

Research Paper

Explainability in wind farm planning: A machine learning framework for automatic site selection of wind farms



Atakan Bilgili, Tümay Arda, Batuhan Kilic*

Department of Geomatic Engineering, Yildiz Technical University, Istanbul 34220, Türkiye

ARTICLE INFO

Keywords:

Wind farm site selection
Machine learning
Explainable artificial intelligence
SHapley Additive exPlanations
Site suitability map

ABSTRACT

The process of wind farm site selection involves complex considerations spanning environmental, economic, and social factors, often handled through multi-criteria decision making (MCDM) methods. However, MCDM approaches face several challenges. To address these, integrating machine learning (ML) techniques offers promise. Nevertheless, existing studies have yet to employ a standalone ML-based approach for this purpose. This study seeks to pioneer an explainable ML-based framework for automated wind farm site selection. The framework incorporates feature selection techniques, seven ML algorithms, statistical tests, and a local explainability method. The experimental area was chosen as Balikesir province, which hosts 13% of Türkiye's installed wind energy capacity. According to the prediction performances of the ML models, Extreme Gradient Boosting-XGBoost has the highest accuracy (0.9607), followed by Light Gradient Boosting Machine-LightGBM (0.9580), Random Forest-RF (0.9518), Histogram-based Gradient Boosting-HGB (0.9387), Classification and Regression Trees-CART (0.8946), Logistic Regression-LR (0.8856), and Naive Bayes-NB (0.8456). Additionally, McNemar's tests revealed statistically significant differences among ML models. Based on the explanations provided by SHapley Additive exPlanations, wind speed, distance to transmission lines, distance to protected zones, and elevation were the top contributing criteria. Moreover, analysis revealed regions with elevated wind speeds, higher elevations, and closer proximity to transmission lines and protected areas as favorable for wind farm installation. The study's outcomes indicate strong agreement (97%) between current wind turbine locations and those identified as suitable by ML models. Within Balikesir province, our research has identified a significant number of highly suitable areas where wind turbines have not yet been installed. These findings underscore the critical importance of these locations for future investments in wind farm development. The proposed framework could be highly efficient for wind farm siting studies in provinces that share similar characteristics with the study area used in this study offering a robust tool for future wind farm siting studies. Finally, the study contributes to achieving UN SDG 7 by employing ML to refine wind farm site selection, thereby facilitating the transition towards sustainable and universally accessible energy.

1. Introduction

In today's era, the concept of energy assumes a crucial role in the societal and economic progress of nations, as well as in sustaining the existence of life for all species inhabiting the planet. According to the latest data released by the International Energy Agency (IEA), the global population without access to electricity is expected to increase by approximately 20 million individuals in 2022, bringing the total to 775 million people [1]. Furthermore, the demand for energy has also recently increased further because of the global recovery from the COVID-19 pandemic, geopolitical conflicts between the Russian

Federation and Ukraine that threaten gas supply, and extreme weather patterns caused by climate change [2].

In parallel with global industrialization, nations' excessive and prolonged use of fossil fuels brings environmental pollution and health risks. Currently, these limited sources fulfill over 80 % of the worldwide energy demand and are accountable for roughly 90 % of the CO₂ emissions associated with energy production [3,4]. According to prominent global energy agencies, there is a widespread consensus that renewable energy, particularly wind and solar technologies, will assume a dominant position in addressing the growing power demand in the near future. The analysis conducted by the IEA projects that renewable

* Corresponding author.

E-mail address: batuhank@yildiz.edu.tr (B. Kilic).

energy sources are expected to contribute to nearly 98 % of the additional 2,518 TWh of electricity generation that is anticipated to be integrated into the global energy portfolio from 2022 to 2025 [5]. In a nutshell, there is a need for an optimal and effective transition from fossil fuels to renewable energy sources in order to fulfill the growing demand for sustainable and clean energy [6].

Wind energy, one of the most utilized renewables, is regarded as a very captivating form of renewable energy owing to its economically efficient energy conversion technologies, vast availability of prospective sites, and continuous wind flow throughout the day. By the middle of the latest year, wind energy is expected to achieve the momentous milestone of 1 TW of installed capacity. According to the Global Wind Energy Council's report, China is expected to continue being the leader in onshore wind installations from 2023 to 2027, with a capacity of 300 GW. During the same period, it is projected that Europe will have approximately 100 GW [5]. For the efficient utilization of wind energy, it is required to conduct a comprehensive assessment of various factors, including, but not limited to, wind speed and air density. This is due to the non-uniform distribution of wind speed worldwide, and the accessibility of suitable sites varies significantly across different locations, encompassing both mechanical and electrical considerations [7]. On the other hand, there is an ongoing necessity to identify appropriate locations for wind turbine (WT) placement, which presents numerous challenges related to minimizing environmental, technical, and social constraints. To overcome these constraints, it is essential to adopt an optimal approach to the placement of wind farms, particularly by employing land suitability analysis.

The initial approaches to the problem of selecting suitable sites for wind farms involve the use of traditional methods for site suitability modeling or site screening. These methodologies depend on the subjective evaluation of professionals in order to determine the most suitable criteria and parameters, and the assessment encompasses a wide range of factors [8,9]. Thus, the incorporation of artificial weights and theoretical criteria into appropriateness models introduces ambiguity and intrinsic issues [6]. The increasing advancements in Geographic Information Systems (GISs), which provide important contributions as a decision support tool, have led to a rise in their popularity and applicability in addressing specific land suitability analyses, such as the problem of selecting suitable sites for wind farms [10]. Generally, the integration of multi-criteria decision making (MCDM) methods with GIS facilitates the amalgamation of geographical data, cost-benefit information, and decision makers' preferences. It also enables decision makers to effectively prioritize and rank several project choices that may possess contradictory and incompatible attributes [11]. Over the past fifteen years, a multitude of studies in various countries have been conducted to address and evaluate the wind farm siting problem through the use of MCDM-integrated GIS methodologies [12–22].

Within the domain of MCDM methodologies, evaluating land suitability involves quantifying the importance of individual factors and organizing them into a hierarchical structure based on their respective scores. Despite their extensive application in the selection of alternative sites for wind farms, these methods are accompanied by various inherent challenges. In the initial stage, it is important to ascertain the relevant criteria that impact the analysis of land suitability as well as to get accurate input data. The collection of geographical data at various scales will have a direct impact on the study. The method of allocating weights to various factors is inherently subjective and relies on the expertise of the individuals participating in the study [6]. The presence of biased comparisons among criteria will have a direct impact on their respective weights, resulting in a significant level of uncertainty in the resulting model. Moreover, the results of MCDM techniques are altered by the inclusion of several new candidates in the set of parameters influencing the selection of appropriate site locations. This implies that achieving a universally applicable level of performance is unfeasible.

One potential approach for examining conformity analysis as an alternative to MCDM procedures involves the utilization of artificial

intelligence methods that emulate the cognitive decision-making abilities exhibited by humans. Machine learning (ML) models have been extensively employed in the literature pertaining to wind energy for the purposes of wind energy forecasting, wind generation, and wind speed prediction [4]. Within the realm of wind farm site selection, Asadi et al. [7] employed a GIS-assisted modeling approach that relied on support vector regression to identify possible sites inside Iran. To date, to our knowledge, there has been a lack of research specifically examining the use of ML approaches in the process of wind farm siting. Therefore, the main objective of this study is to improve the accuracy and reliability of wind farm site suitability maps (SSMs) by utilizing ML algorithms. The Balikesir province situated in Türkiye was chosen as the empirical test area. The region is widely acknowledged for its significant capacity to effectively exploit wind energy resources. In addition, an additional objective of this study is to assess the efficacy of both traditional and state-of-the-art ML algorithms in the task of identifying appropriate locations for the establishment of wind farms.

To achieve these objectives, an ML-based framework was established in a systematic manner, employing seven ML models: Logistic Regression (LR), Naive Bayes (NB), Classification and Regression Trees (CART), Random Forest (RF), Histogram-based Gradient Boosting (HGB), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM). Initially, a comprehensive review of the pertinent literature was conducted to establish a set of 12 factors that influence the selection of wind farm sites. Subsequently, the significance of these criteria was determined by employing the Information Gain (IG) approach to calculate their respective importance scores. Furthermore, the authors employed multicollinearity and correlation tests to examine the associations between the criteria. In order to evaluate the effectiveness of machine learning models in the context of site selection, six measures were employed: accuracy, recall/sensitivity, specificity, precision, F1 score, and area under the curve (AUC). To assess the presence of any statistically significant differences among the models, Cochran's Q and McNemar's statistical tests were employed. Subsequently, the SHapley Additive exPlanations (SHAP) technique was employed for the purpose of interpreting the machine learning models. Finally, wind farm SSMs were constructed using the outputs of ML models.

2. Study area and data

The pursuit of wind energy, which is considered a renewable energy source in Türkiye, has been underway since 2006. Based on the findings of the Türkiye Wind Energy Potential Atlas, it is postulated that the establishment of a wind power facility with a capacity of 5 MW per square kilometer is feasible in viable regions characterized by an elevation of 50 m above ground level and an annual average wind velocity exceeding 7.5 m/s. Furthermore, the analysis reveals that the cumulative capacity of wind power installations that can be feasibly deployed throughout Türkiye amounts to 47,849.44 MW [23]. In late 2022, the proportion of wind power plants in the overall electricity generation stands at roughly 9.97 %. Notably, the western region of the Türkiye accounts for 75 % of this generation [24]. The province of Balikesir is situated in the northwestern portion of Türkiye and is recognized as an area characterized by a significant supply of renewable wind energy. The aforementioned geographical area constitutes around 10.8 % of the total number of operational wind power facilities in Türkiye, as well as 12.9 % of the overall installed capacity [25]. Furthermore, this area has a coastal boundary along both the Marmara and Aegean Seas, hence augmenting the potential for harnessing renewable wind energy. As a result, the Balikesir province was chosen as the study area (Fig. 1) to demonstrate our approach.

2.1. Wind farm inventory

One of the most crucial factors in selecting suitable sites for wind farms is the presence of existing wind turbines, installed by commercial

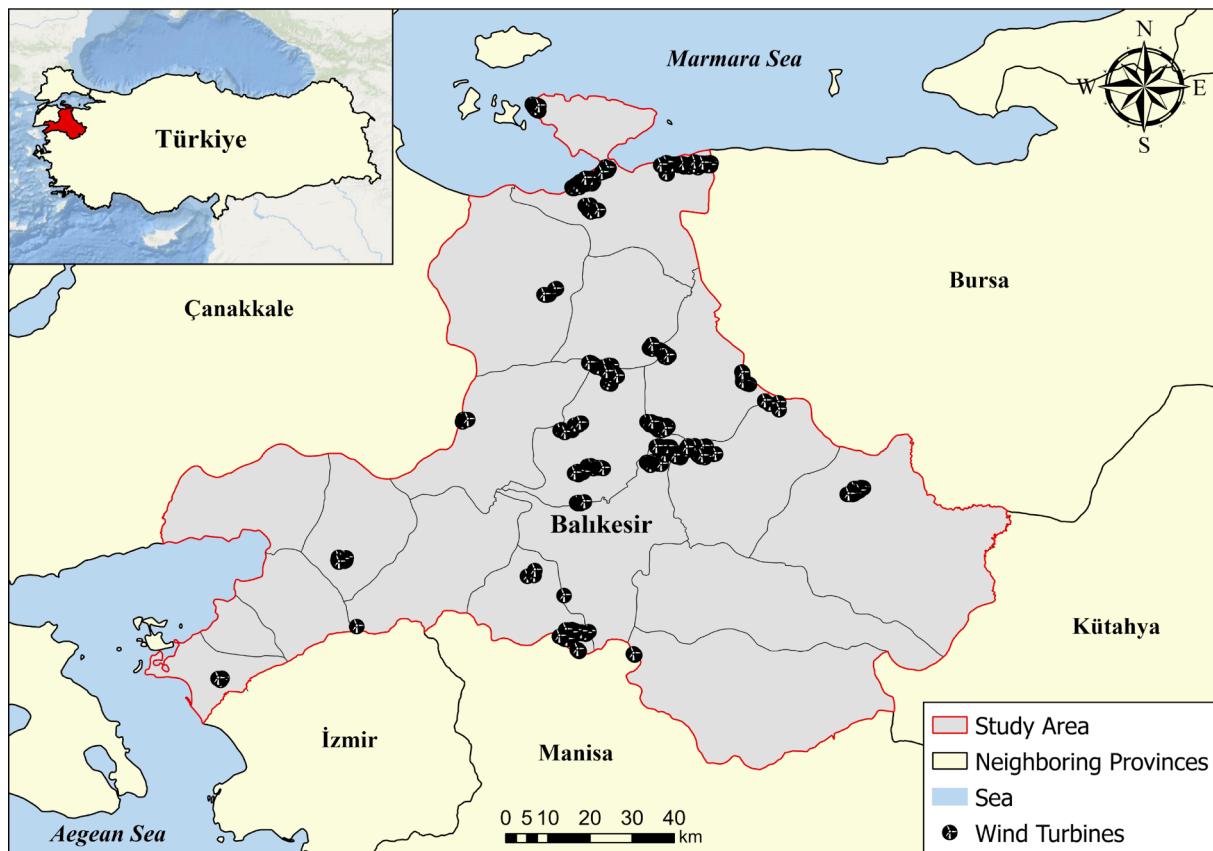


Fig. 1. Study area and available wind turbines.

and public corporations. These turbines provide essential ground truth data for machine learning models, thereby serving as key tools for future wind farm installations. Furthermore, they offer valuable insights into specific zones, as these areas have already been evaluated and deemed suitable by experts in the field. In this present work, the existing wind turbines were obtained from the Overpass turbo [26], a web-based data mining tool for OpenStreetMap. There were 454 available wind turbines scattered throughout the study area (Fig. 1). The existing wind farms were obtained in the Keyhole Markup Language (KML) format and further processed in a GIS environment to convert them into point (shapefile) format. However, the wind turbines occupy more space on land than just a single point. Considering the approximate rotor diameter of a wind turbine, we manually digitized the wind turbines in polygon format for a more fine-grained representation. Furthermore, the performance of the ML models largely depends on the quality and quantity of input data, which is crucial for uncovering vital relationships and optimizing performance [27]. Consequently, we obtained a total of 2418 pixels representing the existing wind farms (labeled as 1). On the other hand, ML models require non-suitable areas for wind farm installation to distinguish between suitable and non-suitable areas. This was done by excluding areas restricted by governmental regulations and those indicated in relevant literature. Subsequently, 2418 random sites, each represented by individual pixels, were selected as non-suitable areas (labeled as 0) from the remaining land.

2.2. Site selection criteria

Highlighting the potential candidate sites is a crucial pre-condition for wind farm siting. This process demands the determination of a variety of assessment and restriction factors that increase, decrease, or limit the suitability of a candidate site [14,20,28]. In this current work, twelve site selection criteria, namely wind speed, elevation, slope,

distance to fault lines, distance to landslide zones, distance to bird migration paths, distance to transmission lines, distance to main roads, distance to protected zones, distance to water bodies, distance to settlements, and distance to main roads, were identified through the relevant literature. The identified criteria were classified based on their environmental, economic, and social impacts on wind farm siting [3,12,14,15,22,29–32]. The categories of the criteria, data format, and resolution of each criterion are given in Table 1, and the relevant literature for the identified criteria is summarized in Table 2.

Restricted areas are sites where the installation of wind farms is restricted or prohibited due to environmental, social, technical, or regulatory factors. It is important to differentiate these areas from potential candidate sites for a proper installation process. Typically, the exclusion of restricted areas is achieved through the use of buffer zones. The size and range of the buffer zones for the related criteria were chosen considering the relevant literature (Table 3) and aggregated in a GIS environment to exclude these sites from the candidate sites (Fig. 2).

2.2.1. Environmental criteria

Environmental criteria refer to the factors that reflect the wind energy potential of an area and the safety of the installations in that area.

Wind Speed: Denotes the average speed of the wind over a specified time period. It is considered the most significant criterion for wind farm installations [19,20,22], has higher wind speeds allow for the generation of greater energy outputs from wind turbines [21]. In this study, the wind speed data was obtained from the Global Wind Atlas and further processed in a GIS environment (Fig. 3(a)).

Elevation: Refers to the height of the terrain where WTs are installed. Areas at higher elevations may capture more wind speed; therefore, WTs are often installed in these elevated areas. However, this can lead to increased installation and maintenance costs [21], so it should be carefully evaluated. In this study, the TanDEM-X Digital Elevation

Table 1
Sources of the site selection criteria.

ID	Category	Criterion	Format	Source
C1	Environmental	Wind Speed	Raster data/ 200 m	Global Wind Atlas [33]
C2		Elevation	Raster data/ 12 m	Türkiye General Directorate of Mapping [34]
C3		Slope	Raster data/12 m	Türkiye General Directorate of Mapping [34]
C4		Distance to Fault Zones	Vector data	Türkiye General Directorate of Mineral Research and Exploration [35]
C5		Distance to Landslide Zones	Vector data	Türkiye General Directorate of Mineral Research and Exploration
C6		Distance to Bird Migration Paths	Vector data	Birdmap- 5D Vision [36]
C7	Economic	Distance to Transmission Lines	Vector data	OpenStreetMap [26]
C8		Distance to Main Roads	Vector data	OpenStreetMap [26]
C9	Social	Distance to Protected Zones	Vector data	OpenStreetMap [26]
C10		Distance to Water Bodies	Raster data/10 m	ESA Worldcover 2021 V200 [37]
C11		Distance to Settlements	Raster data/10 m	ESA Worldcover 2021 V200 [37]
C12		Distance to Airports	Vector data	OpenStreetMap [26]

Model (DEM) [34] was used to generate an elevation map of the study area (Fig. 3(b)).

Slope: Denotes the steepness of the terrain where WTs are installed. It impacts wind flow as well as the costs of transportation, construction, and maintenance. Flat areas, characterized by lower slope values, are considered more suitable for siting WTs, as areas with higher slopes may introduce greater turbulence [28]. However, a certain degree of slope can induce the Venturi effect, which has the potential to increase wind flow speed [20]. In this work, the TanDEM-X DEM was utilized to create a slope map of the study area (Fig. 3(c)).

Distance to Fault Zones: Refers to the distance from a particular site

in the study area to any active fault. WTs should be installed at a safe distance from active fault lines due to the seismic risks they pose. These risks can potentially damage the WT infrastructure by causing terrain deformation, damaging the stability of access roads, and potentially increasing maintenance costs [29]. In this study, the active fault zones were obtained in vector format [35] and further processed in a GIS environment to create the distance to fault zones map (Fig. 3(d)).

Distance to Landslide Zones: It refers to the distance from a specific site within the study area to any active or historical landslide zones. For wind farm siting, maintaining a safe distance from these zones is crucial, as landslides occur on destabilized lands and carry significant safety concerns. Additionally, overlooking this factor could potentially lead to financial losses [29]. In this study, landslide zones were manually digitized using the landslide maps provided by the Türkiye General Directorate of Mineral Research and Exploration. These were then processed to create a landslide map for the study area, as shown in Fig. 3(e).

Distance to Bird Migration Paths: It refers to the distance from a specific site within the study area to the bird migration channels that pass through the area [42]. Considering these migration channels is

Table 3
Threshold values for the restricted sites.

Criterion	Threshold value	Reference
Wind Speed	< 4.4 m/s	Ayodele et al. [28]; Tercan [20]
Elevation	> 1500 m	Atıcı [46]
Slope	> 25 %	Latinopoulos and Kechagia [14]; Tercan [20]
Distance to Fault Lines	< 500 m	Díaz-Cuevas [17]; Moradi et al. [19]; Noorollahi et al. [15]; Tercan [20]
Distance to Landslide Zones	< 1000 m	Tercan [20]
Distance to Bird Migration Paths	< 500 m	Aydin et al. [10]; Ekiz et al. [48]
Distance to Transmission Lines	< 250 m	Gigović et al. [49]; Ayodele et al. [28]; Tercan [20]
Distance to Main Roads	< 150 m	Latinopoulos and Kechagia [14]; Tercan [20]
Distance to Protected Zones	< 1000 m	Díaz-Cuevas [17]; Tercan [20]
Distance to Water Bodies	< 500 m	Eroğlu [29]; Noorollahi et al. [15]
Distance to Settlements	< 550 m	Díaz-Cuevas [17]; Höfer et al. [45]
Distance to Airports	< 3000 m	Gigović et al. [49]; Tercan [20]; Zalhaf et al. [21]

Table 2
The relevant literature of the site selection criteria.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
Sotiropoulou et al. [38]	✓			✓				✓	✓	✓	✓	✓
Asadi et al. [7]	✓			✓				✓	✓	✓	✓	✓
Hoang et al. [22]	✓		✓					✓	✓	✓		✓
Yousefi et al. [39]	✓			✓	✓			✓	✓	✓		✓
Zalhaf et al. [21]	✓		✓	✓	✓			✓	✓	✓		✓
Asadi and Pourhossein [40]	✓			✓				✓	✓			✓
Eroğlu [29]	✓			✓	✓		✓	✓	✓	✓		✓
Saraswat et al. [41]	✓		✓	✓				✓	✓	✓	✓	✓
Tercan [20]	✓			✓	✓			✓	✓	✓	✓	✓
Moradi et al. [19]	✓		✓	✓	✓			✓	✓	✓	✓	✓
Xu et al. [42]	✓			✓			✓		✓	✓		✓
Ali et al. [3]	✓		✓	✓				✓	✓	✓		✓
Ayodele et al. [28]	✓			✓				✓	✓	✓	✓	✓
Díaz-Cuevas [17]	✓		✓	✓	✓			✓	✓	✓	✓	✓
Villacreses et al. [43]	✓		✓					✓	✓	✓		✓
Noorollahi et al. [15]	✓		✓	✓	✓			✓	✓	✓		✓
Sánchez-Lozano et al. [44]	✓			✓				✓	✓	✓		✓
Höfer et al. [45]	✓			✓			✓	✓	✓	✓		✓
Atıcı et al. [46]	✓		✓	✓	✓			✓	✓	✓		✓
Sánchez-Lozano et al. [47]	✓			✓				✓	✓	✓		✓

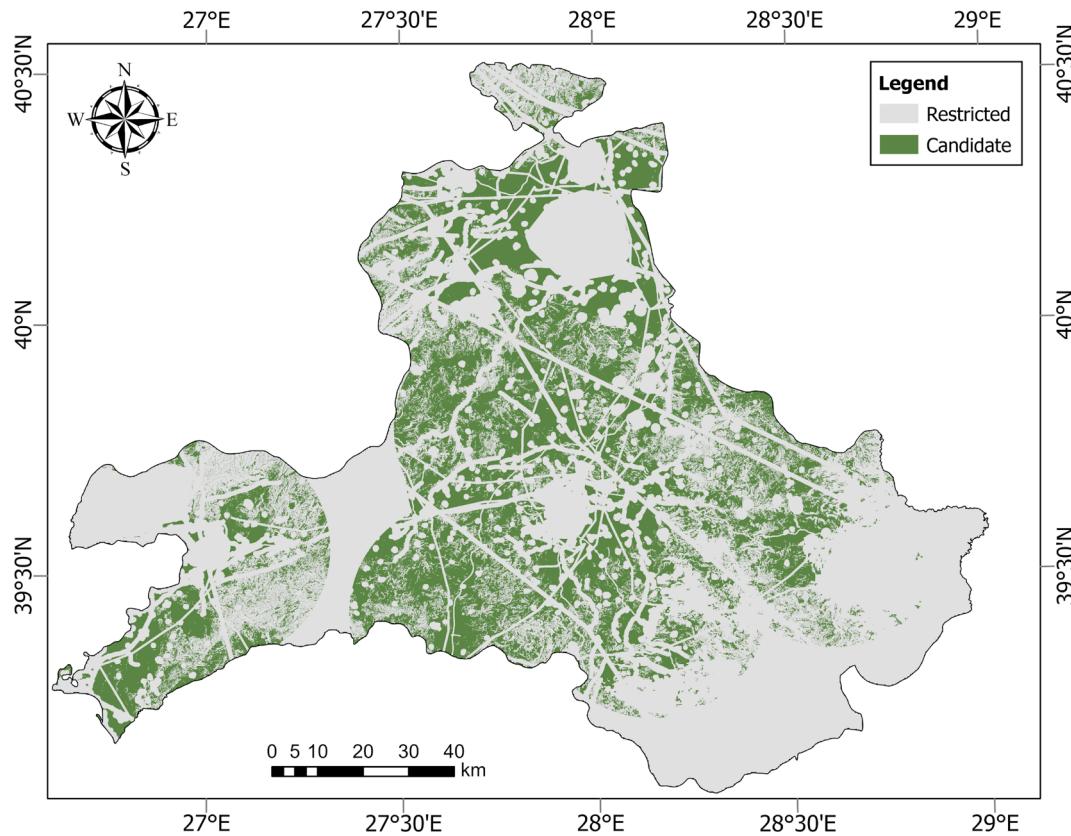


Fig. 2. Restricted and candidate sites in the study area.

crucial when installing WTs, as they may lead to bird collisions, resulting in bird mortality [29]. In this work, bird migration paths were obtained from the Birdmap 5D-Vision [36] and manually digitized in a GIS environment (Fig. 3(f)).

2.2.2. Economic criteria

Economic criteria are factors such as initial investment, potential return on investment, operating and maintenance costs that assess the financial viability and cost-effectiveness of installing a wind farm at a particular site.

Distance to Transmission Lines: The distance between a specific site and existing power or transmission lines in the study area is a crucial criterion widely adopted in studies focused on suitable site selection for WTs (see Table 2). The closer a WT is to transmission lines, the less power loss will occur [19]. Furthermore, wind farms located far from the transmission lines increase the cost of electricity transmission [20]. Additionally, extra connections may be required to transmit electricity generated from a distant wind farm [28,43]. The existing transmission lines were extracted from OpenStreetMap to create a distance to transmission lines map (Fig. 4(a)).

Distance to Main Roads: Refers to the distance between a specific site and major road networks in the study area. WTs located close to main roads facilitate maintenance and reduce transportation costs [19,20,49]. Moreover, the installation of WTs farther away may necessitate the construction of new roads, which would increase the total cost of the project. On the contrary, WTs should be installed at a safe distance from roads to avoid creating visual disturbances, loud noises, and electrical hazards [21,28]. The road network of the study area was extracted from OpenStreetMap and processed in a GIS environment (Fig. 4(b)).

2.2.3. Social criteria

Social criteria denote factors such as community acceptance, cultural

site preservation, and public safety that assess how a wind farm affects the local community and the cultural texture of a specific location.

Distance to Protected Zones: Refers to the distance between a particular location and protected zones such as national parks, wildlife protection zones, nature parks, important bird zones, and archeological sites. WTs should be installed at a certain distance from these protected zones to avoid harming nature through noise and visual disturbance [20,22]. In this study, protected zones were identified using data from OpenStreetMap and by communicating with local municipalities, as well as the Türkiye Ministry of Environment, Urbanization, and Climate Change. The final map showing protected zones is presented in Fig. 5(a).

Distance to Water Bodies: Denotes the distance between a site in the study area and water bodies, including rivers, lakes, ponds, dams, and wetlands. WTs should not be installed too close to these water bodies to preserve the natural balance and protect animal habitats [45]. Additionally, installing WTs close to water bodies may obstruct the waterways, and the cost of the construction may rise [28]. In this work, the water bodies were obtained from ESA Worldcover 2021 V200 [37]. The distance to water bodies map of the study area is given in Fig. 5(b).

Distance to Settlements: Denotes the proximity between WTs and existing settlements in the study area, which include cities, countries, villages, and their associated infrastructure like buildings, parks, and recreation areas. This criterion is crucial, as WTs installed too close to settlements can negatively impact urban aesthetics [28] and decrease real estate values [20]. Additionally, noise emissions from WTs can disturb both the local community and wildlife. Moreover, the blades of WTs can create shadow flickering and a glow effect, leading to visual disturbances [21]. The settlements in the study area were obtained from ESA Worldcover 2021 V200 [37], and the final map for the distance to settlements in the study area is shown in Fig. 5(c).

Distance to Airports: Expresses the distance between a site in study area and existing airports. For safer airport transportation, WTs should be located farther away from the airports, as they can disturb the

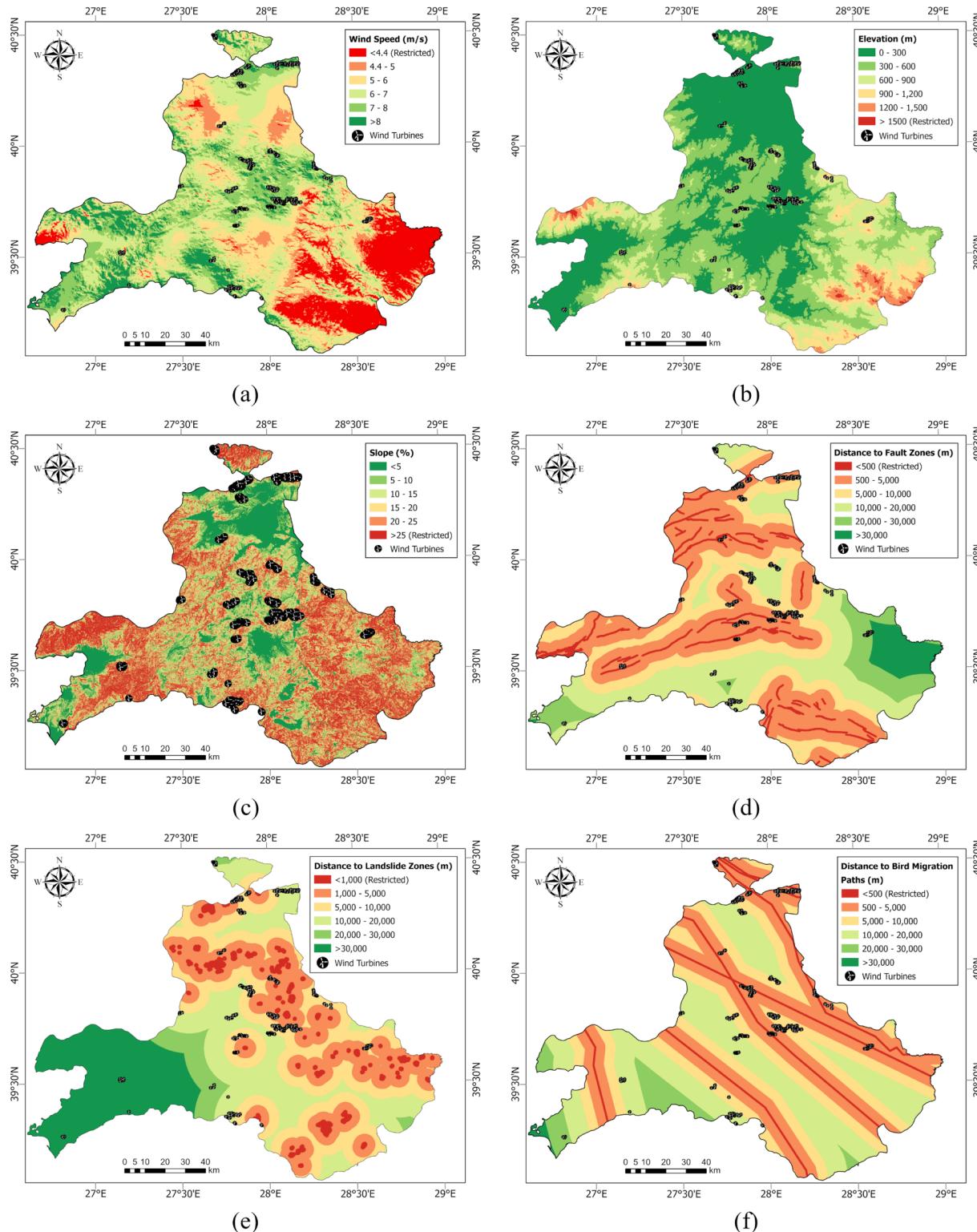


Fig. 3. Environmental criteria maps of the study area: (a) wind speed, (b) elevation, (c) slope, (d) distance to fault zones, (e) distance to landslide zones, (f) distance to bird migration paths.

aviation routes, airport surveillance radar signals, and communication [3,20,21]. The airports in the study area were extracted from OpenStreetMap and processed in a GIS environment. Fig. 5(d) presents the distance to airports map of the study area.

After initial processing, each criterion was rasterized to a 30 m spatial resolution. The pixel values corresponding to each criterion were then aggregated with the wind farm inventory (as described in Section

2.1), to create a data frame for wind farms suitable for use in ML models.

3. Methodology

In this paper, the problem of suitable site selection for wind farm installation is formulated as follows:

Problem: Given candidate sites for wind farm installation $S = \{s_1, s_2, \dots\}$,

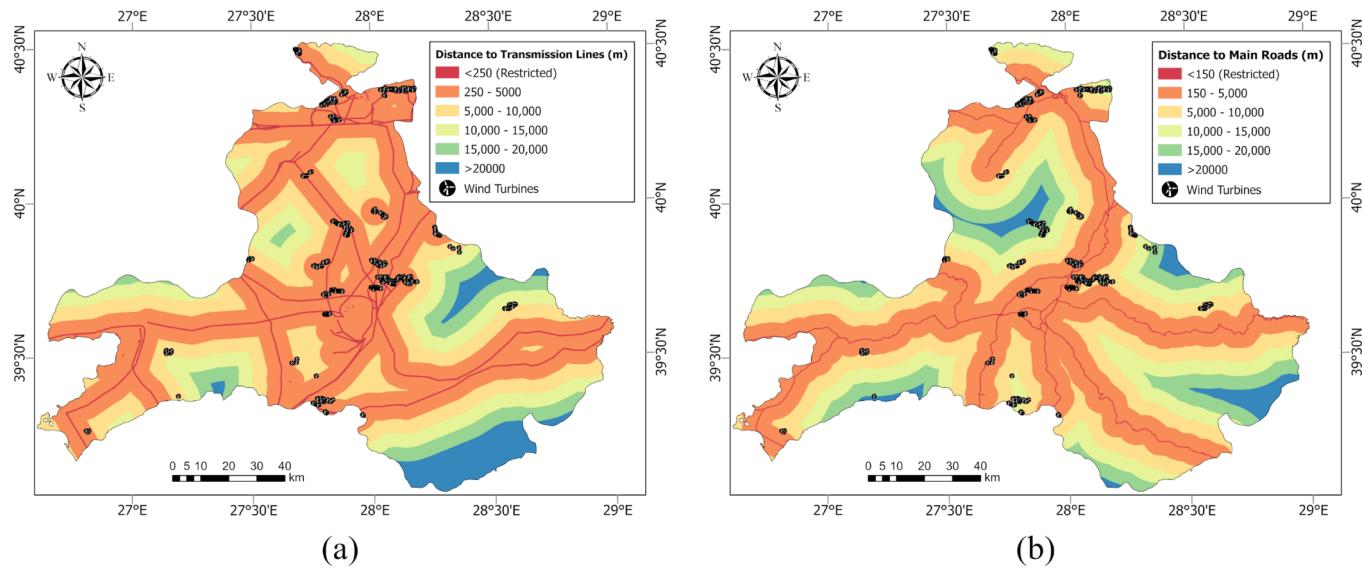


Fig. 4. Economic criteria maps of the study area: (a) distance to transmission lines (b) distance to main roads.

$\dots, s_n \}$, while each site candidate is defined by a vector of site selection criteria $s_i = \{s_i^1, s_i^2, \dots, s_i^n\}$ where s_i^n is a site selection criterion. A soft binary classification returns a probability score, where the probability of a site candidate being suitable is in the range $0 \leq P(s_i) \leq 1$ indicating the degree of confidence that a given site candidate is suitable. The higher the probability score, the more likely the site is to be suitable for wind farm installation. Conceptualizing the automatic selection of suitable site selection for wind farm installation as a soft binary classification problem allows for a more fine-grained representation of site suitability, addressing a significant limitation of the standard binary classification framework. This approach facilitates finer-grained differentiation among sites based on their suitability criteria. This approach involves using probability scores to give suitability levels to available sites as 'very highly suitable,' 'highly suitable,' 'moderately suitable,' 'lowly suitable,' and 'very lowly suitable.'

The study consists of five main phases, which are as follows: (1) data collection which involves collecting data regarding site selection criteria and existing wind turbines to serve as ground truth data, and wind farm dataset generation involves creating a data frame that contains samples of sites, represented by their corresponding site selection criteria and labeled as either suitable or non-suitable sites, and the splitting of the dataset into training and test subsets, (2) feature selection that utilizes correlation and multicollinearity tests to check for interrelations between the features (i.e., site selection criteria) and IG method to calculate feature importance scores, ensuring that there are no unrelated features present in the dataset, (3) model training for seven models (LR, NB, DT, RF, HGB, XGBoost, and LightGBM) and hyperparameter tuning for the models, and model evaluation through accuracy assessment metrics and model comparisons for significant differences with statistical tests, (4) generation of the SSMs for seven ML models, and (5) interpretation of the ML models through SHAP method. An overview of the proposed methodology is given in Fig. 6.

3.1. Investigation of the site selection criteria

The effectiveness of the generated site suitability levels via ML models is strongly influenced by the accurate determination of the contributing factors (i.e., site selection criteria) to the automatic selection of suitable sites. However, the current literature does not provide a framework to select relevant criteria before the initial model training. Hence, training models with unrelated features may lower the accuracy of a model, which hinders the usability of generated SSMs. Feature se-

lection techniques can be used to solve this issue caused by redundant or irrelevant features, increasing the model's overall performance by reducing the noise and enhancing its capacity to generalize to new data. In this study, we have utilized multicollinearity tests and IG techniques to evaluate the contributions of the criteria to suitable site selection for wind farms. A multicollinearity test detects if any interrelated measures exist in the dataset. The multicollinearity was detected using the metrics variance inflation factor (VIF) (Eq. (1)) and tolerance (TOL). Typically, it indicates that multicollinearity occurs when the VIF is greater than 10 or the TOL is less than 0.1. If multicollinearity occurred, the Pearson correlation coefficient was used to detect interrelated site selection criteria (Eq. (2)). Generally, the two variables are considered correlated if the Pearson correlation coefficient is greater than 0.7 or lower than -0.7. The closer the value is to 1, the more strongly positive the correlation is between the two site selection criteria. Conversely, values closer to -1 suggest a strong negative correlation. The criteria are considered independent if the coefficient values are close to 0.

$$VIF_i = \frac{1}{1 - R_i^2} \quad (1)$$

where VIF_i refers to Variance Inflation Factor for the i -th site selection criterion, R_i^2 is the coefficient of determination of a regression model.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

where r is the Pearson correlation coefficient, x_i, y_i are the i -th samples for a site selection criterion pair, and \bar{x}, \bar{y} are the means of a site selection criterion pair.

The IG technique aims to detect the most informative features in a dataset [50]. It measures the entropy loss caused by a feature when predicting a target class. A higher IG value indicates that a site selection criterion has a stronger predictive impact on the target variable. The IG value for criteria C_i and suitability class Y can be formulated as follows:

$$IG(Y; C_i) = H(Y) - H(Y|C_i) \quad (3)$$

where $H(Y)$ denotes the entropy of the suitability label Y and $H(Y|C_i)$ is the conditional entropy for Y given C_i .

3.2. Machine learning models

In this study, seven ML models (LR, NB, DT, RF, HGB, XGBoost, and

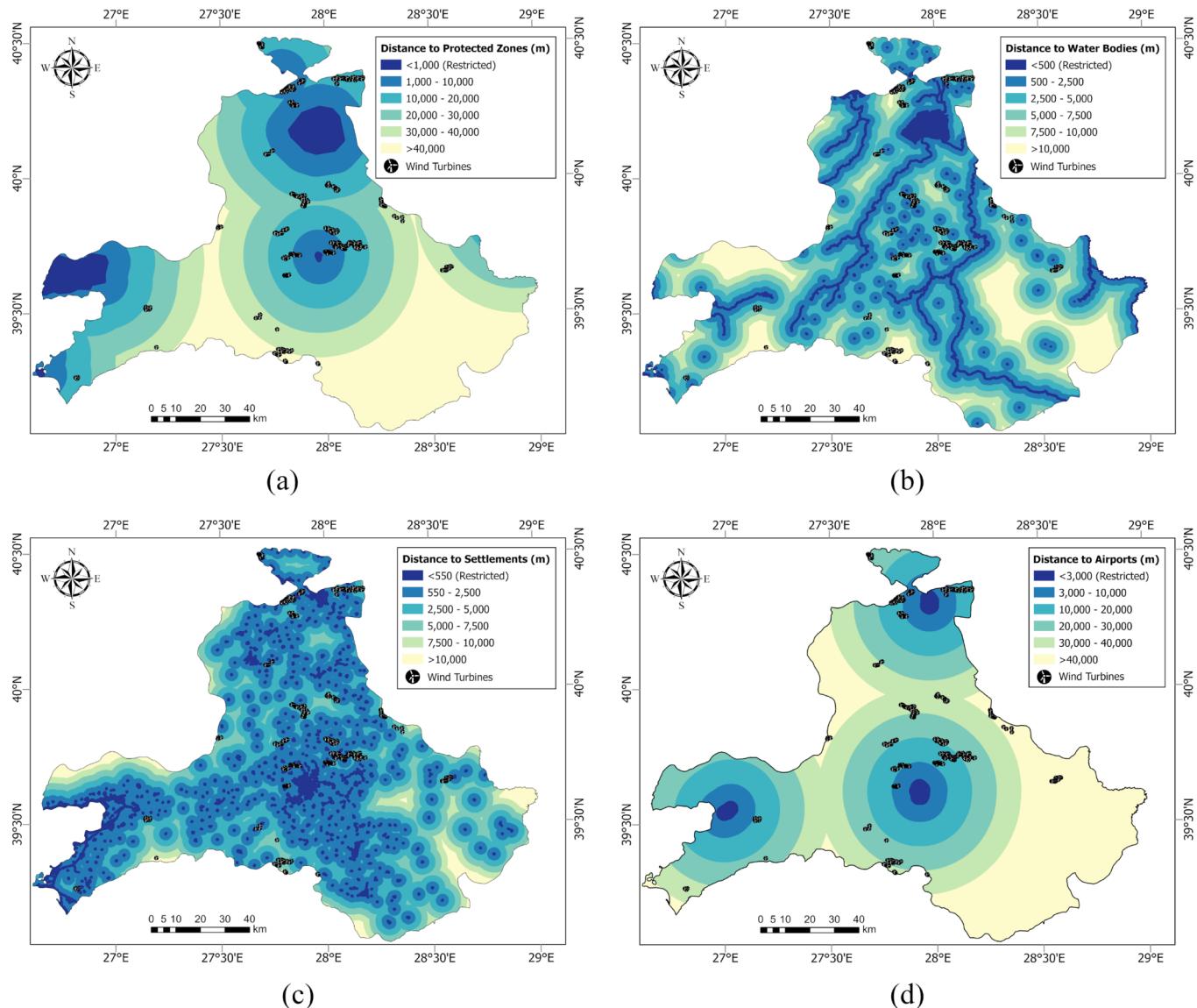


Fig. 5. Social criteria maps of the study area: (a) distance to protected zones, (b) distance to water bodies, (c) distance to settlements, (d) distance to airports.

LightGBM) were utilized to evaluate their predictive performances for the automatic selection of suitable sites for wind farms. A two-way holdout split was used to divide the dataset into train and test subsets. The classifiers were trained with the randomly selected 70 % of the available data, while the remaining 30 % was used for validation. Before training the data, an ML model's hyperparameters need to be tuned for the specific problem. In the present work, the Grid Search Cross Validation (GridSearchCV) method was utilized, which uses a set of hyperparameters to find the best combination through k-fold cross-validation on the training data. After the tuning process, the best combinations of hyperparameters for each classifier were used to generate a final classifier. All operations regarding data splits, model training, hyperparameter tuning, and performance evaluations were performed utilizing the Python's scikit-learn library [51]. The following subsections explain the working principles of the ML classifiers that were employed in the study.

3.2.1. Logistic regression (LR)

The LR is a widely adopted statistical approach that is a supervised classification algorithm. The model utilizes an s-shaped logistic function to model the relationship between independent variables (i.e., site selection criteria) and the dependent variable (i.e., suitability labels). The

logistic function converts the values from independent variables to a probability value between 0–1. During the training, the model updates the weights of each measure to minimize the difference between the predicted probabilities and the actual labels using a loss function. Despite the model's simplicity and effectiveness, it is affected by multicollinearity, which requires a careful examination of the independent features [52,53].

3.2.2. Naive Bayes (NB)

The NB is a popular machine learning algorithm that is based on the Bayes theorem. The algorithm calculates the probability of a particular class label given a set of features for each sample in the dataset. The algorithm makes the assumption that the features used for classification are conditionally independent. Due to this assumption, the calculation of class probabilities can be simplified significantly. Hence, the algorithm is efficient and easy to implement [54].

3.2.3. Classification and regression trees (CART)

The CART are a type of decision tree algorithm that can be used for both classification and regression. The algorithm is commonly utilized in various tasks due to its simple and intuitive nature. The goal of the CART is to split the given dataset into as many homogenous subsets as

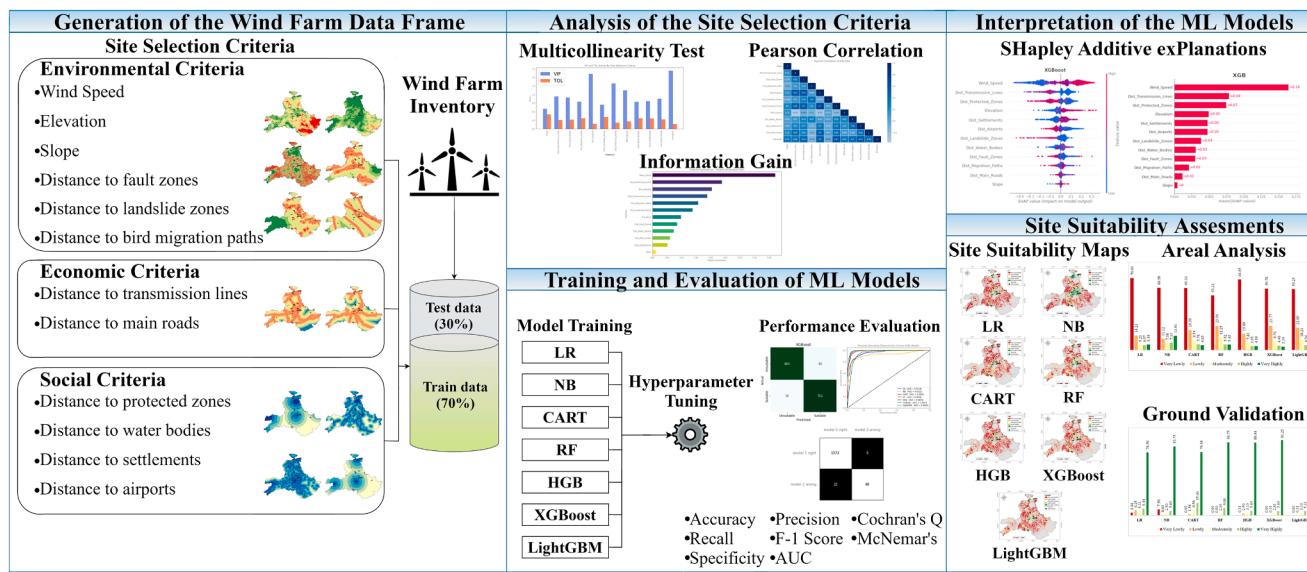


Fig. 6. Overall workflow of the study.

possible, considering the target classes. The CART begins the splitting process by considering the entire training dataset and making binary splits to split each node into two child nodes. It uses Gini Impurity to choose the best splits for each node in the tree. The splitting process continues for each child node until a predefined criterion is met, such as no further decrease in Gini Impurity or reaching the maximum depth of the tree. The resulting end points of the branches are called leaves, each representing the most frequent class label within that leaf. After the tree building process, a new instance is classified by passing its input features into the tree, resulting in a class label [55].

3.2.4. Random Forest (RF)

The RF is a powerful ensemble learning method that aggregates many individual decision trees to make its predictions. Hence, the RF can reduce the overfitting issue of individual decision trees. RF is a popular and widely used algorithm in the literature due to its ability to resist overfitting and noise, achieve high accuracy, and handle missing values without requiring imputation. In addition, it can perform multiple tasks simultaneously, such as feature selection, regression, and classification. The RF follows a process called bootstrap aggregation or bagging to make its predictions. Through bootstrapping and feature sampling, each tree in the RF is built using a different subset of data and features. Specifically, the model randomly samples the data with replacement to create different subsets for each tree and selects a random subset of features for each split in the tree-building process. The samples that are not in the bootstrapped dataset are called out-of-bag samples, and they are used to estimate the performance of the model. Then, to make a prediction, the model runs the new data through all decision trees and aggregates the individual predictions through a majority voting process where the class label that receives the most votes is used to create the final predictions [56].

3.2.5. Histogram-based Gradient Boosting (HGB)

Similar to its predecessor, Gradient Boosting Machines (GBM) [57], HGB builds an ensemble of decision trees sequentially, combining the predictions of each weak learner to form a stronger predictor. The algorithm trains a sequence of weak learners, with each new tree focusing on correcting the residual errors made by the ensemble thus far. The key difference in HGB is the way the algorithm handles continuous features. In the splitting process, HGB creates histograms of each input feature as discrete bins instead of considering each unique value of a feature. This process reduces the number of splits to be considered by the algorithm,

which increases the speed of the training process. For each iteration, the best split is determined by considering the reduction in the overall loss function. After the tree-building process, the algorithm combines the outputs from each tree to predict a final label for a new instance [58].

3.2.6. Extreme Gradient Boosting (XGBoost)

The XGBoost is a very popular boosting-based ensemble machine learning algorithm that can be used for multiple tasks such as classification, regression, and ranking that is based on GBM [57]. Like GBM, it iteratively trains a sequence of weak learners, where each weak learner is created based on the mistakes of the previous one (i.e., boosting). One major advantage of XGBoost over GBM is the normalization of the loss function. It uses a second-order approximation of the loss function, which can improve the overall accuracy of the model and help to converge the optimal solution faster. Other improvements over GBM include parallel processing and distributed computing, making it faster and more scalable [59].

3.2.7. Light Gradient Boosting Machine (LightGBM)

The LightGBM is another popular gradient boosting-based ensemble machine learning algorithm that originates from GBM. It can also be used for tasks such as classification, regression, and ranking, similar to XGBoost. LightGBM specifically gained interest due to its performance and efficiency on large and high-dimensional datasets. The key differences of LightGBM over GBM are its histogram-based technique for finding optimal splits, Exclusive Feature Bundling to reduce the number of features for training, Gradient-based One-Side Sampling that focuses on instances with larger gradients while randomly sampling instances with small gradients, and leaf-wise growth instead of level-wise, choosing the leaf that minimizes the loss during tree growth [60].

3.3. Model evaluation

Evaluating the predictive performance of an ML model is a crucial part of the site selection process. In the literature, there are several metrics to evaluate the predictive performance of an ML classifier. In this study, six well-established evaluation metrics, namely accuracy, recall/sensitivity, specificity, precision, F1 score, and area under the curve (AUC), were utilized to evaluate the performances of seven ML models explained in Section 3.2. The evaluation metrics of a classifier are typically computed through a confusion matrix that consists of four key components: true positive (TP), true negative (TN), false positive

(FP), and false negative (FN). TP denotes the number of sites correctly classified as suitable, while TN represents the number of sites correctly classified as unsuitable. FP refers to the number of sites incorrectly classified as unsuitable, and FN indicates the number of sites incorrectly classified as suitable. The formulas to compute six evaluation metrics are as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$\text{Recall/Sensitivity} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (6)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{F1 Score} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad (8)$$

$$\text{AUC} = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \quad (9)$$

To analyze the performances of ML classifiers in greater detail, it is necessary to conduct statistical significance tests in addition to accuracy metrics [50]. Hence, it can be analyzed whether there are significant differences between the predicted outputs of seven ML classifiers. In the present work, Cochran's Q test was used as an omnibus test to detect any statistically significant difference between ML models. Then, on the basis of the rejection of the null hypothesis, pairwise McNemar's test [61] was applied as a post-hoc test to detect statistically significant differences between ML models.

The McNemar's test, which is based on a 2×2 contingency table, utilizes Chi-square (χ^2) statistics to compare two ML models for statistically significant differences. The test with continuity correction [62] can be formulated as follows:

$$\chi^2 = \frac{(|n_{ij} - n_{ji}| - 1)}{n_{ij} + n_{ji}} \quad (10)$$

where n_{ij} refers to the number of suitable sites misclassified by the model i , but not by model j , while n_{ji} represents the number of suitable sites misclassified by the model j , but not by model i . If the estimated value of the χ^2 statistics exceeds a given threshold such as 3.84 at 95 % confidence interval, it can be concluded that the difference in predictive performance of two ML models is statistically significant.

3.4. Model interpretation

In the field of explainable artificial intelligence (XAI), the ability to provide explanations for machine learning models is a crucial aspect. Permutation feature importance and IG are among the global approaches used in XAI. These methods help to provide an overall understanding of how the features contribute to the model's predictions. On the other hand, local explainability is another approach in XAI that focuses on identifying the specific features that have a greater impact on individual predictions and how they influence those predictions [63]. One of the state-of-art techniques for local explainability is SHAP. In this study, the SHAP method was utilized to interpret the ML models proposed in Section 3.2 by using the Python-based SHAP library [64].

SHAP has the capacity to determine if a factor has a positive or negative impact on the performance of the model, in contrast to other machine learning algorithms that only provide feature importance scores. The SHAP method's ability to generate SHAP values for each sample in a dataset, enabling both local and global model interpretation, is one of its distinctive characteristics. This feature enables SHAP to

provide valuable insights into how individual data points affect the model's output and how the model performs overall [50].

The SHAP method is based on cooperative game theory [65]. The SHAP values computed by the method are a way to fairly distribute the total payout in a game among its participants. In the case of ML, the predictions of an ML model represent the total payout in the game, and the players of the model represent the features of a model (i.e., site selection criteria) who cooperate to gain this payout. This cooperation is quantified based on the weighted average of the marginal contribution to this payout. The SHAP values of a feature (i.e., site selection criterion) can be formulated as follows:

$$\phi_i(F, v) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)}{|F|!} [v(S \cup \{i\}) - v(S)] \# \quad (11)$$

where F is the set of all features, i is an element of F , S is a subset of F without i , v is the value function that maps elements of S to a number, $\phi_i(F, v)$ is the SHAP value of feature i , $|S|$ is the number of features in set S , $|F|$ is the total number of features, $v(S \cup \{i\})$ is the model prediction with the features in set S plus feature i , and $v(S)$ is the model prediction with the features in set S .

Finally, the mean absolute SHAP values ($|SHAP|$) which gives the overall impact of a feature on the model predictions are calculated by Eq. (12).

$$|SHAP|_i = \frac{1}{N} \sum_{j=1}^N |\phi_{ij}| \quad (12)$$

where N is the total number of data samples, $|\phi_{ij}|$ is the absolute SHAP value of feature i for the j -th data sample.

3.5. Generation of site suitability maps

SSMs are the maps of the study area that express restricted zones and the degree of suitability of a given site to install a WT based on the associated site selection criteria. In this current work, the SSMs were generated using the predictions of each ML model. By tradition, a binary classifier outputs an integer value of 0 or 1 as a prediction. However, this violates the fact that each individual site has a degree of suitability for WT installation. Therefore, this study employed a soft binary classification approach, which produces a score ranging between 0 and 1 (Eq. (13)), to more accurately reflect the varying degrees of site suitability. Subsequently, these scores were categorized into five classes to represent different degrees of suitability, utilizing the Jenks Natural Breaks method, which aims to form classes with minimal variance within each class and maximal variance between classes.

$$0 \leq P(s_i) \leq 1 \quad (13)$$

where $P(s_i)$ denotes the site suitability score of a site s_i .

After generating SSMs, the spatial distribution of each suitability class was examined. Ground validation involved verifying the distribution of existing wind farms among these classes for each ML model. This process aimed to gain further insights into the outputs of the models.

4. Results

4.1. Analysis of site selection criteria

Evaluating the importance of each criterion is a fundamental stage in the process of selecting suitable sites for wind farms. The IG ratio was calculated to assess the feature importance of each site selection criterion in the wind farm data frame, as shown in Fig. 7. According to the results, the wind speed is the most important site selection criterion (0.421), followed by the distance to protected zones (0.239), distance to airports (0.230), distance to transmission lines (0.190), distance to bird migration paths (0.154), distance to landslide zones (0.098), distance to

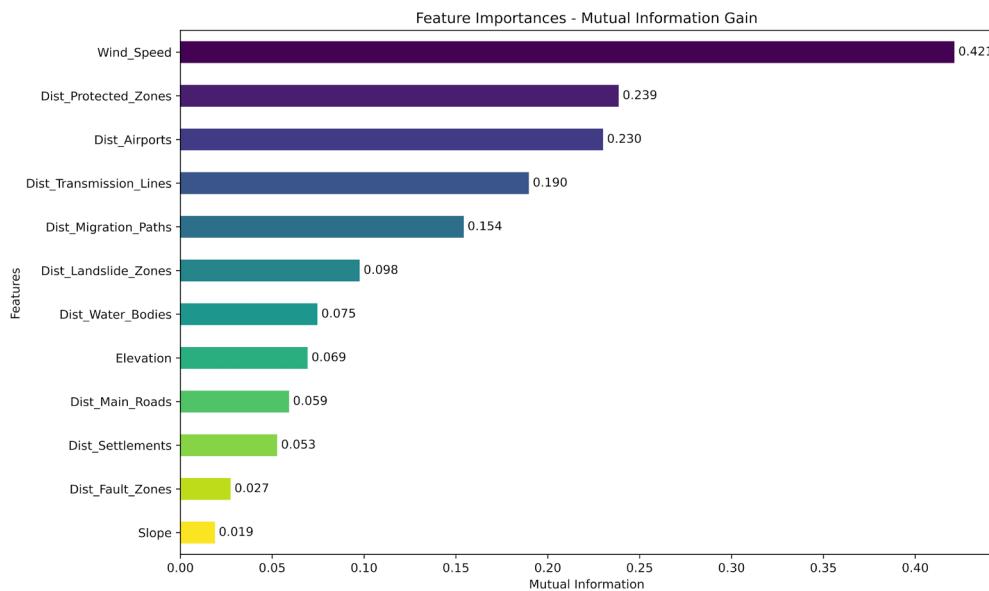


Fig. 7. Importance scores of site selection criteria.

water bodies (0.075), elevation (0.069), distance to main roads (0.059), distance to settlements (0.053), and distance to fault zones (0.027). On the other hand, the slope was found to be the least important criterion, with a score of 0.019.

For a deeper investigation of the association between site selection criteria, a multicollinearity test was conducted. According to the results of the multicollinearity test, the highest VIF was 3.401, while the least TOL value was 0.294, as presented in [Table 4](#). Since all VIF values are less than 10 and TOL values are greater than 0.1, no site selection criteria can be considered collinear with each other.

Pairwise Pearson correlation coefficients were also computed to further validate the results of the multicollinearity test. Typically, values exceeding the threshold values of 0.7 and -0.7 indicate a multicollinearity problem in the data frame. The results of the correlation tests are presented in [Fig. 8](#) as a correlation matrix. Based on these results, none of the correlation coefficients exceed the threshold value of 0.7. The highest coefficient value, computed at 0.7, is between the distance to protected zones and the distance to airports. Both multicollinearity and correlation tests suggest that there is no need to eliminate any site selection criterion from the wind farm data frame.

4.2. Prediction performances of ML models

A comprehensive predictive performance assessment was conducted in this study utilizing traditional ML algorithms such as LR, NB, and CART and some state-of-the-art algorithms such as HGB, XGBoost, and LightGBM. The confusion matrices of each ML model are shown in

Table 4
Results of the multicollinearity test.

Site Selection Criteria	VIF	TOL
Slope	1.172	0.853
Dist_Transmission_Lines	1.895	0.528
Dist_Fault_Zones	1.831	0.546
Dist_Migration_Paths	1.586	0.631
Dist_Airports	3.215	0.311
Dist_Landslide_Zones	1.416	0.706
Dist_Protected_Zones	2.656	0.376
Wind_Speed	2.245	0.446
Dist_Water_Bodies	1.583	0.632
Dist_Settlements	1.628	0.614
Dist_Main_Roads	1.767	0.566
Elevation	3.401	0.294

[Fig. 9](#). The evaluation metrics accuracy, recall, specificity, precision, and F1 score computed through these confusion matrices are presented in [Table 5](#), and [Fig. 9](#) shows the ROC curves and AUC scores of seven ML models. For overall accuracy, the XGBoost model outperformed all other models (0.9607). It was followed by LightGBM (0.9580), RF (0.9518), HGB (0.9387), CART (0.8946), LR (0.8856), and NB (0.8456). The same trend can be observed with the other evaluation metrics, except for the recall where LR and NB switched places.

When the ROC curves and AUC scores are examined ([Fig. 10](#)), it can be deemed that all ML models showed a decent performance for suitable site selection of wind farms for our case study, as AUC scores were higher than 90 % for all models while XGBoost, RF, LightGBM, and HGB models nearly converged to 100 %. XGBoost had the best AUC score (0.9907), followed by RF (0.9892), LightGBM (0.9890), HGB (0.9845), LR (0.9514), CART (0.9438), and NB (0.9211).

Moreover, Cochran's Q and pairwise McNemar's tests were conducted to investigate whether there are statistically significant differences among ML models. According to the results of Cochran's Q test, there was a statistically significant difference between models ($\chi^2(6) = 415.240, p < .000$) as it exceeded the threshold value of 12.592 at the 95 % confidence interval. Hence, the null hypothesis was rejected. The results of pairwise McNemar's tests are given in [Table 6](#). Upon examining [Table 6](#), it can be clearly seen that the output of ML models is statistically significant in general. The greatest significant difference was between XGBoost and NB (148.951), and the least significant difference was between XGBoost and RF (4.645), while the outputs of some ML models are quite similar (LR-CART, RF-LightGBM, XGBoost-LightGBM).

4.3. Model explanations with SHAP method

It is crucial to explain how a particular site sample or criterion impacts the predictions of ML models in order to gain valuable insights for the suitable site selection problem of wind farms. The SHAP method can provide both sample-wise and criterion-wise explanations with its summary plots (beeswarm and global bar plots), respectively. [Fig. 11](#) shows the beeswarm plots of seven ML models. The x-axis of the plot implies the SHAP values, and the y-axis shows the site selection criteria ordered based on their overall importance. Each dot in the figure represents a site sample, with pink indicating a higher value, whereas blue indicates a lower SHAP value. The SHAP values shown on the horizontal axis imply the positive or negative impact of a criterion for the

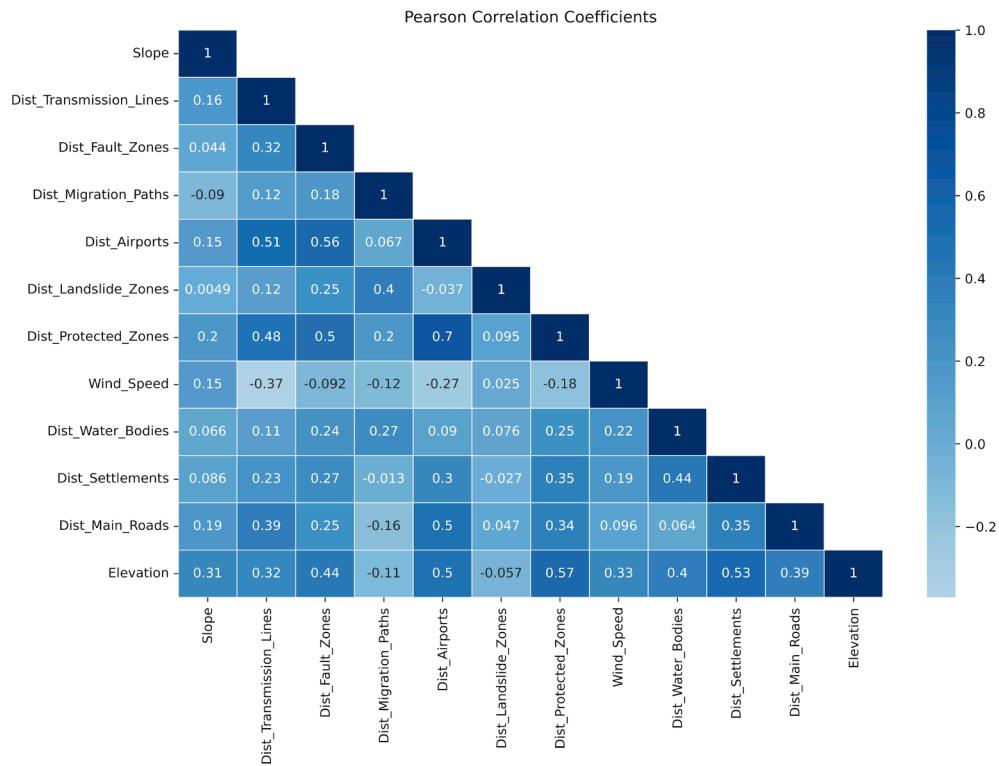


Fig. 8. Correlation matrix of the site selection criteria.

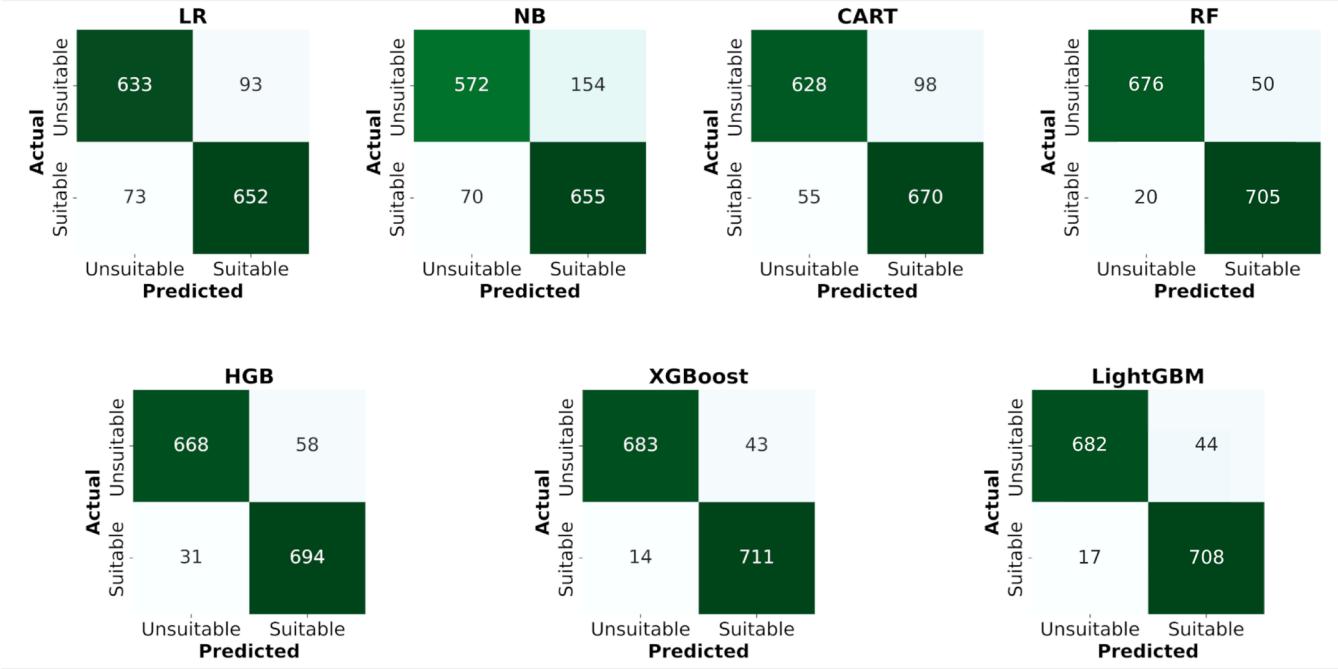


Fig. 9. Confusion matrices of seven ML models.

prediction. For example, the wind speed criterion has a positive impact on the predictions, and it is the most important criterion for all ML models. Therefore, the likelihood of a sample site being suitable increases as the wind speed rises, or vice versa. According to the SHAP plot of XGBoost, wind speed, distance to transmission lines, distance to protected zones, and elevation are the top contributing criteria, while distance to settlements, distance to airports, distance to landslide zones,

and distance to water bodies stand in the middle. The same trend can be seen for the LightGBM in general. For RF, the distance to landslide zones switched places, with distance to protected zones being the top contributing criteria. The slope, distance to main roads, distance to bird migration paths, distance to fault zones, and distance to water bodies are the least contributing criteria for the top 3 performing models (XGBoost, LightGBM, and RF). The same trend can be seen in other traditional ML

Table 5
Performance metrics of ML models.

ML Model	Accuracy	Recall	Specificity	Precision	F1 Score
LR	0.8856	0.8993	0.8719	0.8752	0.8871
NB	0.8456	0.9034	0.7879	0.8096	0.8540
CART	0.8946	0.9241	0.8650	0.8724	0.8975
RF	0.9518	0.9724	0.9311	0.9339	0.9527
HGB	0.9387	0.9572	0.9201	0.9229	0.9397
XGBoost	0.9607	0.9807	0.9408	0.9430	0.9615
LightGBM	0.9580	0.9766	0.9394	0.9415	0.9587

models (LR, NB, and CART) in general. Although the distance to landslide zones is not one of the top contributing criteria, it has a strong negative impact on the HGB model. This implies that the likelihood of a sample site being suitable decreases as the distance to landslide zones increases. Fig. 12 shows the mean absolute SHAP plots of each ML model. A mean absolute SHAP plot indicates the magnitude of the impact of a site selection criterion. Unlike a beeswarm plot, it does not show the direction of the impact or sample-wise SHAP values; instead, it illustrates the overall contribution of a criterion to the suitable site selection process with the SHAP scores, similar to the permutation feature importance. As shown in Fig. 11 earlier, the wind speed, distance to transmission lines, distance to protected zones, elevation, and distance to landslide zones are the most contributing criteria for the predictions

of XGBoost, LightGBM, and RF.

4.4. Generated site suitability maps

The resulting SSMs, generated through the outputs of each ML model, are presented in Fig. 13. Upon examining Fig. 13, it can be observed that the maps show different distributions and patterns for the same sites. For better comprehension, Fig. 14 presents the areal distribution (%) of each suitability class, and Fig. 15 shows the existing wind farm inventory situated in these classes as part of a ground validation process. According to the results (Fig. 14), Balıkesir province has a substantial number of suitable sites where wind farms can be installed. For example, the XGBoost model concluded that 3.34 % and 4.40 % of the study area are very highly or highly suitable for WT installation, followed by 7.41 % moderate, 23.77 % low, and 59.78 % very low suitability. A similar trend can be observed for LightGBM and RF as well. The differences in distribution among classes are more apparent between XGBoost, LightGBM, RF, and the remaining four methods (HGB, CART, NB, and LR). Moreover, Fig. 15 shows that the vast majority of the existing WTs are installed in the very high or high suitability classes for ML models, especially for XGBoost, LightGBM, and RF models, where the percentages are 96.88 %, 96.55 %, and 97.81 %, respectively.

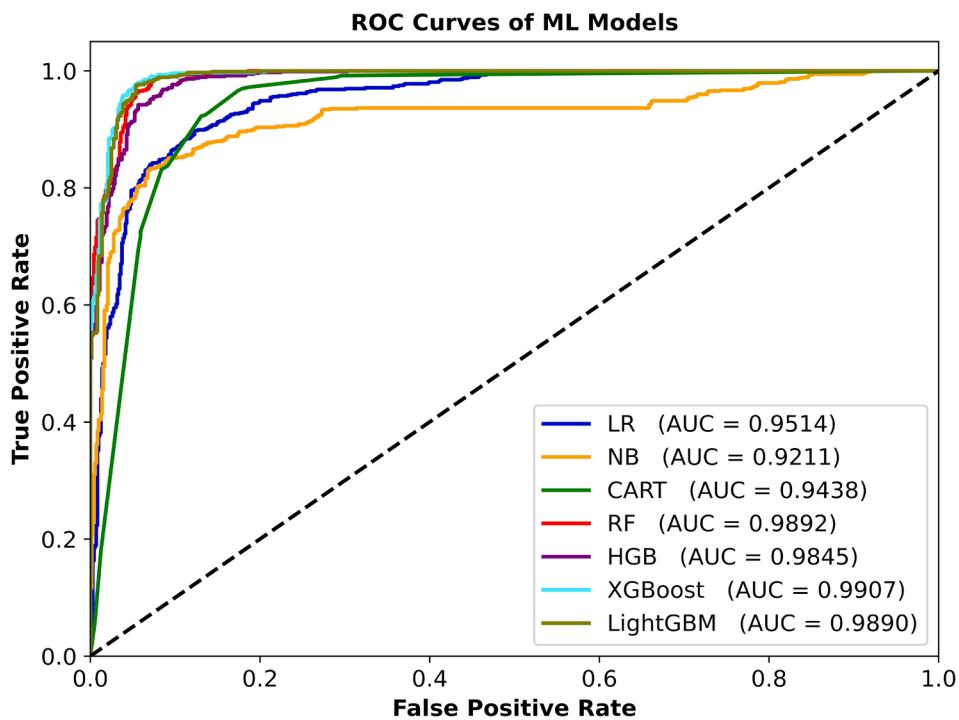
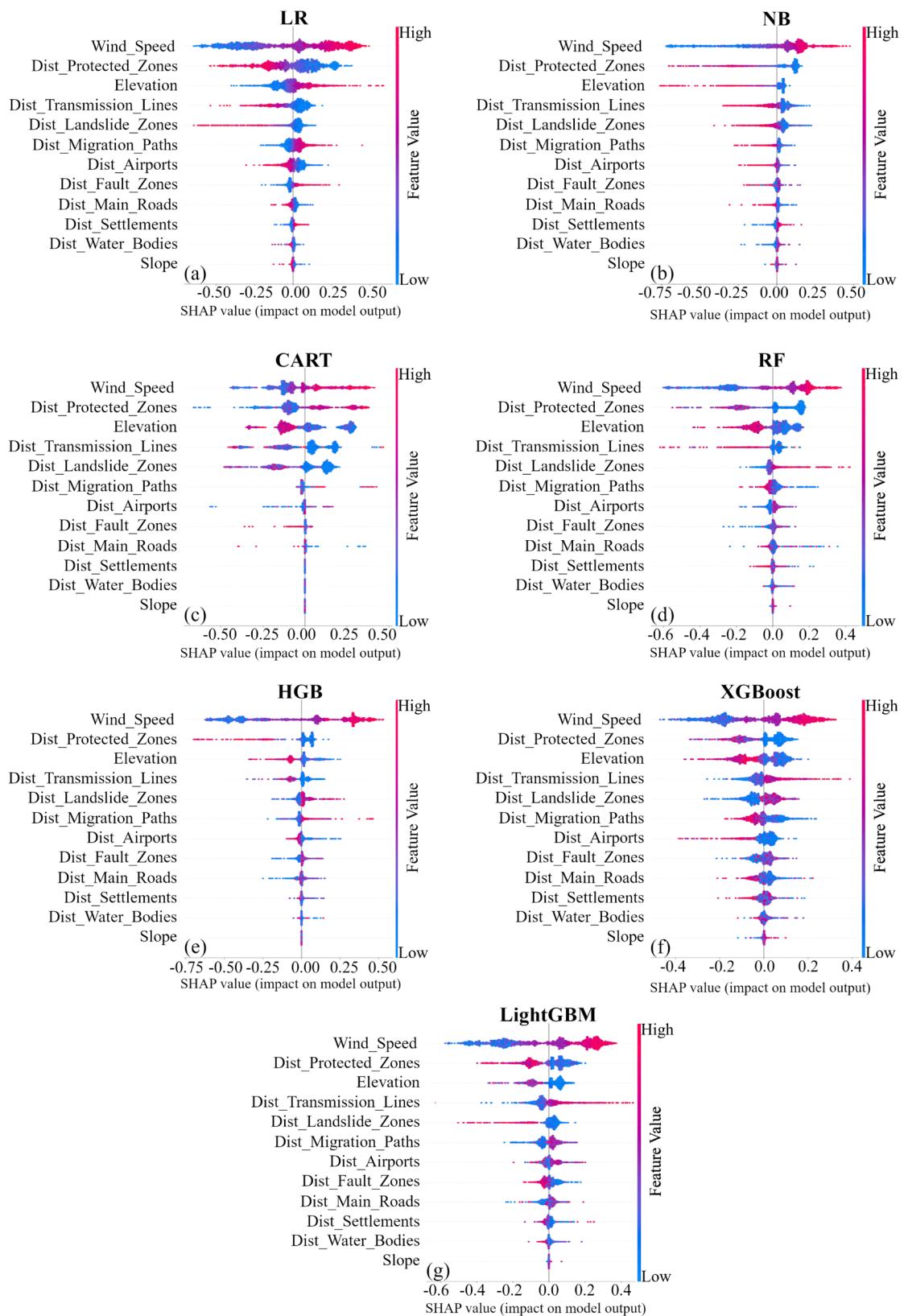


Fig. 10. ROC curves and AUC scores of ML models.

Table 6
Results of the pairwise McNemar's tests (with continuity correction).

	LR	NB	CART	RF	HGB	XGBoost	LightGBM
LR	–	25.786	0.894	75.208	51.115	90.419	87.935
NB		–	24.138	130.050	102.606	148.951	143.410
CART			–	45.128	26.460	62.674	56.719
RF				–	6.113	4.645	1.641
HGB					–	19.220	19.184
XGBoost						–	0.500
LightGBM							–

Note: Statistically significant pairs at 95 % confidence interval is shown in bold.

**Fig. 11.** Beeswarm summary plots of seven ML models.

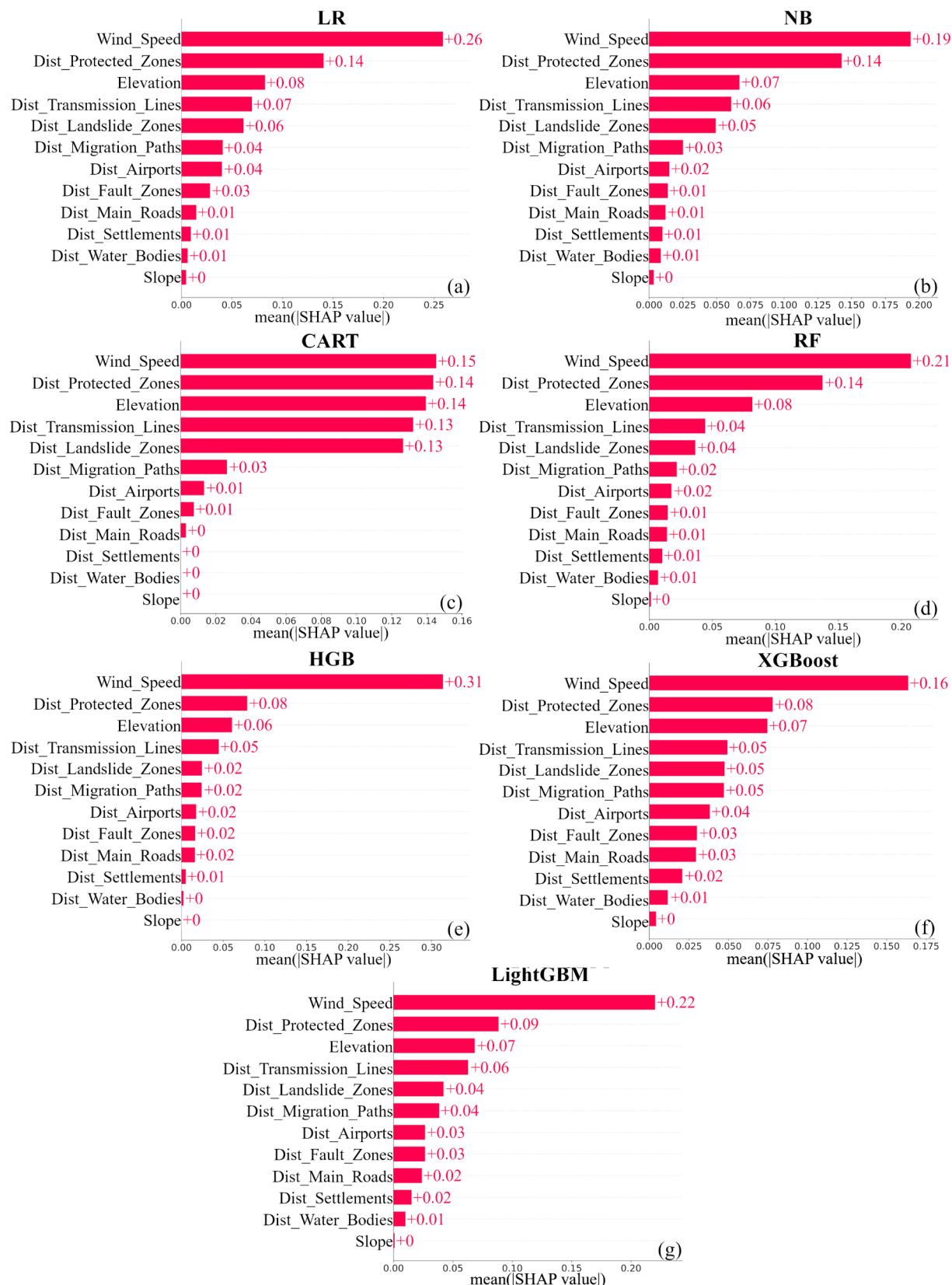


Fig. 12. Mean absolute SHAP summary plots of seven ML models.

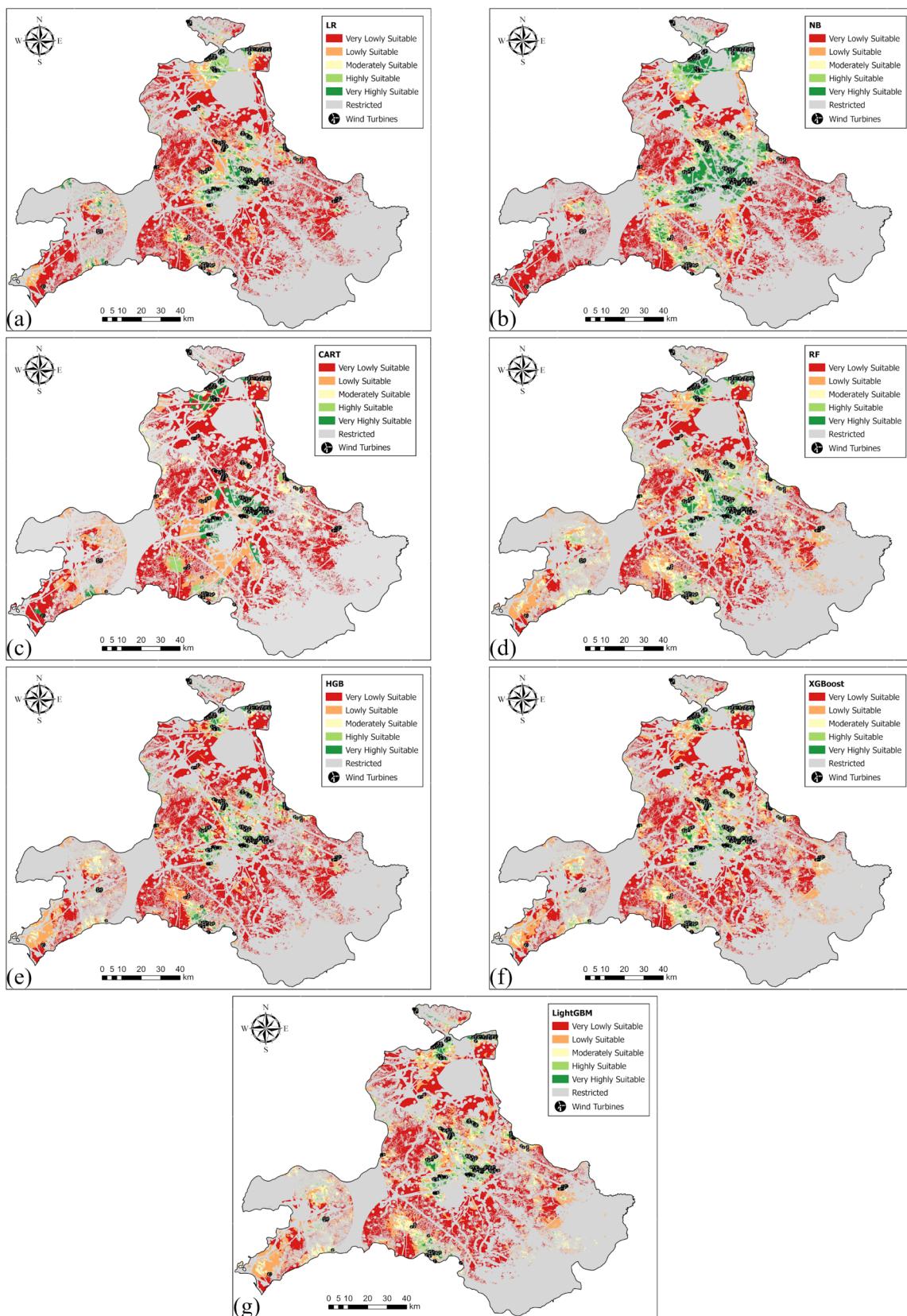


Fig. 13. SSMs of ML models: (a) LR, (b) NB, (c) CART, (d) RF, (e) HGB, (f) XGBoost, (g) LightGBM.

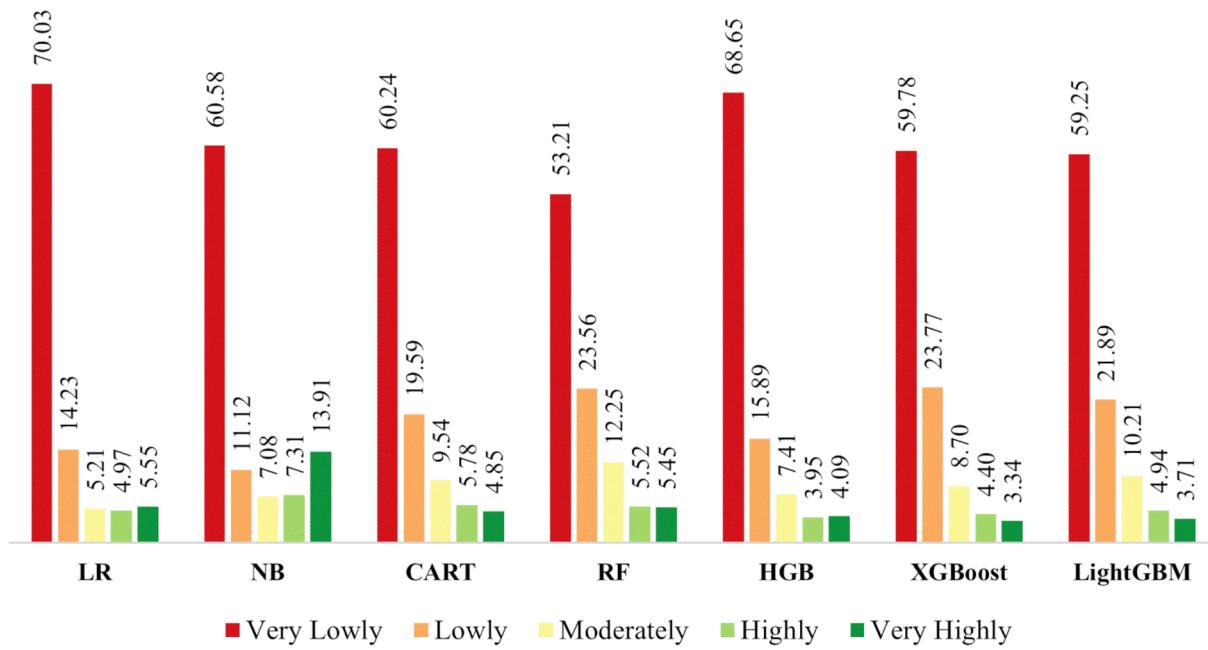


Fig. 14. Areal distribution of suitability classes for ML models.

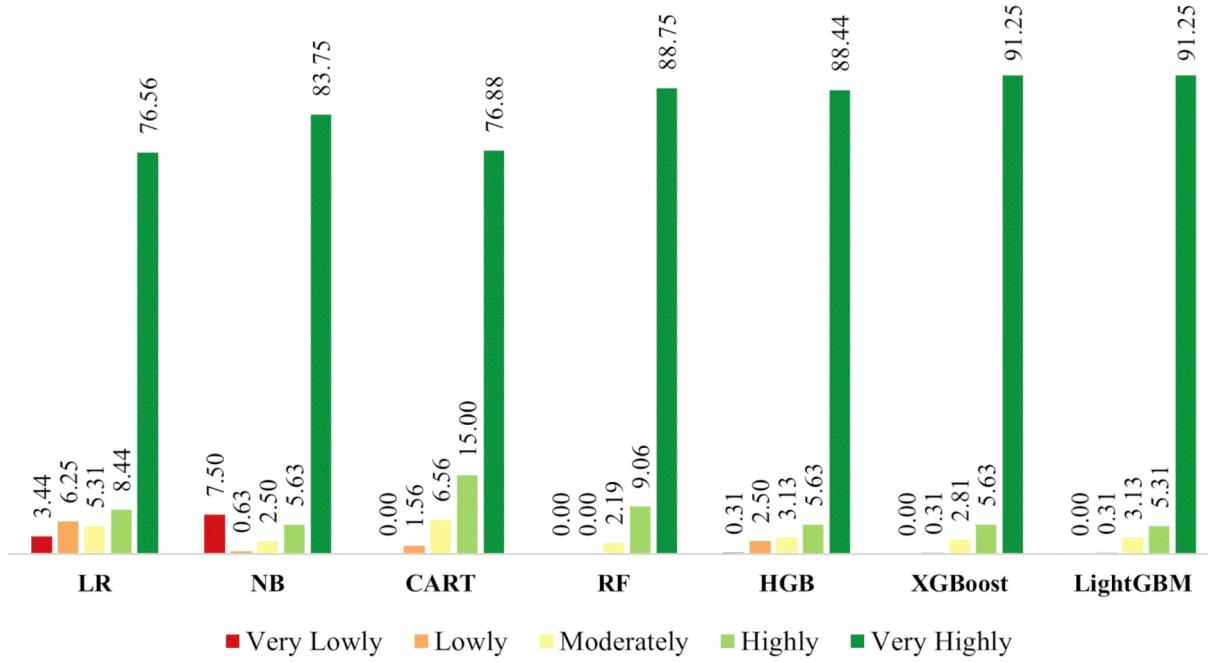


Fig. 15. Overlapping percentages of existing wind turbines and suitability classes provided by ML models.

5. Discussions

5.1. Performances of ML models

The suitable site selection for wind farm installation has gained increased attention in the last decade, as global focus has shifted from fossil fuels to renewable energy sources such as wind and solar energy. In this context, researchers mainly adopted MCDM methods such as the analytical hierarchy process [3,17,19,66], the fuzzy analytical hierarchy process [21,28,29], the best-worst method [16,20,22], or combined methods [67,68]. These MCDM-based methods all rely heavily on the expertise of individual decision-makers thus, biased weight may be resulted. Additionally, the introduction of several new site selection

criteria alters the outcomes of MCDM procedures. On the contrary, ML algorithms can learn the relationships among criteria internally without any domain knowledge from individual experts. Despite its several advantages, a standalone ML-based framework has not been established yet for suitable site selection for wind farms. In this study, a comprehensive ML-based framework that utilizes several ML algorithms along with statistical and explainable ML methods was established.

In ML research, the effectiveness of classification models is significantly influenced by the architecture of the algorithm, the quantity of input variables, and the interactions between these variables. Therefore, predicting which algorithm will be most suitable for a specific problem is a complex task [63,69]. For this reason, both traditional (LR, NB, CART) and state-of-the-art (RF, HGB, XGBoost, LightGBM) ML models

were utilized in our framework to select suitable sites for wind farms in the Balıkesir province. Based on our quantitative comparison utilizing several evaluation metrics, ensemble ML algorithms (XGBoost, LightGBM, RF, HGB) outperformed stand-alone ones (LR, NB, CART), with XGBoost and LightGBM emerging as the top-performing models across all metrics. However, despite the superior performance of XGBoost and LightGBM, the other algorithms are not rendered entirely worthless, as they demonstrated commendable performance in the classification task of candidate sites. The differences in the evaluation metrics of ML models could be attributed to the nature of the metrics themselves and the distinct structures of the ML models utilized in this study. For example, a high AUC or accuracy does not necessarily imply that other metrics, such as precision or recall, should be high at the same time as they consider different cases.

Moreover, according to the results of Cochran's Q and pairwise McNemar's tests, we found statistically significant differences among most ML models (Table 6). Based on these findings, we suggest only using the top-performing tree-based ensemble learning classifiers (XGBoost, LightGBM, and RF) for selecting suitable sites for wind farm installation problems while the differences between XGBoost-LightGBM and LightGBM-RF were statistically insignificant. Additional research is required to determine whether these findings can be generalized to new data utilizing different study areas. In summary, our recommendation is to employ ensemble tree-based machine learning models for the purpose of suitable site selection in wind farm studies, given their demonstrated superior performance in the context of the classification of suitable sites.

5.2. Explanations of the site selection criteria

Another critical component of our ML-based framework is the chosen site selection criteria, as they are the input factors to classify whether a site is suitable or not. Hence, they have a direct impact on the generated SSMS. However, there is no rule of thumb regarding which criteria should be involved in studies. The researchers mainly relied on the findings of previous studies to select which criteria should be incorporated, as was also utilized in this study. We specifically adopted 12 criteria that were used in suitable site selection for wind farm studies conducted in the last decade. We endeavored to systematically aggregate pivotal criteria to enhance the robustness of our framework. Firstly, to ensure the absence of irrelevant criteria in the wind farm data frame, we employed multicollinearity and correlation tests. Our findings suggest that the selected criteria were not interrelated. However, scholars might need to conduct additional tests to determine whether the same applies to other study regions apart from ours. Secondly, we employed IG, a filter-based feature selection method, to assess the preliminary relevance of the site selection criteria. However, the IG method can only offer global explanations regarding the importance of a feature. The SHAP method, on the other hand, possesses the capability to provide sample-wise explanations for criteria, enabling a deeper interpretation of ML models.

Based on the importance scores from IG, the top four contributing criteria are wind speed, distance to protected zones, distance to airports, and distance to transmission lines, respectively (Fig. 7). However, this contradicts the SHAP values provided by XGBoost, LightGBM, and RF (Figs. 11 and 12). The top two contributing criteria are consistent. On the other hand, IG found that the distance to airports is the third most important one and missed the elevation criterion as among the most important criteria. However, the distance to airports criterion is not among the top five for XGBoost, LightGBM, and RF. The reason for this discrepancy may be attributed to the IG method being a simple filter-based technique.

According to SHAP plots for all models (Figs. 11 and 12), the most critical criterion appears to be wind speed, aligning with findings from other studies [21,41,45,47]. As wind speed increases, there is a corresponding rise in the likelihood of a candidate site being classified as suitable. This aligns with expectations, as it directly reflects the energy

generation potential of a wind turbine in tandem with the wind energy potential of Balıkesir province. The distance to transmission lines emerges as the second most significant criterion, as indicated by XGBoost and RF, with mean absolute SHAP values of 0.08 and 0.14, respectively. Other scholars have also identified this criterion among the most important ones [3,17,22,28,39,46]. Based on the models, shorter distances to existing transmission lines enhance the suitability of a candidate site. The proximity to existing transmission lines is a crucial economic criterion, as the primary objective of wind farm installation projects is to maximize energy potential while minimizing costs and environmental impacts. Hence, the domain knowledge is in parallel with the outputs of the ML models. The distance to protected zones is identified as another crucial criterion according to the ML models, consistent with findings in other research studies [20,22,41,45]. According to our domain knowledge, WTs should be situated far from protected zones to prevent harm to wildlife or historical sites. However, the ML models suggest that lower distances from protected zones increase the suitability of a site sample. The reason for this apparent contradiction can be uncovered through a deeper investigation of the spatial relationship between existing WTs and protected zones. In the study area, most existing WTs are located in the ranges of 1000–10000 m or 10001–20000 m (110 and 239, respectively). Consequently, all the WTs were installed at distances greater than 1000 m from the restricted zones. Therefore, although they are situated far enough from protected zones, the models portray them differently as than the initial impression. This finding suggests that using the distance to protected zones as a restriction criterion rather than an assessment one is more appropriate, as utilized by Ayodele et al. [28]. Another important criterion revealed by the models is the elevation. According to the SHAP plots of ML models (Fig. 11), as elevation rises, the suitability of a candidate site also increases. Some researchers backed up this claim [21] as the higher elevations tend to capture more wind speed, while others stated lower elevations are better to reduce installation and maintenance costs [3,19,22]. In our case, the higher elevations tend to increase the suitability of a site. The final important criteria, according to the ML models, is the distance to settlements. The suitability of a site enhances as the distance to settlements rises, in line with other scholars [3,22,45,47]. It is expected since lower distances to the settlements may cause visual and noise emissions, as also stated by Tercan [20]. The final remark that can be concluded from the explanation of the site selection criteria is the case of slope. In our case, ML models revealed that slope is the least contributing criterion for the suitable site selection of wind farms, contrary to studies that assert slope as an important criterion, suggesting that areas with flat slopes are better for WT siting [17,21,22,39]. Some studies also state that slope is not of much importance [19,45]. In our case, the installed WTs are almost equally distributed among suitability classes of slope, with a slight favor toward < 5 % and 5–10 % classes. An in-depth investigation revealed that a trade-off between high wind speed and high slope was made in the study area by local experts. Other criteria can be regarded as less important, so no further investigation was carried out for them.

5.3. Usability of site suitability maps for wind turbine installation

The final output of our ML-based framework is the generated SSMS. These maps showcase the suitability of sites in varying degrees and highlight restricted zones at a 30 m spatial resolution. The SSMS are highly valuable for decision makers in industrial or governmental entities associated with energy companies. Specifically, these maps offer a detailed representation of potential WT installation sites that have not yet been developed, providing valuable insights at a fine-grained spatial scale. The generated SSMS can significantly assist wind experts in future wind farm installations by maximizing the energy generation potential of Balıkesir province [24] and minimizing the costs associated with wind energy projects. Additionally, local inhabitants stand to benefit from these maps, as our framework also incorporates social criteria for SSM

generation to minimize the impact of WTs on the community.

The spatial distribution of site suitability classes reveals that Balikesir province has a vast number of suitable sites for future WT installation, considering its total size of 14,604 km². These locations, at a fine resolution, could substantially improve wind energy generation and assist Türkiye in addressing challenges introduced by fossil fuel consumption. Moreover, the ground validation, which accounts for the overlapping percentage between existing WTs and suitability classes, showed that the ML models utilized in this study are feasible for automatic suitable site selection of wind farms. A significant percentage of existing WTs overlap with the recommended suitability sites, especially ensemble tree-based ML models (RF, HGB, XGBoost, and LightGBM). This situation validates that our framework, which utilizes ML models, is trustworthy, and the generated SSMs could be comfortably incorporated by the wind experts.

5.4. Limitations and future research directions

There are certain limitations to the framework introduced in this study. Firstly, we made efforts to utilize openly available data sources to compose our ML-based framework (except for the DEM). Although this supports accessibility and encourages further improvements, it introduces some challenges regarding data completeness, quality, and integrity. Specifically, the sources obtained from OpenStreetMap should be further investigated, as it also contains the existing WTs (wind farm inventory), which is the key component of our framework that serves as the ground-truth data. Future studies could replace the TanDEM-X DEM with the SRTM DEM for a fully open-source framework since our generated SSMs have a 30 m spatial resolution. Secondly, the proposed framework is fully based on the data of Balikesir province. Further research should be conducted to investigate how well our framework generalizes to other provinces. However, we believe our proposed framework could perform well in provinces with similar characteristics to Balikesir, such as Çanakkale and İzmir. Third, the restrictions applied in this study are completely based on previous research. However, there is no consensus in the current literature on how restrictions should be applied. Expert opinions might be needed for restrictions before utilizing the framework in another province. Fourth, we put substantial effort into including the most significant site selection criteria. More criteria may be added to the framework, and more powerful feature selection techniques, such as recursive feature elimination or exhaustive search, might be employed in future studies for a better preliminary investigation of the site selection criteria. Furthermore, although we believe there will not be much difference between XGBoost and LightGBM, other boosting-based state-of-the-art ML models like Category Boosting-CatBoost [70], and Natural Gradient Boosting-NGBoost [71] could be incorporated into future studies for ML-based suitable site selection studies of wind farms.

6. Conclusions

Nowadays, marked by the harmful consequences and extended use of fossil fuels, which lead to environmental pollution and health risks, the importance of renewable energy sources is increasingly rising to meet the growing demand for sustainable and eco-friendly energy. It is now apparent that the phasing out of fossil fuels needs to be accompanied by the proliferation of renewable energy sources. Currently, wind energy emerges among the principal significant renewable energy sources because of its economically efficient energy conversion technology, numerous suitable sites, and continuous wind flow throughout the day. Nonetheless, the process of choosing suitable locations is complex, requiring the integration of a complete framework to assess various factors that will minimize environmental, technical, and social limitations. At this point, MCDM methods are commonly utilized in this research domain. While there are many advantages to assessing the suitability of WT sites, the weightings assigned to the conditioning

factors are subjective, leading to a state of uncertainty. The objective of this study is to enhance the accuracy and reliability of wind farm SSMs by developing an ML-based framework for Balikesir province in Türkiye. This study seeks to evaluate the predictive performance of the seven ML models, namely Logistic Regression (LR), Naive Bayes (NB), Classification and Regression Trees (CART), Random Forest (RF), Histogram-based Gradient Boosting (HGB), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM), for the first time in wind farm SSM works. These methods were analyzed using twelve site selection criteria for generating wind farm SSMs, divided into three categories: environmental, economic, and social. To evaluate the importance of each criterion, the criteria were subjected to an information gain analysis. Then, a multicollinearity test was applied followed by Pearson correlation test to determine whether there were any unsuitable site selection criteria. Six performance metrics, such as accuracy, recall, specificity, precision, F1 score, and AUC, were utilized to evaluate the effectiveness of ML models. Furthermore, Cochran's Q and pairwise McNemar's tests were employed to determine the statistically significant differences in the outputs of model predictions. To describe how a site sample or criterion affects ML model predictions, we implemented a local explanation strategy, namely SHAP. The main findings are summarized as follows:

- The XGBoost, LightGBM, RF, and HGB models achieved the greatest F1 scores (0.9615, 0.9587, 0.9527, and 0.9397) respectively, followed by CART (0.8975), LR (0.8871), and NB (0.8540).
- Statistically significant differences were found in pairwise McNemar's tests for all pairs, except for the LR-CART, RF-LightGBM, and XGBoost-LightGBM pairings.
- The overall findings of the SHAP analysis indicate that regions characterized by elevated wind speeds, higher elevations, lower slopes, closer distances to transmission lines, protected areas, and greater distances to settlements are more favorable for the selection of WT sites.
- Balikesir province offers numerous appropriate locations for wind farm installations that have not been installed yet. The XGBoost model found that 3.34 % and 4.40 % of the study area are highly or very highly suitable for WT installation, 7.41 % moderate, 23.77 % low, and 59.78 % very low. Similar trends are seen with LightGBM and RF. The differences in distribution between classes are especially noticeable between the three methods (XGBoost, LightGBM, and RF) and the remaining four methods (HGB, CART, NB, and LR).
- While the accuracy outcomes derived from various ML methods appear to be applicable in the context of selecting suitable locations for WTs, the majority of operational WTs are situated in areas classified as high or higher suitability, with the RF, LightGBM, and XGBoost models being particularly prevalent in this regard. This finding is particularly noteworthy as it indicates that around 97 % of the wind energy generated in Balikesir province occurs in regions classified as high or higher suitable areas, according to the research.

Given that wind farms account for around 10 % of Türkiye's total power generation, with the western area representing 75 % of this share, it is evident that such an ML framework might be employed in provinces that have similar characteristics.

The ML-based framework presented in this study can facilitate the establishment of a sustainable energy infrastructure by supporting the wind farm installation process. Thus, it can lead the transition process from fossil fuels to clean and accessible energy in line with the SDG 7 of UN. Future work will focus on the analysis of various clustering algorithms to automatically generate suitable field polygons to aid the WT setup process and the performance of tabular data deep learning algorithms.

- [44] Sánchez-Lozano JM, García-Cascales MS, Lamata MT. GIS-based onshore wind farm site selection using Fuzzy Multi-Criteria Decision Making methods. Evaluating the case of Southeastern Spain. *Appl Energy* 2016;171:86–102. <https://doi.org/10.1016/j.apenergy.2016.03.030>.
- [45] Höfer T, Sunak Y, Siddique H, Madlener R. Wind farm siting using a spatial Analytic Hierarchy Process approach: A case study of the Städteregion Aachen. *Appl Energy* 2016;163:222–43. <https://doi.org/10.1016/j.apenergy.2015.10.138>.
- [46] Atici KB, Simsek AB, Ulucan A, Tosun MU. A GIS-based Multiple Criteria Decision Analysis approach for wind power plant site selection. *Util Policy* 2015;37:86–96. <https://doi.org/10.1016/j.jup.2015.06.001>.
- [47] Sánchez-Lozano JM, García-Cascales MS, Lamata MT. Identification and selection of potential sites for onshore wind farms development in Region of Murcia. Spain. *Energy* 2014;73:311–24. <https://doi.org/10.1016/j.energy.2014.06.024>.
- [48] Ekiz S, Şirin A, Erener A. En uygun rüzgar enerji santrali yerlerinin coğrafi bilgi sistemleri ile belirlenmesi: Kocaeli ili örneği. *Jeodezi ve Jeoinformasyon Dergisi* 2022;9(1):59–79. <https://doi.org/10.9733/JGG.2022R0005.T>.
- [49] Gigović L, Pamučar D, Božanić D, Ljubojević S. Application of the GIS-DANP-MABAC multi-criteria model for selecting the location of wind farms: A case study of Vojvodina, Serbia. *Renew Energy* 2017;103:501–21. <https://doi.org/10.1016/j.renene.2016.11.057>.
- [50] Kavzoglu T, Teke A. Predictive Performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost). *Arab J Sci Eng* 2022;47:7367–85. <https://doi.org/10.1007/s13369-022-06560-8>.
- [51] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 2011;12:2825–30.
- [52] Wright RE. Logistic regression. Reading and Understanding Multivariate Statistics, American Psychological Association; 1995, p. 217–44.
- [53] Iban MC, Sekertekin A. Machine learning based wildfire susceptibility mapping using remotely sensed fire data and GIS: A case study of Adana and Mersin provinces. *Turkey Ecol Inform* 2022;69:101647. <https://doi.org/10.1016/j.ecoinf.2022.101647>.
- [54] Hand DJ, Yu K. Idiot's Bayes—Not So Stupid After All? *Int Statistical Rev* 2001;69 (3):385–98. <https://doi.org/10.1111/j.1751-5823.2001.tb00465.x>.
- [55] Breiman L, Friedman JH, Olshen RA, Stone CJ. *Cart: classification and regression trees* (1984). Belmont, CA: Wadsworth; 1993.
- [56] Breiman L. Random Forests. *Mach Learn* 2001;45(1):5–32. <https://doi.org/10.1023/A:1010933404324>.
- [57] Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* 2001;29(5):1189–232. <https://doi.org/10.1214/aos/1013203451>.
- [58] Guryanov A. Histogram-Based Algorithm for Building Gradient Boosting Ensembles of Piecewise Linear Decision Trees. In: Van Der Aalst WMP, Batagelj V, Ignatov DI, Khachay M, Kusкова V, Kutuzov A, et al., editors. *Analysis of Images, Social Networks and Texts*, vol. 11832, Cham: Springer International Publishing; 2019, p. 39–50. doi: 10.1007/978-3-030-37334-4_4.
- [59] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM; 2016, p. 785–94. doi: 10.1145/2939672.2939785.
- [60] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: A highly efficient gradient boosting decision tree. *Adv Neural Inf Proces Syst* 2017:30.
- [61] McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1947;12:153–7. <https://doi.org/10.1007/BF02295996>.
- [62] Edwards AL. Note on the “correction for continuity” in testing the significance of the difference between correlated proportions. *Psychometrika* 1948;13:185–7. <https://doi.org/10.1007/BF02289261>.
- [63] Aydin HE, Iban MC. Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with SHapley Additive exPlanations. *Nat Hazards* 2023;116:2957–91. <https://doi.org/10.1007/s11069-022-05793-y>.
- [64] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* 2017:30.
- [65] Shapley LS. Stochastic Games*. *Proceedings of the National Academy of Sciences* 1953;39:1095–100. doi: 10.1073/pnas.39.10.1095.
- [66] Al-Shabeeb AR, Al-Adamat R, Mashqabah A. AHP with GIS for a preliminary site selection of wind turbines in the North West of Jordan. *Int J Geosci* 2016;7(10):1208–21. <https://doi.org/10.4236/ijg.2016.710090>.
- [67] Nagababu G, Puppala H, Pritam K, Kantipudi MP. Two-stage GIS-MCDM based algorithm to identify plausible regions at micro level to install wind farms: A case study of India. *Energy* 2022;248:123594. <https://doi.org/10.1016/j.energy.2022.123594>.
- [68] Wu Y, He F, Zhou J, Wu C, Liu F, Tao Y, et al. Optimal site selection for distributed wind power coupled hydrogen storage project using a geographical information system based multi-criteria decision-making approach: A case in China. *J Clean Prod* 2021;299:126905. <https://doi.org/10.1016/j.jclepro.2021.126905>.
- [69] Iban MC, Bilgilioglu SS. Snow avalanche susceptibility mapping using novel tree-based machine learning algorithms (XGBoost, NGBoost, and LightGBM) with eXplainable Artificial Intelligence (XAI) approach. *Stoch Env Res Risk A* 2023;37:2243–70. <https://doi.org/10.1007/s00477-023-02392-6>.
- [70] Dorogush AV, Ershov V, Gulin A. CatBoost: gradient boosting with categorical features support 2018.
- [71] Duan T, Anand A, Ding DY, Thai KK, Basu S, Ng A, et al. Ngboost: Natural gradient boosting for probabilistic prediction. *International conference on machine learning*, PMLR; 2020, p. 2690–700.