Integrating Analytical Hierarchical Process with Random Forest for Geospatial Optimization of Renewable Energy Sites

Victor C. NNAM, Nigeria, Joseph O. ODUMOSU, Namibia, Uche IKWEZU, Nigeria

Key words: *A*HP; MCDA; Machine Learning; Random Forest Model; Renewable Energy Site

SUMMARY

This study presents a comprehensive approach for optimizing site selection for renewable energy installations using a combination of Multi-Criteria Decision Analysis (MCDA) and machine learning techniques. The study focuses on the installation of hybrid solar-wind systems, with an emphasis on solar photovoltaic (PV) as the primary energy source due to its high regional potential within the study area. The key factors considered in this study were wind speed, solar potential, slope, elevation, land use, distance to road networks, and distance to transmission lines. The Analytic Hierarchy Process (AHP) was employed to prioritize these criteria based on expert judgment, generating a weighted overlay map that showed potential sites for renewable energy development. To enhance prediction accuracy and validate the AHP results, a machine learning model, specifically the Random Forest classifier, was implemented. Hence, the methodological approach was such that AHP was used to derive weights from expert judgment, GIS was used for the spatial data processing, while RF was used to refine predictions and also validate the AHP outputs. The model achieved high accuracy (96%), demonstrating its effectiveness in refining site suitability analysis. Feature importance analysis revealed that land use, solar potential, and proximity to roads were the most influential factors. Also, the statistical evaluations (Receiver Operating Characteristic -Area Under the Curve (ROC-AUC) score, confusion matrix, and regression analysis), further validated the model's robustness and predictive capabilities. The findings of the study revealed the potential of combining expert-driven methods with data-driven techniques to identify optimal sites for renewable energy projects. This hybrid approach not only improves the precision of site selection but also provides information about the factors influencing renewable energy development. The study offers a practical framework for policymakers and planners to support sustainable energy initiatives.

Integrating Analytical Hierarchical Process with Random Forest for Geospatial Optimization of Renewable Energy Sites

Victor C. NNAM, Nigeria, Joseph O. ODUMOSU, Namibia, Uche IKWEZU, Nigeria

1.0 INTRODUCTION

Consequent upon the rapidly increasing global population leading to increasing energy needs, there is now a growing demand for renewable energy across the world more than ever. This demand is further intensified by the global push for sustainable energy sources (IRENA, 2023). Most geographic locations with abundant solar, wind, and biomass resources portray significant potential for renewable energy development. However, this potential cannot be harnessed without proper geospatial optimization in the selection of suitable sites for such activities (Adaramola, Agelin-Chaab, & Paul, 2021). Optimizing site selection for renewable energy infrastructure remains a complex challenge that requires balancing environmental, technical, and socio-economic criteria (Malik, Chauhan, & Kumar, 2020). Traditional approaches to site selection often rely on static methods, which fail to capture the dynamic and multidimensional nature of the decision-making process (Al Garni & Awasthi, 2017; Zou, Zhao, Zhang, & Xiong, 2020).

In a bid to provide a near-real-life situation for this determination, the Analytical Hierarchical Process (AHP) has been extensively utilized as a Multi-Criteria Decision-Making (MCDM) tool for renewable energy site selection. This is due to AHP's ability to prioritize competing criteria through pairwise comparisons (Jahan & Edwards, 2015; Ahmad & Tahar, 2018). Despite its popularity, AHP, being a subject and expert-based analysis method, has limitations, such as sensitivity to judgment inconsistencies and challenges in scaling for large datasets (Al Garni & Awasthi, 2017; Malik et al., 2020). Machine Learning (ML) approaches on the other hand are rather objective (data driven) methods, and are capable of uncovering complex, non-linear relationships among variables, processing large volumes of heterogeneous data, and offering higher generalizability. When applied to renewable energy site selection, ML models such as Random Forest (RF) can complement the subjectivity of AHP by validating expert-derived weights with empirical evidence thus, improving the robustness and accuracy of suitability assessments (Ali, Kumar, & Singh, 2022; Choubey, Sharma, & Babu, 2022). The integration of AHP with ML methods therefore provides a hybrid decision-making framework that balances expert knowledge with data-driven insights, allowing for more reliable and scalable solutions in geospatial energy planning.

For these reasons, integrating AHP with advanced machine learning (ML) techniques, such as the RF algorithm, has become a major focus in geospatial optimization for renewable energy site selection in recent times (Ali et al., 2022; Choubey et al., 2022). Such integration is often preferred because it produces results that incorporate the advantages of both techniques. Random Forest, is a robust ML method that builds multiple decision trees and aggregates their prediction to improve accuracy and reduce overfitting. It is an often-preferred ML method because of its efficiency in handling geospatial big data and identifying the unique patterns within such large volumes of data (Ali et al., 2022; Choubey et al., 2022).

Park et al. (2019) had applied RF for the assessment of wind turbine site suitability in South Korea, incorporating factors such as wind speed, slope, and distance to roads. The study demonstrated that Random Forest outperforms traditional methods in terms of prediction accuracy and robustness. Ahmad and Tahar (2018) combined AHP with ML techniques, including RF, to identify optimal locations for renewable energy projects. They noted that RF's ability to process large geospatial datasets complements AHP's subjective weighting, leading to robust decision-making frameworks. These findings were later validated by Choubey et al. (2022), who reviewed various applications of machine learning in renewable energy and highlighted the synergy between AHP and RF. Their study emphasized how RF can address inconsistencies in AHP’s pairwise comparisons by iteratively refining criteria weights based on real-world data, particularly for wind farm site selection in areas with variable wind patterns. Also, Ali et al. (2022) demonstrated the integration of AHP with RF to optimize solar power plant locations. AHP was used to assign weights to criteria such as solar irradiation, proximity to infrastructure, and land use, while RF validated the results by analyzing historical and spatial data. This approach improved predictive accuracy and reduced decision-making biases. Earlier studies by Zou et al. (2020) had applied hybrid AHP-RF models for land suitability assessments and showed their effectiveness in integrating socio-economic and environmental criteria. The study found that RF significantly enhanced the spatial resolution and accuracy of AHP-derived suitability maps, particularly in heterogeneous landscapes. Also, Malik et al. (2020) conducted a comparative analysis of MCDM approaches, including AHP-RF integrations. They identified that while AHP is effective in prioritizing criteria, its integration with RF helps mitigate over-reliance on expert judgments by introducing data-driven validation mechanisms. Mehrian, Qelichi, & Tahouri (2024) used ML techniques, including Random Forest, to assess the suitability of solar energy sites in Iran. The study highlighted how ML can enhance site selection by incorporating complex interactions between multiple criteria and providing more reliable predictions. These studies revealed that AHP-RF integration offers several advantages, including improved scalability, enhanced predictive capabilities, and reduced biases in renewable energy site selection.

Given the advantages of this integration, a methodological workflow for its adoption within the study area is herein proposed for geospatial optimization of renewable energy sites. The methodology proposed combines the strengths of AHP in multi-criteria evaluation with the predictive power of RF. The proposed approach relies on Geographic Information Systems (GIS) to facilitate the spatial analysis of critical factors, including land use, solar irradiation, proximity to infrastructure, and socio-economic considerations (Chiemelu et al., 2021). This study adapts these existing methodologies in a unique regional context by evaluating site suitability for a hybrid solar-wind system, with a focus on solar PV as the primary energy source due to its higher regional potential (Zou et al., 2020; Sun et al., 2024).

Hybrid solar-wind modeling, which integrates both solar and wind energy resources in site selection, enhances energy reliability by taking advantage of the complementary nature of their temporal patterns. This is because solar peaks during the day, while wind often intensifies at night or in different seasons (Ali et al., 2022). This approach ensures more consistent energy output, reduces reliance on storage systems, and improves system efficiency through shared infrastructure and lower costs of energy (Choubey et al., 2022). However, it introduces methodological complexity, requiring advanced spatial and statistical techniques to balance resource-specific criteria (Ahmad & Tahar, 2018; Malik et al., 2020).

2.0 STUDY AREA

Southeastern Nigeria (Figure 1), the focus of this study, is a geographical region comprising five states: Abia, Anambra, Ebonyi, Enugu, and Imo. The region is known for its diverse landscape, demographic characteristics, and significant potential for renewable energy development, particularly in wind and solar energy (Nzekwe, Oladejo, & Emoh, 2018; Okay, Oluwakunmi, & Jonathan, 2017). The region lies between approximately 4° 40' N to 7° 10' N. It also spans from 6° 30' E to 8° 30' E. These coordinates encompass a variety of terrains, from coastal plains in the south to hilly areas and plateaus in the northern parts. The topography of Southeastern Nigeria is varied, with a mix of lowland plains, rolling hills, and highlands.

The region experiences a tropical rainforest climate with two distinct seasons: the wet season (April to October) and the dry season (November to March). Average annual rainfall ranges from 1,500 mm to 2,500 mm, with the highest rainfall recorded in coastal areas. The average annual temperature is around 26°C to 28°C, with minimal seasonal variation (Igboekwe & Nwankwo, 2011)

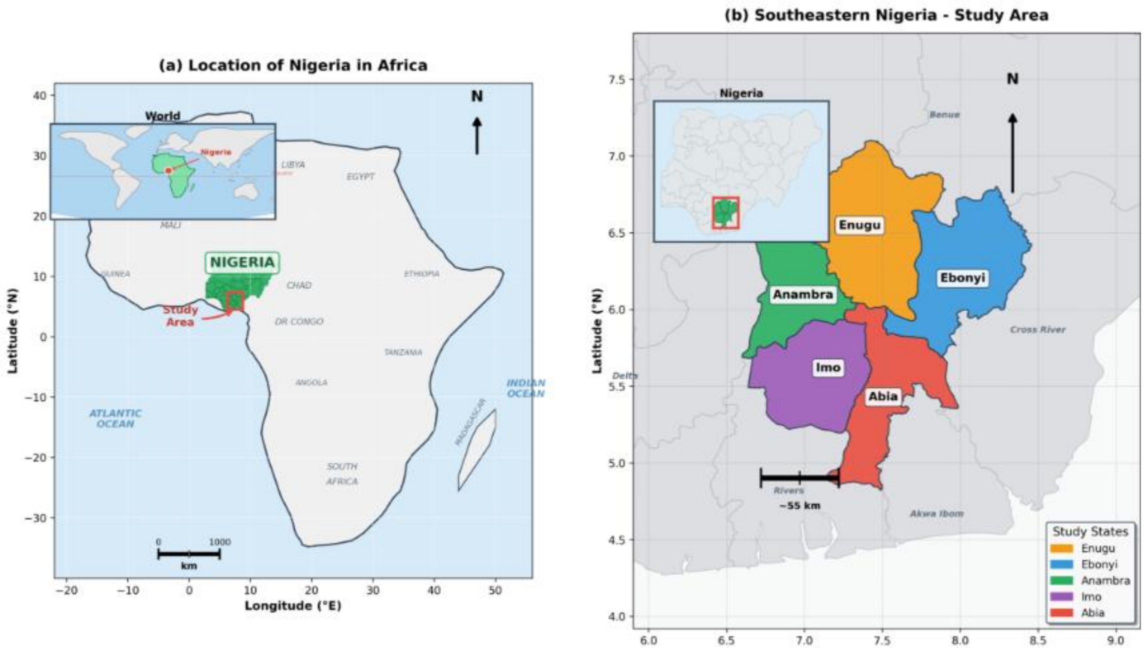

Figure 1: (left) location of Nigeria in Africa and the world (inset), (right) study area showing the five southeastern states of Nigeria.

3.0 MATERIALS AND METHODS

3.1 Materials

In determining optimal sites for renewable energy, environmental factors such as wind speed, solar potential, land use/land cover (LULC), road networks, and existing transmission lines were considered, and their individual roles as well as pair-wise analysis were considered. Wind speed is a key parameter in assessing the feasibility of wind energy projects, as sustained and consistent wind velocities are essential for economic energy production. Studies have shown that areas with average wind speeds above 3 m/s are suitable for small-scale wind energy, while utility-scale projects require speeds exceeding 6 m/s (Adaramola et al., 2021; Malik et al., 2020). Solar potential, measured in terms of solar irradiation, is equally significant for solar power projects, with regions receiving more than 5 kWh/m²/day being ideal for photovoltaic systems (Ali et al., 2022).

Land use and land cover (LULC) assessments ensure the identification of feasible sites while minimizing conflicts with agricultural lands, forests, or urban areas. This factor is crucial in avoiding land-use conflicts and preserving ecological balance (Ahmad & Tahar, 2018). Road networks and proximity to existing infrastructure, such as transmission lines, are also vital considerations. Sites closer to roads and power grids reduce the cost of logistics and energy transmission, enhancing the overall economic viability of renewable energy projects (IRENA, 2023). Slope and elevation significantly influence renewable energy site selection by impacting energy generation potential, construction feasibility, and costs. Higher elevations often provide better wind speeds for wind energy projects due to reduced surface friction, while moderate slopes ensure easier installation and maintenance (Baban & Parry, 2001). For solar energy, south-facing slopes in the Northern Hemisphere optimize sunlight exposure, though steep slopes may increase installation costs (Uyan, 2013). Overall, sites with gentle slopes and appropriate elevations are preferred to balance energy output and construction logistics (Malczewski, 2004).

All these identified environmental conditions were obtained from heterogeneous sources and combined in this study. A list of the data sets, their sources, and specific uses is provided in Table 1 below:

Table 1: Materials used

Data	Source	Use
Wind Speed	Global Wind Atlas	Provides information on the wind speeds.
Solar Potential (PV Output)	Global Solar Atlas	Indicates the amount of solar energy that a particular area receives.
DEM	USGS Earth Explorer	To derive Slope and Elevation for assessing terrain suitability
Landsat 8 (2024)	USGS Earth Explorer	Extracting land use and land cover
Road network	HDX	Proximity to roads for accessibility and logistics in constructing and maintaining renewable energy sites
Transmission lines	World Bank Catalog	Proximity to existing electrical infrastructure reduces the cost of connecting renewable energy sites to the grid

3.2 Methods

This study has been carried out using a two-step solution approach. The solution workflow implemented in this study is provided in Figure 2. The workflow diagram provides an implementable framework for integrating AHP and RF in the optimal selection of suitable sites for dual-resource modeling (hybrid wind-solar) energy in any given location.

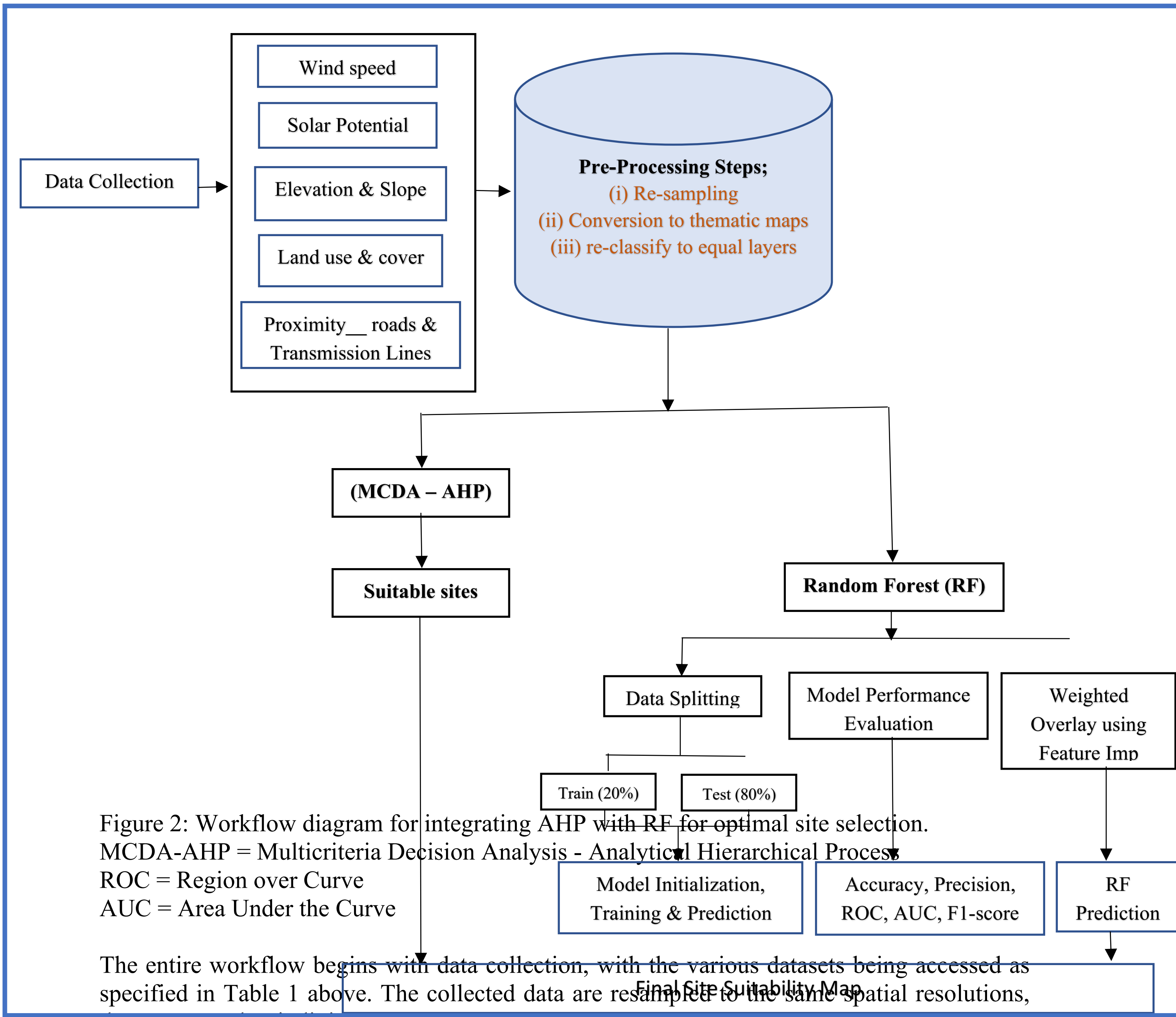

Figure 2: Workflow diagram for integrating AHP with RF for optimal site selection.
MCDA-AHP = Multicriteria Decision Analysis - Analytical Hierarchical Process
ROC = Region over Curve
AUC = Area Under the Curve

The entire workflow begins with data collection, with the various datasets being accessed as specified in Table 1 above. The collected data are resampled to the same spatial resolutions, then converted to individual thematic maps with an equal number of classes. Upon conversion to thematic maps, MCDA-AHP is first used for the optimization of suitable sites. Thereafter, the results obtained from MCDA-AHP are validated using the RF technique, which similarly used to optimize suitable sites for renewable energy within the study area.

Upon completion of site selection, statistical analyses were performed to validate the results and assess the relationships between different criteria. The correlations between various criteria were analyzed to identify potential collinearity or redundancy in the data. A multiple regression analysis was conducted to model the relationship between the suitability score (dependent variable) and the independent criteria. This analysis helped to quantify the influence of each criterion on site suitability. The final output of this study was a geospatial map that shows optimal sites for renewable energy development. This map integrated the results of both GIS-based MCDA and machine learning techniques, providing a more robust and data-driven decision-making tool for site selection.

3.2.1 Multicriteria Decision Analysis via Analytical Hirarchical Process (MCDA – AHP)

The Analytic Hierarchy Process (AHP) was employed to assign relative weights to each of the seven criteria considered in the site suitability analysis. The entire AHP process workflow is shown in Fig. 3. These weights reflect the importance of each factor in determining suitability for hybrid solar-wind energy systems, with a primary focus on solar photovoltaic (PV) deployment due to the region's high solar potential (Chiemelu et al., 2021). The criteria were selected based on their relevance to either or both technologies, as outlined below:

(i) Solar Potential (SP): This is the most critical factor for PV systems and strongly influences energy yield. Higher irradiance directly increases photovoltaic output (Ahmad & Tahar, 2018).

(ii) Wind Speed (WS): Included to assess the potential for supplementary wind energy generation. Although wind speeds in Southeastern Nigeria are moderate, this criterion supports the hybrid system rationale (Adaramola et al., 2021).

(iii) Slope (S): Slope is particularly important for PV installations. Flat to moderately sloped terrain simplifies panel installation, reduces construction costs, and optimizes panel orientation (Uyan, 2013).

Elevation (E): Elevation affects wind resource availability, as higher altitudes typically experience stronger and more consistent winds. It also indirectly influences solar exposure in complex terrain.

Land Use/Land Cover (LULC): Critical for both technologies to avoid unsuitable areas such as forests, wetlands, or densely populated regions. PV systems especially require large, unshaded open land.

Distance to Roads (DR): Accessibility is vital for construction and maintenance of energy infrastructure. It supports logistics for both PV and wind turbine deployment.

Distance to Transmission Lines (DTL): Proximity to the grid enhances economic viability for both technologies by reducing infrastructure extension costs.

The pairwise comparison matrix was constructed based on these technology-specific justifications. For instance, the higher weight assigned to solar potential and LULC reflects their critical roles in PV deployment, while wind speed and elevation were assigned moderate weights to account for their relevance in a potential hybrid system. The consistency ratio (CR) of the matrix was verified to be < 0.1, confirming acceptable judgment consistency (Messaoudi, et al, 2019).

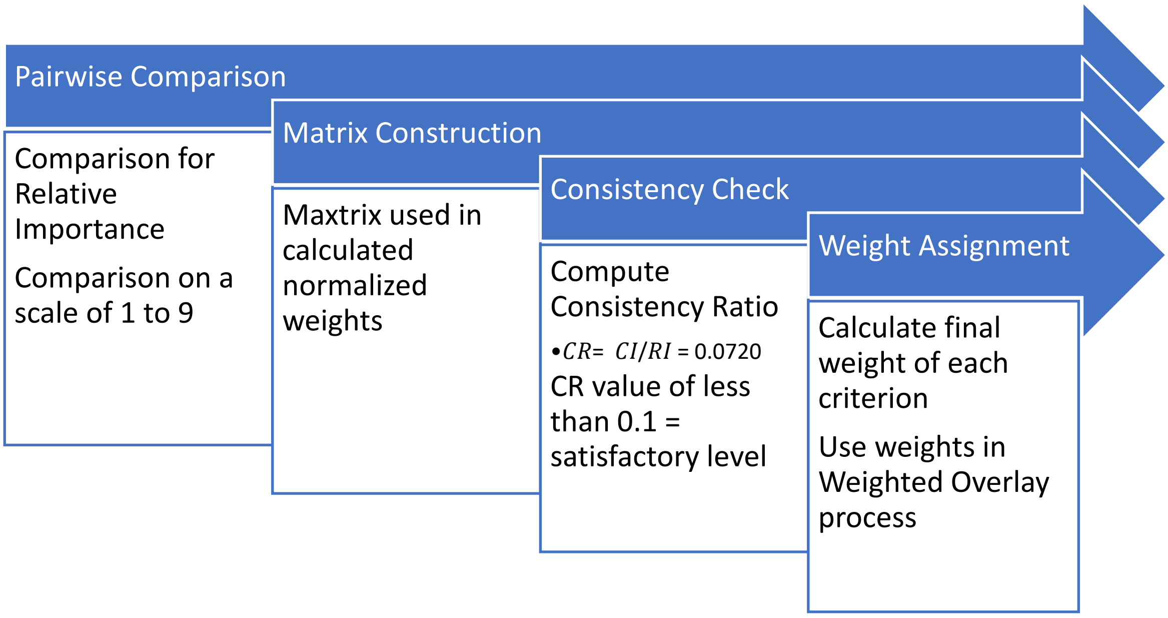

Figure 3: The AHP process

Following the assignment of technology-informed AHP weights, a weighted overlay analysis was conducted within a Geographic Information System (GIS) environment to operationalize the Multi-Criteria Decision Process (MCDP). This approach enabled the integration of multiple reclassified raster layers, each representing a key suitability criterion. The weights derived from AHP (reflecting the relative importance of each factor for solar PV and hybrid applications), were systematically applied to the corresponding layers. The weighted overlay process produced a composite suitability map, where each cell value represents a weighted sum of the underlying criteria, highlighting areas most favorable for renewable energy development. The pairwise comparison matrix and the resulting normalized and final criterion weights are presented in Tables 2 and 3, respectively.

Table 2: Pairwise Comparison Matrix

Criteria	WS	SP	S	E	LULC	DR	DTL
WS	1	3	5	7	4	6	7
SP	1/3	1	4	5	3	7	6
S	1/5	¼	1	3	2	5	4
E	1/7	1/5	1/3	1	1/2	4	3
LULC	1/4	1/3	1/2	2	1	3	5
DR	1/6	1/7	1/5	¼	1/3	1	2
DTL	1/7	1/6	1/4	1/3	1/5	1/2	1

Table 3: Normalized Weights and Criterion Weights

Criteria	WS	SP	S	E	LULC	DR	DTL

WS	0.447	0.589	0.443	0.376	0.362	0.226	0.25
SP	0.149	0.1963	0.3545	0.269	0.271	0.2641	0.214
S	0.089	0.0491	0.088	0.161	0.1813	0.1886	0.1428
E	0.0638	0.0392	0.0295	0.0538	0.0453	0.151	0.107
LULC	0.1118	0.0654	0.0443	0.1076	0.0906	0.1132	0.178
DR	0.0745	0.0280	0.0177	0.0134	0.029	0.0377	0.0714
DTL	0.0638	0.0327	0.0221	0.0177	0.0181	0.0188	0.0357
Weights	**0.385**	**0.1287**	**0.245**	**0.0699**	**0.101**	**0.038**	**0.0298**

Wind Speed (WS), Solar Potential (SP) Slope (S), Elevation (E), Land Use/Land Cover (LULC), Distance to Road Network (DR), Distance to Transmission Lines (DTL)

3.2.2 Random Forest Modeling

To enhance the robustness and predictive accuracy of the site suitability assessment, a RF machine learning model was integrated into the workflow. This model was trained using the same input variables employed in the AHP-GIS analysis, namely reclassified raster layers representing solar potential, wind speed, slope, elevation, land use/land cover, and proximity to infrastructure, while the target variable was derived from the suitability classes generated by the AHP-based weighted overlay. The dataset was randomly split into training (80%) and testing (20%) subsets to ensure unbiased evaluation. The RF algorithm constructed an ensemble of 200 decision trees, combining their outputs through majority voting to improve classification stability and reduce overfitting (Raji et al., 2023). Beyond prediction, the RF model also provided a quantitative measure of feature importance, identifying the most influential criteria in determining site suitability for hybrid solar-wind energy deployment. These quantitative measures complement the expert-derived AHP weights and allow for a data-driven validation of the multi-criteria decision process. The configuration parameters used in the RF model are summarized in Table 4.

Table 4: Random Forest Model Parameters and Configurations

Parameter	**Description**	**Default Value**	**Chosen Value**
n_estimators	Number of trees in the forest.	100	200
max_depth	Maximum depth of each tree.	None	15
min_samples_split	Minimum number of samples required to split an internal node.	2	5
min_samples_leaf	Minimum number of samples required to be at a leaf node.	1	2

random_state	Seed for random number generation to ensure reproducibility.	None	42

The feature importance metric in the Random Forest model was computed based on the mean decrease in Gini impurity. The Gini impurity is a measure of how each variable contributes to the homogeneity of the nodes and leaves across all decision trees in the ensemble. It quantifies how each feature contributes to reducing classification uncertainty in the ensemble of decision trees. For every split in each tree, the algorithm records how much the impurity is reduced by using a particular feature. These reductions are then averaged over all trees and normalized to yield a relative importance score for each input variable (equ 1).

$$G = 1 - \sum_{i=1}^{n} {p_1}^2 \qquad (1)$$

Where

p_1 = probability of class I at a given node,

n = number of classes.

When a feature is used to split a node, the reduction in impurity is recorded. The feature importance score for a feature x_j, is then computed as;

$$FI(x_j) = \frac{1}{T} \sum_{i=1}^{T} \sum_{n \in S_i(x_j)} \Delta G_{n,t} \qquad (2)$$

Where

T = total number of trees

$S_i(x_j)$ = set of all splits of feature x_j in tree t

$\Delta G_{n,t}$ = reduction in Gini impurity at split s in tree t.

Features with higher scores are those that contributed more significantly to accurate predictions of site suitability. This process enables the identification of dominant factors, being land use, solar potential, and slope. The high influence of these features was corroborated by both domain expertise (via AHP) and empirical learning (via RF). Finally, the accuracy of the predictions were examined using the following metric;

$$Accuracy = \frac{(TP+TN)}{Total} \qquad (3.1)$$

$$Precision = \frac{TP}{(TP+FP)} \qquad (3.2)$$

$$Recall = \frac{TP}{(TP+FN)} \qquad (3.3)$$

$$F1 - score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \qquad (3.4)$$

Where

TP = True positive
FP = False positive
FN = False negative
F1-score = harmonic mean of precision and recall
The results provide a dual validation framework, thus confirming the reliability of the site selection methodology.

4.0 RESULTS

The results from the AHP-based Multi-Criteria Decision Analysis (MCDA) and the Random Forest (RF) modeling are illustrated in Figures 4(a) and 4(b), respectively. Figure 4(a) presents the suitability map generated through the AHP-GIS weighted overlay, where each spatial criterion, i.e., solar potential, wind speed, slope, elevation, land use/land cover (LULC), distance to road networks, and distance to transmission lines, was assigned a weight based on its relevance to solar PV and hybrid system deployment. The overlay results reveal that large areas in Anambra, Imo, and Abia states are unsuitable, primarily due to the dominance of urban settlements, forested landscapes, and wetlands, which limit the feasibility of installing ground-mounted renewable energy infrastructure.

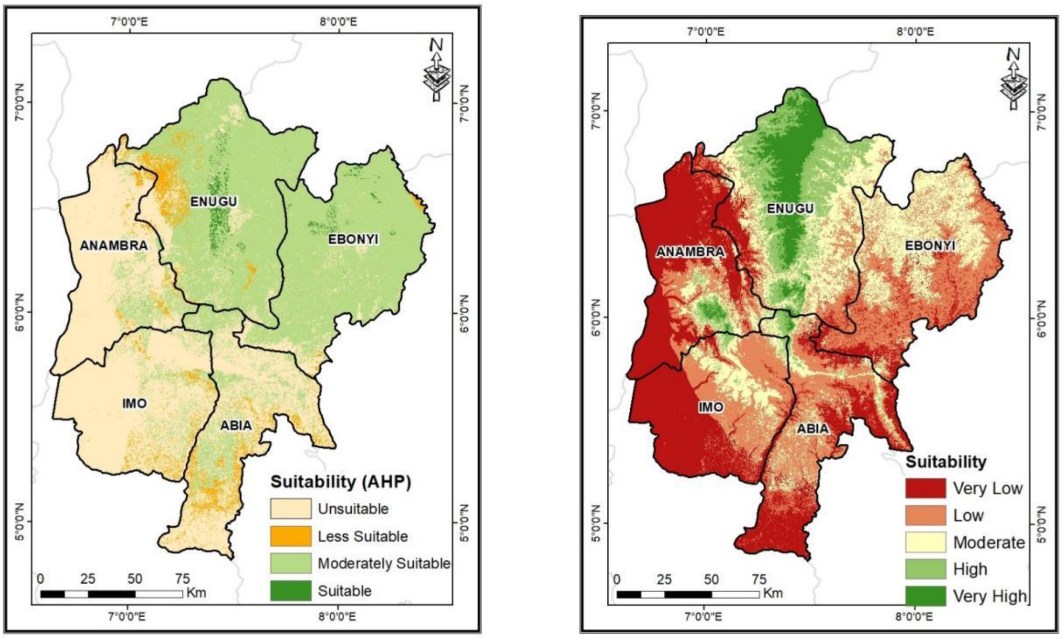

Figure 4: Suitability maps for renewable energy site selection in Southeastern Nigeria. (a) Multi-Criteria Decision Analysis via Analytical Hierarchical Process (MCDA-AHP) output showing four suitability classes (b) Random Forest (RF) machine learning classification output showing five suitability classes.

Areas classified as highly suitable were typically located in parts of Enugu and northern Ebonyi states. These areas exhibited favorable combinations of strong solar irradiance, flat terrain, non-restrictive land cover, and proximity to infrastructure, making them optimal for solar PV or hybrid installations.

Moderately suitable zones reflect locations where one or more key criteria (e.g., wind speed or slope) were suboptimal, though such areas may still be viable for small-scale or dual-resource systems. Conversely, areas with low composite suitability scores—such as mountainous regions, dense forests, or flood-prone zones—were deemed less appropriate due to environmental sensitivities, access limitations, and reduced resource availability. These classifications are further quantified in Table 5(a).

Table 5(a): Percentage Composition from MCDA - AHP

Suitability	***Percentage %***
Not Suitable	46.07
Less Suitable	4.95
Moderately Suitable	48.08
Suitable	0.884

Similar to Figure 4(a), the RF results are shown in Figure 4(b). The RF model identified clusters of high and very high suitability primarily in Enugu and southern Anambra, indicating them as regions with optimal conditions for solar PV deployment and potential for hybrid energy systems. In contrast, moderate suitability zones were more widespread across Ebonyi and Imo, where solar irradiance and infrastructure proximity are favorable, but other limiting factors—such as land cover or slope—may reduce viability. According to Table 5(b), approximately 15% of the study area was classified as having high to very high suitability (High: 9.81% + Very High: 5.43% = 15.24%), emphasizing a relatively constrained but promising zone for renewable energy development.

To assess the model's predictive performance, a confusion matrix was generated using a validation dataset. The RF classifier correctly identified 38 instances as True Positives (i.e., locations predicted and confirmed as suitable) and 58 instances as True Negatives (i.e., correctly classified unsuitable locations). These results demonstrate the model's strong alignment with ground-truth suitability labels and support its reliability in distinguishing between viable and non-viable sites for renewable energy infrastructure.

Table 5(b): Percentage Composition from RF Modelling

Suitability	***Percentage %***
Very Low	30.82
Low	31.01
Moderate	22.91
High	9.81
Very High	5.43

The RF model demonstrated high predictive performance with a low misclassification rate. Specifically, only four instances were incorrectly labeled as suitable when they were, in fact, unsuitable, representing false positives (Unsuitable Predicted as Suitable). Importantly, the model registered no false negatives, meaning all truly suitable sites were accurately identified. This outcome reflects a perfect recall (100%) for the suitable class, indicating the model's strong sensitivity in detecting viable locations. The overall performance is summarized in the confusion matrix (Figure 5), from which key evaluation metrics were derived: an accuracy of 96.67%, a precision of 90.48%, and a balanced F1-score of 95%. These values confirm the model's robustness and suitability for supporting renewable energy site selection through data-driven classification.

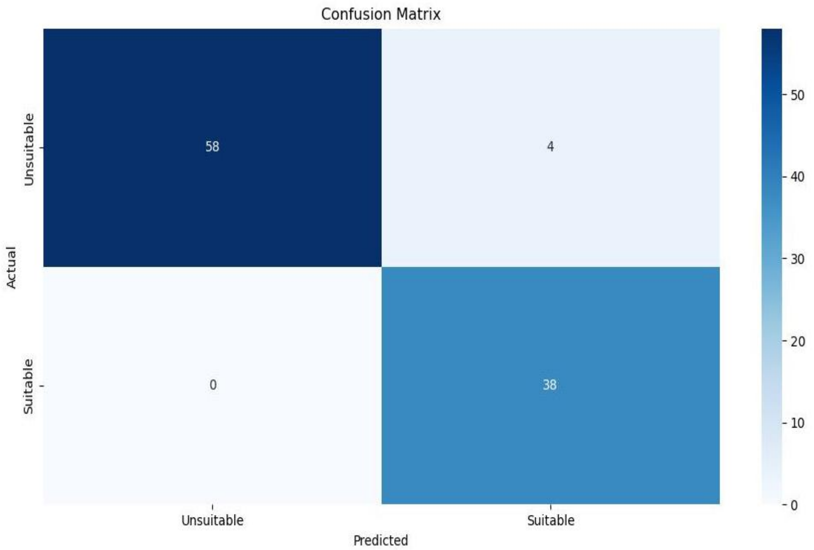

Figure 5: Confusion matrix

The balanced F1-score is calculated as the harmonic mean of precision and recall, providing a balanced measure of classification performance. The confusion matrix (Figure 5) displays the absolute counts of predictions, with 38 true positives, 58 true negatives, 4 false positives, and 0 false negatives. To complement the classification performance metrics, a feature importance analysis was conducted (Figure 6) to assess the relative influence of each criterion used in the model. This helped identify the most impactful variables driving suitability predictions, further validating the integration of expert judgment with data-driven information.

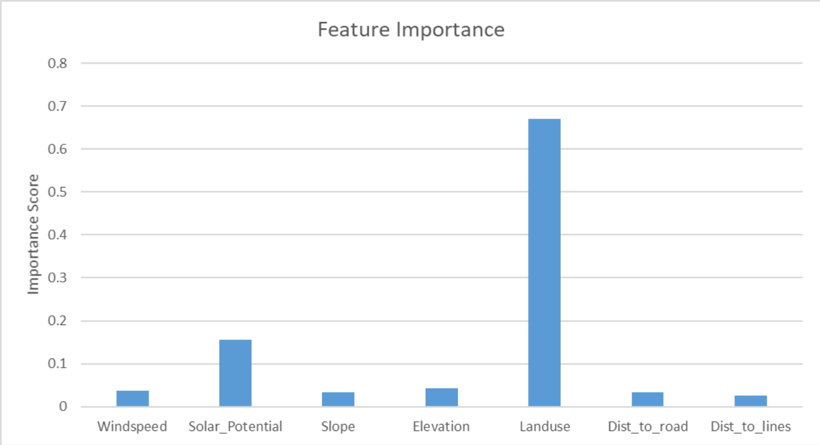

Figure 6: Feature Importance

The feature importance analysis (Figure 6) revealed that Land Use/Land Cover (LULC) was the most influential criterion in the Random Forest model, contributing 67.09% to the overall prediction. This reveals the critical role of land suitability in determining the feasibility of renewable energy infrastructure (Adebimpe & Usman, 2022). Solar potential followed with 15.54%, reflecting the importance of high solar irradiance in driving photovoltaic system performance (Chiemelu et al., 2021). Other variables such as elevation (4.26%), wind speed (3.71%), and slope and proximity to roads (each around 3.4%) had moderate influence, supporting their secondary roles in site logistics and resource availability. Distance to transmission lines, at 2.61%, was the least impactful, likely due to the relative uniformity and availability of grid infrastructure across the region.

To further assess classification performance, the model's Receiver Operating Characteristic (ROC) curve was plotted, with the Area Under the Curve (AUC) reaching a perfect score of 1.0 (Figure 7). This indicates excellent discriminative ability, with the model achieving a high true positive rate and minimal false positive rate across thresholds—confirming its reliability and robustness in differentiating between suitable and unsuitable locations.

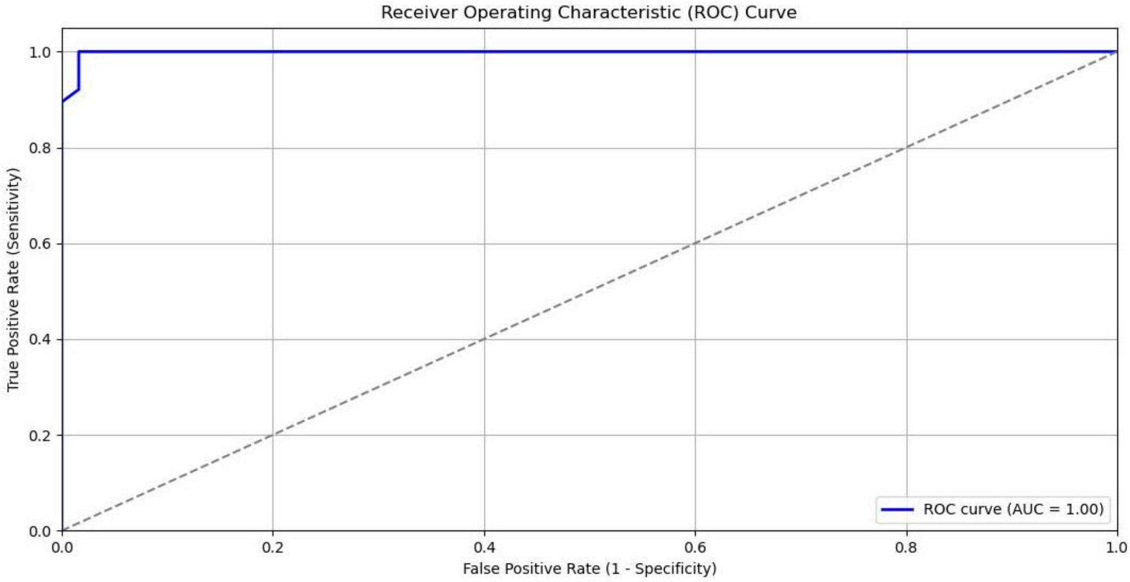

Figure 7: Model Validation

4.1 Discussion of Results

4.1.1 Correlation Analysis of Input Criteria

To ensure the absence of multicollinearity among the input variables used in both AHP and Random Forest modeling, a Pearson correlation analysis was conducted (Figure 8). The results showed low to moderate correlation coefficients across the variables. The highest observed correlation was between land use and solar potential (r = 0.42), indicating a moderate association likely due to the open land types favoring both solar exposure and ease of infrastructure development. A similar moderate relationship was found between distance to roads and distance to transmission lines (r = 0.39), reflecting overlapping infrastructure corridors. Importantly, all pairwise correlations remained below the multicollinearity threshold of 0.7, supporting the independence of criteria and validating their inclusion in the model. This analysis strengthens the reliability of both the AHP weight assignments and the predictive consistency of the RF classifier.

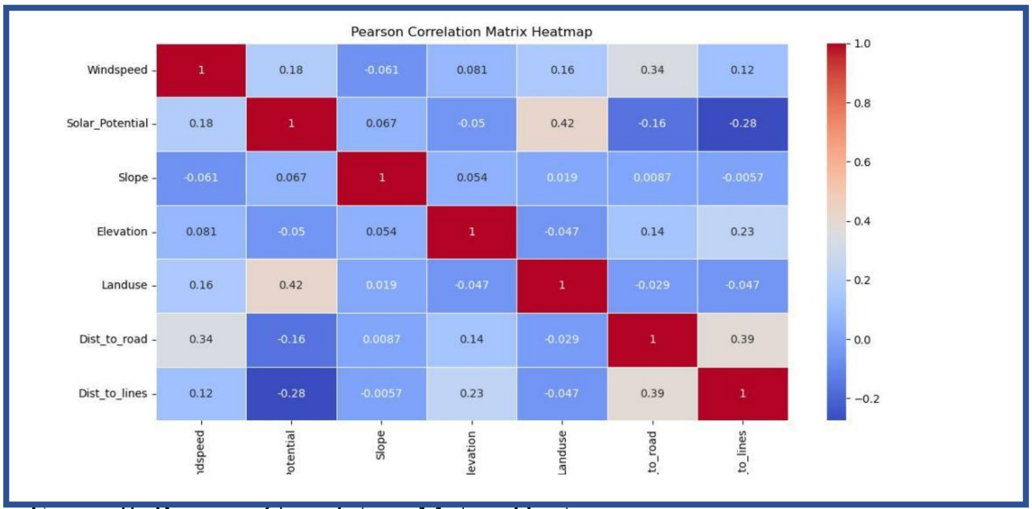

Figure 8: Pearson Correlation Matrix Heatmap

4.1.2 Regression Analysis of FR Model Outputs

To quantify the predictive power of the Random Forest model, a multiple linear regression was performed using the model's suitability scores as the dependent variable and the seven conditioning factors as independent variables. The regression (Table 6) yielded an R-squared value of 0.510, indicating that approximately 51% of the variability in suitability scores could be explained by the selected criteria. While this reflects a moderate level of explanatory power, it also suggests that other unmodeled factors, such as economic cost, land tenure, or environmental impact, may influence site suitability and could be explored in future research. The adjusted R-squared value of 0.503 accounts for the number of predictors, implying model parsimony with minimal overfitting. A statistically significant F-statistic (73.04) further confirms the overall validity of the regression model.

Table 6: Regression analysis results

Metric	Value
R-Squares	0.510
Adjusted R-squared	0.503
F-statistic	73.04

4.1.3 Comparative Evaluation of AHP and RF Results

The integration of AHP and Random Forest provided a dual-layered approach to site suitability evaluation, thus, providing a balance between expert-driven and data-driven methods (Faisal, et al., 2021). The AHP method facilitated structured weighting of factors based on relevance to solar PV and hybrid energy deployment, while the RF model refined these assessments using empirical relationships in the data (Kehinde, et al., 2018). Comparatively, the AHP-derived suitability map emphasized areas with favorable land use, flat terrain, and strong solar potential, while the RF model further quantified the predictive influence of these variables. Notably, Land Use/Land Cover emerged as the most critical factor in both methods, reinforcing its dominance in determining renewable energy site feasibility. The strong alignment between AHP and RF outputs (Table 7) not only validates the analytical framework but also shows the utility of hybrid methodologies for improving the accuracy and transparency of geospatial energy planning.

Table 7: Comparison of AHP and RF Outputs

Criterion	RF Importance (%)	AHP Weight (%)
Land Use / Land Cover (LULC)	67.09	67.41
Solar Potential (SP)	15.54	12.87
Elevation	4.26	6.99
Wind Speed (WS)	3.71	3.5
Slope	3.4	2.45
Distance to Roads (DR)	3.39	3.8
Distance to Transmission Lines (DTL)	2.61	2.98
Sum	100	100

5.0 CONCLUSION

This study developed and applied an integrated geospatial framework that combines Analytic Hierarchy Process (AHP) and Random Forest (RF) machine learning to optimize site selection for renewable energy development in Southeastern Nigeria. The study integrated expert judgment with data-driven validation and addressed key spatial, environmental, and infrastructural factors influencing the feasibility of solar PV and hybrid energy systems.

The AHP method enabled the systematic weighting of criteria relevant to renewable energy siting, while the RF model refined these outputs through classification accuracy and feature importance analysis. The high model performance, reflected in a 96.67% accuracy and a perfect recall, confirms the reliability of the hybrid method. Notably, land use/land cover and solar potential emerged as the most critical factors, showing the need for land planning policies that support renewable infrastructure.

Conclusively, the study offers a scalable, replicable, and transparent decision-support framework for policymakers, planners, and investors aiming to advance sustainable energy deployment. It also highlights the value of integrating geospatial intelligence and artificial intelligence in solving complex environmental and energy planning challenges.

REFERENCES

(IRENA), International Renewable Energy Agency. (2023). *Renewable energy roadmap for Nigeria. Abu Dhabi: IRENA.*

Adaramola, M. S., Agelin-Chaab, M., & Paul, S. S. (2021). Techno-economic analysis of wind energy in Nigeria: The case of a rural community in the southeastern region. *Renewable Energy, 169. https://doi.org/10.1016/j.renene.2020.12.055*, 1044-1052. .

Adebimpe, O. A., & Usman, S. O. (2022). A GIS-AHP Approach for Evaluating Site Suitability for Solar Power Plant : A Case Study of Ewekoro LGA, Nigeria. *Nigerian Journal of Technology, 41(4)*, 680–691.

Ahmad, S., & Tahar, R. M. (2018). A GIS-based multi-criteria decision analysis for land suitability assessment for solar farms in Malaysia. *Renewable and Sustainable Energy Reviews, 82*, 1633–1644. https://doi.org/10.1016/j.rser.2017.07.034.

Al Garni, H. Z., & Awasthi, A. (2017). A hybrid AHP-PROMETHEE decision-making approach for renewable energy site selection. *Renewable Energy, 121*, 45-57. https://doi.org/10.1016/j.renene.2017.01.027.

Ali, H., Kumar, M., & Singh, P. (2022). Integration of Random Forest and GIS for site suitability of solar power plants. *Energy Reports, 8*, 491–499. https://doi.org/10.1016/j.egyr.2022.03.067.

Baban, S. M., & Parry, T. (2001). Developing and applying a GIS-assisted approach to locating wind farms in the UK. *Renewable Energy, 24(1)*, 59-71. https://doi.org/10.1016/S0960-1481(00)00169-5.

Chiemelu, N. E., Anejionu, O. C., Ndukwu, R. I., & Okeke, F. I. (2021). Assessing the potentials of largescale generation of solar energy in Eastern Nigeria with geospatial technologies. *Scientific African, 12*, e00771, https://doi.org/10.1016/j.sciaf.2.

Choubey, A., Sharma, A., & Babu, R. (2022). Machine learning techniques for renewable energy site optimization: A review of current trends and future prospects. *Journal of Cleaner Production, 350, 131515*, https://doi.org/10.1016/j.jclepro.2022.131515.

Faisal, A. A., Kafy, A. A., Al Rakib, A., Akter, K. S., Jahir, D. M., Sikdar, M. S., . . . Rahman, M. M. (2021). Assessing and predicting land use/land cover, land surface temperature and urban thermal field variance index using Landsat imagery for Dhaka Metropolitan area. *Environmental Challenges, 4(June)*, 100192. https://doi.org/10.1016/.

Igboekwe, M. U., & Nwankwo, C. N. (2011). Geostatistical Correlation of Aquifer Potentials in Abia State, South-Eastern Nigeria. *International Journal of Geosciences, 02 (4)*, 541–548. https://doi.org/10.4236/ijg.2011.24057.

Jahan, A., & Edwards, K. L. (2015). Multi-criteria decision analysis for supporting the selection of engineering materials in product design. *Advanced Engineering Materials, 17*, 219-239. https://doi.org/10.1002/adem.201400178.

Kehinde, O., Babaremu, K. O., Akpanyung, K. V., Remilekun, E., Oyedele, S. T., & Oluwafemi, J. (2018). Renewable energy in Nigeria - A review. *International Journal of Mechanical Engineering and Technology, 9(10)*, 1085–1094.

Kok, S. H., Abdullah, A., Jhanjhi, N. Z., & Supramaniam, M. (2019). A review of intrusion detection system using machine learning approach. *International Journal of Engineering Research and Technology, 12(1)*, 8–15.

Malczewski, J. (2004). GIS-based land-use suitability analysis: A critical overview. *Progress in Planning, 62(1)*, 3-65. https://doi.org/10.1016/j.progress.2003.09.002.

Malik, N., Chauhan, D. S., & Kumar, D. (2020). Multi-criteria decision-making approaches for renewable energy site selection: A review. *Energy Reports, 6*, 1508-1523. https://doi.org/10.1016/j.egyr.2020.06.001.

Mehrian, M. R., Qelichi, M. M., & Tahouri, H. (2024). Solar power plant site selection using fuzzy inference system: a case study in Iran. *International Journal of Environmental Science and Technology*, 1-24.

Messaoudi, D., Settou, N., Negrou, B., & Settou, B. (2019). GIS based multi-criteria decision making for solar hydrogen production sites selection in Algeria. *International Journal of Hydrogen Energy, 44(60)*, 31808–31831. https://doi.org/10.1016/j.ijhyden.

Nzekwe, J. U., Oladejo, E. I., & Emoh, F. I. (2018). Assessment of Factors Responsible for Successful Project Implementation in Anambra State , Nigeria. *Civil and Environmental Research, 7(8)*, 39–58.

Okay, O. O., Akinbile, C. O., & Adeyemo, A. J. (2017). Soil erosion in south eastern Nigeria: a review. *Scientific Research Journal*, 5(9), 30-37.

Raji, S. A., Akintuyi, A. O., Wunude, E. O., & Fashoto, B. (2023). A machine learning-based spatial statistical method for modelling different phases of gully development in South-Eastern Nigeria. *Ecological Informatics, 75*, https://doi.org/10.1016/j.ecoi.

Sun, Y., Li, Y., Wang, R., & Ma, R. (2024). Modelling potential land suitability of large-scale wind energy development using explainable machine learning techniques: Applications for China, USA and EU. *Energy Conversion and Management*, *302*, 118131.

Uyan, M. (2013). GIS-based solar farms site selection using analytic hierarchy process (AHP) in Karapinar region, Konya/Turkey. *Renewable and Sustainable Energy Reviews, 28*, 11-17. https://doi.org/10.1016/j.rser.2013.07.042.

Zou, C., Zhao, Q., Zhang, G., & Xiong, B. (2020). Application of hybrid decision-making techniques for renewable energy planning. *Renewable Energy, 145,*, 260-274. https://doi.org/10.1016/j.renene.2019.05.097.

BIOGRAPHICAL NOTES

CONTACTS

Name:	Professor Victor C. NNAM
Organisation:	University of Nigeria, Enugu Campus (UNEC)
Address	Department of Geoinformatic and Surveying Faculty of Environmental Sciences
City:	Enugu
COUNTRY:	Nigeria

Tel.:	+234 803 276 0910
Email:	victor.nnam@gmail.com

Name:	Dr. Joseph O. ODUMOSU
Organisation:	Namibia University of Science and Technology (NUST)
Address	Department of Land and Spatial Sciences
City:	Windhoek
COUNTRY:	Namibia
Tel.:	+264 8184 31932
Email:	jodumosu@nust.na

Name:	Surv. Uche IKWUEZE
Organisation:	University of Nigeria, Enugu Campus (UNEC)
Address	Department of Geoinformatic and Surveying Faculty of Environmental Sciences
City	Enugu
COUNTRY:	Nigeria
Tel.:	+234 810 043 4424
Email:	jerryuc2@gmail.com

APPENDIX

APPENDIX