Uncovering Optimal Solar Site Locations in India using Unsupervised Learning Approaches

Smitesh Nitin Patil

June 29, 2023

School of Computer Science

University of Galway

Supervisor

Dr Karl Mason

In partial fulfillment of the requirements for the degree of

MSc in Computer Science (Artificial Intelligence)

Abstract

This project explores application of unsupervised learning approaches in context of GIS (Geograhpical Information System) data for recognising optimal locations for Solar PV (Photo-Voltaic) plants. The main focus of the project is to identify new techniques unique to the studies previously done by researchers in this field. Different techniques like Kohonens Model and Clustering techniques would be compared and evaluated including some self-supervised learning techniques like the Stack Auto-Encoder which have not been tried previously.

Keywords: Geospatial Information, Unsupervised Learning, Self-supervised Learning, Analytical-Hierarchical Process, Renewable energy, Site Selection, Spatial Analysis, Sustainability.

Contents

List of Figures

List of Abbreviations

1 Introduction

1.1 Motivation

The process of transitioning the global energy supply from fossil fuel-based sources to sustainable energy sources like wind and solar will be crucial for mankind to make in the 21st century. The incentive in deployment of these renewable sources is huge as this resouces are natural, free, available in abundance and replenishable. Solar energy is generated from photovoltaic cells which need a high amount of solar irradiance throughout the year to be profitable. Countries in tropical regions like parts of India tend to receive abundant sunlight throughout the year.

Regarding India, the energy demands are increasing rapidly, it is third largest producer of eletricity around the after the United States and China[5]. Currently, India's energy sector is dominated by fossil fuels with fossil fuels like coal fullfilling three quaters of countries energy demands. However, India is investing heavily in solar and hydropower projects as it has pledged to increase renewable energy generation sources to account for 50 percent of energy consumption by 2030 and reach net zero by 2070 during COP26 summit in 2021[6]. And it is evident that there has been a push in increasing solar power production, between 2017 and 2021, India's solar power production capacity more than tripled to rank third globally in terms of solar capacity[7].

Naturally, given the importance of the task, it is imperative to find new

locations to set-up renewable energy generation plants at a fact pace to make the transition from conventional sources of energy. Solar and Hydro-eletric power generation methods are been preferred in goverment of India's national energy policy for various reasons. India lies in between latitude 20.5937° N and 78.9629° E. Therefore, the country tends to receive quite a high amount of solar irradiance owing to its temperate and tropical climatic conditions.

However, many aspects need to be studied before identifying promising regions where solar farms could be built to tap into that region's solar potential, like the slope gradient of the terrain, proximity to urban centers, and nature, wildlife preserve areas. Scientific studies that have been done previously for installation of solar pv plants using GIS data preferred using Multi-Criteria Decision-Making Methods (MCDMs) to calculate the weight of these aspects[7, 1, 8, 9, 2]. In these studies Analytical Hierarchy Process (AHP) a MCDMs technique is used for determining the criteria. This study focuses on finding novel Unsupervised learnining techniques previously unused for this task, but carried out by other researchers like Chang et al for monitoring landslide susceptibility with Geospatial data[10]. Although, such studies have been carried in India previously by many authors they were limited in scope because of low resolution spatial data[11, 2, 12] and most of them use the aforementioned MCDMs for classification. The data used in this study is gathered from various sources like National Renewable Energy Laboratory (NREL) for solar irradiance, Digital Elevation Model (DEM) for topgraphical infomation generated by Shuttle Radar Topography Mission (SRTM) carried out by United States Geological Survey (USGS). Along

with OpenStreeMap (OSM) data for other important attributes like landuse, protected nature reserves, water bodies, urban centres, rail and road networks.

1.2 Purpose

Following points are the main objectives of this study:

- Create an original approach based on unsupervised learning and self-supervised learning methodologies to identify optimal geolocations for setting up solar pv plants.
- Previous studies for solar pv site selection in India were done with limited scope and limited data with low spatial resolution (>1000meters). This study aims to use datasets with better spatial resolution (<10meters to <30meters)
- To the best of student's knowledge, Unsupervised learning based classification for pv sites has not been done at such a scale.

1.3 Research Questions

1. Can Unsupervised learning and self-supervised learning techniques give better results that MCDMs with higher spatial resolution data.
2. Can these techniques complement AHP

2 Background and Related Work

2.1 Criteria and factors affecting the decision-making

Selecting appropriate features for our predicitive models is an important step for making informed decision on the feasibility of certain geolocation. There have been extensive studies that have been carried out by researchers in determining the factors important for classification of PV solar plant sites.

Colak et al carried out a study to identify suitable locations for setting up of photovoltaic power plants in Malatya province in Turkey[1]. The authors use 11 layers of GIS data for identifying suitable sites. These layers included factors that affect the decision making process like:

1. Solar energy potential: Identification the solar potential of a region is of utmost importance as it determines the amount of energy producted by the region if a photovoltaic power-plant is set there.

2. Slope: The slope of the terrain is an important factor as a solar pv plant requires a even terrain for setting up the PV panels

3. Transformer centers and energy transmission: Carrying electricity over huge distances without energy infrastructure results in loss of energy output through leakage thus there should be a power transmission system inplace for developing a new plant.

4. Land Cover: The land designated for nature reserves, tribal population, and other uses cannot be used for energy generation by law thus this factor must be considered beforehand.

5. Residential areas: Building a solar plant near a urban center can be detrimental for the region as urban sprawl tends to expand with time, On the other hand if a PV solar plant is close by a urban center it would lead to less transmission loss this tradeoff needs to considered.

For data preprocessing, various hardset conditions were set to restrict certain areas like slope elevation of land cannot be more than 20 percent, distance to road, rail network should be more than 0.1 km, no residential areas nearby and proximity to energy transmission network.

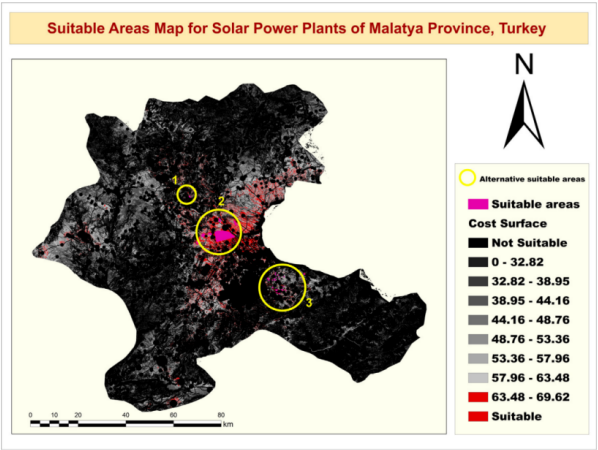


Figure 1: Suitable sites for Solar Powerplant with cost Factor

Figure 1: Suitable areas for solar power plants of Malatya province, Turkey by author Colak et at.[1]

Similar study was carried out by Al Garni et al in Saudi Arabia[8]. The available land land was categorised in five categories: least suitable, marginally suitable, moderately suitable, highly suitable, most suitable. The decision-making process for site selection was done in three phases.

1. Setting decision criteria and restriction for site selection study.

2. Prioritizing certain sites with high solar potential

3. Analysis on the prioritised region for decision making

Like the previous study, authors here have taking GIS data formulated by NREL and selecting attributes that determine criteria for site selection which include DEM, Solar irradiation, Air Temperature. These factors could be divided into roughly two categories technical (factors affectiong energy production) and economical (factors affecting economical viability of the project).

Zoghi et al proposed dividing factors in four major categories for their case study carried out in Ifshan province, Iran[9].

1. Enviromental: Landuse, Protected Areas, Wetlands and Water Resource.

2. Geomorphological: Elevation, Slope, Aspect.

3. Location: Distance to City, Distance to Power line, Distance to Transport network

4. Climatic: Sun shine, Cloudy Days, Dusty Days, Solar Radiation, Rainy and Snowy Days, Humidity.

Study carried out by Saraswat2021 et al is the most elaborate case study for site selection of solar pv plants in India to the best of student's knowledge[2]. A major limitation of the study lies in the data as the spatial resolution of data (around 1000m) is really less for DEM modelling and other attributes.

As a result the solar farms suitablity map generated in this study is not intricate on a spatial level. There are various databases available with spatial resolution of 30m from USGS and NREL that can be utilised to make more precise prediction.

Data was sourced from various govermental bodies like NREL for solar radiation, DIVA-GIS for roads, inland water bodies, DEM model used was provided by United States Geological Survey (USGS). The factors were divided into three categories technical, socio-enviromental and economical.

1. Technical: Solar Radiation, Slope, Aspect, elevation
2. Socio-Enviroment: Distance from coastline, Distance from waterbodies, airports, Landuse.
3. Economic: Distance from urban areas, roads, transmission lines, power plants

2.2 Data Gathering and Pre-Processing

For this project we need needed various layers of GIS Data that would form important features necessary for identification of suitable locations for solar farms. Terrain information is a really important feature for this study because for setting up a solar farm we require large tracts of land with little changes in elevation. Solar irradiance is another important feature to be considered, Solar irradiance is defined as the amount of energy that could be generated from solar radiation incident on that certain place it is measured in watts per meter square W/m2. Other important factors to consider are land cost, population density, land use, protected wildlife sanctuaries, etc

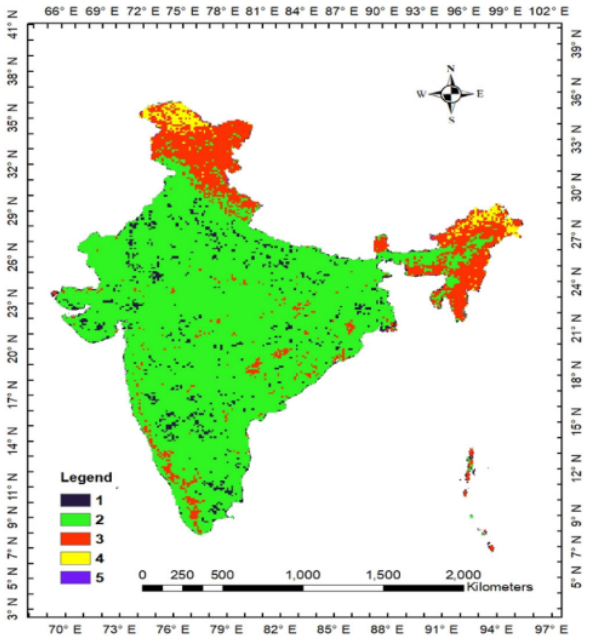
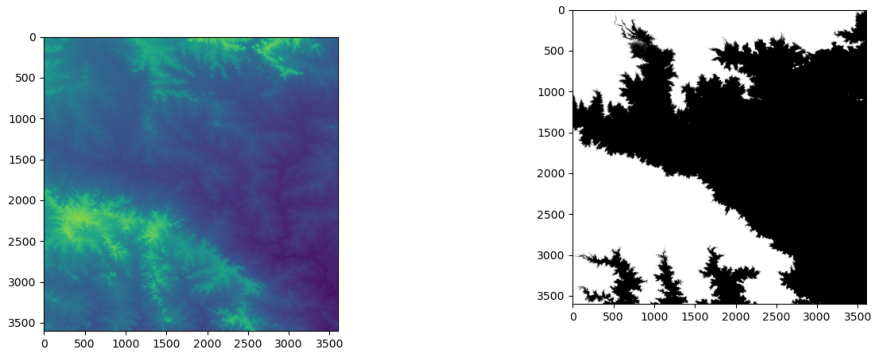


Figure 2: Optimal sites by level of importance by author Saraswat et al.[2]

2.2.1 Terrain Data

USGS is an agency of United States government that works across disciplines like geology, geography and hydrology. SRTM (Shuttle Radar Topography Mission) was undertaken to created digital elevation models (DEM) of earth surface in collaboration with NASA (National Aeronautics and Space Agency). This resulted in two Digital Elevation Models available for research with spatial resolutions of 1 arc-second (30 meters) and 3 arc-second (90 meters). For this study we will be using the DEM model with 1 arc second of spatial resolution[13]



(a) Elevation map

(b) Elevation values with a mask on the average

Figure 3: Elevation map for coordinates N 20' E 78

2.2.2 Solar Irradiance Data

National Solar Irradiance Database (NSRDB) is a database of solar irradiance calculated on hourly and half hourly bases[14]. It is created and maintained by NREL, U.S. Department of Energy and many other contributors. Solar irradiance is measured in three types of measurement- Global Horizontal Irradiance (GHI), Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DHI).

DNI refers to the amount of solar radiation recevied per unit area on the surface that is perpendicular to the sun rays incident on the surface, whereas GHI refers to the total amount of solar radiation received per unit area on earth's surface. It represents cumulation of diffused horizontal irradiance, ground-reflected radiation and diffused sky radiation.

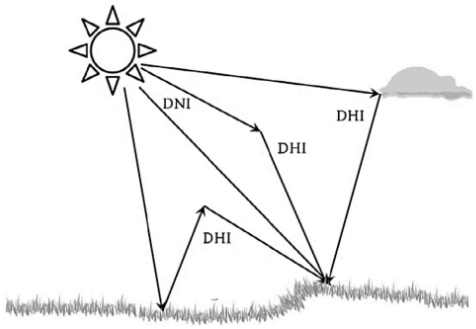
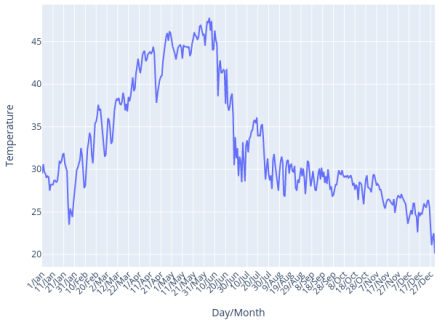
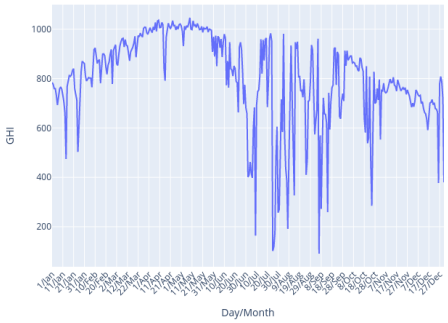


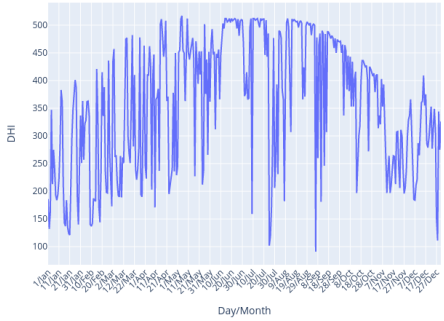
Figure 4: Solar Irradiance components[3]



(a) Temperature



(b) Global Horizontal Irradiance



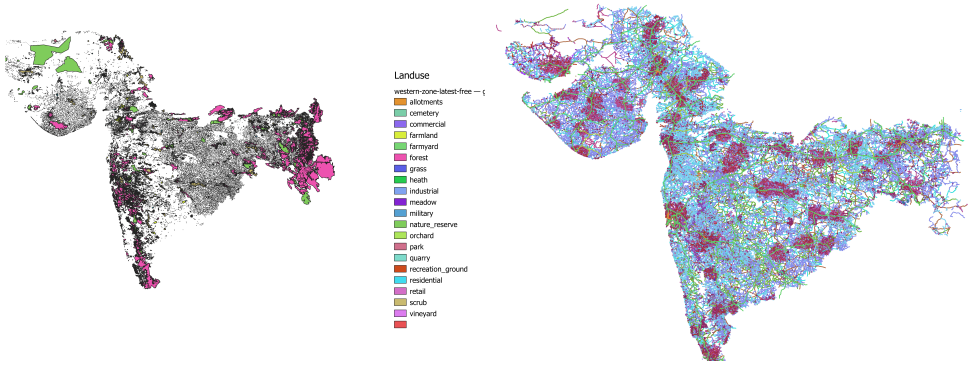
(c) Direct Horizontal Irradiance

Figure 5: Solar irradiance data for co-ordinates for coordinates N 20' E 78'

2.2.3 Other Important Attributes

Solar irradiance and Elevation are probably two of the important features to be considered before setting up of a solar farm, but there are many other factors that need to be considered before selecting a site for solar farm. These factors include the financial viability of the project i.e there is enough demand for energy in the region to sunstain a solar farm, enviromental impact of developing a solar plant in an ecologically sensitive region, land-use guidlines, skilled labour, etc.

OSM is an open source collaboration project that aims to create free geographic data, launched in 2004 it has grown into a global community-driven initiative. Maps and data created by collaborators at OSM would be used for miscellaneous attributes needed[15].



(a) Land Use for Western India

(b) Road and Railways

Figure 6: Attributes for Western India

2.3 Analytical Hierarchy Process

MCDMs have been used extensively in the literature for the task of identifying optimal solar PV plant sites. Unlike machine learning techniques, MCDMs are primarily concerned with making decisions based on a set criteria. The process of ranking criteria that facilitate decision making is set manually. Whereas, ML algorithms enable decision-making models to learn biases without explicitly been programmed.

AHP is a commonly used MCDMs used by many authors in literature[8, 1, 9, 2]. AHP was developed by Prof.Thomas Satty[16]. The AHP at its core involves ranking the criterias that would weigh in the decision-making process. The process of AHP can be summed up in three stages. First stage, define the problem and create a hierarchy. Identify the goal, in the case of this study, To rank suitablity of a location for solar PV plant and define a hierarchy based on the criteria/factors in this case elevation, slope, solar irradiance, land use, land value, etc. These criteria can be futher divided into sub-criteria to create a hierarchical structure.

After setting up a hierarchy, the we need to define importance of criteria/factors relative to one another. This can be done using pairwise comparison. Pairwise comparison involves comparing each factor relative to every other criteria. These comparisions can be stored in a maxtrix called pairwise comparision matrix. Now that the pairwise comparision of each factor relative to another is calculated the next step is to determine weights for each criteria. The matrix is first normalised and a weighted sum is calculated of normalised criteria weights to give a score. From the normalised vector values, the Consistency ratio is calculated to check the validity of the

hierarchy established. Consistency ratio is a measure that helps ensure the reliability of the decision-making process. When the Consistency ratio is below 0.1 the process and the weights generated can be considered acceptable and consistent[16]. By Formula, Consistency Ratio (CR), Consistency Index (CI), Random Consistency Index (RI) formulates

$CR = \frac{CI}{RI}$

RI is a reference value used to assess the consistency of pairwise comparisons. It provides benchmark for validating the consistency achieved for the defined hierarchy. CI is calculated based on the eigen value of pairwise comparision matrix.

$CI = \frac{\lambda_{\max} - n}{(n-1)}$

where n is the number of criteria involved and $\lambda_{\max}$ represents largest eigenvalue.

2.4 Kohonens model

The Kohonen model or the Kohonen neural network also known as Self-Organizing Map (SOM) is one of the unsupervised clustering algorithms, it was developed by Kohonen et al in 1982[17]. Typical it is used for Clustering and it was used extensively used by Chang et al for identifying location with high landslide susceptibility[10].

Dimensionality Reduction is one of the goals of Kohonen model. Creating low dimensional representation while preserving the properties of the data is done by assigning each neuron with a weight vector of the same dimension as the input data. The weights are iteratively aligined to the distribution of the input data.

Initally, all the weight vectors of the neurons are assigned random values based on a normal distribution. Iteratively input vector are selected from training data. A distance or similarity measuring measure like euclidean distance or cosine similarity is used to calculate the distance of the input vector to the weighted vector. Best Matching Unit (BMU) is the neuron whose weighted vector is the closest to the input vector. The weights are updates for the neurons to move closer to the selected input vector. This process is repeated until convergence.

After training, Kohonen model outputs a lower dimension vector space for the input data. The number of neurons in the model represent number of classes/clusters defined for classification. Kohonen model can be feed with vectorised GIS data with multiple criteria as performed by Chang et al for landslide susceptibility[10].

2.5 Auto Encoder and Multi-Layer Perceptron

Ahmadlou et al carried out an extensive research in Iran and India regarding flood susceptibility[4]. Geospatial data with layers like slope, aspect, altitude, landuse, rainfall were used as deciding criteria for the model[4].

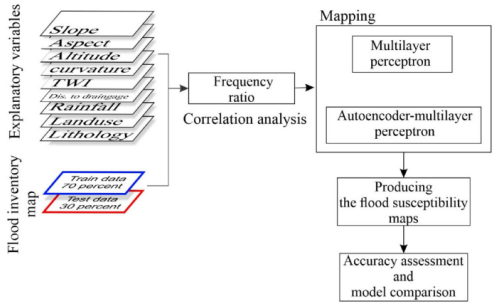


Figure 7: Flowchart of the model proposed by[4]

They used a hybrid model ofMulti-Layer Perceptron (MLP) and Auto-Encoder. Auto-Encoder is a type of neural network architecture proposed by Hinton and Rumelhart[18]. It composes of an encoder and decoder which work together for reducing the dimensionality of the input vector and representing it in a lower dimensional vector space called latent space. Auto encoder over here acts as a feature extractor, later on these features are used for training a MLP to make predicitons.

An auto-encoder is basically a neural network trying to regenerate the input. Structurally, it consists of three components.

1. Encoder: An Encoder that compresses data to a lower dimensional vector space using a neural network consisting of transformational layers like fully connected, convolutional and dropout.

2. Latent Space: The compressed representation of input vector that encoder produced, it is a bottleneck that preserves essential features of the data.

3. Decoder: Another neural network that reconstructs the original input

vector. It is used to calculate the reconstruction cost.

While training an auto-encoder the decoder output is compared to the original input. The error in the reconstructed output is called reconstruction error. This error is minimized using backpropogation algorithm until convergence is reached. The trained encoder is then used to generate lower dimensional data that can be used for different machine learning tasks.

References

[1] H. E. Colak, T. Memisoglu, and Y. Gercek, "Optimal site selection for solar photovoltaic (PV) power plants using GIS and AHP: A case study of malatya province, turkey," *Renewable Energy*, vol. 149, pp. 565–576, Apr. 2020. [Online]. Available: https://doi.org/10.1016/j.renene.2019.12.078

[2] S. Saraswat, A. K. Digalwar, S. Yadav, and G. Kumar, "MCDM and GIS based modelling technique for assessment of solar and wind farm locations in india," *Renewable Energy*, vol. 169, pp. 865–884, May 2021. [Online]. Available: https://doi.org/10.1016/j.renene.2021.01.056

[3] F. Vignola, J. Michalsky, and T. Stoffel, *Solar and Infrared Radiation Measurements, Second Edition.* CRC Press, 2023.

[4] M. Ahmadlou, A. Al-Fugara, A. R. Al-Shabeeb, A. Arora, R. Al-Adamat, Q. B. Pham, N. Al-Ansari, N. T. T. Linh, and H. Sajedi, "Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder

neural networks," *Journal of Flood Risk Management*, vol. 14, no. 1, Dec. 2020. [Online]. Available: https://doi.org/10.1111/jfr3.12683

[5] BP, "BP Statistical Review of World Energy 2021," 2021, accessed: 2023-06-26. [Online]. Available: http://www.indiaenvironmentportal.org.in/files/file/bp%20statistical%20review%20of%20world%20energy%202021.pdf

[6] BBC News, "COP26: India PM Narendra Modi pledges net zero by 2070," Online, 2023, accessed: 2023-06-26. [Online]. Available: https://www.bbc.com/news/world-asia-india-59125143

[7] Reuters, "India's solar boom reverses gas momentum, cements coal use: Maguire," Online, December 2022, accessed: 2023-06-26. [Online]. Available: https://www.reuters.com/world/india/indias-solar-boom-reverses-gas-momentum-cements-coal-use-maguire-2022-12-14/

[8] H. Z. A. Garni and A. Awasthi, "Solar PV power plant site selection using a GIS-AHP based approach with application in saudi arabia," *Applied Energy*, vol. 206, pp. 1225–1240, Nov. 2017. [Online]. Available: https://doi.org/10.1016/j.apenergy.2017.10.024

[9] M. Zoghi, A. H. Ehsani, M. Sadat, M. javad Amiri, and S. Karimi, "Optimization solar site selection by fuzzy logic model and weighted linear combination method in arid and semi-arid region: A case study isfahan-IRAN," *Renewable and Sustainable Energy Reviews*, vol. 68, pp. 986–996, Feb. 2017. [Online]. Available: https://doi.org/10.1016/j.rser.2015.07.014

[10] Z. Chang, Z. Du, F. Zhang, F. Huang, J. Chen, W. Li, and Z. Guo, "Landslide susceptibility prediction based on remote sensing images and GIS: Comparisons of supervised and unsupervised machine learning models," *Remote Sensing*, vol. 12, no. 3, p. 502, Feb. 2020. [Online]. Available: https://doi.org/10.3390/rs12030502

[11] A. Jain, R. Mehta, and S. K. Mittal, "Modeling impact of solar radiation on site selection for solar PV power plants in india," *International Journal of Green Energy*, vol. 8, no. 4, pp. 486–498, May 2011. [Online]. Available: https://doi.org/10.1080/15435075.2011.576293

[12] S. Sindhu, V. Nehra, and S. Luthra, "Investigation of feasibility study of solar farms deployment using hybrid AHP-TOPSIS analysis: Case study of india," *Renewable and Sustainable Energy Reviews*, vol. 73, pp. 496–511, Jun. 2017. [Online]. Available: https://doi.org/10.1016/j.rser.2017.01.135

[13] T. G. Farr and M. Kobrick, "Shuttle radar topography mission produces a wealth of data," *Eos Trans. AGU*, vol. 81, pp. 583–583, 2000.

[14] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby, "The national solar radiation data base (nsrdb)," *Renewable and Sustainable Energy Reviews*, vol. 89, p. 51–60, 2018.

[15] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org ," https://www.openstreetmap.org, 2017.

[16] T. L. Saaty, "What is the analytic hierarchy process?" in *Mathematical*

Models for Decision Support. Springer Berlin Heidelberg, 1988, pp. 109–121. [Online]. Available: https://doi.org/10.1007/978-3-642-83555-1_5

[17] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982. [Online]. Available: https://doi.org/10.1007/bf00337288

[18] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986. [Online]. Available: https://doi.org/10.1038/323533a0