

Enhanced Solar Power Prediction using Polynomial and Gradient-based Variational Techniques

¹ Preeti V. Kapoor, ² Anish Bhalerao, ³ Uday B. Mujumdar

Shri Ramdeobaba College of Engineering and Management, Nagpur

Email: ¹kapoorpv@rknec.edu ²bhaleraoas@rknec.edu ³mujumdarub@rknec.edu

Article History:

Received: 12-12-2024

Revised: 28-01-2025

Accepted: 09-02-2025

Abstract: Accurate prediction of solar photovoltaic (PV) energy output plays a pivotal role in enhancing energy management, planning, and grid reliability. While conventional regression techniques provide interpretability, they often struggle with modelling the nonlinear dependencies present in real-world PV systems. Meanwhile, purely black-box models may yield high accuracy but lack the transparency needed for practical engineering deployment. This study explores the use of polynomial regression for modelling the AC power output from rooftop PV system, trained on measured environmental and PV system parameters. To further refine the predictions and enhance generalizability, we integrate a Projected Gradient Descent (PGD) approach for regularization and optimization. The combination of feature expansion and iterative optimization offers a robust modelling framework with strong predictive capability. Experimental results demonstrate that the polynomial regression model achieves excellent agreement with actual measurements, capturing fine-grained variations in power output with high accuracy. This highlights the proposed approach as a computationally efficient and interpretable method suitable for real-time solar forecasting applications.

Keywords: Solar PV energy, Machine Learning, Polynomial Regression, Variational Inequality

1. Introduction

Solar photovoltaic (PV) energy has emerged as a cornerstone of global renewable energy strategies, offering a sustainable alternative to conventional fossil fuel-based generation [1]. With the total installed global PV capacity exceeding 1 terawatt (TW) by 2022, and major contributions from countries like China, India, and the United States [2], accurate forecasting of PV power output has become a paramount priority [3][4]. Forecasting helps in efficient grid integration, optimising energy dispatch, and enhancing overall system reliability [5]. Traditional solar power forecasting models often rely on statistical techniques or black-box machine learning algorithms [6][7][8][9][10][11], which can lag in explainability or under changing environmental conditions [12]. Motivated by the need for reliable yet mathematically well-grounded prediction frameworks, this study models the PV power forecasting task as a variational inequality (VI) problem. We employ polynomial feature expansion to model the nonlinear relationships inherent in solar power generation [13], followed by two distinct solution strategies: a closed-form Ridge regularized regression method and an iterative

Projected Gradient Descent (PGD) solver [14]. The VI framework provides a rigorous mathematical foundation for regression under regularization [15], enhancing model stability and robustness. Unlike conventional polynomial regression approaches [16][17], the VI-based formulation explicitly controls model complexity, improving generalization performance. Through the use of real-world data collected from a rooftop PV installation, this study aims to display the advantages of novel variational inequality frameworks for renewable energy forecasting applications. Figure 1 displays the stages the data passes through before it was analysed and pre-processed.

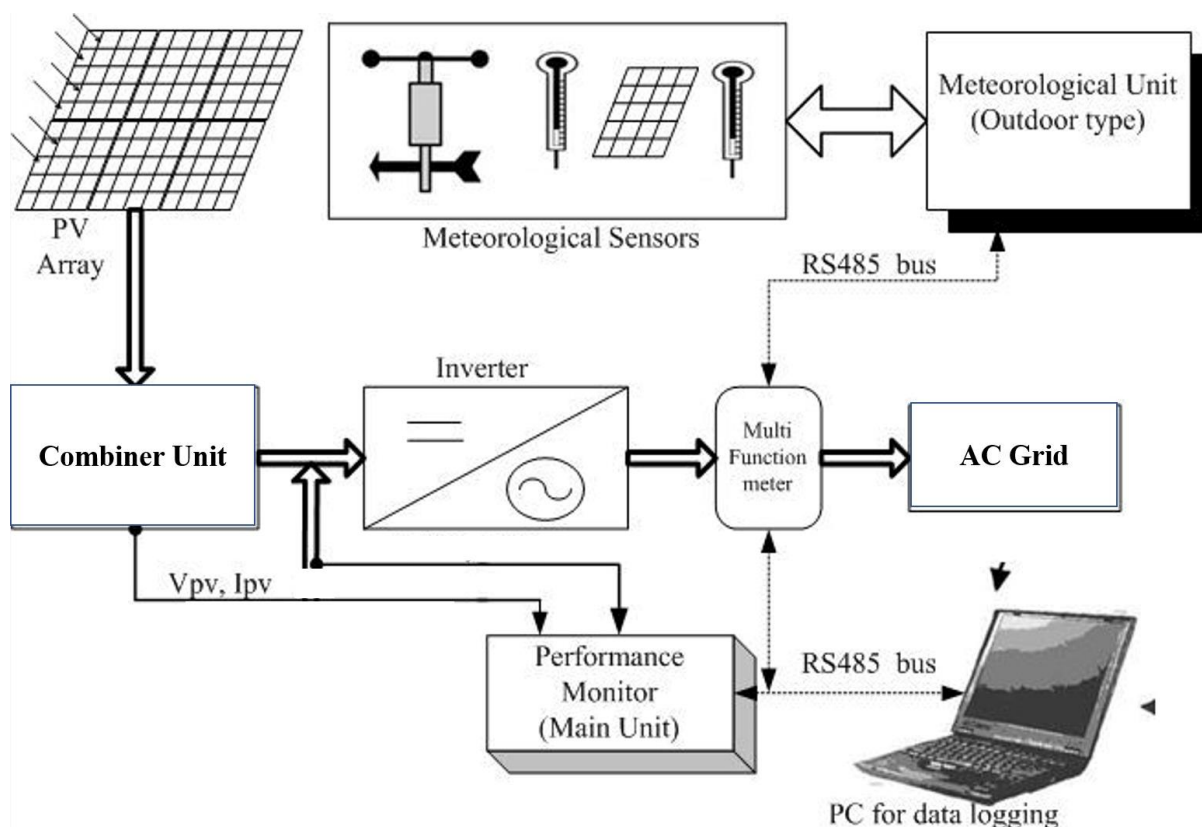


Fig. 1: Data Generation and Normalisation

2. Data Collection and Methodology

The experimental setup used in this study is located at the Energy Research Centre, Ramdeobaba University Campus, Nagpur, India (21.1458°N, 79.0882° E). The system consists of:

- 8 units of WAAREE solar panels, each with a rated capacity of 335 W, giving a total power output of 2.68 kW.
- A data logging system using microcontrollers is implemented to collect real-time operational data.
- Sensors are installed to measure environmental and electrical parameters.
- Data of every single day is recorded at 1-minute intervals and stored in a .csv format for further analysis. These .csv files are then combined using Python to conduct mathematical

analysis. This automation ensured consistency and eliminated manual errors in data compilation.

The dataset used in this study was collected over a 60-day period from January to March 2025 at a laboratory in Nagpur, India. The recorded parameters include ambient temperature, irradiance, relative humidity, panel temperature, Panel voltage, array current and the inverter output AC Power readings. The measurements are logged using a custom-designed microcontroller-based data acquisition system, storing timestamped values into daily .csv files. The raw collected data underwent pre-processing, including outlier removal, missing data handling, and feature engineering to derive additional variables such as mean voltage and panel temperature averages. The integrated dataset formed the foundation for the subsequent regression model training processes. This work suggests a two-stage approach to predict the AC power generation of the solar PV system. The underlying principle is to cast the regression problem into a Variational Inequality (VI) framework, which allows for increased versatility and generalization. Two different solution approaches were experimented with: a closed-form Ridge-based polynomial regression and an iterative Projected Gradient Descent (PGD) solver.

1. Variational Inequality Formulation

The initial step in the implementation of our model involves the construction of an expanded feature space using polynomial transformations of the VI-based input parameters. By applying a third-degree polynomial expansion to key features such as irradiance, PV current, mean voltage, ambient, and panel temperature, the model is able to capture complex nonlinear interactions inherent in photovoltaic system behaviour. This transformation provides the foundation for the subsequent regression analysis and enables more accurate prediction of AC power output under varying environmental conditions. To estimate the AC power output from the input environmental and electrical features, we construct a polynomial regression model within a variational inequality (VI) framework. The regression problem is regularized using ridge (λ) penalty to handle multicollinearity and overfitting due to polynomial feature expansion. Although both the conventional polynomial regression and the proposed VI-based solver utilize the same expanded feature set $\Phi(X)$ derived through polynomial transformations, their learning strategies diverge significantly in both formulation and execution.

Closed-form Polynomial Regression

The conventional polynomial regression model solves for weights w using the ordinary least squares (OLS) formulation. Given training data $X \in \mathbb{R}^{n \times d}$ and targets $y \in \mathbb{R}^n$, the feature-expanded matrix is $\Phi(X) \in \mathbb{R}^{n \times m}$. The OLS minimization objective is:

$$\min_w \|\Phi(X)w - y\|_2^2$$

The optimal weight vector w^* is obtained analytically:

$$w^* = (\Phi(X)^T \Phi(X))^{-1} \Phi(X)^T y$$

This solution assumes full-rank feature matrices and no explicit constraints, and is solved in a single step.

VI-based Solver with Projected Gradient Descent

In contrast, the VI-based solver treats learning as a **variational inequality problem**. The objective is to find a vector $w^* \in K \subset \mathbb{R}^m$ such that:

$$\text{Find } w^* \in W \text{ such that } \langle F(w^*), w - w^* \rangle \geq 0, \quad \forall w \in W \quad (1)$$

Where:

- w is the model coefficient vector,
- $F(w)$ is the operator derived from the objective function.

This condition asserts that the operator $F(w)$ defines a monotone relationship whose fixed point corresponds to the minimizer of the regression objective. In the unconstrained case, the solution to this VI is equivalent to minimizing a convex objective.

The solution is obtained via **Projected Gradient Descent (PGD)**:

The objective function is given by:

$$\min_w \|X_w - y\|^2 + \lambda \|w\|^2$$

Where:

- X represents the polynomial-expanded feature matrix,
- y is the vector of observed AC power output,
- λ is the ridge regularization parameter controlling the penalty on the weights.

The corresponding closed-form solution to this regularized least squares problem is:

$$w^* = (X^T X + \lambda I)^{-1} X^T y$$

Here, I is the identity matrix. This formulation ensures a stable and unique solution even in the presence of multicollinearity among expanded features. The regularization term $\lambda \|w^2\|$ acts as a constraint, helping prevent overfitting when the polynomial degree is high.

To formally define the proposed regression approach, we begin by framing the objective as a Variational Inequality (VI) problem. This provides a general framework for optimization in the presence of regularization and potential constraints.

i. Regularised Loss Formulation

The regularised loss function $L(w)$ is defined as:

$$L(w) = \|Xw - y\|_2^2 + \lambda \|w\|_2^2 \quad (2)$$

Where:

- X is the feature matrix,
- y is the target vector,
- $\lambda > 0$ is the regularization parameter

Here, the first term measures the squared error between predicted and true target values, and the second term is the ridge penalty, enforcing smoothness by discouraging large weights. $\lambda > 0$ is the regularization hyperparameter.

Taking the gradient of Eq. (2) with respect to w gives the VI operator.

ii. Derivation of Operator $F(w)$

The corresponding VI operator $F(w)$ is obtained by taking the gradient of the loss:

$$F(w) = 2X^T(Xw - y) + 2\lambda w \quad (3)$$

This operator is strongly monotone under mild conditions, ensuring the existence and uniqueness of the solution.

iii. Closed-Form Solution (Ridge Regression)

Setting the gradient of the loss to zero yields:

$$2X^T(Xw - y) + 2\lambda w = 0 \quad (4)$$

Solving this yields the classical ridge regression estimator:

$$\begin{aligned} (X^T X + \lambda I)w &= X^T y \\ w^* &= (X^T X + \lambda I)^{-1} X^T y \end{aligned} \quad (5)$$

Alternatively, we may avoid explicit inversion and adopt an iterative solution strategy via Projected Gradient Descent (PGD), defined by the following update rule.

2. Projected Gradient Descent Update Rule

In the PGD approach, at each iteration k , the update rule is:

$$w^{(k+1)} = w^{(k)} - \eta F(w^{(k)}) \quad (6)$$

Where:

- $\eta > 0$ is the step size (learning rate),
- $F(w^{(k)})$ is the VI operator at the current iterate.

This iteration continues until convergence. The first solution approach involved solving a regularized least squares regression problem framed as a Variational Inequality (VI). The corresponding closed-form solution is obtained by:

$$w^x = (X^T X + \lambda I)^{-1} X^T y$$

where I is the identity matrix. This approach ensures a unique and stable solution, even in the presence of multicollinearity among expanded features. In the VI framework, the operator $F(w)$ is defined, and optimality conditions are satisfied.

Convergence Condition

The iterative updates terminate when:

$$\|w^{(k+1)} - w^{(k)}\|_2 < \epsilon \quad (7)$$

Stopping at the best iteration for maximum efficiency.

3. Data Preprocessing and Model Training

Having established the convergence condition for our VI-based solver, we now describe the data processing and training workflow used for empirical evaluation. The dataset was put through thorough cleaning and transformation processes, including:

- Removal of missing values,
- Outlier elimination using the Interquartile Range (IQR) method,
- Standardization of features to zero mean and unit variance.

The dataset was subsequently split into training (80%) and testing (20%) subsets. Both closed-form and PGD models were trained on the expanded and scaled training data. Predictions were then generated and evaluated on the testing data. Figure 2 demonstrates the phases of data pre-processing and evaluation tests models are put through after being deployed on dataset.

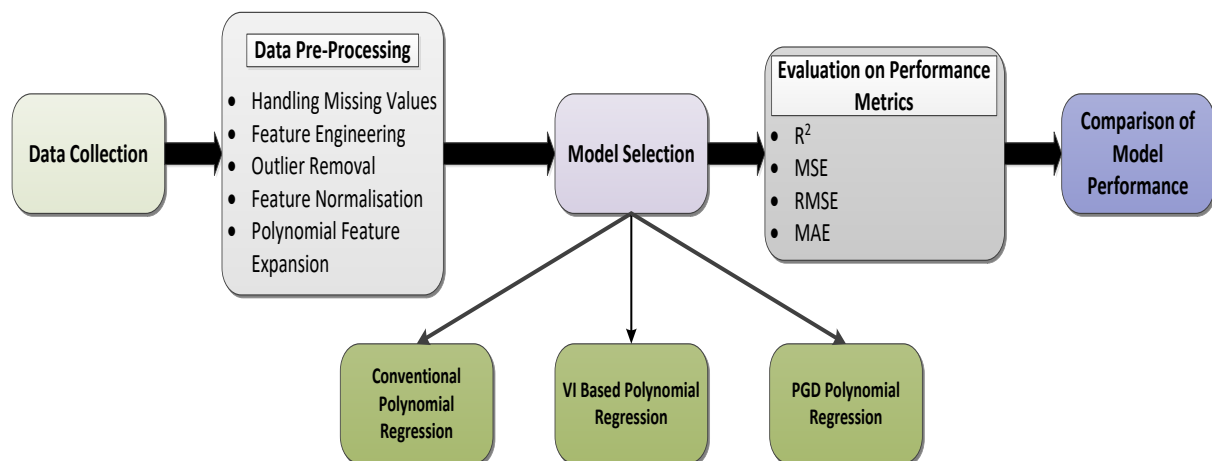


Fig. 2 Pre-processing and evaluation procedure

3. Results

The proposed Variational Inequality (VI)-based regression models were evaluated, both the closed-form Ridge regularized solver and the iterative Projected Gradient Descent (PGD) solver were assessed for their predictive performance.

The performance metrics considered were:

- Mean Squared Error (MSE),
- Root Mean Squared Error (RMSE),
- Mean Absolute Error (MAE),
- Coefficient of Determination (R^2),
- Adjusted R^2 .

Table 1 represents the comparison of our VI-based polynomial regression model and a standard ridge based polynomial regression model.

Table 1: Model Performance Comparison

Model	Performance Metrics				
	MSE	RMSE	MAE	R ²	Adjusted R ²
Conventional Polynomial Regression	26937.00	164.12	116.13W	0.9089	0.9091
VI Based Polynomial Regression	204.73	14.31	7.29 W	0.999268	0.999260

The high R² and low error values confirm that the closed-form VI model is impressively effective in capturing the nonlinear dependencies between environmental variables and AC power generation. In contrast, the PGD solver encountered convergence challenges, likely due to the complex feature space induced by the polynomial expansion. Gradient instabilities and slow learning rates affected the convergence quality, resulting in suboptimal predictive performance compared to the closed-form solution. Nevertheless, the PGD approach remains a valuable method for larger-scale or constraint-ridden problems where closed-form inversion becomes computationally expensive or infeasible.

Figure 3 illustrates the comparative performance of three regression models—Ridge Regression, Conventional Polynomial Regression, and the PGD-based Variational Inequality (VI) Solver—against the actual AC power output from the solar photovoltaic system recorded on 22 April 2025. The X-axis represents the timestamp (in minutes) while the y-axis shows the AC power output in watts. The black curve represents the actual measured AC power values. The red curve corresponds to predictions from the Ridge Regression model, the purple curve shows the predictions from the conventional polynomial regression model (degree 3), and the green curve reflects the predictions obtained via the PGD-based VI approach. All three models follow the true power trend closely, demonstrating high predictive capability. However, subtle variations are visible during power peaks and transitions. The Ridge and PGD models appear to generalize better around higher output values, whereas the conventional polynomial model occasionally exhibits overshooting or underfitting tendencies in those regions. This visual comparison confirms that while all models show promising results, Ridge and PGD approaches offer slightly more robust and stable predictions, especially when considering noise, outliers, and system nonlinearities.

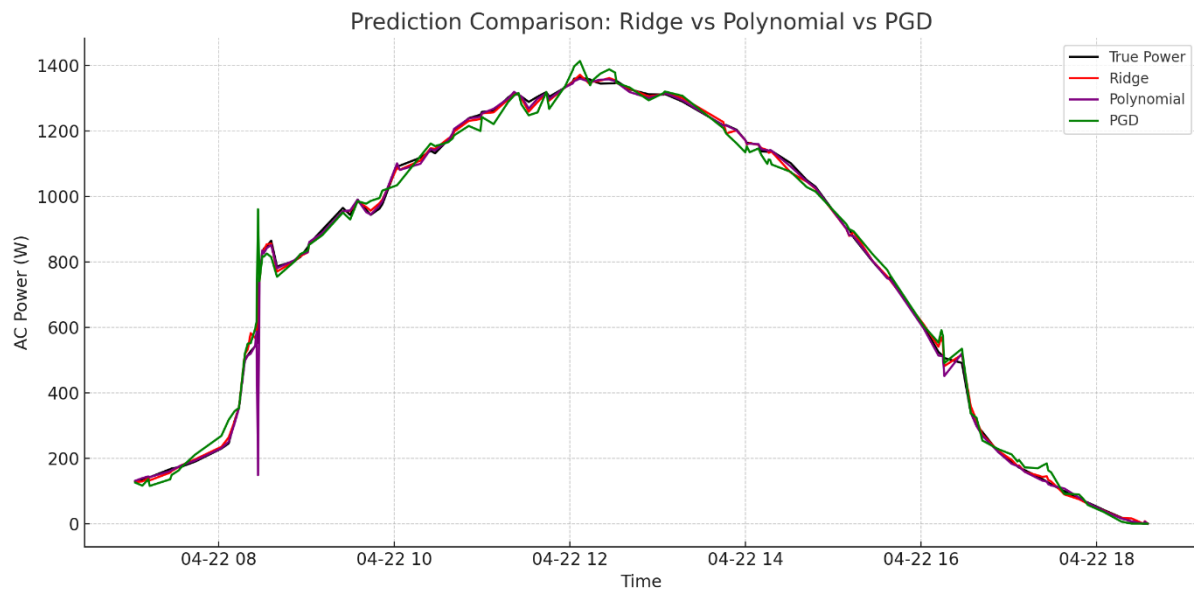


Figure 3: Comparison of model performance for a single day power output

A scatter plot comparing the predicted and actual AC power values was generated for both methods as seen in Figure 4. The closed-form VI model closely follows the ideal 45° reference line, indicating near-perfect predictions across the full range of operational power levels up to 2700 W. In contrast, the PGD predictions showed larger variance and deviations from the ideal line, particularly at higher power outputs.

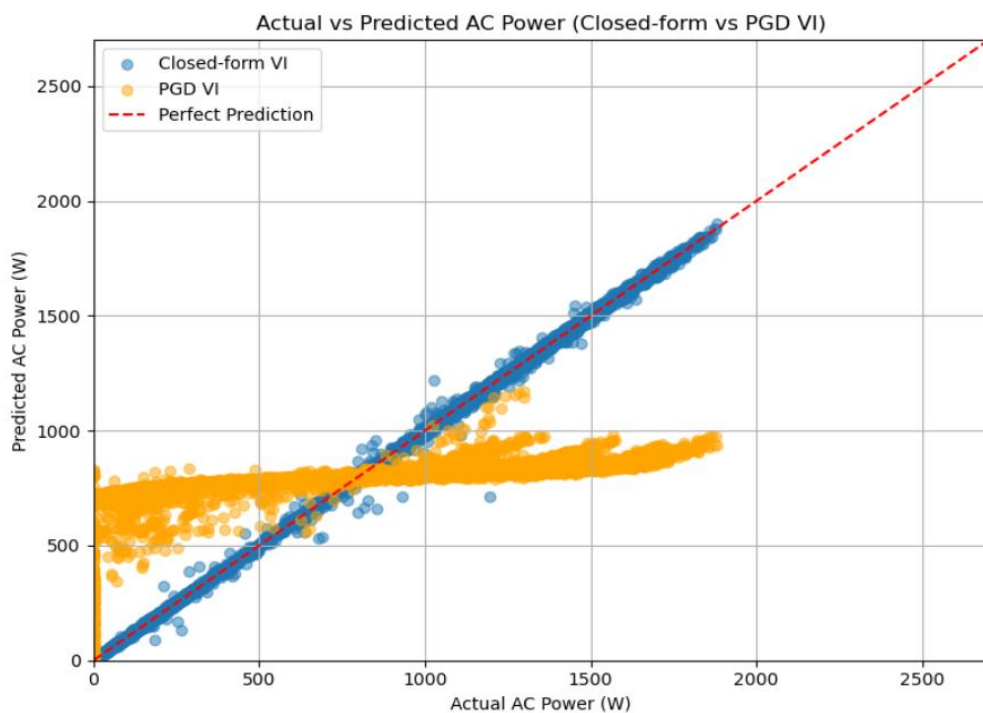


Fig. 4: Scatter plot of both proposed models

4. Conclusion

For this dataset, the comparison results show that the closed-form Ridge technique is preferable. The variational inequality-based formulation with regularisation directly leads to its ability to strike a balance between model simplicity and fitting accuracy. The limitations with the PGD solver demonstrate the real-world challenges of employing iterative gradient-based techniques to solve high-degree polynomial regression problems, especially when feature scaling and step-size tuning are not optimal. The study does, however, confirm that VI-based framing offers a strong and broadly applicable avenue for enhancing machine learning regression models in renewable energy systems.

5. Acknowledgements

The authors express their sincere gratitude to Shri Ramdeobaba College of Engineering and Management (RCOEM), Nagpur, India, for providing the infrastructure and facilities necessary for the successful execution of this research.

Special thanks are extended to the Energy Research Centre of RCOEM for granting access to the rooftop solar photovoltaic (PV) installation and associated data logging systems. The support and technical guidance from the faculty and research staff have been instrumental in the project's data collection, instrumentation calibration, and experimental validation phases.

References

- [1] Ugli, T. J. T. (2019). The importance of alternative solar energy sources and the advantages and disadvantages of using solar panels in this process.
- [2] Franke, K., Garcia, J. F., Kleinschmitt, C., & Sensfuß, F. (2024). Assessing worldwide future potentials of renewable electricity generation: installable capacity, full load hours and costs. *Renewable Energy*, 226, 120376.
- [3] Ajayi, A. O., Agupugo, C. P., Nwanevu, C., & Chimziebere, C. (2024). Review of penetration and impact of utility solar installation in developing countries: policy and challenges. *International Journal of Frontiers in Engineering and Technology Research*, 7(2), 11-24.
- [4] Iheanetu, K. J. (2022). Solar photovoltaic power forecasting: A review. *Sustainability*, 14(24), 17005.
- [5] Singla, P., Duhan, M., & Saroha, S. (2021). A comprehensive review and analysis of solar forecasting techniques. *Frontiers in Energy*, 1-37.
- [6] AlKandari, M., & Ahmad, I. (2024). Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. *Applied Computing and Informatics*, 20(3/4), 231-250.
- [7] Ye, H., Yang, B., Han, Y., & Chen, N. (2022). State-of-the-art solar energy forecasting approaches: Critical potentials and challenges. *Frontiers in Energy Research*, 10, 875790.
- [8] Zazoum, B. (2022). Solar photovoltaic power prediction using different machine learning methods. *Energy Reports*, 8, 19-25.
- [9] Jebli, I., Belouadha, F. Z., Kabbaj, M. I., & Tilioua, A. (2021). Prediction of solar energy guided by pearson correlation using machine learning. *Energy*, 224, 120109.
- [10] Munawar, U., & Wang, Z. (2020). A framework of using machine learning approaches for short-term solar power forecasting. *Journal of Electrical Engineering & Technology*, 15(2), 561-569.
- [11] Balal, A. T., Jafarabadi, Y. P. T., Demir, A. T., Igene, M. T., Giesselmann, M. T., & Bayne, S. T. (2023). Forecasting solar power generation utilizing machine learning models in Lubbock.
- [12] Anuradha, K., Erlapally, D., Karuna, G., Srilakshmi, V., & Adilakshmi, K. (2021). Analysis of solar power generation forecasting using machine learning techniques. In *E3S web of conferences* (Vol. 309, p. 01163). EDP Sciences.

- [13] Ahsan SM, Khan HA, Hussain A, Tariq S, Zaffar NA. Harmonic Analysis of Grid-Connected Solar PV Systems with Nonlinear Household Loads in Low-Voltage Distribution Networks. *Sustainability*. 2021; 13(7):3709. <https://doi.org/10.3390/su13073709>
- [14] Vella, C., Gosselet, P., & Prudhomme, S. (2024). An efficient PGD solver for structural dynamics applications. *Advanced Modeling and Simulation in Engineering Sciences*, 11(1), 15.
- [15] Noor, M. A., Noor, K. I., & Rassias, M. T. (2020). New trends in general variational inequalities. *Acta Applicandae Mathematicae*, 170, 981-1064.
- [16] Jafari, F., Moradi, K., Shafiee, Q., & Moghaddam, F. (2024, May). Data-driven performance modeling of solar panels using polynomial regression. In *2024 International Conference on Control, Automation and Diagnosis (ICCAD)* (pp. 1-8). IEEE.
- [17] Arce, J. M. M., & Macabebe, E. Q. B. (2019, November). Real-time power consumption monitoring and forecasting using regression techniques and machine learning algorithms. In *2019 IEEE international conference on internet of things and intelligence system (IoTals)* (pp. 135-140). IEEE.