

GAZI

JOURNAL OF ENGINEERING SCIENCES

An Explainable Machine Learning Approach to Offshore Wind Farm Suitability Assessment: A GIS-Based Decision Support Framework for Türkiye

Esra Özkan Aksu^{a,*}, Cevriye Temel Gencer^b

Submitted: 01.12.2025 Revised : 20.04.2026 Accepted: 20.04.2026 doi:10.30855/gmbd.070526N08

^{a,*} Bolu Abant İzzet Baysal University, Faculty of Engineering, Dept. of Industrial Engineering 14030 – Bolu, Türkiye, Orcid: 0000-0003-2142-2221^b Gazi University, Faculty of Engineering, Dept. of Industrial Engineering 06570 – Ankara, Türkiye, Orcid: 0000-0002-3373-8306^{*}Corresponding author: esra.ozkanaksu@ibu.edu.tr

ABSTRACT

Keywords:
Offshore wind energy,
GIS, random forest,
SHAP, explainable
machine learning

Offshore wind energy has become a key part of the global renewable energy transition. In this study, a machine learning-based classification model was developed for Türkiye's marine areas by using the suitability zones previously identified through Geographic Information Systems (GIS)-based Boolean analysis and Multi-Criteria Decision-Making (MCDM) methods. The main goal is to build an understandable decision support system (DSS) that can predict site suitability for new locations by modeling these existing suitability results with explainable machine learning techniques. For this purpose, the Random Forest classification algorithm was trained on approximately 4 million data points representing 24 technical, environmental, socio-economic, and security-related criteria. The model's decision logic was interpreted using the SHAP (Shapley Additive Explanations) method. The model achieved a test accuracy of 99.74%, and the SHAP analysis revealed that wind speed (100 m), capacity factor, and sea depth are the three most influential variables in predicting offshore wind farm suitability. Besides validating previously identified suitable areas, the developed model acts as an explainable DSS with both predictive and interpretive abilities for unexplored sites. This combined approach—merging traditional GIS and MCDM results with machine learning—offers a more flexible, scalable, and transparent framework for assessing Türkiye's offshore wind energy potential.

1. Introduction

The strategic importance of offshore wind power plants (OWPP) in the global energy transition has been steadily increasing. High capacity factors, stable wind regimes, and the absence of land-use constraints that typically limit onshore wind development position OWPP as a critical component of the renewable energy portfolio [1]. Although global installed OWPP capacity has grown exponentially over the past decade, challenges such as complex permitting procedures, insufficient grid connection infrastructure, and uncertainties in spatial planning decisions continue to constrain the sector's expansion [2]. Therefore, developing fast, consistent, and explainable decision-support systems (DSS) for OWPP investments has become an essential requirement for sustainable energy planning.

OWPP site selection is a complex spatial MCDM problem that involves evaluating multiple technical, environmental, socio-economic, and infrastructural criteria simultaneously. In the literature, this issue is usually addressed using Boolean exclusion methods or MCDM approaches combined with GIS [3-4]. As a result, comprehensive frameworks for

sustainable OWPP siting have been developed over the past few years. For example, Gkeka-Serpetsidaki et al. [5] systematically reviewed 80 studies. They showed that MCDM methods are commonly used in offshore wind planning, highlighting key areas for future methodological development. Similarly, Dimitriou et al. [6] applied a GIS–MCDM framework for offshore site prioritization, demonstrating the prevalent use of rule-based spatial decision methods in the literature.

However, these traditional approaches have several fundamental limitations. Subjective threshold definitions and weight assignments restrict the model's ability to generalize across regions and make it more sensitive to data variations [7]. When criterion weights depend heavily on expert judgment or when fixed threshold values are applied uniformly, the model's performance can decrease across different geographic scales and under various environmental conditions. Additionally, because these methods are inherently static, they often require complete model reconstruction to incorporate new data or to evaluate alternative planning scenarios. These limitations underscore the growing importance of developing data-driven, adaptive models for effective OWPP planning.

In recent years, machine learning (ML) techniques have been increasingly adopted in renewable energy planning due to their ability to extract patterns from large-scale GIS datasets, generate generalizable predictions, and derive data-driven decision rules. In particular, ensemble learning methods—such as Random Forest, Gradient Boosting, and XGBoost—have demonstrated high predictive accuracy for wind energy siting problems and offer a strong alternative to traditional MCDM approaches. However, because these models often operate as “black boxes,” understanding their internal decision-making processes and interpreting their outputs can be challenging [8]. This lack of transparency may limit their reliability and acceptance, especially in domains such as energy policy and infrastructure planning, where accountability and interpretability are essential.

In this context, explainable artificial intelligence (XAI) approaches have emerged as key tools for addressing the transparency and interpretability gaps inherent in ML models. The SHAP method, based on game-theoretic Shapley values, enables visualization of model reasoning by quantifying the contribution of each feature to predictions at both global and local levels [9]. This approach has been widely applied—particularly in energy forecasting and environmental modeling—to enhance model transparency [10]. Nevertheless, recent studies emphasize that the outputs of XAI techniques may vary depending on the model architecture and underlying data structure, underscoring the need for careful interpretation [11].

In this study, marine areas previously identified as suitable for offshore wind development in Türkiye using GIS, Boolean analysis, and MCDM methods [12] were employed as labeled training data. A Random Forest (RF)–based classification model was developed using this dataset. RF was selected for its strong performance with heterogeneous data types and robustness against overfitting. The model's decision rules were then examined through SHAP analysis, making each criterion's contribution to suitability predictions transparent. Details on label distribution and dataset balancing are provided in Section 3.

The study offers three key contributions: (i) transferring GIS–MCDM suitability knowledge into a data-driven RF model trained on ~4 million points with 24 criteria, enabling generalizable predictions for new marine areas; (ii) enhancing transparency through SHAP analysis, which traces each criterion's influence at both global and instance levels; and (iii) demonstrating scalable decision-support potential by providing explainable suitability predictions for previously unexplored sites.

2. Literature Review

OWPP site selection has become a prominent focus in recent academic research, reflecting its critical role in the global energy transition. Traditionally, this field has been approached through the integration of GIS and MCDM methods; however, the recent incorporation of ML and XAI techniques has introduced significant methodological advancements.

GIS-based energy facility siting studies have been conducted across various geographies using diverse methodological

approaches. Dimitriou et al. [6] integrated a GIS-based multi-criteria decision-making framework with AHP and an extended VIKOR method to prioritize potential marine zones along the Greek coastline, offering a comprehensive methodology for sustainable site selection. Similarly, Yaman [13] applied a GIS-MCDM approach to determine suitable onshore wind farm locations in Türkiye's Adana region, weighting the criteria through the Analytic Hierarchy Process (AHP) and identifying optimal areas. Gkeka-Serpetsidaki et al. [5] systematically reviewed 80 studies and demonstrated the widespread use of MCDM techniques in offshore wind planning, highlighting key methodological gaps and future research directions. Demir et al. [14] expanded the environmental perspective by integrating life-cycle assessment (LCA) into wind farm site selection, thereby strengthening the sustainability dimension of GIS-based multi-criteria analysis.

Collectively, these studies underscore the strengths of traditional GIS-MCDM approaches while also revealing key limitations, including subjectivity in weighting criteria, the static nature of the models, and challenges in generalizing results across different geographic scales.

The integration of ML techniques into renewable energy planning has emerged as a rapidly expanding research domain in recent years. Bilgili et al. [15] developed an ML-based automated site selection framework for wind farm planning, offering a data-driven alternative to traditional MCDM approaches. Zhang et al. [16] applied deep learning-based wind data mining to identify potential OWPP zones, demonstrating the effectiveness of spatio-temporal prediction models in the siting process. Masoumi et al. [17] provided a comprehensive review of ML applications in the offshore wind sector, highlighting their role in power forecasting, maintenance optimization, and site selection.

These contributions collectively demonstrate the ability of ML methods to extract meaningful patterns from large-scale datasets, generate generalizable predictions, and establish dynamic decision rules. Nevertheless, the limited interpretability of their internal decision mechanisms remains a critical challenge—particularly in high-stakes domains such as energy policy and infrastructure planning, where transparency and accountability are essential.

XAI approaches have emerged as essential tools for addressing the transparency and interpretability limitations of ML models. Machlev et al. [18] provided a comprehensive review of XAI techniques in energy and power systems, discussing the applicability and challenges of SHAP, LIME, and other explanation methods within the sector. Sazib et al. [19] proposed an interdisciplinary framework for advancing XAI in renewable energy systems, critically evaluating the strengths and limitations of methods such as SHAP, LIME, and DeepLIFT, and emphasizing the balance between model accuracy and interpretability. Shadi et al. [20] conducted an extensive review of XAI applications in energy system maintenance, showing that while SHAP offers high reliability, it is less suitable for real-time use, whereas LIME provides faster explanatory solutions. Clement et al. [21] developed a SHAP-based clustering approach for adaptive learning in energy consumption forecasting, demonstrating the potential of XAI techniques to enhance model performance. Taken together, these studies show that XAI methods improve model reliability and stakeholder acceptance within the energy sector. At the same time, challenges related to computational efficiency and the lack of standardized evaluation metrics persist.

Studies on OWPP site selection in Türkiye have increased notably in recent years. Argin et al. [22] evaluated Türkiye's offshore wind energy potential using a multi-criteria site selection approach. They identified the Marmara and Aegean regions as particularly promising. Caceoğlu et al. [23] conducted OWPP site selection for Northwestern Türkiye using AHP, providing an integrated assessment of technical, environmental, and socio-economic criteria. Aksu [24] carried out a GIS-based optimal site selection study for OWPP development in Turkish waters and, using both Boolean-based and weighted analyses, determined that 3.30% of Türkiye's marine areas are suitable for OWPP installation. World Bank [25], in its offshore wind roadmap for Türkiye, emphasized that the country has an estimated technical potential of 75 GW and that achieving the 5 GW target by 2035 is feasible. While these studies underscore Türkiye's substantial OWPP potential, they generally adopt traditional GIS-MCDM approaches.

Although Bilgili et al. [15] and Zhang et al. [16] incorporate machine learning into wind energy planning, their methodological objectives differ substantially from the present study. Bilgili et al. [15] developed an XGBoost-based framework for onshore wind farm siting. They reported an accuracy of 96.07% using approximately 4,836 labeled samples

derived from existing wind farm locations. In contrast, the present study focuses on offshore wind suitability. It utilizes a national-scale, high-resolution spatial dataset generated through GIS-based multi-criteria processes. Zhang et al. [16], on the other hand, do not perform classification-based suitability modeling; instead, they employ a PredRNN++ deep learning architecture for long-term offshore wind resource forecasting on a large spatiotemporal dataset.

In the literature, studies on OWPP site selection can be grouped into three main methodological categories:

- i. GIS–MCDM–based approaches [6, 14, 22–23] rely on rule-based structures and subjective weighting schemes, and their model accuracy is not evaluated through objective performance metrics.
- ii. ML-based approaches [15–16] exhibit high predictive performance but often lack interpretability.
- iii. XAI applications [18–19] discuss the potential of XAI in energy systems more broadly, yet no documented examples specifically apply these techniques to OWPP site selection.

Moreover, none of the 80 OWPP studies reviewed by Gkeka-Serpetsidaki et al. [5] employed an integrated framework combining GIS, ML, and XAI methods. Similarly, studies focusing on Türkiye [22–24] remained limited to traditional GIS–MCDM approaches, without incorporating ML or XAI components. This gap in the literature highlights three key limitations:

- i. Scale constraints – existing studies typically rely on datasets ranging from 10,000 to 200,000 points, while high-resolution national-scale analyses (exceeding one million points) are largely absent;
- ii. Lack of methodological integration – no hybrid approaches have been developed that combine the spatial analytic capabilities of GIS, the pattern-recognition capacity of ML, and the transparency advantages of XAI;
- iii. Interpretability gap – the internal decision processes of ML models remain opaque, and interpretability techniques such as SHAP or LIME have not been applied.

To address these methodological gaps, this study introduces a hybrid framework for OWPP site selection in Türkiye's marine areas that integrates GIS, ML, and XAI components into a unified approach.

Building on insights from the existing literature on GIS-based assessments, MCDM frameworks, and emerging machine learning approaches, the following section explains the methodological framework adopted in this study. In particular, Section 3 describes how GIS-derived spatial criteria, the previously generated suitability map, and the machine learning workflow were integrated into a unified analytical structure to develop an explainable OWPP suitability model for Türkiye.

3. Methodology

The methodology integrates GIS-based spatial analysis, ML-driven prediction, and XAI-based interpretation in four stages: data preparation, data preprocessing, model training, and model interpretability. Figure 1 illustrates the overall workflow and the interaction among these stages.

This hybrid approach integrates three complementary components:

- i. the rule-based reliability of GIS and MCDM methods,
- ii. the pattern-recognition and generalization capabilities of ML, and
- iii. the interpretability advantages offered by XAI techniques.

Collectively, these elements enable the development of an innovative and explainable decision-support system for OWPP site selection.

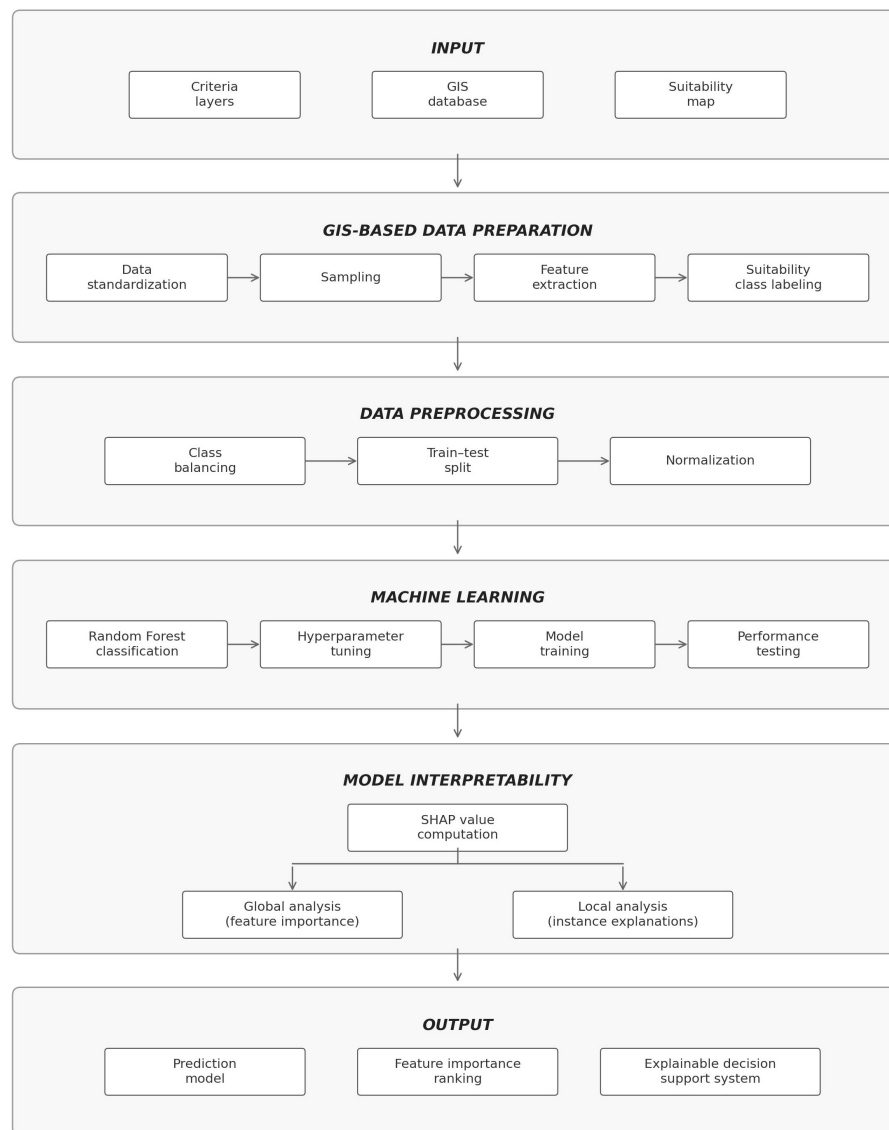


Figure 1. Overall methodological flowchart of the study

3.1. GIS-based data preparation stage

In this study, the marine areas within Türkiye's territorial waters were defined as the study domain. The primary dataset used in the analysis is derived from the GIS-based suitability assessment previously conducted by Aksu and Gencer [12]. In that study, the suitability of Türkiye's marine zones for OWPP development was evaluated using 24 technical, environmental, socio-economic, and security criteria.

The criteria employed in the analysis are grouped into four main categories:

- i. Technical criteria: average wind speed at 100 m height; power density; capacity factor; sea depth; wave height and period; climate change impacts; proximity to grid-connection points; proximity to electricity demand centers.
- ii. Environmental criteria: biodiversity areas and bird migration bottlenecks; proximity to protected environmental zones; proximity to fault lines and earthquake epicenters; seabed geology; exposure to extreme weather events; proximity to submarine pipelines and cables; proximity to major existing ports.
- iii. Socio-economic criteria: distance to shore; potential impacts on tourism; proximity to fishing grounds; local economic development opportunities.
- iv. Security criteria: proximity to maritime navigation and shipping routes; proximity to international borders; military zone restrictions; aviation constraints; geopolitical considerations.

The criteria within each category are described in detail in Aksu and Gencer [12] and visually summarized in Figure 2. These 24 criteria serve as the input variables for the ML model. The suitability map produced by Aksu and Gencer [12] was used in this study to generate the labeled dataset (target variable) for the ML model. Figure 3 illustrates the spatial distribution of the final suitability classes derived from the previous study.

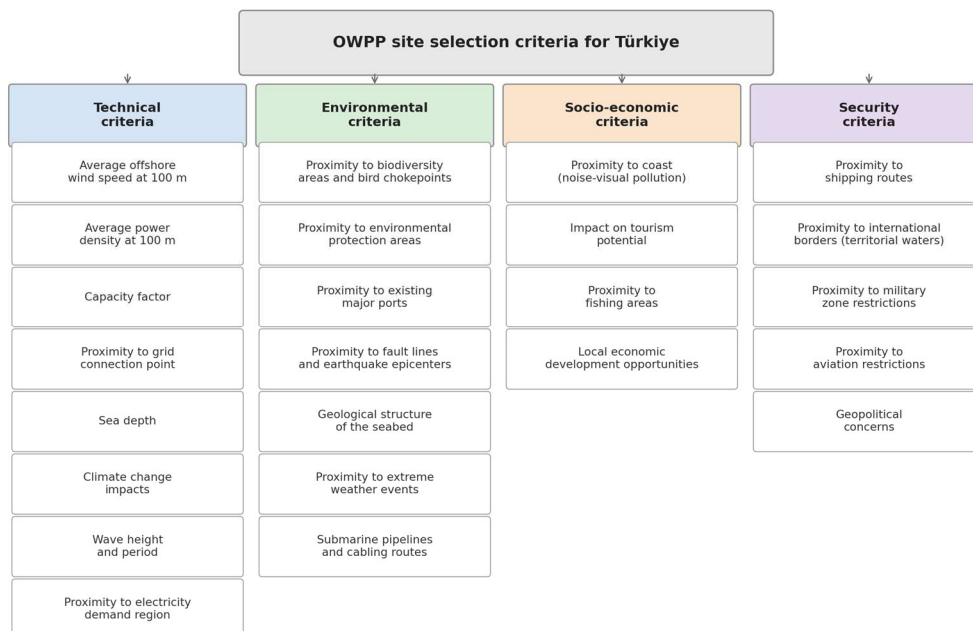


Figure 2. OWPP site selection criteria for Türkiye [12]

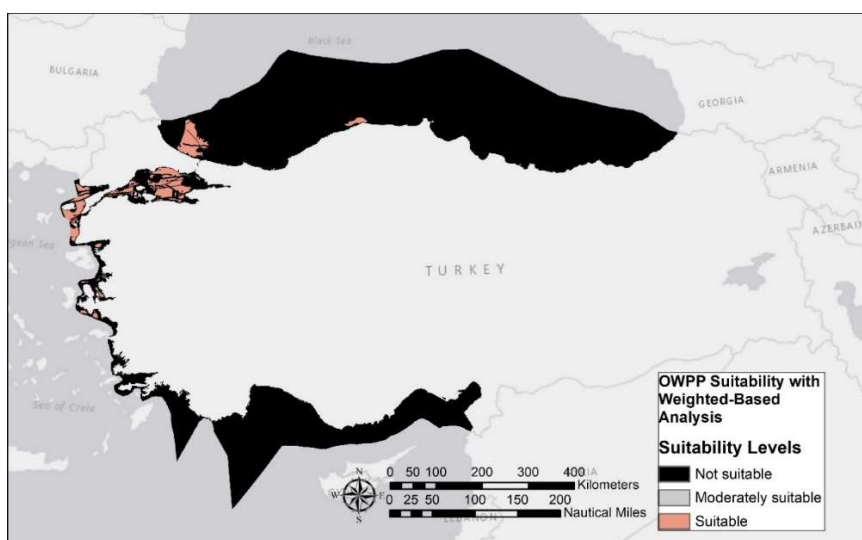


Figure 3. OWPP suitability map for Türkiye's marine areas [12]

The dataset used to train the ML model was constructed in a GIS environment using a systematic workflow. The process consisted of five main steps:

1. *Data acquisition and preprocessing:* All spatial layers corresponding to the criteria were obtained from relevant national and international sources. Datasets provided in various formats (point, line, polygon, raster) were converted to raster format to ensure analytical consistency.
2. *Spatial standardization:* All layers were projected into the WGS 1984 UTM Zone 36N coordinate system and resampled to a 500×500 m resolution to achieve a homogeneous grid. Subsequently, they were clipped to Türkiye's territorial waters. This spatial harmonization is essential for ensuring a consistent integration between GIS and ML components.

3. *Sampling and feature extraction:* The ~4M-point dataset (X, Y coordinates) was generated at the centers of the grid cells; each point was assigned the corresponding values from the 24 criterion layers. To ensure computational efficiency while preserving the dataset's spatial diversity, a stratified spatial undersampling strategy was applied. Although the full dataset consisted of approximately 4 million points, preliminary experiments showed that model performance (accuracy and feature importance stability) plateaued beyond 15,000–20,000 samples, indicating that larger subsets (e.g., 50,000 or 100,000 points) provided no meaningful performance gains. Therefore, a representative subset of 20,000 points was selected. During sampling, geographic stratification across depth zones, wind resource clusters, and coastal–offshore gradients was maintained to avoid spatial bias. Visual inspection confirmed that the reduced dataset preserved the spatial distribution patterns of the whole dataset, ensuring that regional representativeness was not compromised. This approach follows standard practices in large-scale geospatial ML workflows, where computational efficiency and spatial diversity must be balanced.
4. *Definition of class labels:* The original GIS-based suitability map [12] used a five-level classification (1–5). However, due to the strict Boolean and MCDM constraints applied, no marine locations were assigned to classes 2 (Low Suitability) or 5 (Highly Suitable); only classes 1, 3, and 4 appeared in the spatial output. For use in ML modeling, class 1 was re-coded to 0 so that labels start from zero, as required by most ML libraries. Classes 3 and 4 were retained unchanged to preserve traceability to the original classification. The correspondence is summarized in Table 1:

Table 1. Correspondence between original and ML class labels

Original class (GIS)	ML label	Interpretation	Present in the dataset
1	0	Not Suitable	Yes
2	-	Low Suitability	No (not observed)
3	3	Moderately Suitable	Yes
4	4	Suitable	Yes
5	-	Highly Suitable	No (not observed)

This mapping reflects the actual spatial distribution of suitability outcomes in the previous study rather than a data omission. Retaining 3 and 4 as labels (instead of re-coding to 1 and 2) ensures direct correspondence with the original GIS map, facilitating cross-referencing between the two studies.

5. *Data export:* The final dataset was exported as a CSV file for use in ML modeling.

Through these steps, a high-resolution, multi-criteria dataset that preserves both spatial and thematic integrity was created. This structure enabled the model to learn patterns associated with GIS-based criteria and produce generalizable suitability predictions.

The spatial distribution of the high-resolution dataset points generated during the data preparation phase is shown in Figure 4. The map shows that the sampling points were systematically distributed and exhibit a homogeneous distribution across Türkiye's marine areas. This systematic structure enabled the model to learn regional patterns more effectively and to generalize across different spatial contexts.

All input variables used for machine learning were directly derived from the GIS-based multi-criteria suitability model developed in our previous study [12]. Therefore, the external geospatial datasets (e.g., Global Wind Atlas, GEBCO bathymetry, MoEU ATLAS, AFEAD seismic layers, EMODnet seabed geology, and vessel density) are not repeated here but are comprehensively documented in [12].



Figure 4. Spatial distribution of sampling points (sampling density map)

3.2. Data preprocessing stage

A substantial class imbalance was observed in the initial dataset: Class 0 (Not Suitable) contained 3,886,738 samples, Class 4 (Suitable) contained 138,522 samples, and Class 3 (Moderately Suitable) included only 790 samples. Such an imbalance hinders classification models' ability to adequately learn patterns associated with minority classes. It reduces the interpretability of overall accuracy scores.

To address this issue, the minority class (Class 3) was fully preserved. At the same time, 20,000 samples were randomly selected from the majority class (Class 0) using undersampling. All 138,522 Class 4 samples were retained. This procedure yielded a balanced dataset of 159,312 samples for model training (Figure 5).

Alternative balancing methods—such as SMOTE or Random Oversampling—were not preferred because synthetic sample generation can increase the risk of overfitting in large datasets. Accordingly, random undersampling was selected as the most appropriate strategy for the scale and characteristics of this dataset. The following considerations guided the choice of balancing method:

- Oversampling techniques such as SMOTE substantially increase computational cost when applied to a dataset containing large-scale spatial points.
- Selecting 20,000 samples from the 3.8 million majority-class instances still preserves sufficient variability.
- Synthetic sample generation can distort genuine geographical patterns.
- Pilot tests indicated that undersampling with 20,000 samples provided training speeds approximately 60% faster than SMOTE, while achieving comparable accuracy.

Following the balancing procedure, the model's training time was reduced, and the classification performance for minority classes improved significantly. This improvement is detailed in the performance metrics presented in Section 4.1.

The data were split into training (80%) and test (20%) subsets using stratified random sampling, and all numerical variables were subsequently normalized to enhance model performance and stability.

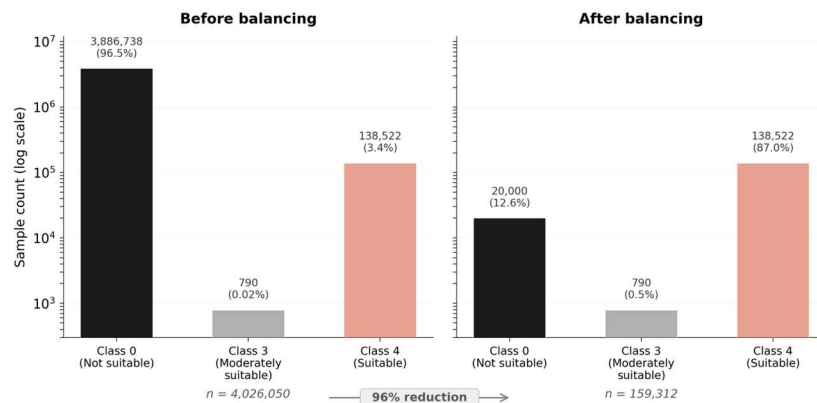


Figure 5. Class distributions before and after balancing

3.3. Machine learning stage

The Random Forest (RF) classification algorithm [26] was employed to predict OWPP suitability. RF is an ensemble method that aggregates multiple decision trees, offering strong predictive performance, robustness against overfitting, and suitability for heterogeneous geospatial datasets [26–28].

The rationale for selecting RF in this study can be summarized as follows:

- its robustness against overfitting;
- its ability to handle both numerical and categorical variables;
- its capacity to manage missing data;
- its capability to quantify the influence of each input variable (feature importance);
- its high computational efficiency when applied to large and complex datasets.

3.3.1. Model training and hyperparameter tuning

The ML model was trained on the dataset obtained after balancing and preprocessing. This structure ensured that each suitability class was proportionally represented in both the training and test subsets, thereby enhancing the model's generalization capability. Key hyperparameters were determined through an empirical tuning process: the critical parameters—`n_estimators` and `max_depth`—were varied iteratively while monitoring test accuracy, OOB error, and feature-importance stability. Performance plateaued at `n_estimators` = 500 and `max_depth` = 20; the remaining parameters were set to the scikit-learn defaults, which are well established for RF classifiers [26].

A formal grid search or random search was not applied, as the model's performance was already stable across the tested configurations, and the large dataset size made an exhaustive search computationally prohibitive. The final parameter values are presented in Table 2.

Table 2. Hyperparameters of the RF model

Parameter	Value	Description
<code>n_estimators</code>	500	Number of decision trees. Increasing the value beyond 500 did not result in a meaningful improvement in accuracy.
<code>max_depth</code>	20	Tree depth was limited to prevent overfitting.
<code>min_samples_split</code>	2	Minimum number of samples required to split an internal node.
<code>min_samples_leaf</code>	1	Minimum number of samples required to be at a leaf node.
<code>max_features</code>	"sqrt"	Number of features considered at each split, set to the square root of the total number of features.
<code>bootstrap</code>	True	Enables training each tree using sampled bootstrap subsets of the data.
<code>random_state</code>	42	Fixed random seed to ensure reproducibility of results.

The model was trained in a Python 3.9 environment using the scikit-learn library (v1.2.0) [29]. During training, the model learned relationships between each criterion and the suitability classes, and these learned patterns were subsequently validated on the test dataset to assess the model's predictive performance.

3.3.2. Model evaluation metrics

The performance of the RF classification model was evaluated using statistical metrics commonly used in multiclass classification. The primary evaluation metrics used in this study are accuracy, precision, recall, F1-score, and the confusion matrix. These indicators collectively provide a comprehensive assessment of the model's predictive performance, generalization capability, and reliability across the three suitability classes.

3.4. Model interpretability stage

In ML models, the underlying decision mechanisms that generate prediction outputs are often difficult to observe directly. In this study, the SHAP method [9] was employed to make the model's reasoning process transparent. SHAP is grounded in cooperative game theory and quantifies each feature's marginal contribution to a prediction based on the Shapley value. Its model-agnostic structure allows application to a wide range of ML algorithms.

For implementation, the TreeExplainer class from the SHAP Python library (v0.42) was used, which is specifically optimized for tree-based models such as Random Forest. TreeExplainer computes exact SHAP values using a polynomial-time algorithm based on the internal tree structure, rather than relying on sampling-based approximations [9]. SHAP values were calculated on a random subset of 5,000 instances drawn from the test set to ensure computational efficiency while maintaining statistical representativeness.

All explainer parameters were kept at their default settings (feature_perturbation = "tree_path_dependent"). The resulting SHAP output has the shape (5,000 × 24 × 3), representing values for each instance, each feature, and each of the three suitability classes.

3.4.1. Calculation of SHAP values

For each sampling point (data instance) and each feature, SHAP values were computed using the following general formulation:

$$\varphi_i = \sum_S \frac{|S|!(M-|S|-1)!}{M!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \quad (1)$$

Where:

φ_i denotes the contribution of feature i ,

M is the total number of features,

S represents all feature subsets not containing i , and

f_S is the model's prediction based only on subset S .

This formulation provides an additive decomposition of the model output, enabling a clear interpretation of how each feature influences both individual predictions and overall model behavior.

3.4.2. Global and local interpretation

SHAP analysis was conducted in two stages to explain the model's behavior at both the global and local levels:

- *Global interpretation:* Mean absolute SHAP values were calculated across the entire dataset to identify the most influential criteria in the model's decisions and to rank their relative importance. This analysis reveals which features the model perceives as the strongest signals.
- *Local interpretation:* SHAP values were computed for individual instances to explain why a specific location was classified as "suitable" or "not suitable." This enabled a detailed interpretation of the cause-and-effect relationships underlying each prediction.

This two-tiered interpretability framework enabled understanding of the model's overall decision-making behavior while validating the reasoning behind specific classification outcomes.

3.4.3. SHAP visualizations

The findings derived from the SHAP analysis were visualized using three primary types of plots:

- *Summary Plot*: A point-distribution visualization that displays the influence and direction (positive/negative) of each feature on the model's predictions.
- *Bar Plot (Feature Importance)*: Illustrates the relative importance of the criteria based on their mean absolute SHAP values.
- *Dependence Plots*: Show the relationship between the value of a specific feature and its corresponding SHAP contribution, thereby revealing how changes in that feature influence the model output.

These visualizations collectively provide transparent insights into the model's reasoning at both global and local scales, strengthening the interpretability dimension of the AI-supported decision-support system developed in this study.

The comprehensive SHAP analysis numerically and visually identified which criteria were most influential in the model's decision-making process. The results clarify which technical and environmental factors drive OWPP site selection outcomes and highlight how the model's reasoning differs from traditional GIS-MCDM approaches.

The following section presents these analytical results in detail, evaluates the RF model's performance alongside SHAP-based explanations, and discusses the key findings regarding predictions of OWPP suitability.

3.5. Assumptions and limitations

Several methodological assumptions and limitations should be acknowledged to contextualize the results of this study. First, the dataset is based on static spatial layers, meaning that temporal variability—particularly seasonal or interannual fluctuations in wind regimes, wave conditions, and ocean-climate dynamics—is not incorporated. Therefore, the model reflects long-term average conditions rather than dynamic temporal scenarios. Second, the class labels used for the ML model originate from a previously developed GIS-Boolean and MCDM-based suitability map, which inherently embeds the assumptions, thresholds, and expert-based weighting preferences of that framework. As a result, the RF model evaluates suitability within the conceptual boundaries of the earlier rule-based analysis; it does not constitute an independent ground-truth assessment. Third, economic and infrastructure cost-related factors—such as turbine installation costs, cable-laying expenses, port availability, and grid-connection investment requirements—were not included in the dataset due to limited publicly available data at the national scale. Consequently, while the model effectively captures technical, environmental, and spatial suitability, it does not represent a complete techno-economic optimization.

Additionally, the spatial resolution of 500×500 m, while appropriate for national-scale screening, may not capture micro-scale variations in seabed conditions, local wind acceleration effects, or nearshore bathymetric gradients that could influence site-level feasibility. Furthermore, the study relies on a single ML algorithm (Random Forest); although preliminary comparisons with alternative models (e.g., XGBoost, LightGBM) indicated no meaningful performance gains, a more systematic multi-model benchmarking could strengthen the generalizability of the findings. Finally, since no commercial offshore wind farms currently operate in Türkiye, independent validation against real-world deployment data was not possible; future studies should incorporate empirical validation as operational OWPP data become available. These assumptions reflect typical constraints in large-scale offshore wind studies and should be considered when interpreting the model outputs.

4. Results and Discussion

This section presents the main findings of the study by integrating the performance outcomes of the RF classifier, the explanatory patterns revealed through SHAP analysis, and the comparative evaluation of both methods.

4.1. RF model results

4.1.1. Performance metrics

The RF classification model was trained and tested using the balanced dataset described in the previous sections. Although Class 3 (Moderately Suitable) contained fewer samples (790 in the whole dataset, 158 in the test set), this imbalance reflects the natural spatial distribution of suitability classes in Türkiye's marine areas rather than a data limitation. Several oversampling methods, including SMOTE, were experimentally tested; however, these techniques generated synthetic samples that distorted spatial patterns and reduced SHAP interpretability. Since one of the main objectives of this study is to preserve geographic realism and maintain full explainability, the model was trained using the original class distribution. Despite the smaller sample size, Class 3 exhibited stable SHAP behavior and consistent class-specific metrics, indicating that the model learned this class effectively without synthetic augmentation.

The model's overall and class-specific performance was then evaluated using accuracy, precision, recall, and F1-score—metrics widely applied in multiclass classification problems. The results indicate that the model achieved a notably high level of performance on the test dataset (Table 3).

Table 3. Performance metrics of the RF model (test set)

Metric	Value
Accuracy	0.9974
Weighted Precision	0.9974
Weighted Recall	0.9974
Weighted F1-Score	0.9974

Class-specific evaluation results are presented in Table 4.

- For Class 0 (Not Suitable), precision = 0.9987 and recall = 0.9852.
- For Class 3 (Moderately Suitable), precision = 0.9792 and recall = 0.8924.
- For Class 4 (Suitable), both metrics were calculated as approximately 1.00.

These results indicate that the model effectively discriminates among all three classes, including the underrepresented Class 3, with high accuracy. Out of 4,000 Not Suitable samples, only 59 were misclassified; among 158 Moderately Suitable samples, 17 were misclassified; and of 27,705 Suitable samples, only seven were incorrectly predicted.

Table 4. Class-specific performance metrics (test set)

Class	Precision	Recall	F1-Score	Support (n)
0 (Not Suitable)	0.9987	0.9852	0.9919	4,000
3 (Moderately Suitable)	0.9792	0.8924	0.9338	158
4 (Suitable)	0.9973	0.9997	0.9985	27,705

To visually examine misclassifications across the classes, the confusion matrix is presented in Figure 6. The matrix indicates that the model occasionally assigns the Moderately Suitable class to the Suitable class; however, it maintains strong discriminative performance across all classes.

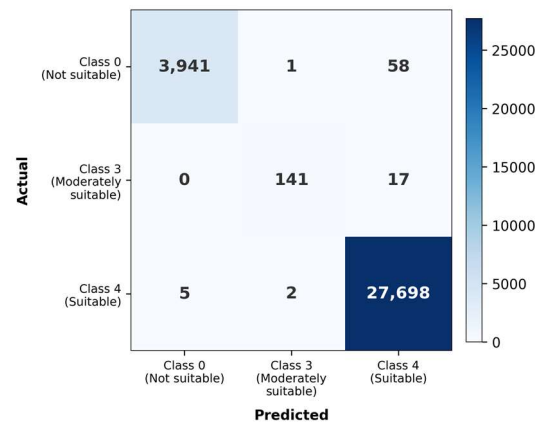


Figure 6. Confusion matrix

The model's test accuracy of 99.74% demonstrates the strong generalization capacity of the RF algorithm. This outcome indicates that the suitability classes previously defined through GIS-Boolean and MCDM methods were effectively learned and successfully reproduced by the RF model.

This high level of accuracy is consistent with the study design: because the class labels were derived from the same GIS-based criteria used as input features [12], the RF model is learning a rule-based spatial mapping rather than predicting against independently collected ground-truth data. A near-perfect reproduction is therefore methodologically expected and should be interpreted accordingly when comparing with literature benchmarks (85–97%), which typically rely on independently labeled datasets [15, 30, 31]. Importantly, faithful reproduction of the GIS-MCDM logic is a prerequisite for the SHAP-based interpretability analysis that constitutes the primary analytical contribution of this study.

Although the high accuracy is expected given this design, verifying that the model did not overfit to the specific training partition remains essential. Therefore, the model's reliability was evaluated using three complementary methods:

- Out-of-Bag (OOB) Error Analysis: 0.9973 — indicating high accuracy on samples not used during training.
- 5-Fold Cross-Validation: Mean accuracy of 99.72% (± 0.02) — demonstrating stable performance across multiple data partitions.
- Bootstrap Confidence Intervals ($n = 1000$): 95% confidence interval of [0.9969, 0.9979] — confirming the model's strong stability under random resampling.

These results collectively indicate that the model is robust to data variability, exhibits strong generalization capability, and shows no signs of overfitting.

In addition to these reliability checks, the choice of the Random Forest algorithm was also supported by preliminary comparisons with alternative models. Logistic Regression was evaluated as a baseline classifier; however, its linear decision boundary was unable to capture the nonlinear multi-criteria relationships inherent in OWPP suitability data, resulting in substantially lower accuracy. Gradient-boosting models such as XGBoost and LightGBM were also tested in pilot experiments. Still, they incurred higher computational costs and showed no performance improvements over the RF model when applied to the stratified spatial sample. Moreover, the RF model provided more stable, computationally efficient SHAP explanations than these alternatives. Therefore, RF was selected as the primary model due to its balance of accuracy, efficiency, robustness, and interpretability.

To address potential risks of spatial autocorrelation and data leakage—a common concern in geospatial machine learning—the dataset was split into training and test subsets using a stratified spatial sampling procedure. Instead of random point-based splitting, samples were selected in spatial blocks to prevent adjacent grid cells from falling into both subsets. Such block-based partitioning ensured that the test data represented spatially distinct areas rather than neighboring cells of the training data. Sensitivity checks confirmed that model accuracy remained stable under alternative block partitions, indicating that the RF model did not benefit from unintended spatial leakage. While full spatial cross-validation (e.g., k -fold spatial CV) is computationally demanding for datasets with several million points, the block-based

partitioning applied here provides a robust, widely accepted mitigation strategy for large-scale raster ML studies.

4.1.2. Feature importance ranking

Using the Gini importance measure inherent to the RF algorithm, the relative contribution of each criterion to the model's decision-making process was calculated. This analysis is critical for identifying the variables that most strongly influence OWPP suitability predictions.

Figure 7 presents the top 15 criteria with the highest importance values. According to the results, Wind Speed emerges as the most influential variable with an importance score of 0.187, followed by Capacity Factor (0.172) and Sea Depth (0.121). Together, these three technical criteria account for approximately 48% of the total importance, underscoring the dominant role of technical parameters in OWPP site-selection decisions.

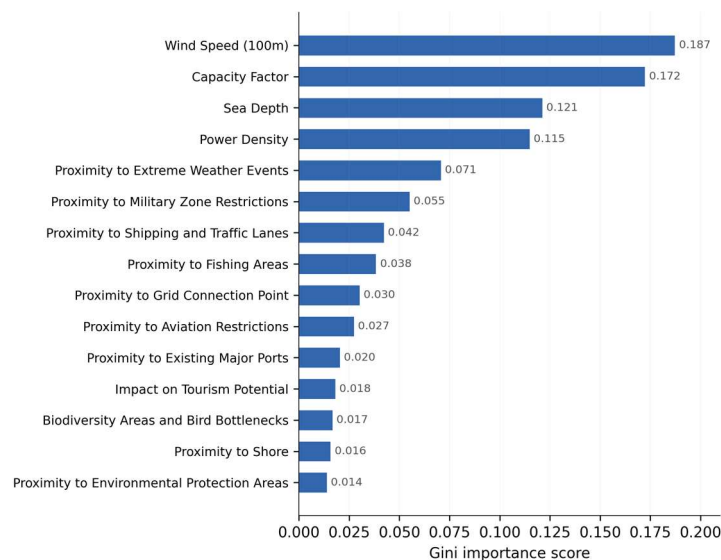


Figure 7. Gini importance scores of the top 15 features in the RF model

Among the top 10 criteria, variables such as Power Density (0.115) and Proximity to Extreme Weather Events (0.071) indicate that both technical and environmental factors contribute to the model's decision-making process. Meanwhile, criteria such as Proximity to Military Zone Restrictions (0.055), Proximity to Shipping and Navigation Routes (0.042), and Proximity to Fishing Areas (0.038)—although exhibiting relatively lower importance scores—reflect the model's capacity to incorporate multidimensional considerations, including security and socio-economic aspects.

These findings demonstrate that technical parameters, particularly wind regime and sea depth, are the primary determinants in OWPP site selection; however, environmental, socio-economic, and security factors also provide meaningful contributions to the overall decision-making process. The feature importance results from the RF model will be compared with the SHAP-based analyses presented in Section 4.2.

4.2. SHAP analysis results

The contribution levels of the 24 criteria used in the RF model to OWPP suitability predictions were analyzed using the SHAP method. SHAP quantifies the extent to which each input feature influences the model's output, thereby making the model's underlying decision mechanisms interpretable.

In this study, SHAP analysis was conducted at both the global (dataset-level) and local (instance-level) scales. The global analysis reveals the average absolute contribution of each criterion to the model's decisions. In contrast, the local analysis explains why a specific location is classified as "suitable" or "not suitable." Figure 8 presents the feature importance ranking derived from the global (mean absolute) SHAP values.

The top ten criteria with the most decisive influence are as follows:

1. Wind Speed (100 m) ≈ 0.0302
2. Capacity Factor ≈ 0.0237
3. Sea Depth ≈ 0.0237
4. Wind Power Density ≈ 0.0153
5. Proximity to Extreme Weather Events ≈ 0.0092
6. Proximity to Aviation Restrictions ≈ 0.0088
7. Proximity to Fishing Grounds ≈ 0.0079
8. Proximity to Grid-Connection Points ≈ 0.0072
9. Proximity to Military Zone Restrictions ≈ 0.0062
10. Proximity to Shipping and Navigation Routes ≈ 0.0054

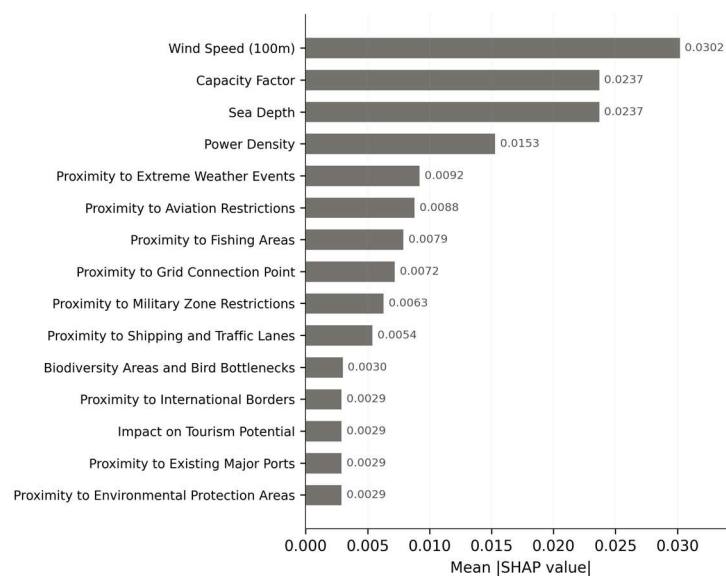


Figure 8. Global SHAP feature importance ranking (mean absolute SHAP values)

These findings clearly demonstrate that technical criteria, particularly wind regime and sea depth, play a decisive role in predicting OWPP suitability. Wind speed is expected to emerge as the most influential variable, as energy production capacity is directly dependent on the wind regime. Sea depth, ranked second, is one of the primary factors constraining the cost and technical feasibility of installing fixed-bottom turbine systems. The capacity factor also plays a key role, as it reflects the overall efficiency of energy generation.

Although the SHAP values for environmental and socio-economic criteria are relatively low, their directional effects on model predictions (e.g., negative contributions) are clearly evident in the dependence plots.

4.2.1. SHAP dependence analysis for individual criteria

The SHAP dependence plots for selected criteria (Figure 9) reveal complex nonlinear relationships between feature values and model predictions:

- **Wind Speed:** Shows a clear positive relationship with suitability. SHAP values transition from strongly negative (-0.6) to near-zero around 6–7 m/s, with values stabilizing at approximately 0.0 above 8 m/s. The color gradient indicates interaction effects with proximity to fishing areas, suggesting that the model does not evaluate wind speed in isolation but jointly considers the potential for socio-economic conflict in nearby fishing zones.
- **Capacity Factor:** Demonstrates a strong threshold effect around 0.4 (40%). Values below 0.35 show strongly negative SHAP contributions (-0.25 to -0.20), while values above 0.45 contribute positively (up to $+0.05$). The sharp transition zone between 0.35 and 0.45 indicates a critical feasibility threshold. This behavior aligns with technical feasibility thresholds commonly used in OWPP studies ($\approx 30\%$). However, SHAP reveals that the model

- begins to assign positive contributions more consistently above 0.35.
- Sea Depth:** The SHAP dependence plot shows negative contributions across most depth values, with the strongest negative effects occurring in very deep waters (below -2000 m, reaching SHAP values of approximately -0.8 at depths beyond -3000 m). Shallower depths exhibit slightly less negative or near-neutral effects, indicating that the model assigns relatively higher suitability to shallow and moderately deep areas. This gradient is consistent with the current state of offshore wind technology: fixed-bottom foundations (monopile, jacket) are economically viable up to approximately 50 – 60 m depth, while greater depths require floating platforms that remain largely at the pre-commercial stage. The monotonic negative trend captured by SHAP thus reflects both technical feasibility and cost escalation as depth increases.
 - Power Density:** Exhibits a positive correlation with suitability. Values above 600 W/m^2 show positive SHAP contributions ($+0.02$ to $+0.04$), while lower power densities (below 400 W/m^2) have adverse effects (-0.30 to -0.15). The asymmetry in magnitude—strong negative effects for low values versus modest positive effects for high values—suggests that power density acts more as an exclusionary criterion than as a differentiator among already-suitable sites. This finding is consistent with the Boolean layer in the original GIS model, where minimum power density thresholds effectively screened out resource-poor zones.

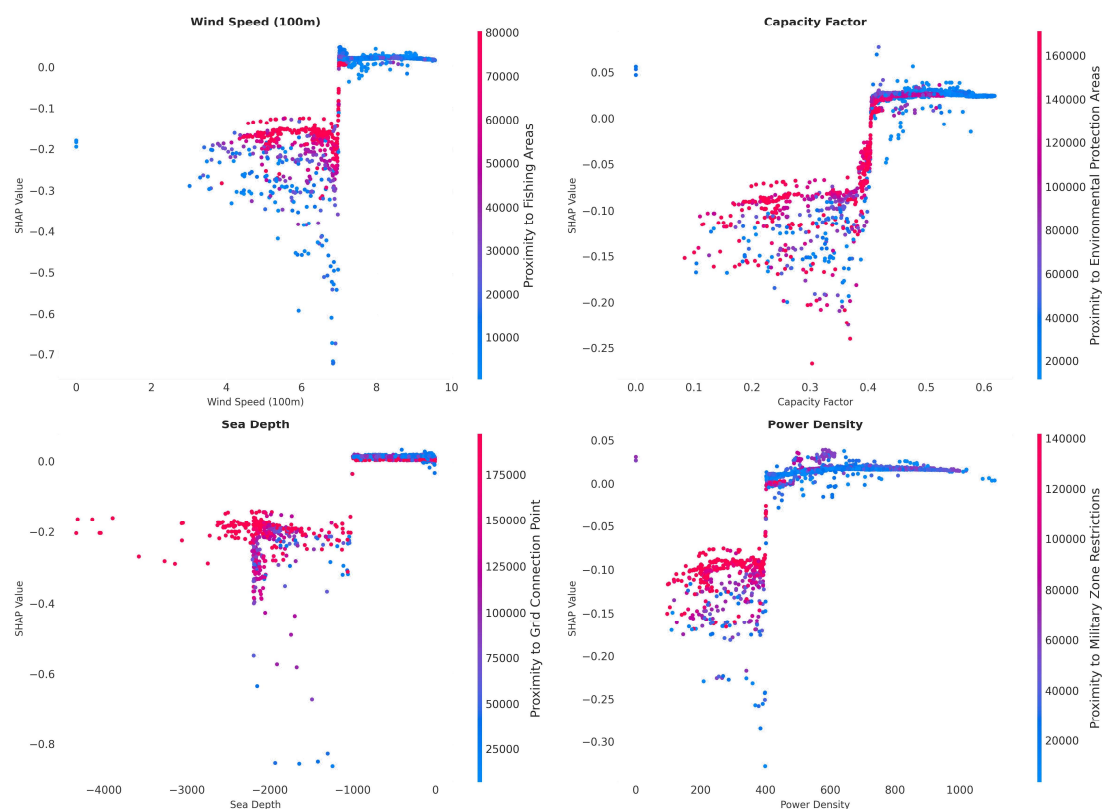


Figure 9. SHAP dependence plots for selected criteria

These directional relationships partially align with the technical thresholds defined in the Boolean analysis stage of Aksu and Gencer [12]. The wind speed threshold (≥ 7 m/s) is clearly validated, showing near-zero SHAP contributions above this value, in contrast to the strongly negative values observed below it. Similarly, the capacity factor threshold demonstrates a critical transition around 35 – 40% , supporting feasibility requirements. However, the sea depth pattern reveals interesting insights: while the Boolean analysis specified an optimal range of 50 – 60 m, the SHAP analysis shows consistently negative contributions for deeper waters, with the least adverse effects near shallow depths. This pattern observed in the SHAP analysis suggests that the model has learned a preference for shallower installations, possibly reflecting recent technological advances in shallow-water foundations or economic considerations not fully captured in the rule-based assessment. The power density criterion shows near-zero to slightly positive contributions above 600 W/m^2 , in sharp contrast to the strongly negative values observed below 400 W/m^2 , confirming the importance of adequate

energy resource levels for site suitability.

Overall, the SHAP analysis demonstrates that while the data-driven model generally supports the rule-based inferences, it also reveals nuanced relationships and the potential regions where the Boolean thresholds might benefit from refinement based on the statistical patterns learned from the data.

4.2.2. Class-based SHAP contributions

The mean SHAP contribution values for each class are presented in Figure 10, revealing distinct patterns that explain the model's classification decisions:

- *For Class 0 (Not Suitable):* Sea Depth and Wind Speed (100m) are the two strongest contributors, both positive, indicating that deep waters and weak wind regimes are the primary factors that push locations into this class. Capacity Factor shows the strongest negative contribution, meaning that even in otherwise unsuitable areas, moderate capacity values partially counteract the Not Suitable prediction. Proximity to Extreme Weather Events contributes negatively, while Geopolitical Concerns contributes positively—the latter suggesting that geopolitically sensitive zones tend to overlap with technically unfavorable areas.
- *For Class 3 (Moderately Suitable):* Proximity to Grid Connection Point dominates with the highest positive contribution, followed by Proximity to Extreme Weather Events, indicating that these locations benefit from good infrastructure access and sheltered weather conditions. However, Wind Speed (100m), Power Density, and Proximity to Environmental Protection Areas all contribute negatively, revealing that favorable accessibility alone is insufficient when technical parameters are suboptimal and environmental constraints are present. Sea Depth and Proximity to Fishing Areas also exert negative effects, further limiting full suitability.
- *For Class 4 (Suitable):* Capacity Factor exhibits the strongest positive contribution, followed by Wind Speed (100m), confirming that high energy yield is the primary driver of the Suitable classification. Power Density and Proximity to Extreme Weather Events also contribute positively, reinforcing the role of strong technical resources and favorable weather conditions. Sea Depth, by contrast, remains the largest negative contributor even for suitable sites, reflecting the cost and engineering challenges associated with deeper installations. Proximity to International Borders contributes negatively, whereas Proximity to Military Zone Restrictions shows a slight positive effect, suggesting that suitable sites tend to be located away from restricted military areas.

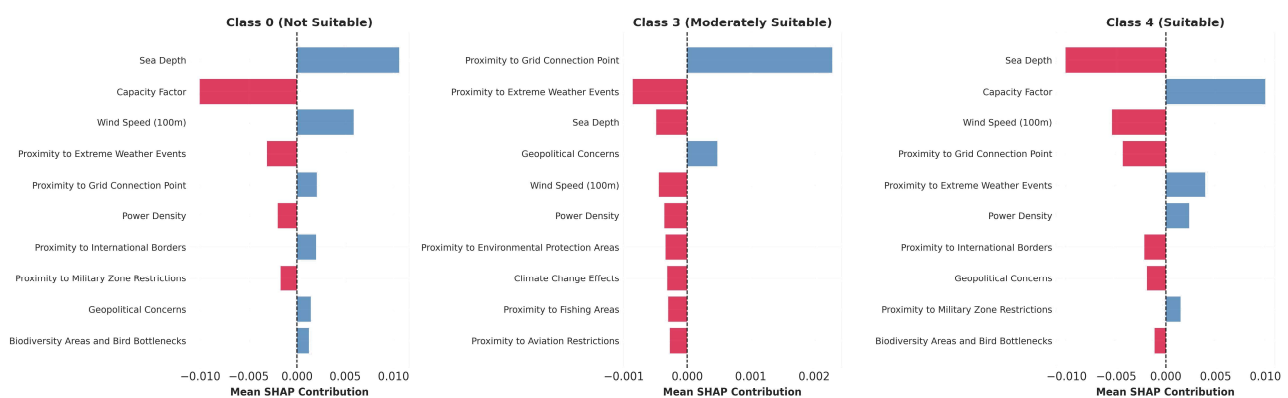


Figure 10. Class-specific SHAP contribution distributions for the top 10 criteria per suitability class

From a planning perspective, the distinct SHAP profiles of these three classes suggest that targeted policy interventions could be designed for each: Class 0 zones should be deprioritized under current technological conditions, Class 3 zones could become viable through infrastructure investment or environmental mitigation measures, and Class 4 zones may proceed to detailed feasibility studies. This multidimensional evaluation framework is essential for comprehensive OWPP site assessment [5].

4.3. Comparison of RF and SHAP method's feature importance

Figure 11 presents a comprehensive comparison between RF Gini importance and SHAP-based feature importance methods. The left panel displays the top 15 criteria ranked by RF importance, with corresponding SHAP values shown for direct comparison. Wind Speed (100m) dominates both methods, with importance values of approximately 0.187 (RF Gini) and 0.030 (mean |SHAP|), followed by Capacity Factor, with values around 0.172 (RF) and 0.024 (SHAP).

The right panel illustrates the ranking correlation between the two methods, revealing a strong Spearman rank correlation coefficient of $\rho = 0.963$ ($p < 0.001$), indicating excellent agreement between RF and SHAP in identifying important features. Most features cluster near the perfect agreement line, though some minor deviations are observed. Notably, "Proximity to Existing Major Ports" and "Proximity to Aviation Restrictions" show slightly different rankings between methods, with differences of approximately 3-4 positions.

The high correlation suggests that both methods consistently identify the same critical factors for offshore wind farm suitability, with technical parameters (wind resources and capacity) being paramount, followed by economic feasibility factors (sea depth, power density) and accessibility considerations (grid connection, extreme weather events). This agreement strengthens the reliability of the feature importance analysis and validates the model's interpretability.

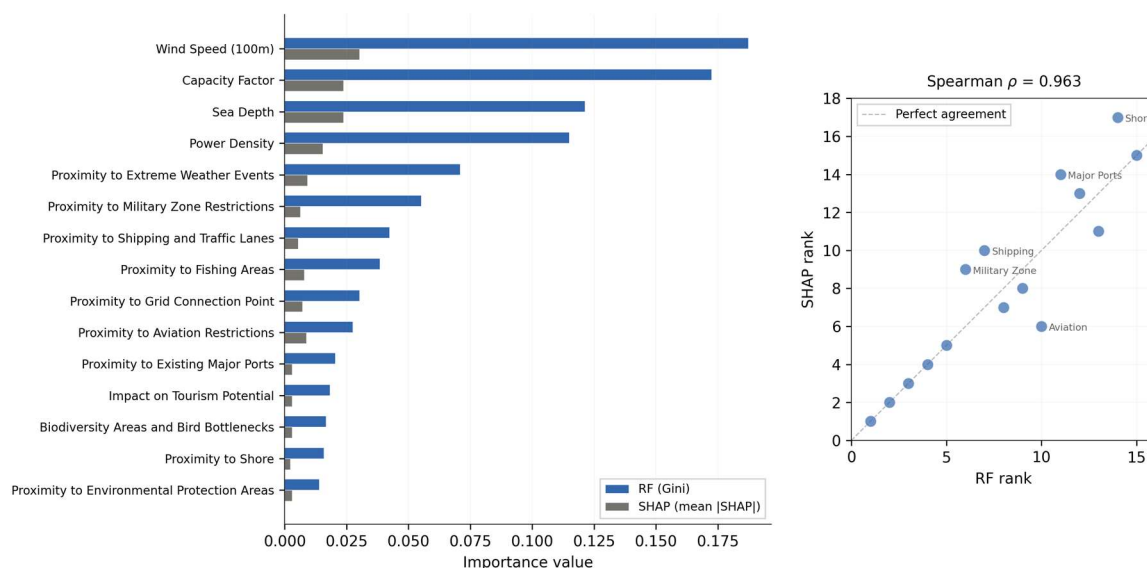


Figure 11. Comparison of RF (Gini) and SHAP feature importance values with ranking correlation analysis

4.4. Overall evaluation of findings

The findings demonstrate that the proposed GIS-ML-XAI framework provides a methodologically coherent, internally consistent framework for assessing OWPP suitability. The SHAP-based explanations aligned closely with the RF's feature-importance rankings and with the rule-based thresholds applied in the Boolean/MCDM stages (e.g., wind speed ≥ 7 m/s, depth between 50–600 m). This convergence indicates that the ML component successfully captured the underlying spatial logic of the earlier multi-criteria assessment.

A second observation concerns the consistency of SHAP behavior across suitability classes, despite the unbalanced class structure inherited from the original GIS-based map. Technical variables—particularly wind speed, capacity factor, and sea depth—showed stable and directionally coherent SHAP contributions, confirming their dominant influence on suitability patterns.

The ML model developed in this study was trained to reproduce the GIS-Boolean and MCDM-based suitability classification produced in the previous analysis [12]. Therefore, the model's predictions reflect the structure of the earlier rule-based assessment rather than an entirely independent measure of 'true' offshore wind suitability. This dependence

represents a methodological constraint, as the RF model learns the decision boundaries embedded in the prior study, including subjective elements such as threshold selections and weighting assumptions. However, this design was intentional, as the objective of the present study is not to redefine suitability criteria but to evaluate whether a data-driven ML algorithm can (i) generalize the rule-based patterns across a large geospatial domain, (ii) provide transparent explanations through SHAP values, and (iii) enable the development of an explainable decision-support tool. Although no independent validation dataset—such as operational OWPP locations—could be used due to the absence of commercial offshore wind farms in Türkiye, the model serves as a robust analytical framework for evaluating the internal consistency of the previous multi-criteria assessment. Future studies may incorporate empirical validation using realized OWPP deployments as they become available, or integrate additional physical, economic, and environmental datasets to generate independently verifiable suitability predictions.

5. Conclusion

This study developed a GIS–ML–XAI hybrid framework to identify suitable offshore wind zones within Türkiye's territorial waters. The RF model faithfully reproduced the GIS–MCDM suitability classes (99.74% accuracy), confirming its capacity to learn the spatial decision logic of the original rule-based assessment. More importantly, the SHAP analysis provided analytical insights that go beyond simple reproduction. First, while the previous Boolean analysis set a minimum capacity factor threshold of 30%, the SHAP dependence analysis revealed a more refined transition zone between 35% and 45%, suggesting that 35% may serve as a more empirically grounded threshold for positive suitability contributions. Second, the depth-related SHAP gradient showed a monotonic negative trend consistent with current foundation-type feasibility, an insight not explicitly captured in the rule-based framework. Third, the class-specific SHAP profiles revealed that the three suitability classes are driven by distinctly different combinations of criteria—technical dominance for Class 4, infrastructure accessibility for Class 3, and fundamental resource deficiency for Class 0—providing a level of diagnostic detail that traditional MCDM outputs do not offer.

By combining the spatial analytical capabilities of GIS, the pattern-recognition strengths of ML, and the interpretability of XAI, this study presents a transparent, data-driven decision-support framework for OWPP site selection. For policymakers and energy planners, the framework offers several practical advantages: SHAP-based explanations make site-selection criteria auditable and communicable to non-technical stakeholders, the class-specific profiles enable differentiated investment strategies (e.g., prioritizing Class 4 zones for immediate development while targeting Class 3 zones for enabling infrastructure), and the model's data-driven thresholds can inform the calibration of future regulatory standards for offshore wind licensing. Consequently, the study offers a scientifically grounded and accountable evaluation tool for offshore wind planning in Türkiye.

In future research, this framework may be extended by benchmarking performance against alternative ML algorithms (e.g., XGBoost, LightGBM, CNNs, GNNs). Incorporating temporal variability in wind regimes and ocean conditions could enable the production of dynamic suitability maps. Integrating additional economic and infrastructural parameters—such as turbine transport costs, grid connection distances, and construction logistics—would enable the development of a more comprehensive planning tool to support investment decisions.

Overall, this study demonstrates the strong potential of explainable AI-based models in guiding offshore wind energy planning in Türkiye. Since the proposed framework relies on universally available spatial data types (wind atlases, bathymetry, protected-area boundaries, infrastructure networks), it is transferable to other maritime regions facing similar site-selection challenges, such as the Mediterranean, Black Sea, or Southeast Asian coastlines. Moreover, the methodology is not limited to offshore wind; with appropriate criteria substitution, the same GIS–ML–XAI pipeline could support suitability assessments for other renewable energy technologies, including onshore wind, solar PV, wave energy, and floating solar installations.

Conflict of Interest Statement

The authors declare that there is no conflict of interest

Authorship Contribution Statement

E.Ö.A: Conceptualization, Methodology, Software, Formal Analysis, Visualization, Writing – Original Draft; **C.T.G:** Supervision, Writing – Review & Editing

References

- [1] H. Díaz and C. Guedes Soares, "Review of the current status, technology and future trends of offshore wind farms," *Ocean Engineering*, vol. 209, 107381, Aug. 2020. doi: 10.1016/j.oceaneng.2020.107381
- [2] Global Wind Energy Council (GWEC), "Global Offshore Wind Report 2024," Brussels, 2024. [Online]. Available: <https://www.gwec.net/reports/globaloffshorewindreport/2024>. [Accessed: Nov. 24, 2024].
- [3] D. G. Vagiona and M. Kamilakis, "Sustainable site selection for offshore wind farms in the South Aegean—Greece," *Sustainability*, vol. 10, no. 3, 749, Mar. 2018. doi: 10.3390/su10030749
- [4] A. Chaouachi, C. F. Covrig, and M. Ardelean, "Multi-criteria selection of offshore wind farms: Case study for Baltic States," *Energy Policy*, vol. 103, pp. 179-192, Apr. 2017. doi: 10.1016/j.enpol.2017.01.018
- [5] P. Gkeka-Serpetsidaki, G. Skiniti, S. Tournaki, and T. Tsoutsos, "A review of the sustainable siting of offshore wind farms," *Sustainability*, vol. 16, no. 14, 6036, Jul. 2024. doi: 10.3390/su16146036
- [6] I. C. Dimitriou, E. Sarmas, G. P. Trachanas, V. Marinakis, and H. Doukas, "Multi-criteria GIS-based offshore wind farm site selection: Case study in Greece," *Renewable and Sustainable Energy Reviews*, vol. 207, 114962, Jan. 2025. doi: 10.1016/j.rser.2024.114962
- [7] Y. Wu, J. Zhang, J. Yuan, S. Geng, and H. Zhang, "Study of decision framework of offshore wind power station site selection based on ELECTRE-III under intuitionistic fuzzy environment: A case of China," *Energy Conversion and Management*, vol. 113, pp. 66-81, Apr. 2016. doi: 10.1016/j.enconman.2016.01.020
- [8] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206-215, May 2019. doi: 10.1038/s42256-019-0048-x
- [9] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4765-4774.
- [10] V. Belle and I. Papantonis, "Principles and practice of explainable machine learning," *Frontiers in Big Data*, vol. 4, 688969, Jul. 2021. doi: 10.3389/fdata.2021.688969
- [11] W. Liao, J. Fang, L. Ye, B. Bak-Jensen, Z. Yang, and F. Porté-Agel, "Can we trust explainable artificial intelligence in wind power forecasting?" *Applied Energy*, vol. 376, 124273, Dec. 2024. doi: 10.1016/j.apenergy.2024.124273
- [12] E. Ö. Aksu and C. T. Gencer, "GIS-based optimum location selection for offshore wind energy: A case study for Turkey," *Ocean Engineering*, vol. 320, 120292, Mar. 2025. doi: 10.1016/j.oceaneng.2025.120292
- [13] A. Yaman, "A GIS-based multi-criteria decision-making approach (GIS-MCDM) for determination of the most appropriate site selection of onshore wind farm in Adana, Turkey," *Clean Technologies and Environmental Policy*, vol. 26, no. 12, pp. 4231-4254, May. 2024. doi: 10.1007/s10098-024-02866-3
- [14] A. Demir, A. E. Dinçer, C. Çiftçi, S. Gülçimen, N. Uzal, and K. Yılmaz, "Wind farm site selection using GIS-based multicriteria analysis with life cycle assessment integration," *Earth Science Informatics*, vol. 17, pp. 1591-1608, Jan. 2024. doi: 10.1007/s12145-024-01227-4
- [15] A. Bilgili, T. Arda, and B. Kilic, "Explainability in wind farm planning: A machine learning framework for automatic site selection of wind farms," *Energy Conversion and Management*, vol. 309, 118441, Jun. 2024. doi: 10.1016/j.enconman.2024.118441
- [16] J. Zhang, T. Zhang, Y. Li, X. Bai, and L. Chang, "Study on mining wind information for identifying potential offshore wind farms using deep learning," *Frontiers in Energy Research*, vol. 12, 1419549, Aug. 2024. doi: 10.3389/fenrg.2024.1419549

- [17] M. Masoumi, "Machine learning solutions for offshore wind farms: A review of applications and impacts," *Journal of Marine Science and Engineering*, vol. 11, no. 10, 1855, Sep. 2023. doi: 10.3390/jmse11101855
- [18] R. Machlev, L. Heistrene, M. Perl, K. Y. Levy, J. Belikov, S. Mannor, and Y. Levron, "Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities," *Energy and AI*, vol. 9, 100169, Aug. 2022. doi: 10.1016/j.egyai.2022.100169
- [19] A. M. Sazib, S. Al Farabi, F. Rayhan, and M. A. Karim, "Advancing renewable energy systems through explainable artificial intelligence: A comprehensive review and interdisciplinary framework," *Journal of Computer Science and Technology Studies*, vol. 7, no. 2, pp. 56-70, Apr. 2025. doi: 10.32996/jcsts.2025.7.2.5
- [20] M. R. Shadi, H. Mirshekali, H. R. Shaker, "Explainable artificial intelligence for energy systems maintenance: A review on concepts, current techniques, challenges, and prospects," *Renewable and Sustainable Energy Reviews*, vol. 216, 115668, Jul. 2025. doi: 10.1016/j.rser.2025.115668
- [21] T. Clement, H. T. T. Nguyen, N. Kemmerzell, M. Abdelaal, and D. Stjelja, "Coping with data distribution shifts: XAI-based adaptive learning with SHAP clustering for energy consumption prediction," in *AI 2023: Advances in Artificial Intelligence, Lecture Notes in Computer Science*, vol. 14472, T. Liu, G. Webb, L. Yue, and D. Wang, Eds. Singapore: Springer, 2024, pp. 147–159. doi: 10.1007/978-981-99-8391-9_12
- [22] M. Argin, V. Yerci, N. Erdogan, S. Kucuksari, and U. Cali, "Exploring the offshore wind energy potential of Turkey based on multi-criteria site selection," *Energy Strategy Reviews*, vol. 23, pp. 33-46, Jan. 2019. doi: 10.1016/j.esr.2018.12.005
- [23] E. Caceoğlu, H. K. Yildiz, E. Oğuz, N. Huvaj, and J. M. Guerrero, "Offshore wind power plant site selection using Analytical Hierarchy Process for Northwest Turkey," *Ocean Engineering*, vol. 252, 111178, May 2022. doi: 10.1016/j.oceaneng.2022.111178
- [24] E. Ö. Aksu, "Türkiye deniz üstü rüzgâr enerjisi santralleri için Coğrafi Bilgi Sistemleri destekli Boolean ve ağırlıklı analiz yöntemleriyle yer seçimi," Ph.D. dissertation, Dept. of Industrial Eng., Gazi Univ., Ankara, Türkiye, 2025.
- [25] World Bank, "Offshore Wind Roadmap for Türkiye," Washington, D.C., 2024. [Online]. Available: <http://documents.worldbank.org/curated/en/099110624015521050>. [Accessed: Oct. 3, 2025].
- [26] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, Oct. 2001. doi: 10.1023/A:1010933404324
- [27] T. Hengl, M. Nussbaum, M. N. Wright, G. B. M. Heuvelink, and B. Gräler, "Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables," *PeerJ*, vol. 6, e5518, Aug. 2018. doi: 10.7717/peerj.5518
- [28] S. Georganos, T. Grippa, A. Niang Gadiaga, C. Linard, M. Lennert, S. Vanhuyse, N. Mboga, E. Wolff, and S. Kalogirou, "Geographical random forests: A spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling," *Geocarto International*, vol. 36, no. 2, pp. 121-136, Jun. 2019. doi: 10.1080/10106049.2019.1595177
- [29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, Oct. 2011. [Online]. Available: <https://jmlr.org/papers/v12/pedregosa11a.html>
- [30] P. V. Sireesha and S. Thotakura, "Wind power prediction using optimized MLP-NN machine learning forecasting model," *Electrical Engineering*, vol. 106, pp. 7643–7666, May. 2024. doi: 10.1007/s00202-024-02440-6
- [31] Y. Sun, Y. Li, R. Wang, and R. Ma, "Modelling potential land suitability of large-scale wind energy development using explainable machine learning techniques: Applications for China, USA and EU," *Energy Conversion and Management*, vol. 302, 118131, Feb. 2024. doi: 10.1016/j.enconman.2024.118131

