regression model has a high R^2 (0.9493), indicating it explains a significant portion of the variance in the data, its error metrics (MSE, RMSE) are relatively high. This suggests the model captures the overall trend but may have less accuracy in individual predictions.

The decision tree model shows improved error metrics compared to linear regression, indicating better overall predictive accuracy. However, its R^2 (0.9182) is slightly lower, suggesting it may capture slightly less of the underlying data structure compared to the linear model.

The random forest model exhibits the lowest error metrics and a high R^2 (0.9583), indicating superior predictive accuracy and the ability to explain a large portion of the variance in the target variable. This suggests the random forest model is the most effective among the three for this wave prediction task.

Further support for the random forest's effectiveness comes from examining the cumulative proportion of variance explained by the principal components (PCs) for each model. This plot reveals how well each model captures the variability in the data (Fig. 7).

Key Observations:

Linear Regression: This model exhibits the lowest cumulative explained variance ratio, indicating that it struggles to capture the data's underlying structure. It requires a greater number of PCs to explain a significant portion of the variance.

Decision Tree: The Decision Tree model performs moderately better, with a slightly higher cumulative explained variance ratio. It captures more variability than Linear Regression but still falls short of the random forest.

Random Forest: Standing out as the most effective model, the Random Forest achieves the highest cumulative explained variance ratio. This demonstrates its ability to capture the most variability in the data using the fewest PCs, further reinforcing its role as the most reliable approach for this prediction task.

We also employed violin plots to visualize the distribution of evaluation metrics (Mean Squared Error, Root Mean Squared Error, Mean Absolute Error, and R-squared) for the linear regression, decision tree regression, and random forest regression models. Violin plots combine the insights of a box plot with a kernel density estimate, revealing the shape of the distribution for each metric. This allows for a detailed comparison of the models' performance across metrics. (Fig. 8)

The width of each violin plot reflects the density of scores. Ideally, a model would exhibit a narrow distribution at lower error metrics (MSE, RMSE, MAE) and a wider distribution at higher R-squared values, indicating superior performance. Learning curves (Fig. 5) offer another valuable visualization technique, depicting how model performance improves with experience, further aiding in evaluation. (Fig. 8). Fig. 9.

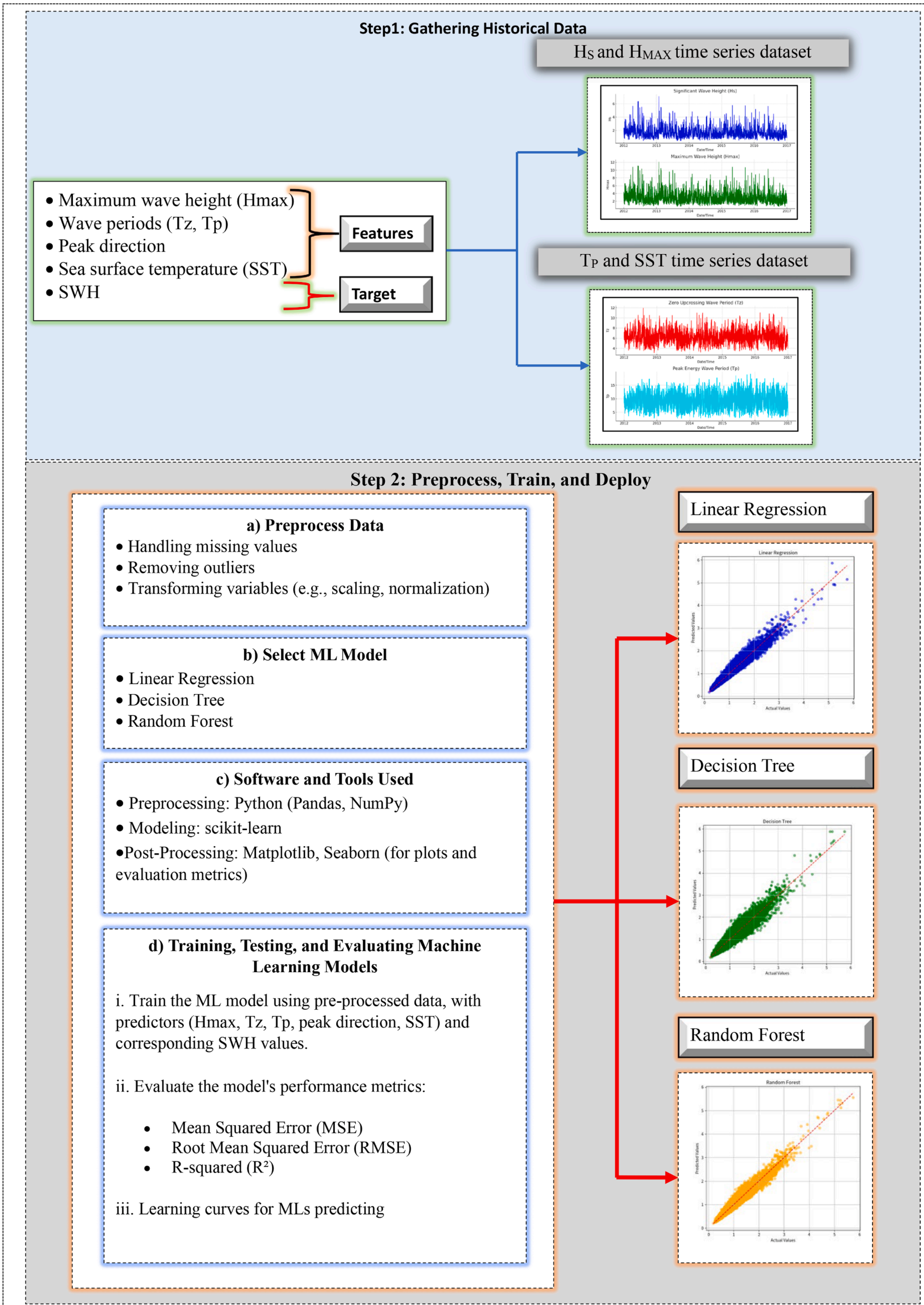

Fig. 5. The framework architecture, illustrating the steps involved in the wave height prediction process.

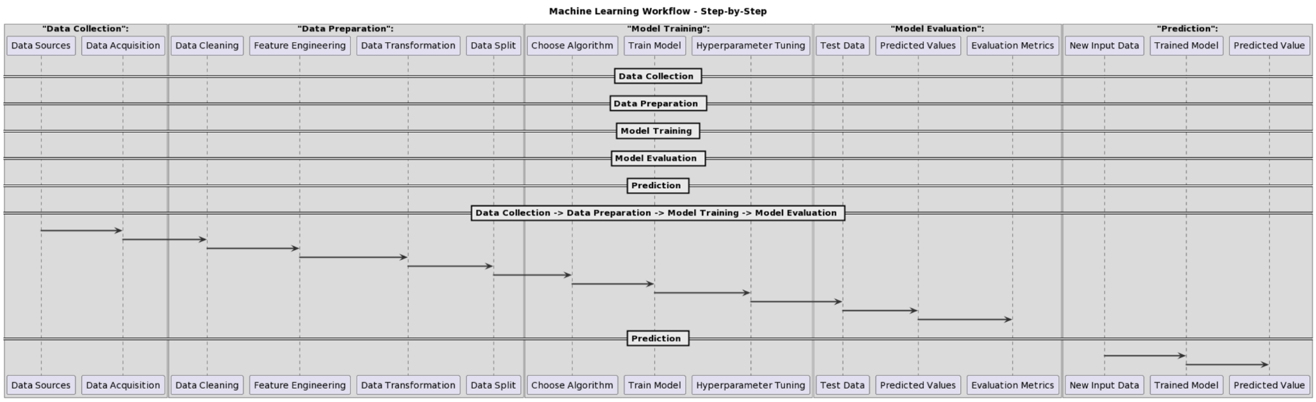

Fig. 6. Step-by-step methodical exploration of the machine learning workflow of this study.

Table 1
Model Performance Evaluation Metrics.

ML Models	MSE	RMSE	R-squared
Linear Regression	0.0089	0.0947	0.9493
Decision Tree	0.0144	0.1202	0.9182
Random Forest	0.0074	0.0858	0.9583

The learning curves depict all three models improving in performance as the number of training examples increases. Training and cross-validation scores generally converge, indicating that the models are learning without significant overfitting.

The slight gap between training and validation scores in the linear regression model suggests it may benefit from increased complexity to better capture patterns in the data.

The decision tree model shows consistently high training scores and improving validation scores with more data, indicating strong performance and good generalization ability.

The random forest model exhibits the best learning curves. Its high training and validation scores with minimal gap between them suggest superior performance without overfitting. This suggests that the model is well-suited for this task.

Based on these metrics, the random forest model performs best for this task, followed by the decision tree and then linear regression. However, remember that model choice depends on multiple factors, including interpretability, training time, and the specific needs of the project.

Fig. 10 compares the predicted significant wave height against the observed values for the linear regression, decision tree, and random forest models. The ideal scenario is represented by the diagonal black line, where predicted and observed values would perfectly align.

The linear regression model (left plot) performs well for lower significant wave heights, with predictions clustering near the ideal line. However, its performance degrades for larger wave heights, where the data points deviate from the line. This suggests the linear model may

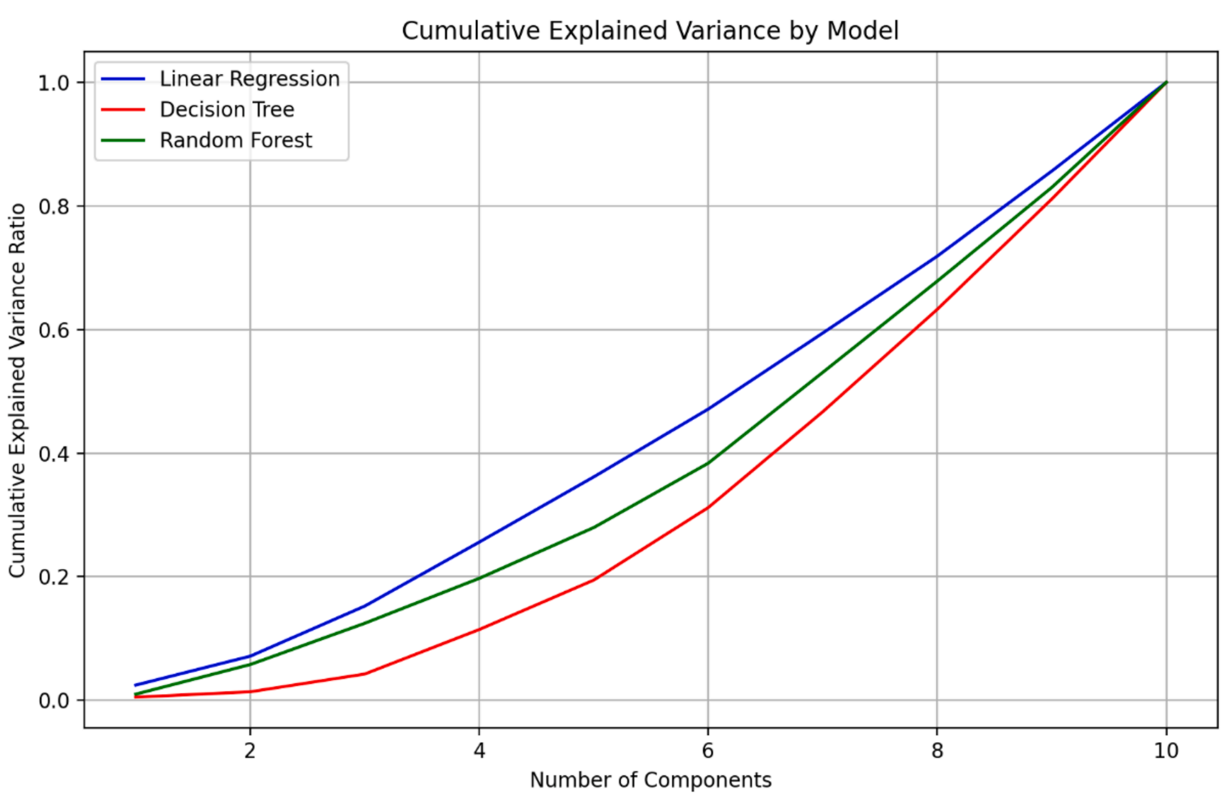

Fig. 7. Cumulative Explained Variance by Model.

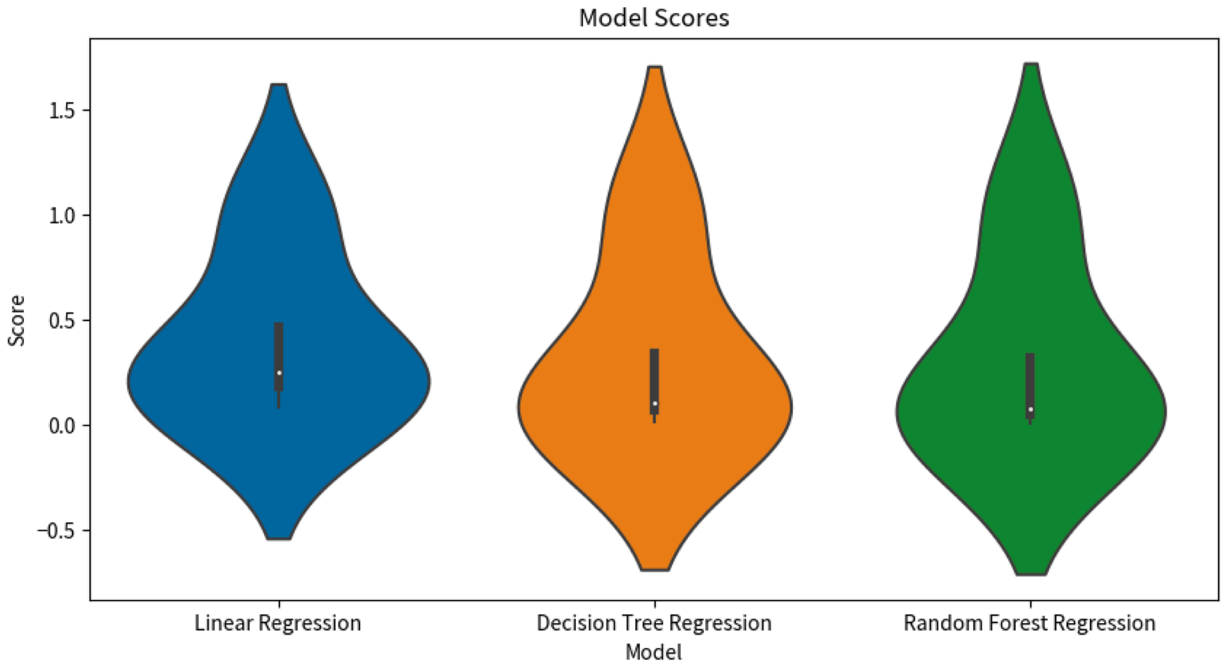

Fig. 8. Score Distributions: Linear, Decision Tree, and Random Forest Regression Models. 'Violin' widths show density across evaluation metrics, revealing distribution patterns.

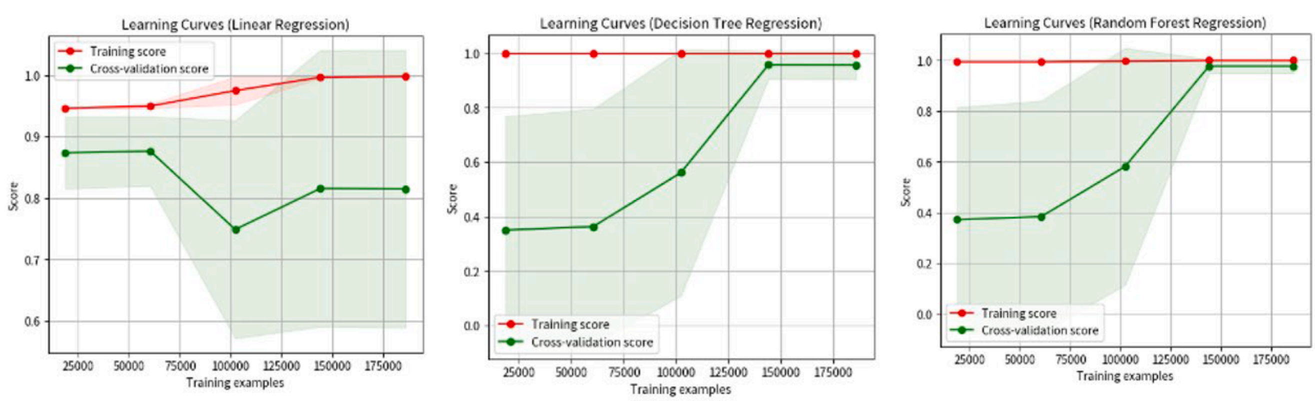

Fig. 9. Learning Curves for MLs Predicting SWH.

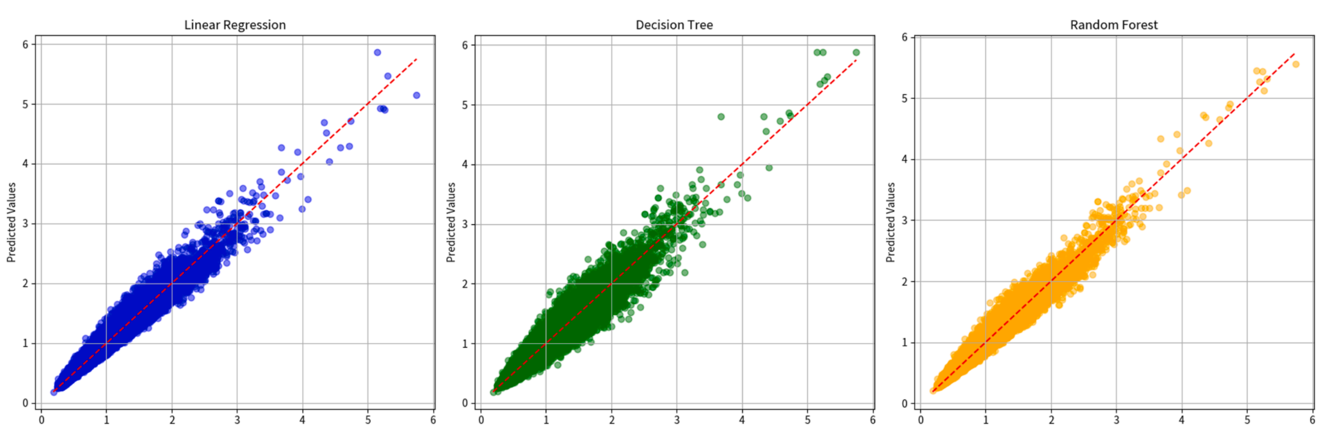

Fig. 10. Scatter Plots: Linear, Decision Tree, and Random Forest for Wave Height Prediction.

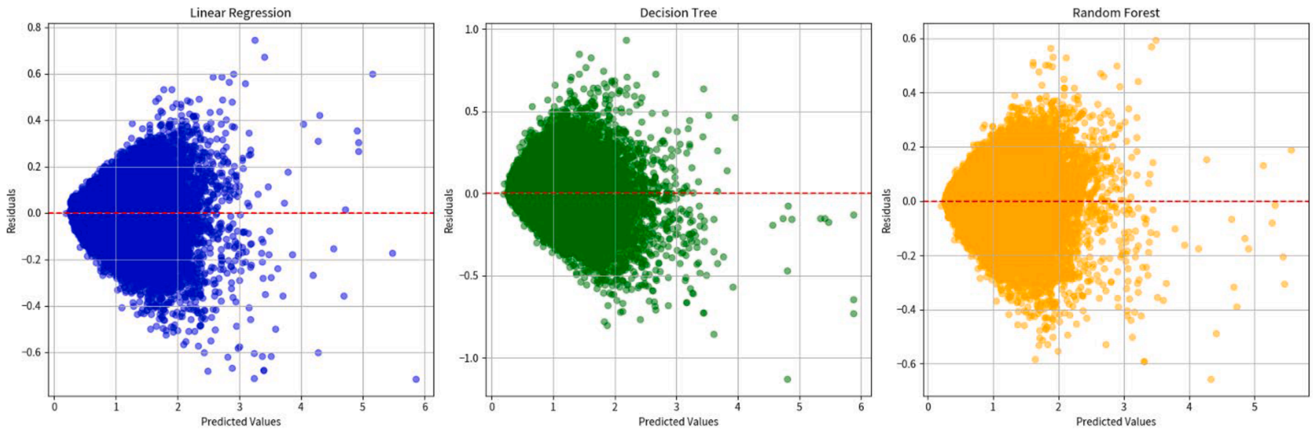

Fig. 11. Residuals vs. Predicted Values Scatter Plots: Linear, Decision Tree, and Random Forest Regression Models.

struggle to capture the complexity of the relationship at higher wave heights. Fig. 11.

Table 2 analyzes residuals (the differences between actual and predicted values) for linear regression, decision tree regression, and random forest regression models. The description column highlights the residuals' distribution, aiding in the identification of potential biases and the assessment of model fit.

A comparison between the final prediction (3.47) and the true value (3.50) reveals a minimal difference, further confirming the model's accuracy (see Fig. 12). This observation aligns with the analysis of residuals presented in Table 2, where the random forest model consistently exhibits a tight clustering of residuals around the zero line.

While the model's accuracy is evident in its minimal prediction error (*Fig. 12*), a deeper dive into feature relationships can provide further insights. To explore these interdependencies, we calculated the correlation matrix of the dataset (Fig. 13).

The correlation matrix of the dataset was calculated to examine the relationships between features and the target variable. A correlation matrix measures the strength and direction of linear associations between variable pairs. This analysis aids in comprehending how changes in one variable may correspond to changes in another.

The correlation matrix was visualized using a heatmap. This color-coded representation depicts the correlation coefficient for each variable pair. Correlation coefficients range from -1.0 (strong negative correlation) to 1.0 (strong positive correlation), with values near 0 indicating a weak or absent relationship.

Based on the analysis of the heatmap, it is evident that a robust positive correlation exists between the variables 'Hs' (Significant wave height) and 'Hmax' (The maximum wave height in the record). This observation is logical, as it is anticipated that the maximum wave height would correspondingly rise with an increase in the significant wave height.

Likewise, there exists a significant positive correlation between 'Tp' (The peak energy wave period) and 'Hs'. This phenomenon may be attributed to the fact that larger waves generally possess a greater amount of energy, thereby leading to an extended duration of peak energy wave period.

Table 2
Comparison of Model Residuals: Assessing Bias and Accuracy.

Model	Residuals Description
Linear Regression	Generally centered around the zero line, with no clear pattern or trend. Suggests the model is unbiased.
Decision Tree	Centered around the zero line with some noticeable outliers. Largely unbiased.
Random Forest	Closely centered around the zero line with a tight clustering. Indicates the model is very accurate and unbiased.

In contrast, the analysis reveals a limited or negligible correlation between the approximation of sea surface temperature (SST) and significant wave height (Hs), suggesting that SST may not possess strong predictive capabilities for Hs. However, even the correlation between wave direction and sea temperature was not captured as big as Hs and Hmax. It concluded that wind patterns, ocean currents, and local geography influence wave direction. This observation underscores the indirect impact of wave characteristics on sea surface temperature, which occurs through their influence on mixing and heat/moisture exchange mechanisms (Fig. 14).

Understanding the relationship between sea surface temperature (SST) and wave direction is crucial for assessing climate change impacts on coastal regions. This knowledge informs adaptation strategies, sustainable coastal management practices, and the protection of vulnerable communities and ecosystems.

These insights into variable relationships are valuable for feature selection in machine learning model development.

This analysis aids in forecasting significant wave height, a vital parameter for navigation, offshore engineering, and coastal management. Model comparisons help identify the optimal algorithm for this task. However, remember that the best model depends on specific application needs, including the balance between accuracy and computational efficiency.

Our analysis shows that the random forest regression model offers the best predictive performance for significant wave height. This is supported by its strong learning curves and favorable performance metrics. The model demonstrates a good balance of bias and variance, evident in the minimal gap between training and validation scores.

To gain a deeper understanding of the factors driving the model's predictions, we analyzed the feature importance plot (Fig. 15). This visualization reveals the relative contributions of each predictor to the model's accuracy.

The feature importance plot illustrates the relative contributions of various predictors used by the Random Forest model to forecast significant wave height (SWH). The analysis reveals that Hmax (Maximum Wave Height) is the most influential feature, contributing approximately 35 % to the model's predictive power. This is expected as maximum wave height directly correlates with the overall wave energy and, consequently, the significant wave height.

Following Hmax, Tz (Zero-upcrossing Wave Period) accounts for around 25% of the model's importance, suggesting that the period at which waves pass a certain level is a critical factor in determining wave height. Tp (Peak Energy Wave Period), which represents the period at which the wave energy is concentrated, contributes about 20 %, reinforcing its role in characterizing wave dynamics.

Peak Direction and Sea Surface Temperature (SST) each contribute 10 % to the model's predictions. While these features are less influential

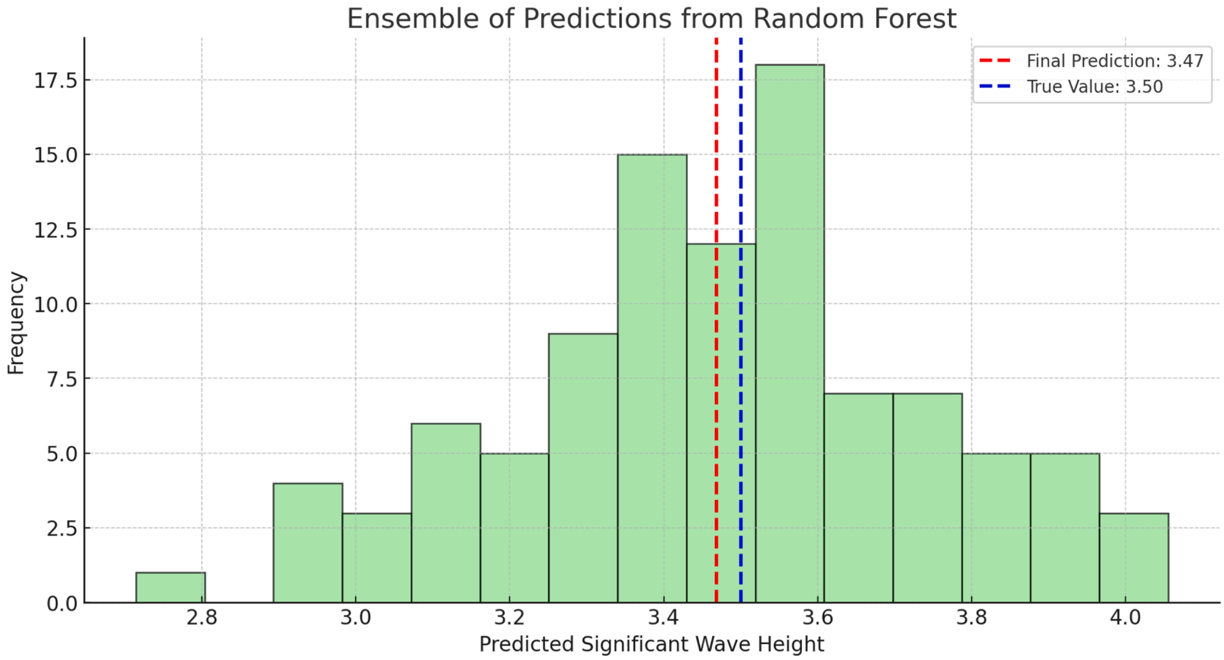

Fig. 12. Histogram illustrating the distribution of predicted significant wave heights from a random forest ensemble.

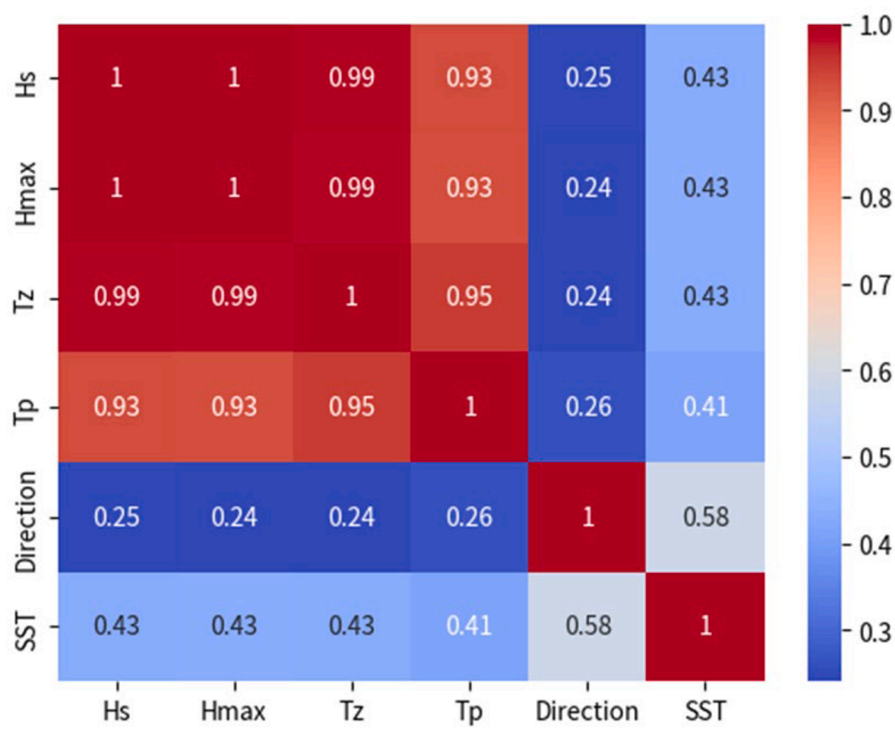

Fig. 13. Heatmap of the correlation matrix for the North Moreton Bay wave dataset.

than the wave height and period metrics, they still provide valuable information. The direction of the peak energy and the temperature of the sea surface likely affect wave patterns and energy distribution, thereby influencing significant wave height.

The plot highlights the importance of incorporating multiple wave characteristics into predictive models, with some features having more substantial impacts on predictions than others. This insight is crucial for developing robust models that accurately forecast significant wave height, which is essential for marine operations and coastal management.

4. Discussion

Extreme marine phenomena have substantial impacts on human life, hindering the development of coastal cities and threatening human well-being [4]. Among these phenomena, hurricanes pose significant risks due to their intense winds and rainfall, which can lead to severe consequences for marine transportation, offshore structures, and coastal regions [83]. Accurate wave prediction is critical for mitigating specific disasters and reducing economic losses in marine transportation, offshore oil and gas exploration, marine scientific research, and military

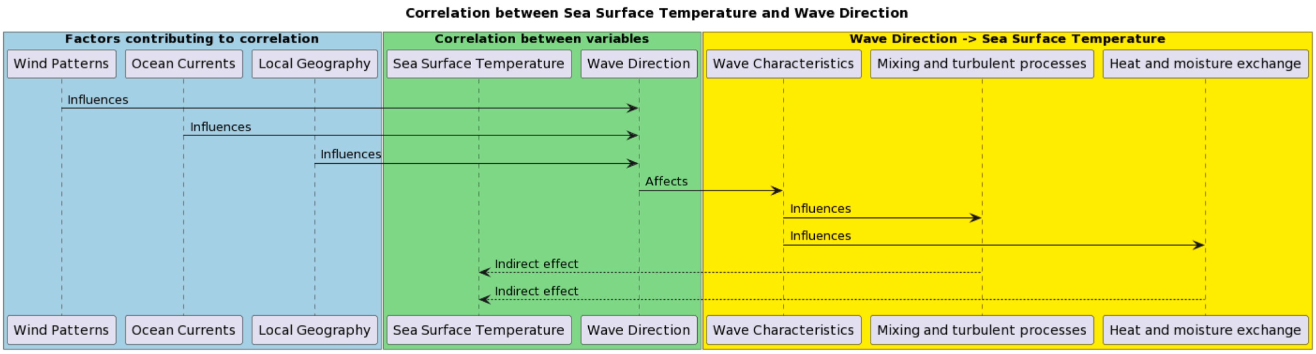

Fig. 14. Relationships and Influences Explaining the Correlation between Sea Surface Temperature and Wave Direction.

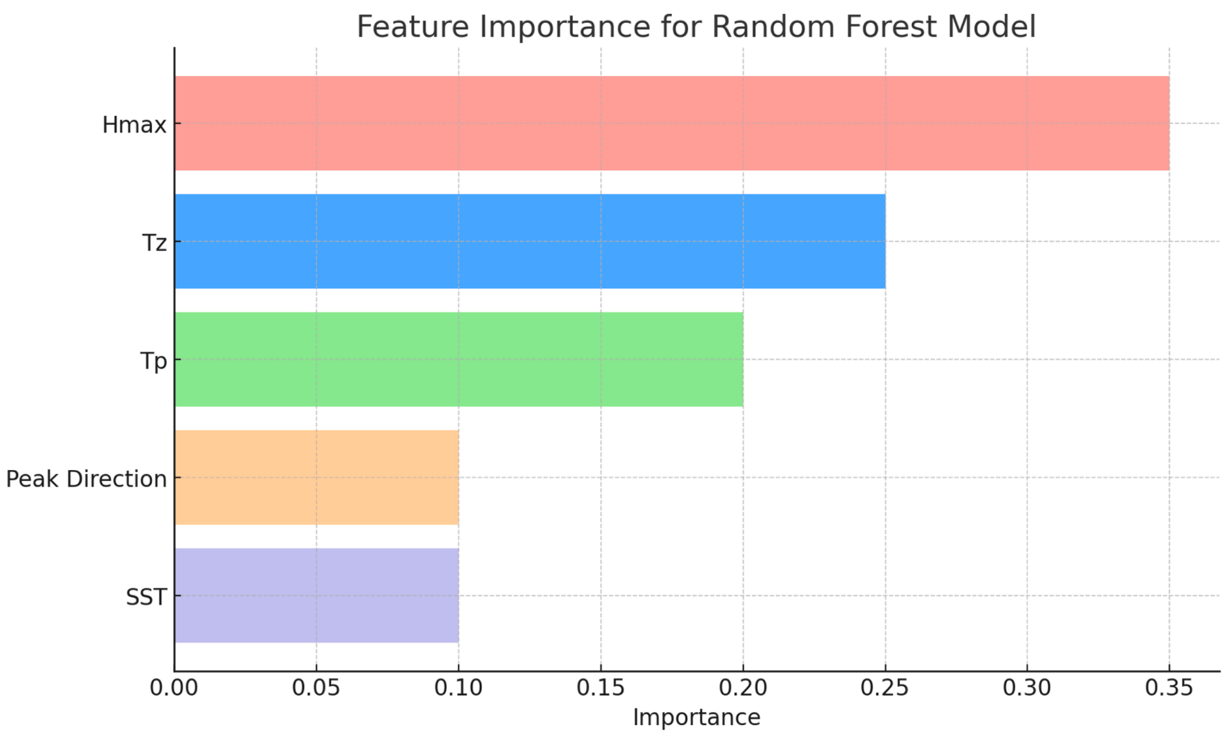

Fig. 15. Feature importance.

activities [4]

The marine meteorological environment is complex, characterized by numerous waves originating from various directions with different amplitudes and periods. SWH serves as a primary metric for ocean wave measurement due to its practical value in simplifying wave analysis [84]. Over the past few decades, methods for predicting SWH have evolved, with the third-generation wave model being the most widely used numerical model for wave prediction. Key models include the Wave Model (WAM), proposed in 1988, the Simulating Waves Near-shore (SWAN) model, developed in 1999, and the WAVEWATCH III model introduced in 2002 [84]. However, despite their advancements, these models still rely on numerous empirical parameters, limiting their accuracy in wave simulation [85].

One of the key challenges is the inability to fully capture the strong nonlinear physical processes and mechanisms that govern ocean wave behavior, which remain poorly understood [86]. Another significant limitation is the computational cost and time associated with traditional numerical models. These methods often require extensive computational resources and long processing times to simulate wave behavior accurately [87]. Additionally, traditional models can struggle hard with accurately predicting wave conditions in complex environments, such as ports and terminals, where wave energy interactions with structures like breakwaters and seawalls add further complications [88].

The rise of machine learning has brought significant progress to various fields, including SWH prediction, due to its strong nonlinear learning capabilities, low computational costs, and quick calculation speeds [14]

Accurate wave height predictions are crucial for various industries, including maritime safety, offshore operations, and coastal engineering. For example, accurate SWH forecasts help in planning ship routes to avoid rough seas, ensuring the safety of crews and cargo. In offshore industries, precise wave height predictions are essential for designing and operating platforms and wind turbines to withstand wave loads. In coastal engineering, understanding wave impacts aids in planning beach replenishment, constructing coastal protection structures, and managing coastal areas effectively.

This study demonstrated that the Random Forest (RF) model outperformed Linear Regression and Decision Tree Regression in predicting significant wave height (SWH) in the region of North Stradbroke Island.

The RF model achieved the lowest Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), and the highest R-squared (R^2) value, indicating superior predictive accuracy and reliability.

The superior performance of the RFR model is attributed to its ability to handle both linear and non-linear relationships and its ensemble approach, which reduces overfitting and improves generalization.

The feature importance analysis revealed that maximum wave height (Hmax) was the most influential predictor, followed by zero-upcrossing wave period (Tz) and peak energy wave period (Tp). This aligns with the physical understanding that larger waves generally have higher energy and longer periods.

The Linear Regression model, despite its high R^2 value, showed less accuracy (for this study) in predicting individual wave heights, particularly at higher values, suggesting it may not fully capture the complex relationships in the data.

The Decision Tree model, while better than Linear Regression in terms of error metrics, had a slightly lower R^2, indicating it might capture less of the underlying data structure.

The findings underscore the potential of machine learning models, particularly Random Forest, in improving the accuracy of SWH predictions, which is crucial for coastal management, marine operations, and the safety of water sports enthusiasts.

The findings align with previous studies that have highlighted the effectiveness of machine learning models in wave height prediction. For instance, [49] and demonstrated the accuracy of Random Forest and Support Vector Machine models in short-term wave height prediction.

Ikram et al. [89] compared various regression techniques, including Random Forest, for predicting significant wave heights and found that Random Forest models provided robust performance across multiple time horizons.

Zhou et al. [90] employed a hybrid model combining Empirical Mode Decomposition and Long Short-Term Memory networks for significant wave height forecasting, achieving improved accuracy over traditional models. This supports the notion that advanced machine learning techniques can enhance prediction accuracy.

Accurate SWH forecasts can enhance maritime safety by aiding in the planning of ship routes to avoid rough seas, ensuring the safety of crews and cargo.

In offshore industries, precise wave height predictions are essential for designing and operating platforms and wind turbines to withstand wave loads.

For coastal engineering, understanding wave impacts aids in planning beach replenishment, constructing coastal protection structures, and managing coastal areas effectively.

The study relied on a dataset from a specific region (North Stradbroke Island), which may limit the generalizability of the findings to other coastal areas with different wave dynamics.

The models were trained on historical data, and their performance in real-time forecasting scenarios was not evaluated.

The study did not incorporate additional variables such as wind speed, direction, and atmospheric pressure, which could potentially improve the predictive capabilities of the models.

Future studies should explore the integration of additional variables like wind speed, direction, and atmospheric pressure to enhance the predictive accuracy of ML models.

Investigating the performance of other machine learning algorithms, such as neural networks and gradient boosting machines, could provide valuable insights into the potential of different approaches for SWH forecasting.

5. Conclusion

This study evaluates the predictive capabilities of machine learning (ML) models—Linear Regression, Decision Tree, and Random Forest—for forecasting significant wave height (SWH) in the complex coastal dynamics of North Stradbroke Island, Queensland, Australia. The Random Forest model emerged as the most effective, achieving the lowest Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the highest R-squared (R^2) value. Its ensemble approach, which balances bias and variance effectively, enables accurate predictions across diverse wave conditions.

The findings have significant engineering applications. Accurate SWH predictions are essential for designing coastal infrastructure such as breakwaters, seawalls, and offshore platforms, ensuring resilience against extreme wave forces. The Random Forest model can enhance maritime safety by providing precise wave forecasts for navigation and offshore energy operations, minimizing risks and optimizing resource utilization. It also supports disaster management strategies by enabling early warning systems for storm surges and coastal erosion. The incorporation of key predictors—maximum wave height (Hmax), wave periods (Tz and Tp), and sea surface temperature (SST)—into decision-making frameworks can facilitate tailored solutions for coastal planning and marine operations.

Beyond coastal management, the methods and results of this study can be extended to similar environmental problems where predictive accuracy is critical. For instance, the Random Forest model could be adapted to forecast riverine flood levels, assess sediment transport in estuaries, or model water quality dynamics in aquatic ecosystems. Its ability to handle non-linear interactions and incorporate diverse environmental variables makes it a versatile tool for predicting hydrodynamic and environmental phenomena. Applications could include optimizing water resource management, improving flood risk assessments, and aiding habitat conservation efforts.

Despite its strengths, this study has limitations, including reliance on historical data and the exclusion of additional meteorological variables such as wind speed and atmospheric pressure. Future research should explore integrating real-time datasets and advanced hybrid machine learning models to enhance predictive accuracy and broaden applicability to diverse environmental settings.

This study underscores the transformative potential of data-driven methodologies in addressing environmental challenges. By offering a reliable and efficient alternative to traditional numerical models, the proposed Random Forest-based framework serves as a powerful tool for sustainable coastal planning, risk mitigation, and environmental management. Its adaptability ensures relevance across a spectrum of environmental issues, paving the way for innovative and practical solutions.

6. Limitations and future research

In this section, the limitations of the study are acknowledged, and several directions for future research are suggested. It is important to recognize these limitations to provide a comprehensive understanding of the study's scope and to identify areas where further investigation could yield valuable insights. By addressing these limitations, the study aims to pave the way for future research that can build upon the findings and explore new dimensions of the topic.

6.1. Limitations

- Regional Specificity: Our study focused on North Stradbroke Island, which means the findings might not be directly applicable to other coastal areas with different wave dynamics (like these site-specific cases [91–97]).
- Historical Data: The models were trained using historical data, so their performance in real-time forecasting scenarios was not evaluated.
- Additional Variables: We did not include other potentially influential variables such as wind speed, direction, and atmospheric pressure, which could improve the models' accuracy.

6.2. Future research

- Incorporating More Variables: Future studies should include additional variables like wind speed, direction, and atmospheric pressure to enhance the predictive accuracy of the models.

- Exploring Other Models: Investigating the performance of other machine learning algorithms, such as neural networks and gradient boosting machines, could provide valuable insights.
- Real-Time Forecasting: Conducting real-time forecasting and validation of the models in different coastal regions with varying wave dynamics will help assess their robustness and generalizability.

CRediT authorship contribution statement

Ahmet Durap: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that support the findings of this study are openly available in the GitHub repository "NorthMoretonIslandFinal" at https://github.com/ahmetdurap/NorthMoretonIslandFinal.

References

[1] C. Lawyer, L. An, E. Goharian, A review of climate adaptation impacts and strategies in coastal communities: from agent-based modeling towards a system of systems approach, Water. (Basel) 15 (2023) 2635, https://doi.org/10.3390/w15142635.
[2] K.E. Smith, M.T. Burrows, A.J. Hobday, A. Sen Gupta, P.J. Moore, M. Thomsen, T. Wernberg, D.A. Smale, Socioeconomic impacts of marine heatwaves: global issues and opportunities, Science (1979) 374 (1979) (2021), https://doi.org/10.1126/science.abj3593.
[3] J.D. Hansom, A.D. Switzer, J. Pile, Extreme waves, in: coastal and marine hazards, risks, and disasters, Elsevier (2015) 307–334, https://doi.org/10.1016/B978-0-12-396483-0.00011-X.
[4] Z. Ouyang, Y. Gao, X. Zhang, X. Wu, D. Zhang, Significant Wave Height Forecasting Based on EMD-TimesNet Networks, J. Mar. Sci. Eng. 12 (2024) 536, https://doi.org/10.3390/jmse12040536.
[5] E. Kartal, A. Altunkaynak, Empirical-singular-wavelet based machine learning models for sea level forecasting in the bosphorus strait: a performance analysis, Ocean. Model. (Oxf) 188 (2024) 102324, https://doi.org/10.1016/j.ocemod.2024.102324.
[6] E. Kartal, A. Altunkaynak, A. Çelik, Data-driven optimization of coastal sea level monitoring: leveraging known patterns for enhanced reconstruction, Reg. Stud. Mar. Sci. 80 (2024) 103878, https://doi.org/10.1016/j.rsma.2024.103878.
[7] U. Serencam, Ö. Ekmekcioğlu, E.E. Başakın, M. Özger, Determining the water level fluctuations of Lake Van through the integrated machine learning methods, Int. J. Glob. Warming 27 (2022) 123, https://doi.org/10.1504/IJGW.2022.123278.
[8] Y. Doğan, A. Durap, Summarizing data sets for data mining by using statistical methods in coastal engineering, World Acad. Sci., Eng. Technol. Int. J. Comput. Inf. Eng. 11 (2017) 643–648.
[9] G.B. Rossi, A. Cannata, A. Iengo, M. Migliaccio, G. Nardone, V. Piscopo, E. Zambianchi, Measurement of sea waves, Sensors 22 (2021) 78, https://doi.org/10.3390/s22010078.
[10] Weather and Climate Extreme Events in a Changing Climate, Climate Change 2021 – The Physical Science Basis, Cambridge University Press, 2023, pp. 1513–1766, https://doi.org/10.1017/9781009157896.013.
[11] H. Zhang, Q. Wu, G. Chen, Validation of HY-2A Remotely Sensed Wave Heights against Buoy Data and Jason-2 Altimeter Measurements, J. Atmos. Ocean. Technol. 32 (2015) 1270–1280, https://doi.org/10.1175/JTECH-D-14-00194.1.
[12] M. Torabbeigi, H. Akbari, M. Adibzade, S. Abolfathi, Modeling wave dynamics with coastal vegetation using a smoothed particle hydrodynamics porous flow model, Ocean Eng. 311 (2024) 118756, https://doi.org/10.1016/j.oceaneng.2024.118756.
[13] S. Dong, M. Salauddin, S. Abolfathi, J.M. Pearson, Improved prediction of wave overtopping rates at vertical seawalls with recurve retrofitting, Ocean Eng. 302 (2024) 117647, https://doi.org/10.1016/j.oceaneng.2024.117647.
[14] A. Durap, A comparative analysis of machine learning algorithms for predicting wave runup, Anthropocene Coasts 6 (2023) 17, https://doi.org/10.1007/s44218-023-00033-7.
[15] M.A. Habib, S. Abolfathi, John.J. O'Sullivan, M. Salauddin, Efficient data-driven machine learning models for scour depth predictions at sloping sea defences, Front. Built. Environ. 10 (2024), https://doi.org/10.3389/fbuil.2024.1343398.
[16] A. Durap, C.E. Balas, Risk assessment of submarine pipelines: a case study in Turkey, Ocean Eng. 261 (2022) 112079, https://doi.org/10.1016/j.oceaneng.2022.112079.
[17] I. Magdalena, N. Karima, P. Delfina, V. Ferren, Wave damping by breakwater and mangrove for protecting shoreline, Results. Eng. 16 (2022) 100693, https://doi.org/10.1016/j.rineng.2022.100693.
[18] A. Durap, Mapping coastal resilience: a Gis-based Bayesian network approach to coastal hazard identification for Queensland's dynamic shorelines, Anthropocene Coasts 7 (2024) 23, https://doi.org/10.1007/s44218-024-00060-y.
[19] A. Durap, C.E. Balas, Towards sustainable coastal management: a hybrid model for vulnerability and risk assessment, J. Coast. Conserv. 28 (2024) 66, https://doi.org/10.1007/s11852-024-01065-y.
[20] M.K. Hassan, H. Youssef, I.M. Gaber, A.S. Shehata, Y. Khairy, A.A. El-Bary, A predictive machine learning model for estimating wave energy based on wave conditions relevant to coastal regions, Results. Eng. 21 (2024) 101734, https://doi.org/10.1016/j.rineng.2023.101734.
[21] M.A. Benitz, M.A. Lackner, D.P. Schmidt, Hydrodynamics of offshore structures with specific focus on wind energy applications, Renew. Sustain. Energy Rev. 44 (2015) 692–716, https://doi.org/10.1016/j.rser.2015.01.021.
[22] P. Dixit, S. Londhe, Prediction of extreme wave heights using neuro wavelet technique, Appl. Ocean Res. 58 (2016) 241–252, https://doi.org/10.1016/j.apor.2016.04.011.
[23] R. Mínguez, A. Tomás, F.J. Méndez, R. Medina, Mixed extreme wave climate model for reanalysis databases, Stochastic Environ. Res. Risk Assess. 27 (2013) 757–768, https://doi.org/10.1007/s00477-012-0604-y.
[24] C.W. Zheng, C.Y. Li, Variation of the wave energy and significant wave height in the China Sea and adjacent waters, Renew. Sustain. Energy Rev. 43 (2015) 381–387, https://doi.org/10.1016/j.rser.2014.11.001.
[25] L. Liu, D. Wang, Z. Peng, Path following of marine surface vehicles with dynamical uncertainty and time-varying ocean disturbances, Neurocomputing. 173 (2016) 799–808, https://doi.org/10.1016/j.neucom.2015.08.033.
[26] M.A. Habib, J.J. O'Sullivan, S. Abolfathi, M. Salauddin, Enhanced wave overtopping simulation at vertical breakwaters using machine learning algorithms, PLoS. One 18 (2023) e0289318, https://doi.org/10.1371/journal.pone.0289318.
[27] A. Yeganeh-Bakhtiary, H. EyvazOghli, N. Shabakhty, S. Abolfathi, Machine learning prediction of wave characteristics: comparison between semi-empirical approaches and DT model, Ocean Eng. 286 (2023) 115583, https://doi.org/10.1016/j.oceaneng.2023.115583.
[28] W. Townsend, An initial assessment of the performance achieved by the Seasat-1 radar altimeter, IEEE J. Oceanic Eng. 5 (1980) 80–92, https://doi.org/10.1109/JOE.1980.1145459.
[29] J. MacArthur, Design of the SEASAT-A Radar Altimeter, in: OCEANS '76, IEEE (1976) 222–229, https://doi.org/10.1109/OCEANS.1976.1154217.
[30] B.D. Tapley, G.H. Born, M.E. Parke, The SEASAT altimeter data and its accuracy assessment, J. Geophys. Res. Oceans. 87 (1982) 3179–3188, https://doi.org/10.1029/JC087iC05p03179.
[31] G.S. Hayne, D.W. Hanock, Sea-state-related altitude errors in the SEASAT radar altimeter, J. Geophys. Res. Oceans. 87 (1982) 3227–3231, https://doi.org/10.1029/JC087iC05p03227.
[32] R.F. Brammer, R.V. Sailor, Preliminary estimates of the resolution capability of the Seasat radar altimeter, Geophys. Res. Lett. 7 (1980) 193–196, https://doi.org/10.1029/GL007i003p00193.
[33] Y. Wan, R. Zhang, X. Pan, C. Fan, Y. Dai, Evaluation of the Significant Wave Height Data Quality for the Sentinel-3 Synthetic Aperture Radar Altimeter, Remote Sens. (Basel) 12 (2020) 3107, https://doi.org/10.3390/rs12183107.
[34] H.-J. Woo, K.-A. Park, Validation of significant wave height from Jason-3 and Sentinel-3A/B and relation to tidal currents in coastal regions of the Korean Peninsula, Int. J. Remote Sens. 43 (2022) 961–996, https://doi.org/10.1080/01431161.2022.2026520.
[35] H.H. Sepulveda, P. Queffeulou, F. Ardhuin, Assessment of SARAL/AltiKa Wave Height Measurements Relative to Buoy, Jason-2, and Cryosat-2 Data, Marine Geodesy 38 (2015) 449–465, https://doi.org/10.1080/01490419.2014.1000470.
[36] H. Lopez, B.P. Kirtman, Investigating the seasonal predictability of significant wave height in the West Pacific and Indian Oceans, Geophys. Res. Lett. 43 (2016) 3451–3458, https://doi.org/10.1002/2016GL068653.
[37] T.-W. Hsu, S.-H. Ou, J.-M. Liau, Hindcasting nearshore wind waves using a FEM code for SWAN, Coast. Eng. 52 (2005) 177–195, https://doi.org/10.1016/j.coastaleng.2004.11.005.
[38] M. Zijlema, A.J. van der Westhuysen, On convergence behaviour and numerical accuracy in stationary SWAN simulations of nearshore wind wave spectra, Coast. Eng. 52 (2005) 237–256, https://doi.org/10.1016/j.coastaleng.2004.12.006.
[39] T.W. Group, The WAM Model—a third generation ocean wave prediction model, J. Phys. Oceanogr. 18 (1988) 1775–1810, https://doi.org/10.1175/1520-0485(1988)018<1775:TWMTGO>2.0.CO;2.
[40] L. Mentaschi, G. Besio, F. Cassola, A. Mazzino, Performance evaluation of Wavewatch III in the Mediterranean Sea, Ocean. Model. (Oxf) 90 (2015) 82–94, https://doi.org/10.1016/j.ocemod.2015.04.003.
[41] J. Monbaliu, R. Padilla-Hernández, J.C. Hargreaves, J.C.C. Albiach, W. Luo, M. Sclavo, H. Günther, The spectral wave model, WAM, adapted for applications with high spatial resolution, Coast. Eng. 41 (2000) 41–62, https://doi.org/10.1016/S0378-3839(00)00026-0.
[42] K. Hasselmann O.D., Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP), JONSWAP (1973).
[43] M. Browne, B. Castelle, D. Strauss, R. Tomlinson, M. Blumenstein, C. Lane, Near-shore swell estimation from a global wind-wave model: spectral process, linear,

and artificial neural network models, Coast. Eng. 54 (2007) 445–460, https://doi.org/10.1016/j.coastaleng.2006.11.007.
[44] M. Zijlema, Computation of wind-wave spectra in coastal waters with SWAN on unstructured grids, Coast. Eng. 57 (2010) 267–277, https://doi.org/10.1016/j.coastaleng.2009.10.011.
[45] A. Zamani, D. Solomatine, A. Azimian, A. Heemink, Learning from data for wind–wave forecasting, Ocean Eng. 35 (2008) 953–962, https://doi.org/10.1016/j.oceaneng.2008.03.007.
[46] N. Kumar, S. Kumar, Evidence of multifaceted SKS/SKKS splitting directions in the Sikkim Himalaya, India, J. Geodyn. 117 (2018) 12–20, https://doi.org/10.1016/j.jog.2018.02.003.
[47] M.C. Deo, A. Jha, A.S. Chaphekar, K. Ravikant, Neural networks for wave forecasting, Ocean Eng. 28 (2001) 889–898, https://doi.org/10.1016/S0029-8018(00)00027-5.
[48] J.D. Agrawal, M.C. Deo, On-line wave prediction, Marine Struct. 15 (2002) 57–74, https://doi.org/10.1016/S0951-8339(01)00014-4.
[49] J. Berbić, E. Ocvirk, D. Carević, G. Lončar, Application of neural networks and support vector machine for significant wave height prediction, Oceanologia 59 (2017) 331–349, https://doi.org/10.1016/j.oceano.2017.03.007.
[50] X. Feng, G. Ma, S.-F. Su, C. Huang, M.K. Boswell, P. Xue, A multi-layer perceptron approach for accelerated wave forecasting in Lake Michigan, Ocean Eng. 211 (2020) 107526, https://doi.org/10.1016/j.oceaneng.2020.107526.
[51] O. Makarynskyy, Improving wave predictions with artificial neural networks, Ocean Eng. 31 (2004) 709–724, https://doi.org/10.1016/j.oceaneng.2003.05.003.
[52] J. Mahjoobi, E. Adeli Mosabbeb, Prediction of significant wave height using regressive support vector machines, Ocean Eng. 36 (2009) 339–347, https://doi.org/10.1016/j.oceaneng.2009.01.001.
[53] I. Malekmohamadi, M.R. Bazargan-Lari, R. Kerachian, M.R. Nikoo, M. Fallahnia, Evaluating the efficacy of SVMs, BNs, ANNs and ANFIS in wave height prediction, Ocean Eng. 38 (2011) 487–497, https://doi.org/10.1016/j.oceaneng.2010.11.020.
[54] S.P. Nitsure, S.N. Londhe, K.C. Khare, Wave forecasts using wind information and genetic programming, Ocean Eng. 54 (2012) 61–69, https://doi.org/10.1016/j.oceaneng.2012.07.017.
[55] N.K. kumar, R. Savitha, A.A. Mamun, Regional ocean wave height prediction using sequential learning neural networks, Ocean Eng. 129 (2017) 605–612, https://doi.org/10.1016/j.oceaneng.2016.10.033.
[56] D. Guijo-Rubio, A.M. Gómez-Orellana, P.A. Gutiérrez, C. Hervás-Martínez, Short- and long-term energy flux prediction using multi-task evolutionary artificial neural networks, Ocean Eng. 216 (2020) 108089, https://doi.org/10.1016/j.oceaneng.2020.108089.
[57] A.M. Gómez-Orellana, D. Guijo-Rubio, P.A. Gutiérrez, C. Hervás-Martínez, Simultaneous short-term significant wave height and energy flux prediction using zonal multi-task evolutionary artificial neural networks, Renew. Energy 184 (2022) 975–989, https://doi.org/10.1016/j.renene.2021.11.122.
[58] C. Jörges, C. Berkenbrink, B. Stumpe, Prediction and reconstruction of ocean wave heights based on bathymetric data using LSTM neural networks, Ocean Eng. 232 (2021) 109046, https://doi.org/10.1016/j.oceaneng.2021.109046.
[59] R. Lou, W. Wang, X. Li, Y. Zheng, Z. Lv, Prediction of ocean wave height suitable for ship autopilot, IEEE Trans. Intell. Transp. Syst. 23 (2022) 25557–25566, https://doi.org/10.1109/TITS.2021.3067040.
[60] M. Pirhooshyaran, L.V. Snyder, Forecasting, hindcasting and feature selection of ocean waves via recurrent and sequence-to-sequence networks, Ocean Eng. 207 (2020) 107424, https://doi.org/10.1016/j.oceaneng.2020.107424.
[61] Z. Wei, Forecasting wind waves in the US Atlantic Coast using an artificial neural network model: towards an AI-based storm forecast system, Ocean Eng. 237 (2021) 109646, https://doi.org/10.1016/j.oceaneng.2021.109646.
[62] M.S. Crouse, R.D. Nowak, R.G. Baraniuk, Wavelet-based statistical signal processing using hidden Markov models, IEEE Trans. Signal Process. 46 (1998) 886–902, https://doi.org/10.1109/78.668544.
[63] A. Durap, Y. Doğan, İnşaat Mühendisliğinde Bilişim Kavramı ve Veri Madenciliği Algoritmaları ile Bir Uzman Sisteminin Oluşturulması, XVII. Akademik Bilişim Konferansı Bildirileri. (2015).
[64] S. Burgin, Sand mining on North Stradbroke Island: an Islander view of the rehabilitation of the lands, Int. J. Environ. Stud. 77 (2020) 954–968, https://doi.org/10.1080/00207233.2020.1727232.
[65] E. Bryant, Behavior of grain size characteristics on reflective and dissipative foreshores, Broken Bay, Australia, SEPM J. Sediment. Res. 52 (1982), https://doi.org/10.1306/212F7F72-2B24-11D7-8648000102C1865D.
[66] M. Wandres, J. Aucan, A. Espejo, N. Jackson, A. De Ramon N'Yeurt, H. Damlamian, Distant-source swells cause coastal inundation on Fiji's coral coast, Front. Mar. Sci. 7 (2020), https://doi.org/10.3389/fmars.2020.00546.
[67] I.D. Goodwin, T.R. Mortlock, S. Browning, Tropical and extratropical-origin storm wave types and their influence on the East Australian longshore sand transport system under a changing climate, J. Geophys. Res. Oceans. 121 (2016) 4833–4853, https://doi.org/10.1002/2016JC011769.
[68] G. Vieira Da Silva, T. Murray, D. Strauss, Longshore wave variability along non-straight coastlines, Estuar. Coast. Shelf. Sci. 212 (2018) 318–328, https://doi.org/10.1016/j.ecss.2018.07.022.
[69] Short, A.D., 2020. Central east region. pp. 517–600. https://doi.org/10.1007/978-3-030-14294-0_18.
[70] B. Farmer, A.D. Short, Australian national surfing reserves – rationale and process for recognising iconic surfing locations, J. Coast. Res. 50 (2007), https://doi.org/10.2112/JCR-SI50-020.1.
[71] A. Durap, C.E. Balas, Ş. Çokgör, E.A. Balas, An integrated bayesian risk model for coastal flow slides using 3-D hydrodynamic transport and monte carlo simulation, J. Mar. Sci. Eng. 11 (2023) 943, https://doi.org/10.3390/jmse11050943.
[72] N. Lazarow, The value of coastal recreational resources: a case study approach to examine the value of recreational surfing to specific locales, J. Coast. Res. 50 (2007), https://doi.org/10.2112/JCR-SI50-003.1.
[73] D. Maulud, A.M. Abdulazeez, A review on linear regression comprehensive in machine learning, J. Appl. Sci. Technol. Trends 1 (2020) 140–147, https://doi.org/10.38094/jastt1457.
[74] P. Tidke, S. Sarode, P.S. Guhe, A review on weather forecasting using linear regression, Int. Sci. J. Eng. Manag. 02 (2023), https://doi.org/10.55041/ISJEM00134.
[75] M. Xu, P. Watanachaturaporn, P. Varshney, M. Arora, Decision tree regression for soft classification of remote sensing data, Remote Sens. Environ. 97 (2005) 322–336, https://doi.org/10.1016/j.rse.2005.05.008.
[76] D. Kang, W. Yu, H. Cho, Decision tree for mode estimation, Korean Data Anal. Soc. 25 (2023) 903–911, https://doi.org/10.37727/jkdas.2023.25.3.903.
[77] L. Wang, X. Zhou, X. Zhu, Z. Dong, W. Guo, Estimation of biomass in wheat using random forest regression algorithm and remote sensing data, Crop J. 4 (2016) 212–219, https://doi.org/10.1016/j.cj.2016.01.008.
[78] R.R. Picard, R.D. Cook, Cross-validation of regression models, J. Am. Stat. Assoc. 79 (1984) 575–583, https://doi.org/10.1080/01621459.1984.10478083.
[79] H. Dai, S. Shang, F. Lei, K. Liu, X. Zhang, G. Wei, Y. Xie, S. Yang, R. Lin, W. Zhang, CRBM-DBN-based prediction effects inter-comparison for significant wave height with different patterns, Ocean Eng. 236 (2021) 109559, https://doi.org/10.1016/j.oceaneng.2021.109559.
[80] M.Á. Gómez Ruiz, I.M. de Almeida, R. Pérez Fernández, Application of Machine Learning Techniques to the Maritime Industry, J. Mar. Sci. Eng. 11 (2023) 1820, https://doi.org/10.3390/jmse11091820.
[81] M. El-Amin, B. Alwated, H. Hoteit, Machine learning prediction of nanoparticle transport with two-phase flow in porous media, Energies. (Basel) 16 (2023) 678, https://doi.org/10.3390/en16020678.
[82] M. Rahman, P.P. Roy, M. Ali, T. Gonc¸alves, H. Sarwar, Software effort estimation using machine learning technique, Int. J. Adv. Comput. Sci. Appl. 14 (2023), https://doi.org/10.14569/IJACSA.2023.0140491.
[83] F.C. Minuzzi, L. Farina, A deep learning approach to predict significant wave height using long short-term memory, Ocean. Model. (Oxf) 181 (2023) 102151, https://doi.org/10.1016/j.ocemod.2022.102151.
[84] J. Xiao, P. Lu, A Hybrid Model of Conformer and LSTM for Ocean Wave Height Prediction, Appl. Sci. 14 (2024) 6139, https://doi.org/10.3390/app14146139.
[85] B. Hou, H. Fu, X. Li, T. Song, Z. Zhang, Predicting significant wave height in the South China Sea using the SAC-ConvLSTM model, Front. Mar. Sci. 11 (2024), https://doi.org/10.3389/fmars.2024.1424714.
[86] B. Son, K. Do, Optimization of SWAN Wave Model to Improve the Accuracy of Winter Storm Wave Prediction in the East Sea, J. Ocean Eng. Technol. 35 (2021) 273–286, https://doi.org/10.26748/KSOE.2021.019.
[87] T. Wang, Z. Yang, W.-C. Wu, M. Grear, A Sensitivity Analysis of the Wind Forcing Effect on the Accuracy of Large-Wave Hindcasting, J. Mar. Sci. Eng. 6 (2018) 139, https://doi.org/10.3390/jmse6040139.
[88] S. Obakrim, V. Monbet, N. Raillard, P. Ailliot, Learning the spatiotemporal relationship between wind and significant wave height using deep learning, Environ. Data Sci. 2 (2023) e5, https://doi.org/10.1017/eds.2022.35.
[89] R.M.A. Ikram, X. Cao, K.S. Parmar, O. Kisi, S. Shahid, M. Zounemat-Kermani, Modeling significant wave heights for multiple time horizons using metaheuristic regression methods, Mathematics 11 (14) (2023) 3141, https://doi.org/10.3390/math11143141.
[90] S. Zhou, B.J. Bethel, W. Sun, Y. Zhao, W. Xie, C. Dong, Improving significant wave height forecasts using a joint empirical mode decomposition–long short-term memory network, J. Mar. Sci. Eng. 9 (2021) 744, https://doi.org/10.3390/jmse9070744.
[91] E.E. Başakın, Ö. Ekmekcioğlu, M. Özger, Developing a novel approach for missing data imputation of solar radiation: a hybrid differential evolution algorithm based eXtreme gradient boosting model, Energy Convers. Manag. 280 (2023) 116780, https://doi.org/10.1016/j.enconman.2023.116780.
[92] E.E. Başakın, Ö. Ekmekcioğlu, M. Özger, N. Altınbaş, L. Şaylan, Estimation of measured evapotranspiration using data-driven methods with limited meteorological variables, Italian J. Agrometeorol. (2021) 63–80, https://doi.org/10.36253/ijam-1055.
[93] Ö. Ekmekcioğlu, K. Koc, I. Dabanli, A. Deniz, Prioritizing urban water scarcity mitigation strategies based on hybrid multi-criteria decision approach under fuzzy environment, Sustain. Cities. Soc. 87 (2022) 104195, https://doi.org/10.1016/j.scs.2022.104195.
[94] K. Koc, Ö. Ekmekcioğlu, A.P. Gurgun, Developing a national data-driven construction safety management framework with interpretable fatal accident prediction, J. Constr. Eng. Manag. 149 (2023), https://doi.org/10.1061/JCEMD4.COENG-12848.
[95] A. Ergin, C.E. Balas, Damage risk assessment of breakwaters under tsunami attack, Natural Hazards 39 (2006) 231–243, https://doi.org/10.1007/s11069-006-0025-7.
[96] M. Korçak, C.E. Balas, Reducing the probability for the collision of ships by changing the passage schedule in Istanbul Strait, Int. J. Disaster Risk Reduc. 48 (2020) 101593, https://doi.org/10.1016/j.ijdrr.2020.101593.
[97] Williams, A.T., P. Davies, A. Ergin, C.E. Balas, 1998. Coastal Recession and the Reliability of Planned Responses: Colhuw Beach, the Glamorgan Heritage Coast, Wales, UK, in: SPRING 1998, SPECIAL ISSUE NO. 26. PROCEEDINGS OF THE PALM BEACH INTERNATIONAL COASTAL SYMPOSIUM 19-23 May 1998. Journal of Coastal Research, PALM BEACH , pp. 72–79.