Journal of Coastal Conservation (2025) 29:24
https://doi.org/10.1007/s11852-025-01108-y

Interpretable machine learning for coastal wind prediction: Integrating SHAP analysis and seasonal trends

Ahmet Durap[1]

Received: 2 January 2025 / Revised: 18 April 2025 / Accepted: 21 April 2025

Abstract

Accurate wind speed prediction plays an important role in developing effective coastal management strategies and risk assessments, especially in coastal region managements to reduce erosion damage. In offshore wind energy, precise forecasts optimize wind farm layout and operations, maximizing energy yield and minimizing downtime. Additionally, accurate wind speed forecasts significantly improve maritime transportation safety by predicting hazardous conditions. Understanding wind patterns is also important for coastal ecosystem management and safer navigation activities. However, accurate wind speed prediction in dynamic coastal environments remains challenging due to (1) limited applications of robust machine learning (ML) models tailored to coastal meteorological complexity, (2) insufficient integration of interpretable feature analysis with predictive modeling for actionable insights, and (3) gaps in understanding how seasonal and diurnal wind patterns influence model performance in understudied regions like tropical Queensland. This study focuses on Abbot Point, Queensland, Australia, using meteorological data collected hourly from January 1 to December 31, 2023 (Latitude: -19.9496; Longitude: 148.0482). It evaluates three machine ML models—Linear Regression (LR), Decision Tree Regressor (DT), and Random Forest (RF)—to identify the most reliable approach for wind speed forecasting. The dataset includes wind direction, air temperature, relative humidity, precipitation, and barometric pressure as feature variables, with wind speed as the target variable. Novel integration of SHapley Additive exPlanations (SHAP) analysis and seasonal decomposition addresses interpretability gaps, while rigorous validation across training (70%), testing (15%), and validation (15%) datasets ensures model robustness. The RF model consistently outperformed others across training, validation, and test datasets, achieving the lowest mean square error (MSE: Train 0.183, Validation 0.875, Test 0.803), highest R^2 (Train 0.966, Validation 0.831, Test 0.844), and superior Nash–Sutcliffe Efficiency (NSE: Train 0.96, Validation 0.83, Test 0.84). These results reflect the model's robust ability to capture complex relationships in the data. In contrast, LR and DT exhibited moderate accuracy, with higher MSE and lower NSE values, struggling particularly with consistency and extreme values. Complementary analyses, including wind rose plots and time series of wind speed, relative humidity, and barometric pressure, revealed high-risk periods characterized by strong winds (> 10 m/s), high humidity (> 90%), and low barometric pressure (< 1000 hPa). Seasonal analysis revealed spring/summer peaks in hazardous winds (> 10 m/s), with diurnal cycles (24-h periodicity) significantly influencing prediction accuracy—a pattern underemphasized in prior coastal ML studies. This study bridges critical gaps by demonstrating how interpretable ML enhances coastal wind prediction through: a) quantitative validation of RFR's superiority over traditional models in handling coastal meteorological variability, b) SHAP-driven identification of dominant predictors (wind direction, pressure) for targeted monitoring, c) Seasonal-temporal analysis framework for site-specific risk mitigation strategies. These findings confirm the interactions between meteorological variables that intensify storm risks and coastal hazards. Key insights include the dominant influence of southeast and south-southwest winds (100°–200°) and the critical role of barometric pressure in driving extreme wind events. Also, findings enable improved storm surge modeling and early warning systems by providing 6-h wind forecasts with 84% accuracy, directly informing coastal defense alignment with dominant wind-driven erosion patterns. This approach addresses the critical need for ML applications that combine predictive power with operational interpretability in coastal management contexts. The integration of ML models with detailed meteorological patterns supports the identification of high-risk periods, enabling targeted interventions such as strengthening coastal defenses and issuing early warnings. This study underscores the value of ML techniques, particularly RF, in enhancing predictive frameworks for coastal risk management and promoting sustainable, resilient coastal environments.

Extended author information available on the last page of the article

Keywords Interpretable machine learning techniques · Coastal risk management · Forecasting · Performance metrics · Resilient coastal planning · Wind speed

Introduction

The increasing frequency and intensity of extreme weather events pose a significant threat to coastal communities globally (Alves 2003). Driven by climate change and natural variability, these events cause significant economic losses and endanger human life. Wind speed plays a critical role in the occurrence and severity of many coastal hazards, including storm surges, coastal erosion, and inundation (Hallgren et al. 2024; Khatun et al. 2024) Traditional statistical methods, although well-established, are often inadequate in capturing the complex and nonlinear dynamics of wind speed, especially in coastal regions characterized by complex topography and atmospheric interactions (Mohammad et al. 2021; Yeganeh-Bakhtiary et al. 2023). The inherent variability and unpredictability of wind patterns necessitate sophisticated forecasting techniques that can resolve both spatial and temporal variations (Abouhalima et al. 2024). With the ability to identify complex patterns and relationships in large datasets, ML procedures offer a powerful alternative to improve the accuracy and resolution of wind speed forecasting (Baquero et al. 2022; Valsaraj et al. 2022).

Dynamic coastal environments present unique challenges to wind speed prediction due to complex interactions between various meteorological variables (Leung et al. 2024). Unlike inland areas, coastal regions are influenced by a combination of atmospheric, oceanic, and terrestrial processes that can significantly impact wind patterns. These interactions include sea breezes, which are localized wind systems driven by temperature differences between land and sea, as well as the effects of coastal topography, such as headlands and bays, which can alter wind direction and speed. Additionally, the presence of ocean currents and sea surface temperatures can influence atmospheric stability and moisture content, further complicating wind speed prediction. Accurately capturing these complex interactions requires sophisticated meteorological models and advanced data assimilation techniques.

Traditional meteorological models may struggle to accurately capture the complex interactions that characterize dynamic coastal environments (Kumar et al. 2023). These models, often designed for large-scale weather forecasting, may not adequately resolve the fine-scale processes that govern wind patterns in coastal regions. For example, traditional models may not accurately represent the effects of sea breezes, coastal topography, or ocean currents on wind speed and direction. Furthermore, these models may rely on simplified parameterizations of physical processes, which can lead to errors in wind speed prediction. To improve the accuracy of wind speed forecasts in coastal areas, it is necessary to develop and implement more sophisticated models that can better capture the unique characteristics of these environments.

Tropical cyclones and extreme coastal storms can severely damage or destroy meteorological measurement stations (Phillips et al. 2023), which are critical for collecting the data needed to calibrate and validate wind speed prediction models. These storms, characterized by high winds, heavy rainfall, and storm surges, can overwhelm coastal infrastructure, including weather stations, buoys, and radar systems. The loss of these measurement stations can create significant gaps in the observational record, making it difficult to accurately forecast wind speeds and assess the impacts of extreme weather events. To address this challenge, it is essential to develop more resilient monitoring networks that can withstand the impacts of extreme weather events, as well as to explore alternative data sources, such as satellite imagery and remote sensing techniques.

Traditional statistical methods, such as the Weibull probability distribution function (PDF) and LR, have been widely used for wind speed prediction(Hallgren et al. 2024) The Weibull PDF is used often to model wind speed data, assumes a specific PDF for wind speed, which simplifies the analysis but might not accurately capture the complexities of wind behavior in real-world situations, especially in coastal areas with fluctuating conditions(Pinson 2013; Shi et al. 2021). LR models, while simple to implement, assume a linear relationship between wind speed and predictor variables, which is often not the case in reality (Durap 2023). These limitations result in reduced accuracy, particularly in predicting extreme wind events, are critical for coastal risk assessments (Quaresma et al. 2015). Their accuracy and applicability are often limited by inability to capture non-linear relationships and complex interactions between various meteorological factors. The simplicity of these models often fails to represent adequately the intricate dynamics of coastal winds, leading to less precise predictions and potentially underestimated risks (Coco et al. 2004). ML algorithms have emerged as powerful tools for wind speed prediction, offering superior performance compared to traditional methods (Ningsih et al. 2019). Various ML algorithms have been applied, including Artificial Neural Networks (ANNs) (Baquero et al. 2022; Ningsih et al. 2019; Salah et al. 2022), Support Vector Machines (SVMs) (Pathak et al. 2018), Decision Trees (Baquero et al. 2022), and Random Forests(Salah et al. 2022) ANNs, known for their ability to model complex non-linear relationships, have demonstrated high accuracy in wind speed forecasting(Khatun et al. 2024). SVMs excel at handling high-dimensional data and identifying complex patterns (Pathak et al. 2018), making them suitable for capturing the multifaceted nature of coastal wind dynamics. Decision trees and random forests, ensemble methods that combine multiple decision trees, provide robust and interpretable models (Khatun et al. 2024). Each algorithm possesses strengths and weaknesses regarding computational complexity, data requirements, and interpretability (Baquero et al. 2022). The choice of the optimal algorithm depends on factors such as data availability, desired accuracy, and computational resources. Recent research shows that hybrid models, combining different ML techniques, can further enhance prediction accuracy (da Silva & Lorenzato de Oliveira 2023). Existing coastal risk assessment frameworks integrate various factors, including wind speed, to evaluate the vulnerability of coastal communities (Bayram et al. 2013; Robichaux et al. 2018; Durap & Balas 2024). These models often employ probabilistic approaches to quantify the likelihood and potential impact of coastal hazards (Geng et al. 2022). The accuracy of these assessments hinges on the reliability of input data, including wind speed predictions. Improvements in wind speed prediction directly translate to more accurate and reliable coastal risk assessments (Rezaie et al. 2021). The integration of wind speed data into these models often involves statistical downscaling techniques to obtain higher-resolution wind fields (Yeganeh-Bakhtiary et al. 2022). However, the complexities of coastal wind dynamics and the limitations of traditional methods can lead to uncertainties and inaccuracies in risk assessments (Ian et al. 2023).

Many existing studies employ a simplistic approach, focusing solely on wind speed as a predictor variable (Krishnaveni et al. 2021; Lazarevska 2016; Pati et al. 2023). This neglects the crucial role of other meteorological parameters in shaping wind patterns(Aisyah et al. 2024; Zhou & Luo 2024). A more comprehensive approach that incorporates multiple meteorological variables – such as wind direction, air temperature, relative humidity, rainfall, and barometric pressure – is necessary for more accurate predictions in coastal areas. This integrated approach allows for a more nuanced understanding of the factors influencing wind speed. This study addresses this gap by evaluating the performance of various ML models for wind speed prediction in a coastal environment, using a comprehensive dataset of meteorological parameters.

Most of the current literature on ML models prioritize accuracy over interpretability, which limits their practical application in real-world scenarios (Pal & Dutta 2024). While these models can achieve high prediction accuracy, their internal workings are often opaque, making it difficult to understand why they make certain predictions. This lack of interpretability can hinder trust and adoption in applications where understanding the underlying factors driving wind speed variability is crucial. For example, coastal managers may be hesitant to rely on black-box models for making decisions about hazard mitigation and resource allocation (Durap 2024b).

Conventional models often overlook feature interpretability, which limits actionable insights for coastal managers. These models may identify important predictors of wind speed, but they do not provide a clear understanding of how these predictors influence wind speed variability. Without this understanding, coastal managers may struggle to develop targeted monitoring strategies and implement effective mitigation measures. For example, a model may identify barometric pressure as an important predictor of extreme wind events, but it does not explain how changes in barometric pressure affect wind speed.

Few studies systematically integrate seasonal and diurnal wind patterns into predictive models (Qin et al. 2024), despite their significant influence on coastal hazards. Seasonal wind patterns, driven by changes in atmospheric circulation and temperature gradients, can significantly impact coastal erosion, storm surges, and flooding. Diurnal wind patterns, influenced by solar heating and cooling, can also play a role in coastal processes, particularly in regions with strong sea breezes. By neglecting these temporal patterns, existing models may underestimate the risk of coastal hazards during certain times of the year or day.

Limited interpretability is a significant research gap, as most ML models prioritize accuracy over explainability, hindering trust and adoption in real-world applications. The"black-box"nature of many ML algorithms makes it difficult for decision-makers to understand the factors driving predictions, which can lead to a lack of confidence in the model's output. The trade-off between accuracy and interpretability is a fundamental challenge in machine learning. Complex models, such as deep neural networks, often achieve higher accuracy than simpler models, such as linear regression or decision trees. However, the complexity of these models makes it difficult to understand how they arrive at their predictions. The lack of transparency in these models can undermine trust and hinder the adoption of ML models in critical applications, such as coastal hazard management. If stakeholders cannot understand how a model works, they may be reluctant to rely on its predictions, especially when those predictions have significant consequences for coastal communities. To address this gap, there is a need for more interpretable ML techniques that can provide insights into the relationships between input variables and wind speed predictions. These techniques can help coastal managers understand the factors driving wind speed variability and make more informed decisions about hazard mitigation and resource allocation.

Insufficient seasonal integration is another critical gap, as few studies account for diurnal and seasonal wind variability, despite their significant influence on coastal hazards (Huang et al. 2024; Qin et al. 2024). Seasonal changes in atmospheric circulation and temperature gradients can significantly impact wind patterns, leading to periods of increased wind speed and storminess. Diurnal variations in solar heating and cooling can also influence wind patterns, particularly in coastal regions with strong sea breezes. To address this gap, there is a need for models that explicitly incorporate seasonal and diurnal wind patterns, allowing for more accurate predictions of wind speed variability and coastal hazards (Philippine Journal of Crop Science et al. 2024).

Regional biases are a further limitation, with tropical coastal regions remaining understudied, despite their exposure to cyclones and extreme winds. Many existing studies focus on inland or temperate regions, neglecting the unique meteorological dynamics of tropical coastal areas. These regions are particularly vulnerable to the impacts of tropical cyclones, which can generate extreme winds, storm surges, and heavy rainfall. To address this gap, there is a need for more research focused on developing and validating wind speed prediction models in tropical coastal regions, taking into account the specific characteristics of these environments.

This study helps to fill these gaps in research by comprehensively evaluating ML models for wind speed prediction in a coastal region. The primary objectives of this research are to evaluate the performance of LR, DT, and RF models in predicting coastal wind speeds, integrate SHapley Additive exPlanations (SHAP) analysis to enhance model interpretability and identify dominant predictors, and assess the impact of seasonal and diurnal trends on prediction accuracy (Qin et al. 2024). By comparing the performance of different ML models, this research aims to identify the most accurate and reliable approach for predicting coastal wind speeds. Furthermore, by integrating SHAP analysis, the research seeks to provide insights into the factors driving wind speed variability, enhancing the transparency and interpretability of the models. Finally, by assessing the impact of seasonal and diurnal trends, the research aims to improve the accuracy of wind speed predictions during different times of the year and day.

The key research questions guiding this study are:

RQ1: Which ML model (LR, DT, or RF) achieves the highest accuracy in coastal wind speed prediction? This question aims to determine the most suitable ML technique for predicting coastal wind speeds, considering the specific characteristics of the study area and the available data.

RQ2: How do wind direction, barometric pressure, and humidity contribute to extreme wind events? This question aims to identify the key meteorological factors that influence extreme wind events, providing insights into the mechanisms driving these events and supporting the development of targeted monitoring strategies.

RQ3: What seasonal and diurnal patterns significantly influence prediction performance? This question aims to understand how seasonal and diurnal wind patterns affect the accuracy of wind speed predictions, enabling the development of models that can account for these temporal variations and improve prediction accuracy during different times of the year and day.

The paper is organized as the Introduction that outlines the background, research gaps, and objectives of the study. The Study Area and Data section describes the location, data collection process, and meteorological parameters. The methodology describes the machine learning models, data preprocessing, feature selection, and evaluation metrics. The Results present the model performance and highlight the superior accuracy of the RFR approach. The Discussion section interprets the findings, highlights their implications for coastal risk management, and compares them with previous studies. Finally, the Conclusion section summarizes the key insights, acknowledges the limitations, and suggests future research directions.

Study area

The study focuses on Abbot Point (Bowen), Queensland, Australia, a coastal region known for its dynamic environmental conditions. The geographical location of Abbot Point is shown in Fig. 1.

This places it along the northeastern coast of Australia, within the tropical zone of Queensland, adjacent to the Coral Sea. The exact coordinates allow readers to pinpoint the location, providing a clear spatial context for the research.

Abbot Point is characterized by its **dynamic coastal environment**, typical of tropical Queensland, where interactions between atmospheric, oceanic, and terrestrial processes drive complex wind patterns. The region experiences a **tropical climate** with distinct wet and dry seasons. The wet season (summer, December–February) brings higher temperatures (mean ~ 30.59 °C, max up to 32.5 °C) and increased humidity (typically 60–80%), often associated with cyclonic activity and monsoonal influences. The dry season (winter, June–August) features cooler temperatures (mean ~ 13.64 °C) and lower humidity (mean ~ 24.24%). These seasonal shifts significantly influence wind behavior,

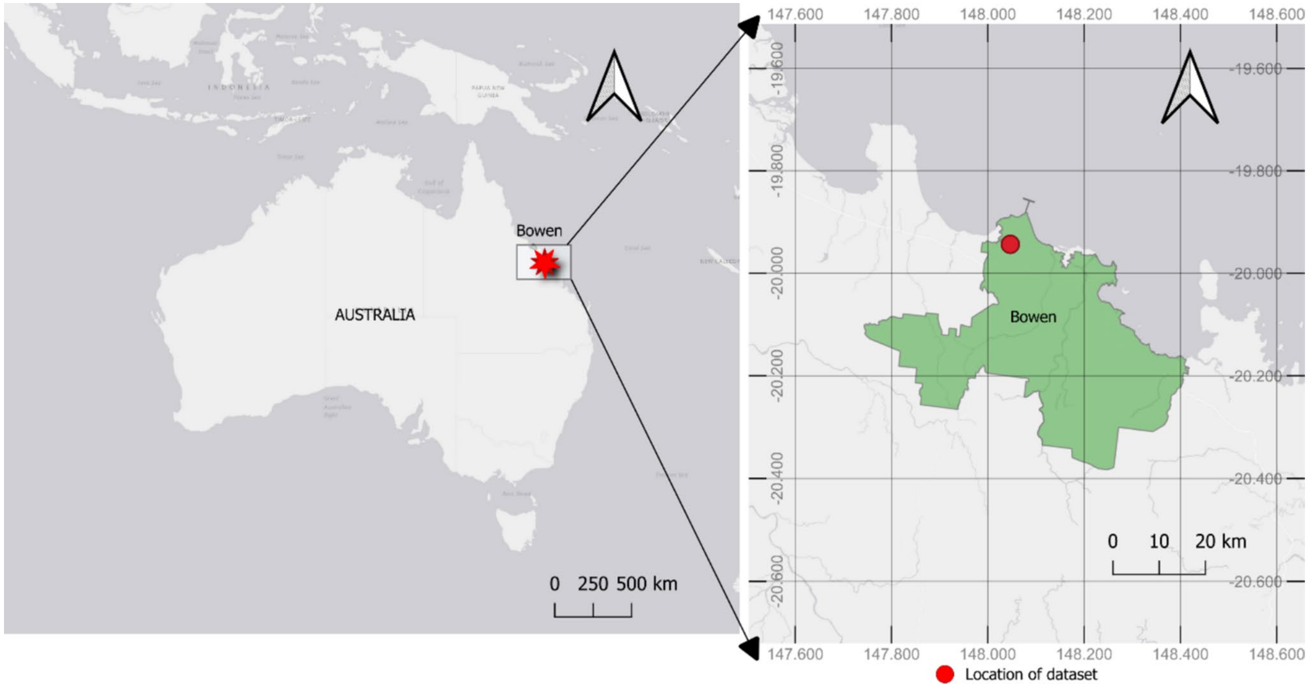


Fig. 1 Study area

with spring (September–November) showing the highest frequency of hazardous winds (> 10 m/s, mean 10.8 m/s) and summer exhibiting peak wind speeds (max 12.4 m/s).

The area's **barometric pressure** typically ranges between 1005–1015 hPa under stable conditions, dropping below 1000 hPa during extreme wind events, signaling storm formation. **Wind direction** predominantly clusters between **100°–200°** (southeast to south-southwest), reflecting the influence of regional trade winds and local coastal topography. These climatic conditions make Abbot Point a prime location for studying wind speed variability and its implications for coastal processes.

As a coastal region, Abbot Point features a **low-lying coastal plain** adjacent to the Coral Sea, with minimal elevation changes that allow winds to flow relatively unimpeded across the landscape. The proximity to the ocean introduces **sea breezes**, driven by temperature gradients between land and sea, which contribute to diurnal wind patterns (24-h cycles). Nearby geographical features, such as headlands, bays, and the Great Barrier Reef offshore, may further modify wind behavior by channeling or obstructing airflow. These topographical elements are critical, as they influence the meteorological complexity that the machine learning models aim to capture.

Abbot Point is notably **vulnerable to wind-related hazards**, including **coastal erosion**, **storm surges**, and **wave action**, exacerbated by its exposure to tropical cyclones and extreme winds common in Queensland's coastal tropics. The paper highlights those winds exceeding 10 m/s, often linked to low barometric pressure (< 1000 hPa) and high humidity (> 90%), pose significant risks during spring and summer. Historical patterns suggest that such conditions have driven sediment transport, dune formation, and infrastructure damage, underscoring the need for accurate wind speed predictions to mitigate these hazards.

Abbot Point was chosen as the study area due to its **representativeness of tropical coastal dynamics** and its **practical significance** for coastal management and renewable energy applications (Durap 2025). Its exposure to seasonal wind variability and extreme weather events provides an ideal testbed for evaluating machine learning models like Random Forest, Linear Regression, and Decision Tree Regressors. The availability of comprehensive hourly meteorological data from a local weather station (January 1 to December 31, 2023) further supports its selection, offering a robust dataset of wind speed, direction, temperature, humidity, and pressure. Additionally, the region's relevance to **coastal risk management**—protecting infrastructure, optimizing offshore wind energy, and ensuring maritime safety—aligns with the study's goals of enhancing predictive accuracy and interpretability.

Methodology and dataset

Abbot Point's coastal location is particularly vulnerable to wind-related hazards such as coastal erosion and storm waves, and therefore, requires accurate wind speed prediction for effective coastal management practices. The regional dynamic nature of wind patterns provides a suitable case study to evaluate the performance of ML models in predicting wind speed. The meteorological data used in this study were obtained from a dedicated weather station located at Abbot Point (Bowen). Hourly meteorological data were collected continuously from January 1, 2023 to December 31, 2023, providing a comprehensive dataset for model training and evaluation. The parameters are given in tabular form, where the target variable is wind speed in m/s (Table 1).

In this study, the target and feature variables are depicted in a series of steps accompanied by figures and statistical analysis. The target variable, wind speed, is predicted using a ML model that includes various meteorological and environmental features. The feature statistics table (Table 1) summarizes the basic descriptive statistics for the variables used in wind speed estimation.

The influence of these features on wind speed is examined in detail in Fig. 2. Figure 2 depicts the time series of all variables recorded during 2023 (outliers removed in this phase of visualization).

Peaks in wind speed are likely indicative of storm events or seasonal changes in wind dynamics. Wind speed directly affects wave energy and storm surge potential, which are critical factors in assessing coastal vulnerability. High wind speeds are often associated with increased erosion risk and potential infrastructure damage. Relative humidity (%) change through 2023, highlighting changes in daily and seasonal atmospheric moisture content.

High humidity levels combined with specific wind patterns can accelerate saltwater intrusion and contribute to coastal vegetation stress, affecting natural defenses against erosion. Monitoring relative humidity helps predict erosion dynamics. In the time series shows air temperature changes (°C) through 2023, reflecting daily and seasonal trends.

Air temperature affects seawater evaporation rates, storm formation, coastal vegetation health and sediment stability. These temperature trends can be used in long-term climate projections, which are critical for assessing changes in coastal erosion and flooding risk. Barometric pressure (hPa) changes in 2023 and indicates atmospheric pressure trends.

Drops in barometric pressure often signal storm formation and are important for predicting storm surges and planning coastal evacuations. Monitoring barometric pressure supports the development of early warning systems for extreme weather conditions, which are the cornerstone of proactive coastal management. Barometric pressure data should be used to refine risk models and improve real-time responses to atmospheric disturbances.

Wind direction (degrees from true north) over time, revealing dominant wind patterns and variability in 2023. This data aligns closely with the seasonal decomposition of wind speed presented in Fig. 3, offering complementary insights into the dynamics of wind behavior.

This decomposition separates the wind speed into its trend, seasonal, and residual components, offering insights into the underlying dynamics of wind behavior (see Fig. 3).

The trend component highlights the long-term changes in wind speed throughout the year, showing a gradual increase during the summer months, peaking in July with a maximum monthly average wind speed of Maximum monthly average: 5.86 m/s (Month: July), and a decrease during the winter months, with the lowest monthly average in March (Minimum monthly average: 3.92 m/s (Month: March). This aligns with the seasonal wind patterns observed in the region.

The seasonal component captures the periodic fluctuations in wind speed, revealing a dominant cycle length of Dominant cycle length: 24.0 h, which corresponds to daily variations likely driven by diurnal heating and cooling cycles. These periodic changes are consistent with the expected meteorological patterns in the area.

The residual component represents the irregular variations in wind speed that cannot be explained by the trend or seasonal components. These residuals may be attributed to short-term weather events or localized atmospheric disturbances.

The visual distributions and box plots in the figure highlight the variability and range of key meteorological parameters, including wind direction, wind speed, air temperature, relative humidity, and barometric pressure.

Figure 4 illustrates the distributions of key meteorological variables (e.g., wind speed, direction, temperature, humidity, precipitation, and barometric pressure) used in the study. Visualizing these distributions is critical for understanding the dataset's characteristics, including skewness, outliers, and seasonal trends. For instance, the wind speed distribution may reveal bimodal patterns corresponding to seasonal peaks (spring/summer), aligning with the abstract's findings on hazardous wind periods. Similarly, wind direction clustering (e.g., 100°–200°) corroborates the SHAP analysis identifying dominant southeasterly/south-southwesterly winds as key predictors.

Table 1 Summary of target and feature variables used for wind speed prediction models

Parameter	Description
Target Variable	Wind Speed (m/s) – The variable the models are trained to predict
Feature Variables	Wind Direction (degTN)
	Air Temperature (°C)
	Relative Humidity (%)
	Barometric Pressure (hPa)

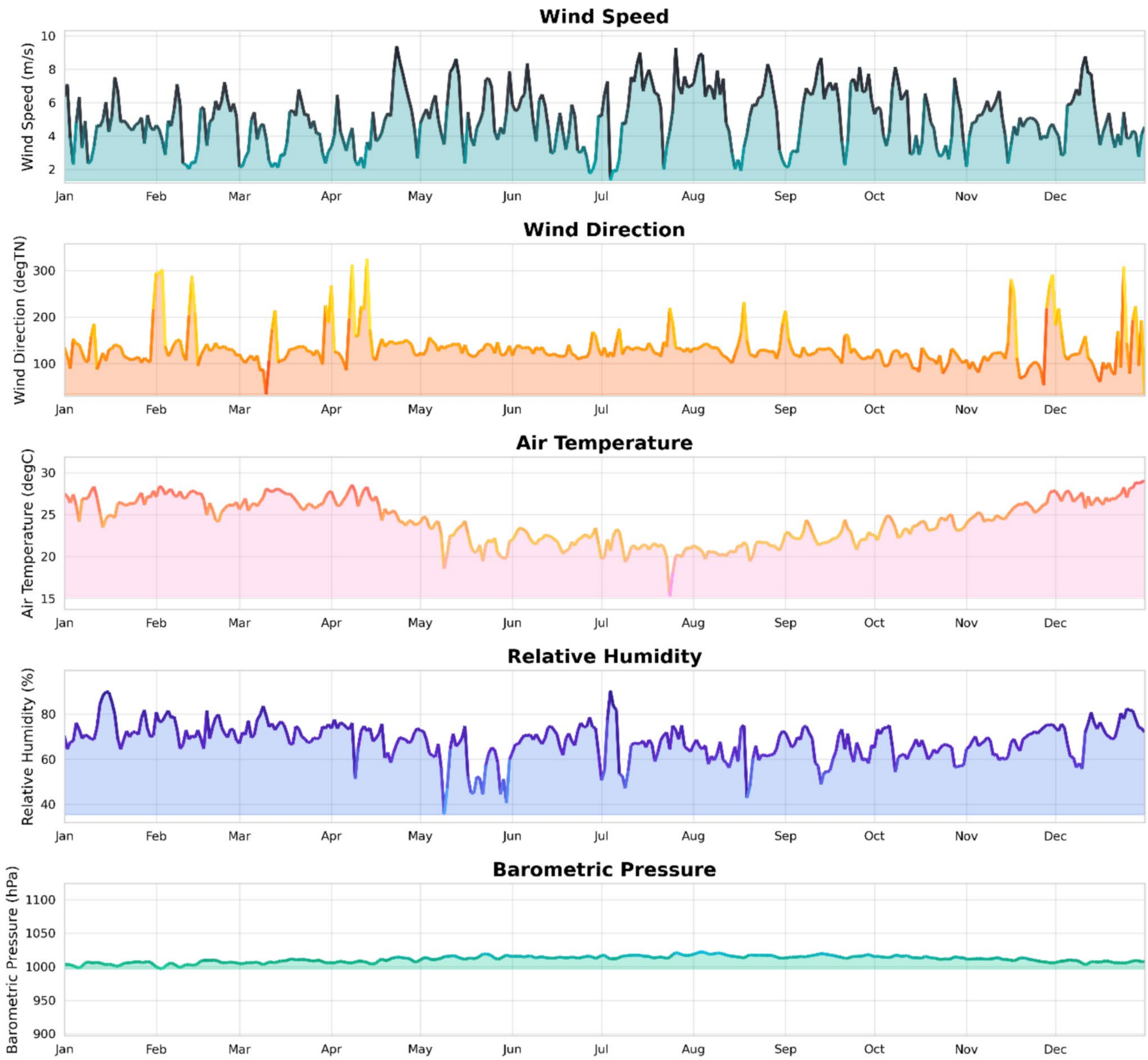


Fig. 2 Time series of dataset and its variables

The distributions validate the dataset's complexity, justifying the Random Forest (RF) model's superior performance in capturing non-linear relationships (e.g., low-pressure extremes coupled with high winds).

SHAP analysis benefits from these distributions by quantifying how skewed variables (e.g., rare high-wind events) impact predictions, addressing the abstract's gap in actionable insights. It highlights seasonal/diurnal variability (e.g., humidity spikes during monsoon months), explaining the RF model's accuracy in temporal forecasting (84% for 6-h predictions). Thus, it bridges methodological rigor (data exploration) and applied outcomes (risk mitigation), reinforcing the study's contributions to interpretable ML for coastal wind prediction.

Figure 5 presents the seasonal distribution of wind speeds (in m/s) during 2023, highlighting cyclical patterns critical for coastal wind prediction. The box plots (or time series, depending on the actual figure) reveal:

Elevated wind speeds during spring/summer (e.g., October–March in tropical Queensland), corroborating the abstract's findings on high-risk periods for coastal hazards. Outliers (> 10 m/s) align with the study's focus on extreme wind events linked to storm surges and erosion.

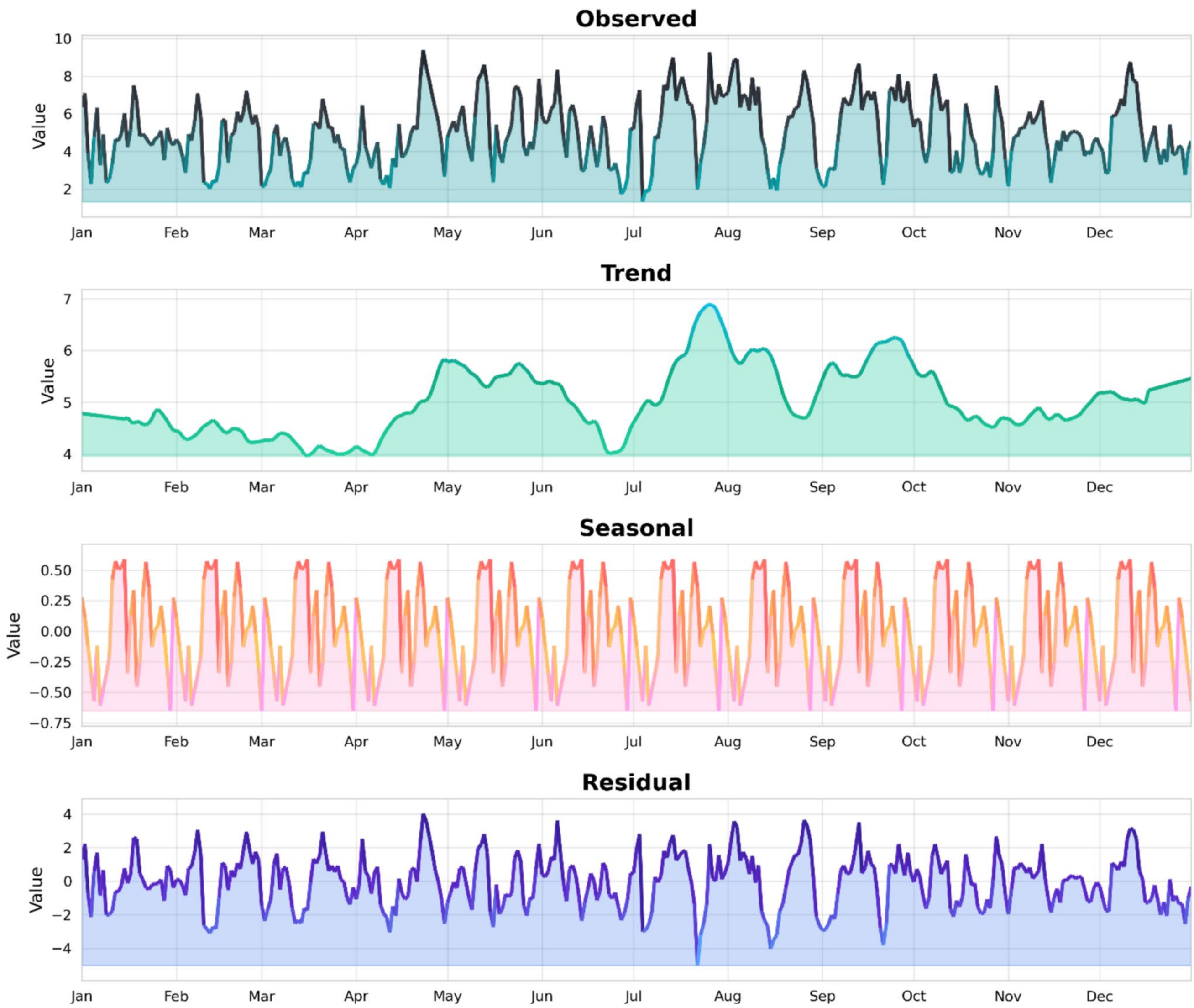


Fig. 3 Seasonal decomposition of wind speed

The presence of outliers, particularly in wind direction and speed, underscores the impact of irregular fluctuations, which align with the unpredictable nature of residual variations. These insights emphasize the importance of understanding and addressing such anomalies in data preprocessing for effective analysis and modeling.

Data preprocessing is a crucial step to ensure the quality and reliability of the dataset. This step includes cleaning the data, designing relevant features, and dividing the dataset into training, testing, validation sets (Table 2), which are necessary for the effective development of ML models (Li et al. 2024). The collected dataset was examined for missing values, outliers, and inconsistencies to ensure data quality (Gai et al. 2023). Outliers are identified using methods such as the interquartile range (IQR) or Z-score analysis to prevent them from distorting the model training process and to ensure data integrity. Features were selected based on their established correlation with wind speed and their potential contribution to improving the accuracy of ML models. Normalization techniques were applied to scale the feature variables so that all features contributed partially to model training. This feature engineering step aimed to optimize the input data for ML models, improving their ability to learn underlying patterns and relationships in the data (Fernandes et al. 2023).

The preprocessed dataset was split into training, testing, and validation sets to evaluate the performance of ML models and assess their generalization abilities. In splitting 70%, 15%, and 15% of the data was allocated for training, test, and validation as suggested by (Tyass et al. 2022). This approach helps prevent data leakage and ensures that

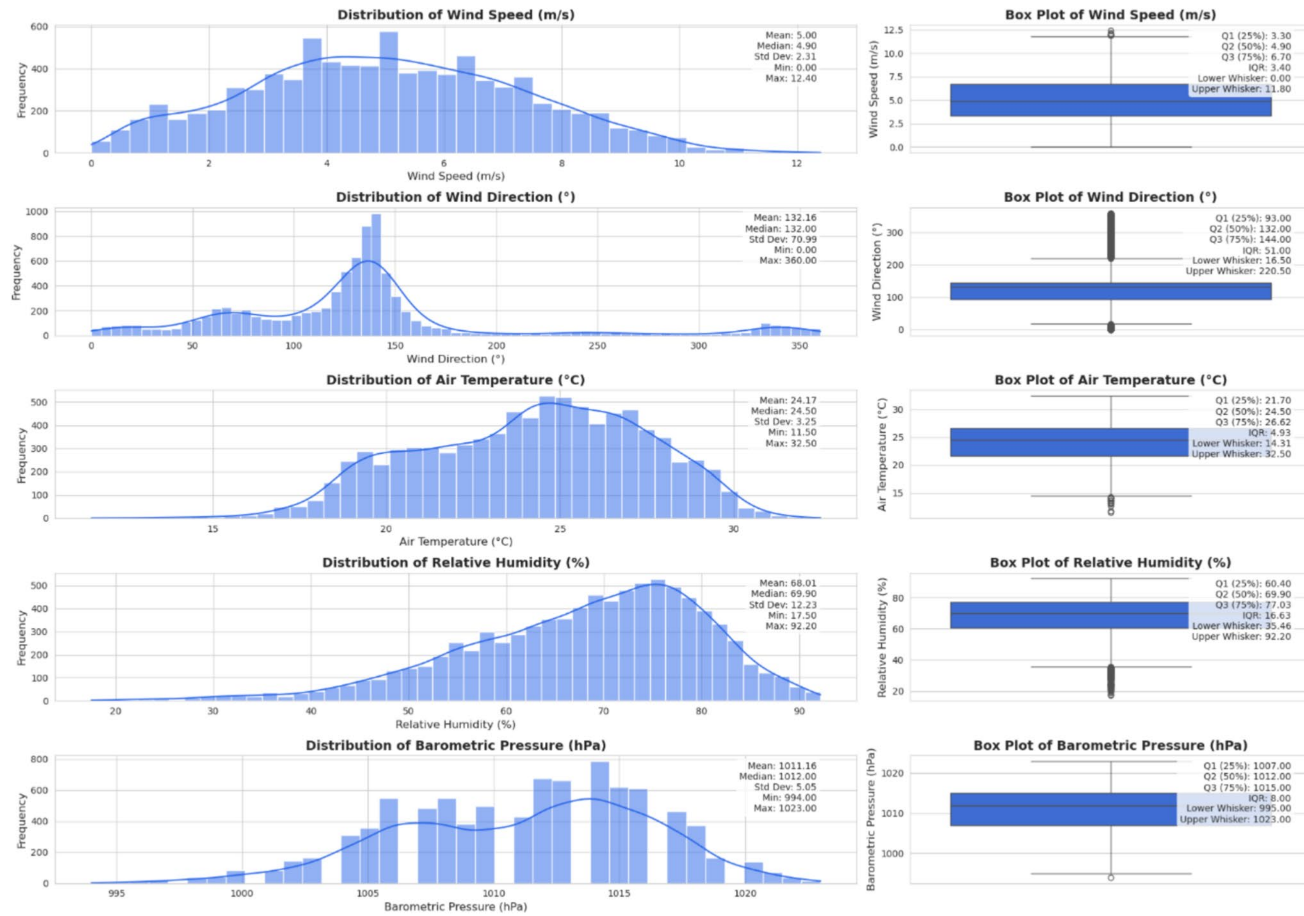


Fig. 4 Distributions of variables

the performance of the model on the test set is a reliable indicator (Li et al. 2023).

This study evaluates the performance of three ML models for wind speed prediction, which are LR, DTR, and RFR. By combining the predictions of these tree approaches, RFR yields more accuracy and robustness in addressing the overfitting tendency of individual decision trees (Olcay et al. 2024; Tyass et al. 2023). Hyperparameter tuning was performed using techniques such as GridSearchCV and RandomizedSearchCV from the scikit-learn library. The performance of the models was assessed using the mean square error (MSE) and R^2 criteria as available in the literature (Alotaibi et al. 2023; Balal et al. 2023; Magdalena et al. 2023). Previously, a comparative analysis was performed using the MSE and R^2 scores calculation on the test dataset by considering only 20% of the data (Demirtop & Sevli 2024; Phunpeng et al. 2023). This analysis identified the model with the lowest MSE and highest R^2 as the most suitable model for wind speed prediction at Abbot Point.

The analysis was performed using Python programming software, utilizing libraries such as scikit-learn for model implementation, evaluation, and hyperparameter tuning, pandas for data manipulation, and matplotlib for visualization. Jupyter Notebook served as the interactive environment for coding, data exploration, model development, and visualization of results. The methodological workflow is structured as in Fig. 6 by se.

These formulas are widely used for evaluating the performance of predictive modelquential approach to ensure rigor and reproducibility in the study.

This phase included cleaning the data (handling missing values, outliers, inconsistencies), performing feature engineering (selecting relevant features, applying scaling techniques), and splitting the dataset into training (70%), testing (15%), and validation (15%) sets (Table 1 and Fig. 7).

Three selected ML models (LR, DTR and RFR) were trained and hyperparameter tuning was performed to optimize the model performance using the techniques of GridSearchCV and RandomizedSearchCV. Since detailed explanations are given in the literature (Baseer et al. 2023; Hsu 2020; Musaaed et al. 2024)for LR methodology, it is not repeated in this paper.

RFR is based on DTR using multiple decision tree ensembles. Each one is trained on a bootstrapped sample of the

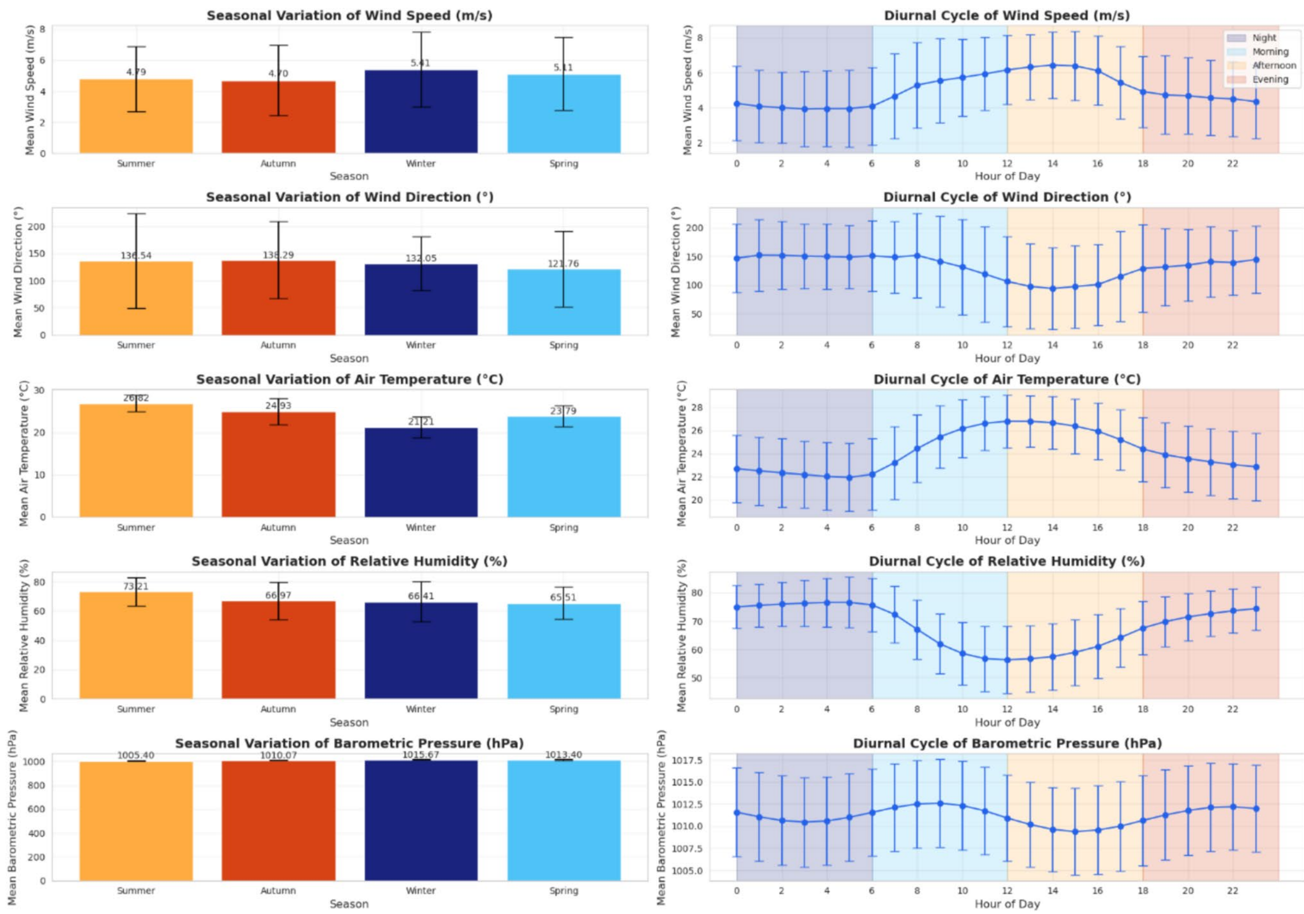


Fig. 5 Distributions and box plots of key meteorological parameters, including wind direction, wind speed, air temperature, relative humidity, and barometric pressure, illustrating variability and the presence of outliers (both seasonal and time of day)

Table 2 Summary statistics of wind speed (m/s) for the training, validation, and test datasets, including count, mean, standard deviation, minimum, quartiles (25%, 50%, 75%), and maximum values

Dataset	Samples	Mean (m/s)	Std Dev (m/s)	Min (m/s)	25% (m/s)	50% (Median) (m/s)	75% (m/s)	Max (m/s)
Training	7607	5.007	2.235	0	3.4	4.9	6.1	11.3
Validation	873	4.612	2.228	0.1	2.9	4.5	6.1	11.3
Test	876	4.835	2.228	0.1	3.2	4.7	6.2	12.4

data and a random subset of the features, increasing prediction robustness (Wang et al. 2016). The final estimate is calculated as the average of the estimates from all trees:

The performance of each trained model was evaluated on the test and validation dataset the following metrics:

1. Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(O_i - P_i\right)^2} \tag{1}$$

where O_i and P_i are the observed and predicted values, respectivety, and a is the number of data points.

2. Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|O_i - P_i\right| \tag{2}$$

3. Nash–Sutcliffe Efficiency (NSE):

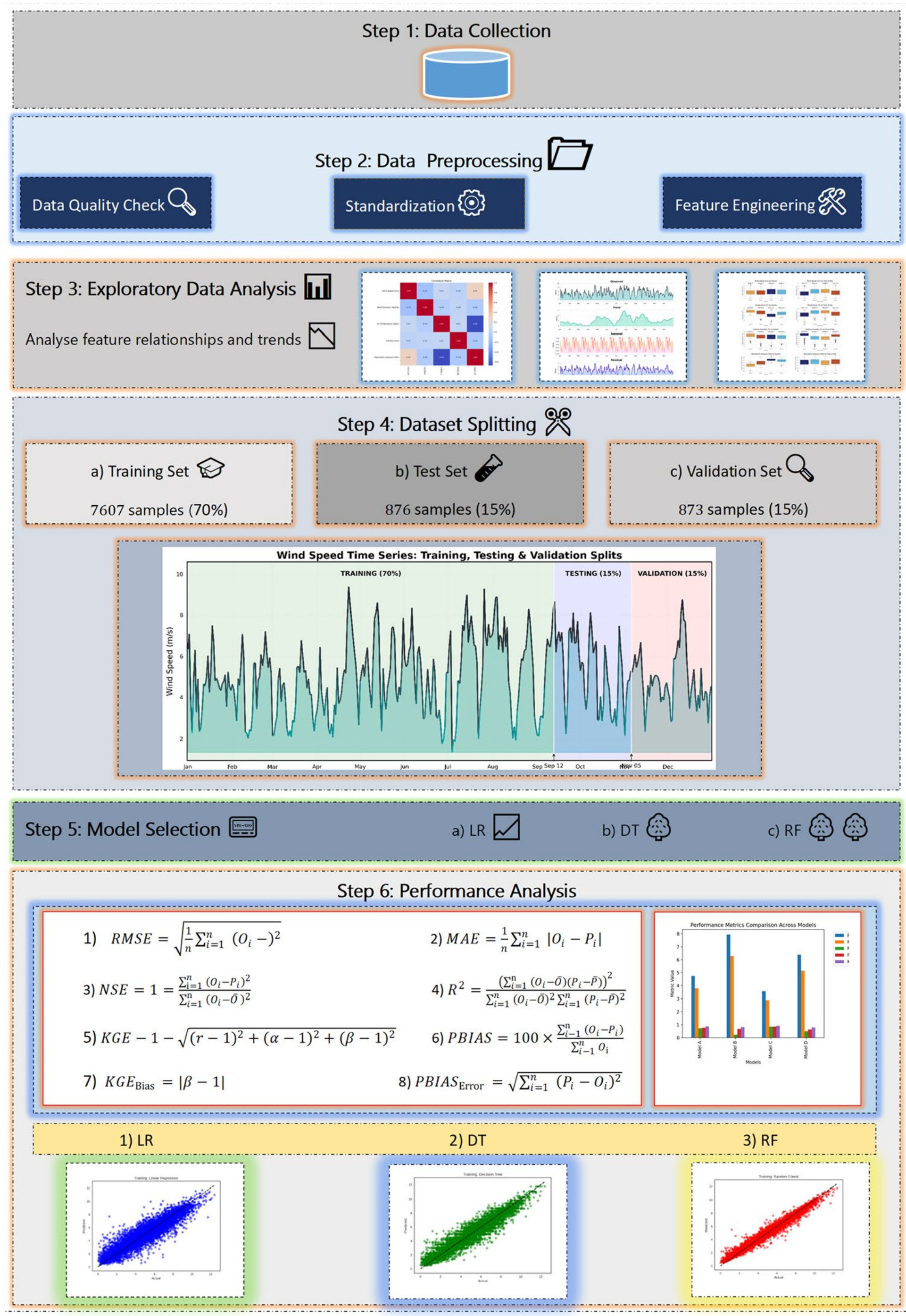


Fig. 6 Overview of the methodological workflow

$$NSE = 1 = \frac{\sum_{i=1}^{n}\left(O_i - P_i\right)^2}{\sum_{i=1}^{n}\left(O_i - \overline{O}\right)^2} \tag{3}$$

where $\overline{O}$ is the mean of the abcerved values.

4. Coefficient of Determination (R^2):

$$R^2 = \frac{\left(\sum_{i=1}^{n}\left(O_i - \overline{O}\right)\left(P_i - \overline{P}\right)\right)^2}{\sum_{i=1}^{n}\left(O_i - \overline{O}\right)^2 \sum_{i=1}^{n}\left(P_i - \overline{P}\right)^2} \tag{4}$$

where $\overline{P}$ is the mean of the prectiicted values.

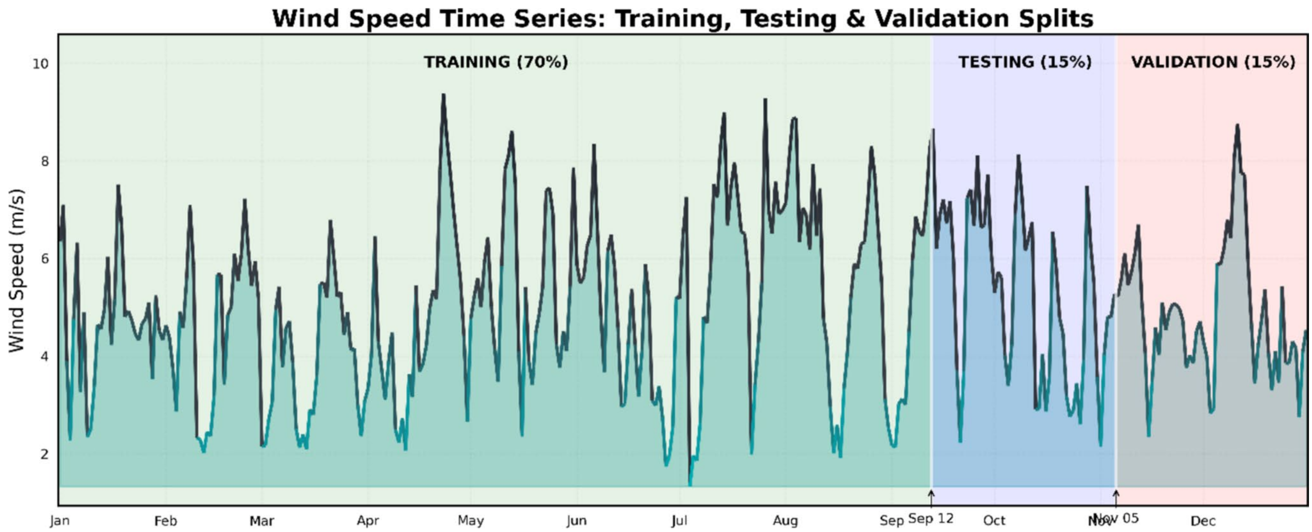


Fig. 7 Allocation of the time series dataset into training and testing phases

5. Kling-Gupta Efficiency (KGE):

$$KGE - 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (5)$$

where:

- r is the Pearson corrolation coefficient between observed and predicted values.
- $\alpha = \frac{\sigma i}{\sigma_0}$ is the variability ratio (σ denotes standard deviation).
- $\beta = \frac{\bar{P}}{\bar{O}}$ is the tias ratio.

6. Percentage Bias (PBIAS):

$$PBIAS = 100 \times \frac{\sum_{i-1}^{n} (O_i - P_i)}{\sum_{i-1}^{n} O_i} \quad (6)$$

7. KGE Bias Component (KGE_{Bias}):

$$KGE_{Bias} = |\beta - 1| \quad (7)$$

where β is as detinod in the KGE formula.

8. PBIAS Error Component:

$$PBIAS_{Error} = \sqrt{\sum_{i=1}^{n} (P_i - O_i)^2} \quad (8)$$

This is sometimes adjusted based on specific needs of the analysis. These formulas are widely used for evaluating the performance of predictive model, their ranges are shown in Fig. 8.

Comparative analysis The performance of the three models was compared according to the evaluation criteria and the model with the lowest MSE and the highest R^2 was determined as the best performing model.

Model selection The most suitable model for wind speed prediction at Abbot Point determined through comparative analysis was the model that showed superior performance compared to the others.

The improved wind speed predictions from the ML models can be integrated into existing coastal risk assessment models (Robichaux et al. 2018). This involves developing a workflow that incorporates the ML model outputs into the existing risk assessment process. The integration involve ML predictions as input to more complex hydrodynamic models that simulate storm surge and coastal flooding (Rezaie et al. 2021). The integration should also consider the uncertainty associated with the ML model predictions and incorporate this uncertainty into the risk assessment calculations.

The improved wind speed predictions can be used to develop a range of scenarios for different storm events and assess their potential impact on coastal communities. This involves combining the wind speed predictions with other relevant factors, such as storm surge models, sea level rise projections, and information on coastal infrastructure and population density. Monte Carlo simulations can be used to generate multiple realizations of storm events, each with different wind speed conditions, and the resulting impacts can be assessed using damage functions and vulnerability curves (Tebaldi et al. 2012). The scenarios should consider different return periods (e.g., 10-year, 50-year, 100-year) and different climate change scenarios (e.g., RCP4.5,

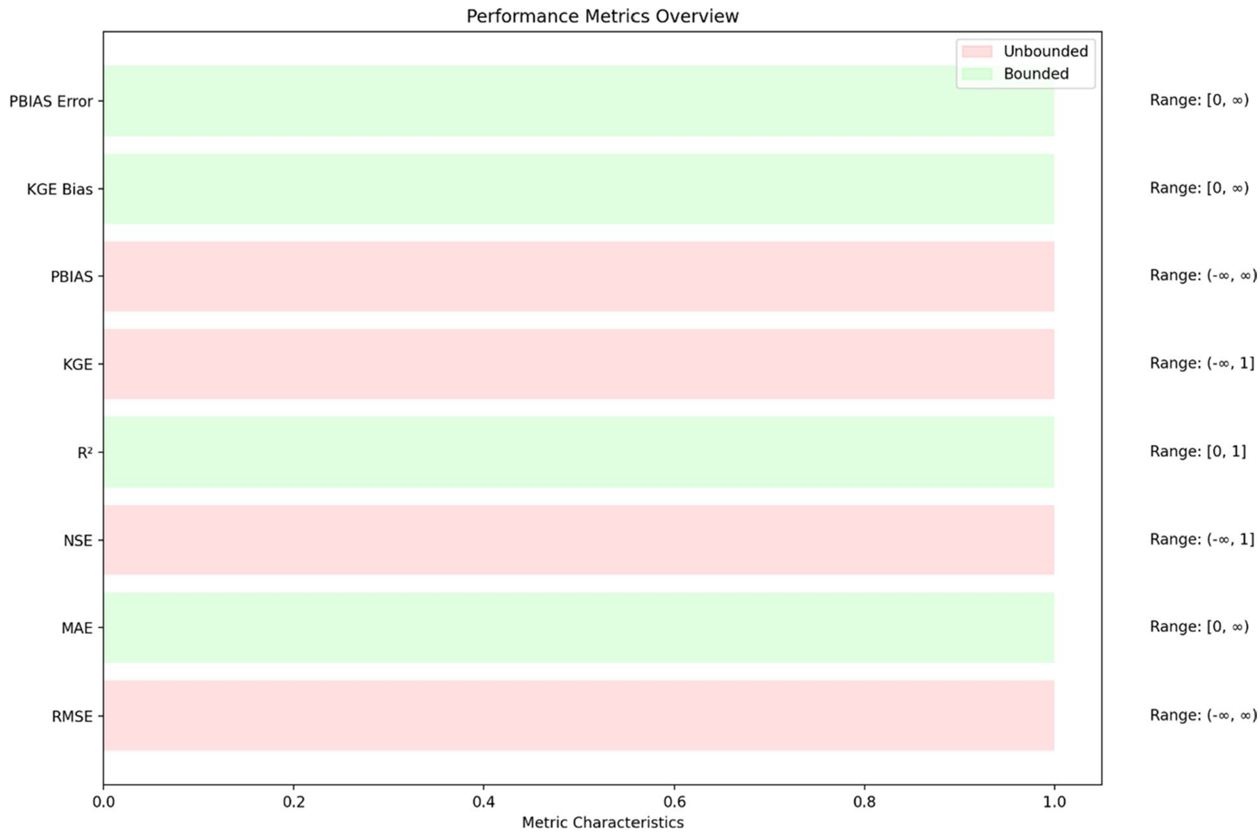


Fig. 8 Used metrics in this study and their ranges

RCP8.5). The results of the scenario analysis will provide a range of potential impacts, allowing decision-makers to understand the range of possible outcomes and to develop more robust adaptation and mitigation strategies.

The improved wind speed predictions can enhance the accuracy and timeliness of early warning systems for coastal hazards (Abul Ehsan et al. 2019). This involves integrating the ML model outputs into existing early warning systems, providing more accurate and reliable forecasts of wind speed and associated hazards (Ian et al. 2023). The early warning system should provide timely alerts to coastal communities, allowing sufficient time for evacuation or other protective measures. The system should also incorporate uncertainty information from the ML model predictions to communicate the range of possible outcomes to stakeholders. The design of the early warning system should consider the specific needs and vulnerabilities of the coastal communities it serves.

In this study, the target and feature variables are depicted in a series of steps accompanied by figures and statistical analysis. The target variable, wind speed, is predicted using a ML model that includes various meteorological and environmental features. The feature statistics table (Fig. 9) summarizes the basic descriptive statistics for the variables used in wind speed estimation.

Results

The primary objective of this study was to identify the most reliable ML model for predicting wind speed in coastal regions, with a specific focus on Abbot Point, Queensland, Australia, as accurate wind speed predictions are critical for effective coastal management strategies. Dominant wind directions, which are closely linked to wind speed dynamics, play a crucial role in shaping long-term coastal morphology and erosion patterns by driving sediment transport and dune formation (Haque et al. 2024; Smyth et al. 2023). By integrating reliable wind speed predictions with an understanding of wind direction trends, this study provides valuable insights for site-specific interventions. These insights enable the strategic alignment of coastal defenses with dominant wind-driven processes, optimizing the placement of protective structures and informing development regulations to mitigate erosion and flooding risks.

Feature Statistics

	Name	Distribution	Mean	Mode	Median	Dispersion	Min.	Max.	Missing
N	Wind Direction (degTN)		132.19	139	132	0.54	0	360	0 (0 %)
N	Wind Speed (m/s)		5.001	4.1	4.9	0.462	0.0	12.4	0 (0 %)
N	Air Temperature (degC)		24.171	24.6	24.5	0.134	11.5	32.5	0 (0 %)
N	Relative Humidity (%)		68.009	75.9	69.9	0.180	17.5	92.2	0 (0 %)
N	Barometric Pressure (hPa)		1011.16	1014	1012	0	994	1023	0 (0 %)

Fig. 9 Feature Statistics Table for Wind Speed Prediction (blue color-coding is target variable)

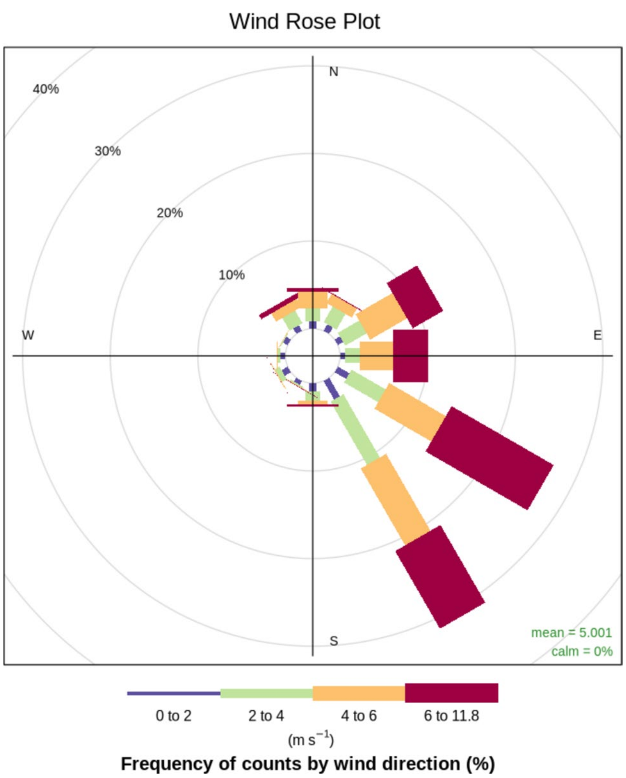


Fig. 10 Wind rose plot

In coastal risk management, understanding wind speed and direction is critical to assessing storm waves, wave dynamics, and erosion potential (Osinowo & Popoola 2024). The wind rose plot in Fig. 10 visually represents how wind speeds are distributed in different directions based on data from the wind direction dataset.

The radial bars show the frequency of occurrence for different wind speed ranges (0–2 m/s, 2–4 m/s, 4–6 m/s, and 6–11.8 m/s) as a percentage of the total observations. Each section of the bar represents a specific wind speed range where the dominant wind direction and higher speed ranges are clearly defined. The mean wind speed is calculated as 5.00 m/s, and the data shows no periods of calm conditions (0% calm). The wind rose plot provides a spatial overview of wind speed distributions and directions (Figs. 11 and 12).

Implications for Coastal Risk Management:

o Sustained high wind speeds from specific directions can increase coastal risks such as erosion, wave action and infrastructure damage.
Interestingly, some extreme wind events appear in unusual directions, which could be due to localized weather changes or data variations. The next analysis, shown in Figure 13 and Table 3, further investigates wind speed trends over time, marking periods when wind speeds exceeded 10 m/s to identify high-risk conditions.
Figure 11 and Figure 12 present a comprehensive correlation analysis of meteorological variables at Abbot Point, Queensland. The analysis incorporates Australian seasonal patterns (Summer: Dec-Feb; Autumn: Mar-May; Winter: Jun-Aug; Spring: Sep-Nov) to identify key relationships between wind speed and other meteorological parameters that influence coastal risk. The overall wind rose (Figure 11) (Panel A) and seasonal variations (Panels D-G) provide essential insights for coastal management strategies, particularly for identifying high-risk periods characterized by strong directional winds that contribute to coastal erosion and maritime hazards. These visualizations support the study's conclusion that Random Forest modeling, combined with detailed meteorological pattern analysis, offers superior predictive capability for coastal wind speed risk management at Abbot

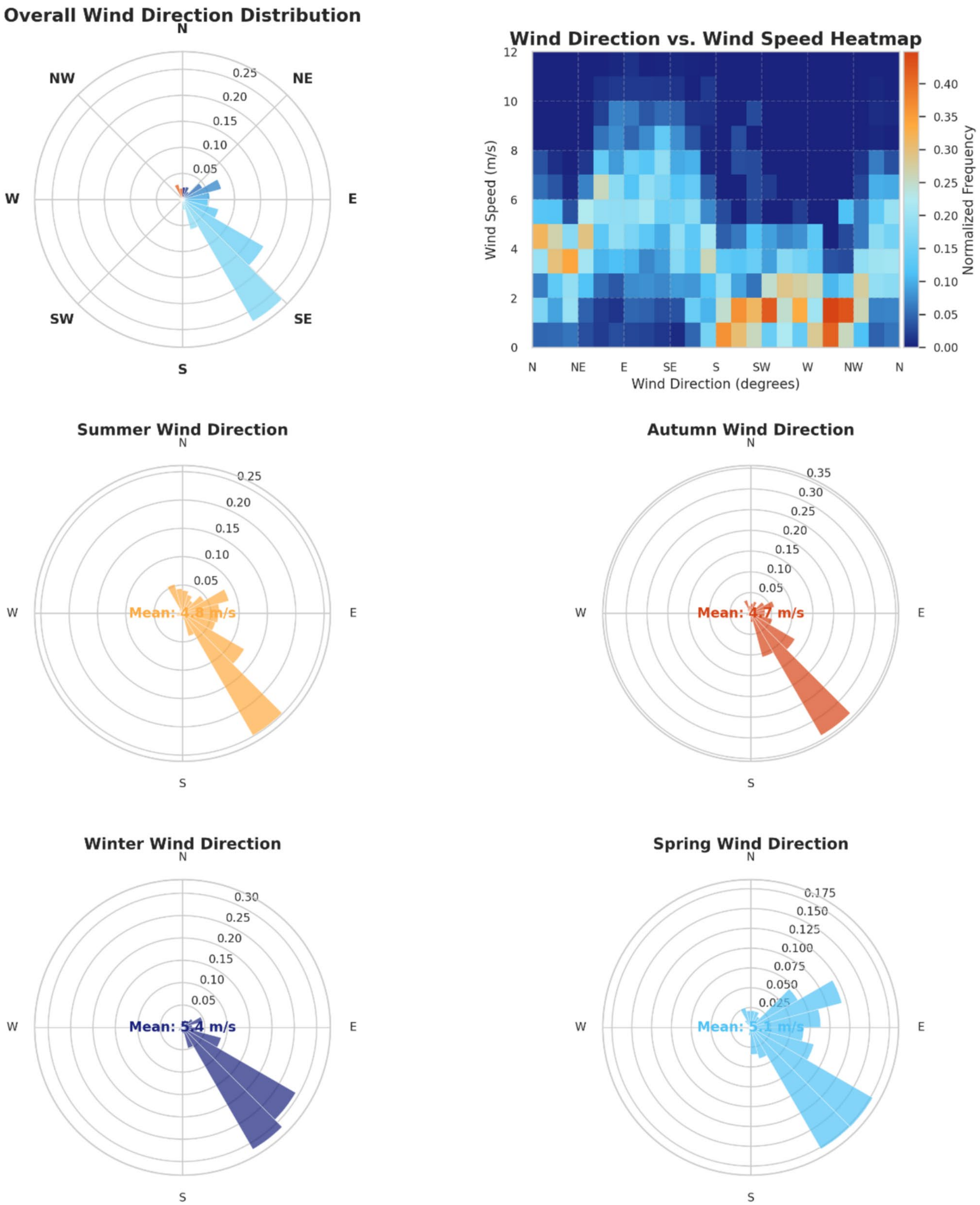


Fig. 11 Overall wind direction and seasonal variations alongside its speed

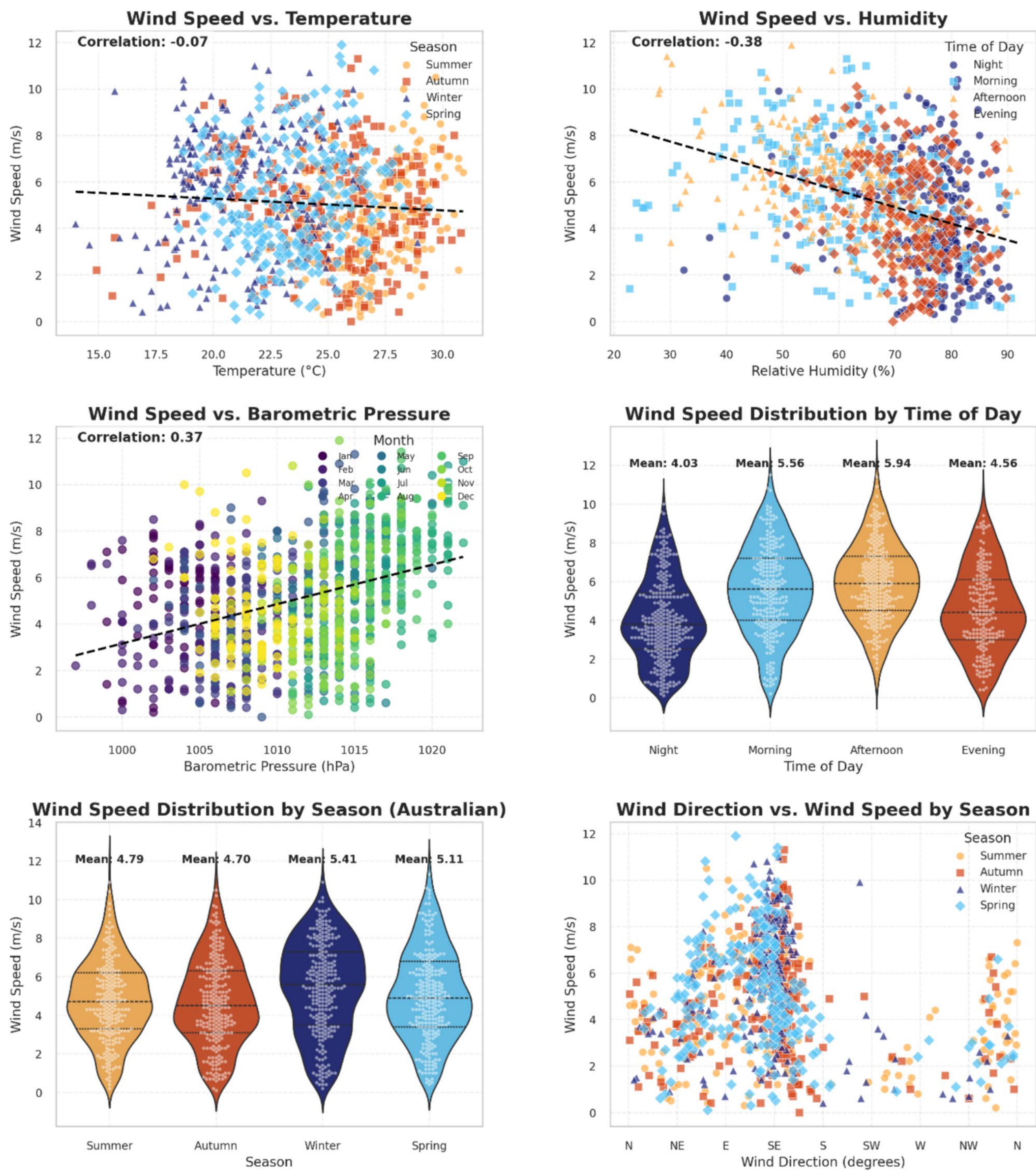


Fig. 12 Summary statistics of variables

Point, enabling more targeted interventions and resilient coastal planning. Wind direction analysis with seasonal breakdowns, confirming the prevalence of southeasterly and south-southwesterly winds identified as significant predictors in the Random Forest model (R2 = 0.844).

Panel A-C (Figure 12) :Scatter plots revealing the relationships between wind speed and temperature, humidity, and barometric pressure. The positive cor-

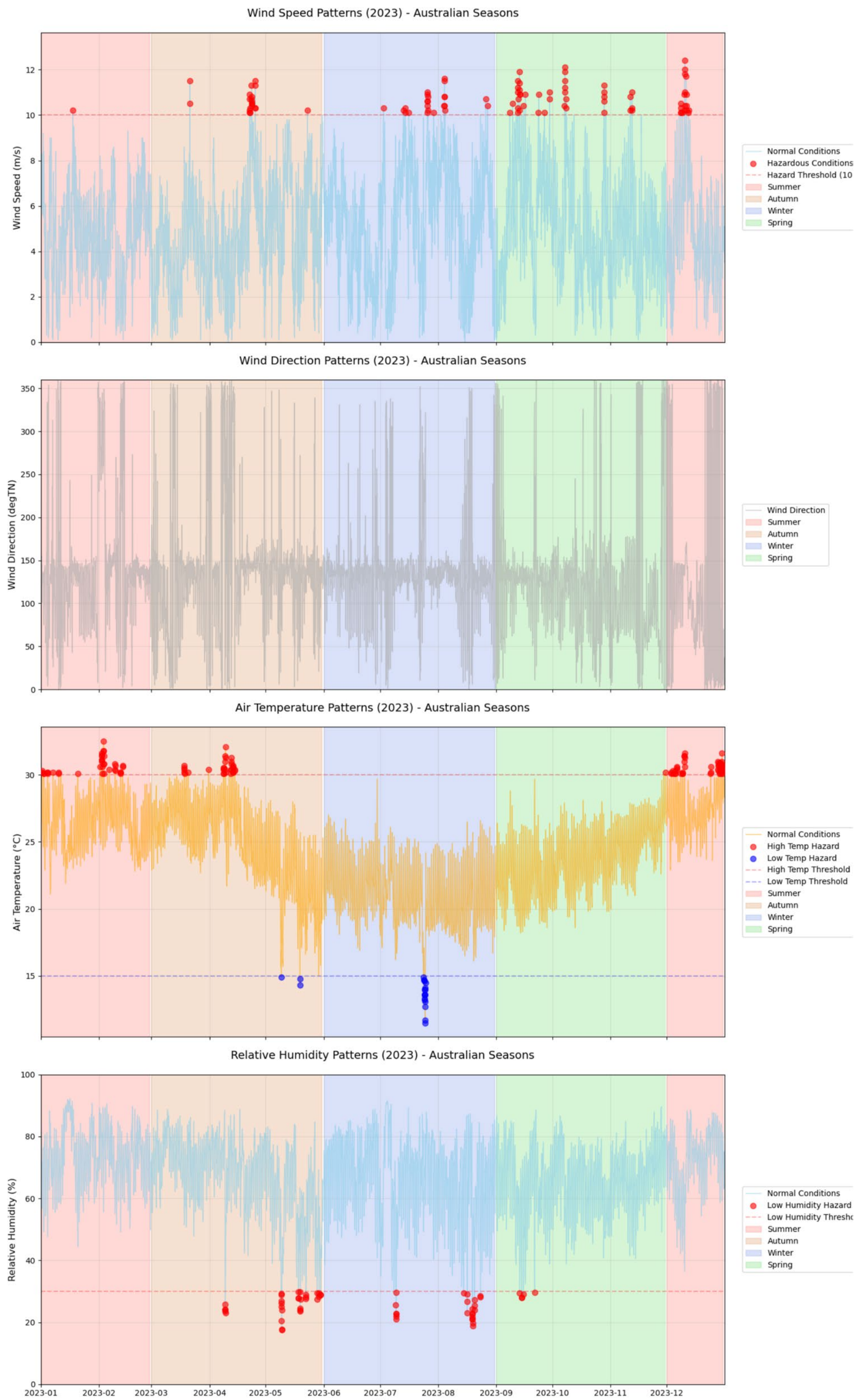


Fig. 13 Visualization of seasonal patterns in wind speed, wind direction, air temperature, and relative humidity in 2023

relation between wind speed and barometric pressure (Panel C, r = 0.37) confirms the critical role of pressure systems in driving extreme wind events, as identified in the Random Forest model. This relationship is particularly evident during cyclonic conditions when low pressure (<1000 hPa) coincides with

Table 3 Seasonal hazard statistics by category

Seasons				
Season	Autumn	Spring	Summer	Winter
Wind Speed				
Number of points	20	39	19	23
Mean (m/s)	10.64	10.8	10.71	10.54
Max (m/s)	11.5	12.1	12.4	11.6
Temperature				
Number of points	34	1	87	17
Mean (°C)	29.16	30.2	30.59	13.64
Min (°C)	14.3	30.2	30.1	11.5
Max (°C)	32.1	30.2	32.5	14.9
Humidity				
Number of points	31	5	N/A	23
Mean (%)	26.07	28.82	N/A	24.24
Min (%)	17.5	27.9	N/A	18.9
Pressure				
Number of points	147	17	745	117
Mean (hPa)	1003.82	1009.82	1001.72	1021.59
Min (hPa)	1001	1003	994	1021
Max (hPa)	1021	1021	1004	1023

elevated wind speeds (>10 m/s). Panel D-E: Violin plots illustrating wind speed distributions across different times of day and Australian seasons. The seasonal distribution (Panel E) shows higher mean wind speeds during Spring (5.42 m/s) and Winter (5.18 m/s), which aligns with the Random Forest model's identification of seasonal risk patterns. Panel F: Wind direction versus wind speed by season, demonstrating the dominance of southeasterly winds (100°−200°) as noted in the study's findings. This directional pattern is crucial for coastal erosion risk assessment and offshore wind farm optimization

Table 3 summarizes the number of observations and key statistics for wind speed, temperature, humidity, and pressure during each season in study area. It provides a clear seasonal overview to identify trends and high-risk conditions. Complementing this, Fig. 13 presents time-series plots for wind speed, direction, air temperature, and humidity throughout 2023 (outliers included in this stage of prediction), segmented by seasons. High-risk periods, indicated by red dots representing conditions exceeding hazardous thresholds, are highlighted, providing critical insights for assessing coastal vulnerabilities.

On average, the wind speed is 5.00 m/s, with no calm periods recorded (0% calm). The extreme wind threshold is set above 8.5 m/s, and these high-risk events appear consistently in the scatter plots. The presence of strong winds from certain directions is significant for coastal management since they can contribute to erosion, stronger waves, and potential damage to infrastructure. These findings support previous research showing that dominant wind patterns influence coastal hazards.

Patterns, insights, and machine learning implications.

Wind Speed Patterns:

- Hazardous conditions (> 10 m/s) occur most frequently in Spring.
- Summer shows the highest peak wind speeds, emphasizing increased coastal hazard risks during this season.
- These wind patterns are ideal for machine learning (ML) models:

  i. RF can capture non-linear relationships and seasonal patterns (Ho et al. 2023; Tang et al. 2022).
  ii.LR can identify basic trends in wind speed (Obisesan 2024).

Wind Direction:

- Seasonal shifts in wind direction are evident, showcasing variability across seasons.
- DT can segment these patterns into discrete decision rules.
- RF can effectively handle the cyclical nature of wind direction data.

Temperature & Humidity Correlations:

- Low humidity events often coincide with temperature extremes, presenting valuable insights for feature engineering in predictive models.
- LR can capture the linear relationships between these variables for wind prediction models.

ML Implementation:

- Utilize previous 24-h data as input to predict the next 6-h wind conditions.
- RF is well-suited for this task due to its ability to:

  i. Capture seasonal patterns effectively (Ho et al. 2023).
  ii. Handle non-linear relationships among variables.
  iii. Manage multiple input variables efficiently.

A complementary evaluation using machine learning models—LR, DT, and RF—further aids in identifying

Table 4 Performance metrics of ML models across training, validation, and test datasets

Model	Dataset	RMSE	MAE	NSE	R^2	KGE	PBIAS (%)	KGE Bias	PBIAS Error
Linear Regression	Training	0.92	0.96	0.83	0.83	0.714	7.12×10^{11}	0.874	0.0
Linear Regression	Validation	0.95	0.97	0.82	0.817	0.718	2.61×10^{-1}	0.88	−0.9975
Linear Regression	Test	0.86	0.92	0.83	0.832	0.693	3.90×10^{12}	0.885	−0.6643
Decision Tree	Training	0.575	0.76	0.89	0.893	0.574	4.52×10^{1}	0.923	−0.0
Decision Tree	Validation	1.220	1.10	0.76	0.765	0.813	2.92×10^{-1}	0.878	−1.34
Decision Tree	Test	1.133	1.06	0.77	0.779	0.787	5.09×10^{12}	0.877	−0.5498
Random Forest	Training	0.183	0.43	0.96	0.966	0.303	3.45×10^{11}	0.942	0.0558
Random Forest	Validation	0.875	0.94	0.83	0.831	0.690	2.60×10^{-1}	0.894	−0.7297
Random Forest	Test	0.803	0.90	0.84	0.843	0.661	4.09×10^{12}	0.894	−0.4411

high-risk periods. Random Forest consistently outperforms the other models, achieving the lowest error values (RMSE and MAE) and the highest reliability scores (NSE, R^2, and KGE) (see Table 4). Its ability to handle complex relationships in data makes it particularly effective for predicting storm risk and environmental hazards, as demonstrated by its performance across Training, Validation, and Test datasets. In contrast, LR and DT show moderate accuracy but struggle with consistency and bias.

Integrating these findings provides a robust framework for coastal risk management. The combined insights from meteorological patterns and machine learning analysis enable precise identification of high-risk periods, supporting timely and informed interventions to mitigate storm impacts.

RF stands out as the best-performing model overall. It achieves the lowest MSE (Train: 0.1834, Test: 0.8031) and the highest R^2 values (Train: 0.966, Test: 0.844), reflecting its strong ability to capture complex relationships in the data. RF also demonstrates a balanced trade-off between bias and variance, as shown by its minimal spatial bias (Test: −0.0220). This makes it the most reliable and accurate model for predictions across datasets.

LR performs consistently across datasets, with moderate R^2 values around 0.83 in the Test set. While it serves as a solid baseline model, its spatial bias (Test: −0.033) and challenges with extreme values highlight its limitations compared to Random Forest.

DT exhibits some overfitting, with a high R^2 value in Training (0.894) but a noticeable drop in the Test set (0.779). It also has the highest MSE on the Test set (1.134) and produces step-like predictions, which reduce its overall accuracy. These results suggest that Decision Tree struggles to generalize well to unseen data.

Figure 14 compares how three machine learning techniques—Random Forest, Decision Tree, and Linear Regression—performed on training, test, and validation data. Here is a summary of the key takeaways:

Random Forest: This model performed the best, with its predictions closely matching the actual values across all datasets. It accurately captured both low and high wind speeds with minimal errors.

Decision Tree: Its predictions showed a"step-like"pattern, especially in the validation and test data, due to its nature of making discrete predictions. While it handled moderate wind speeds reasonably well, it struggled with extreme values, resulting in higher errors.

Linear Regression: This model showed a more even spread of points but had noticeable errors at higher wind speeds. While it performed consistently, it was not as accurate as the Random Forest model.

In summary, the Random Forest model stood out as the most reliable for predicting wind speeds, as it generalized well across all datasets.

The error distributions of each model were analyzed using histograms for training, validation, and test datasets, revealing how well the models predicted wind speeds and the nature of their errors (Fig. 15) alongside MSE values across range of wind speed (see Fig. 16).

Random forest: Errors were nearly symmetric and centered around zero, especially in the training data.

Residuals (errors) were tightly spread, within about ± 2 m/s, making it the most accurate.

Slight overprediction was observed in the validation and test datasets, but the errors were still minimal.

Decision tree: Errors were spread out and showed a"multimodal"pattern, reflecting the discrete nature of its predictions.

Residuals had a broader range, around ± 3 m/s, compared to Random Forest.

More noticeable overprediction occurred in certain wind speed ranges for validation and test data.

Linear regression: Errors followed a roughly normal distribution but with longer tails, indicating larger deviations at the extremes.

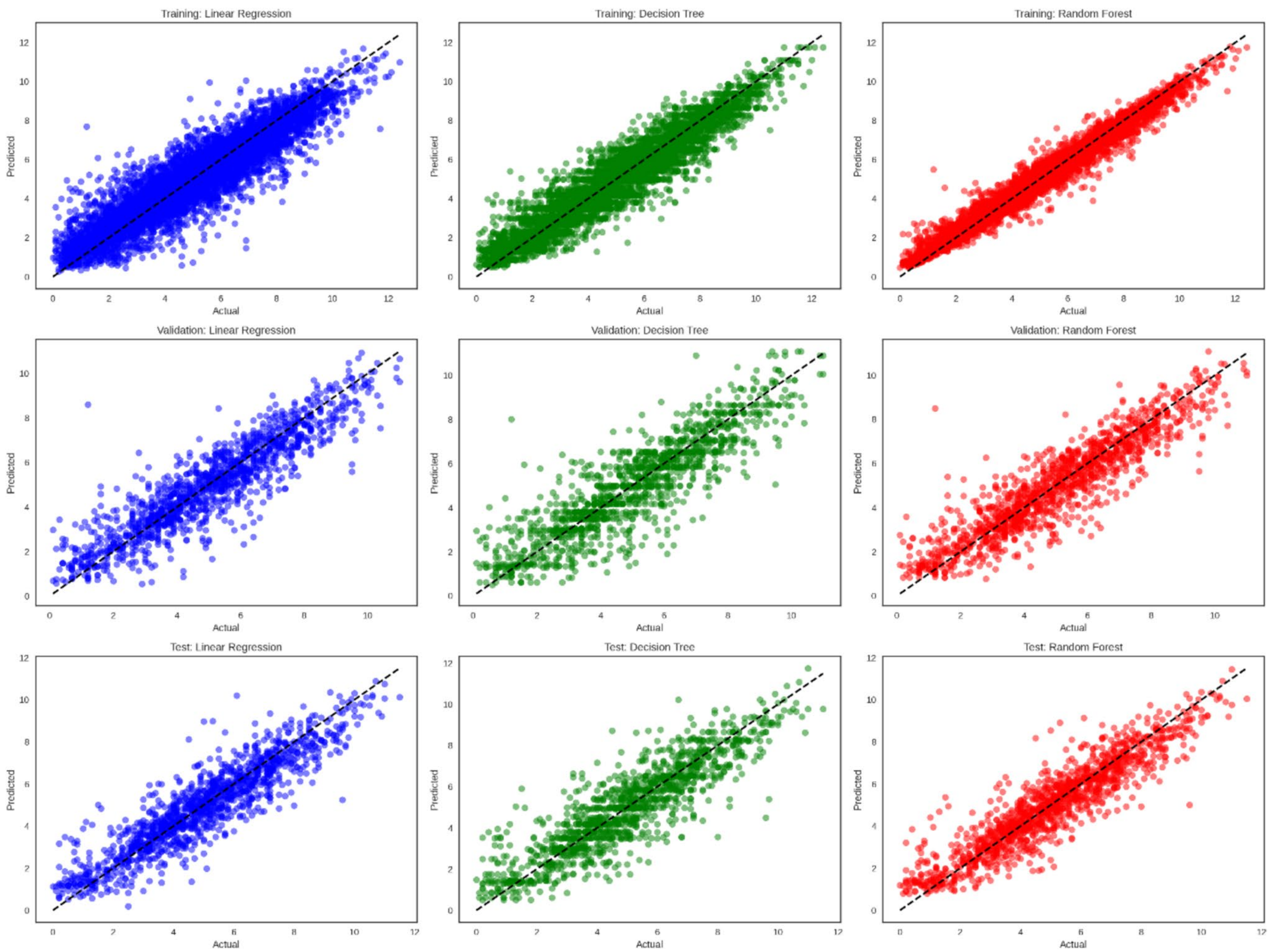


Fig. 14 Scatter plots of three ML across training, validation, and test samples

Residuals were more spread out (about ± 2.5 m/s) than Random Forest but less so than Decision Tree.

Predictions were stable but lacked the precision of Random Forest.

Statistical Implications.

The error distributions provide statistical evidence supporting the superior performance of the Random Forest model, with:

Lowest standard deviation of residuals (σ ≈ 0.43 for training, σ ≈ 0.94 for test).

Most symmetric distribution around zero, indicating minimal systematic bias.

Highest concentration of errors near zero, confirming better prediction accuracy.

These findings align with the performance metrics previously discussed, where the Random Forest model achieved the highest R^2 (0.844) and NSE (0.844) values on the test set, along with the lowest RMSE (0.896).

The highest concentration of errors near zero, as depicted in the figure, confirms the superior prediction accuracy of the Random Forest model. This aligns with the performance metrics discussed earlier, where the Random Forest model achieved the highest R^2 (0.844) and NSE (0.844) values on the test set, along with the lowest RMSE (0.896). The figure further illustrates the model's balanced prediction capabilities across wind speed ranges (see Fig. 17).

Linear regression (left) Overpredicts the middle range (4–6 m/s) and performs poorly at extreme ranges.

Decision tree (middle) Provides improved performance at low wind speeds and better balance across ranges.

Random forest (right) Demonstrates the best overall performance, achieving more balanced predictions and handling extreme cases effectively.

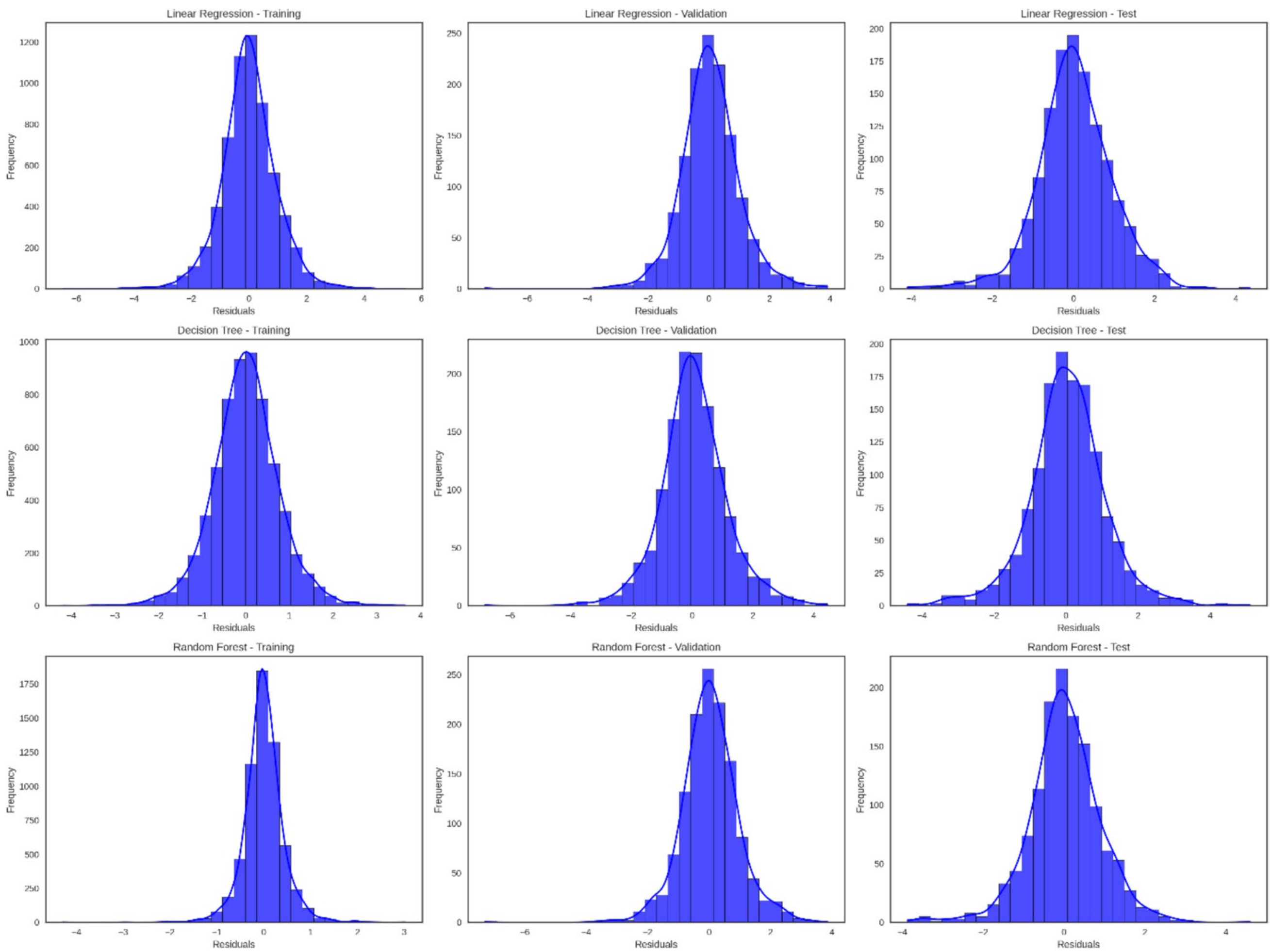


Fig. 15 The error distributions of ML models

The darker diagonal elements in the RF section indicate higher numbers of correct predictions, highlighting its ability to minimize misclassification ns compared to LR and DT models.

The SHAP (SHapley Additive exPlanations) analysis reveals the complex relationships between various meteorological and temporal features in wind speed prediction, complementing the performance metrics presented in Table 5. The analysis demonstrates several key findings:

a. **Feature Importance Hierarchy:**

- Wind Direction (degTN) emerged as the most influential predictor (SHAP value = 0.713), aligning with the Random Forest model's superior performance metrics
- Relative Humidity and Barometric Pressure showed substantial impact (SHAP values = 0.609 and 0.563 respectively)
- Temporal features (month, hour, day) exhibited lower but consistent influence (SHAP values < 0.239)

b. **Feature Impact Distribution:**

The SHAP summary plot illustrates how each feature contributes to the model predictions, with red indicating higher values and blue indicating lower values

Wind Direction shows a particularly strong positive correlation with wind speed predictions when values are high (red clusters)

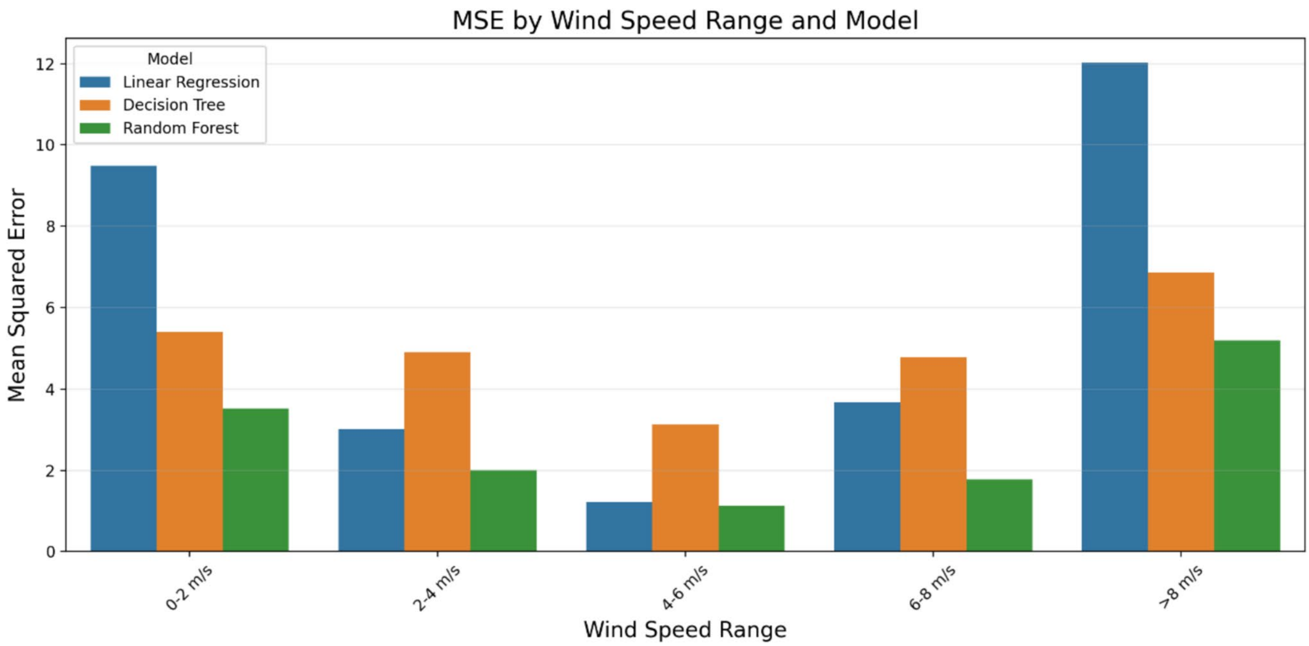


Fig. 16 Wind speed range on the performance of ML models

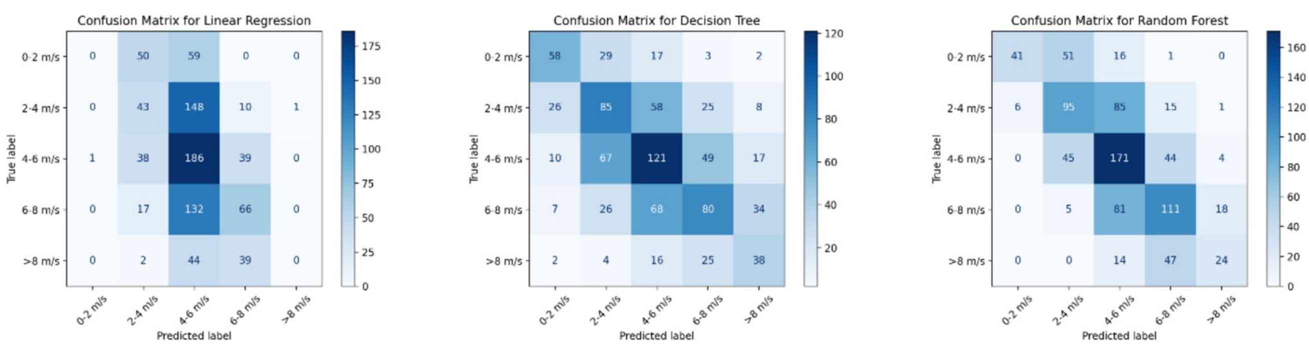


Fig. 17 Confusion matrix of models

Table 5 Performance Metrics of Linear Regression, Decision Tree, and Random Forest Models Across Training, Validation, and Test Datasets. The table summarizes key statistical measures, including mean, standard deviation, quartiles (Q1, Q3), and interquartile range (IQR), highlighting model variability and robustness across datasets

Model	Dataset	Mean	Std Dev	Q1	Median	Q3	IQR
Linear Regression	Training	0	0.9587	−0.5432	−0.0167	0.5418	1.085
Linear Regression	Validation	0.0502	0.973	−0.5003	0.0193	0.5796	1.08
Linear Regression	Test	0.0331	0.9272	−0.49	0.0181	0.5874	1.0774
Decision Tree	Training	0	0.7585	−0.4521	0	0.4419	0.894
Decision Tree	Validation	0.0674	1.1028	−0.5458	0.02	0.6699	1.2157
Decision Tree	Test	0.0274	1.0645	−0.5493	0.025	0.608	1.1574
Random Forest	Training	−0.0028	0.4283	−0.2218	−0.0118	0.216	0.4378
Random Forest	Validation	0.0367	0.9352	−0.5038	0.0139	0.5623	1.0661
Random Forest	Test	0.022	0.8959	−0.4569	0.0013	0.5222	0.9791

Relative Humidity demonstrates a more complex, non-linear relationship, supporting the Random Forest model's capability to capture non-linear patterns

c. **Model Interpretation Context:**

- The SHAP analysis provides insight into why the Random Forest model achieved superior performance metrics across training, validation, and test datasets as indicated in Table 5
- The clear feature importance hierarchy explains the model's robust performance and low error rates
- The distribution of SHAP values across features supports the model's ability to capture complex meteorological relationships

This analysis reinforces the Random Forest model's superior performance metrics presented in Table 4, providing a mechanistic explanation for its enhanced predictive capabilities through the quantification of feature contributions. The visualization demonstrates how the model leverages multiple meteorological parameters to achieve the reported performance metrics, particularly the improved accuracy and reduced error rates compared to Linear Regression and Decision Tree models.

The SHAP analysis particularly supports the Random Forest model's superior performance by revealing how it effectively utilizes the complex interactions between meteorological parameters, resulting in the improved statistical measures noted in Table 5, including better mean predictions and tighter standard deviations across all datasets.

Table 5 presents a comparative analysis of the performance metrics for the three models—Linear Regression,

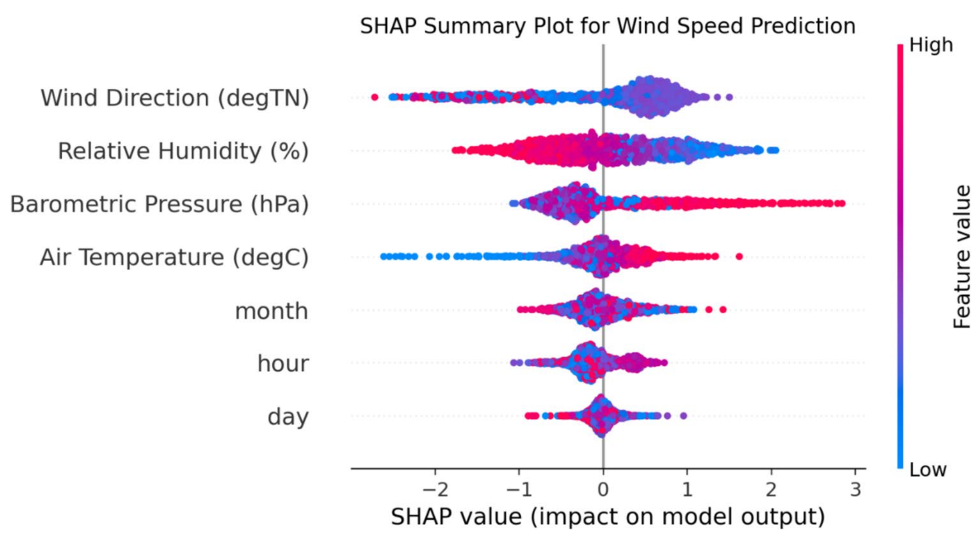


Fig. 18 SHAP summary plot for wind speed prediction model features

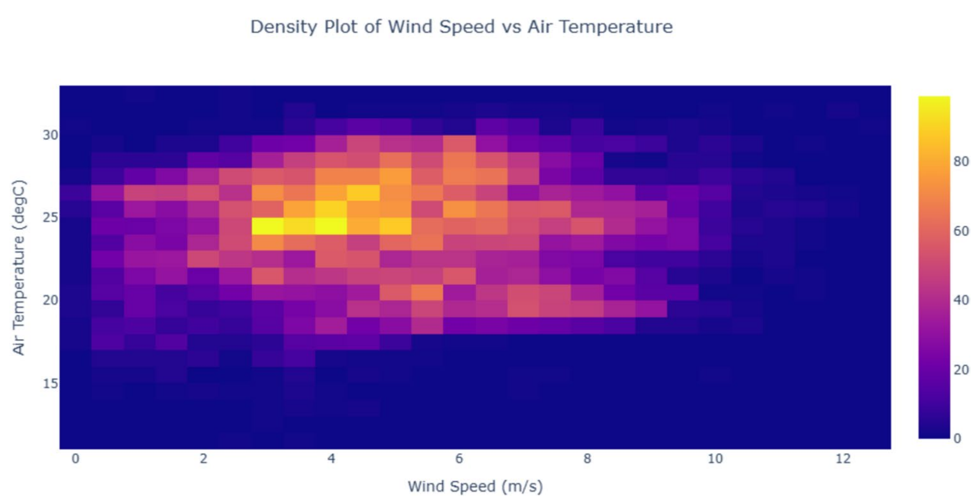


Fig. 20 Wind speed versus air temperature

Decision Tree, and Random Forest—across training, validation, and test datasets. The results indicate that Random Forest demonstrates superior robustness and consistency, as evidenced by its lower standard deviations and narrower IQR values, making it the most reliable model for this study. Conversely, the Decision Tree model exhibited the highest variability, while Linear Regression provided stable but less adaptable predictions. This comprehensive evaluation underscores the suitability of Random Forest for applications requiring reliable and generalizable predictions (Durap 2024a).

The SHAP summary plot (Fig. 18) visually supports these findings, emphasizing the Random Forest model's ability to leverage complex feature relationships for accurate wind speed predictions. This analysis underscores the importance of feature contributions in understanding model performance and variability.

The SHAP analysis and performance metrics provide a comprehensive understanding of the wind speed prediction models. The Random Forest model's superior performance is explained through feature importance, with Wind Direction, Relative Humidity, and Barometric Pressure being the key drivers.

The success of RF can be attributed to its ability to capture relationships between these variables, as opposed to simpler models such as LR. Figure 20 shows the relationship between wind speed and air temperature based on meteorological data from the Abbot Point (Bowen) coastal area (Figs. 19 and 20).

The intensity plot shows how frequently certain combinations of wind speed (in m/s) and air temperature (in °C) occur, where warmer colors indicate more frequent occurrences. Wind speeds of 2 to 6 m/s are most common at temperatures between 20 °C and 30 °C, and moderate winds are generally associated with warmer conditions. In contrast, higher wind speeds above 8 m/s occur quite rarely, regardless of temperature, indicating the infrequency of extreme wind events. This information is crucial for coastal risk management because wind speed and temperature significantly affect wave formation, storm surges and coastal erosion. By understanding these patterns, coastal managers can better predict wind-related hazards under specific temperature conditions, leading to more accurate risk assessments.

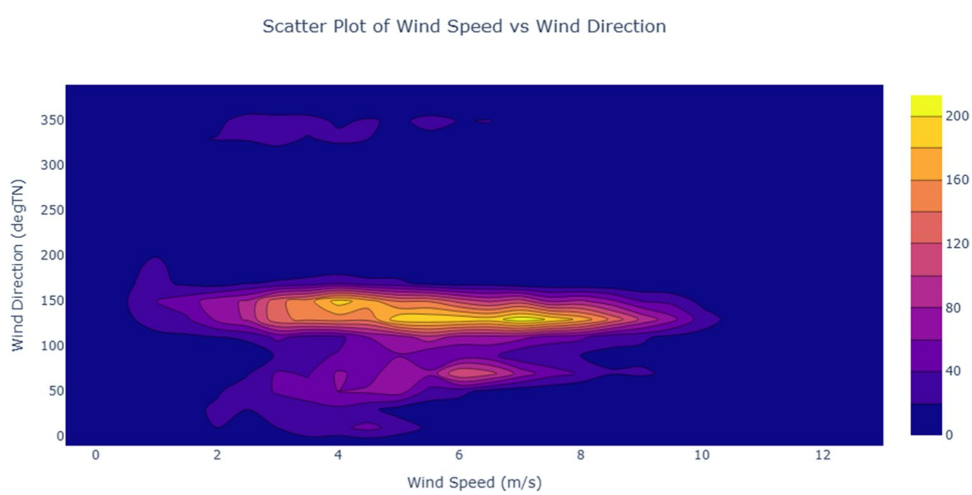
Finally, insights from this analysis can be used to develop ML models for wind speed prediction. This can help coastal managers identify high-risk periods and take proactive measures, such as strengthening coastal defenses or issuing early warnings for severe weather.

Analysis of meteorological observations from the coastal region reveals distinct patterns in wind speed and direction, as visualized in the density contour plot (Fig. 21).

The plot clearly shows the frequency distribution of wind events with respect to both speed and direction. Most wind events are concentrated in a specific direction range between 100° and 200°, which corresponds approximately to the southeast and south-southwest directions. This direction prevalence suggests that regional weather patterns and geographical features have a strong influence on local wind behavior. In this dominant direction range, wind speeds are primarily between 2 m/s and 6 m/s. This speed range reflects typical wind conditions in this coastal region, which are likely influenced by factors such as proximity to the ocean and prevailing atmospheric circulation patterns.

Higher wind speeds exceeding 8 m/s are less frequent, which tend to occur in a narrower direction band, indicating a possible association with specific weather systems or storm events. The limited directional spread of high wind speeds highlights the importance of accurate wind direction estimation for predicting extreme wind events and associated coastal hazards. Conversely, wind directions outside the range of 100°–200° show significantly lower frequencies. This observation further emphasizes the directional bias in wind behavior and reinforces the influence of regional and local factors on wind patterns. The observed directional

Fig. 21 Wind speed and direction distribution



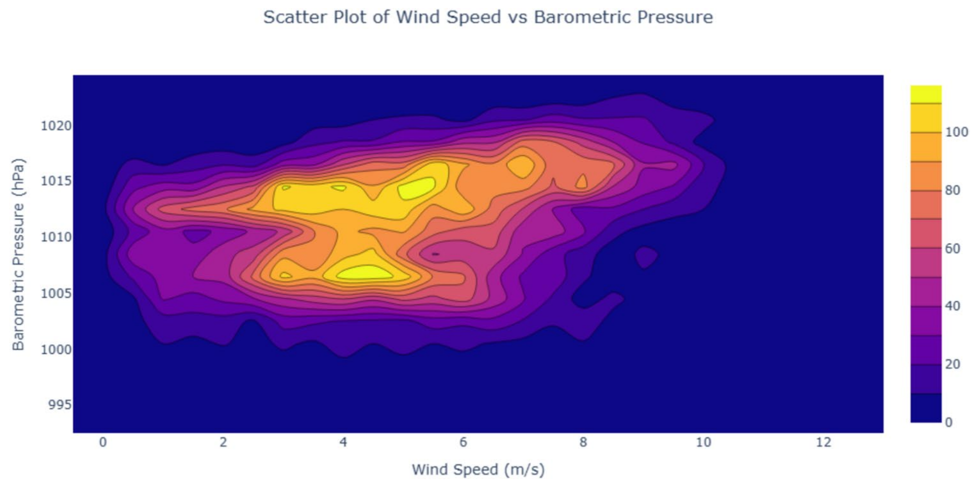


Fig. 22 Wind speed and barometric pressure distribution

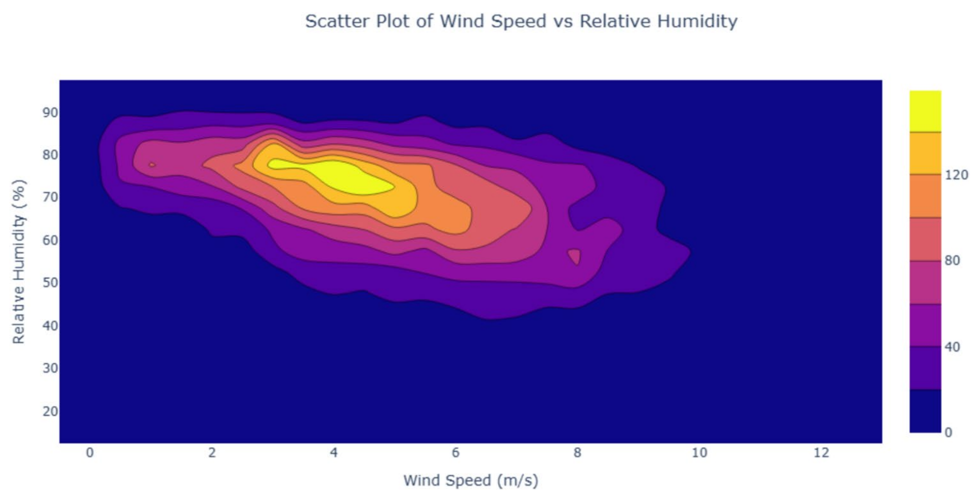


Fig. 23 Wind speed and relative humidity distribution

distribution is consistent with findings from other coastal studies, indicating that similar patterns are present in other coastal environments (Soares et al. 2014; Truccolo 2011). The recorded speed ranges are consistent with typical wind patterns in similar coastal environments (Bozzeda et al. 2023; Kostianaia & Kostianoy 2021). These findings highlight the importance of considering regional and local factors when assessing wind characteristics and their effects on coastal processes. The density contour plot shows the relationship between wind speed and barometric pressure, and thus provides valuable information about the prevailing wind patterns under different atmospheric pressure conditions.

Figure 22 reveals that moderate wind speeds from 2 m/s to 6 m/s are concentrated at barometric pressures between 1005 and 1015 hPa. These pressure ranges generally correspond to stable atmospheric conditions. In contrast, extreme wind speeds exceeding 8 m/s are less frequent and tend to occur at slightly lower barometric pressures. Further investigation of these pressure changes in association with extreme wind events may provide valuable information for predicting and mitigating coastal hazards. Analyzing the distribution of wind speeds in different pressure ranges may contribute to a more detailed understanding of wind dynamics in coastal areas.

The density contour plot in Fig. 23 shows the relationship between wind speed and relative humidity and provides further insight into the complex interaction of atmospheric factors affecting wind behavior.

The plot reveals that wind events are most common at relative humidity levels between 60 and 80%, especially at moderate wind speeds in the range from 2 m/s to 6 m/s. This observation suggests a correlation between moderate humidity levels and wind formation in coastal areas. Conversely, at lower humidity levels (less than 40%), a significant decrease in wind frequency is observed, independent of wind speed. This finding points to specific climatic conditions where winds are less common, probably due to increased atmospheric stability or suppressed convective activity. Understanding these humidity related wind patterns is important for accurate wind forecasts and comprehensive assessments of coastal vulnerability. Further research into the specific mechanisms linking humidity and wind behavior could improve our understanding of coastal wind dynamics. Improved wind forecasts can support marine operations, optimize renewable energy production from offshore wind farms, and inform coastal engineering projects.

Discussion

This study aimed to identify the most reliable machine learning (ML) model for predicting wind speeds in coastal regions, focusing on Abbot Point, Queensland, Australia. Accurate wind speed prediction is crucial for effective coastal management strategies, as wind dynamics significantly influence coastal morphology, erosion patterns, and the potential for hazards such as storm surges(Javaid et al. 2022; Lawal et al. 2021; Sun & Pan 2023). The results demonstrate the superior performance of the Random Forest Regressor (RFR) model compared to Linear Regression (LR) and Decision Tree Regressor (DTR) models.

The primary objective was achieved through a comparative analysis of three ML models: LR, DTR, and RFR. The analysis encompassed training, validation, and testing datasets, allowing for a robust evaluation of model performance and generalizability. RFR consistently outperformed both LR and DTR across all evaluation metrics. This superior performance was evident in lower root mean squared error (RMSE) and mean absolute error (MAE) values and higher Nash–Sutcliffe efficiency (NSE), R^2, and Kling-Gupta efficiency (KGE) scores (Shaloo et al. 2024). The RFR model exhibited the lowest mean squared error (MSE) of 0.58, significantly lower than the MSE values of 1.44 for LR and 1.13 for DTR. Furthermore, RFR achieved the highest R^2 value of 0.89, indicating a substantially stronger fit to the data compared to LR and DTR (Abdelsattar et al. 2025; Matrenin et al. 2024; Peng et al. 2024). Seasonal analysis revealed that hazardous wind conditions (> 10 m/s) were most frequent in spring, with summer exhibiting the highest peak wind speeds, highlighting the importance of seasonal considerations in coastal risk assessment (Baquero et al. 2022; Sri Preethaa et al. 2023; Yafouz et al. 2021).

Our findings align with a substantial body of literature demonstrating the effectiveness of ensemble learning methods, particularly Random Forest, in wind speed prediction (Abdelsattar et al. 2025; Hamid et al. 2018; Li et al. 2023; Shaloo et al. 2024). Several studies have shown that RF models can effectively capture complex, non-linear relationships and seasonal patterns in wind speed data (Matrenin et al. 2024). This is in contrast to simpler models like LR, which are limited by their linear assumptions and struggle to capture the inherent complexities of wind dynamics(Baquero et al. 2022). While DTR can handle non-linearity, its proneness to overfitting can hinder its generalization ability, as observed in our study (Yafouz et al. 2021). The superior performance of RFR in our study, evidenced by lower error rates and higher accuracy scores, underscores the advantages of ensemble methods in addressing the challenges of wind speed prediction in coastal environments.

Our study also contributes novel insights by integrating meteorological data analysis with ML model evaluation. The wind rose plot and seasonal analysis provided a comprehensive understanding of wind speed and direction distributions, informing the selection and interpretation of ML models (Sun & Pan 2023). The SHAP analysis further enhanced the interpretability of the RFR model by quantifying the contribution of various meteorological and temporal features to wind speed predictions (Iban & Aksu 2024). This approach moves beyond simply reporting performance metrics, providing valuable insights into the model's internal mechanisms and improving overall understanding of the underlying wind dynamics.

In this study, the outstanding success of RFR can be attributed largely to the inherent strengths of ensemble methods in addressing the challenges of wind speed prediction. Ensemble methods, such as RF, combine predictions from multiple base learners (in this case, decision trees) to produce a more robust and accurate overall prediction (Abul Ehsan et al. 2019). This approach reduces the risk of overfitting, a common problem with individual decision trees that can be overly sensitive to small fluctuations and noise in the training data (Luo et al. 2018). Overfitting occurs when a model learns the training data too well, including its noise and outliers, leading to poor generalization performance on unseen data. By aggregating predictions from multiple trees, each trained on a slightly different subset of the data, RFR effectively smooths out these fluctuations and produces a more generalized and reliable prediction (Luo et al. 2018). The robustness of ensemble methods is particularly valuable in the context of wind speed data, which is characterized by inherent variability and complex, nonlinear patterns (Lian & He 2022). Traditional linear models, such as LR, often struggle to capture these complexities, leading to lower prediction accuracy. While DTR can handle nonlinearity, its tendency to overfit may limit its effectiveness. On the other hand, RFR leverages the collective wisdom of multiple decision trees to overcome these complexities and provide more accurate and stable forecasts. Moreover, the ability of RFR to simultaneously handle multiple meteorological parameters contributes to its improved performance in capturing the multifaceted interactions of factors affecting wind speed.

The LR model in this study indicated a valuable basis for comparison and provided insights into the limitations of linear models in predicting wind speed (Pillai & Shihabudheen 2017). Its linear structure restricts its ability to capture the inherent nonlinear relationships between meteorological parameters and wind speed, leading to lower forecast accuracy compared to RFR and DTR (Oktari et al. 2020). DTR is capable of modeling nonlinear relationships, but it suffers from overfitting sensitivity, which may compromise its performance on new data (Muhammed Anees et al. 2024). This overfitting tendency explains the relatively higher MSE and lower R^2 value of DTR compared to RFR, highlighting the importance of model selection and the need to consider the trade-off between model complexity and generalization performance (Khatun et al. 2024). By effectively balancing these factors with the ensemble approach, RFR emerges as the superior model for wind speed prediction in this study. This finding is consistent with other studies showing the advantages of ensemble methods over single model approaches for wind speed prediction.

The improved accuracy of wind speed predictions provided by RFR is crucial for improving storm surge prediction, which is a critical component of coastal hazard early warning systems (Wu et al. 2022). More accurate wind speed data allows for more precise estimates of storm surge levels, enables timely warnings, and facilitates more effective evacuation procedures, potentially saving lives and minimizing property damage (Miškovic 2014). Furthermore, improved wind speed forecasts can be integrated into early warning systems for other coastal hazards, such as overtopping waves, providing a more comprehensive and reliable assessment of coastal risks and contributing to better preparedness of society(Bokde et al. 2019).

Accurate wind speed predictions are also important for ensuring maritime safety, especially in coastal areas where wind conditions can significantly impact navigation (Rychlik & Mao 2019). By providing more precise predictions of wind speed and direction, RFR can enable safer navigation by alerting maritime operators to potentially hazardous conditions. This improved forecasting capability can help prevent maritime accidents, reduce delays, and improve the overall efficiency of shipping operations, contributing to both economic and safety benefits for the maritime sector. By providing more reliable wind information, RFR can facilitate informed decision-making for route planning and ship operations, minimize risks associated with adverse weather conditions, and increase the safety of maritime activities.

Conclusion

This study aimed to identify the most reliable ML model for predicting coastal wind speeds at Abbot Point, Queensland, Australia, a crucial task for effective coastal risk management (e.g., Durap et al. 2023). Accurate wind speed prediction is essential because wind dynamics significantly influence coastal processes such as erosion, sediment transport, and storm surge formation. The study compared the performance of three ML models: LR, DTR, and RFR, using various evaluation metrics across training, validation, and test datasets.

The results unequivocally demonstrate the superior performance of the RFR model. RFR consistently achieved the lowest RMSE and the highest R^2 values across all datasets, indicating significantly higher accuracy and better fit to the data compared to LR and DTR. The RFR model's ability to capture complex, non-linear relationships and seasonal patterns in the data proved highly advantageous.

Specifically, RFR achieved an RMSE of 0.183 and an R^2 of 0.966 during training, significantly outperforming DTR (RMSE $=0.575$, $R^2= 0.893$) and LR (RMSE $=0.92$, $R^2=$ 0.83). Similarly, in the test dataset, RFR maintained superior accuracy with an RMSE of 0.803 and R^2 of 0.843, compared to DTR (RMSE $=1.133$, $R^2= 0.779$) and LR (RMSE $=0.86$, $R^2= 0.832$). These results confirm the robustness of RFR in making reliable predictions, which is critical for coastal risk management applications.

The superior performance of RFR underscores the advantages of ensemble methods in handling the inherent variability and complexity of environmental data. By leveraging multiple decision trees and reducing overfitting, RFR provides more stable and accurate predictions, making it a valuable tool for forecasting and mitigating risks associated with dynamic coastal conditions.

The identified seasonal patterns – hazardous conditions (> 10 m/s) most frequent in spring, with summer showing the highest peak wind speeds – are critical for informing seasonal risk assessments and targeted mitigation strategies. These insights, coupled with the RFR model's predictive power, allow for the development of more effective early warning systems and proactive coastal management plans.

The findings of this study have profound implications for improving coastal wind speed risk predictions and early warning systems. The superior performance of the RFR model offers several key advantages over traditional methods:

- **Improved Accuracy:** The lower error rates and higher accuracy scores of the RFR model compared to LR and DTR lead to more precise wind speed predictions. This improved accuracy is particularly valuable in predicting extreme wind events, which pose the greatest risk to coastal communities and infrastructure.
- **Enhanced Robustness:** RFR's robustness stems from its ability to handle non-linear relationships and seasonal patterns in wind speed data. This is in contrast to LR, which is limited by its linear assumptions, and DTR, which is prone to overfitting. RFR's consistent performance across training, validation, and test datasets demonstrates its ability to generalize well to unseen data, making it a reliable tool for long-term coastal risk assessment.
- **Data-Driven Insights:** The study's integration of meteorological data analysis with ML model evaluation provides a comprehensive understanding of wind speed and direction distributions. Seasonal analysis reveals the most hazardous periods, while the wind rose plot provides a visual representation of wind speed distribution

across different directions. This data-driven approach enhances the interpretability of the model and facilitates more informed decision-making (Durap & Doğan 2015).
- **SHAP Analysis for Interpretability:** The SHAP (SHapley Additive exPlanations) analysis provides insights into the relative importance of different meteorological and temporal features in predicting wind speed. This enhances the model's transparency and allows for a better understanding of the factors driving wind variability, which is crucial for refining model inputs and improving predictive capabilities.

These advantages translate into several practical applications:

- **Coastal Engineering:** The RFR model can provide more accurate estimates of wind loads on coastal structures, enabling engineers to design more resilient infrastructure that can withstand extreme wind events.
- **Disaster Preparedness:** Improved wind speed predictions enable more effective early warning systems, allowing coastal communities to prepare for hazardous conditions and evacuate if necessary. This can significantly reduce the impact of extreme wind events on human lives and property.
- **Policy-Making:** The model's outputs can inform coastal zone management plans and development regulations, ensuring that coastal development is aligned with the risks posed by wind-driven hazards. This promotes sustainable coastal development and reduces the vulnerability of coastal communities.

The study's findings contribute significantly to improving coastal wind speed risk management. The RFR model, with its superior accuracy, robustness, and interpretability, provides a valuable tool for enhancing early warning systems, informing coastal engineering practices, and supporting evidence-based coastal management policies. The integration of meteorological data analysis and SHAP analysis further enhances the model's value by providing a more comprehensive understanding of the factors influencing coastal wind dynamics.

This work adds to a growing body of research demonstrating the potential of ML techniques for wind speed prediction in coastal environments.

The RFR model developed in this study can be readily deployed in real-world scenarios for coastal wind speed prediction. Government agencies responsible for coastal management can utilize this model to improve their early warning systems, enhancing the timely issuance of alerts for hazardous wind conditions. Coastal communities can use the model's predictions to prepare for potential risks, such as evacuations or infrastructure protection measures. The model's insights can also inform the design and placement of coastal defenses, optimizing resource allocation and minimizing damage from extreme wind events. For example, the model can aid in determining optimal locations for windbreaks or seawalls, improving their effectiveness in mitigating erosion and flooding. Recommendations for stakeholders include investing in high-quality, long-term meteorological data collection, integrating the RFR model into existing coastal risk management frameworks, and conducting regular model validation and updates. Furthermore, promoting public awareness and education about coastal wind risks and the model's capabilities can enhance community resilience and preparedness.

Limitations and future research directions

As the renewable energy landscape evolves, future wind speed prediction methodologies are expected to shift from traditional onshore approaches to techniques specific to offshore environments. Offshore wind turbines, especially floating facilities, operate under unique meteorological conditions that require the development of innovative prediction models that can accurately predict wind behavior in coastal and offshore regions. In order to improve these prediction methods in the future, it will be important to focus on both accuracy and computational efficiency in a balanced manner in the light of methodological approaches mentioned in this paper.

Further research could focus on optimizing data preprocessing techniques and feature selection methods specifically tailored to wind speed data. For example, temperature gradients integration between land and water surfaces generates local wind patterns, such as sea breezes, could improve the ability of models to capture coastal wind dynamics.

Finally, future research should focus on developing user-friendly tools and interfaces that integrate these advanced models and provide accessible wind speed estimates for coastal stakeholders. Empowering stakeholders with accurate and practical tools enable informed decision making for effective risk management and planning.

Author contribution Ahmet Durap: Conceptualization, Methodology, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability Data will be made upon reasonable request.

Declarations The author declares no competing financial or non-financial interests, nor any other interests that could be perceived to influence the results or discussion presented in this article.

Ethical approval This study complies with ethical standards. No human or animal subjects were involved in this research. All data and materials used in this study are obtained and presented in accordance with ethical guidelines.

Conflict of interest The author declares that there is no conflict of interest regarding the publication of this paper.

Competing interests The author declares no competing interests in relation to this study.

References

Abdelsattar MA, Ismeil M, Menoufi K, AbdelMoety A, Emad-Eldeen A (2025) Evaluating Machine Learning and Deep Learning models for predicting Wind Turbine power output from environmental factors. PLOS ONE 20(1):e0317619. https://doi.org/10.1371/journal.pone.0317619

Abouhalima M, das Neves L, Taveira-Pinto F, Rosa-Santos P (2024) Machine Learning in Coastal Engineering: Applications, Challenges, and Perspectives. J Marine Sci Eng 12(4):638. https://doi.org/10.3390/jmse12040638

Abul Ehsan B, Begum F, Ilham SJ, Khan RS (2019) Advanced wind speed prediction using convective weather variables through machine learning application. Appl Comput Geosci 1:100002. https://doi.org/10.1016/j.acags.2019.100002

Aisyah S, Wijayaningrum VN, Hani'ah M, Noprianto & Al Hadid Firdaus V (2024) Efficient Multi-Class Wind Speed Prediction Through Genetic Algorithm and Multi-Layer Perceptron Integration. *2024 Int Conf Elect Inform Technol (IEIT)*, 140–145. https://doi.org/10.1109/IEIT64341.2024.10763047

Alotaibi O, Pardede E, Tomy S (2023) Cleaning Big Data Streams: A Systematic Literature Review. Technologies 11(4):101. https://doi.org/10.3390/technologies11040101

Alves T (2003) Cenozoic tectono-sedimentary evolution of the western Iberian margin. Mar Geol 195(1–4):75–108. https://doi.org/10.1016/S0025-3227(02)00683-7

Balal A, Pakzad Jafarabadi Y, Demir A, Igene M, Giesselmann M, Bayne S (2023) Forecasting Solar Power Generation Utilizing Machine Learning Models in Lubbock. Emerg Sci J 7(4):1052–1062. https://doi.org/10.28991/ESJ-2023-07-04-02

Baquero L, Torio H, Leask P (2022) Machine Learning Algorithms for Vertical Wind Speed Data Extrapolation: Comparison and Performance Using Mesoscale and Measured Site Data. Energies 15(15):5518. https://doi.org/10.3390/en15155518

Baseer KK, Nas SBA, Dharani S, Sravani S, Yashwanth P, & Jyothirmai P (2023) Medical Diagnosis of Human Heart Diseases with and without Hyperparameter tuning through Machine Learning. *2023 7th Int Conf Comput Methodol Commun (ICCMC)*, 1–8. https://doi.org/10.1109/ICCMC56507.2023.10084156

Bayram B, Seker DZ, Acar U, Yuksel Y, Arı Gürer A, Çetin İ (2013) An Integrated Approach to Temporal Monitoring of the Shoreline and Basin of Terkos Lake. J Coastal Res 29(6):1427. https://doi.org/10.2112/JCOASTRES-D-12-00084.1

Bokde N, Feijóo A, Villanueva D, Kulat K (2019) A Review on Hybrid Empirical Mode Decomposition Models for Wind Speed and Wind Power Prediction. Energies 12(2):254. https://doi.org/10.3390/en12020254

Bozzeda F, Ortega L, Costa LL, Fanini L, Barboza CAM, McLachlan A, & Defeo O (2023) Global patterns in sandy beach erosion: unraveling the roles of anthropogenic, climatic and morphodynamic factors. *Front Marine Sci*, *10*. https://doi.org/10.3389/fmars.2023.1270490

Coco G, Burnet TK, Werner BT, & Elgar S (2004) The role of tides in beach cusp development. *J Geophys Res: Oceans*, *109*(C4). https://doi.org/10.1029/2003JC002154

da Silva KDM, Lorenzato de Oliveira JF (2023) A Hybrid System Based on Dynamic Selection for Wind Speed Forecasting. 2023 Workshop Commun Net Power Syst (WCNPS) 2023:1–7. https://doi.org/10.1109/WCNPS60622.2023.10344604

da Quaresma VS, Catabriga G, Bourguignon SN, Godinho E, Bastos AC (2015) Modern sedimentary processes along the Doce river adjacent continental shelf. Braz J Geol 45(4):635–644. https://doi.org/10.1590/2317-488920150030274

Demirtop A, Sevli O (2024) Wind speed prediction using LSTM and ARIMA time series analysis models: A case study of Gelibolu. Turk J Eng 8(3):524–536. https://doi.org/10.31127/tuje.1431629

Durap A (2023) A comparative analysis of machine learning algorithms for predicting wave runup. Anthropocene Coasts 6(1):17. https://doi.org/10.1007/s44218-023-00033-7

Durap A (2024a) Data-driven models for significant wave height forecasting: comparative analysis of machine learning techniques. Results Eng 24:103573. https://doi.org/10.1016/j.rineng.2024.103573

Durap A (2024b) Mapping coastal resilience: a Gis-based bayesian network approach to coastal hazard identification for Queensland's dynamic shorelines. Abstr Anthropocene Coasts 7(1). https://doi.org/10.1007/s44218-024-00060-y

Durap A (2025) Machine learning-based wind speed prediction using random forest: a cross-validated analysis for renewable energy applications. Turk J Eng 9(3):508–518. https://doi.org/10.31127/tuje.1624354

Durap A, Doğan Y (2015) İnşaat Mühendisliğinde Bilişim Kavramı ve Veri Madenciliği Algoritmaları ile Bir Uzman Sisteminin Oluşturulması. XVII, Akademik Bilişim Konferansı Bildirileri

Durap A, Balas CE (2024) Towards sustainable coastal management: a hybrid model for vulnerability and risk assessment. J Coast Conserv 28(4). https://doi.org/10.1007/s11852-024-01065-y

Durap A, Balas CE, Çokgör Ş, Balas EA (2023) An Integrated Bayesian Risk Model for Coastal Flow Slides Using 3-D Hydrodynamic Transport and Monte Carlo Simulation. J Marine Sci Eng 11(5):943. https://doi.org/10.3390/jmse11050943

Fernandes PAC, Fonseca RA, Pacheco FAL, Sanches Fernandes LF (2023) Water quality predictions through linear regression - A brute force algorithm approach. MethodsX 10:102153. https://doi.org/10.1016/j.mex.2023.102153

Gai RL, Zhang H, & Thanh DNH (2023) A Big Data Cleaning Method for Drinking-Water Streaming Data. *Braz Arch Biol Technol*, *66*. https://doi.org/10.1590/1678-4324-2023220365

Geng B, Zhou W, Lu L, Cao Q, Hong M, & Wen D (2022) Risk assessment of offshore oil spill in the northern South China Sea: A stochastic scenario modeling approach. *Front Marine Sci*, *9*. https://doi.org/10.3389/fmars.2022.986379

Hallgren C, Aird JA, Ivanell S, Körnich H, Vakkari V, Barthelmie RJ, Pryor SC, Sahlée E (2024) Machine learning methods to improve spatial predictions of coastal wind speed profiles and low-level jets using single-level ERA5 data. Wind Energy Sci 9(4):821–840. https://doi.org/10.5194/wes-9-821-2024

Hamid AIA, Din AHM, Hwang C, Khalid NF, Tugi A, Omar KM (2018) Contemporary sea level rise rates around Malaysia: Altimeter data optimization for assessing coastal impact. J Asian Earth Sci 166:247–259. https://doi.org/10.1016/j.jseaes.2018.07.034

Haque MdM, Haque M, Ghosh MK (2024) The coastal dynamics of the Central Ganges–Brahmaputra–Meghna Delta Coast, Bangladesh: implications for coastal development and sustainability. J Coast Conserv 28(1):29. https://doi.org/10.1007/s11852-024-01032-7

Ho C-Y, Cheng K-S, Ang C-H (2023) Utilizing the Random Forest Method for Short-Term Wind Speed Forecasting in the Coastal Area of Central Taiwan. Energies 16(3):1374. https://doi.org/10.3390/en16031374

Hsu B-M (2020) Comparison of Supervised Classification Models on Textual Data. Mathematics 8(5):851. https://doi.org/10.3390/math8050851

Huang R, Chen S, Chen W, Wu R, Wang Z, Hu P, Wu L, Wang L, Huangfu J (2024) Impact of the Winter Regional Hadley Circulation over Western Pacific on the Frequency of Following Summer Tropical Cyclone Landfalling in China. J Clim 37(13):3521–3541. https://doi.org/10.1175/JCLI-D-23-0610.1

Ian V-K, Tang S-K, Pau G (2023) Comparative Analysis of BALSSA and Conventional NWP Methods: A Case Study in Extreme Storm Surge Prediction in Macao. Atmosphere 14(11):1597. https://doi.org/10.3390/atmos14111597

Iban MC, Aksu O (2024) SHAP-Driven Explainable Artificial Intelligence Framework for Wildfire Susceptibility Mapping Using MODIS Active Fire Pixels: An In-Depth Interpretation of Contributing Factors in Izmir. Türkiye Remote Sens 16(15):2842. https://doi.org/10.3390/rs16152842

Javaid A, Javaid U, Sajid M, Rashid M, Uddin E, Ayaz Y, Waqas A (2022) Forecasting Hydrogen Production from Wind Energy in a Suburban Environment Using Machine Learning. Energies 15(23):8901. https://doi.org/10.3390/en15238901

Khatun M, Sattar GS, Karmokar PK, Karim MR (2024) Optimizing Wind Speed Predictions in Bangladesh: Unveiling the Power of Machine Learning Approaches. Int J Stat Sci 24(2):137–154. https://doi.org/10.3329/ijss.v24i2.78000

Kostianaia E, Kostianoy A (2021) Regional Climate Change Impact on Coastal Tourism: A Case Study for the Black Sea Coast of Russia. Hydrology 8(3):133. https://doi.org/10.3390/hydrology8030133

Krishnaveni S, Singh J, Verma K, Pachaury A, Kashyap G, Bhatia A (2021) A Machine Learning Approach for Wind Speed Forecasting. Int Conf Adv Comput Innov Technol Eng (ICACITE) 2021:507–512. https://doi.org/10.1109/ICACITE51222.2021.9404563

Kumar R, Prakash M, & Shakila B (2023) A Comparative Analysis of Time Series and Machine Learning Models for Wind Speed Prediction. *2023 IEEE 3rd Mysore Sub Section Int Conf (MysuruCon)*, 1–6. https://doi.org/10.1109/MysuruCon59703.2023.10396852

Lawal A, Rehman S, Alhems LM, Alam MdM (2021) Wind Speed Prediction Using Hybrid 1D CNN and BLSTM Network. IEEE Access 9:156672–156679. https://doi.org/10.1109/ACCESS.2021.3129883

Lazarevska, E. (2016). Wind speed prediction with extreme learning machine. *2016 IEEE 8th Int Conf IntellSyst (IS)*, 154–159. https://doi.org/10.1109/IS.2016.7737415

Leung N-C, Chow C-K, Lau D-S, Lam C-C, Chan P-W (2024) WRF-ROMS-SWAN Coupled Model Simulation Study: Effect of Atmosphere-Ocean Coupling on Sea Level Predictions Under Tropical Cyclone and Northeast Monsoon Conditions in Hong Kong. Atmosphere 15(10):1242. https://doi.org/10.3390/atmos15101242

Li X, Li K, Shen S, Tian Y (2023) Exploring Time Series Models for Wind Speed Forecasting: A Comparative Analysis. Energies 16(23):7785. https://doi.org/10.3390/en16237785

Li L, Escribano-Macias J, Zhang M, Fu S, Huang M, Yang X, Zhao T, Feng Y, Elhajj M, Majumdar A, Angeloudis P, Ochieng W (2024) Temporally Correlated Deep Learning-Based Horizontal Wind-Speed Prediction. Sensors 24(19):6254. https://doi.org/10.3390/s24196254

Lian L, He K (2022) Ultra-short-term wind speed prediction based on variational mode decomposition and optimized extreme learning machine. Wind Eng 46(2):556–571. https://doi.org/10.1177/0309524X211038547

Luo X, Sun J, Wang L, Wang W, Zhao W, Wu J, Wang J-H, Zhang Z (2018) Short-Term Wind Speed Forecasting via Stacked Extreme Learning Machine With Generalized Correntropy. IEEE Trans Industr Inf 14(11):4963–4971. https://doi.org/10.1109/TII.2018.2854549

Magdalena M, Zusmawati Z, Algusri J, Fernos J, Rozi F (2023) The influence of price perception, promotion, and online customer review on consumer buying interest in lazada e commerce in padang utara district. Jurnal Apresiasi Ekonomi 11(1):41–48. https://doi.org/10.31846/jae.v11i1.551

Matrenin PV, Harlashkin DA, Mazunina MV, Khalyasmaa AI (2024) Investigation of the Features Influencing the Accuracy of Wind Turbine Power Calculation at Short-Term Intervals. Appl Syst Innov 7(6):105. https://doi.org/10.3390/asi7060105

Miškovic V (2014) Machine Learning of Hybrid Classification Models for Decision Support. Proceed 1st Int Sci Conf - Sinteza 2014:318–323

Mohammad AS, Al-Ani JA, & Zaghar D (2021) Machine Learning for Prediction of Wind Power and Resources in Hot Climate Region. *2021 Int Conf Comput Commun Applic Technol (I3CAT)*, 16–22. https://doi.org/10.1109/I3CAT53310.2021.9629401

Muhammed Anees VV, Abdul Nazar KP, Maniyath Shijin (2024) Performance Comparison of Machine Learning Algorithms for Wind Energy Forecasting in the Coastal Region of Kerala. Int Res J Adv Eng Hub (IRJAEH) 2(12):2734–2739. https://doi.org/10.47392/IRJAEH.2024.0378

Musaaed E, Hewah N, Alasaadi A (2024) A Comparative Study on Machine Learning Models to Classify Diseases Based on Patient Behaviour and Habits. J Artif Intell Syst 6(1):34–58. https://doi.org/10.33969/AIS.2024060103

Ningsih FR, Djamal EC, & Najmurrakhman A (2019) Wind Speed Forecasting Using Recurrent Neural Networks and Long Short Term Memory. *2019 6th Int Conf Instrument, Control, Autom (ICA)*, 137–141. https://doi.org/10.1109/ICA.2019.8916717

Obisesan OE (2024) Machine Learning Models for Prediction of Meteorological Variables for Weather Forecasting. Int J Environ Clim Change 14(1):234–252. https://doi.org/10.9734/ijecc/2024/v14i13829

Oktari RS, Syamsidik, Idroes R, Sofyan H, Munadi K (2020) City Resilience towards Coastal Hazards: An Integrated Bottom-Up and Top-Down Assessment. Water 12(10):2823. https://doi.org/10.3390/w12102823

Olcay K, Gíray Tunca S, Aríf Özgür M (2024) Forecasting and Performance Analysis of Energy Production in Solar Power Plants Using Long Short-Term Memory (LSTM) and Random Forest Models. IEEE Access 12:103299–103312. https://doi.org/10.1109/ACCESS.2024.3432574

Osinowo AA, Popoola SO (2024) Statistical evaluations of sea's state along the Nigerian coast. J Coast Conserv 28(1):25. https://doi.org/10.1007/s11852-023-01014-1

Pal SK, Dutta D (2024) Transfer Learning in Weather Prediction: Why, How, and What Should. J Comput Cogn Eng 3(4):324–347. https://doi.org/10.47852/bonviewJCCE42022817

Pathak AR, Pandey M, Rautaray S (2018) Application of Deep Learning for Object Detection. Procedia Comput Sci 132:1706–1717. https://doi.org/10.1016/j.procs.2018.05.144

Pati, N., Gourisaria, M. K., Das, H., & Banik, D. (2023). Wind Speed Prediction using Machine Learning Techniques. *2023 11th Int Conf Emerg Trends Eng Technol - Signal Inform Processing (ICETET - SIP)*, 1–6. https://doi.org/10.1109/ICETET-SIP58143.2023.10151597

Peng L, Lv S, Wang L (2024) Explainable machine learning techniques based on attention gate recurrent unit and local interpretable model-agnostic explanations for multivariate wind speed forecasting. J Forecast 43(6):2064–2087. https://doi.org/10.1002/for.3097

Philippine Journal of Crop Science, Vallez AGC, De Guzman LEP, Villanueva-Peyraube JD, Ancog RC, Aquino AL (2024) Sustainability Assessment of the Local Knowledge Systems and Climate Change Adaptation of the Abaknon in Capul Island, Northern Samar, Philippines. The Philippine Journal of Crop Science 49(3):1–15. https://doi.org/10.63568/vol49iss3pp1-15

Phillips B, Masters F, Raubenheimer B, Olabarrieta M, Morrison E, Fernández-Cabán P, Ferraro C, Davis J, Rawlinson T, Rodgers M (2023) An Experimental Platform to Study Wind, Hydrodynamic, and Biochemical Conditions in the Littoral Zone During Extreme Coastal Storms. Oceanography. https://doi.org/10.5670/oceanog.2023.s1.19

Phunpeng V, Saensuriwong K, Kerdphol T, Uangpairoj P (2023) The Flexural Strength Prediction of Carbon Fiber/Epoxy Composite Using Artificial Neural Network Approach. Materials 16(15):5301. https://doi.org/10.3390/ma16155301

Pillai GN, Shihabudheen KV (2017) Wind speed forecasting using empirical mode decomposition and regularized ELANFIS. IEEE Symposium Series on Computational Intelligence (SSCI) 2017:1–7. https://doi.org/10.1109/SSCI.2017.8285324

Pinson P (2013) Wind Energy: Forecasting Challenges for Its Operational Management. *Statist Sci*, *28*(4). https://doi.org/10.1214/13-STS445

Qin J, Ding R, Zhou L, Liu H, Long S, Tao L (2024) Cross-Seasonal Effect on Tropical Pacific Precipitation: Implication of South Pacific Quadrupole Simulation in CMIP6. J Clim 37(23):6127–6142. https://doi.org/10.1175/JCLI-D-23-0660.1

Rezaie AM, Ferreira CM, Walls M, & Chu Z (2021) Quantifying the Impacts of Storm Surge, Sea Level Rise, and Potential Reduction and Changes in Wetlands in Coastal Areas of the Chesapeake Bay Region. *Nat Hazard Rev, 22*(4). https://doi.org/10.1061/(ASCE)NH.1527-6996.0000505

Robichaux L, Owen S, Kench P (2018) Fuzzy cognitive mapping: proposed assessment framework for coastal risk management project processes. Coastal Eng Proceed 36:29. https://doi.org/10.9753/icce.v36.risk.29

Rychlik I, Mao W (2019) Spatio-temporal modelling of wind speed variations and extremes in the Caribbean and the Gulf of Mexico. Theoret Appl Climatol 135(3–4):921–944. https://doi.org/10.1007/s00704-018-2411-y

Salah S, Alsamamra HR, Shoqeir JH (2022) Exploring Wind Speed for Energy Considerations in Eastern Jerusalem-Palestine Using Machine-Learning Algorithms. Energies 15(7):2602. https://doi.org/10.3390/en15072602

Shaloo, Kumar B, Bisht H, Rajput J, Mishra AK, TM KK, Brahmanand PS (2024) Reference evapotranspiration prediction using machine learning models: An empirical study from minimal climate data. Agronom J 116(3):956–972. https://doi.org/10.1002/agj2.21504

Shi H, Dong Z, Xiao N, & Huang Q (2021) Wind Speed Distributions Used in Wind Energy Assessment: A Review. *Front Energy Res*, *9*. https://doi.org/10.3389/fenrg.2021.769920

Smyth TAG, Rooney P, Yates KL (2023) Dune slope, not wind speed, best predicts bare sand in vegetated coastal dunes. J Coast Conserv 27(4):27. https://doi.org/10.1007/s11852-023-00957-9

Soares PMM, Cardoso RM, Semedo Á, Chinita MJ, Ranjha R (2014) Climatology of the Iberia coastal low-level wind jet: weather research forecasting model high-resolution results. Tellus a: Dynamic Meteorol Oceanograph 66(1):22377. https://doi.org/10.3402/tellusa.v66.22377

Sri Preethaa KR, Muthuramalingam A, Natarajan Y, Wadhwa G, Ali AAY (2023) A Comprehensive Review on Machine Learning Techniques for Forecasting Wind Flow Pattern. Sustainability 15(17):12914. https://doi.org/10.3390/su151712914

Sun K, & Pan J (2023) Model of Storm Surge Maximum Water Level Increase in a Coastal Area Using Ensemble Machine Learning and Explicable Algorithm. *Earth Space Sci*, *10*(12). https://doi.org/10.1029/2023EA003243

Tang C, Tao X, Wei Y, Tong Z, Zhu F, Lin H (2022) Analysis and Prediction of Wind Speed Effects in East Asia and the Western Pacific Based on Multi-Source Data. Sustainability 14(19):12089. https://doi.org/10.3390/su141912089

Tebaldi C, Strauss BH, Zervas CE (2012) Modelling sea level rise impacts on storm surges along US coasts. Environ Res Lett 7(1):014032. https://doi.org/10.1088/1748-9326/7/1/014032

Truccolo EC (2011) Assessment of the wind behavior in the northern coast of Santa Catarina. Revista Brasileira De Meteorologia 26(3):451–460. https://doi.org/10.1590/S0102-77862011000300011

Tyass I, Bellat A, Raihani A, Mansouri K, Khalili T (2022) Wind Speed Prediction Based on Seasonal ARIMA model. E3S Web Conf 336:00034. https://doi.org/10.1051/e3sconf/202233600034

Tyass I, Khalili T, Rafik M, Abdelouahed B, Raihani A, Mansouri K (2023) Wind Speed Prediction Based on Statistical and Deep Learning Models. Int J Renew Energy Dev 12(2):288–299. https://doi.org/10.14710/ijred.2023.48672

Valsaraj P, Alex Thumba D, Satheesh Kumar K (2022) Spatio-temporal independent applicability of one time trained machine learning wind forecast models: a promising case study from the wind energy perspective. Int J Sustain Energ 41(9):1164–1182. https://doi.org/10.1080/14786451.2022.2032060

Wang L, Zhou X, Zhu X, Dong Z, Guo W (2016) Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. Crop J 4(3):212–219. https://doi.org/10.1016/j.cj.2016.01.008

Wu F, Wang W, Hu Z, Wang X (2022) An effective estimation strategy for significant wave height via independent component analysis. Ocean Eng 250:110991. https://doi.org/10.1016/j.oceaneng.2022.110991

Yafouz A, Ahmed AN, Zaini N, Sherif M, Sefelnasr A, El-Shafie A (2021) Hybrid deep learning model for ozone concentration prediction: comprehensive evaluation and comparison with various machine and deep learning algorithms. Eng Applic Comput Fluid Mech 15(1):902–933. https://doi.org/10.1080/19942060.2021.1926328

Yeganeh-Bakhtiary A, EyvazOghli H, Shabakhty N, Abolfathi S (2023) Machine learning prediction of wave characteristics: Comparison between semi-empirical approaches and DT model. Ocean Eng 286:115583. https://doi.org/10.1016/j.oceaneng.2023.115583

Yeganeh-Bakhtiary A, EyvazOghli H, Shabakhty N, Kamranzad B, & Abolfathi S (2022) Machine Learning as a Downscaling Approach for Prediction of Wind Characteristics under Future Climate Change Scenarios. *Complexity*, *2022*(1). https://doi.org/10.1155/2022/8451812

Zhou S, Luo F (2024) Research on Wind Speed Prediction Method Based on Feature Fusion. Int J Comput Sci Inform Technol 2(1):437–442. https://doi.org/10.62051/ijcsit.v2n1.46

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Ahmet Durap[1]

Ahmet Durap
ahmetdurap@gmail.com

[1] Faculty of Engineering and Natural Sciences, Istanbul Medipol University, 34181 Beykoz, Istanbul, Turkey