EEET ECOLOGICAL ENGINEERING & ENVIRONMENTAL TECHNOLOGY

Ecological Engineering & Environmental Technology, 2025, 26(9), 109–120
https://doi.org/10.12912/27197050/209482
ISSN 2719–7050, License CC-BY 4.0

Received: 2025.07.15
Accepted: 2025.08.16
Published: 2025.09.01

Forecasting and optimization of a residential off-grid solar photovoltaic-battery energy storage system

Riad Mollik Babu[1*], Md Shahidul Islam[1], Enamul Basher[1]

[1] Department of Electrical and Electronic Engineering, University of Asia Pacific, 74/A Green Rd, Dhaka 1205, Bangladesh

* Corresponding author's e-mail: riadmallickbabu@gmail.com

ABSTRACT

Achieving full energy independence in residential settings remains a challenge with conventional grid-connected solar photovoltaic (PV) systems. This study presents a data-driven approach to designing a grid-independent solar PV and battery energy storage system (BESS) using machine learning techniques. Actual residential load profiles and solar generation data, combined with weather and calendar-based inputs, were analyzed using four regression models: Polynomial Regression, Support Vector Regression, Gradient Boosting, and Random Forest. These models were evaluated based on R^2, mean absolute error (MAE), and root mean square error (RMSE) to determine the most accurate predictor. The Random Forest model delivered the highest accuracy, achieving an R^2 of 0.92 for generation and 0.90 for load forecasting. The optimized forecast supported the expansion of the PV system from 11 kW to 30 kW and the integration of a 120 kWh BESS, ensuring complete residential grid independence. This framework offers a reliable and adaptable methodology for intelligent energy management and autonomous PV–BESS system design.

Keywords: off grid solar PV system, BESS, machine learning forecasting, renewable energy prediction models, system optimization.

INTRODUCTION

Global energy demand is rising rapidly due to the growing human population, urbanization, and technological advancement. A strong correlation exists between urbanization and energy consumption, as modern urban lifestyles require continuous access to electricity and energy-intensive infrastructure (Rathore et al., 2019). Historically, fossil fuels have been the dominant source of energy, meeting the majority of global demand over the past four decades (Ghasemian et al., 2024). However, in recent years, the share of renewable energy in the global electricity mix has grown substantially. Currently, approximately 34.5% of global electricity is generated from renewable sources. Projections indicate that by 2030, this figure will increase to 45.6%, with solar photovoltaic (PV) technology emerging as the leading contributor. Solar PV alone is expected to account for 16.1% of the global electricity supply, surpassing other forms of renewable energy (Renewables Energy System - IEA, n.d.). While grid-connected solar PV remains the dominant technology, integrating PV into the distribution system introduces power quality challenges, including voltage fluctuations, load harmonics, and reactive power, all of which require dedicated control and compensation. Off-grid solar PV systems operate independently of the utility, avoiding grid-related power quality issues and electricity charges. Excess generation can be stored in batteries under intelligent energy management to supply demand during low-generation periods and smooth the output (Karthikeyan et al., 2017; Obi and Bass, 2016; Shafiullah et al., 2022; Sharma et al., 2016). Consequently, the increasing global demand for sustainable and decentralized energy systems has accelerated the adoption of solar photovoltaic (PV) technologies in residential applications (Khan, 2020).

In Bangladesh, the solar energy sector has made significant progress over the past two decades, particularly in off-grid applications (Shapna et al., 2025). The country has implemented one of the largest solar home systems (SHS) programs globally. At its peak in 2013, the SHS program installed more than 861,000 off-grid solar PV units. By 2018, it had provided renewable electricity to approximately 20 million people, accounting for nearly 12.5% of the national population at the time (Hellqvist and Heubaum, 2023). This achievement has been largely driven by national policy support and initiatives led by institutions such as the Infrastructure Development Company Limited (IDCOL) (Nahid et al., 2024). With an average daily solar radiation ranging from 4.5 to 5 kWh/m², Bangladesh possesses strong potential for solar energy harvesting across most regions (Babu and Basher, 2024).

However, most SHS deployments follow conventional design approaches that lack adaptive control and intelligent management. In terms of enabling productive use of energy, these systems face clear limitations. To ensure reliability, the systems are typically over-sized, which often results in substantial excess capacity. This excess energy is effectively wasted, reducing the overall efficiency of the system. Such design limitations often result in inefficient utilization of solar resources, underperformance of storage systems, and reliability issues during extended periods of low solar generation (Chowdhury et al., 2015). Moreover, as household energy demands evolve and climate conditions become increasingly variable, there is a pressing need to transition from static design methodologies to smart, data-driven solutions (Ahangar et al., 2023). Integrating forecasting techniques and optimization algorithms into solar PV systems can significantly enhance their autonomy, efficiency, and resilience in practical conditions (Shafiei et al., 2025).

In previous studies, the study in (Patel and Swathika, 2024) aims to predict solar energy output using four machine learning methods: Support Vector Machines, Ensemble of Trees, Gaussian Process Regression, and Neural Networks. The study aligns with the United Nations' Sustainable Development Goals and uses meteorological parameters and hourly global solar radiation to predict electricity output from photovoltaic panels. Bayesian Optimization is used to optimize the models. The study shows Ensemble of Trees performs best across all datasets and requires shorter training time. However, the analysis is limited to a single geographic location, which may affect the generalizability of the results across different climatic zones. (Liu and Gou, 2025) proposes a hybrid framework for residential photovoltaic systems, combining physical energy flow constraints with XGBoost-based machine learning for robust forecasting. Two optimization strategies, proximal policy optimization (PPO) and rule-based control (RBC), are developed for charge-discharge scheduling, incorporating grid stability metrics. While the proposed framework significantly improves prediction accuracy, energy self-sufficiency rate, and power fluctuations. The analysis evaluates only two control strategies without real-world validation, which may limit the practical applicability of the results. Khaoula et al., 2025 presents a hybrid deep learning model for accurate long-term electricity demand forecasting, combining BiLSTM networks and Convolutional Neural Networks. The model, based on hourly data from Morocco and Spain, captures seasonal, meteorological, and socioeconomic factors influencing power usage. The model provides reliable 30-day forecasts, demonstrating its versatility and potential for utilities and regulators to address long-term demand uncertainties and promote renewable integration (Boumais and Messaoudi, 2025). Abdullah et al., 2017 uses support vector regression (SVR), Polynomial Regression, and Lasso for hour-ahead solar PV power forecasting. The SVR forecasting model outperforms other models in terms of accuracy, based on features like weather conditions, power generation, and day and time information. However, the accuracy of SVR tends to decline when more features are added, including temperature, humidity, visibility, wind speed, and wind direction, as this introduces noise into the dataset and increases the risk of overfitting (Alfadda et al., 2017). Muhammad et al., 2018 compares the accuracy, stability, and computational cost of random forest (RF) and extra trees (ET) models for predicting hourly PV generation output. All models have similar predictive power and are equally applicable. However, ET outperforms RF and SVR in terms of computational cost. The stability and algorithmic efficiency of ETs make them ideal for wider PV output forecasting deployment. A key limitation of the current study is that it does not incorporate fault detection, weather classification, or multi-timescale forecasting, which may limit its practical deployment under diverse operating conditions (Ahmad et al., 2018). To improve prediction accuracy, Liu and Sun (2019) proposed a new model using Principal Component Analysis, K-means clustering, and Random Forest algorithm. The model uses filtered input data and Random Forest parameters to avoid artificial filtering. Comparative experiments show the model has higher prediction accuracy and robustness. However, there are several limitations that should be addressed in future work. If the lead time is extended too far or the input variables have weak correlation with the output, the algorithm's performance may deteriorate. The hybrid clustering algorithm may also place excessive emphasis on differences within a single variable while overlooking subtle but important changes in more stable variables (Liu and Sun, 2019). Jesus et al., 2023 investigates power forecasting for building-integrated photovoltaic (BIPV) systems installed on vertical façades using decision tree-based machine learning algorithms. Utilizing the Python-based skforecast library, both deterministic and probabilistic forecasting approaches were implemented. Deterministic forecasts using XGBoost and Random Forest showed improved accuracy with the inclusion of exogenous variables. Probabilistic forecasting was conducted using XGBoost with Bootstrap. Results demonstrate that both models are effective for BIPV power prediction, achieving mean absolute errors of approximately 40% for the south-facing array and below 30% for the east-facing array. However, the combination of different machine learning algorithms and forecasting strategies, such as recursive multi-step forecasting, hyperparameter tuning, and back-testing, could lead to more accurate results (Polo et al., 2023).

This study introduces a novel data driven framework for the forecasting and optimization of a residential off grid solar PV and BESS using multiple ML techniques. While existing approaches often rely on either conventional sizing methods or single model forecasting, this work integrates four distinct regression models: polynomial regression, support vector regression, gradient boost, and random forest to ensure robust performance evaluation and optimal model selection.

The key novelty and contributions of this research are summarized as follows:

- A comparative machine learning approach is proposed for both solar PV generation and residential electricity demand forecasting. This allows for a comprehensive evaluation of predictive performance using real world weather and consumption data.
- The model is trained and validated using actual field data from a residential site in Bangladesh, combined with weather parameters from NASA and local meteorological sources. This enhances prediction accuracy and practical relevance.
- The Random Forest model is identified as the most accurate predictor based on R^2, MAE, and RMSE values, achieving R^2 scores of 0.92 and 0.90 for generation and demand forecasting, respectively.
- The optimized forecasts are applied to design and size an off grid solar PV BESS system, scaling the PV capacity from 11 kW to 30 kW and integrating a 120 kWh BESS to achieve full grid independence.
- The proposed framework serves as a foundation for autonomous residential energy systems, enabling smart control, better resource utilization, and reduced energy wastage in off grid applications.

By addressing both forecasting accuracy and system optimization in a unified framework, this study contributes an intelligent methodology for enhancing the resilience and autonomy of residential solar energy systems in data constrained environments like Bangladesh.

This study addresses the lack of integrated machine learning forecasting and optimization frameworks for off-grid PV–BESS systems in Bangladesh, where prior works often rely on single predictive models or simulated datasets. It is based on the premise that evaluating and comparing multiple machine learning regression models using real residential load and weather data will produce forecasts with higher accuracy than single-model approaches, thereby enabling more precise PV–BESS sizing and enhancing system autonomy. The novelty lies in combining predictive modeling with technical optimization to design a fully grid-independent system. The study aims to demonstrate that enhanced forecasting accuracy directly results in more reliable and efficient PV–BESS configurations for residential applications in resource-constrained regions.

METHODOLOGY

This section presents a comprehensive method of forecasting generated power from the solar PV plant and energy demand for the residential house. After collecting the required data, four machine learning algorithms developed in Python are used to accomplish the forecasting on the Google Colab platform. After predicting data using distinct machine learning algorithms such as Polynomial Regression (PR), Support Vector Regression (SVR), Gradient Boost (GB) and Random Forest (RF), the performance is evaluated by R2 score, mean absolute error and root mean square error values. The algorithm delivers the best prediction among others, is selected as a model to optimize the solar plant and develop the BESS system. Figure 1 illustrates the overall methodology used in this study, starting from data collection and pre-processing, followed by the implementation of four different machine learning regression models, model evaluation using statistical metrics, and final system optimization for PV–BESS sizing.

Data collection

To forecast power generation and demand using machine learning models, a complete dataset is essential for training and testing. For solar PV generation forecasting, weather parameters including direct horizontal irradiance (DHI), direct normal irradiance (DNI), global horizontal irradiance (GHI), and temperature were used as independent features. For electricity demand forecasting, additional inputs included relative humidity, time-periodical data such as season, and day of the week.

Hourly weather data for the study site (coordinates: 23.8103° N, 90.4125° E) covering the period 1 January 2023 to 31 December 2023 were obtained from the NASA Prediction of Worldwide Energy Resources (POWER) database and cross-verified with the Bangladesh Meteorological Department (BMD) records. The dataset contained 8.760 hourly records for each variable. The dependent features, like electricity consumption and solar power generation, are collected from the real field.

The dataset can be explored via the NASA POWER Data Access Viewer. This API call automatically returns the 2023 hourly data for the specified coordinates and variables, ensuring full reproducibility for other researchers.

Data availability – the dataset used in this study is publicly available through the direct API link provided above. The processed PV generation and load demand datasets from the study site will be made available in an open-access repository (e.g., Zenodo) upon acceptance of the manuscript, with a DOI provided in the final version.

Polynomial regression

A simple and fundamental regression model is used to predict continuous data when a non-linear relation exists between the target variable and independent features. It is a modified version of the linear regression model, which provides

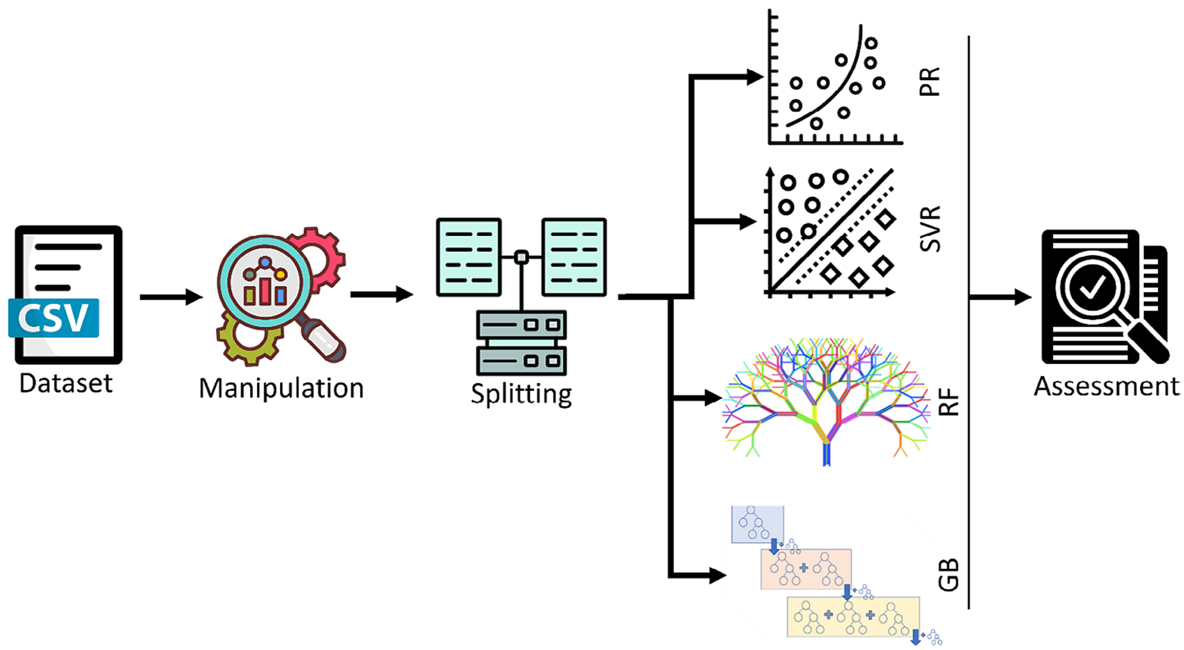

Figure 1. Methodology of machine learning modelling for the proposed system

better results than linear regression in case of a nonlinear pattern of data. To estimate the predicted value, the regression model calculates several parameters in a specific process. The process initiates with a number of random coefficients and then calculates the prediction errors resulting from applying those coefficients. Subsequently, the algorithm updates the coefficients till it ensures the minimum error in each point. After this process, the algorithm develops a best-fit curve on which the predicted points are located and closely match the real data. The mathematical equation of a polynomial regression curve is expressed as equation 1.

$$y = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \cdots + b_n x^n \quad (1)$$

where: y is the point of the fitted curve, x is the independent variable, b_0, b_1, b_2 …b_n are the coefficients of the regression model and n represents the degree of the polynomial curve. A lower degree reduces accuracy by an underfitted curve and a higher degree also fails to predict precisely due to overfitting of the curve. Since a nonlinear relationship exists between the independent data and the target variables of the proposed system, the PR algorithm is used in forecasting.

Support Vector Regression

Instead of reducing error between a predicted line and actual data as like conventional regression models, SVR reduces error between a tube or margin area and the actual data. The tube or margin area is specified around the best-fit curve of the SVR model. In this case, the error is ignored when the point of actual data exists in the area of the margin of tolerance. Although this algorithm spends more time on training data, it makes the model robust against minor noise. The tube that allocates the margin of tolerance is called the epsilon tube. The data points lying outside the tube are used to build a model. The width of the epsilon tube depends on a parameter ε where a smaller value of ε tends to overfit the curve. On the other hand, a larger value of ε tends to underfit the curve. To develop the curve, various kernel functions are used in different cases. In the proposed system, the Gaussian kernel, also known as the radial basis function (RBF), is generally used for capturing complex nonlinear patterns in high-dimensional space data. The equation of the Gaussian kernel is given below.

$$K(x - x') = e^{(-\gamma \|x - x'\|^2)} \quad (2)$$

where: x and x' are two vectors of data and $\gamma\gamma$ is called the tunable parameter. The equation calculates the similarity score of the input vectors. The tunable parameter is adjusted before training, which can be chosen by cross-validation.

Random Forest

An ensemble learning algorithm is applicable in both classification and regression. Ensemble learning is a hybrid model that combines multiple weak learning models to enhance the predictive result. Among the three types of ensemble models, known as Bootstrap Aggregating, Stacking and Boosting, a Bootstrap Aggregating algorithm RF is selected for the system. The algorithm trains multiple decision tree models simultaneously, then provides a combined result as the average of all decision tree models. Consequently, the RF is a forest of decision trees that aggregates the results of all trees. However, the fundamental block of the RF decision tree acts like a flowchart. The flow begins with selecting the best feature based on the splitting criterion. For the regression model, the mean square error (MSE) is applied as a splitting criterion. After selecting the best feature, the dataset is split into multiple groups. The splitting process continues repeatedly creating multiple nodes (subgroups) from each group until each node contains the minimum number of sample data of the same class. To calculate MSE, the model begins with mean of the data, then calculates MSE using the following equation.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^2 \quad (3)$$

where: n indicates the number of data points in node, y_i and $\overline{y}$ presents the actual value of the nth sample and the mean of all target values in the node.

The node of the lowest MSE is selected as the best feature. During the splitting process, every node contains the mean value of that node and MSE calculation process is kept continuous. Figure 2 illustrates the architecture of the Random Forest regression model used in this study. It shows how multiple decision trees are trained on

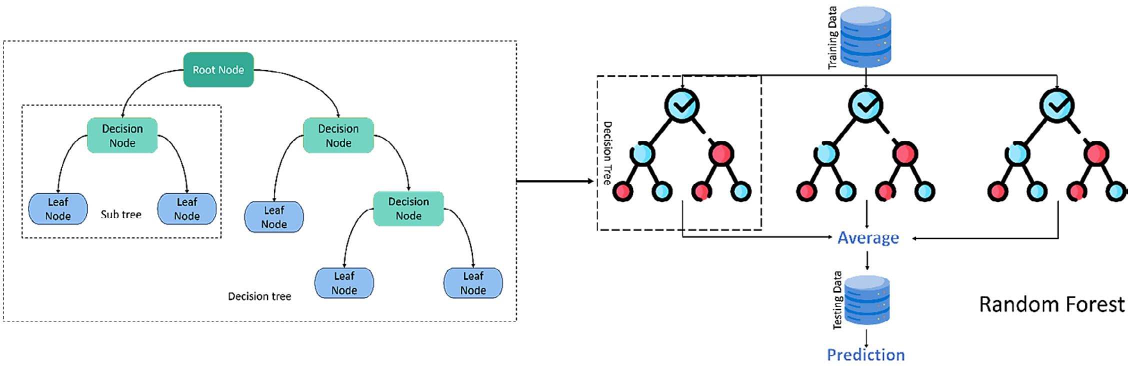

Figure 2. Architecture of random forest regression model

bootstrapped subsets of the dataset, with predictions aggregated to produce a final output. This ensemble approach reduces overfitting, improves prediction accuracy, and was found to be the most effective among the models tested in this work.

Gradient Boost

The GB machine learning algorithm is another ensemble boosting-type learning model that trains data using several decision tree models. The main distinction between RF and GB is that RF trains multiple decision tree models simultaneously, while GB trains them sequentially. The working principle of the decision tree model has been briefed in the RF section; therefore, adding those decision tree (DT) models is the remaining portion of the GB process, which minimizes the function loss and increases the accuracy of the result. The addition of weak model outputs is made by following equation.

$$F_m(x) = F_{m-1}(x) + \eta \sum_{m=1}^{M} h_m(x) \quad (4)$$

where: $F_m(x)$ and $h_m(x)$ are the outputs of the GB model and the DT model, m is the number of DT blocks and η is the learning rate.

Figure 3 shows the architecture of the Gradient Boost regression model applied in this study. In this sequential ensemble approach, each decision tree is trained to correct the residual errors of the previous tree, with outputs combined through a learning rate to minimize overall prediction loss. This method enables the model to capture complex non-linear relationships in the data, making it suitable for solar generation and load forecasting.

Evaluation

Since four algorithms are used to forecast the generation and energy demand of the system, the

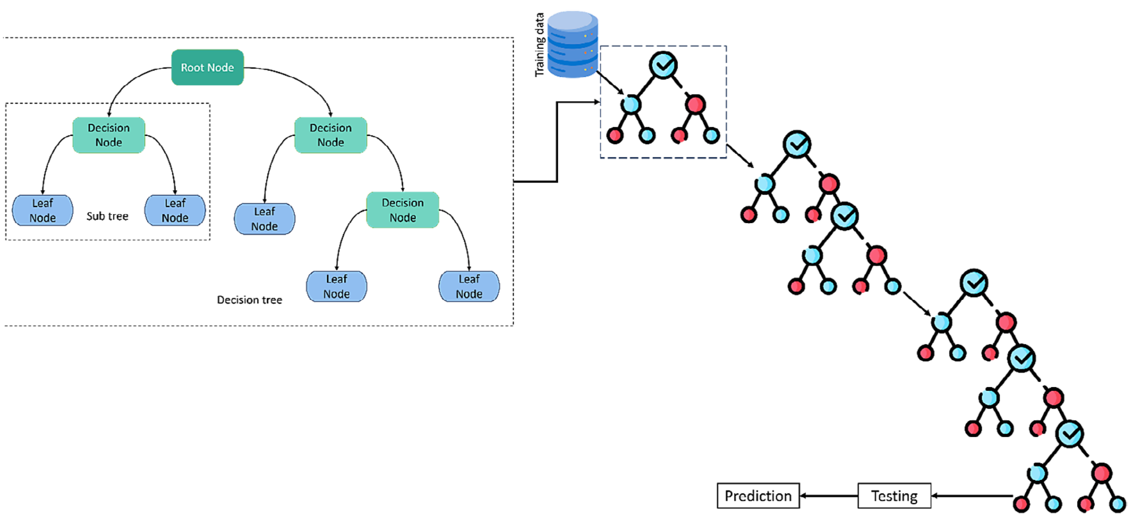

Figure 3. Architecture of gradient boost regression model

best model has to be selected through an evaluation process for system optimization. The evaluation is based on measuring three parameters called R^2 scores, MAE and root mean square error (RMSE), where the R^2 score measures the accuracy of capturing the pattern of target values by the regression model. To make better prediction, R^2 score must be close to 1. The formula is used to calculate the score is expressed in Equation 5.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{5}$$

where: y_i, $\hat{y}_i$ and $\bar{y}\bar{y}$ are the actual values, predicted values and mean of actual values.

The average of the absolute difference between predicted and actual values is known as MAE. To make a perfect prediction, the value of MAE should be close to zero. To calculate the MAE value, the following formula is used, where n is the number of samples.

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i| \tag{6}$$

A slightly different from RAE, the RMSE is defined as the square root of mean of squared error between the predicted and actual values. The formula is used to calculate this parameter is given below.

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \tag{7}$$

RESULTS AND DISCUSSION

In this section, the performance of models, the relation between the target and the features variable as well as the solar plant and the battery size estimation according to the predicted load profile are presented with a comprehensive analysis.

Data visualization

Before testing and training, the dataset should be analyzed to identify the patterns and relationships within the data. It is useful to select the machine learning model and extract the independent features used in prediction. Figure 4 presents the scatter plots of all weather parameters used in model training, excluding the target variables. These visualizations highlight the seasonal and daily variations in irradiance, wind speed, temperature, and humidity, which are critical for identifying patterns relevant to forecasting accuracy. The figure shows time vs DNI, DHI, GHI, wind speed, temperature and humidity. From observing the plots, DNI varies highly with a dense vertical band, indicating spikes in irradiance during sunny days. GHI demonstrates seasonal waves throughout the year and a seasonal and daily pattern is found in the DHI plot. Wind speed varies highly with numerous spikes. Temperature shows an annual sinusoidal variation, indicating warmer days in the middle and cooler days at the beginning and end of the year. Finally plot of humidity expresses almost an inverse relation with the temperature data.

To measure the variation between two variables, covariance is applied to them. In the case of multiple variables, a covariance matrix is suitable for measuring their variation together. As shown in Figure 5, the covariance matrix quantifies the relationships between all variables in the dataset. Strong positive correlations are observed between DNI and PV production, and between GHI and PV production, confirming their importance as key predictive features. Each cell of the matrix expresses the covariance of a pair of variables, which means how they vary together in off-diagonal cells. The diagonal cells show the variance of each variable. The positive and negative values of covariance indicate proportionality and inverse proportionality of two variables. According to the matrix, it can be summarized that a strong positive relation exists between the DNI and PV production pair, with a covariance value of 1329566. The same relation can be explained within GHI by the 1186927 covariance value. 21040 is the covariance value of the temperature and PV production pair. A relation is observed between electric demand and season as well as week, besides the weather data.

Performance assessment

Among the four models of machine learning algorithms, the optimal model is selected based on three evaluation parameters, such as the R^2 score, MAE and RMSE value. Table 1 compares the performance of all four regression models for PV generation and electricity demand forecasting, using R^2, MAE, and RMSE as evaluation metrics. The Random Forest model achieves the highest R^2 and lowest error values for both generation and demand prediction, indicating its superiority for this application. Figure 6 shows scatter regression plots comparing predicted and actual values for all four models. The random forest model has points

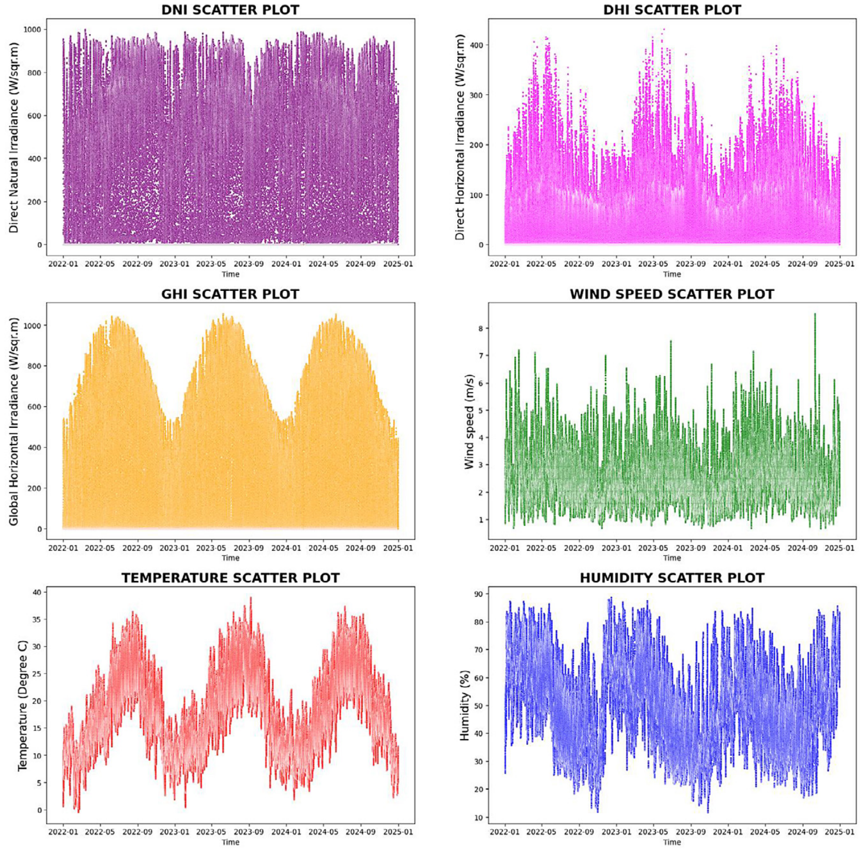

Figure 4. Scatter plot of all weather data used in the training and testing process

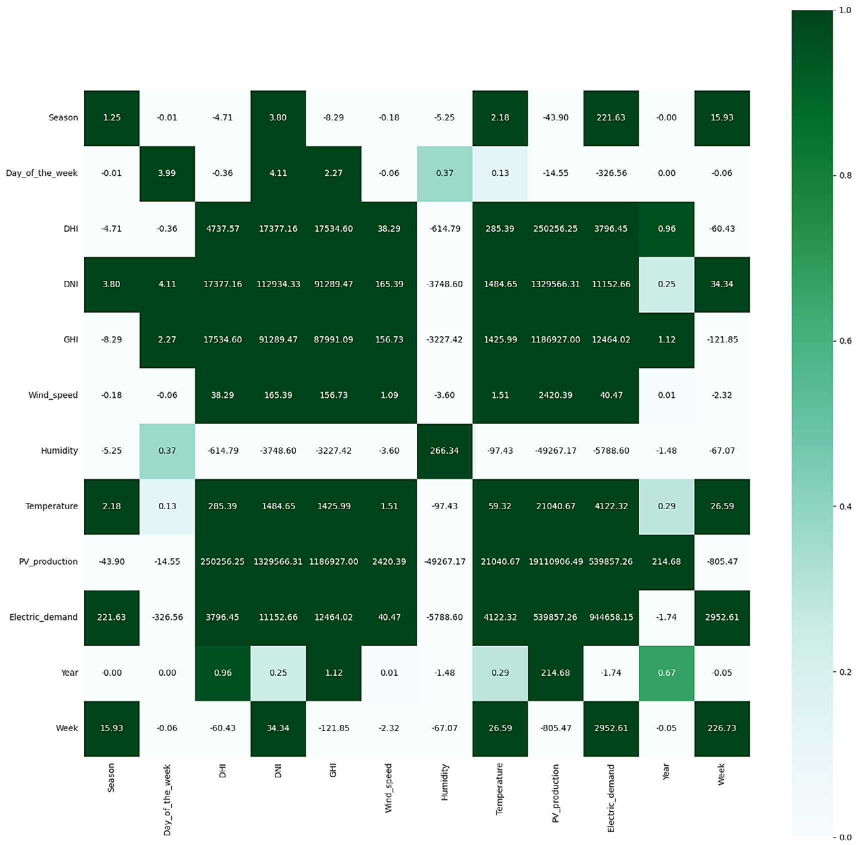

Figure 5. Covariance matrix of the whole dataset used to train and test models

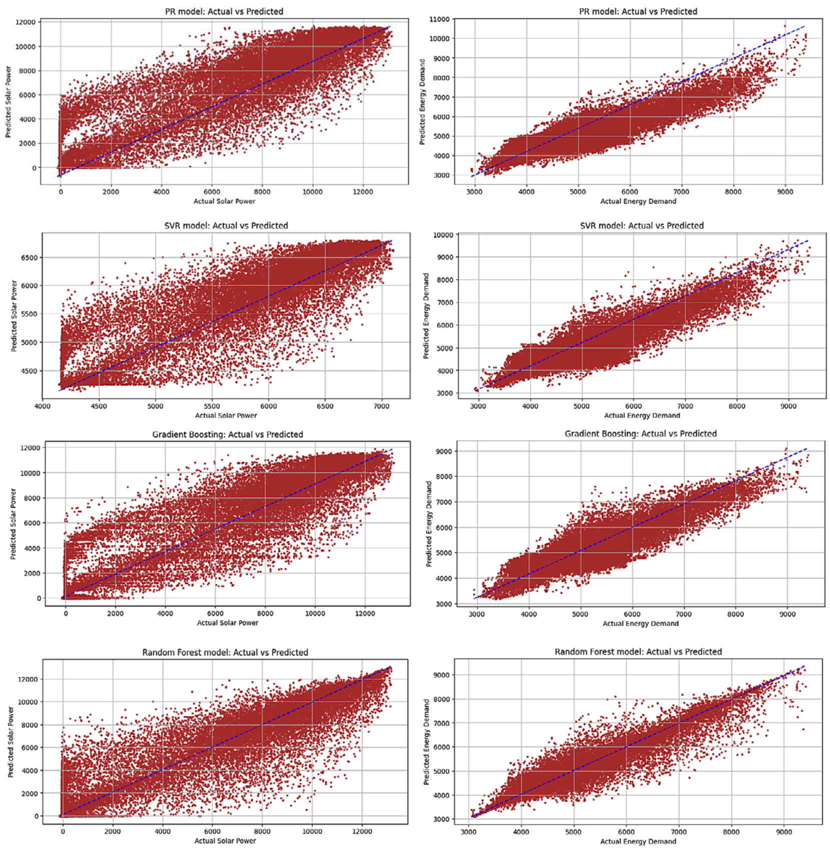

Figure 6. The scatter regression plots of all models for electricity generation and prediction

Table 1. R^2 score, MAE and RMSE value of all regression models for energy production and consumption

Model	PV production			Electric demand		
	R^2 Score	MAE	RMSE	R^2 Score	MAE	RMSE
Polynomial Regression	0.89	793.23	1383.47	0.774	368.5	463.95
Random Forest	0.92	644.8	1262.14	0.90	175.1	306.96
Gradient Boost	0.90	764.65	1357.2	0.76	378.12	471.68
Support Vector Regression	0.899	200.94	309.15	0.796	329.13	437.18

closest to the diagonal, indicating the highest accuracy, while Polynomial Regression shows clear underfitting. From observing the scatter regression plots, it is clear that the RF model fits better than others, points closely follow the diagonal line with reliable prediction and low error. On the other side, the PR model shows the most underfitted curve and fails to tune nonlinear data. The data in the table also clearly reveals that the RF model performs best among them, with the highest value of R^2 score and the lower value of error.

Forecast with optimal model

As the RF model is selected as the best performer among the considered models, this subsection presents the forecasting of the selected model, as the forecasting data can be used for subsequent operation of the plant optimization. Figure 7 compares the Random Forest model's predictions with the actual measured PV generation and load demand. The close alignment between the two curves demonstrates the model's ability to accurately capture variations over time.

After forecasting the electric production and consumption almost perfectly, a load profile is required in the optimization process of the solar plant. Figure 8 depicts the 24-hour load profile of the PV system as forecasted by the Random Forest model, showing peak generation during midday and peak demand in the evening, which informs the PV–BESS sizing strategy. The figure

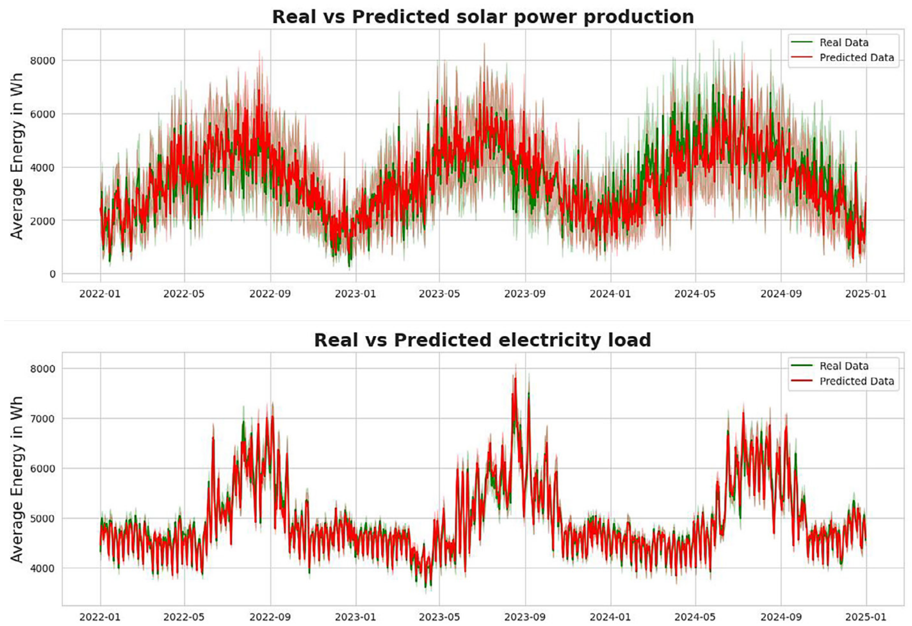

Figure 7. Real tested vs predicted data of the random forest algorithm

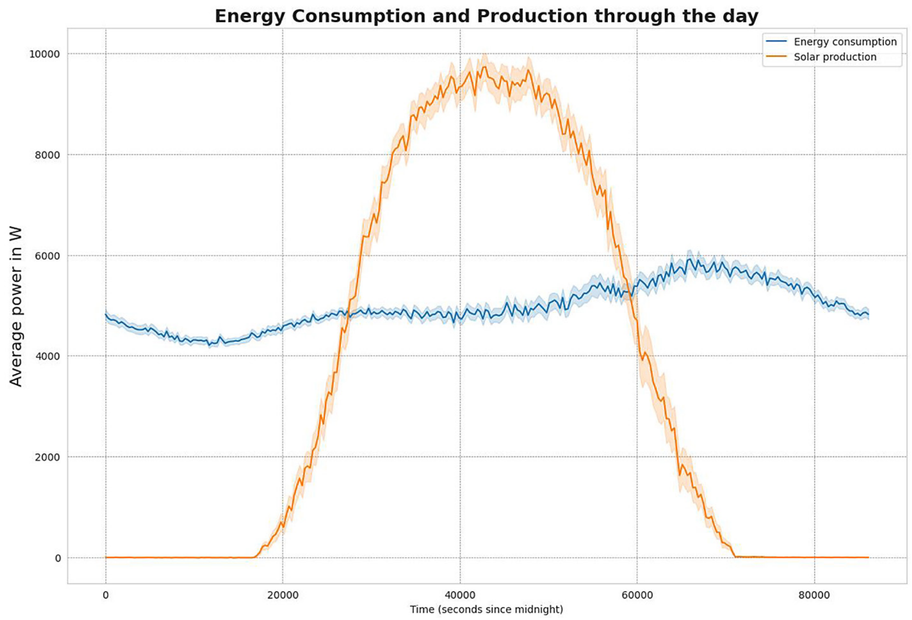

Figure 8. The load profile of the solar PV plant estimated through machine learning model

shows time (in seconds) vs power graph over a 24-hour day. Here, the maximum mean energy consumption and the solar production can be recorded from the figure as about 6 kW and 10 kW.

Plant and BESS sizing

From the load profile, the maximum supply (from PV) and demand are recorded as 10 kW and 6 kW. Therefore, to remove the grid dependency of the system, the verification of the PV plant capacity and the suggested storage capacity is accomplished in this subsection. According to the achieved data, the total average consumption of electric energy is about 6 kW × 24 h = 144 kWh, and the production from the solar PV plant of 11 kW capacity is about 10 kW × 5 h = 50 kWh (considering the sunny hour is 5 h). Therefore, to overcome the grid dependency, new capacity of the solar plant will be $P_{new} = 2.8 \times 11 = 30$ kW almost.

Here, 2.8 is the scaling factor $k = \frac{144}{50} = 2.8$

When the power production of the new PV plant is E_{pv} = 30 kW × 5 h = 150 kWh and the total load during the charging period is E_L = 6 kW × 5h = 30 kWh. Then, to store all PV-generated power, the energy storage capacity should be E_E = 150 kWh – 30 kWh = 120 kWh (approximate).

CONCLUSION

The research achieved its objective by demonstrating that a Random Forest based forecasting approach yields superior prediction accuracy (R^2 = 0.92 for PV generation, R^2 = 0.90 for demand) compared to Polynomial Regression, Gradient Boost, and Support Vector Regression. This improved forecasting accuracy directly informed an optimized PV–BESS design, resizing the PV array from 11 kW to 30 kW and specifying a 120 kWh storage capacity for complete grid independence. The study fills a critical gap by providing a replicable framework that integrates multi-model forecasting with technical optimization using real-world data from Bangladesh. This approach offers significant potential for autonomous residential energy systems in resource-constrained regions and can be extended to hybrid renewable configurations with real-time IoT-based control.

REFERENCES

1. Ahangar, P. A., Lone, S. A., Gupta, N. (2023). Combining data-driven and model-driven approaches for optimal distributed control of standalone microgrid. *Sustainability 15*(16), 12286. https://doi.org/10.3390/SU151612286
2. Ahmad, M. W., Mourshed, M., Rezgui, Y. (2018). Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression. *Energy*, *164*, 465–474. https://doi.org/10.1016/J.ENERGY.2018.08.207
3. Alfadda, A., Adhikari, R., Kuzlu, M., Rahman, S. (2017). Hour-ahead solar PV power forecasting using SVR based approach. *2017 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2017*. https://doi.org/10.1109/ISGT.2017.8086020
4. Babu, R. M., Basher, E. (2024). Performance evaluation and economic analysis of a grid-connected solar power plant: A case study of engreen sarishabari solar plant Ltd. in Bangladesh. *Ecological Engineering & Environmental Technology*, *25, 5*(5), 220–234. https://doi.org/10.12912/27197050/185934
5. *Bangladesh Meteorological Department*. (n.d.). Retrieved August 5, 2025, from https://live6.bmd.gov.bd/
6. Boumais, K., Messaoudi, F. (2025). *BiLSTM-CNN with Bayesian Optimization for Accurate Long-Term Load Forecasting: Cross-Regional Insights from Morocco and Spain*. 1–6. https://doi.org/10.1109/ESTREAM66938.2025.11016829
7. Chowdhury, S. A., Aziz, S., Groh, S., Kirchhoff, H., Leal Filho, W. (2015). Off-grid rural area electrification through solar-diesel hybrid minigrids in Bangladesh: resource-efficient design principles in practice. *Journal of Cleaner Production*, *95*, 194–202. https://doi.org/10.1016/J.JCLEPRO.2015.02.062
8. Ghasemian, S., Faridzad, A., Abbaszadeh, P., Taklif, A., Ghasemi, A., Hafezi, R. (2024). An overview of global energy scenarios by 2040: identifying the driving forces using cross-impact analysis method. *International Journal of Environmental Science and Technology*, *21*(11), 7749–7772. https://doi.org/10.1007/S13762-020-02738-5/TABLES/4
9. Hellqvist, L., Heubaum, H. (2023). Setting the sun on off-grid solar?: policy lessons from the Bangladesh solar home systems (SHS) programme. *Climate Policy*, *23*(1), 88–95. https://doi.org/10.1080/14693062.2022.2056118
10. Karthikeyan, V., Rajasekar, S., Das, V., Karuppanan, P., Singh, A. K. (2017). Grid-connected and off-grid solar photovoltaic system. *Green Energy and Technology*, *0*(9783319501963), 125–157. https://doi.org/10.1007/978-3-319-50197-0_5
11. Khan, I. (2020). Impacts of energy decentralization viewed through the lens of the energy cultures framework: Solar home systems in the developing economies. *Renewable and Sustainable Energy Reviews*, *119*, 109576. https://doi.org/10.1016/J.RSER.2019.109576
12. Liu, D., Sun, K. (2019). Random forest solar power forecast based on classification optimization. *Energy*, *187*, 115940. https://doi.org/10.1016/J.ENERGY.2019.115940
13. Liu, X., Gou, Z. (2025). Hybrid forecasting and optimization framework for residential photovoltaic-battery systems: Integrating data-driven prediction with multi-strategy scenario analysis. *Building Simulation*, 1–23. https://doi.org/10.1007/S12273-025-1319-6/METRICS
14. Nahid, N. R., Hossain, S., Ahammad, J., Ibne Anowar, T. (2024). Cost-Effective Off-Grid Solar Power System, Integrating MPPT and Smart Battery Management System. *2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication, IConSCEPT 2024 - Proceedings*. https://doi.org/10.1109/ICONSCEPT61884.2024.10627870
15. NASA POWER, Homepage. (n.d.). Retrieved

August 5, 2025, from https://power.larc.nasa.gov/

16. Obi, M., Bass, R. (2016). Trends and challenges of grid-connected photovoltaic systems – A review. *Renewable and Sustainable Energy Reviews*, *58*, 1082–1094. https://doi.org/10.1016/J.RSER.2015.12.289
17. Patel, A., Gnana Swathika, O. V. (2024). Off-grid small-scale power forecasting using optimized machine learning algorithms. *IEEE Access*, *12*, 107406–107419. https://doi.org/10.1109/ACCESS.2024.3430385
18. Polo, J., Martín-Chivelet, N., Alonso-Abella, M., Sanz-Saiz, C., Cuenca, J., de la Cruz, M. (2023). Exploring the PV power forecasting at building façades using gradient boosting methods. *Energies 16*(3), 1495. https://doi.org/10.3390/EN16031495
19. Rathore, P. K. S., Chauhan, D. S., Singh, R. P. (2019). Decentralized solar rooftop photovoltaic in India: On the path of sustainable energy security. *Renewable Energy*, *131*, 297–307. https://doi.org/10.1016/J.RENENE.2018.07.049
20. *Renewables - Energy System - IEA*. (n.d.). Retrieved July 30, 2025, from https://www.iea.org/energy-system/renewables
21. Shafiei, K., Seifi, A., Hagh, M. T. (2025). A novel multi-objective optimization approach for resilience enhancement considering integrated energy systems with renewable energy, energy storage, energy sharing, and demand-side management. *Journal of Energy Storage*, *115*, 115966. https://doi.org/10.1016/J.EST.2025.115966
22. Shafiullah, M., Ahmed, S. D., Al-Sulaiman, F. A. (2022). Grid Integration Challenges and Solution Strategies for Solar PV Systems: A Review. *IEEE Access*, *10*, 52233–52257. https://doi.org/10.1109/ACCESS.2022.3174555
23. Shapna, K. J., Li, J., Hossain, M. L. (2025). Solar energy integration in off-grid communities: empowering remote areas in Bangladesh for sustainable development. *Discover Sustainability*, *6*(1), 1–30. https://doi.org/10.1007/S43621-025-01214-9/FIGURES/13
24. Sharma, P., Bojja, H., Yemula, P. (2016). Techno-economic analysis of off-grid rooftop solar PV system. *2016 IEEE 6th International Conference on Power Systems, ICPS 2016*. https://doi.org/10.1109/ICPES.2016.7584208