Automatic solution for solar cell photo-current prediction using machine learning

Mohammed Azza^{1,*}, Jabran Daaif¹, Adnane Aouidate³, El Hadi Chahid^{1,2} and Said Belaaouad¹

¹Laboratory of Physical Chemistry of Materials LPCM, Ben M'Sick Faculty of Sciences, P. B. 7955.Bd. Driss El Harti. Hassan II University of Casablanca. Morocco.

²Regional Center for Education and Training CRMEF-Beni-Mellal Khenifra. Mohamed V Street. BeniMellal,. Morocco.

³School of sciences, Moulay Ismail University, Meknes, Morocco.

Abstract. In this paper, we discuss the prediction of future solar cell photo-current generated by the machine learning algorithm. For the selection of prediction methods, we compared and explored different prediction methods. Precision, MSE and MAE were used as models due to its adaptable and probabilistic methodology on model selection. This study uses machine learning algorithms as a research method that develops models for predicting solar cell photo-current. We create an electric current prediction model. In view of the models of machine learning algorithms for example, linear regression, Lasso regression, K Nearest Neighbors, decision tree and random forest, watch their order precision execution. In this point, we recommend a solar cell photocurrent prediction model for better information based on resistance assessment. These reviews show that the linear regression algorithm, given the precision, reliably outperforms alternative models in performing the solar cell photo-current prediction I_{ph}

Key words: solar cell, machine learning, prediction, photo-current.

1 Introduction

In recent years the whole world is moving towards the production of renewable energies due to the depletion of

fossil energy resources. It is a strategy to avoid the energy crisis and to partially end the pollution of our planet triggered by the combustion of fossil fuels, the advantage of this kind of energy exists from its respectable relationship with the environment. Solar energy is the promising source of energy in industry,

* Corresponding author: mohammedazza81@gmail.com

which is why we are interested in solar radiation that arrives at the earth's surface to seek information for the implementation of photovoltaic system, is optimal.

The photovoltaic effect is a physical phenomenon where there is a conversion of light energy into electrical energy in a semiconductor [1], through a device which is the photovoltaic cell. The absorption of light photons results in the generation of excitons by recombination of electrons and the holes thus formed as a result of the jump of electrons from the valence band to the conduction band, in recent years the world has put this technology primarily because of the increase in energy requirements, this is why we are interested in studying the electrical parameters of this technology. [2]

In this work we know how to focus on the photocurrent prediction of the photovoltaic cell from the measured results of shunt and series resistance using machine learning which plays a major role in image detection, spam reorganization, normal voice control, product recommendation and medical diagnosis. The current machine learning algorithm helps us improve safety alerts, ensure public safety, and improve medical improvements. The machine learning system also provides better customer service and safer automotive systems. The algorithms used are linear regression, Lasso regression, K Nearest Neighbors, decision tree and random forest, aim to predict future photocurrent values, as we have compared the Precision between the algorithms used, this prediction will allow us to specify the most valid method for the prediction of the photocurrent of the photovoltaic (PV) system studied. Our solar cell is modeled by the electrical circuit shown in Figure 1.

Fig. 1. The equivalent circuit of the PV cell.

2 Material and method

This section provides details about the data and tools used in the study. The choice of algorithms to perform the solar cell photo-current forecast took into account the nature of the data set. Linear regression, KNN, decision tree and random forest are the different methods that have been applied for the regression. For this we present the values of the parameters to be used to solve this problem in the table represented in table 1. These parameters are calculated by the method of least squares and the method of Newton – Raphson which consists in extracting the parameters of the solar cell.

Table 1. Features of the solar cell dataset [3].

No.	I _{ph}	R _{sh}	RS
1	0.0090	900.0	1.74
2	0.0174	582.8	1.38

3	0.0246	460.7	1.70
4	0.0316	428.3	1.66
5	0.0378	376.7	1.73
6	0.0451	300.0	1.62
7	0.0510	299.4	1.67
8	0.0315	337.4	1.73
9	0.0317	428.3	1.66
10	0.0322	465.4	1.36
11	0.0334	403.8	2.27
12	0.0339	347.3	2.90
13	0.0357	316.2	1.73
14	0.0361	311.4	1.66
15	0.0090	874.8	1.69

2.1 Least squares method

The extraction of the electrical parameters by the Least Squares algorithm (LMS) to determine the electrical parameters using the one-diode model which is very sufficient to describe the electrical operation of the solar cell [4].

$$I(V) = I_{ph} - I_s \left[\exp\left(\frac{V + IR_s}{nKT}\right) - 1 \right] - \frac{V + IR_s}{R_p} \quad (1)$$

The principle of the least squares method consists in minimizing the sum of the squares of the deviations E called the objective function, between the N experimental measurements.

2.2 The loss functions

Before running a regression model, a loss function must be chosen. Although the mean squared error (MSE) is one of the best known and most used regression measures [5], here we have evaluated our models with MSE, RMSE and MAE.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2 \quad (2)$$

Where y_i , i and n are, respectively, the true value, the predicted one and n the number of samples. $\tilde{y}$ the root mean square error (RMSE) is defined as the square root of the MSE [6].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2} \quad (3)$$

The mean absolute error (MAE) is another standard way to measure a model's error in predicting quantitative data. Mathematically, it is defined as follows [6] :

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \tilde{y}_i| \quad (4)$$

2.3 Linear regression

2.3.1 Simple linear regression

A simple linear regression statistical method allows us to summarize and study the relationship between two continuous quantitative variables [7]:

$$Y = BX + a$$

Where X, Y are, respectively, input data, our prediction, a and b are the learning coefficients and which allow us to produce the most accurate predictions.

2.3.2 Multivariate regression

A more complex multivariate linear equation is used to check whether there is a statistically noticeable association among the sets of variables [7-8].

$$Y(x_1, x_2, x_3) = w_1x_1 + w_2x_2 + w_3x_3 + w_0 \quad (5)$$

Or the variables x_1, x_2, x_3 represent the attributes or distinct information for each observation, Y is our prediction and w represents the coefficients or weights, our model will try to learn. [6]

2.4 K Nearest Neighbors Regression

K Nearest Neighbors (KNN) regression is a simple algorithm that stores all available cases and predicts the digital target based on a similarity measure. The distance functions used by the KNN regression are the same functions as the KNN classification [9].

Euclidean:

$$\sum_{i=1}^k (x_i - y_i)^2 \quad (6)$$

Manhattan:

$$\sum_{i=1}^k |x_i - y_i| \quad (7)$$

Minkowski:

$$\left[\sum_{i=1}^k (|x_i - y_i|)^q \right]^{\frac{1}{q}} \quad (8)$$

2.5 Lasso Régession

We usually define the design matrix X as a matrix having n rows and p columns, representing the p variables for n instances. β is the coefficient vector [10]. We also define the target vectory as a column length vector n containing the corresponding values of the target variable.

The lasso regression technique attempts to produce a sparse solution, in that several of the slope parameters will be set to zero. The Lasso is also formulated with respect to the matrix X. In addition, the penalty L1 is only applied to the slope coefficients, and therefore the intercept, β_0 , is excluded from the penalty term. Therefore, the Lasso can be expressed as a constrained minimization problem [11].

$$B_{lasso} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n (Y_i - X_i^T \beta)^2 \quad (9)$$

Subject to

$$\sum_{j=1}^p |\beta_j| \leq t$$

For $t \geq 0$; this can again be rephrased using the Lagrangian for the L1- penalty, as follows.

$$\hat{B}_{lasso} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n (Y_i - X_i^T \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (10)$$

Or $\lambda \geq 0$ and, as before, there is a one-to-one correspondence between t and λ . The L1 penalty makes the solution nonlinear in the y_i . The constrained minimization above is a quadratic programming problem, the solution of which can be efficiently approximate

2.6 Decision tree

The decision tree builds regression or classification models in the form of a tree structure. It decomposes a dataset into smaller and smaller subsets while gradually developing an associated decision tree. The end result is a tree with decision nodes and leaf nodes [12].

2.7 Random Forest

Bootstrap refers to random sampling with replacement. This is a general procedure that can be used to reduce variance for algorithms that have high variance, usually decision trees [13].

Random Forest is part of Bootstrap techniques. Trees in random forests are executed in parallel. There is no interaction between these trees during construction [14]. It works by building a multitude of decision trees at training time and displaying the class which is the mode of classes (classification) or mean prediction (regression) of individual trees [15].

3 Results and discussion

3.1 Prediction of solar cell current photo values

In data science, the analyst might be interested in the estimated values for the objects in the training set, predicting the responses for the new objects. However, identifying the most important input variables for making good predictions is difficult due to the complex relationships between these variables [16].

This study presents the prediction of photo-current I_{ph} using different algorithms such as linear regression, lasso algorithm, KNN, decision tree, random forest based on data such as shunt resistance and series

resistance calculated by least squares and Newton-Raphson methods.

Fig. 1. Prediction of Iph by the decision tree

In figure 3 we have evaluated our model using the decision tree algorithm, in this case the algorithm is reliable in the first observations, up to node number 3 in which it has a large difference between the values predicted and actual values.

Fig. 2. Prediction of Iph values by LinearRegression

Fig. 2. Prediction of Iph values by random forest.

In this graph the actual values and the predicted values are moderately far apart in node number 3, on the other hand in the other nodes the forecast is optimal due to a minimum distance which can be seen between the values to be optimized.

Fig. 3. The values predicted by contribution to the real values by lasso regression.

Figure 5 presents the prediction of Iph treated by lasso Regression which also gives a good contribution between the real values and the predicted values; this algorithm evokes valid values for RSME, MSE and MAE.

Fig. 4. Prediction of Iph values by KNN.

Graph 6 shows that the predicted values do not coincide with the actual values due to high RSME value which says that this algorithm is unreliable in predicting the photocurrent values of the solar cell studied so the description of results obtained by this method remains invalid.

3.2 Comparison of algorithms

Fig. 5. Precision of the algorithms used.

From this graph, we can see that the linear regression and the lasso regression and the random drill algorithm give good precision in the prediction of the photocurrent values.

Fig. 6. RSME test given by the algorithms used.

To interpret these results, we first compare RSME of these algorithms, such that the KNN algorithm has a very high RSME compared to the other algorithms, which justifies that the prediction of Iph by this method is low since it evokes a large error, and we find that the linear regression gives a very low RSME which shows that this algorithm is reliable due to its low error values.

Fig. 7. Test MSE given by the algorithms used

This figure works out the values of MSE of the algorithms used; the KNN gives a very high value of MSE and a very low value for the linear regression.

Fig. 8. Test MAE given by the algorithms used

The following result confirms that KNN has a high value of MAE which justifies that it is weak in the prediction of parameters studied in a valid way, on the other hand the linear regression algorithm always remains weak during the tests carried out.

Table 1. Illustration of algorithms.

Algorithm	Score	MSE	RSME	MAE
Linear Regression	0.959437	8.504986e-07	0.000922	0.000911
Lasso Regression	0.958530	7.365432e-06	0.002714	0.004706
KNN	0.593065	4.689389e-05	0.016220	0.005800
Decision tree	0.617507	1.231507e-05	0.003509	0.002171
Random Forest	0.848903	4.864835e-06	0.002206	0.001737

This table summarizes the model precision values, MSE, RSME and MAE, so the classification will be made from the most reliable algorithm to the less reliable one, first the linear regression has a value of 0.959437 followed by the lasso regression with a value of 0.958530, then the random forest with a value of 0.848903, then the decision tree algorithm with a value of 0.617507 and finally the KNN with a low value of 0.593065, and for RSME, MSE and MAE have the same classification as the precision of the model, so the linear regression algorithm is very reliable compared to the other algorithms studied.

4 Conclusion

In this work we used five algorithms for the prediction of the values of I_{ph} of the photovoltaic cell, such as the linear regression which presents a good contribution between the real values and the predicted values, this

method presents optimal values concerning MSE, RSME and MAE, the same description is carried out on the lasso regression and the random forest algorithm which also give better results for the prediction of this electrical parameter which characterizes the solar cell studied.

The other two KNN algorithms and the decision tree algorithm are moderately low in accuracy, because there is a gap between the predicted and actual values. At this point it can be concluded that the The reliable algorithm in this study is that of linear regression.

References:

- Othmani, H., Lachiheb, M. A., Bsiri, N., Zrir, M. A., Khirouni, K., Abbes, O., & Bouaicha, M. (2020). Si-nanocrystals embedded in SiNWs to reduce thermalisation in solar cells. *Optik*, 221, 165299.
- Khelifi, M., Mejatty, M., Berrehar, J., & Bouchriha, H. (1985). Effet photovoltaïque dans des couches minces de phtalocyanines. *Revue de physique appliquée*, 20(7), 511-5
- Abdulrazzaq, A. K., Bognár, G., & Plesz, B. (2020). Evaluation of different methods for solar cells/modules parameters extraction. *Solar Energy*, 196, 183-195.
- Abdulrazzaq, A. K., Bognár, G., & Plesz, B. (2020). Accurate method for PV solar cells and modules parameters extraction using I-V curves. *Journal of King Saud University-Engineering Sciences*.
- Christoffersen, P. F., & Diebold, F. X. (1997). Optimal prediction under asymmetric loss. *Econometric theory*, 808-817.
- Feng, Y., Hao, W., Li, H., Cui, N., Gong, D., & Gao, L. (2020). Machine learning models to quantify and map daily global solar radiation and photovoltaic power. *Renewable and Sustainable Energy Reviews*, 118, 109393.
- Kouassi, A. M., Mamadou, A., Ahoussi, K. E., & Biemi, J. (2014). Conception de modèles statistiques à variables hydrochimiques pour la prédiction de la conductivité électrique des eaux souterraines. *LARHYSS Journal P-ISSN 1112-3680/E-ISSN 2521-9782*, (20).
- Chiteka, K., Arora, R., & Sridhara, S. N. (2020). A method to predict solar photovoltaic soiling using artificial neural networks and multiple linear regression models. *Energy Systems*, 11(4), 981-1002.
- Nugrahaeni, R. A., & Mutijarsa, K. (2016, August). Comparative analysis of machine learning KNN, SVM, and random forests algorithm for facial expression classification. In *2016 International Seminar on Application for Technology of Information and Communication (ISemantic)* (pp. 163-168). IEEE.

10. regression, bagging and boosting. *Cogent Economics & Finance*, 8(1), 1729569
11. Shrivastava, S., Jeyanthi, P. M., & Singh, S. (2020). Failure prediction of Indian Banks using SMOTE, Lasso.
Hans, C. (2009). Bayesian lasso regression. *Biometrika*, 96(4), 835-845.
12. Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences*, 51, 350-365.
13. Saporta, G. (2006). Probabilités, analyse des données et statistique. Editions Technip.
14. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
15. Afonso, B., Melo, L., Oliveira, W., Sousa, S., & Berton, L. (2019, October). Housing Prices Prediction with a Deep Learning and Random Forest Ensemble. In *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional* (pp. 389-400). SBC.
16. Sajedi-Hosseini, F., Malekian, A., Choubin, B., Rahmati, O., Cipullo, S., Coulon, F., & Pradhan, B. (2018). A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. *Science of the total environment*, 644, 954-962.
17. Wen, C., Liu, S., Yao, X., Peng, L., Li, X., Hu, Y., & Chi, T. (2019). A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. *Science of the total environment*, 654, 1091-1099.
18. Rajalakshmi, S., & Titus, S. (2020). Optimal Slewing Mode Converter-based Energy Management System for Renewable Energy sources. *International Journal*, 9(3).