ORIGINAL RESEARCH Open access

# Data-Driven Maximum Power Point Prediction for Photovoltaic Systems Using Machine Learning Regression Models

Elena Petrova[1*], Ivan Georgiev[1]

## Abstract

Accurately estimating the Maximum Power Point (PMPP) in photovoltaic (PV) systems is essential for maximizing energy production and improving solar power capture. This investigation examines how data-centric methodologies can be applied to refine PMPP estimation, using sophisticated regression approaches such as Ridge Regression, Lasso Regression, Decision Tree Regression, and Random Forest Regression. By evaluating a dataset that includes irradiance, temperature, and PMPP values, the study contrasts the capacity of these algorithms to represent intricate nonlinear connections among critical factors. The outcomes demonstrate that tree-driven techniques, most prominently Random Forest Regression, outperform linear alternatives, exhibiting improved estimation precision and resilience. An evaluation of feature significance also confirms the overriding impact of irradiance (GPOA) on PMPP, stressing the need for accurate irradiance measurements. These discoveries point to the considerable potential of machine learning methods for refining PV system operation. Future work should focus on incorporating additional parameters, including meteorological variables and panel specifications, as well as exploring deep learning strategies to improve estimation accuracy further.

**Keywords** Photovoltaic systems, Maximum power point (PMPP), Data driven approaches, Regression techniques, Tree-based models, Nonlinear relationships

*Correspondence:
Elena Petrova
elena.petrova@uni-sofia.bg
[1] Department of AI-Based Electrical Networks, Faculty of Engineering, Sofia University, Sofia, Bulgaria

## Introduction

The significance of sustainable energy, particularly photovoltaic (PV) technology, has drawn considerable focus owing to its ability to counteract climate change and curtail dependency on fossil-based fuels. Photovoltaic installations transform sunlight directly into electric power, delivering an eco-friendly response to the world's escalating energy requirements. **Figure 1** is an example of a photovoltaic system. Embedding PV technology within the energy portfolio not only helps lower greenhouse gas output but also bolsters energy self-sufficiency and economic advancement. Investigations show that the rollout of PV infrastructure can yield meaningful decreases in carbon emissions, with calculations implying that a broad-based migration toward solar power might cut worldwide CO2 discharges by as much as 4.9 gigatons annually [1]. In addition, breakthroughs in PV technology, including improved conversion efficiencies and declining costs, have made solar power more accessible, encouraging widespread uptake across residential, commercial, and industrial settings [2]. The pivot toward renewable energy provision, specifically through the deployment of photovoltaic systems, stands as a cornerstone for realizing sustainability objectives and guaranteeing a robust energy landscape.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

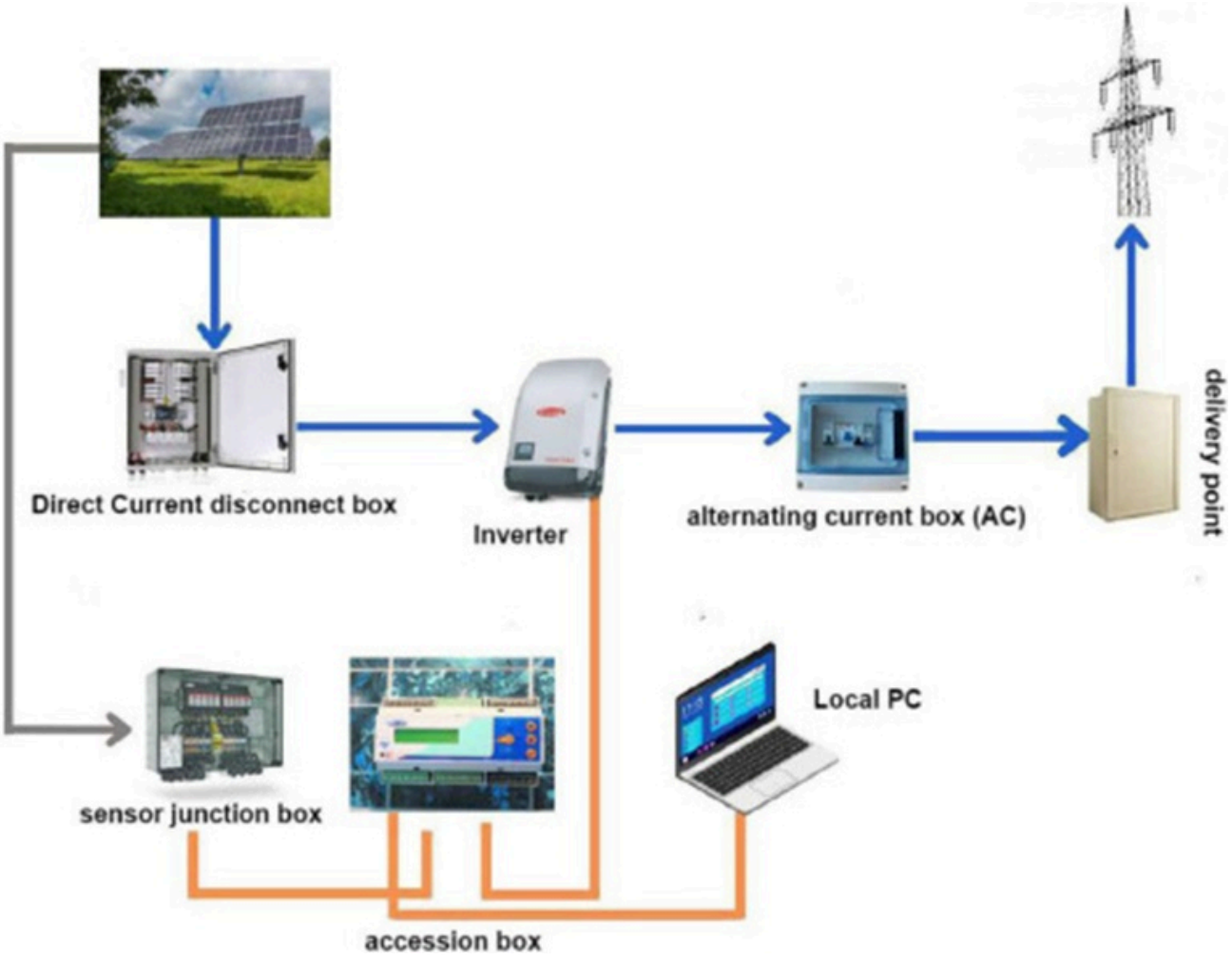



**Figure 1.** A photovoltaic system.

A fundamental obstacle confronting renewable energy infrastructure, especially photovoltaic (PV) systems, is achieving peak energy extraction amid shifting environmental conditions. Elements such as varying solar irradiance, thermal fluctuations, and localized shading profoundly influence the operational effectiveness of PV modules. These hurdles call for the development of sophisticated control and optimization frameworks to ensure reliable and maximized energy retrieval [3]. Research underscores that surmounting these obstacles demands robust algorithms capable of responding to dynamic operating conditions and minimizing energy dissipation.

Forecasting the maximum power point (PMPP) is integral to elevating the operational efficacy of PV systems. The concept involves locating the precise spot on the solar panel's I-V characteristic where the multiplication of current and voltage attains its peak, thus securing the highest power delivery. Reliable PMPP forecasting enables the PV system to continually recalibrate its operating state toward the most favorable PMPP despite variable environmental factors such as shifts in sunlight intensity and thermal conditions [4]. This objective is realized through various Maximum Power Point Tracking (MPPT) strategies, which modify the electrical load on the solar modules to maintain operation at the PMPP, thereby optimizing the power drawn from the solar assembly [5]. Progressive PMPP forecasting approaches incorporating machine learning or artificial intelligence can foresee environmental variations and preemptively modify system configurations, boosting both the productivity and dependability of PV installations. Indeed, addressing the challenges inherent in maximizing energy output and the contribution of PMPP estimation to system enhancement is indispensable for the progression and integration of solar electricity into the global energy system. Efficient MPPT methodologies that rely on accurate PMPP forecasting are vital for maximizing the financial returns on solar energy ventures.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

The operational behavior of photovoltaic (PV) infrastructure is intrinsically tied to surrounding conditions, chiefly sunlight intensity and temperature. PV modules manifest a negative thermal coefficient, customarily approximated at −0.4 °C [4]. This denotes that as the operational temperature of the PV unit climbs, its voltage production and resultant Maximum Power Point (PMPP) diminish. From a physical standpoint, this degradation of PMPP with escalating temperature stems from amplified electron-phonon interactions within the semiconducting medium, which produces greater electrical resistivity and a diminished open-circuit voltage [5]. Sunlight intensity, in contrast, has a direct bearing on the electrical current output of the PV unit. Higher irradiance means more photons impacting the photovoltaic cells, leading to an increased short-circuit current and a commensurately larger PMPP. That said, the influence of irradiance is intertwined with the unit's thermal state. During periods of strong irradiance, the module's temperature typically climbs, partially negating the increases in PMPP attributable to the adverse thermal coefficient. Hence, the combined effect of irradiance and temperature governs the net output and realizable PMPP of a PV system. Our machine learning frameworks are designed to represent these multifaceted, frequently nonlinear interdependencies to furnish reliable PMPP forecasts under assorted operating scenarios.

The core aim of this undertaking is to improve the forecasting fidelity of the maximum power point (PMPP) in photovoltaic (PV) installations through data-driven strategies. This research probes the associations linking environmental parameters, notably irradiance and temperature, to PMPP by employing advanced regression techniques. Through comparative evaluation of various methodologies, spanning linear formulations and tree-structured models, alongside the appraisal of feature relevance, the inquiry endeavors to identify the most effective means of representing the complex interactions present in the dataset. The insights generated are meant to support the advancement of solar energy harvesting efficiency by refining PV system functionality.

The ramifications of this study carry broad significance for the sustainable energy domain. Dependable PMPP estimations not only strengthen the efficiency and steadiness of PV installations but also facilitate smoother grid assimilation and more effective energy regulation schemes. Furthermore, understanding the relationship between environmental variables provides practical guidance for the planning and fine-tuning of PV deployments across a variety of geographical settings and climatic zones.

This study evaluates the predictive capability of multiple regression frameworks, including Ridge Regression, Lasso Regression, Decision Tree Regression, and Random Forest Regression, in forecasting PMPP. Each technique brings distinct strengths: Ridge and Lasso regressions are proficient at dealing with multicollinearity, whereas Decision Tree and Random Forest are adept at modeling nonlinear dependencies. An essential component of the investigation entails scrutinizing the relevance of principal inputs, such as irradiance and temperature, in shaping PMPP, thereby affording a deeper comprehension of the physical principles underpinning PV system dynamics.

The organization of this paper proceeds as follows: the introductory section surveys the obstacles and backdrop associated with forecasting the maximum power point (PMPP) in photovoltaic systems, accentuating the value of employing cutting-edge regression models. The methods section outlines the data provenance, preparation steps and regression techniques used in the analysis. The findings segment supplies an extensive comparison of model efficacy via statistical metrics and graphical representations. Lastly, the discussion unpacks the significance of the results, maps out avenues for subsequent inquiry, and highlights the transformative capacity of machine learning in optimizing photovoltaic system operation.

# Literature Review

Determining the maximum power point (PMPP) with precision is essential for operating photovoltaic (PV) installations, given its immediate effect on power capture and system-wide efficiency. Established methodologies have played a notable role in advancing this area by offering functional means to approximate PMPP [6-8]. Yet these techniques are often constrained by their inability to handle variable and intricate climatic scenarios, underscoring the need for more refined solutions [9-11].

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Long-standing procedures for approximating the maximum power point (PMPP) within photovoltaic (PV) configurations largely consist of approaches like Perturb and Observe (P&O) [8, 9], Incremental Conductance (IncCond) [8], and Fractional Open Circuit Voltage (FOCV) [12, 13]. The P&O strategy has gained extensive acceptance thanks to its uncomplicated nature and straightforward execution [13]. The concept involves introducing a disturbance to the PV array's voltage setpoint and then tracking the resulting shift in generated power. A rise in power prompts continued perturbation along the same trajectory; a decline triggers a reversal of that trajectory. A drawback, however, is that this technique tends to generate oscillations near the maximum power point and may falter when confronted with abruptly varying atmospheric conditions [14].

Incremental conductance builds upon the P&O foundation by deriving the instantaneous slope of the power-voltage characteristic [15], enabling a sharper identification of the maximum power point. This alternative can achieve quicker settling times and greater accuracy, particularly in fluctuating environments, though it adds sophistication and demands greater computational resources. The fractional open-circuit voltage technique infers the maximum power point by applying a constant multiplier to the open-circuit voltage. While commendable for its simplicity and low computational footprint, its practical success hinges heavily on the particular attributes of the PV hardware and the prevailing environmental circumstances [16].

Burgeoning developments in data-informed and machine-learning-based paradigms are transforming the domain of PMPP estimation. These state-of-the-art strategies capitalize on large data repositories and processing horsepower to improve the accuracy and speed of maximum power point tracking. Machine learning algorithms, exemplified by Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs), can sift through historical records, anticipate ideal operating setpoints, and recalibrate PV system controls on the fly. Such data-driven frameworks demonstrate greater agility in responding to shifting environmental factors than traditional linear constructs, thereby delivering more dependable operation across a range of contexts [17-19].

Notwithstanding the benefits offered by classical techniques, they habitually lean on linear formulations that may prove insufficient for encoding the subtleties of actual field conditions. The weak points of these linear constructs include an intrinsic failure to accommodate nonlinear behaviors and the swift variations characteristic of environmental drivers like temperature swings and irradiance transients. Accordingly, there is growing interest in state-of-the-art mechanisms capable of integrating nonlinear representations with streaming data interpretation.

Machine-learning-based paradigms offer a compelling response by leveraging algorithms that extract structure from datasets, discern regularities, and formulate projections anchored in multifaceted linkages, thereby raising the bar for holistic PMPP forecasting performance in photovoltaic contexts [20, 21].

To recapitulate, while time-honored methods of PMPP prediction have constructed the scaffolding for enhancing photovoltaic system output, the fusion of data-centric and machine-learning philosophies constitutes a transformative leap. These forward-looking innovations surmount the confines of linear constructs, affording greater agility and precision during live deployment, culminating in more productive solar energy extraction [20, 21].

What sets this study apart is its comprehensive scrutiny of cutting-edge regression algorithms for projecting the maximum power point (PMPP) in photovoltaic installations, with a focus on integrating machine-learning methodologies. While orthodox procedures rely heavily on linear schemas, this inquiry highlights the indispensability of nonlinear representation to mirror the nuance inherent in real-world settings. By deploying tools such as lasso regression, ridge regression, and random forest regression, the work proposes an original approach that improves both forecasting reliability and responsiveness under operational conditions. This trailblazing orientation not only remedies the inadequacies present in established practice but also feeds into the broader maturation of data-powered tactics aimed at refining photovoltaic system efficacy.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

# Materials and Methods

## Overview of regression methods

Regression formalisms serve as cornerstones for inferring continuous quantities and teasing out dependencies linking input variables to output targets. Across the scope of this investigation, four distinct regression paradigms were harnessed to project the Maximum Power Point (PMPP) in photovoltaic installations drawing upon environmental descriptors such as irradiance and temperature. The quartet encompasses ridge regression, lasso regression, decision tree regression, and random forest regression, each marked by its own profile of properties and merits.

## Ridge regression (L2 regularization)

Ridge Regression constitutes a sturdy vehicle for faithfully deducing solar cell metrics from field-generated datasets. Its architecture rests on a linear formulation supplemented by an L2 cost term, whose role is to anchor coefficient estimates and diminish susceptibility to overfitting. The central mission of the L2 penalty, frequently labeled the “ridge penalty,” is to draw coefficient magnitudes toward the origin. The intensity of this contraction scales in step with the regularization constant, λ: larger settings of λ enforce more aggressive shrinkage. Beyond trimming model variance, the L2 penalty nudges the representation toward parsimony, yielding a construct that is sturdier and less apt to overfit.

Ridge Regression is singularly adept at reining in model intricacy, thus curtailing overfitting and fortifying the capacity to extend to fresh observations. These traits make it an excellent candidate for anticipating maximum power points (PMPP) in photovoltaic contexts, where the underlying data may embody tangled and heterogeneous linkages connecting climatic drivers (exemplified by temperature and irradiance) to system-level responses. Through coefficient compression, Ridge Regression makes the fitted model less reactive to minor perturbations in the training corpus, culminating in more stable and credible forecasts. Moreover, Ridge Regression copes effectively with multicollinearity, a recurrent phenomenon in photovoltaic datasets wherein covariates such as temperature, irradiance, and module specifications habitually exhibit strong mutual correlation [22-24].

Among the distinguishing virtues of Ridge Regression are its ability to combat overfitting via complexity restraint, its delivery of solid and reproducible parameter inferences, and its deft management of multicollinearity. The L2 cost component also helps the model transfer learned patterns to out-of-sample data, a property of considerable utility when projecting PMPP across heterogeneous weather regimes and hardware layouts. This suite of properties ensures that Ridge Regression furnishes accurate and trustworthy outputs in applied settings, where gauging the response of photovoltaic installations under capricious environmental influences is of paramount importance [24, 25].

Before the commencement of model fitting, all input covariates were subjected to standardization so that each exhibited a mean of zero and a standard deviation of one, thereby guaranteeing that the regularization penalty bore uniformly on every feature. This preparatory measure averts distortion of coefficient estimates by disparities in raw feature scales.

## Lasso regression (L1 regularization)

Lasso Regression operates as a constrained estimation technique that imposes a bound on the summed absolute values of the model coefficients, deliberately inducing sparse solution vectors. This enforced sparsity means a substantial portion of coefficient estimates collapse exactly to zero, thereby executing automated predictor elimination. The L1 penalty term embedded in Lasso regression plays a decisive role in shrinking certain coefficients to zero, a mechanism that autonomously recognizes and jettisons inconsequential inputs. What emerges is a parsimonious and readily intelligible fitted model. Additionally, the L1 penalty helps improve computational performance by zeroing in on the subset of variables that exert the greatest influence on the target estimate—in this context, the maximum power point (PMPP) of photovoltaic arrays [25, 26].

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

The principal benefits of Lasso regression include self-acting variable selection, structural economy, and enhanced clarity. By penalizing absolute coefficient magnitudes, the algorithm inherently surfaces the most salient attributes while sidelining redundant ones, simultaneously acting as a bulwark against overfitting. This reduction in model scope elevates its overall predictive caliber. Furthermore, the parsimony generated by the L1 cost component yields a specification with markedly fewer nonzero parameters, considerably easing interpretation and illuminating how the explanatory variables map onto the response, PMPP. Such lucidity is especially prized in decision-driven settings. Additionally, Lasso regression demonstrates competence in navigating multicollinearity—the scenario where predictor variables share strong linear dependencies. In contradistinction to Ridge regression, which uniformly compresses all coefficient estimates, Lasso regression possesses the capacity to entirely neutralize the contribution of superfluous inputs by forcing their associated weights to zero, thereby reinforcing the model's consistency and insensitivity to perturbations [27, 28].

Consistent with the preprocessing applied to Ridge Regression, all input features underwent standardization to place coefficient magnitudes on a common scale and to enable the L1 penalty to function as intended.

## Decision tree regression

Decision Tree Regression belongs to the family of supervised learning procedures devised for estimating continuous-valued response variables given a collection of explanatory measurements. Its structure takes the form of binary trees, wherein internal nodes embody decision predicates, and branches represent the alternative paths flowing from those predicates. When harnessed to predict the maximum power point (PMPP) in photovoltaic (PV) configurations, decision trees distinguish themselves by representing complex, nonlinear interdependencies that connect ambient factors—most notably irradiance and temperature—with the operational traits of the PV unit. Unlike conventional classification trees (which handle categorical outcomes), Decision Tree Regression is purpose-built for real-valued target quantities, making it a natural methodological choice for projecting PMPP readings [28-30].

Among the principal assets of decision trees is their natural facility for modeling nonlinear input-output mappings, a property of substantial relevance for PV system characterization, given that meteorological conditions and device-specific attributes commonly interact in highly convoluted ways. Beyond this, decision trees afford pronounced transparency; their branching architecture renders explicit the sequence of inferential steps, enabling practitioners to trace precisely how covariates shape the resulting forecast. Such openness is instrumental for informed deliberation and for building confidence in the model's inner workings. Compared to their more opaque machine learning counterparts, decision trees also remain relatively straightforward to deploy and reason about, positioning them as a practical instrument for initial exploratory analyses and for diagnosing fundamental associations among predictors [30, 31].

An additional practical advantage is that decision trees can accommodate incomplete records without recourse to sophisticated gap-filling techniques, a valuable trait in PV system analytics where data acquisition can occasionally produce patchy coverage. During tree construction, the algorithm performs embedded predictor screening, consistently favoring the most discriminative variable at each partitioning step to best separate the observations with respect to PMPP. A notable drawback, however, is their susceptibility to memorizing noise and peculiarities of the training sample—particularly when the dataset is limited or contaminated—potentially eroding their performance on previously unseen cases. As a countermeasure, ensemble frameworks like random forests are widely adopted to reinforce the reliability and steadiness of tree-derived models [31].

## Random forest regression

Random Forest Regression is a powerful ensemble aggregation method that extends the basic decision tree construct. It generates a multitude of decision trees during the fitting phase and, for predictive problems involving continuous targets, outputs the arithmetic mean of the individual tree-level estimates. This strategy proves especially efficacious for anticipating continuous quantities such as the maximum power point (PMPP) in photovoltaic (PV) installations. By

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

capitalizing on the combined judgment of a diverse collection of trees, the technique overcomes many of the shortcomings inherent to single-tree formulations [32, 33].

A defining strength of Random Forest Regression resides in its capacity to counteract overfitting. Because the final estimate is obtained by averaging across an ensemble of trees, each grown on a distinct bootstrap sample of the original dataset, the procedure tempers the variance that might otherwise arise from fitting too closely to incidental regularities in the training corpus. The resulting model is both more resilient and better able to extend its accuracy to new data. Moreover, Random Forest Regression typically delivers appreciably higher precision than any solitary decision tree, particularly when applied to datasets characterized by intricate structure. The ensemble architecture permits the model to capture a richer diversity of latent patterns within the observations—a critical edge in PMPP forecasting, where the interplay between environmental drivers like irradiance and temperature can be highly nuanced.

Random forests also handle high-dimensional feature spaces and elaborate data configurations with considerable dexterity, rendering them exceptionally well-matched to the demands of PV system modeling. Their design confers a degree of immunity to measurement noise and anomalous records, since the ensemble-wide averaging process dilutes the footprint of any single erratic observation. In parallel, the algorithm performs an intrinsic feature-prioritization step. At each node split, a random subset of candidate predictors is evaluated, thereby spotlighting those attributes most germane to PMPP determination. This mechanism simultaneously enhances computational efficiency and sharpens the model's focus on the most impactful inputs. Another pragmatic benefit is the algorithm's innate tolerance for absent data entries, obviating the need for elaborate imputation protocols—a quality that proves highly advantageous in field settings where sensor records may be intermittently unavailable [34-36].

In summary, Random Forest Regression magnifies the inferential power and structural stability of individual decision trees by orchestrating a coalition of multiple base learners, boosting predictive fidelity, curtailing the risk of overfitting, and upholding the dependability of forecasts within the application domain of PMPP estimation in photovoltaic systems.

## Hyperparameter tuning

To improve both predictive accuracy and out-of-sample generalizability, a methodical hyperparameter optimization protocol was implemented. This protocol employed an exhaustive grid search coupled with 10-fold cross-validation to locate the parameter configuration that reduces predictive discrepancy while guarding against data memorization. In the case of the Decision Tree Regression model, the hyperparameters subjected to systematic exploration were the following:

• 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 parameterization that yielded the strongest cross-validated metrics was:

• Max_depth = 10

• Min_samples_split = 2

• Min_samples_leaf = 1

For the Random Forest Regression model, the search space encompassed the following hyperparameters:

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

• 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 configuration that achieved the best overall performance was:

• n_estimators = 200

• Max_depth = 20

• Min_samples_split = 2

• Min_samples_leaf = 1

These hyperparameter values were chosen based on cross-validated Mean Squared Error (MSE). To further substantiate the integrity of the Random Forest model—especially in view of its perfect coefficient of determination ($R^2 = 1.0$)—both learning curve trajectories and cross-validation diagnostics were examined. The learning curves revealed a progressively shrinking divergence between the training and validation performance measures, a pattern indicative of sound generalization behavior rather than an overfitted solution.

The tuned configurations described above were adopted uniformly for all subsequent model performance comparisons and for the feature importance assessments reported.

# Description of the Dataset and RELATIONSHIP ANALYSIS of PMPP, TMOD, and GPOA

The body of measurements harnessed in the present inquiry was collected at the photovoltaic experimental rig operated by the SIRTA (Site Instrumental de Recherche par Télédétection Atmosphérique) observatory [1, 2], hosted in Palaiseau, France (48.7 N, 2.2E) within the Ecole Polytechnique campus grounds. Brought into service during 2014, this testing setup houses a set of five commercially marketed solar modules spanning different technological families. The panels sit in an unenclosed mount, oriented southward at a 27° incline, and the logged parameters cover three central quantities: incoming solar radiation, thermal readings, and the resultant maximum power point (PMPP). Global Plane of Array irradiance (GPOA), measured in watts per square meter ($W/m^2$), captures the flux density of solar energy impinging upon a collector surface and operates as the primary driver of electrical generation in photovoltaic (PV) devices. The temperature variable, whether representing ambient air or the module substrate itself, exerts a commanding influence over PV operational behavior, given that elevated thermal states characteristically depress cell-level voltage and thereby shift PMPP. PMPP itself, denoted in watts (W), signifies the peak electrical output attainable from a PV unit under any particular combination of environmental circumstances, and it stands as a fundamental barometer of system capability and conversion effectiveness. This dataset provides an extensive empirical basis for analyzing how irradiance and temperature jointly modulate PMPP, as well as for developing forecasting frameworks. Leveraging these empirical records, the investigation compares conventional estimation paradigms with algorithmically driven approaches, aiming to show how the latter excel at representing nonlinear functional forms, responding nimbly to transient environmental shifts, and delivering higher prognostic precision. **Table 1** provides a condensed synopsis of the dataset.

**Table 1**. Data overview.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

| DATA | Max | Med | Standard Deviation | Mean | Count |
|---|---|---|---|---|---|
| **GPOA (W/m2)** | 1106.85 | 502.02 | 288.61 | 511.18 | 36,908 |
| **TMOD ( °C)** | 48.87 | 22.18 | 10.26 | 23.06 | 36,908 |

Delving deeply into this measurement archive is highly significant for developing high-fidelity emulators of photovoltaic cell response. By capitalizing on the dependencies that tether irradiance levels, temperature readings, and maximum power point (PMPP) observations together, the ambition is to engineer resilient models capable of projecting PMPP values across widely varying atmospheric regimes. Acquiring such a predictive capability is vital for fine-tuning PV installation operation, amplifying energy capture, and seamlessly assimilating solar-generated electricity into broader power networks. The latent value embedded in this dataset resides in its potential to propel the transition toward a more productive and durable energy paradigm. Refined PMPP estimation allows system architects to sharpen design parameters, empowers grid operators to devise superior energy dispatch strategies, and catalyzes the pervasive rollout of renewable generation assets.

Examination of **Figure 2** discloses that the correlation index linking PMPP with TMOD climbs to an exceedingly high level, hinting at a near-ideal positive rectilinear coupling between these two quantities. The association observed between TMOD and GPOA likewise registers as moderately elevated, with a coefficient of 0.84—confirming a positive rectilinear tie, albeit one characterized by a looser point distribution. These patterns throw into sharp relief the intimate mutual bonds woven among the triad of variables. The intervariable dynamics give the impression of unfolding along predominantly straight-line progressions, notwithstanding the likelihood that supplementary influences may impose a more layered modulation on the ultimate PMPP magnitude.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

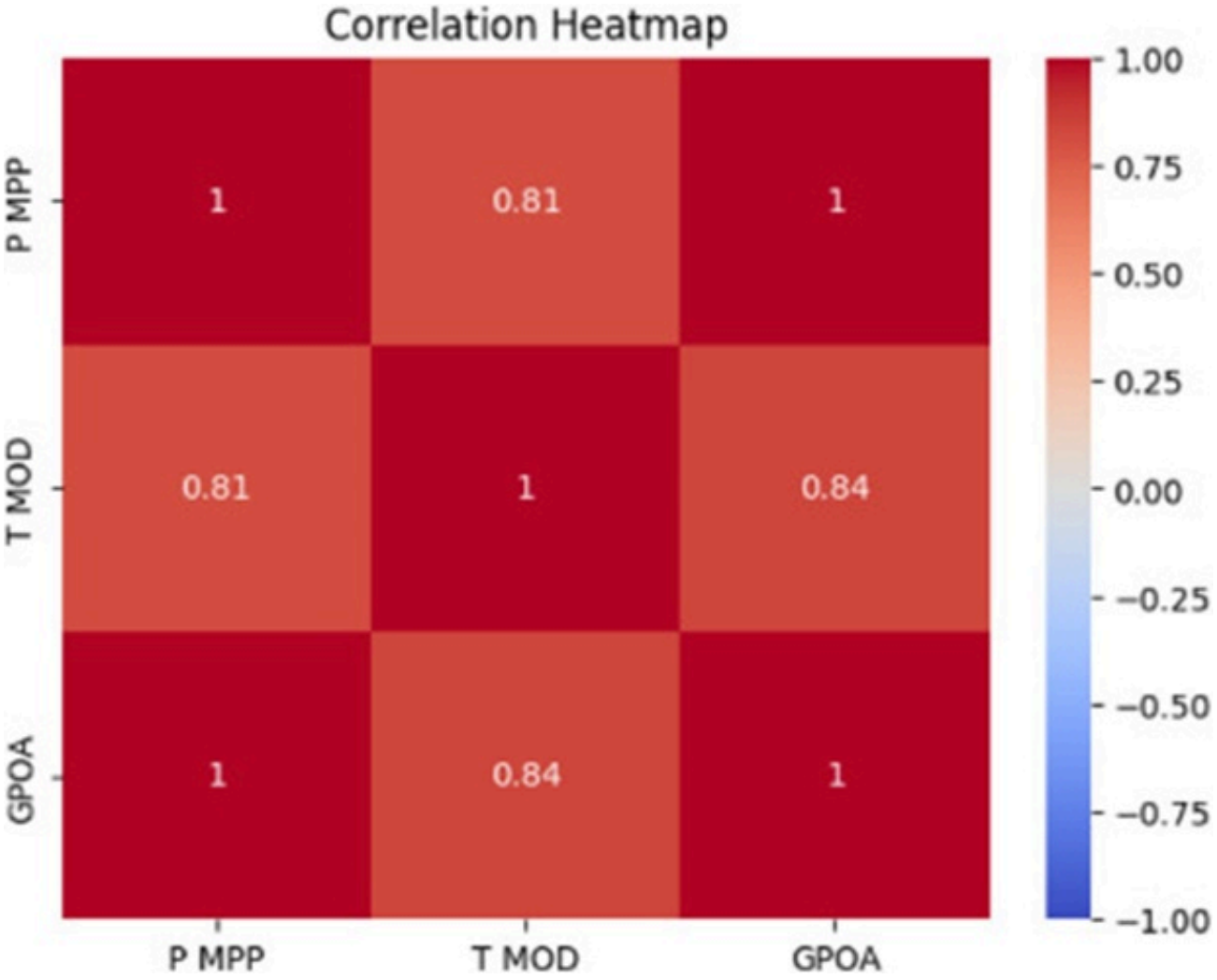



**Figure 2.** Correlation heatmap.

The set of numeric correlation values shown in **Figure 3** is visually corroborated by the paired scatter diagrams in **Figure 4**, which reveal upward-sloping rectilinear affinities among the variables. The PMPP-to-TMOD mapping exhibits a rising trajectory, mirroring the behavior of the PMPP-to-GPOA mapping, both of which suggest that corresponding increases in PMPP match increases in TMOD. Even so, the dispersal of individual observations around the superimposed trend paths betrays the operation of additional contributory factors that inject meaningful variability into PMPP, thereby preventing the linkage from conforming to a purely rectilinear ideal. The scatter representation in **Figure 5** conveys how thermal and radiative covariates simultaneously map onto the Maximum Power Point (PMPP) space.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

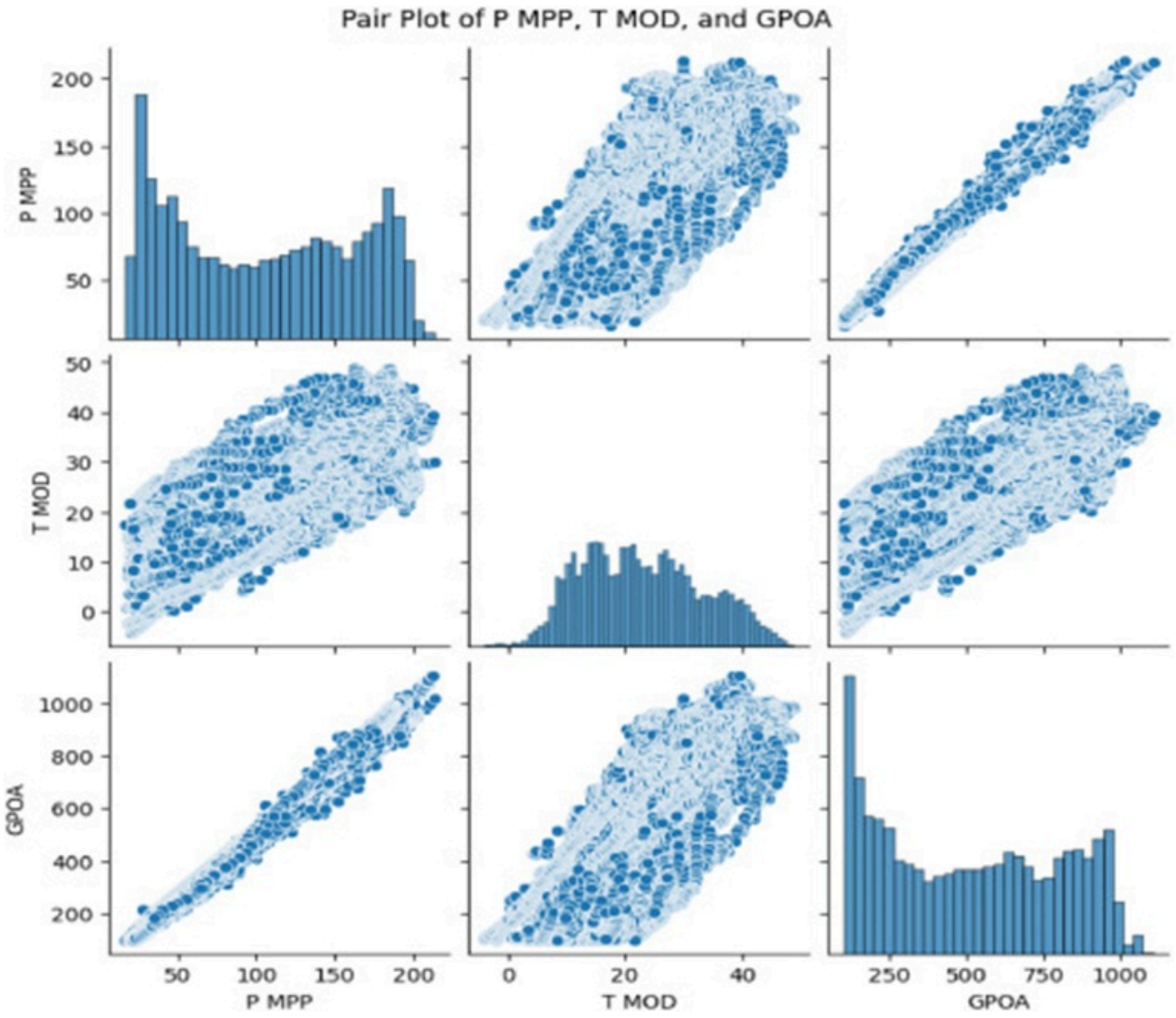



**Figure 3**. Correlation coefficients.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

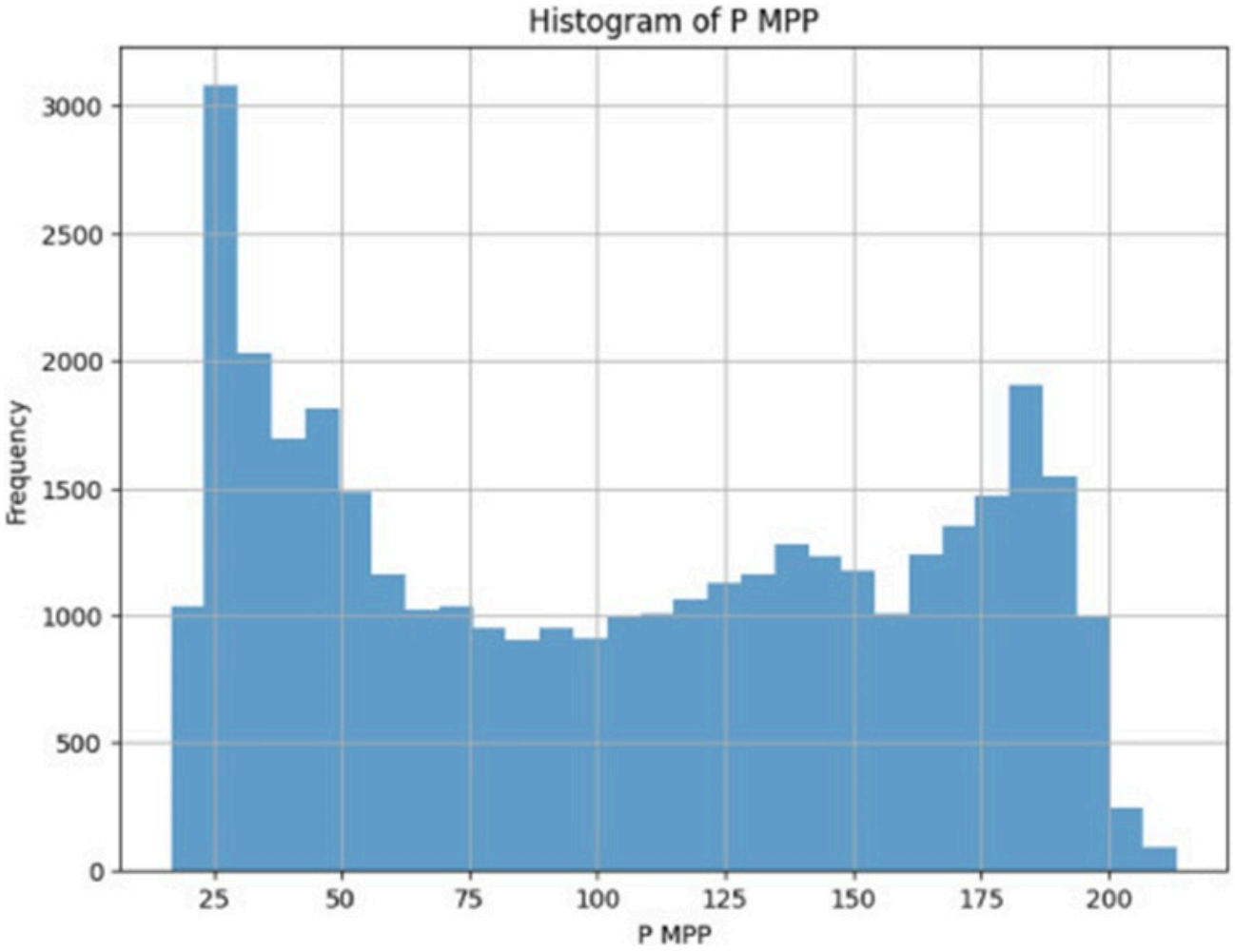



**Figure 4.** Scatter plots.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

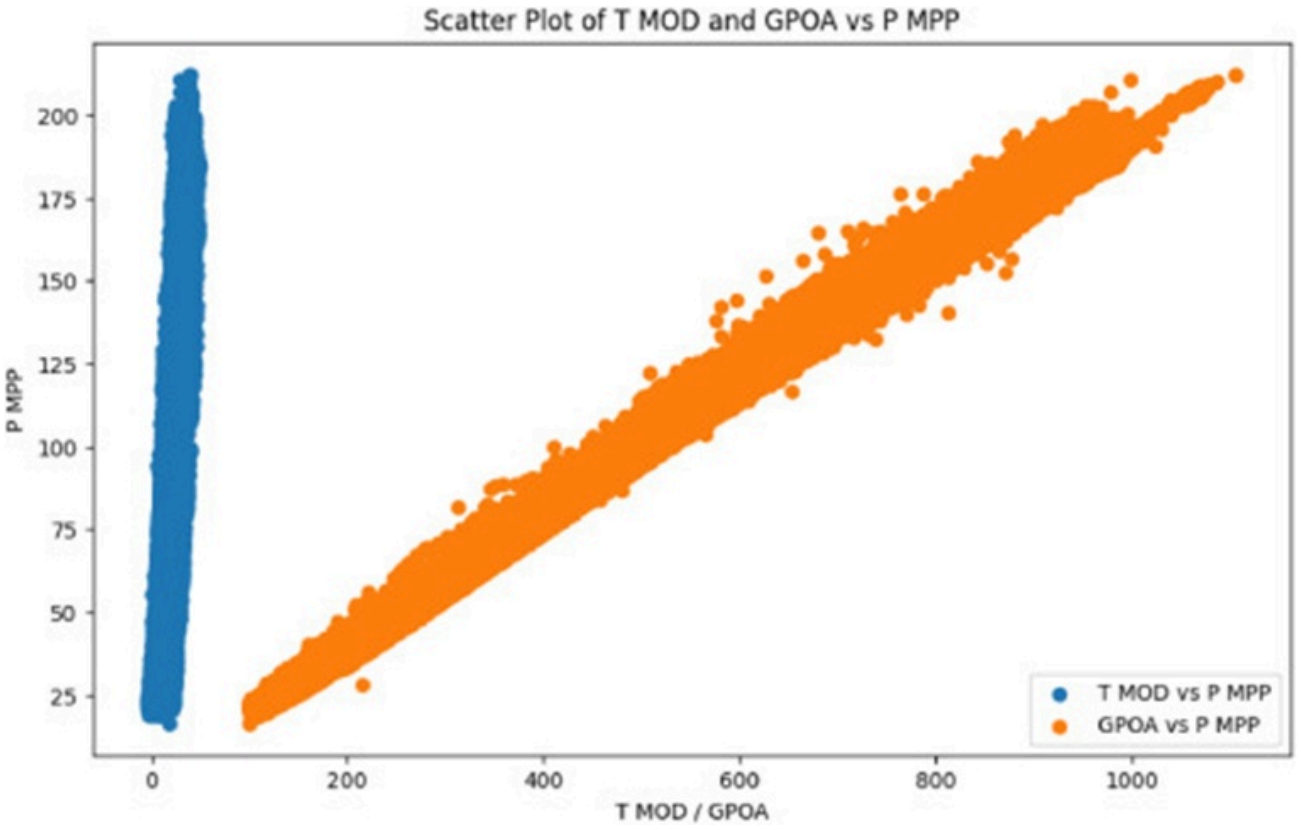



**Figure 5.** Scatter plot of TMOD and GPOA vs PMPP.

TMOD vs PMPP: renders the dependency path running from module temperature (TMOD) toward the Maximum Power Point (PMPP). The cerulean markers crowd into the leftmost sector of the graphic field, a configuration that communicates the fact that TMOD covers a considerably narrower numeric range relative to GPOA. The overwhelming share of TMOD readings congregate near the zero mark, a pattern that conveys the impression of only modest excursions in module thermal state. The compressed vertical extent of the TMOD-PMPP marker cloud feeds the interpretation that, confined to this specific observational window, TMOD exerts only a faint pull on PMPP.

GPOA vs PMPP: charts the pathway connecting Global Plane of Array irradiance (GPOA) to the Maximum Power Point (PMPP). The amber markers traverse an expansive interval of GPOA intensities, ranging from the origin out beyond 1000 W/m². An unmistakable affirmative trend crystallizes when examining GPOA alongside PMPP: every increment in GPOA is shadowed by a commensurate rise in PMPP. This unequivocal covariational signature indicates that global plane-of-array irradiance is a preeminent determinant of the maximum power point achieved by photovoltaic modules.

Drawing meaning from the measurement collection brings into focus divergent dynamical motifs characterizing the respective linkages of module temperature (TMOD) and global plane-of-array irradiance (GPOA) with the maximum power point (PMPP). In the case of TMOD, the tight vertical envelope and the leftward massing of cerulean markers jointly imply that, within the envelope of conditions captured, oscillations in module thermal condition do not materially perturb PMPP—a scenario consistent with either naturally muted temperature variance or active thermal regulation. Shifting attention to GPOA, the interrogation lays bare a strong positive linear relationship with PMPP, a conclusion underwritten by the amber markers that ascend methodically in lockstep with swelling GPOA magnitudes. This finding highlights the decisive role of incident radiation in the electricity generation sequence, a result wholly in keeping with the physical axiom that more abundant solar illumination begets proportionally larger electrical yield from photovoltaic collectors.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

# Learning Curves Analysis and Evaluation of Prediction Accuracy

## Learning curves analysis

The decision tree specification yields a training performance metric that remains anchored at a lofty and unvarying ceiling (1000), as shown in **Figure 6**, indicating an exhaustive accommodation to the estimation sample—behavior emblematic of algorithmic constructs with ample flexibility to internalize convoluted data topographies. Meanwhile, the validation trace, while consistently occupying a plane beneath the training trace, displays a gentle upward inclination as the pool of observations expands, indicating satisfactory generalization capability.

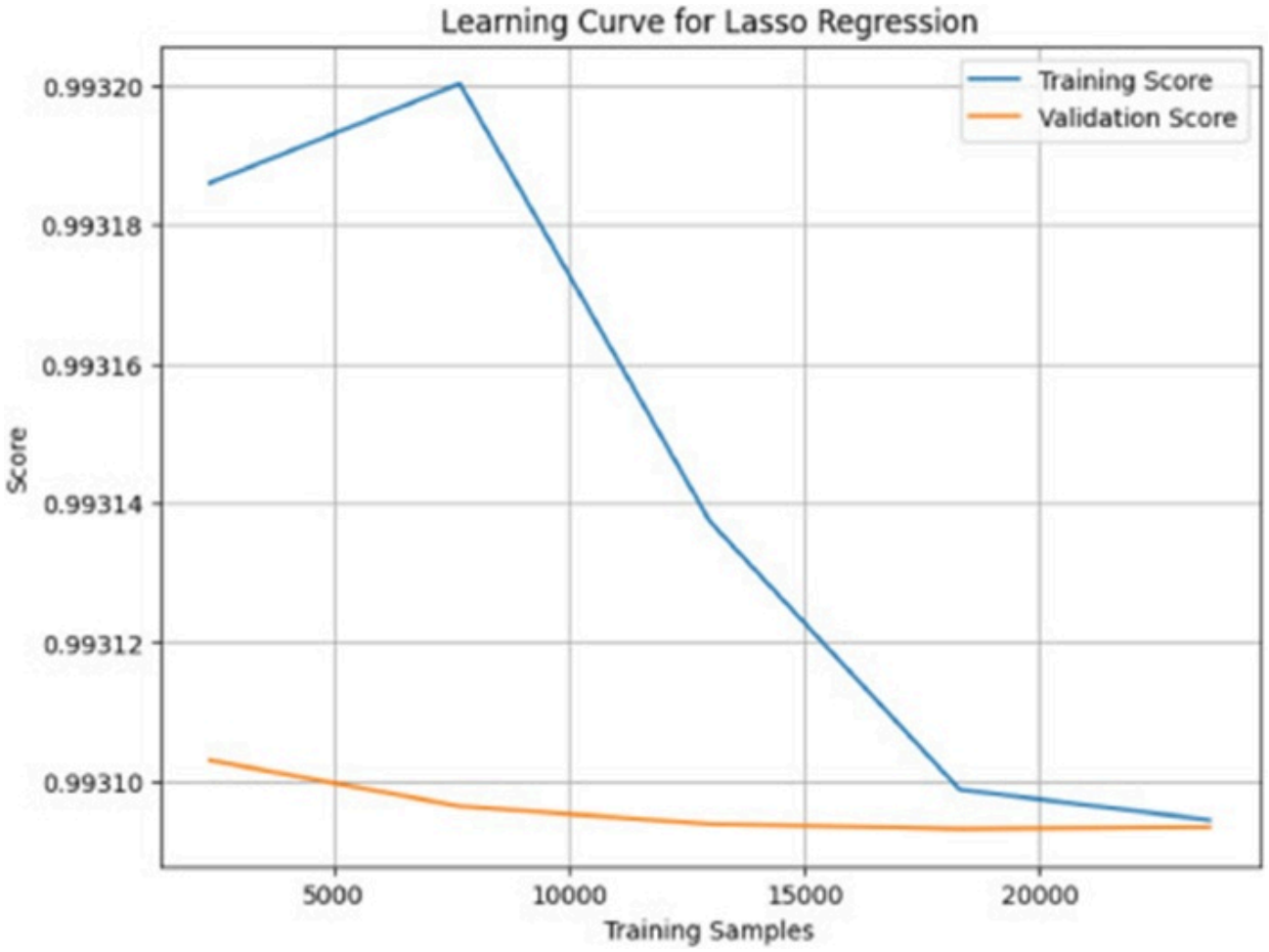



**Figure 6**. Decision tree learning curve.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

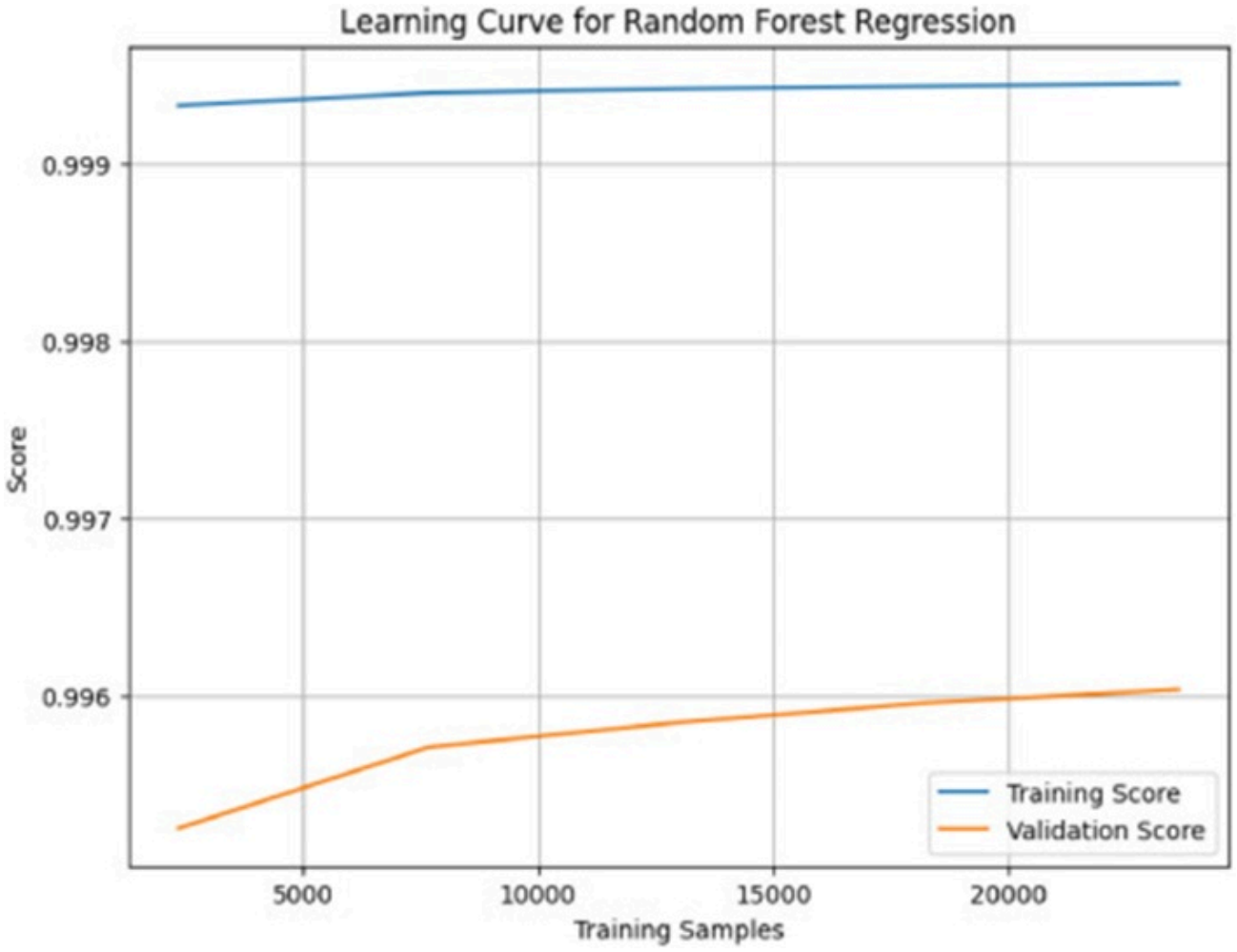



**Figure 7.** Lasso regression.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

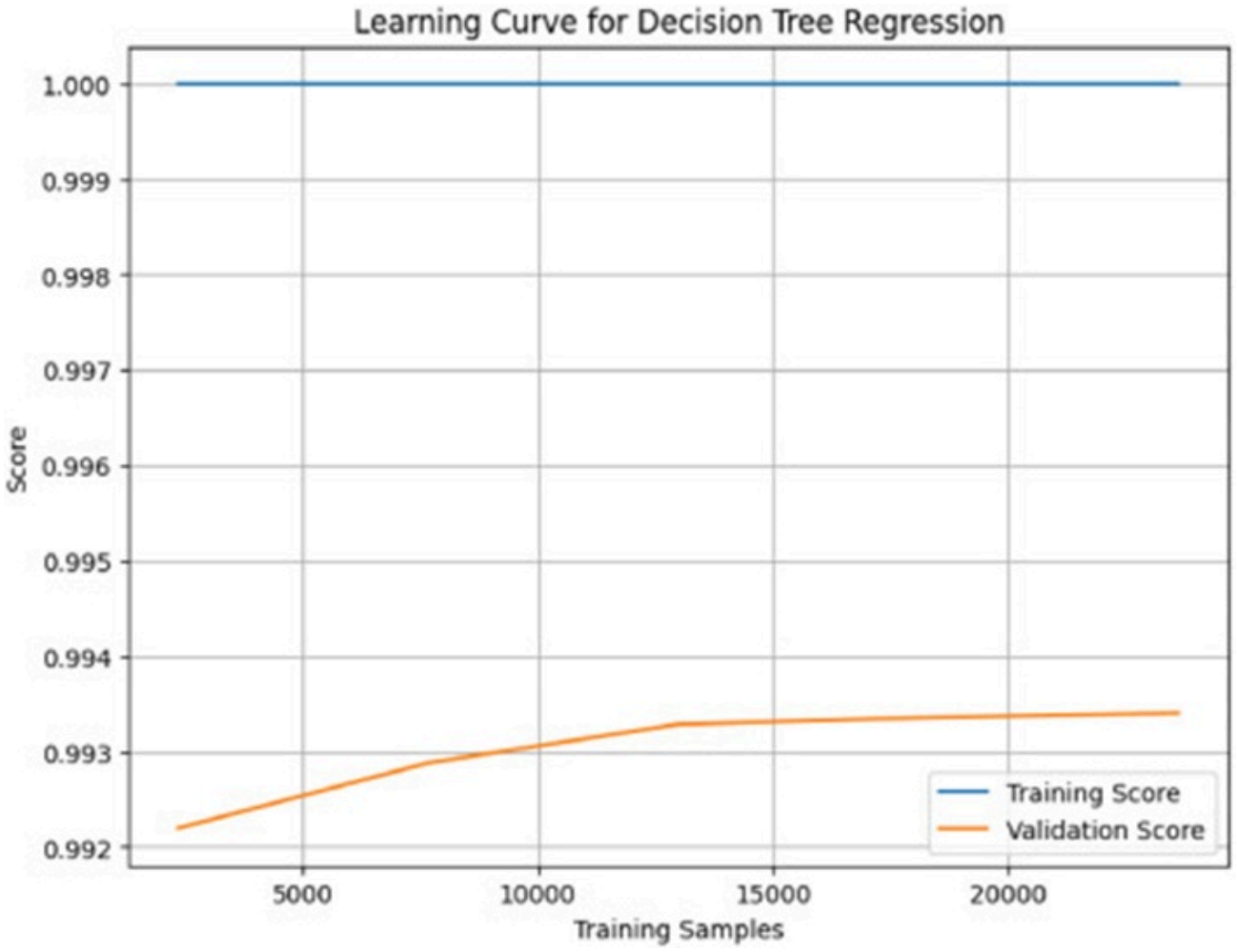



**Figure 8.** Ridge regression learning curve.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

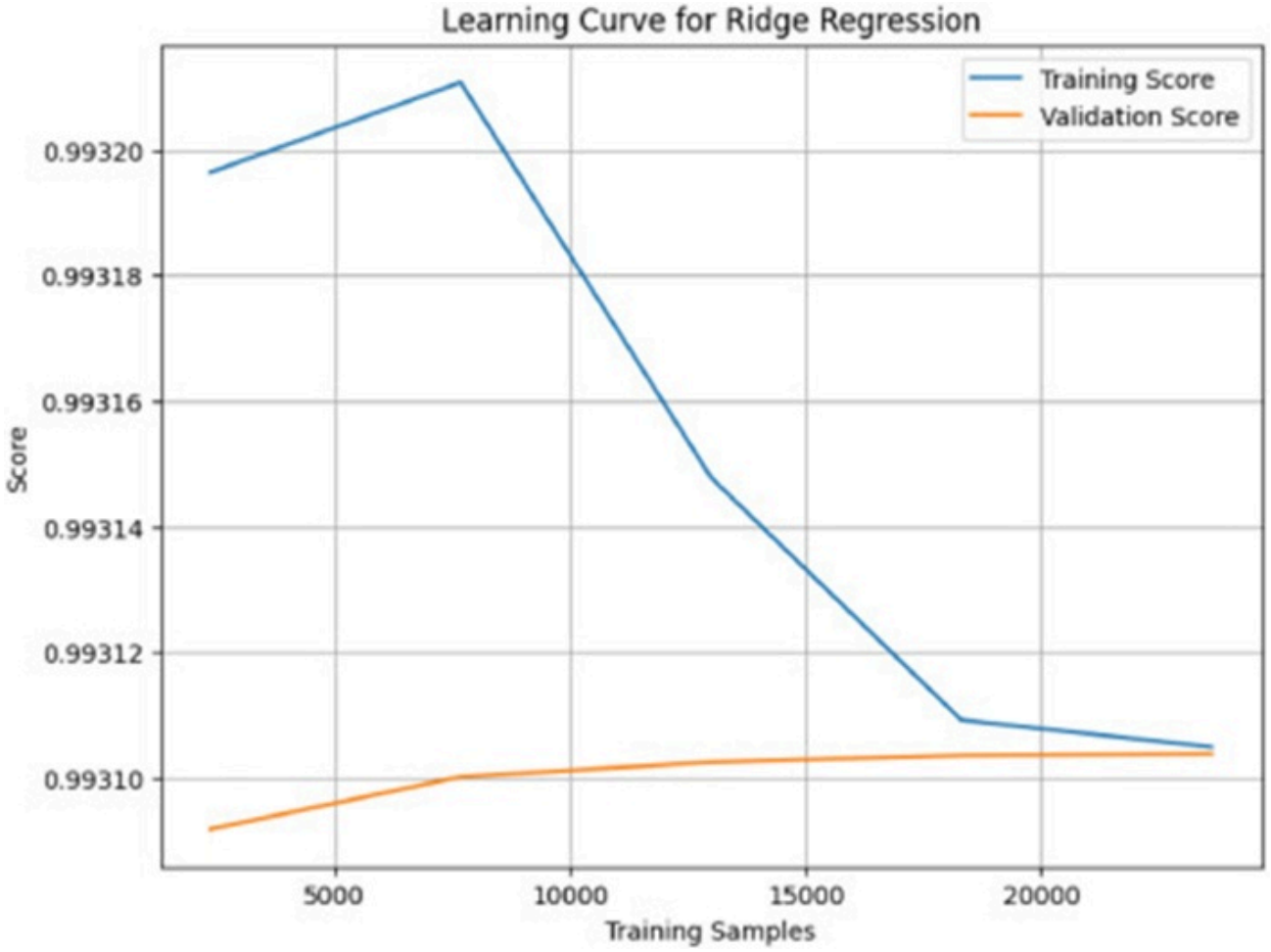



**Figure 9.** Random forest regression learning curve.

The Lasso regression construct exhibits a training score trajectory that eases downward in a measured fashion as supplementary data are absorbed, mirroring an inbuilt self-regulating dynamic that operates to forestall excessive tailoring to the training corpus. Along complementary lines, **Figure 7** discloses that the validation score likewise undergoes a slight retraction while nonetheless preserving a serviceable performance tier, a pattern redolent of persistent fine-tuning and incremental refinement as fresh training exemplars are supplied.

**Figure 8** shows that Ridge regression starts from an initially inflated training score that progressively erodes as data volume accumulates, a direct outgrowth of its penalization mechanism that militates against coefficient estimates of large absolute size. The validation score, though occupying a lower register, climbs steadily as the sample size swells, signaling an ongoing improvement in generalization prowess and a gradual contraction of the divergence separating the estimation and holdout performance metrics.

**Figure 9** reveals that Random Forest regression secures a training score brushing the theoretical maximum. This circumstance testifies to a formidable capacity for absorbing the estimation dataset while simultaneously sounding a cautionary note regarding possible overfitting. With the successive addition of more extensive training corpora, the validation trajectory climbs higher. It migrates closer to the training trajectory, a behavioral signature that augurs healthier generalization and leaves open the possibility of further performance dividends were the tally of available samples augmented still further.

## Evaluation of prediction accuracy

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Examining **Figure 10**, one finds that the cerulean data markers exhibit near-total coincidence with the unity-slope reference. Even so, a delicate pattern of structured undervaluation appears to set in toward the upper end of the PMPP continuum, where markers begin to peel away ever so slightly from the crimson diagonal. Given its formulation as a coefficient-constrained linear estimator, Ridge regression may lack the expressive range needed to faithfully disentangle the deeper nonlinear motifs present in the measurement record—territory where tree-based learners and ensemble aggregators hold a distinct edge.

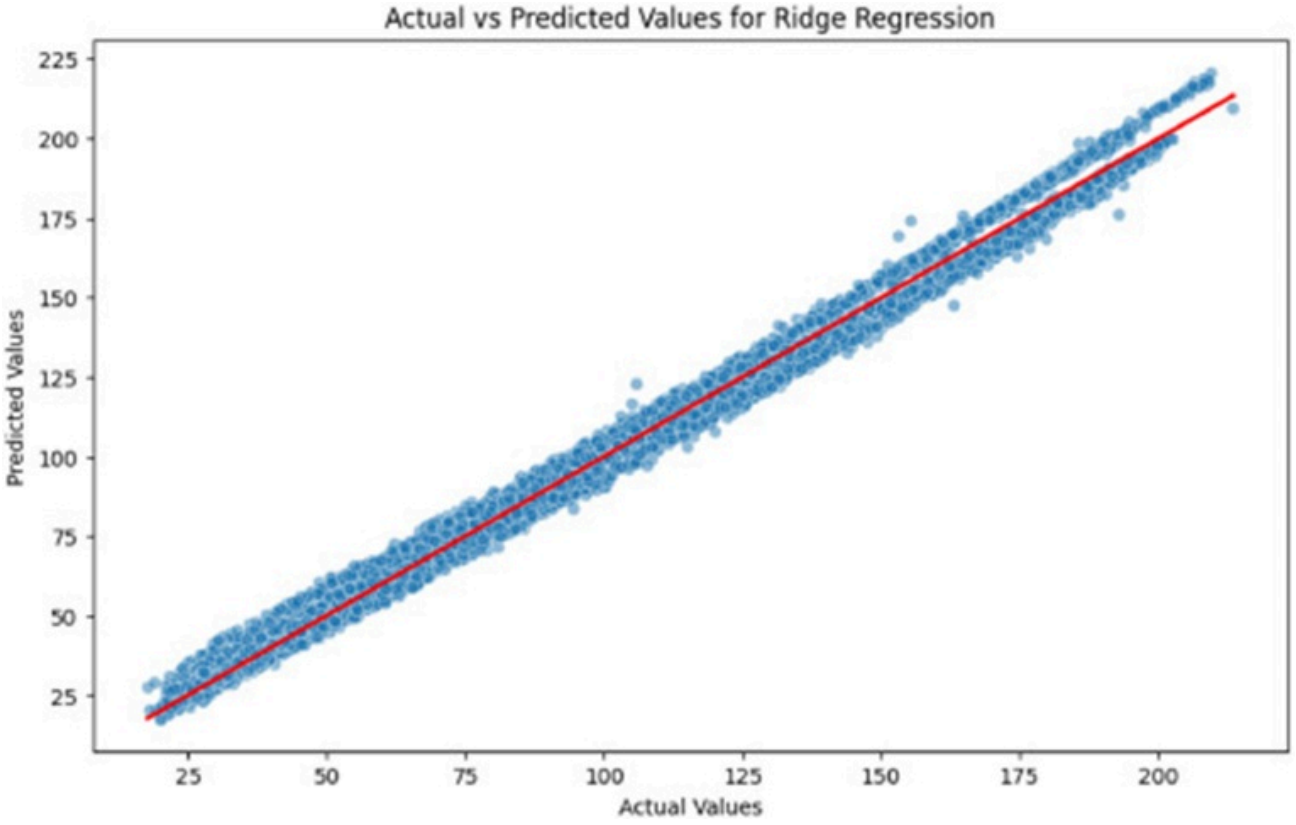



**Figure 10.** Ridge regression.

Shifting to **Figure 11**, the overwhelming share of plotted markers nestles within a narrow envelope surrounding the crimson trace. This arrangement broadcasts the Lasso regression specification's competence at generating usable PMPP projections from thermal and radiative predictors. Several readings land at essentially identical coordinates. Anomalous cases do present themselves, most visibly among the uppermost value strata. Encircling the reference diagonal, a modest handful of markers drift outward, though the overall diffusion stays tightly contained—a sign that the predictive distribution remains largely well-anchored with a restrained spread.

With respect to **Figure 12**, the cerulean markers achieve near-seamless congruence with the crimson trace. The swarm of points occupies a visibly more condensed footprint than its decision tree equivalent, underscoring the heightened prognostic power of the random forest ensemble. This configuration distinguishes itself through a pronounced knack for tracking undulations in PMPP readings, a facility born of its capacity to encode intricate, multiply-branching variable linkages.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

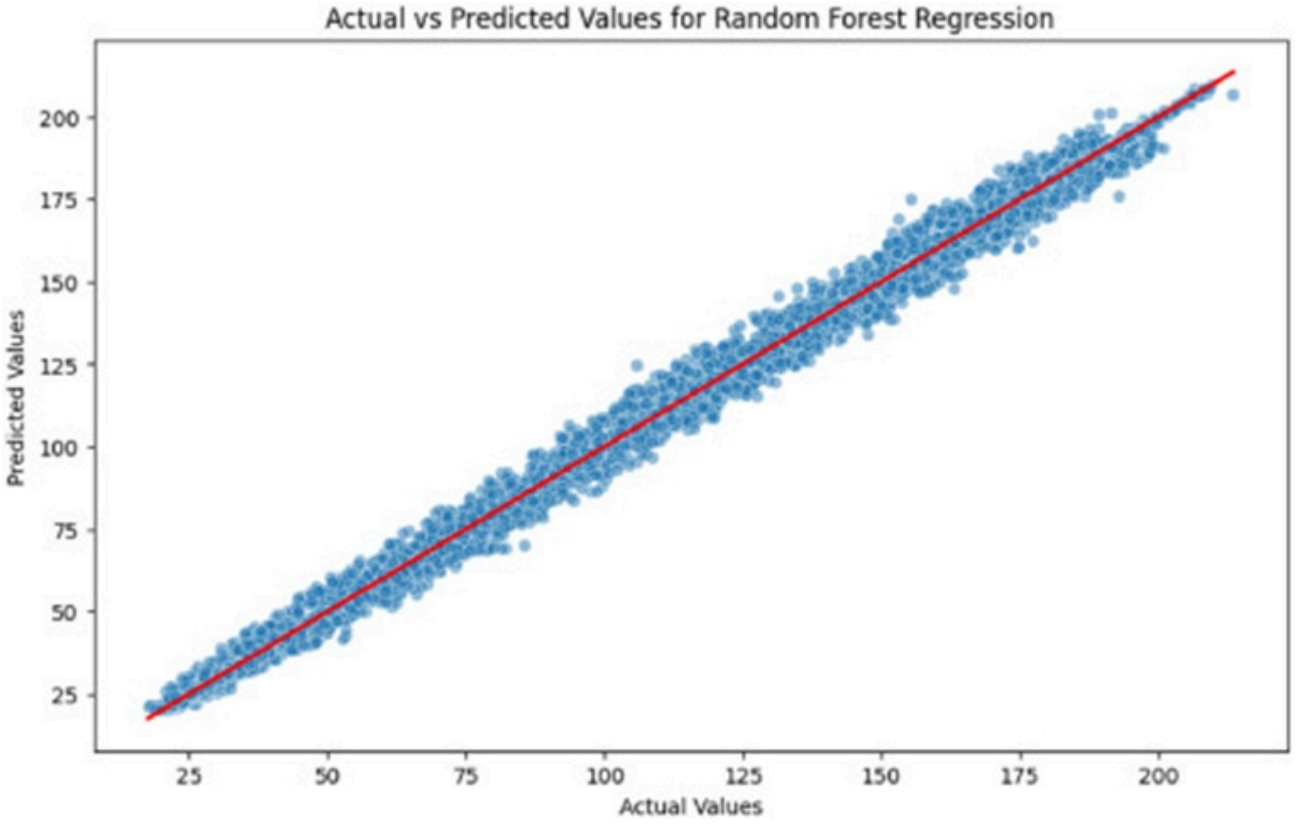
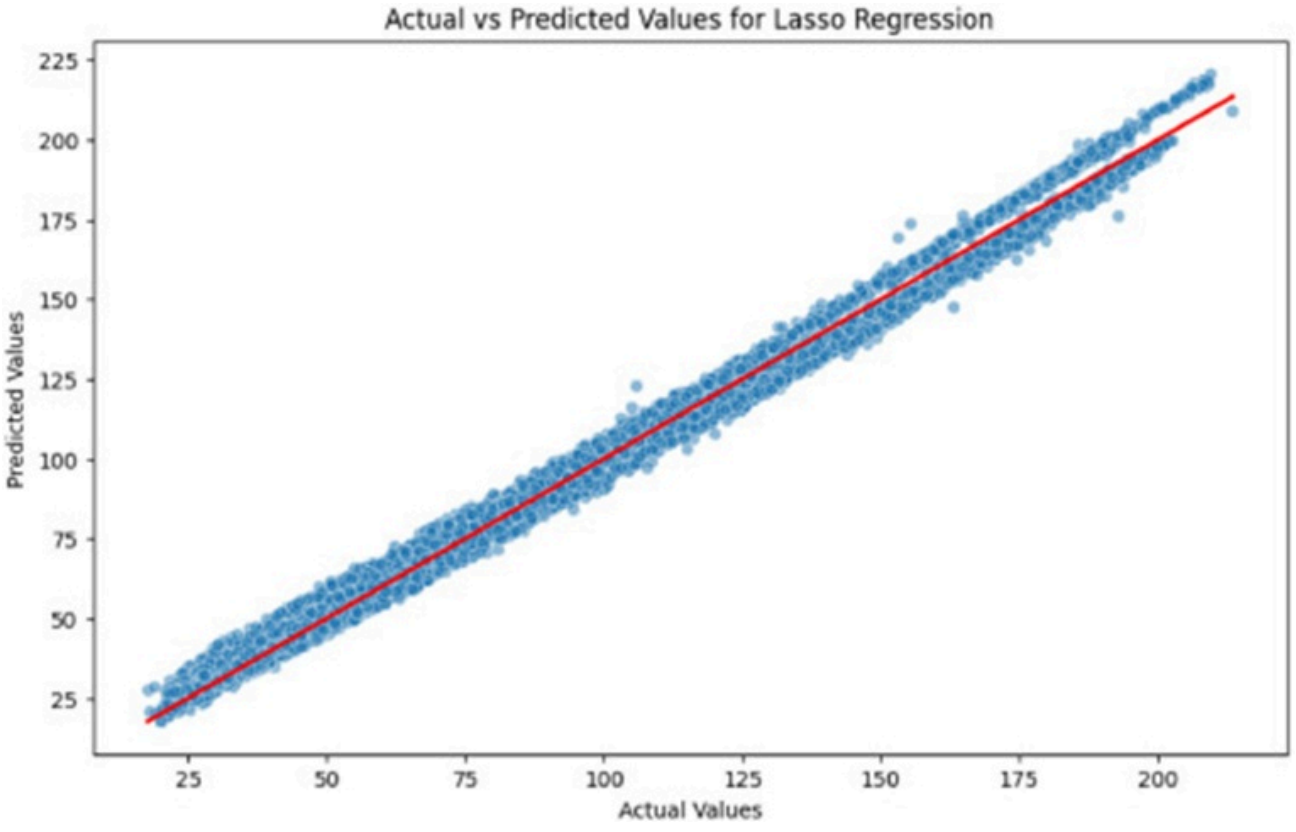



**Figure 11.** Lasso regression.




© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

**Figure 12.** Random forest regression.

As rendered in **Figure 13**, the observations dispose themselves in disciplined alignment with the crimson trace, a geometry that testifies to considerable model proficiency. That said, a subtle loosening of the point concentration becomes discernible, particularly within the domain of larger PMPP figures, lending weight to the suspicion that the model may occasionally lean toward conservative forecasts under particular ambient regimes.

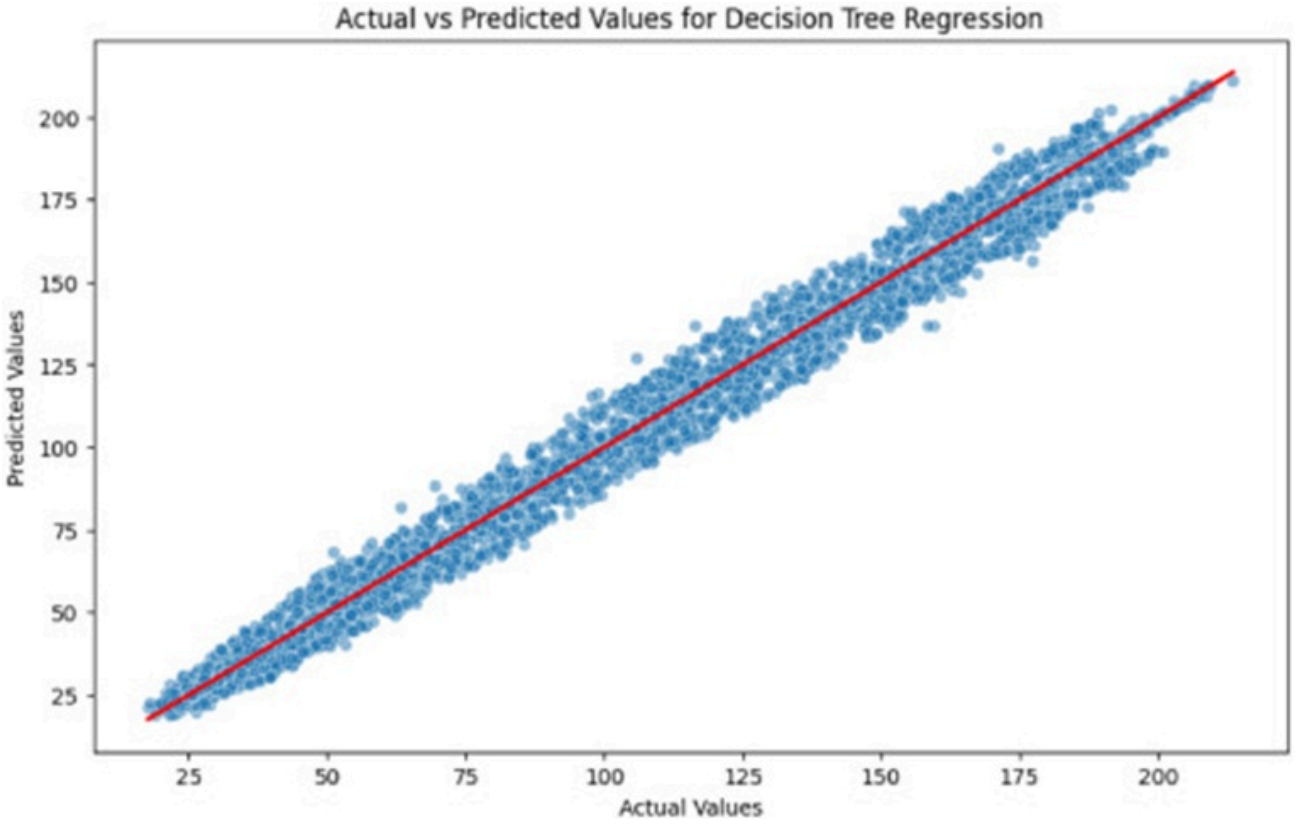



**Figure 13.** Decision tree regression.

# Results and Discussion

## Feature importance analysis

This segment dissects the relative salience of input variables as revealed by the suite of regression frameworks deployed—Ridge, Lasso, Decision Tree, and Random Forest. The accompanying visualizations map out the comparative influence exerted by each predictor in the endeavor to anticipate maximum power output (PMPP) from module temperature (TMOD) and irradiance (GPOA).

**Figures 14 and 15** make plain that Ridge and Lasso regression each confer outsized weight to module temperature (TMOD) while relegating irradiance (GPOA) to a distinctly secondary plane. This outcome suggests that, across both paradigms, TMOD shoulders the greater share of predictive responsibility for PMPP relative to GPOA. The introduction of the Ridge penalty compresses coefficient estimates globally, yet TMOD maintains its commanding stature throughout. Similarly, Lasso regularization inherently gravitates toward predictor triage, forcing the weights attached to less salient inputs to zero—with TMOD persisting as the overwhelmingly preeminent driver of PMPP.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

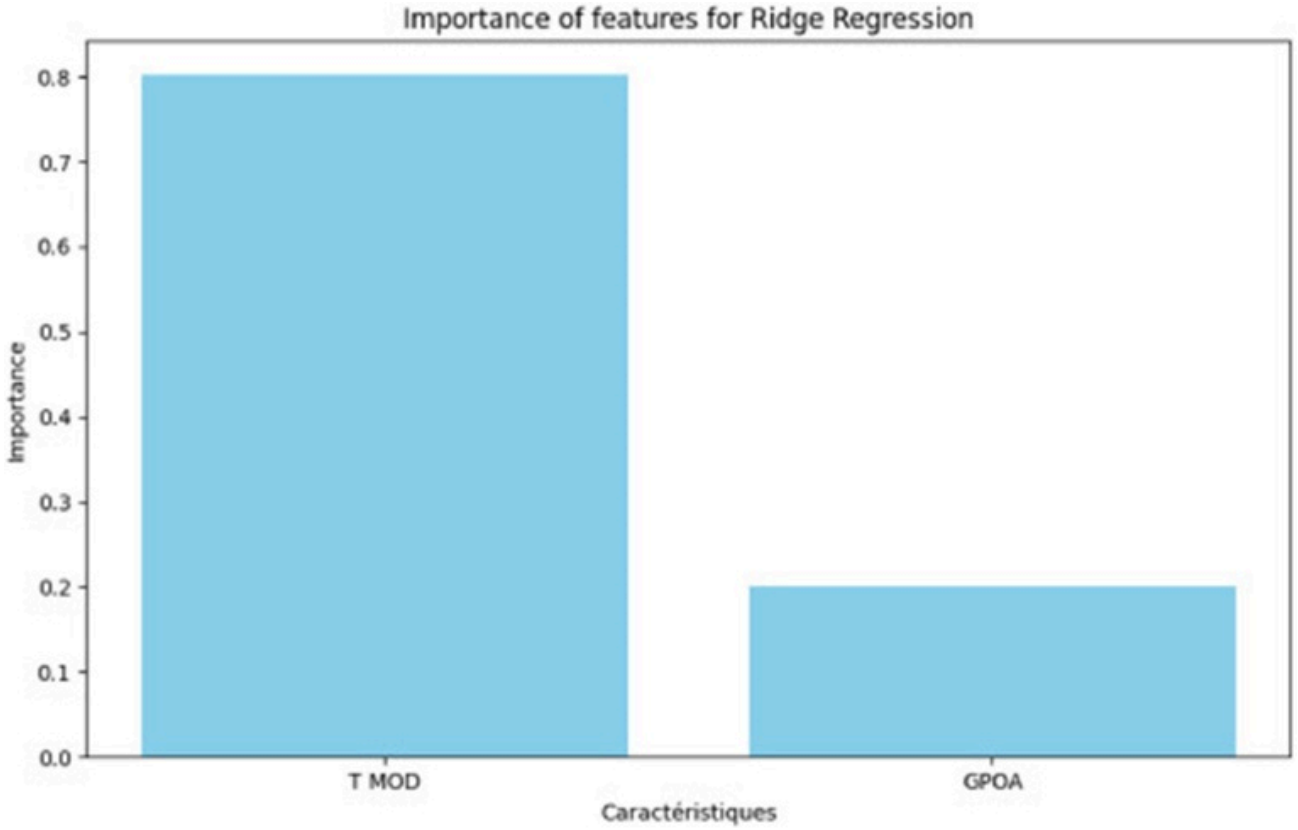
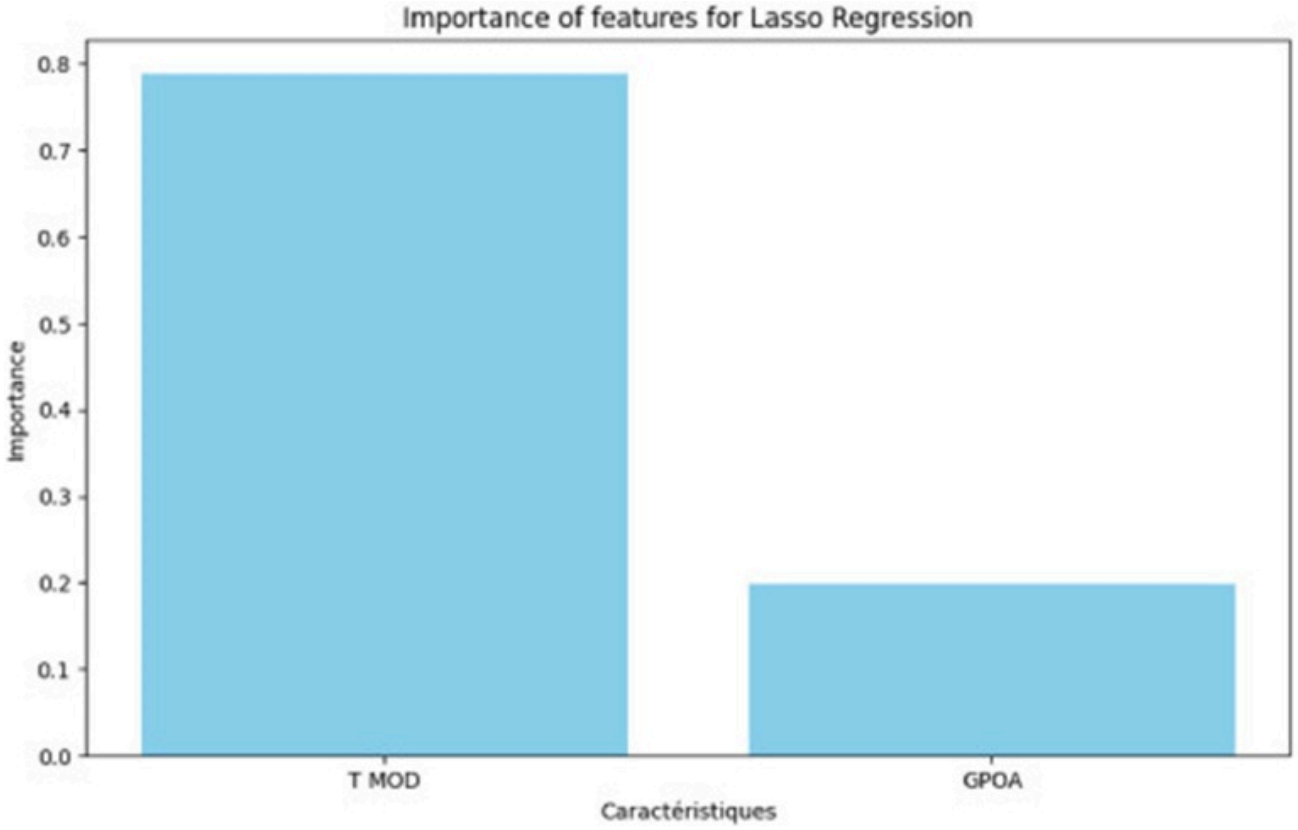



**Figure 14.** importance of features for Ridge Regression.




© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

**Figure 15.** Importance of features for Lasso Regression.

**Figures 16 and 17** show that Decision Tree and Random Forest regression each assign irradiance (GPOA) dramatically greater importance when juxtaposed against module temperature (TMOD). Inside the Decision Tree construct, GPOA monopolizes the lion's share of attribute relevance, a revelation that swings in GPOA overwhelmingly govern the branching choices encoded throughout the tree hierarchy. The Decision Tree algorithm apportions substantial credit to the most discriminative splitting variables; here, GPOA steps forward as the decisive operative factor. In lockstep, Random Forest regression, which fuses the outputs of a large collection of constituent trees, reaffirms the centrality of GPOA as the foremost determinant sculpting PMPP behavior. Variation across the individual tree rankings points toward a broadly shared consensus in assigning primacy to GPOA.

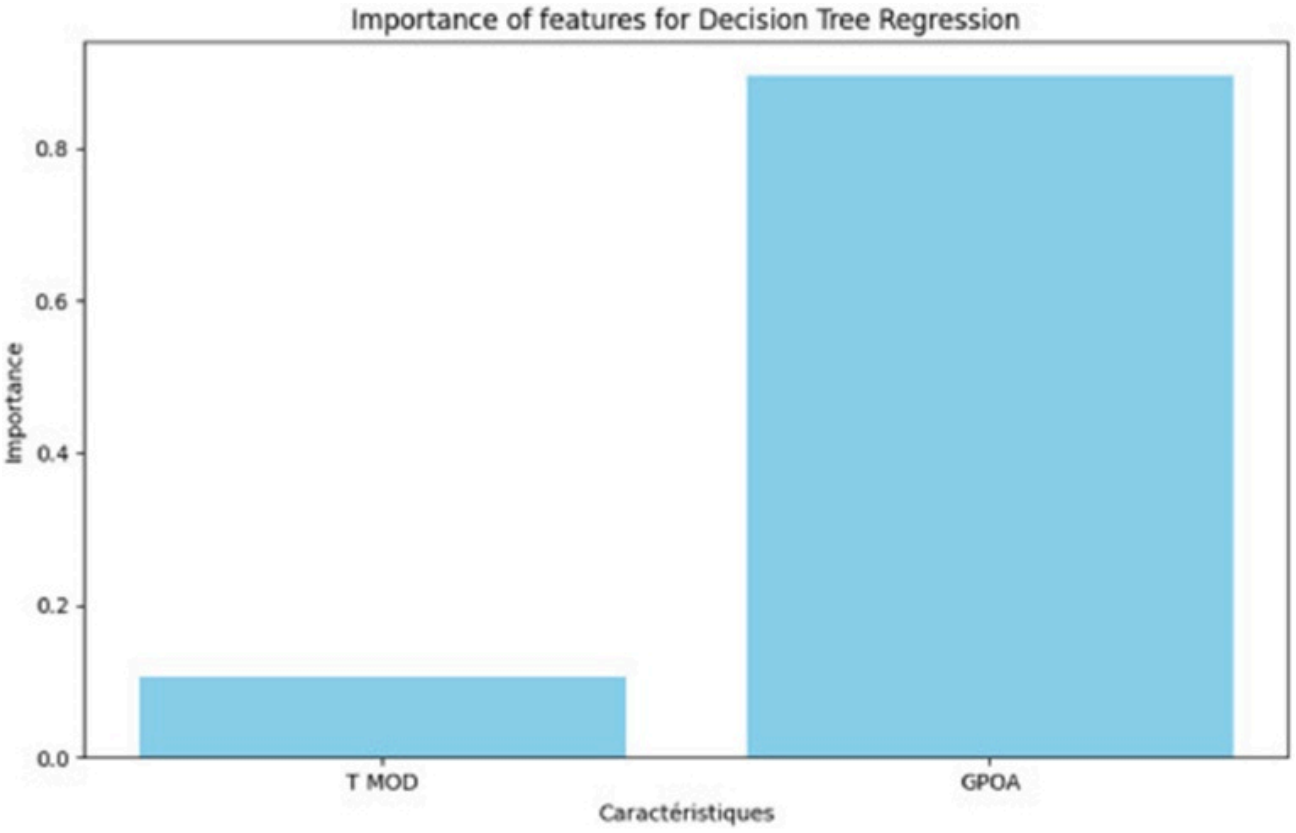



**Figure 16**. Importance of features for Decision Tree regression.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

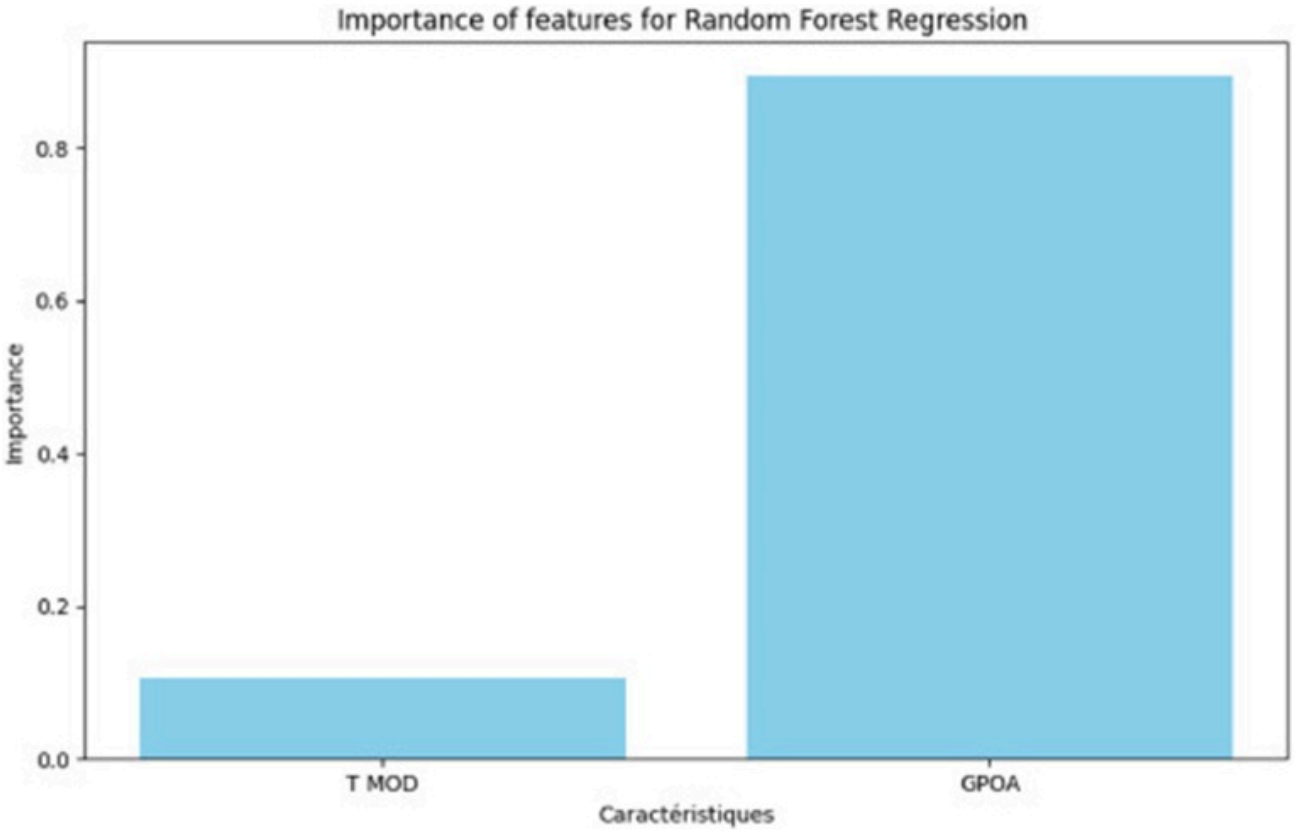



**Figure 17.** Importance of features for Random Forest regression.

It should be noted that the feature importance appraisal undertaken for the Ridge and Lasso regression frameworks was performed only after predictor standardization had been performed. This precaution ensures that the coefficient magnitudes presented are mutually comparable and provide an honest portrayal of the true relative impact of each covariate.

Placing the regression methodologies side by side unmasks sizable gulfs in predictive quality, as quantified through the Mean Squared Error (MSE) yardstick. The visualization in **Figure 18** shows that Lasso regression bears the heaviest MSE penalty among the full array of techniques scrutinized, peaking at 22.82, with Ridge regression practically mirroring it at an MSE of 22.76. Both of these approaches turn in broadly analogous—though perceptibly weaker—results when measured against the tree-powered alternatives: Decision Tree logs an MSE value of 21.59, whereas Random Forest sweeps in with the most competitive MSE of 12.8. This statistic graphically dramatizes its runaway superiority.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

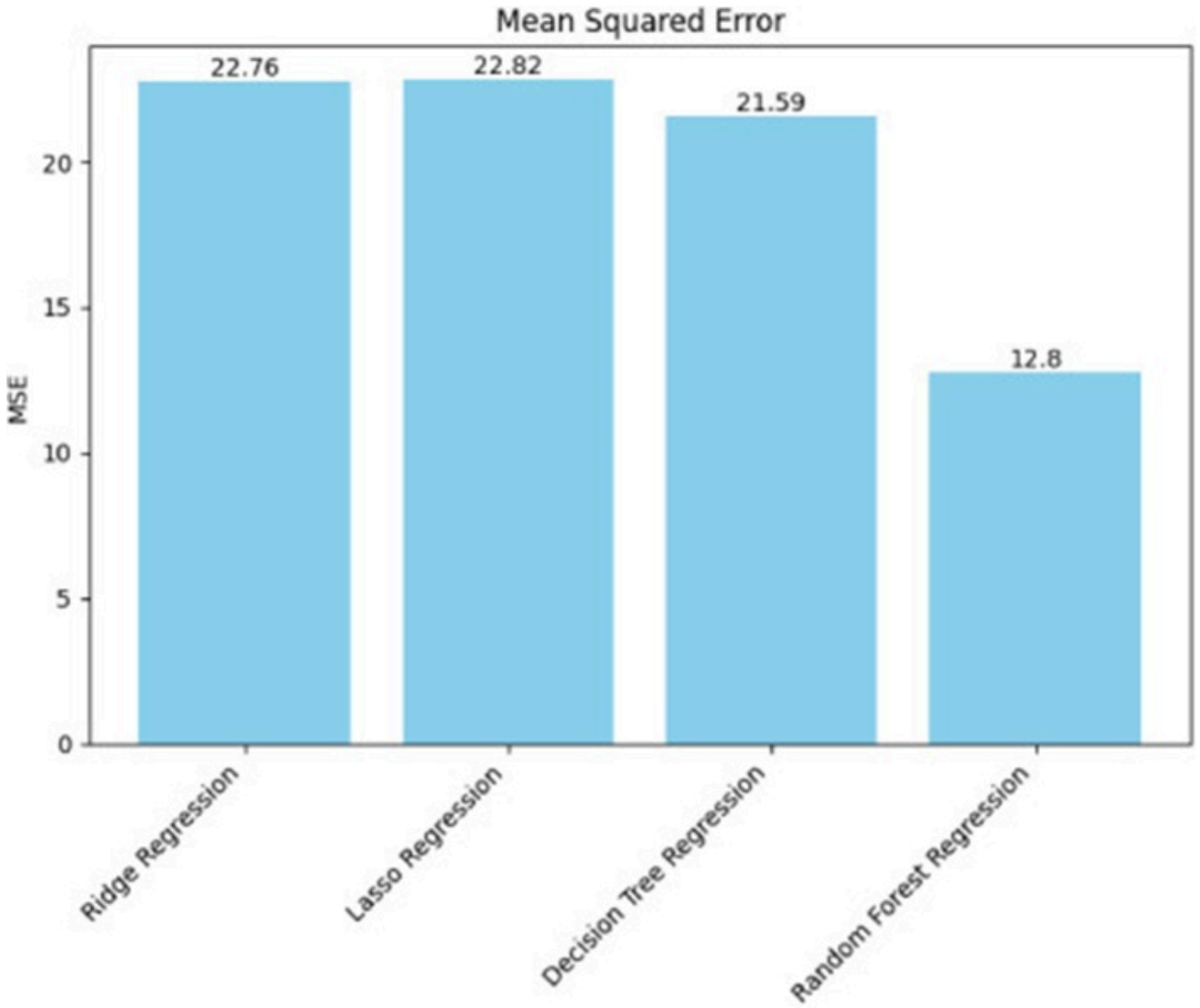



**Figure 18**. MSE result.

Strikingly, Ridge Regression and Lasso Regression each muster a coefficient of determination $R^2$ pegged at 0.99, with Decision Tree likewise reaching the 0.99 threshold. At the same time, Random Forest climbs to a pristine $R^2$ of 1, as documented in **Figure 19**. This juxtaposition brings into stark relief a consequential performance chasm wherein aggregated tree-driven constructs outclass their coefficient-shrunken linear cousins (Ridge and Lasso) in explanatory reach over the variance embedded in the dataset, a divergence rendered tangible in the accompanying graphic. A head-to-head plotting of the regression diagnostics appears in **Figure 20.**

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

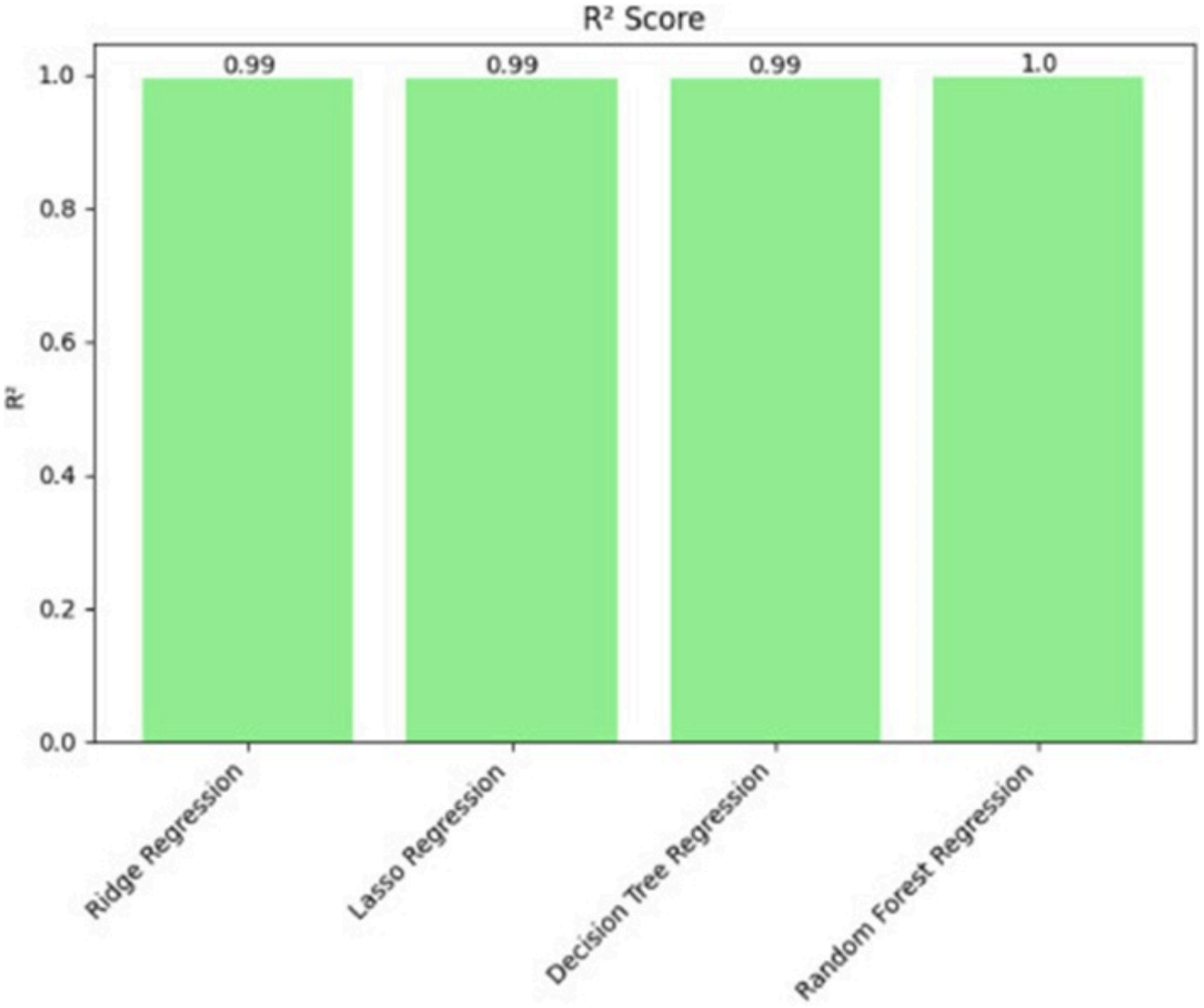



**Figure 19**. Coefficient of determination.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

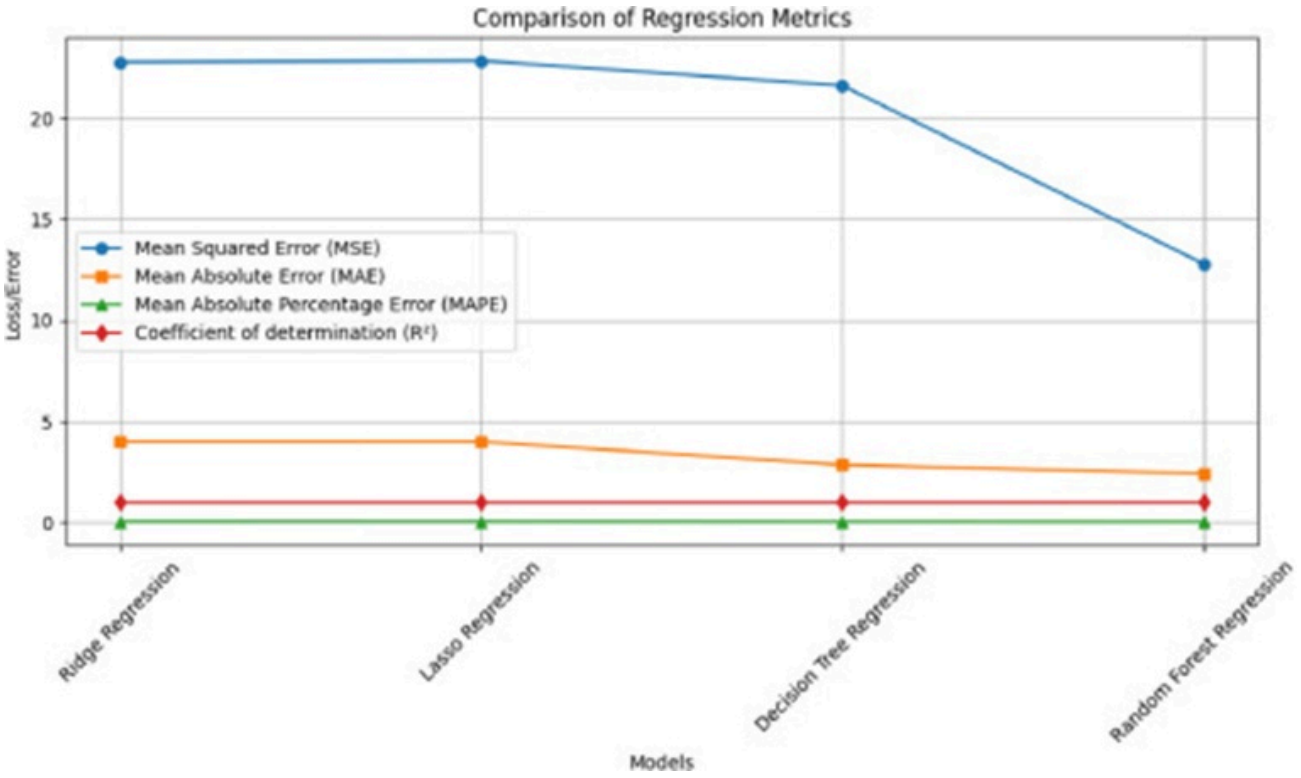



**Figure 20.** Comparison plot.

The Random Forest specification produced a test-partition $R^2$ of 0.99, a result that vouches for its considerable capacity to transfer learning beyond the fitting sample. The on-paper training $R^2$ reading of 1.0 should be understood as an artifact of the ensemble's practice of averaging estimates across its population of trees and, in light of the rigorous hyperparameter search conducted, does not constitute evidence of pathological memorization.

Delving further into the constellation of diagnostic indices (MAE, MAPE, and $R^2$) assembled in **Table 2**, an unmistakable through-line materializes that uniformly tips the balance toward tree-grounded architectures. Ridge Regression and Lasso Regression each clock an equivalent MAE of 4, while Decision Tree pares this down to an MAE of 2.86, and Random Forest whittles it still further to an MAE of 2.43. The MAPE readings sketch a congruent narrative: Ridge Regression and Lasso Regression each post a MAPE of 0.05, standing in contrast to the leaner MAPE of 0.03 turned in by both Decision Tree and Random Forest. Beyond these figures, Random Forest seizes a perfect $R^2$ of 1. In contrast, Ridge Regression, Lasso Regression, and Decision Tree converge upon a still-formidable $R^2$ of 0.99, a benchmark that collectively confirms their aptitude for absorbing a dominant slice of the observed variation. This unevenness in measured performance underscores the importance of allowing nonlinear functional forms when attempting to faithfully reproduce the tangled web of interactions between input covariates and the PMPP target. While Ridge Regression and Lasso Regression undoubtedly shed useful light on rectilinear tendencies, their toolkit for fully rendering the texture of nonlinear couplings and interaction phenomena may prove fundamentally circumscribed. Tree-anchored formulations, by marked contrast, have exhibited a far richer vocabulary for encoding these complexities, a fact that materializes directly in their elevated forecast accuracy.

**Table 2.** Comparison of regression model performance.

| Method | $R^2$ | MAPE | MAE | MSE |
|---|---|---|---|---|
| Ridge Regression | 0.99 | 0.05 | 4.00 | 22.76 |
| Lasso Regression | 0.99 | 0.05 | 4.00 | 22.82 |

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

| | | | | |
|---|---|---|---|---|
| Decision Tree Regression | 0.99 | 0.03 | 2.86 | 21.59 |
| Random Forest Regression | 1.00 | 0.03 | 2.43 | 12.80 |

Within this analytical subsection, comparative benchmarking cemented the finding that tree-structured modeling paradigms (Decision Tree and Random Forest) systematically outstripped their linear equivalents (Ridge and Lasso), lending further weight to the necessity of embracing nonlinear representational schemes when seeking to capture the governing dynamics of photovoltaic response.

The knowledge harvested from this investigation points to concrete pathways for upgrading energy orchestration, increasing system throughput, and easing the technical frictions of grid interconnection through dependable PMPP nowcasting. Fertile directions for follow-on study encompass the absorption of auxiliary predictor dimensions (exemplified by panel-level parameters, atmospheric descriptors, and senescence trajectories), together with the pursuit of frontier machine learning formalisms—deep neural architectures occupying a prominent position—to amplify both the granularity and the correctness of model outputs.

Stepping beyond scalar accuracy benchmarks, a qualitative audit of predictive conduct was mounted to diagnose the structural properties of the model residuals. The series of observed-against-predicted validation charts **(Figures 10-13)** show tight adherence to the identity baseline and exhibit no organized departure patterns. This finding suggests residuals are distributed evenly about the origin and display no pronounced heteroscedastic signatures. The tree-based variants, in particular, sustain even-tempered error profiles across the complete sweep of PMPP magnitudes, a trait that further cements their appropriateness for the target application. These regularities intimate that the bedrock assumptions of residual normality and uniform variance hold to a workable degree.

While Random Forest Regression claimed top-tier accuracy honors on the dataset forming the basis of this study, due recognition must be accorded to its native constraints regarding extrapolation beyond the training envelope. Conceived as a nonparametric ensemble aggregator, Random Forest thrives when generating estimates for instances residing within the multidimensional hull described by its calibration data. Yet, it is not structurally engineered for reliable projection into uncharted reaches of the covariate space. This inherent limitation could lead to complications if the fitted model is transplanted to unfamiliar PV module categories, divergent climatic zones, or operational contexts not represented during training. To hedge against such fragilities, upcoming investigations would do well to probe remedies including domain-adaptive incremental retraining regimens, the purposeful injection of more heterogeneous data streams spanning diverse contexts, and the hybridization of purely data-driven learning with physics-anchored constraints to stiffen generalizability. Such methodological enrichments would act to fortify the model's trustworthiness under the unpredictable conditions of real-world field operation.

Alongside considerations of predictive exactness, the burden on computational resources emerges as a decisive factor when weighing the feasibility of embedding machine-learned predictors into live, real-time PV control loops. Random Forest Regression, its estimable accuracy notwithstanding, can exact a nontrivial computational toll arising from its reliance on a large assembly of constituent trees. Where deployment scenarios demand low latency or must function within tight hardware envelopes, the model's inference turnaround time demands careful vetting against the capabilities of the intended processing substrate. Candidate strategies for mitigating this overhead involve dialing back the tally of base learners, enforcing stricter ceilings on tree depth, or resorting to model compaction techniques such as weight quantization or student-teacher distillation. In select circumstances, deliberately pared-down architectures—single shallow decision trees among them—may become the option of choice when the operational imperative of minimal response lag overrides the desire for the last increment of point-estimate fidelity. Navigating such design compromises assumes substantial practical weight when embedding algorithmic forecasting modules into functioning solar energy management frameworks.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

A further matter meriting serious reflection pertains to the endurance of estimation frameworks across protracted operational timelines, given the inexorable performance attrition that characterizes photovoltaic hardware over its service life [37]. With the passage of time, solar modules succumb to gradual conversion efficiency losses—a rule-of-thumb benchmark places this at approximately 0.5% per annum—a phenomenon that introduces steady-state drift into PMPP signatures. Predictive constructs trained exclusively upon commissioning-phase measurements stand to suffer mounting inaccuracies as the equipment matures. Remedial pathways meriting exploration in future rollouts encompass cyclic model refreshing informed by contemporaneous operational logs, or the deliberate incorporation of aging-attentive regressors—such as chronological age or aggregate lifetime generation—within the predictor matrix. A supplementary route involves self-updating learning frameworks that perpetually refresh their internal parameterization, thereby preserving prognostic validity throughout the asset's deployment lifespan.

Notwithstanding that the fitted models drew upon observational records encompassing a generous span of radiative and thermal regimes, dedicated evaluation under conditions characterized by sharp transient perturbations—typified by localized shading events or swiftly transiting cloud formations—was not undertaken as a discrete experimental phase. Forthcoming inquiries will prioritize the vetting of model response under such turbulent operational tableaus to substantiate dependability in bona fide PV field environments.

Turning to the question of transferability, although the present modeling framework posts strong predictive metrics on the corpus harvested from the SIRTA installation, its off-the-shelf applicability to PV assets employing alternative module architectures or deployed in geographically distinct settings may encounter constraints. Discrepancies arising from panel construction particulars, mounting inclinations, and region-specific climatic cadences each possess the capacity to perturb model fidelity. It therefore stands as prudent practice, whenever the model is to be introduced into novel environments, to perform either full retraining or parameter refinement anchored to locally representative datasets. Investigations moving forward should also weigh the merit of folding in auxiliary descriptor variables—module technology categorization, array azimuth and tilt, and latitude-longitude coordinates among them—to amplify the model's elasticity and span of relevance across disparate PV populations.

Embedding the formulated PMPP estimation apparatus within operational MPPT regulation loops represents an enticing avenue for continued scholarly pursuit. Although Random Forest constructs exact a more demanding computational toll, they nonetheless retain the capacity to serve predictions at operational cadences on state-of-the-art embedded processors. In settings where resource parsimony takes precedence, more streamlined specifications such as Ridge or Lasso furnish expedited inference with tolerable sacrifices in point-estimate exactitude. Anticipated research will tackle these systems-engineering obstacles head-on in the quest to enable instantaneous PV operational refinement. Future scholarly activity ought to methodically chart the multidimensional compromise plane spanned by model richness, estimation fidelity, and processing overhead, to isolate the most favorable parameterization for live PMPP deduction within photovoltaic installations. In parallel, subsequent inquiries could well venture into harnessing deep neural network paradigms—Long Short-Term Memory (LSTM) architectures and Convolutional Neural Networks (CNNs) figuring prominently—for the express purpose of PMPP forecasting in solar energy contexts. LSTM configurations are innately geared toward capturing temporally extended dependency structures embedded in serially ordered records, a capability that dovetails neatly with the imperative of modeling the mercurial behavior of PV systems subjected to fluctuating atmospheric forcings. CNNs, for all their provenance in visual recognition tasks, have amply demonstrated their utility on structured sequential prediction problems by deriving progressively abstracted feature hierarchies from input streams. Such deep learning formalisms carry with them the promise of lifting the ceiling on prognostic accuracy, most pointedly under operational regimes featuring abrupt vacillations in insolation and thermal conditions. Yet their uptake is shadowed by substantial hurdles, including steeper computational demands, the need for training corpora that are simultaneously more voluminous and more heterogeneous, and an intensified vulnerability to pathological overfitting when restricted to measurements from a single geographic provenance. Arriving at a workable accommodation between these prospective dividends and attendant liabilities emerges as an essential prerequisite for their productive assimilation into the sphere of instantaneous PV system optimization.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

## Cross-validation results

To add structural integrity to the performance assessment, a 10-fold cross-validation regimen was applied to each member of the regression quartet. This technique carves the full dataset into ten equivalently sized partitions, cyclically fits the model on nine of these subdivisions, and gauges its performance on the lone withheld segment—rotating through this procedure ten times so that every partition occupies the validation role exactly once. Aggregate descriptive statistics capturing both central tendency and spread were derived for three principal diagnostic indicators: MSE, MAE, and $R^2$.

The compilation presented in **Table 3** below encapsulates the findings consolidated over the entirety of the folds:

**Table 3.** for Cross-validation results.

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

These outcomes reaffirm that the Random Forest ensemble delivers projections distinguished by superior accuracy and enhanced stability, as reflected in the most competitive error statistics and the highest $R^2$ readings averaged across the cross-validation cycles. The deployment of 10-fold cross-validation helps dampen overfitting risks and lends greater credence to the inter-model ranking exercise.

# Conclusion

The present body of work has illuminated the capability of data-anchored regression methodologies to anticipate the maximum power point (PMPP) realized by photovoltaic (PV) systems. Among the suite of analytical instruments subjected to comparative scrutiny, tree-grounded constructs—preeminently Random Forest Regression—systematically eclipsed linear alternatives, a superiority rooted in their ability to reproduce the nonlinear interdependences that bind irradiance, temperature authentically, and the resultant PMPP. The evidence marshaled herein amplifies the indispensability of faithful irradiance sensing and of nonlinear modeling postures in augmenting both the energy conversion performance and the operational trustworthiness of PV assets. High-fidelity PMPP forecasts can inform sharper energy dispatch strategies, more judicious system dimensioning, and accelerated assimilation of solar-generated power into the broader electrical infrastructure.

Peering into the horizon, forthcoming endeavors will pivot toward reinforcing model generalizability by assimilating supplementary predictor dimensions, including but not limited to module-level descriptors, meteorological variables, and senescence-tracking indicators. Deep neural network formulations, spanning LSTM and CNN morphologies, are slated for systematic inquiry into their proficiency at encoding temporally extended dynamics and at hardening predictive resilience against volatile meteorological transients. Beyond modeling refinements, the tangible incorporation of the prediction engine into operational real-time maximum power point tracking (MPPT) controllers will be actively explored, with exacting attention devoted to both computational frugality and the pragmatic difficulties of seamless embedding within resource-constrained embedded PV platforms.

## List of abbreviations and acronyms

PV: Photovoltaic

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

GPOA: Global Plane of Array Irradiance

MAE: Mean Absolute Error

MAPE: Mean Absolute Percentage Error

MSE: Mean Squared Error

$R^2$: Coefficient of Determination

## Acknowledgements

None

## Conflict of interest

None

## Financial support

None

## Ethics statement

None

Received: 27 May 2022 Revised: 02 Aug 2022 Accepted: 19 Aug 2022
Published online: 15 June 2023

### Rights and permissions

**Open Access** The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

## References

1. Valavanidis A. Can renewable energy sources produce all the global electricity by 2050? Decarbonisation, ambitious climate goal by reducing carbon intensity of the electricity power sector. Athens: National and Kapodistrian University of Athens; 2020.

2. Panagoda LPSS, Sandeepa RAHT, Perera WAVT, Sandunika DMI, Siriwardhana SMGT, Alwis MKSDA, et al. Advancements in photovoltaic (PV) technology for solar energy generation. J Res Technol Eng. 2023;4(3):30-72.

3. Fahim KE, De Silva LC, Hussain F, Yassin H. 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(15):11837.

https://doi.org/10.3390/su151511837.

4. Klugmann-Radziemska E, Wcisło-Kucharek P. Photovoltaic module temperature stabilization with the use of phase change materials. Sol Energy. 2017;150:538-45. https://doi.org/10.1016/j.solener.2017.05.016.

5. Pillai DS, Ram JP, Lopez Garcia J, Kim YJ, Catalao JPS. Experimental studies on a new array design and maximum power tracking strategy for enhanced performance of soiled photovoltaic systems. IEEE Trans Power Electron. 2024;39(1):1596-608. 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.

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 MAB, Zhao D, Jamil H. Forecasting optimal power point of photovoltaic system using reference current based model predictive control strategy under varying climate conditions. Int J Control Autom Syst. 2024;22(10):3117-32. https://doi.org/10.1007/s12555-023-0823-7.

10. Farhat M, Barambones O, Sbita L. A real-time implementation of novel and stable variable step size MPPT. Energies. 2020;13(18):4668. https://doi.org/10.3390/en13184668.

11. Farhat M, Barambones O. Advanced control scheme optimization for stand-alone photovoltaic water pumping systems. Computation. 2024;12(11):224. https://doi.org/10.3390/computation12110224.

12. Azlor M, Ferré E, Gasulla M, Reverter F. FOCV-MPPT power management unit for submilliwatt indoor PV cells. Proceedings. 2024;97(1):99. https://doi.org/10.3390/proceedings2024097099.

13. Farhat M, Barambones O, Sbita L. 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. Nzoundja Fapi CB, Touré ML, Camara MB, Dakyo B. 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); 2024. https://doi.org/10.1109/ISCV60512.2024.10620157.

15. Jony MJA, Islam BU, Roy K, Lamon A, Alom MS, Jami MR. 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); 2024. https://doi.org/10.1109/ICEEICT62016.2024.10534369.

16. Büyük M, İnci M. 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.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

17. Guessoum D, Takruri M, Badawi SA, Farhat M, ElBadawi I. Maximum power point tracking using unsupervised learning for photovoltaic power systems. Int J Sustain Eng. 2024;17(1):397-412. https://doi.org/10.1080/19397038.2024.2356834.

18. Takruri M, Farhat M, Barambones O, Ramos-Hernanz JA, Turkieh MJ, Badawi M, et al. Maximum power point tracking of PV system based on machine learning. Energies. 2020;13(3):692. https://doi.org/10.3390/en13030692.

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.

20. Kalogerakis C, Koutroulis E, Lagoudakis MG. Global MPPT based on machine-learning for PV arrays operating under partial shading conditions. Appl Sci. 2020;10(2):700. https://doi.org/10.3390/app10020700.

21. Yılmaz M, Kaleli A, Çorapsız MF. 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 A, Nayak N, Hannon N, et al. A new real-time maximum power point tracking scheme for PV-based microgrid stability using online deep ridge extreme learning machine algorithm. Results Eng. 2023;20:101590. https://doi.org/10.1016/j.rineng.2023.101590.

23. Marinho FP, Rocha PAC, Rocha Neto AR, Santos VO. Dimensional reduction for solar irradiance forecasting problem using principal components analysis and Turk-Pentland strategy. In: 2024 International Joint Conference on Neural Networks (IJCNN); 2024. https://doi.org/10.1109/IJCNN60899.2024.10651398.

24. Zheng Y, Ge Y, Muhsen S, Wang S, Elkamchouchi DH, Ali E, et al. New ridge regression, artificial neural networks and support vector machine for wind speed prediction. Adv Eng Softw. 2023;179:103426. https://doi.org/10.1016/j.advengsoft.2023.103426.

25. Dai S. Variable selection in convex quantile regression: L1-norm or L0-norm regularization? Eur J Oper Res. 2023;305(1):338-55. https://doi.org/10.1016/j.ejor.2022.05.041.

26. Cardall AC, Hales RC, Tanner KB, Williams GP, Markert KN. LASSO (L1) regularization for development of sparse remote-sensing models with applications in optically complex waters using GEE tools. Remote Sens. 2023;15(6):1670. https://doi.org/10.3390/rs15061670.

27. Wang S, Chen Y, Cui Z, Lin L, Zong Y. Diabetes risk analysis based on machine learning LASSO regression model. J Theory Pract Eng Sci. 2024;4(1):58-64. https://doi.org/10.53469/jtpes.2024.04(01).08.

28. Shi M, Hu W, Li M, Zhang J, Song X, Sun W. Ensemble regression based on polynomial regression-based decision tree and its application in the in-situ data of tunnel boring machine. Mech Syst Signal Process. 2023;188:110022. https://doi.org/10.1016/j.ymssp.2022.110022.

29. Dinesh P, 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. AIP Conf Proc. 2024;2853(1):020140. https://doi.org/10.1063/5.0203746.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

30. Chen W, Yang Z. Landslide susceptibility modeling using bivariate statistical-based logistic regression, naïve Bayes, and alternating decision tree models. Bull Eng Geol Environ. 2023;82(5):190. https://doi.org/10.1007/s10064-023-03216-1.

31. Sa R, Yang T, Zhang Z, Guan F. 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(1):e68-e80.

32. Seyrek EC, Uysal M. 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(2):161-74. https://doi.org/10.15446/esrj.v28n2.105296.

33. Chitteti C, Sree TSD, Madhavi KR, Pooja P, Jayanth S, Reddy MH. 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); 2024. https://doi.org/10.1109/ICITEICS61368.2024.10624954.

34. Xu W, Tu J, Xu N, Liu Z. 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 D, Sun K. Random forest solar power forecast based on classification optimization. Energy. 2019;187:115940. https://doi.org/10.1016/j.energy.2019.115940.

36. Ait Tchakoucht T, Elkari B, Chaibi Y, Kousksou T. Random forest with feature selection and K-fold cross validation for predicting the electrical and thermal efficiencies of air based photovoltaic-thermal systems. Energy Rep. 2024;12:988-99. https://doi.org/10.1016/j.egyr.2024.07.002.

37. Sharma V, Chandel SS. 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.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press