GIS Framework for Optimal Solar Farm Siting

Group Number : 25

IMT2023034	Nainika Agrawal
IMT2023035	Harshita Bansal
MT2024011	Aishwarya J Panampilly

Problem definition

With rapidly growing energy demand, rapid urbanization, and increasing pressure on Bengaluru's existing power infrastructure, the city needs to diversify into clean and reliable energy sources. Solar energy is particularly well-suited for Bengaluru due to its high annual irradiance and the state's strong policy push toward renewable integration. To support this transition, this project aims to develop a framework that predicts solar energy potential and integrates it with spatial, environmental, and infrastructural layers. The outcome is an optimized Solar Farm Suitability Map for the Bengaluru region, which provides spatial intelligence to guide sustainable energy planning and strengthen the city's long-term clean energy strategy.

Objectives

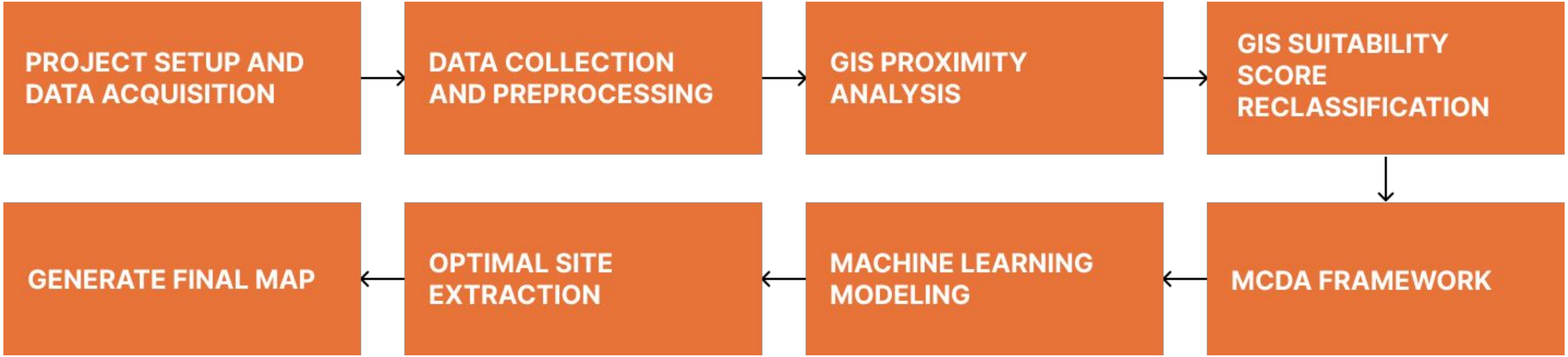
- To acquire, preprocess, and integrate relevant GIS thematic layers, including terrain characteristics, land use/land cover, environmental constraints, infrastructure accessibility and solar irradiance.
- To design, train, and validate a Machine Learning model (XGBoost) capable of accurately forecasting temporal solar irradiance based on historical meteorological data.
- To perform multi-criteria decision analysis(MCDA) by applying weighted scoring to environmental, technical, and socio-economic factors to quantify site feasibility.
- To generate a Solar Farm Suitability Map that identifies and ranks optimal zones for utility-scale solar development within the Bengaluru district.

Literature Review and Research Questions

- Rodrigues, S., et al. (2017). "Suitability analysis of solar photovoltaic farms: A Portuguese case study." International Journal of Renewable Energy Research
- Uyan, M. (2013). "GIS-based solar farms site selection using analytic hierarchy process (AHP) in Karapinar region, Konya/Turkey." *Renewable and Sustainable Energy Reviews*
- Al Garni, H. & Awasthi, A. (2017). "Solar PV power plant site selection using a GIS-AHP based approach with application in Saudi Arabia."
- Bhanja, R. and Roychowdhury, K.: A MULTI-CRITERIA GIS BASED ANALYTICAL HIERARCHICAL APPROACH FOR SOLAR PHOTOVOLTAIC FARM SITE SELECTION IN THE KOLKATA METROPOLITAN AREA, INDIA, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVIII-4/W5-2022, 31–36,

- Which environmental and spatial constraints most significantly influence solar site suitability in Bengaluru when weighted via MCDA?
- Does integrating topographic modeling with temporal irradiance forecasting offer a more robust feasibility assessment than static mapping alone?
- How does enforcing a minimum contiguous land area (e.g., >5 Ha) impact the district's total utility-scale solar capacity?

Flow of the Project



Methodology

1. Data Collection : Define study area, set up QGIS and python, collect all required data.

2. Preparation of Layers : Standardize CRS (**EPSG:32643, WGS 84 / UTM zone 43N**), clip all layers to study area, create slope and aspect layers from DEM data, generate distance-to-roads and distance-to-grid rasters.

3. GIS Reclassification & MCDA: Reclassify all layers to **1–5 suitability scale** using literature and statistics of the layer. Assign weights (e.g., Solar 35%, Grid 25%, Roads 15%, Slope 15%, LULC 10%) for the MCDA calculation. framework.

4. Weighted Overlay & Suitability Map: Apply MCDA using Raster Calculator, produce **Final Suitability Map** showing optimal high-score zones.

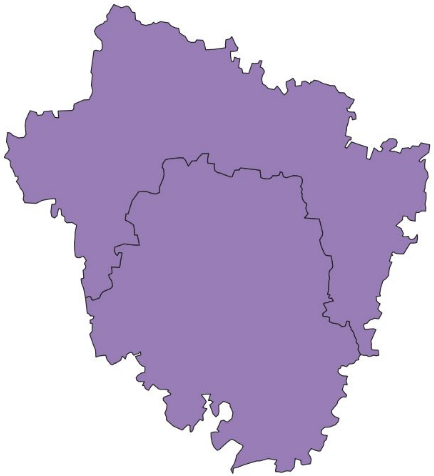
5. Machine Learning Modeling: Perform preprocessing and train XGBoost model on irradiance time-series.

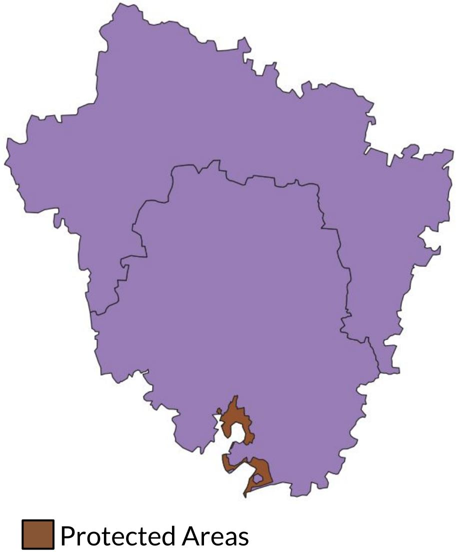
6. Final Site Selection: Select top sites by area and mean solar energy received. Use ML model to forecast expected energy output for chosen locations.

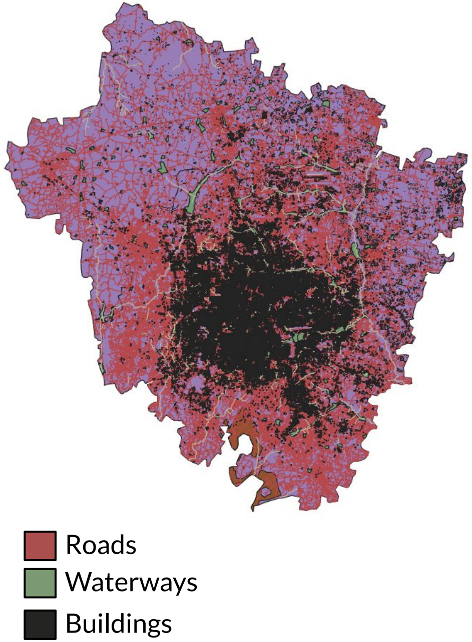
Methodology : Data Collection

- **Geofabrik – OpenStreetMap Extracts** (for downloading OSM layers)
- **GADM – Global Administrative Boundaries** (for country, state, and district boundaries)
- **NASA POWER Data Access Viewer** (for solar irradiance and meteorological data, 2016–2025)
- **ArcGIS Living Atlas – Land Cover Explorer** (for Land Use/Land Cover data)
- **NRSC Bhoonidhi Portal** (for Digital Elevation Model data)
- **Environmental Justice Atlas – Protected Areas of India** (for protected and conservation areas)
- **QuickOSM Plugin (QGIS)** (for extracting infrastructure layers such as substations, power lines)
- **SAGA (System for Automated Geoscientific Analyses)** algorithm *"Potential Incoming Solar Radiation"* was executed within the QGIS Processing Toolbox with DEM as the input layer to obtain total insolation energy.

Methodology: Preparation of Vector layers





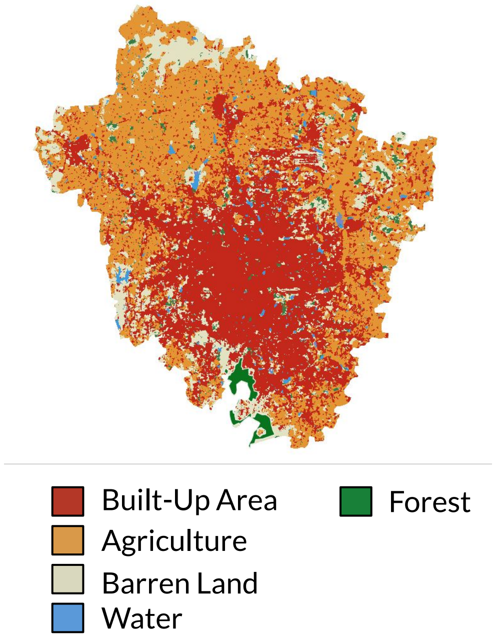


1. Add the Bengaluru urban and Bengaluru rural district boundary data from global administrative boundary data collected

2. Add protected areas and clip to Bengaluru boundaries

3. Add roads, buildings, water, waterways from OpenStreetMap data and clip them to Bengaluru boundary.

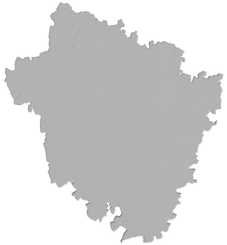
Methodology: Adding Raster layers



Raster map of the Land Use/Land Cover layer at 10 m resolution, resampled to 30 m using the nearest-neighbour algorithm.

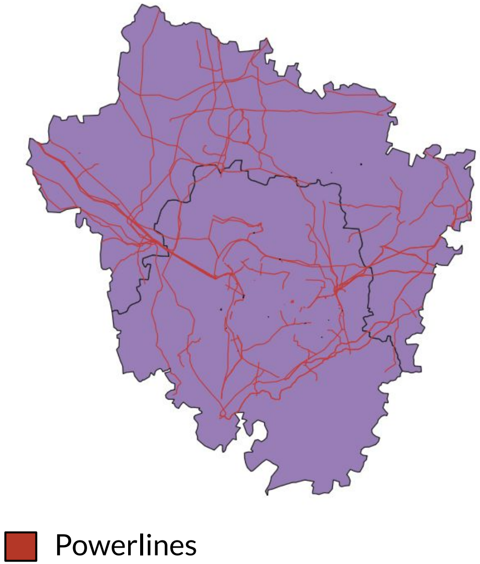
4. Raster map of DEM(Digital Elevation Model) 30m resolution

From DEM we made:
1) Slope



2) Aspect

Methodology: Infrastructure and Energy



Use QuickOSM in QGIS to download powerline and substation data by querying the OSM `power` features

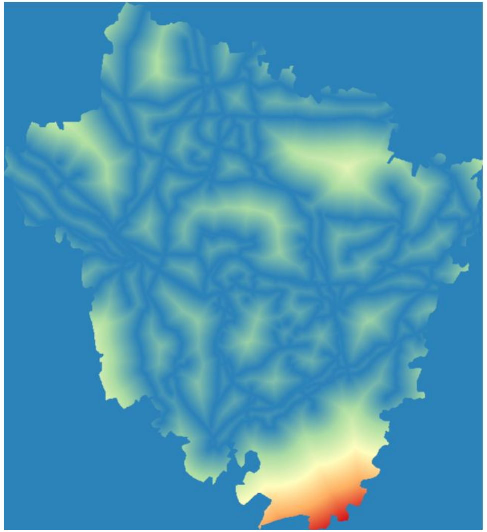
Using the SAGA 'Potential Incoming Solar Radiation' tool, we computed total annual solar energy from the DEM.

The model was run for 0–365 days with a 5-day interval and a 0.5-hour temporal resolution. This setup captures seasonal changes and terrain shadowing in high detail, producing a raster of annual solar energy per 30 m cell.

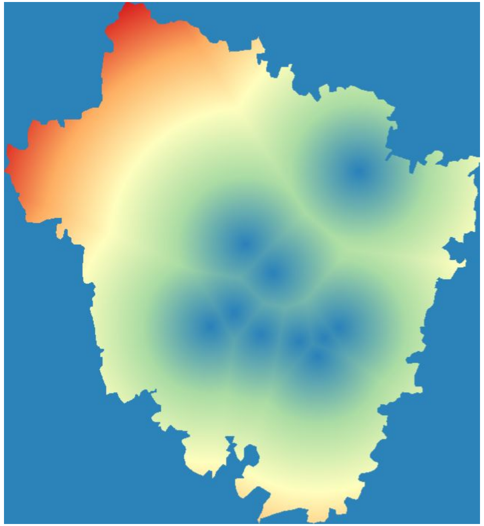
Methodology: Proximity Analysis

We rasterized the road and grid layers at a 30 m resolution, and then used these rasters to compute the distance of each point from the nearest road/grid, expressed in meters.

An inverted Spectral color ramp is used, where **blue** denotes the shortest distances and **red** represents the greatest distance.



Distance to powerlines



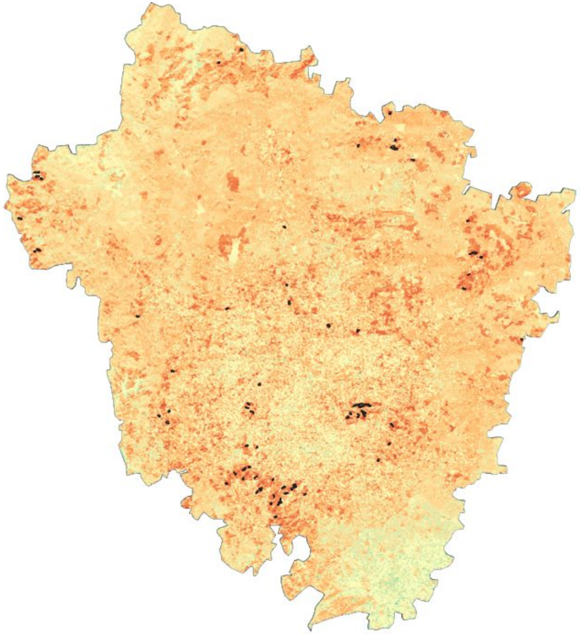
Distance to substations

Methodology: GIS Reclassification and MCDA Framework

- Using data from QGIS → Properties → Information → Statistics, each final raster layer was reclassified into **5** suitability classes from **1 (very poor) to 5 (Excellent)** based on min–max values to create score layers:
  - **Distance to Roads** (Closer = Better)
  - **Distance to Power Lines** (Closer = Better)
  - **Distance to Power Substations** (Closer = Better)
  - **Total Solar Energy (**Higher = Better)
  - **Slope** (Low = Better)
  - **Aspect** (South-facing = Better)
  - **Land Use/Land Cove**r (Open land = Better)

- **MCDA Score** given using Analytical Hierarchy Process(**AHP**)**:**
  - **TotalEnergy**: 35%
  - **Distance to Power Lines**: 15%
  - **Distance to Power Substations**: 10%
  - **Aspect**: 5%
  - **Distance to Roads**: 15%
  - **Slope**: 10%
  - **LandUse/LandCover:** 10%

Results : Suitability Map



■ Suitable Solar Siting Areas

Zonal Statistics — Features Total: 67, Filtered: 67, Selected: 0

	fid	DN	area_ha	_mean
1	873	1	14	7.028079734...
2	1202	1	13	6.992815128...
3	577	1	9	6.990055141...
4	23	1	13	6.989258456...
5	1213	1	6	6.868131070...
6	443	1	6	6.784465179...
7	561	1	10	6.768981117...
8	568	1	13	6.750128910...
9	105	1	6	6.705191227...
10	340	1	9	6.462593387...
11	496	1	33	6.451885334...
12	4784	1	6	6.270383521...
13	4629	1	10	6.241167474...
14	3395	1	6	6.198043071...
15	1296	1	6	6.172980645...
16	4752	1	8	6.157953087...
17	2755	1	9	6.135996918...

Using Raster calculator we generated a final suitability map.

We applied a threshold (Score > 4.8) to isolate only the top-tier "Optimal Zones" and converted these into vector polygons.

Area Constraints: We filtered candidate sites to retain only contiguous places larger than 5 Hectares (~2.5 MW capacity)

Methodology: Machine Learning Model

1. **Load NASA POWER solar-irradiance GIS data** from CSV (header row 18), clean column names, remove sentinel values, and construct a proper hourly DatetimeIndex.

2. **Perform data preprocessing**, including interpolating missing values, converting timestamp fields, generating month/hour fields, and preparing the dataset for feature engineering and model training.

3. **Generates GIS-based advanced engineered features**, such as cyclical encodings (hour/month sin–cos), irradiance lags (1–168 hours), rolling window statistics, and meteorological predictors (temperature, humidity, wind, pressure, UV indices).

4. **Define the prediction target** (ALLSKY_SFC_SW_DWN) and perform an 80/20 time-aware split into training and testing sets, and standardize all numeric input features.

5. **Train an XGBoost regression model**, and export the final Booster using optimized hyperparameters and the engineered feature input space.

6. **Evaluate the model** using RMSE, MAE, and R^2 metrics, and visualizes predicted vs. actual irradiance using Matplotlib to confirm temporal accuracy. Our output values are: Test RMSE=4.0566, Test MAE= 2.0139, Test R2=0.9998.

7 **Save outputs**, including the engineered datasets, and prediction plots, and use the model to forecast solar energy for the year 2026. We got an average of **5.98** kWh/m^2/day for bangalore annually.

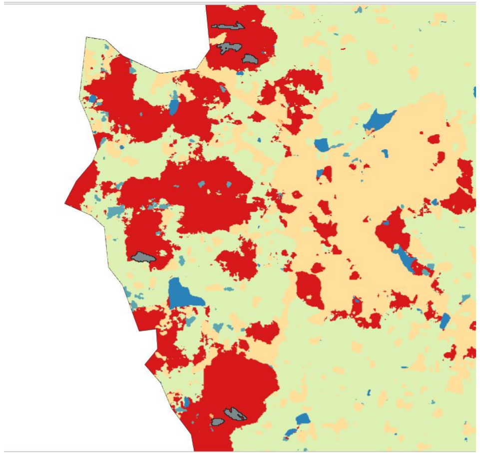
Results

Actions taken to identify and validate the single best location for development

Ranking: Ranked the final candidate polygons based on Mean Solar Potential (via Zonal Statistics) and Total Area.

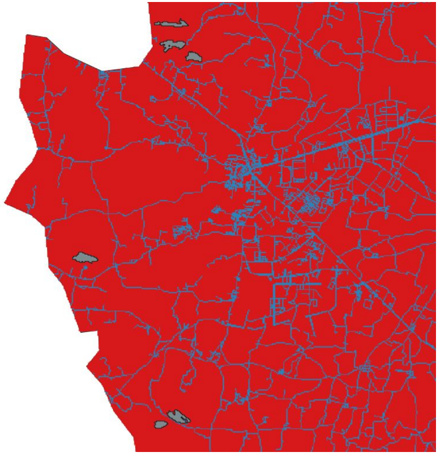
Validation: Performed a visual "ground truth" check using latitude/longitude to ensure land availability(converter)

Forecasting: Applied the ML model to generate a 2026 Energy Production Forecast, confirming long-term viability.



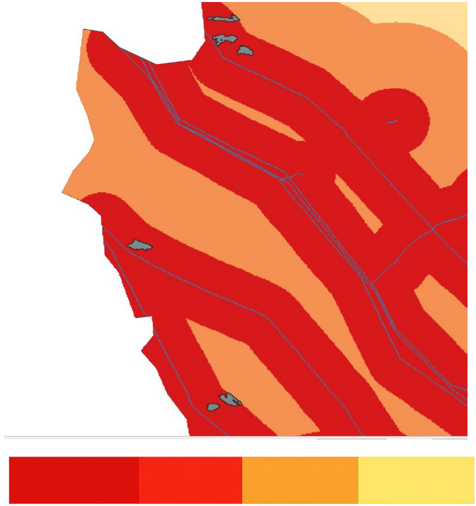
Built-Up Area
Water
Barren Land
Agriculture

LULC



Roads

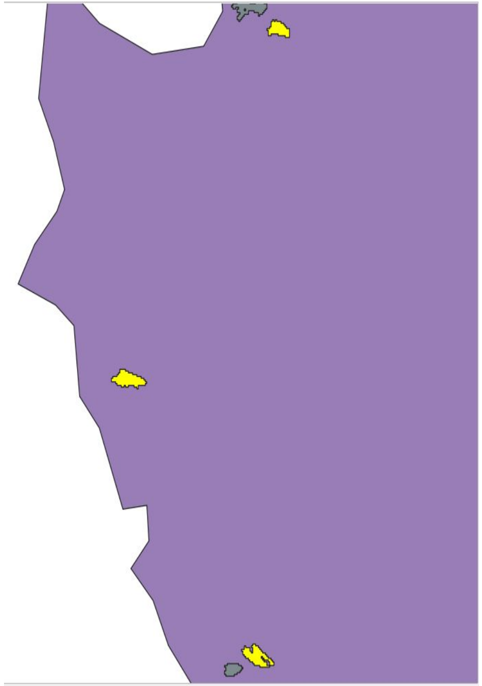
Suitable Site for Solar Farm



Closest **Farthest**

Distance from Power Lines

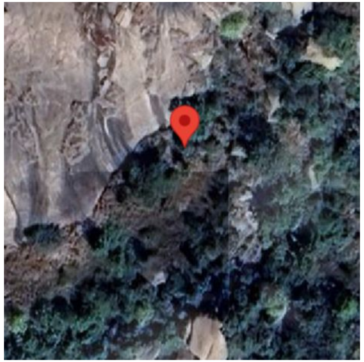
Top 3 sites



Suitable Solar Plant Site

First:
13°12'17.4"N,
77°11'46.6"E

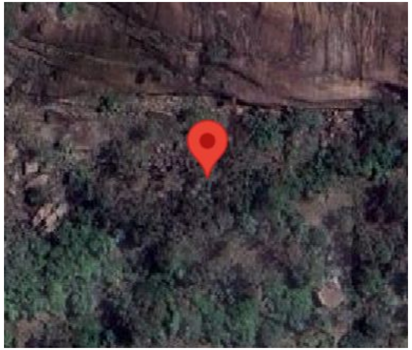
Area: 14 hectares
Mean Solar energy: 7.03 kWh/m²/day
(17.6% better than predicted)



Area: 13 hectares

Solar Energy: 6.993 kWh/m²/day (16.9% better)

2nd: 13°09'44.9"N, 77°13'17.1"E



Area: 9 hectares

Solar Energy: 6.990 kWh/m²/day (16.9% better)

3rd: 13°15'28.5"N, 77°13'35.5"E

Results: Accuracy Assessment

The XGBoost solar irradiance model achieved **very high predictive accuracy** (R^2 = 0.9998, RMSE = 4.06, MAE = 2.01), reliably capturing temporal and seasonal variability.

High-quality GIS inputs (LULC, DEM–30 m, infrastructure layers) and **AHP-based MCDA** ensured consistent and logically weighted spatial evaluation (CR < 0.1).

MCDA-generated high-suitability zones showed **strong agreement with actual solar potential**, with selected sites receiving >6.9 kWh/m²/day.

Ground-truth validation confirmed land availability, minimal environmental conflicts, and compliance with the **>5 ha contiguity constraint**.

Overall, the integrated **GIS–Machine Learning framework** provides a **robust, accurate, and scalable decision-support approach** for utility-scale solar farm siting in Bengaluru

Future Work

- Include socio-economic and policy constraints such as land ownership, population density, and regulatory zones for more realistic site selection.

- Incorporate grid capacity and substation limits

- Develop an interactive decision-support tool or web-based GIS platform for policymakers, planners, and energy developers to access suitability maps easily.

- Expand the framework to other renewable sources (e.g., hybrid solar–wind farm suitability) to support multi-energy planning.