Strategic Solar Deployment in India: Leveraging Machine Learning for Site Localization

Bunyad Chaudhary,
Dept. of Information Technology,
KJ. Somaiya College of Engineering, Somaiya Vidyavihar University, Mumbai, India
bunyad.c@somaiya.edu

Nilkamal More,
Dept. of Information Technology,
KJ. Somaiya College of Engineering, Somaiya Vidyavihar University, Mumbai, India
neelkamalsurve@somaiya.edu

Suchitra Patil,
Dept. of Information Technology,
KJ. Somaiya College of Engineering, Somaiya Vidyavihar University, Mumbai, India
suchitrapatil@somaiya.edu

Pankaj Mishra
Dept. of Information Technology,
KJ. Somaiya College of Engineering, Somaiya Vidyavihar University, Mumbai, India
pankaj.mishra@somaiya.edu

***Abstract*—Accurate and scalable site selection is critical for maximizing the impact of utility-scale solar projects in India. Existing approaches—ranging from GIS-based AHP and MCDM to opaque machine-learning models—often rely on subjective weightings, ignore key demand-side factors, or sacrifice transparency for performance. To overcome these limitations, we introduce a fully data-driven framework using a custom dataset compiled across all Indian states and union territories over a five-year period. This dataset integrates five complementary criteria—solar irradiance, a composite solar potential index, regional energy demand, population distribution, and prevailing wind conditions. An autoencoder extracts robust latent features from this high-dimensional data, reducing overfitting and capturing subtle spatial-temporal interactions that traditional PCA or manual indices miss. We then apply elbow-method clustering to establish objective, statistically derived thresholds and employ a transparent predicate-count ranking to categorize regions into most, moderate, and least favorable classes with clear explanations for each placement. When benchmarked against conventional GIS-AHP and black-box ML approaches, our method demonstrates improved alignment with national deployment targets while offering interpretable insights for policymakers. The analysis highlights Rajasthan, Gujarat, and Maharashtra as top candidates and demonstrates that this modular framework can be readily adapted to other regions or renewable technologies, providing a powerful blueprint for data-driven energy planning.**

***Keywords*—Solar energy, Machine learning, autoencoder, clustering algorithms,**

I. Introduction

Aligning with India vision, renewable resources, especially Solar energy, also serves as a promising solution for pollution mitigation, as India tries to establish a sustainable and stable energy system. Urbanization and climate change are driving rapid growth in demand for electrical power across the nation, making the effective harnessing of solar energy essential as illustrated via machine learning-based solar site selection using SVMs in Turkey [1]. But in India, the diverse terrains and climatic conditions of the country make detection of ideal sites for solar panel means difficult.This strong emphasis on technical feasibility leads to disregard for important components — like infrastructure, economics, and environmental issues — that are crucial for lasting success. The current study seeks to overcome this barrier by developing a large-scale, comprehensive data-driven method to score and rank Indian states based on their suitability for large-scale solar power projects. The goal is to balance local energy demand, make optimal use of resources, and conform to environmental limits.

Many recent studies of all-around choice of site for solar energy systems utilize machine learning and similar methods, often using multi-criteria decision making (MCDM). GIS and Analytic Hierarchy Processes (AHP) are examples of these tools used in some studies to integrate spatial and climatic data for instance use of AHP based process to evaluate potential sites for solar farms in Maharashtra [2]. Yet, these approaches tend to obfuscate the potency of decisions, because decision-makers depend on complex algorithms, and thus it is hard for users to explain site favoritism as explained with the help of GIS and MCDM-based approaches in Algeria for finding solar sites from both technical and economic point of view [3]). And although clustering algorithms are adept at classifying similar regions, they rarely explain how particular factors affect the suitability of a site for solar energy. Recent developments highlight what is important including solar irradiance, population density, and energy consumption as key factors. However, a systematic way to evaluate the accuracy of AI with a level of statistical rigor that can also be communicated meaningfully to users is still in its infancy in spite of some excellent advances like LSTM based approach for forecasting outperforms other models in colocated solar and wind plants [4]. In this study we addressed these problems by leveraging a predicate-driven approach. This approach relies on statistical methods to set thresholds for determining what is an important area and is therefore more quantitative and objective in nature. This approach allows the acquired results to be evidence-based and practical, allowing for informed decisions to be made about the site selection of solar energy.

This study looks at a new method to judge locations for solar power projects. It combines specific range-based criteria with a close look at traits to see how good each place is for solar. The study considers five important things: global horizontal irradiance (GHI), population size, energy consumption, solar potential along with wind speed. These things are checked at low, medium as well as high amounts. The study uses real statistics from the area, so the study shows the different places and people in India correctly. For example, states above 824.6 W/m^2 for GHI and over 94,874.99 MW for solar potential work better for big solar projects. But states in the middle ranges fit better for smaller, local solar projects. The system is better than older grouping methods. It uses rules to check how well each state fits the best conditions for solar. The main objective is to give policymakers understandable information about states appropriate for solar projects. It will tell them why some states are appropriate, for example, lots of solar energy, few problems with wind, or a good match with local energy needs.

The research problem centers on a lack of clear structure for guiding solar investments in India. Many methods assess technical elements separate from social plus economic settings or use complex formulas that are not easily understood. This investigation addresses problems by focusing on two central questions: First, what simple directions will help find good places for solar energy projects? The directions must be understandable and reliable. How does a fair system rate the states based on these directions? For these questions, the investigation examines data from twenty-nine states and eight union territories over five years. It looks at weather, population size along with energy use. The method uses processes to show concealed structures in the data, then grades the states according to measurable standards. As an example Rajasthan, possessing considerable global horizontal irradiance (GHI) plus reduced wind speed, becomes a good option for extensive solar projects. Bihar, having diminished solar capability, needs focused government actions to improve its solar energy power. Through the combination of full data examination and distinct directions, this strategy gives a repeatable design for solar energy planning. This design can grow and change for different situations.

The research offers a notable change to how renewable energy is planned. It replaces reliance on complicated computer programs with a simple, organized method for evaluating renewable energy choices. This method helps decision-makers, companies along with other involved parties make good choices concerning resource allocation as shown in Saudi Arabia for planning solar PV projects with the help of multi-layer GIS-AHP models [5] . The research identifies states best for solar energy and gives customized plans for each area's specific needs. For example the research suggests big solar facilities in high-sunlight areas, or solar panels on roofs in crowded areas with little open ground. Through this action, the study backs India's solar energy targets by finding a balance between energy needs and environmental protection. It acts as a good example of using data to make decisions about renewable energy growth around the planet.

II. LITERATURE REVIEW

Finding good places to put solar power plants helps get the most from renewable energy setups. A lot of work examines items that change how much solar energy a place produces, such as the site's location, typical weather along with people's way of life. Machine learning plus deep learning methods applied to renewable energy showed potential for making site choices more precise. For instance studies used grouping methods and ways to make datasets smaller to look at big collections of information. These methods found patterns as well as the patterns then help in plans for infrastructure.

Methods to enhance solar power installation to satisfy increasing energy needs. The search for optimal solar power plant locations requires research because it combines multiple methods and technologies to achieve peak performance while minimizing environmental harm. Research shows machine learning effectiveness in this area through the use of support vector machines (SVMs) to select optimal solar power plant sites in Adıyaman province, Turkey by analyzing geographical and climatic characteristics [1]. A study from the Far North of Russia evaluated Solar Radiation Concentration Systems (SRCS) equipped with Fresnel lenses and found that these systems increased electricity generation by 16–17% during clear weather conditions [6]. The Analytical Hierarchy Process (AHP) identifies Satara, Ahmednagar, Kolhapur, and Latur as the optimal locations for solar photovoltaic farms in Maharashtra, India by evaluating diverse geographic, climatic, and economic factors [2]. The research team in Sarajevo introduced a fuzzy logic-based algorithm to determine both the best power levels for PV plants and their connection locations within the distribution network while highlighting the importance of synchronizing PV plant power with network conditions [7].

The Double MCDM-GIS approach which combines Data Envelopment Analysis (DEA) and AHP within a Geographic Information System (GIS) framework assists New Zealand's disaster-prone regions like Hawkes Bay in evaluating potential solar sites based on environmental and catastrophe risk factors. A comparative evaluation examined Support Vector Regression (SVR), Random Forest (RF), and Long Short-Term Memory (LSTM) networks for solar power forecasting at a co-located wind and photovoltaic plant. The research demonstrates that LSTM networks deliver superior accuracy because they effectively capture long-term dependencies within the data [4]. By estimating ASP capacity based on load requirements and working dwells, a systematic method to dimension autonomous solar power plants (ASPs) at off-grid sites is established, and considers gaps in the literature [8]. To identify the potential locations for solar PV systems, an Algerian municipality employed a two-stage optimization model based on GIS and MCDM. This study exemplifies combining technical analysis and economic analysis for site matching [3]. With a keen focus on a solar power tower in Kuwait, this study tracks the effects of aerosols for drier biomass on Concentrated Solar Power (CSP) systems. Aerosols (dust, particulates, etc.) can reduce total solar irradiation significantly which will impact the efficacy of power production. To quantify the impacts, the researchers consider both satellite data with ground measurements, demonstrating the need for mitigating factors from environmental considerations to get the maximum performance and economic return from CSP systems in difficult to manage climatic conditions [9]. The authors published this article in 2020 which delivers a full review of India's solar energy policies. This initiative focuses on residential, commercial and industrial sectors which have not yet fully embraced solar power.To understand the reasons for the policies and their effects, it took a close look at incentives, the regulatory framework, and the financial systems that accompany solar power development. One clear message coming through—this is the good kind of clear message that we in the sustainable development business love to hear—is that solar power development is also development that contributes to achieving sustainable development goals. [10].

A recent study introduces a C++ tool meant to help figure out where best to set up renewable energy projects in Myanmar by tapping into the Analytic Hierarchy Process (AHP). It jumps right in, using factors like resource potential, local climate, and some socioeconomic quirks to pinpoint promising sites. In most cases, this tool acts as a solid decision aid that nudges sustainable growth and reinforces strategic energy planning in Myanmar's sector [11]. Switching gears a bit, another part of the work aims to boost data analysis methods for forecasting production at solar power plants. Data taken over 34 days – recorded every 15 minutes – is run through a mix of machine learning algorithms, such as support vector machines, neural nets, and various regression techniques. The findings, in most cases, show clear improvements in both operational efficiency and prediction accuracy, which really underlines how these methods might up the overall reliability of solar energy systems [12].Then there's an effort down in India where an AI dataset is being built from satellite snapshots to better predict outcomes and grasp the environmental and socioeconomic impacts of solar farm installations [13]. Over in Saudi Arabia, decisions for large-scale photovoltaic projects are getting a tech boost from Geographic Information System (GIS) technology. They mix in AHP to weigh factors like land availability and solar radiation by merging several GIS layers, offering a more nuanced picture for decision-makers [5].Advanced data analysis techniques, i.e., artificial neural networks and time series analysis, are used in smart grids to predict solar PV power generation, indicating the importance of big data and analytics to optimize solar PV systems [14].

Approach to designing uneven distribution networks with solar power using historical irradiance data analysis to counteract the problem of output uncertainty in solar photovoltaic panels. On the IEEE 13-bus system, it utilizes the application of load flow analysis using OPEN-DSS software to optimize the placement of solar PV modules, increasing voltage levels and decreasing line losses for better network performance [15]. This extensive study of Taiwan's energy industry utilizes two essential multi-criteria decision-making (MCDM) methods: the Analytic Hierarchy Process (AHP) and the use of Data Envelopment Analysis (DEA). It evaluates decision-making frameworks to achieve maximum utilization of energy management methods, with a focus placed heavily on the effectiveness of the distribution of resources and the implementation of legislation for Taiwan's sustainable development of the energy sector [16]. Since they can utilize reflected solar irradiance, SHJ bifacial photovoltaic cells in Qatar achieve an 11% greater energy output compared to monoracial cells [17]. According to a comparative study of the two developed under desert climate conditions. Installing solar powered power plants atop canals is a new method that has been suggested for Pakistan. It is capable of minimizing water evaporation as well as the costs of transmission, using the available land in an efficient manner [18]. In the case of Russia, a study of the impact of pollution in the environment on the performance of solar panels revealed that pollutants considerably decrease the efficiency, with the implication that the region lacks the financial capacity as well as suitable sites to establish solar power power stations [19]

Studies explore novel solar deployment strategies, such as canal-top plants, while addressing pollution and environmental factors.

III. Methodology

The overall intention of the research is to secure maximum efficiency in the generation of energy by designing a holistic system to determine the most appropriate Indian states for the establishment of solar power plants. In doing so, five significant factors are considered; wind speed, solar potential, population density, sun irradiation, and energy demand. Additionally, to classify the states based on how suitable they are for the installation of solar power plants, each of their characteristics is thoroughly studied. This provides a holistic and accurate study capable of informing strategic investment ideas.

In the current study, a specially designed multi year dataset of five years(2019 to 2023) has been used to evaluate based on the official weather data. Specifically, the dataset includes data for eight union territories and twenty-nine states of India. The large dataset with nine columns and 37 rows includes a vast range of environment and meteorological parameters, including zones, population, wind speed (m/s), GHI (W/m2), temperature (degree Celsius), solar potential (MW), and energy use (per capita in kWh). The dataset provides the fundamentals of the data analysis and clustering methodologies that are outlined in the subsequent subsections.

The data employed in the current study is pre-processed taking into account several essential measures in order to maintain integrity. To start with, all the entries in the 'States' column that contained missing values were dropped. This is a necessary step since incomplete information can heavily undermine the accuracy of the rank and clustering algorithms. In the current study, five essential parameters are considered: solar irradiance, population density, energy demand, solar potential, and wind speed. The quantity of solar radiation, i.e., the available potential solar energy, is represented by the quantity of the volume of solar radiance, measured in watts per square meter (W/m2) or GHI. Population data helps in better comprehension of the demand-side of the use of energy. The same is the case with solar potential, given in megawatts (MW), that signifies the maximum possible capacity of solar power generation. The effect of wind speed in increasing the efficiency of solar panels is accounted for, with measurements in the unit of meters per second (m/s).

A. Defining predicates and ranges

Predicates were constructed for each parameter, classifying the data into low, moderate, and high ranges to analyze the data efficiently, as shown in Table 1. The statistical distribution of the data was used to define these ranges, guaranteeing that the classification is both useful and applicable,

TABLE I
PREDICATES AND ITS RANGE

Parameter	Low Range	Moderate Range	High Range
Sun Irradiation (W/m^2)	< 642.2	642.2 - 824.6	> 824.6
Population	< 78,608,333.33	78,608,333.33 - 157,147,666.66	> 157,147,666.66
Energy Consumption (kWh)	< 66,821,300,000	66,821,300,000 - 133,643,000,000	> 133,643,000,000
Solar Potential (MW)	< 47,439.98	47,439.98 - 94,874.99	> 94,874.99
Wind Speed (m/s)	< 3.76	3.76 - 5.38	> 5.38

B. Data Normalization and Autoencoder model Data preprocessing

StandardScaler was used to equalize the chosen parameters to guarantee that each feature contributes equally to the clustering process. To remove any bias resulting from variations in the parameter scales, normalization is an essential step in the preprocessing of data. It converts the data to have a mean of zero and a standard deviation of one. Because an autoencoder model can learn compressed representations of the data, it was selected. An autoencoder—a widely used unsupervised deep learning technique—extracts robust latent features from this high-dimensional data, reducing overfitting and capturing subtle spatial-temporal interactions. This approach, in contrast to traditional dimensionality reduction methods like PCA, reflects one of several machine learning perspectives that balance interpretability and performance.The autoencoder is composed of an input layer, a two- dimensional latent space representation, two hidden layers for encoding, and two hidden levels for decoding. Because of this architecture, the model can reduce the number of dimensions in the data while capturing its most important aspects. The model was trained for 50 epochs using the SGD optimizer, which is renowned for its effectiveness when working with big datasets. The data were then represented in latent space using the encoder component of the autoencoder. To manage any missing values, these representations were imputed using the SimpleImputer, guaranteeing that data gaps would not have an impact on the analysis.

C. Clustering with K means

The ideal number of clusters was found using the elbow approach. K-Means clustering is chosen for its simplicity and scalability to establish objective, statistically derived thresholds. While other clustering methods (e.g., hierarchical or DBSCAN) offer different strengths, K-Means aligns well with our structured, high-dimensional input and supports transparent, policy-driven interpretation With the help of this method, the within-cluster sum of squares (WCSS) is plotted against the number of clusters to find out the elbow point, or the point at which the WCSS decreases at a slower rate. The elbow approach was adopted after the silhouette score was first taken into account because it did not produce results with our dataset and model that were satisfactory. The elbow point was utilized to determine the ideal number of four clusters to use for K-Means clustering. We used K-Means clustering to cluster the states to classify them into groups that share similar characteristics and to identify trends in the data. The clusters indicate which were the best solar power plant states in terms of solar energy potential along with population density, energy demand and wind speed.

While the Elbow Method is a popular and computationally efficient approach for determining the optimal number of clusters, it is not without limitations. A major drawback is its reliance on visual inspection of the distortion score curve, which can be ambiguous when the curve lacks a pronounced "elbow" point. This subjectivity can lead to inconsistent cluster selection, especially in datasets with smooth or gradual decreases in inertia. Moreover, the method assumes well-separated, spherical clusters of roughly equal size—a condition that real-world geospatial and energy datasets may not always satisfy. These structural assumptions can reduce clustering accuracy in complex, high-dimensional feature spaces. Despite these challenges, the Elbow Method was chosen for its simplicity, interpretability, and compatibility with the unsupervised nature of the problem addressed in this study.

The Parallel Coordinate Plot offers a clear representation of how the different parameters of energy vary among the clusters. Energy consumption (kWh) is the most variable, reflecting its dominance in discriminating the clusters. Population is also considerably variable, showing that more populated states tend to consume more energy, an aspect that affects cluster formation. In contrast, wind speed is relatively uniform in all the clusters, implying that wind speed has a minor role in clustering. Also evident from the differences in solar potential (MW) and GHI (W/m^2) is that high solar potential states are

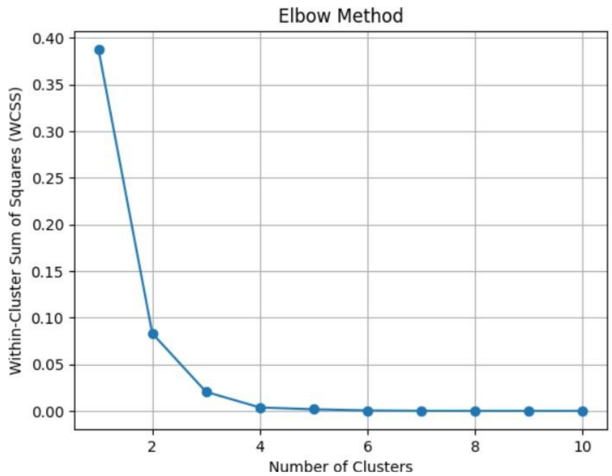

Fig. 1. Elbow method to find no of clusters

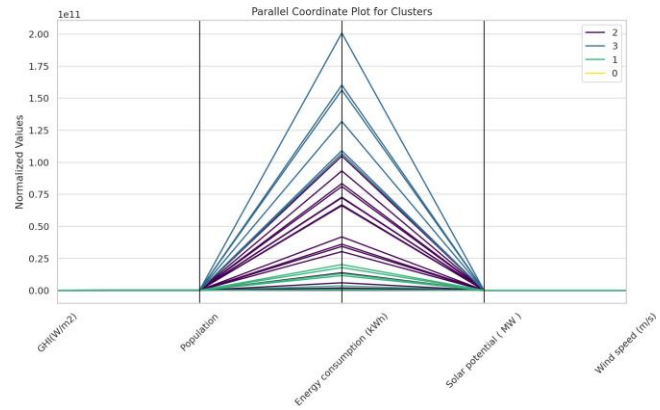

Fig. 2. Parallel Coordinate Plot for Clusters

clustered apart. The analysis points out that population density and energy consumption are important parameters in clustering the states for renewable energy potential and planning.

While clustering was not the sole basis for deriving final recommendations, it played a critical role in offering preliminary insights into data groupings. The K-Means algorithm, guided by the elbow method, helped identify underlying patterns among states based on five normalized parameters: GHI, solar potential, wind speed, energy consumption, and population. The clusters offered an overview of which states share similar energy and environmental characteristics. For example, certain clusters contained states with consistently high solar irradiance and potential, while others reflected more diverse or constrained conditions. To better understand the influence of each parameter in these clusters, a parallel coordinate plot and feature correlation heatmap were used. These visualizations indicated that energy consumption and population contributed significantly to differentiating the groups, while wind speed showed less variation. The clustering thus helped validate key assumptions but also revealed the need for a more granular, interpretable method. This motivated the use of statistically defined predicate ranges, which offered a more structured and transparent framework for evaluating state-level suitability for solar power installations.

D. Expected category ranges for suitability

In order to qualify as suitable for the installation of solar power plants, a state should fulfill the following criteria: high solar irradiance (GHI (W/m2)), a moderate to high population, moderate to high levels of energy consumption (kWh), high solar potential (MW), and low wind speed (m/s). These anticipated groups were established in light of the best circumstances for producing solar energy and the requirement to balance supply and demand for energy. Significant energy demand is indicated by moderate to high population density and energy consumption, while high sun irradiation guarantees maximum energy generation. Low wind speed reduces the influence on solar panel efficiency, whereas high solar potential indicates the potential for solar power generation.

E. Evaluating and ranking states based on clusters formed

To determine if the parameter values of each state and union territory were within the anticipated ranges, a thorough evaluation of each was conducted. Every state was thoroughly compared across the predicates that were defined, and the number of predicates that each state satisfied was tallied. Their ranking clearly indicated states' suitability for solar power plant installations based on the total number of satisfied predicates. By assigning resources to the states with the greatest potential for producing solar energy, this ranking aids in setting priorities for development and investment.

The Feature Correlation Heatmap identifies important connections between different energy-related factors. It shows a strong link (+0.79) between Population and Energy Consumption. This means that states with more people generally use more energy. There is also a moderate link (+0.45) between Energy Consumption and Solar Potential. This suggests that places with higher energy use often have a better chance for solar power. When we look at Solar Potential and Wind Speed, their connection is weak (+0.23). This means that having higher solar potential does not always mean having stronger winds. The link between GHI (Global Horizontal Irradiance) and Wind Speed is modest (+0.36). This means areas with more sunlight could possibly see slightly stronger winds. Another point is the weak negative link (-0.18) between Population and Wind Speed. This indicates that densely populated areas generally don't have great wind resources. Lastly, there is no significant link (-0.10) between Wind Speed and Energy Consumption. This implies that wind availability doesn't directly affect how much energy a state uses.

IV. Results

The study conducted a rigorous evaluation of all states and union territories of India, using five basic parameters: Global Horizontal Irradiance (GHI), solar potential, wind speed, energy demand patterns, and population density indicators. Every

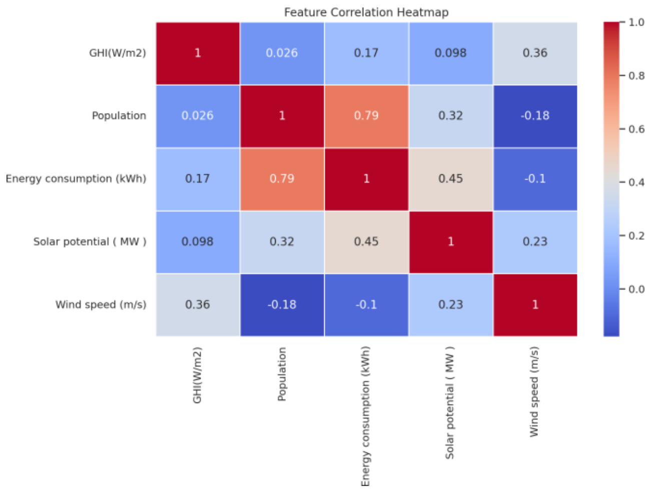

Fig. 3. Feature Correlation Heatmap between Variables

state was analyzed using a predicate-based analysis approach, under which each of the parameters was graded into low, moderate, or high level of suitability. The final grade of a state was determined based on the number of satisfactory predicate conditions met by the state, with higher-ranking states being ranked as more suitable for the installation of solar power units. The grading system was carefully designed to combine both supply-side and demand-side factors: states with high irradiance and potential got higher marks for availability of the resource, and moderate to high population densities and energy demands had a positive effect on demand feasibility. Additionally, low wind speed, because of its role in minimizing soiling loss for solar panels, was positively surveyed. The outcomes were divided into six categories, depending on how many predicates each state met.

TABLE II
STATE CATEGORIZATION BASED ON PREDICATE SATISFACTION

Categories	Number of Predicates Satisfied
Most Favourable	5
Highly Favourable	4
Moderately Favourable	3
Favourable	2
Less Favourable	1
Least Favourable	0

A. *Most Favourable States*

Rajasthan, Gujarat, and Maharashtra emerged as highly promising candidates for large-scale solar deployment, given their favorable alignment with all five key parameters. These states exhibit a strategic balance of high solar irradiance(more than 824.6 W/m^2) and solar potential(more than 824.6 W/m^2), alongside significant energy demand and manageable wind conditions(less than 3.76 m/s). While these attributes suggest strong infrastructural feasibility, regional policy support, land availability, and transmission infrastructure will further influence the practical success of solar initiatives.”

B. *Highly Favourable States*

The Karnataka, Tamil Nadu, Andhra Pradesh, and Telangana regions are most suitable for solar projects. These regions satisfy four of the five most important parameters, falling barely in the next best category for solar potential. These regions possess a proper balance in their energy portfolios, including a proper amount of sun (GHI of 642.2 – 824.6 W/m^2) and high potential of solar power (from 47,439.98 to 94,874.99 MW). Their high energy demand, as seen from the high use of energy, indicates high demand for electricity. The population of these regions is neither too high nor too low, indicating a proper number of people who can utilize the energy. One consideration in these states is occasional variation in wind speed(> 3.76 m/s), which may slightly influence solar panel maintenance needs due to potential dust accumulation. However, this factor does not significantly diminish their overall viability, especially given the robust solar potential and favorable energy demand profiles. Barring that, the proper sun, availability of infrastructure, and high demand for the same make the respective regions the most suitable for all types of solar systems. These range from rooftop solar panels to community-based solar projects, as well as blends of solar and wind power. These regions possess developed markets for energy as well as advanced renewable energy policies, increasing their potential as a suitable region for solar investments.

C. *Moderately Favourable States*

Madhya Pradesh, Uttar Pradesh, Odisha, and Chhattisgarh have all landed in the Moderately Favorable category, having met three out of five important criteria. These areas generally experience moderate solar irradiance and energy consumption, which means they have some resources, but not the best. While the population density tends to be acceptable, the solar potential is on the lower end (under 47,439.98 MW), and wind speeds can sometimes go beyond the ideal range, especially in inland or semi-arid regions. The varying infrastructure and climate conditions in these states present both hurdles and opportunities. While these regions may currently present moderate feasibility for utility-scale solar farms, their expanding rural electrification programs, varied topography, and available land suggest potential for scalable, decentralized solar solutions such as mini-grids and community-based systems. Long-term viability in these areas may depend on targeted policy incentives and infrastructure development. With targeted government support, capacity-building efforts, and collaboration with the private sector, these states could greatly enhance their capacity for solar energy adoption.

D. *Favourable and Less Favourable States*

Some states, such as Haryana, Punjab, Kerala, Jharkhand, West Bengal, and Bihar, are categorized as either "Favorable" or "Less Favorable" for solar energy, depending on how they meet certain criteria. These states encounter certain environmental and infrastructural constraints, such as relatively lower solar irradiance(below 642.2 W/m^2) and limited large-scale solar potential. As a result, achieving optimal returns on solar investments may require innovative deployment models, such as rooftop installations, microgrids, or targeted subsidies to mitigate these limitations. Bihar and West Bengal, for example, are densely populated with high energy requirements. However, their low Global Horizontal Irradiance (GHI) levels and high incidence of clouds make it difficult for solar panels to function efficiently. On the other hand, Kerala has moderate energy requirements but poor availability of open space and high humidity, both of them reducing the efficiency of solar panels. Despite that, the big cities in Kerala offer good scope to instal rooftop solar panels, battery storage, and intelligent microgrid systems. Well-planned subsidy schemes and tie-ups by the government with private firms can assist in countering these natural challenges to make solar power feasible.

E. Least Favourable States

Meghalaya, Nagaland, and Arunachal Pradesh are not suitable for large-scale solar projects yet. Current infrastructural and climatic conditions—such as frequent cloud cover, low solar irradiance, and limited grid connectivity—pose challenges to conventional large-scale solar deployment. Nonetheless, these regions offer unique opportunities for innovation through portable solar solutions, autonomous microgrids, and off-grid technologies, which can significantly improve energy access in underserved areas. However, the same challenges also create opportunities to benefit these regions in other ways. We can apply small and innovative solutions such as portable solar home power, solar lanterns, and autonomous microgrids suitable for villages not connected to the central power system. Cutting-edge technology such as satellite solar forecasting and self-adjusting weather-dependent energy systems can make a huge positive impact here. Even if the regions are unsuitable for large-scale solar power projects, such innovative ideas can deliver much-needed power to the region.

The map Below defines the Indian states by their potential for solar power plant locations, in a color-coded format. Deep red marks the Most Favorable states, and bright red marks the Highly Favorable ones, orange marks the Moderately Favorable, light orange marks the Favorable, the Less Favorable are marked by peach, and pale yellow marks the Least Favorable. The gradient provides a quick visual analysis of regional potential, an essential aid for making rational choices in renewable energy planning and investment decisions

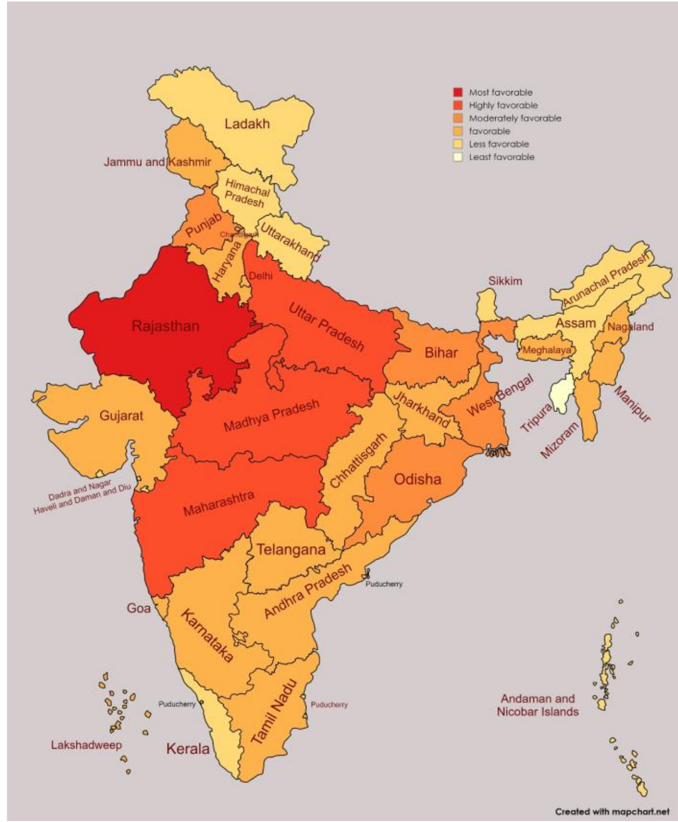

Fig. 4. Ranking of States

V. Conclusion

The study provides a wide-ranging, data-driven framework for the evaluation of Indian states as a potential place to install solar power plants based on five predicates: solar irradiance (GHI), population density, energy usage, solar potential, and wind speed. Through the systematic comparison of each state based on these measures, the study established a step-by-step classification system that declares regions from 'Most Favorable' to 'Least Favorable' for the installation of solar energy. The findings highlight that the states of Rajasthan, Gujarat, and Maharashtra, meeting all five predicates, offer ideal conditions for mega-solar projects due to their high solar irradiance, high solar potential, and high demand for energy. States of Karnataka, Tamil Nadu, Andhra Pradesh, and Telangana, meeting four predicates, also exhibit high potential, particularly for decentralized solar schemes. Regions with lower alignment to suitability predicates may face limitations for conventional large-scale solar deployment. However, their specific contexts may lend themselves well to adaptive solar solutions, including localized or hybrid models, which can still contribute meaningfully to energy equity and decentralized access. The stratified approach provides beneficial insight to policymakers, investors, and planners, assisting them to make informed decisions regarding the prioritization of regions for the establishment of solar infrastructure.The proposed framework, though developed for Indian states, is adaptable to other regions with diverse climatic and energy profiles. By adjusting parameter thresholds, it can support solar planning in countries such as Brazil, South Africa, or Indonesia, where decentralized and data-driven renewable energy strategies are increasingly critical. By aligning solar power projects with the unique strengths and weaknesses of each state, the approach promotes informed, context-sensitive decision-making by considering both environmental resource availability and regional demand factors. Recognizing that suitability is not uniform but rather shaped by a combination of technical, demographic, and infrastructural conditions, this framework supports adaptive planning strategies to advance India's solar energy goals in a sustainable and inclusive manner.

References

[1] B. Bostanci, T. Kaynak, and Z. Çapkurt, *Location selection for solar power plants by using support vector machines*, Adıyaman Province, Turkey," Renewable Energy, 2024.

[2] P. Jamodkar and A. Mitra, "Determination of Optimal Locations for Installation of Solar Photovoltaic Farm in Maharashtra, India," in *10th IEEE International Conference on Power Systems (ICPS)*, 2023, pp. 1–6.

[3] B. Settou, "Geographic information-driven two- stage optimization model for location decision of solar power plant: A case study of an Algerian municipality," *Sustainable Cities and Society*, vol. 77, pp. 103 567–103 567, 2022.

[4] A. G. Kavaz and A. Karazor, "Solar Power Forecasting by Machine Learning Methods in a Co-located Wind and Photovoltaic Plant," in *2023 12th International Conference on Power Science and Engineering (ICPSE)*, 2023, pp. 55–59.

[5] K. A. Alhamdan, *Locating Utility-Scale PV Plants in Power System: Case Study of Saudi Arabia*, Riyadh, Saudi Arabia, 2022.

[6] N. Mestnikov, A. Alzakkar, and Y. Samofalov, "The Impact of Introducing Solar Radiation Concentration System for Solar Power Plants Under Conditions of the Russian Far North," in *2023 International Russian Automation Conference (RusAutoCon)*, 2023, pp. 48–53.

[7] N. Dautbašić, F. Likić, A. Mujezinović, I. Turković, M. M. Dedović, and A. Alihodžić, "Selection of Location and Power of Photovoltaic Plant in Distribution Network using Fuzzy Logic," in *2023 XXIX International Conference on Information, Communication and Automation Technologies*. ICAT, 2023, pp. 1–7.

[8] R. Oksenych, O. Moroz, M. Qawaqzeh, I. Trunova, R. Buinyi, and S. Dudnikov, "Methodology for Designing the Capacity of Solar Power Plants for an Offline Home Network," in *2023 IEEE 5th International Conference on Modern Electrical and Energy System (MEES)*, 2023, pp. 1–5.

[9] M. S. Alfailakawi, S. Michailos, D. B. Ingham, K. J. Hughes, L. Ma, and M. Pourkashanian, "Multi-temporal resolution aerosols impacted techno-economic assessment of concentrated solar power in arid regions: Case study of solar power tower in Kuwait," *Sustainable Energy Technologies and Assessments*, vol. 52, pp. 102 324–102 324, 2022.

[10] A. Kharesaxena, S. Saxena, and K. Sudhakar, "Solar energy policy of India: An overview," *CSEE Journal of Power and Energy Systems*.

[11] Y. M. Thu, T. A. Shestopalova, M. G. Tyagunov, and H. Haiyang, "Program for the Selection of Sites to construct Renewable Energy Plants based on the Method of Analytic Hierarchy Process (AHP)," in *2022 VI International Conference on Information Technologies in Engineering Education (Inforino)*, 2022, pp. 1–5.

[12] M. B. Arab, M. Rekik, and L. Krichen, "Solar power plant data analysis and prediction using different techniques of machine learning," in *2022 IEEE 21st international Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, pp. 549–554.

[13] A. Ortiz, D. Negandhi, and S. R. Mysorekar, "An artificial intelligence dataset for solar energy locations in India," *Scientific Data*, vol. 9, no. 1, 2022.

[14] S. Jose and R. L. Itagi, "Data Analytics in Solar Photovoltaics Power Forecasting for Smart Grid Applications," *2021 International Conference on Intelligent Technologies (CONIT), Hubli*, pp. 1–5, 2021.

[15] S. Saijeet, S. R. Ghatak, and S. C. Swain, "Planning of Unbalanced Distribution Network Using Solar Energy Sources under Uncertain Environment," in *2021 1st International Conference on Power Electronics and Energy (ICPEE)*, 2021, pp. 1–5.

[16] C. -N Wang, N. -A. -T Nguyen, T. -T Dang, and J. Bayer, "A Two-Stage Multiple Criteria Decision Making for Site Selection of Solar Photovoltaic (PV) Power Plant: A Case Study in Taiwan," *IEEE Access*, vol. 9, pp. 75 509–75 525, 2021.

[17] A. A. Abdallah, A. Abotaleb, B. Figgis, M. M. Kivambe, B. Aïssa, and V. B. Benito, "Thermal Behavior and Photovoltaic Performance of Monofacial and Bifacial Silicon Heterojunction Modules Under Desert Conditions," in *2021 IEEE 48th Photovoltaic Specialists Conference (PVSC)*, 2021, pp. 843–0845.

[18] F. Javaid and Z. Islam, "Proposed Location and Proposal for Canal Top Solar PV Plant," in *2020 7th International Conference on Energy Efficiency and Agricultural Engineering (EE&AE)*, 2020, pp. 1–3.

[19] Y. Zatsarinnaya, A. Logacheva, and D. Amirov, "Contamination of Solar Panels as Factor in Selecting Location for Construction and Operation of Solar Power Plants in Russia," in *2019 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, 2019, pp. 1–5.