

Article

Explainable GeoAI for Photovoltaic Site Suitability Assessment in Rajasthan, India: A Rule-Derived, Spatially Validated Decision-Support Framework

Chinmay Nischal ¹, Jagriti Gupta ¹, Shri Krishna Mishra ², Saurabh Singh ^{3,4} , Ram Avtar ^{4,5} ,
Fahdah Falah Ben Hasher ⁶, Zoe Kanetaki ⁷ , Antreas Kantaros ⁸  and Mohamed Zhran ^{9,*} 

¹ Department of Civil Engineering, Suresh Gyan Vihar University, Jaipur 302017, Rajasthan, India; chinmay.2345477@mygyanvihar.com (C.N.); jagriti.gupta@mygyanvihar.com (J.G.)

² Department of Mechanical Engineering, Suresh Gyan Vihar University, Jaipur 302017, Rajasthan, India; krishna.mishra@mygyanvihar.com

³ Department of Civil Engineering, Poornima University, Sitapura 302022, Rajasthan, India; saurabh.singh@poornima.edu.in or saurabhsingh@ees.hokudai.ac.jp

⁴ Faculty of Environmental Earth Science, Hokkaido University, Sapporo 060-0810, Japan; ram@ees.hokudai.ac.jp

⁵ Department of Civil Engineering, Chennai Institute of Technology, Chennai 600069, Tamil Nadu, India

⁶ Department of Geography and Environmental Sustainability, College of Humanities and Social Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia; fahdhasher@pnu.edu.sa

⁷ Department of Mechanical Engineering, University of West Attica, 12241 Athens, Greece; zoe kanet@uniwa.gr

⁸ Department of Industrial Design and Production Engineering, University of West Attica, 12244 Athens, Greece; akantaros@uniwa.gr

⁹ Public Works Engineering Department, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt

* Correspondence: mohamedzhran@mans.edu.eg

Abstract

The rapid transition toward renewable energy requires transparent and spatially explicit methods for identifying suitable photovoltaic (PV) development areas. This study develops a geospatial artificial intelligence (GeoAI) decision-support framework for PV site suitability assessment in Rajasthan, India. Eleven harmonized predictors were used: global horizontal irradiance (GHI), photovoltaic power output (PVOUT), temperature, wind speed, aerosol optical depth (AOD), elevation, slope, albedo, land use/land cover (LULC), distance to roads, and distance to power lines. Reference labels were generated from an explicit rule-derived suitability index, class thresholds, and exclusion logic; therefore, the machine-learning task was to reproduce a transparent suitability framework rather than to predict observed PV yield or project-level performance. Extreme Gradient Boosting (XGBoost) was compared with simpler baseline models, evaluated using random and spatial-block validation, and interpreted using SHapley Additive exPlanations (SHAP). Independent overlays with known solar-installation records, presence-background robustness testing, and uncertainty/sensitivity analysis were used to examine spatial plausibility, spatial autocorrelation, deterministic label effects, and parameter uncertainty. The resulting outputs include pixel-level suitability zones, contiguous candidate polygons, district-level capacity-oriented summaries, and planning-priority classes. The framework is intended as a risk-aware regional screening tool: high model agreement indicates consistency with the constructed suitability labels, while final project decisions require parcel-scale land, grid, environmental, social, and economic assessment.



Academic Editors: Sofia Spyridonidou and Vasiliki Tsoukala

Received: 28 April 2026

Revised: 12 June 2026

Accepted: 14 June 2026

Published: 18 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: GeoAI; solar photovoltaic; site suitability; XGBoost; SHAP; machine learning; Rajasthan; renewable energy; capacity estimation; spatial decision support

1. Introduction

Global energy demand is steadily increasing, and at the same time, concerns about climate change and the depletion of fossil fuels are pushing countries toward cleaner and more sustainable energy sources [1,2]. Among the available options, solar photovoltaic (PV) energy has become one of the most attractive solutions because it is widely available, environmentally friendly, and increasingly cost-effective. The success of solar energy projects depends heavily on where they are installed [3]. Selecting the right location is not straightforward, as it requires considering multiple factors such as solar radiation, terrain conditions, land availability, and access to infrastructure [4]. Therefore, identifying suitable sites for solar development has become a key step in energy planning.

PV site-suitability research has developed through three overlapping methodological streams: geographic information system and multi-criteria decision-making (GIS-MCDM) overlay analysis [4–7], hybrid geographic information system-artificial intelligence (GIS-AI) modeling [8–11], and explainable artificial intelligence (AI) or uncertainty-aware decision support [12–14]. The earliest and still most widely used stream combines GIS with MCDM methods such as analytic hierarchy process (AHP) and fuzzy AHP [5,7], weighted overlay [4], technique for order preference by similarity to ideal solution (TOPSIS), and simple additive weighting (SAW) [6,15]. These approaches are attractive because they are transparent, easy to communicate to planners, and well-suited for integrating heterogeneous criteria such as irradiance, slope, land cover, distance to roads, and distance to transmission lines [4–7]. However, recent reviews and comparative studies show that their outputs can vary substantially with criteria selection, expert weights, normalization rules, exclusion thresholds, class-boundary definitions, and the chosen MCDM algorithm, meaning that two technically defensible workflows may produce different priority zones for the same region [6,13–16].

Recent international work has shifted from purely additive GIS overlay toward hybrid approaches that combine spatial criteria with optimization, machine learning, deep learning, explainable AI, and scenario analysis [8–12]. Across these studies, the main methodological advance is not simply the use of a more complex algorithm but the attempt to represent nonlinear interactions among resource quality, terrain, land cover, accessibility, and climate-sensitive planning assumptions [8–12]. For example, nonlinear MCDM and GIS-AI studies show that suitability rankings can change when criteria interact rather than being treated as independent additive weights, while explainable-AI studies demonstrate the value of reporting which predictors drive mapped decisions [9–12]. The uncertainty associated with exclusion-threshold selection, suitability-class definition, and weighting has therefore been extensively examined in earlier GIS-MCDM, hybrid PV-siting, and broader geospatial-assessment studies [6,13–16]. At the same time, this literature also exposes unresolved weaknesses: many studies provide limited information on how reference suitability labels are generated, how random pixel splits may inflate reported accuracy through spatial autocorrelation, whether simpler models can reproduce the same decision logic, and how source-dataset uncertainty affects final suitability classes.

A related body of work uses satellite remote sensing and AI for PV potential assessment and generation-oriented analysis. Ghaedrahmati et al. [17] combined satellite-based remote sensing with artificial-intelligence techniques to assess PV power-plant potential and energy prediction, while Reddy et al. [18] used machine learning and climate-change

projections for long-term solar potential assessment. Such studies are valuable because they connect solar-resource mapping with future energy-planning questions. However, their objectives differ from site-screening studies: energy-prediction models require observed generation, plant design, and climate time-series information, whereas suitability-screening models primarily classify land and resource conditions. This distinction is important for the present manuscript because the XGBoost model is trained to reproduce transparent suitability labels, not to predict measured electricity production.

India-specific PV site-selection studies have also expanded, but most remain concentrated in either GIS-MCDM or hybrid decision-support frameworks. Rane et al. [19] used a GIS-based multi-influencing factor approach for photovoltaic plant siting in Nashik, India, emphasizing the role of terrain, land cover, and accessibility. Guntupalli et al. [13] developed a hybrid artificial neural network (ANN)-AHP-GIS framework with dimensionality reduction and uncertainty quantification for Southern India, showing that uncertainty-aware hybrid models can improve robustness relative to single-method overlay. These studies are important for the Indian context, yet they do not fully address the Rajasthan-scale planning problem considered here, where a large solar-resource region must be translated into contiguous candidate polygons, district-level capacity summaries, and development-priority classes while retaining transparent assumptions about labels and validation.

The literature therefore reveals a clear progression but also a set of unresolved gaps. Traditional GIS-MCDM studies are transparent but often sensitive to subjective weighting and algorithm choice [6,15,16]. Recent machine-learning and deep-learning studies improve classification flexibility but may provide limited detail on the origin of reference labels, the effect of spatial autocorrelation, and the difference between internal classification accuracy and independent field validation [8,11,12]. India-specific studies increasingly combine GIS, AI, and MCDM methods, but they less commonly connect suitability classes to contiguous candidate-site extraction, indicative capacity estimation, and district-level planning outputs [13,19]. Finally, many studies present final suitability maps without testing simpler model baselines or comparing mapped outputs against independent solar-installation records.

This study addresses these limitations by positioning geospatial artificial intelligence (GeoAI) as a transparent spatial decision-support framework rather than as a black-box predictor of PV yield. The proposed workflow explicitly employs the Extreme Gradient Boosting (XGBoost) model: it first constructs rule-based suitability labels from harmonized solar, climatic, terrain, land-cover, and infrastructure variables and then uses XGBoost to reproduce and map this defined suitability framework. The manuscript further compares XGBoost with simpler models, evaluates both random and spatial-block validation designs, applies SHapley Additive exPlanations (SHAP) to interpret predictor influence, overlays known solar-installation records as presence-based plausibility evidence, and converts High and Very High suitability pixels into contiguous candidate-site polygons for capacity-oriented district prioritization. In this way, the novelty is not the isolated use of XGBoost or SHAP, which has appeared in recent solar-suitability literature, but the integration of transparent label construction, explainable modeling, spatial validation, candidate-region extraction, and planning-ready capacity summaries in a single Rajasthan-focused workflow.

In this study, machine learning is therefore used to classify rule-derived suitability categories from geospatial predictors; it is not used to predict measured electricity generation or solar yield. SHAP is used only to interpret the relative contribution of predictors to the suitability classification. This distinction is important because a high classification score indicates agreement with the defined suitability framework, whereas real project performance would require independent production, grid, and economic data.

To address these challenges, this study proposes an integrated GeoAI-based framework for photovoltaic site suitability assessment in Rajasthan, India. The study has four specific research objectives: (1) to integrate multi-source solar, climatic, terrain, land-cover, and infrastructure datasets into a common geospatial feature space; (2) to generate transparent suitability labels and classify PV suitability zones using an XGBoost-based machine-learning model; (3) to interpret the contribution of each predictor using SHAP-based explainable AI; and (4) to extract contiguous candidate sites, estimate potential capacity, and summarize district-level development priorities. By linking these objectives, the study seeks to move from descriptive suitability mapping toward interpretable and planning-oriented spatial decision support.

The specific contribution of the study is fourfold: (i) it states the origin of the suitability labels and treats them as constructed planning labels rather than field-observed PV performance; (ii) it compares XGBoost with simpler baseline models and reports a spatial-block validation design to reduce over-reliance on random pixel splits; (iii) it combines SHAP interpretation with an external presence-based overlay of known solar installations; and (iv) it transforms pixel-level suitability classes into contiguous candidate regions and district-level capacity summaries. These elements directly respond to the main methodological weaknesses identified in the recent literature and make the workflow more useful for strategic solar-planning applications.

The remainder of the paper is organized as follows. Section 2 describes the study area, input datasets, label-generation procedure, XGBoost model, validation metrics, and candidate-site extraction method. Section 3 presents model performance, spatial suitability patterns, SHAP-based feature importance, candidate sites, and district-level capacity estimates. Section 4 discusses the planning implications, validation evidence, uncertainties, and limitations. Section 5 concludes the study and outlines future research directions.

2. Methodology

2.1. Study Area

The study area is Rajasthan, located in northwestern India between approximately 23°03' N–30°12' N and 69°30' E–78°17' E. It covers about 342,239 km² and is the largest state in India (Figure 1). The state includes arid desert landscapes in the west, semi-arid plains in the central region, and the Aravalli hill range toward the southeast [20,21].

Rajasthan was selected as the study area because official state policy identifies the state as a major solar-energy region and reports a National Institute of Solar Energy has assessed the solar potential of 142 GW [22]. Instead of relying on generalized climate statements, this manuscript uses the Global Solar Atlas GHI and PVOUT layers as the spatially explicit solar-resource evidence for the suitability analysis [23]. Topographically, western and central Rajasthan include extensive low-slope terrain, which can reduce land preparation and construction constraints for large-scale PV development. In contrast, the Aravalli region has higher relief and slope variability that can constrain suitability. The LULC and terrain layers are therefore used as explicit predictors rather than as unsupported assumptions about developability.

The LULC map (Figure 2) shows that large parts of Rajasthan are dominated by barren land, shrub-scrub, and sparsely vegetated surfaces, especially in western districts such as Jaisalmer, Barmer, and Jodhpur. These land-cover types were treated as relatively favorable in the suitability framework because previous PV siting studies commonly prefer barren, sparsely vegetated, or otherwise low-conflict land-cover classes over built-up, forested, or water-covered areas [7,10,19]. Agricultural and vegetated areas are more common in eastern and southeastern districts, while built-up areas are mainly concentrated around urban centers. The LULC data were derived from the Dynamic World dataset at 10 m

resolution, which provides near-real-time global land-cover information from Sentinel-2 imagery [24].

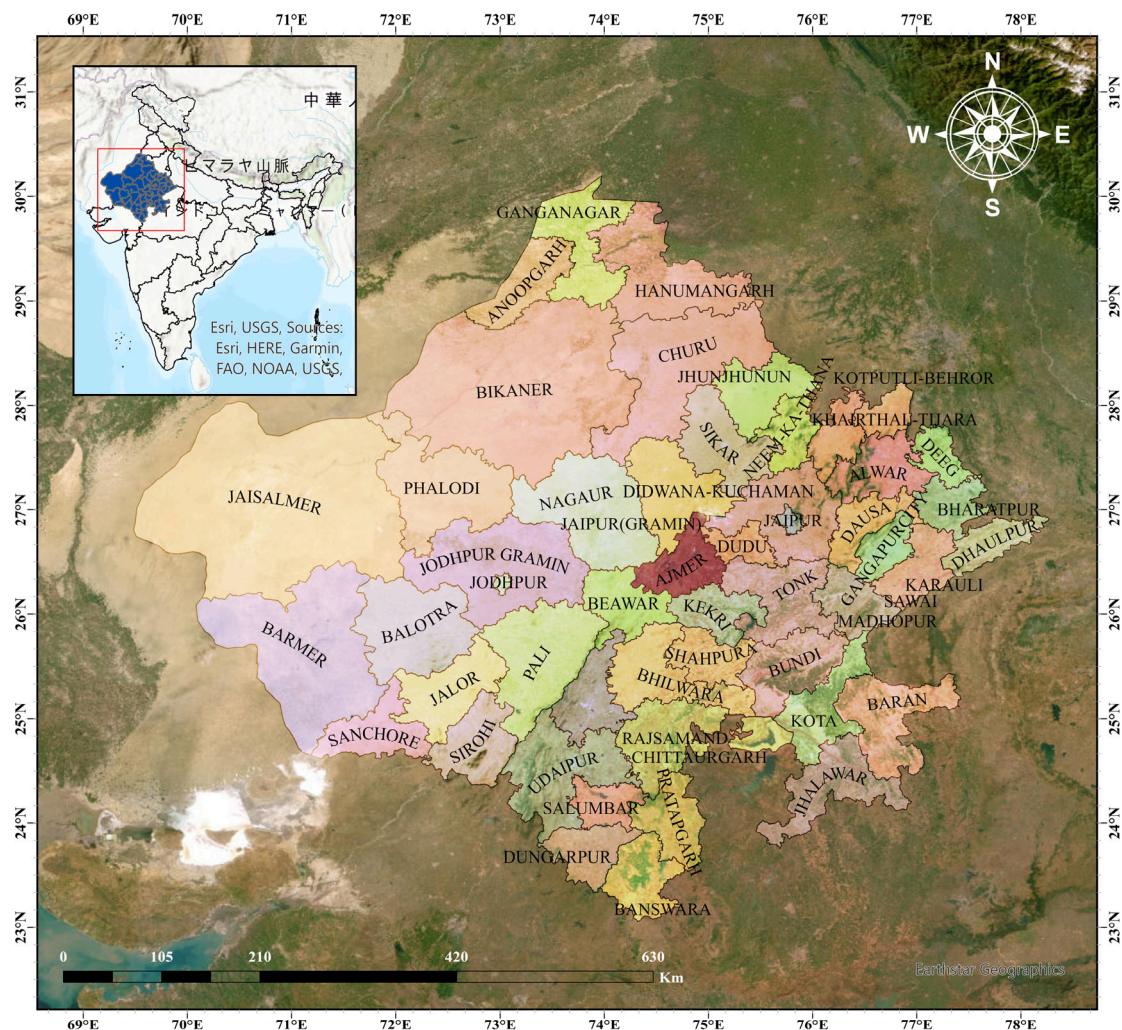


Figure 1. Geographic location and district-level administrative boundaries of Rajasthan, India. District-boundary GIS data are documented from the administrative-boundary source used for this study [25].

2.2. Integrated GeoAI Framework for Solar Suitability Assessment

This study applies a data-driven GeoAI framework to identify suitable locations for large-scale solar photovoltaic deployment. The workflow is organized into five phases: (1) data acquisition and spatial harmonization; (2) feature engineering and suitability-label construction; (3) XGBoost-based multi-class modeling with simpler-model benchmarks; (4) statistical evaluation, spatial-block validation, SHAP explainability, and external plausibility overlay; and (5) suitability mapping, candidate-site extraction, capacity estimation, and district-level prioritization (Figure 3).

2.3. Data Integration and Feature Space Construction

A multi-criteria dataset was prepared by combining solar resource variables, climate parameters, terrain attributes, land-surface characteristics, and infrastructure accessibility indicators. These variables were selected to represent both physical generation potential and practical feasibility for utility-scale deployment. GHI is a solar-resource input expressing incoming horizontal irradiance, whereas PVOUT is a performance-adjusted photovoltaic output indicator expressed per installed kWp; the two variables are correlated but not

identical and therefore capture complementary resource and system-output information. Temperature, wind speed, AOD, and albedo represent climatic and atmospheric controls on PV performance; DEM and slope represent terrain complexity and construction feasibility; LULC represents land availability and potential land-use conflict; and distance to roads and power lines approximates construction access and grid-connection feasibility.

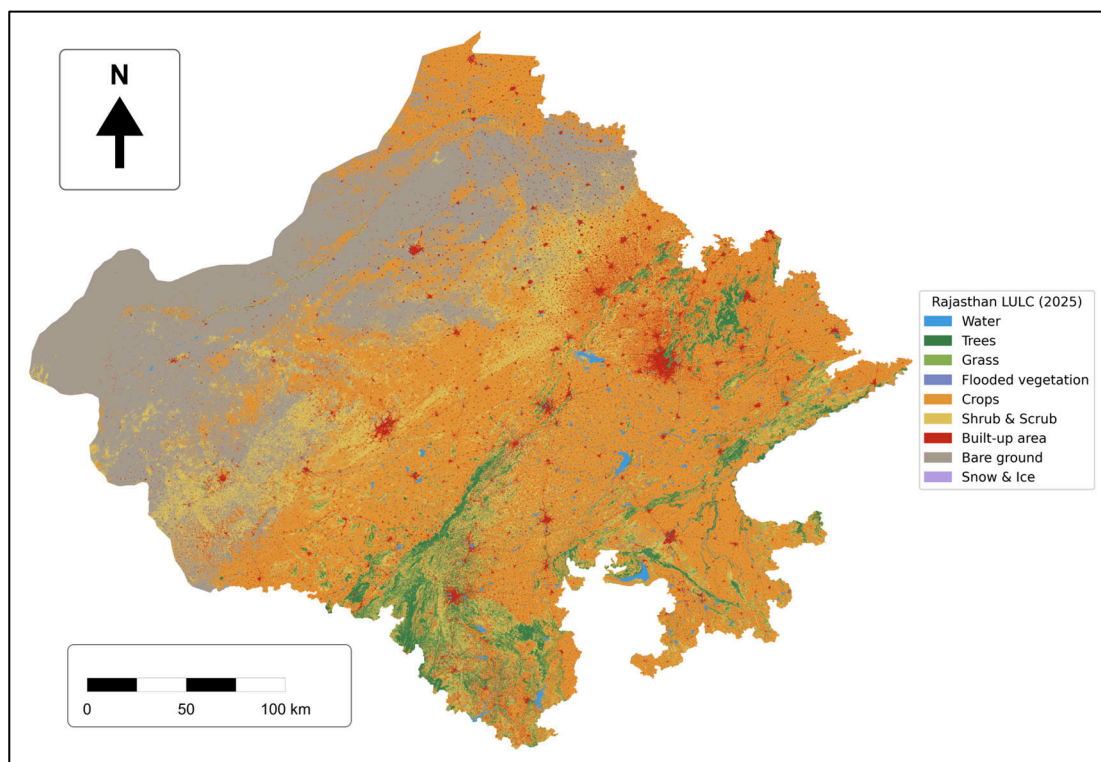


Figure 2. Land Use Land Cover (LULC) distribution of Rajasthan derived from the Dynamic World/Sentinel-2 dataset (2025) [24]. District boundaries are based on the administrative-boundary GIS source used in this study [25]. The map includes a north arrow and scale bar for cartographic reference.

All datasets were harmonized to a common analysis grid of approximately 250 m and aligned to a common spatial reference for overlay. Continuous variables, such as solar irradiance and temperature, were resampled using bilinear interpolation, while categorical variables, such as LULC, were resampled using nearest-neighbor interpolation to preserve class values. Vector-derived road and power-line distance layers were stored as kilometer-based distance rasters, and candidate-region areas were summarized in km². The predictor variables used in the model are summarized in Table 1.

Table 1. Integrated predictor variables used for GeoAI-based solar suitability modeling.

Variable	Role in Solar Suitability	Unit	Data Source	Native Resolution	Reference Period
GHI	Solar-resource benefit criterion; incoming irradiance	kWh/m ² /day	Global Solar Atlas/Solargis solar-resource raster [23]	Approx. 250 m after harmonization	Long-term climatological solar resource layer
PVOUT	PV-system output benefit criterion; performance-adjusted yield per kWp	kWh/kWp/day	Global Solar Atlas/Solargis PV output raster [23]	Approx. 250 m after harmonization	Long-term climatological PV resource layer

Table 1. Cont.

Variable	Role in Solar Suitability	Unit	Data Source	Native Resolution	Reference Period
Temperature	Thermal cost criterion; excessive heat reduces module efficiency	deg C	Copernicus Climate Data Store ERA5-Land climate layer [26]	Approx. 250 m after resampling	2016–2025 mean
Wind Speed	Contextual climate criterion; moderate wind can support cooling	m/s	Copernicus Climate Data Store ERA5-Land wind layer [26]	Approx. 250 m after resampling	2016–2025 mean
AOD	Atmospheric cost criterion; aerosols reduce irradiance	Unitless	NASA Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), aerosol optical depth layer [27]	Approx. 250 m after resampling	Study-period mean
DEM	Terrain context criterion; elevation indirectly affects accessibility	m	USGS/NASA Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) [28]	Approx. 30 m native; 250 m harmonized	Static terrain layer
Slope	Terrain cost criterion; lower slope is preferred	degrees	Derived from USGS/NASA Shuttle Radar Topography Mission (SRTM) DEM [28]	Approx. 30 m native; 250 m harmonized	Static terrain layer
Albedo	Surface-reflectivity context criterion	Unitless	Moderate Resolution Imaging Spectroradiometer (MODIS) bidirectional reflectance distribution function (BRDF)/Albedo product (MCD43A3) [29]	Approx. 250 m after resampling	Study-period mean
LULC	Land-availability constraint and exclusion criterion	Categorical	Dynamic World/Sentinel-2 LULC [24]	10 m native; 250 m harmonized	2024–2025 mode/2025 map
DistRoad	Accessibility cost criterion; shorter distance preferred	km	OpenStreetMap (OSM) road network [30]	Vector-derived distance raster	OSM extract dated 15 March 2026
DistPower	Grid-accessibility proxy; shorter distance preferred	km	OpenStreetMap (OSM)/compiled power-line vector layer [30]	Vector-derived distance raster	2026 compiled layer

Other potentially relevant factors, such as land ownership, land price, protected-area restrictions, water availability, local policy incentives, and detailed grid-hosting capacity, were not included because consistent, high-resolution, state-wide layers were not available for all districts. These exclusions are treated as limitations and priorities for future work rather than as assumptions that the omitted factors are unimportant.

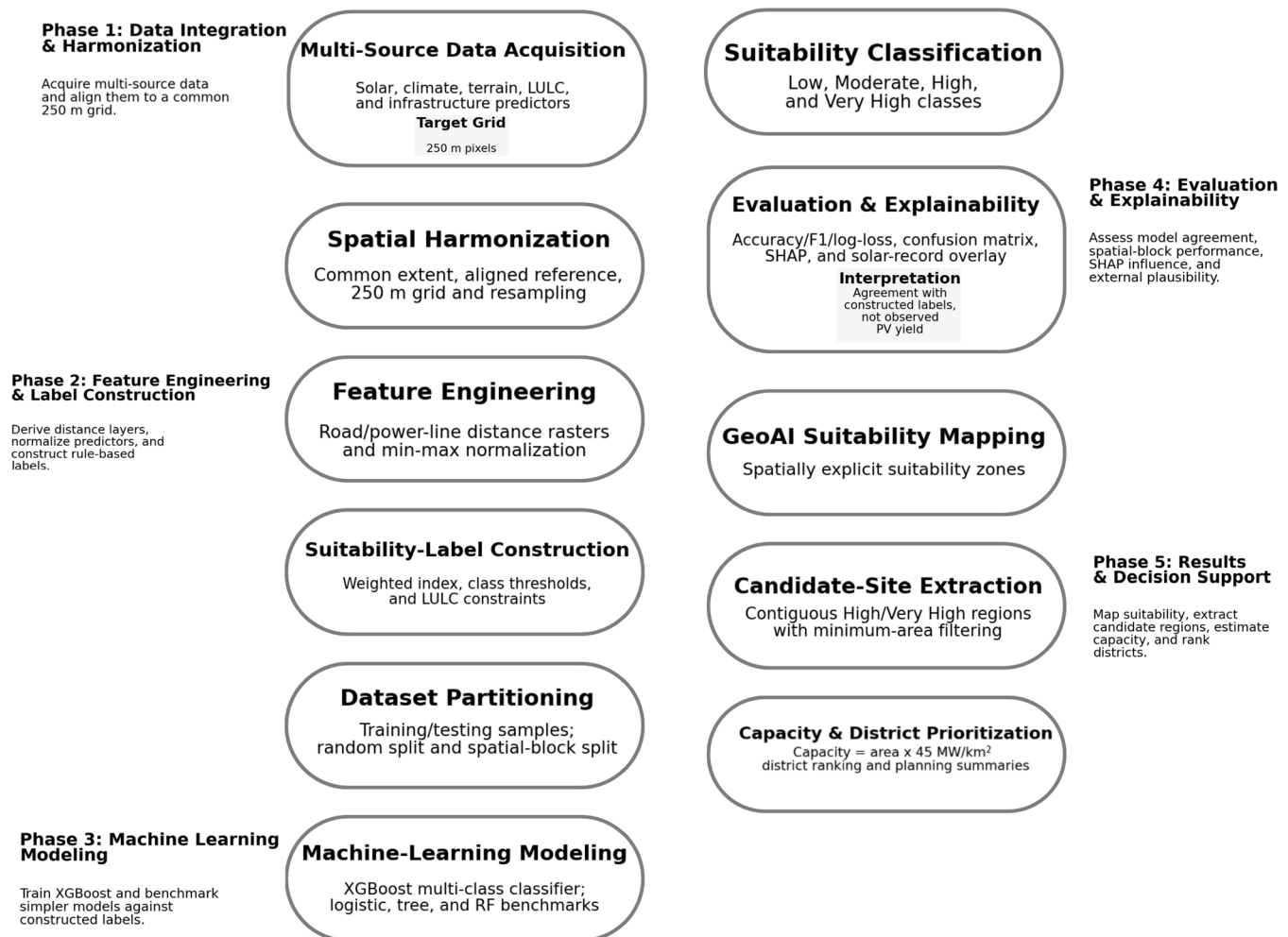


Figure 3. Five-phase methodological workflow of the integrated GeoAI framework for solar photovoltaic site suitability assessment, including data harmonization, rule-based label construction, machine-learning modeling, spatial validation, SHAP explainability, candidate-site extraction, and capacity-oriented prioritization.

2.4. Infrastructure Accessibility Modeling

Infrastructure accessibility was represented using distance-to-road and distance-to-power-line layers. These variables are important because solar farms require access for construction, operation, maintenance, and grid connection, and similar accessibility criteria are commonly included in GIS-based PV siting studies [10,19]. Euclidean distance was used to calculate the shortest distance between each pixel and the nearest infrastructure feature. The distance calculation is given in Equation (1):

Distance to power lines was used as a planning-level proxy for grid accessibility. This proxy does not fully represent high-voltage transmission availability, substation capacity, transformer requirements, right-of-way constraints, or grid-hosting capacity. Accordingly, the variable should be interpreted as an accessibility indicator rather than a complete grid-interconnection feasibility model.

$$D(x, y) = \min \left(\sqrt{(x - x_i)^2 + (y - y_i)^2} \right) \quad (1)$$

where $D(x, y)$ is the minimum distance from the pixel (x, y) to the nearest infrastructure feature (x_i, y_i) . Lower distance values indicate better accessibility and lower expected installation or connection cost.

2.5. Feature Normalization and Scaling

Because the predictors have different units and value ranges, normalization was applied before model training. Min–Max normalization was used to rescale continuous variables between 0 and 1. The normalization procedure is shown in Equation (2):

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

where X' is the normalized value, X is the original value, and $X_{\min}$ and $X_{\max}$ are the minimum and maximum values of the variable. This step helps reduce scale-related bias and improves numerical stability during model training.

2.6. Suitability-Label Generation and XGBoost Model Training

Solar suitability was treated as a four-class classification problem: Low, Moderate, High, and Very High. The reference labels were not derived from existing solar farms; they were generated before model training from a transparent rule-based suitability index. Predictor values were first converted to 0–1 suitability scores according to their expected role as benefit or cost criteria. The composite score was calculated as $S = 0.25(\text{GHI}) + 0.20(\text{PVOUT}) + 0.15(\text{Slope suitability}) + 0.10(\text{AOD suitability}) + 0.10(\text{Temperature suitability}) + 0.10(\text{Road-distance suitability}) + 0.05(\text{Power-line-distance suitability}) + 0.05(\text{LULC suitability})$. The final classes were assigned using fixed thresholds: Low ≤ 0.45 , Moderate > 0.45 to ≤ 0.60 , High > 0.60 to ≤ 0.75 , and Very High > 0.75 . Water and built-up LULC categories were constrained to Low suitability. This procedure makes the target variable a constructed planning label; therefore, model performance measures how well XGBoost reproduces the defined suitability logic, not how accurately it predicts independently observed PV project performance.

$$L = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (3)$$

$l(y_i, \hat{y}_i) y_i \hat{y}_i \Omega(f_k)^{k^{th}}$ where the loss term measures the difference between the constructed reference suitability label and predicted class, while the regularization term controls model complexity. The labeled dataset was divided into training and testing subsets using an 80:20 stratified split to preserve class representation. Hyperparameter tuning was performed using validation log-loss and weighted F1-score, with conservative tree depth and subsampling to reduce overfitting. The specific settings and optimization procedures for the XGBoost model are detailed in Table 2.

Table 2. XGBoost model settings and optimization procedure.

Setting	Value/Strategy
Model	XGBoost multi-class classifier
Objective	multi:softprob
Number of classes	4
Train-test split	80:20 stratified split
Evaluation sample	300,000 sampled pixels for validation comparison; 240,000 train and 60,000 test in the random split
Tuning strategy	Fixed parameter set selected and monitored using validation log-loss and weighted F1-score; no external solar-installation records were used for tuning
Regularization approach	Tree-depth control, learning-rate control, subsampling, column subsampling, and default L2 regularization
Primary validation outputs	Accuracy, weighted precision, weighted recall, weighted F1-score, confusion matrix, and log-loss curves

Table 2. *Cont.*

Setting	Value/Strategy
Random validation design	300,000 sampled pixels; 240,000 training and 60,000 testing pixels
Spatial validation design	0.5-degree spatial-block split; complete geographic blocks held out for testing
Benchmark models	Multinomial logistic regression, decision tree, random forest, and XGBoost
n_estimators	200
max_depth	8
learning_rate	0.05
subsample	0.80
colsample_bytree	0.80
eval_metric	mlogloss
random_state	42

2.7. Model Evaluation and Statistical Validation

The model was evaluated using accuracy, precision, recall, F1-score, balanced accuracy, confusion matrices, and log-loss curves. These metrics are commonly used to assess classification performance and are derived from the confusion matrix. Because the dataset consists of spatially adjacent pixels, two validation designs were used. First, a random stratified pixel split was applied to match the initial model assessment, using 300,000 sampled pixels with 240,000 pixels for training and 60,000 pixels for testing. Second, a spatial-block validation was conducted by grouping pixels into 0.5-degree geographic blocks and holding out complete blocks for testing. This second design reduces the likelihood that neighboring pixels with similar predictor values appear simultaneously in the training and testing subsets.

To test whether the complexity of XGBoost was justified, the same sampled dataset and validation designs were also evaluated using three simpler baselines: multinomial logistic regression, a decision tree, and a random forest. In addition, the original rule-based weighted suitability score and exclusion rules were reproduced as a transparency check. This rule-based baseline is not an independent predictive model; it documents the deterministic procedure used to construct the final suitability labels and therefore provides an upper-bound reference for label reproducibility.

A presence-based external plausibility assessment was conducted using independent records of existing solar installations that were not used to generate the suitability labels or train the models. Solar plant and solar farm locations from the World Resources Institute (WRI) Global Power Plant Database (GPPD), the Global Energy Monitor (GEM) Global Solar Power Tracker, and OpenStreetMap (OSM) were filtered to the Rajasthan boundary and overlaid with the final suitability raster and the extracted candidate site mask. Because these datasets provide presence locations rather than a complete sample of both suitable and unsuitable land, the overlay was interpreted as supporting evidence for spatial plausibility rather than as a conventional accuracy assessment.

As an additional robustness test, an India-wide presence-background model was constructed following the logic of presence-background/maximum-entropy-style distribution modeling [31,32]. Known solar-site records from OSM solar farms/sites, WRI/GPPD solar plants, and operating GEM Global Solar Power Tracker projects were snapped to the nearest harmonized grid cell and deduplicated, producing 3542 presence grid cells from 7437 source records. Background cells were sampled from non-excluded grid cells that did not contain a known solar record and were at least two grid cells away from any known solar presence, corresponding to approximately 10–12 km at the grid spacing

used here. Using a fixed random seed of 42, a 5:1 background-to-presence ratio was used, resulting in 17,710 background cells. A regularized logistic classifier was then fitted using the available solar-resource, terrain, land-cover, infrastructure-distance, atmospheric, and surface-reflectance predictors, and evaluated using five spatial folds based on geographic blocks [33]. This analysis was kept separate from the main XGBoost suitability-label model; it was used only as supplementary presence-background evidence because background cells are not confirmed unsuitable sites.

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

Precision measures of how many predicted positive samples were actually correct, as shown in Equation (5):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

Recall, also known as sensitivity, measures how many actual positive samples were correctly identified, as shown in Equation (6):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

The F1-score provides a balanced measure of precision and recall and is calculated using Equation (7):

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

where TP , TN , FP , and FN represent true positives, true negatives, false positives, and false negatives, respectively. A confusion matrix was also used to examine class-wise performance, where diagonal values represent correct classifications and off-diagonal values represent misclassifications.

2.8. Model Explainability and Feature Contribution Analysis

To improve model interpretability, SHapley Additive exPlanations (SHAP) were used to estimate the contribution of each predictor to the model output. SHAP is based on game theory and assigns each feature an importance value according to its contribution to prediction. The SHAP value for a feature is calculated using Equation (8):

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (8)$$

2.9. Suitability Classification and Statistical Characterization

After training and validation, the model was applied to the full study area to classify each valid pixel into the four suitability classes. The rule-based label thresholds used to define the reference classes are summarized in Table 3, while the resulting full-map class distribution is reported in the Results section.

2.10. Candidate-Site Extraction and Capacity Estimation

Pixels classified as High and Very High suitability were extracted from the final suitability raster and converted into contiguous candidate regions using connected-neighbor aggregation. In this manuscript, a candidate site means a raster-derived contiguous suitability polygon, not a cadastral land parcel or a plot belonging to a single owner. Regions smaller than 1 km² were removed to focus on utility-scale screening opportunities and reduce isolated single-pixel artifacts, consistent with the need to consider land-use intensity in utility-scale PV planning [34]. Low and Moderate pixels were excluded from

candidate-site delineation. Built-up/residential, settlement, and water pixels were constrained through the LULC rule summarized in Table 4; therefore, the workflow does not treat settlement land as suitable. Road and power-line accessibility were retained as continuous distance predictors rather than imposed as hard buffer exclusions. No additional settlement-distance buffer, protected-area buffer, land-ownership buffer, parcel-boundary filter, substation-capacity filter, or high-voltage-transmission exclusion buffer was applied because harmonized state-wide datasets were not available for all districts. Candidate-region centroids were used only for labels and summary tables; the candidate regions themselves are represented and analyzed as contiguous polygons.

Table 3. Rule-based suitability-label thresholds and exclusion logic.

Class	Suitability-Score Rule	Exclusion/Constraint Rule	Planning Interpretation
Low	$S \leq 0.45$	Pixels classified as Water or Built-up/Residential (Settlement) were restricted to the Low class	Unsuitable or highly constrained
Moderate	$0.45 < S \leq 0.60$	No additional hard buffer	Potentially suitable only after local screening; may be limited by resource, terrain, access, or land-cover conditions
High	$0.60 < S \leq 0.75$	No additional hard buffer	Suitable screening priority; requires project-level checks
Very High	$S > 0.75$	No additional hard buffer	Highest screening priority, subject to parcel, grid, permitting, socio-environmental conditions, and field verification

Table 4. Sensitivity analysis scenario design used to test source-data uncertainty, exclusion-mask parameters, capacity assumptions, and background-sampling choices.

Objective	Parameter or Source of Uncertainty Tested	Magnitude or Scenario Design
Input-data uncertainty propagation	Normalized predictor scores for GHI, PVOUT, temperature, wind, AOD, DEM/slope, albedo, LULC, and infrastructure distances	500 Monte Carlo runs on a 300,000-pixel sample using correlated source-level bias and pixel-level residual noise
Exclusion-mask sensitivity	Slope screening and LULC/built-up constraint	Additional hard slope caps (>5 , >10 , and >15 degrees); no hard LULC exclusion; baseline water/built-up constraint; stricter LULC-score thresholds (>0.20 and >0.50)
Capacity-assumption sensitivity	Area-density conversion from candidate-polygon area to indicative installed capacity	35, 45, and 50 MW/km ² , equivalent to approximately 2.86, 2.22, and 2.00 ha/MW
Background-sampling sensitivity	Definition of non-plant background samples in the supplementary presence-background model	Uniform buffered, environmental-stratified, and preference-weighted plausible-siting background samples under five spatial folds
Candidate-ranking sensitivity	Relative emphasis on area, solar-resource quality, terrain, and accessibility in candidate ranking	$\pm 20\%$ weight perturbations and area-dominant, resource-dominant, and access-dominant ranking variants

For map presentation and district summaries, each candidate polygon was assigned a post-processing development-priority class from a normalized candidate-rank score. This score was calculated after XGBoost mapping and candidate-polygon extraction from

the mean normalized suitability of five candidate-level indicators: GHI, PVOUT, slope suitability, road-distance accessibility, and power-line-distance accessibility. It did not incorporate the XGBoost class probability or the preceding four-class pixel label as an additional factor; those outputs were used only to define the candidate mask. Polygon area was used separately for indicative capacity estimation and capacity-oriented top-candidate ranking. Candidate scores were grouped into five approximately equal-frequency priority classes: Very High (0.45–0.75), High (0.38–0.44), Moderate (0.33–0.38), Low (0.29–0.33), and Very Low (0.15–0.29), based on rounded score ranges in the final candidate layer. These five classes are therefore planning-priority categories for candidate polygons, not additional suitability classes.

Planning-level installable capacity for each candidate region was estimated with an area-density approach, which avoids dimensional inconsistency associated with multiplying daily irradiance directly by area:

$$C_{\text{installed}} = A_{\text{site}} \times D_{\text{cap}} \quad (9)$$

where $C_{\text{installed}}$ is the estimated installed capacity (MW), A_{site} is the contiguous candidate-site area (km^2), and D_{cap} is the assumed deployment density (MW per km^2). In this study, capacity estimates were derived using a planning-level density of 45 MW per km^2 [34]. This value is used only as a transparent area-density assumption; utility-scale PV land-use intensity varies with plant design, technology, spacing, and whether direct or total land area is measured. The resulting values should be interpreted as indicative technical potential rather than final project capacities because detailed parcel ownership, grid-hosting capacity, environmental permitting, and economic feasibility were outside the present scope.

2.11. Uncertainty Propagation and Parameter-Sensitivity Assessment

To quantify uncertainty in the suitability workflow, a scenario-based uncertainty assessment was implemented following recent recommendations that geospatial assessments should explicitly examine source-data uncertainty, error propagation, sensitivity to calculation choices, and the interpretation of uncertainty in planning decisions [14]. Because formal per-pixel error covariance information was not available for all input products, the analysis is reported as a planning-level uncertainty and sensitivity assessment rather than as a fully field-calibrated probabilistic error model.

Five uncertainty components were tested. First, normalized predictor scores were perturbed using a Monte Carlo design with correlated source-level bias and pixel-level residual noise, reflecting the fact that errors in products such as GHI, LULC, DEM-derived slope, and infrastructure-distance layers may be spatially heterogeneous and partly systematic. The simulation used 500 runs on a reproducible 300,000-pixel sample with fixed random seed 42, and both High/Very High area intervals and pixel-level class-stability probabilities were calculated. Second, key mask parameters were tested by imposing additional hard slope caps and alternative LULC/built-up exclusion variants. Because the harmonized workflow used categorical Dynamic World LULC rather than a continuous built-up-intensity raster, built-up sensitivity was evaluated through the baseline water/built-up exclusion and stricter LULC suitability-score thresholds. Third, the capacity-density assumption was varied between 35 and 50 MW/ km^2 , equivalent to approximately 2.86–2.00 ha/MW and explicitly including the 2.00 ha/MW scenario. Fourth, the presence-background validation was repeated under uniform buffered, environmental-stratified, and preference-weighted plausible-siting background samples. Fifth, the uncertainty outputs were interpreted as planning-level intervals rather than construction-ready confidence intervals.

3. Results

3.1. Model Performance and Validation

Table 5 reports the actual class counts used in the validation workflow. The full Rajasthan raster is intentionally imbalanced because High and Very High classes occupy large areas. The random validation split was stratified to preserve this distribution; it was not an artificially balanced four-class dataset.

Table 5. Class distribution of the full raster, validation sample, and random train-test split.

Class	Full-Map Pixels	300,000-Pixel Sample	Train Pixels	Test Pixels
Low	229,012	13,938	11,151	2787
Moderate	260,762	15,915	12,732	3183
High	2,456,946	149,423	119,538	29,885
Very High	1,985,972	120,724	96,579	24,145

The XGBoost model showed high apparent agreement with the constructed reference labels under the random stratified pixel split, achieving an overall accuracy of 99.14% and a weighted F1-score of 99.15% using a 60,000-pixel test sample. However, because neighboring pixels can share similar predictor values, this result was not treated as sufficient evidence of spatial generalization. A spatial-block validation was conducted to test sensitivity to spatial autocorrelation. Under the 0.5-degree block holdout design, XGBoost accuracy decreased to 97.73% and the weighted F1-score decreased to 97.73%. This reduction indicates that random pixel splitting likely benefits from local spatial similarity, while the block test indicates that the model still reproduces much of the constructed suitability-label structure in geographically separated areas.

Because the suitability labels were derived from a weighted composite index and exclusion rules based on the same predictor family used for model training, the high accuracy should be interpreted cautiously. The rule-based score and exclusion logic reproduced the final labels exactly, indicating that the classification target is partly deterministic by design. Therefore, the reported machine-learning accuracy indicates how well each model approximates the constructed suitability framework, not independent proof of observed PV project performance.

The rule-based weighted suitability score and exclusion rules reproduced the generated final labels with 100% agreement. This result is reported to clarify that the model metrics describe the reproduction of a constructed suitability framework rather than independent field validation (Table 6).

Table 6. Model comparison under random pixel and spatial-block validation designs.

Validation Design	Model	Train Pixels	Test Pixels	Accuracy	Weighted F1
Random stratified pixel split	Logistic regression	240,000	60,000	0.9177	0.9065
Random stratified pixel split	Decision tree	240,000	60,000	0.9748	0.9748
Random stratified pixel split	Random forest	240,000	60,000	0.9875	0.9875
Random stratified pixel split	XGBoost	240,000	60,000	0.9915	0.9915
Spatial block split	Logistic regression	239,621	60,379	0.9178	0.9106
Spatial block split	Decision tree	239,621	60,379	0.9334	0.9334
Spatial block split	Random forest	239,621	60,379	0.9646	0.9645
Spatial block split	XGBoost	239,621	60,379	0.9773	0.9773

To examine the spatial plausibility of the suitability outputs against independent ground evidence, known solar installations were overlaid with the final suitability map. The final raster classifies 90.07% of the Rajasthan study area as High or Very High suitability; therefore, the overlay was interpreted relative to this high area-wide baseline rather than as a standalone accuracy value. The official WRI/GPPD records showed the strongest agreement, with 95.89% of Rajasthan solar plants and 94.16% of their recorded capacity falling in High or Very High classes. OSM solar-farm/site points showed a similarly high record-level agreement, while GEM operating records were more conservative because the tracker includes a broader range of project phases, scales, and locational records (Table 7).

Table 7. Presence-based external plausibility assessment of the suitability and candidate-site outputs using known solar installations in Rajasthan.

Dataset	Records in Rajasthan	High/Very High Records (%)	High/Very High Capacity (%)	Inside Candidate Mask (%)	Candidate Capacity (%)
WRI/GPPD India solar [35]	146	95.89	94.16	38.36	56.86
GEM Solar Tracker, operating [36]	448	80.58	72.69	17.41	15.78
OSM solar farms/sites [30]	6955	98.07	97.36	2.65	62.87
All records, deduplicated from WRI/GPPD, GEM, and OSM sources [30,35,36]	5875	94.42	76.90	6.71	15.08

The High/Very High class covers 90.07% of the final suitability raster inside Rajasthan, whereas the contiguous candidate-site mask covers only 5.77% of the study area. Candidate-mask overlap therefore represents a stricter spatial test than suitability-class overlap. Because the OSM dataset includes many small or capacity-unknown solar features, capacity-weighted agreement is more informative than raw record count for this source.

The candidate-site overlay provides a stricter check because the extracted candidate mask occupies only 5.77% of Rajasthan. Under this more selective mask, 38.36% of WRI/GPPD solar plants and 56.86% of their recorded capacity intersected candidate regions, while 17.41% of GEM operating projects and 15.78% of their capacity intersected the same mask. These results support the spatial plausibility of the candidate-site extraction for major utility-scale locations, but they also show that the candidate layer is a planning screen rather than a complete inventory of all existing solar developments.

For clearer comparison, Figure 4 visualizes the two overlay levels. Panel A shows agreement with broad High/Very High suitability classes, while Panel B uses the much more selective contiguous candidate-site mask. The baseline lines indicate the corresponding map-area shares, allowing the observed solar-installation overlap to be interpreted against the spatial prevalence of each mapped class.

The supplementary India-wide presence-background spatial cross-validation provided an additional robustness check that is independent of the rule-derived suitability labels (Table 8). Across five geographic folds, the regularized logistic presence-background model achieved a mean receiver operating characteristic area under the curve (ROC-AUC) of 0.724 ± 0.039 and a mean precision-recall area under the curve (PR-AUC) of 0.398 ± 0.108 . The top decile of held-out background scores captured $38.5 \pm 10.4\%$ of held-out presence grid cells, indicating enrichment of observed solar sites in high-scoring predictor space relative to buffered background locations. When the same held-out scoring logic was applied to Rajasthan solar records, the top-decile presence-background-score records intersected the extracted candidate-site mask more frequently than the full solar-record set (26.3%

versus 6.15% by record count). These results support the spatial plausibility of the mapped candidate regions, but they are interpreted only as supplementary evidence because the background sample does not represent verified absence or officially rejected project sites.

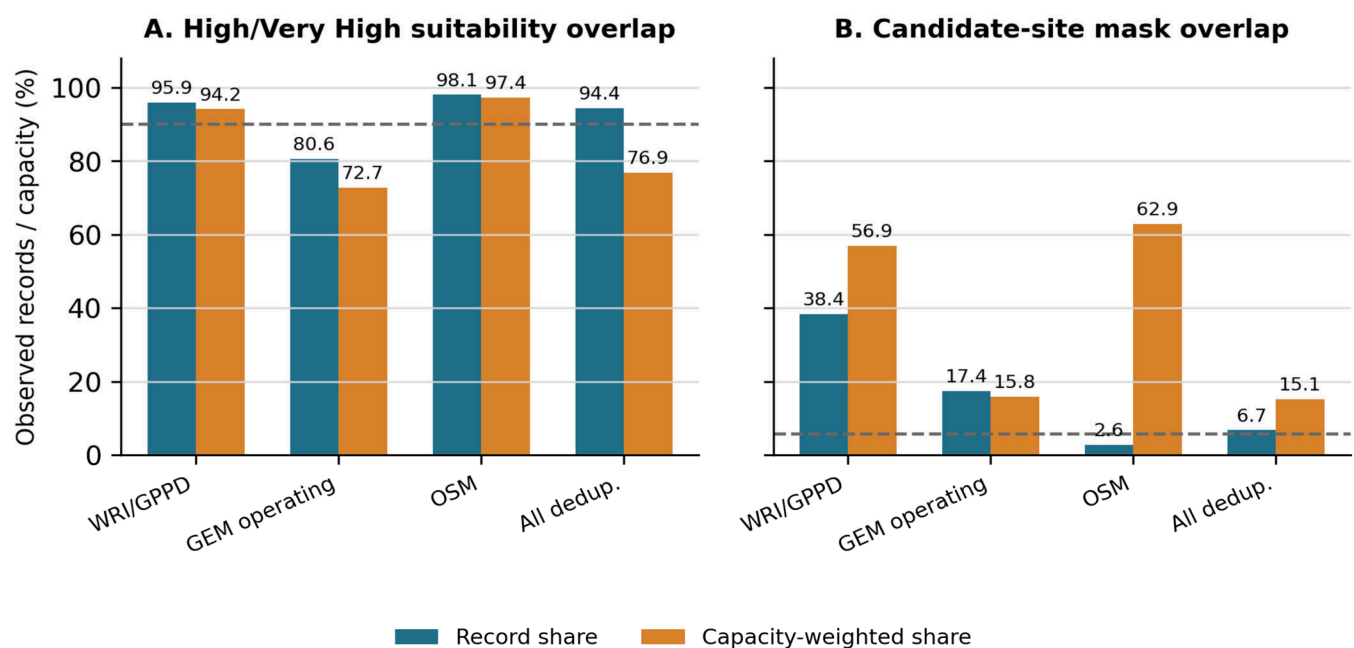


Figure 4. Presence-based external plausibility assessment using known solar installations in Rajasthan. Panel (A) compares the share of observed records and capacity falling in High or Very High suitability classes against the area-wide High/Very High baseline. Panel (B) compares the same ground records against the extracted contiguous candidate-site mask, which covers a much smaller share of the state and therefore provides a stricter spatial test.

Table 8. Supplementary India-wide presence-background spatial cross-validation using known solar-site records and buffered background cells.

Component	Result/Interpretation
Presence data	7437 India-wide source records from OSM solar farms/sites [30], WRI/GPPD solar plants [35], and operating GEM solar projects [36]; deduplicated to 3542 grid cells
Background data	17,710 non-excluded grid cells sampled with fixed random seed 42 at a 5:1 background-to-presence ratio and buffered approximately 10–12 km from known solar records
Validation design	Five spatial folds based on geographic blocks; solar-site records were not used to generate the main suitability labels
Mean ROC-AUC	0.724 ± 0.039
Mean PR-AUC	0.398 ± 0.108
Presence enrichment	38.5 ± 10.4% of held-out presence cells exceeded the held-out background top-decile threshold
Rajasthan candidate-mask comparison	Top-decile presence-background-score records intersected candidate polygons at 26.3% by record count, compared with 6.15% for all scored Rajasthan solar records
Interpretation	Supplementary presence-background robustness evidence only; not true absence-based validation

The confusion matrix (Figure 5) shows that most pixels were correctly classified, as indicated by the dominant diagonal values. Misclassification was limited and mainly occurred between neighboring suitability classes, especially *Moderate–High* and *High–Very*

High. This is reasonable because suitability transitions in geographic space are gradual rather than sharply separated.

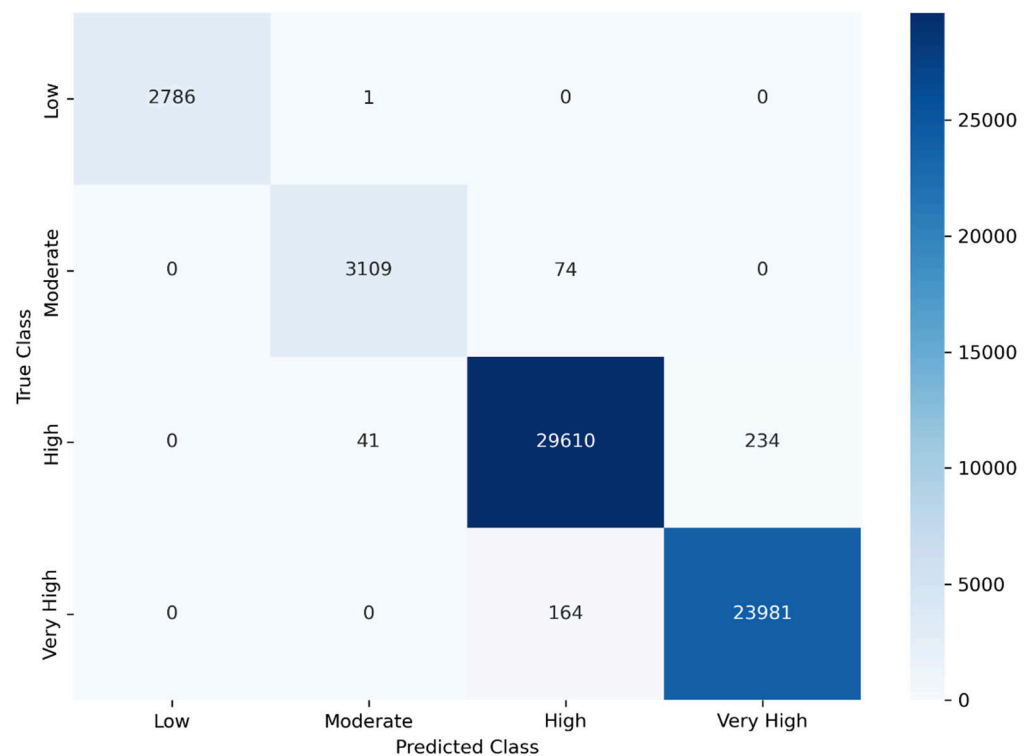


Figure 5. Confusion matrix and classification performance curves showing agreement between XGBoost predictions and constructed reference labels.

The training and validation loss curves (Figure 6) show stable convergence, with both curves decreasing consistently across boosting iterations. Their close alignment suggests that the model maintained a balanced bias-variance trade-off and did not show clear signs of overfitting within the tested validation designs. Overall, the results indicate that XGBoost reproduced the constructed suitability-label structure consistently. This supports its use as a transparent surrogate for the defined suitability framework, while not implying independent prediction of observed PV performance.

3.2. Spatial Characteristics of Topographic Constraints

The topographic variables show clear spatial differences across Rajasthan. As shown in Figure 7, western Rajasthan is mostly characterized by low slope values, particularly below 5° , which makes it favorable for large-scale solar installations. Flat terrain reduces construction difficulty, land preparation costs, and potential shading problems. In contrast, higher slope values are mainly concentrated along the Aravalli range, where terrain complexity may limit solar deployment.

Elevation supports this pattern. Although elevation alone does not directly prevent solar installation, it is often linked with slope, accessibility, and local terrain complexity. Therefore, elevation acts as an indirect constraint rather than a primary limiting factor. These findings are consistent with solar site selection studies that consider slope, land cover, and accessibility as key siting constraints [8].

3.3. Solar Resource Potential and Climatic Variability

The solar resource variables show a clear west–east gradient in solar energy potential. Western and southwestern Rajasthan show higher values of Global Horizontal Irradiance

(GHI) and Photovoltaic Output (PVOOUT), indicating favorable resource conditions for solar energy generation.

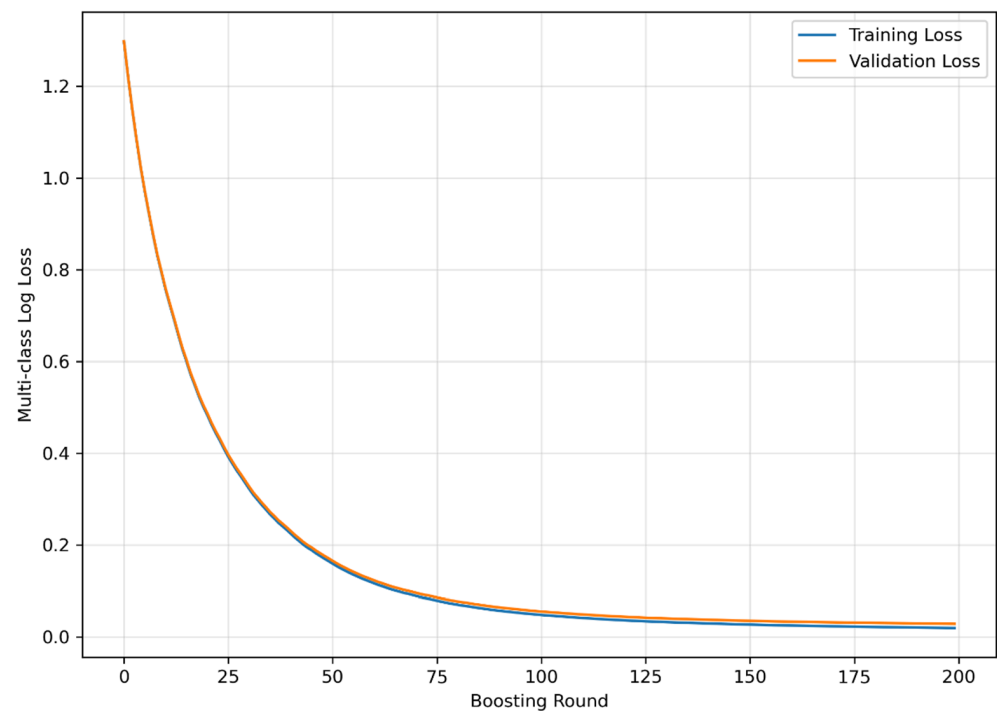


Figure 6. Training and validation loss curves showing convergence behavior of the XGBoost model.

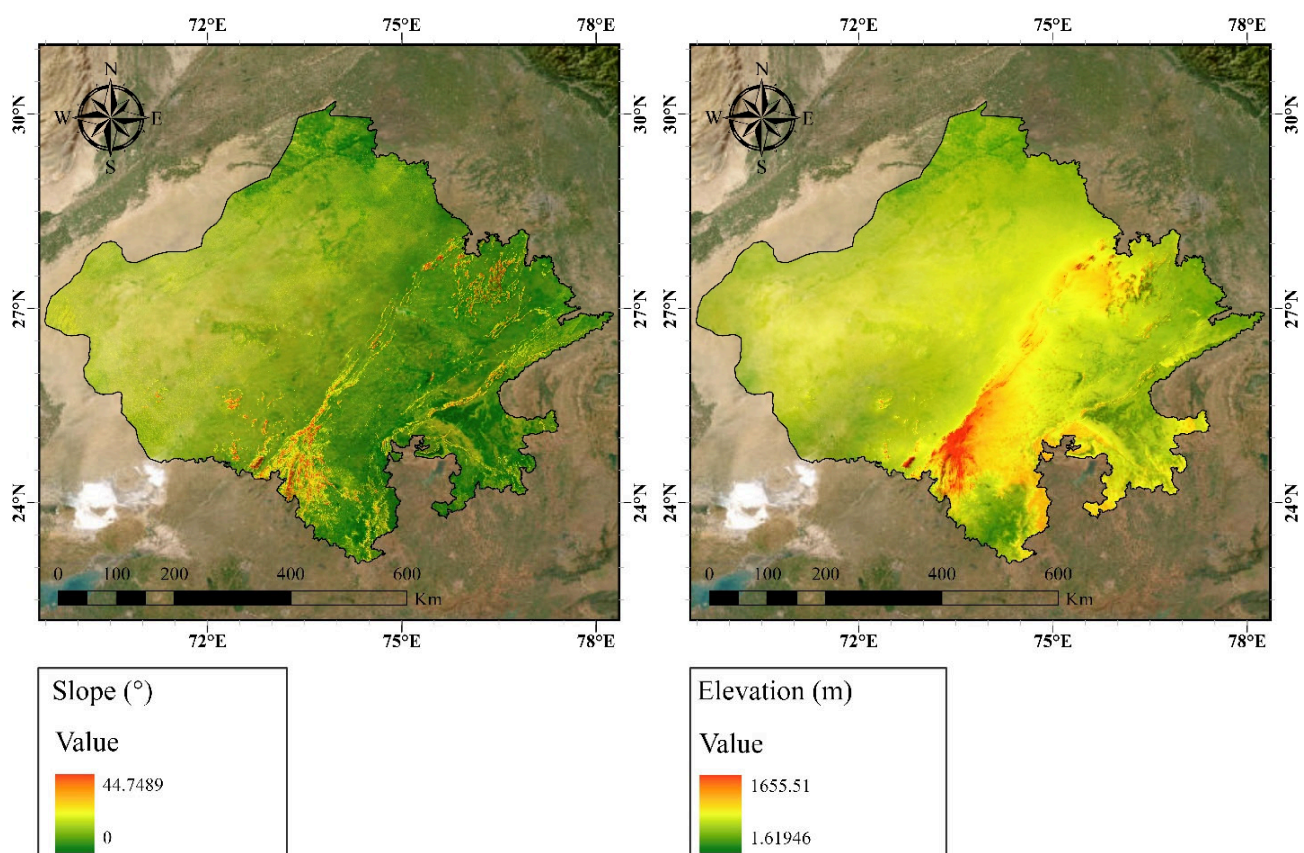


Figure 7. Distribution of terrain slope and elevation highlighting topographic constraints for photovoltaic deployment. Elevation was derived from SRTM DEM data, and slope was calculated from the same DEM source [28].

Temperature varies moderately across the state, with higher values observed in arid western districts. Elevated temperature can reduce PV module efficiency through temperature-coefficient losses; however, western Rajasthan still shows high PVOUT because its Very High GHI provides a large gross solar-energy input. Thus, high irradiance increases total modeled energy yield, not by physically offsetting temperature-related efficiency losses (Figure 8). These findings are consistent with studies that identify solar irradiance as a primary factor controlling photovoltaic performance and site suitability [18,23].

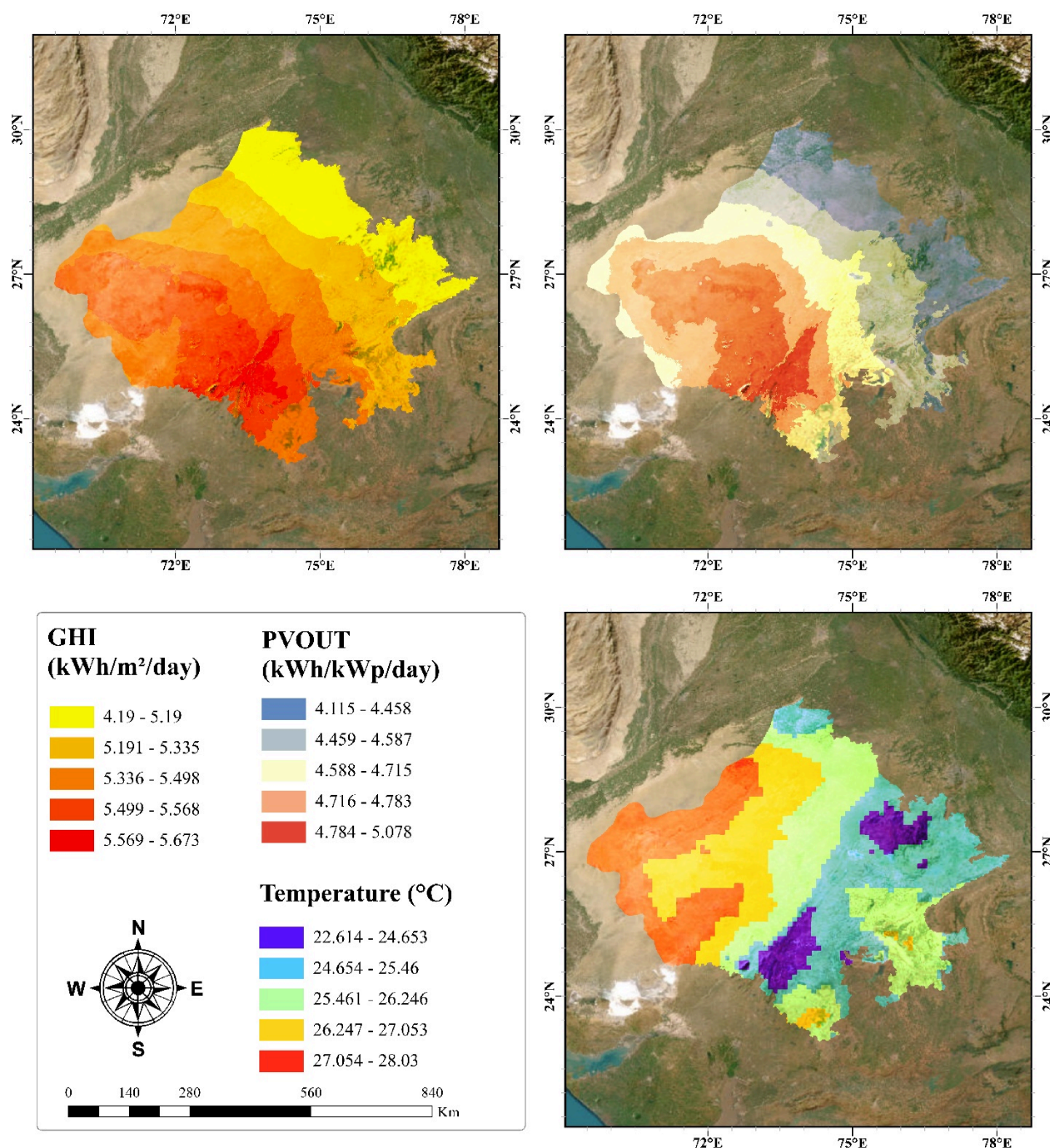


Figure 8. Distribution of solar resources and climate variables. GHI and PVOUT are based on Global Solar Atlas/Solargis data [23], while temperature is based on the reanalysis climate layer used in this study [26].

3.4. Infrastructure Accessibility Assessment

Infrastructure accessibility is an important planning factor for photovoltaic deployment. The distance-to-road and distance-to-power-line layers (Figure 9) indicate that central and eastern Rajasthan generally have stronger network proximity, whereas some western areas are farther from major infrastructure corridors.

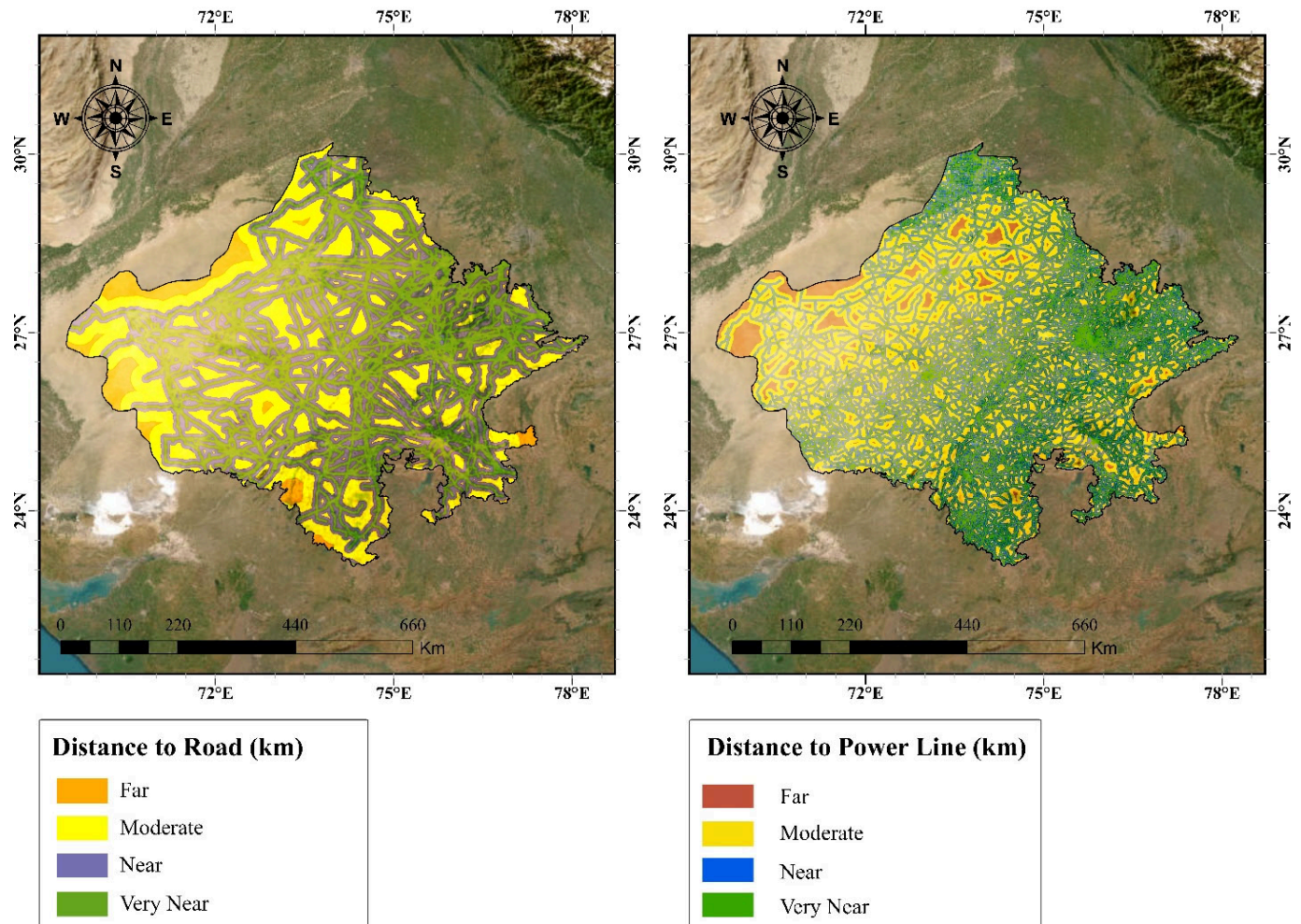


Figure 9. Spatial distribution of accessibility indicators, including distance to road networks and power transmission lines. Road and power-line vector inputs were derived from Open-StreetMap/compiled infrastructure layers used in this study [30].

Western Rajasthan may remain attractive at the regional screening scale where large contiguous land parcels and strong solar-resource conditions compensate for longer distances to some infrastructure. Nevertheless, proximity to roads and transmission lines can reduce construction, maintenance, and grid-connection costs, making accessibility an important site-selection criterion [10].

3.5. Feature Importance and Model Interpretability

The SHAP-based feature importance analysis provides a clear understanding of how each predictor influenced the model output. As shown in Figure 10, PVOUT and GHI were the most influential variables, followed by LULC, distance to roads, and distance to power lines.

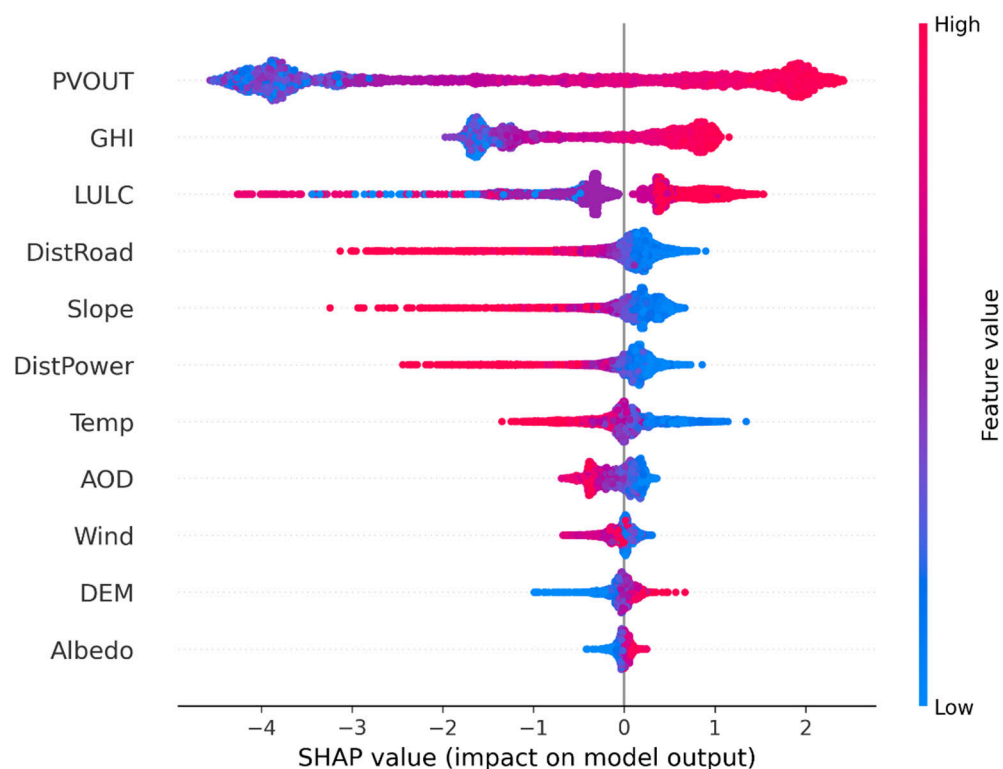


Figure 10. SHAP-based feature importance ranking showing the dominant predictors influencing solar suitability.

Slope also contributed to the model, confirming its role as a physical constraint for solar installation. In contrast, wind speed and elevation had relatively lower contributions, suggesting that they played a secondary role in this analysis. This feature hierarchy is scientifically reasonable because solar irradiance, land availability, and infrastructure access are among the most important factors in PV site selection [8].

3.6. Distribution of Solar Suitability Classes

The suitability classification shows that a large proportion of Rajasthan falls within favorable solar development zones under the adopted scoring thresholds. As shown in Table 9, High and Very High suitability classes together account for nearly 90% of the study area. This result reflects the input GHI/PVOUT layers, dryland land-cover classes, low terrain constraints, and infrastructure-access criteria included in the constructed suitability framework; it should not be interpreted as an independent estimate of immediately developable land.

Table 9. Distribution of solar suitability classes across the study area.

Class	Pixel Count	Percentage (%)
Low	229,012	~4.6
Moderate	260,762	~5.3
High	2,456,946	~49.8
Very High	1,985,972	~40.3

3.7. Candidate-Site Distribution and Development Priority

The extracted candidate solar sites (Figure 11) show clustered, pixel-based polygon patterns rather than generalized point symbols. These candidate polygons are concentrated mainly in western and west-central Rajasthan, where high solar resource availability, rela-

tively flat terrain, and suitable land-cover conditions combine to create large contiguous development zones. Across all 192 final candidate polygons used for the development-priority classes, mean GHI ranges from 5.507 to 5.669 kWh/m²/day, mean PVOUT ranges from 4.735 to 4.960 kWh/kWp/day, and mean slope ranges from 1.53 to 3.33 degrees. Within the Very High priority subset, the corresponding ranges are 5.571–5.669 kWh/m²/day for GHI, 4.795–4.960 kWh/kWp/day for PVOUT, and 1.81–3.32 degrees for slope; this low-slope range is important because increasing slope can raise grading, mounting, road-access, and installation costs.

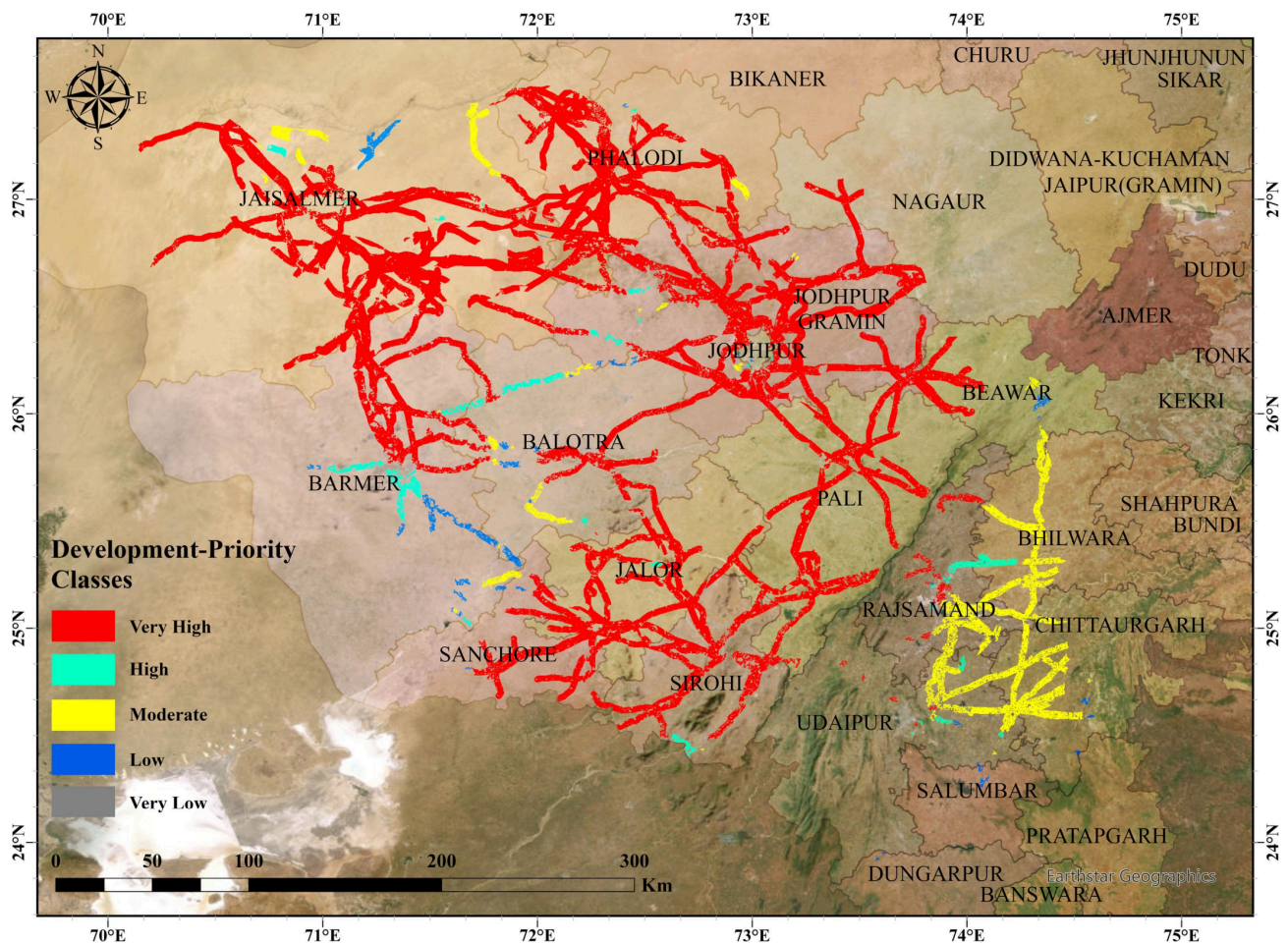


Figure 11. Raster-derived contiguous candidate photovoltaic-site polygons in Rajasthan classified into post-processing development-priority classes. The priority classes are derived from candidate-region ranking scores and are distinct from the four pixel-level suitability classes used in the XGBoost classification.

The development-priority classes shown in Figure 11 are post-processing planning categories derived from the mean normalized suitability of the same five candidate-level indicators: GHI, PVOUT, slope suitability, road-distance accessibility, and power-line-distance accessibility. They should not be confused with the four pixel-level suitability classes used in the XGBoost classification. The five priority classes were produced after candidate-polygon extraction by grouping rounded candidate-rank scores into approximately equal-frequency classes: Very High (0.45–0.75), High (0.38–0.44), Moderate (0.33–0.38), Low (0.29–0.33), and Very Low (0.15–0.29). This transformation was used only to communicate relative development priority among candidate polygons.

Central Rajasthan shows a mixture of High and Moderate priority candidate polygons, while eastern and southeastern areas are relatively more constrained by terrain, vegetation,

and land-use conditions. The five priority levels provide a practical way to distinguish stronger contiguous screening units from more marginal candidate polygons.

3.8. District-Wise Solar Site Potential and Capacity Assessment

The district-level assessment shows strong spatial variation in solar site potential across Rajasthan. As summarized in Table 10, western districts dominate in terms of both suitable area and estimated capacity, highlighting their importance for large-scale solar development.

Table 10. District-wise summary of candidate solar sites, area, capacity, and priority classes. VH, H, M, L, and VL denote Very High, High, Moderate, Low, and Very Low post-processing priority classes, respectively.

District	Sites	Area (km ²)	Capacity (MW)	Best Rank	Mean Score	VH	H	M	L	VL
JAISALMER	24	4979.21	224,064.36	16	0.372	6	4	6	2	6
JODHPUR GRAMIN	23	3525.778	158,659.96	16	0.394	4	7	9	3	0
PHALODI	8	2864.036	128,881.59	16	0.365	2	1	2	1	2
PALI	5	1730.818	77,886.84	6	0.561	5	0	0	0	0
BARMER	36	1659.6	74,681.99	16	0.311	1	7	4	11	13
JALOR	6	1443.294	64,948.19	15	0.435	4	1	0	0	1
BALOTRA	34	1112.043	50,041.95	16	0.364	6	5	10	9	4
SIROHI	10	1065.444	47,944.93	6	0.5	7	2	1	0	0
SANCHORE	5	744.871	33,519.2	26	0.35	1	0	1	3	0
CHITTAURGARH	11	600.172	27,007.78	79	0.302	0	0	3	3	5
RAJSAMAND	21	574.053	25,832.46	4	0.475	11	7	2	1	0
BHILWARA	3	562.758	25,324.11	8	0.458	1	1	1	0	0
BEAWAR	12	476.278	21,432.58	16	0.282	1	0	3	2	6
UDAIPUR	20	421.896	18,985.29	1	0.448	7	6	4	2	1
NAGPUR	1	280.701	12,631.52	16	0.515	1	0	0	0	0
JODHPUR	6	42.546	1914.61	16	0.384	1	2	0	3	0
SALUMBAR	3	15.594	701.78	134	0.293	0	0	0	2	1
DUNGARPUR	2	3.452	155.35	117	0.32	0	0	0	2	0
PRATAPGARH	1	1.141	51.33	149	0.291	0	0	0	1	0

Jaisalmer shows the highest estimated capacity, with approximately 224,064 MW across 4979.21 km², making it the strongest district in the modeled planning output. This high score is mainly associated with extensive barren-land pixels, low slope constraints, and high GHI/PVOUT values in the input layers. Jodhpur Gramin and Phalodi also form a major high-potential solar cluster, with estimated capacities of approximately 158,660 MW and 128,882 MW, respectively.

In contrast, districts such as Dungarpur, Pratapgarh, and Salumbar show much lower potential, with estimated capacities below 1000 MW. These areas are more constrained by terrain variability, vegetation cover, and relatively lower solar resource availability.

An important finding is that the number of candidate sites does not necessarily indicate higher energy potential. For example, Barmer and Balotra have many candidate sites, but their total estimated capacity is lower than that of districts with fewer but larger contiguous suitable areas. This suggests that site size and spatial continuity are more important than site count for utility-scale solar deployment.

The mean rank score also highlights differences in site quality. Districts such as Pali, Nagaur, and Sirohi show relatively high mean rank scores despite having smaller total areas, suggesting that these districts contain compact but highly suitable locations.

The results indicate three major solar development corridors:

- Western corridor: Jaisalmer–Barmer–Phalodi
- Central corridor: Jodhpur–Pali–Nagaur
- Southern corridor: Udaipur–Rajsamand–Sirohi

These corridors show that the best solar locations are controlled by a combination of high solar resource availability, flat terrain, land availability, and infrastructure access.

3.9. Top Candidate Polygons and Capacity-Oriented Ranking

The top candidate polygons were extracted from the contiguous High and Very High suitability regions. For Figure 12, the ten largest capacity-oriented candidates were ranked using polygon area and the 45 MW/km² area-density method; the representative points are used only for cartographic clarity, while capacity was still calculated from the original polygon area. This capacity-oriented ranking is therefore separate from the five-indicator development-priority classes shown in Figure 11.

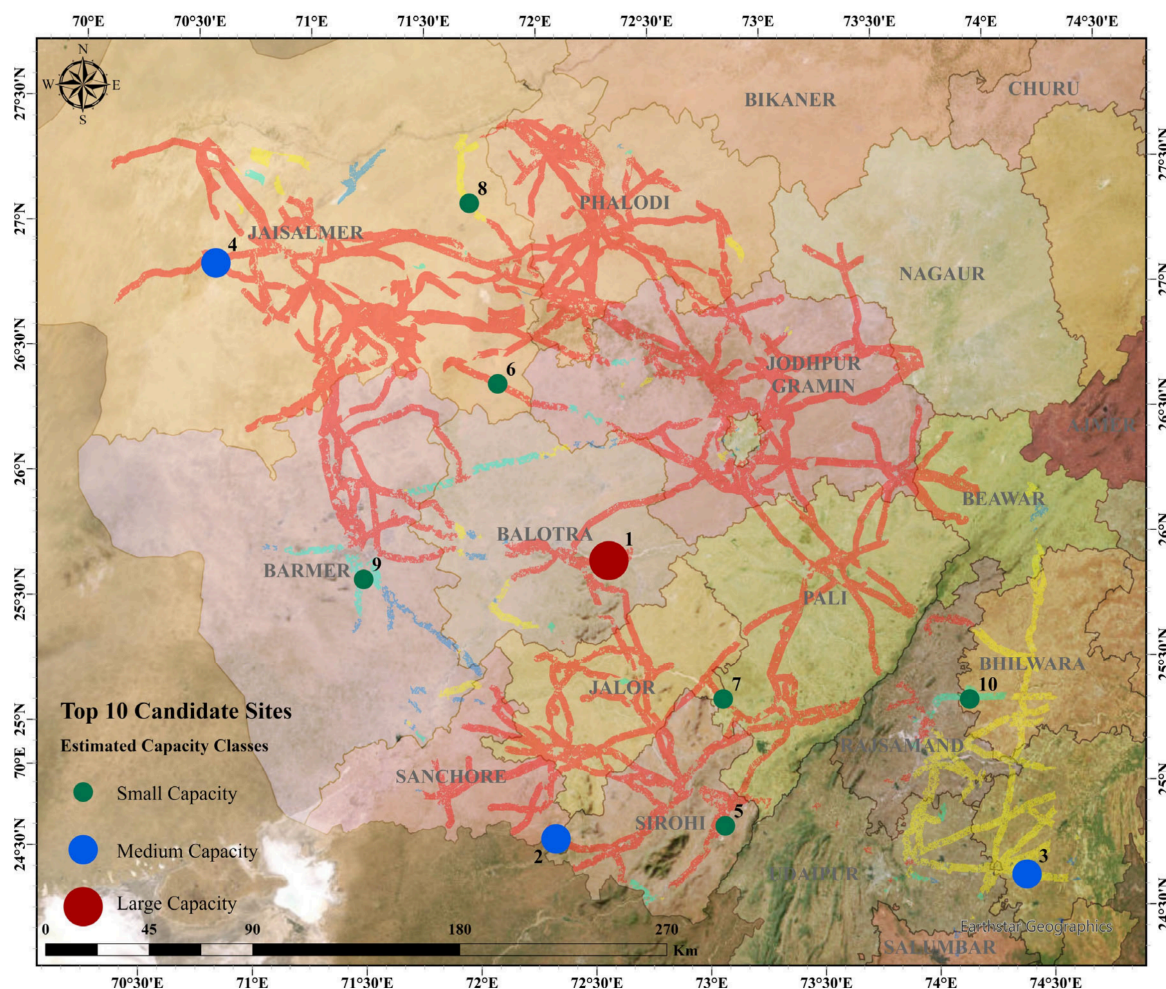


Figure 12. Representative points of the top 10 contiguous candidate-site polygons ranked by indicative installed capacity. Points were generated inside the original polygons for cartographic clarity, while capacity was calculated from polygon area using the area-density method. Point size indicates estimated capacity class, and numeric labels indicate capacity rank.

These candidate regions should be interpreted as screening units for further feasibility assessment. Their indicative capacities do not immediately imply buildable project sizes because cadastral boundaries, ownership, substation capacity, environmental permitting, water availability, and economic feasibility were not modeled at the parcel scale.

Several candidate zones show high indicative technical potential, especially in western Rajasthan. The overlap between candidate zones and known solar installations was examined using three reference solar-park polygons compiled from KMZ files: Bhadla Solar Park, Clean Solar Dedasari Solar, and Mahindra Solar Company. Overlay analysis showed that these reference sites fall overwhelmingly within High and Very High suitability classes, with Bhadla Solar Park showing 99.92% Very High pixels, Clean Solar Dedasari Solar 98.69% Very High pixels, and Mahindra Solar Company 94.61% Very High pixels. This comparison provides additional evidence of spatial plausibility while remaining limited to a small reference set.

3.10. Robustness and Sensitivity Analysis

To reduce dependence on a single set of assumed suitability weights and class thresholds, a sensitivity analysis was conducted. The pixel-level analysis recalculated the composite suitability score under alternative weighting and threshold scenarios and compared each High/Very High mask with the baseline mask using percentage-area change and Jaccard spatial overlap. The candidate-level analysis then tested whether the top capacity-oriented candidate polygons remained stable when the separate capacity-ranking emphasis was made area-dominant, resource-dominant, or access-dominant (Table 11).

Table 11. Robustness and sensitivity analysis of suitability classes and candidate-site ranking.

Scenario	Perturbation Tested	High/Very High Area (%)	Change vs. Baseline (pp)	Stability Result
Baseline	Original weights and thresholds	90.07	0.00	Reference case
Solar +20%	GHI and PVOUT weights increased 20%	87.60	−2.47	Jaccard = 0.973
Solar −20%	GHI and PVOUT weights decreased 20%	92.93	+2.86	Jaccard = 0.969
Infrastructure +20%	Road and power-line weights increased 20%	91.56	+1.49	Jaccard = 0.984
Terrain/LULC +20%	Slope and LULC weights increased 20%	91.33	+1.26	Jaccard = 0.986
Equal weights	All eight score components weighted equally	94.32	+4.25	Jaccard = 0.950
Relaxed thresholds	High and Very High thresholds lowered by 0.05	95.05	+4.97	Jaccard = 0.948
Strict thresholds	High and Very High thresholds raised by 0.05	80.18	−9.89	Jaccard = 0.890
Area-dominant ranking	Candidate ranking emphasizes polygon area	-	-	8/10 top candidates retained
Resource/access ranking	Candidate ranking emphasizes solar resource or access	-	-	10/10 top candidates retained

The $\pm 20\%$ weight perturbations produced relatively small changes in High/Very High area, ranging from -2.47 to $+2.86$ percentage points for solar-resource weights and from $+1.26$ to $+1.49$ percentage points for terrain/LULC and infrastructure weights. Spatial overlap with the baseline High/Very High mask remained high for these scenarios (Jaccard = 0.969 – 0.986), indicating that the main suitability pattern is not highly sensitive to moderate weight changes. Threshold changes produced larger differences, especially under strict thresholds, showing that class boundaries remain an important source of uncertainty (Figure 13). For candidate ranking, 8 of the baseline top 10 polygons were retained under the area-dominant scenario, while all 10 were retained under resource- and access-dominant rankings; the leading districts remained concentrated in Jodhpur Gramin, Jaisalmer, Phalodi, Pali, and Barmer. These results support the robustness of the broad spatial pattern while also justifying the manuscript's caution that suitability classes and candidate ranks should be treated as planning-screening outputs rather than final project decisions.

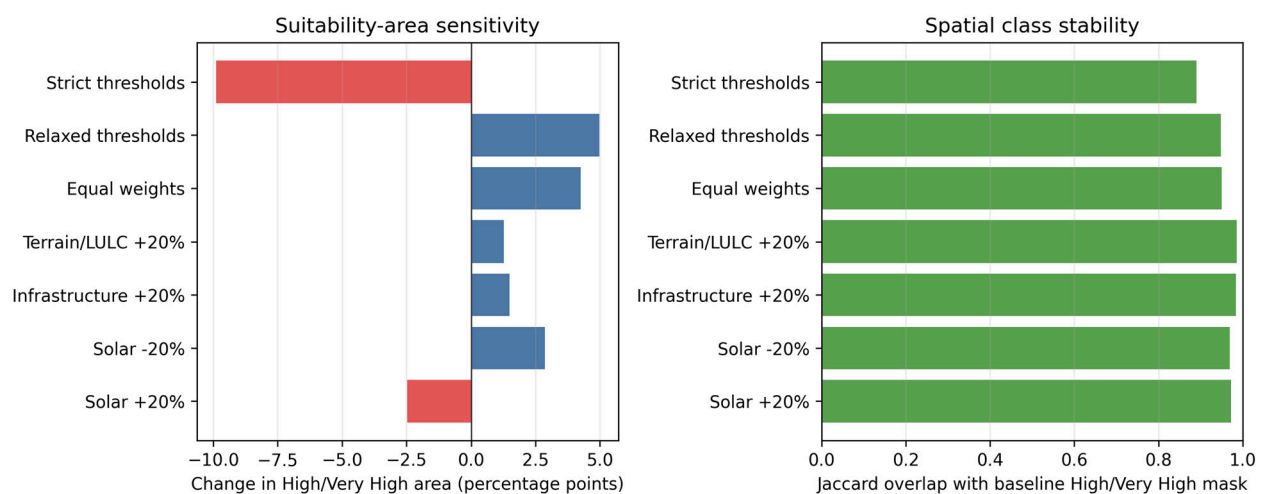


Figure 13. Sensitivity of High/Very High suitability area and spatial class overlap under alternative weight and threshold scenarios.

Additional uncertainty checks are summarized in Table 12. The correlated input-score perturbation produced a High/Very High area interval of 87.39 – 91.72% (5th–95th percentile), indicating that the broad suitability pattern is stable but not deterministic. Pixel-level class stability was also high: the mean same-class probability was 0.915 , 82.33% of sampled pixels retained their baseline class with probability ≥ 0.80 , and only 3.17% fell within an ambiguous High/Very High boundary-risk band ($p = 0.25$ – 0.75). The slope-cap scenarios did not change the mapped High/Very High area because the high-suitability pixels were already concentrated on low-slope terrain. In contrast, the LULC/built-up proxy showed stronger sensitivity: removing the hard LULC constraint increased the High/Very High area to 94.33% , while stricter low-LULC-score constraints reduced it to 50.64% . Capacity estimates were also sensitive to land-density assumptions, ranging from $773,629$ to $1,105,184$ MW under 35 – 50 MW/km². These results provide quantitative planning-level uncertainty ranges while reinforcing that the outputs are screening-level decision-support products.

The background-sampling row in Table 10 should be interpreted cautiously. The added analysis does not convert the presence-background validation into true presence-absence validation, and it does not eliminate potential preference-sampling bias. However, it now tests three alternative background definitions. The uniform buffered and environmental-stratified samples produced similar spatial ROC-AUC values (0.675 and 0.664), while

the preference-weighted plausible-siting background produced a slightly higher mean ROC-AUC of 0.688 and stronger presence enrichment (36.5%). These results show that the decision boundary is sensitive to background design, but observed solar locations remain enriched in high-scoring cells across strategies (Figure 14). Therefore, the presence-background scores should be used as supplementary plausibility evidence, not as an invariant or definitive accuracy estimate.

Table 12. Additional uncertainty propagation and parameter-sensitivity checks for the suitability and capacity assessment workflow.

Component	Test Design	Quantitative Result and Interpretation
Input data propagation	Monte Carlo perturbation of normalized predictor scores with correlated source-level bias and pixel-level residual noise; 500 runs on a 300,000-pixel sample; fixed seed 42. Class-stability probabilities were calculated from the same simulations.	High/Very High area mean = 89.72%; 5th–95th percentile = 87.39–91.72%. Mean same-class probability = 0.915; 82.33% of sampled pixels retained the baseline class with probability ≥ 0.80 ; 3.17% fell in the High/Very High boundary-risk band.
Slope mask	Additional hard slope caps were applied to the baseline High/Very High mask.	>5 , >10 , and >15 degree caps each retained 90.07% High/Very High area, equal to the baseline, because high-suitability pixels were already concentrated on very low-slope terrain.
Built-up/LULC mask	Categorical LULC/built-up exclusion sensitivity was tested using no hard LULC exclusion, the baseline water/built-up constraint, and stricter low-LULC-score thresholds as a built-up-intensity proxy.	No hard LULC exclusion = 94.33%; baseline = 90.07%; LULC_score $> 0.20 = 50.64\%$; LULC_score $> 0.50 = 50.64\%$. The LULC/built-up proxy is therefore a major source of uncertainty.
Capacity density	Area-density scenarios equivalent to 2.86–2.00 ha/MW were tested for the 22,103.68 km ² candidate area; the upper-density scenario corresponds to 2.00 ha/MW.	Capacity range = 773,629–1,105,184 MW under 35–50 MW/km ² ; baseline 45 MW/km ² = 994,666 MW. These are indicative planning ranges only.
Background sampling	Presence-background validation was repeated under uniform buffered, environmental-stratified, and preference-weighted plausible-siting background samples using spatial folds.	Complete-feature analysis used 3276 presence grid cells and 16,380 background cells per strategy. Spatial ROC-AUC = 0.664–0.688; PR-AUC = 0.340–0.428; presence enrichment above the background top decile = 26.6–36.5%.

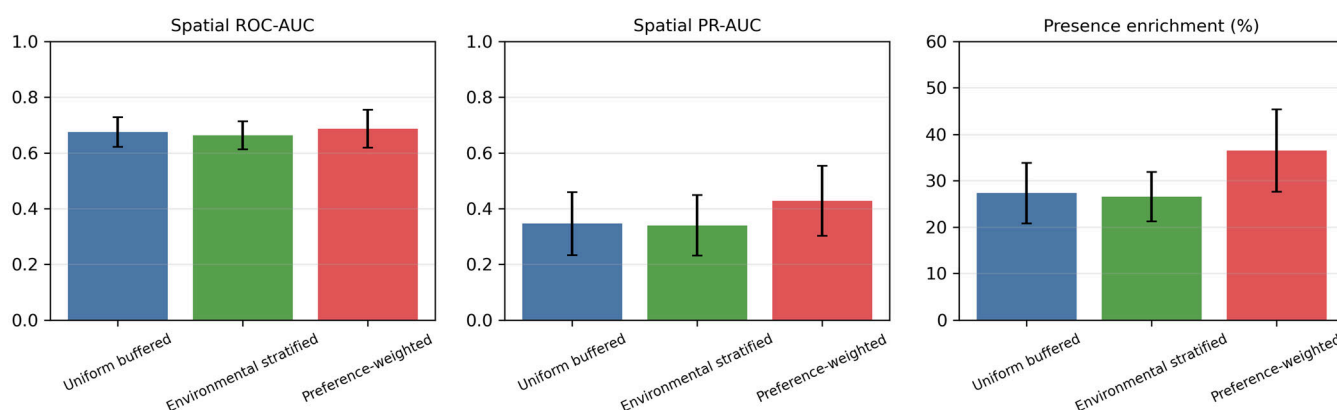


Figure 14. Background-sampling sensitivity of the supplementary presence-background validation under uniform buffered, environmental-stratified, and preference-weighted plausible-siting background samples. Bars show five-fold spatial cross-validation means, and error bars show standard deviations.

4. Discussion

Three points guide the discussion: the mapped High/Very High area is a screening construct, candidate polygons are more decision-relevant than statewide class totals, and model accuracy must be interpreted together with spatial validation and uncertainty analysis.

The results indicate that integrating geospatial analysis with machine learning can provide a useful framework for solar photovoltaic (PV) site suitability assessment, but the spatial extent of High and Very High suitability classes must be interpreted from a planning perspective rather than as immediately developable land. The finding that approximately 90% of valid pixels are High or Very High reflects the combination of the input GHI/PVOUT layers, dryland land-cover conditions, terrain constraints, infrastructure accessibility, and the adopted suitability thresholds, rather than a claim that 90% of Rajasthan is ready for project development. This interpretation is consistent with official recognition of Rajasthan's large solar potential [22] and with the spatial solar-resource products used in the analysis [23], but practical deployment would be much more selective after accounting for land tenure, protected areas, detailed grid capacity, water availability, local environmental constraints, social acceptance, and project economics.

Among all variables, solar resource indicators such as Global Horizontal Irradiance (GHI) and PVOUT emerged as the most influential predictors in the model. This is physically consistent with PV planning literature and with the Global Solar Atlas resource evidence used in this study, because irradiance is a primary driver of photovoltaic generation potential [8,18,23]. The SHAP-based feature importance results therefore indicate that the model reproduces interpretable components of the constructed suitability framework rather than relying only on obscure statistical artifacts.

Topographic conditions, particularly slope, affected suitability because flatter terrain generally reduces grading, construction complexity, and installation constraints for utility-scale PV development [19]. Elevation had lower SHAP importance in this analysis, suggesting that its role was mainly indirect through terrain complexity and accessibility rather than a dominant suitability driver.

Infrastructure accessibility adds practical planning context by indicating where road access and transmission proximity may reduce construction, maintenance, and grid-connection costs [8]. At the same time, the results show that areas with weaker immediate accessibility may remain relevant for early-stage screening if they contain large contiguous candidate polygons and strong solar-resource conditions.

The district-level analysis provides additional insights for regional planning. Western districts such as Jaisalmer, Jodhpur Gramin, and Phalodi emerged as favorable zones because they combine high irradiance, relatively flat terrain, and limited land-use constraints. This result is consistent with the presence of major solar developments in western Rajasthan, including the Bhadla region. The external overlay with WRI/GPPD, GEM, and OSM solar-installation records provides additional plausibility support, although it should not be interpreted as a complete field validation because existing plant locations are presence-only records influenced by land ownership, policy, grid access, and investment decisions.

A practical contribution of this study is the translation of pixel-level suitability mapping into planning-oriented outputs. The workflow extracts contiguous candidate polygons, ranks them, estimates indicative installed capacity using a dimensionally consistent area-density method, compares XGBoost with simpler models, and evaluates mapped outputs using independent solar-installation records. This makes the framework more informative for early-stage screening than a standalone suitability map.

The integration of machine learning with explainable AI (SHAP) strengthens the analysis by improving model transparency and interpretability. At the same time, the validation results show why caution is necessary when interpreting Very High pixel-level

accuracy. The random split produced higher performance than the spatial-block split, which is consistent with the presence of spatial autocorrelation among neighboring pixels. In addition, the final labels were generated from a suitability score that uses several of the same variables supplied to the model, meaning that some predictors partially encode the target definition. The comparison with logistic regression, decision tree, and random forest baselines shows that XGBoost provided the best overall performance, but it also shows that simpler tree-based models can approximate much of the constructed suitability logic. Therefore, XGBoost is justified primarily by its stronger performance, nonlinear modeling capacity, and compatibility with SHAP interpretation, rather than by accuracy alone.

Despite these strengths, several limitations should be acknowledged. First, the suitability labels were generated from a composite rule-based index, and the same predictor family was used to train the machine-learning models. This design is appropriate for learning and mapping the constructed suitability framework, but it can inflate classification accuracy if interpreted as independent empirical validation. Second, random pixel splitting can overestimate performance because of spatial autocorrelation; the spatial-block validation reduces this risk but does not replace independent field validation. Third, assumed parameter values, including variable scoring rules, class thresholds, the 1 km² minimum candidate-area threshold, and the 45 MW per km² capacity density, introduce uncertainty into the candidate-site and capacity estimates. Fourth, the 250 m analysis resolution may miss small suitable patches and local exclusion features. Fifth, the analysis does not include land ownership, land price, protected-area restrictions, water availability, detailed transmission hosting capacity, or policy constraints. Sixth, the added presence-background analysis reduces reliance on random pixel-split accuracy, but it remains a presence-background test rather than a true presence-absence validation because background cells are not verified as unsuitable or rejected solar-development locations. Finally, most inputs are static or long-term mean layers, whereas solar resources, aerosols, temperature, and land-cover conditions vary seasonally and interannually. Future research should incorporate time-series solar and climate data, independent project-performance observations, finer grid-infrastructure constraints, and socio-economic feasibility layers.

The additional uncertainty assessment partially responds to some of these limitations by providing scenario-based intervals for input-score perturbation, class-stability probability, LULC/built-up masking, slope screening, capacity density, and background-sampling design. However, these intervals should not be read as definitive confidence intervals for construction-ready project decisions. A full uncertainty-propagation framework would require source-specific error distributions for each raster, spatial covariance structures, LULC class-probability layers, independent field observations of rejected and accepted project sites, and repeated project-level model fitting under externally documented sampling designs.

The findings indicate that GeoAI-based approaches can support scalable solar site screening. The agreement with independent solar-installation records and the supplementary India-wide presence-background spatial cross-validation provide supporting evidence for spatial plausibility, while the spatial-block validation and external overlay also clarify the limits of interpreting pixel-level accuracy as real-world predictive certainty.

5. Conclusions

This study developed an explainable GeoAI decision-support framework for photovoltaic (PV) site suitability assessment by integrating multi-source geospatial data, rule-derived suitability labels, machine-learning classification, SHAP interpretation, spatial validation, presence-based plausibility checks, and uncertainty/sensitivity analysis. The central contribution is methodological: the workflow makes the origin of suitability la-

bels explicit, treats model metrics as agreement with a constructed decision framework, and translates pixel-level suitability into contiguous candidate polygons and district-level planning summaries.

The modeling innovation lies in combining transparent label construction with XGBoost benchmarking, simpler-model comparison, spatial-block validation, and SHAP-based interpretation. This design reduces the risk of presenting high random-split accuracy as independent field validation and provides a more reproducible framework for regional PV screening.

At the regional level, the mapped outputs indicate that western districts such as Jaisalmer, Jodhpur Gramin, and Phalodi contain large contiguous screening zones under the adopted input data and thresholds. These areas combine high GHI/PVOUT values, low slope, and suitable land-cover conditions, but they should be interpreted as planning-screening priorities rather than confirmed developable parcels.

The uncertainty analysis provides planning-level risk information by reporting input-perturbation intervals, class-stability probabilities, LULC/built-up mask sensitivity, capacity-density ranges, and background-sampling sensitivity. The results suggest that the broad state-scale pattern is stable, but candidate boundaries, capacity estimates, and validation scores remain sensitive to exclusion rules and sampling design.

Future research should extend the framework using finer-resolution parcel and grid datasets, seasonal solar and climate time series, explicit land-ownership and protected-area constraints, substation-hosting-capacity data, project-cost modeling, and larger field-verified validation samples. These additions would support the transition from state-scale screening to project-level feasibility assessment.

Overall, the framework is intended for early-stage, risk-aware spatial screening and can be transferred to other regions when local data sources, exclusion rules, validation records, and uncertainty assumptions are clearly documented.

Author Contributions: Conceptualization, C.N. and S.S.; methodology, C.N. and S.S.; software, C.N.; validation, C.N., S.K.M., J.G., R.A. and S.S.; formal analysis, C.N.; investigation, C.N., S.K.M. and M.Z.; resources, S.S., Z.K., A.K., F.F.B.H. and M.Z.; data curation, C.N.; writing—original draft preparation, C.N.; writing—review and editing, S.K.M., S.S., J.G., R.A., A.K., F.F.B.H., Z.K. and M.Z.; visualization, C.N.; supervision, S.S. and M.Z.; project administration, M.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R675), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Data Availability Statement: Public source datasets used in this study are available from their provider portals: Global Solar Atlas for GHI/PVOUT (<https://globalsolaratlas.info/download/world>, (accessed on 12 June 2026)), Copernicus Climate Data Store ERA5 for climate variables (<https://cds.climate.copernicus.eu>), USGS EarthExplorer/NASA SRTM for DEM data (<https://earthexplorer.usgs.gov>), Dynamic World via Google Earth Engine for LULC (https://developers.google.com/earth-engine/datasets/catalog/GOOGLE_DYNAMICWORLD_V1 (accessed on 12 June 2026)), OpenStreetMap for road, power-line, and solar-farm features (<https://www.openstreetmap.org>), WRI Global Power Plant Database (<https://www.wri.org/research/global-database-power-plants> (accessed on 12 June 2026)), and Global Energy Monitor Global Solar Power Tracker (<https://globalenergymonitor.org/projects/global-solar-power-tracker/> (accessed on 12 June 2026)). Processed rasters, candidate-site layers, district summaries, and validation outputs generated during the study are available from the corresponding author upon reasonable request, subject to data-provider licensing conditions. The additional uncertainty-analysis outputs, including the Monte Carlo class-stability summary, deterministic mask-sensitivity counts, capacity-density sensitivity

table, and background-sampling strategy sensitivity table and figure, are stored with the revised manuscript outputs and are available from the corresponding author upon request.

Acknowledgments: The authors extend their appreciation to Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R675), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This paper is based on work supported by the Faculty of Engineering, Mansoura University, Project number ENGFAC-RSU-2026. We are also thankful to the JST SICORP Grant Number JPMJSC23A1.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
DL	Deep Learning
EO	Earth Observation
RS	Remote Sensing
GEE	Google Earth Engine
EPSG	European Petroleum Survey Group (Coordinate System Code)
kWh	Kilowatt-hour
kWp	Kilowatt-peak
DNI	Direct Normal Irradiance
DHI	Diffuse Horizontal Irradiance
LCOE	Levelized Cost of Energy
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
PCA	Principal Component Analysis
AHP	Analytic Hierarchy Process
AOD	Aerosol Optical Depth
BRDF	Bidirectional Reflectance Distribution Function
DEM	Digital Elevation Model
F1-score	Harmonic mean of precision and recall
GEM	Global Energy Monitor
GeoAI	Geospatial Artificial Intelligence
GHI	Global Horizontal Irradiance
GIS	Geographic Information System
GPPD	Global Power Plant Database
H	High priority class
L	Low priority class
LULC	Land Use/Land Cover
M	Moderate priority class
MCDM	Multi-Criteria Decision-Making
MERRA-2	Modern-Era Retrospective Analysis for Research and Applications, Version 2
MIF	Multi-Influencing Factor
MODIS	Moderate Resolution Imaging Spectroradiometer
MW	Megawatt
NASA	National Aeronautics and Space Administration
OSM	OpenStreetMap
PR-AUC	Precision-Recall Area Under the Curve
PV	Photovoltaic
PVOUT	Photovoltaic Power Output
ROC-AUC	Receiver Operating Characteristic Area Under the Curve
SAW	Simple Additive Weighting
SHAP	SHapley Additive exPlanations

SRTM	Shuttle Radar Topography Mission
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
USGS	United States Geological Survey
VH	Very High priority class
VL	Very Low priority class
WRI	World Resources Institute
XGBoost	Extreme Gradient Boosting

References

1. International Energy Agency. *Renewables 2024: Analysis and Forecast to 2030*; IEA: Paris, France, 2024. Available online: <https://www.iea.org/reports/renewables-2024> (accessed on 12 June 2026).
2. Wang, X.; Sun, X.; Ahmad, M.; Chen, J. Energy transition, ecological governance, globalization, and environmental sustainability: Insights from the top ten emitting countries. *Energy* **2024**, *292*, 130551. [\[CrossRef\]](#)
3. Andrade-Arias, A.S.; Kabir, G.; Elizondo-Noriega, A.; Mirmohammadsadeghi, M. Public Acceptability of Solar Energy Implementation in Mexico. *Eng. Proc.* **2024**, *76*, 7. [\[CrossRef\]](#)
4. de Luis-Ruiz, J.M.; Salas-Menocal, B.R.; Pereda-García, R.; Pérez-Álvarez, R.; Sedano-Cibrián, J.; Ruiz-Fernández, C. Optimal Location of Solar Photovoltaic Plants Using Geographic Information Systems and Multi-Criteria Analysis. *Sustainability* **2024**, *16*, 2895. [\[CrossRef\]](#)
5. Yilmaz, İ.; Kocer, A.; Aksoy, E. Site selection for solar power plants using GIS and fuzzy analytic hierarchy process: Case study of the western mediterranean region of Türkiye. *Renew. Energy* **2024**, *237*, 121799. [\[CrossRef\]](#)
6. Jong, F.C.; Ahmed, M.M. Multi-Criteria Decision-Making Solutions for Optimal Solar Energy Sites Identification: A Systematic Review and Analysis. *IEEE Access* **2024**, *12*, 143458–143484. [\[CrossRef\]](#)
7. Almasad, A.; Pavlak, G.; Alquthami, T.; Kumara, S. Site suitability analysis for implementing solar PV power plants using GIS and fuzzy MCDM based approach. *Sol. Energy* **2023**, *249*, 642–650. [\[CrossRef\]](#)
8. Ashraf, H.A.; Li, J.; Li, Z.; Sohail, A.; Ahmed, R.; Butt, M.H.; Ullah, H. Geographic Information System and Machine Learning Approach for Solar Photovoltaic Site Selection: A Case Study in Pakistan. *Processes* **2025**, *13*, 981. [\[CrossRef\]](#)
9. Ghiyas, S.; Sikarudi, D.; Jafari, M. A novel integrated GIS-AI framework for optimal CSP plant site selection: A multi-criteria approach under climate change scenarios in Bushehr, Iran. *Adv. Space Res.* **2025**, *76*, 5146–5167. [\[CrossRef\]](#)
10. Khan, A.; Ali, Y.; Pamucar, D. Solar PV power plant site selection using a GIS-based non-linear multi-criteria optimization technique. *Environ. Sci. Pollut. Res.* **2023**, *30*, 57378–57397. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Khazael, S.M.; Maulud, K.N.A.; Gibril, M.B.A.; Sachit, M.S.; Karim, O.A. Enhancing solar PV suitability mapping in the Middle East using an optimized deep learning framework. *Alex. Eng. J.* **2025**, *129*, 553–571. [\[CrossRef\]](#)
12. Shen, J.; Zheng, F.; Tesema, F.B.; Deng, W.; Bellotti, A.G.; Xie, J. Deep learning and explainable Artificial Intelligence for large-scale photovoltaic suitability analysis in Australian cities. *Renew. Energy* **2026**, *259*, 125025. [\[CrossRef\]](#)
13. Guntupalli, R.; Pradeepkumar, S.K.B.; Duddeti, B.B.; Ankireddy, N.; Meena, V.P.; Jadoun, V.K. A hybrid ANN-AHP-GIS framework with dimensionality reduction and uncertainty quantification for solar site selection in Southern India. *Energy Convers. Manag. X* **2025**, *28*, 101280. [\[CrossRef\]](#)
14. Zhong, C.; Peng, L.; Yu, J.; Swan, I.; Li, H. Toward more reliable, complete, and equitable global urban land use efficiency assessments. *Commun. Earth Environ.* **2025**, *6*, 1055. [\[CrossRef\]](#)
15. Shehab, Z.N.; Faisal, R.M.; Ahmed, S.W. Multi-criteria decision making (MCDM) approach for identifying optimal solar farm locations: A multi-technique comparative analysis. *Renew. Energy* **2024**, *237*, 121787. [\[CrossRef\]](#)
16. Liang, L.; Wang, X.; Zhang, X.; Jia, X.; Chen, Z.; Zheng, X. A systematic review of photovoltaic power plant site selection approaches: Literature retrieval and analysis method based on ChatGPT and DeepSeek. *Energy Rep.* **2025**, *14*, 1993–2014. [\[CrossRef\]](#)
17. Ghaedrahmati, H.; Talebi, S.; Moradi, A.; Eskandari, A.; Parvin, P.; Aghaei, M. Potential analysis and energy prediction of photovoltaic power plants using satellite-based remote sensing and artificial intelligence techniques. *Front. Energy Res.* **2025**, *13*, 1611429. [\[CrossRef\]](#)
18. Reddy, B.N.S.; Gautam, K.; Pachauri, N. Solar potential assessment using machine learning and climate change projections for long-term energy planning. *Sci. Rep.* **2025**, *15*, 39935. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Rane, N.L.; Günen, M.A.; Mallick, S.K.; Rane, J.; Pande, C.B.; Giduturi, M.; Bhutto, J.K.; Yadav, K.K.; Tolche, A.D.; Alreshidi, M.A. GIS-based multi-influencing factor (MIF) application for optimal site selection of solar photovoltaic power plant in Nashik, India. *Environ. Sci. Eur.* **2024**, *36*, 5. [\[CrossRef\]](#)
20. Singh, S.; Avtar, R.; Jain, A.K.; Alkhuraji, W.S.; Zhran, M. Influence of Land Use/Land Cover Dynamics on Urban Surface Metrics in Semi-Arid Heritage Cities. *Land* **2025**, *14*, 1834. [\[CrossRef\]](#)

21. Singh, S.; Pandey, S.; Jain, A. Integrating Temporal Change Detection and Advanced Hybrid Modeling to Predict Urban Expansion in Jaipur, a UNESCO World Heritage City. *Rev. Int. Géomat.* **2025**, *34*, 899–914. [[CrossRef](#)]
22. Government of Rajasthan. *Rajasthan Integrated Clean Energy Policy*; Government of Rajasthan: Jaipur, India, 2024. Available online: <https://jankalyanfile.rajasthan.gov.in/Content/UploadFolder/DepartmentMaster/197/2024/Dec/30409/197b7999796-4748-4300-8898-61b86f301947.pdf> (accessed on 12 June 2026).
23. World Bank Group. Energy Sector Management Assistance Program (ESMAP), and Solargis. Global Solar Atlas 2.0 Data: Global Horizontal Irradiation and Photovoltaic Power Potential. 2020. Available online: <https://globalsolaratlas.info/download/world> (accessed on 12 June 2026).
24. Brown, C.F.; Brumby, S.P.; Guzder-Williams, B.; Birch, T.; Hyde, S.B.; Mazzariello, J.; Czerwinski, W.; Pasquarella, V.J.; Haertel, R.; Ilyushchenko, S.; et al. Dynamic World, Near real-time global 10 m land use land cover mapping. *Sci. Data* **2022**, *9*, 251. [[CrossRef](#)]
25. Data[Meet] Community Maps. India Administrative Boundaries/District Boundaries. GitHub Repository. Available online: <https://github.com/datameet/maps> (accessed on 12 June 2026).
26. Munoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **2021**, *13*, 4349–4383. [[CrossRef](#)]
27. Gelaro, R.; McCarty, W.; Suárez, M.J.; Todling, R.; Molod, A.; Takacs, L.; Randles, C.A.; Darmenov, A.; Bosilovich, M.G.; Reichle, R.; et al. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Clim.* **2017**, *30*, 5419–5454. [[CrossRef](#)] [[PubMed](#)]
28. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **2007**, *45*, RG2004. [[CrossRef](#)]
29. NASA LP DAAC. MODIS/Terra+Aqua BRDF/Albedo Daily L3 Global 500 m SIN Grid V061 (MCD43A3). 2021. Available online: <https://lpdaac.usgs.gov/products/mcd43a3v061> (accessed on 12 June 2026).
30. OpenStreetMap Contributors. OpenStreetMap Planet Data and Derived Vector Features. 2026. Available online: <https://www.openstreetmap.org> (accessed on 12 June 2026).
31. Phillips, S.J.; Anderson, R.P.; Schapire, R.E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **2006**, *190*, 231–259. [[CrossRef](#)]
32. Elith, J.; Phillips, S.J.; Hastie, T.; Dudík, M.; Chee, Y.E.; Yates, C.J. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* **2011**, *17*, 43–57. [[CrossRef](#)]
33. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Aroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [[CrossRef](#)]
34. Ong, S.; Campbell, C.; Denholm, P.; Margolis, R.; Heath, G. *Land-Use Requirements for Solar Power Plants in the United States*; Technical Report NREL/TP-6A20-56290; National Renewable Energy Laboratory: Golden, CO, USA, 2013. [[CrossRef](#)] [[PubMed](#)]
35. World Resources Institute. Global Power Plant Database. 2024. Available online: <https://www.wri.org/research/global-database-power-plants> (accessed on 12 June 2026).
36. Global Energy Monitor. Global Solar Power Tracker. February 2026 Release. Available online: <https://globalenergymonitor.org/projects/global-solar-power-tracker> (accessed on 12 June 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.