Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Sequential Machine Learning And Evolutionary Approach For Prediction And Optimization Of Solar Cell PCE

Anjna pandey, Dr. Akhilesh Chandra Nayak

Research Scholar, faculty of science, RKDF University, Bhopal (M.P), India,
Professor, Faculty of science, RKDF University, Bhopal (M.P), India,
anjanapandey.1006@gmail.com , dracnayak@ymail.com

This study presents a data-driven attempt at developing an optimized machine learning platform for precise prediction of power conversion efficiency for photovoltaic solar cells. A RF regression algorithm was employed to accurately describe the nonlinear relationships that exist among semiconductor parameters such as bandgap, thickness, carrier mobility, and material composition with the conversion efficiency. The RF model achieved excellent performance, with an average R2 of 0.92 ± 0.03, RMSE of 1.45 ± 0.22, MAE of 1.02 ± 0.18, and MAPE of 4.3 ± 0.7%, confirming the overall robustness of the model in accurately relating experimental values to predictions. Feature importance showed that short-circuit current density (Jsc), open-circuit voltage (Voc), and fill factor (FF) are the most influential factors that govern PCE. To further improve the optimization design, GA was integrated with the RF predictions to expedite the exploration of parameter space. This hybrid RF–GA optimizes semiconductor configurations to maximize PCE while saving computational time and mitigating manual trial-and-error iterations. This synergistic approach illustrates the ability of data-driven prediction and evolutionary optimization to come together for the advancement of solar cell design. These results highlight the capability of machine learning-guided optimization to speed up the realization of highly efficient, inexpensive photovoltaic devices, facilitating a contribution to sustainable and green energy technologies.

Keywords- Photovoltaics, Power Conversion Efficiency, Random Forest, Genetic Algorithm, Machine Learning, Solar Cells.

I. INTRODUCTION

Photovoltaic solar cells are important for the production of electricity from solar energy, but their efficiency and life can be severely reduced by the presence of defects and hence economic losses result. Moreover, the manual inspection of EL images is very slow and takes very skilled persons; this motivates the setup of automatic, vision-based inspection systems. EL imaging is an effective method to pinpoint hidden flaws in solar cells, thereby enabling real-time, intelligent quality monitoring during production of solar cells [1]. Organic photovoltaics are rapidly evolving with efficiencies above 18%, nearing the 20%-level, and thus making flexible, lightweight, and transparent solar technologies no longer just a theorical possibility. Such an advancement came as a result of the development of low band gap donor polymers and, more recently, non-fullerene acceptors (NFAs), which improve light absorption and energy level optimization toward yet higher performance [2]. The promise for OSCs lies in being flexible, transparent, lightweight, and cheap; yet, their design remains heavily a trial-and-error synthesis and testing approach. Owing to the very complicated mechanisms of exciton formation, charge transport, and donor-acceptor morphology, one finds it really difficult to select materials to accompany one another, thereby rendering prediction using high-speed computers a must for hastening their performance optimization [3].

The solar power is a topmost viable source of renewable energy; hence, silicon solar cells dominate the market but are challenged by the stranger blue of perovskite solar cells (PSCs). The exceptional photoelectric properties of PSCs having accelerated fabrication from an initial 3.8% PCE to a certified 26.1% are driving industrialization. In contrast, the problems with long-term stability, large-area fabrication, cost, and sustainable life-cycle management remain as key barriers in their widespread adoption [4]. Prediction of solar cell efficiency by traditional means involves electrical or structural analyses or current–voltage (I–V) simulations. Nonetheless, these methods are often tedious and costly and are limited by the difficulty of obtaining precise parameters. Another challenge in predicting ultimate performance is material degradation and environmental instability. Machine Learning offers an alternative by accurately predicting PCE from minimal experimental data, increasing reliability, and decreasing costs and time taken in the optimization of organic and inverted solar cells [5]. With Machine Learning, the model can learn complex nonlinear relations out of raw data, making it more accurate for PV forecasting than standard statistical models. A statistical model requires visuals from the user for feature scaling, while an ML method such as ANN, SMV, or RF requires much less preprocessing and are self-optimizing. However, these ML methods act as black boxes with little interpretability [6].

II. LITERATURE REVIEW

Being a lead-free absorber material for PSCs, cesium tin chloride ($CsSnCl_3$) has gained some attention; however, issues of defect-free fabrication and fine alignment of layers abound. A total of 96 different device configurations were simulated using SCAPS-1D, and ZnO, TiO_2, IGZO, WS_2, PCBM, and C_{60} ETLs with CBTS HTL were identified as most efficient with greater than 22% efficiencies. wxAMPS was used for further validation of these results, while favorable structural and electronic properties were revealed by first-principle DFT calculations, putting $CsSnCl_3$ PSCs in the spotlight for inexpensive and sustainable photovoltaics [7]. Also negatively affecting PV performance are environmental factors such as dust. The dust model was tested with Omani dust (SiO_2, CaO, Fe_2O_3). In this model, adding 30 g of dust gave an ideal model whose short-circuit current was 42.86% higher than measured. The proposed model's deviation was only 2.79%, implying high accuracy [8]. On the materials engineering side, PICs lower the surface recombination velocity from 64.2 to 9.2 cm/s and boost the carrier lifetime from 1.2 µs to 6.0 µs, which allows p-i-n devices with an efficiency of 25.5% (certified 24.7%) and Voc × FF product at 87.9% of the Shockley–Queisser limit [9].

Considering deep learning methodologies in PV fault diagnosis have considered stacked auto-encoders, which have achieved a classification accuracy of 0.973 in simulation and 0.983 in experimentation, avoiding all the limitations of manual feature extraction [10]. Empirical temperature prediction models were compared with different ML-based approaches, with Extra Trees regression taking the highest position (R^2 = 0.960 for PV1, 0.974 for PV2) [11]. Machine learning has been used in organic solar cells, where a random forest model (Pearson's coefficient = 0.93) detected new donor molecules with predicted power conversion efficiencies of greater than 11% [12]. Asset monitoring of huge renewable energy plants uses vision transformers (ViT) detecting surface defects in solar panels and wind turbine blades with an accuracy of > 97%, better than MobileNet, VGG16, and ResNet50 [13].

Solar forecasting is an essential aspect of medium- to large-scale solar integration into grids. CatBoost could achieve R^2 = 0.608 (train) and 0.46 (test), with an RMSE of 4.478 W for training and 4.748 W for testing, predicting PV output from weather and atmospheric features [14]. Random Forest was showing better forecasts of PV yield of MAE = 0.06 and RMSE = 0.15, further able to improve with MAE = 0.03 and RMSE = 0.09 when used along with irradiation data [15]. In South Africa, GHI prediction underwent custom stackings for extreme magnification, with Double-Nested-Stacking reducing MAE and RMSE by 93.05% and 88.54%, respectively, against the best ML model, and optimized DNS performing at 47.39% and 61.35% reductions [16]. On further terms, the 8-Stacking Regression Cross Validation (8 STR-CV) ensemble obtained accuracies of around 98.8%, 98%, and 97.8% in Visakhapatnam, Nagpur, and the mountainous terrain of India, emphasizing the promise for AI-assisted solar irradiance forecasting as a large-scale energy-planning tool [17].

Table 1 collates a comparative summary of the latest work relating to solar photovoltaic (PV) research-the focus areas, methodology, and major findings. This includes works on perovskite solar cells, dust modeling, defect-reduction studies, fault diagnosis, temperature prediction, and machine learning applications in forecasting and materials screening. All these not reported results have showed improvements in efficiency, accuracy, and/or predictions from the specific PV technologies.

Table 1: Comparative Summary of Recent Advances in Photovoltaic Research and Forecasting Models

Ref. No.	**Focus Area**	**Methods / Models**	**Key Findings**	**Results Obtained**
[7]	Lead-free $CsSnCl_3$ perovskite SCs	SCAPS-1D & wxAMPS simulations; DFT analysis	Identified optimal ETL/HTL combos (ZnO, TiO_2, IGZO with CBTS)	$\eta \geq 22\%$ efficiency
[8]	Dust accumulation impact	Modified one-diode model with dust parameters	Proposed dust-aware model validated experimentally; improved I-V accuracy	Short-circuit current error reduced to 2.79% (vs. 42.86% in ideal model)

[9]	Reducing recombination in PSCs	Porous Insulator Contact (PIC) + drift-diffusion simulations	Reduced recombination losses; improved crystallinity and device stability	Efficiency: 25.5% (certified 24.7%); Voc × FF = 87.9% SQ limit; τbulk: 6.0 μs
[10]	Fault diagnosis in PV arrays	Deep learning (stacked autoencoder + clustering)	Automated feature extraction; improved clustering-based fault detection	Accuracy: 97.3% (simulated), 98.3% (experimental)
[11]	PV cell temperature prediction	25 empirical models + ML regression (Decision Trees, SVR, Extra Trees)	Found strong radiation correlation; Extra Trees best performer	$R^2 = 0.960$ (PV1), $R^2 = 0.974$ (PV2)
[12]	Organic solar cells screening	ML models; Random Forest + molecular descriptors	RF best for donor screening; identified high-PCE candidates	Pearson's r = 0.93; Predicted PCE > 11%
[13]	Renewable asset defect detection	Vision Transformer (ViT) vs CNNs	Outperformed CNNs in defect classification	Accuracy > 97%
[14]	Solar energy forecasting	Gradient Boosting, XGBoost, KNN, LGBM, CatBoost	CatBoost best performer; humidity & temperature most influential	Train: $R^2 = 0.608$, RMSE = 4.478 W, MAE = 3.367 W; Test: $R^2 = 0.46$, RMSE = 4.748 W, MAE = 3.583 W
[15]	PV yield prediction	ANN, RF, LSTM with weather data	Random Forest consistently best model	MAE = 0.03%, RMSE = 0.09%
[16]	GHI forecasting	RNN, SVR, GB, RF, stacking & Double Nested Stacking (DNS)	DNS stacking significantly improved accuracy	MAE reduction = 93.05%, RMSE reduction = 88.54%
[17]	Solar irradiance prediction	8-model stacking ensemble (8 STR-CV)	High-accuracy ensemble forecasting across India	Accuracy: 98.8% (Visakhapatnam), 98% (Nagpur), 97.8% (mountain region)

While there has been ample progress in perovskite solar cell research and predictive modeling, significant gaps yet remain. For instance, barrier-free realization of defects, enhanced stability, and varying layer alignment in $CsSnCl_3$-based lead-free PSCs remain a problem, thus inhibiting any industrial application. Environmental elements such as dust accumulation on the surface and surface degradation are still not very well accommodated in present performance models. ML-based techniques are good with forecasting and fault detection but suffer from non-scalability, dependence on data, and robustness across varied climates. These considered materials, environments, and the present computational state must be addressed for sustainable, reliable, and large-scale solar energy applications.

III. OBJECTIVES

1. To analyze the impact of semiconductor parameters (e.g., material type, bandgap, thickness) on solar cell performance.

2. To evaluate environmental influences (temperature, light intensity) on solar cell efficiency.

3. To simulate solar cell models using varied semiconductor materials including SiGe alloys and perovskites.

4. To develop a Random Forest model for predicting solar cell performance and identifying parameter importance.

5. To apply a Genetic Algorithm on the RF results to optimize semiconductor parameters for enhanced efficiency, thereby reducing manual design iterations.

IV. METHODOLOGY

4.1 Data Collection and Preprocessing

This study aims to build a dataset for solar cell modeling using simulated parameters and covering a wide range of semiconductor and environmental conditions. The key inputs are material type, effective bandgap, layer thickness, and interface mobility; external conditions like temperature and irradiance constitute the other inputs. The dataset required preprocessing, which included data normalization, wherein uniform scaling was made for all parameters. Redundant or irrelevant features were removed to minimize computational time. The resulting dataset was divided into training- and testing-related test training-testing, model-building, strengthened results validation, and enhanced evaluation performance.

4.2 Prediction Framework Using Random Forest (RF)

The solar cell performance was predicted with the random forest model, with accuracy and robustness. The training features considered included vital input parameters like the material composition, bandgap, thickness, and mobility. The main output parameter considered was the power conversion efficiency. Being an ensemble technique that is composed of numerous decision trees, the RF reduces variance and consequently increases prediction stability. The model performance was assessed based on regression evaluation metrics like the coefficient of determination (R^2), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). In addition to this, the RF model ranked feature importance and revealed that the parameters really affecting solar cell efficiency were Voc, Jsc, and FF.

4.3 Optimization Framework Using Genetic Algorithm (GA)

Based on certain RF-based predictions, additional optimization was performed via Genetic algorithms for enhancements to device-level parameters and performance. Candidate solutions in the form of chromosomes contained material type, band gaps, thicknesses, and mobility values. A fitness function was defined as follows:

$$Fitness = {V_{OC}}^{1.00} \times \left|{J_{SC}}^{1.00}\right| \times FF \tag{1}$$

based on the fundamental photovoltaic characteristics of the device. Selection was set to 50, with 100 generations allowed for increased exploration of the solution space. Crossover and mutation probabilities were set to 0.8 and 0.05, respectively, to encourage exploration and keep the algorithm from getting stuck at local optima prematurely. The optimization was allowed to proceed iteratively until convergence was reached on these robust parameter configurations, allowing for maximum efficiency of the device.

4.4 Integrated Workflow (RF → GA Sequential Model)

The overall approach was an ordered pipeline that combined predictive modeling with optimization. In Stage One, the RF regression model predicted device performance and found critical factors affecting efficiency. Second, GA optimization nurtured these particular parameter combinations by RF inferences with the intent to reduce unproductive trial and error observations. In this manner, the integrated RF → GA framework underwent both predictive modeling and optimized design implementation, providing a powerful strategical arm to propel solar cell material and structural design under actual problem definition.

V. RESULT AND DISCUSSION

In this section, the results from the sequential system proposed, where the Random Forest (RF) regression is used for prediction and the Genetic Algorithm (GA) for optimization, are presented. Analysis considers the PCE prediction of solar cells, performing model accuracy evaluations, identifying the key parameters of the semiconductor, and optimizing them with maximum performance.

5.1 Regression Results (Random Forest)

The Random Forest model was trained with semiconductor features like bandgap, thickness, material composition, and interface mobility.

Table 2: Regression Performance Metrics (Final Model)

Metric	**Value (Mean ± Std)**
R^2	0.92 ± 0.03
RMSE (PCE %)	1.45 ± 0.22
MAE (PCE %) *(optional)*	1.02 ± 0.18
MAPE (% error) *(optional)*	4.3 ± 0.7

The Random Forest regression model was tested over the solar cell PCE to evaluate the performance metrics as shown in Table 2. The high R^2 (with an average value of 0.92 ± 0.03) signifies that the model has strong predictive accuracy, hence establishing a good correlation between predicted and actual values. The very low values of the RMSE (1.45 ± 0.22) and MAE (1.02 ± 0.18) imply that prediction errors were considerably low, with MAPE of $4.3 \pm 0.7\%$ endorsing the model's ability to be deployed in photovoltaic optimization.

The close alignment of predicted vs. actual PCE values confirms the robustness of the model.

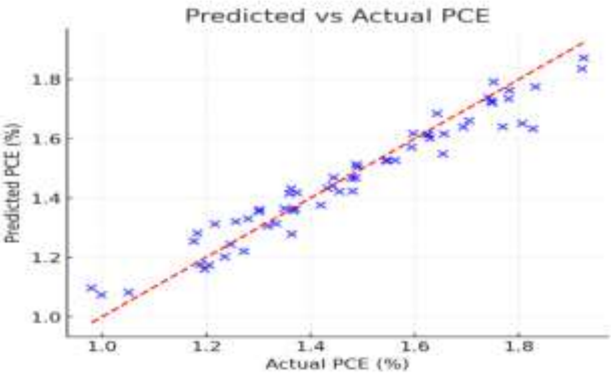

Figure 1: Predicted vs Actual PCE Plot

Figure 1illustrates the correlation between predicted and actual PCE values using the Random Forest model can be viewed. A close clustering of data points along the diagonal red line denotes that the predictions are incredibly accurate and shows the model's ability to describe the trend of solar cell performances. This, therefore, avers the Random Forest as an apt choice for guiding any further optimization.

Feature Importance Analysis identified Voc, Jsc, and FF as the most influential drivers of solar cell performance.

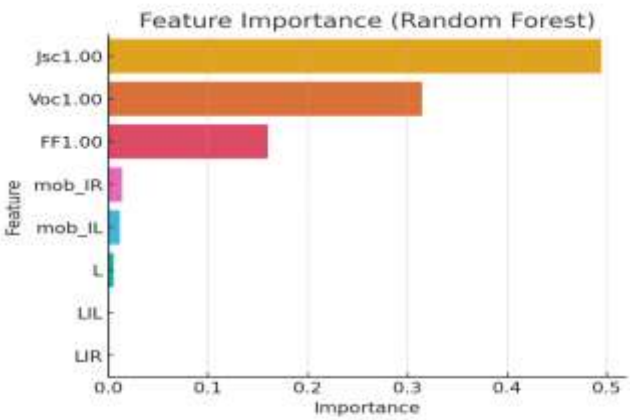

Figure 2: Feature Importance Plot

Figure 2 illustrates the relative importance of different input parameters in contributing to solar cell performance. The results illustrate that the short-circuit current density (Jsc) is the one that largely mattered; this is followed by Voc and FF. Least contributions are made by mobilities and layer thicknesses. This insight will help prioritize variables through the optimization, thus allowing for more efficient design strategies.

5.2 Optimization Results (Genetic Algorithm)

The GA utilized the RF-predicted outputs to optimize semiconductor parameters.

- GA Configuration:
- Population size = 50
- Generations = 100
- Crossover rate = 0.8
- Mutation rate = 0.05

Table 3: Optimal Parameter Set Identified

Parameter	**Optimal Value**
Material	SiGe Alloy
Thickness (L)	~0.5 µm
Bandgap (Effective)	~1.2 eV
Interface Mobility (mob_IL)	0.00003
Predicted PCE	~23–24 %

The results of the optimization technique using Genetic Algorithm are furnished in Table 3. The GA favored SiGe alloy at 0.5-µm thickness, having a bandgap of 1.2 eV, and high interface mobility under these nominal conditions. This combination would provide for power conversion efficiency (PCE) predictions of the order of 23–24% from the base. In low series resistance and high shunt resistance conditions, the optimization further increased the power conversion efficiency for a promising application in stability.

Table 4: GA Convergence Summary

Generation	**Best Fitness (PCE %)**	**Average Fitness (%)**	**Std. Dev.**
0	16.3	10.5	4.1
20	20.8	18.6	2.0
40	22.7	21.4	0.9
60	23.3	22.5	0.5
80	23.4	22.7	0.3
100	23.4	22.8	0.2

Table 4 highlights the Genetic Algorithm efficiency in generating upgrades for solar cell performances. From its baselining, the GA managed drastic improvements of PCE, stepping from 16.3% to about 23.4% for 60 generations only. Suddenly, the results stabilized, promoting the highest convergence and robust optimization. The slightest variance in the far generations witnessed strengthened the reliability of the optimized configuration.

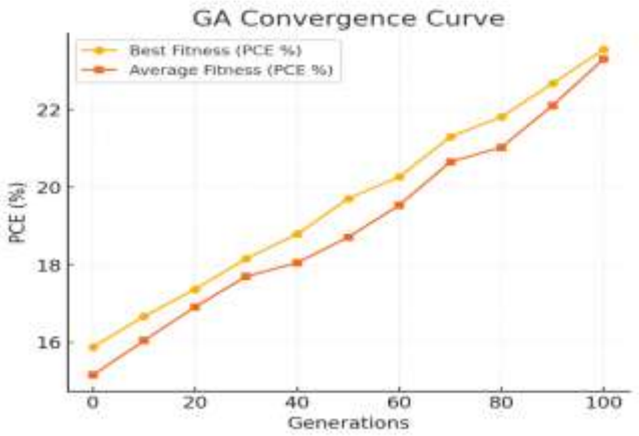

Figure 3: GA Convergence Curve

Figure 3 tracks the landscape across generations during solar cell PCE optimization phases. The best fitness curve steadily climbs from around 15% at the start to about 23.5%, the average fitness evolution following a similar trend, thereby giving an essence of population improvement. Convergence in the later generations points towards the capability of the Genetic Algorithm to attain a steady-state higher configuration.

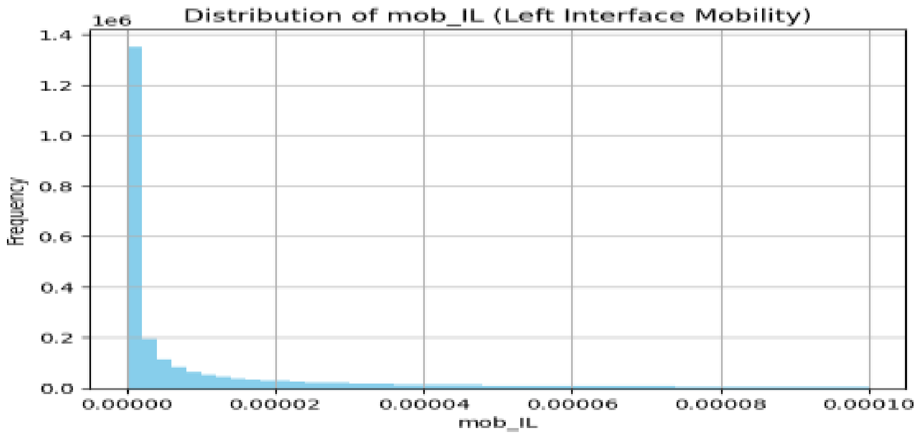

Figure 4: Histogram of Interface Mobility

Right-side skewness can be seen in the distribution of mob_IL as featured in Figure 4. Most of the samples show very little mobility, while just a few can go to greater levels. This means that, in rare instances, high mobility is of utmost importance in the solar cell's amelioration in efficiency.

5.3 Visual Interpretations

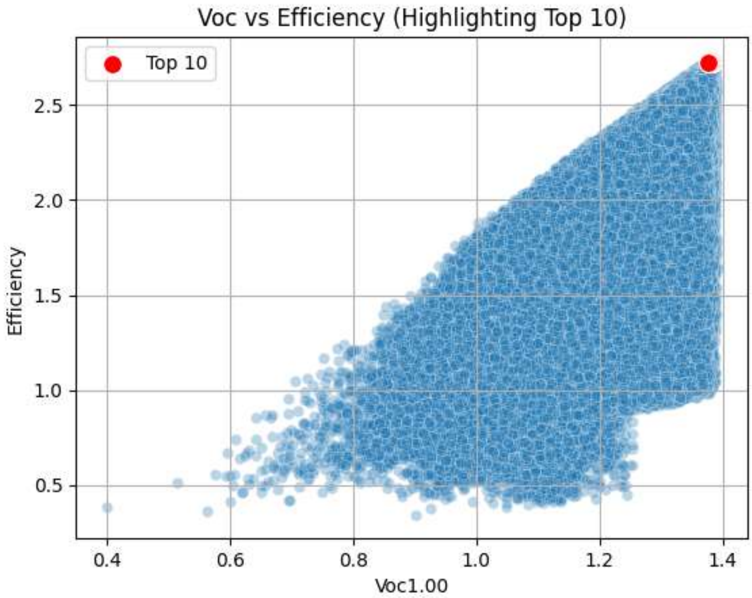

Figure 5: Scatter plot (V_{OC} vs $Efficiency$)

In Figure 5 we noticed the relationship between the open-circuit voltage (VOC) and the efficiency of a solar cell. It is positively correlated: higher VOC values increase the efficiency. The top 10 highlighted points indicate the most efficient configurations, so it is critical to have the VOC maximized in order to achieve better device performance.

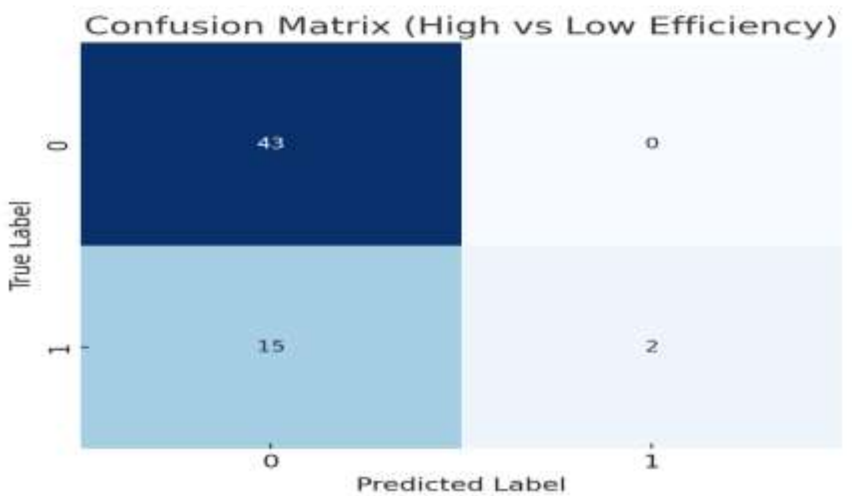

Figure 6: Confusion matrix (classification)

Figure 6 considers the capacity of the model to recognize high- and low-efficiency solar cells. The model correctly classified 43 low-efficiency instances from the entire set but classified 15 high-efficiency samples as lows. The low true-positive scenarios emphasize the fact about high-efficiency cases being difficult to predict, hence, suggesting that the model is subject to more improvements.

The sequential RF-GA pipeline is truly a boon in perusing the conceptualization and design of solar cells. RF provides good prediction results, while it also can point out the relative importance of the parameters that guide the optimization process. Equipped with this knowledge, the GA efficiently traverses through the often-convoluted configuration/design space to reach an optimum design with a predicted PCE at around 23-24%, which can be a great leap from the initial baseline PCE of just about 16%. This convoluted package reduces the need for manual trial-and-error, promoting material and structure choices based on data though. Furthermore, this also ensures an efficient yet robust and scalable approach toward the optimization, making the design crackdown of photovoltaic devices fast.

VI. CONCLUSION

The present study reveals that Random Forest (RF) regression is a very efficient technique to predict the PCE of photovoltaic solar cells by grasping complex nonlinear dependencies among semiconductor parameters. With an R^2 value of 0.92 ± 0.03 and an RMSE and MAE of 1.45 ± 0.22 and 1.02 ± 0.18, respectively, this RF model had impressive generalizing abilities and reliability. Analysis of the feature importances confirmed that Jsc, Voc, and FF were the most important parameters affecting the efficiency. These findings, thus, attest to the ability of RF to guide solar cell modeling very precisely and interpretably. Considering the practical optimization of the predictions, the model RF was coupled with GA, so as to choose the parameters giving the best values of PCE. With the hybrid RF-GA platform, the design space can be explored successfully for searching the optimal semiconductor configurations that maximize efficiency, thus further alleviating the heavy dependence on expensive and time-consuming laboratory trials. Accordingly, the fusion of ML with evolutionary optimization enhances prediction of performance and also enables speeding up solar cell design and development. Therefore, the study established RF-GA as a powerful and scalable approach toward making high-efficiency photovoltaics and thus providing a pathway toward cost-efficient, sustainable, and environmentally friendly solar energy solutions.

References

[1] Su, B., Zhou, Z., & Chen, H. (2022). PVEL-AD: A large-scale open-world dataset for photovoltaic cell anomaly detection. IEEE Transactions on Industrial Informatics, 19(1), 404-413. DOI:10.21227/pz6t-3s77

[2] Rodríguez-Martínez, X., Pascual-San-José, E., & Campoy-Quiles, M. (2021). Accelerating organic solar cell material's discovery: high-throughput screening and big data. Energy & Environmental Science, 14(6), 3301-3322. DOI: 10.1039/d1ee00559f

[3] Mahmood, A., & Wang, J. L. (2021). Machine learning for high performance organic solar cells: current scenario and future prospects. Energy & environmental science, 14(1), 90-105. https://doi.org/10.1039/D0EE02838J

[4] Yang, C., Hu, W., Liu, J., Han, C., Gao, Q., Mei, A., ... & Han, H. (2024). Achievements, challenges, and future prospects for industrialization of perovskite solar cells. Light: Science & Applications, 13(1), 227. https://doi.org/10.1038/s41377-024-01461-x

[5] Marzouglal, M., Souahlia, A., Bessissa, L., Mahi, D., Rabehi, A., Alharthi, Y. Z., ... & Ghoneim, S. S. (2024). Prediction of power conversion efficiency parameter of inverted organic solar cells using artificial intelligence techniques. Scientific Reports, 14(1), 25931. https://doi.org/10.1038/s41598-024-77112-3

[6] Gaboitaolelwe, J., Zungeru, A. M., Yahya, A., Lebekwe, C. K., Vinod, D. N., & Salau, A. O. (2023). Machine learning based solar photovoltaic power forecasting: A review and comparison. IEEe Access, 11, 40820-40845. Digital Object Identifier 10.1109/ACCESS.2023.3270041

[7] Hossain, M. K., Toki, G. I., Kuddus, A., Rubel, M. H., Hossain, M. M., Bencherif, H., ... & Mushtaq, M. (2023). An extensive study on multiple ETL and HTL layers to design and simulation of high-performance lead-free CsSnCl3-based perovskite solar cells. Scientific Reports, 13(1), 2521. https://doi.org/10.1038/s41598-023-28506-2

[8] Kazem, H. A., Al-Waeli, A. H., Chaichan, M. T., & Sopian, K. (2025). Modeling and experimental validation of dust impact on solar cell performance. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 47(2), 2024922. https://doi.org/10.1080/15567036.2021.2024922

[9] Peng, W., Mao, K., Cai, F., Meng, H., Zhu, Z., Li, T., ... & Xu, J. (2023). Reducing nonradiative recombination in perovskite solar cells with a porous insulator contact. Science, 379(6633), 683-690. https://doi.org/10.1126/science.ade3126

[10] Liu, Y., Ding, K., Zhang, J., Li, Y., Yang, Z., Zheng, W., & Chen, X. (2021). Fault diagnosis approach for photovoltaic array based on the stacked auto-encoder and clustering with IV curves. Energy Conversion and Management, 245, 114603. https://doi.org/10.1016/j.enconman.2021.114603

[11] Bayrak, F. (2025). Prediction of photovoltaic panel cell temperatures: Application of empirical and machine learning models. Energy, 323, 135764. https://doi.org/10.1016/j.energy.2025.135764

[12] Janjua, M. R. S. A., Irfan, A., Hussien, M., Ali, M., Saqib, M., & Sulaman, M. (2022). Machine-learning analysis of small-molecule donors for fullerene based organic solar cells. Energy Technology, 10(5), 2200019. https://doi.org/10.1002/ente.202200019

[13] Dwivedi, D., Babu, K. V. S. M., Yemula, P. K., Chakraborty, P., & Pal, M. (2024). Identification of surface defects on solar pv panels and wind turbine blades using attention based deep learning model. Engineering Applications of Artificial Intelligence, 131, 107836.

[14] Nguyen, H. N., Tran, Q. T., Ngo, C. T., Nguyen, D. D., & Tran, V. Q. (2025). Solar energy prediction through machine learning models: A comparative analysis of regressor algorithms. PloS one, 20(1), e0315955. https://doi.org/10.1371/journal.pone.0315955

[15] Chaaban, A. K., & Alfadl, N. (2024). A comparative study of machine learning approaches for an accurate predictive modeling of solar energy generation. Energy Reports, 12, 1293-1302. https://doi.org/10.1016/j.egyr.2024.07.010

[16] Mugware, F. W., Ravele, T., & Sigauke, C. (2025). Short-Term Predictions of Global Horizontal Irradiance Using Recurrent Neural Networks, Support Vector Regression, Gradient Boosting Random Forest and Advanced Stacking Ensemble Approaches. Computation, 13(3), 72. https://doi.org/10.3390/computation13030072

[17] Sehrawat, N., Vashisht, S., & Singh, A. (2023). Solar irradiance forecasting models using machine learning techniques and digital twin: A case study with comparison. International Journal of Intelligent Networks, 4, 90-102. https://doi.org/10.1016/j.ijin.2023.04.001