COMPUTATIONAL RESEARCH PROGRESS IN APPLIED SCIENCE & ENGINEERING (CRPASE)

CRPASE: TRANSACTIONS OF INDUSTRIAL ENGINEERING

Journal homepage: http://www.crpase.com

CRPASE: Transactions of Industrial Engineering 9 (3) Article ID: 2862, 1–8, September 2023

ISSN 2423-4591

Research Article

Predicting Power Output of Solar Panels Using Machine Learning Algorithms

Lutfu S. Sua[1*], Figen Balo[2]

[1] *Department of Management and Marketing, Southern University and A&M College, USA*
[2] *Department of METE, Engineering Faculty, Firat University, Turkey*

Keywords

Machine learning, Solar panel, Output prediction, Renewable energy, Solar energy.

Abstract

Prediction of the energy output of solar energy systems is a field of research drawing significant attention due to its increasing share in the energy market among other factors. By integrating many techniques and simulating complicated dynamics, machine learning algorithms have emerged as an option to conventional approaches to handle these problems and offer resolutions that enhance the efficiency of photovoltaic systems. Research on machine learning and photovoltaic systems, in particular, has advanced rapidly in the last five years thanks to the involvement of deep learning in photovoltaic systems. Today, more potent models are being used to analyze structured data, including multidimensional time series, pictures, and videos. This necessitates the review of novel approaches that tackle issues in photovoltaic systems utilizing cutting-edge machine learning modellings. Machine learning methods are proving to be very efficient in various energy-related applications including consumption prediction, intrusion detection, and output prediction. This research aims to compare a number of machine learning algorithms in predicting the power output of solar panels using 13 different parameters. The paper contributes to the field by building a comprehensive model of output prediction providing an accurate comparison of several methods.

1. Introduction

Electricity is a crucial component of any nation's industrialization, urbanization, and economic expansion. Electricity is produced using a variety of traditional and unconventional energy sources. The demand for electricity is increasing every day and nonrenewable power resources cannot keep up with this demand. Because they are adaptable and environmentally beneficial, solar energy systems are becoming more and more popular. The adoption of interactive solar systems has increased significantly over the past 20 years in the area of solar energy. Solar power is an abundant and environmentally friendly source of electricity. In theory, its potency far exceeds the global power need. Rather, the energy generation from solar energy stations depends on the sun irradiation and ecological circumstances (clouds' presence, latitude, shading and terrain, aerosol amount in the content of the atmosphere, temperature, and air humidity). However, installing photovoltaic systems still entails high prices and problems with efficiency that must be fixed. The prices of installing photovoltaic systems are still being reduced as efforts are undertaken to boost their effectiveness, facilitate their installation, and couple them to electrical grids for solar power systems, with a particular emphasis on deep learning. It looks at how machine learning is used in photovoltaic systems for control, islanding identification, management, problem detection, and diagnostics, as well as sizing, site adaptation, and forecasting irradiance and power generation.

* Corresponding Author: Lutfu S. Sua
E-mail address: Lutfu.sagbansua@subr.edu, ORCID: https://orcid.org/0000-0003-4395-865X

Received: 1 September 2023; Revised: 21 September 2023; Accepted: 26 September 2023
https://doi.org/10.61186/crpase.9.3.2862

Academic Editor: **Mohammad Mahdi Ahmadi**

Please cite this article as: L.S. Sua, F. Balo, Predicting Power Output of Solar Panels Using Machine Learning Algorithms, Computational Research Progress in Applied Science & Engineering, CRPASE: Transactions of Industrial Engineering 9 (2023) 1–8, Article ID: 2862.

One of the renewable energy sources with the fastest growth rates in the world is sun power. Solar energy might be one of the best sources for future energy demand. Given that the renewable resources are plentiful and offered without charge, the implementation of sustainable power sources like sun power has a far larger potential. Solar power is the energy that is obtained from sunlight and then transformed into electricity [1]. This conversion may occasionally be carried out via concentrated sun energy, photovoltaics, or a combination of both. Since sun energy does not emit any greenhouse gases, it is a clean, self-sufficient resource that also lessens the carbon impact.

The World Energy Council claims that the solar constant represents the amount of solar energy as 1367 W/m2. The overall sun flux obtaining the superficies of earth is computed to be 1.08108 GW, and the whole amount of power obtaining the superficies is 3400000 EJ each year, taking into account both absorption and scattering. This is around 7000–8000 times the annual primary energy usage for the entire world. Only 0.1 percent of this power can be transformed with a performance as low as 10%, producing roughly 10000 GW of electricity, rounding the total to about 6000GW. According to the International Energy Agent, 11 percent of the energy requirements of the world will be met through solar energy in thirty years. It is anticipated that the use of sustainable power resources would greatly grow by 2030, with an annual growth rate of 7.6 percent [2]. All the same, installing photovoltaic mechanisms still entails high prices and problems with performance that must be fixed. The prices of installing photovoltaic mechanisms are still being reduced as efforts are undertaken to boost their effectiveness, facilitate their installation, and couple them to electrical grids.

The rest of the paper is organized as follows: Section 2 presents a review of the existing literature followed by Section 3 which provides an overview of the methodology employed. The data sources and the summary of results are presented in Section 4. Finally, concluding remarks are provided in Section 5.

2. Literature Review

The abundance of the studies comparing various models and putting forth novel hybrid techniques demonstrates that forecasting solar energy production is still a difficult task to do because the outcomes depend on many distinct variables. Reviews on the use of artificial intelligence in power mechanisms have been made in this direction and some of the most pertinent and fascinating ones are those put forth by Tina et al. [3], Youssef et al. [4], Forootan et al. [5], and Kurukuru et al. [6]. The battery storage capacity, optimum amount of solar panels, azimuth, and tilt angles needed in photovoltaic mechanisms have all been determined using machine learning approaches. Additionally, a number of methods for measuring the photovoltaic systems installed by residential customers have been established.

Using data from behind-the-meter, Kumar et al. utilized Artificial Neural Networks to calculate the photovoltaic size, inclination, and azimuth. The Pecan-Street dataset, which records the behavior of more than 1.3K customer loads over the course of a year, served as the source of the dataset used [7, 8]. The PCLib toolkit and the mechanism advisor modelling simulator were used to simulate photovoltaic generation [9]. The minimal annual net load levels were utilized as the input values. The modellings utilized to forecast photovoltaic size were put to the test in a variety of author-generated scenarios, such as photovoltaic estimations for 1K analyze customers, prediction with ranking azimuth and tilt taking dataset errors into account, diverse net load information resolutions, and the added information with incorrect labels. Khatib and Elmenreich employed a generalized Recurrent Neural Networks with the primary goal of estimating the photovoltaic array and battery sizing ratio. The load probability index loss, longitude, and latitude were used to calculate this sizing. The modelling was validated utilizing a simulation of sun irradiation on an hourly basis and load need, resulting in an average real error percentage of 0.6% [10].

Malof et al. suggested a machine learning technique to map photovoltaic systems in high-solution aerial photography by defining single-track photovoltaic arrays, their extent, and energy generating capacity over vast geographic fields [11]. Convolutional Neural Networks semantic segmentation, which offers pixel-by-pixel labels for an input image, was used to produce it. On the basis of the identification made possible by the classifier, a prediction to determine the sun panels' sun performance is also suggested. The object-sourced performance metric gave the model a precision of 0.76, indicating accuracy. The researchers initially evaluated the superficies field of the photovoltaic array placed using the segmentation procedure in order to compute the installed solar capacity. Then, to anticipate the installed photovoltaic array capacity, basic linear regression parameters based on surface area were utilized. By estimating the parameters for each array using color imagery, the model was able to attain a correlation coefficient of 0.91.

Solar irradiance and temperature are two weather variables that have an important effect on the production of photovoltaic energy. As a result of this energy source's variable production levels, it is challenging for energy firms to balance electricity consumption and production when utilizing photovoltaic systems. So, a number of Machine Learning algorithms have been put into practice to predict sun irradiation and the energy generation from photovoltaic mechanisms. Data from Temixo were used by Tovar et al. to anticipate photovoltaic power using a five-layer Convolutional Neural Networks Long Short-Term Memory model [12, 13]. Forecasting horizons from 10 minute to 180 minute were used [14].

The objective of Suresh et al. was to use Convolutional Neural Networks to predict the production of solar photovoltaic. For medium- and short-term prediction, a Convolutional Neural Networks-Long Short-Term Memory, multi-headed Convolutional Neural Networks, and ordinary Convolutional Neural Networks methods were used. An auto-regressive moving average model and multiple linear regression were then put up against the models for comparison. Solar irradiance, ambient temperature, wind speed, and the photovoltaic module temperature taken in 5-minute windows were the variables used for prediction. These characteristics were selected based on the international energy agency report [15]. According to the findings of that study, a straightforward Convolutional Neural Networks and a Convolutional Neural Networks combined with a Long Short-Term Memory network applied admirably for prediction time horizons of one hour, one day, and one week [16].

Gated Recurrent Unit and Encoder-Decoder Long Short-Term Memory networks were compared by Narvaez et al. for weekly and daily time horizon estimations. Both in daily and weekly timeframes, the Long Short-Term Memory performed better than the Gated Recurrent Unit networks [17].

For the forecast of every half hour diffuse horizontal sun radiation using solely the worldwide horizontal radiation and a geographical position, Miranda et al. compared various Machine Learning techniques, containing Artificial Neural Network. Data for six different sites in Colombia were acquired from the national sun irradiation database [18]. Over several sites, the researchers obtained coefficients varying from 0.9983 to 0.9974 [19].

As a short-term forecaster for photovoltaic power, Park and Ahmed used the Recurrent Neural Networks. Power data that was gathered in real-time was used to train this model. The module temperature, sun radiation, wind speed, outside temperature, and relative moisture were the data that were gathered. The predicting was done for time horizons of 5-minute, 15-minute, 1 hour, and 3 hours. The results of the trials showed that, for the short-term forecast, the suggested model outperformed the auto-regressive combined moving mean and support-vector regression with Random Forest modellings in terms of prediction accuracy. For forecasting across time horizons of 5 minutes and 15 minutes, the Recurrent Neural Networks model's accuracy was 99.1% and 98.6%, respectively, while for forecasting over time horizons of 1 hour and 3 hours, it was 97.4% and 96.2% [20].

Mahmood and Hossain classified historical radiation information into various sky classes using a K-nearest-neighbors technique. The researchers displayed that the suggested modelling surpassed modellings such as Recurrent Neural Networks, extreme learning machines, and general regression neural network [21]. Pan et al. utilized a short-term sun-based production predicting methodology based on Long Short-Term Memory with temporal attention system [22].

In spite of the applicability of machine learning algorithms to the renewable energy field, there is a need to build practical approaches for the practitioners as part of their feasibility studies. To the best of our knowledge, this is the first study that provides a comparison of several machine learning algorithms in predicting the power output of all major solar panel brands providing 350W energy output based on 13 electrical properties. The paper contributes to the field by building a comprehensive model of output prediction providing an accurate comparison of several methods.

3. Machine-Learning Methodology

The machine-learning is a combination of computer science and artificial intelligence which concentrates on the utilization of algorithms and data to imitate the way that people learn, gradually developing its accuracy. Machine learning is a significant element of the data science. By the utilization of the statistical methodologies, algorithms are trained to make predictions or classifications, and to unearth main perceptions in planning of data mining. These insights followed by the use of decision-making methods in business applications, help reaching optimal decisions. As big data proceeds to grow and expand, the market demand will also raise for information scientists. They will be needed to aid identifying the information necessary to answer the most important business questions.

Machine learning algorithms are typically constructed using frameworks that expedite solution development, such as PyTorch and TensorFlow. Various machine learning algorithms find widespread use in this field. Some of these include: Logistic Regression: A supervised learning algorithm used for making predictions involving binary outcomes, such as "yes" or "no" answers to questions. The linear regression method is employed to estimate digital values and is based on a linear relationship between multiple factors. Neural networks, which have integrated a very high number of processing nodes are planned to apply likewise to the people brain. Implementations that benefit from capacity of neural networks for sample finding contain natural language translation, imagery production, speech recognition, and picture identification.

The classification of data into categories and the regression of numerical values are both possible uses for decision trees. The decision-trees employ a branching series of connected decisions that can be visualized as a tree. Unlike the black box of the neural network, decision-trees are simple to audit and validate, which is one of their advantages. Data patterns can be found via unsupervised learning using clustering algorithms, allowing the data to be classified. Through spotting distinctions between data points that people have missed, computers can aid information scientists. By aggregating the findings from various decision trees, the machine learning algorithm in a random forest predicts either a value or category [23].

The most prevalent Machine Learning methods that have been used on photovoltaic systems are the ones with a focus on Deep Learning techniques. Artificial Neural Networks (ANN) are modellings that are encouraged in the fact that neurons, as a processing unit, are connected in the brain of people [24]. Several examples of supervised learning and reinforcement learning methods used with photovoltaic systems are given in Table 1 [25].

An artificial neural network involves a process that translates an n-dimensional input to an r-dimensional output. It is built using the serial and parallel connections of neurons, which are the building blocks of a basic cognitive processes. Each of the neurons calculates a weighted sum of its own inputs that is then converted utilizing an enabling process. The information operation brings up the operation of finding the weights for each of the neurons in a way that the artificial neural network modelling is able to exactly anticipate the output for a given input [26]. An agent can learn check policies by trial-and-error utilizing rewards offered through an interactive environment, which is referred to as a

"Markov Decision Operation", thanks to the Reinforcement Learning technique [27].

Table 1. Several examples of supervised and reinforcement learning methods used with photovoltaic systems.

Machine Learning Techniques	Supervised Learning	Deep Learning Modellings	Convolutional Neural Networks
			Recurrent Neural Networks
			Artificial Neural Networks
			Gated Recurrent Unit
			Long Short-Term Memory
		Support Vector Machines	
		Random Forest	
	Reinforcement Learning	Tabular	State–action–reward–state–action
			Q-Learning
		Deep Learning	Soft-Actor-Critic
			Deep Deterministic Policy Gradient
			Twin-delayed Deep Deterministic Policy Gradient
			Deep Q-Learning

A Reinforcement Learning problem must have three key components: the state (the agent's current status), the reward (the environment's feedback), and the strategy (the procedure that establishes the relationship between the agent's states and its actions). Finding a set of actions that maximizes the reward from the agent's interactions with the environment must be the algorithm's main objective. State–action–reward–state–action, Q-learning, and their variations are some of the methods now employed to carry out the learning process in Reinforcement Learning [28].

Some methods, referred to as Q-tables, partition the continual states and field of action into a finite number of areas. Another group of techniques employs DL modellings and Artificial Neural Networks to simulate a policy function. The Q-table is substituted through a nonlinear continual process modeled in an artificial neural network [29], enabling continuous state and action spaces [30].

Deep Deterministic Policy Gradient, Soft-Actor Critic, Deep Q-Learning, and Twin Delayed Deep Deterministic Policy Gradient are a few of these methods. Convolutional Neural Networks are a type of Artificial Neural Networks using filters of a specific length and a hierarchical approach to extract key characteristics from a 1-dimensional, 2-dimensional, or 3-dimensional systems. Recurrent Neural Networks are a class of networks that can classify events or predict the future outcomes on information arrays of any length. By analyzing the frame's connections that point backwards, interconnections between the frames can be inferred [31].

A component of the analyzed frame and the output of the earlier estimation, which was based on the last analyzed frame factor, are both inputs to each of the neurons in the Recurrent Neural Network layer. According to this structure, Recurrent Neural Networks allow the prior outputs' utilization as inputs while having secret units that use hidden states to account for the past [32]. The goal of Gated Recurrent Units is to retain data from the past for a long time [33].

This is accomplished through placing a series of characteristic between the edges of neighboring hidden units in Recurrent Neural Networks. This characteristic would function as a gate, controlling how much data the network can retain over time. More memory is preserved if this parameter value is close to one, but memory of earlier stages would be lost if it were close to zero. A unique variety of Recurrent Neural Network termed Short- and Long-term Memory is able to learn long-term dependencies [34].

It can retain data for extended time periods due to a memory cell called cell state, which stores past information and is updated using two separate parameters that are comparable to those found in Gated Recurrent Units. These characteristics enable former data to be forgotten or remembered, and novel data to be added to an exact degree. The cell state is then utilized to update the secret status of the widespread Recurrent Neural Network through utilizing another characteristic that permits exact parts of its data flow by the secret unit.

These characteristics are the results of three simple Artificial Neural Network gates—the output gate, input gate, and forget gate—whose input is the earlier secret state. A method known as Support Vector Machines obtains a linear modelling for grouping [31]. The Support Vector Machine attribute is based on the idea of greatest margin, which refers to the separation between the nearest observations in either class and the separating hyperplane. A model called Random Forest is based on the decision-trees that are built concurrently with input variables' random subsets. The class with the greatest votes is assigned by a Random Forest model using an example that has been run through all of the trained decision trees [35].

4. Summary of Results

4.1. Data Summary

For the purpose of this study, a list of major solar panel brands providing 350W of output is created. The characteristics (DC electrical properties) of these solar panels are given in Table 2. The Matrix in Table 3 shows the correlation relations among thirteen electrical properties.

The histogram for PTC is displayed in Figure 1. The PTC-STC scatter plot is shown in Figure 2.

Table 2. 350 W Solar panel characteristics (DC electrical properties)

	PTC	STC	Peak	Power	Cells	Imp	Vmp	Isc	Voc	NOCT	CoPow	CoVol	Fuse
1	309,10	162,20	16,22	0,00	60,00	8,13	42,98	8,93	51,47	50,00	-0,48	-0,19	15,00
2	321,32	180,40	18,04	1,50	72,00	9,07	38,60	9,57	47,20	45,00	-0,41	-0,14	15,00
3	326,55	191,80	19,18	1,50	120,00	10,48	33,40	11,04	40,20	44,00	-0,36	-0,12	20,00
4	318,80	180,90	18,09	2,50	72,00	9,08	38,58	9,53	47,01	46,00	-0,38	-0,13	15,00
5	321,15	180,40	18,04	1,50	72,00	9,10	38,48	9,81	47,37	46,70	-0,38	-0,13	15,00
6	319,00	180,40	18,04	1,50	72,00	9,13	38,34	9,58	46,89	45,00	-0,40	-0,14	20,00
7	321,33	180,60	18,06	0,50	72,00	9,09	38,70	9,60	47,00	45,00	-0,41	-0,16	15,00
8	323,85	174,40	17,44	1,00	144,00	9,11	38,43	9,58	46,27	45,00	-0,37	-0,14	20,00
9	322,72	180,40	18,04	1,50	72,00	8,94	39,10	9,38	47,50	45,00	-0,39	-0,14	15,00
10	324,40	189,50	18,95	1,50	120,00	10,42	33,60	11,19	40,10	45,00	-0,35	-0,11	20,00
11	319,20	180,40	18,04	1,00	72,00	9,18	38,14	9,70	47,40	45,00	-0,39	-0,14	15,00
12	322,72	180,40	18,04	1,50	72,00	8,94	39,10	9,38	47,50	45,00	-0,39	-0,14	20,00
13	324,70	180,00	18,00	1,00	72,00	9,03	38,80	9,50	46,40	45,00	-0,39	-0,14	15,00
14	326,60	188,00	18,80	1,50	120,00	10,48	33,40	11,04	40,20	44,00	-0,36	-0,12	20,00
15	317,30	175,50	17,55	1,00	72,00	9,03	38,80	9,62	47,40	46,00	-0,45	-0,16	15,00
16	317,10	180,40	18,04	1,00	72,00	9,01	39,13	9,65	47,57	45,00	-0,39	-0,15	15,00
17	329,15	195,10	19,51	2,50	120,00	10,37	33,76	10,97	41,11	43,00	-0,34	-0,11	20,00
18	318,70	174,40	17,44	1,00	72,00	9,00	38,90	9,72	46,70	44,60	-0,36	-0,14	20,00
19	321,32	180,40	18,04	2,50	72,00	9,16	38,20	9,56	46,70	45,00	-0,41	-0,14	15,00
20	322,50	179,50	17,95	0,50	72,00	8,89	39,40	9,50	48,80	45,00	-0,43	-0,15	15,00
21	326,60	188,00	18,80	1,50	120,00	10,48	33,40	11,04	40,20	44,00	-0,36	-0,12	20,00

Table 3. Correlation Matrix

	PTC	STC	Peak	Power	Cells	Imp	Vmp	Isc	Voc	NOCT	CoPow	CoVol	Fuse
PTC	1,000	0,854	0,854	0,485	0,694	0,787	-0,811	0,717	-0,789	-0,841	0,762	0,851	0,518
STC	0,854	1,000	1,000	0,612	0,574	0,890	-0,905	0,843	-0,860	-0,807	0,761	0,906	0,456
Peak	0,854	1,000	1,000	0,612	0,574	0,890	-0,905	0,843	-0,860	-0,807	0,761	0,906	0,456
Power	0,485	0,612	0,612	1,000	0,250	0,461	-0,493	0,380	-0,451	-0,478	0,551	0,672	0,232
Cells	0,694	0,574	0,574	0,250	1,000	0,787	-0,781	0,775	-0,809	-0,512	0,667	0,672	0,722
Imp	0,787	0,890	0,890	0,461	0,787	1,000	-0,998	0,986	-0,991	-0,684	0,749	0,877	0,643
Vmp	-0,811	-0,905	-0,905	-0,493	-0,781	-0,998	1,000	-0,976	0,988	0,719	-0,771	-0,902	-0,641
Isc	0,717	0,843	0,843	0,380	0,775	0,986	-0,976	1,000	-0,977	-0,602	0,720	0,834	0,640
Voc	-0,789	-0,860	-0,860	-0,451	-0,809	-0,991	0,988	-0,977	1,000	0,675	-0,772	-0,876	-0,683
NOCT	-0,841	-0,807	-0,807	-0,478	-0,512	-0,684	0,719	-0,602	0,675	1,000	-0,747	-0,779	-0,497
CoPow	0,762	0,761	0,761	0,551	0,667	0,749	-0,771	0,720	-0,772	-0,747	1,000	0,909	0,654
CoVol	0,851	0,906	0,906	0,672	0,672	0,877	-0,902	0,834	-0,876	-0,779	0,909	1,000	0,588
Fuse	0,518	0,456	0,456	0,232	0,722	0,643	-0,641	0,640	-0,683	-0,497	0,654	0,588	1,000

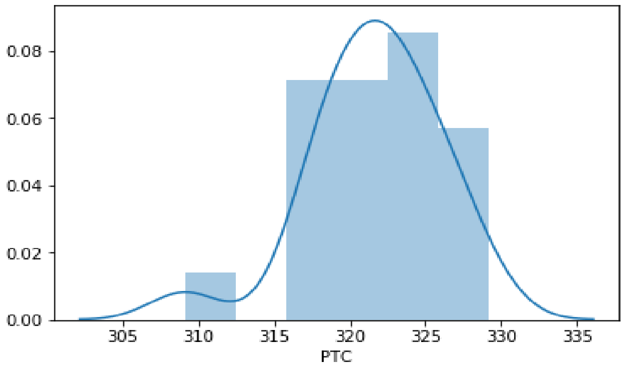

Figure 1. Histogram PTC

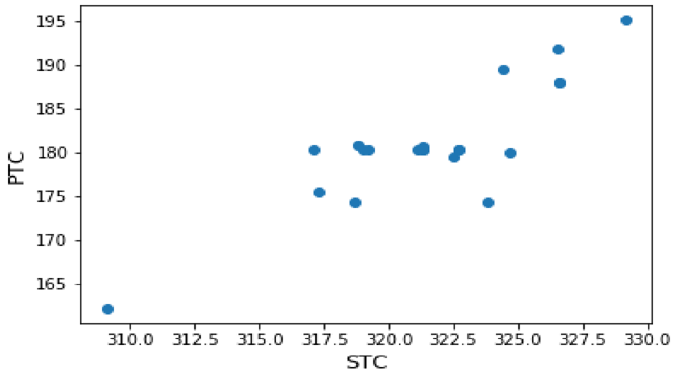

Figure 2. PTC-STC Scatter Plot

4.2. Prediction of PTC with Machine Learning Methods

This section provides the prediction performance of various machine learning methods. The results obtained by applying Linear Regression, XGBRegressor, Gradient Boosting Regressor, Decision Tree Regressor, and Random Forest Regressor are presented along with the error terms.

a. Linear Regression:

Regression model of one dependent and twelve independent variables is as it shown in Eqs. (1-2)

$$y\_pred = regressor.predict(X\_test) \tag{1}$$

$$PTC=STC.x1 + Peak.x2 + \ldots.. + Fuse.x12 \tag{2}$$

Table 4 presents the factor coefficients obtained from linear regression. Figure 3 illustrates a comparison of the actual versus predicted values of PTC when linear regression is used for the prediction.

Table 4. Regression Coefficients

	STC	Peak	Power	Cells	Imp	Vmp	Isc	Voc	NOCT	CoPow	CoVol	Fuse
Coefficient	0,9504	0,095	-0,7759	0,1135	-55,559	-7,8051	5,9062	-2,7331	-0,2439	-30,841	-15,682	-0,1271

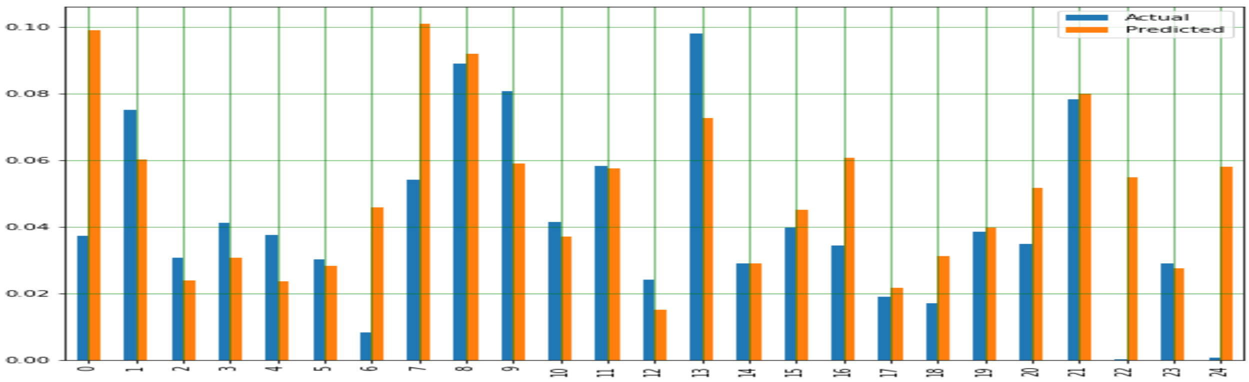

Figure 3. Actual vs. Predicted PTC

Average absolute error, average squared error, root average squared error, and R^2 values reveal the performance of the method.

Average Absolute Error:	2.16804
Average Squared Error:	8.60751
Root Average Squared Error:	2.93385
R^2:	-0.79481

R^2 evaluates how well the selected modelling fits in comparison to the null hypothesis, which is a horizontal straight line. R^2 is negative since the selected modelling fits the data less well than a horizontal line. R^2 can have a negative value without breaking any mathematical principles, as it is important to note that it is not always the square of anything. Only when the chosen modelling deviates from the data's trend and provides a worse fit then R^2 becomes negative. Given these facts, the model has no meaning. R^2 must be zero (or positive) and equals the correlation coefficient's square, r, in linear regression with no restrictions. Only when the intercept or slope are confined in such a way that the "best-fit" line fits worse than a horizontal line is a negative R^2 with linear regression conceivable. Whenever the best-fit modelling fits the data less well than a horizontal line, the R^2 might be negative in a nonlinear regression. In conclusion, a negative R^2 indicates how badly the selected model matches the data compared to a horizontal line.

b. XGBRegressor, Gradient Boosting Regressor, Decision Tree Regressor, Random Forest Regressor

Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, and R2 values of the remaining ML methods are provided in Table 5.

It can be observed from Table 5 that XGB Regressor and Gradient Boosting Regressor have the highest R^2 values indicating their high performance.

Table 5. Comparison of ML Methods

	Regr	**XGB**	**GBR**	**DTR**	**RFR**
MAE	2.1680432660917	1.7463457031250	2.50316838146860	2.0920000000000	1.9322400000002
MSE	8.6075140578517	3.5851640925968	7.51061838005685	7.4176400000000	5.4656317931196
RMSE	2.9338565162345	1.8934529549468	2.74055074393028	2.7235344682966	2.3378690709959
R2	-0.7948115295317	**0.9999999772198**	**0.99993212473186**	-0.5467027650999	-0.1396762052939

5. Conclusions

Sun power generation is perhaps one of the most popular renewable power resources. For instance, it is projected to produce 48% of the energy demand in the United States by 2050 [36]. Additionally, as a result of capacity expansions by Japan, India, US, Germany, and China, it is predicted that the world's solar photovoltaic capacity would increase up to 1582.9 GW in 2030 [37].

In this study, accuracy of five machine learning methods in predicting the power output of 350W solar panels are compared using thirteen electrical properties. The empirical results on machine learning algorithms provide several insights for decision-makers. First, it surveys the solar panel brands providing 350W energy output and collects their values for a comprehensive list of electrical properties. Second, it provides a comprehensive model of output prediction providing an accurate comparison of several methods. Five different models are test using machine learning algorithms and their accuracy levels along with error terms are presented. Decision-makers might use our framework to help businesses on investment decisions and find the best combination of system components to design their solar systems.

There are inherent limitations in our study mainly due to the availability of data. Other factors such as price and physical properties can be added to the model in the future, when data become available for all the brands. Another venue for future research is to devise a system approach by applying the prediction algorithms for various components of the renewable energy systems.

Conflict of Interest Statement

The authors declare no conflict of interest.

References

[1] L. Sagbansua, F. Balo, Photovoltaic panel selection: AHP approach, International Journal of Engineering and Technical Research 6 (2016) 95–100.
[2] World Energy Council, Energy Resources: Solar, World Energy Counc. World Energy Resour. Sol., (2013) 1–28.
[3] G. M. Tina, C. Ventura, S. Ferlito, S. DeVito. A state-of-art-review on machine-learning based methods for PV. Applied Sciences 11 (2021).
[4] A. Youssef, M. El-Telbany, A. Zekry, The role of artificial intelligence in photo-voltaic systems design and control: a review, Renewable and Sustainable Energy Reviews 78 (2017) 72–79.
[5] M. M. Forootan, I. Larki, R. Zahedi, A. Ahmadi, Machine learning and deep learning in energy systems: a review, Sustainability 14 (2022) 4832.
[6] V.S.B. Kurukuru, A. Haque, M A. Khan, S. Sahoo, A. Malik, F. Blaabjerg, A review on artificial intelligence applications for grid-connected solar photovoltaic systems, Energies 14 (2021) 4690.
[7] S.A. Kumar, M.S.P. Subathra, N.M. Kumar, M. Malvoni, N.J. Sairamya, S.T. George, E.S. Suviseshamuthu, S.S. Chopra, A novel islanding detection technique for a resilient photovoltaic-based distributed power generation system using a tunable-q wavelet transform and an artificial neural network, Energies 13 (2020) 4238.
[8] PecanStreet. PecanStreet inc. dataport load data, URL, https://www.pecanstreet.org/dataport/.
[9] Home - system advisor model (SAM), URL, https://sam.nrel.gov/.
[10] T. Khatib, W. Elmenreich, An Improved Method for Sizing Standalone Photovoltaic Systems Using Generalized Regression Neural Network, (2014).
[11] J. M Malof, B. Li, B. Huang, K. Bradbury, A. Stretslov. Mapping Solar Array Location, Size, and Capacity Using Deep Learning and Overhead Imagery. arXiv (2019) page 6.
[12] ESOLMET-IER instituto de energías renovables, http://esolmet.ier.unam.mx/Tipos_consulta.php.
[13] M.Tovar.Esolmet,https://github.com/mariotovarrosas/ESOLMET2019. original-date: 2020-09-21T22:54:55Z.
[14] M. Tovar, M. Robles, F. Rashid. PV power prediction, using CNN-LSTM hybrid neural network model. case of study: Temixco-morelos, mexico. Energies 13 (2020) 6512.
[15] S. Pelland, J. Remund, J. Kleissl, T. Oozeki, and K. De Brabandere. Photovoltaic and Solar Forecasting: State of the Art. International Energy Agency Photovoltaic Power Systems Programme, Report IEA-PVPS T14-01 (2013) ISBN 978-3- 906042-13-8.
[16] V. Suresh, P. Janik, J. Rezmer, and Z. Leonowicz. Forecasting Solar PV Output Using Convolutional Neural Networks with a Sliding Window Algorithm. Energies 13 (2020) 15.
[17] G. Narvaez, L.P. Giraldo, M. Bressan, and A. Pantoja. Machine Learning for Site-Adaptation and Solar Radiation Forecasting. Renewable Energy 167 (2021) 333–342.
[18] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, S. James, The national solar radiation data base (NSRDB), Renewable and Sustainable Energy Reviews 89 (2018) 51–60.
[19] E. Miranda, J.F.G. Fierro, N. Gabriel, L.F. Giraldo, M. Bressan, Prediction of site-specific solar diffuse horizontal irradiance from two input variables in Colombia, Heliyon 7 (2021) 08602.
[20] H.K. Ahn, N. Park, Deep RNN-based photovoltaic power short-term forecast using power IoT sensors, Energies 14(2) (2021) 436.

[21] M.S. Hossain, H. Mahmood, Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast, IEEE Access 8 (2020) 172524e172533.
[22] T. Khatib, W. Elmenreich, An Improved Method for Sizing Standalone Photovoltaic Systems Using Generalized Regression Neural Network, International Journal of Photoenergy (2014). https://www.hindawi.com/journals/ijp/2014/748142/. ISSN: 1110-662X Pages: e748142 Publisher: Hindawi
[23] https://www.ibm.com/topics/machine-learning
[24] K. Gurney. Introduction to Neural Networks. UCL Press Limited Oxford, 1997. OCLC: 892785047. ISBN: 9780203451519
[25] J.F. Gaviria, G. Narvaez, C. Guillen, L.F. Giraldo, M. Bressan, Machine learning in photovoltaic systems: A review, Renewable Energy 196 (2022) 298–318.
[26] R. Hecht-Nielsen. Theory of the Backpropagation Neural Network. International 1989 Joint Conference on Neural Networks, Washington, DC, USA, 1 (1989) 593–605.
[27] S. Bhatt, Reinforcement learning 101, URL, https://towardsdatascience.com/reinforcement-learning-101-e24b50e1d292.
[28] R.S. Sutton, A.G. Barto. Reinforcement learning: an Introduction. Adaptive Computation and Machine Learning Series. The MIT Press, second ed. Edition, 2018. ISBN 978-0-262-03924-6.
[29] Deep deterministic policy gradient spinning up documentation,https://spinningup.openai.com/en/latest/algorithms/ddpg.html.
[30] J. Fan, Z. Wang, Y. Xie, Z. Yang, A theoretical analysis of deep q-learning, ArXiv (2019) http://arxiv.org/abs/1901.00137
[31] A. Geron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, second Edition, 2019.
[32] A. Amidi, Recurrent neural networks cheatsheet, https://stanford.edu/shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks.
[33] S. Kostadinov, Understanding GRU networks. https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be.
[34] OLah, Understanding LSTM networks. https://web.stanford.edu/class/cs379c/archive/2018/class_messages_listing/content/Artificial_Neural_Network_Technology_Tutorials/OlahLSTM-NEURAL-NETWORK-TUTORIAL-15.pdf.
[35] T. Yiu. Understanding random forest. https://towardsdatascience.com/understanding-random-forest-58381-0602d2.
[36] Renewable energy, https://www.c2es.org/content/renewable-energy/.
[37] G. Data, Global solar photovoltaic (PV) market update, 2019 with historic (2006-2018) and forecast (2019-2030), https://www.businesswire.com.