Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 9 (2023) 512–518
www.elsevier.com/locate/egyr

7th International Conference on Renewable Energy and Conservation, ICREC 2022, November 18–20, 2022, Paris, France

Photovoltaic sizing using machine learning

Haytham M. Dbouk[a,*], Mahdi Chehimi[b], Aya Khalaf[a]

[a] *American University of Beirut, Beirut, Lebanon*
[b] *Virginia Tech, VA, United States of America*

Received 19 August 2023; accepted 3 September 2023
Available online 22 September 2023

Abstract

Renewable energy is the future of energy in the world. Solar energy is a major renewable energy source which addresses the energy deficiency problem in various countries. This endless energy source is captured using photovoltaic systems. However, in order to make the best out of photovoltaic systems, the size of the system must be optimized. Sizing has always been a critical issue that leads to huge losses once not properly designed. Existing methods consider sizing problem as a case dependent optimization problem that requires data from the place where the photovoltaic system is to be installed. This study introduces machine learning algorithms into the photovoltaic systems' sizing problem for the first time. The proposed case-independent model automates the sizing process and is generalized to be applied at any location. Linear regression, polynomial regression, neural network, random forest, and decision tree were implemented and compared in terms of accuracy. Results show that random forest is the best suited algorithm for this application.
© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Peer-review under responsibility of the scientific committee of the 7th International Conference on Renewable Energy and Conservation, ICREC, 2022.

Keywords: Machine learning; Photovoltaic sizing; Linear regression; Polynomial regression; Neural network; Random forest; Decision tree; Energy optimization

1. Introduction

Nowadays, the world is facing a global energy crisis affecting the provision of electricity for households, industries, and entire countries [1]. This is mainly due to the reliance on non-renewable energy sources, which poses environmental and energy security challenges. Furthermore, due to political and economic circumstances, various countries are facing shortages in conventional energy sources, which threatens their energy security. Lebanon is an exemplar of these countries since it has been suffering a serious energy deficiency problem over the past decades. Daily power cuts in state-provided electricity now last 22 h in some parts of the country [2]. In 2012, around 98% of the energy produced in such countries, e.g., Lebanon, was derived from using imported petroleum products and coal, while only 2% came from renewable energy [3]. Recently, renewable energy technologies, especially solar

* Corresponding author.
E-mail address: hmd13@mail.aub.edu (H.M. Dbouk).

https://doi.org/10.1016/j.egyr.2023.09.025
2352-4847/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Peer-review under responsibility of the scientific committee of the 7th International Conference on Renewable Energy and Conservation, ICREC, 2022.

energy sources, have drawn greater global attention due to their capability to meet the growing energy demand [4]. Various suffering countries are rich in solar radiation, where the potential of photovoltaic (PV) energy is multiple orders of magnitude larger than the demand of the countries [5]. Since 2014, installing PV farms has been one of the most promising solutions for the energy problem in different countries that are economically and politically suffering [6]. Moreover, by increasing the dependence on renewable energy for low-income families, the energy poverty problem will be overcome [7]. In fact, harnessing solar energy using PV farms would save millions of dollars in developing countries on the long run. For instance, PV farms would save around $500 million/year in a long-term vision in Lebanon [8].

In order to address the energy crisis locally and globally, governments, companies, and individuals are investing in renewables. Over the past few years, renewables, especially solar energy, have been increasingly contributing to electricity provision. However, to maximize the benefit of PV systems, it is essential to optimize the corresponding system size. In fact, improper sizing of PV systems causes tremendous energy losses and thus majorly higher costs [9]. An important parameter that relates to the size of PV systems is the load factor, also called plant factor. This is the ratio between the generated and rated energy of the plant during a specific period. When the load factor of the PV system is oversized, the rated energy is decreased. Thus, the generated PV power is truncated at the rated energy of the plant, causing huge power losses [10]. An example is a PV plant located in UK with 39% PV array oversizing. Data shows that the plant has an annual load factor of 15.43%. It is shown that if oversizing was neglected, the yearly load factor of the plant would be 11.11%. This would result in reducing the wasted energy and positively impacting the levelized cost of energy (LCOE) [10,11]. Both, over-sizing and under-sizing of PV systems increase the encountered energy losses and thus the CAPEX and OPEX [12] of the corresponding system. Proper PV sizing leads to efficient PV systems with minimized operating costs and energy losses as well [13]. Thus, PV sizing optimization is becoming more and more crucial considering that is aligned with the Lebanese vision to promote and support installations of Rooftop Solar PV Systems [14], as well as promoting and supporting Concentrated Solar Power (CSP) Technologies [14].

This study presents a case-independent machine learning model that automates the PV sizing process.

2. Related work

PV sizing is considered an optimization problem in the literature. The majority of existing PV sizing methods are heuristic, such as Particle Swarm Optimization (PSO) and Genetic Algorithm (GA), which do not guarantee reaching optimal solution [15], in addition to few other studies that use mathematical programming models to optimally size PV systems. Khare et al. (2014) optimizes the size of a solar photovoltaic (SPV) system and its lifetime cost [16]. The objective is to minimize the annual cost of an on-grid SPV system over its operation lifetime. This cost includes the investment, replacement, and maintenance costs, in addition to profits. Thus, the sizing problem is modeled as a multivariable nonlinear optimization problem and the adopted solving method is PSO algorithm.

Berbaoui et al. (2020) optimizes the size of hybrid power systems [17]. The authors find the optimal number of solar panels that must be added to diesel generator systems with and without batteries to study the effects of having batteries in these hybrid systems. Optimization search algorithm adopted is the Virus Colony algorithm (VCS) which is more effective than both PSO and Modified PSO (MPSO) algorithms. Simulations prove that using batteries is the most gainful hybrid system. Sharafi and ELMekkawy (2014) optimized the design of hybrid renewable energy systems [18]. Considers minimizing the system cost, CO2 emissions, and the probability of load loss. Adopted method is PSO and the proposed approach is computationally efficient. Maleki et al. (2014) applied artificial bee swarm optimization (ABSO) algorithm to minimize the total annual cost in order to obtain the proper size of a hybrid energy system by including maximum allowable loss of power supply probability (LPSP- max) as a constraint to guarantee the system performance [19]. Engel (2016) applied the multi-agent adaptive fuzzy neuronet method to properly size PV systems minimizing the cost. The objective function considered in the optimization problem consists of the system's lifetime cost and the PV module power [20].

Analyzing the existing PV sizing methods, it is noted that the majority of existing studies consider hybrid systems e.g. PV/wind systems. On top of that, existing methods are case dependent; meaning that proposed methods are based on data from a certain city, and in order to apply the methods in other cities, extensive amount of data from that city must be provided first. Khenissi et al. (2021) optimized the sizing of a grid connected PV system in Tunisia. The authors implemented and compared GA with PSO to guarantee the reliability and feasibility of the system by optimizing the number of PV modules, batteries, and inverters [21]. The work of Ramli et al. (2015) considered

optimizing the size of a hybrid PV/wind/diesel system with battery storage [22]. The developed model is based on data of the city of Yanbu, Saudi Arabia. Okoye and Solyal ı (2017) optimized the size of stand-alone PV systems in residential buildings [23]. Their proposed approach optimizes the number of PV panels and batteries taking into consideration the annual cost of the system to assure its economic feasibility. The model is case-dependent, as it is based on site-specific data taken from Bursari, Nigeria. Also, integer programming is employed to solve the optimization problem. The model resulted in 30% reduction of cost compared to diesel generators used at Bursari.

On the other hand, all existing machine learning (ML) models which address PV systems do not tackle the sizing problem. Machine Learning algorithms are heuristically deployed in the literature to assist determining the solar irradiates at given locations and predicting the PV loads. Global Horizontal Solar Irradiance (GHI) is one of the main pillars for appropriate PV sizing. Jiang (2017) applies sparse quadratic RBF neural network [24], while Jiang and Dong (2017) fapplied Kernel Support Vector Machine (SVM) regression algorithm for GHI forecasting [25]. Similarly, Li et al. (2016) apply Hidden Markov Model and SVM regression algorithms to forecast solar irradiance as well as weather/temperature conditions affecting PV systems [26]. Recently, Mahmoud et al. (2019) developed a ML platform using regression trees to assess energy losses for PV systems [27]. Similarly, Voyan et al. (2017) provided an overview of forecasting methods of solar irradiation using ML approaches, it shows that regression tree, random forest, gradient boosting and many other methods begin to be used in this context of prediction, on top of methodologies like neural networks or support vector regression [28]. This work applies ML algorithms for solving the problem of PV sizing. The main contributions of this work are summarized as follows:

- Introduce ML into PV systems' sizing process.
- Develop a case independent model that fulfills the needs of any required PV farm
- Optimize PV systems' sizing process

3. Methodology

The overall system design is presented in Fig. 1. Initially, information about the site location where the PV system will be installed as well as the total power load needed are given as inputs to the proposed system. Then, data preparation is performed to set the required inputs for the ML model and the system provides as output the PV size. The details of the dataset used for developing the model are described in the following subsection.

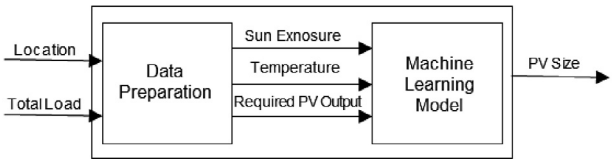

Fig. 1. System model showing inputs and outputs.

3.1. Data collection and preparation

The dataset used to train and test the ML models is collected using two tools. The first tool is Photovoltaic Geographical Information System [29], it takes as input the location data with the required total electrical load to give three parameters:

- Sun exposure for the location in which the corresponding PV unit is implemented
- Average temperature of the location in which the corresponding PV unit is implemented
- The actual corresponding electrical load that should be covered by the PV system

These output parameters are inputted into the second tool, PVOutput [30], to provide the size of the corresponding implemented PV system. The collected data describes the performance of various operating PV systems of different sizes in different locations in the world.

The size of the implemented systems is expressed both in Kilo Watt (KW) and as number of PV cells, where the number of PV cells used is determined by dividing the system capacity in KW by the capacity of each PV cell in KW. One thousand different data entries were collected and used for training and testing the designed ML model.

4. Proposed methods

This project applies supervised learning algorithms. Thus, the dataset includes output labels for every input instance. The output of the designed ML model is the optimal size of the corresponding PV system. The size of a PV system is considered a continuous variable since the size of PV systems varies dramatically based on the required power load, the site location, and the environmental conditions. In this paper, linear regression is investigated to model the system, and its hypothesis function is represented in Eq. (1), in addition, to polynomial regression, a non-linear regression algorithm, whose hypothesis is represented in Eq. (2). In both models, θ's represents the model parameters to be optimized by minimizing the cost function (Eq. (3)) using the gradient descent algorithm.

$$h_\theta(x) = \theta_0 + \theta_1.x, \tag{1}$$

$$h_\theta(x) = \theta_0 + \theta_1.x + \theta_2.x^2 + \theta_3.x^3 \ldots, \tag{2}$$

$$.J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - y^{(i)}\right)^2. \tag{3}$$

Other non-linear regression algorithms are tested in this paper including Decision Trees which create regression models in the form of a tree structure. Then, they break down the dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. Also, the further extension of Decision Trees is considered which is the Random Forest Regression. Finally, a Neural Network (NN) was built to model the system. The general hypothesis of NNs is presented in Eq. (4)., where θ are the NN weights that are tuned:

$$h_\theta(x) = g\left(\theta^{(L-1)^T} a^{(L-1)}\right) = \frac{1}{1 + e^{-(L-1)^T a^{(L-1)}}} \tag{4}$$

5. Experimental results

In this paper, we adopt 310 W PV cells to determine the optimal size of the PV system, which has higher efficiency compared to other PV cells of the same dimensions [31].

5.1. Experimental setup

The designed models were coded in Python language using Spyder and Google Colab platforms. The major libraries used are Scikit Learn [32] and Tensorflow (Keras), in addition to pandas, numpy and matplotlib. The features were first Standardized by removing the mean and scaling to unit variance. Next, the data was split between training and testing data and the split applied is 0.8:0.2. After that, k-fold Cross-Validation was applied to the training data to minimize sampling bias in the data and make sure the model had unbiased generalization performance. The hyper-parameters that were tuned are:

- The random state of the dataset
- The number of estimators and the random state of the Random Forest Regressor
- The number of folds (k) used for validation
- The number of layers of the neural network

The adopted measure of performance is the mean absolute percentage error (MAPE) whose formula is the following:

$$error = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y_{test} - y_{predicted}}{y_{test}} \right| \tag{5}$$

6. Results and discussion

Six models using the six proposed algorithms were built and trained on the training data. Then the performance of the models was tested on the testing data, where the MAPE error was calculated for each model and the accuracy was considered to be 100%-MAPE error. The number of folds used for validation was tuned and the best performance was obtained from the 5-fold cross validation case. The accuracy values obtained for all models are shown in Fig. 2.

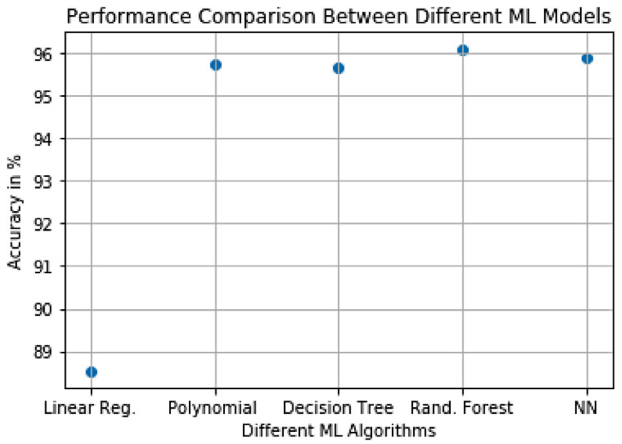

Fig. 2. Comparison between the performance of different algorithms.

From the accuracy results, it is noticed that all models have relatively high accuracy, which shows that the features considered as input for the system are satisfactory representatives of the PV sizing problem. The resulted weight of the linear regression model shows that the weight of the PV load feature is four times that of the sun exposure feature, which shows the importance of PV sizing optimization. The model that has the best performance is the Random Forest Regression. This is not surprising because of the tabular nature of the generated dataset. It is well-known that NNs do not achieve the best performance when dealing with tabular data, and they require the data to have more complexities in order to outperform other models. However, the performance of Random Forest can be improved by tuning its hyper-parameters. In fact, after tuning the number of estimators and the random state of the Random Forest Regressor, in addition to the random state of the dataset itself, the accuracy of the model increased to 98.31% as shown in Fig. 3.

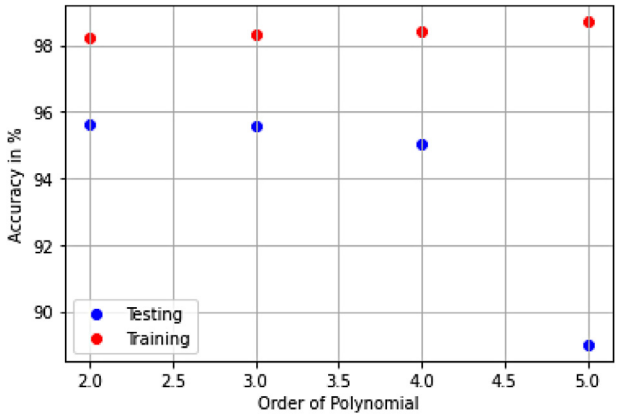

Fig. 3. Effect of increasing order of polynomial regression on overfitting.

Another point that was expected is that linear regression had the worst performance while all other nonlinear regression models performed better. This is a clear indication of the non-linearity of the sizing problem of PV system. On the other hand, neural networks performance on the data requires further tuning of its corresponding hyper-parameters to further improve its performance. Regarding polynomial regression, the model performed better than linear regression when its order was of second and third degree. However, after further increase in the order of the polynomial (e.g. n = 4, 5), the model performed badly and the error rose to around 12%. This is due to

overfitting which happens as the order of the polynomial becomes large, meaning that the model corresponds closely to the training data but does not generalize to new data.

7. Conclusion and future work

This paper presents the first machine learning model for optimizing the size of PV systems. The presented work shows a generalized case-independent model that automates the sizing process and is generalized to be applied at any location and for any targeted load. The designed model takes as input the sun exposure, temperature, and load to be fulfilled by PV and accordingly gives the optimal size of the system as the number of the 310 W cells needed. This paper compares between several machine learning designed models including linear regression, polynomial regression, random forest regression, decision trees and Neural networks. Future work will focus on upgrading this designed model to the sizing of concentrated solar power (CSP), which will revolutionize the solar domain and will be boosted by several countries and initiatives all over the world [9].

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgment

This work was funded and supported by the Research and Development Department at Innovating Green Technology (I.G.T.) LLC, Lebanon.

References

[1] Birol F. What does the current global energy crisis mean for energy investment?. 2022, [Online]. Available: https://www.iea.org/commentaries/what-does-the-current-global-energy-crisis-mean-for-energy-investment.
[2] Lauphlin S. Why Lebanon needs renewable energy now. 2021, [Online]. Available: https://www.thenewhumanitarian.org/opinion/2021/11/11/why-Lebanon-needs-renewable-energy-now.
[3] Kinab E, Elkhoury M. Renewable energy use in lebanon: barriers and solutions. Renew Sustain Energy Rev 2012;16(7):4422–31.
[4] Borovik MR, Albers JD. Participation in the illinois solar renewable energy market. Electr J 2018;31(2):33–9.
[5] Haidar M, Baliozian P, Mustafa A, Mourad S, Preu R. Technical and economic potential of pv in lebanon and jordan aiming for regional readiness level development. In: 33rd European photovoltaic solar energy conference and exhibition. The Netherlands: At Amsterdam; 2017.
[6] Bayeh CZ, Moubayed N. Comparison between pv farm solar chimney and csp tower in lebanon: Economical study for 100 mw power plant. In: International conference on renewable energies for developing countries 2014. IEEE; 2014, p. 205–10.
[7] Lee J, Shepley MM. Benefits of solar photovoltaic systems for low-income families in social housing of Korea: Renewable energy applications as solutions to energy poverty. J Build Eng 2020;28:101016.
[8] Berjawi AEH, Najem S, Faour G, Abdallah C, Ahmad A. Assessing solar pv's potential in lebanon. In: Issam fares institute for public policy and international affairs. American University of Beirut; 2017.
[9] Chowdhury S, Matlokotsi T. Optimal placement and sizing of renewable distributed generation in electricity networks considering different load models. In: 2017 52nd international universities power engineering conference (UPEC). 2017, p. 1–6.
[10] Müller N, Kouro S, Zanchetta P, Wheeler P, Bittner G, Girardi F. Energy storage sizing strategy for grid-tied pv plants under power clipping limitations. Energies 2019;12(9).
[11] BEIS. Digest of United Kingdom energy statistics 2018. Department for Business, Energy Industrial Strategy; 2018, p. 1–260.
[12] Ioannou A, Angus A, Brennan F. Parametric capex, opex, and lcoe expressions for offshore wind farms based on global deployment parameters. Energy Sources B 2018;13(5):281–90.
[13] Khawaja Y, Giaouris D, Patsios H, Dahidah M. Optimal cost- based model for sizing grid-connected pv and battery energy system. In: 2017 IEEE jordan conference on applied electrical engineering and computing technologies (AEECT). 2017, p. 1–6.
[14] Decision to support installations of rooftop solar pv systems in lebanon. 2019, http://lcec.org.lb/en/LCEC/NewsAndMediapage=2, accessed: 2019-12-13.
[15] Modra'k V, Pandian RS. Operations management research and cellular manufacturing systems: Innovative methods and approaches. Business Science Reference; 2012.
[16] Khare A, Rangnekar S. Optimal sizing of a grid integrated solar photovoltaic system. IET Renew Power Gener 2014;8(1):67–75.
[17] Berbaoui B, Dehini R, Hatti M. An applied methodology for optimal sizing and placement of hybrid power source in remote area of south algeria. Renew Energy 2020;146:2785–96.

[18] Sharafi M, ELMekkawy TY. Multi-objective optimal design of hybrid renewable energy systems using pso-simulation based approach. Renew Energy 2014;68:67–79, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148114000408.
[19] Maleki A, Askarzadeh A. Artificial bee swarm optimization for optimum sizing of a stand-alone pv/wt/fc hybrid system considering lpsp concept. Sol Energy 2014;107:227–35.
[20] Engel EA. Sizing of a photovoltaic system with battery on the basis of the multi-agent adaptive fuzzy neuronet. In: 2016 International conference on engineering and telecommunication (EnT). 2016, p. 49–54.
[21] khenissi I, Fakhfakh MA, Sellami R, Neji R. A new approach for optimal sizing of a grid connected PV system using PSO and GA algorithms: Case of Tunisia. Appl Artif Intell 2021;35(15):1930–51.
[22] Ramli MA, Bouchekara H, Alghamdi AS. Optimal sizing of pv/wind/diesel hybrid microgrid system using multi- objective self-adaptive differential evolution algorithm. Renew Energy 2018;121:400–11, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148118300648.
[23] Okoye CO, Solyalı O. Optimal sizing of stand-alone photovoltaic systems in residential buildings. Energy 2017;126:573–84.
[24] Jiang H. A novel approach for forecasting global horizontal irradiance based on sparse quadratic rbf neural network. Energy Convers Manage 2017;152:266–80.
[25] Jiang H, Dong Y. Forecast of hourly global horizontal irradiance based on structured kernel support vector machine: A case study of tibet area in china. Energy Convers Manage 2017;142:307–21.
[26] Machine learning for solar irradiance forecasting of photovoltaic sys- tem. Renew Energy 2016;90:542–53.
[27] Mahmoud K, Abdel-Nasser M. Fast yet accurate energy-loss- assessment approach for analyzing/sizing pv in distribution systems using machine learning. IEEE Trans Sustain Energy 2019;10(3):1025–33.
[28] Voyant C, Notton G, Kalogirou S, Nivet M-L, Paoli C, Motte F, Fouilloy A. Machine learning methods for solar radiation forecasting: A review. Renew Energy 2017;105:569–82.
[29] Pv output data. 2019, https://pvoutput.org/outputs.jsp, accessed: 2019-10-6.
[30] Photovoltaic geographical information system. 2019, https://re.jrc.ec.europa.eu/, accessed: 2019-10-7.
[31] 310W pv cell. 2019, jasolar.com/uploadfile/2018/0803/20180803102956815.pdf, accessed: 2019-12-13.
[32] Pedregosa Fabian, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825–30.