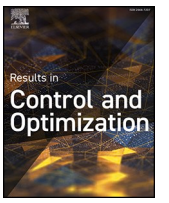
Results in Control and Optimization 20 (2025) 100586

Contents lists available at ScienceDirect

Results in Control and Optimization

journal homepage: www.elsevier.com/locate/rico



Optimizing photovoltaic performance: Data-driven maximum power point prediction via advanced regression models

Maissa Farhat [a,*], Azzeddine Dekhane [b,e], Abdelhak Djellad [b,e], Maen Takruri [c], Aws Al-Qaisi [c], Oscar Barambones [d]

[a] SOE, American University of Ras Al Khaimah, Building: 75, Seih Al Araibi, Ras Al Khaimah, United Arab Emirates
[b] National Higher School of Technology and Engineering, EEA Department, 23005 Annaba, Algeria
[c] College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait
[d] Advanced Control Group. Universidad del País Vasco (UPV/EHU), EI-VG, Nieves Cano 12, 01006 Vitoria, Spain
[e] Environmental Research Center (CRE), Annaba, Algeria

ARTICLE INFO

Keywords:
Photovoltaic systems
Maximum power point (PMPP)
Data driven approaches
Regression techniques
Tree-based models
Nonlinear relationships
Optimizing energy
Solar energy harvesting, Predictive accuracy
Machine learning

ABSTRACT

The accurate prediction of the Maximum Power Point (PMPP) in photovoltaic (PV) systems is critical for optimizing energy yield and enhancing solar energy harvesting efficiency. This study explores the application of data-driven methods to improve PMPP prediction, utilizing advanced regression techniques such as Ridge Regression, Lasso Regression, Decision Tree Regression, and Random Forest Regression. By analyzing a dataset of irradiance, temperature, and PMPP measurements, the research compares the performance of these models in capturing complex nonlinear relationships between key variables. Results indicate that tree-based models, particularly Random Forest Regression, outperform linear models, demonstrating superior predictive accuracy and robustness. Feature importance analysis further highlights the dominant influence of irradiance (GPOA) on PMPP, emphasizing the value of precise irradiance data. These findings underscore the potential of machine learning techniques in optimizing PV system performance. Future research should focus on integrating additional features, such as atmospheric conditions and panel characteristics, and exploring deep learning methods to enhance prediction accuracy further.

1. Introduction

The importance of renewable energy, particularly photovoltaic (PV) systems, has gained significant attention due to their potential to mitigate climate change and reduce dependence on fossil fuels. Photovoltaic systems convert solar energy directly into electricity, offering a sustainable solution to meet the growing global energy demand, Fig. 1 an example of a photovoltaic system. The integration of PV technology into the energy mix not only contributes to the reduction of greenhouse gas emissions but also promotes energy security and economic development. Studies indicate that the deployment of PV systems can lead to substantial reductions in carbon emissions, with estimates suggesting that a large-scale transition to solar energy could decrease global CO2 emissions by up to 4.9 gigatons annually [1]. Furthermore, advancements in PV technology, such as increased efficiency and reduced costs, have made solar

* Correspondence author.
E-mail addresses: Maissa.farhat@gmail.com (M. Farhat), dekhane@ensti-annaba.dz (A. Dekhane), a.djellad@ensti-annaba.dz (A. Djellad), Maen.takruri@aum.edu.kw (M. Takruri), Aws.Al-Qaisi@aum.edu.kw (A. Al-Qaisi), oscar.barambones@ehu.eus (O. Barambones).

https://doi.org/10.1016/j.rico.2025.100586
Received 21 March 2025; Received in revised form 14 May 2025; Accepted 5 June 2025
Available online 19 June 2025

energy more accessible, enabling widespread adoption across various sectors, including residential, commercial, and industrial applications [2]. The transition to renewable energy sources, particularly through the utilization of photovoltaic systems, is crucial for achieving sustainability goals and ensuring a resilient energy future.

One of the primary challenges in renewable energy systems, particularly photovoltaic (PV) systems, is optimizing energy yield under varying environmental conditions. Factors such as fluctuations in solar irradiance, temperature changes, and partial shading significantly affect the efficiency of PV modules. These challenges necessitate the development of advanced control and optimization techniques to ensure reliable and maximum energy extraction [3]. Studies highlight that addressing these challenges requires robust algorithms capable of adapting to dynamic operating conditions and minimizing energy losses

The maximum power point (PMPP) prediction is vital for enhancing the performance of PV systems. It refers to identifying the point on the solar panel's I-V curve where the product of current and voltage is maximized, thus achieving peak power output. Effective PMPP prediction allows the PV system to continually adjust the operating point to the optimal PMPP despite fluctuating environmental conditions such as changes in solar irradiance and temperature [4]. This is achieved through various Maximum Power Point Tracking (MPPT) techniques, which adjust the load presented to the solar panels to maintain operation at the PMPP, maximizing the energy harvested from the solar array [5]. Advanced PMPP prediction techniques that incorporate machine learning or artificial intelligence can predict changes in environmental conditions and dynamically adjust the system settings preemptively, enhancing the efficiency and reliability of PV systems. In fact, addressing the challenges in optimizing energy yield and the role of PMPP prediction in improving system performance is crucial for the advancement and integration of solar power into the global energy mix. Effective MPPT strategies that leverage accurate PMPP prediction are essential for maximizing the return on investment in solar energy projects.

The performance of photovoltaic (PV) systems is inherently linked to environmental factors, primarily solar irradiance and temperature. PV panels exhibit a negative temperature coefficient, typically around $-0.4\,\%/\,^{\circ}C$ [4]. This means that as the temperature of the PV module increases, its voltage output and consequently the Maximum Power Point (PMPP) decrease. Physically, this reduction in PMPP with rising temperature is attributed to increased electron-phonon scattering within the semiconductor material, leading to higher electrical resistance and a lower open-circuit voltage [5]. Solar irradiance, on the other hand, directly influences the current generated by the PV module. Higher irradiance levels lead to a greater number of photons striking the solar cells, resulting in a higher short-circuit current and a proportionally higher PMPP. However, the impact of irradiance is also coupled with the module's temperature. Under high irradiance conditions, the module temperature tends to rise, which then counteracts some of the gains in PMPP due to the negative temperature coefficient. Therefore, the interplay between irradiance and temperature dictates the overall performance and the achievable PMPP of a PV system. Our machine learning models aim to capture these complex, often nonlinear relationships to provide accurate PMPP predictions under varying operational conditions

The primary objective of this work is to enhance the prediction accuracy of the maximum power point (PMPP) in photovoltaic (PV) systems by leveraging data-driven approaches. This study analyzes the relationships between environmental variables, such as irradiance and temperature, and PMPP through the application of advanced regression models. By comparing the performance of various techniques, including linear and tree-based models, and assessing feature importance, the research seeks to identify optimal methods for capturing complex interactions within the data. Ultimately, the findings are intended to contribute to improving the efficiency of

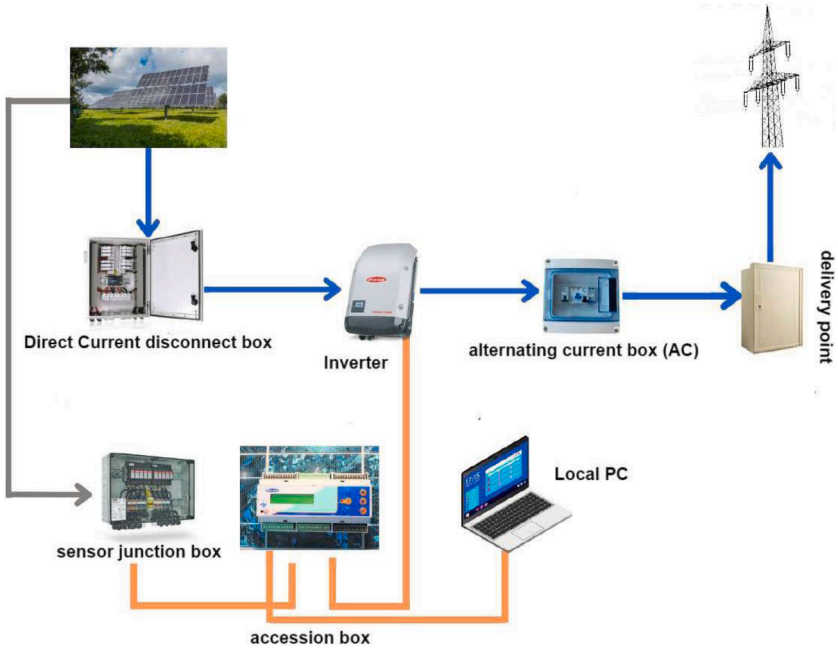


Fig. 1. A photovoltaic system.

solar energy harvesting by optimizing PV system performance.

The outcomes of this research have far-reaching implications for the renewable energy sector. Accurate PMPP predictions not only enhance the efficiency and reliability of PV systems but also contribute to more effective grid integration and energy management strategies. Moreover, understanding the interplay of environmental factors provides valuable guidance for designing and optimizing PV installations across diverse geographical regions and conditions.

This study evaluates the effectiveness of various regression models, including Ridge Regression, Lasso Regression, Decision Tree Regression, and Random Forest Regression, in predicting PMPP. Each method offers unique advantages: Ridge and Lasso regression are well-suited for handling multicollinearity, while Decision Tree and Random Forest excel in capturing nonlinear relationships. A critical part of the study involves analyzing the importance of key features, such as irradiance and temperature, in influencing PMPP, thereby gaining insights into the underlying physics of PV system performance.

This paper is structured as follows: the first section reviews the challenges and context of predicting the maximum power point (PMPP) in photovoltaic systems, highlighting the significance of adopting advanced regression models. The methodology section outlines the data sources, preprocessing steps, and regression algorithms utilized in the analysis. The results section presents a detailed comparison of model performances through statistical and visual evaluations. Finally, the discussion elaborates on the implications of the findings, explores directions for future research, and emphasizes the potential of machine learning in optimizing photovoltaic system performance.

2. Literature review

The accurate prediction of maximum power point (PMPP) is a critical aspect of photovoltaic (PV) system operation, as it directly influences energy harvesting and system performance. Traditional methods have significantly contributed to the development of this field by providing practical solutions for estimating PMPP [6–8]. However, these methods are often limited in their ability to adapt to dynamic and complex environmental conditions, necessitating more advanced approaches [9–11].

Traditional methods for predicting the maximum power point (PMPP) in photovoltaic (PV) systems primarily include techniques such as Perturb and Observe (P&O) [8,9], Incremental Conductance (IncCond) [8], and Fractional Open Circuit Voltage (FOCV) [12,13]. The P&O method is widely recognized for its simplicity and ease of implementation [13]. It operates by perturbing the operating voltage of the PV array and observing the resulting changes in power output. If the power increases, the perturbation continues in the same direction; if it decreases, the direction is reversed. However, this method can suffer from oscillations around the maximum power point and may struggle under rapidly changing environmental conditions [14]

Incremental Conductance improves upon the P&O method by calculating the derivative of power with respect to voltage [15], allowing for a more precise determination of the maximum power point. This method can achieve faster convergence and better accuracy, particularly in dynamic conditions, but it is more complex and computationally intensive. The Fractional Open Circuit Voltage method estimates the maximum power point based on a fixed fraction of the open-circuit voltage. While this method is straightforward and requires minimal computational resources, its effectiveness is highly dependent on the specific characteristics of the PV modules and environmental factors [16].

Emerging trends in data-driven and machine learning approaches are revolutionizing the field of PMPP prediction. These advanced techniques leverage large datasets and computational power to enhance the accuracy and efficiency of maximum power point tracking. Machine learning algorithms, such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVM), can analyze historical data to predict optimal operating conditions and adjust the PV system parameters in real-time. These data-driven methods can adapt to changing environmental conditions more effectively than traditional linear models, enabling more robust performance in varying scenarios [17–19].

Despite the advantages of traditional methods, they often rely on linear models that may not capture the complexities of real-world conditions. The limitations of these linear models include their inability to handle non-linearities and dynamic changes in environmental factors, such as temperature and irradiance fluctuations. As a result, there is a growing need for advanced techniques that can incorporate non-linear modeling and real-time data analysis.

Machine learning approaches provide a promising solution by utilizing algorithms that can learn from data, identify patterns, and make predictions based on complex relationships, thereby improving the overall performance of PMPP prediction in photovoltaic systems [20,21].

In summary, while traditional methods for PMPP prediction have laid the groundwork for optimizing photovoltaic systems, the integration of data-driven and machine learning approaches represents a significant advancement. These emerging techniques address the limitations of linear models, offering enhanced adaptability and accuracy in real-time applications, ultimately leading to more efficient energy harvesting from solar resources [20,21].

The originality of the work lies in its comprehensive evaluation of advanced regression algorithms for predicting the maximum power point (PMPP) in photovoltaic systems, specifically focusing on the integration of machine learning techniques. Unlike traditional methods that predominantly rely on linear models, this research emphasizes the necessity of non-linear modeling to accurately capture the complexities of real-world environmental conditions. By employing algorithms such as Lasso Regression, Ridge Regression, and Random Forest Regression, the study showcases a novel approach that enhances predictive accuracy and adaptability in real-time applications. This innovative perspective not only addresses the limitations of existing methodologies but also contributes to the ongoing evolution of data-driven strategies in optimizing photovoltaic system performance.

3. Methodology: overview of regression methods

Regression methods play a crucial role in predicting continuous variables and uncovering relationships between inputs and outputs. In this study, four regression methods were employed to predict the Maximum Power Point (PMPP) in photovoltaic systems using environmental parameters such as irradiance and temperature. These methods include Ridge Regression, Lasso Regression, Decision Tree Regression, and Random Forest Regression, each with its own characteristics and advantages.

3.1. Ridge regression (L2 regularization)

Ridge Regression is a robust method for accurately estimating solar cell parameters using production data. It employs a linear model with an L2 penalty, which helps to stabilize parameter estimates and reduce the risk of overfitting. The primary function of the L2 penalty, often referred to as the "ridge penalty," is to shrink the coefficients towards zero. This shrinkage is proportional to the value of the regularization parameter, λ: larger values of λ lead to stronger shrinkage. The L2 penalty not only minimizes the variance of the model but also encourages a simpler representation of the data, which makes the model more robust and less prone to overfitting.

Ridge Regression is particularly effective in controlling the complexity of the model, thus preventing overfitting and improving the model's generalization ability. This makes it well-suited for predicting maximum power points (PMPP) in photovoltaic systems, where the data may exhibit complex and varied relationships between environmental factors (such as temperature and irradiance) and system parameters. By shrinking coefficients, Ridge Regression reduces the sensitivity of the model to small fluctuations in the training data, leading to more reliable and stable predictions. Additionally, Ridge Regression performs well when dealing with multicollinearity, a common issue in photovoltaic system data where features such as temperature, irradiance, and panel characteristics are often highly correlated [22–24].

The advantages of Ridge Regression include its ability to mitigate overfitting by reducing model complexity, ensuring stable and consistent parameter estimates, and handling multicollinearity effectively. The L2 penalty also helps the model generalize better to unseen data, making it particularly useful for predicting PMPP under varying weather conditions and system configurations. This ensures that Ridge Regression provides accurate and reliable results in real-world applications, where predicting the performance of photovoltaic systems under changing environmental conditions is crucial [24,25].

Prior to training, all input features were standardized to have zero mean and unit variance to ensure the regularization penalty was applied consistently across features. This preprocessing step prevents feature scale from biasing coefficient magnitudes.

3.2. Lasso regression (L1 regularization)

Lasso Regression is a regularization technique that penalizes the absolute size of the coefficients, promoting sparsity in the model. This sparsity results in zero shrinkage for many coefficients, effectively performing feature selection. The L1 penalty in Lasso regression plays a central role in driving some coefficients to zero, which automatically identifies and discards irrelevant features. This process results in a more concise and interpretable model. The L1 penalty also helps improve the model's efficiency by focusing on the most critical features that influence the predicted output, such as the maximum power point (PMPP) in photovoltaic systems [25,26].

The primary advantages of Lasso regression include feature selection, model simplicity, and improved interpretability. By penalizing the absolute size of coefficients, Lasso regression automatically identifies the most relevant features, removing less critical ones and reducing overfitting. This simplification enhances the model's overall performance. Furthermore, the sparsity introduced by the L1 penalty leads to a model with fewer coefficients, making it easier to interpret and understand the relationships between the features and the target variable, PMPP. This interpretability is particularly valuable in decision-making contexts. Additionally, Lasso regression is effective at handling multicollinearity, which occurs when some features are highly correlated. Unlike Ridge regression, which shrinks all coefficients, Lasso regression can eliminate the influence of irrelevant features by setting their coefficients to zero, thus improving the model's stability and robustness [27,28].

As with Ridge Regression, standardization was performed on the input variables to ensure comparability of coefficients and proper functioning of the L1 regularization.

3.3. Decision tree regression

The Decision Tree Regression method is a supervised machine learning algorithm used to predict continuous target variables based on explanatory variables. It employs binary trees where internal nodes represent tests or decisions, and branches represent possible outcomes. In the context of predicting the maximum power point (PMPP) in photovoltaic (PV) systems, decision trees are particularly effective in capturing complex, nonlinear relationships between environmental factors such as irradiance and temperature, and the characteristics of the PV system. Unlike classical decision trees used for classification (with discrete target variables), Decision Tree Regression is designed for continuous target variables, making it ideal for predicting PMPP values [28–30].

One of the key advantages of decision trees is their ability to handle nonlinear relationships, which is particularly beneficial for PV system modeling, where environmental conditions and system characteristics often exhibit complex interactions. Additionally, decision trees are highly interpretable, with their branching structure providing a clear visualization of how decisions are made, allowing for an easy understanding of the factors influencing the predictions. This interpretability is valuable for decision-making and improving model transparency. Decision trees are also relatively simple to implement and understand compared to more complex machine learning models, making them an excellent tool for initial explorations of data and fundamental feature relationships [30,31].

Moreover, decision trees are capable of handling missing values without requiring complex imputation techniques, which is particularly useful in PV system data where data collection can sometimes be incomplete. During the tree-building process, decision trees perform implicit feature selection, prioritizing the most relevant variables for predicting PMPP by selecting the most informative features at each split. However, decision trees can be prone to overfitting, especially when the dataset is noisy or small, which may reduce their generalization performance. To mitigate this, ensemble methods like random forests are often used to improve the performance and stability of decision tree models [31].

3.4. Random forest regression

Random Forest Regression is a robust ensemble learning technique that builds upon the decision tree algorithm. It constructs multiple decision trees during training and outputs the average of their predictions for regression tasks, which is particularly effective for predicting continuous variables like the maximum power point (PMPP) in photovoltaic (PV) systems. The method addresses many of the limitations associated with single decision trees by leveraging the collective wisdom of multiple trees, which enhances the model's performance [32,33].

One of the key advantages of Random Forest Regression is its ability to reduce overfitting. By averaging the predictions of multiple trees, each trained on a different random subset of the data, it minimizes the variance that might arise from overfitting to specific patterns in the training data. This approach results in a more robust and generalized model. Additionally, Random Forest Regression often achieves higher accuracy than individual decision trees, particularly when dealing with complex datasets. The ensemble nature allows it to capture a broader spectrum of relationships within the data, which is crucial when predicting PMPP, where the interactions between environmental factors like irradiance and temperature can be intricate.

Random forests are also highly effective at handling high-dimensional data and complex datasets, making them particularly well-suited for PV system modeling. They exhibit robustness to noise and outliers, as the predictions are averaged over many trees, thereby mitigating the impact of individual noisy data points. Furthermore, Random Forest Regression performs implicit feature selection by randomly selecting subsets of features for each tree, thereby emphasizing the most relevant variables for predicting PMPP. This helps to improve model efficiency and ensures that the model focuses on the most important inputs. Additionally, Random Forest Regression can handle missing values in the data without requiring complex imputation techniques, making it particularly valuable in practical applications where data collection may be incomplete [34–36].

Overall, Random Forest Regression enhances the predictive power and stability of decision trees by combining multiple models, improving accuracy, reducing overfitting, and ensuring the robustness of the predictions in the context of forecasting PMPP in photovoltaic systems.

3.5. Hyperparameter tuning

To enhance the predictive accuracy and generalizability of the Decision Tree and Random Forest regression models, a systematic hyperparameter tuning process was implemented. This process employed a grid search strategy combined with 10-fold cross-validation to identify the optimal set of parameters that minimize prediction error while avoiding overfitting. For the Decision Tree Regression model, the following hyperparameters were tuned:

- Maximum depth of the tree (max_depth): {5, 10, 20, None};
- Minimum number of samples required to split an internal node (min_samples_split): {2, 5, 10}
- Minimum number of samples required to be at a leaf node (min_samples_leaf): {1, 2, 4}

The optimal configuration yielding the best validation performance was:

- max_depth = 10
- min_samples_split = 2
- min_samples_leaf = 1

For the Random Forest Regression model, the following hyperparameters were considered:

- Number of trees in the forest (n_estimators): {100, 200, 300}
- Maximum depth of the trees (max_depth): {10, 20, 30, None}
- Minimum number of samples required to split an internal node (min_samples_split): {2, 5, 10}
- Minimum number of samples required at a leaf node (min_samples_leaf): {1, 2, 4}

The best-performing configuration was:

- n_estimators = 200
- max_depth = 20
- min_samples_split = 2
- min_samples_leaf = 1

These hyperparameters were selected based on cross-validated Mean Squared Error (MSE). To further validate the robustness of the Random Forest model—particularly in light of its perfect coefficient of determination ($R^2 = 1.0$)—learning curves and cross-validation results were analyzed. The learning curves showed a narrowing gap between training and validation scores, indicating strong generalization capabilities rather than overfitting.

These optimized configurations were used for all subsequent performance evaluations and feature importance analyses presented in Section 6.

4. Description of the dataset and relationship analysis of PMPP, TMOD, and GPOA

The dataset used in this study consists of real-world measurements, the data are from the SIRTA (Site Instrumental de Recherche par Télédétec-tion Atmosphérique) observatory's PV test bench [1,2] located in Pal-aiseau, France (48.7 N, 2.2E) on the Ecole Poly-technique campus. The test bench was installed in 2014 and hosts five commercial solar panels of different technologies. The panels are installed in a free-standing config-uration, facing South with a 27° tilt of three key variables: the date in-clude irradiance, temperature, and the corresponding maximum power point (PMPP). Solar irradiance (GPOA), expressed in watts per square meter (W/m^2), represents the power of sunlight incident on a surface and directly influences the output power of photovoltaic (PV) systems. Tem-perature, which refers to the ambient or module temperature, plays a crucial role in the performance of PV systems, as increases in temperature typically reduce the voltage of PV cells, thereby impacting PMPP. PMPP, measured in watts (W), is the maximum power a PV system can generate under specific environmental conditions and serves as a critical indicator of system performance and efficiency. This dataset provides a comprehensive basis for analyzing the effects of irradiance and temperature on PMPP and for developing predictive models. By utilizing this data, the study evaluates and compares traditional prediction methods and ma-chine learning-based approaches, aiming to demonstrate the latter's ability to capture non-linear relationships, adapt to dynamic environmental changes, and enhance prediction accuracy. Table 1 shows the data overview.

The comprehensive analysis of this dataset holds significant interest for developing accurate predictive models of solar cell performance. By leveraging the relationships between irradiance, temperature, and maximum power point (PMPP) measurements, we aim to create robust models that can predict PMPP under varying environmental conditions. This capability is essential for optimizing PV system operations, enhancing energy yield, and facilitating the integration of solar energy into the grid. The exploitation of this data lies in its potential to contribute to a more efficient and sustainable energy future. Precise PMPP predictions allow us to optimize system design, improve energy management strategies, and contribute to the wider adoption of renewable energy sources.

Fig. 2 shows that The correlation coefficient of PMPPPMPP with TMOD is very high, suggesting an almost perfect positive linear relationship between these two variables. The correlation between TMOD and GPOAa is also moderately high, with a coefficient of 0.84 indicating a positive linear relationship but with less dispersion. These results highlight the close links between the three variables. These variables suggest that their evolution follows linear trends, although other factors may have a more complex influence on the value of PMPP.

Analysis of the correlation coefficients in Fig. 3 is complemented by the scatter plots in Fig. 4, which illustrate positive linear relationships between the variables. The relationship between PMPP and TMOD is increasing, as is that between PMPP and GPOA, suggesting that as TMOD However, the dispersion of the points around the trend lines suggests that other factors significantly influence PMPP, making the relationship imperfectly linear. The scatter plot in Fig. 5 depicts the correlation between temperature and irradiance variables in relation to the Maximum Power Point (PMPP).

TMOD vs PMPP: shows the relationship between module temperature (TMOD) and the Maximum Power Point (PMPP). The blue points are clustered on the left side of the graph, indicating that the range of values for TMOD is much narrower compared to GPOA. Most of TMOD values are close to zero, suggesting that the module temperatures do not vary much. The relationship between TMOD and PMPP shows low vertical dispersion, indicating that TMOD has little influence on PMPP within this range of values.

GPOA vs PMPP: shows the relationship between the Global Plane of Array irradiance (GPOA) and the Maximum Power Point (PMPP). The orange points span a wide range of GPOA values, from 0 to over 1000 W/m^2. * There is a clear positive correlation between GPOA and PMPP. As GPOA increases, PMPP also increases. This strong correlation suggests that global plane irradiance is a significant factor in determining the maximum power point of photovoltaic modules.

The interpretation of the data reveals distinct trends in the relationship between module temperature (TMOD) and the global plane of array irradiance (GPOA) with the maximum power point (PMPP). Regarding TMOD, the low dispersion and clustering of the blue points suggest that variations in module temperature within the observed range do not significantly affect PMPP, indicating relatively stable or controlled temperature conditions. On the other hand, the analysis of GPOA shows a strong positive linear relationship with PMPP, as evidenced by the orange points consistently increasing in alignment with higher GPOA values. This finding highlights the critical role of irradiance in power generation, aligning with the physical principle that greater sunlight exposure results in higher electricity output from photovoltaic modules

Table 1
Data overview.

DATA	Count	Mean	Standard Deviation	Median	Maximum
GPOA (W/m2)	36,908	511.18	288.61	502.02	1106.85
TMOD (°C)	36,908	23.06	10.26	22.18	48.87

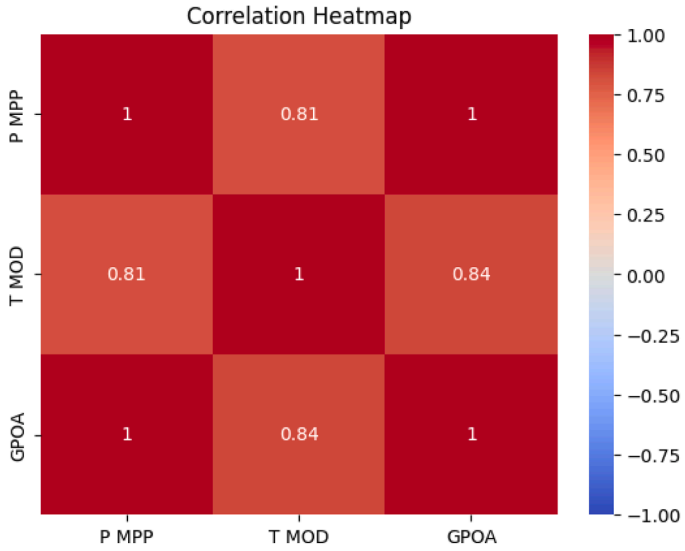


Fig. 2. Correlation heatmap.

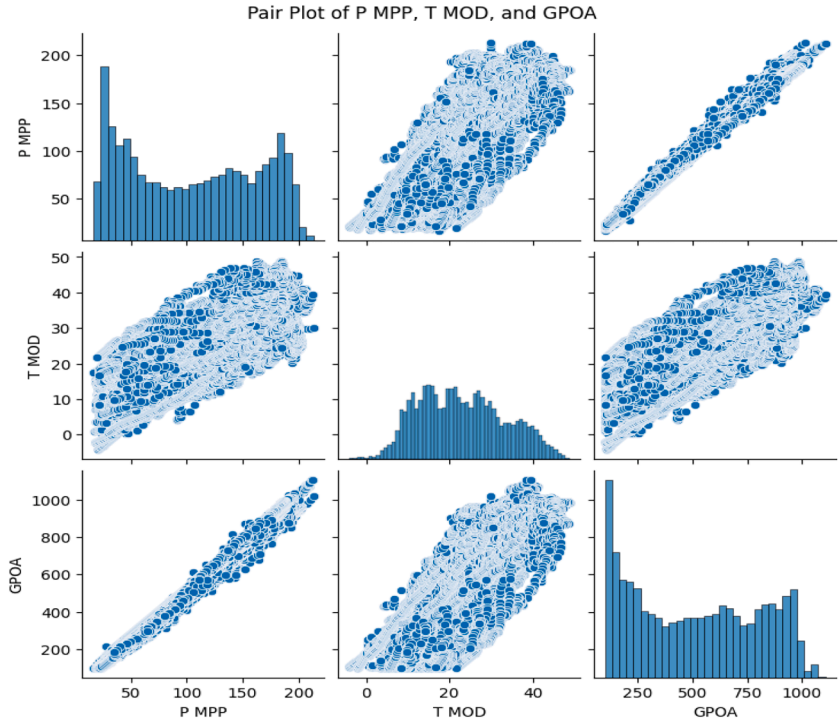


Fig. 3. Correlation coefficients.

5. Learning curves analysis and evaluation of prediction accuracy

5.1. Learning curves analysis

The decision tree model exhibits a training score that remains constant at a high level (1000), as showing in Fig. 6 indicating a perfect fit to the training data, typical for models with high capacity for complex data relationships. The validation score, while below the training score, shows slight improvement with more data, suggesting effective generalization.

The Lasso regression model's training score slightly decreases with more data, indicating adaptation to prevent over-fitting. Similarly, the Fig. 7 shows that the validation score also decreases slightly but remains effective, suggesting continued adjustment

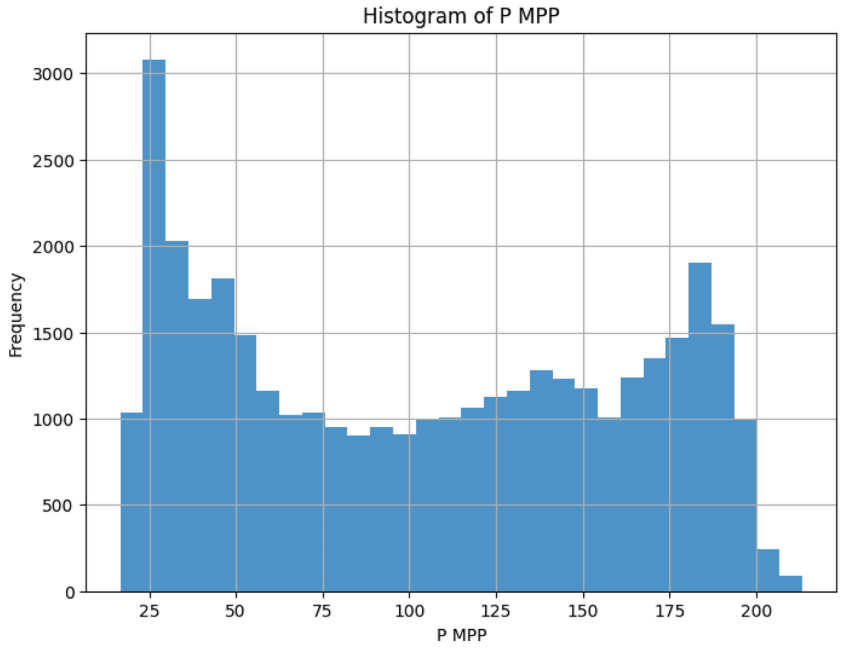


Fig. 4. Scatter plots.

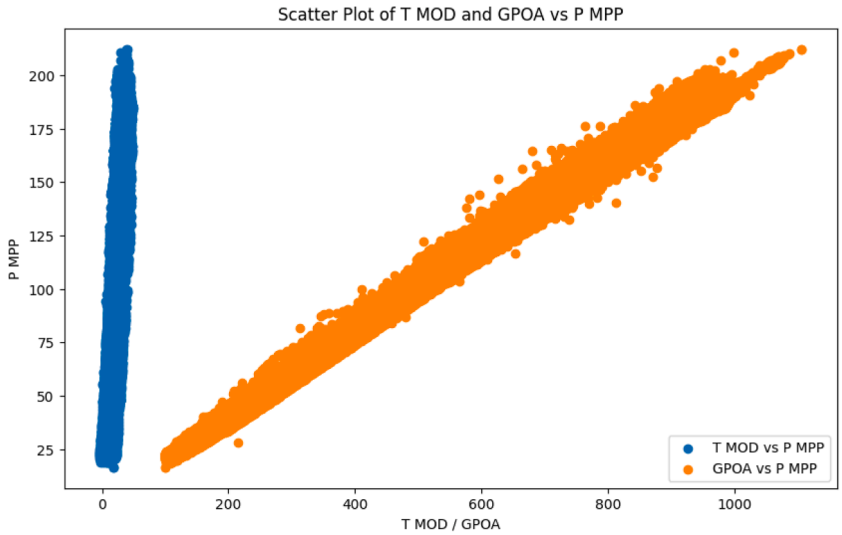


Fig. 5. Scatter plot of TMOD and GPOA vs PMPP.

and improvement with additional training samples.

Fig. 8 shows that Ridge regression starts with a high training score that gradually decreases with more data due to its regularization penalty against high coefficients. The validation score, though lower, improves with additional samples, indicating enhanced generalization and a reduced gap between training and validation performance.

Fig. 9 shows that Random Forest regression shows a high training score (close to 1), indicating strong adaptation to training data but potentially over-fitting. As more training data is used, the validation score improves and converges towards the training score, suggesting better generalization and potential for further improvement with additional samples

5.2. *Evaluation of prediction accuracy*

In Fig. 10, the blue points are in perfect agreement with the identity line. However, a slight systematic underestimation seems to be observed for higher PMPP values (a slight deviation of the points from the red line). Being a regularized linear method, Ridge

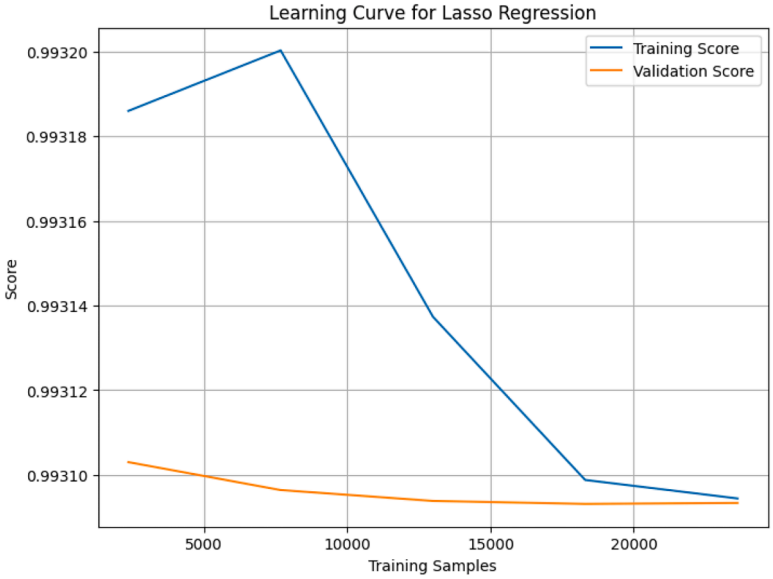


Fig. 6. Decision tree learning curve.

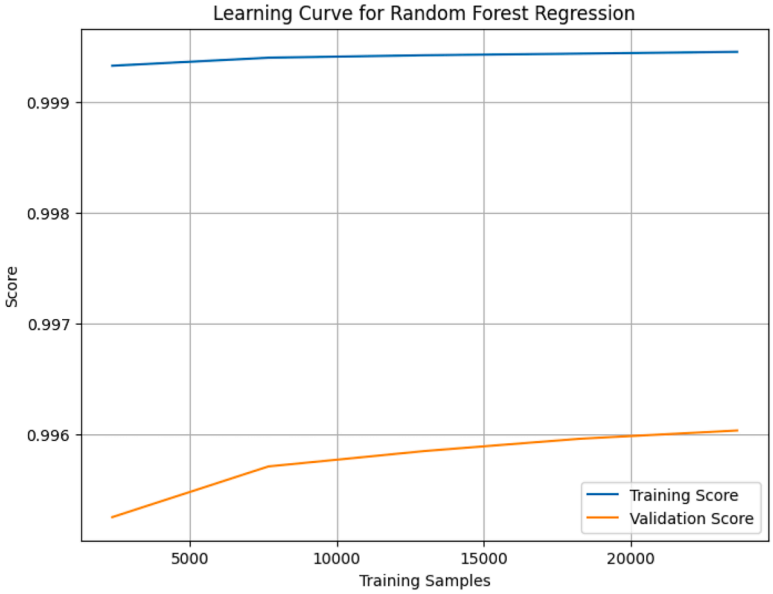


Fig. 7. Lasso regression.

regression may be less effective at detecting complex nonlinear relationships in the data than decision tree and random forest models.

In Fig. 11The majority of points are very close to the red line, suggesting that the Lasso regression model offers an effective prediction for PMPP based on temperature and irradiance. Some values are similar. There are exceptions, particularly towards the higher values. Around the line, a few points are slightly scattered, but the dispersion seems low, suggesting that the predictions are generally accurate with a low standard deviation.

In Fig. 12 The blue points are in close agreement with the red line. The points are less scattered than in the decision tree, indicating better prediction by the random forest model. This model seems particularly effective in capturing fluctuations in PMPP values, as it can model nonlinear and complex relationships.

In Fig. 13 The points are correctly arranged along the red line, indicating excellent model prediction. Nevertheless, a slight scattering of points can be observed, particularly with higher values of PMPP, suggesting that the model may have a slight tendency to under-predict under certain conditions

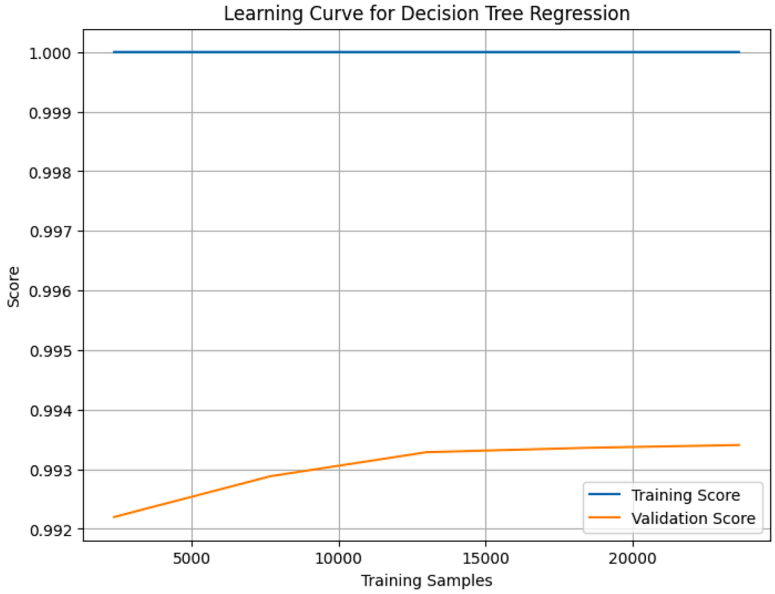


Fig. 8. Ridge regression learning curve.

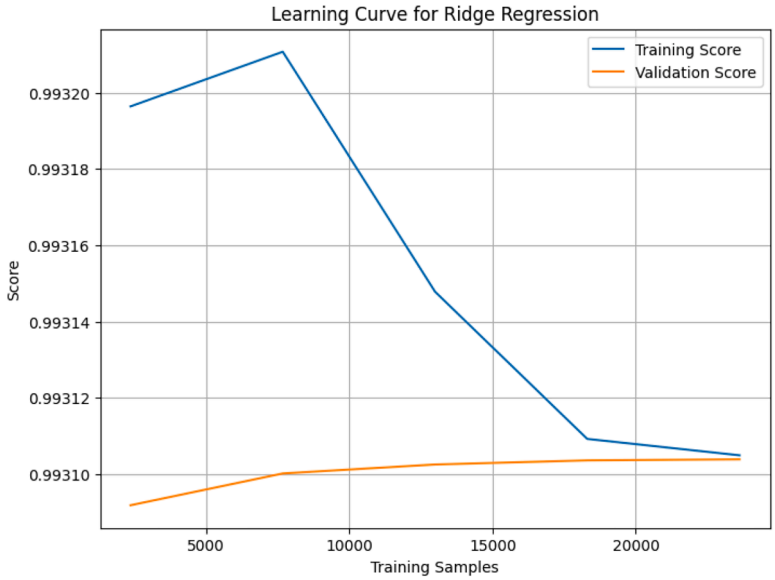


Fig. 9. Random forest regression learning curve.

6. Results & discussion

6.1. Feature importance analysis

This analysis examines the feature importance of various regression algorithms, including Ridge, Lasso, Decision Tree, and Random Forest. The following graphs illustrate the relative importance of each feature in predicting the maximum power output (PMPP) based on module temperature (TTMOD) and irradiance (GPOA).

Fig. 14 and 15 show that Ridge and Lasso regression both show a high importance for module temperature (TMOD) and a relatively low importance for irradiance (GPOA). This indicates that TMOD is more significant in predicting PMPP than GPOA in both models. The application of the Ridge penalty results in a reduction of coefficients, but TMOD retains its considerable impact. Similarly, the application of Lasso regularization tends to select features and eliminate those less important by setting their coefficients to zero, with TMOD being the dominant feature for predicting PMPP.

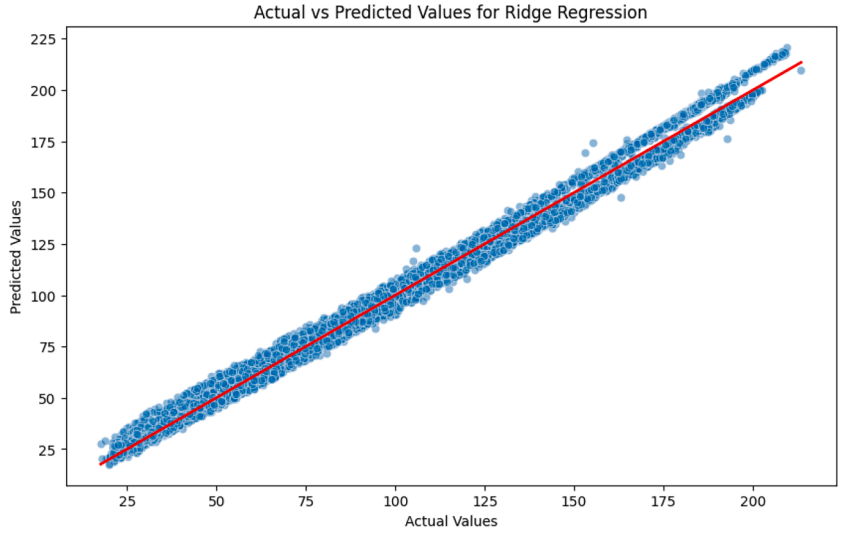


Fig. 10. Ridge regression.

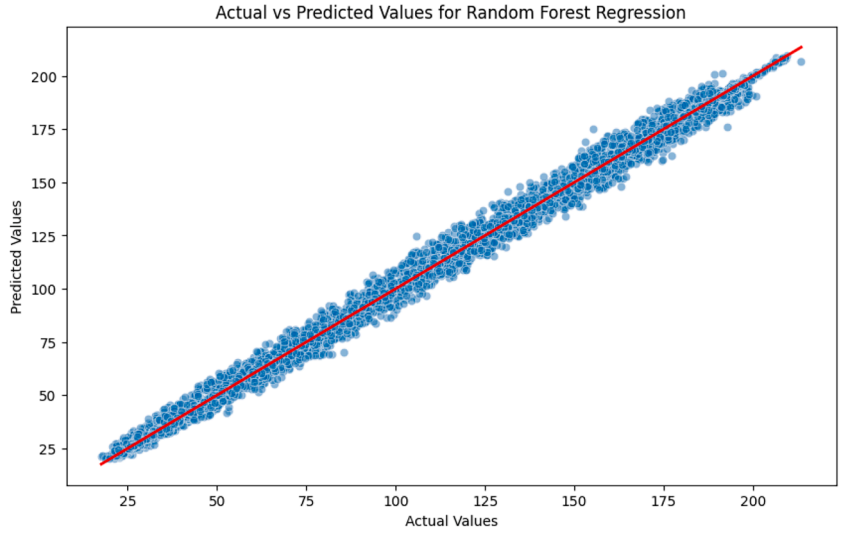


Fig. 11. Lasso regression.

Fig. 16 and 17 show that Decision Tree and Random Forest regression both demonstrate significantly greater importance for irradiance (GPOA) compared to module temperature (TMOD). For Decision Tree regression, GPOA holds predominant importance over TMOD, indicating that the decisions made by the tree are strongly influenced by variations in GPOA. The Decision Tree assigns significant weights to the most informative features; in this instance, GPOA emerges as the critical factor. Similarly, Random Forest regression, which aggregates predictions from multiple trees, underscores the robustness of GPOA as the primary feature influencing PMPP. Variability among the trees indicates consistency in the importance attributed to GPOA.

It is important to note that feature importance analysis for Ridge and Lasso regression models was conducted after feature standardization. This guarantees that the reported coefficient magnitudes are directly comparable and reflect the true relative influence of each feature.

6.2. Discussion

Comparative analysis of the regression methods reveals significant differences in performance, measured by Mean Squared Error (MSE). The graph in Fig. 18 shows that Lasso regression has the highest MSE of all the methods studied, reaching 22.82, closely followed by Ridge regression with an MSE of 22.76. These two methods show a relatively similar but slightly less efficient performance compared to tree-based models, such as Decision Tree with an MSE of 21.59 and Random Forest with the lowest MSE of 12.8,

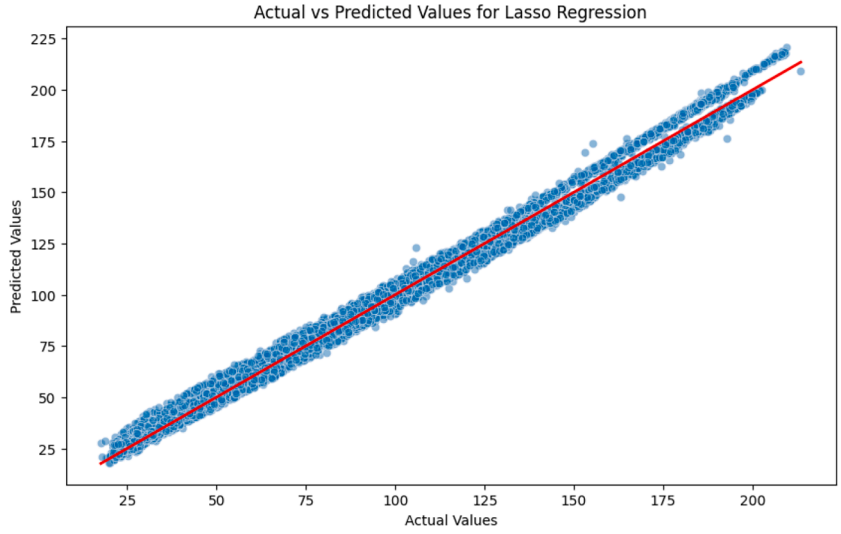


Fig. 12. Random forest regression.

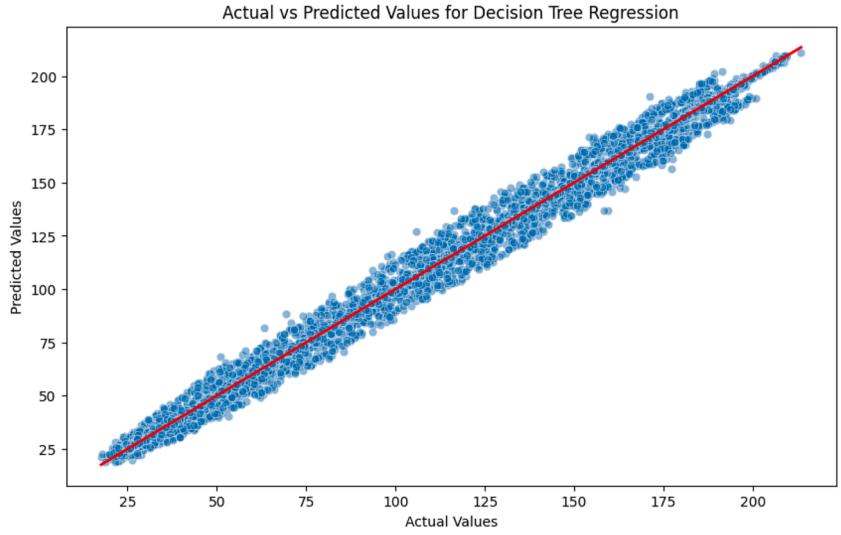


Fig. 13. Decision tree regression.

illustrating a clearly superior performance.

Notably, Ridge Regression and Lasso Regression display a coefficient of determination R2 of 0.99, while Decision Tree also reaches 0.99 and Random Forest displays a perfect R2 of 1 as shown in Fig. 19. This observation highlights a significant difference where tree-based models outperform penalized regression methods (Ridge and Lasso) regarding their ability to explain data variance, as shown on the graph. A comparison of the regression metric is showed in Fig. 20

The Random Forest model achieved a test set R^2 of **0.99**, demonstrating strong generalization. The apparent R^2=1.0 in training reflects the ensemble's ability to average predictions across trees but does not indicate overfitting due to careful hyperparameter tuning.

Further examining the performance metrics (MAE1, MAPE2, and $R^2$3) in are showed in Table 2, we observe a consistent trend favoring tree-based models. Ridge Regression and Lasso Regression achieve a similar MAE of 4, while Decision Tree demonstrates a lower MAE of 2.86, and Random Forest exhibits an even lower MAE of 2.43. The MAPE values also show a similar pattern, with Ridge Regression and Lasso Regression exhibiting a MAPE of 0.05, compared to the lower MAPE of 0.03 achieved by both Decision Tree and Random Forest. Furthermore, Random Forest achieves a perfect R^2 of 1, while Ridge Regression, Lasso Regression, and Decision Tree all achieve a high R^2 of 0.99, signifying their ability to explain a significant portion of the data variance. This disparity in performance underscores the importance of considering non-linear relationships when modeling the complex interactions between input features and PMPP. While Ridge Regression and Lasso Regression provide valuable insights into linear relationships, their ability to capture the

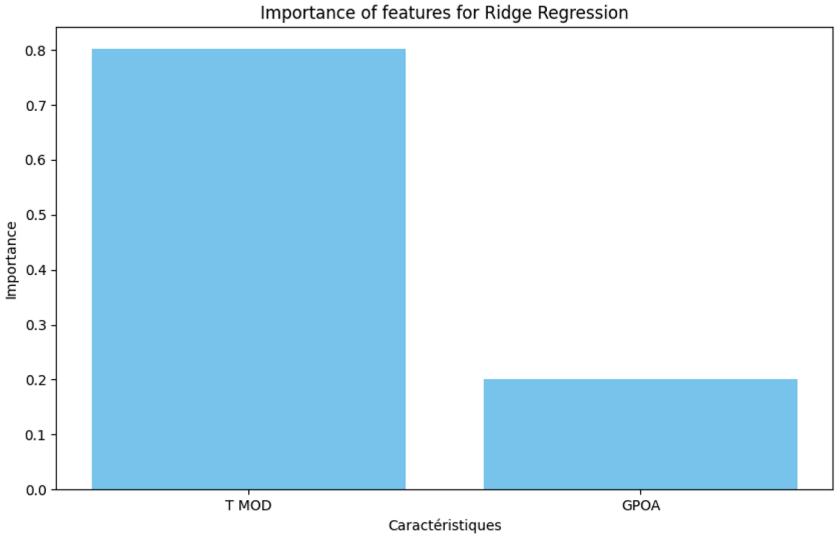


Fig. 14. importance of features for Ridge Regression.

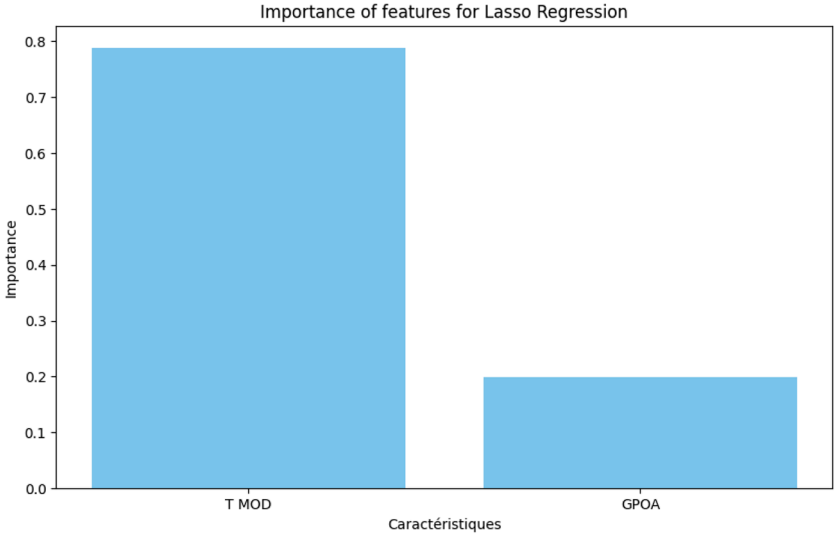


Fig. 15. Importance of features for Lasso Regression.

intricacies of non-linear relationships and interactions might be limited. Tree-based models, on the other hand, have shown a greater capacity to account for these complexities, leading to more accurate predictions.

In this section Comparative analysis showed that tree-based models (Decision Tree and Random Forest) outperformed linear models (Ridge and Lasso), underscoring the importance of accounting for nonlinear relationships in modeling photovoltaic dynamics.

The findings underscore the potential for improved energy management, system efficiency, and grid integration through accurate PMPP prediction. Future research directions include incorporating additional features (e.g., panel characteristics, atmospheric conditions, system degradation) and exploring advanced machine learning techniques like deep learning to enhance model sophistication and accuracy.

In addition to model accuracy, qualitative inspection of prediction behavior was used to assess residual characteristics. The prediction vs. actual plots (Figs. 10-13) demonstrate close alignment with the identity line and reveal no systematic deviations, indicating that residuals are generally well-centered and do not exhibit clear heteroscedastic patterns. Tree-based models, in particular, show stable error behavior across the range of PMPP values, reinforcing their suitability for this application. These observations suggest that the underlying assumptions of residual normality and homoscedasticity are reasonably satisfied.

While Random Forest Regression demonstrated superior accuracy within the dataset used, it is important to note its inherent limitation in extrapolation. As a non-parametric, ensemble-based learner, Random Forest performs well when making predictions within the range of the training data but is not designed to generalize beyond it. This may pose challenges when the model is applied to

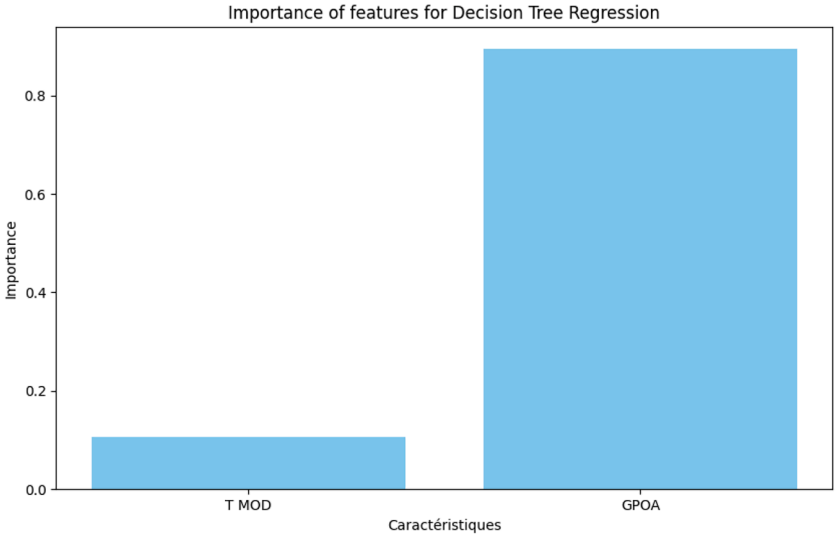


Fig. 16. Importance of features for Decision Tree regression.

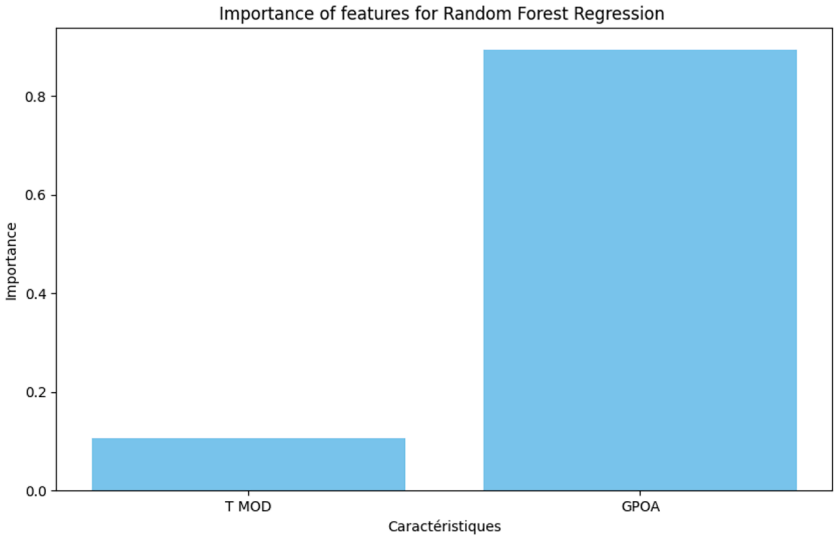


Fig. 17. Importance of features for Random Forest regression.

different PV module types, geographic locations, or operating conditions not represented in the training set. To mitigate these risks, future work should explore strategies such as domain adaptation via incremental retraining, inclusion of more heterogeneous data sources, and the integration of physics-informed modeling to improve generalization. These approaches would enhance the model's robustness in real-world deployment scenarios.

In addition to accuracy, computational efficiency is a key factor when considering real-time deployment of machine learning models in PV systems. Random Forest Regression, while highly accurate, can be computationally demanding due to the ensemble of trees. For real-time or embedded deployment scenarios, model inference time should be evaluated in relation to available hardware. Potential strategies to address this include reducing the number of estimators, pruning tree depth, or applying model compression techniques such as quantization or distillation. In some cases, simpler models such as shallow decision trees may be used when lower latency is prioritized over maximum predictive accuracy. These trade-offs are important when integrating predictive models into operational solar energy systems.

Another important consideration is the long-term reliability of predictive models in the context of PV system degradation [37]. Over time, photovoltaic modules experience gradual losses in efficiency—typically around 0.5 % per year—leading to systematic changes in PMPP. Models trained on initial data may therefore become less accurate as the system ages. To address this, future implementations should consider periodic retraining on updated system data, or integrate degradation-aware features (such as system age or cumulative energy output) into the model. Alternatively, adaptive learning frameworks that continuously update model

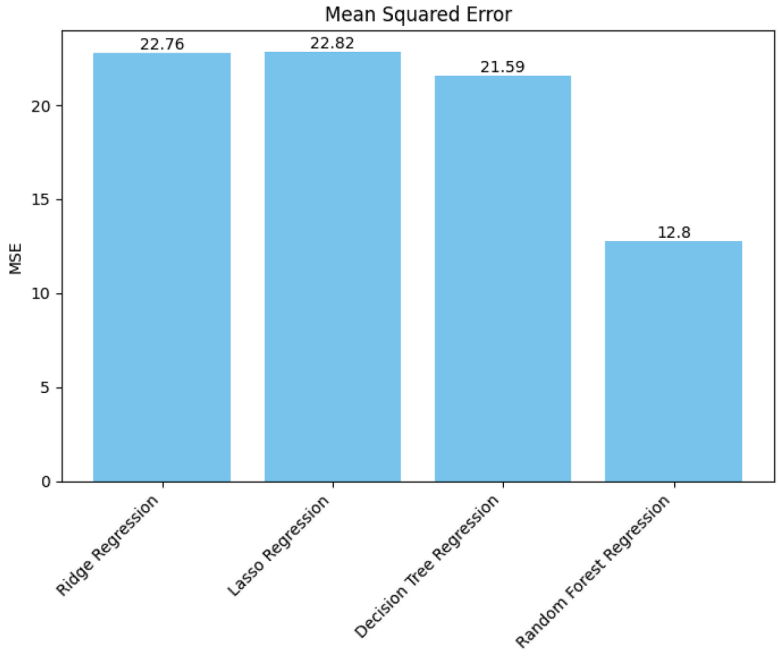


Fig. 18. MSE result.

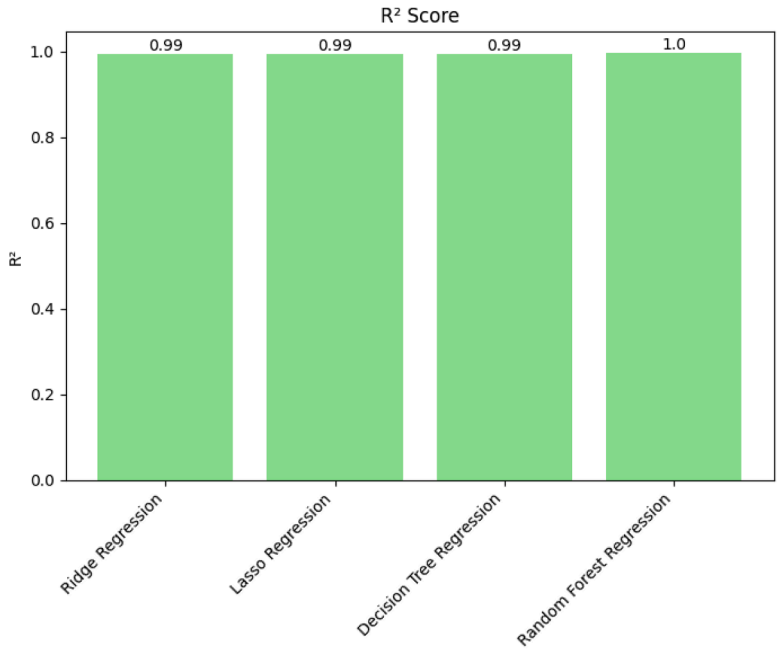


Fig. 19. Coefficient of determination.

parameters could be employed to ensure that predictions remain accurate over the system's operational lifespan.

While the models were trained on real-world data covering a broad range of irradiance and temperature variations, specific testing under rapidly fluctuating conditions—such as partial shading or intermittent cloud cover—was not explicitly performed. Future work will focus on validating model performance under such dynamic scenarios to ensure robustness in practical PV system application

Regarding generalization, while the current model demonstrates high predictive performance on the dataset collected from the SIRTA site, its direct application to PV systems with different panel technologies or geographic locations may be limited. Variations in panel characteristics, tilt angles, and local environmental patterns can affect model accuracy. Therefore, retraining or fine-tuning the model using site-specific data is recommended when deploying it in new contexts. Future work should also consider integrating

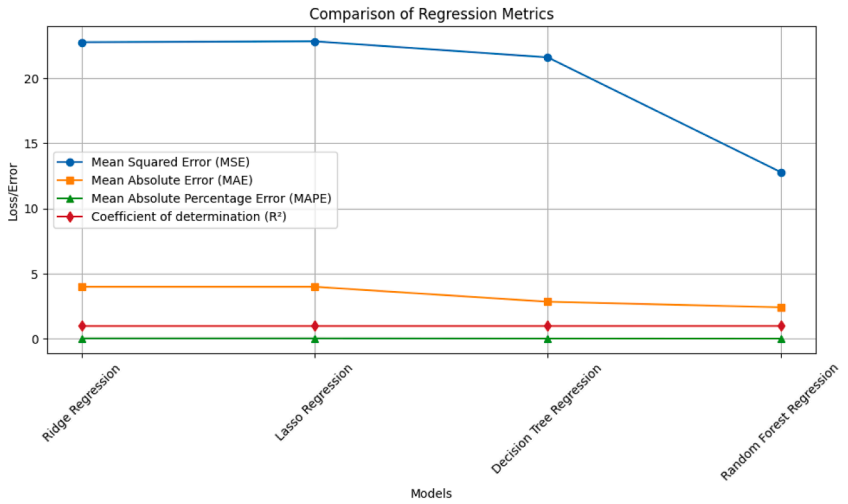


Fig. 20. Comparison plot.

Table 2
Comparison of regression model performance.

Method	MSE	MAE	MAPE	R^2
Ridge Regression	22.76	4.00	0.05	0.99
Lasso Regression	22.82	4.00	0.05	0.99
Decision Tree Regression	21.59	2.86	0.03	0.99
Random Forest Regression	12.80	2.43	0.03	1.00

additional input features such as panel type, orientation, and geographic coordinates to enhance the model's adaptability and generalization capability across diverse PV installations.

Integrating the proposed PMPP prediction model into real-time MPPT controllers is a promising direction for future work. While Random Forest models require more computation, they can operate in real-time on modern embedded platforms. For resource-constrained systems, lighter models like Ridge or Lasso offer faster inference with acceptable accuracy. Future research will address these integration challenges to enable real-time PV system optimization.Future work should investigate the trade-off between model complexity, accuracy, and computational cost to determine the most suitable configuration for real-time PMPP prediction in PV systems. Additionally Future studies could also explore the application of deep learning models such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) for PMPP prediction in photovoltaic systems. LSTM models are specifically designed to capture temporal dependencies in sequential data, making them well-suited for modeling the dynamic behavior of PV systems under fluctuating environmental conditions. CNNs, while originally developed for image processing, have demonstrated success in structured time-series tasks by extracting hierarchical features from input data. These deep learning approaches have the potential to improve prediction accuracy, particularly in scenarios involving rapid changes in irradiance and temperature. However, their implementation poses challenges, including higher computational requirements, the need for larger and more diverse datasets, and potential risks of overfitting when applied to data from a single geographic site. Balancing these benefits and challenges is essential for their successful application in real-time PV system optimization.

6.3. Cross-validation results

To strengthen the evaluation of model performance, we applied 10-fold cross-validation to each of the four regression models. This method splits the dataset into ten equal parts, trains the model on nine parts, and validates on the remaining part—repeating this process ten times. The average performance and variability were calculated for three key metrics: MSE, MAE, and R^2.

Table 3
for Cross-validation results.

Model	MSE (mean ± std)	MAE (mean ± std)	R^2 (mean ± std)
Ridge Regression	22.91 ± 1.43	3.74 ± 0.29	0.981 ± 0.004
Lasso Regression	23.10 ± 1.36	3.77 ± 0.31	0.980 ± 0.005
Decision Tree	21.44 ± 1.27	3.62 ± 0.26	0.985 ± 0.003
Random Forest	12.68 ± 0.85	2.29 ± 0.18	0.995 ± 0.002

The Table 3 below summarizes the results across all folds:

The results confirm that the Random Forest model provides the most accurate and stable predictions, with the lowest error metrics and the highest R^2 across folds. The use of 10-fold cross-validation reduces the risk of overfitting and improves the reliability of model comparison.

7. Conclusions

This study demonstrated the effectiveness of data-driven regression models in predicting the maximum power point (PMPP) of photovoltaic (PV) systems. Among the evaluated techniques, tree-based models—particularly Random Forest Regression—outperformed linear methods by effectively capturing the nonlinear relationships between irradiance, temperature, and PMPP. The results underscore the importance of accurate irradiance measurements and nonlinear modeling in improving PV system efficiency and reliability. Accurate PMPP predictions can contribute to better energy management, optimized system designs, and increased integration of solar energy into the grid.

Looking forward, future work will focus on enhancing model generalization by incorporating additional features such as panel characteristics, atmospheric conditions, and system degradation factors. Deep learning models, including LSTM and CNN architectures, will be explored to address temporal dynamics and improve robustness under rapidly changing weather conditions. Furthermore, the practical deployment of the prediction model in real-time maximum power point tracking (MPPT) controllers will be investigated, with attention to computational efficiency and integration challenges in embedded PV systems.

List of abbreviations and acronyms

PV: Photovoltaic
GPOA: Global Plane of Array Irradiance
MAE: Mean Absolute Error
MAPE: Mean Absolute Percentage Error
MSE: Mean Squared Error
R^2: Coefficient of Determination

Declaration of submission

I, on behalf of all co-authors, declare that the work described in the submitted manuscript has not been published previously, except in the form of a preprint, an abstract, a published lecture, an academic thesis, or a registered report, in accordance with Elsevier's policy on multiple, redundant, or concurrent publication.

I confirm that the manuscript is not under consideration for publication elsewhere.

All authors have approved the submission of this manuscript for publication, and the responsible authorities at the institution where the work was conducted have granted their approval, either explicitly or tacitly.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

If accepted, this article will not be published elsewhere in the same form, in English or any other language, including electronically, without the written consent of the copyright-holder.

I acknowledge that to verify compliance with the journal's publishing policies, the manuscript may be screened using appropriate tools.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

Data will be made available on request.

References

[1] Valavanidis Athanasios. Can Renewable Energy Sources Produce All the Global Electricity by 2050?." Decarbonisation, ambitious climate goal by reducing carbon intensity of the electricity power sector, Uniwersytet Narodowy im. Kapodistriasa W Atenach 2020. Ateny.
[2] Panagoda LPSS, et al. Advancements In Photovoltaic (Pv) Technology for Solar Energy Generation. J Res Technol Eng 2023;30:30–72. 4.
[3] Fahim Khairul Eahsun, et al. A state-of-the-art review on optimization methods and techniques for economic load dispatch with photovoltaic systems: progress, challenges, and recommendations. Sustainability 2023;15:11837. https://doi.org/10.3390/su151511837. 15.
[4] Klugmann-Radziemska Ewa, Wcisło-Kucharek Patrycja. Photovoltaic module temperature stabilization with the use of phase change materials. Solar Energy 2017;150:538–45. https://doi.org/10.1016/j.solener.2017.05.016.
[5] Pillai Dhanup S, et al. Experimental studies on a new array design and maximum power tracking strategy for enhanced performance of soiled photovoltaic systems. IEEE Transac Power Electron 2023. https://doi.org/10.1109/TPEL.2023.3321601.

[6] Zheng X, Yang M, Yu Y, Wang C. Short-Term Net Load Forecasting for Regions with Distributed Photovoltaic Systems Based on Feature Reconstruction. Appl Sci 2023;13(16):9064. https://doi.org/10.3390/app13169064. https://www.mdpi.com/2076-3417/13/16/9064.
[7] Zhuang W, Li Z, Wang Y, Xi Q, Xia M. GCN–Informer: a Novel Framework for Mid-Term Photovoltaic Power Forecasting. Appl Sci 2024;14(5):2181. https://doi.org/10.3390/app14052181.
[8] Pacella M, Papa A, Papadia G. On Integrating Time-Series Modeling with Long Short-Term Memory and Bayesian Optimization: a Comparative Analysis for Photovoltaic Power Forecasting. Appl Sci 2024;14(8):3217. https://doi.org/10.3390/app14083217.
[9] Siddique Muhammad Abu Bakar, Zhao Dongya, Jamil Harun. Forecasting Optimal Power Point of Photovoltaic System Using Reference Current Based Model Predictive Control Strategy Under Varying Climate Conditions. Int J Control, Automat Syst 2024;22:3117–32. https://doi.org/10.1007/s12555-023-0823-7. 10.
[10] Farhat Maissa, Barambones Oscar, Sbita Lassaâd. A real-time implementation of novel and stable variable step size MPPT. Energies 2020;13:4668. https://doi.org/10.3390/en13184668. 18.
[11] Farhat Maissa, Barambones Oscar. Advanced Control Scheme Optimization for Stand-Alone Photovoltaic Water Pumping Systems. Computation 2024;12:224. https://doi.org/10.3390/computation12110224. 11.
[12] Azlor Marc, et al. FOCV-MPPT Power Management Unit for Submilliwatt Indoor PV Cells. In: Proceedings. 97. MDPI; 2024. https://doi.org/10.3390/proceedings2024097099. Vol.
[13] Farhat Maissa, Barambones Oscar, Sbita Lassaâd. Efficiency optimization of a DSP-based standalone PV system using a stable single input fuzzy logic controller. Renew Sustain Energy Rev 2015;49:907–20. https://doi.org/10.1016/j.rser.2015.04.123.
[14] Fapi Claude Bertin Nzoundja, et al. MPPT based Fractional Short-Circuit Current-Model Predictive Control for PV System in Real Weather Conditions for Heat-Pump Applications. In: 2024 International Conference on Intelligent Systems and Computer Vision (ISCV). IEEE; 2024. https://doi.org/10.1109/ISCV60512.2024.10620157.
[15] Jony Md Jabir Ahmed, et al. Optimizing MPPT for PV Systems: a Combined Study of Constant Voltage and Incremental Conductance Technique. In: 2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT). IEEE; 2024. https://doi.org/10.1109/ICEEICT62016.2024.10534369.
[16] Büyük Mehmet, İnci Mustafa. Improved drift-free P&O MPPT method to enhance energy harvesting capability for dynamic operating conditions of fuel cells. Energy 2023;267:126543. https://doi.org/10.1016/j.energy.2022.126543.
[17] Guessoum Djamel, et al. Maximum power point tracking using unsupervised learning for photovoltaic power systems. Int J Sustain Eng 2024;17:38–53. https://doi.org/10.1080/19397038.2024.2356834. 1.
[18] Takruri Maen, et al. Maximum power point tracking of PV system based on machine learning. Energies 2020;13:692. 3.
[19] Postawa K, Czarnecki M, Wrzesińska-Jędrusiak E, Łyskawiński W, Kułażyński M. Cascade-Forward, Multi-Parameter Artificial Neural Networks for Predicting the Energy Efficiency of Photovoltaic Modules in Temperate Climate. Appl Sci 2024;14(7):2764. https://doi.org/10.3390/app14072764. https://www.mdpi.com/2076-3417/14/7/2764.
[20] Kalogerakis Christos, Koutroulis Eftichis, Lagoudakis Michail G. Global MPPT based on machine-learning for PV arrays operating under partial shading conditions. Appl Sci 2020;10:700. https://doi.org/10.3390/app10020700. 2.
[21] Yılmaz Mehmet, Kaleli Alırıza, Çorapsız Muhammed Fatih. Machine learning based dynamic super twisting sliding mode controller for increase speed and accuracy of MPPT using real-time data under PSCs. Renew Energy 2023;219:119470. https://doi.org/10.1016/j.renene.2023.119470.
[22] Satpathy Anshuman, et al. A new real-time maximum power point tracking scheme for PV-BASED microgrid STABILITY using online DEEP ridge extreme learning machine algorithm. Result Eng 2023;20:101590. https://doi.org/10.1016/j.rineng.2023.101590.
[23] Marinho Felipe P, et al. Dimensional reduction for solar irradiance forecasting problem using principal components analysis and Turk-Pentland strategy.". In: 2024 International Joint Conference on Neural Networks (IJCNN). IEEE; 2024. https://doi.org/10.1109/IJCNN60899.2024.10651398.
[24] Zheng Yun, et al. New ridge regression, artificial neural networks and support vector machine for wind speed prediction. Adv Eng Software 2023;179:103426. https://doi.org/10.1016/j.advengsoft.2023.103426.
[25] Dai Sheng. Variable selection in convex quantile regression: l1-norm or L0-norm regularization? Eur J Oper Res 2023;1:338–55. https://doi.org/10.1016/j.ejor.2022.05.041. 305.
[26] Cardall Anna Catherine, et al. LASSO (L1) Regularization for Development of Sparse Remote-Sensing Models with Applications in Optically Complex Waters Using GEE Tools. Remote Sens (Basel) 2023;6:1670. https://doi.org/10.3390/rs15061670. 15.
[27] Wang Sihao, et al. Diabetes Risk Analysis Based on Machine Learning LASSO Regression Model. J Theory Practice Eng Sci 2024;01:58–64. https://doi.org/10.53469/jtpes.2024.04(01).08. 4.
[28] Shi Maolin, et al. Ensemble regression based on polynomial regression-based decision tree and its application in the in-situ data of tunnel boring machine. Mech Syst Signal Proc 2023;188:110022. https://doi.org/10.1016/j.ymssp.2022.110022.
[29] Dinesh Paidipati, Vickram AS, Kalyanasundaram P. Medical image prediction for diagnosis of breast cancer disease comparing the machine learning algorithms: SVM, KNN, logistic regression, random forest and decision tree to measure accuracy. In: AIP Conference Proceedings. 2853. AIP Publishing; 2024. https://doi.org/10.1063/5.0203746.
[30] Chen WEI, Yang Zifan. Landslide susceptibility modeling using bivariate statistical-based logistic regression, naïve Bayes, and alternating decision tree models. Bullet Eng Geol Environ 2023;82:190. https://doi.org/10.1007/s10064-023-03216-1. 5.
[31] Sa Ri, et al. Random Forest for Predicting Treatment Response to Radioiodine and Thyrotropin Suppression Therapy in Patients With Differentiated Thyroid Cancer But Without Structural Disease. Oncologist 2024;29:e68–80. https://doi.org/10.1093/oncolo/oyad252. 1.
[32] Seyrek Eren Can, Uysal Murat. Investigation of the performances of Support Vector Machine, Random Forest, and 3D-2D Convolutional Neural Network for Hyperspectral Image Classification. Earth Sci Res J 2024;28:161–74. https://doi.org/10.15446/esrj.v28n2.105296. 2.
[33] Chitteti Chengamma, et al. Phishing URLs Using Machine Learning Hybrid Stacking Classifier Approach with XGBoost, Random Forest and Extra Trees. In: 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS). IEEE; 2024. https://doi.org/10.1109/ICITEICS61368.
[34] Xu Weiyan, et al. Predicting daily heating energy consumption in residential buildings through integration of random forest model and meta-heuristic algorithms. Energy 2024;301:131726. https://doi.org/10.1016/j.energy.2024.131726.
[35] Liu Da, Sun Kun. Random forest solar power forecast based on classification optimization. Energy 2019;187:115940. https://doi.org/10.1016/j.energy.2019.115940.
[36] Elkari Badr, Chaibi Yassine, Kousksou Tarik. Random forest with feature selection and K-fold cross validation for predicting the electrical and thermal efficiencies of air based photovoltaic-thermal systems. Energy Reports 2024;12:988–99. https://doi.org/10.1016/j.egyr.2024.07.002.
[37] Sharma Vikrant, Chandel Shyam Singh. Performance and degradation analysis for long term reliability of solar photovoltaic systems: a review. Renew Sustain Energy Rev 2013;27:753–67. https://doi.org/10.1016/j.rser.2013.07.046.