SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

A Spatially Validated AI Framework for Predicting Optimal Solar-PV Tilt Angles Across Egypt

Mohammed Al shareef[1], M.M. El-Halwany[2,3], Islam ismael[4], Abdelhady Ghanem[4], Ramadan Madi Ali Bakir[5]

[1] Mechanical Engineering Department, Lecturer in the Mechanical Engineering Department, Najran College of Technology, 61411, KSA.
[2] Department of Mathematics and Physics Engineering, Faculty of Engineering, Mansoura University, Elgomhouria St, El-Mansoura 35516, Egypt.
[3] Faculty of Engineering, Mansoura National University, Gamsa, Egypt
[4] Department of Electrical Engineering, Faculty of Engineering, Mansoura University, Elgomhouria St, El-Mansoura 35516, Egypt.
[5] Semiology Department, National Research Institute of Astronomy and Geophysics (NRIAG), Cairo 4037101, Egypt
*Corresponding author: Ramadan Madi Ali Bakir, Email: ahr_bakir@yahoo.com

Abstract
Selecting an appropriate photovoltaic (PV) tilt angle is essential for maximizing energy production from fixed-tilt systems. However, the conventional assumption that the optimum tilt equals the site latitude does not adequately represent the spatial and climatic variability of solar resources across Egypt. This study develops a physics-informed spatial machine-learning framework for predicting optimal fixed-PV tilt angles at annual, biannual, seasonal, and monthly time scales. A national ground-truth dataset was generated for 1,487 locations distributed over Egypt using a 0.25°spatial grid. Hourly PVGIS-SARAH3 irradiance and meteorological data were processed for each site, and a physics-based PV engine evaluated 91 candidate tilt angles from 0° to 90° using Perez irradiance transposition. Geographic, irradiance, temperature, wind, and spatial-climatic features were then used to train and compare Linear Regression, Random Forest, Gradient Boosting, and HistGradientBoosting models against the conventional latitude rule. Model evaluation employed spatial block cross-validation and energy-based metrics, including Energy Regret. The final framework was further tested on 12 geographically unseen Egyptian sites. The physics-derived annual optimum ranged from 20° to 30°, with a mean of 26.06°. Linear Regression performed best for most prediction targets, while HistGradientBoosting provided improved performance for selected seasonal and monthly targets. External validation achieved a mean annual absolute error of 0.623°, mean multi-target MAE of 0.879° , and mean RMSE of 1.09°. Most importantly, the mean annual Energy Regret was only 0.00613%, remaining below 0.01%. A GUI deployment tool was also developed to provide location-specific tilt recommendations, energy estimates, schedules, and panel-orientation visualization. The results demonstrate that the proposed framework can support accurate and practical fixed-tilt PV deployment across Egypt.

Keywords: photovoltaic systems; optimal tilt angle; physics-informed machine learning; spatial cross-validation; Energy Regret; PVGIS; Egypt; fixed-tilt optimization.

1. Introduction

The rapid expansion of photovoltaic (PV) energy has increased the need for accurate and practical design methods capable of maximizing energy production under different geographic and climatic conditions. Among the main design parameters of a PV system, the module tilt angle has a significant influence on the incident solar irradiance, plane-of-array energy, and final electrical yield. Even a relatively small deviation from the optimum angle may result in a measurable annual energy loss, particularly in locations characterized by strong seasonal variations in solar position and solar radiation.

The conventional design practice commonly uses the site latitude as an approximation of the optimum annual tilt angle. Although this rule is simple and useful for preliminary engineering calculations, it does not necessarily maximize the energy yield at all locations. The optimum tilt depends not only on latitude, but also on longitude, direct and diffuse irradiance, atmospheric conditions, temperature, wind speed, surface reflection, and the selected optimization period. Previous investigations in Egypt have demonstrated that the optimum angle changes significantly between monthly, seasonal, half-year, and annual operating conditions [1].

Egypt provides an important case study for the development of a nationwide PV-tilt optimization framework. Its large geographic extent includes the Mediterranean coast, the Nile Delta, Greater Cairo, Upper Egypt, the Eastern Desert, the Western Desert, and the Sinai Peninsula. These regions differ in solar-resource availability, atmospheric conditions, temperature profiles, and seasonal irradiance distributions. Therefore, applying one fixed national rule or simply equating the tilt angle to latitude may lead to suboptimal designs in several parts of the country [1], [3].

Solar-tracking systems can reduce the angular mismatch between the PV surface and the incoming solar radiation by continuously changing the module orientation. However, tracking systems require additional mechanical structures, actuators, control systems, maintenance, and installation costs. Their practical and economic benefits therefore depend on system scale, local solar conditions, land availability, and maintenance capabilities. Previous studies have shown that tracking can increase energy production compared with fixed-tilt systems, but the additional cost and complexity may limit its suitability in many applications [2]. Consequently, optimized fixed-tilt systems remain attractive for residential, agricultural, remote, and utility-scale installations.

Another limitation of conventional PV design is that it generally produces only one annual tilt angle. In practice, several operating schedules may be considered, including one annual angle, two biannual angles, four seasonal angles, or twelve monthly angles. These schedules represent different compromises between energy gain, mechanical adjustment, installation complexity, and operational cost. Therefore, the tilt-angle problem should be formulated as a multi-target optimization task rather than as the estimation of a single annual value.

The availability of high-resolution solar and meteorological datasets provides an opportunity to construct physically based training data for machine-learning models. The Photovoltaic Geographical Information System (PVGIS) provides geographic photovoltaic and solar-radiation information for locations worldwide [3]. Satellite-based climate data records such as SARAH-3 provide long-term surface solar-radiation variables suitable for irradiance and PV-performance analysis [4]. In addition, the pvlib Python framework provides reproducible methods for solar-position calculation, irradiance transposition, plane-of-array irradiance estimation, and photovoltaic energy modeling [5].

A physics-informed machine-learning approach is particularly appropriate for this problem. In such an approach, a physical PV model is first used to calculate the energy produced at multiple candidate tilt angles. The angle that maximizes the simulated energy is then considered the ground-truth optimum. Machine-learning models can subsequently learn the relationship between geographic, meteorological, and solar-resource features and the physics-based optimum angle. This approach combines the physical consistency of irradiance modeling with the computational efficiency of machine-learning prediction.

The spatial structure of the problem also requires special consideration during model evaluation. Neighboring locations generally have similar climate and solar-resource characteristics. If a dataset is divided randomly, geographically close sites may appear in both the training and test subsets, resulting in information leakage and overly optimistic performance estimates. Spatially structured validation is therefore more appropriate for evaluating the ability of a model to generalize to geographically separated areas [6].

Despite the availability of previous studies on PV tilt-angle optimization, an integrated framework covering the entire territory of Egypt is still lacking. Existing studies have mainly focused on individual cities, limited regions, or conventional analytical equations. A unified system that combines a nationwide physics-based dataset, multiple machine-learning models, spatial block cross-validation, energy-based evaluation, external validation on unseen locations, and a practical deployment interface has not been sufficiently developed.

Accordingly, this study proposes a physics-informed artificial-intelligence framework for predicting optimum fixed PV tilt angles at arbitrary locations in Egypt. The proposed framework generates a regular geographic grid covering Egypt and evaluates the energy response of candidate tilt angles from 0∘0^\circ0∘ to 90∘90^\circ90∘. It produces annual, biannual, seasonal, and monthly optimum-tilt targets and compares several machine-learning models with the conventional latitude rule.

The main contributions of this study are as follows:

1. Development of a nationwide physics-based ground-truth dataset for PV tilt optimization across Egypt.
2. Generation of annual, biannual, seasonal, and monthly optimum tilt-angle targets.
3. Training and comparison of Linear Regression, Random Forest, Gradient Boosting, and HistGradientBoosting models against the conventional latitude rule.
4. Use of spatial block cross-validation to provide a realistic assessment of geographic generalization.
5. Evaluation using MAE, RMSE, and energy regret rather than angular error alone.
6. External validation using geographically unseen locations distributed across Egypt.
7. Development of a graphical user interface that provides annual, seasonal, and monthly tilt recommendations, expected PV energy, energy regret, and panel-orientation visualization.

The proposed system is intended to support engineers, researchers, and PV-system designers in selecting fixed panel orientations that are more representative of local Egyptian conditions than a single latitude-based approximation. By integrating physical simulation, spatial machine learning, and practical deployment, the framework provides a reproducible pathway from solar-resource data to an operational PV design recommendation.

2. Related Work

2.1. Traditional PV Tilt-Angle Rules

The selection of the PV-panel tilt angle has traditionally been based on simplified geometric rules. The most widely used approximation is to set the annual fixed tilt angle equal to the site latitude. This rule provides a convenient first estimate because the latitude is related to the annual solar-altitude distribution. However, it does not explicitly account for local irradiance distribution, atmospheric conditions, ground albedo, temperature, soiling, or the selected optimization period.

Hamza presented a solar-power design study for Iraq and emphasized the importance of adapting PV-system design parameters to local climatic and geographic conditions [1]. Although the study provided a useful regional perspective, it was not designed as a nationwide machine-learning prediction framework and did not use spatial block validation or an energy-regret objective.

Manzoor *et al.* investigated optimized PV tilt angles for ten cities in Pakistan using analytical solar-radiation calculations [2]. Their results showed substantial seasonal differences between winter and summer tilt angles and demonstrated that seasonally optimized panels can outperform the latitude rule. However, the study focused on a limited number of cities and relied primarily on mathematical calculations rather than a machine-learning model capable of predicting the optimum angle at arbitrary locations.

Lin *et al.* proposed a hybrid machine-learning and PVLib framework for predicting monthly PV tilt angles across global climate zones [3]. This study is particularly relevant because it combined PV simulation with machine learning and considered monthly optimization. Nevertheless, its analysis was based on a limited set of globally distributed representative cities rather than a dense national grid covering a geographically diverse country such as Egypt.

Yadav *et al.* experimentally investigated the effects of temperature, soiling, humidity, wind, and rainfall on rooftop PV performance [4]. Their findings demonstrate that the optimum installation angle cannot be interpreted independently of environmental and operational conditions. However, the study focused on performance losses rather than developing a national tilt-angle prediction model.

Shen *et al.* developed deep-learning and explainable-artificial-intelligence methods for large-scale PV-suitability analysis in Australian cities [5]. Their work demonstrated the value of spatial deep learning and feature-importance analysis for renewable-energy planning. However, PV suitability mapping is different from predicting the optimum fixed tilt angle, and the study did not evaluate annual, seasonal, and monthly tilt targets using physics-based energy regret.

Muhamad *et al.* proposed an AI-enabled predictive-analytics framework for solar-farm performance and wind-turbine health using distributed sensor networks [6]. Their work supports the broader use of AI for renewable-energy monitoring and performance management. However, the main objective was predictive maintenance and operational analytics rather than geographic tilt-angle optimization.

For Egypt specifically, Abdelaal and El-Fergany evaluated optimum PV tilt angles with experimental verification in Suez [7]. The study confirmed that optimum angles vary according to the optimization period and that monthly and seasonal adjustment may provide higher energy yield than a single fixed angle. Its main limitation is its localized geographic scope, which motivates the development of an Egypt-wide prediction framework.

A comprehensive review by Yadav and Chandel summarized methods for tilt-angle optimization and concluded that no single analytical rule is universally optimal for all locations and operating conditions [8]. The review identified latitude-based, empirical, isotropic, anisotropic, and numerical methods, but it preceded the recent development of spatial machine learning and physics-informed surrogate models.

2.2. Solar Geometry and Irradiance Transposition

The physical foundation of PV tilt optimization is the calculation of solar position and irradiance on an inclined surface. Liu and Jordan established one of the earliest statistical relationships between direct, diffuse, and global solar radiation [9]. Their work provided the basis for estimating diffuse radiation from global horizontal radiation and remains influential in solar-resource modeling.

Hay and Davies developed a model for calculating solar radiation incident on inclined surfaces by separating beam, isotropic diffuse, and circumsolar components [10]. This approach improved the representation of anisotropic diffuse radiation compared with purely isotropic models.

Perez et al. proposed widely used models for estimating diffuse irradiance and daylight components on tilted surfaces with arbitrary orientation [11]. The Perez family of models is particularly important for PV modeling because it accounts for circumsolar brightening, horizon brightening, and sky anisotropy.

Ineichen, Perez, and Seals demonstrated that ground-reflected radiation and surface albedo can influence the total radiation received by inclined surfaces [12]. This is important for desert environments, where reflective sand and soil conditions may affect the optimum angle.

Yang examined corrections and benchmark procedures for solar radiation on inclined surfaces and highlighted that different transposition models can produce different estimates under varying sky conditions [13]. Noorian, Moradi, and Kamali compared twelve models for estimating hourly diffuse irradiance on inclined surfaces and showed that model performance depends on the climate and orientation of the surface [14].

These studies establish that the optimum PV angle is not determined by geometry alone. It depends on the irradiance transposition model, the quality of direct and diffuse radiation inputs, the temporal resolution of the data, the surface albedo, and the climate of the target location.

2.3. PVGIS-, Satellite-, and pvlib-Based Optimization

Large-scale tilt optimization requires reliable weather and irradiance data. PVGIS has become one of the most widely used platforms for estimating solar radiation and photovoltaic energy yield at geographic locations. Šúri et al. demonstrated the potential of PVGIS for assessing solar-electricity generation across European countries using spatially distributed solar-resource information [15].

The European Commission's PVGIS status report describes the continued development of PVGIS datasets, photovoltaic-performance models, and geographic solar-energy services [16]. These resources make it possible to create reproducible hourly weather inputs for large geographic areas.

The SARAH-3 satellite-based climate data record provides surface solar-radiation variables with broad spatial coverage and long-term consistency [17]. Such data are particularly useful in countries where ground-based irradiance stations are sparse or unevenly distributed.

Holmgren, Hansen, and Mikofski introduced pvlib Python as an open-source framework for modeling solar position, irradiance, PV-module behavior, and energy production [18]. Its modular structure enables researchers to combine weather data, solar geometry, irradiance transposition, temperature models, and PV-performance models in a reproducible workflow.

Haroun, Issa, and Rahme developed a machine-learning framework for predicting optimum tilt angles for monofacial and bifacial PV systems using high-resolution irradiance data from 184 U.S. locations [19]. The study tested multiple machine-learning models and demonstrated the potential for rapid large-scale tilt prediction. However, it focused on U.S. sites and considered both monofacial and bifacial systems, whereas the present study focuses on fixed monofacial PV deployment across Egypt and evaluates monthly, seasonal, biannual, and annual targets.

Zurfi, Altahir, and Ibrahim investigated tilt, orientation, and tracking scenarios across all Iraqi provinces using SAM simulations [20]. They reported that monthly optimized fixed tilts can improve energy extraction compared with latitude-based settings. However, their work did not develop a supervised machine-learning model for unseen-site prediction or use spatial block cross-validation.

Al-Abadi et al. used GIS, entropy weighting, AHP, TOPSIS, and spatial interpolation to identify suitable locations for large PV farms in Basrah, Iraq [21]. This study illustrates the value of combining geographic information and multi-criteria decision analysis, but PV-farm siting is conceptually different from predicting the optimum tilt angle at a given location.

2.4. Machine Learning for Solar-Energy Prediction

Machine learning has been increasingly used for solar-resource forecasting, PV-power prediction, system-performance assessment, and renewable-energy site selection. Ensemble methods are particularly useful because they can model nonlinear interactions among geographic, meteorological, and solar-resource variables.

Breiman introduced Random Forests as an ensemble method based on multiple randomized decision trees [22]. Random Forests are robust to nonlinear relationships and feature interactions, but their performance may be affected by spatial leakage when geographically close observations are split randomly between training and testing subsets.

Friedman introduced gradient boosting as a stage-wise additive modeling technique that combines weak learners to form a strong predictive model [23]. Gradient Boosting is suitable for nonlinear regression problems such as predicting optimum PV tilt angles from irradiance and geographic features.

Chen and Guestrin proposed XGBoost, an efficient and scalable gradient-boosting implementation that has become widely used in structured-data prediction [24]. Although the present study uses conventional Gradient Boosting and HistGradientBoosting, the XGBoost literature supports the broader use of tree-based ensemble learning for renewable-energy prediction.

Recent studies have applied machine learning to solar-power forecasting, PV suitability analysis, fault diagnosis, soiling detection, and energy-yield prediction. However, most of these studies predict energy or power directly. Fewer studies formulate the optimum tilt angle itself as the prediction target. Even fewer evaluate the predicted angle using the physical energy penalty caused by using that angle instead of the true optimum.

The current study differs from conventional PV-power forecasting because the target variable is not instantaneous power. Instead, the target is the fixed tilt angle that maximizes physically simulated annual, seasonal, biannual, or monthly energy. This formulation makes the model directly useful for PV installation design.

2.5. Spatial Validation in Geospatial Machine Learning

Spatial prediction requires validation strategies that respect geographic dependence. Roberts et al. showed that standard random cross-validation can substantially underestimate prediction error when data contain spatial, temporal, hierarchical, or phylogenetic structure [25]. Their work is directly relevant to PV tilt prediction because geographically neighboring sites generally have similar solar and climatic conditions.

Valavi et al. developed the blockCV framework for generating spatially separated and environmentally separated validation folds [26]. Their analysis demonstrated that spatial blocks can provide a more realistic estimate of a model's ability to generalize to new areas.

Meyer et al. showed that spatially autocorrelated predictors such as latitude and longitude can lead to overfitting and overly optimistic performance when ordinary random validation is used [27]. They recommended combining spatial cross-validation with careful treatment of spatial predictor variables.

These findings motivate the use of spatial block cross-validation in the present study. The Egyptian grid is divided into spatial blocks, and each fold contains geographically separated groups of sites. This design evaluates whether a model can generalize beyond nearby training locations rather than simply interpolating between highly similar neighboring sites.

2.6. Physics-Informed and Hybrid AI Models

Purely data-driven machine-learning models may provide accurate predictions within the range of the training data, but they may produce physically implausible outputs when applied to new environments. Physics-informed and hybrid approaches address this problem by incorporating physical equations, simulation outputs, conservation laws, or physical constraints into the learning process.

Raissi, Perdikaris, and Karniadakis introduced physics-informed neural networks that combine data-driven learning with governing physical equations [28]. Although the present study does not use a conventional PINN architecture, it follows the same general principle of embedding physical knowledge into the AI workflow.

Karniadakis et al. reviewed physics-informed machine learning and emphasized that hybrid models can combine noisy observations, numerical simulations, and known physical relationships [29]. This approach is especially valuable when real-world labeled data are limited or expensive.

Reichstein et al. argued that machine learning for environmental systems should be connected to process understanding rather than used only as a black-box prediction tool [30]. In the context of PV systems, this means that the predicted tilt should remain linked to solar geometry, irradiance transposition, and energy production.

Willard et al. presented a taxonomy of methods for integrating scientific knowledge with machine learning in engineering and environmental applications [31]. Their work distinguishes between sequential hybrid models, residual correction, physics-guided feature engineering, physics-based loss functions, and simulation-to-ML surrogate models.

The present study adopts a sequential physics-informed hybrid strategy. First, the physical PV engine calculates the plane-of-array energy for every candidate tilt angle. Second, the maximizing angle is extracted as the ground-truth target. Third, machine-learning models are trained using geographic and weather features to predict the ground-truth angle. Finally, the prediction is evaluated not only by angular error but also by the energy regret resulting from the predicted angle.

2.7. Research Gap

The reviewed literature reveals several important limitations.

First, traditional latitude-based rules are simple but do not adequately represent local climate and seasonal irradiance conditions. Second, many tilt-angle studies are limited to a small number of cities and do not produce a continuous national prediction surface. Third, several machine-learning studies focus on PV power or site suitability rather than directly predicting optimum fixed tilt angles. Fourth, many studies use random train-test splitting, which may produce optimistic results for spatially correlated datasets. Fifth, angular error alone does not show the practical effect of a prediction error on energy yield. Finally, few studies combine nationwide physics-based simulation, multiple tilt schedules, spatial validation, energy regret, unseen-site validation, and an operational GUI.

The proposed Egypt PV Tilt AI framework addresses these gaps by:

- covering the entire Egyptian territory using a regular geographic grid;
- generating physics-based annual, biannual, seasonal, and monthly optimum-tilt targets;

- comparing several machine-learning algorithms with the latitude rule;
- using spatial block cross-validation;
- evaluating both angular accuracy and energy regret;
- validating the final models on geographically unseen Egyptian sites; and
- providing a GUI for practical deployment.

Table 1. Summary of Related Literature

Ref.	Study scope	Main method/data	Main contribution	Main limitation relative to this study
[1]	Iraq	Solar-power design study	Regional solar-system design	No national ML model or spatial CV
[2]	Pakistan, 10 cities	Analytical solar-radiation model	Seasonal tilt optimization	Limited locations; no ML prediction
[3]	Global climate zones	Hybrid ML and PVLib	Monthly tilt prediction	Representative cities rather than dense national grid
[4]	India	Experimental temperature and soiling analysis	Quantified environmental PV losses	Does not predict optimum tilt
[5]	Australian cities	Deep learning and XAI	Spatial PV suitability analysis	Suitability mapping, not tilt prediction
[6]	Renewable-energy farms	Distributed AI analytics	Solar-farm monitoring and predictive analytics	Focuses on performance monitoring
[7]	Suez, Egypt	Simulation and experimental verification	Egypt-specific seasonal and annual angles	Single-region study
[8]	Global review	Review of tilt-angle methods	Classified analytical and numerical methods	Predates recent spatial ML developments
[9]	Solar-radiation modeling	Liu–Jordan radiation relationships	Direct/diffuse radiation estimation	General radiation model, not PV ML
[10]	Inclined surfaces	Hay–Davies transposition	Anisotropic diffuse-radiation modeling	Requires irradiance inputs and model selection
[11]	Tilted surfaces	Perez model	Circumsolar and horizon-brightening representation	Model uncertainty varies by climate
[12]	Tilted surfaces	Ground-reflected radiation analysis	Importance of albedo	Does not provide a national prediction model
[13]	Solar transposition	Inclined-surface corrections	Benchmarks for radiation models	Focuses on irradiance estimation
[14]	Multiple climates	Comparison of 12 diffuse models	Demonstrated climate-dependent model performance	No AI-based optimum-angle prediction
[15]	Europe	PVGIS solar-resource modeling	Geographic solar-electricity assessment	European scope
[16]	Global PVGIS	PVGIS status and data services	Supports reproducible solar-resource retrieval	Not an angle-prediction study
[17]	Global satellite data	SARAH-3 climate record	Long-term surface radiation data	Data source rather than optimization method
[18]	PV modeling	pvlib Python	Reproducible PV simulation framework	Does not provide ML prediction
[19]	United States	13 ML models for mono/bifacial PV	Fast tilt prediction at many sites	U.S.-specific and not spatially blocked
[20]	Iraq, 18 provinces	SAM and fixed/tracking scenarios	Compared monthly, annual, and tracking modes	No ML model or unseen-site test
[21]	Basrah, Iraq	GIS–Entropy/TOPSIS–AHP	PV-farm suitability mapping	Siting problem rather than tilt optimization
[22]	General ML	Random Forest	Robust nonlinear ensemble prediction	Not developed specifically for PV tilt
[23]	General ML	Gradient Boosting	Stage-wise nonlinear regression	Does not address spatial dependence
[24]	General ML	XGBoost	Scalable tree-based learning	General method, not PV-specific
[25]	Spatial modeling	Spatially structured CV	Demonstrated random-CV optimism	Developed outside PV applications
[26]	Spatial modeling	Block cross-validation	Spatial and environmental folds	Requires appropriate block-size selection

[27]	Spatial prediction	Spatial predictor analysis	Identified spatial overfitting	Focused on ecological applications
[28]	Physics-informed AI	PINNs	Integrates physical equations and ML	Primarily PDE-based applications
[29]	Physics-informed AI	Review of hybrid physics–ML	Taxonomy of physics-guided learning	General framework rather than PV tilt
[30]	Environmental AI	Process-aware deep learning	Links ML to physical processes	Broad Earth-system focus
[31]	Engineering AI	Scientific-knowledge/ML integration	Hybrid-model taxonomy	Does not address Egypt PV deployment directly

3. Methodology

The proposed methodology follows a physics-informed machine-learning framework for estimating the optimum fixed tilt angle of photovoltaic panels across Egypt. The complete workflow is illustrated in **Figure 1**, beginning with spatial-grid construction and weather-data acquisition, followed by physics-based energy simulation, feature generation, machine-learning training, spatial validation, and external-site testing.

3.1. Study Area and Spatial Grid

The study covers the Arab Republic of Egypt, including the mainland, Nile Delta, Western Desert, Eastern Desert, and Sinai. The national boundary was obtained from the Natural Earth administrative-boundary dataset and processed using GeoPandas.

A regular geographic grid with a spatial resolution of 25° was generated. Only grid points located inside or on the Egyptian boundary were retained. The resulting grid contained 1,487 valid locations:

$$\mathcal{G} = \{(\phi_i, \lambda_i) : (\phi_i, \lambda_i) \in \Omega_{\mathrm{Egypt}}\} \quad (1)$$

where $\emptyset_i$ and λ_i represent latitude and longitude, respectively, and Ω_{Egypt} denotes the Egyptian geographic domain.

Each grid point was assigned a unique identifier and a preliminary geographic region for subsequent spatial analysis.

3.2. Weather Data

Hourly meteorological data were obtained from the PVGIS-SARAH3 radiation database. For each grid location, a complete annual series containing 8,760 hourly records was used. The main variables were:

- Global horizontal irradiance (GHI).
- Direct normal irradiance (DNI).
- Diffuse horizontal irradiance (DHI).
- Ambient temperature.
- Wind speed.
- Wind direction and atmospheric pressure where available.

For an unseen-site validation, PVGIS Typical Meteorological Year (TMY) data were downloaded using the site latitude and longitude. These data were used to generate the same weather descriptors required by the trained models.

3.3. Physics-Based Energy Engine

The physics engine was implemented using pvlib. For every site, panel tilt angles between 0° and 90° were evaluated at 1° increments. The panel azimuth was fixed at 180°, corresponding to a south-facing configuration in the Northern Hemisphere.

The plane-of-array irradiance was calculated using the Perez transposition model, with Perez-Driesse implementation where supported:

$$G_{\mathrm{POA}}(t,\beta) = f_{\mathrm{Perez}}\left[GHI(t), DNI(t), DHI(t), \theta_z(t), \gamma_s(t), \beta, \gamma_p\right] \quad (2)$$

where β is the panel tilt, θ_z is the solar-zenith angle, $\gamma_s(t)$ is the solar azimuth, and γ_p= 180° is the panel azimuth.

The hourly plane-of-array irradiance was converted into photovoltaic output using the PVWatts DC model. The energy produced at a given tilt was calculated as:

$$E_s(\beta) = \sum_{t \in s} \frac{P_{\mathrm{DC}}(t,\beta)}{1000} \quad (3)$$

where s represents the annual, biannual, seasonal, or monthly period.

The optimum tilt for each period was then obtained by:

$$\beta_s^* = \underset{\beta\in\{0,1,\ldots,90\}}{\arg\max}\; E_s(\beta) \tag{4}$$

Thus, the physics engine generated the following ground-truth targets:
1. One annual optimum.
2. Two biannual optima.
3. Four seasonal optima.
4. Twelve monthly optima.

These results formed the physics-based ground-truth dataset used for machine-learning development.

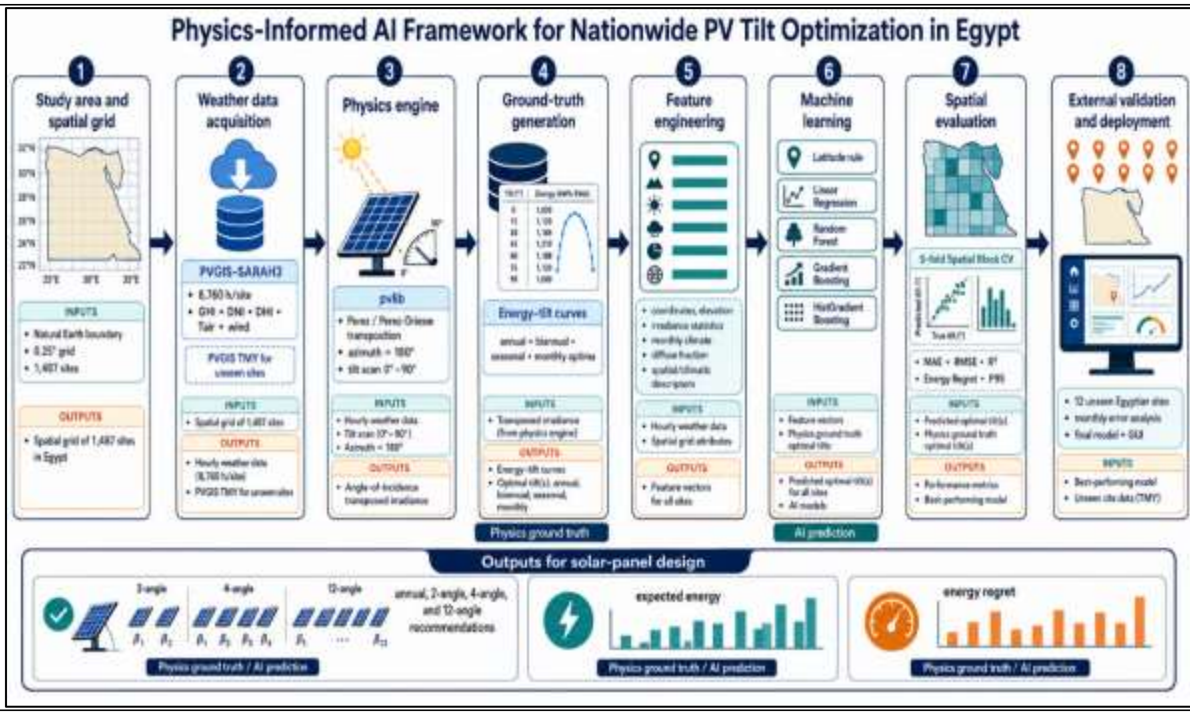


Figure. 1. End-to-end physics-informed AI methodology for nationwide photovoltaic tilt optimization in Egypt, including spatial-grid construction, PVGIS weather-data acquisition, physics-based ground-truth generation, feature engineering, machine-learning prediction, spatial cross-validation, external-site validation, and final deployment.

3.4. AI Feature Engineering

For every site, a feature vector was constructed from geographical, meteorological, irradiance, and spatial descriptors:

$$\mathbf{x}_i = [\phi_i, \lambda_i, h_i, \mathbf{r}_i, \mathbf{c}_i, \mathbf{s}_i] \tag{5}$$

where:
- $\emptyset_i, \lambda_i$ are latitude and longitude.
- h_i is elevation.
- r_i contains annual and monthly irradiance statistics.
- c_i contains temperature and wind descriptors.
- s_i contains spatial and climatic descriptors.

The engineered features included:
- Latitude, longitude, and elevation.
- Mean, standard deviation, minimum, and maximum irradiance.
- Annual and monthly GHI, DNI, and DHI.
- Temperature statistics.
- Wind-speed statistics.
- Diffuse-fraction indicators.
- Squared geographic coordinates.
- Spatial-block and regional descriptors.

The target variables consisted of 19 tilt values: annual, two biannual, four seasonal, and twelve-monthly targets.

3.5. Machine-Learning Models

Five prediction approaches were compared.

3.5.1. Latitude Rule

The conventional baseline was defined as:

$$\beta_{lat} = |\phi| \tag{6}$$

This rule was used as a reference rather than as a trained model.

3.5.2. Linear Regression

Linear regression provides an interpretable relationship between the engineered features and the optimum tilt.

$$\hat{y} = \beta_0 + \sum_{j=1}^{p} \beta_j x_j \tag{7}$$

3.5.3. Random Forest

Random Forest uses an ensemble of decision trees, with the final prediction calculated as:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^{B} T_b(\mathbf{x}) \tag{8}$$

where B is the number of trees.

3.5.4. Gradient Boosting

Gradient Boosting constructs an additive model of sequential weak learners:

$$\hat{y}(\mathbf{x}) = \sum_{m=1}^{M} \eta h_m(\mathbf{x}) \tag{9}$$

where M is the number of boosting stages and η is the learning rate.

3.5.5. HistGradientBoosting

HistGradientBoosting was also evaluated because its histogram-based implementation provides efficient nonlinear regression for large feature sets while retaining strong predictive performance.

3.6. Spatial Block Cross-Validation

Random train-test splitting was avoided because geographically close sites may have highly similar climate and solar conditions. Random splitting could therefore produce overly optimistic results.

Instead, the Egyptian grid was divided into spatial blocks. The block identifier was used as the grouping variable in a five-fold GroupKFold procedure. In each fold, complete spatial blocks were assigned exclusively to either the training or testing subset.

This procedure evaluates whether a model can generalize to geographically separated locations rather than merely interpolating between nearby sites.

3.7. Evaluation Metrics

The models were evaluated using angular, statistical, energetic, and external-validation metrics.

Mean Absolute Error

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{10}$$

Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{11}$$

Coefficient of Determination

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{12}$$

Energy Regret

Angular error alone does not fully represent engineering impact. Therefore, each predicted tilt was evaluated using the physics-based energy curve:

$$\text{Energy Regret}(\%) = \frac{E_{\text{optimal}} - E_{\text{predicted}}}{E_{\text{optimal}}} \times 100 \tag{13}$$

where $E_{optional}$ is the maximum physics-based energy and $E_{predicted}$ is the energy obtained at the model-predicted tilt.

The 95th-percentile energy regret was also calculated to identify high-loss cases:

$$P95_{\text{regret}} = \text{Percentile}_{95}(\text{Energy Regret}_1, \ldots, \text{Energy Regret}_n) \tag{14}$$

Monthly prediction errors were calculated separately for all twelve months. Finally, the selected models were evaluated on 12 geographically distributed unseen Egyptian sites using PVGIS TMY data. This external validation assessed the practical ability of the framework to predict optimum tilts for locations not included in the original grid.

4. Results

4.1. Ground-Truth Dataset

The physics-based simulation produced a complete ground-truth dataset covering the whole Egyptian territory. As described in **Table 2**, the final spatial grid contained **1,487 valid locations** distributed at a spatial resolution of 0.25°. For each location, the physics engine evaluated 91 candidate tilt angles ranging from 0° to 90°at 1° intervals.

Energy–Tilt records. Each record represents the simulated photovoltaic energy corresponding to one site–tilt combination. All locations contained exactly 91 tilt values, with no incomplete curves or missing annual optimum values.

The annual optimum tilt was extracted as the angle that maximized the simulated annual PV energy:

$$\beta^{*}_{\text{annual}} = \underset{\beta \in \{0,\ldots,90\}}{\arg\max}\ E_{\text{annual}}(\beta) \qquad (15)$$

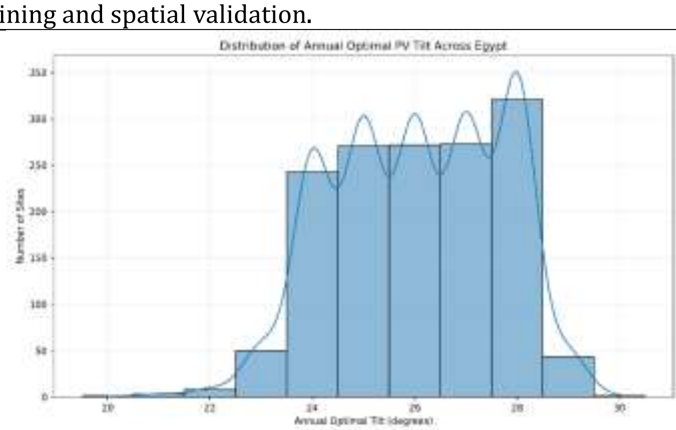
As described in **Figure 2(a)** The resulting annual optimum angles ranged from 20° to 30°, with a mean value of 26.06° and a standard deviation of 1.62°∘. The median value was 26°, while the first and third quartiles were 25°∘ and 27°, respectively.

Table 2. Summary of the physics-based ground-truth dataset

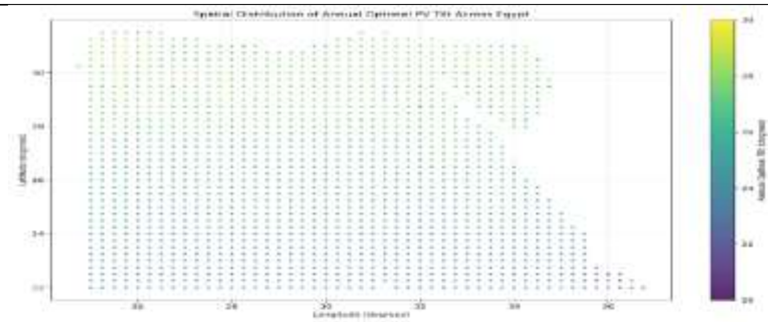
Quantity	Value
Geographic domain	Egypt
Grid resolution	0.25°
Number of sites	1,487
Candidate tilt range	0°–90°
Tilt resolution	1°
Tilts per site	91
Total Energy–Tilt records	135,317
Annual optimum tilt minimum	20°
Annual optimum tilt maximum	30°
Mean annual optimum tilt	20.06°
Median annual optimum tilt	26°
Standard deviation	1.62°
Incomplete site curves	0

The relatively narrow annual tilt range indicates a coherent national pattern, while the spatial variation between

20° and 30° confirms that a single fixed rule cannot fully represent all Egyptian locations. The distribution also demonstrates that the physics engine generated physically consistent and spatially continuous targets suitable for machine-learning training and spatial validation.



(a)



(b)

Figure. 2. Distribution and spatial variation of the physics-derived annual optimum photovoltaic tilt across Egypt: (a) histogram of annual optimum tilt values for 1,487 grid sites, and (b) geographic distribution of the optimum tilt angles over the 0.25°spatial grid.

The spatial distribution of the physics-derived annual optimum tilt is presented in **Figure 2(b)**. The results demonstrate that the optimum angle is not spatially constant across Egypt. Instead, it varies continuously across the national grid, with annual optimum values ranging from 20° to 30°.

A general increase in the optimum tilt can be observed toward the northern parts of Egypt, while lower values are more common in the southern regions. However, this variation is not determined by latitude alone. Differences in irradiance composition, seasonal solar-resource distribution, temperature, and atmospheric conditions produce local deviations from a simple north–south trend.

The results also show that the conventional latitude rule cannot be assumed to provide the same accuracy throughout Egypt. For a given location, the latitude-based estimate is calculated as:

$$\beta_{\text{latitude},i} = |\phi_i| \tag{16}$$

whereas the physics-derived optimum is:

$$\beta^*_{\text{physics},i} = \arg\max_{\beta} E_i(\beta) \tag{17}$$

The location-specific difference can therefore be expressed as:

$$\Delta\beta_i = \beta^*_{\text{physics},i} - \beta_{\text{latitude},i} \tag{18}$$

This spatial difference confirms that the latitude rule is useful as a preliminary design approximation, but it does not fully capture the climatic and irradiance-dependent behavior of fixed photovoltaic systems across Egypt. Therefore, a location-aware prediction model is required to provide more precise annual, seasonal, and monthly tilt recommendations.

4.2. Model Comparison

The five prediction approaches were compared using the spatial cross-validation results for the annual, biannual, seasonal, and monthly tilt targets. The comparison was based primarily on the mean MAE across the targets within each schedule group. Lower values indicate more accurate predictions.

Table 3. Comparison of models according to prediction schedule

Prediction group	Linear Regression	Random Forest	Gradient Boosting	HistGradientBoosting	Latitude Rule	Best model
Annual	**0.27°**	0.35°	0.34°	0.33°	1.16°	Linear Regression
Biannual	**0.27°**	0.28°	0.29°	0.28°	19.78°	Linear Regression
Seasonal	**0.26°**	0.28°	0.27°	0.26°	18.53°	Linear Regression
Monthly	0.25°	0.25°	0.26°	**0.25°**	18.86°	HistGradientBoosting

Linear Regression provided the best overall performance for the annual and biannual targets and remained highly competitive for the seasonal and monthly targets. Its performance indicates that the relationship between the engineered geographic–climatic features and the optimum tilt is largely smooth and structured across Egypt.

HistGradientBoosting achieved the best aggregate performance for the monthly targets and was also highly competitive for selected seasonal targets. This advantage is attributed to its ability to represent nonlinear interactions between monthly irradiance, temperature, wind, and geographic variables.

Random Forest and conventional Gradient Boosting produced stable but slightly higher errors than Linear Regression and HistGradientBoosting. Their performance remained substantially better than the latitude-based baseline.
The Latitude Rule produced the weakest results, particularly for seasonal and monthly predictions. Its mean MAE exceeded 18° for both groups, demonstrating that the simple relationship β=|ϕ| does not adequately represent the temporal variation of the optimal fixed tilt. This confirms that the optimum angle depends not only on latitude but also on the seasonal distribution of solar irradiance and local climatic conditions.
Accordingly, Linear Regression was selected for most annual, biannual, seasonal, and monthly targets, while HistGradientBoosting was selected for the specific targets where it achieved lower spatial-validation error. This target-specific selection strategy was retained in the final model package.

4.3. Feature Importance Analysis

Feature importance was evaluated using stable permutation importance rather than the original model-dependent importance values. For each spatial cross-validation fold, one feature was randomly permuted while all other features were kept unchanged. The increase in prediction error was then calculated as:

$$PI_j = \frac{1}{K}\sum_{k=1}^{K}\left(MAE_{j,k}^{\text{permuted}} - MAE_{k}^{\text{baseline}}\right) \quad (19)$$

where PI_j is the importance of feature j, K is the number of spatial folds, and $MAE_{j,k}^{permuted}$ represents the error after permuting that feature in fold k.
The stable permutation analysis identified the following groups as the most influential predictors:

1. Mean GHI.
2. Annual DNI.
3. Annual DHI.
4. Mean and seasonal temperature.
5. Mean and seasonal wind speed.
6. Latitude and absolute latitude.
7. Monthly irradiance variables, particularly monthly GHI and DNI.

These results indicate that the optimum tilt is primarily controlled by the amount and composition of incident solar radiation. GHI represents the overall available solar resource, while DNI and DHI describe the direct and diffuse components that determine the angular response of the PV surface. Temperature and wind also contribute indirectly through their influence on module operating temperature and energy conversion efficiency. Latitude remains important because it represents the underlying solar-geometry effect, but its influence is not sufficient to explain the complete spatial and temporal variation.
Monthly irradiance variables were particularly relevant for seasonal and monthly tilt prediction, confirming that short-term tilt recommendations require climate descriptors that preserve the temporal structure of the solar resource.
The stable permutation-importance plot was used as the final feature-importance figure in the manuscript. The earlier importance plot containing extremely large, poorly interpretable values was excluded because its scale was affected by the preprocessing representation and did not provide a reliable physical interpretation.

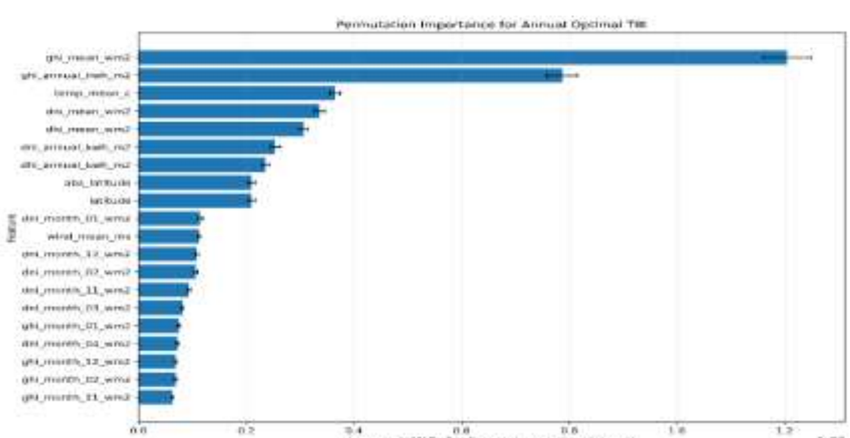


Figure. 3. Stable permutation importance of the main geographic and climatic features for annual optimum photovoltaic tilt prediction. Importance values represent the increase the increase in spatial-cross-validation MAE after feature permutation, averaged across the validation folds.

4.4. Energy-Based Evaluation

Angular prediction error alone does not fully describe the engineering performance of a photovoltaic tilt model. A small angular deviation may have almost no practical effect when the Energy–Tilt curve is relatively flat near its maximum. Therefore, every AI-predicted tilt was re-evaluated using the physics-based energy curve, and the resulting energy was compared with the maximum physically achievable energy.

For the 12 geographically distributed unseen Egyptian sites, the external-validation results were:

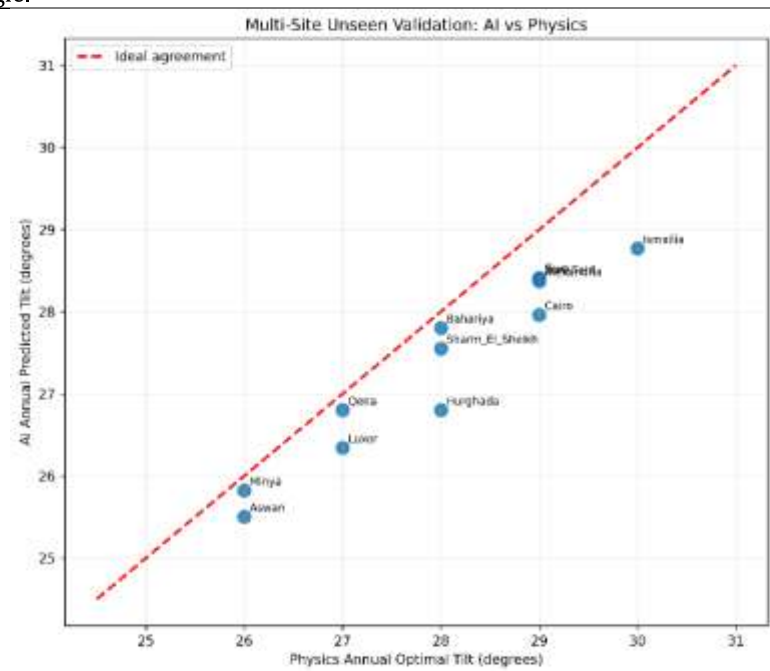
Table 4. validation results

Metric	Result
Mean annual absolute tilt error	0.623°
Mean external MAE	0.879°
Mean external RMSE	1.09°
Mean annual energy regret	0.00613%

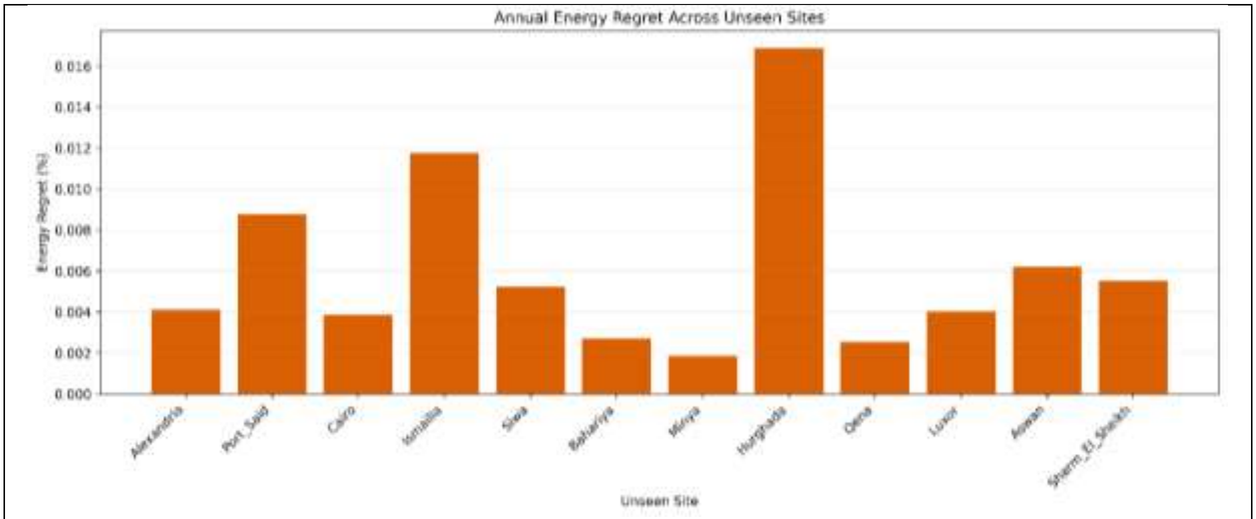
The extremely low mean energy regret demonstrates that the predicted tilts operated very close to the physics-derived optimum. Thus, even where a small angular difference existed between the AI prediction and the physics optimum, the corresponding reduction in annual energy was practically negligible.

This result is particularly important for fixed-tilt photovoltaic systems, where the main engineering objective is not necessarily to reproduce the exact optimum angle, but to achieve near-maximum annual energy without requiring mechanical tracking. The results therefore confirm that the proposed model provides both accurate tilt estimates and robust energy performance for unseen locations.

The energy-based evaluation also demonstrates the advantage of using Energy Regret as a performance metric. Unlike angular error, Energy Regret directly measures the practical consequence of installing a panel at the model-predicted angle.



(a)



(b)

Figure. 4. Energy-based external validation on unseen Egyptian sites: (a) comparison between AI-predicted and physics-derived annual optimum tilts, and (b) annual energy regret associated with the AI-predicted fixed-panel tilt at each unseen site.

4.5. GUI Deployment and Practical Applicability

To demonstrate the practical usability of the proposed Physics–AI framework, a graphical user interface (GUI) was developed as a deployment layer for the final trained models. The GUI is not considered an independent scientific model; rather, it provides an accessible interface through which users can apply the validated model to new locations in Egypt.

The user enters the site latitude, longitude, elevation, and panel azimuth. The system then retrieves the corresponding PVGIS Typical Meteorological Year (TMY) data, generates the required geographic and climatic features, and applies the final target-specific prediction models. The interface reports the annual optimum tilt, biannual and seasonal tilt schedules, and monthly tilt recommendations.

In addition, the GUI provides an Energy–Tilt curve that compares the physics-derived optimum with the AI-predicted angle. This visualization allows the user to assess the effect of the predicted tilt on the expected annual POA energy. The corresponding energy regret is also displayed as a practical performance indicator.

The deployment interface includes the following functions:

- Input of latitude, longitude, elevation, and panel azimuth.
- Automatic loading of PVGIS TMY weather data.
- Prediction of the annual fixed-panel tilt.
- Generation of biannual and seasonal tilt schedules.
- Generation of the 12-month tilt schedule.
- Visualization of the Energy–Tilt relationship.
- Calculation of expected POA energy.
- Calculation of annual energy regret.
- Visualization of panel direction and inclination relative to north and south.
- Generation of a compact site-specific prediction report.

The orientation module displays the panel azimuth measured clockwise from geographic north and expresses the corresponding rotation relative to south. A side-view diagram illustrates the recommended panel inclination, allowing the output to be interpreted directly during installation.

The GUI therefore converts the research outputs into a practical decision-support tool for fixed-tilt photovoltaic deployment. Its main value is operational: it enables engineers and users to obtain location-specific tilt recommendations without manually executing the weather-processing, physics simulation, or machine-learning steps. The scientific validity of the approach remains established by the spatial cross-validation and unseen-site validation experiments, whereas the GUI represents the final implementation and deployment layer of the proposed framework.

Figure 5. Deployment interface of the proposed Egypt PV Tilt AI system showing site inputs, annual optimum tilt, seasonal and monthly schedules, Energy–Tilt curve, energy regret, expected POA energy, and panel orientation visualization.

5. Discussion

5.1. Why Does the Optimum Differ from the Latitude Rule?

The commonly used latitude rule assumes that the optimum fixed tilt is approximately equal to the site latitude. Although this provides a useful first-order estimate, it does not explicitly account for the actual distribution of solar irradiance throughout the year. The physics-based results obtained in this study show that the optimum annual tilt across Egypt ranges from 20∘20^\circ20∘ to 30∘30^\circ30∘, with a mean value of 26.06∘26.06^\circ26.06∘, despite the wider Egyptian latitude range.

This difference is caused by several factors. First, the annual optimum is determined by the weighted contribution of direct, diffuse, and ground-reflected irradiance rather than latitude alone. Second, atmospheric conditions, cloudiness, aerosol loading, temperature, and wind influence the effective irradiance received by the module. Third, the optimum tilt depends on the temporal distribution of solar radiation. For example, winter-dominated radiation favors steeper tilts, whereas summer-dominated radiation favors lower tilts. Consequently, two sites with similar latitudes may have different optimum angles because their climatic and irradiance profiles are not identical.

The latitude rule also becomes less reliable for seasonal and monthly schedules. The spatial cross-validation results showed substantially larger errors for the latitude rule in biannual, seasonal, and monthly targets than for the machine-learning models. This confirms that latitude alone is insufficient for representing the temporal and climatic variability of solar-resource conditions across Egypt.

5.2. Why Did Linear Regression Perform Well?

Linear Regression achieved the best performance for most annual, biannual, seasonal, and monthly targets. This result is physically interpretable because the optimal tilt varies smoothly across geographic space and is strongly related to a limited number of dominant variables, particularly latitude, annual and monthly irradiance, temperature, and wind conditions.

The ground-truth targets were generated by a deterministic physics engine, and the resulting optimum angles do not exhibit highly discontinuous behavior across neighboring locations. Therefore, a relatively simple model can capture much of the underlying relationship between the geographic-climatic features and the optimal tilt. The strong performance of Linear Regression also indicates that the relationship is approximately low-dimensional over the geographical extent of Egypt.

However, the results do not imply that the relationship is perfectly linear. HistGradientBoosting provided the best performance for selected nonlinear targets, particularly some seasonal and monthly angles. These targets are more sensitive to changes in seasonal irradiance and therefore may contain threshold effects or local nonlinearities that are not fully captured by a global linear function. The appropriate model therefore depends on the prediction target: Linear Regression is highly effective for the dominant smooth relationships, whereas HistGradientBoosting is useful when stronger nonlinear behavior is present.

5.3. Why Is Energy Regret More Important Than Angular Error?

Angular error alone does not directly measure the practical effect of a prediction on photovoltaic energy production. A prediction that differs from the physics optimum by 1∘1^\circ1∘ may still produce almost the same annual energy, whereas a larger angular error may have a more noticeable effect if it occurs in a steep region of the Energy–Tilt curve.

For this reason, the study evaluates the models using Energy Regret in addition to MAE and RMSE. Energy Regret directly quantifies the percentage of theoretically optimal energy that is lost when the AI-predicted tilt. The unseen-site validation produced a mean annual absolute tilt error of 0.623, a mean multi-target MAE of 0.879, and a mean RMSE of 1.09. Nevertheless, the mean annual energy regret was only 0.00613%. This very small regret demonstrates that the predicted angles are located within the high-energy plateau surrounding the optimum.

The Energy–Tilt curves confirm that the energy response is relatively flat near the maximum. Therefore, a small angular difference, such as 1 deg, does not necessarily result in a meaningful energy loss. This is particularly important for engineering applications, where installation tolerances, structural constraints, maintenance requirements, and construction practicality may be more important than reproducing the exact optimum angle to the nearest degree.

5.4. Practical Suitability of the Proposed Framework

The results support the practical use of the proposed framework for fixed-tilt photovoltaic systems across Egypt. The model was trained on a dense spatial grid containing 1,487 locations and was evaluated using spatial block cross-validation to reduce the optimistic bias associated with random splitting of geographically correlated data. It was then tested independently on 12 unseen sites distributed across different Egyptian climatic regions.

The external validation demonstrated consistent performance across coastal, desert, Nile Valley, Red Sea, and Sinai locations. The small energy regret indicates that the model can provide technically useful tilt recommendations even when the predicted angle is not identical to the physics-derived optimum. This makes the framework suitable for preliminary PV design, site screening, regional planning, and rapid engineering recommendations.

The deployment GUI further improves practical usability by allowing users to enter latitude and longitude, retrieve PVGIS TMY data, and obtain annual, seasonal, and monthly tilt recommendations together with the Energy–Tilt curve, expected POA energy, energy regret, and panel orientation. The GUI should therefore be understood as an implementation and deployment layer for the validated model, rather than as an independent scientific method.

Overall, the results indicate that the proposed Physics–AI framework provides a useful balance between physical consistency, predictive accuracy, computational efficiency, and practical interpretability. Its main advantage is not merely predicting an angle close to the physics optimum, but identifying a fixed-panel configuration that preserves nearly all of the available photovoltaic energy while allowing rapid prediction for new locations.

Conclusion

This study developed and validated a physics-informed spatial machine-learning framework for predicting optimal fixed photovoltaic tilt angles across Egypt. A physics-based ground-truth dataset was generated for 1,487 locations using hourly irradiance and meteorological data, with tilt angles evaluated from 0 to 90. The framework produced annual, biannual, seasonal, and monthly tilt recommendations while accounting for geographic and climatic variability.

The results demonstrated that the optimum tilt angle is not adequately represented by the conventional latitude rule. Instead, the physics-derived optimum varies spatially across Egypt, with annual values ranging from 20 to 30 and a national mean of 26.06. Spatial block cross-validation showed that machine-learning models substantially outperformed the latitude rule, with Linear Regression providing the best performance for most targets and HistGradientBoosting improving selected seasonal and monthly predictions.

The key result is that angular accuracy was translated into energy-based performance. External validation on 12 geographically unseen Egyptian sites produced a mean annual absolute error of 0.623, a mean multi-target MAE of 0.879, and a mean RMSE of 1.09. More importantly, the mean annual Energy Regret was only 0.00613%, remaining below 0.01%. This confirms that the AI-predicted angles preserve nearly all of the energy obtained at the physics-derived optimum, even when small angular differences occur.

Therefore, the proposed Physics–AI framework provides an accurate, spatially validated, and computationally efficient solution for fixed-tilt photovoltaic design in Egypt. The developed GUI further demonstrates the deployment potential of the framework by allowing users to enter a location, retrieve PVGIS TMY data, obtain annual and seasonal recommendations, generate a monthly schedule, visualize the Energy–Tilt curve, calculate energy regret, and inspect panel orientation.

The main conclusion of the research is:

The proposed physics-informed spatial machine-learning framework accurately predicts optimal fixed-tilt angles across Egypt while maintaining extremely low energy regret. External validation on geographically unseen Egyptian sites confirmed a mean angular error below 1 deg and an average annual energy regret below 0.01%, demonstrating the suitability of the framework for practical fixed-tilt photovoltaic deployment.

Future work may extend the framework to additional years of weather data, uncertainty-aware predictions, site-specific shading and soiling effects, terrain-aware irradiance modeling, and deployment as a web or mobile decision-support platform.

References

[1] G. Hamza, "Solar power system solution for Iraq," Jun. 2024, doi: 10.14293/PR2199.000930.v1.

[2] H. U. Manzoor, S. M. Aaqib, T. Manzoor, F. Azeem, M. W. Ashraf, and S. Manzoor, "Effect of optimized tilt angle of PV modules on solar irradiance for residential and commercial buildings in different cities of Pakistan: Simulation-based study," *Energy Science & Engineering*, vol. 13, pp. 1831–1845, 2025, doi: 10.1002/ese3.70004.

[3] O. Z. Lin, L. Štěpanec, E. Koutroulis, D. Juchelkova, and H. Y. Aye, "Optimizing monthly solar PV tilt angles and energy yield across global climate zones: A hybrid machine learning and PVLib approach," *Renewable Energy*, Art. no. 125163, 2025, doi: 10.1016/j.renene.2025.125163.

[4] D. Yadav, S. K. Yadav, V. Varma, *et al.*, "Experimental analysis of elevated temperature and soiling loss on the rooftop PV modules performance under composite climatic conditions," *Scientific Reports*, vol. 15, Art. no. 42035, 2025, doi: 10.1038/s41598-025-25846-z.

[5] J. Shen, F. Zheng, F. B. Tesema, W. Deng, A. G. Bellotti, and J. Xie, "Deep learning and explainable artificial intelligence for large-scale photovoltaic suitability analysis in Australian cities," *Renewable Energy*, vol. 259, Art. no. 125025, 2026, doi: 10.1016/j.renene.2025.125025.

[6] N. Muhamad, X. Li, S. Imtiaz, M. R. M. Altimania, W. M. Abdelfattah, and H. M. Munir, "AI-enabled predictive analytics for wind turbine health and solar farm performance using distributed sensor networks," *IET Renewable Power Generation*, vol. 20, no. 1, Art. no. e70200, 2026, doi: 10.1049/rpg2.70200.

[7] A. K. Abdelaal and A. El-Fergany, "Estimation of optimal tilt angles for photovoltaic panels in Egypt with experimental verifications," *Scientific Reports*, vol. 13, Art. no. 3268, 2023, doi: 10.1038/s41598-023-30375-8.

[8] A. K. Yadav and S. S. Chandel, "Tilt angle optimization to maximize incident solar radiation: A review," *Renewable and Sustainable Energy Reviews*, vol. 23, pp. 503–513, 2013, doi: 10.1016/j.rser.2013.02.027.

[9] B. Y. H. Liu and R. C. Jordan, "The interrelationship and characteristic distribution of direct, diffuse and total solar radiation," *Solar Energy*, vol. 4, no. 3, pp. 1–19, 1960, doi: 10.1016/0038-092X(60)90062-1.

[10] J. E. Hay and J. A. Davies, "Calculation of the solar radiation incident on an inclined surface," in *Proc. First Canadian Solar Radiation Data Workshop*, Ottawa, Canada, 1980, pp. 59–72.

[11] R. Perez, P. Ineichen, R. Seals, J. Michalsky, and R. Stewart, "Modeling daylight availability and irradiance components from direct and global irradiance," *Solar Energy*, vol. 44, no. 5, pp. 271–289, 1990, doi: 10.1016/0038-092X(90)90055-H.

[12] P. Ineichen, R. Perez, and R. Seals, "The importance of correct albedo determination for adequately modeling energy received by tilted surfaces," *Solar Energy*, vol. 39, no. 4, pp. 301–305, 1987, doi: 10.1016/S0038-092X(87)80016-6.

[13] D. Yang, "Solar radiation on inclined surfaces: Corrections and benchmarks," *Solar Energy*, vol. 136, pp. 288–302, 2016.

[14] A. M. Noorian, I. Moradi, and G. A. Kamali, "Evaluation of 12 models to estimate hourly diffuse irradiation on inclined surfaces," *Renewable Energy*, vol. 33, no. 6, pp. 1406–1412, 2008.

[15] J. Šúri, T. A. Huld, E. D. Dunlop, and H. A. Ossenbrink, "Potential of solar electricity generation in the European Union member states and candidate countries," *Solar Energy*, vol. 81, no. 10, pp. 1295–1305, 2007, doi: 10.1016/j.solener.2006.12.007.

[16] European Commission, Joint Research Centre, "Photovoltaics geographical information system: Status report 2024," 2024, doi: 10.2760/2979595.

[17] U. Pfeifroth, J. Drücke, S. Kothe, J. Trentmann, M. Schröder, and R. Hollmann, "SARAH-3—Satellite-based climate data records of surface solar radiation," *Earth System Science Data*, vol. 16, pp. 5243–5265, 2024, doi: 10.5194/essd-16-5243-2024.

[18] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, "pvlib Python: A Python package for modeling solar energy systems," *Journal of Open Source Software*, vol. 3, no. 29, Art. no. 884, 2018, doi: 10.21105/joss.00884.

[19] H. Haroun, J. S. Issa, and P. Rahme, "Machine learning-based prediction of optimal tilt angles for monofacial and bifacial PV systems," *Solar Energy*, vol. 301, Art. no. 113924, 2025, doi: 10.1016/j.solener.2025.113924.

[20] A. Zurfi, A. A. R. Altahir, and A. Ibrahim, "Investigation of optimal tilt, orientation, and tracking of a solar PV system in Iraq," *International Journal of Power Electronics and Drive Systems*, vol. 16, no. 3, pp. 1914–1925, 2025, doi: 10.11591/ijpeds.v16.i3.pp1914-1925.

[21] A. M. Al-Abadi, A. M. Handhal, M. A. Abdulhasan, W. L. Ali, J. J. Hassan, and A. H. Al-Aboodi, "Optimal siting of large photovoltaic solar farms at Basrah governorate, southern Iraq using hybrid GIS-based Entropy-TOPSIS and AHP-TOPSIS models," *Renewable Energy*, vol. 241, Art. no. 122308, 2025, doi: 10.1016/j.renene.2024.122308.

[22] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[23] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[24] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[25] D. R. Roberts, V. Bahn, S. Ciuti, *et al.*, "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," *Ecography*, vol. 40, no. 8, pp. 913–929, 2017, doi: 10.1111/ecog.02881.

[26] R. Valavi, J. Elith, J. J. Lahoz-Monfort, and G. Guillera-Arroita, "blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models," *Methods in Ecology and Evolution*, vol. 10, no. 2, pp. 225–232, 2019, doi: 10.1111/2041-210X.13107.

[27] H. Meyer, C. Reudenbach, S. Wöllauer, and T. Nauss, "Importance of spatial predictor variable selection in machine-learning applications—Moving from data reproduction to spatial prediction," *Ecological Modelling*, vol. 411, Art. no. 108815, 2019, doi: 10.1016/j.ecolmodel.2019.108815.

[28] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019, doi: 10.1016/j.jcp.2018.10.045.

[29] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, *et al.*, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, pp. 422–440, 2021, doi: 10.1038/s42254-021-00314-5.

[30] M. Reichstein, G. Camps-Valls, B. Stevens, *et al.*, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, pp. 195–204, 2019, doi: 10.1038/s41586-019-0912-1.

[31] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar, "Integrating scientific knowledge with machine learning for engineering and environmental systems," *ACM Computing Surveys*, 2022.