Maximizing solar photovoltaic system efficiency by multivariate linear regression based maximum power point tracking using machine learning

Introduction. In recent times, there has been a growing popularity of photovoltaic (PV) systems, primarily due to their numerous advantages in the field of renewable energy. One crucial and challenging task in PV systems is tracking the maximum power point (MPP), which is essential for enhancing their efficiency. One of the main challenges is that PV systems face two main challenges. Firstly, they exhibit low efficiency in generating electric power, particularly in situations of low irradiation. Secondly, there is a strong connection between the power output of solar arrays and the constantly changing weather conditions. This interdependence can lead to load mismatch, where the maximum power is not effectively extracted and delivered to the load. This problem is commonly referred to as the maximum power point tracking (MPPT) problem.

Various control methods for MPPT have been suggested to optimize the peak power output and overall generation efficiency of PV systems. This article presents a novel approach to maximize the efficiency of solar PV systems by tracking the MPP and dynamic response of the system is investigated. Originality. The technique involves a multivariate linear regression (MLR) machine learning algorithm to predict the MPP, which is necessary for irradiance level and temperature, based on data collected from the solar PV generator specifications. This information is then used to calculate the duty ratio for the boost converter. Results. MATLAB/Simulink simulations and experimental results demonstrate that this approach consistently achieves a mean efficiency of over 96 % in steady-state operation of the PV system, even under variable irradiance level and temperature. Practical value. The improved efficiency of 96 % of the proposed MLR based MPP in the steady-state operation extracting maximum from PV system, adds more value. The same is evidently proved by the hardware results. References 24, table 4, figures 14.

Key words: machine learning, maximum power point trackers, solar photovoltaic systems.

Вступ. Останнім часом зростає популярність фотоелектричних (ФЕ) систем, насамперед через їх численні переваги в галузі відновлюваної енергії. Одним з найважливіших і складних завдань у ФЕ системах є відстеження точки максимальної потужності (MPP), яка необхідна для підвищення їх ефективності. Мета. ФЕ системи стикаються із дніма основними проблемами. По-перше, вони демонструють низьку ефективність вироблення електроенергії, особливо в умовах низького випромінювання. По-друге, існує сильний зв'язок між вихідною потужністю сонячних батарей і погодними умовами, що постійно змінюються. Ця взаємозалежність може призвести до ненадійності навантаження, коли максимальна потужність не ефективно відбирається і передаватиметься в навантаження. Ця проблема відбувається після відстеження точки максимальної потужності (МРТ). Для оптимізації низької вихідної потужності та загальної ефективності генерації ФЕ системи було запропоновано різні методи керування МРТ. Методологія. У цій статті представлено новий підхід до максимізації ефективності сонячних ФЕ систем шляхом відстеження МРТ на основі динамічної реакції системи. Оригіналність. Цей підхід включає алгоритми машинного навчання багатоваріантної лінійної ресерві (MLR) для прогнозування МРТ для будь-якого рівня освітленості та температури на основі даних, зібраних зі спеціфікації сонячних ФЕ генераторів. Ця інформація потім використовується для розрахунку коефіцієнта заповнення перетворювача, що підсилює. Результати. Моделювання MATLAB/Simulink та експериментальні результати показують, що цей підхід посідає забезпечує середньо ефективність понад 96 % в режимах роботи ФЕ системи, що встановлює, наочні при змінах рівня освітленості і температурі. Практична цінність. Підвищення ефективності 96 % ТРМР на основі MLR в режимах роботи, що вистачає максимум з ФЕ системи, підсилює цінність. Те саме, очевидно, підтверджає і аргументи результати. Біблія, 24, табл. 4, рис. 14.

Ключові слова: машинне навчання, відстежування максимальної потужності, сонячні фотоелектричні системи.

Introduction. Solar photovoltaic (PV) generator energy systems have become increasingly popular as a source of renewable energy. However, one of the main challenges is, achieving maximum power extraction from the PV generator as it is typically not operated at its maximum power point (MPP) known as MPPT techniques, which aim to improve the efficiency of PV generator. The most common conventional methods for MPPT of a PV generator are Perturb & Observe (P&O) and Incremental Conductance (IC) algorithms. These methods involve adjusting the voltage of the PV generator to calculate the required change in voltage for maximum power extraction. Other methods include mathematical-based approaches like the curve-fitting algorithm, which indirectly tracks the MPP using the power-voltage curve of the panel. Constant-parameter algorithms like fractional open-circuit voltage require periodic measurement of the open-circuit voltage, while the fractional short-circuit current algorithm requires periodic measurement of the short-circuit current.

Intelligent prediction algorithms like fuzzy logic control (FLC) and artificial neural networks (ANN) can predict MPP by adjusting the weights of different layers through error-based methods like gradient descent. These algorithms are designed to operate the PV generator at the MPP to extract the maximum available power for delivery to the load.

Machine learning (ML) algorithms can predict unknown data with a high degree of accuracy by learning from known data. By training a ML algorithm with existing data and testing it with new data, a ML model is created. Typically, 75 % of the data is used for training, and the remaining 25 % for testing the model. Image-based ML and reinforcement learning algorithms have been used for MPPT in PV generator. To operate the PV generator at the MPP, a converter is required.
The literature reports the use of various types of converters, including DC-DC buck converters, boost converters, buck-boost converters, single-ended primary inductor converters, and controlled inverters.

Although the conventional P&O and IC methods are simple and require fewer sensing elements, they have a low MPPT speed for rapid changes in irradiances. Intelligent prediction algorithms like ANN and FLC can address this issue. The performance of the ANN model depends on the correlation between the training and validation data, the number of iterations used for training, and the number of layers and neurons. The accuracy of the FLC is dependent on the rule-based design, which requires human expertise and experience. The Cuckoo Search (CS) technique is considered one of the fastest and most reliable optimization techniques but has a high failure rate and high oscillations in the steady state.

Achieving fast-tracking of the MPP is crucial for efficient solar PV generator, as irradiance and temperature change rapidly. ML algorithms offer a promising solution to improve MPPT speed without requiring an iterative approach or controller. To evaluate this approach, a new multivariate linear regression (MLR) algorithm is proposed in this study, and its performance is compared to conventional techniques like P&O and IC, intelligent methods like ANN and FLC, and optimization algorithms.

The block diagram shown in Fig. 1 for a complete system, where $P_{mp}$ is maximum power available at MPP, $V_{mp}$ is the voltage of the solar PV generator at MPP, $I_{mp}$ is the current through the solar PV generator at MPP, $D$ is duty cycle, $R_{mp}$ is the resistance at MPP and $R_{s}$ is the load resistance. The mean efficiency is calculated under different irradiance level (IL) and temperature $T$ to validate the effectiveness of the MLR method.

\[ I_{pv} = I_L - I_D(e^{\frac{V + IR_s}{nV_T}} - 1) - \frac{V + I_{pv}R_s}{R_{sh}}, \tag{1} \]

where $I_{pv}$ is the solar PV generator current; $I_L$ is the photocurrent as a function of IL and T; $I_D$ is the diode saturation current; $V$ is the solar PV generator voltage; $R_s$ is the series resistance; $n$ is the diode ideal factor ($1 \leq n \leq 2$); $V_T$ is the thermal voltage equivalent; $R_{sh}$ is the shunt resistance.

Figure 3 illustrates a boost converter with pulse width modulation control, which is powered by a solar PV generator. The MOSFET switch and duty cycle ($D$) is responsible for controlling the amount of power that is delivered to the load from the solar PV generator. The inductor $L$ present in the circuit boosts the solar PV generator voltage to the required output voltage level. Additionally, the load current $I_L$ flow through the load and input and output capacitors $C_i$ and $C_o$ are utilized to minimize the ripple content in the voltages [8-10].

Fig. 1. System block diagram

**System description. Characteristics of PV generator and DC-DC boost converter.** Solar PV generator convert sunlight into electricity, and several cells are connected to form a PV generator. The one-diode equivalent circuit [7-11] of a PV generator is depicted in Fig. 2 and represented mathematically in (1). The number of solar PV generator in a panel determines the specifications for voltage, current, and power.

Fig. 2. The one-diode equivalent circuit of a PV generator

The solar panel specifications used for the simulation include a maximum power of 250 W, short-circuit current of 9.38 A, open-circuit voltage of 36 V, voltage at MPP of 28.8 V, and current at MPP of 8.68 A. The current-voltage and power-voltage characteristics of the solar PV generator under different temperature and irradiances are illustrated in Fig. 4.

**Multivariate linear regression.** The linear regression method is a simple ML technique that is suitable for predicting real numbers from available data. It works by predicting unknown data, which is also known as dependent data, from the features, which are referred to as independent data [12, 13]. If the data has a single feature, then the univariate linear regression algorithm gives a straight line that predicts the data in a two-dimensional space. On the other hand, if there are multiple features, the MLR algorithm provides a plane in multidimensional space.
space. The general form of the multiple linear regression planes [12] can be expressed as:

\[ y = \beta_0 + \beta_0 x_1 + \ldots + \beta_{n-1} x_{n-1} + \beta_n x_n \]  
(2)

where \( y \) is the data to be predicted in an \( n \)-dimensional space \( x_1, x_2, \ldots, x_{n-1}, x_n \) are the feature with \( \beta_0, \beta_0, \ldots, \beta_{n-1}, \beta_n \) as regression coefficients.

ANN-based MPP [14-18] is shown in Fig. 5 for an example of finding the duty at MPP (\( D_{mpp} \)) based on the training provided for the ANN. The results of \( D_{mpp} \) are taken as output and are used for comparisons.

**Data in linear regression.** ML algorithms acquire knowledge by analyzing data, allowing them to identify patterns, make informed decisions, and assess their level of certainty based on the information provided. The quality of the training data plays a critical role in determining the effectiveness of the model. Figure 6 indicates the learning model. Three-dimensional MLR model is shown in Fig. 7.

Learning is data refers to raw and unprocessed facts, values, texts, sounds, or images that are yet to be analyzed. It is a crucial component in the fields of ML and artificial intelligence, and without it, cannot train any models. Information, on the other hand, is data that has been interpreted and manipulated to provide final results. Knowledge is a combination of inferred information, experiences, learning, and insights that result in awareness.

**Data preprocessing. Training data.** The part of data used to train the model. This is the data that the MLR model sees (both input and output) and learns from this data. In the proposed work, 70% of data is given for training purpose and the records were chosen randomly (Fig. 8).

**Validation data.** The part of data that is used to do a frequent evaluation of the model, fits on the training dataset along with improving involved hyper parameters (initially set parameters before the model begins learning). This data plays its part when the model is training. For validation of data, only 20% of the data is given and the records were random.

![Fig. 5. Neural network example](image)

![Fig. 6. Learning model](image)

![Fig. 7. MLR model in a three-dimensional space](image)

![Fig. 8. Data preprocessing](image)

![Fig. 9. Flowchart for the proposed MMPT using MLR](image)

\[
SSE = \sum_{K=1}^{n_i} \left( Y_{A,K} - Y_{P,K} \right)^2 ;
\]  
(3)

\[
R^2 = 1 - \frac{\sum_{K=1}^{n_i} \left( Y_{A,K} - Y_{P,K} \right)^2}{\sum_{K=1}^{n_i} \left( Y_{A,K} - Y_{P,Avg} \right)^2} ;
\]  
(4)

\[
RMSE = \sqrt{\frac{1}{n_i} \sum_{K=1}^{n_i} \left( Y_{A,K} - Y_{P,K} \right)^2} ,
\]  
(5)

where \( Y_A \) represents the actual data; \( Y_P \) is the predicted data; \( n_i \) is the number of samples; \( Y_{Avg} \) is the average
values of \( Y_t \). The value of \( R^2 \in [0, 1] \) specifies the prediction strength of models, and an \( R^2 \) value closer to 1 ensures the best fit of the model. Likewise, the SSE and RMSE values measure the residual or error among \( Y_t \) and \( Y_r \). Therefore, SSE and RMSE values closer to 0 represent the models’ superior prediction.

In the proposed methodology, the third stage involved using the MLR model to perform MPPT. The MLR model predicted the maximum power available at MPP (\( P_{mp} \)) and the voltage of the solar PV generator at MPP (\( V_{mp} \)) for a given IL and temperature \( T \). The predictions were used to determine the required \( D \) for the boost converter to operate the PV generator at MPP. The corresponding resistance at MPP (\( R_{mp} \)) was computed using these predicted values as in (6). The \( R_{mp} \) was reflected between nodes of boost converter by controlling the \( D \) of the boost converter. The \( D \) in terms of \( R_{mp} \) and load resistance \( R_0 \) is given in (7):

\[
R_{mp} = V_{mp}^2 / P_{mp} ;
\]

\[
D = 1 - \sqrt{R_{mp} / R_0} .
\]  

The maximum and minimum values of the load resistance were determined using the method proposed in [8]. The boost converter is designed using the procedure explained in [7]. The required boost converter inductance \( L \) and capacitance \( C \) are as follows:

\[
L = V_{inp} \cdot (V_{out} - V_{inp}) / f_{sw} \cdot A V \cdot I_{out} ;
\]

\[
C = I_{out} \cdot (V_{out} - V_{inp}) / f_{sw} \cdot AV \cdot V_{out} ;
\]

where \( V_{inp} \) is the input voltage; \( V_{out} \) is the output voltage; \( f_{sw} \) is the switching frequency; \( AV \) is the current ripple; \( I_{out} \) is the output current; \( AV \) is the voltage ripple.

The fourth step of the methodology involved a comparative analysis of the MLR methodology with existing conventional, intelligent, and optimization MPPT methods.

**Simulation results and discussion. Data collection.**
The simulated dynamic result for the IL changed from 900 to 500 W/m\(^2\) is shown in Fig. 10. In that corresponding solar power, voltage, and current were demonstrated that the maximum power can track using the proposed method.

The data collected for this study includes four variables: \( I_r, T, P_{mp, r} \) and \( V_{mp, r} \). The values of \( P_{mp} \) and \( V_{mp} \) depend on \( I_r \) and \( T \).

![Fig. 10. Simulation results of \( V_{mp, r}, I_{mp, r}, P_{mp, r} \) and \( D \) for change in \( I_r \) from 900 to 300 W/m\(^2\), \( T = 25 \^\circ\)C](image)

To predict \( P_{mp} \) and \( V_{mp} \), \( I_r \) and \( T \) are used as features. The MPP of changes in variables for the installed roof solar PV generator and its specification of 250 W Zy-TECH 250P [19-21] are given in Table 1.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rated power, W</td>
<td>250</td>
</tr>
<tr>
<td>Voltage at maximum power, V</td>
<td>28.8</td>
</tr>
<tr>
<td>Current at maximum power, A</td>
<td>8.68</td>
</tr>
<tr>
<td>Open circuit voltage, V</td>
<td>36</td>
</tr>
<tr>
<td>Short-circuit current, A</td>
<td>9.38</td>
</tr>
<tr>
<td>Voltage temperature coefficient</td>
<td>–0.36901</td>
</tr>
<tr>
<td>Current temperature coefficient</td>
<td>0.086988</td>
</tr>
</tbody>
</table>

**Performance of the proposed MLR model.** The MLR machine learning models created using MATLAB/Simulink involves two independent and one dependent variable. These models can predict the values of \( P_{mp} \) and \( V_{mp} \) based on specific values of \( I_r \) and \( T \). The data were collected as described earlier, based on the specification of the PV generator. The MLR model developed is presented mathematically in (10) and (11):

\[
P_{mp} = 0.8994 + 0.01001 \cdot I_r - 0.03685 - T ;
\]

\[
V_{mp} = 19.21 + 0.0007073 \cdot I_r - 0.08946 - T .
\]

The developed MATLAB MLR machine learning technique consists of two input variables and one output variable. These techniques can predict \( P_{mp} \) and \( V_{mp} \) at various irradiance \( I_r \) and temperature \( T \).

The regression coefficients of (10) define a plane in \( I_r, T \) and \( P_{mp} \), as shown in Fig. 11.a. The residuals in the prediction for these parameters are shown in Fig. 11.b. The numerical analysis of SSE, \( R^2 \), and RMSE are 0.0197, 0.9999 and 0.0405, respectively. The SSE and RMSE values are close to 0, and the \( R^2 \) value is close to 1, indicating the best prediction of the models and the results given in Table 2, 3.

![Fig. 11. a – \( P_{mp} \) plane defined by regression coefficients; b – residuals in prediction](image)

**Table 2**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>( 1.0347 \times 10^{-12} )</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>1</td>
</tr>
<tr>
<td>MSE</td>
<td>( 1.0762 \times 10^{-28} )</td>
</tr>
<tr>
<td>Prediction speed</td>
<td>9800 obs/s*</td>
</tr>
<tr>
<td>Training time</td>
<td>4.9252 s</td>
</tr>
</tbody>
</table>

*obs/s – refers to number of observations processed per second.
Performance comparison of various methods. The performance of the MLR model was compared to other models, and the results were summarized in Table 4 for the time range of 0 to 0.5 s. The comparison indicated that the P&O and IC methods exhibited oscillations in steady-state, while the other models did not [22-24]. According to Table 4 the MLR model settled in less than half the time with a high steady-state value of 230 W and almost zero overshoot compared to the P&O method. Similarly, the MLR model settled in less than half the time with a high steady-state value and nearly zero overshoot compared to the IC method. Overall, the MLR model outperformed the P&O and IC algorithms in terms of settling time, steady-state value, and overshoot.

Table 3

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>$3.6016 \times 10^{-14}$</td>
</tr>
<tr>
<td>$R^2$</td>
<td>1</td>
</tr>
<tr>
<td>MSE</td>
<td>$1.2972 \times 10^{-27}$</td>
</tr>
</tbody>
</table>

Table 4

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLR</th>
<th>P&amp;O</th>
<th>IC</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rise time, s</td>
<td>0.1409</td>
<td>0.0463</td>
<td>0.0352</td>
<td>0.1314</td>
</tr>
<tr>
<td>Settling time, s</td>
<td>0.2410</td>
<td>0.5000</td>
<td>0.4994</td>
<td>0.2144</td>
</tr>
<tr>
<td>Overshoot, %</td>
<td>0.0023</td>
<td>9.2364</td>
<td>39.294</td>
<td>0</td>
</tr>
<tr>
<td>Peak time, s</td>
<td>0.4999</td>
<td>0.0829</td>
<td>0.2300</td>
<td>0.5</td>
</tr>
</tbody>
</table>

According to the power response numerical values, the MLR model’s performance is comparable to that of the intelligent methods, such as ANN and FLC, while the CS method exhibits an undesirable undershoot. Moreover, the MLR model outperforms the CS optimization method in terms of rise time and overshoot. Based on this analysis, it can be concluded that the MLR control method is suitable for MPPT in PV generator, as it can track the MPP under varying $I_r$ and $T$ conditions in a stable state and ensure that the PV generator operate at the MPP.

Experimental results and discussion. To further substantiate the dynamic performance, the experiments have been conducted using the solar PV generator of 250 W Zy-TECH 250P where considered for this work shown in Fig. 12. Under standard test conditions of $I_r = 1000$ W/m$^2$ and $T = 25$ °C solar PV generator produce power of 250 W. MLR algorithm tested for solar PV generator under various $I_r$ and $T$ profiles.

The experimental panel shown in Fig. 13 consists of a solar PV panel, a designed boost converter and a program kit ESP-32. The IL is changed from 900 W/m$^2$ to 500 W/m$^2$ at $t_{IL}$ result shown in Fig. 14.

Fig. 12. Solar PV panel power comparison for various methods

Fig. 13. Experimental setup

Fig. 14. Dynamic performance of proposed MPPT controller. IL changed from 900 W/m$^2$ to 500 W/m$^2$

Note. $T = 25$ °C, $V_{pv} = 29$ V, time axis: 20 ms/div, and $t_{IL}$ is the instant at which step change in $I_r$ of solar PV generator initiated.

Conclusions. A new approach based on multivariate linear regression machine learning was implemented in this study to achieve high accuracy in tracking the maximum power point of a solar photovoltaic generator using a pulse width modulation control boost converter. The mean efficiency was found to be over 96.18 % in steady-state, which validates the effectiveness of the multivariate linear regression algorithm. Simulation with experimental hardware results showed that the multivariate linear regression algorithm had a high level of accuracy in maximum power point tracking in steady-state compared to conventional perturb & observe, incremental conductance algorithms, intelligent prediction artificial neural networks algorithm, and cuckoo search optimization method. Moreover, the multivariate linear regression algorithm proved to be effective even in the presence of varying irradiance and temperature.

As a part of future work, the effect of partial shading on photovoltaic generator will be analyzed with the help of hardware implementation.

Acknowledgment. The authors would like to thank Charitesh Raja K., Sameer S., Sravankumar Reddy K. and Senthamizh Selvan S. for their support in the preparation of this work.

Conflict of interest. The authors declare that they have no conflicts of interest.

REFERENCES


Linear Models: Theory and Applications with SAS. 2nd ed

Electronics, and Communications Boost Converter.


Shavelkin A.A., Gerlici J., Shvedchykova I.O., Kravchenko K., Kruhlik H.V. Management of power consumption in a photovoltaic system with a storage battery connected to the network with multi-zone electricity pricing to supply the local facility own needs. Electrical Engineering & Electromechanics, 2021, no. 2, pp. 36-42. doi: https://doi.org/10.20998/2074-272X.2021.2.06.


Janardhan G., Surendra Babu N.N., Srinivas G.N. Single phase transformerless inverter for grid connected photovoltaic system with reduced leakage current. Electrical Engineering & Electromechanics, 2022, no. 5, pp. 36-40. doi: https://doi.org/10.20998/2074-272X.2022.5.06.


How to cite this article: