Optoelectronic devices informatics: optimizing DSSC performance using random-forest machine learning algorithm

Omar Al-Sabana and Sameh O. Abdellatif*

Electrical Engineering Department, Faculty of Engineering and FabLab in the Centre for Emerging Learning Technology (CELT), The British University in Egypt, Cairo 11387, Egypt

(Received 12 July 2021; Revised 5 September 2021)

This paper provides an attempt to utilize machine learning algorithm, explicitly random-forest algorithm, to optimize the performance of dye sensitized solar cells (DSSCs) in terms of conversion efficiency. The optimization is implemented with respect to both the mesoporous TiO_2 active layer thickness and porosity. Herein, the porosity impact is reflected to the model as a variation in the effective refractive index and dye absorption. Database set has been established using our data in the literature as well as numerical data extracted from our numerical model. The random-forest model is used for model regression, prediction, and optimization, reaching 99.87% accuracy. Perfect agreement with experimental data was observed, with 4.17% conversion efficiency.

Document code: A **Article ID:** 1673-1905(2022)03-0148-4

DOI https://doi.org/10.1007/s11801-022-1115-9

Silicon-based solar cells are the most well spread commercial light harvesters worldwide, especially in large-scale applications[1]. However, silicon-based cells showed limitations including but not limited to, fairy high manufacturing cost, non-transparency, large area occupation, and low performance under diffused light. Accordingly, new generations of solar cells started to glow up[2]. Among various new structures and materials, dye sensitized solar cells (DSSCs) have been introduced in the literature as a low cost, simply fabricated and efficiently operated under low-light intensity solar cells[3,4]. In DSSCs, the active thin film layer is fabricated using mesoporous materials that include empty gaps, where the dye can penetrate[5-7].

Previous investigations in the literature have been reported toward the optimization of these DSSCs, seeking for the maximum conversion efficiency[8,9]. Such optimization processes were mostly based on experimental investigations, where very few theoretical attempts were demonstrated[4]. In both cases, experimental or numerical optimizations, the efficiency maximization process was limited to a set of restricted boundaries without any prediction capabilities.

Recently, the term material informatics has shined in the literature as a booming optimization technique based on machine learning algorithms[10]. Materials informatics is a study field that focuses on investigating and applying the techniques of informatics to materials science and engineering[10]. Previous studies dealing with integrating machine learning model in optimizing solar cell performance were presented in Refs.[11] and [12]. Specifically, the works in Refs.[13—16] were introduced for machine learning algorithms related to DSSC optimization. In Ref.[14], a valuable attempt was demonstrated to optimize the sensitizers used in DSSC fabrication. Alternatively, Ref.[15] illustrated an artificial integument model for exploring various organic dyes capable of being integrated in organic DSSCs. However, by screening the literature, no previous investigation on the optimization of the active mesoporous layer for maximizing DSSC efficiency was addressed. We believe that the mesoporous TiO_2 layer is the main critical layer in DSSC design[5,6]. TiO_2, in its porous structure, is not only functionalized as the electron transport medium in DSSCs, but also it is considered as the hosting medium to the dye, through its pours. Additionally, the scattering mechanisms in the porous structure can attribute to the optical path length in the active layer, which reflects in the absorption enhancement.

Consequently, this study presents an optimization process for the cell conversion efficiency through utilizing machine learning algorithm. Random-forest algorithms is used through python scripting code[17]. As inputs, the mesoporous TiO_2 thickness and porosity are considered, while the conversion efficiency is promoted as the main targeted output. The porosity variation is studied as variation in the effective thickness and the effective permittivity of the TiO_2-dye effective medium (Fig.1). The machine learning model is seeded with a dataset extracted from our previous experimental work in Refs.[3—6], along with numerical simulation results conducted using scanner

* E-mail: Sameh.osama@bue.edu.eg

application software (SCAPS) numerical tool[18].

As an essential step in a machine learning model, a dataset of inputs/output matrix is implemented. To ensure the reliability of the constructed model, the core dataset is imported from a real experimental data for segmented and complete DSSCs[3-6]. However, the limitation in the number of fabricated samples associated with any experiential process leads to the integration of numerical data to provide the acceptable dataset range for the machine learning model.

Herein, we utilize SCAPS model, which is already verified with respect to experiential data in our work in Ref.[4]. SCAPS allows us to examine the variation in the overall efficiency against each design parameter. In our model the mesoporous layer and the dye are combined in an effective medium layer (layer1). The optical effective permittivity of such medium can be addressed through the effective medium theory, given by

$$\varepsilon_{\text{eff}}=\varepsilon_{\text{m}}\frac{2\delta_{\text{i}}(\varepsilon_{\text{i}}-\varepsilon_{\text{m}})+\varepsilon_{\text{i}}+2\varepsilon_{\text{m}}}{\varepsilon_{\text{i}}+2\varepsilon_{\text{m}}-2\delta_{\text{i}}(\varepsilon_{\text{i}}-\varepsilon_{\text{m}})}, \tag{1}$$

where ε_{m} is the medium permittivity (bulk TiO_2), ε_{i} is the inclusion permittivity (dye), and δ_{i} is the fraction volume of inclusion (porosity).

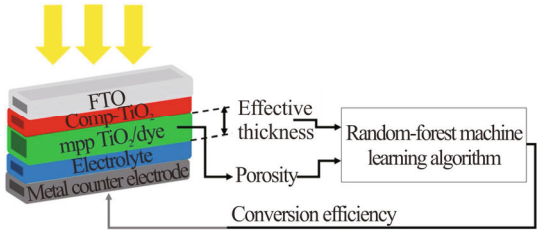

Fig.1 Demonstrative schematic for DSSC with the effective mesoporous (mpp) TiO_2/dye layer (The figure also shows the machine learning model with its input and output)

On the other hand, the variation in the effective layer thickness can be attributed to either variation in the mesoporous TiO_2 layer thickness or in the porosity. In other words, fluctuation in porosity is contributed to the model as a variation in the effective permittivity (in terms of δ_{i}) as well as change in the overall effective thickness of the TiO_2/dye layer. The idea beside that is very simple, as the porosity increase, the capability of absorbing dye in TiO_2 voids increases, proportionally. Accordingly, the overall absorption increases. This increase in the absorption is modeled as an increase in the effective layer, while keeping the absorption coefficient of the dye (the imaginary part of the permittivity ε) fixed. In the current study, we consider the extinction behavior of the effective TiO_2/dye layer as a combination of the volume scattering due to the mesoporous structure, and the absorption spectrum of the dye[4]. The volume scattering in the TiO_2 is modelled using Mie scattering model[5].

Experimentally, TiO_2 nanoparticles were prepared using the sol-gel method described in Ref.[7]. Consequently, deposition process took place using screen printing method. Commonly used N719 dye was utilized as a visible dye, while iodine-based electrolyte was inserted[4]. DSSCs were sealed using a hot press at a temperature of 120 °C. For the sake of characterization, indoor C-band light emitting diode (LED)-based solar simulator was used, with Keithley 2401 current-voltage source meter[19]. The overall cell efficiency is calculated as

$$\eta_{\text{E}}=\frac{V_{\text{OC}}J_{\text{SC}}FF}{I_{1\text{S}}\cdot\lambda_{1\text{S}}\cdot A}, \tag{2}$$

where V_{OC} stands for the open-circuit voltage, J_{SC} represents the short-circuit current, FF labels the fill factor, $I_{1\text{S}}$ is equal to 1 kW/m² which is approximately the AM1.5 Sun intensity[20], and A is the cell area. Herein, we introduce the term $\lambda_{1\text{S}}$ as the spectral mismatching correction factor. To investigate such a mismatching coefficient, a V-770 ultraviolet-visible/near infrared (UV-Vis/NIR) spectrophotometer was used, where the wavelength from 190 nm to 2 700 nm can be investigated. The process was carried out by simply placing the optical LED array source in aliment with the optical fiber detector then the spectrometer sweeps the sample and displays the result on its compatible software. Logically, mismatching is considered with respect to the AM1.5G, where 9% deviation was observed.

For the sake of optimization, the dataset implemented is seeded in a random-forest machine learning algorithm[17]. Random-forest has gained high popularity due to its simplicity to use, flexibility, and ability to handle both regression, and classification problems[17]. The input data is divided into two sets, the training set, and the validation set. The training data set is the core of the machine learning process, where the algorithms analyses and learns on. Alternatively, the validation data set is created to ensure that no overfitting or biasing in the machine learning process occurs. Validation data is mostly a part of the main dataset, which is kept from the algorithm until it finishes training, then it is applied on the trained module to analyses its performance on such new data from the machine learning model prospective. In this investigation, random-forest is chosen as an accurate classification algorithm that uses decision trees for learning method[17]. In random-forest, as the algorithm trains the data, a number of decision trees are constructed, resulting in a modal made by the prediction of each of the individual trees[17]. The flow of the algorithm can be demonstrated in the flow chart shown in Fig.2[16].

Following the procedure described above, the random-forest algorithm is applied on the generated dataset. It is observed in Fig.3 that the prediction performance of the module applied on the dataset is perfectly matches the actual inputted efficiencies, where the model estimated accuracy is found to be 99.87%. By tracing the model, both the thickness and porosity of the TiO_2/dye effective layer have a great effect on the efficiency of the solar cell. Nearly 49.53% of the total increase in the cell

efficiency being due to the thickness and about 50.47% is due to the porosity. This means that the influence of the porosity is a bit greater than that of the thickness. The prediction of the efficiency variation against the TiO_2/dye effective thickness and porosity reached its optimum value of 4.17% at effective thickness of 7 μm and porosity of 63%.

Finally, the optimized *J-V* curve for the simulated DDSC is plotted against experimental data captured from our recipe in Ref.[4] (Fig.4). A perfect matching is detected between both curves with an overall conversion efficiency of 4.17%. It is worth to highlight that the efficiency observed in this work is not considered as the highest efficiency reported for DSSC in general. This limitation is a part of the constrains associated to the recipe and material used. The reason of selecting such relatively low efficiency solar cell is attributed to the existence of a fabricated set of samples[3-6], from which our database has been constructed. However, we believe that the know-how presented in this work can be extended to other high efficiency DSSCs as well as perovskite solar cells, where we consider this part as a future work.

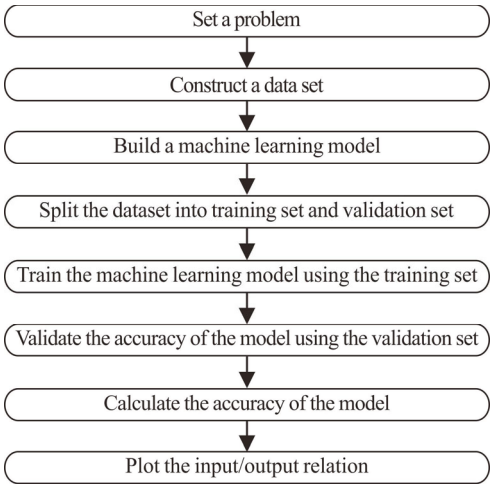

Fig.2 Random-forest machine learning algorithm used in optimizing the power conversion efficiency of DSSC[16]

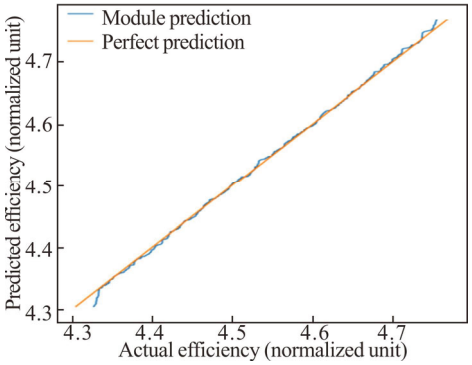

Fig.3 Actual efficiency against predicted efficiency resulted from the random forest machine learning model (The module prediction is plotted with respect to the perfect (hypothetical) prediction, where slope tends to one)

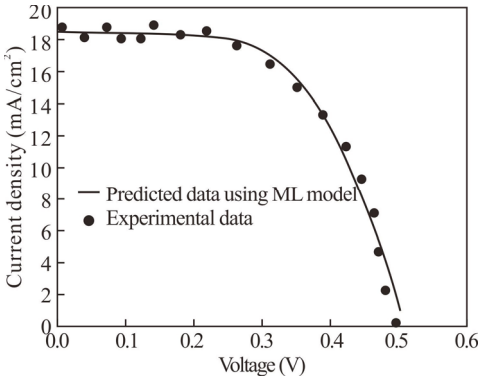

Fig.4 Predicted optimum *J-V* curve under effective thickness of 7 μm and porosity of 63% against experimental data

In conclusion, for enhancing and predicting the performance of DSSCs, the random-forest algorithm provides both efficiency and flexibility to be applied. Optimization for the mesostructured thin film layers were modelled using experimental data, and SCAPS 1-D simulation model. The machine learning modules was built using python random forest regressor algorism. Perfect matching between optimized predicted data and experimental results was observed with model accuracy tends to 99.87%. An optimized DSSC with 4.17% efficiency was concluded under effective thickness of 7 μm and porosity of 63%.

Statements and Declarations

The authors declare that there are no conflicts of interest related to this article.

References

[1] SÁNCHEZ M, FRANCISCO J, STEINER M, et al. Worldwide energy harvesting potential of hybrid CPV/PV technology[J]. Joule, 2021, 24(7): 971-987.

[2] AHMAD K S, NAQVI K S, JAFFRI S B. Systematic review elucidating the generations and classifications of solar cells contributing towards environmental sustainability integration[J]. Reviews in inorganic chemistry, 2021, 41(1): 21-39.

[3] ABDELLATIF S O, JOSTEN S, KHALIL A S G, et al. Transparency and diffused light efficiency of dye-sensitized solar cells: tuning and a new figure of merit[J]. IEEE journal of photovoltaics, 2020, 10(2): 522-530.

[4] MOUSTAFA M M, ISMAIL Z S, HASHEM E M, et al. Investigating the tradeoff between transparency and efficiency in semitransparent bifacial mesosuperstructured solar cells for millimeter-scale applications[J]. IEEE journal of photovoltaics, 2021, 11(5): 222-235.

[5] ABDELLATIF S, SHARIFI S, KIRAH K, et al. Refractive index and scattering of porous TiO_2 films[J]. Microporous and mesoporous materials, 2018, 264: 84-91.

[6] ABDELLATIF S, JOSTEN S, SHARIFI P, et al. Optical

investigation of porous TiO_2 in mesostructured solar cells[C]//Proceedings of Physics and Simulation of Optoelectronic Devices XXVI, International Society for Optics and Photonics, January 27-February 1, 2018, San Francisco, CA, USA. Washington：SPIE, 2018：105260A.

[7] HATEM T, ISMAIL Z, ELMAHGARY Z, et al. Optimization of organic meso-superstructured solar cells for underwater IoT² self-powered sensors[J]. IEEE transactions on electron devices, 2021, 68(10)：5319-5321.

[8] MARRI A R, MARCHINI E, CABANES V D, et al. Record power conversion efficiencies for iron (II)-NHC-sensitized DSSCs from rational molecular engineering and electrolyte optimization[J]. Journal of materials chemistry A, 2021, 9(6)：3540-3554.

[9] SHAH S A A, GUO Z, SAYYAD M H, et al. Optimizing zinc oxide nanorods based DSSC employing different growth conditions and SnO coating[J]. Journal of materials science：materials in electronics, 2021, 32(2)：2366-2372.

[10] WARD L, WOLVERTON C. Atomistic calculations and materials informatics：a review[J]. Current opinion in solid state and materials science, 2017, 21(3)：167-176.

[11] MAHMOOD A, WANG J L. Machine learning for high performance organic solar cells：current scenario and future prospects[J]. Energy & environmental science, 2021, 14(1)：90-105.

[12] CHOUDHARY K, BERCX M, JIANG J, et al. Accelerated discovery of efficient solar cell materials using quantum and machine-learning methods[J]. Chemistry of materials, 2019, 31(15)：5900-5908.

[13] SAHU H, MA H. Unraveling correlations between molecular properties and device parameters of organic solar cells using machine learning[J]. The journal of physical chemistry letters, 2019, 10(22)：7277-7284.

[14] AROOJ Q, WANG F. Switching on optical properties of D-π-A DSSC sensitizers from π-spacers towards machine learning[J]. Solar energy, 2019, 188：1189-1200.

[15] WEN Y, FU L, LI G, et al. Accelerated discovery of potential organic dyes for dye-sensitized solar cells by interpretable machine learning models and virtual screening[J]. Solar RRL, 2020, 4(6)：2000110.

[16] AL-SABAN O, ABDELLATIF S O. Optoelectronic materials informatics：utilizing random-forest machine learning in optimizing the harvesting capabilities of mesostructured-based solar cells[C]//Proceedings of 2021 IEEE International Telecommunications Conference (ITC-Egypt), July 13-15, 2021, Alexandria, Egypt. New York：IEEE, 2021：1-4.

[17] RODRIGUEZ-GALIANO V, SANCHEZ-CASTILLO M, CHICA-OLMO M, et al. Machine learning predictive models for mineral prospectivity：an evaluation of neural networks, random forest, regression trees and support vector machines[J]. Ore geology reviews, 2015, 71：804-818.

[18] EID A A, ISMAIL Z S, ABDELLATIF S O. Optimizing SCAPS model for perovskite solar cell equivalent circuit with utilizing Matlab-based parasitic resistance estimator algorithm[C]//Proceedings of 2020 IEEE 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), October 24-26, 2020, Giza, Egypt. New York：IEEE, 2020：503-507.

[19] HASSAN M M, SAHBEL A, ABDELLATIF S O, et al. Toward low-cost, stable, and uniform high-power LED array for solar cells characterization[C]//Proceedings of New Concepts in Solar and Thermal Radiation Conversion III, International Society for Optics and Photonics, August 24-September 4, 2020, California, USA. Washington：SPIE, 2020：114960Q.

[20] KIM J Y, LEE K, COATES N E, et al. Efficient tandem polymer solar cells fabricated by all-solution processing[J]. Science, 2007, 317(5835)：222-225.