

Article

A Two-Stage Unsupervised–Supervised Machine Learning Framework for Maximum Power Point Tracking of Photovoltaic Systems

Maissa Farhat ¹, Maen Takturi ^{2,*}, Oscar Barambones ³, Mohammad Rabih ¹, Aws Al-Qaisi ², Ghaya Alhebsi ¹, Mariam Al Ali ¹, Reem Al Nuaimi ¹, Naema Alharmi ¹ and Muna Alshehhi ¹

- ¹ Department of Mechanical, Electrical and Intelligent Systems Engineering, School of Engineering and Computing, American University of Ras Al Khaimah, Ras Al Khaimah 72603, United Arab Emirates; maissa.farhat@gmail.com (M.F.); mohammadrabih.aziz@gmail.com (M.R.); ghayaalhebsi@gmail.com (G.A.); maryamalhadi@yahoo.com (M.A.A.); reeemaalneaimi@gmail.com (R.A.N.); naemaabdullaalshahin@hotmail.com (N.A.); munaam.19@hotmail.com (M.A.)
- ² College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait; aws.al-qaisi@aum.edu.kw
- ³ Systems Engineering and Automatic Control Department, Faculty of Engineering Vitoria-Gasteiz, University of the Basque Country, 01006 Vitoria-Gasteiz, Spain; oscar.barambones@ehu.es
- * Correspondence: maen.takturi@aum.edu.kw

Abstract

Maximum Power Point Tracking (MPPT) is essential for maximizing photovoltaic (PV) energy extraction under changing operating conditions. This paper investigates a two-stage machine-learning framework combining K-means or fuzzy C-means (FCM) clustering of irradiance–temperature conditions with region-specific support vector regression (SVR), Gaussian process regression (GPR), and bagged ensemble trees (BET) to estimate the maximum-power-point reference voltage. The predicted voltage is supplied to a proportional–integral–derivative (PID)-controlled boost converter, and closed-loop performance is evaluated in MATLAB R2024/Simulink under changing irradiance, temperature, and load. Under the original evaluation protocol, two-cluster K-means—BET achieved the lowest MAE among the K-means prediction configurations, whereas integrated-energy rankings did not consistently follow voltage-prediction rankings. An additional leakage-controlled assessment used 70% of data for training and 30% for independent test observations, with clustering, standardization, and five-fold cross-validated Bayesian hyperparameter optimization performed using training observations only. In this stricter comparison, unclustered SVR achieved the lowest overall MAE, while FCM with three clusters improved BET MAE by approximately 9.7%. These results show that clustering benefits depend on the regression model and evaluation protocol. The closed-loop energy findings apply to the original simulated configurations; the new strict hold-out models were not separately tested in closed-loop simulation. Hardware implementation and broader-dataset validation remain future work.



Academic Editor: Adel El-Shahat

Received: 27 August 2026

Revised: 11 October 2026

Accepted: 14 October 2026

Published: 16 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: photovoltaic systems; maximum power point tracking; machine learning; clustering; gaussian process regression; ensemble learning; support vector regression; renewable energy

1. Introduction

The increasing global demand for clean and sustainable energy has accelerated the deployment of photovoltaic (PV) systems as one of the most promising renewable energy technologies. Solar energy offers a viable solution for reducing greenhouse gas emissions, mitigating climate change, and decreasing dependence on fossil fuels. Despite these advantages, the relatively low energy conversion efficiency of PV systems remains one of the major challenges limiting their widespread utilization [1–4].

The output characteristics of a PV module are highly dependent on environmental conditions, particularly solar irradiance and cell temperature. Variations in these parameters caused by cloud cover, atmospheric particles, seasonal changes, and other weather conditions continuously alter the operating point of the PV system, leading to significant fluctuations in the generated power [5]. Consequently, ensuring that the PV system operates continuously at its Maximum Power Point (MPP) is essential for maximizing energy harvesting and improving the overall efficiency and economic viability of solar installations [3,5,6].

Maximum Power Point Tracking (MPPT) algorithms have therefore become an integral component of modern PV systems. Their primary objective is to continuously identify the optimal operating voltage corresponding to the MPP under changing environmental conditions. Over the past decades, numerous MPPT techniques have been proposed, ranging from conventional methods to advanced optimization and artificial intelligence approaches [3,4,7]. Conventional techniques are generally simple to implement but often suffer from steady-state oscillations, slow convergence, or degraded performance under rapidly changing environmental conditions. These limitations have motivated researchers to investigate intelligent data-driven alternatives capable of modeling the nonlinear characteristics of PV systems more accurately.

Recent advances in Machine Learning (ML) have demonstrated considerable potential for improving MPPT performance [5,8–14]. Supervised learning methods have been successfully employed to estimate the optimal reference voltage directly from environmental measurements, thereby reducing tracking errors and improving convergence speed. Meanwhile, unsupervised learning techniques have been explored to identify hidden structures within PV operating conditions without requiring labeled data. Collectively, these approaches demonstrate the capability of ML to address the nonlinear relationship between environmental variables and the optimal operating point of PV systems.

Among supervised learning techniques, Support Vector Regression (SVR) has received considerable attention because of its strong generalization capability and effectiveness in nonlinear regression problems [15–17]. Gaussian Process Regression (GPR) provides a probabilistic framework that combines high prediction accuracy with uncertainty modeling, making it particularly attractive for complex nonlinear estimation problems [18,19]. Ensemble learning methods, such as Bagging Ensemble Trees (BET), have also demonstrated excellent predictive performance by combining multiple decision trees to improve robustness while reducing prediction variance [20–23].

Although significant progress has been achieved in ML-based MPPT, most existing approaches employ a single global prediction model that attempts to represent the entire irradiance–temperature operating space. Since PV operating conditions exhibit substantial variability and nonlinear behavior, learning a single global mapping can unnecessarily increase model complexity and reduce prediction accuracy. Partitioning the operating space into smaller and more homogeneous regions before regression has the potential to simplify the learning problem and improve prediction performance. The computational requirements of this architecture, however, depend on the selected regression model, model

size, hardware architecture, and implementation platform and are not quantified in the present study.

Although hybrid and machine-learning-based MPPT approaches have increasingly been investigated, the role of machine learning within these frameworks varies considerably. Existing approaches commonly employ machine-learning models for direct prediction of the maximum-power operating point or combine intelligent prediction with conventional or optimization-based MPPT techniques. In contrast, the present study investigates a two-stage unsupervised–supervised learning architecture in which unsupervised clustering is first used to partition the photovoltaic operating space into distinct operating regions, followed by region-specific supervised regression for estimation. Specifically, K-means and Fuzzy C-Means (FCM) are investigated for operating-region identification, while SVR, GPR, and BET are evaluated as local regression models. The corresponding regression models without clustering are also evaluated as global baselines, allowing the effect of operating-region decomposition on prediction to be examined explicitly. The estimated reference voltage is then employed by a PID-controlled DC–DC boost converter to regulate the PV operating point and maximize power extraction V_{MPP} .

The present framework should also be distinguished from the K-Nearest Neighbors (KNN)-based hybrid MPPT method previously reported in [24]. In [24], K-Nearest Neighbors is employed to identify operating conditions neighboring the current irradiance condition and their associated maximum-power-point information, which is subsequently used to restrict the duty-cycle search range for a local MPPT search under partial-shading conditions. In contrast, the present study does not use KNN to define a restricted search interval. Instead, K-means and FCM are employed as an unsupervised first stage to partition the irradiance–temperature operating space into operating regions, after which SVR, GPR, and BET are trained as region-specific regression models for direct estimation of V_{MPP} . Furthermore, the corresponding global regression models without clustering are evaluated as baselines to quantify the effect of the operating-region decomposition. Therefore, the two studies employ different learning architectures and use machine learning for different functions within the MPPT process.

The principal contribution of this work lies in the systematic investigation of operating-region decomposition followed by region-specific supervised regression for V_{MPP} estimation, rather than in the individual clustering or regression algorithms themselves.

The main contributions of this work can be summarized as follows:

- A hybrid two-stage MPPT framework that combines unsupervised clustering with supervised regression to improve the estimation of the optimum reference voltage.
- A comprehensive comparison between two clustering techniques (K-means and Fuzzy C-Means) and three regression models (SVR, GPR, and BET) under multiple clustering configurations.
- An investigation of the impact of operating-space partitioning on regression accuracy, demonstrating that selected clustering–regression configurations improve prediction performance relative to their corresponding regression-only baselines.
- A closed-loop MATLAB/Simulink evaluation of the trained machine-learning models within the complete MPPT system.

The remainder of this paper is organized as follows. Section 2 reviews the existing literature on MPPT techniques and machine learning approaches for photovoltaic systems. Section 3 presents the proposed methodology, including the PV model, machine learning framework, and controller design. Section 4 discusses the results and comparative performance evaluation. Finally, Section 5 concludes the paper and outlines possible directions for future research.

2. Literature Review

The performance of PV systems is strongly influenced by environmental conditions, particularly solar irradiance and ambient temperature. Variations in these factors continuously alter the current–voltage and power–voltage characteristics of PV modules, causing the operating point to deviate from the MPP. Consequently, considerable research efforts have been devoted to developing MPPT techniques capable of maintaining optimal operation under varying environmental conditions and maximizing the energy harvested from PV systems [5].

Over the past several decades, numerous MPPT techniques have been proposed, ranging from conventional control algorithms to advanced machine learning approaches. These techniques differ considerably in terms of implementation complexity, convergence speed, computational requirements, and robustness under dynamic operating conditions. The following subsections provide an overview of the most widely adopted approaches and highlight the motivation for the proposed hybrid framework.

2.1. Conventional MPPT Techniques

Conventional MPPT algorithms remain widely used because of their relatively simple implementation and modest computational requirements. Among these methods, Perturb-and-Observe (P&O) is arguably the most extensively investigated and practically implemented algorithm. The basic operating principle relies on perturbing the operating voltage and observing the resulting change in output power. If the power increases, the perturbation continues in the same direction; otherwise, the direction is reversed. Owing to its simplicity, P&O does not require prior knowledge of the PV characteristics or direct measurements of irradiance and temperature [4]. Reported efficiencies often exceed 97% under steady operating conditions [25,26]. Nevertheless, the method suffers from steady-state oscillations around the MPP and its performance deteriorates under rapidly changing irradiance conditions.

Several improvements to the conventional P&O algorithm have been proposed to overcome these limitations. Modified P&O techniques have demonstrated improved tracking speed and accuracy during sudden changes in irradiance [27], while other studies introduced enhanced variants capable of reducing oscillations and improving convergence characteristics [28]. Dual-mode implementations have also been investigated and experimentally validated, offering improved performance while maintaining a relatively simple hardware implementation that requires only voltage and current measurements [29,30].

Besides P&O, other conventional MPPT approaches have been extensively explored, including Incremental Conductance (INC), Parasitic Capacitance (PC), Constant Voltage (CV), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) [25,31–33]. Although these methods have demonstrated satisfactory performance in many applications, they generally remain sensitive to abrupt environmental changes and often exhibit oscillatory behavior around the operating point, motivating the development of more adaptive and intelligent MPPT techniques.

2.2. Machine Learning-Based MPPT

Recent advances in machine learning have significantly expanded the range of techniques available for MPPT. Unlike conventional algorithms that iteratively search for the optimum operating point, machine learning methods attempt to directly estimate the optimal reference voltage from environmental measurements, thereby reducing tracking time while improving prediction accuracy under nonlinear operating conditions [34].

Artificial Neural Networks (ANNs) were among the earliest machine learning techniques applied to MPPT. Their ability to approximate highly nonlinear functions enables

accurate estimation of the reference voltage under varying irradiance and temperature conditions [35]. Furthermore, several studies have demonstrated that ANN-based MPPT systems exhibit rapid response during testing while maintaining satisfactory prediction accuracy [13].

SVR has also attracted considerable attention because of its strong generalization capability and its ability to converge toward global rather than local optima. Consequently, SVR has been successfully employed to estimate the optimal reference voltage in several MPPT applications [5,11,36].

Ensemble learning techniques have likewise demonstrated promising performance in MPPT applications. Random Forest (RF), one of the most widely used ensemble methods, has shown faster convergence and lower prediction errors than several ANN-based approaches [37–39]. Other ensemble tree algorithms, including Boosted Trees and BETs, have also been investigated for PV modeling and MPPT [12,40]. These approaches benefit from combining multiple decision trees, thereby reducing prediction variance while improving robustness and generalization capability [41,42].

Similarly, GPR provides a probabilistic regression framework capable of modeling complex nonlinear relationships while simultaneously quantifying prediction uncertainty. Hybrid GPR-based MPPT strategies have demonstrated improved convergence speed and tracking efficiency compared with several optimization algorithms [43]. GPR has also been successfully employed for PV power forecasting [9,44] and for estimating the reference voltage under partial shading conditions using fuzzy logic-based controllers [45].

Collectively, these studies demonstrate that supervised machine-learning algorithms can model the nonlinear relationship between environmental conditions and the optimum operating point. Recent studies have continued to expand prediction-based and intelligent MPPT, including comparative evaluation of multiple regression models for maximum-power prediction and broader comparisons between classical and artificial-intelligence MPPT strategies [46,47]. These developments reinforce the importance of evaluating learning architecture, model choice, and system-level MPPT performance rather than relying on prediction error alone.

2.3. Unsupervised Learning for MPPT

Compared with supervised learning, unsupervised learning has received less attention in MPPT research, although data partitioning can be useful for identifying more homogeneous photovoltaic operating regions. Earlier work combined a genetic K-means procedure with a Radial Basis Function (RBF) network for photovoltaic MPP prediction [48]. More recently, K-means clustering has been applied directly to irradiance–temperature data to define homogeneous atmospheric regions and reduce the local MPPT search space [49].

These studies show that clustering can be used either to support a subsequent prediction model or to restrict the operating region searched by another MPPT method. The role of clustering in the present work is different: K-means and FCM are used to partition the irradiance–temperature space, after which separate supervised regression models are trained within the identified operating regions for direct estimation of the optimum reference voltage.

Accordingly, the present study evaluates whether operating-space partitioning followed by region-specific regression improves prediction performance relative to the corresponding global regression models, while also comparing hard and fuzzy partitioning within the same evaluation framework.

2.4. Reinforcement Learning Approaches

Reinforcement Learning (RL) has also emerged as a promising alternative for MPPT because of its ability to learn optimal control policies directly through interaction with the environment. Several RL-based MPPT algorithms have demonstrated fast convergence, stable operation, and reduced dependence on prior knowledge of PV characteristics [10,50–53]. Nevertheless, RL methods often require extensive training, careful reward design, and considerable computational resources, which may limit their practical implementation in resource-constrained embedded PV controllers.

2.5. Research Gap and Motivation

The literature clearly demonstrates the effectiveness of both conventional and machine learning-based MPPT techniques. Supervised learning algorithms have shown excellent capability for modeling the nonlinear relationship between environmental conditions and the optimum operating voltage, while unsupervised clustering has proven effective in partitioning PV operating conditions into more homogeneous regions.

However, most existing machine learning approaches employ a single global regression model to represent the entire irradiance–temperature operating space. Since PV operating conditions are inherently nonlinear and continuously changing, learning a single global mapping may increase model complexity, reduce prediction accuracy, and limit generalization under diverse operating regimes.

Integrating unsupervised clustering with supervised regression offers an attractive alternative. By partitioning the operating space into smaller and more homogeneous regions, clustering can reduce intra-cluster variability and allow the regression models to learn more localized mappings. The extent to which this architecture reduces online computational burden, however, depends on the selected regression model, model size, hardware architecture, and implementation platform. Accordingly, computational efficiency and embedded feasibility are not inferred from the present simulation results and require dedicated hardware profiling.

Motivated by these observations, this paper proposes a hybrid two-stage MPPT framework that combines unsupervised clustering with supervised regression to estimate the optimum reference voltage. Two clustering algorithms, namely K-means and FCM, are investigated to partition the irradiance–temperature operating space, while three supervised regression models—SVR, GPR, and BET—are evaluated to determine the effect of operating-space partitioning on prediction accuracy and subsequent closed-loop MPPT performance.

3. Methodology

This section presents the proposed hybrid two-stage machine learning framework for MPPT. The framework integrates unsupervised clustering, supervised regression, and closed-loop control to accurately estimate and continuously track the optimum operating voltage of PV system. Unlike conventional MPPT techniques that iteratively search for the maximum power point, the proposed approach directly predicts the optimum reference voltage from the measured irradiance and temperature. The predicted reference voltage is subsequently supplied to a PID-controlled DC–DC boost converter, enabling the PV system to operate continuously at the MPP.

Figure 1 illustrates the overall architecture of the proposed framework. The methodology consists of four main stages:

1. Experimental data acquisition and preprocessing.
2. Unsupervised clustering of the operating conditions.
3. Supervised learning for reference voltage prediction.
4. Closed-loop MPPT using a PID-controlled boost converter.

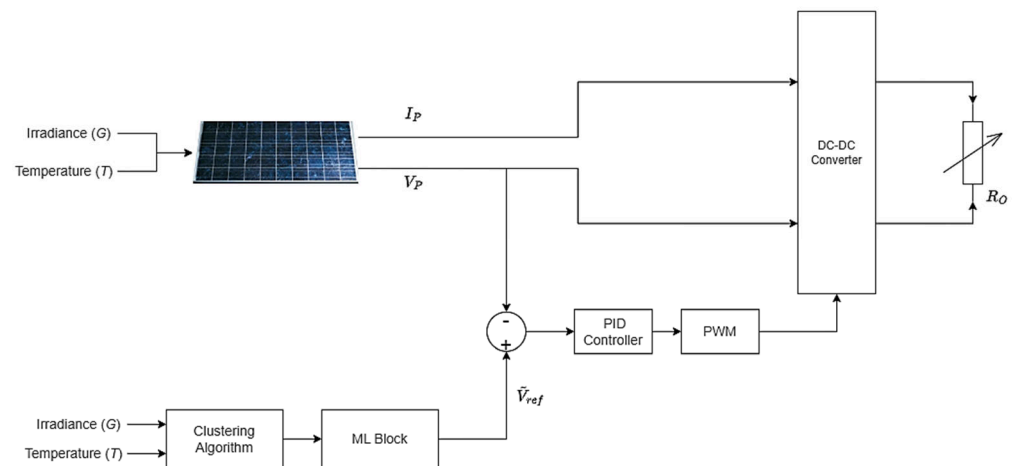


Figure 1. The proposed MPPT system.

Each stage is described in detail in the following subsections.

3.1. Overall Framework

The proposed MPPT system, illustrated in Figure 1, combines machine learning with conventional feedback control to maximize the power extracted from the photovoltaic array. The system consists of a PV panel, an ML-based reference voltage estimator, a Proportional–Integral–Derivative (PID) controller, a Pulse Width Modulation (PWM) generator, and a DC–DC boost converter.

Unlike conventional MPPT algorithms, which continuously perturb the operating voltage to locate the MPP, the proposed framework predicts the optimum reference voltage V_{MPP} directly from the environmental conditions. Solar irradiance and cell temperature are first processed by the machine learning block, which estimates the optimum operating voltage V_{MPP} corresponding to the maximum power point. This estimated voltage is then used as the reference input ($V_{ref} = V_{MPP}$) of the PID controller. The controller regulates the duty cycle of the boost converter through PWM switching, forcing the PV system to operate at the predicted optimum voltage and thereby maximizing the extracted power.

The proposed architecture separates the estimation problem from the control problem. Machine learning is responsible for accurately estimating the optimum operating voltage, while the PID controller ensures accurate voltage tracking despite variations in irradiance, temperature, and load conditions. This modular architecture improves flexibility and allows different learning algorithms to be evaluated without modifying the underlying control system.

3.2. Experimental Dataset and Data Preprocessing

The performance of any machine learning model depends heavily on the quality and representativeness of the training data. Rather than relying on synthetic simulations, this study employs an experimentally acquired photovoltaic dataset collected over an entire year in Vitoria-Gasteiz, Spain [5,54]. The dataset was generated using the experimental setup shown in Figure 2 and includes measurements of solar irradiance, ambient temperature, PV voltage, output current, and generated power under a wide range of operating conditions.



Figure 2. The data acquisition system, reproduced with permission from [54].

Prior to model development, the collected data underwent a preprocessing stage to improve consistency and reliability. Duplicate measurements and anomalous observations were removed, resulting in a clean dataset suitable for machine learning. For the regression-only baseline configurations, the processed dataset was divided into training and testing subsets using a 70:30 split. For the clustering-based configurations, the partitioning procedure described below was applied within the identified clusters. The training dataset was employed to construct the machine learning models, while the remaining samples were reserved for independent performance evaluation (testing).

The training data were further subjected to five-fold cross-validation for model development and model evaluation and selection. Specifically, the training subset was divided into five folds, with four folds used for model fitting and the remaining fold used for validation in each iteration. This process was repeated so that each fold served once as the validation fold. Importantly, the 30% hold-out test data were not included in the cross-validation or model-selection process and were reserved for the final evaluation of the trained models.

From each experimentally obtained I–V characteristic curve, the corresponding Maximum Power Point was identified, yielding paired values of irradiance, temperature, and optimum operating voltage. These quantities constitute the input–output pairs used throughout the learning process, where irradiance and temperature serve as predictor variables and the optimum reference voltage represents the target variable.

For the clustering-based configurations, the unsupervised clustering stage was first applied to the irradiance and temperature variables of the available dataset. The V_{MPP} target variable was not used during clustering. The observations within each resulting cluster were subsequently divided into 70% training and 30% testing subsets, with five-fold cross-validation applied exclusively to the training subset for model development and model evaluation and selection. Therefore, although the regression target values of the test

samples were not used during clustering or model training, the predictor variables of the complete dataset contributed to the unsupervised identification of the cluster structure.

3.3. Proposed Two-Stage Machine Learning Framework

The proposed MPPT strategy employs a two-stage machine learning architecture that combines unsupervised and supervised learning to estimate the optimum operating voltage of the photovoltaic system. The overall objective is to simplify the nonlinear mapping between the environmental conditions and the Maximum Power Point (MPP) voltage while maintaining high prediction accuracy.

Unlike conventional machine learning approaches that construct a single global regression model for the entire operating space, the proposed framework first partitions the irradiance–temperature space into more homogeneous regions before performing regression. This decomposition reduces the complexity of the learning problem by allowing each regression model to learn only the local behavior of the PV system within its assigned operating region. Consequently, improved prediction accuracy can be achieved without increasing the complexity of the individual regression models.

The proposed learning framework therefore consists of two sequential stages: an unsupervised clustering stage followed by a supervised regression stage.

3.3.1. Unsupervised Learning Stage

The first stage of the proposed framework aims to partition the photovoltaic operating conditions into homogeneous clusters according to their environmental characteristics. The clustering process is performed using irradiance and temperature measurements, which represent the two primary variables influencing the operating point of the PV system.

Two clustering techniques were investigated in this work, namely K-means clustering and Fuzzy C-Means (FCM) clustering. K-means partitions the dataset into mutually exclusive clusters by assigning each observation to the nearest cluster centroid. In contrast, FCM performs soft clustering by assigning each observation a degree of membership to multiple clusters, allowing overlapping operating regions to be represented. The present study compares these hard and fuzzy partitioning strategies within the same regression framework; hardware-specific computational performance is not evaluated.

For the investigated K-means configurations, cluster assignment is based on the Euclidean distance between the irradiance–temperature input vector and the corresponding cluster centroids, with each observation associated with the cluster having the minimum distance. Accordingly, once the cluster centers have been learned, a new operating point can be assigned to the nearest learned centroid without repeating the clustering procedure. In FCM, the MATLAB FCM algorithm determines the cluster centers together with the fuzzy membership matrix U , which represents the degree of membership of each observation in each cluster. Thus, whereas K-means provides a hard partition of the operating space, FCM represents the association of an observation with the identified operating regions through fuzzy membership degrees. The investigated configurations comprised two, three, and four clusters.

The rationale behind incorporating clustering before regression is that PV operating conditions are inherently nonlinear and nonstationary. Constructing a single regression model capable of accurately representing the entire irradiance–temperature space can be unnecessarily complex and may reduce generalization capability. By partitioning the operating space into smaller and more homogeneous regions, the intra-cluster variability is reduced, allowing each regression model to learn a simpler and more representative local mapping.

Consequently, the clustering stage serves two important purposes. First, it reduces the complexity of the subsequent learning problem by limiting the operating range seen by each regression model. Second, it improves prediction accuracy by allowing each model to specialize in a particular operating regime rather than attempting to approximate the complete nonlinear behavior of the photovoltaic system.

Cluster Validity Assessment

To quantitatively evaluate the quality of the clustering solutions and investigate the appropriate number of operating regions, the Silhouette coefficient and Davies–Bouldin (DB) index were employed. These complementary internal validation measures assess cluster compactness and separation without using the regression target variable.

For observation i , the Silhouette coefficient is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (1)$$

where $a(i)$ is the average distance between observation i and the other observations within the same cluster, while $b(i)$ is the minimum average distance between observation i and observations belonging to another cluster. The coefficient ranges from -1 to 1 , with larger values indicating better cohesion within clusters and greater separation between clusters. The mean Silhouette coefficient over all observations was used to evaluate each clustering configuration.

The DB index provides a complementary measure based on the ratio between within-cluster dispersion and between-cluster separation and is defined as:

$$DB = \frac{1}{K} \sum_{i=1}^K \max_{j \neq i} \frac{S_i + S_j}{M_{ij}} \quad (2)$$

where K is the number of clusters, S_i and S_j represent the within-cluster dispersion of clusters i and j , respectively, and M_{ij} denotes the distance between their centroids. Unlike the Silhouette coefficient, lower DB values indicate better clustering, corresponding to compact and well-separated clusters.

Both metrics were evaluated for $K = 2, 3$, and 4 to provide an objective assessment of the clustering structures in addition to the subsequent regression performance.

3.3.2. Supervised Learning Stage

Once the operating conditions have been partitioned, an independent supervised regression model is trained for each cluster to estimate the optimum reference voltage corresponding to the Maximum Power Point. The regression models receive irradiance and temperature as inputs and predict the optimum operating voltage, which subsequently serves as the reference signal for the PID controller.

Three supervised regression algorithms were investigated in this study: SVR, GPR, and BET. These algorithms were selected because they represent three complementary families of nonlinear regression techniques and have demonstrated excellent performance in energy forecasting and photovoltaic applications [15,18,19,55].

Support Vector Regression was selected because of its strong generalization capability and its ability to model nonlinear relationships using kernel functions while maintaining good prediction accuracy for relatively small training datasets. GPR provides a probabilistic regression framework capable of modeling highly nonlinear systems while simultaneously quantifying prediction uncertainty. BETs employ bootstrap aggregation to combine multiple decision trees, thereby reducing prediction variance and improving predictive robustness.

For each clustering configuration, an independent regression model is trained using only the samples belonging to the corresponding cluster. During operation, new irradiance–temperature measurements are first assigned to the appropriate cluster before the corresponding regression model predicts the optimum operating voltage. This hierarchical learning strategy allows each regression model to focus on a limited operating region and enables the effect of localized regression on prediction accuracy to be evaluated relative to a single global regression model.

Regression Performance Metrics

The predictive performance of each model is evaluated using the Mean Absolute Error (MAE) defined in (3). MAE measures the average magnitude of the prediction errors and is therefore employed as the primary measure of prediction accuracy. Additionally, Absolute Bias (AB), defined in (4), is used as a complementary measure of systematic prediction offset. While MAE quantifies the average magnitude of the individual prediction errors, AB quantifies the magnitude of the mean signed prediction error. Thus, the two metrics provide complementary information regarding overall prediction accuracy and systematic prediction bias.

$$MAE = \frac{1}{n} \sum_{i=1}^n \left| V_{MPP} - \tilde{V}_{MPP} \right| \quad (3)$$

$$AB = \left| \frac{1}{n} \sum_{i=1}^n \left(V_{MPP} - \tilde{V}_{MPP} \right) \right| \quad (4)$$

where n is the number of test data points.

Lower values of both metrics are desirable; however, MAE is used as the principal measure for comparing prediction accuracy because positive and negative errors may partially cancel when calculating AB.

3.4. MPPT Control System

The predicted reference voltage generated by the proposed machine learning framework is integrated into a closed-loop MPPT system to ensure continuous operation of the photovoltaic panel at its Maximum Power Point. The control system consists of a photovoltaic module, a DC–DC boost converter, and a PID controller. While the machine learning block determines the optimum operating voltage, the controller and power converter are responsible for maintaining the PV module at this operating point under varying environmental and loading conditions.

The interaction between the learning framework and the control system forms the complete MPPT architecture illustrated in Figure 1. The machine learning model estimates the optimum reference voltage directly from the measured irradiance and temperature, whereas the PID controller continuously adjusts the converter duty cycle to minimize the voltage tracking error. As a result, the PV system is driven toward its maximum power operating point without relying on conventional perturbation-based search algorithms.

3.4.1. Photovoltaic Model

The photovoltaic module considered in this study is the Mitsubishi TD185MF5 panel, which consists of fifty polycrystalline cells connected in series. The electrical characteristics of the module are summarized in Table 1.

Table 1. PV panel specifications.

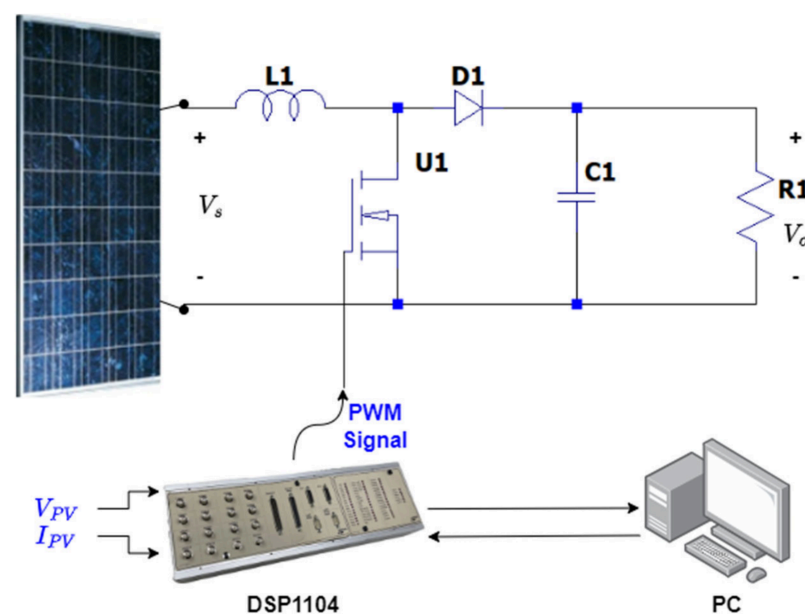
Cell Type	Polycrystalline Silicon
Maximum power	185 W
Open-Circuit Voltage (V_{oc})	30.60 V
Short-Circuit Current (I_{sc})	8.13 A
Maximum Power Voltage (V_{mpp})	24.40 V
Max Power Current (I_{mpp})	7.58 A
Normal Operating Cell Temperature	47.5 °C

To evaluate the proposed MPPT strategy under realistic operating conditions, experimentally acquired current–voltage characteristics covering a wide range of irradiance and temperature levels were employed. These characteristics define the nonlinear relationship between the environmental conditions and the corresponding Maximum Power Point, thereby providing the basis for training and evaluating the proposed machine learning framework.

Rather than modeling only the electrical behavior of the photovoltaic module, this work focuses on accurately estimating the operating voltage corresponding to the Maximum Power Point under varying environmental conditions. Consequently, the PV model serves as the physical plant whose operating point is continuously regulated by the proposed control system.

3.4.2. DC–DC Boost Converter

A DC–DC boost converter is employed to regulate the operating voltage of the photovoltaic panel and to match the electrical requirements of the load. The converter receives the voltage generated by the PV module and adjusts its operating point according to the duty cycle produced by the PID controller. The overall converter configuration is shown in Figure 3.

**Figure 3.** The DC–DC boost converter used for the study, reproduced with permission from [54].

The converter operates in two complementary switching modes. During the first interval, the switching device is turned on, allowing energy to be stored in the inductor while the load is supplied by the output capacitor. During the second interval, the switch

is turned off, causing the stored magnetic energy to be transferred to both the capacitor and the load through the forward-biased diode. By continuously adjusting the switching duty cycle, the converter regulates the photovoltaic operating voltage while maintaining the desired output voltage.

Assuming ideal converter operation, the average voltage across the inductor over one switching period is zero, leading to the well-known relationship expressed in (5). Rearranging (5) yields (6), which establishes the dependence of the output voltage on the converter duty cycle. This relationship forms the basis for regulating the photovoltaic operating point through closed-loop control.

$$V_{L,avg} = DV_s + (1 - D)(V_s - V_o) = 0 \quad (5)$$

$$V_o = \frac{V_s}{1 - D} \quad (6)$$

where V_s and V_o are the input and output voltages, respectively.

3.4.3. PID-Based Voltage Tracking

The predicted reference voltage generated by the machine-learning framework is supplied to a PID controller, which regulates the duty cycle of the boost converter. The controller acts on the error between the predicted reference voltage and the measured photovoltaic voltage, with the objective of tracking the reference voltage generated by the machine-learning model. The resulting control signal is applied to the boost converter through pulse-width modulation.

Unlike conventional MPPT algorithms in which the controller itself performs the search for the optimum operating point, the proposed framework separates reference-voltage estimation from closed-loop voltage tracking. The machine-learning model estimates the optimum operating voltage, whereas the PID controller tracks the resulting reference by adjusting the converter duty cycle. In the present study, the PID controller serves as the voltage-tracking component and was not itself the subject of optimization or comparative investigation.

The same controller configuration was maintained for all investigated machine-learning configurations so that the comparative closed-loop evaluation reflects differences in the predicted reference voltage rather than changes in the controller configuration.

3.5. Algorithm Selection and Model-Development Configuration

The predictive performance of machine-learning models depends on the selected learning algorithm and model configuration. In this study, the supervised models were developed using MATLAB R2024 Regression Learner, and the comparison was performed using standard Regression Learner model configurations rather than a separately defined manual hyperparameter-search procedure.

Three supervised regression algorithms—SVR, GPR, and BET—were investigated because they represent complementary machine learning paradigms. SVR provides excellent generalization capability for nonlinear regression problems through the use of kernel functions, making it particularly effective when the available training data are limited. GPR offers a probabilistic learning framework capable of modeling highly nonlinear relationships while simultaneously providing uncertainty estimates for its predictions. BET combines the outputs of multiple decision trees through bootstrap aggregation, thereby reducing prediction variance and improving predictive robustness. Evaluating these diverse learning strategies enables a comprehensive assessment of the proposed two-stage framework under identical operating conditions.

The supervised regression models were developed using the MATLAB Regression Learner app. Five-fold cross-validation was applied exclusively to the training data for model evaluation and selection. The Regression Learner configurations investigated included Fine Gaussian Support Vector Machine (SVM) for SVR, Exponential GPR for Gaussian process regression, and Bagged Trees for ensemble regression. The standard model configurations and automatic parameter settings provided by MATLAB Regression Learner were retained rather than defining a separate manual hyperparameter search space. Model selection was based on the five-fold cross-validation performance obtained from the training data, while the hold-out test data were excluded from model selection and used only for final performance evaluation. Together with the data-preprocessing and clustering procedures described in Sections 3.2 and 3.3, these settings define the model-development workflow used to obtain the reported machine-learning results.

The clustering stage evaluated two-, three-, and four-cluster configurations. The suitability of these partitions was assessed using the Silhouette coefficient and Davies–Bouldin index together with the subsequent regression performance, as described in Section 3.3.1. For FCM, the MATLAB implementation was used to determine the cluster centers and membership matrix; no separate manual optimization of the fuzzification coefficient or convergence criteria was performed in this study.

The adopted algorithmic configuration therefore provides a structured basis for evaluating the prediction performance of the investigated clustering and regression combinations, while the computational requirements of real-time embedded implementation remain to be quantified experimentally.

4. Discussion

This section evaluates the proposed two-stage hybrid machine-learning framework for MPPT through prediction analysis using the experimentally acquired photovoltaic dataset and closed-loop evaluation in MATLAB/Simulink. The evaluation is performed at three complementary levels. First, the prediction capability of the proposed clustering–regression framework is assessed using statistical performance measures to compare the investigated clustering and regression configurations for optimum-voltage estimation. Second, the trained machine-learning models are integrated within the closed-loop MPPT system to evaluate their influence on photovoltaic power and energy extraction under the investigated dynamic operating profile. Third, the practical implications and limitations of the proposed framework are examined, including generalization, computational requirements, cluster-assignment sensitivity, and the limitations associated with simulation-based validation.

The MPPT framework illustrated in Figure 1 was implemented in MATLAB/Simulink. Initially, the machine learning models were trained, validated, and tested offline using the experimental photovoltaic dataset collected in Vitoria-Gasteiz, Spain [5,54]. During this stage, the regression models were developed independently of the remaining control system to evaluate their ability to accurately estimate the optimum reference voltage ($\tilde{V}_{ref}$). Once the models had been trained, the selected configurations were implemented in MATLAB Function blocks and integrated with the photovoltaic module, the DC–DC boost converter, the PID controller, and the variable load to form the complete MPPT system.

The integrated framework was evaluated under varying irradiance, temperature, and load conditions. Consequently, the presented results assess both the statistical prediction accuracy of the proposed learning framework and its closed-loop ability to maximize photovoltaic power extraction under the simulated operating conditions considered in this study.

The discussion begins with the prediction accuracy results obtained from the initial evaluation, followed by an additional strict hold-out evaluation designed to assess the

generalization capability of the proposed framework while preventing information leakage during clustering. Subsequently, the closed-loop MATLAB/Simulink results are discussed, followed by the practical implications and limitations of the study. The strict hold-out evaluation was limited to prediction accuracy and did not include additional closed-loop simulations.

4.1. Prediction Accuracy Evaluation

The primary objective of the proposed hybrid learning framework is to estimate the optimum photovoltaic operating voltage as accurately as possible under varying environmental conditions. Since the estimated reference voltage is supplied directly to the PID controller, any prediction error may cause the photovoltaic operating point to deviate from the actual Maximum Power Point (MPP), thereby reducing the generated output power. Consequently, improving voltage prediction accuracy represents the first step towards improving the overall MPPT performance.

The predictive capability of the investigated machine learning models was evaluated using the Mean Absolute Error (MAE). MAE provides the average magnitude of the prediction error and is used as the primary measure of prediction accuracy. Lower MAE values indicate more accurate estimation of the optimum operating voltage.

To demonstrate the contribution of the proposed two-stage learning framework, the evaluation was performed progressively. First, the three regression algorithms were trained using the complete dataset without clustering to establish a baseline for comparison. Subsequently, K-means and Fuzzy C-Means (FCM) clustering were introduced as preprocessing stages to investigate whether partitioning the photovoltaic operating space could simplify the regression task and improve prediction accuracy.

Baseline Regression Performance

The prediction performance obtained using SVR, GPR, and BET without clustering is summarized in Table 2. These results represent the conventional machine learning approach, where a single regression model is responsible for approximating the complete nonlinear relationship between irradiance, temperature, and the optimum operating voltage.

Table 2. Results of all three algorithms when no clustering is applied.

Algorithm	MAE	AB
SVR	1.672	1.421
GPR	1.412	1.111
BET	1.298	1.171

All three regression algorithms successfully estimated the optimum reference voltage with satisfactory accuracy. Nevertheless, noticeable differences exist among the investigated models. BET achieved the lowest MAE values, indicating superior prediction accuracy, while GPR ranked second and SVR produced comparatively larger prediction errors. The improved performance of BET can be attributed to its ensemble learning strategy, which combines multiple decision trees to approximate highly nonlinear relationships while simultaneously reducing prediction variance. In contrast, SVR relies on a single nonlinear mapping to represent the complete operating space, making it more difficult to capture the diverse operating conditions represented within the experimental dataset.

Although the obtained prediction accuracy is encouraging, the baseline results also reveal the inherent difficulty of representing the complete photovoltaic operating space using a single global regression model. The operating characteristics of photovoltaic modules vary considerably with irradiance and temperature, resulting in a highly nonlinear

mapping between the environmental conditions and the optimum operating voltage. This observation motivates the introduction of an unsupervised clustering stage to partition the operating space into smaller and more homogeneous regions before regression.

Performance of the K-means-based Hybrid Framework

To investigate the effect of partitioning the operating space before regression, K-means clustering was incorporated as the first stage of the proposed hybrid learning framework. Three clustering configurations comprising two, three, and four clusters were investigated. Before evaluating their influence on regression performance, the quality of the resulting partitions was quantitatively assessed using the Silhouette coefficient and DB index. These internal validation measures provide complementary information regarding cluster cohesion and separation and enable an objective assessment of the different clustering configurations. The resulting values are presented in Table 3.

Table 3. Cluster-validity measures for the investigated K-means configurations.

Number of Clusters	Silhouette Coefficient	Davies-Bouldin Index
2	0.803	0.499
3	0.775	0.470
4	0.779	0.449

As shown in Table 3, the two-cluster configuration achieved the highest Silhouette coefficient (0.803), compared with 0.775 and 0.779 for three and four clusters, respectively, indicating the strongest overall combination of intra-cluster cohesion and inter-cluster separation according to this measure. In contrast, the Davies–Bouldin index decreased from 0.499 for two clusters to 0.470 and 0.449 for three and four clusters, respectively, with the four-cluster configuration therefore being favored according to the DB criterion. Consequently, the two internal validation measures do not identify a single unequivocal optimum. Rather than selecting the number of clusters based on one validity measure alone, the cluster-validity results were considered together with the subsequent regression performance and implementation complexity.

Having assessed the internal quality of the clustering solutions, their influence on prediction performance was subsequently investigated. Independent SVR, GPR, and BET regression models were trained using the corresponding clustered datasets. The resulting MAE and Absolute Bias values are summarized in Tables 4–6, while Figure 4 compares the MAE obtained for the investigated regression algorithms and clustering configurations.

Table 4. Results of all three algorithms using 2 clusters in K-means clustering.

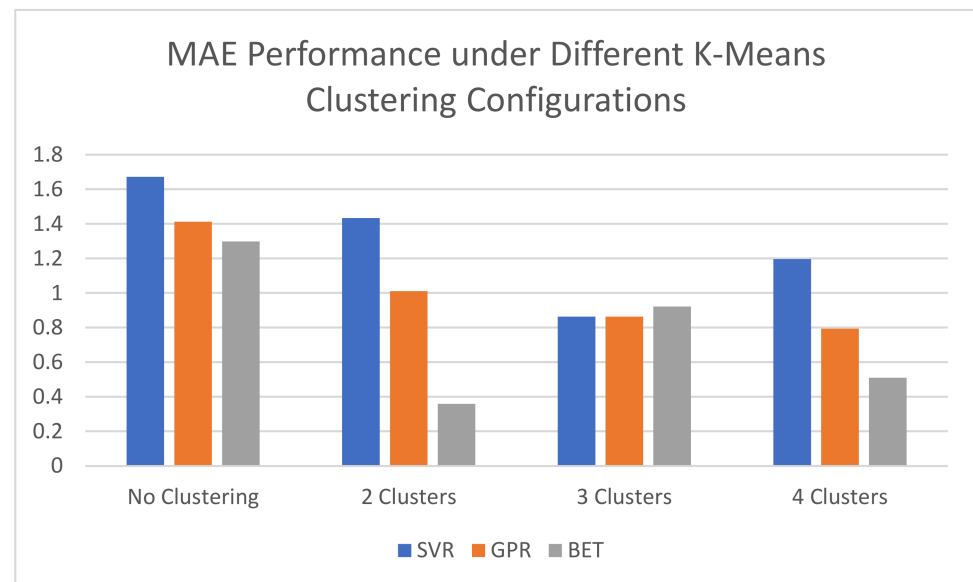
Algorithm	MAE	AB
SVR	1.433	1.239
GPR	1.011	1.004
BET	0.359	0.088

Table 5. Results of all three algorithms using 3 clusters in K-means clustering.

Algorithm	MAE	AB
SVR	0.863	0.725
GPR	0.863	0.705
BET	0.922	0.284

Table 6. Results of all three algorithms using 4 clusters in K-means clustering.

Algorithm	MAE	AB
SVR	1.196	1.041
GPR	0.794	0.743
BET	0.510	0.289

**Figure 4.** Bar Chart representing the performance of the three tested algorithms using K-means clustering.

Referring to Tables 4–6 and Figure 4, the effect of increasing the number of clusters on prediction accuracy was found to depend on the regression algorithm. SVR achieved its lowest MAE of 0.863 V using three clusters, whereas GPR achieved its lowest MAE of 0.794 V using four clusters, compared with 0.863 V for the three-cluster configuration. In contrast, BET achieved its lowest MAE of 0.359 V using two clusters, while its MAE increased to 0.922 V and 0.510 V for three and four clusters, respectively. These results indicate that increasing the number of clusters does not necessarily lead to improved prediction performance and that the appropriate degree of partitioning depends on the characteristics of the subsequent regression model.

When the regression results are considered together with the cluster-validity analysis, the two-cluster configuration provides a particularly attractive solution. It achieves the highest Silhouette coefficient while enabling BET to attain the lowest MAE (0.359 V) among all investigated K-means configurations. The corresponding Absolute Bias of 0.088 V further indicates a small systematic prediction offset. Although the Davies–Bouldin index favors a larger number of clusters, the additional partitioning does not result in a consistent improvement in regression accuracy and requires a greater number of local regression models. Considering cluster validity, prediction accuracy, systematic prediction offset, and implementation complexity collectively, the two-cluster K-means—BET configuration was selected for the subsequent closed-loop MPPT evaluation.

Performance of the Fuzzy C-Means-based Hybrid Framework

To further investigate the influence of the clustering strategy, the hard partitioning employed by K-means was replaced with Fuzzy C-Means (FCM) clustering. Unlike K-means, which assigns each operating condition exclusively to a single cluster, FCM allows each observation to belong simultaneously to multiple clusters through fuzzy membership

values. Such soft partitioning provides graded membership information for observations located near the transition regions between neighboring operating conditions.

The prediction performance obtained using the FCM-based framework is summarized in Tables 7–9, while the corresponding MAE values are presented in Figure 5.

Table 7. Results of all three algorithms using 2 clusters in Fuzzy C-means clustering.

Algorithm	MAE	AB
SVR	1.454	1.167
GPR	0.974	0.968
BET	0.355	0.113

Table 8. Results of all three algorithms using 3 clusters in Fuzzy C-means clustering.

Algorithm	MAE	AB
SVR	1.226	1.174
GPR	0.974	0.968
BET	0.826	0.809

Table 9. Results of all three algorithms using 4 clusters in Fuzzy C-means clustering.

Algorithm	MAE	AB
SVR	0.949	0.589
GPR	0.741	0.689
BET	0.606	0.405

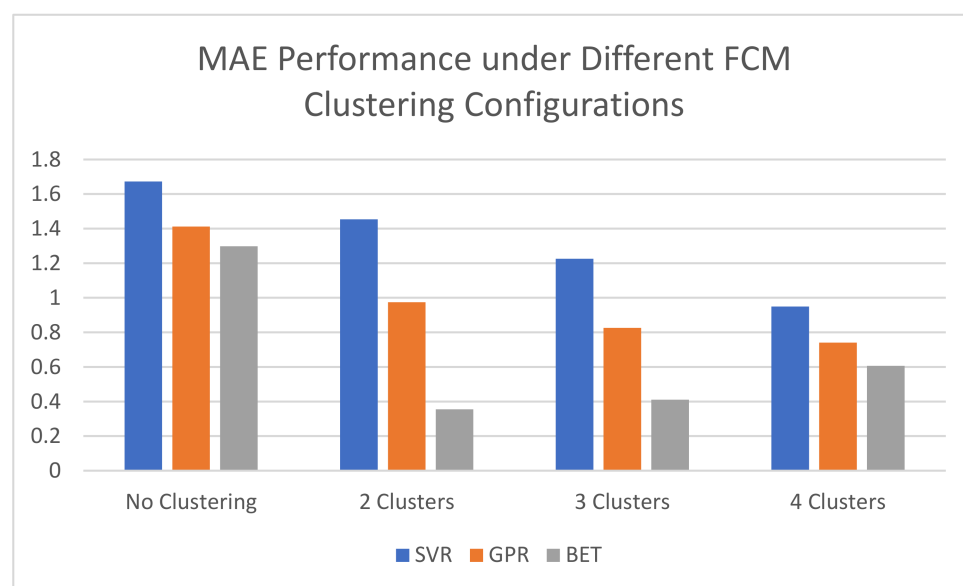


Figure 5. Bar Chart representing the performance of the three tested algorithms using Fuzzy C-means clustering.

Referring to Tables 7–9 and Figure 5, the influence of the number of clusters on prediction accuracy under FCM differs among the regression algorithms. SVR and GPR exhibit progressive reductions in MAE as the number of clusters increases, achieving their lowest MAE values of 0.949 V and 0.741 V, respectively, with four clusters. In contrast, BET

exhibits its best performance with two clusters, achieving an MAE of 0.355 V, compared with 0.411 V and 0.606 V for three and four clusters, respectively. Thus, while finer partitioning benefits SVR and GPR, it does not improve BET performance beyond the two-cluster configuration. This further demonstrates that increasing the number of clusters does not inherently improve prediction accuracy; rather, its effect depends on the characteristics of the subsequent regression model.

The Absolute Bias results provide complementary information regarding systematic prediction offset. The two-cluster FCM–BET configuration achieves an AB of 0.113 V, substantially lower than the corresponding values of 1.167 V and 0.968 V obtained using SVR and GPR, respectively. Therefore, among the investigated FCM configurations, the two-cluster BET model provides the lowest overall MAE while also exhibiting a small systematic prediction offset.

The average prediction results presented in Figure 5 confirm these findings. BET achieves the lowest MAE among the investigated FCM configurations, while the relative performance of SVR and GPR depends on the number of clusters. The strong performance of BET is consistent with the variance-reduction properties of bagged tree ensembles and with previous applications of tree-based ensemble learning in photovoltaic and renewable-energy modelling [12,42,55].

Comparison of the Hybrid Learning Frameworks

The overall prediction performance of the investigated hybrid learning frameworks is summarized in Figure 6, which directly compares the average Absolute Bias (AB) values obtained using K-means and FCM clustering.

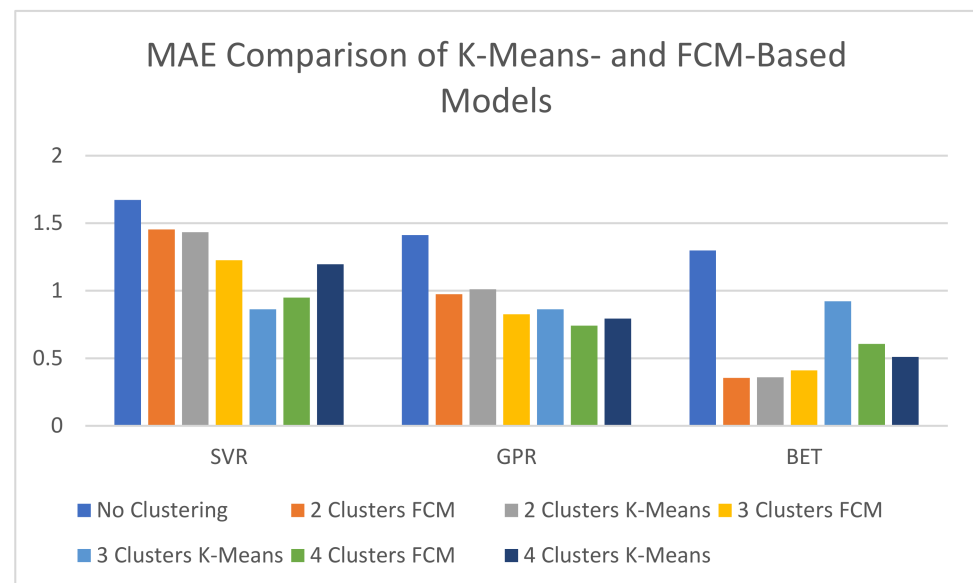


Figure 6. A comparison of the performances of the FCM-based and K-means-based hybrid algorithms.

Figure 6 compares the MAE obtained using the K-means- and FCM-based hybrid frameworks across the investigated regression algorithms and cluster configurations. The results show that neither clustering technique consistently outperforms the other across all regression models and numbers of clusters. Instead, the effect of clustering is strongly dependent on both the regression algorithm and the degree of partitioning. For SVR, the lowest MAE is obtained using three K-means clusters, whereas for GPR the lowest MAE is achieved using four FCM clusters. In contrast, BET achieves its lowest MAE using only two clusters for both clustering techniques.

In particular, the two-cluster FCM–BET and K-means–BET configurations achieve MAE values of 0.355 V and 0.359 V, respectively, indicating nearly identical prediction accuracy. The corresponding Absolute Bias values are 0.113 V and 0.088 V, respectively, with the K-means configuration exhibiting the smaller systematic prediction offset. These findings indicate that the choice of regression algorithm and the number of clusters have a greater influence on prediction performance than the choice between K-means and FCM itself. Considering the comparable MAE of the two BET configurations, the lower Absolute Bias, and the simpler hard-assignment mechanism of K-means, the two-cluster K-means–BET configuration was chosen as the best candidate for integration into the subsequent closed-loop MPPT system.

The prediction accuracy results indicate that incorporating clustering improved the performance of several regression configurations under the initial evaluation procedure. In particular, the combination of K-means clustering with two clusters and BET achieved favorable prediction accuracy compared with the other investigated configurations. To further examine the generalization capability of the proposed framework, an additional evaluation using a strict hold-out procedure is presented in the following subsection.

Generalization Assessment Using Strict Hold-Out Evaluation

To examine generalization without allowing independent test observations to influence clustering, a strict hold-out experiment was performed on PV data. A fixed random seed of 4 was used to select 70% of the data for training and 30% for testing. The same partition was used for all clustered and no-clustering regression configurations.

Irradiance and temperature were standardized using only the training-subset means and standard deviations. K-means and FCM were fitted to the training predictors for $K = 2, 3$, and 4. Unseen test observations were assigned to the fixed K-means centroids by minimum Euclidean distance or to the fixed FCM centers by maximum calculated membership. Neither clustering algorithm was refitted on the test observations, and the target voltage was not used to form clusters.

SVR, GPR, and BET were trained either globally (without clustering) or separately within each training-derived cluster. Bayesian hyperparameter optimization used 30 objective evaluations and five-fold cross-validation exclusively on the corresponding training observations. Predictions from the local regressors were recombined in their original test-observation order, and MAE and absolute bias (AB) were computed over all the unseen test observations. This strict hold-out experiment uses an additional model-development protocol and is reported in Table 10 separately from the original MATLAB Regression Learner results.

The strict hold-out evaluation demonstrates that incorporating clustering does not consistently improve prediction accuracy across all regression algorithms. SVR without clustering achieved the lowest overall MAE (0.724 V), while the unclustered GPR model also outperformed its clustered counterparts. Nevertheless, selected clustered configurations demonstrated improvements. In particular, FCM with three clusters reduced the MAE of BET from 0.934 V to 0.843 V, corresponding to an improvement of approximately 9.7%. These findings indicate that the effectiveness of clustering depends on the regression algorithm and the number of clusters.

The results of the strict hold-out evaluation differ from those of the initial evaluation, in which the two-cluster K-means–BET configuration achieved the best prediction accuracy among the K-means configurations. This difference highlights the influence of the evaluation procedure on the relative performance of the investigated models. The closed-loop MATLAB/Simulink results presented in Section 4.2 correspond to the configurations developed under the initial evaluation procedure and provide a separate assessment of

system-level performance. Since the models developed under strict hold-out procedure were not evaluated in closed-loop simulations, no additional conclusions regarding their energy-extraction performance can be drawn from the prediction accuracy results alone.

Table 10. Prediction performance under the strict hold-out evaluation protocol (all errors in V).

Clustering	K	Regression	MAE (V)	AB (V)
None	–	SVR	0.724	0.324
None	–	GPR	0.740	0.297
None	–	BET	0.934	0.335
K-means	2	SVR	0.731	0.337
K-means	2	GPR	1.026	0.448
K-means	2	BET	0.881	0.391
K-means	3	SVR	0.983	0.395
K-means	3	GPR	0.957	0.396
K-means	3	BET	0.890	0.384
K-means	4	SVR	0.932	0.229
K-means	4	GPR	0.956	0.360
K-means	4	BET	0.934	0.260
FCM	2	SVR	0.743	0.301
FCM	2	GPR	1.045	0.516
FCM	2	BET	0.910	0.425
FCM	3	SVR	0.835	0.392
FCM	3	GPR	0.972	0.377
FCM	3	BET	0.843	0.304
FCM	4	SVR	0.908	0.326
FCM	4	GPR	0.945	0.416
FCM	4	BET	1.041	0.433

4.2. Closed-Loop MPPT Performance

Although the prediction accuracy analysis presented in the previous subsection demonstrates the effectiveness of the proposed hybrid learning framework, the ultimate objective of any Maximum Power Point Tracking algorithm is to maximize the electrical power extracted from the photovoltaic module under varying operating conditions. Consequently, the best-performing regression models identified during the prediction stage were integrated with the complete MPPT system illustrated in Figure 1 to evaluate their performance under closed-loop simulation.

The complete MPPT system consists of the trained machine learning estimator, the photovoltaic module, the DC–DC boost converter, the PID controller, and a variable load. During operation, the machine learning model estimates the optimum reference voltage directly from the measured irradiance and temperature. This predicted reference voltage is subsequently supplied to the PID controller, which regulates the converter duty cycle to maintain the photovoltaic operating voltage as close as possible to the predicted optimum operating point. Therefore, the accuracy of the estimated reference voltage influences the resulting closed-loop MPPT performance, together with the dynamic response of the converter and controller.

The closed-loop MPPT system was evaluated over an 8 s dynamic simulation profile incorporating step variations in environmental and load conditions. During 0–2 s, the irradiance, temperature, and load resistance were 981.83 W/m², 312.63 K, and 10 Ω , respectively. At $t = 2$ s, the irradiance and temperature were changed to 452.83 W/m² and 301.14 K, respectively, while the load resistance remained at 10 Ω . At $t = 4$ s, the load resistance was increased from 10 Ω to 20 Ω while the environmental conditions remained unchanged. Finally, at $t = 6$ s, the irradiance and temperature were changed to 985.58 W/m² and 326.35 K, respectively, with the load resistance maintained at 20 Ω .

These step changes were intentionally employed to evaluate the transient response of the closed-loop MPPT system under abrupt environmental and load disturbances rather than to reproduce the natural temporal evolution of irradiance and temperature. Because the environmental variations were implemented as ideal step changes, finite irradiance and temperature rates of change were not prescribed; the corresponding step magnitudes were $\Delta G = -529.00 \text{ W/m}^2$ and $\Delta T = -11.49 \text{ K}$ at $t = 2 \text{ s}$, and $\Delta G = +532.75 \text{ W/m}^2$ and $\Delta T = +25.21 \text{ K}$ at $t = 6 \text{ s}$. These step changes were intentionally employed to evaluate the transient response of the closed-loop MPPT system under abrupt environmental and load disturbances rather than to reproduce the natural temporal evolution of irradiance and temperature.

The generated output power obtained is presented in Figures 7–10. Unlike the prediction analysis discussed previously, which evaluates the statistical accuracy of the estimated reference voltage, the following discussion investigates how these prediction differences influence the actual electrical power extracted from the photovoltaic system.

Comparison of the Best Regression Models

Figure 7 compares the generated output power obtained using the selected two-cluster FCM configuration for each regression algorithm, namely SVR2, GPR2, and BET2. All three configurations produce closely comparable photovoltaic power trajectories throughout the investigated dynamic simulation profile, demonstrating their ability to provide suitable reference voltages for closed-loop MPPT operation.

Small differences can be observed among the investigated regression models, but the instantaneous power curves overlap over much of the simulation interval. Therefore, the power plots should be interpreted together with the integrated-energy analysis presented in the *Quantitative Energy-Extraction Analysis* subsection (Figures 11 and 12), rather than used alone to rank the regression models.

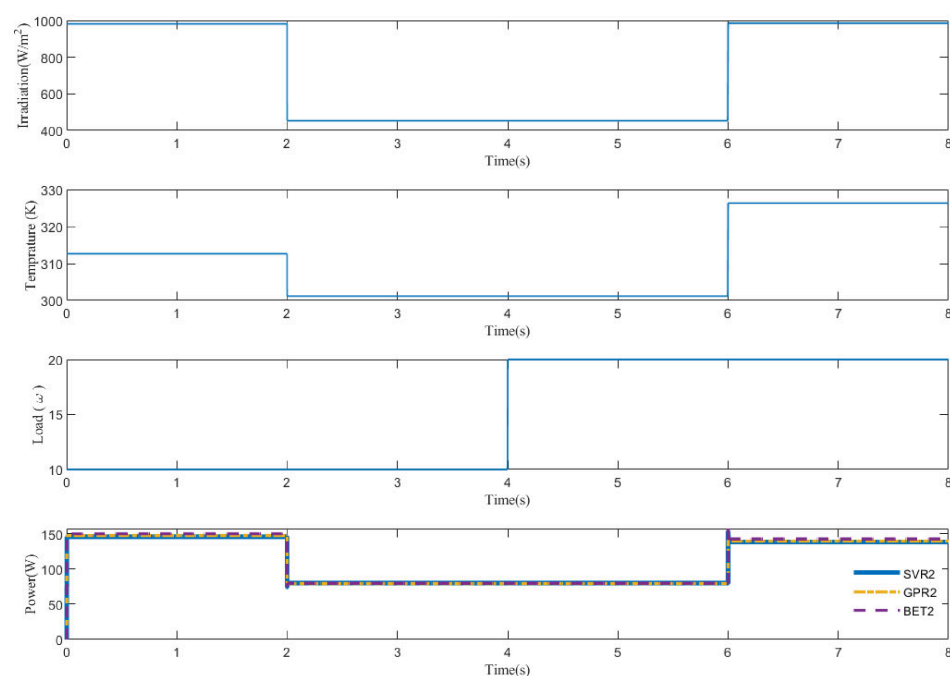


Figure 7. A comparison of generated power as a result of employing the three supervised ML algorithms with four FCM clusters.

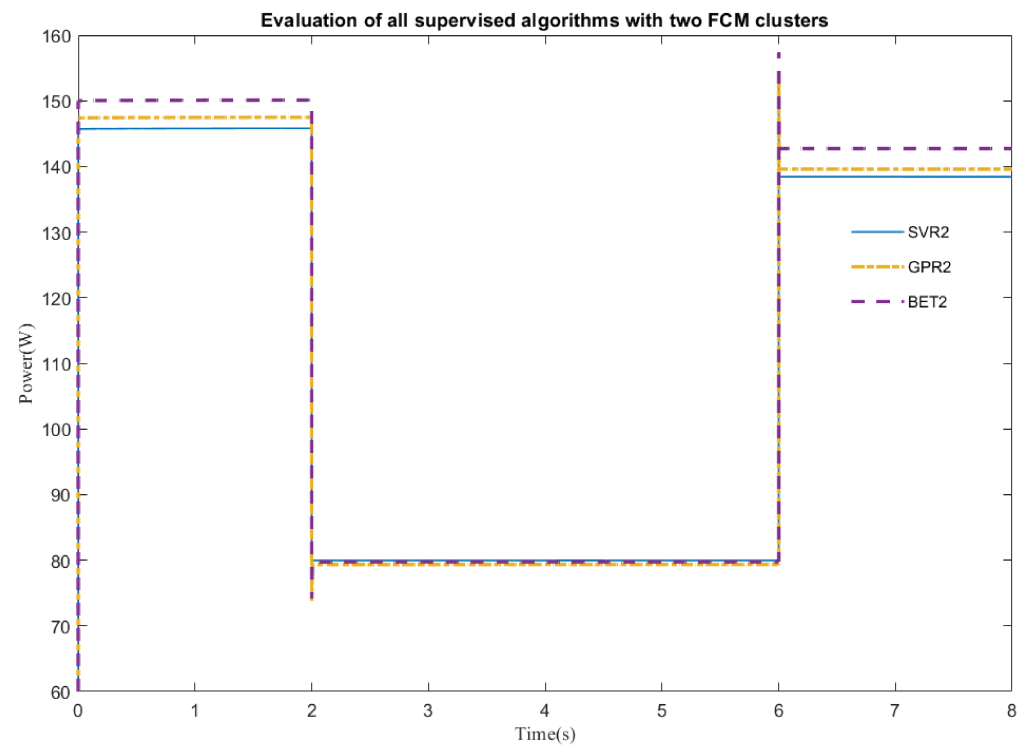


Figure 8. Zoomed version of the last subplot in Figure 7.

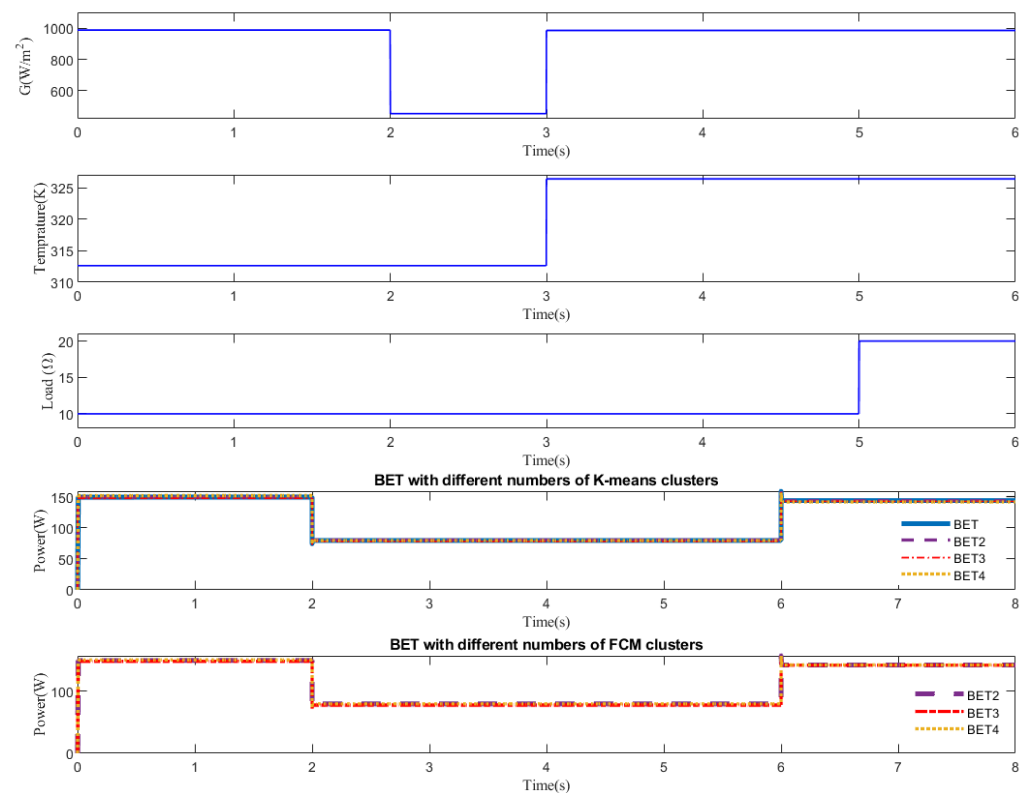


Figure 9. A comparison of generated power as a result of employing the BET-based hybrid algorithm for different clustering scenarios.

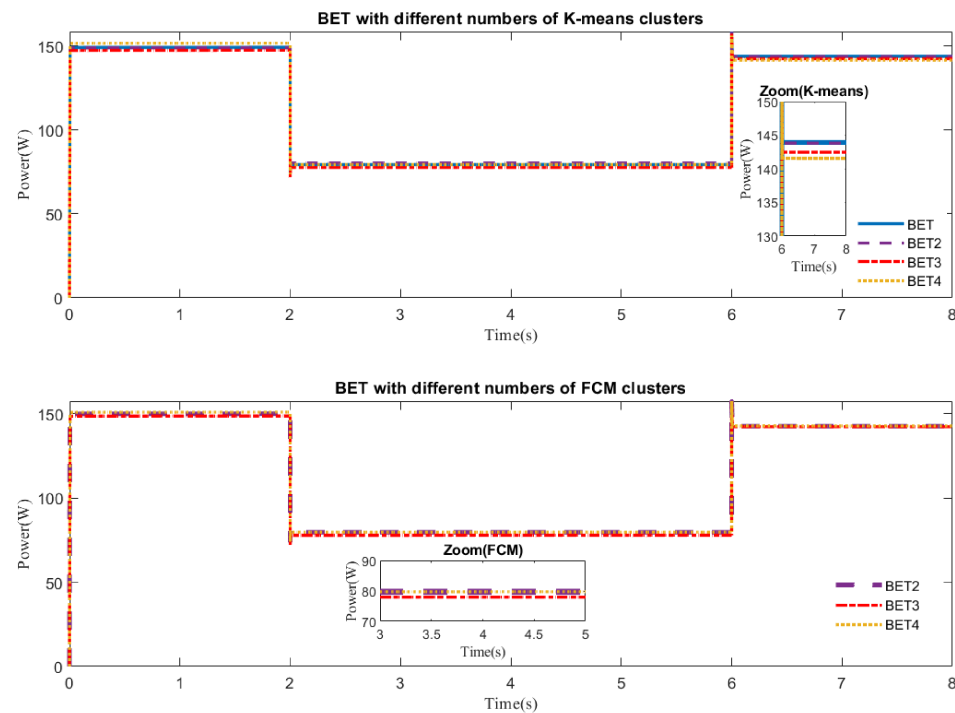


Figure 10. Zoomed-in section from Figure 9 showing the power achieved using BET-FCM and BET-K-means, respectively.

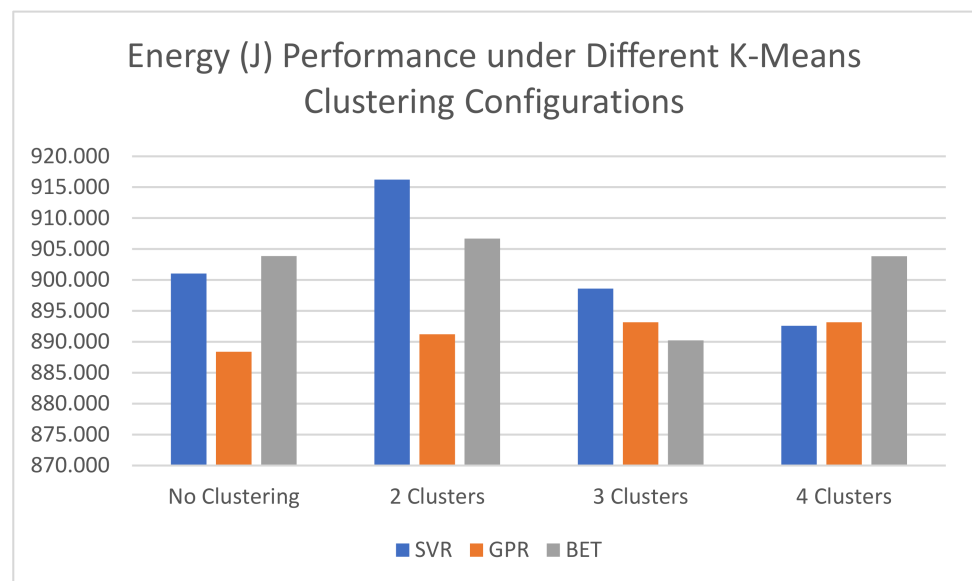


Figure 11. Total extracted photovoltaic energy for the investigated K-means configurations over the 8 s closed-loop simulation interval.

Because the three power curves almost overlap throughout most of the operating range, the differences between the investigated regression models are difficult to distinguish from Figure 7 alone. Therefore, an enlarged view of the same operating interval is presented in Figure 8.

The enlarged comparison highlights the relatively small instantaneous-power differences between the three machine-learning models. Because the local power ranking changes over the simulation interval, these curves do not establish a single model as uniformly superior in power extraction. The integrated-energy comparison in Figures 11 and 12 provides the corresponding quantitative system-level assessment.

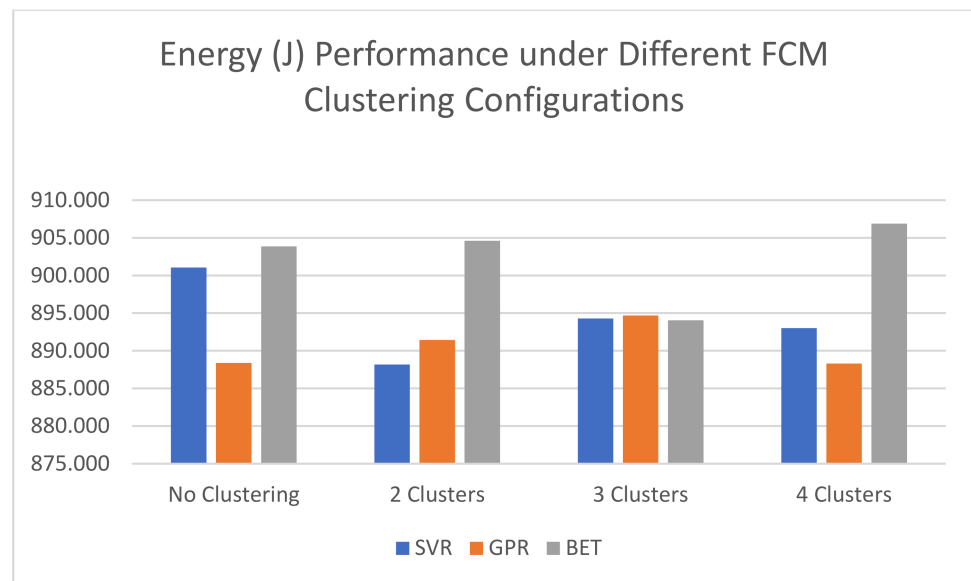


Figure 12. Total extracted photovoltaic energy for the investigated FCM configurations over the 8 s closed-loop simulation interval.

Influence of the Number of Clusters

Having identified BET as the regression technique achieving the lowest prediction errors for the selected configurations, the influence of the clustering configuration on the closed-loop MPPT performance was subsequently investigated. Figure 9 compares the generated photovoltaic power obtained using BET with two, three, and four clusters for the K-means- and FCM-based frameworks.

The generated-power trajectories for the investigated clustering configurations remain closely grouped throughout much of the dynamic simulation profile. Nevertheless, small differences are observed among the configurations, indicating that the choice of clustering method and number of clusters influences the resulting closed-loop power extraction.

Unlike the comparison presented in Figures 7, 9 and 10 are intended to examine the influence of the clustering configuration on closed-loop power extraction. At certain operating points, the three- and four-cluster configurations produce slightly higher instantaneous power than the corresponding two-cluster configuration. However, the relative ranking varies over the simulation interval, and these local differences alone are insufficient to determine the overall energy-extraction performance.

Since the differences among the investigated clustering configurations are relatively small, Figure 10 presents an enlarged view of the same operating interval to facilitate a more detailed comparison. The enlarged view confirms that the generated-power trajectories remain closely grouped, while their relative ranking varies locally over the simulation interval. Therefore, the instantaneous-power curves alone do not establish a single clustering configuration as uniformly superior in power extraction. The corresponding overall performance is assessed quantitatively using the integrated-energy results presented in the *Quantitative Energy-Extraction Analysis* subsection (Figures 11 and 12).

Engineering Interpretation of the Power Generation Results

The local variations observed in Figures 9 and 10 can be explained by considering the nonlinear relationship between the photovoltaic operating voltage and the generated electrical power. Since the predicted reference voltage is supplied directly to the PID controller, deviations between the estimated reference voltage $\tilde{V}_{ref}$ and the actual (V_{MPP})

can cause the photovoltaic module to operate away from its optimum operating point, thereby affecting the generated output power.

However, the resulting difference in generated power is not determined solely by the magnitude of the voltage-estimation error. It also depends on the local characteristics of the photovoltaic power–voltage (P–V) curve. In operating regions where the P–V curve is relatively flat, comparatively large voltage-estimation errors may produce only small changes in generated power. Conversely, in regions where the P–V characteristic exhibits a steeper gradient, smaller deviations from the optimum operating voltage may produce more noticeable changes in power. Furthermore, the resulting power depends on whether the estimated reference voltage lies above or below the actual V_{MPP} , since the two sides of the P–V curve exhibit different characteristics.

Consequently, the instantaneous power generated at a particular operating point does not necessarily follow the same ranking as the statistical prediction accuracy measured by MAE. This explains why configurations with higher overall MAE may occasionally produce slightly higher instantaneous power over limited operating intervals. Such behavior arises because the effect of a reference-voltage prediction error on generated power depends not only on its magnitude but also on its direction and on the local characteristics of the P–V curve. Accordingly, prediction-error metrics and closed-loop power-extraction performance provide complementary rather than interchangeable measures of the proposed framework.

Quantitative Energy-Extraction Analysis

The preceding instantaneous-power comparisons indicate that the ranking of the investigated configurations in terms of generated power does not necessarily follow their ranking in terms of reference-voltage prediction error. To quantify this effect over the complete 8 s closed-loop dynamic simulation profile, the generated-power trajectories were integrated over the simulation interval to determine the total electrical energy extracted by each configuration, according to

$$E_{PV} = \int_0^8 P_{PV}(t) dt \quad (7)$$

For each regression algorithm, the corresponding configuration without clustering was used as the baseline, and the relative change in extracted energy was calculated as

$$\Delta E(\%) = \frac{E_{clustered} - E_{baseline}}{E_{baseline}} \times 100 \quad (8)$$

This relative measure is used to quantify the effect of introducing clustering and should not be interpreted as absolute MPPT efficiency, since the theoretical maximum available photovoltaic energy is not used as its denominator. The resulting energy values are therefore used as a comparative system-level measure under the common dynamic simulation profile and are not intended to represent long-term photovoltaic energy yield.

At the prediction stage, the two-cluster K-means–BET configuration was identified as the preferred candidate based on the combined consideration of MAE, AB, cluster validity, and the number of local regression models required. It was therefore anticipated that its improved V_{MPP} estimation would translate into correspondingly improved energy extraction when employed as the reference-voltage generator in the closed-loop MPPT system. The subsequent system-level evaluation, however, demonstrates that the ranking based on extracted energy does not exactly follow the ranking based on voltage-prediction accuracy. For the K-means configurations, as seen in Figure 11, the two-cluster BET model extracted 906.683 J over the 8 s interval, representing a 0.311% increase relative to the no-clustering BET baseline of 903.874 J. However, the largest relative energy improvement

was obtained using the two-cluster K-means–SVR configuration, which extracted 916.237 J compared with 901.052 J for the corresponding no-clustering SVR baseline, representing an increase of 1.685%.

The FCM configurations exhibited a similarly non-monotonic relationship between prediction accuracy and energy extraction, as shown in Figure 12. Among the FCM configurations, the largest relative energy improvement was obtained using the three-cluster GPR configuration, which extracted 894.700 J compared with 888.396 J for the corresponding no-clustering GPR baseline, representing an increase of approximately 0.710%.

These findings confirm that improved V_{MPP} prediction accuracy does not necessarily translate proportionally into increased photovoltaic energy extraction. The resulting energy depends not only on the magnitude of the reference-voltage prediction error but also on its direction, the local characteristics of the nonlinear P–V curve, and the dynamic response of the closed-loop converter and controller. Consequently, MAE and AB characterize the performance of the machine-learning reference-voltage estimator, whereas integrated extracted energy provides a direct system-level measure of the resulting MPPT performance. The two levels of evaluation should therefore be considered complementary rather than interchangeable.

Collectively, the closed-loop simulation results demonstrate that the proposed learning configurations can effectively generate reference voltages for photovoltaic power tracking, while also showing that improvements in voltage-prediction metrics do not translate uniformly into improvements in extracted energy. The quantitative energy analysis therefore provides an important system-level complement to the statistical prediction analysis and enables the investigated configurations to be evaluated directly in terms of their resulting photovoltaic energy extraction.

The proposed framework belongs to the growing class of prediction-based MPPT techniques, in which the optimum operating point is estimated using data-driven models rather than determined solely through iterative perturbation. Recent studies have evaluated multiple machine-learning predictors for maximum-power estimation and compared classical and artificial-intelligence MPPT strategies under common simulation frameworks [46,47]. The present results show that operating-space partitioning can improve reference-voltage prediction for the evaluated dataset, while the quantitative energy analysis demonstrates that prediction-level improvements and system-level energy extraction must be assessed separately.

4.3. Practical Implications and Study Limitations

Beyond evaluating prediction accuracy and photovoltaic power extraction, the obtained results provide several important engineering insights regarding the application of hybrid machine-learning techniques to intelligent MPPT.

The principal contribution of the proposed framework lies in combining unsupervised clustering with supervised regression to simplify the estimation of the optimum operating voltage. Rather than approximating the complete nonlinear photovoltaic operating space using a single global regression model, the proposed methodology first partitions the operating conditions into smaller and more homogeneous regions before constructing localized regression models. The obtained results demonstrate that this decomposition improves prediction accuracy over the operating conditions represented by the evaluated dataset.

An important characteristic of the proposed framework is the separation between offline model training and online inference. The machine-learning models are trained offline using historical photovoltaic operating data, whereas online operation requires cluster identification followed by evaluation of the corresponding trained regression model to estimate the optimum reference voltage. This architecture avoids performing the model-training

stage during online MPPT operation. However, the present study did not quantitatively evaluate online execution time, memory requirements, model size, processor utilization, or other hardware-dependent computational metrics. Consequently, the feasibility of implementing the proposed framework on resource-constrained embedded PV controllers cannot be established from the present simulation-based results alone. Quantitative computational profiling on representative embedded hardware is therefore required as part of future implementation and hardware-in-the-loop validation.

The additional robustness experiment applies a fixed 70–30% training–test partition, training-only standardization and clustering, and five-fold cross-validated hyperparameter optimization using training observations only. The same test observations are used for all 21 configurations. Nevertheless, this remains a single hold-out split from a relatively small dataset; repeated independent splits and additional external datasets would better characterize variability and generalization.

The closed-loop results show that reference-voltage prediction accuracy and photovoltaic energy extraction are related but not directly proportional. The integrated-energy analysis demonstrates that the configuration with the lowest prediction error does not necessarily produce the largest extracted energy over the dynamic simulation profile. Consequently, prediction metrics such as MAE and AB and system-level energy extraction are treated as complementary performance measures.

Another noteworthy characteristic of the proposed framework is its modular architecture. Since the clustering stage, regression model, and control system operate as independent functional modules, alternative clustering algorithms, regression techniques, or feedback controllers may be incorporated without fundamentally modifying the remaining components of the MPPT system. This modular structure provides a suitable basis for investigating alternative machine-learning and control techniques in future work.

The initial evaluation demonstrated that incorporating K-means and FCM clustering improved prediction accuracy for selected regression configurations, with BET achieving particularly favorable results. However, the strict hold-out evaluation revealed that these improvements were dependent on the regression algorithm and clustering configuration. Specifically, SVR without clustering achieved the lowest overall MAE, while clustering improved BET performance under selected configurations. The closed-loop MATLAB/Simulink energy results presented in Section 4.2 correspond to the models developed under the initial evaluation procedure and should not be interpreted as evidence of improved energy extraction by the models developed under the strict hold-out procedure.

The comparative analysis in the present study is primarily intended to evaluate the effect of introducing unsupervised clustering before supervised regression and to identify the most suitable configuration within the proposed framework. Accordingly, the quantitative comparisons were performed among the investigated regression-only, K-means-based, and FCM-based configurations under the same dataset and simulation conditions. Direct quantitative benchmarking against conventional MPPT techniques such as Perturb and Observe (P&O), Incremental Conductance (INC), or other intelligent approaches such as reinforcement-learning-based MPPT was not performed under identical operating conditions. Therefore, the results presented in this study should not be interpreted as demonstrating universal superiority of the proposed framework over these alternative MPPT techniques. A systematic benchmark under identical PV models, environmental profiles, converter conditions, and evaluation criteria represents an important direction for future investigation, particularly in light of recent comparative studies spanning classical and AI-based MPPT methods [47] and recent DRL-based partial-shading approaches [53].

Despite the encouraging results obtained throughout this investigation, several limitations should be acknowledged. First, the proposed framework was trained and evaluated

using an experimental photovoltaic dataset collected from a single geographical location and a single PV module technology. Although the dataset covers a full annual cycle and a wide range of irradiance and temperature conditions, the results demonstrate the performance of the proposed framework only within the operating domain represented by the investigated dataset. Therefore, the present results should not be interpreted as demonstrating universal generalization across different climates, geographical locations, or PV technologies. Further validation using datasets acquired from different climatic regions and PV module technologies is required to establish the broader generalization capability of the proposed framework.

Another limitation concerns the sensitivity of the proposed framework to cluster assignment. For K-means, the selection of the corresponding local regression model depends on the Euclidean distance between the current operating point and the learned cluster centroids. Consequently, operating points located near cluster boundaries may be sensitive to variations in irradiance and temperature, potentially affecting the selection of the local regression model and the resulting reference-voltage estimate. Similarly, the fuzzy membership degrees in FCM may become less distinct for observations located near overlapping operating regions. The present study did not explicitly quantify the influence of cluster-assignment uncertainty on reference-voltage prediction or closed-loop MPPT performance, nor was a dedicated fault-tolerant mechanism investigated. Future work should therefore examine cluster-boundary sensitivity and appropriate mitigation strategies, particularly under measurement noise and uncertainty.

A methodological distinction is necessary between the initial prediction evaluation and the additional robustness assessment. In the original clustering-based protocol, unsupervised clustering used predictor values from the full dataset before within-cluster training–testing splits, potentially allowing test-predictor information to influence cluster structure. To address this limitation, the additional strict hold-out experiment learned all cluster centers and preprocessing parameters from training observations only and assigned unseen test observations to fixed training-derived clusters. The new results are therefore reported separately and should not be numerically conflated with the initial protocol or its closed-loop simulations.

In addition, the present study did not systematically investigate the influence of PID tuning, controller sampling time, duty-cycle constraints, anti-windup implementation, or switching-frequency selection on the closed-loop response; these controller-level implementation aspects should be examined as part of future hardware-oriented validation.

Finally, the present investigation employed an experimentally acquired photovoltaic dataset for model development and evaluation, while the closed-loop MPPT system was implemented and evaluated in MATLAB/Simulink. Therefore, the closed-loop results presented in this study constitute simulation-based validation rather than hardware-in-the-loop or physical experimental validation. Consequently, practical implementation effects such as converter non-idealities, sensor noise and uncertainties, computational latency, and hardware-related constraints were not directly evaluated in the present study. Future work will therefore focus on hardware-in-the-loop implementation followed by experimental validation using a physical photovoltaic system to quantify the influence of these practical factors and further assess the applicability of the proposed framework under real operating conditions [56].

Furthermore, the present work considered two clustering algorithms and three supervised regression techniques. Although these algorithms represent well-established machine learning methods, future investigations may extend the proposed framework by considering adaptive clustering techniques, adaptive optimization algorithms, deep

learning, reinforcement learning, and other intelligent MPPT strategies that have recently demonstrated promising performance under dynamic operating conditions [50,51,53].

Finally, the current study evaluates the proposed framework under operating conditions characterized by variations in irradiance and temperature without partial-shading conditions. Partial shading can produce multiple local maxima in the photovoltaic P–V characteristic and therefore represents a substantially more challenging MPPT scenario. Such conditions were not represented in the dataset or evaluated in the closed-loop simulations performed in this study. Consequently, the present results should not be interpreted as demonstrating the effectiveness of the proposed framework under partial-shading conditions. Evaluation under partial shading is therefore identified as an important direction for future work.

In addition, the present study assumes the measured irradiance and temperature inputs to be representative of the actual operating conditions and does not explicitly evaluate robustness against measurement noise, sensor ageing, communication delays, or long-term PV module degradation. These effects may introduce deviations between the measured inputs and the actual PV operating conditions and may consequently affect the accuracy of the machine-learning-based reference-voltage estimation. Therefore, systematic robustness analysis under sensor uncertainty, measurement noise, communication delays, and PV module degradation remains an important subject for future investigation.

Despite these limitations, the results support the value of operating-space partitioning followed by local regression for the dataset and simulation conditions evaluated in this study. Broader practical claims require additional validation, including computational profiling, hardware-in-the-loop testing, and physical implementation. Recent HIL-based MPPT studies illustrate the type of validation needed to bridge this gap [56].

5. Conclusions

This study investigated a two-stage photovoltaic MPPT framework combining K-means or FCM clustering of irradiance–temperature operating conditions with local SVR, GPR, or BET regression for reference-voltage estimation. The predicted reference was used by a PID-controlled boost converter in the original MATLAB/Simulink closed-loop evaluation. The contribution lies in comparing prediction-based operating-region decomposition with regression-only alternatives and examining the distinction between prediction error and extracted energy.

Under the initial prediction evaluation, the two-cluster K-means–BET configuration exhibited the strongest K-means MAE performance and was selected for the original system-level investigation. An additional strict hold-out evaluation, using training-only clustering and identical independent test observations for all 21 configurations, yielded a different ranking: unclustered SVR attained the lowest overall MAE (0.724 V). Clustering nevertheless benefited BET in selected cases; FCM with three clusters reduced BET MAE from 0.934 V to 0.843 V, an improvement of approximately 9.7%. Thus, clustering is a model-dependent design option rather than a universally superior preprocessing step.

The trained models were subsequently evaluated within a closed-loop MATLAB/Simulink MPPT system under an 8 s dynamic profile containing step changes in irradiance, temperature, and load resistance. Integrated energy extraction provided a direct system-level complement to the voltage-prediction metrics. The energy ranking did not exactly follow the MAE/AB ranking: the two-cluster K-means–BET configuration extracted 906.683 J, representing a 0.311% increase relative to its no-clustering BET baseline, whereas the two-cluster K-means–SVR configuration produced the largest relative energy increase among the investigated K-means configurations, extracting 916.237 J, or 1.685% above its cor-

responding no-clustering baseline. These results confirm that prediction accuracy and extracted energy are related but are not interchangeable performance measures.

Together, the results emphasize that predictive accuracy, operating-region partitioning, and simulated energy extraction must be evaluated separately. The closed-loop energy findings apply to the initially trained and simulated models; the additional strict hold-out models were not rerun through the closed-loop system. The present results therefore do not establish corresponding energy gains for the new strict hold-out configurations or demonstrate hardware performance.

Future work should investigate repeated and external hold-out evaluations across climates and PV technologies, examine cluster-boundary sensitivity and partial shading, and independently assess the strict hold-out models within closed-loop simulation. Hardware-in-the-loop and physical implementation studies should also quantify computational requirements, measurement uncertainty, tracking dynamics, and real-world energy extraction.

Author Contributions: Conceptualization, M.T., M.F., O.B. and A.A.-Q.; methodology, M.T., M.F., O.B., M.R., A.A.-Q., G.A., M.A.A., R.A.N., N.A. and M.A.; software, M.T., M.F., M.R., G.A., M.A.A., R.A.N., N.A. and M.A.; validation, M.T., M.F., O.B., M.R., A.A.-Q., G.A., M.A.A., R.A.N., N.A. and M.A.; formal analysis, M.T., M.F., O.B., M.R., A.A.-Q., G.A., M.A.A., R.A.N., N.A. and M.A.; investigation, M.T., M.F., O.B., M.R., A.A.-Q., G.A., M.A.A., R.A.N., N.A. and M.A.; resources, M.F. and O.B.; writing—original draft preparation, M.T., M.F., G.A., M.A.A., R.A.N., N.A. and M.A.; writing—review and editing, M.T., M.F., O.B., M.R. and A.A.-Q.; visualization, M.T., M.F., M.R., G.A., M.A.A., R.A.N., N.A., M.A.; supervision, M.T., M.F. and O.B.; project administration, M.T., M.F. and O.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Praharaj, C. Photovoltaics and Energy Conversion Devices. In *Group III-Nitride Semiconductor Optoelectronics*, 1st ed.; Wiley: Hoboken, NJ, USA, 2023; pp. 129–145. [\[CrossRef\]](#)
2. Koev, K. An Investigation of the Energy Efficiency of a Grid Connected Photovoltaic System. In Proceedings of the 2022 8th International Conference on Energy Efficiency and Agricultural Engineering (EE&AE), Ruse, Bulgaria, 30 June–2 July 2022; pp. 1–4. [\[CrossRef\]](#)
3. Deghfel, N.; Badoud, A.E.; Merah, F.; Bajaj, M.; Zaitsev, I. A new intelligently optimized model reference adaptive controller using GA and WOA-based MPPT techniques for photovoltaic systems. *Sci. Rep.* **2024**, *14*, 6827. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Mazumdar, D.; Sain, C.; Biswas, P.K.; Sanjeevikumar, P.; Khan, B. Overview of Solar Photovoltaic MPPT Methods: A State of the Art on Conventional and Artificial Intelligence Control Techniques. *Int. Trans. Electr. Energy Syst.* **2024**, *2024*, e8363342. [\[CrossRef\]](#)
5. Takruri, M.; Farhat, M.; Barambones, O.; Ramos-Hernanz, J.A.; Turkieh, M.J.; Badawi, M.; AlZoubi, H.; Sakur, M.A. Maximum Power Point Tracking of PV System Based on Machine Learning. *Energies* **2020**, *13*, 692. [\[CrossRef\]](#)
6. 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*, 907–920. [\[CrossRef\]](#)
7. Mokhlis, M.; Ferfra, M.; Vall, H.A.; Idrissi, R.E.; Ahmed, C.C.; Taouni, A. Comparative Study Between the Different MPPT Techniques. In Proceedings of the 2020 5th International Conference on Renewable Energies for Developing Countries (REDEC), Marrakech, Morocco, 29–30 June 2020; pp. 1–6. [\[CrossRef\]](#)
8. Nkambule, M.S.; Hasan, A.N.; Ali, A.; Hong, J.; Geem, Z.W. Comprehensive Evaluation of Machine Learning MPPT Algorithms for a PV System Under Different Weather Conditions. *J. Electr. Eng. Technol.* **2021**, *16*, 411–427. [\[CrossRef\]](#)

9. Mukherjee, D.; Chakraborty, S.; Guchhait, P.K.; Bhunia, J. Machine Learning Based Solar Power Generation Forecasting with and Without MPPT Controller. In Proceedings of the 2020 IEEE 1st International Conference for Convergence in Engineering (ICCE), Kolkata, India, 4–6 September 2020; pp. 44–48. [\[CrossRef\]](#)
10. Phan, B.C.; Lai, Y.-C.; Lin, C.E. A Deep Reinforcement Learning-Based MPPT Control for PV Systems Under Partial Shading Condition. *Sensors* **2020**, *20*, 3039. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Takruri, M.; Farhat, M.; Sunil, S.; Ramos-Hernanz, J.A.; Barambones, O. Support Vector Machine for Photovoltaic System Efficiency Improvement. *J. Sustain. Dev. Energy Water Environ. Syst.* **2020**, *8*, 441–451. [\[CrossRef\]](#)
12. Tina, G.M.; Ventura, C.; Ferlito, S.; De Vito, S. A State-of-Art-Review on Machine-Learning Based Methods for PV. *Appl. Sci.* **2021**, *11*, 7550. [\[CrossRef\]](#)
13. Mishra, S.K.; Mohanty, D.; Sahu, J.K.; Mohanty, S. Artificial Neural Network Based MPPT Controller for Stand-Alone Solar PV System. In Proceedings of the 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), Puducherry, India, 30–31 July 2021; pp. 1–6. [\[CrossRef\]](#)
14. Mahesh, P.V.; Meyyappan, S.; Alla, R.R. Maximum power point tracking using decision-tree machine-learning algorithm for photovoltaic systems. *Clean Energy* **2022**, *6*, 762–775. [\[CrossRef\]](#)
15. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [\[CrossRef\]](#)
16. Zhang, F.; O'Donnell, L.J. Chapter 7—Support vector regression. In *Machine Learning*; Mechelli, A., Vieira, S., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 123–140. [\[CrossRef\]](#)
17. Awad, M.; Khanna, R. Support Vector Regression. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*; Awad, M., Khanna, R., Eds.; Apress: Berkeley, CA, USA, 2015; pp. 67–80. [\[CrossRef\]](#)
18. Schulz, E.; Speekenbrink, M.; Krause, A. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *J. Math. Psychol.* **2018**, *85*, 1–16. [\[CrossRef\]](#)
19. Takruri, M.; Abubakar, A.; Jallad, A.-H.; Altawil, B.; Marpu, P.R.; Bermak, A. Machine Learning-Based Estimation of PM2.5 Concentration Using Ground Surface DoFP Polarimeters. *IEEE Access* **2022**, *10*, 23489–23496. [\[CrossRef\]](#)
20. Dietterich, T.G. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. *Mach. Learn.* **2000**, *40*, 139–157. [\[CrossRef\]](#)
21. Kotsiantis, S.B.; Tsekouras, G.E.; Pintelas, P.E. Bagging Model Trees for Classification Problems. In *Advances in Informatics*; Bozani, P., Houstis, E.N., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 328–337. [\[CrossRef\]](#)
22. Altman, N.; Krzywinski, M. Ensemble methods: Bagging and random forests. *Nat. Methods* **2017**, *14*, 933–935. [\[CrossRef\]](#)
23. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [\[CrossRef\]](#)
24. Guessoum, D.; Takruri, M.; Rabi, M.; Farhat, M.; Badawi, S.A. K-Nearest Neighbors hybrid method for maximum power point tracking under partial shading for photovoltaic power systems. *Results Eng.* **2025**, *27*, 106694. [\[CrossRef\]](#)
25. Hohm, D.P.; Ropp, M.E. Comparative study of maximum power point tracking algorithms. *Prog. Photovolt. Res. Appl.* **2003**, *11*, 47–62. [\[CrossRef\]](#)
26. Femia, N.; Petrone, G.; Spagnuolo, G.; Vitelli, M. Optimization of perturb and observe maximum power point tracking method. *IEEE Trans. Power Electron.* **2005**, *20*, 963–973. [\[CrossRef\]](#)
27. Abdel-Salam, M.; El-Mohandes, M.-T.; Goda, M. An improved perturb-and-observe based MPPT method for PV systems under varying irradiation levels. *Sol. Energy* **2018**, *171*, 547–561. [\[CrossRef\]](#)
28. Ali, A.I.M.; Sayed, M.A.; Mohamed, E.E.M. Modified efficient perturb and observe maximum power point tracking technique for grid-tied PV system. *Int. J. Electr. Power Energy Syst.* **2018**, *99*, 192–202. [\[CrossRef\]](#)
29. Ghamrawi, A.; Gaubert, J.-P.; Mehdi, D. A new dual-mode maximum power point tracking algorithm based on the Perturb and Observe algorithm used on solar energy system. *Sol. Energy* **2018**, *174*, 508–514. [\[CrossRef\]](#)
30. Ghamrawi, A.; Gaubert, J.-P.; Mehdi, D. A New Dual-Mode MPPT Algorithm Applied to a Quadratic Converter in a Solar Energy System. In Proceedings of the 2020 22nd European Conference on Power Electronics and Applications (EPE'20 ECCE Europe), Lyon, France, 7–11 September 2020; pp. 1–9. [\[CrossRef\]](#)
31. Ramani, S.U.; Kollimalla, S.K.; Arundhati, B. Comparative study of P&O; incremental conductance method for PV system. In Proceedings of the 2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT), Kerala, India, 20–21 April 2017; pp. 1–7. [\[CrossRef\]](#)
32. Hussein, K.H.; Muta, I.; Hoshino, T.; Osakada, M. Maximum photovoltaic power tracking: An algorithm for rapidly changing atmospheric conditions. *IEE Proc. Gener. Transm. Distrib.* **1995**, *142*, 59–64. [\[CrossRef\]](#)
33. Mao, M.; Cui, L.; Zhang, Q.; Guo, K.; Zhou, L.; Huang, H. Classification and summarization of solar photovoltaic MPPT techniques: A review based on traditional and intelligent control strategies. *Energy Rep.* **2020**, *6*, 1312–1327. [\[CrossRef\]](#)
34. Djellad, A.; Bouderrès, N.; Hamdi, B.A.; Chiheb, S.; Logerais, P.O. A Comprehensive Comparative Study and Optimization of MPPT Algorithms in a Standalone Photovoltaic Plant Using Classical and Machine Learning Techniques. In Proceedings of the 2023 2nd International Conference on Electronics, Energy and Measurement (IC2EM), Medea, Algeria, 28–30 November 2023; pp. 1–6. [\[CrossRef\]](#)

35. Bouarroudj, N.; Boukhetala, D.; Feliu-Batlle, V.; Boudjema, F.; Benlahbib, B.; Batoun, B. Maximum Power Point Tracker Based on Fuzzy Adaptive Radial Basis Function Neural Network for PV-System. *Energies* **2019**, *12*, 2827. [\[CrossRef\]](#)
36. Yu, B.; Abo-Khalil, A.G.; So, J.; Yu, G. Support vector regression based maximum power point tracking for PV grid-connected system. In Proceedings of the 2009 34th IEEE Photovoltaic Specialists Conference (PVSC), Philadelphia, PA, USA, 7–12 June 2009; pp. 2037–2042. [\[CrossRef\]](#)
37. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.
38. Inamdar, S.; Chakravarty, S.; Shi, Y.; Das, S. Improving Maximum Power Point Tracking Efficiency for Partially Shaded PV Arrays by Nested Zone Prediction via Machine Learning. In Proceedings of the 2020 47th IEEE Photovoltaic Specialists Conference (PVSC), Virtual, 16–21 June 2020; pp. 2795–2798. [\[CrossRef\]](#)
39. Shareef, H.; Mutlag, A.H.; Mohamed, A. Random Forest-Based Approach for Maximum Power Point Tracking of Photovoltaic Systems Operating under Actual Environmental Conditions. *Comput. Intell. Neurosci.* **2017**, *2017*, 1673864. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Mojeed, F.A. Comparative Study of Different Photovoltaic MPPT Techniques Under Various Weather Conditions. Ph.D. Thesis, University of Johannesburg, Johannesburg, South Africa, 2017. Available online: <https://www.proquest.com/docview/2525583081/abstract/B8672575FE9F4A2CPQ/1> (accessed on 22 April 2024).
41. Farayola, A.M.; Sun, Y.; Ali, A. Optimization of PV Systems Using Linear Interactions Regression MPPT Techniques. In Proceedings of the 2018 IEEE PES/IAS PowerAfrica, Cape Town, South Africa, 26–29 June 2018; pp. 545–550. [\[CrossRef\]](#)
42. Omer, Z.M.; Shareef, H. Adaptive Boosting and Bootstrapped Aggregation Based Ensemble Machine Learning Methods for Photovoltaic Systems Output Current Prediction. In Proceedings of the 2019 29th Australasian Universities Power Engineering Conference (AUPEC), Nadi, Fiji, 26–29 November 2019; pp. 1–6. [\[CrossRef\]](#)
43. Huang, C.; Wang, L.; Long, H.; Luo, X.; Wang, J.-H. A hybrid global maximum power point tracking method for photovoltaic arrays under partial shading conditions. *Optik* **2019**, *180*, 665–674. [\[CrossRef\]](#)
44. Achouri, F.; Damou, M.; Harrou, F.; Sun, Y.; Bouyeddou, B. Gaussian Processes for Efficient Photovoltaic Power Prediction. In Proceedings of the 2023 International Conference on Decision Aid Sciences and Applications (DASA), Annaba, Algeria, 16–17 September 2023; pp. 290–295. [\[CrossRef\]](#)
45. Yılmaz, M.; Çorapsız, M.R.; Çorapsız, M.F. A novel maximum power point tracking approach based on fuzzy logic control and optimizable Gaussian Process Regression for solar systems under complex environment conditions. *Eng. Appl. Artif. Intell.* **2025**, *141*, 109780. [\[CrossRef\]](#)
46. Hamad, S.A.; Ghalib, M.A.; Munshi, A.; Alotaibi, M.; Ebied, M.A. Evaluating machine learning models comprehensively for predicting maximum power from photovoltaic systems. *Sci. Rep.* **2025**, *15*, 10750. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Sousa, J.T.; Barbosa, R.S. Comparison of Classical and Artificial Intelligence Algorithms to the Optimization of Photovoltaic Panels Using MPPT. *Algorithms* **2025**, *18*, 493. [\[CrossRef\]](#)
48. Liao, C.-C. Genetic k-means algorithm based RBF network for photovoltaic MPP prediction. *Energy* **2010**, *35*, 529–536. [\[CrossRef\]](#)
49. Guessoum, D.; Takruri, M.; Badawi, S.A.; Farhat, M.; ElBadawi, I. Maximum power point tracking using unsupervised learning for photovoltaic power systems. *Int. J. Sustain. Eng.* **2024**, *17*, 397–412. [\[CrossRef\]](#)
50. Lin, D.; Li, X.; Ding, S.; Wen, H.; Du, Y.; Xiao, W. Self-Tuning MPPT Scheme Based on Reinforcement Learning and Beta Parameter in Photovoltaic Power Systems. *IEEE Trans. Power Electron.* **2021**, *36*, 13826–13838. [\[CrossRef\]](#)
51. Avila, L.; De Paula, M.; Carlucho, I.; Sanchez Reinoso, C. MPPT for PV systems using deep reinforcement learning algorithms. *IEEE Lat. Am. Trans.* **2019**, *17*, 2020–2027. [\[CrossRef\]](#)
52. Kofinas, P.; Doltsinis, S.; Dounis, A.I.; Vouros, G.A. A reinforcement learning approach for MPPT control method of photovoltaic sources. *Renew. Energy* **2017**, *108*, 461–473. [\[CrossRef\]](#)
53. Big-Alabo, A. Maximum power point tracking of solar-PV under partial shading and variable load resistance using deep reinforcement learning algorithm. *Discov. Electron.* **2026**, *3*, 41. [\[CrossRef\]](#)
54. Ramos-Hernanz, J.A.; Barambones, O.; Lopez-Guede, J.M.; Zamora, I.; Eguia, P.; Farhat, M. Sliding Mode Real-Time Control of Photovoltaic Systems Using Neural Estimators. *Int. J. Photoenergy* **2016**, *2016*, 5214061. [\[CrossRef\]](#)
55. Hassan, M.A.; Khalil, A.; Kaseb, S.; Kassem, M.A. Exploring the potential of tree-based ensemble methods in solar radiation modeling. *Appl. Energy* **2017**, *203*, 897–916. [\[CrossRef\]](#)
56. Zeb, O.; Ahmad, I. Advanced MPPT analysis of PV systems: A comparative study of robust nonlinear controllers, model-free controllers, and machine learning techniques. *Renew. Energy* **2026**, *270*, 125950. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.