Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predictive Urban Scale Digital Twins For Optimising Solar Panel Energy Yield

Oreofe Jolaolu

jasonjolaolu@gmail.com

Independent Reseacher https://orcid.org/0009-0004-2039-2686

Ajibola Oladiran

Independent Reseacher https://orcid.org/0009-0004-3078-312X

Dayo Osibamowo

Independent Reseacher https://orcid.org/0009-0006-3745-629X

Research Article

Keywords: Digital Twins, Artificial Intelligence, Building Information Modelling, BIM, Solar Energy, Renewable Energy Systems, Machine Learning, Spatial Data

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11617811/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Predictive Urban Scale Digital Twins For Optimising Solar Panel Energy Yield

Contributing Authors:

Jolaolu O.J.[1], Oladiran A.A.[2] & Osibamowo D.[3]

Abstract

The adoption of smart, renewable energy systems in developing regions is often challenging due to the cost associated with obtaining proprietary spatial data and dependence on forecasting models that do not adequately capture localised microclimate phenomena. We present a physics driven AI digital twin to optimise solar energy infrastructure planning using an educational institution as a case study for the pilot application in this research work.

Current solar assessment methodologies often isolate spatial mapping from temporal forecasting. They also utilise machine learning models that neglect thermodynamic laws and in some instances, they ignore region specific energy drawbacks such as periodic harmattan dust soiling, and extreme thermal degradation.

We resolve these limitations by integrating open source geospatial intelligence (OpenStreetMap) and physics constrained machine learning. The methodology utilises a modular physics engine to calculate metrics such as building level shading factors and baseline thermal degradation. These physical representations serve as the features for an optimised LightGBM gradient boosting predictive model trained on hourly ERA5 meteorological data, to accurately forecast 8,760 hour temporal energy yields. We also deploy a spatial nearest neighbour algorithm to triangulate unnamed structures with local road networks/landmarks to generate an automated operational directory.

The resulting framework is an interactive 3D WebGL digital twin which includes financial and environmental dashboards that can be localised based on the region. The model quantifies the exact energy losses attributed to seasonal dust or heat and the deployment strategy targets the highest yield, lowest shading facilities.

Keywords: Digital Twins, Artificial Intelligence, Building Information Modelling, BIM, Solar Energy, Renewable Energy Systems, Machine Learning, Spatial Data.

[1] ORCID ID: 0009-0004-2039-2686 | [2] ORCID ID: 0009-0004-3078-312X |[3] ORCID ID: 0009-0006-3745-629X

Introduction

Unreliable national grids across developing regions force heavy reliance on expensive diesel generators (Olowosejeje et al., 2019) and there is a need to seek cheaper alternative energy sources to address climatic concerns (Bhatti & Jamro, 2025).

Large scale institutional environments, such as university campuses, serve as effective test sites for solar photovoltaic (PV) micro grids due to high, predictable energy demands (Muhtadi et al., 2021) and large rooftop footprints. The transition to a localised smart grid will require detailed spatial and temporal planning to optimise capital expenditure and guarantee long term energy security (Schmidt & Müller, 2018).

Spatial solar potential assessments have relied on high resolution, proprietary datasets such as Light Detection and Ranging (LiDAR) (Zhong & Tong, 2022) and Building Information Modelling (BIM) (Alves et. al, 2025). These data acquisition methods are expensive and largely unavailable in developing nations (Tian et al., 2021).

The increase in the use of Artificial Intelligence (AI) and digital twins has impacted urban energy planning (Koirala et al., 2024). Yet, the integration of these technologies in regions such as the global south remains limited. This project combines open source geospatial data (OpenStreetMap) and machine learning techniques to generate urban digital twins that simulate energy potential without incurring significant data collection cost.

While the cost of solar PV hardware is falling (Kavlak et al., 2025), the analytical frameworks used to plan micro grids in these contexts require improvement. Assessments show limitations such as:

- Spatial data deficit. Planners that lack LiDAR-derived 3D shading models, rely on flat 2D approximations that ignore building heights (Kolečanský et al., 2021) and tree canopy obstruction.
- Localised micro climate challenges. Tools like PVWatts use generic efficiency metrics that fail to capture local environmental factors. In Nigeria solar arrays suffer severe performance drops due to seasonal Harmattan dust accumulation (Kazaure et.al, 2024) and high operating temperatures that exceed the 25°C Standard Test Condition (STC).
- Purely empirical forecasting. Machine learning models process raw meteorological data without incorporating thermodynamic principles (Chu et.al, 2024) and fundamental atmospheric laws (Verma et al., 2024).
- Operational Disconnect. Geospatial models often produce abstract coordinates and building IDs (Azari et.al, 2025) leading to operational disconnects. In

regions with unmapped or unnamed structures, physical installation teams are left without usable deployment guides.

The primary objective of this research is to develop a physics informed AI digital twin that optimises spatial and temporal solar micro grid planning for the University of Ibadan campus. We utilise open source data to create a localised and practical deployment framework.

Research Objectives

- Build an interactive WebGL digital twin using OpenStreetMap geometries and Python libraries to extract building heights and model dynamic temporal shading.
- Develop a physics informed LightGBM model incorporating GOSS sampling to forecast 8,760 hours of annual energy output.
- Implement a spatial nearest neighbour algorithm to cross reference unnamed buildings with local road networks and landmarks. This generates a construction directory with street level metadata.
- Construct dashboards that turn spatial and machine learning outputs into decision making tools, including a 25 year financial ROI model calibrated to local currency.

Literature Review

A digital twin is a smart system that serves as a virtual carbon copy of the physical urban environment, maintaining constant data exchange with its physical counterpart (Hooli, 2025). The prevalent use of solar thermal and photovoltaic technologies makes it vital to map regional solar potential using digital twins

Abolhassani et al. (2026) evaluated the CityDigitalTwin (CDT) platform to address the historical lack of empirical validation in high resolution Urban Building Energy Modeling (UBEM). The validation revealed that 87% of actual measured daily electrical cooling loads fell directly within the digital twin's interquartile simulation envelope, proving its usefulness for neighbourhood scale energy and heat mitigation planning.

To address the challenges of the Urban Heat Island (UHI) effect, researchers developed and validated a 3D City digital twin framework in Padua, Italy. This platform specialised in coupling high resolution 3D spatial mapping of urban soil sealing (impervious surfaces) with real time physical environmental variables (Sukma et.al, 2024).

The core methodologies of digital twin development using Light Detection and Ranging and Building Information Modelling have been demonstrated across scientific literature. For example, a study by Rashidi et al., 2026 provides a comprehensive analysis of the transition from BIM to digital twins in the construction industry. However, the cost and computational requirements of LiDAR and BIM indicates the need for alternatives in urban modelling projects in developing regions.

OpenStreetMap (OSM) is a practical alternative for urban energy planning project. It is an open source, global project designed to provide a free, editable digital map (OpenStreetMap, 2021). Due to OSM's availability and ease of use, it is frequently used for projects at urban level.

Traditional Machine Learning (ML) and Deep Learning (DL) models often ignore fundamental physical principles and stereotypically, they do not offer consistent uncertainty estimates (Rajaperumal & Chinnappan, 2026). As a result, the transition from structural spatial mapping to real time power dispatch might present different operational challenges. To resolve these unconstrained "black-box" anomalies, recent methodology has shifted toward Physics Informed Machine Learning (PIML) for time-series solar forecasting. PIML combines machine learning techniques with physical laws to enhance the accuracy and reliability of machine learning models.

For large scale data manipulation, tree based ensemble methodologies are effective in handling tabular meteorological features without requiring intensive input data transformations. LightGBM can utilise Gradient-Based One Side Sampling (GOSS) to selectively subset data based on large gradient instances, allowing it to map dense

8,760 hour yearly profiles while retaining accuracy and reducing computational memory strain (Vargas et al., 2024).

It should be noted that the evaluation of solar energy potential (using digital twins) requires balancing spatial computational limits with the physical realities of solar cell degradation.

Beyond computational modelling, a gap remains between laboratory performance metrics and operational field yields. Solar photovoltaic modules are traditionally rated under laboratory Standard Test Conditions (STC) at a controlled cell temperature of 25°C, which fails to represent real world field environments where ambient conditions fluctuate dynamically.

In actual outdoor deployment, panels experience high solar irradiance paired with limited natural cooling, forcing cell operating temperatures far above laboratory baselines and causing a 20% to 25% drop in true power output compared to ideal capacities (Oyshei, 2024; Bae, 2020).

This environmental performance gap is severe in tropical or high irradiance zones, where sustained thermal exposure regularly pushes operational cell surfaces past 65°C, and triggers heavy thermal throttling. Standard silicon crystalline modules suffer an efficiency penalty of 0.4% to 0.5% for every single degree Celsius climbed past the 25°C threshold. Over extended periods, this extreme heat irradiance exposure combined with intense ambient humidity fast tracks structural defects like micro cracking, resulting in high annual power degradation rates (Haruna et al., 2024; Islam et al., 2024).

In this research, we show that 3D solar micro grid planning can be achieved purely through open source geospatial data and intelligent modelling, we also encode micro climate factors (such as the West African Harmattan season and tropical thermal degradation) into the machine learning pipeline which provides a template for accurate, region specific energy forecasting. The inclusion of an automated spatial smart directory also addresses the implementation gap, ensuring that academic energy models can be translated into practical operations.

Methodology

The location of the case study (for this research) is geographically bounded to the University of Ibadan, Oyo State, Nigeria (*Lat: 7.444°N, Lon: 3.900°E*). Spatial analysis is restricted to building footprints and road networks indexed within OpenStreetMap (as of the data retrieval date) and it is supplemented by procedural tree canopy generation. Temporal energy forecasting encompasses a simulated 8,760 hour (one year) meteorological cycle utilising ERA5 historical weather datasets, influenced by physical parameters of standard commercial monocrystalline PV panels and a baseline system degradation rate of 0.5% per annum (Jordan et. al, 2016) over a 25 year projection.

Project Framework

The project framework is structured into five phases:

- Spatial data acquisition and 3D geometry processing
- Solar yield and harmattan impact assessment engine
- Solar trajectory and usability analysis engine
- Temporal forecasting
- Automated operational mapping.

Spatial Data Acquisition and 3D Geometry Processing

To retrieve building footprints, we utilise OSM APIs via *osmnx* and *geopandas* libraries to extract the raw building polygons and metadata. The data is projected into UTM Zone 31N (*EPSG:32631*) for accurate metric based spatial operations. A processed spatial vector dataset of building footprints (in the specified boundaries) is saved in a GeoJSON file. To clarify geometric data sources and distinguish measured attributes from procedural rules, the data origins are defined below:

Geometric Attribute	**Data Source and Extraction Method**	**Resolution / Accuracy**	**Processing and Validation Method**
Footprint Geometry (X, Y)	OpenStreetMap (OSM) via osmnx	Vector Polygons ($\pm 1\text{m}$)	Reprojected to UTM 31N metric CRS
Ground Elevation: Z_{ground}	SRTM 30m Global DEM via OpenTopography	30m spatial grid ($\pm 5\text{m}$ vertical)	Centroid point raster sampling
Building Levels	OSM building:levels tag	Explicit tag when available	Numeric parsing, fallback to semantic rules

Geometric Attribute	Data Source and Extraction Method	Resolution / Accuracy	Processing and Validation Method
Estimated Height: H_{building}	Semantic Rule Engine	Procedural rules	Explicit levels x 3.5m ; Academic (10.5m); Residential (3.5m); Fallback (7.0m)
Roof Elevation: Z_{roof}	Vectorised Height Imputation Formula	Derived Calculation	$Z_{\text{roof}} = Z_{\text{ground}} + H_{\text{building}}$
Roof Tilt and Azimuth	Newell's Normal Vector Method	Procedural 3D Plane Normal Vector	Calculated directly from extruded 3D Polygon Z coordinates

Table 1.0: Data Origin Matrix

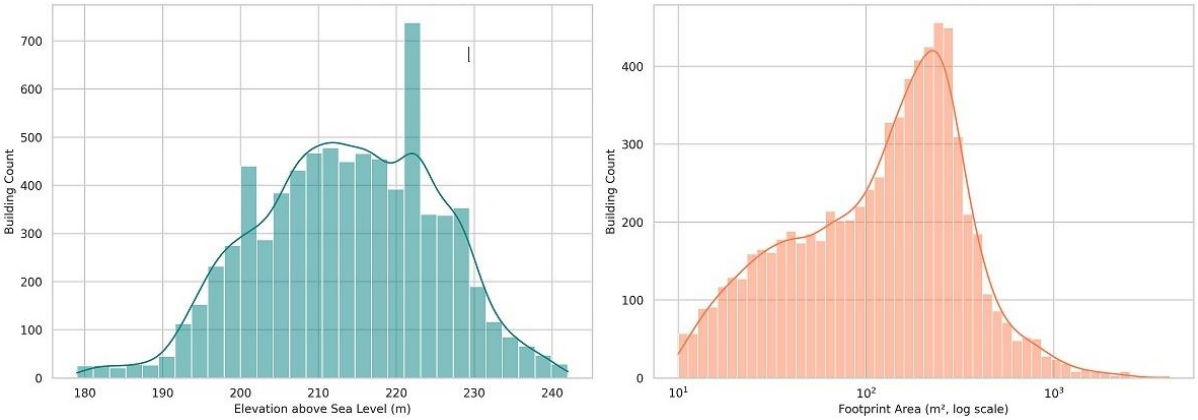


Figure 1: Campus ground elevation profile and building footprint area distribution

	Elevation Z_{ground} (m)	Area $A_{\text{footprint}}$ (m²)
Count	7497	7497
Mean	213.59	185.04
Std	11.47	216.21
Min	179	1.21
25%	206	51.34
50%	214	138.98
75%	222	245.87
Max	242	4108.46

Table 2: Exploratory spatial statistics for 7497 buildings in defined coordinates.

We perform feature engineering and clean the raw spatial data in the generated building dataset. This includes fixing geometric errors and computing building heights/absolute roof elevations. We also filter out noise (for tiny structures) and prepare the dataset for 3D visualisation and solar analysis

The output is saved as the enriched GeoJSON dataset (containing the newly calculated area attributes) for further processing in the pipeline. Contextual semantic rules are applied to refine building heights and extrude standard 2D polygons into true 3D spatial features (*Polygon Z*) configured for GIS visualisation software and format export compatibility. This includes scanning building names for semantic keywords and factors in building area sizes to improve height estimation rules (explicitly defined levels take precedence), followed by inferred academic blocks and residential structures. We also compute absolute roof elevations (Z_{roof}) by adding the new estimated heights to the ground elevations:

$$Z_{roof} = Z_{ground} + H_{building}$$

Where:

- Z_{roof} = Absolute 'top of roof' elevation above sea level
- Z_{ground} = Ground elevation
- $H_{building}$ = Estimated vertical height of the building

The output is the (complete) dataset which is a full 3D GeoJSON file and a streamlined subset as an ESRI Shape file.

Solar Yield and Harmattan Impact Assessment Engine

To evaluate the solar energy potential of the buildings (while accounting for local environmental constraints like seasonal Harmattan dust), we perform:

- Geometry Extraction. The building data is processed and we apply Newell's algorithm to calculate individual roof tilt and azimuth angles.
- Yield Modelling. We use a region specific 12 month climate matrix to calculate transposition irradiance, cell operating temperatures using the Nominal Operating Cell Temperature (NOCT) model, and temperature dependent efficiency losses.
- Environmental Adjustment. We incorporate monthly dust soiling loss factors, particularly accounting for dust degradation during peak Harmattan months.

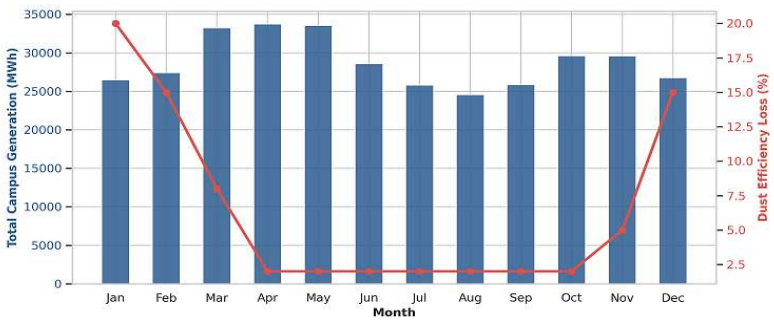


Figure 2: Monthly Potential vs. Harmattan Dust Impact

Solar Trajectory and Usability Analysis Engine

Meteorological solar positions were calculated for the harmattan solstice (worst case shadow length). Solar altitude (α) and shadow length multipliers were derived:

$$\text{Shadow Multiplier} = \frac{1.0}{\tan(\alpha)}$$

Rooftop Net Usable Solar Area

To address real world rooftop constraints (slopes, HVAC installations, parapets, access pathways, and setback requirements), the net usable solar area (A_{usable}) for each facility is defined as:

$$A_{\text{usable}} = A_{\text{footprint}} \times \eta_{\text{roof_type}} \times \left(1.0 - \frac{\text{Shade}_{\text{daytime\%}}}{100}\right) \times \Phi_{\text{setback}}$$

Where:

- $A_{\text{footprint}}$ = Raw metric footprint area (m^2).
- $\eta_{\text{roof_type}}$ = Structural usability factor (0.70 for flat roofs allowing parapet/service access and 0.85 for pitched roofs).
- $\text{Shade}_{\text{daytime}}$ = Mean daytime shading fraction (08:00 to 16:00) cast by adjacent structures and surrounding vegetation.
- Φ_{setback} = Boundary setback and safety buffer coefficient (0.85).

Procedural Vegetation Modelling

Tropical tree canopies were generated procedurally within a perimeter buffer zone (3m to 12m from building edges), assuming a standard height of 12m and canopy radius of 4m to simulate mutual tropical tree shading.

Feature Engineering

To establish the predictive modelling in physical laws, the thermodynamic constraints were embedded directly into the feature space prior to model training. The system retrieves 8,760 hours of historical micro climate data including Global Horizontal Irradiance (GHI), Diffuse Horizontal Irradiance (DHI), ambient temperature, and wind speed via the Open Meteo API (backed by a synthetic data fallback).

For thermodynamic feature engineering, deterministic physical variables were calculated, including PV cell temperature (T_{cell}) derived from ambient heat, wind cooling, and solar flux using Nominal Operating Cell Temperature (NOCT) parameters. Environmental penalties were modelled by combining thermal efficiency degradation

(−0.4% efficiency per degree above 25°C) with a seasonal Harmattan dust soiling penalty peaking between December and February to reflect the dust accumulation.

Feature importance scoring revealed that GHI, and roof top shading fraction were the dominant predictors followed by dust index, DHI and cell temperature (T_{cell}).

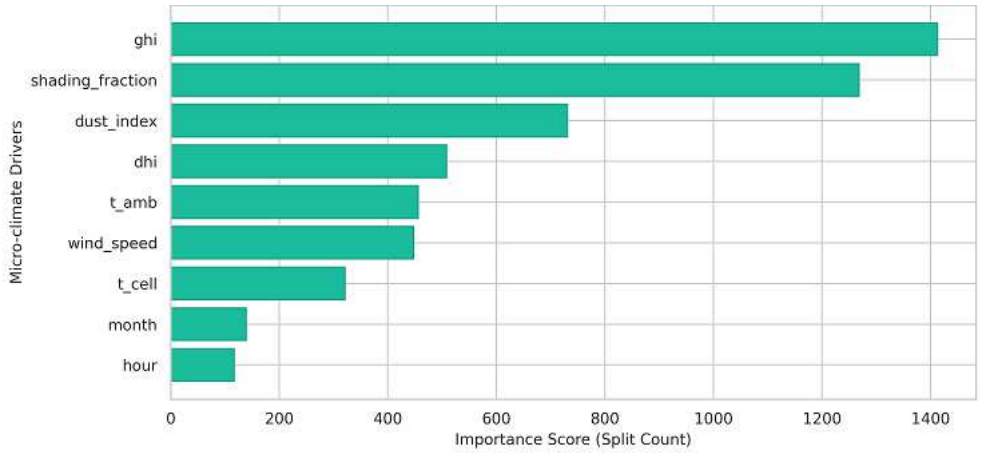


Figure 3: LightGBM feature importance based on real climate data

Temporal Forecasting

We acquire the hourly historical weather analysis data (from the Open-Meteo ERA5 API) for the case study coordinates and design a physics informed model to calculate photovoltaic energy yield by integrating GHI, diffuse radiation, ambient temperature, and wind speed.

- **Cell Temperature Physics** (T_{cell})**:**

$$T_{\text{cell}} = T_{\text{ambient}} + GHI \times \left(\frac{\text{NOCT} - 20}{800}\right) - (W_s \times 0.5)$$

Where $\text{NOCT} = 45^{\circ}\text{C}$ and W_s is wind speed in m/s

- **Instantaneous Thermal Degradation Factor** (η_{temp})**:**

$$\eta_{\text{temp}} = 1.0 - 0.004 \times (T_{\text{cell}} - 25.0)$$

- **Target Variable Calculation** (Reference Array Physics Engine)**:**

$$P_{\text{yield_hourly}} = \left(\frac{GHI}{1000}\right) \times A_{\text{ref}} \times \eta_{\text{STC}} \times (1 - \text{Shade}_{\%}) \times (1 - \text{Soiling}_{\%}) \times \eta_{\text{temp}} \times (1 - L_{\text{sys}})$$

Where:

- GHI = Global Horizontal Irradiance (W/m^2)
- ,A_{ref} = Reference array area (500m^2 / 100 kWp)

- η_{STC} = Standard Test Condition module efficiency (0.20)
- $\mathrm{Shade}_{\%}$ = Hourly rooftop shading fraction
- $\mathrm{Soiling}_{\%}$ = Seasonal Harmattan dust efficiency loss factor
- η_{temp} = Instantaneous temperature dependent efficiency degradation factor
- L_{sys} = Balance of system losses (0.14)

Machine Learning Training

The dataset was split into 80% training and 20% testing subsets. A LightGBM Regressor utilizing GOSS was trained with *n_estimators =180*, *learning_rate=0.07*, and *num_leaves=31*. The LightGBM model is designed to learn and rapidly predict the complex, non-linear thermodynamic interactions across 8,760 hours. The model performance is evaluated using Mean Absolute Error and other performance metrics.

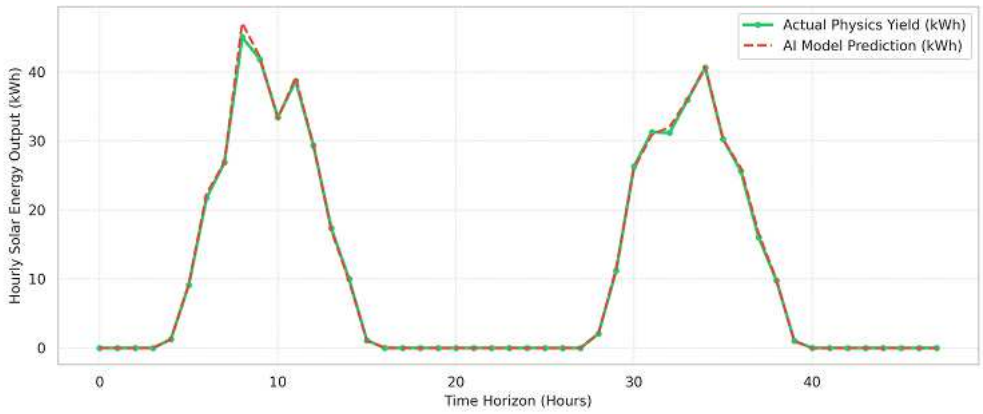


Figure 4: Model accuracy 48 hours continuous yield tracking

We also performed experimental evaluations to compare a baseline Linear Regression model and an advanced Random Forest architecture against the proposed LightGBM model on the same 8,760 hour ERA5-driven thermodynamic dataset with the following results:

Model	MAE (kWh)	RMSE (kWh)	R^2
Linear Regression (Baseline)	1.4058	2.0530	0.9847
Random Forest	0.2156	0.4358	0.9993
LightGBM (GOSS) (Proposed)	**0.1903**	**0.3911**	**0.9994**

Table 3: Predictive Yield Engine Evaluation

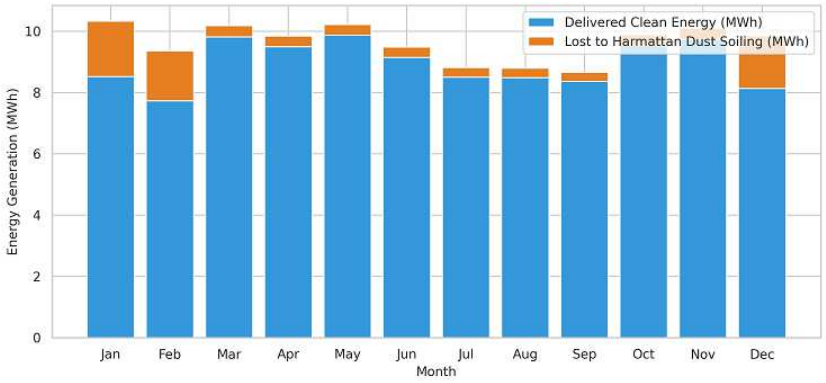


Figure 5: Monthly Energy Yield and Harmattan Dust Impact

Temporal Cross Validation

Standard random train/test splits can introduce temporal data leakage in meteorological and solar yield time series datasets because adjacent hours exhibit high autocorrelation in atmospheric states. When random splitting scatters consecutive hours from the same day across both training and testing partitions, the model is exposed to nearly identical ambient conditions during evaluation. As a result, models can memorise localised weather patterns rather than learning the true underlying physical and thermodynamic dynamics. Standard random splits might lead to artificially inflated accuracy metrics that fail to reflect true out of sample generalisation.

To resolve this (temporal data leakage challenge), we perform a 5-fold temporal block cross-validation implemented by grouping the 12 calendar months into five distinct seasonal blocks. In this setup, the model is strictly trained on four blocks and evaluated on one completely unseen contiguous seasonal block. This approach eliminates time-series autocorrelation bias and ensures true out of sample generalisation. The 5-fold temporal block cross validation yielded the results outlined in Table 4.

Fold / Seasonal Block	Primary Meteorological Characteristic	MAE (kWh)	RMSE (kWh)	R^2
Fold 1 (Dec – Feb)	Peak Harmattan Dust and High Heat	1.9434	3.2496	0.9533
Fold 2 (Mar – Apr)	Early Wet Transition	0.1837	0.3734	0.9995
Fold 3 (May – Jul)	Core Rainy Season	0.1700	0.3336	0.9996
Fold 4 (Aug – Sep)	August Break / Late Rain	0.1653	0.3293	0.9996
Fold 5 (Oct – Nov)	Post Rainy Season Dry Transition	0.1790	0.3437	0.9996
5 Fold Cross Validation Average	**Overall Temporal Generalisation**	**0.5283**	**0.9259**	**0.9903**

Table 4: Cross Validation Results

Automated Operational Mapping

In the final stage we translate abstract spatial geometries into operational recommendations:

- Street and Landmark Indexing. This queries campus road networks and named landmarks within a 2.5 km radius via OpenStreetMap.
- Nearest Neighbour Spatial Matching. This links building centroids to their closest street segment and landmark, computing precise proximity distances.
- Automated Labelling. This resolves missing metadata by assigning descriptive spatial labels (e.g., "Facility near Central Administration (~45m, Campus Way)") to unnamed OSM polygons.
- Interactive 3D Digital Twin. We compile an enriched master GeoJSON, export a deployment directory and render an interactive 3D WebGL dashboard via PyDeck with custom CSS hover tooltips displaying coordinates and routing context.

Discussion

Model Results

The model evaluation compares the baseline Linear Regression model against the Random Forest model and the proposed LightGBM GOSS on the same 8,760 hour dataset.

- Linear Regression (Baseline). While the standard linear model achieves a high baseline R^2 due to the strong linear relationship between raw solar irradiance (GHI) and power output, its error metrics are significantly higher than the non linear models. A Mean Absolute Error (MAE) of 1.4058kWh indicates that a purely linear approach struggles to capture complex, non linear interactive penalties such as the compounded effect of cell overheating T_{cell} and nonlinear shading geometry. The wider gap between MAE (1.4058 kWh) and RMSE (2.0530 kWh) further indicates that the linear model incurs larger residual errors during sudden meteorological changes.
- Random Forest Regressor. Transitioning to a tree based ensemble reduces the prediction error. The Random Forest model captures non linear decision boundaries efficiently, dropping the MAE by over 84% compared to the linear baseline. However, when scaling the predictive yield engine across thousands of individual campus building rooftops, Random Forest's memory footprint and inference overhead might become prohibitively high compared to the LightGBM model.
- LightGBM GOSS (Proposed Model). The Mean Absolute Error of 0.1903 kWh show that the model's hourly energy predictions deviate from the physics simulated target values by less than two tenths of a kilowatt-hour across the evaluation dataset. The alignment between MAE and RMSE (0.3911 kWh) confirms that LightGBM maintains high stability even during volatile weather shifts. The Coefficient of Determination of 0.9994 reveals that the model successfully explains 99.94% of the variance in hourly photovoltaic energy generation. This value confirms that the LightGBM solar yield engine has successfully captured the underlying non linear physical interactions modelled in the pipeline.

Executive Visualisations and Financial Modelling

The pipeline visualisations includes the following important analytical deliverables:

- Energy Loss Waterfall Analysis. A breakdown quantifying gross solar potential versus delivered energy after accounting for shading and thermal losses. The final factor is dust loss due to harmattan soiling.

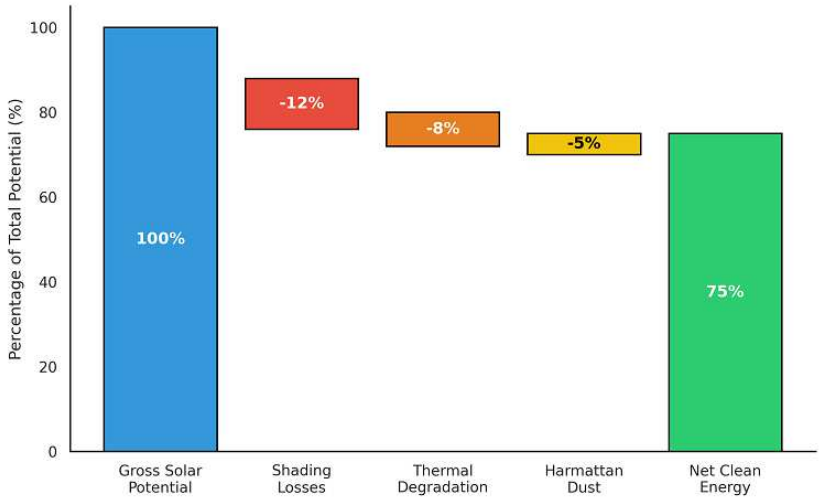


Figure 6: System Efficiency: Where does the solar potential go?

- Gross Solar Potential: 100%
- Rooftop Shading Losses: -12%
- Thermal Degradation Throttling: -8%
- Harmattan Dust Soiling: -5%
- Net Clean Energy Delivered: 75%

- Localised Financial ROI Projections. A 25 year cumulative financial savings trajectory modelled in the local currency, incorporating a 0.5% annual degradation rate and tracking cumulative metric tons of CO_2 offset.

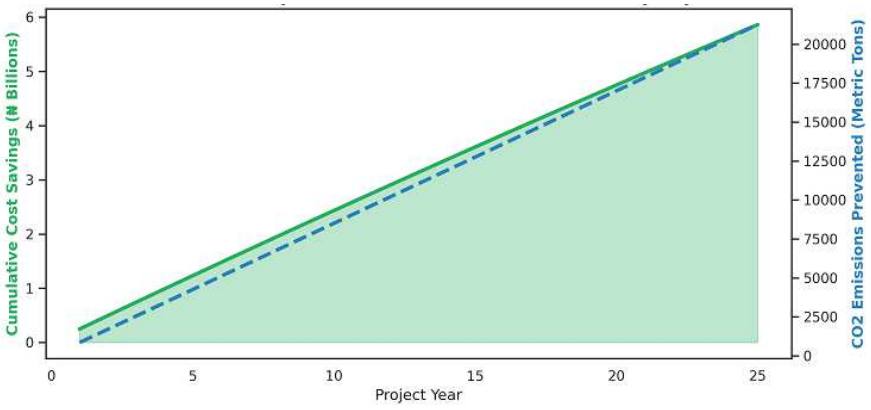


Figure 7: 25 year value projection: Financial ROI and Sustainability Impact

- Micro Grid Supply vs. Demand Curves. Overlays of predicted solar generation profiles against estimated campus power loads to highlight midday clean energy surpluses.

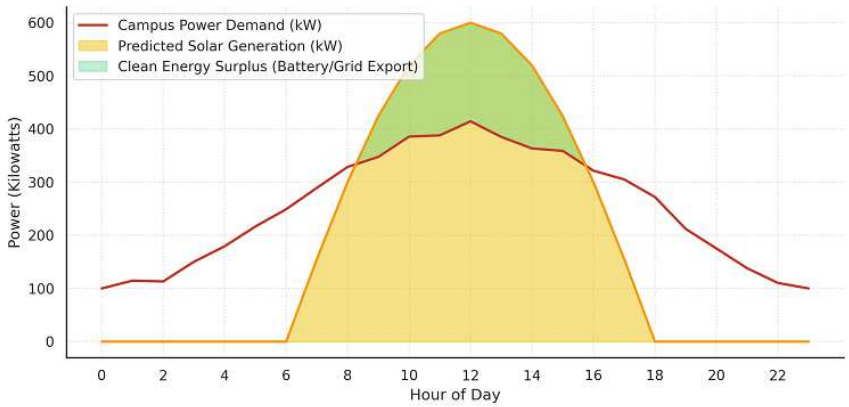


Figure 8: Daily Load Profile (Solar generation vs. Campus Energy Demand)

A primary technical hurdle in urban energy planning has been the financial and computational barrier of proprietary spatial data acquisition, such as LiDAR scans and BIM models (Tian et al., 2021). The successful compilation of the interactive 3D WebGL digital twin using OpenStreetMap (OSM) footprints and procedural building heights proves that digital twins can be built at zero data acquisition cost.



Figure 9: Solar Digital Twin Dashboard

The temporal predictive module indicates the need to avoid naive machine learning models for renewable energy planning. By precalculating deterministic physical constraints the LightGBM GOSS algorithm factored physical certainty. This prevented the overestimation of yields during peak thermal hours.

Standard global calculators (e.g., PVWatts) fail to capture regional weather anomalies (Zolkaffly et al., 2026). The inclusion of the seasonal Harmattan dust soiling penalty (peaking between December and February) addressed a major blind spot in solar yield estimation. OpenStreetMap datasets in developing regions frequently contain unlabelled building polygons (Herfort et al., 2023). The implementation of the nearest

neighbour spatial algorithm resolved this by indexing structures against local road networks and named landmarks. In regions where national grid unreliability forces heavy reliance on expensive diesel generation (Olowosejeje et al., 2019), the 25 year cumulative savings projection (Figure 7) combined with carbon offset tracking (CO_2 metric tons) offers financial and environmental justification.

The framework offers a practical methodology but there are limitations that should be noted: OpenStreetMap footprints provide spatial shapes and estimated heights, but they omit data regarding structural integrity and roof age. Deploying heavy or clamped PV racking systems on aging campus buildings requires audit.

The meteorological pipeline relies on ERA5 historical data and it is supported by a synthetic data fallback, but real time operational forecasting would require direct on site weather station integration. Distribution network constraints is outside the current scope of the project and adding hundreds of kilowatts into legacy campus transformers without assessing conductor capacities risks overloads or protection relay trips. Financial parameters were calibrated to current commercial energy market conditions in the case study region only.

Conclusion

The research contributes a physics informed machine learning solution which merged spatial mapping (where to place infrastructure) and temporal forecasting (when energy is generated). It fed building level 3D shading factors directly into an 8,760 hour temporal yield forecasting engine. It demonstrated that high accuracy, 3D energy digital twins can be constructed entirely from open source data and procedural geometry.

To establish architectural clarity, the digital twin prototype in this research is formally classified as a Level 1.5 Physics Informed Predictive Digital Model on the maturity spectrum (operating above static digital models but operates independently of live SCADA control systems).

We also generated an automated spatial directory complete with coordinates and nearest streets in the project. This minimises installation guesswork in practical scenarios, and reduces soft labour costs.

This project offers direct, practical economic value for institutional stakeholders and investors in regions facing persistent grid unreliability. The digital twin provides an economic justification for transitioning to solar micro grids, and reduces institutional operational expenditure.

Data Availability

The datasets used in this study are available at the following URL:

BaseMap Dataset

- https://kaggle.com/datasets/42bc5b92f9b4f98eeb6f1a7eb2754deaaf01f65260112f2b65d9844214098842

Solar Yield Prediction Dataset

- https://kaggle.com/datasets/42bc5b92f9b4f98eeb6f1a7eb2754deaaf01f65260112f2b65d9844214098842

References

- Abolhassani, S. S., Sedaghat, A., Parsaee, M., Sailor, D. J., & Al-Ghamdi, S. G. (2026). Urban energy–microclimate digital twin: Census-tract validation using mobile traverses and utility billing benchmarking. E3S Web of Conferences, 716, 4038. https://doi.org/10.1051/e3sconf/202671604038
- Albert, I., Nathaniel, B., & Olonilebi, P. (2025). Investigation of Building Information Modelling (BIM) in sustainable construction projects. *FUDMA Journal of Sciences*, *9*(2), 133–139. https://doi.org/10.33003/fjs-2025-0902-3121
- Alves, J. L., Palha, R. P., & de Almeida Filho, A. T. (2025). BIM-Based Framework for Photovoltaic Systems: Advancing Technologies, Overcoming Challenges, and Enhancing Sustainable Building Performance. Sustainability, 17(8), 3695. https://doi.org/10.3390/su17083695
- Azari, P., Li, S., Shaker, A., & Sattar, S. (2025). Georeferencing Building Information Models for BIM/GIS Integration: A Review of Methods and Tools. ISPRS International Journal of Geo-Information, 14(5), 180. https://doi.org/10.3390/ijgi14050180
- Bae, J.-H., Kim, D.-Y., Shin, J.-W., Lee, S.-E., & Kim, K.-C. (2020). Analysis on the features of NOCT and NMOT tests with photovoltaic module. IEEE Access, 8, 151546–151554. doi: 10.1109/ACCESS.2020.3017372
- Bhatti, Z., & Jamro, D. A. (2025). Renewable and alternative energy sources offer sustainable solutions to address the growing demand for power while mitigating environmental impact. Bulletin of Toraighyrov University. Physics & Mathematics Series. doi: 10.48081/OCZT8757
- Chu, Y., Lin, G., Deng, M., Xue, L., Li, W., Shin, H. H., Zhang, J. A., Guo, H., & Wang, Z. (2026). Thermodynamics-guided machine learning model for predicting convective boundary layer height and its multi-site applicability. Atmospheric Chemistry and Physics, 26(2), 1415–1434. doi: 10.5194/acp-26-1415-2026
- Haruna, T., Ndem Ikot, A., & Big-Alabo, A. (2024). A Review of Factors Affecting the Efficiency and Output of PV Systems Applied in the Tropical Climate of South-South Nigeria. African Journal of Advances in Science and Technology Research, 16(1), 01-17. https://doi.org/10.62154/ajastr.2024.016.010328
- Herfort, B., Lautenbach, S., Porto de Albuquerque, J. et al. A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap. Nature Communications 14, 3985 (2023). https://doi.org/10.1038/s41467-023-39698-6
- Hooli M. (2025) Digital Twins and Urban Planning: Designing Smarter, More Inclusive Cities, European Journal of Computer Science and Information Technology, 13 (44), 104-112. doi: https://doi.org/10.37745/ejcsit.2013/vol13n44104112
- Islam, M. I., Jadin, M. S. B., Al Mansur, A., & Alharbi, T. (2024). Electrical performance and degradation analysis of field-aged PV modules in tropical climates: A comparative experimental study. Energy Conversion and Management: X, 24, Article 100719. https://doi.org/10.1016/j.ecmx.2024.100719
- Jordan, D. C., Kurtz, S. R., VanSant, K., & Newmiller, J. (2016). Compendium of photovoltaic degradation rates. Progress in Photovoltaics: Research and Applications, 24(7), 978–989 https://doi.org/10.1002/pip.2744
- Karunaratne, T., Ajiero, I. R., Joseph, R., Farr, E., & Piroozfar, P. (2025). Evaluating the Economic Impact of Digital Twinning in the AEC Industry: A Systematic Review. Buildings, 15(14), 2583. https://doi.org/10.3390/buildings15142583

- Kavlak, G., Klemun, M. M., Kamat, A. S., Smith, B. L., Margolis, R. M., & Trancik, J. E. (2025). Nature of innovations affecting photovoltaic system costs. PLOS ONE, 20(8), Article e0320676. https://doi.org/10.1371/journal.pone.0320676
- Kazaure, J. S., Ibrahim, J. ., & Mohammed, A. (2024). Effect of Environmental Dust on the Performance of Photovoltaic Panels in Northern Nigeria. HAFED POLY Journal of Science, Management and Technology, 5(2), 221–231. https://doi.org/10.4314/hpjsmt.v5i2.16
- Koirala, B., Cai, H., Khayatian, F., Munoz, E., An, J. G., Mutschler, R., Sulzer, M., De Wolf, C., & Orehounig, K. (2024). Digitalization of urban multi-energy systems – Advances in digital twin applications across life-cycle phases. Advances in Applied Energy, 16, Article 100196. https://doi.org/10.1016/j.adapen.2024.100196
- Kolečanský, Š., Hofierka, J., Bogľarský, J., & Šupinský, J. (2021). Comparing 2D and 3D Solar Radiation Modeling in Urban Areas. Energies, 14(24), 8364. https://doi.org/10.3390/en14248364
- Münzinger, M., Willenborg, B., Kolbe, T. H., Burghardt, D., & Behnisch, M. (2026). Citywide assessment of urban tree shading on roof and facade solar radiation using semantic 3D city models. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XII-4/W1-2026, 243–250. https://doi.org/10.5194/isprs-annals-XII-4-W1-2026-243-2026
- Muqeet, H. A., Munir, H. M., Javed, H., Shahzad, M., Jamil, M., & Guerrero, J. M. (2021). An Energy Management System of Campus Microgrids: State-of-the-Art and Future Challenges. Energies, 14(20), 6525. https://doi.org/10.3390/en14206525
- Obada, D. O., Muhammad, M., Tajiri, S. B., Kekung, M. O., Abolade, S. A., Akinpelu, S. B., & Akande, A. (2024). A review of renewable energy resources in Nigeria for climate change mitigation. Case Studies in Chemical and Environmental Engineering, 9, Article 100669. https://doi.org/10.1016/j.cscee.2024.100669
- Olowosejeje, S., Leahy, P., & Morrison, A. (2019). The economic cost of unreliable grid power in Nigeria. African Journal of Science, Technology, Innovation and Development, 11(2), 149–159. https://doi.org/10.1080/20421338.2018.1550931
- OpenStreetMap (2021), url:https://www.openstreetmap.org
- Oyshei, K. R., Hasan, K. M. S., Sadat, N., & Hoque, M. A. (2024). Comparative analysis of solar module configuration and tracking systems for enhanced energy generation in South Sakucia Union, Bhola, Bangladesh: A software based analysis. Heliyon, 10(13), Article e33884. https://doi.org/10.1016/j.heliyon.2024.e33884.
- Rajaperumal, T. A., & Chinnappan, C. C. (2026). Physics-constrained neural networks for reliable solar power forecasting with uncertainty quantification. Results in Engineering, 29, Article 109521,https://doi.org/10.1016/j.rineng.2026.109521
- Rashidi, A., Sarvari, H., Chan, D.W., Olawumi, T.O., & Edwards, D.J. (2026), "A systematic taxonomic review of the application of BIM and digital twins technologies in the construction industry". Engineering, Construction and Architectural Management, Vol. 33 No. 3 pp. 1813–1835, doi: https://doi.org/10.1108/ECAM-02-2024-0239
- Schmidt, M., & Müller, J. (2018). Spatiotemporal modelling for integrated spatial and energy planning. Environmental and Spatial Science Reviews, 12(4), 112–125. doi: 10.1186/s13705-018-0174-z
- Shen, T., Kong, F., Yuan, S., Wang, X., Sun, D., & Ren, Z. (2025). From 2D to 3D Urban Analysis: An Adaptive Urban Zoning Framework That Takes Building Height into Account. Buildings, 15(7), 1182. https://doi.org/10.3390/buildings15071182
- Sukma, A. I., Koeva, M. N., Reckien, D., Bockarjova, M., da Silva Mano, A., Canili, G., Vicentini, G., & Kerle, N. (2024). 3D city digital twin simulation to mitigate heat risk of urban heat

islands. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-4/W11-2024, *129–136* doi: 10.5194/isprs-archives-XLVIII-4-W11-2024-129-2024

- Tian, A., Zünd, D., & Bettencourt, L. M. A. (2021). Estimating rooftop solar potential in urban environments: A generalized approach and assessment of the Galápagos Islands. Frontiers in Sustainable Cities, 3, Article 632109. doi: 10.3389/frsc.2021.632109
- Vargas, J., Martínez, R., & Loo, L. (2024, November). Enhancing photovoltaic energy forecasting with machine learning: A comparison study of XGBoost, LightGBM and LSTM models. 2024 IEEE Latin American Conference on Computational Intelligence (LA-CCI), 1–6. DOI: 10.1109/LA-CCI62337.2024.10814763
- Verma, Y., Heinonen, M., & Garg, V. (2024). ClimODE: Climate and weather forecasting with physics-informed neural ODEs. https://doi.org/10.48550/arXiv.2404.10024
- Wang, H., Chen, B., & Gong, P. (2026). High-resolution 3D solar potential assessment reveals significant shading-driven overestimation in dense urban areas. Landscape and Urban Planning, 265, Article 105521. https://doi.org/10.1016/j.landurbplan.2025.105516
- Zhong, Q., & Tong, D. (2022). Assessing Solar Photovoltaic Potential Using LiDAR and GIS Modeling. In Handbook of Climate Change Mitigation and Adaptation: Third Edition (Vol. 2, pp. 1285-1312). Springer International Publishing. https://doi.org/10.1007/978-3-030-72579-2_114
- Zolkaffly, M. Z., Anuar, N., Noor, S. Z. M., & Mohamed, H. (2026). IOP Conference Series: Earth and Environmental Science, 1587(1), Article 012009. doi:10.1088/1755-1315/1587/1/012009