AI-Driven Site Suitability Modelling for Solar Farms using Regression and SHAP Analysis

Anandita Garg, Trinav Talukdar and Chinmayi Kalapatapu

Abstract

The importance of clean and renewable energy has become evident in the past few decades. This is owing to growing populations, over-consumption of natural resources and rising concerns around global climate change. A shift to reliance on solar energy has emerged as a viable alternative to fossil fuels.

In India, solar energy has significantly contributed to the renewable energy sector in recent years, accounting for 92.12 GW out of the 203.18 GW of clean energy generated. The deployment of solar farms is a key aspect of this effort, driving studies aimed at selecting optimal sites for photovoltaic farm installations.

A major challenge lies in identifying locations that can yield the highest solar output. Factors related to geographical location, terrain and weather become important in this analysis. Therefore, identifying ways of maximising solar potential becomes an area requiring extensive research.

A review of past studies has revealed that current methodologies for predicting solar sites primarily employ statistical methods. These models are based on subjective Multi-Criteria Decision-Making (MCDM) techniques, such as the Analytic Hierarchy Process (AHP) and Weighted Linear Combination (WLC).

These traditional methodologies are limited in their ability to work with a wider spectrum of features and data points. Studies deploying Artificial Intelligence and Machine Learning methods are relatively unexplored in this domain. This highlights a major gap; the absence of robust, scalable models capable of integrating a diverse range of features.

Here, we present a model that is driven by an AI-ML approach to predict optimality of locations for solar farm installation using various regressor architectures (Random Forest, K- Nearest Neighbours, Multi-Layer Perceptron, Support Vector Machines and Extreme Gradient Boosting). Our approach enables highly precise assessments, taking into account a multitude of diverse factors, considering data points from all over India.

Our efforts also advance research into alternatives to AHP-GIS methodologies in this domain, presenting a compelling case for further exploration of the advantages of AI over traditional methods in solar site selection. Our codebase has been released as open source.

Introduction

India stands as the most populated country in the world with a reported population of 1,463.9 million in 2025, according to the UN World Population Dashboard. The need and pursuit of non-conventional energy sources increases parallel to the growing population and technology dependency of a nation [IEA, 2024]. As noted by the IEA, this phenomenon is particularly observed in dense and developing countries like India. Supporting this, India's energy demand is projected to increase by 35% by 2035, necessitating urgent policy shifts towards renewable energy sources [IEA, 2024].

Relevant alternatives to fossil-fuel-based energy sources include solar energy. As of March 20, 2023, India had an installed solar capacity of approximately 66,780 MW, with major plants distributed across various states [US Department of Energy, n.d.].

Installing solar farms is one of the most effective ways of harnessing solar energy. Solar farms can be defined as large-scale installations of solar panels that harness sunlight to generate electricity. These farms consist of large arrays of photovoltaic panels (PV) that convert sunlight into electricity, offering a scalable and sustainable solution to meet substantial energy demands [Panagoda, 2023, p. 32]. In addition to their environmental virtues, solar farms also provide economic benefits by creating jobs and stimulating local economies[Singh et al., 2017, p.270]. A study by Steffen et al., published in Renewable Energy Focus indicates that in 2017, operations and maintenance (O&M) accounted for 20%–25% of life cycle costs of wind and solar plants in Europe, highlighting the relatively low O&M expenses associated with these renewable energy sources, further demonstrating the case for installation of solar farms as a viable, sustainable source of energy.

With the inquiry into the widespread utilisation and harnessing of solar energy gaining prominence, comes the challenge of identifying suitable locations to set up solar farms while considering a wide range of factors. Bamisile et al., in a paper published in February 2025, talked about environmental factors affecting solar photovoltaic output. A study conducted by Onuoha et al. in 2025, in Nigeria concluded that social and economic factors also determine the the optimality of locations the selection of photovoltaic farm sites. farm sites.

Current methods for selection of solar farm locations rely on subjective Multi-Criteria DecisionMaking (MCDM) techniques, such as the Analytic Hierarchy Process (AHP) and Weighted Linear Combination (WLC) [Garni & Awasthi, 2017]. These techniques help rank sites based on various factors, but they have shortcomings. They often suffer from subjectivity and lack of transparency [De Silva, 2024, p.4]. The selection process used has a tendency to be influenced by the researcher's subjectivity, expertise level, and practical experience in the field [Sachit et al., 2022].

When it comes to India specifically, most research on solar energy potential has been restricted to rooftop solar feasibility in very specific locations rather than large-scale solar farms [Kumar, 2019]. Hence, a major gap arises in the lack of robust and scalable models that integrate a wide spectrum of features and data points. Our solution is rooted in adopting a data-driven, AI-ML approach to predict optimality of sites for installation of solar farms.

Discipline And Thematic

Our research lies at the crossroads of Environmental Science, Geospatial Science, and Artificial Intelligence. We aim to combine multiple datasets using precise geographic coordinates, accounting for the dramatic climate and terrain variations across states in an Indian con- text. The shift from static statistical methods to AI-driven predictive frameworks is not just a technological improvement, it also represents a paradigm shift in sustainability research [Antonanzas et al., 2021]. For our model, we explore multiple Machine Learning algorithms and integrate SHAP (SHapley Additive Explanations) to enhance interpretability [Lundberg, 2017, p.1] thereby ensuring greater transparency and accountability in decision-making driven by our AI-ML approach.

Beyond developing a highly accurate and interpretable model, we aim to democratise access to solar energy data by relying on open-source global datasets such as Open Street Map, Earth Engine, and Open Meteo.

This project presents a novel attempt to modernise solar farm research by adopting a rather unexplored approach, especially in an Indian context. With climate change emerging as a critical global crisis, we believe this work is essential for ensuring sustainable development and contributing to UN Sustainable Development Goals (SDG 7: Affordable and Clean Energy, SDG 13: Climate Action) [Frey Energy, n.d.].

Overview of Scholarly Discourse in the field Pre-AI approaches (1960-present)

GIS and MCDM approaches have dominated the field of solar optimisation since their inte- gration in the early 2000s. GIS (Geographical Information Systems) are tools that help enable the creation of exten- sive databases due to their ability to effectively integrate diverse spatial data sources such as maps, aerial photographs and even quantitative and qualitative data. [De Silva, 2024, p.14]. However, as significant as GIS is in the manipulation and analysis of spatial data, GIS tools lack the ability to perform complex decision-making especially when "multiple and conflicting criteria and objectives are concerned" as noted by Stephen Carver in his 1991 paper "Integrating multi-criteria evaluation with geographical information systems". [Carver, 1991, pg.321]. As a result there was an integration of MCDM techniques alongside GIS. MCDM or Multiple Criteria Decision-Making, is an age old technique used in "structuring, decision-making, and planning steps when the domain possesses manifold criteria to reach an optimum solution based on the deciders' preferences" [Taherdoost & Madanchian, 2023, pg.77].

While Stephen Carver's 1991 paper appears to be one of the earliest papers integrating the two techniques, his paper focuses on environmental management and land use. The earliest mention in the domain of solar optimisation appears to be in 2008 by Ara´ n Carrio´ n et al, aiming to solve the solar farm optimisation problem in Spain. Through their research, they developed an Environmental Decision Support System (EDSS) using GIS and AHP (a widely used MCDM technique), and their model was able to effectively point outregionssuitable for PV power plants that were verified with case studies [Ara´ n Carrio´ n et al., 2008, pg. 237].

For almost two decades after this, there have been multiple studies exploring the integration of various MCDM models with GIS thus identifying significant insights to solving the solar farm optimisation problem. In 2014, Chen et al, created a model using GIS and DEMATEL, which unlike common MCDM models, also considered the inter-dependencies and feedback loops among the features. This helped provide insights such as the significant role of orography/land topology in enhancing the performance of solar technologies. Similarly, in 2016 a study conducted by Ahemadi et al, identified climatic factors such as temperature, radiation, precipitation, sundial, evaporation, alongside topological factors that contributed to optimal site selection of solar power plants [Ahmadi, Morshedi, & Azimi, 2016].

These studies are only representative of a small fraction of the vast research in the field. In 2020, a literature review focused on the use of GIS and AHP in solar farm optimisation identified 47 relevant studies over 20 different countries, of which, 24 studies combined GIS and AHP and 6 combined GIS with Fuzzy AHP [Suprova, Zidan, & Rashid, 2022, pg. 3-4].

AI-based approaches (2020-present)

While GIS and MCDM approaches continue to dominate the field of selecting optimal solar farm locations, in the 2020s, another method arose with promising results. In the US researcher De Silva's paper, "Suitability Mapping of Solar Power Plants Using An Explainable AI- based Approach", they provide criticism for the current GIS-MCDM approaches for inherent subjectivity and a lack of transparency. This arises due to the selection parameters and the weights assigned to them being influenced by an expert's subjectivity, expertise level and practical experience in the field [De Silva, 2024, pg. 14].

Another example of a model making use of this Explainable AI technology was proposed by Rudro et 103 al, whose SPXAI model extensively collects power production data from solar farms and makes use of ML and DL models to analyse this data on an hourly basis [Rudro, Na- har, & Al Sohan, 2024, pg. 1]. Theirmodel provides "clearinsightsinto predictions, identifies influential factors, and offers rule-based explanations for complex model decisions"[Rudro, Nahar, & Al Sohan, 2024, pg. 1]. While Rudro et al, provided a framework for a model, De Silva implemented this approach by integrating Explainable AI (XAI) with five different ML models—Random Forest (RF), Support Vector Machines (SVM), Multi-layer Percep- tron (MLP), Decision Tree (DT), and K-Nearest Neighbours (k-NN)—to challenge existing GIS-MCDM models [De Silva, 2024, pg. iv].

Through this study they identified the highest overall accuracy of 88% using the RF model. Further analysis with SHapley Additive exPlanations (SHAP), an XAI technique, showed that Solar Radiation

(13%) and Cloud Index (12%) were the most influential variables affecting model predictions.[De Silva, 2024, pg. iv]. However, this was tailored for a database of US solar plants. Additionally, ground truth valuesfor the coordinates were found by calculating the solar efficiency ofthe existing solar plants that were then attributed to the other coordinates based on their distance from the existing plants.

The papers authored by Sachit et al, and Sun also contributed to the small pool of literature on explainable AI-ML techniques in solar plant optimisation. Both used binary classification and highlighted the importance of considering environmental,socioeconomic, technical, and infrastructural characteristics. However, by doing binary classifications, they lose granularity in the target variable value. Thus, this becomes unreflective of the real world where locations cannot always be classified as optimal or not optimal, thus taking away the nuance of how optimal a location is over another.

Methodology

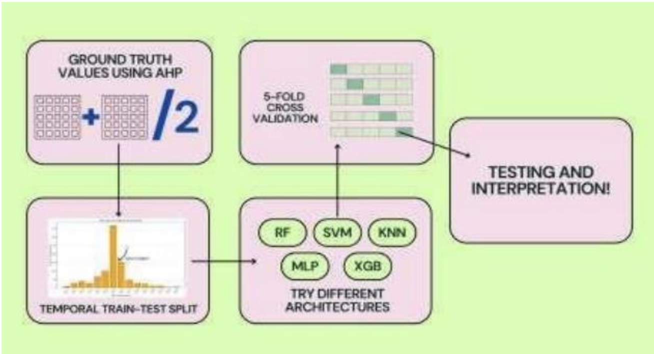

Figure 1: The figure depicts our chosen methodology framework

Initial Processing

We wanted to use various regression models to determine how optimal a location is for a solar farm. We developed a scoring system to help us interpret the obtained values:

Score Range	Interpretation
0–0.2	Very Low Suitability
0.2–0.4	Low Suitability
0.4–0.6	Moderate Suitability
0.6–0.8	High Suitability
0.8–1.0	Very High Suitability

Table 1: Interpretation scale for scores from 0 to 1. It is used to assess the suitability of a location for solar farm development. The role of this scale is to help standardise the evaluation of model outputs and model decision-making

Ground Truth Values

To start our analysis, we randomly sampled 2792 coordinates in India that do not have solar farms at present. A major challenge in this process was determining the ground truth values. To address this, we decided to use Analytical Hierarchy Processing (AHP). AHP is a decision-making methodology, particularly useful where multiple criteria are involved. It organises the decision-making process into a hierarchical structure, starting with the construction of a pairwise comparison matrix, where each element’s importance is weighed against others. This is based on a pair-wise comparison questionnaire designed to collect judgments and opinions from respondents. Once this is done, the principal eigenvector of the matrix is calculated to derive the normalised weights for each element in the hierarchy. The Consistency Ratio (CR) is calculated to determine whether the judgments are consistent (if theCRis higherthan 0.10, the results are deemed unacceptable).In this case, adjustments can be made and reiterations can be carried out. Then, the eigenvector is calculated, representing the normalised weights, and these values are used to prioritise decision elements based on their importance in relation to the overall goal [Taherdoost, 2017, p. 245].

We referenced two different papers that applied AHP and determined their pairwise comparison matrices using expert opinion, serving as a basis for our approach [Taherdoost, 2017, p. 245].

S. No.	Feature (Relative Weight)	Source
1	Solar Radiation (13.56), Wind Speed (12.71), LST (11.86), Relative Humidity (11.02), Vegetation (10.17), Distance from Road (9.32), Elevation (11.02), Aspect (11.02)	Rane, 2024, p.6.
2	Solar Radiation (10), Temperature (15), Cloud Cover (5), Humidity (5), Elevation (12), Slope (35), Land Cover (10), Distance from Roads (8)	Dhadiwal, 2020, p.23.

Table 2: Comparison of feature weights from two studies that use AHP for solar site selection

The variations in the weights appear to highlight the subjective nature of AHP in the table above. We carried out data collection for relevant features (details mentioned in Section 6.3) and applied the AHP methodology to our dataset, processing each data point individually. The resultant values were averaged to generate ground truth values between 0 and 1.0. To do this, we normalised the values of each feature and calculated a ‘Suitability Score’ (which was then normalised). For positively correlated features, the raw values were used. For negatively correlated features, the raw values were subtracted from 1. In case either of the matrices declared a location as very unsuitable, we excluded it from the averaging process and automatically assigned it the lower of the two scores. For coordinates with ground truth values as 1, we selected 1000 locations in India that already have solar farms.

Train-Test Split

We decided to split our dataset in a temporal manner. Splitting at 2017, we considered the earlier points in our training set and the remaining in our test. This would help us evaluate our model’s performance by testing if it could predict the solar farms that were set up later. We ended up with 5050 coordinates in our training set and 1236 coordinates in our test set out of which 400 and 278 were solar farm coordinates respectively.

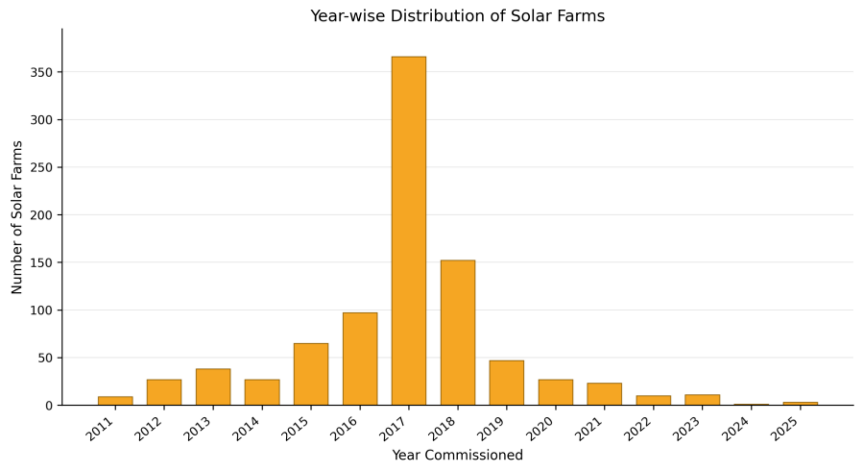

Figure 2: Year-wise distribution of solar farms. It illustrates the deployment trends of solar energy infrastructure across the country

According to this distribution, the dataset was split temporally with locations 2017 and prior forming the test set and those developed after constituting the test set.

Model Selection

We explored a range of architectures for our study: Support Vector Machines, K-Nearest Neighbours, XGBoost, Multilayer Perceptron, and Random Forest. We decided to go with using regression models instead of traditionally used classifiers because we observed that regressors can capture certain nuances that classifiers tend to overlook.

Performance Metrics

For our analysis, we decided to employ the Mean Squared Error, the Root Mean Squared Error and the R^2 Score. More information about these is presented in the Table 3.

Metric	Range	Formula
Mean Squared Error (MSE)	[0, ∞); Smaller values indicate fewer large prediction errors.	MSE = (1/n) Σ_(i=1)^n (y_i − ŷ_i)^2
Root Mean Squared Error (RMSE)	[0, ∞); Reflects average prediction error in actual units, lower is better.	RMSE = √[(1/n) Σ_(i=1)^n (y_i − ŷ_i)^2]
R^2 Score	(−∞, 1]; Higher values show better variance ex- planation by the model.	R^2 = 1 − [Σ_(i=1)^n (y_i − ŷ_i)^2 / Σ_(i=1)^n (y_i − ȳ)^2]

Table 3: Performance Metrics (Rashidi et al., 2023, p. 5)

Dataset

For the purpose of this paper, we chose to focus on geographic and terrain data. The features associated with the geographical and terrain data tend to be most influential and relevant when identifying suitable solar farm sites. Social, economic, and political factors can play a part, but their impact is comparatively less significant and remains outside the primary scope of this particular study.

Our resultant dataset had 23 features and 6286 coordinates. We extracted features from Earth Engine, OpenStreetMap, and Open-Meteo, and obtained solar farm coordinates along with their years in which they were commissioned from the Global Solar Power Tracker.

S.No.	Features	Source
1	Solar irradiance, Aerosols, Pollutants, Water Vapour, Albedo, Snow Cover, Wind	Bamisile et al., 2025
2	Sunshine Duration	Yildiz et al., 2025
3	UV Index	Aman et al., 2025
4	Dew Point	Sarmah et al., 2023
5	Snow Fall	Andrews et al., 2013
6	Carbon Dioxide	Patnaik et al., 2021

Table 4: Literature-supported features considered for solar farm site suitability

From initial AHP analysis, the features incorporated into our study were Solar Radiation, Temperature, Cloud Cover, Humidity, Elevation, Slope, Land Cover, Distance from Major Roads, Wind Speed, LST (Land Surface Temperature), Relative Humidity, Vegetation, LULC (Land Use and Land Cover) and Aspect. These features played a key role in assigning ground truth values to our coordinates.

Results

This section provides a detailed overview of the results obtained after the implementation of the following ML architectures: Random Forest (RF), Support Vector Machines (SVM), Multi- layer Perceptron (MLP), Extreme Gradient Boosting (XGBoost), and K-Nearest Neighbours (k-NN).

Model	MSE	RMSE	R^2 Error
k-Nearest Neighbours (k-NN)	0.0201	0.1417	0.6890
Support Vector Machine (SVM)	0.0231	0.1519	0.6421
Random Forest (RF)	0.0147	0.1212	0.7729
Extreme Gradient Boosting (XGBOOST	0.0135	0.1161	0.7915
Multilayer Perceptron (MLP)	0.0187	0.1367	0.7095

Table 5: Performance comparison of different regression models

As we can see, all the different model architectures that we tried, exhibited comparable and notably strong performance, underscoring the robustness of the dataset and the suitability of using ML approaches for identifying optimal solar farm locations. However, for further analysis, we chose XGBoost as it yielded the best results among them, indicating superior predictive accuracy and generalisation capability.

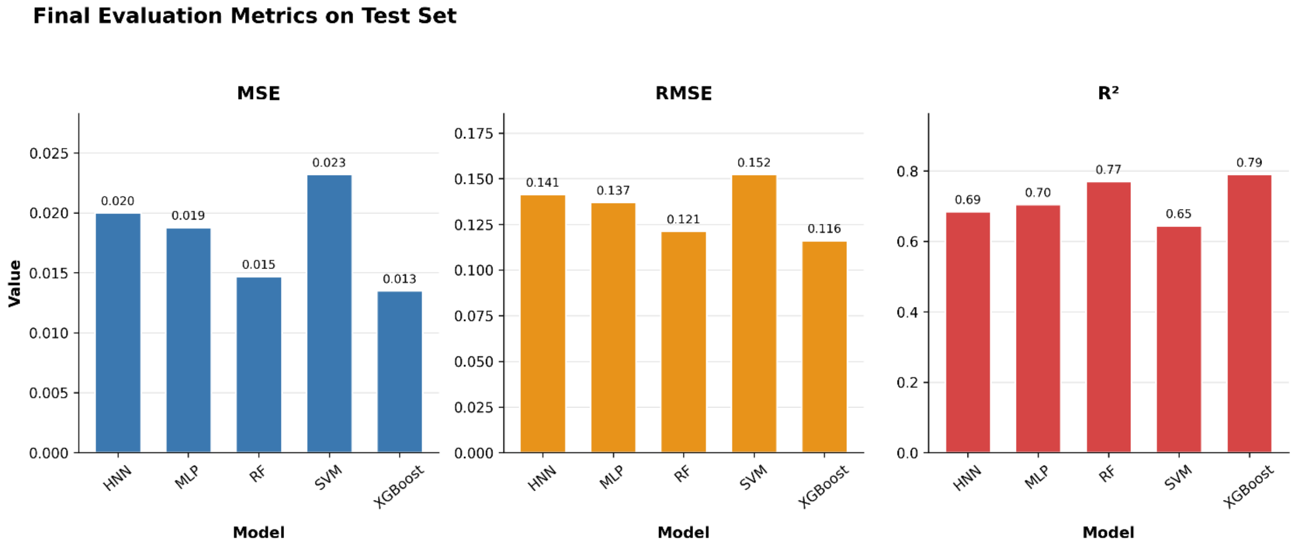

Figure 3: The figure compares the performance of all the tested machine learning models based on important regression metrics. Models appear to have comparable results indicating effective patters in chosen features and pre-processing steps

However, XGBoost was still found to have consistently outperformed its counterparts in both R^2 Score and Mean Squared Error, making its case to be considered the most reliable model for our pipeline. XGBoost was selected for further analysis with regards to final suitability predictions and validation.

SHAP Analysis

An important aspect of our methodology was the ability to easily interpret and explain the results of our model. We found SHAP (SHapley Additive exPlanations) to be an invaluable tool for this very purpose. SHAP is a method that can assign each feature an importance value for a particular prediction; more precisely, SHAP values attribute the change in the expected model prediction to each feature when conditioning on that feature [Lundberg and Lee, 2017, p. 1].

The following plot on the importance scores of each feature was obtained upon SHAP implementation:

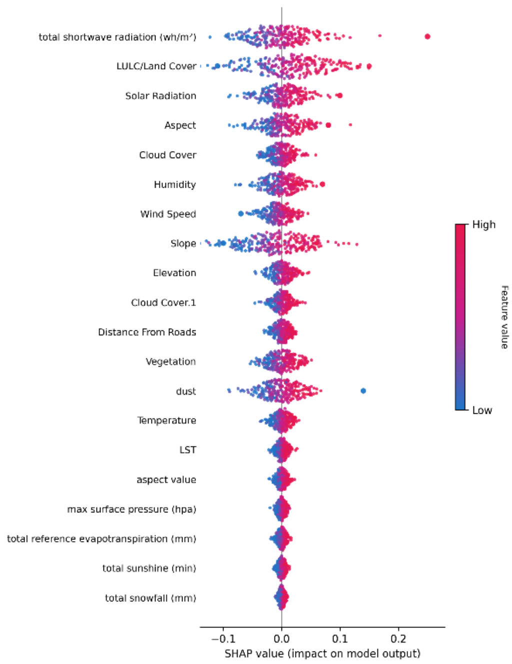

Figure 4: The SHAP summary waterfall plot presents the magnitude and direction of each feature's individual impact on the model's output. The SHAP value lined up under the horizontal axis indicates how much a feature contributes to increasing (positive) or decreasing

A wider spread implies a generally stronger influence on predictions for each range of actual values. The colour adds another layer of interpretation in the form of indicating low feature values (denoted by blue) and high feature values (denoted by red). For example, red dots for solar radiation clustered on the right suggest that high solar radiation values strongly increase suitability scores, on the other hand red values for cloud cover skewed left show that high cloudiness tend to suitability.

This waterfall plot highlights the influence of individual features on the model output, not just pointing out which variables most significantly impact solar suitability predictions, but also the manner in which they do so. The blue-red colour grading provides insight into how the absolute values of each feature for a given coordinate influence the calculations of the predicted scores for said coordinate. For instance, lower LULC values are associated with a positive contribution (an increase) to the predicted score, whereas higher values result in a pronounced decrease.

SHAP was instrumental in determining the causes behind significant discrepancies in the AHP based suitability scores and our model's predicted scores. Using a threshold of less than 0.5 to indicate unfavourable locations and greater than 0.8 to indicate favourable locations for solar farm installation, we found 20 unique coordinates where the two methods produced conflicting results. In all 20 instances, SHAP was used to explain the discrepancy behind AHP indicating suitability ('yes') but our model predicting unsuitability ('no'). A summarised report for one such coordinate is presented below:

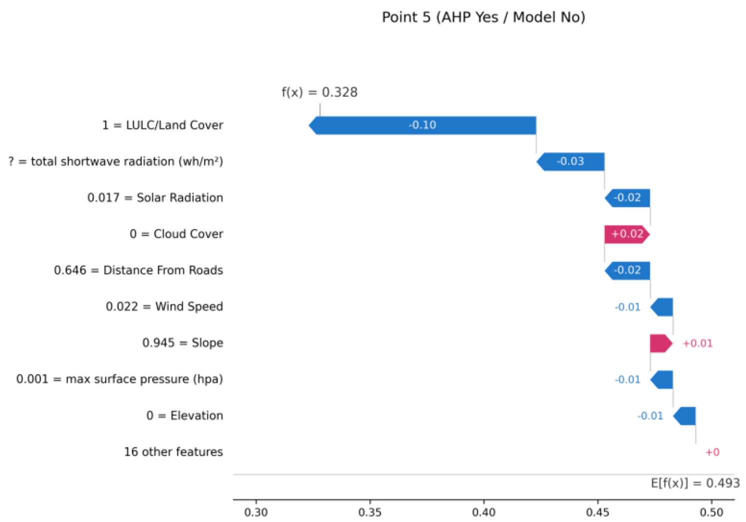

Figure 5: This is a local explanation for a single prediction made by the model. Horizontal axis shows the movement from the model's base value (mean prediction across all dat points) towards the final output. Blue bars represent features that pulled the prediction score down and red bars represent the features pushing it up

1. AHP Score: 1.000
2. Model Prediction: 0.326

Top contributing features(as per SHAP analysis on normalised values):

1. LULC / Land Cover = 1.0 → Decreased prediction by 0.104
2. Total Shortwave Radiation (Wh/m²) = 9.52×10^{-5} → Decreased prediction by 0.026
3. Solar Radiation = 0.0170 → Decreased prediction by 0.021
4. Cloud Cover = 1.66×10^{-5} →Increased prediction by 0.021
5. Distance from Roads = 0.6461 → Decreased prediction by 0.019

This SHAP analysis shows that despite an AHP score of 1.000, the model predicted low suitability (0.326) due to strong negative contributions from LULC (indicating unsuitable land type) , low shortwave radiation, and solar radiation. Minor positive effects from low cloud cover and wind speed were not enough to offset the negative impacts.

Validation of Data

A key focus of our research was to show how our results compared to the traditional GIS- MCDM methods in the domain. To achieve this, we assessed how accurately our model could predict existing solar farm locations in comparison to AHP-MCDM methods.

We ran AHP on all solar farm coordinates instead of directly assigning the Suitability Score as 1, like we did during model training. We then compared the AHP Suitability Scores with the model predictions and compared both methods.

Our results our summarised in the figure below:

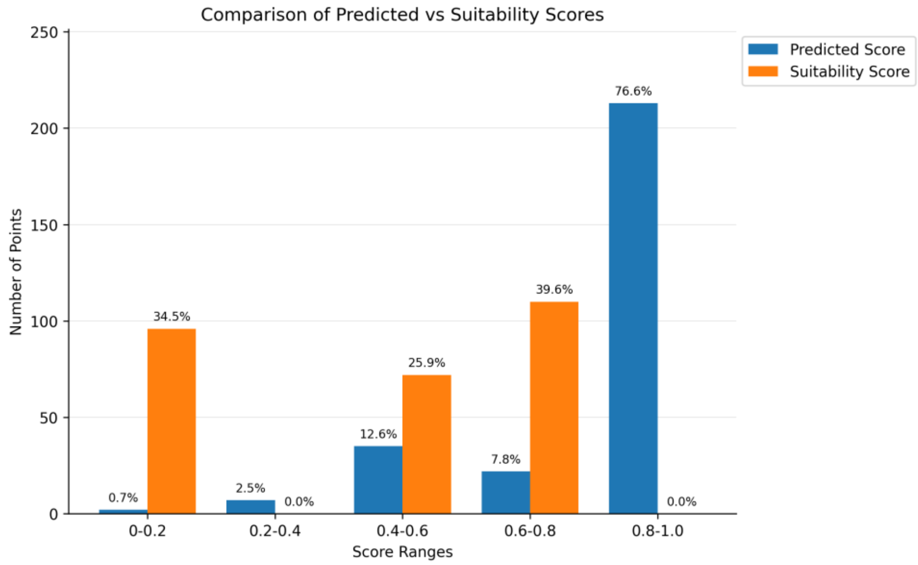

Figure 6: Distribution of predictions of model and AHP for existing solar farm locations. The plot clearly reveals a substantial improvement displayed by our model in more reliably predicting an existing solar farm site as suitable

As observable from Figure 8, 76.6% of solar farm coordinates were predicted to have a score of 0.8-1.0 by our model (Very Highly Suitable), in comparison to null points categorised by the AHP in this bracket. It is also notable that AHP categorised 34.5% of the solar farms in the 0-0.2 bracket (Very Highly Unsuitable), effectively categorising these points as not suitable. It is evident that our model has a higher rate of correctly classifying locations for solar farm optimality than AHP.

Limitations

While our model yielded favorable results, outperforming AHP in the assessment of solar farm suitability, we acknowledge several limitations. Firstly, our model works on the assumption that existing solar farms are placed in optimal locations. We believe that this is a justifiable assumption due to the rigorous planning and substantial capital involved in establishing solar farm.

Another considerable limitation lies in our method of obtaining ground truth values using AHP. While AHP is a well-established method, it is inherently subjective and may introduce bias, despite our use of two separate AHP pair wise matrices to obtain the suitability score.

Additionally, due to AHP scores influencing the target labels, there is an ambiguity in our model's "misclassifications". When the model's prediction deviates from the target label, it is unclear whether the model is making an error or correctly identifying a flaw in the AHP-based label. Although such misclassified points made up an insignificant portion of our dataset, (20 out of 1236 points from the test set) this ambiguity underlies the need for further validation by domain experts.

Our model focused on geographical and meteorological features for the purpose of this paper, as these proved to be the most influential factors in determining suitable locations for solar farm installation. However, based on insights from existing studies, we recognise areas for further research and experimentation by incorporating more social, economic, and political features as we have observed that such features can also influence solar farm constructions.

Conclusion

Solar energy has immense potential to revolutionise the energy sector. At a time when looking at alternatives to fossil fuel-based energy sources has become highly necessary, facilitating large-scale adoption and expanding accessibility of solar power through the installation of solar farms has become equally crucial.

Our study has shown that AI-ML based methodologies are a viable alternative to traditional GIS based approaches for this purpose, if not better. Our models were trained on a robust dataset of 6286 coordinates across India, taking 23 geographical features into account. We split our dataset into train and test sets around the year 2017. Five different regressor methods: Random Forest, k-Nearest Neighbours, Multi-Layer Perceptron, Support Vector Machines and Extreme Gradient Boosting, were carefully tuned and evaluated. Among these, XGBoost achieved the best performance, with an RMSE of 0.1161 and an R^2 score of 0.7915.

The results were comprehensively explained and interpreted with the help of SHAP, and a direct comparison with AHP calculations revealed that our dynamic AI-based approach is more accurate in correctly identifying existing solar farm locations as suitable. In a broader context, the convincing results of our model and its convenient, scalable and dynamic nature, makes our study an important step towards addressing the limitations of AHP-GIS in this domain. Developments in this field that build upon our findings should help streamline the solar site selection process and boost the development of solar infrastructure.

References

1. Ahmadi, H., Morshedi, J., & Azimi, F. (2016). Site selection of solar power plant using Geospatial Information System and climatic data (Case study: Ilam province). Journal of RS and GIS for Natural Resources, 7. https://sanad.iau.ir/en/Article/901911
2. Al Garni, H. Z., & Awasthi, A. (2017). Solar PV power plant site selection using a GIS-AHP based approach with application in Saudi Arabia. Applied Energy, 206, 1225–1240. https://doi.org/10.1016/j.apenergy.2017.10.024
3. Al-Dousari, A., Ahmed, M., Al-Dousari, N., William, T., & Al-Anzi, A. (2021). Spatial management for solar and wind energy in Kuwait. Turkish Journal of Com- puter and Mathematics Education, 12(3). https://www.researchgate.net/publication/ 350808128 Spatial Management for Solar and Wind Energy in Kuwait
4. Genc¸er, Y., & Er, Z. (2022, September). Solar energy forecasting using ma- chine learning models. In Proceedings of the International Conference on Advanced Materials Science & Engineering and High Tech Device Applications. Ankara, Turkey. https://www.researchgate.net/publication/378333569 https://www.researchgate. Net/publication/378333569
5. Ara´n Carrio´n, J., Esp´ın Estrella, A., Aznar Dols, F., Zamorano Toro, M., Rodr´ıguez, M., & Ramos Ridao, A. (2008). Environmental decision-support systems for evaluating the carrying capacity of land areas: Optimal site selection for grid-connected photovoltaic power plants. Renewable and Sustainable Energy Reviews, 12(9), 2358–2380. https://doi.org/10.1016/j.rser.2007.06.011
6. Bamisile, O., Acen, C., Cai, D., Huang, Q., & Staffell, I. (2025). The environmental factors affecting solar photovoltaic output. Renewable and Sustainable Energy Reviews, 208, 115073. https://doi.org/10.1016/j.rser.2024.115073
7. Carver, S. J. (1991). Integrating multi-criteria evaluation with geographical information systems. International Journal of Geographical Information Systems, 5(3), 321–339. https://doi.org/10.1080/02693799108927858
8. Chen, C.-R., Huang, C.-C., & Tsuei, H.-J. (2014). A hybrid MCDM model for improving GISbased solar farms site selection. The Scientific World Journal, 2014, Article 925370. https://doi.org/10.1155/2014/925370
9. De Silva, M. H. R. M. (2024). Suitability mapping of solar power plants using an explain- able AI-based approach (Master's thesis). NOVA Information Management School, Universidad Nova de Lisboa. https://run.unl.pt/bitstream/10362/165508/1/TGEO293.pdf
10. Enphase Energy. (n.d.). How weather conditions impact solar energy production. Retrieved March 10, 2025, from https://enphase.com/en-in/blog/homeowners/ weather-conditions-impact-solar-energy-production
11. Garc´ıa, S., Ram´ırez-Gallego, S., Luengo, J., Ben´ıtez, J. M., & Herrera, F. (2016). Multilayer perceptron: Architecture optimization and training. International Journal of Artificial Intelligence and Interactive Multimedia, 4(1), 36–44. https://www.ijimai.org/ journal/sites/default/files/files/2016/02/ijimai20164 1 5 pdf 30533.pdf
12. International Energy Agency. (2024). World energy outlook 2024. https://www.iea.org/reports/world-energy-outlook-2024
13. Boateng, E. Y., Otoo, J., & Abaye, D. A. (2020). Basic tenets of classification algorithms: Knearest-neighbor, support vector machine, random forest and neural network: A review.

Journal of Data Analysis and Information Processing, 8, 341–357.https://www.scirp.org/pdf/jdaip 2020112011230954.pdf

14. Rane, N. L., Gu¨ nen, M. A., Mallick, S. K., Rane, J., Pande, C. B., Giduturi, M., Bhutto, J. K., Yadav, K. K., Tolche, A. D., & Alreshidi, M. A. (2024). GIS-based multi- influencing factor (MIF) application for optimal site selection of solar photovoltaic power plant in Nashik, India. Environmental Sciences Europe, 36(5).https://doi.org/10.1186/ s12302-023-00832-230954.pdf
15. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpret- ing model predictions. In Advances in Neural Information Processing Systems (NIPS 2017), 30, 4765–4774. https://proceedings.neurips.cc/paper files/paper/2017/ file/8a20a8621978632d76c43dfd28b67767- Paper.pdf
16. Mierzwiak, M., & Ca lka, B. (2017). Multi-criteria analysis for solar farm location suitability. Reports on Geodesy and Geoinformatics, 104(1), 20–32. https://doi.org/10. 1515/rgg-2017-0012
17. Emmanuel, M. (2025). Understanding the Gradient Boosting Algorithm in XGBoost. Unpublished manuscript, Ladoke Akintola University of Technology. https://www.researchgate.net/publication/390137877 Understanding the Gradient Boosting Algorithm in XGBoost
18. Onuoha, H., Denwigwe, I., Babatunde, O., Abdulsalam, K. A., Adebisi, J., Emezirin- wune, M., Okharedia, T., Akindayomi, A., Adisa, K., & Hamam, Y. (2025). Integrating GIS and AHP for photovoltaic farm site selection: A case study of Ikorodu, Nigeria. Processes, 13(1), 164. https://doi.org/10.3390/pr13010164
19. Panagoda, C. (2023). Solar farm design and optimization. In Renewable energy engi- neering (pp. 30–40). Springer.
20. Rudro, R. A. M., Nahar, A., & Al Sohan, M. F. A. (2025). SPXAI: Solar power gener- ation with explainable AI technology. Energy Proceedings, 51. American International University–Bangladesh.
21. Sachit, M. S., Shafri, H. Z. M., Abdullah, A. F., Rafie, A. S. M., & Gibril, M. B. A. (2022). Global spatial suitability mapping of wind and solar systems using an explainable AI-based approach. ISPRS International Journal of Geo-Information, 11(8). https://doi. Org/10.3390/ijgi11080422
22. Singh, D. R., Kumar, P., Kar, A., Jha, G. K., & Kumar, A. (2017). Solar energy use in agriculture for enhancing farmers' income: A case of solar tubewell in North-Western Rajasthan. Agricultural Economics Research Review, 30(Conference Number), 269–277. https://doi.org/10.5958/0974-0279.2017.00040.4
23. Steffen, B., Beuse, M., Tautorat, P., & Schmidt, T. S. (2020). Experience curves for operations and maintenance costs of renewable energy technologies. Joule, 4(2), 359–375. https://doi.org/10.1016/j.joule.2019.11.012
24. Sun, Y. (2023). Spatial modelling the location choice of large-scale solar pho- tovoltaic power plants: Application of interpretable machine learning techniques and the national inventory. Energy Conversion and Management, 289, 117198. https://doi.org/10.1016/j.enconman.2023.117198
25. Suprova, N. T., Zidan, M. A. R., & Rashid, A. R. M. H. (2020). Optimal site selection for solar farms using GIS and AHP: A literature review. In Proceedings of the Interna- tional Mechanical and Industrial Engineering and Operations Management (IMEOM) Conference.

https://www.ieomsociety.org/imeom/253.pdf

26. Taherdoost, H. (2017). Decision making using the Analytic Hierarchy Process (AHP): A stepby-step approach. International Journal of Economics and Management Systems, 2, 244–253. https://www.researchgate.net/publication/322887394 Decision Making Using the Analytic Hierarchy Process AHP A Step by Step Approach

27. Taherdoost, H., & Madanchian, M. (2023). Multi-criteria decision making (MCDM) methods and concepts. Encyclopedia, 3(1), 77–87. https://doi.org/10.3390/ encyclopedia3010006

28. Upadhyay, A., & Chowdhury, A. (2014). Solar energy fundamentals and challenges in Indian restructured power sector. International Journal of Scientific and Research Publications, 4(10), 1–5. https://www.ijsrp.org/research-paper-1014/ijsrp-p3535.pdf

29. U.S. Department of Energy. (n.d.). Literature review: Effects of UV exposure on PV modules. Office of Energy Efficiency & Renewable Energy. Retrieved March 10, 2025, from https://www.energy.gov/eere/solar/articles/ literature-review-effects-uv-exposure-pv-modules

30. Worldometer. (2025). Current world population. https://www.worldometers.info/ worldpopulation/

31. Xu, Y., Yu, J., Zhang, R., Li, K. (2023). Common statistical concepts in the supervised Machine Learning arena. Frontiers in Oncology, 13, 1130229. https://doi.org/10.3389/ fonc.2023.1130229

32. Yadaw, P. K., & Umar, M. M. (2024). Unleashing India's solar potential: A review of the National Solar Mission and the path to sustainable energy in India. Journal of Technology Innovations and Energy, 3(2), 49–63. https://doi.org/10.56556/jtie.v3i2.930