MULTI-CRITERIA GIS-BASED MACHINE LEARNING TECHNIQUES FOR WIND FARM SITE SELECTION

Oluwatosin Ibukunoluwa Ilesanmi[1*] Opeyemi Moses Ogungbenle[2], Aweda Temidayo Olusina[3], Ajiboye Oyesiji Oyeyode[4]

[1]Department of Photogrammetry and Remote Sensing, Federal School of Surveying, Oyo, Nigeria
[2]Department of Cartography and G.I.S, Federal School of Surveying, Oyo, Nigeria
[3]Department of Computer Science, Federal School of Surveying, Oyo, Nigeria
[4]Department of Surveying and Geo-informatics, Federal School of Surveying, Oyo, Nigeria
[*]Corresponding author: hillesanmet@gmail.com

ABSTRACT

This research, conducted in Ekiti state, aimed to identify the optimal location for a wind farm using a Geographic Information System (GIS) -based machine learning approach. Predictive models, specifically Support Vector Machine (SVM) and Random Forest (RF), were employed to enhance the accuracy of wind farm site selection. Six factors were utilized to create training and testing datasets for model verification and validation. The validation process, using the Area under the Curve (AUC) metric, yielded AUC values of 0.75 for SVM and 0.8 for RF. The findings indicate that the Random Forest model demonstrated superior predictive capability. In conclusion, the integration of GIS and machine learning, particularly employing the Random Forest algorithm, proved to be effective in assessing the potential for wind farm sites in the research area.

Keywords: *Geographic Information System, machine learning, predictive model, Support Vector Machine, Random Forest, Area under Curve.*

1. Introduction

Energy stands as a crucial driver for global economic and industrial advancement. The predominant share of the world's energy consumption is derived from conventional fossil fuels such as oil, gas, and coal. (SeyedAlavi et al, 2022). The primary drawbacks of fossil fuels, including the generation of environmental pollutants, limited availability, and elevated costs, have prompted numerous governments to explore alternative energy sources. Among these alternatives, renewable energy sources stand out, as they offer a clean and limitless supply in various regions across the globe without the associated drawbacks of pollution and finite reserves. (Moltames et al, 2022).

Given the growing focus on environmental concerns, the pursuit of clean energy, exemplified by wind and solar energy, is at the forefront of the energy revolution. Solar energy, in particular, has witnessed rapid and substantial development. (Yang et al., 2019). Wind power has garnered significant attention due to its abundant resources and efficient technology for power generation. (Liu et al., 2016). Because of the intermittent characteristics of wind and solar energy, standalone wind and photovoltaic (PV) energy systems typically necessitate energy storage devices or additional generation sources to establish a hybrid system. The storage device may take the form of a battery bank, supercapacitor bank, superconducting magnetic energy storage (SMES), or a fuel cell (FC)–electrolyzer system. (Wang and Nehrir 2008).

Forecasting wind power poses a significant challenge due to the weather-dependent nature of wind speed, characterized by high instability, randomness, and volatility. The inherent instability and uncontrollable nature of wind flows contribute to strong randomness in short periods. This random behavior in wind power generation creates an imbalance between power generation and consumption, leading to increased costs and unpredictability for users of this energy. Consequently, accurate prediction of wind power becomes crucial for effective energy management, including tasks such as appropriate generation, distribution, transmission, planning, and scheduling. (Yurek et al., 2021)

The significant benefit of wind energy lies in its capacity to deploy wind turbines in areas distant from traditional electricity grids. Additionally, it can function as an alternative energy source during periods of peak demand. (Moltames et al, 2022).

A Geographic Information System (GIS) serves as a robust tool for decision-making across diverse business sectors, given its capability to analyze environmental, demographic, and topographic data. The data intelligence derived from GIS applications is instrumental in aiding companies, industries, and consumers in making well-informed decisions. Using digital thematic maps and a conceptual model for data integration, GIS minimizes human error and efficiently identifies optimal locations for wind farms. The selection of suitable sites for wind farm installation hinges on factors encompassing environmental, technical, geographical, and theoretical parameters. (Moltames et al., 2022).

Selecting an appropriate site for the installation of wind farms poses a significant challenge in the development of wind resources. Geographic Information System (GIS) has emerged as a widely employed Decision Support System (DSS) to effectively identify and assess suitable locations for the establishment of wind farms. (Moltames et al., 2022). Utilizing Geographic Information System (GIS) as a Decision Support System (DSS) proves invaluable by supplying extensive spatial data for decision-making processes related to the assessment and development of wind resources.

Moreover, precise wind power prediction enhances the utilization of wind energy. Machine learning (ML) has emerged as a pivotal player in the energy sector. ML techniques are extensively employed to analyze historical data, enabling accurate predictions of future wind power generation. This utilization of ML contributes to enhanced forecasting performance in the realm of wind power generation. (Yurek et al., 2021). Wind power prediction methods can be categorized into deterministic prediction and probabilistic prediction. The majority of methods fall under deterministic prediction, offering specific predicted values at particular times. Common evaluation indices for these deterministic predictions include mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), among others. Classical prediction models like support vector machines (SVM) and artificial

neural networks (ANN) are extensively applied in deterministic wind power predictions. (Li et al., 2020)
The SVM (Support Vector Machine) model is considered ideal for wind power prediction due to its remarkable learning ability, especially when dealing with limited sample data. Consequently, the SVM is selected as the foundational model for wind power prediction in this study. It's worth noting that the selection of parameters plays a crucial role in influencing the prediction performance of the SVM model. (Li et al., 2020).

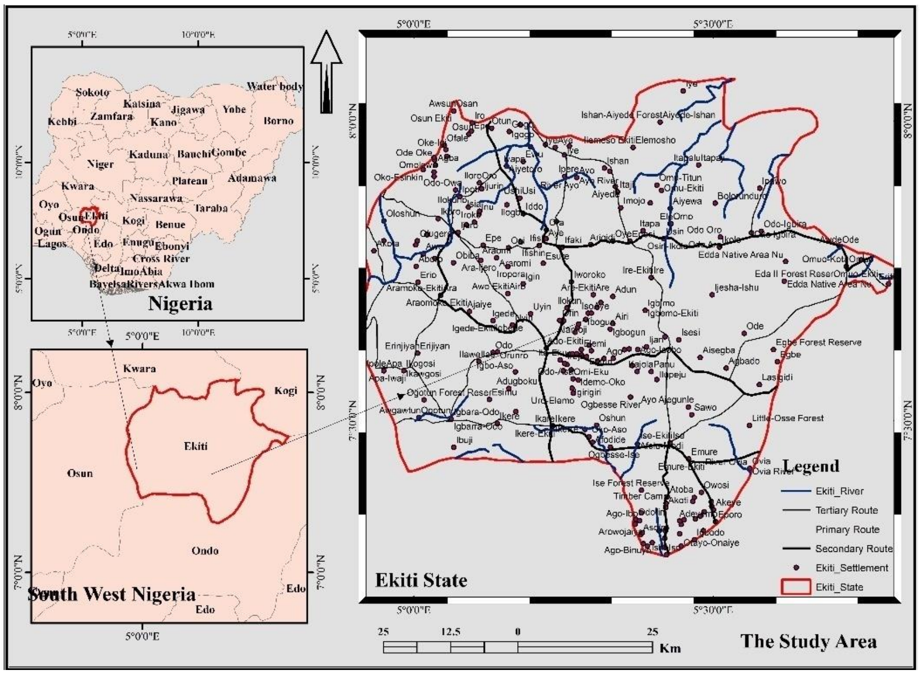


Figure 1: The study area map, Ekiti State

2. Study Area

The research area under consideration is Ekiti State, situated in the southwestern region of Nigeria (refer to Figure 1). Ekiti State is positioned between Latitudes 7° 25′ to 8° 5′ N and Longitudes 4° 45′ to 5° 45′ E, covering an approximate land area of 5435 km². Predominantly an upland zone, the state exhibits elevations ranging from 250 to 540 m above mean sea level (a.m.s.l.). The climate in Ekiti State is tropical, characterized by two distinct seasons: a rainy season extending from April to October and a dry season spanning November to March. The state experiences high humidity levels, and the mean air temperature fluctuates between 21 and 28 °C. (Olorunfemi et al., 2020). The study area encompasses a diverse landscape, comprising both urban and rural components, along with vegetated and non-vegetated areas. Ekiti State has undergone notable changes in land use and land cover (LULC). The population in the region has witnessed consistent growth, escalating from 1.63 million in 1991 to 2.38 million in 2006. Projections indicate a further increase to approximately 3.17 million by the year 2015. This demographic trend suggests ongoing urbanization and potential alterations in the distribution of land use and land cover within Ekiti State. (ekitistate.gov.ng; citypopulation.de, retrieved 2018, Feb. 6). As of 2015, Ekiti State recorded a population density of 498.19 individuals per square kilometer. This was associated with an annual growth rate of +3.13% during the period from 2006 to 2015 (source: citypopulation.de, retrieved on February 6, 2018). The steady rise in the human population has been a driving force behind significant changes in the Land Use and Land Cover (LULC) patterns within the study area. This demographic shift is likely contributing to alterations in the landscape and the utilization of land for various purposes. (Olorunfemi et al., 2020)

3. Methodology

To map suitable sites for wind farm construction in Ekiti State, we adopted the methodology presented in Figure 2

3.1 Data Collection

The research approach involves leveraging a comprehensive dataset that encompasses various factors influencing the planning of wind project placement. The collected data spans across socioeconomic, environmental, and technical parameters, as outlined in Table 1. This multifaceted dataset is likely crucial for conducting a holistic assessment and identifying suitable sites for wind farm construction in Ekiti State. The integration of these diverse parameters reflects a thorough and inclusive methodology aimed at making well-informed decisions regarding wind energy project locations.

Table 1: The data used in the study

Data	File Format
Wind Speed	Grid
Digital Elevation (STRM)	Grid
Landsat 9	Grid
Road Network	Shapefile
Settlements	Shapefile
River Network	Shapefile

The study was initiated by preparing a wind farm plant map, utilizing wind speed data obtained through a raster area that classified wind speed into two classes based on their rates. A total of 5 geo-environmental factors, namely Elevation, Slope, Road network, River network, Settlements, and Land Use/Land Cover, were selected based on a comprehensive literature review.

The processing of these factors was executed in R-Studio, a statistical computing and graphics environment. Two modeling techniques, namely Support Vector Machine (SVM) and Random Forest (RF), were employed and subsequently compared. To train these models, 70% of the borehole data was utilized, and the accuracy of the models was validated using the remaining 30% of the wind speed data.

The model outputs were exported as raster files, and the Accuracy under the Curve (AUC) was calculated to assess the performance of the SVM and RF models. This approach, involving a combination of spatial analysis, machine learning, and validation techniques, provides a robust framework for identifying suitable sites for wind farm construction based on a comprehensive set of geo-environmental factors and wind speed data.

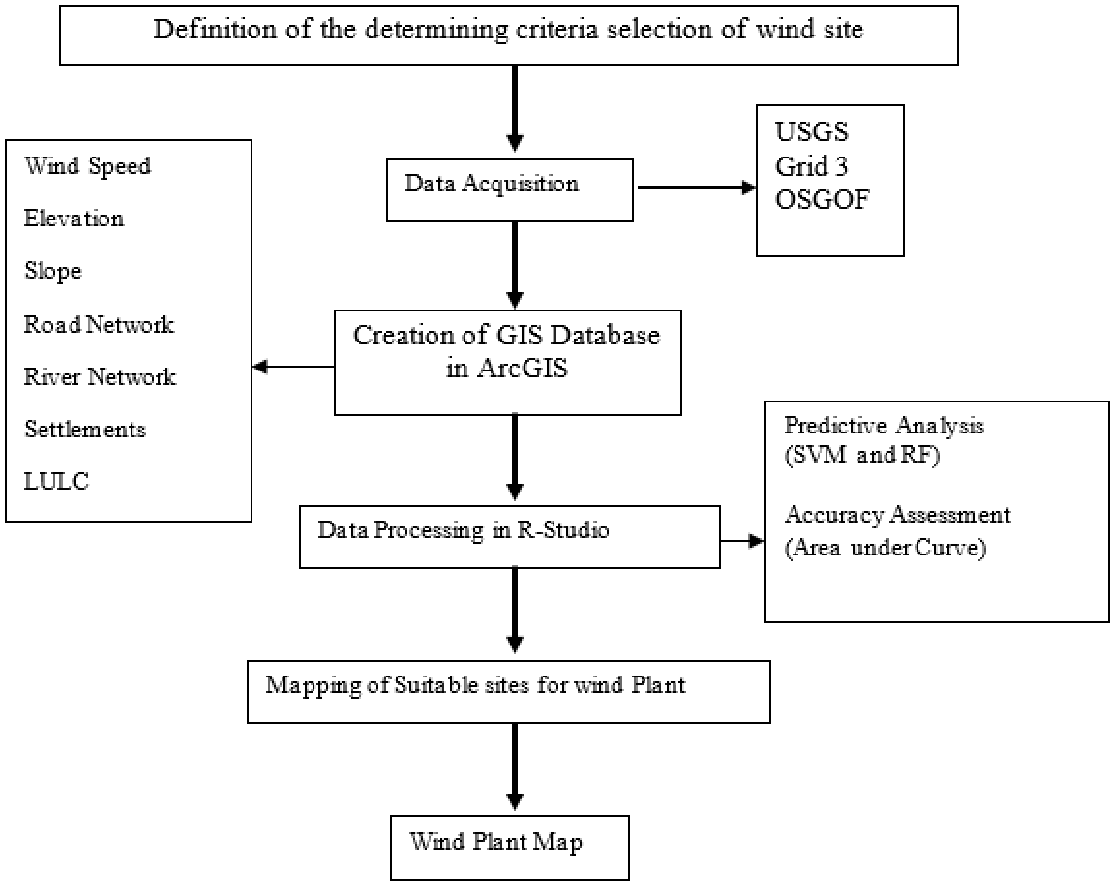


Figure 2: Methodology Flow chart

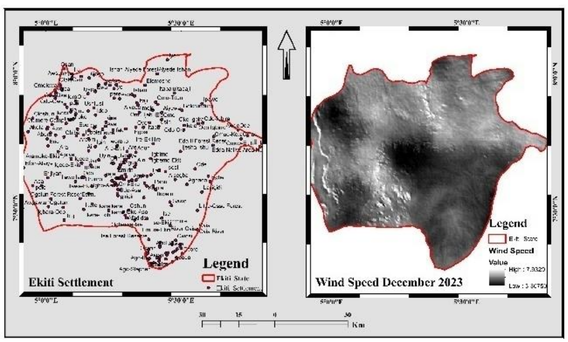
3.2 Criteria for Wind site selection

3.2.1 Wind Speed

The primary determinant for selecting the site for a wind power plant was the wind speed. Most wind turbines commence operation with optimal efficiency at 3 m/s and cease operation at a speed of 25 m/s. Wind speeds exceeding 3.5 m/s were deemed favorable. Consequently, the initial step involved excluding locations where winds speed of 3.5 m/s or higher was not expected. (Benti et al 2023).

3.2.2 Distance to Settlement

The proximity of wind power plants to human settlements can have adverse effects on the local population. Issues such as noise pollution from wind turbine operations, the undesirable effects of shadow flickers on residents, a reduction in wind speed, and the potential hindrance to the development of future residential areas underline the importance of maintaining a reasonable distance from inhabited areas. To mitigate environmental harm, it is recommended to observe a maximum distance of 7.5 kilometers between wind power plants and settlement areas. This guideline aims to safeguard the well-being of local communities and minimize the negative impacts associated with wind energy infrastructure. (Benti et al., 2023).



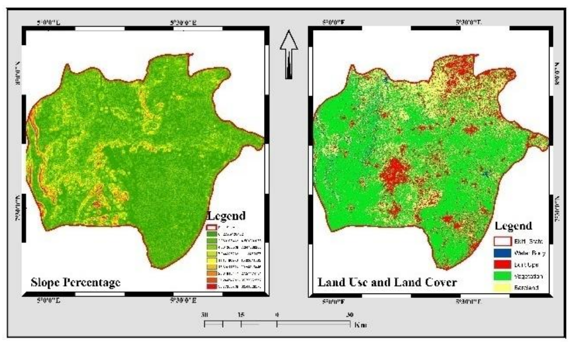


Figure 3 (A) (B) Figure 4 (A) (B)

Figure 2: (A) shows the settlement distribution of the study area. Figure 3:(A) shows the slope percentage of the study area,

(B) Shows the wind distribution of the study area in December 2023. (B) Shows the land use and land cover of the study area

3.2.3 Distance from Roads

The ideal location for wind farms is considered to be as close as possible to main roads. This proximity aims to minimize various discomforts, including the negative impact on road mobility due to loud noises and changes in the visual scene caused by the rotation of wind turbines during operation. Additionally, locating wind farms near main roads is seen as advantageous for reducing transportation costs and facilitating easier access for various employees. To implement this criterion, various studies suggest that the minimum distance between wind farm projects and main roads should be 1.67 km. This distance serves as a guideline to strike a balance between maximizing the benefits of wind energy generation and minimizing potential adverse effects on road infrastructure and visual aesthetics. (Benti et al., 2023).

3.2.4 Distance from rivers

Water bodies such as rivers, lakes, and wetlands are considered unsuitable for hosting wind farm sites due to their vital ecological services. These bodies of water hold significant ecological and economic value, often serving as habitats for diverse flora and fauna species. To enhance the safety of wind farm facilities, it is recommended to maintain a considerable distance from riverbeds. This precaution is essential as river routes are dynamic, subject to constant changes, and there is a potential risk of flooding.

As a guideline, renewable energy projects should not be constructed within 300 meters of water bodies. In this study, a protective buffer of 0.6 kilometers was established around water bodies, ensuring that the areas proximate to these buffered watercourses were excluded from the study area. This approach aims to respect and protect the ecological and dynamic nature of water bodies while avoiding potential hazards associated with flooding or other changes in river dynamics. (Benti et al., 2023).

3.2.5 Slope

The slope of the terrain is a critical technical factor that requires careful consideration in the selection of wind farm sites. Steep slopes pose challenges for access, leading to increased costs for maintenance and equipment installation. In the establishment of wind farms, flat and low-slope areas are often recommended to mitigate the challenges associated with construction. Regions with a slope exceeding 10% are excluded from the final suitability map. This exclusion ensures that areas with steep slopes, which can be difficult to navigate and construct wind farms on, are not considered suitable for development. By focusing on flat and low-slope terrain, the aim is to enhance accessibility, reduce construction complexities, and ultimately optimize the efficiency and cost-effectiveness of wind farm projects. (Benti et al., 2023).

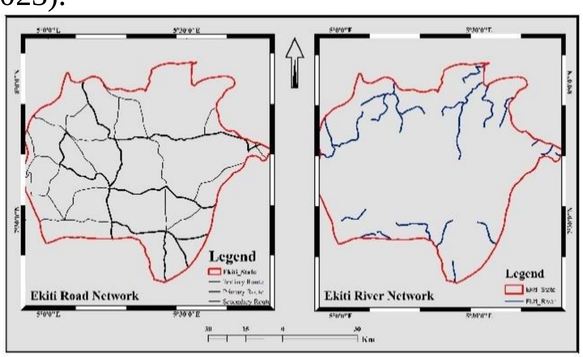


Figure 4:(A) shows the Road Network of the study area, (B) shows the River Network of the study area.

3.2.6 Land use/land cover (LULC)

Land use is a pivotal factor in the decision-making process for energy investments, especially when considering wind energy installations. Priority is given to areas where the impact of wind turbines on current land use is minimal. The choice of a wind farm location is significantly influenced by land use restrictions, prohibiting the construction of wind farms in certain areas despite adequate wind speeds. Unsuitable locations may include forests, wetlands, aviation zones, archaeological sites, and more. The most suitable types of land for wind farm installation are agricultural land, grassland, barren land, and shrubland, while forested land is considered less suitable. Notably, this study excluded wetlands, water sources, and settlements from consideration, recognizing that it would be inappropriate to construct wind farms in these areas. This approach acknowledges the importance of preserving specific land uses and ensures responsible decision-making in the development of wind energy projects. (Benti et al., 2023).

3.3 Machine Learning Model

3.3.1 Support Vector Machine Model

Support Vector Machine (SVM) is a non-linear, data-driven technique that has gained popularity, particularly for its superior performance compared to traditional Error-Reduction Minimization (ERM) methods used in conventional neural networks. The strength of SVM lies in its ability to minimize the upper bound on anticipated risk, which mitigates the impact of reducing training data error. This characteristic provides SVM with a greater capacity to generalize functions compared to neural networks. The non-linear nature of SVM allows it to handle complex relationships in data, making it effective in various applications, including wind power prediction in this study. (Pandit and Kolios, 2020). Support Vector Machines (SVMs) were originally designed for objective or optimal classification, commonly referred to as Support Vector Classification (SVC). However, in more recent applications, SVMs have been adapted for regression tasks and are known as Support Vector Regression (SVR). This section provides a theoretical description of SVM regression models. The fundamental concept behind SVM regression involves mapping the input data, denoted as x, into a high-dimensional feature space using a nonlinear mapping. Once the data is transformed, a linear regression is performed within this feature space. This approach allows SVM to capture complex relationships and patterns in the data, making it suitable for regression tasks where the goal is to predict continuous outcomes rather than discrete classes. The use of SVR extends the versatility of SVM to handle various types of predictive modeling, including regression applications. (Mohandes et al., 2004).

3.3.2 Random Forest

The Random Forest (RF) algorithm is widely recognized and highly effective for addressing both regression and classification problems. Introduced by Breiman in 2001, the algorithm is grounded in the concept of model aggregation. The foundation of RF lies in the combination of "bagging" (an idea presented in Breiman's work in 1996) and the incorporation of random feature selection introduced by Ho in 1998. The central principle of RF involves the creation of an ensemble comprising numerous binary decision trees. These trees are constructed using multiple bootstrap samples derived from the learning dataset (L). At each node in the trees, a random subset of explanatory factors (x) is selected, contributing to the diversity of the individual trees. Crucial parameters for RF models include the number of trees in the ensemble and the choice of predictors used to determine the splits at each node, factors that significantly influence the performance of the Random Forest algorithm (Vorpahl et al., 2012).

4. Results

4.1 Factors Considered for Wind Plants

Wind Speed

Figure 6b indicates that the wind speed in the study area ranges from 3 to 7 km/s. The suitability classification suggests that areas with a wind speed greater than 4.0 km/s are considered suitable for a wind farm plant, while other areas are deemed less suitable. This approach aligns with the common practice in wind energy projects, where areas with higher wind speeds are typically preferred for optimal energy generation. The specified threshold of 4.0 km/s serves as a criterion for determining the suitability of different regions to wind speed in the context of your analysis.

Distance to Settlement

The most suitable region is identified as an area greater than 16 km away from a settlement area. This is represented as "very good" in Figure 6a. Table 2 provides further details, showing the distribution of road proximity to settlement areas. It seems that the study considers a greater distance from settlement areas as more favorable, possibly to minimize potential impacts or conflicts associated with wind farm construction. The specified threshold of 16 km suggests a particular criterion for determining the suitability of regions concerning their proximity to settlements in the context of your analysis.

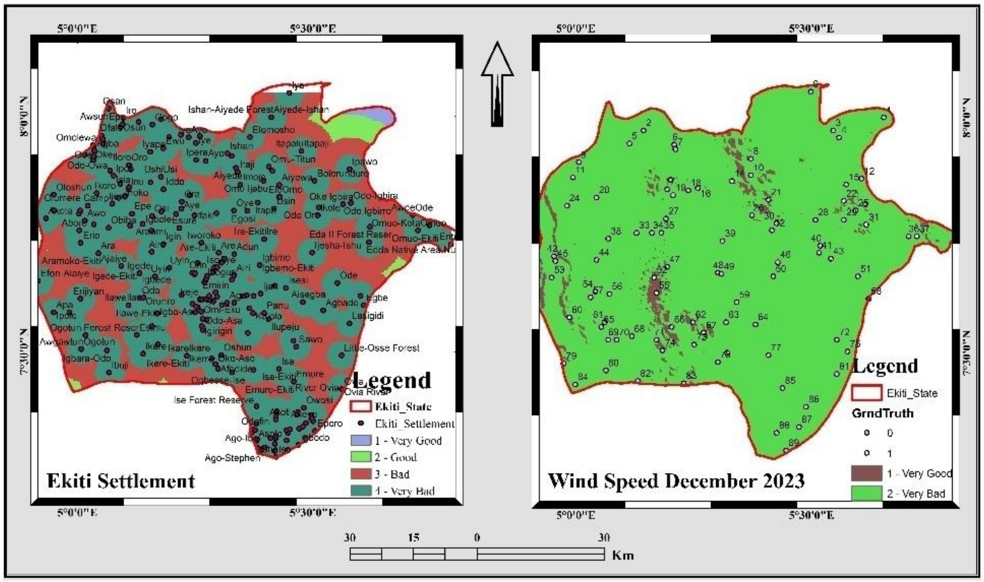


Figure 5: (A) Settlement proximity to suitable wind farm site, (B) wind speed validity

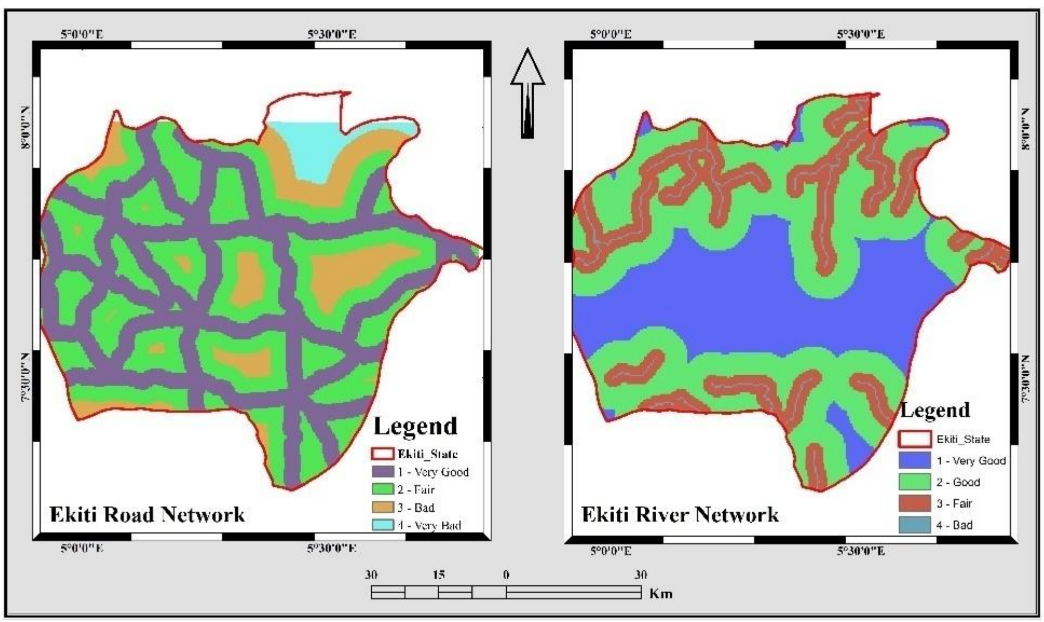


Figure 6: (A) Road Network Proximity, (B) River Network Proximity

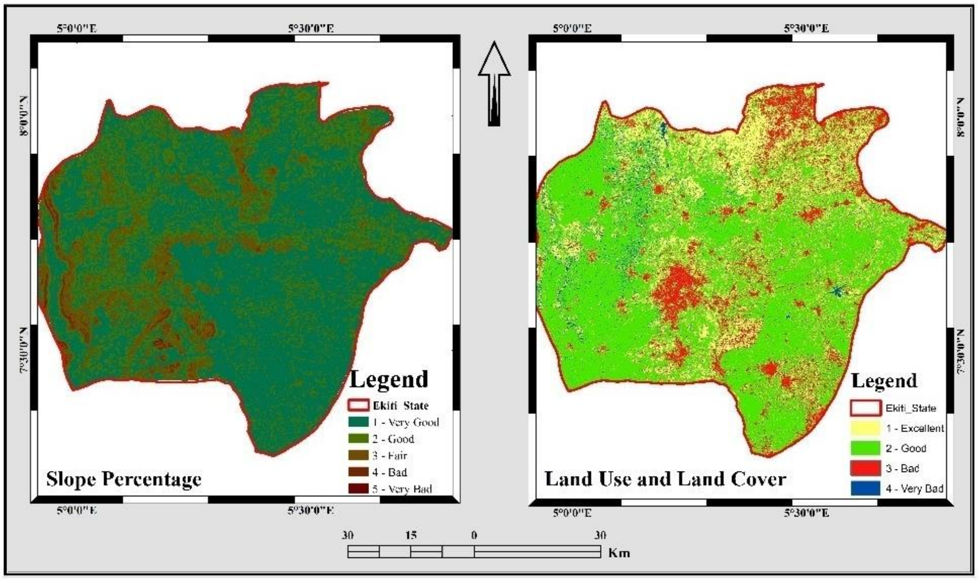


Figure 7: (A)slope distribution for suitable wind farm site, (B) Land use and Land Cover Validity

Table 2: Showing the distribution of factors used in the study

Rank	Distance Settlement	Distance from Road	Distance from River	Slope %	LULC	Remark
1	> 16km	< 1.67 km	> 7 km	< 10	Bare land	**Very Good**
2	7 – 16km	1.67 – 3 km	4 – 7 km	10 – 20	Vegetation	**Good**
3	3- 7 km	3- 5 km	600 m – 4km	20 – 30	Built-up	**Bad**
4	0 – 3 km	> 5 km	< 600 m	> 30	Waterbodies	**Very Bad**

Distance from Road

In this study, distance plays a crucial role in determining the accessibility to the wind farm. Figure 7a indicates that areas with a distance of less than 1.67 km are classified as "very good," suggesting that closer proximity is considered highly favorable. On the other hand, areas with distances greater than 5 km are graded as "not suitable." Table 2 provides further details on the distribution of areas based on these distance classifications. This approach emphasizes the significance of accessibility, with the study favoring regions that are closer to the specified distance threshold for optimal wind farm site selection.

Distance from rivers

In this study, the most suitable region, considering the distance to the river, is identified as being 7 km away from the river. The Euclidean distance metric was employed for this determination, and it is visually represented as "very good" in Figure 7b. Additionally, Table 2 indicates that areas with distances less than 600 meters are considered not suitable for the wind farm. This suggests that, based on the specified criteria, maintaining a minimum distance of 7 km from the river is considered optimal for wind farm site selection, ensuring the suitability of the chosen region. Areas closer than 600 meters from the river are likely excluded due to certain constraints or considerations mentioned in your study.

Slope

In Figure 8a, areas that are depicted as flat are considered suitable for wind farm development. The suitability is measured on a scale ranging from 0^0 to 10^0, where regions with a value greater than 10 are considered less suitable. This indicates that the flat areas, characterized by values below 10, are considered more favorable or appropriate for the implementation of a wind farm plant in the context of your study. The specific threshold of 10 suggests a criterion for determining the suitability of flat areas for wind energy projects in this analysis.

Land use/land cover (LULC)

The results presented in Figure 8b indicate that bare land and vegetation are more suitable for the implementation of a wind farm plant. This suggests that, in the context of this study, areas characterized by bare land and vegetation are deemed more favorable or appropriate for establishing wind farm facilities. This finding aligns with the common practice in wind energy planning, where the choice of land types, such as bare land and vegetation, is often preferred due to factors like ease of construction, lower environmental impact, and better utilization of wind resources.

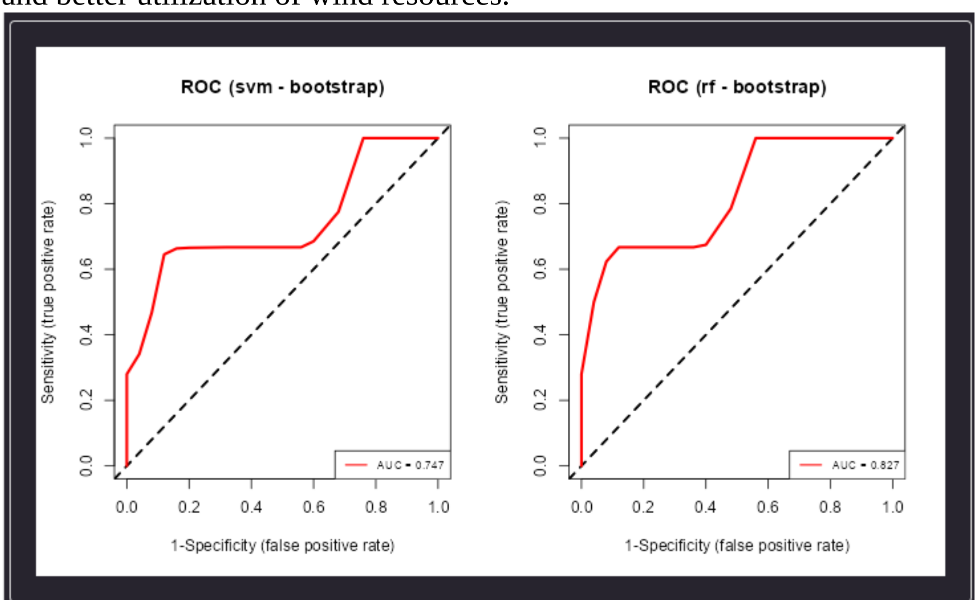


Figure 8: Area under Curve for SVM and RF

4.2 Machine Learning Model

Accuracy of Model

The Area Under the Curve (AUC) is a metric that ranges from 0 to 1, representing the accuracy of a predictive model. In Figure 9, it is observed that the Random Forest model's prediction for wind farm plant site selection exhibits a higher accuracy with an AUC of 0.827. In comparison, the Support Vector Machine (SVM) Model Prediction shows a slightly lower accuracy with an AUC of 0.747. This indicates that, based on the AUC metric, the Random Forest model is performing better in predicting suitable sites for wind farm construction than the Support Vector Machine model in this particular study.

Table 3: Area Under Cover

Methods	AUC
SVM	0.75
RF	0.83

Table 4: show RF and SVM results and percentage of area covered

	Value	RF Area (ha)	RF %	SVM Area(ha)	SVM %
1	Very Good	1174.04	0.23	2758.42	0.55
2	Good	2484.86	0.49	9164.34	1.81
3	Fair	13871.9	2.74	28929.23	5.72
4	Bad	97160.26	19.2	182899.29	36.14
5	Very Bad	391365.28	77.34	282305.05	55.79

0.23% of the random forest model of the entire study area is most suitable for wind plant farm compare to 0.55% of the support vector machine model, while 0.49% and 1.81% of RF and SVM respectively is under the good category while 77.34% of the RF and 55.79% of the SVM is not suitable as labeled in table 4. As shown in Figure 10, the map of the model is being revealed.

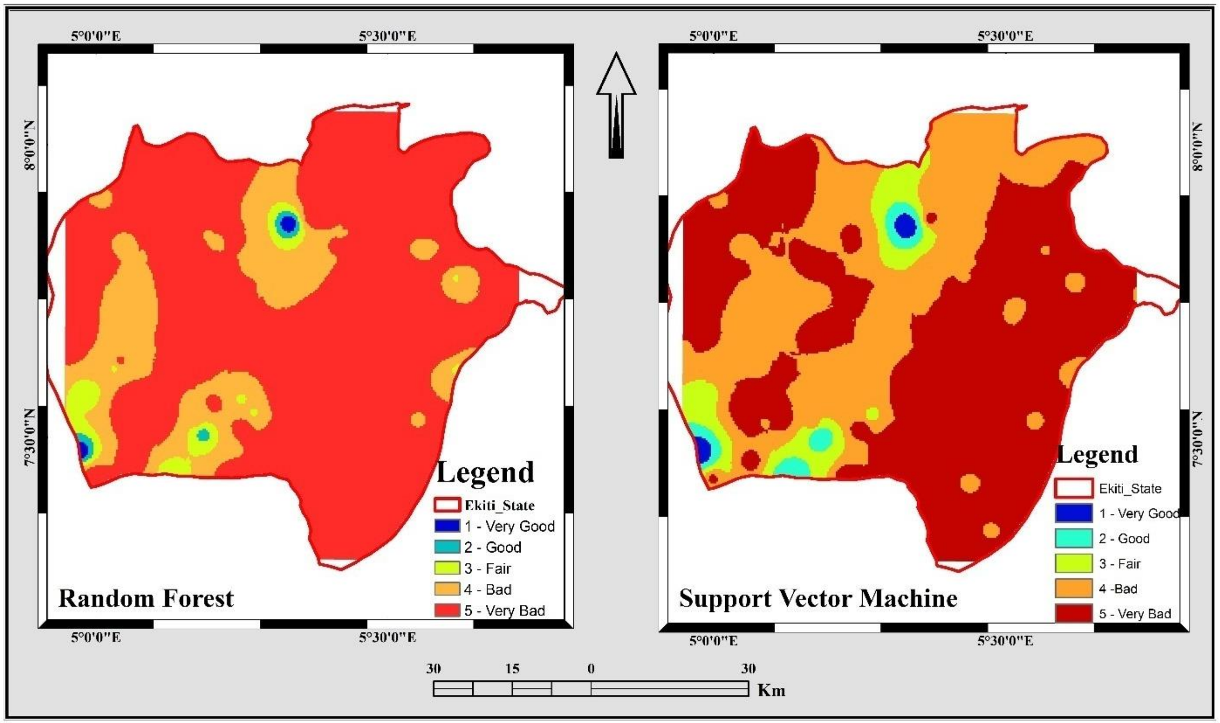


Figure 9: Random Forest and Support Vector Machine Result

5. Conclusion

In conclusion, this study conducted in Ekiti state successfully employed a GIS-based machine learning approach to identify the most suitable sites for wind farm plants. Utilizing predictive models, namely Support Vector Machine (SVM) and Random Forest (RF), enhanced the precision of wind farm site selection. The inclusion of six factors in creating training and testing datasets for model verification and validation contributed to the robustness of the analysis. The validation process, utilizing the Area under the Curve (AUC) metric, revealed AUC values of 0.75 for SVM and 0.8 for RF. Notably, the Random Forest model exhibited the highest predictive capability, suggesting its superiority in this context. Ultimately, the integration of GIS and machine learning models, especially the application of the Random Forest algorithm, proved to be an effective method for assessing the potential sites for wind farm plants in the study area.

References

Benti, N. E., Alemu, Y. B., Chaka, M. D., Semie, A. G., &Mekonnen, Y. (2023). *Site Suitability Assessment for the Development of Wind Power Plant in Wolaita area, Southern Ethiopia: An AHP-GIS Model* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-2474836/v1

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Li, L.-L., Zhao, X., Tseng, M.-L., & Tan, R. R. (2020). Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. *Journal of Cleaner Production*, *242*, 118447. https://doi.org/10.1016/j.jclepro.2019.118447

Mohandes, M. A., Halawani, T. O., Rehman, S., & Hussain, A. A. (2004). Support vector machines for wind speed prediction. *Renewable Energy*, *29*(6), 939–947. https://doi.org/10.1016/j.renene.2003.11.009

Moltames, R., Naghavi, M. S., Silakhori, M., Noorollahi, Y., Yousefi, H., Hajiaghaei-Keshteli, M., &Azizimehr, B. (2022). Multi-Criteria Decision Methods for Selecting a Wind Farm Site Using a Geographic Information System (GIS). *Sustainability*, *14*(22), Article 22. https://doi.org/10.3390/su142214742

Olorunfemi, I., Fasinmirin, J., Olufayo, A., &Komolafe, A. (2020). GIS and remote sensing-based analysis of the impacts of land use/land cover change (LULCC) on the environmental sustainability of Ekiti State, southwestern Nigeria. *Environment, Development and Sustainability*, *22*. https://doi.org/10.1007/s10668-018-0214-z

Pandit, R., &Kolios, A. (2020). SCADA Data-Based Support Vector Machine Wind Turbine Power Curve Uncertainty Estimation and Its Comparative Studies. *Applied Sciences*, *10*(23), Article 23. https://doi.org/10.3390/app10238685

SeyedAlavi, S. M., Maleki, A., &Khaleghi, A. (2022). Optimal site selection for wind power plant using multi-criteria decision-making methods: A case study in eastern Iran. *International Journal of Low-Carbon Technologies*, *17*, 1319–1337. https://doi.org/10.1093/ijlct/ctac009

Vorpahl, P., Elsenbeer, H., Märker, M., &Schröder, B. (2012). How can statistical models help to determine driving factors of landslides? *Ecological Modelling*, *239*, 27–39. https://doi.org/10.1016/j.ecolmodel.2011.12.007

Wang, C., &Nehrir, M. H. (2008). Power Management of a Stand-Alone Wind/Photovoltaic/Fuel Cell Energy System. *Energy Conversion, IEEE Transactions On*, *23*, 957–967. https://doi.org/10.1109/TEC.2007.914200

Yürek, Ö.,BiRant, D., &Yürek, İ. (2021). MakineÖğrenmesiAlgoritmalarınıKullanarakRüzgarEnerjisiÜretimiTahmini. *DeuMuhendislikFakultesi Fen veMuhendislik*, *23*(67), 107–119. https://doi.org/10.21205/deufmd.2021236709