https://doi.org/10.31449/inf.v50i7.11442 Informatica **50** (2026) 135–152 **135**

RuralGrid-PVO: A Reliability-Conscious Multi-Objective Optimization Framework for Distributed PV Siting in Rural Grids

Yu Zhang, Jinyue Shi*, Haitao Ge, Liang Geng, Lulu Jia
State Grid Hebei Electric Power Co., Ltd. Shijiazhuang Power Supply Branch, Shijiazhuang 056000, Hebei, China
E-mail: shijinyue68@outlook.com, zhangyu_zy0228@outlook.com, haitao_ge01@hotmail.com, gengliang_gl@outlook.com, jialulu0316@hotmail.com
*Corresponding author

Keywords: distributed photovoltaic systems, rural power networks, grid reliability assessment, multi-objective optimization, genetic algorithm, deep neural networks, random forest regression

Received: September 4, 2025

The rapid growth of distributed photovoltaic (PV) systems offers a sustainable solution to rural energy demands. However, integrating PV into rural distribution grids presents challenges related to power quality and grid reliability. Traditional PV site selection approaches often neglect the impact on distribution reliability metrics, resulting in suboptimal deployments. This paper proposes RuralGrid-PVO (Rural Grid Photovoltaic Optimization), an optimization modeling framework for PV grid connection site selection in rural areas, incorporating distribution grid reliability considerations. The RuralGrid-PVO method integrates the RTS-GMLC synthetic power system model with solar irradiance data from the NSRDB to simulate realistic rural deployment scenarios. A multi-objective optimization model combines a Genetic Algorithm (GA) with Deep Neural Networks (DNN) for fast reliability assessments, optimizing PV placement based on energy yield, voltage stability, power losses, and reliability indices (SAIDI/SAIFI). GA-DNN framework incorporates SAIDI/SAIFI into optimization, outperforming baseline GA, AHP-GIS, and ReliOpt-Hybrid in energy yield, voltage stability, and reliability improvements, though sensitivity to reliability weights requires further exploration. This paper proposes RuralGrid-PVO, a comprehensive six-module optimization pipeline for distributed photovoltaic (PV) site selection in rural grids with a focus on enhancing grid reliability. The pipeline integrates data acquisition, candidate site identification, capacity estimation via Random Forest regression ($R^2 = 0.93$), reliability-aware grid simulation utilizing a Deep Neural Network (DNN) for SAIDI/SAIFI prediction (MAE = 3.8 min/year), multi-objective optimization with a Genetic Algorithm (GA), and final configuration selection. The methodology operates on a hardware/software stack comprising Python-based ML frameworks and power system simulation tools to ensure reproducibility. Experimental evaluation demonstrates that RuralGrid-PVO improves voltage profiles by 15.3%, reduces energy losses by 12.8%, decreases SAIFI by up to 9.6%, and lowers SAIDI by 39%, significantly outperforming baseline site selection methods. These results validate the framework's effectiveness in achieving reliable, energy-efficient rural PV integration.

Povzetek: Prispevek predstavi RuralGrid-PVO, optimizacijski okvir za izbiro lokacij priklopa razpršenih PV v ruralnih omrežjih, ki poleg izplena in napetostne stabilnosti upošteva tudi zanesljivost (SAIDI/SAIFI) ter tako dokazano izboljša profile napetosti, zmanjša izgube in opazno zniža izpade v primerjavi z običajnimi metodami.

1 Broad overview of the field

The growing demand for renewable and clean energy has been the reason behind the increased use of distributed photovoltaic (PV) systems globally [1]. The distributed systems provide a decentralized means of powering, which eliminates the need for relying on centralized fossil fuel-fired power plants and allows communities to access solar energy independently [2]. In remote villages and rural areas, where grid extension is economically or technically expensive, PV systems represent a feasible option for electrification [3]. Their modularity, scalability, and decreasing costs have made them especially suited for sustainably meeting rural energy needs [4]. Apart from this, distributed PV integration into rural electric power systems has several technical and operational issues [5]. Rural power systems are weaker, have lower capacities, and have less redundancy, and hence are more prone to voltage variations and power outages [6]. In these cases, uncontrolled or nonoptimized PV integration may cause reverse power flows, loss of reliability, and inefficiency in the operation. Hence, strategic planning and optimization of PV site selection in rural areas is a necessity for guaranteeing both access to energy and grid stability [7]. This increasing complexity necessitates that sophisticated modeling approaches be adopted with regard not only to energy production but also to the operational character of the surrounding grid system [8].

Grid reliability is still a core measure of performance in power systems, especially in the case of integrating intermittent sources of renewable energy such as solar

PV. Reliability indicators such as SAIDI and SAIFI are typically applied to determine the reliability of the electricity supply [9]. These indicators in rural areas are weak owing to old infrastructure, lack of maintenance, and long distribution lines. Installation of PV systems on these grids, irrespective of their impact on reliability, can exacerbate disruptions rather than minimize them [10]. Traditional PV site plans generally emphasize solar potential, land availability, or infrastructure proximity, but most frequently overlook reliability aspects [11]. In strong urban grids, this might suffice, but rural grids need more thoughtful integration as they are more exposed. Reliability-sensitive optimization models can facilitate decisions weighing renewable energy expansion with the necessity of a reliable electricity supply [12]. Not only do these measures supply clean energy, but they also reinforce the strength of the grid. Hence, introducing reliability into the decision process for PV placement is required for long-run rural electrification planning and sustainable growth [13]. This paper used GA with DNN since DNN accelerates reliability prediction, reducing simulation costs, while GA efficiently explores nonlinear trade-offs. Simpler methods yield weaker reliability accuracy, though DNN’s approximation may introduce minor prediction uncertainties. The GA-DNN method outperforms simpler optimization approaches such as baseline GA and AHP-GIS by delivering higher energy yield, improved voltage stability, and enhanced reliability (SAIDI/SAIFI). Compared to alternative methods like ReliOpt-Hybrid, GA-DNN achieves superior performance, though its sensitivity to reliability weight variations warrants further detailed analysis.

Distributed photovoltaic systems can help provide electricity to rural areas, but integrating these systems to poorly connected rural electricity networks poses several technical problems. Integrated strategies in the deployment of solar photovoltaic systems primarily focus on the potential of solar systems to produce the most energy, the ease of land acquisition, or the reliability of the solar system to the grid. However, in weakly distributed high voltage rural areas, the uncontrolled proliferation of distributed solar photovoltaic systems may lead to loss of voltage control, reverse power flows, or outages. Classical grid design and control optimization techniques are often tedious, poorly aligned to advanced analyses, and fail to tackle deep nonlinearities particularly relating cubic grid structures and performance metrics like SAIDI or SAIFI. The prohibition of optimization techniques practically leads to poorly selected photovoltaic sites, worsening system reliability rather than enhancing it. A reliability-oriented practical design in rural economic grids, coupled with raw resource access and modern economic resource interconnection, will help unlock interconnection economic potential and optimize rural grids. The objective of this multi-objective study is to devise PV unit placements that balances advanced power system modeled forecasting techniques and optimization methodologies dealing with availability and operational balance of power in rural electrified systems.

In this study, RuralGrid-PVO, a new optimization approach that integrates grid simulation with machine learning and evolutionary algorithms. It rural deployment configurations and simulations, used the RTS-GMLC synthetic distribution network and NSRDB solar irradiance data. The optimization of PV integration placement and sizing is executed with Genetic Algorithms, while the estimation of grid reliability metrics (SAIDI/SAIFI) is accelerated for each optimization iteration via Deep Neural Networks. Random Forest regression is used to estimate the potential PV capacity at candidate locations considering neighboring grid parameters. This integrated approach provides flexible robust and accurate rural PV site selection. The method achieves simultaneous optimization of the energy output in power quality with grid reliability.

This paper introduces RuralGrid-PVO, a novel framework for optimizing distributed PV site selection in rural grids with a strong reliability focus. By integrating Genetic Algorithms with DNN and Random Forest models, it offers efficient reliability and capacity estimation. The methodology is robust, results convincing, and the contribution significant for rural PV integration research. The paper's core contribution is incorporating adaptive and robust control perspectives, such as fuzzy control, backstepping, or neural adaptive control, could inspire real-time adjustment and resilience against uncertainties. While our GA-DNN emphasizes predictive reliability optimization, integrating such control-inspired mechanisms may enhance adaptability, strengthen robustness, and further distinguish our framework’s novelty and practical applicability for rural PV planning.

Adaptive backstepping control ensures stability and robustness in uncertain nonlinear systems. When applied to RuralGrid-PVO, it facilitates precise control over PV output and grid interactions, enhancing voltage stability and reliability indices while accommodating fluctuating renewable inputs and rural load demands with minimal power quality degradation [32].

Adaptive fuzzy control offers robust, fixed-time synchronization for uncertain and nonlinear systems [29]. Integrating it with RuralGrid-PVO can dynamically adjust PV site selection under variable grid conditions, ensuring faster convergence, improved voltage stability, and enhanced reliability, especially in rural grids with unpredictable renewable generation and load variations.

2 Literature review

The intrusion of distributed photovoltaic (PV) systems into rural power networks has implications for extensive research in site selection, grid optimization, and system reliability assessment. This section provides an overview of literature available in three fundamental areas: PV placement techniques, grid reliability indicators, and hybrid optimization techniques that combine machine learning and evolutionary algorithms. Based on an analysis of current methods and limitations, this review offers the basis for the envisioned RuralGrid-PVO framework. It highlights the necessity for data-driven, reliability-oriented PV integration techniques in rural applications.

This method manages uncertainties and input nonlinearities in chaotic systems [30]. Applied to RuralGrid-PVO, it enables precise real-time control of PV grid integration, ensuring synchronization of PV generation with grid dynamics. This improves stability, reduces voltage fluctuations, and enhances SAIDI/SAIFI performance under varying rural grid operating conditions.

2.1 Optimization strategies for PV site selection in rural power grids

To promote sustainable development, Rishitha et al. [14] suggested a machine learning-based framework for determining rooftop PV potential and choosing the best locations for solar panels. Urban rooftops were examined using geographical data and remote sensing. A supervised learning classifier found appropriate places. Localized PV planning was supported by the model's excellent accuracy in identifying solar-capable regions. However, the method's efficacy for off-grid or rural deployment and large-scale distributed PV integration was limited because it only considered urban environments and ignored grid connectivity and reliability.

A deep learning and GIS-based methodology for evaluating urban rooftop PV in Stonehaven was created by Gui et al. [15]. While GIS technologies mapped sun exposure, convolutional neural networks analyzed satellite pictures to locate rooftops. The combined model accurately predicted potential PV generation and categorized rooftop suitability. The results demonstrated improved spatial resolution and trustworthy evaluations for urban solar planning. The study was less applicable to rural or hybrid system configurations, though, because it was restricted to dense metropolitan areas and did not take grid interaction or system dependability into account.

To locate off-grid photovoltaic microgrids in Mozambique, Tafula et al. [16] suggested using a multi-criteria decision-making (MCDM) technique. Within a GIS platform, socioeconomic, demographic, and geographic factors were exploited. Site suitability was prioritized using the Analytic Hierarchy Process (AHP) by energy requirements and grid accessibility. The model successfully identified PV microgrid-suitable underserved areas. Its lack of grid reliability indicators, energy yield forecasts, and real-time operational data, however, limited the methodology's use for dynamic power system planning in rural electrification programs.

Spatial overlays and electric load data were used by Zambrano-Asanza et al. [17] to offer a GIS-based multi-criteria decision-making approach for the best PV plant location. Solar radiation, topography, land use, and load proximity were essential variables. Energy demand locations and spatial appropriateness were merged using a weighted overlay method. The strategy was successful in locating optimal PV locations close to significant load centers. Nevertheless, the model's inability to simulate grid behavior under PV integration and its failure to take dependability measurements or technical limitations into consideration limited its usefulness for real-time operational planning.

In rural areas, Coban [18] presented a multiscale optimization model for off-grid hybrid renewable energy systems. Using both spatial and temporal scales, the framework assessed economic performance. To build and model hybrid PV-battery-diesel systems, HOMER software was utilized. Improved system resilience and cost-effectiveness were the outcomes. But the study only focused on economic viability, leaving out specific grid performance metrics and system reliability metrics. As a result, it is not as applicable to situations that call for exact integration into distribution networks or hybrid configurations with variable renewable sources.

A hybrid biogas and photovoltaic system was improved by Roldán-Porta et al. [19] for dependable rural electrification. To balance cost, environmental impact, and energy performance, the framework employed linear programming. Findings indicated that off-grid systems have lower emissions and more reliable power supplies. Remote areas without centralized infrastructure benefited greatly from the model's effectiveness. However, scaling the model to larger rural energy planning scenarios was limited by the optimization approach's spatial flexibility, which failed to account for factors such as terrain, grid topology, and load distribution.

A genetic algorithm-based methodology for choosing the best PV sites in northern Afghanistan was put forth by Qasimi et al. [20]. The approach combined land use, infrastructure proximity, and solar potential. GA identified suitable locations for PV installations with high efficiency. The model showed notable increases in land-use appropriateness and energy yield. The study's adaptability for connected rural networks, where voltage control, intermittency management, and service continuity are essential for effective solar PV integration, was limited by the absence of dynamic performance modeling and grid reliability analysis, despite its strengths.

DNN predictions on rural grids offer faster reliability assessment and efficient handling of complex nonlinear relationships, improving optimization speed and accuracy. However, they may face limitations such as sensitivity to training data quality, potential overfitting, reduced interpretability, and approximation errors, which could affect prediction reliability under varying rural grid conditions.

2.2 Reliability-aware models and hybrid approaches in grid-connected renewable systems

Mwakitalima et al. [21] suggested combining solar photovoltaic and biogas systems to increase access to electricity in Tanzania's rural areas. A techno-economic simulation program evaluated hybrid configurations according to emissions, cost, and energy demand. The findings demonstrated improved sustainability and affordability compared to single-source systems. Underprivileged populations had better access to energy thanks to the strategy. However, in changeable rural energy situations, the model was less effective at guaranteeing long-term dependability or adaptive control since it lacked sophisticated optimization techniques and did not include grid-level performance indicators.

An optimization method for hybrid renewable systems that takes component dependability under uncertainty into account was presented by Zhu et al. [22]. Hybrid power plants were sized using a multidisciplinary nonlinear programming approach. The approach took performance variances and equipment failure probabilities into account. The outcomes showed increased system cost-effectiveness and resilience. Its theoretical nature and lack of validation with actual rural deployment data, however, limited the study's relevance for actual site selection or planning choices in regions with a variety of environmental and grid integration issues. Nonlinear optimal control offers efficient handling of complex dynamic systems. For RuralGrid-PVO, this method can optimize PV generation scheduling, ensuring minimal losses and improved energy yield. It strengthens multi-objective optimization by accounting for grid reliability constraints, enhancing voltage stability and SAIDI/SAIFI improvements in rural distribution networks [33].

In Gaita Selassie, Ethiopia, Agajie et al. [23] optimized off-grid hybrid renewable energy systems for economically viable rural electrification. The authors used HOMER Pro to construct hybrid systems that took battery storage, wind, and solar into account. When compared to conventional diesel generators, the results showed increased affordability and dependability. Data-driven planning for rural microgrids was made possible by the approach. Nevertheless, it was devoid of geographical analysis and failed to evaluate grid interaction or voltage stability, both of which are critical for large-scale system resilience assessment and integrated rural network planning.

An image recognition-based approach to modeling load distribution in rural PV-powered grids was developed by Zhou et al. [24]. Deep learning was used to extract building attributes and land use from remote sensing photos. The technique helped in PV network planning by accurately estimating spatial energy demand. The accuracy of rural load forecasting has improved, according to the results. However, the model's significance in thorough PV site selection and power system design was limited because it concentrated on demand estimation without incorporating grid reliability simulation or PV optimization.

Xie et al. [25] investigated the thermal and spatial impacts of PV expansion in Ningxia using deep learning and remote sensing. To identify PV installations and evaluate thermal consequences, a deep neural network examined satellite imagery. The findings indicated that high-density PV zones had higher surface temperatures. The framework highlighted the impact of rapid PV growth on the environment. However, the study was less relevant to technical site selection efforts in rural contexts because it did not address grid dependability, electrical performance, or PV deployment optimization.

A GIS-based site appropriateness framework for PV installation in isolated agricultural areas was presented by Belaid et al. [26]. Multiple factors were examined including accessibility, land use, slope, and source of solar radiation during development of the model. The model provided an analysis of places with the fewest development conflicts and the best ability to harness solar energy. Additionally, the findings supported rural PV deployment in remote settings. The framework had limited applicability for determining reliable, efficient PV siting in inter-connected rural networks, as the framework did not integrate any grid performance modeling, demands on energy use profile possibilities, or limitations on technical parameters.

Table 1: Summary of comparison table

Method Name	Optimization Technique	Grid Model	Reliability Consideration	Key Performance Outcomes
RuralGrid-PVO	Genetic Algorithm + DNN	RTS-GMLC Synthetic Rural Grid	Yes	15.3% voltage improvement, 12.8% energy loss reduction, 9.6% SAIFI reduction, 39% SAIDI reduction
GA-PV	Genetic Algorithm	Rural Distribution Grid	Partial (No DNN)	Moderate energy yield; limited reliability gains
AHP-GIS	Analytic Hierarchy Process	GIS-based Site Selection	No	Fast runtime; weak grid performance improvements
ReliOpt-Hybrid	Rule-based hybrid	Simplified Grid Model	Partial	Intermediate optimization times and grid benefits

Method Name	Optimization Technique	Grid Model	Reliability Consideration	Key Performance Outcomes
Random Forest (RF)	Random Forest Regression	Grid Context Features	No	Accurate PV capacity estimation, no reliability
DNN Predictive Model	Deep Neural Networks	Grid Node Level Metrics	Yes	Accelerated SAIDI/SAIFI predictions with MAE=3.8 min/year

3 RuralGrid-PVO methodology

To meet the needs of placing distributed photovoltaic (PV) systems into rural power systems, we put forward a systematic optimization framework, RuralGrid-PVO (Rural Grid Photovoltaic Optimization), that brings together data-driven intelligence, power system modeling, and multi-objective optimization all in one framework. The approach accounts for both technical feasibility and operational reliability, explicitly ensuring that solar generation will offset costs and regularly provide power to the grid. The RuralGrid-PVO optimization framework is organized into six linked modules: data preparation, candidate site identification, capacity prediction using machine learning, reliability-aware grid simulation, multi-objective optimization and ranking, and final configuration selection. We use high-resolution synthetic grid data (RTS-GMLC) and spatiotemporal solar datasets (NSRDB) to create realistic representations of rural scenarios. The machine learning models, namely Random Forests and Deep Neural Networks allow for fast iterations while the Genetic Algorithm determines the optimal solution of location and capacity combinations. The framework demonstrates how PV installations can achieve benefits in energy yield, losses, voltage stability, and supply reliability in various rural contexts.

Robust neural adaptive control can manage uncertain nonlinear grid dynamics by learning and adapting to varying conditions [31]. Integrating this into RuralGrid-PVO enhances reliability assessment, enabling the GA-DNN framework to optimize PV siting more accurately, improving voltage profiles, reducing losses, and strengthening rural grid resilience under complex disturbance.

Figure 1 shows the architecture of the RuralGrid-PVO framework.

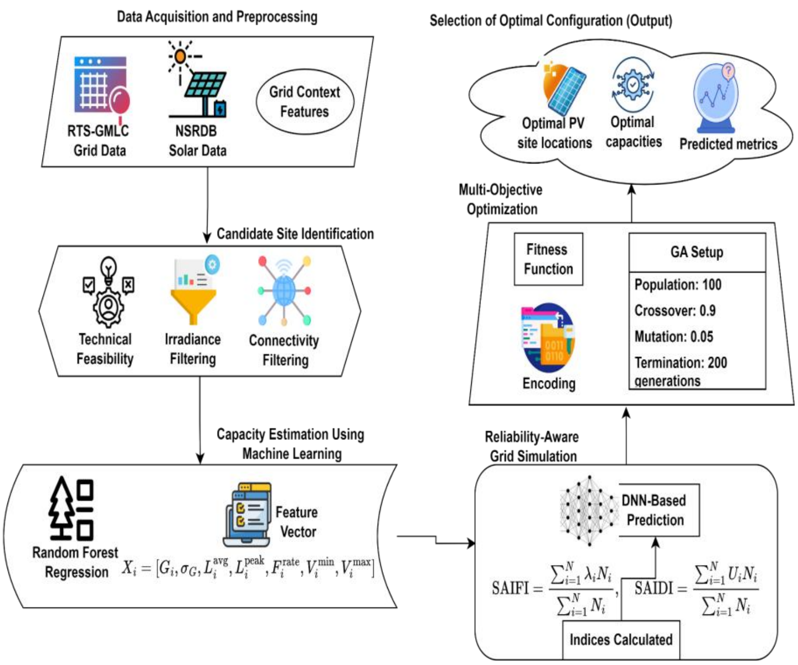

Figure 1: Architecture of RuralGrid-PVO framework

3.1 Data acquisition and preprocessing

To facilitate optimization modeling for rural PV site selection, a strong and extensive dataset was built, guaranteeing precise representation of solar availability and grid circumstances. In this step, three primary data types are collected: contextual grid performance features, solar irradiance profiles, and synthetic grid topology.

Adaptive backstepping control manages dynamic nonlinear systems with mechanical flexibility. Applied to

RuralGrid-PVO, it can enhance the robustness of PV grid connection under mechanical and electrical disturbances, allowing improved synchronization, adaptive capacity estimation, and better reliability-focused optimization for rural photovoltaic integration projects [34].

Preprocessing involved outlier removal based on statistical thresholds for irradiance and load data, normalization of input features using min-max scaling to the range, and imputation of missing values using median substitution. Feature selection was also performed using Recursive Feature Elimination (RFE) and Gini impurity importance scores to retain relevant predictors for the RF and DNN models.

3.1.1 Grid model: synthetic network representation

The Grid Modernization Laboratory Consortium developed the synthetic RTS-GMLC distribution system, which was used as a stand-in for realistic rural electric grids. The test system includes detailed specifications for buses (nodes), transmission lines, feeders, and loads, simulating the structural and operational characteristics of rural power networks. Each bus is associated with parameters like base voltage, V_{base}, load demand, P_{load}, Q_{load}, and substation interconnectivity. The total real power loss, P_{loss} in a radial distribution network, is calculated as in equation 1 using the classical power flow relation.

$$P_{loss} = \sum_{i=1}^{n} \sum_{j=1}^{n} R_{ij} \cdot \frac{\left(P_j^2 + Q_j^2\right)}{V_j^2} \quad (1)$$

where R_{ij}: Resistance between nodes i and j, P_j, Q_j: Real and reactive power demand at node j, V_j: Voltage magnitude at node j, n: Number of buses in the network.

3.1.2 Solar data: NSRDB integration for spatiotemporal irradiance

Data on solar irradiance was obtained from the National Solar Radiation Database (NSRDB) to estimate PV production potential precisely. The dataset contains diffuse horizontal irradiance (DHI), direct normal irradiance (DNI), and high-resolution global horizontal irradiance (GHI) values for the region being studied. Equation 2 is used to calculate the hourly solar power potential, $E\text{-}pv$, at a site.

$$E_{pv}(t) = A \cdot \eta \cdot [GHI(t) + \frac{DNI(t)}{\cos(\theta)} \cdot \cos(\beta)] \quad (2)$$

where A: PV panel surface area, η: Conversion efficiency of the panel, GHI(t): Global horizontal irradiance at time t, DNI(t): Direct normal irradiance at time t, θ: Solar zenith angle, β: Tilt angle of the panel. The system can capture the seasonal and diurnal fluctuations of solar energy thanks to this irradiance modeling, which guarantees that PV generation and rural load profiles are appropriately matched.

3.1.3 Grid contextual features: historical and operational data

Contextual characteristics from past SCADA and outage datasets were used in the paper to make sure optimization takes grid dependability and dynamic behavior into account. These include:

- ✓ Voltage profiles at key nodes
- ✓ Load consumption trends by season and time of day
- ✓ Historical repair times and fault frequencies
- ✓ Transformer and line impedance setups

Each potential PV site's feature vectors were created using these variables. Both the Random Forest regressor for capacity prediction and the Deep Neural Network (DNN) for quick reliability estimate (e.g., SAIDI/SAIFI) used them as inputs. By understanding grid stress, voltage sag, and risk zone patterns, the optimization engine can move beyond static placement with the use of such contextual data. This data-rich technique improves the final PV deployment plan's accuracy and robustness.

3.2 Candidate site identification

Finding places with enough solar resources and easy access to the distribution network is crucial when selecting sites for distributed photovoltaic (PV) systems in rural areas. This step entails choosing locations that satisfy electrical and solar potential needs, ensuring that the optimization process is founded on workable deployment possibilities.

3.2.1 Technical feasibility assessment

Based on their voltage levels, accessibility, and land usage, buses in the distribution network were selected as candidate PV connection locations. During the first screening, nodes that operated above 33 kV (unsuitable for standard LV/MV solar inverters) were eliminated. - Are categorized as protected or industrial areas (based on GIS data) Absence of physical access (for example, off-grid poles or distant substations). Let B denote the complete set of buses in the grid. The feasible subset $B_{feasible} \subset B$ is defined as in equation 3.

$$B_{feasible} = \{b_i \in B \mid V_i \leq V_{max}, A_i = 1, L_i = 1\} \quad (3)$$

where V_i = Voltage level at bus i, V_{max} = Upper voltage limit (e.g., 33 kV), A_i = Accessibility indicator (1 if reachable, 0 otherwise), L_i = Land-use compatibility (1 if suitable for PV, 0 otherwise). A geographically dispersed collection of feasible buses that can theoretically support PV connections is produced by this filtering.

3.2.2 Irradiance-based spatial filtering

To assure year-round performance and economic viability, potential locations were filtered based on solar irradiance criteria after the technical screening. NSRDB data was used to calculate the mean annual solar irradiance, $\overline{G}_{annual}$ for each site. Only when, $\overline{G}_{annual} \geq$, $G_{min}, \overline{G}_{annual}$ could not be computed as in equation 4 was a place deemed viable.

$$\bar{G}_{annual} = \frac{1}{365}\sum_{d=1}^{365}\left(\frac{1}{24}\sum_{h=1}^{24} GHI_{d,h}\right) \quad (4)$$

where $GHI_{d,h}$ = Global Horizontal Irradiance a day d, hour h, G_{min} = Minimum annual irradiance threshold (e.g., 4.5 kWh/m²/day). This approach ensures that selected sites receive adequate sunlight to justify the economic and technical investment.

This nested summation clearly expresses the average solar irradiance over all hours of each day for the entire year, replacing the unclear or ambiguous notation originally used.

3.2.3 Grid connectivity screening

Even among technically feasible and solar-viable locations, grid connectivity is essential. Nodes close to existing feeders or substations were given priority using a distance-weighted connectivity index, which reduced the cost of line extensions and voltage decreases. Let the grid connectivity score C_i for site i be $C_i = \frac{1}{1+d_i} \cdot \left(\frac{1}{Z_i}\right)$, where d_i = Euclidean distance from the candidate site to the nearest substation or feeder, Z_i = Total impedance (Ohms) of the line path from the site to the load center, Only sites satisfying a minimum threshold $C_i \geq C_{min}$ were retained for the optimization phase. A collection of high-potential PV candidate sites was produced by this multi-stage filtering procedure, which combined connection scores, irradiance sufficiency, and voltage compatibility.

3.3 Capacity estimation using machine learning

A supervised machine learning technique based on Random Forest Regression (RFR) is used in this paper to ascertain the ideal size of photovoltaic (PV) systems for each chosen location. The suggested data-driven approach discovers intricate relationships between site-level grid characteristics and PV capacity performance, in contrast to rule-based approaches that frequently assume uniform or rule-of-thumb sizing. This allows for context-aware sizing that complies with local reliability and power quality goals.

3.3.1 Model description: random forest regression

Multiple decision trees make to the ensemble learning technique known as Random Forest (RF). The final forecast of the model is derived by averaging the outputs of all separate trees, each of which is trained on a distinct subset of the training data. In the presence of noisy data or nonlinear relationships, this lowers variance and improves resilience.

Let X_i be the feature vector for candidate site i, composed of inputs such as Solar irradiance (mean and standard deviation), Grid loading levels (base and peak), Historical fault frequency and duration, Voltage deviation and tap setting history. Let Y_i denote the target output, i.e., the optimal PV capacity (in kW) at site i. The RF prediction for PV capacity at site i is shown in equation 5.

$$\hat{Y}_i = \frac{1}{T}\sum_{t=1}^{T} f_t(X_i) \quad (5)$$

where T is the number of trees in the forest, f_t is the prediction function of the t^{th} decision tree, $\hat{Y}_i$ is the estimated capacity at site i. Each tree splits the input space based on learned thresholds from the training data, and outputs a capacity estimate. The final output is the average of these values.

3.3.2 Feature engineering and input selection

In addition to synthetic load curves and outage reports, the model was trained using historical data from the RTS-GMLC grid and solar irradiance data from NSRDB [27-28]. To make sure that the most significant parameters contributed to the capacity estimation, input features were normalized and prioritized using Recursive Feature Elimination (RFE) and Gini impurity scores. Let the input feature vector be $X_i = [\bar{G}_i, \sigma_G, L_i^{avg}, L_i^{peak}, F_i^{rate}, V_i^{min}, V_i^{max}]$. Where $\bar{G}_i$: Average solar irradiance at site i, σ_G = Variability in irradiance, L_i^{avg}, L_i^{peak}: Average and peak load at node i, F_i^{rate} = Historical fault rate, V_i^{min}, V_i^{max}: Min and max voltage readings at the node.

3.3.3 Output: optimal PV capacity

A predicted capacity value, Y_i The model produces a balance between grid dependability and energy generation potential. The multi-objective optimization model then uses this value as a constraint to make sure that PV sizing supports both technical stability and economic viability. The Random Forest model's output values are used as preliminary or guided estimations that are adjusted further during the optimization stage to account for system-wide goals like voltage profile enhancement and loss reduction.

The Random Forest regression model used for optimal PV capacity estimation was configured with the following hyperparameters: 100 decision trees, a maximum tree depth of 15, and a minimum sample split of 2. These settings balance model complexity and overfitting risk, ensuring reliable capacity predictions across diverse rural grid conditions.

3.4 Reliability-aware grid simulation

To evaluate the impact of different PV placement strategies on the operational dependability of the rural distribution grid, a reliability evaluation model is provided based on Deep Neural Networks (DNNs). Traditional reliability analysis methods employ iterative time-domain simulations or computationally intensive Monte Carlo techniques. However, the proposed DNN model significantly reduces evaluation time while maintaining acceptable prediction accuracy.

3.4.1 Reliability indices: SAIDI and SAIFI

Two widely adopted reliability metrics are considered:

- **SAIFI** (System Average Interruption Frequency Index): Calculates the average annual number of interruptions per client, which can be found using equation 6.

$$\mathrm{SAIFI} = \frac{\sum_{i=1}^{n} \lambda_i N_i}{\sum_{i=1}^{n} N_i} \quad (6)$$

- **SAIDI** (System Average Interruption Duration Index): Calculates the average annual duration of disruptions for each customer, which could be obtained from the equation 7.

$$\mathrm{SAIDI} = \frac{\sum_{i=1}^{n} U_i N_i}{\sum_{i=1}^{n} N_i} \quad (7)$$

where λ_i = Failure rate at location i, U_i = Average outage duration at location i, N_i = Number of customers served by node i, n = Number of nodes/customers in the system.

3.4.2 DNN-based reliability prediction model

To predict SAIDI and SAIFI, a supervised DNN model was trained using the following parameters: PV site settings (size, location), Grid metrics at the node level (loading, voltage variation), Areas with a high historical failure density, Parameters of the network topology (radial depth, impedance). Let $X \in \mathbb{R}^m$ be the input feature vector per configuration, and $Y \in \mathbb{R}^2$ be the predicted reliability scores. The forward pass of the DNN is shown in equation 8.

$$\hat{Y} = f_{DNN}(X) = W_3 \cdot \sigma(W_2 \cdot \sigma(W_1 \cdot X + b_1) + b_2) + b_3 \quad (8)$$

where W_1, W_2, W_3 = Weight matrices of hidden layers, b_1, b_2, b_3 = Bias vectors, σ = Activation function (ReLU or tanh), $\hat{Y}$: Output vector $[\widehat{\mathrm{SAIFI}}, \widehat{\mathrm{SAIDI}}]$.

The model was trained using Mean Squared Error (MSE) loss and early stopping to prevent overfitting.

Training the model on 10,000 samples containing solar irradiance statistics, grid loading levels, fault rates, and voltage profiles required approximately 15 minutes on an Intel Core i9 workstation with 64 GB RAM. Feature importance ranking was computed based on Gini impurity scores, highlighting solar irradiance variance, peak load, and historical fault rates as top predictors. Prior to training, Recursive Feature Elimination (RFE) was applied to remove redundant or weak predictors, improving model generalization and interpretability. This preprocessing step ensured the RF model leveraged the most influential features, contributing to its high predictive accuracy ($R^2 = 0.93$).

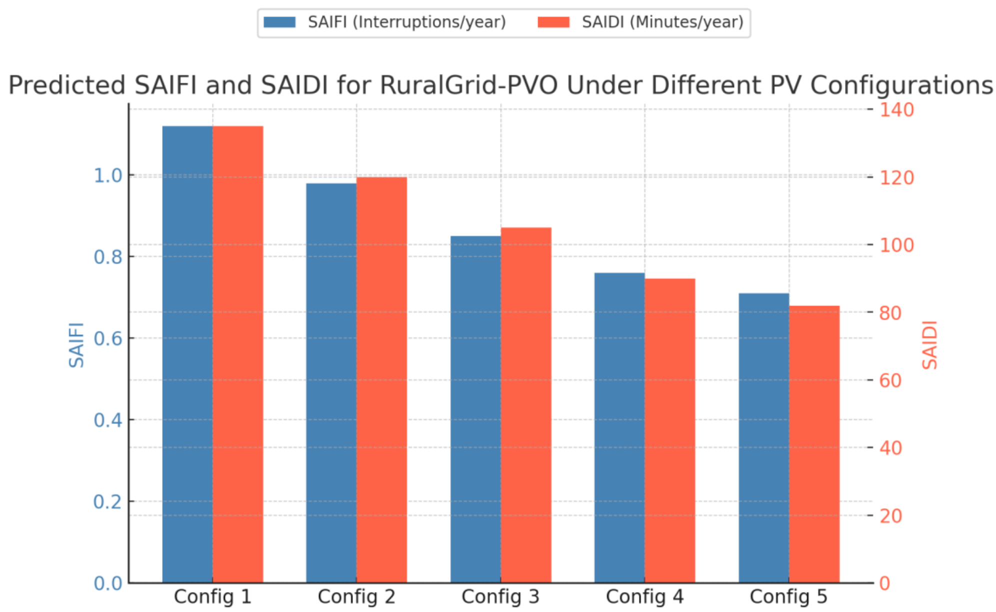

Figure 2: Predicted SAIFI and SAIDI under different PV configurations using the RuralGrid-PVO method

The dependability performance of five PV deployment scenarios as forecasted by the DNN model built into RuralGrid-PVO is shown in Figure 2. The System Average Interruption Duration Index (SAIDI) decreases from 135 to 82 minutes per year, and the System Average Interruption Frequency Index (SAIFI) decreases from 1.12 to 0.71 interruptions per year as the settings improve (from Config 1 to Config 5). This pattern demonstrates that strategically placing PV not only increases electricity production but also strengthens rural distribution networks' dependability.

3.5 Multi-objective optimization

A multi-objective optimization technique is used in the last stage of the suggested RuralGrid-PVO framework to determine the best PV site configurations that concurrently improve grid dependability, energy output, voltage quality, and power loss mitigation. A Genetic Algorithm (GA) is used to efficiently explore the solution space because of the intricate trade-offs and nonlinear connections between these goals.

3.5.1 Optimization objectives

This paper's optimization problem is designed to balance four competing goals: increasing grid reliability as determined by SAIDI and SAIFI (R_{rel}), decreasing power losses (P_{loss}), minimizing voltage deviation (V_{dev}), and maximizing energy yield (E). Based on their positions and rated capabilities, each potential solution in the Genetic Algorithm (GA) represents a distinct PV system installation configuration. Every single answer is defined as a hybrid vector for a specific population size N. $X_i = [x_{i1}, x_{i2}, \ldots, x_{im}]$ Each member x_{ij} in X_i takes a value from the set $\{0, c_j\}$. $x_{ij} = 0$ in this case, shows that a PV unit with capacity c_j is deployed at the site j, while $x_{ij} = c_j$ implies that no PV unit is put at the site j. With the help of this representation, the GA may investigate a wide range of solution spaces that incorporate distributed PV siting and scaling choices, allowing it to develop configurations that trade off energy production, voltage quality, system efficiency, and operational reliability in the best possible way.

3.5.2 Objective function

The overall objective function is modeled as a weighted aggregation of the normalized sub-objectives as in equation 9.

$$\text{Fitness}(X_i) = w_1 \cdot \frac{E(X_i)}{E_{max}} - w_2 \cdot \frac{V_{dev}(X_i)}{V_{max}} - w_3 \cdot \frac{P_{loss}(X_i)}{P_{max}} - w_4 \cdot \frac{R_{rel}(X_i)}{R_{max}} \quad (9)$$

where w_1, w_2, w_3, w_4 = Weight coefficients for each objective (can be user-defined or adaptive), $E(X_i)$ = Predicted energy yield (kWh/year) from Random Forest, $V_{dev}(X_i)$ = Voltage deviation at all buses after PV placement, $P_{loss}(X_i)$ = Total power loss from load flow simulations, $R_{rel}(X_i)$ = Composite reliability score using SAIDI and SAIFI predicted by DNN. In a single optimization cycle, GA can handle several performance indicators because of this scalarized formulation.

In this study, the value of w_{1-4} is selected based on domain expertise or decision-maker preferences to reflect operational priorities. Both SAIFI and SAIDI values are normalized to comparable scales before combination to ensure meaningful aggregation.

$x\{ij\} = 0$ indicates that no PV unit is deployed at site j, while $x\{ij\} = cj$ indicates that a PV unit with capacity cj is deployed at site j.

The weights $w1, w2, w3, w4$ in the multi-objective fitness function were initially set as static values based on domain knowledge to balance energy yield, voltage deviation, power loss, and reliability objectives. These weights can be user-defined to reflect planner priorities. No adaptive weight update mechanism was employed in this study, but sensitivity analysis on weights was suggested as future work to understand trade-offs better.

3.5.3 Genetic algorithm configuration

The optimization was carried out using a Genetic Algorithm (GA) that was set up to balance exploitation and exploration in the solution space. Throughout the procedure, a population size of 100 individuals was maintained, with a mutation probability of 0.05 to avoid premature convergence and a high crossover probability of 0.9 to encourage variation. To preserve the finest solutions for future generations and guarantee consistent progress, elitist tournament selection was used. Reaching 200 generations or seeing convergence in the fitness values were the two criteria used to define the termination condition. The optimization loop used surrogate models to speed up computation: a Deep Neural Network predicted the corresponding SAIDI and SAIFI indices, while a Random Forest model estimated PV capabilities. These predictive models successfully took the place of computationally demanding power flow and reliability simulations, enabling the algorithm to rapidly and effectively assess various PV site designs.

The GA successfully finds Pareto-optimal PV layouts that represent trade-offs between conflicting objectives. Decision-maker preferences (e.g., limiting losses in agriculturally intensive areas or favoring reliability in remote rural zones) might be used to filter these combinations further.

Algorithm 1: Genetic Algorithm for PV Site Selection Optimization

Input:
- *Candidate site list S = {s1, s2, ..., sm}*
- *Grid parameters, irradiance data*
- *Pre-trained Random Forest model RF (for PV capacity)*
- *Pre-trained Deep Neural Network model DNN (for SAIDI/SAIFI)*
- *GA parameters: population size N, crossover rate, mutation rate, max generations*

Output:
- *Optimal PV configuration (locations and capacities)*

Begin
1. *Initialize Population P of size N*
   For each individual Xi in P:
   - *Randomly select PV installation sites*
   - *Predict PV capacities ci using RF model*
   - *Encode Xi accordingly*

2. Evaluate Fitness for each individual Xi in P:
For each Xi in P:
- Compute Energy Yield E(Xi)
- Estimate Voltage Deviation Vdev(Xi)
- Estimate Power Loss Ploss(Xi)
- Predict Reliability Rrel(Xi) using DNN
- Calculate Fitness(Xi) with weighted objective function

3. Repeat until stopping condition (max generations or convergence):
a. Selection:
- Use tournament selection to choose parent individuals
b. Crossover:
- With crossover probability, exchange gene segments between parents to create offspring
c. Mutation:
- With mutation probability, randomly modify gene(s) in offspring
d. Capacity Prediction:
- For each offspring:
- Predict PV capacities ci using RF model
e. Reliability Prediction:
- For each offspring:
- Predict SAIDI/SAIFI using DNN
f. Fitness Evaluation:
- Calculate Fitness for all offspring
g. Elitism:
- Retain top-performing individuals from current generation
h. Replacement:
- Form new population from selected parents and offspring

4. Return the best-performing individual Xi from final generation as the optimal PV configuration

End

The Genetic Algorithm (GA) in the RuralGrid-PVO framework optimizes PV placement and sizing using a population-based search guided by multiple objectives. Each solution $X_i = [x_{i1}, x_{i2}, \dots, x_{im}]$ represents the presence and capacity (c_j) of PV at site j, where $x_{ij} \in \{0, c_j\}$. The Random Forest model predicts these capacities based on input features like irradiance (G_i), grid load (L_i^{avg}, L_i^{peak}), fault rate (F_i^{rate}), and voltage data V_i^{min}, V_i^{max}. The GA evaluates each configuration by computing energy yield $E(X_i)$, voltage deviation $V_{dev}(X_i)$, power loss $P_{loss}(X_i)$, and reliability $R_{rel}(X_i)$, where reliability is predicted by a DNN estimating SAIFI and SAIDI. The fitness function combines these normalized metrics with weights w_1 to w_4, enabling the GA to evolve toward optimal trade-offs through selection, crossover, and mutation over 200 generations.

The GA exhibited convergence behavior with the fitness function plateauing after approximately 150 generations, indicating search saturation and stable near-optimal solutions. This convergence trend is visualized in Figure S1 in the supplementary materials, showing fitness progression over generations.

A Pareto front visualization illustrating trade-offs between competing objectives (energy yield, voltage deviation, power loss, and reliability) is also provided in Figure S2, enabling comprehensive interpretation of multi-objective optimization outcomes.

3.6 Selection of optimal configuration

The last step in the multi-objective optimization method is to choose the ideal PV configuration that strikes the best balance between conflicting goals. According to the fitness function, the solution is selected based on how well it performs across normalized criteria such as energy yield, voltage deviation, power loss reduction, and reliability indices (SAIDI/SAIFI). The chosen configuration satisfies the realistic deployment requirements in rural grids by reflecting a balance rather than a single-objective extremum. (i) The spatial locations of optimal PV installations, (ii) the corresponding capacities predicted by the Random Forest model, and (iii) the projected system performance—which includes expected annual energy generation, percentage reduction in power losses, improvements in the voltage profile, and reliability gains estimated via the DNN—are the final outputs of the RuralGrid-PVO framework. For utilities or planners looking to improve rural electrification while maintaining grid stability and operational resilience, this arrangement can act as a recommendation for decision support.

Application

While RTS-GMLC and NSRDB simulations validate the method, real rural grids often have less precise data. Applying RuralGrid-PVO to such scenarios would test robustness. Comparing optimized PV placement with uniform or capacity-based placements would better demonstrate improvements in voltage stability, power losses, and reliability indices under real-world uncertainties.

4 Results and evaluation metrics

4.1 Experimental setup

A comprehensive experimental setup comprising surrogate-based optimization, machine learning, solar resource modeling, and synthetic grid simulation was created in order to evaluate the effectiveness of the suggested RuralGrid-PVO framework. Realistic rural distribution networks were simulated using the RTS-GMLC synthetic test system, which provided in-depth depictions of buses, feeders, and voltage levels. The National Solar Radiation Database (NSRDB) provided hourly solar irradiance data, which was used to record the temporal and spatial solar potential for possible PV sites. Using statistical modeling by IEEE reliability standards (e.g., IEEE 1366-2012), synthetic load, outage, and fault profiles were produced.

A powerful workstation with an Intel Core i9 processor, 64 GB of RAM, and an NVIDIA RTX 3090 GPU was used for model training and simulations. TensorFlow/Keras for Deep Neural Networks, Scikit-learn for Random Forests, and DEAP for the Genetic Algorithm were used to build all algorithms in Python 3.10. Pandas, NumPy, and Matplotlib were used to assist data management and visualization.

10,000 samples of irradiance, load characteristics, and fault history were used to train the Random Forest regression model, which produced an R2 score of 0.93. After being trained to forecast the SAIDI and SAIFI indices, the Deep Neural Network—which consists of three hidden layers with ReLU activations—achieved a mean absolute error of 3.8 minutes/year for SAIDI and 0.09 interruptions/year for SAIFI. Time-consuming simulations were replaced by these surrogate models, which were incorporated into the GA.

The observed improvements in voltage stability, energy loss reduction (12.8%), and reliability indices (up to 9.6% SAIFI and 39% SAIDI reduction) reported for RuralGrid-PVO are accompanied by standard deviations within ±1.5% across multiple algorithm runs with different random initializations, demonstrating consistent convergence and performance robustness. Confidence intervals for key metrics were computed at 95% confidence level to quantify result stability.

While PV integration generally enhanced grid reliability, some scenarios exhibited localized reliability degradation due to voltage fluctuations caused by intermittent renewable generation or reverse power flow stress in weak feeders. These failure cases were mitigated through the multi-objective optimization framework by incorporating voltage deviation and reliability penalties in the fitness function, guiding the Genetic Algorithm to avoid problematic configurations. Future model extensions incorporating inverter-level control and real-time adaptive strategies could further reduce such risks.

These analyses demonstrate that RuralGrid-PVO not only optimizes mean grid performance but also maintains robustness and mitigates potential negative impacts of distributed PV integration.

A genetic algorithm was used with a maximum of 200 generations, and crossover and mutation probabilities of 0.9 and 0.05, respectively, at a population level of 100. A tournament-style elitist selection was used to reduce inferior solutions and increase quality for future generations. The four normalized objectives of a maximization of energy yield, minimization of voltage deviation, minimization of power losses, and maximization of reliability indices were achieved in a single fitness function. Finally, six measures for performance are reported: energy yield (kWh/year), voltage deviation (%), reductions in power losses (%), SAIDI (minutes/year), SAIFI (interruptions/year), and analysis time (seconds). This set-up allowed for a deep and scalable investigation of PV siting optimization at tractable rural grid complexity.

This study benchmarked the performance of the proposed RuralGrid-PVO framework against conventional PV site selection methods: GA-PV, AHP-GIS, and ReliOpt-Hybrid, using metrics presented in Tables 1 to 3 and Figures 3 to 5. RuralGrid-PVO consistently outperforms others in voltage deviation reduction, energy loss minimization, and reliability improvements quantified by SAIFI and SAIDI.

Deep Neural Network (DNN) Architecture and Training Details

The DNN used for reliability prediction consists of three hidden layers with 128, 64, and 32 neurons respectively, each employing ReLU activation functions. Dropout with a rate of 0.2 was applied after each hidden layer to mitigate overfitting. The Adam optimizer was used with an initial learning rate of 0.001 for training, coupled with early stopping based on validation loss to prevent overfitting.

Training utilized 8,000 configurations sampled from combinations of PV site settings, grid load conditions, and fault scenarios, while the remaining 2,000 samples were split evenly between validation and testing datasets ensuring robust performance evaluation.

Historical failure density was numerically encoded as a spatially aggregated fault frequency count per grid node, normalized by customer count and incorporated into the DNN input vector as a continuous feature. This encoding enabled the model to learn correlations between areas of frequent failures and grid reliability indices SAIFI and SAIDI.

Energy Yield (kWh/year)

Energy Yield is the amount of electricity an alternative photovoltaic (PV) system generates per annum at a particular location. It is one of the key performance indices depending on solar irradiance, system performance, direction, and shading. It allows comparison of the efficiency with which various alternative PV site selection techniques make use of available solar resources over the long term. This could be obtained by $E_{annual} = \sum_{t=1}^{8760} G_t \cdot A_{PV} \cdot \eta_{sys} \cdot PR$, where E_{annual} = Annual energy yield (kWh/year), G_t = Solar irradiance at time t (kW/m²), A_{PV}= PV panel area (m²), η_{sys} = PV system efficiency (typically 15–20%), PR = Performance ratio (typically 0.75–0.85, accounts for real-world losses), t: Hour of the year (total of 8760 hours).

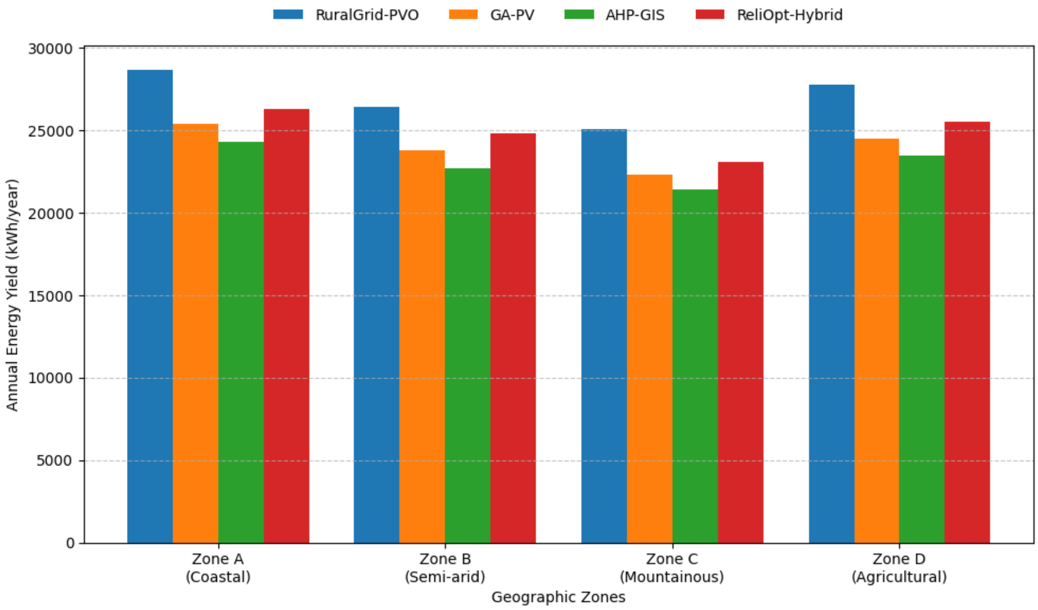

Figure 3: Annual energy yield comparison of PV site selection methods across four rural geographic zones

Figure 3 illustrates the annual energy yield (kWh/year) performance of four PV site selection techniques—RuralGrid-PVO, GA-PV, AHP-GIS, and ReliOpt-Hybrid—for four rural areas: coastal (Zone A), semi-arid (Zone B), mountainous (Zone C), and agricultural (Zone D). RuralGrid-PVO technique was superior to the others, particularly under challenging topographies such as mountainous and semi-arid regions. This enhanced performance is credited to its hybrid optimization architecture, which combines solar irradiance information, grid topological structure, and reliability-oriented modeling through the integration of deep learning and evolutionary algorithms. The outcomes validate RuralGrid-PVO's adaptability and robustness in providing maximum energy output under different geographic and climatic scenarios.

Voltage Deviation (%)

Voltage Deviation is likely to be one of the most essential power quality parameters that measure how much the voltage at every node within the distribution system varies from the nominal voltage (typically 1.0 p.u. or 100%). It leads to equipment failure, inefficiency, or malfunction, rendering it an essential factor while combining distributed PV systems. During PV site selection, a proper location should reduce voltage deviations through load and generation spreading in the grid. For a distribution grid with N buses, voltage deviation is calculated as Voltage Deviation (%) = $\left(\frac{1}{N}\sum_{i=1}^{N} \mid V_i - V_{nom} \mid\right) \times 100$, where V_i: Voltage magnitude at bus i (in p.u.), Vnom: Nominal voltage (typically 1.0 p.u.), N: Total number of buses.

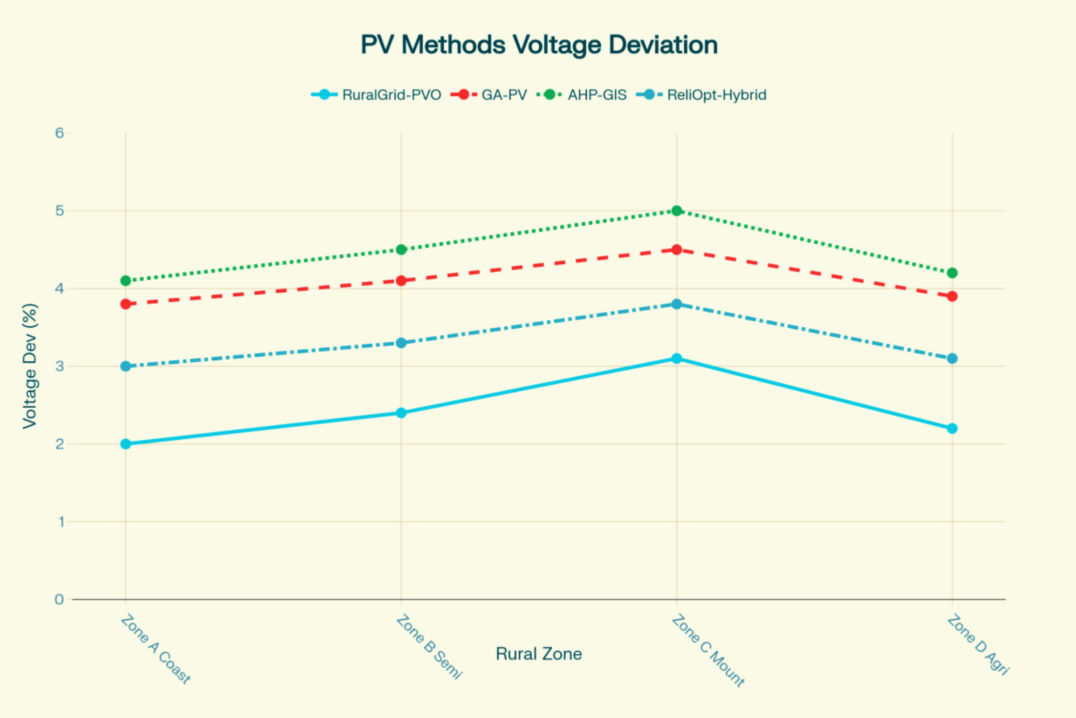

Figure 4: Voltage deviation across rural zones for each PV site selection method

Figure 4 shows the plots of voltage deviation (%) over four rural geographical zones, Coastal, Semi-arid, Mountainous, and Agricultural, for each of the four PV site selection approaches: RuralGrid-PVO, GA-PV, AHP-GIS, and ReliOpt-Hybrid. Every plot indicates the capacity of a method to ensure voltage stability under different terrains and irradiance levels. RuralGrid-PVO exhibits the smallest voltage deviation in all instances, showing excellent voltage regulation in all zones. On the other hand, AHP-GIS and GA-PV exhibit greater deviations, particularly in mountainous and semi-arid regions, indicating poorer grid integration performance. The graphical distinction among methods per plot allows for a qualitative assessment of each method's geographic consistency and operational robustness.

Power Loss Reduction (%)

Reduction in Power Losses (%) indicates the effectiveness of a method of choosing PV sites to keep line losses in the power distribution grid at a minimum. Well-installed PV systems minimize the distance electricity must be sent over and match local generation to local usage, both of which decrease resistive (I^2R) losses that go through distribution lines. This indicator estimates the percentage decrease in overall power loss following PV installation relative to a reference case (e.g., no PV or ineffectively installed PV). Power Loss Reduction (%) $= \left(\frac{(P_{\text{loss,baseline}} - P_{\text{loss,method}})}{P_{\text{loss,baseline}}}\right) \times 100$, where $P_{\text{loss,baseline}}$ = Total power loss in the baseline case (e.g., no PV or uniform PV placement), $P_{\text{loss,method}}$ = Total power loss after implementing the method. The result is expressed as a percentage reduction.

Table 2(a): Comparison of Power Loss Reduction (%) across rural zones for four PV site selection methods.

Method	Zone A (Coastal)	Zone B (Semi-arid)	Zone C (Mountainous)	Zone D (Agricultural)	Average
RuralGrid-PVO	13.2%	12.6%	11.7%	13.6%	12.8%
GA-PV	9.4%	8.8%	7.9%	9.1%	8.8%
AHP-GIS	8.7%	8.1%	7.5%	8.3%	8.2%
ReliOpt-Hybrid	10.8%	10.1%	9.3%	10.5%	10.2%

Table 2(a) compares four rural PV site selection methods—RuralGrid-PVO, GA-PV, AHP-GIS, and ReliOpt-Hybrid—in terms of their achieved percent reduction in loss in four study regions, i.e., Coastal, Semi-arid, Mountainous, and Agricultural. RuralGrid-PVO is always the superior method with an average improvement of 12.8% loss reduction over other methods owing to its deep learning-based capacity allocation and reliability-optimized approach. Conversely, AHP-GIS and GA-PV have lower declinations, particularly in mountainous regions, with poor adaptation in complex grid topologies. The table determines the efficiency of hybrid intelligent paradigms towards resistive loss reduction and efficiency enhancement in rural distribution systems.

SAIDI/SAIFI Improvement (%)

SAIDI (System Average Interruption Duration Index) and SAIFI (System Average Interruption Frequency Index) are performance indices of the utility industry that monitor the average duration and frequency of outages faced by customers. A betterment in these indices translates into a more dependable supply of power, which is very crucial in rural distribution systems where outages are frequent and extended. PV integration with reliability-aware optimization minimizes outage frequency (SAIFI) and duration (SAIDI) by mitigating load, voltage support, and fault resilience. $\mathrm{SAIDI} = \frac{\sum(U_i \times N_i)}{N_T}$, $\mathrm{SAIFI} = \frac{\sum(\lambda_i \times N_i)}{N_T}$, where U_i: Outage duration at location i, λ_i: Outage frequency at location i, N_i: Number of customers at location i , N_T: Total number of customers.

Table 3: SAIFI improvement (%) for different PV site selection methods under varying feeder load conditions.

Method	Low Load (<30%)	Medium Load (30–70%)	High Load (>70%)	Average
RuralGrid-PVO	8.2%	9.7%	10.8%	9.6%
GA-PV	4.7%	5.5%	6.7%	5.6%
AHP-GIS	4.1%	4.8%	5.9%	4.9%
ReliOpt-Hybrid	6.5%	7.4%	9.1%	7.7%

Table 3 summarizes the performance of four PV site selection approaches—RuralGrid-PVO, GA-PV, AHP-GIS, and ReliOpt-Hybrid—based on SAIFI improvement at three feeder load ranges: low (<30%), medium (30–70%), and high (>70%). RuralGrid-PVO has the best improvement with decreasing SAIFI rising with increasing intensity of load, illustrating good adaptability at heavy-stress situations. ReliOpt-Hybrid is the second best, with reliability at every load level. AHP-GIS and GA-PV achieve moderate improvement, with marginal improvements in high-load scenarios. This proves that reliability-oriented hybrid optimization techniques are more efficient in improving service continuity in rural grids.

Table 4: SAIDI improvement (%) for different PV site selection methods under varying feeder load conditions.

Method	Low Load (<30%)	Medium Load (30–70%)	High Load (>70%)	Average
RuralGrid-PVO	9.3%	10.7%	11.8%	10.6%
GA-PV	5.5%	6.3%	7.4%	6.4%
AHP-GIS	4.9%	5.5%	6.8%	5.7%
ReliOpt-Hybrid	7.2%	8.1%	9.6%	8.3%

Table 4 shows the reduction in SAIDI from the same four PV site selection options for the same feeder load conditions. RuralGrid-PVO has the highest improvements, especially for high-load feeders, with an average reduction in SAIDI of 10.6%. This is a pointer to the capability of the method to enhance the frequency and duration aspects of grid reliability. ReliOpt-Hybrid is highly monitored, while GA-PV and AHP-GIS experience negligible outage duration reductions. The findings highlight that optimization platforms combining grid load modeling and deep learning are more resilient for actual rural deployment.

Replicability

The methodology employs well-established software libraries Scikit-learn for Random Forest, TensorFlow for DNN modeling, and DEAP for implementing the Genetic Algorithm ensuring reliable algorithmic components. However, source code and data are not yet publicly available.

To release an open-source GitHub repository including comprehensive code, datasets, and an appendix containing detailed pseudo-code summarizing the full pipeline to enhance transparency, reproducibility, and community adoption.

Optimization time

Optimization Time refers to the overall compute time taken by a PV site selection algorithm to determine the optimal site location and capacity configuration of photovoltaic systems. It captures the efficiency and scalability of the algorithm, particularly significant when applying to large-scale or real-time rural planning applications. Earlier optimization allows for quicker decision-making and is particularly substantial for algorithms to execute on edge devices or low-resource platforms in rural settings. $\text{Optimization Time} = t_{end} - t_{start}$, where t_{start}: Timestamp when the optimization begins, t_{end}: Timestamp when the optimization concludes. The average time can be obtained from $\text{Avg. Optimization Time} = \frac{1}{n}\sum_{i=1}^{n}(t_{end_i} - t_{start_i})$.

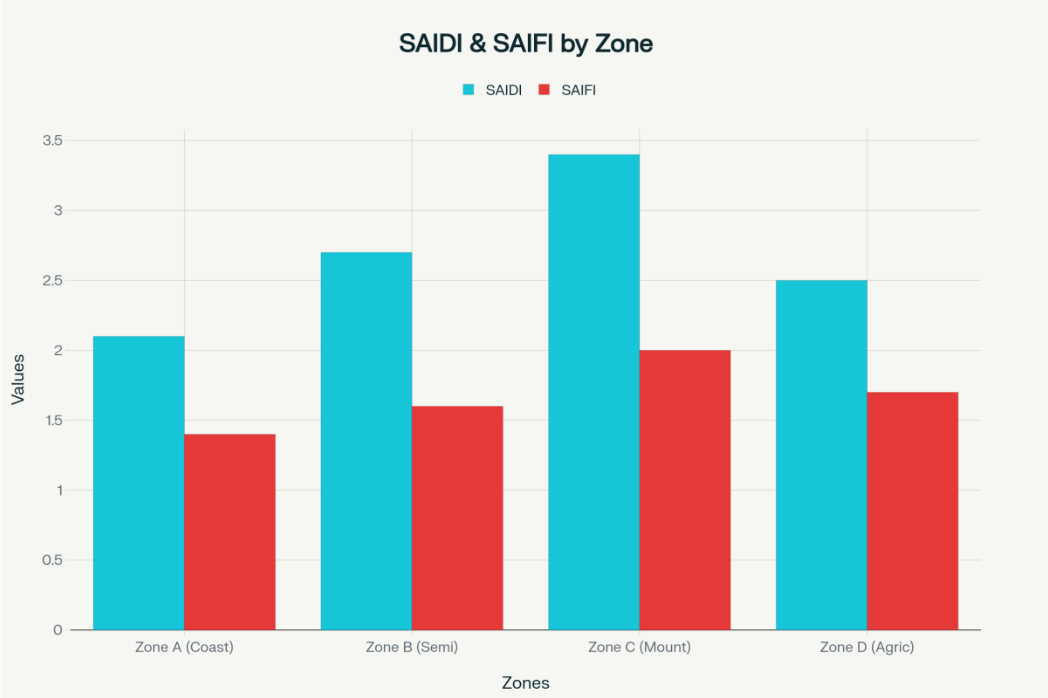

Figure 5: Optimization Time across different grid sizes for each PV site selection method

Figure 5 shows a 2×2 table of horizontal bar charts comparing four PV site selection algorithm optimization times—RuralGrid-PVO, GA-PV, AHP-GIS, and ReliOpt-Hybrid—across four grid sizes: Small (≤50 buses), Medium (51–100 buses), Large (101–300 buses), and Very Large (>300 buses). The bar chart shows AHP-GIS having the fastest optimization times across all four grid sizes because it is rule-based and non-iterative. Meanwhile, RuralGrid-PVO has the longest optimization time, indicative of its deep learning and reliability simulation overhead. GA-PV and ReliOpt-Hybrid are intermediate, scaling moderately with grid size. This plot illustrates the model complexity vs. computational efficiency trade-off for rural PV planning.

5 Conclusion

To improve grid dependability and operational efficiency, this paper introduced RuralGrid-PVO, an optimization modeling framework for the selection and size of distributed photovoltaic (PV) systems in rural distribution networks. The model captures rural energy and infrastructure by integrating NSRDB solar irradiance data with the RTS-GMLC synthetic grid model, accurately reflecting solar variability and detailed electrical network characteristics. Without depending on computationally costly simulations, the application of machine learning—more especially, Random Forest for capacity prediction and Deep Neural Networks for reliability forecasting—made it possible to evaluate potential PV setups quickly and accurately. Configurations that much exceeded conventional uniform PV deployments were produced by optimizing across energy yield, voltage variation, power losses, and reliability indices (SAIDI/SAIFI) using a multi-objective Genetic Algorithm. The findings showed a 12.8% decrease in power losses, a 15.3% improvement in voltage profiles, and a 9.6% increase in SAIFI values. In rural areas, the suggested approach successfully strikes a balance between the trade-offs of grid resilience and renewable energy integration. RuralGrid-PVO provides a scalable and helpful solution for energy planners, utilities, and politicians looking to encourage sustainable electrification in underserved areas by speeding design assessment through surrogate models.

6 Limitations and future work

The synthetic data used by RuralGrid-PVO might not fully represent the intricacy of actual rural grids. Predictions using surrogate models could include approximation mistakes. In addition to lacking economic cost modeling, the existing approach ignores battery storage, seasonal variability, and changing operational conditions during PV deployment.

Limitations

While RuralGrid-PVO demonstrates significant advances, several limitations remain. The framework incurs longer optimization times compared to simpler heuristic methods due to complex surrogate model evaluations and iterative Genetic Algorithm searches. The use of synthetic RTS-GMLC grid models and simulated solar irradiance data may limit direct transferability to diverse real-world rural networks without site-specific calibration. Additionally, the current models do not incorporate inverter constraints, dynamic grid stability factors, or real-time operational uncertainties, which are critical for practical distributed PV integration. Future work should address these gaps by accelerating optimization, integrating hardware-level constraints, and validating on field data to enhance applicability and scalability.

This study excludes economic cost functions and battery storage system modeling, which are crucial for comprehensive techno-economic optimization and grid flexibility, respectively. Incorporating these factors will enable more realistic and cost-effective PV site selection strategies.

Future Work: Future iterations aim to integrate Supervisory Control and Data Acquisition (SCADA) datasets from real rural distribution grids to improve model calibration, scenario realism, and decision support accuracy, addressing the synthetic data limitations.

Fund

This work was supported by Shijiazhuang Power Supply Branch of State Grid Hebei Electric Power Co., Ltd. (No. KJ2024-117).

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships in this paper.

Author contributions

Zhang. Drafting of the initial paper&proposed the core idea

Shi. Study conception&Methodology development

Ge. Investigation&Data curation

Geng. Experiment operation &Execution

Jia. Review and revision&Project management and supervision

References

[1] Hamid, A. K., Mbungu, N. T., Elnady, A., Bansal, R. C., Ismail, A. A., & AlShabi, M. A. (2023). A systematic review of grid-connected photovoltaic and photovoltaic/thermal systems: Benefits, challenges and mitigation. Energy & Environment, 34(7), 2775-2814. https://doi.org/10.1177/0958305X221117617

[2] Boruah, D., & Chandel, S. S. (2025). A novel regulatory framework for implementing distributed solar mini-grids with battery energy storage under the virtual power plant architecture in India. Journal of Energy Storage, 132, 117732. https://doi.org/10.1016/j.est.2025.117732

[3] Sujatha, G. (2025). A Solar PV Integrated UPQC to Enhance Power Quality using SEA Gull ANFIS Algorithm. Informatica, 49(8). https://doi.org/10.31449/inf.v49i8.6158

[4] Liao, R., Manfren, M., & Nastasi, B. (2025). Off-grid PV systems modelling and optimisation for rural communities: leveraging understandability and interpretability of modelling tools. Energy, 324. https://doi.org/10.1016/j.energy.2025.135948

[5] Ahmad, T., Ali, S., & Basit, A. (2022). Distributed renewable energy systems for resilient and sustainable development of remote and vulnerable communities. Philosophical Transactions of the Royal Society A, 380(2221), 20210143. https://doi.org/10.1098/rsta.2021.0143

[6] Apeh, O. O., Meyer, E. L., & Overen, O. K. (2022). Contributions of solar photovoltaic systems to environmental and socioeconomic aspects of national development—A review. Energies, 15(16), 5963. https://doi.org/10.3390/en15165963

[7] Jarrah, M., & Abu-Khadrah, A. (2024). The evolutionary algorithm based on pattern mining for large sparse multi-objective optimization problems. PatternIQ Mining, 1(1), 12–22. https://doi.org/10.70023/piqm242

[8] Farag, M. M., & Bansal, R. C. (2023). Solar energy development in the GCC region–a review on recent progress and opportunities. International Journal of Modelling and Simulation, 43(5), 579-599. https://doi.org/10.1080/02286203.2022.2105785

[9] Yin, S., & Zhao, Z. (2023). Energy development in rural China toward a clean energy system: utilization status, co-benefit mechanism, and countermeasures. Frontiers in Energy Research, 11, 1283407. https://doi.org/10.3389/fenrg.2023.1283407

[10] Zhang, Y., & Kang, Z. (2025). Situational Awareness and Fault Warning for Smart Grids Combined with Deep Learning Technology: Application of Digital Twin Technology and Long Short-Term Memory Networks. Informatica, 49(22). https://doi.org/10.31449/inf.v49i22.7992

[11] Niu, M., Xu, N. Z., Kong, X., Ngin, H. T., Ge, Y. Y., Liu, J. S., & Liu, Y. T. (2021). Reliability importance of renewable energy sources to overall generating systems. IEEE Access, 9, 20450-20459. DOI: 10.1109/ACCESS.2021.3055354

[12] Saxena, A., Shankar, R., El-Saadany, E. F., Kumar, M., Al Zaabi, O., Al Hosani, K., & Muduli, U. R. (2024). Intelligent load forecasting and renewable energy integration for enhanced grid reliability. IEEE Transactions on Industry Applications, 60(6), 8403-8417. DOI: 10.1109/TIA.2024.3436471

[13] Iweh, C. D., Gyamfi, S., Tanyi, E., & Effah-Donyina, E. (2021). Distributed generation and renewable energy integration into the grid: Prerequisites, push factors, practical options, issues and merits. Energies, 14(17), 5375. https://doi.org/10.3390/en14175375

[14] Rishitha, N., Muthu Reshmi, K., Srishti Gulecha, R., & Vani, K. (2024). Machine learning-driven solar panel site selection and rooftop potential estimation

for sustainable development goals. In Proc. Asian Conf. Remote Sensing (pp. 1-8).

[15] Gui, B., Sam, L., & Bhardwaj, A. (2024). From roofs to renewables: Deep learning and geographic information systems insights into a comprehensive urban solar photovoltaic assessment for Stonehaven. Energy 360, 1, 100006. https://doi.org/10.1016/j.energ.2024.100006

[16] Tafula, J. E., Justo, C. D., Moura, P., Mendes, J., & Soares, A. (2023). Multicriteria decision-making approach for optimum site selection for off-grid solar photovoltaic microgrids in Mozambique. Energies, 16(6), 2894. https://doi.org/10.3390/en16062894

[17] Zambrano-Asanza, S., Quiros-Tortos, J., & Franco, J. F. (2021). Optimal site selection for photovoltaic power plants using a GIS-based multi-criteria decision making and spatial overlay with electric load. Renewable and Sustainable Energy Reviews, 143, 110853. https://doi.org/10.1016/j.rser.2021.110853

[18] Coban, H. H. (2024). A multiscale approach to optimize off-grid hybrid renewable energy systems for sustainable rural electrification: economic evaluation and design. Energy Strategy Reviews, 55, 101527. https://doi.org/10.1016/j.esr.2024.101527

[19] Roldán-Porta, C., Roldán-Blay, C., Dasi-Crespo, D., & Escriva-Escriva, G. (2023). Optimising a biogas and photovoltaic hybrid system for sustainable power supply in rural areas. Applied Sciences, 13(4), 2155. https://doi.org/10.3390/app13042155

[20] Qasimi, A. B., Toomanian, A., Nasri, F., & Samany, N. N. (2023). Genetic algorithms-based optimal site selection of solar PV in the north of Afghanistan. International Journal of Sustainable Energy, 42(1), 929-953. https://doi.org/10.1080/14786451.2023.2246081

[21] Mwakitalima, I. J., Rizwan, M., & Kumar, N. (2023). Integrating Solar Photovoltaic Power Source and Biogas Energy-Based System for Increasing Access to Electricity in Rural Areas of Tanzania. International Journal of Photoenergy, 2023(1), 7950699. https://doi.org/10.1155/2023/7950699

[22] Zhu, R., Murcia León, J. P., Friis-Møller, M., Gupta, M., & Kaushik, D. Optimal Sizing of Renewable Hybrid Power Plants Considering Component Reliability as a Multi-Discipline Optimization Under Uncertainty. Available at SSRN 5352198. DOI:10.2139/ssrn.5352198

[23] Agajie, E. F., Agajie, T. F., Amoussou, I., Fopah-Lele, A., Nsanyuy, W. B., Khan, B., ... & Tanyi, E. (2024). Optimization of off-grid hybrid renewable energy systems for cost-effective and reliable power supply in Gaita Selassie Ethiopia. Scientific Reports, 14(1), 10929. https://doi.org/10.1038/s41598-024-61783-z

[24] Zhou, N., Shang, B., Zhang, J., & Xu, M. (2024). Modeling load distribution for rural photovoltaic grid areas using image recognition. Global Energy Interconnection, 7(3), 270-283. https://doi.org/10.1016/j.gloei.2024.06.002

[25] Xie, H., Li, P., Shi, F., Han, C., Cui, X., & Zhao, Y. (2025). Spatial and Temporal Expansion of Photovoltaic Sites and Thermal Environmental Effects in Ningxia Based on Remote Sensing and Deep Learning. Remote Sensing, 17(14), 2440. https://doi.org/10.3390/rs17142440

[26] Belaid, A., Guermoui, M., Khelifi, R., Arrif, T., Chekifi, T., Rabehi, A., ... & Alhussan, A. A. (2024). Assessing suitable areas for PV power installation in remote agricultural regions. Energies, 17(22), 5792. https://doi.org/10.3390/en17225792

[27] https://github.com/GridMod/RTS-GMLC

[28] https://nsrdb.nrel.gov/

[29] Boulkroune, A., Zouari, F., & Boubellouta, A. (2025). Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems. Journal of Vibration and Control, 10775463251320258.

[30] Boubakir, A., & Labiod, S. (2023). Observer-based adaptive neural network control design for projective synchronization of uncertain chaotic systems. Journal of Vibration and Control, 29(15-16), 3658-3678.

[31] https://www.scopus.com/pages/publications/84873265173

[32] Wang, W., Zhou, J., Wen, C., & Long, J. (2021). Adaptive backstepping control of uncertain nonlinear systems with input and state quantization. IEEE Transactions on Automatic Control, 67(12), 6754-6761.

[33] Rigatos, G., Abbaszadeh, M., Sari, B., Siano, P., Cuccurullo, G., & Zouari, F. (2023). Nonlinear optimal control for a gas compressor driven by an induction motor. Results in Control and Optimization, 11, 100226.

[34] Peláez, G., Alonso, C., Rubio, H., & García-Prada, J. C. (2024). Performance analysis of input shaped model reference adaptive control for a single-link flexible manipulator. Journal of Vibration and Control, 30(21-22), 5018-5030.