Original Article

IJCIH
Converging Healthcare & Technology

INTERNATIONAL JOURNAL OF CONVERGENCE IN HEALTHCARE

Published by
IJCIH & Pratyaksh Medicare LLP

www.ijcih.com

Comparative Study of Five Classifiers for Predicting Power Conversion Efficiency in Perovskite Solar Cells

Nita Samantaray[1], Anu Tonk[2], Arjun Singh[3]

[1]Research Scholar, Department of Applied Sciences, The Northcap University, Gurugram, India, [2]Assistant Professor, Department of Multidisciplinary Engineering, The Northcap University, Gurugram, India, [3]Assistant Professor, Department of Applied Sciences, The Northcap University, Gurugram

Abstract

Perovskite solar cells are promising candidates for low-cost and efficient photovoltaic devices. However, the experimental design based on trial and error consumes lots of time and money. Machine learning (ML) techniques can predict perovskite material design and quickly optimize material fabrication parameters and device structure. In this work, we compare machine learning models to predict the power conversion efficiency of perovskite solar cells. A 60-point experimental data set with various experimental conditions and device structures are used to train and test machine learning (ML) models. We trained each of the five classifiers on the 60 data points using 10-fold cross-validation. We split each dataset into ten equal-sized folds, trained the classifier on 9 -folds, and tested it on the remaining fold. We repeat this process ten times so that each fold is used for testing once. The machine learning algorithms or classifiers discussed in this paper include ZeroR, Linear Regression, Gaussian Processes, Random Tree and Random Forest. By comparing the performance of five different classifiers on 60 data points of perovskite solar cell data, we can better understand which methods are most effective for predicting PCE. This article also emphasizes the application of symbolic regression and machine learning to design robust and effective halide perovskite materials. Additionally, it acts as a foundation for additional experimental perovskite material optimization. This work has been performed using Weka the machine learning software.

Keywords: *Machine learning, perovskite solar cell, classifier, ZeroR, Linear Regression, Gaussian Processes, Random Tree and Random Forest.*

Introduction

Due to their high-power conversion efficiency, low material cost, and simplicity of processing, perovskite solar cells are a promising class of photovoltaic devices that have attracted much attention recently. These cells typically consist of a thin film of perovskite material sandwiched between two charge-transporting layers and electrodes. Under sunlight, the perovskite layer generates electron-hole pairs, which can be collected to respective electrodes via electron- and hole-transporting layers. Despite their potential advantages, perovskite solar cells face challenges related to stability and scalability, which must be addressed before they can be commercially viable. Ongoing research aims to develop new materials and fabrication methods to improve perovskite solar cells' performance and reliability.

Corresponding Author:
Dr. Arjun Singh
Assistant Professor, Department of Applied Sciences,
The NorthCap University, Gurugram (India)
e-mail: arjunsingh@ncuindia.edu

Perovskite solar cells (PSCs) have demonstrated remarkable efficiency of >25% [(1)] in recent years, surpassing conventional silicon solar cells. However, several vital challenges must be addressed before going for commercialization. One major obstacle is the absence of long-term stability, as PSCs are prone to moisture, heat, and light exposure degradation. Additionally, PSCs require a high-quality perovskite layer, which can be challenging to achieve through scalable fabrication methods. Furthermore, there are concerns regarding using lead-based perovskite materials, which can pose environmental and health hazards. Addressing these challenges will require further research and development to improve PSCs' stability, scalability, and sustainability.

Machine learning (ML) has emerged as a powerful tool for optimizing perovskite solar cell (PSC) performance. By leveraging large datasets and complex models, ML algorithms can identify patterns and relationships in PSC materials and fabrication processes that are difficult to discern through formal experimentation. The machine learning methods help researchers to accelerate the work progress in the vast parameter space of PSC design and fabrication, hence accelerating the discovery of optimal PSC configurations. ML has been applied to various PSC optimization tasks like material compositions, electronic properties, processing conditions, and developing new device architectures [(2)]. Furthermore, ML has the potential to aid in the development of more durable, sustainable PSC materials and processes by enabling a more comprehensive understanding of the factors that contribute to degradation. Overall, ML holds great promise for advancing PSC research and accelerating the development of efficient, stable, and scalable PSCs for renewable energy applications.

Several machine learning (ML) classifiers have been applied in this paper to optimize perovskite solar cells (PSCs) such as ZeroR, Linear Regression, Gaussian Processes, Random Tree and Random Forest. However, one common approach is using regression models, such as linear regression or support vector regression, to predict the performance of PSCs [(3)] based on input parameters such as precursor concentrations or processing conditions. Another technique is Bayesian optimization, which uses probabilistic models to identify the most promising parameter combinations for PSC fabrication [(4)]. Deep learning, a type of ML that uses artificial neural networks, has also been applied to PSC optimization tasks, such as predicting perovskite crystal structures or optimizing device architectures [(5)]. In addition, unsupervised learning techniques, such as clustering and principal component analysis, have been used to identify patterns and relationships in large PSC datasets [(6)]. Overall, the diversity of ML techniques applied to PSC optimization highlights the versatility of these methods in tackling a wide range of materials science challenges.

The work presented in this paper is highly significant as it addresses a critical challenge in developing perovskite solar cells - the time-consuming and expensive process of trial and error for optimizing material fabrication parameters and device structure. By leveraging machine learning techniques, we have demonstrated a more efficient and accurate way of predicting the power conversion efficiency of perovskite solar cells. Comparing five classifiers based on 60 experimental data points provides valuable insights into which methods are most effective for predicting PCE. The use of machine learning (ML) and symbolic regression models are also highlighted for designing durable and high-performance perovskite solar cell materials. This research is a vital platform for further experimental optimization of perovskite materials, potentially accelerating the development of low-cost and efficient photovoltaic devices.

Methods

Machine learning classifiers can be used in various applications, including analyzing and optimizing perovskite solar cells. Perovskite solar cells are a type of solar cell that use materials with a perovskite crystal structure as the active layer. These solar cells have shown great potential due to their high efficiency and low production costs. One application of machine learning classifiers in perovskite solar cells is to predict the solar cell's performance based on various parameters such as the composition of the perovskite layer, the thickness of the layers, and the choice of electrode materials. This can help researchers optimize the design and fabrication of perovskite solar cells, leading to higher efficiency and stability. The overall methodology is shown in the figure given below in Figure 1.

Figure 1: Methodology of using classifiers in Machine Learning

The first requirement is a dataset with input features and corresponding output values to predict power conversion efficiency. The input features could include the physical properties of the materials used in the conversion process, the operating conditions of the conversion system, and other relevant variables. The output values would be the corresponding power conversion efficiency for each set of input features.

Once the dataset is ready, you can train a machine-learning model using a supervised learning algorithm. You would split the dataset into training and testing sets and use the training set to train the model, and you would then use the testing set to evaluate the accuracy of your model. The overall flow chart is shown in figure 1The cross-validation, and independent test results serve as the primary foundation for the algorithm selection criteria. Mean absolute error (MAE), root mean square error (RMSE), determination coefficient (R^2), and correlation coefficient (r) for regression are some of the frequently used evaluation metrics. This study evaluated the PCE of perovskite solar cells using five different classifiers. The classifiers used in this work are ZeroR, Linear Regression, Gaussian Processes, Random Tree, and Random Forest. This work has been performed using Weka, the machine learning software.

(a) ZeroR: ZeroR is a straightforward classification algorithm that foretells the most prevalent class from the training set of data. In our study, we used ZeroR as a baseline model to compare the performance of other classifiers. Although ZeroR is a simple algorithm, it can be used to determine the accuracy of more complex models. ZeroR has been used as a baseline model for comparison with other machine learning algorithms for predicting the efficiency of perovskite solar cells based on various parameters such as the device architecture, film thickness, and composition. [7] However, it is unsuitable for complex datasets and may not provide accurate predictions for more challenging problems.[8]

(b) Linear Regression: A popular algorithm for modeling the linear relationship between the independent and dependent variables is linear regression. Linear regression assumes a linear relationship between the variables and can produce accurate predictions when the data follow a linear trend. However, if the data is non-linear, linear regression may not be the best option.(9) The linear regression model was able to accurately predict the efficiency of perovskite solar cells based on the molecular descriptors of the materials, showing the potential of QSPR models for designing and optimizing perovskite solar cells.[10]

(c) Gaussian Processes: Gaussian processes are a Bayesian machine learning technique that models the distribution of functions over a continuous domain. Gaussian processes are flexible and can handle non-linear data. They can also produce accurate predictions with limited data. Gaussian Processes have been used as a machine learning algorithm for predicting the efficiency of perovskite solar cells based on various parameters such as the device architecture, film thickness, and composition.[11] However, they can be computationally expensive and may not scale well to large datasets.[12]

(d) Random Tree: Random Tree is a decision tree-based algorithm that constructs multiple decision trees and aggregates the results to make predictions. Random Tree is a simple and interpretable algorithm that can handle both categorical and continuous data. It can also handle non-linear relationships between variables. RT is being used for predicting the power conversion efficiency of perovskite solar cells based on the device architecture and composition.[13] However, it may not perform well with noisy data.[14]

(e) Random Forest: An ensemble learning algorithm called Random Forest (RFs) combines various decision trees to produce predictions. Random Forest is a highly accurate algorithm that can handle non-linear relationships between variables, noisy data, and missing values. It can also provide estimates of feature importance, which can aid in feature selection. However, it can be computationally expensive and perform poorly with highly imbalanced datasets. RFs have been used to predict the efficiency of perovskite solar cells based on various parameters such as the device architecture, film thickness, and composition.[15]

Results and Discussions

This section will compare the analysis and performance of all examined classification algorithms using datasets based on perovskite solar cell performance parameters. This study uses a dataset of 60 data points to evaluate the performance of various classifiers in predicting PCE. The dataset was pre-processed and split into training and testing sets, and five different algorithms (ZeroR, Linear Regression, Gaussian Processes, Random Tree, and Random Forest) were used to predict PCE based on the values of the independent variables.

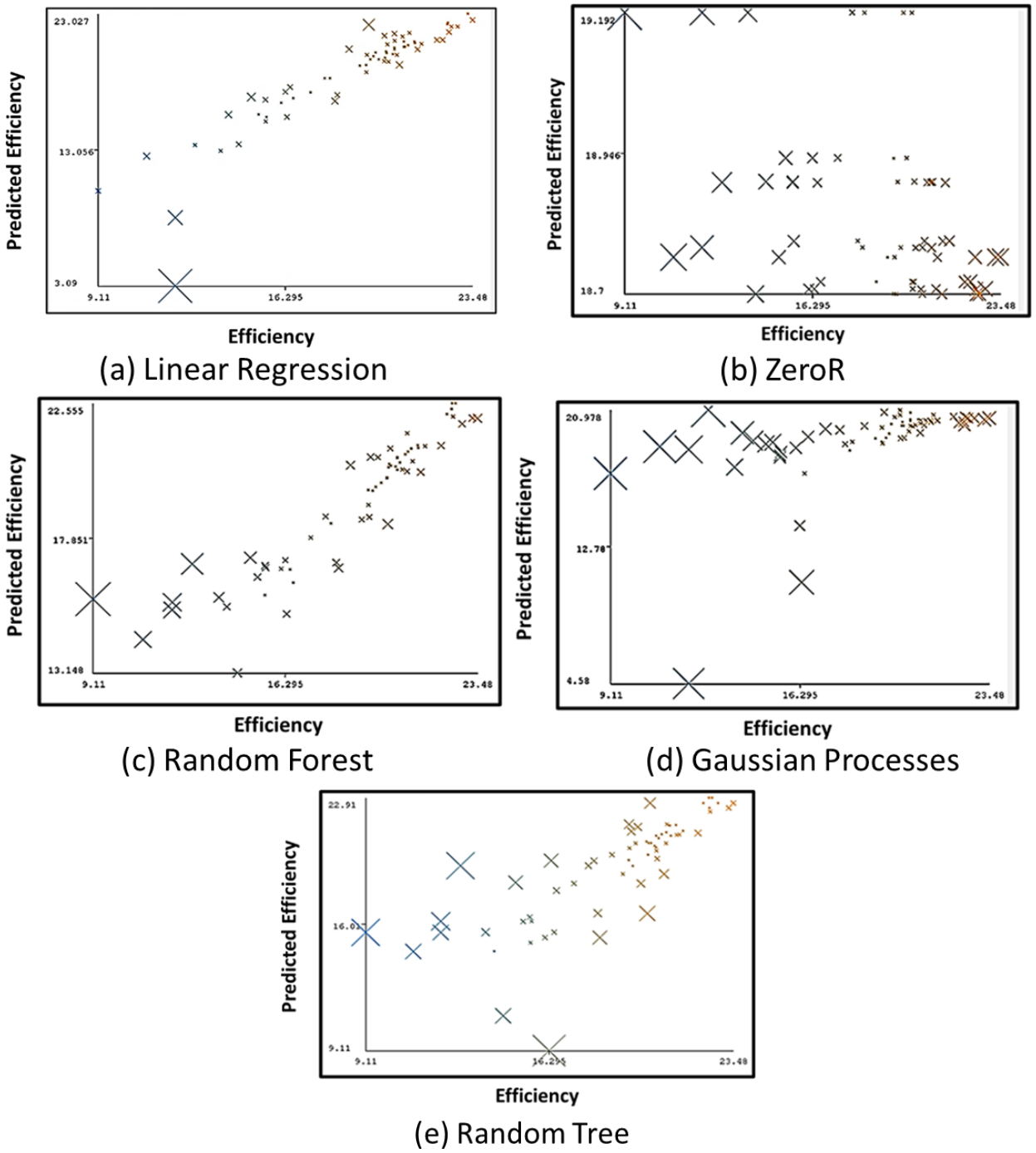

Figure 2: Predicted power conversion efficiency using different classifiers (a) Linear Regression (b) ZeroR (c) Random Forest (d) Gaussian Processes (e) Random Tree. The machine learning software Weka has been used for this work.

Critical performance metrics such as correlation coefficient, mean absolute error, and root means squared error was used to compare each classifier's accuracy and reliability. The graphical comparison results of each classifier are shown in figure 2. In these graphs, the x-axis indicates the actual efficiency, and the y-axis indicates the predicted efficiency. The obtained performance results using all the classifiers for predicting PCE are presented in Table 1.

Table 1: Comparative Analysis of different classifiers used for predicting PCE of perovskite solar cell

ML Classifiers	Correlation Coefficient (r)	Coefficient of determination (R^2) ($R^2 = r^2$)	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)	(Actual PCE (%)	Predicted PCE (%)	Remarks
Linear Regression	0.914	0.835	0.849	1.476	23.48	23.027	Good correlation
Random Forest	0.916	0.849	0.908	1.449	23.48	22.555	Good correlation
Random Tree	0.782	0.611	1.325	2.042	23.48	22.910	Moderate correlation
Gaussian Processes	0.519	0.269	1.965	2.583	23.48	20.978	Moderate/Less Correlation
ZeroR	-0.393	0.154	2.628	3.247	23.48	19.192	Less correlation

The Correlation coefficient(r) is a statistical parameter used to check the relationship between two chosen variables[16]. The correlation coefficient ranges between -1 to +1, representing the strength between the two chosen variables. A strong relationship is one having the r value closer to 1 between the chosen variables, and -1 represents the negative relationship between the variables. The r values near zero or equal to zero represent no correlation between the variables. Here in this work, the Linear regression and Random Forest give r values of 0.914 and 0.916, respectively, and these two represents the excellent correlation between predicted PCE and experimental PCE of perovskite solar cell. The coefficient of determination(R^2) is also high for Linear regression and Random Forest, i.e., R^2 ~0.84. It shows good outcomes between the dependent and independent variables. [17] The machine learning classifiers Random Tree gives the coefficient of correlation(r) and coefficient of determination (R^2) values of 0.782 and 0.611, respectively. This classifier gives a moderate outcome. The Gaussian and ZeroR give either less correlation between the chosen variables. Hence from table 1.0 above, we can say that linear progression and Random Forest algorithms give a good outcome for predicting the PCEs of perovskite solar cells.

A standard metric for calculating the average discrepancy between predicted and actual values in a regression problem is the Mean Absolute Error (MAE). The absolute difference between the PCEs' predicted and actual values and the average of these absolute differences are used to calculate the MAE. The model makes more accurate predictions when the MAE is lower. The Linear regression (MAE value ~ 0.849) and Random forest (0.908) also give lower values of MAE as compared to Random Tree (1.325), Gaussian Processes (1.965), and ZeroR(2.628). The MAE values also predicted that the linear regression and random forest give the best outcome.[18]

A popular metric for assessing the efficacy of regression models is the Root Mean Squared Error (RMSE). The average squared difference between the predicted and actual values is what is measured. Both the linear regression and the random forest have RMSE values of 1.476 and 1.449, respectively. The model is making more accurate predictions when the RMSE is lower. The RMSE for others models is given in table 1. This comparison can provide insight into whether the model is making meaningful predictions or is simply predicting the average value. It is worth noting that the RMSE penalizes significant errors more than the MAE because of the squaring operation, and this means that the RMSE is more sensitive to outliers in the data than the MAE.[19]

Conclusion

Predicting power conversion efficiency using machine learning (ML) classifiers is a standard process in renewable energy. Several machine learning classifiers can be used to make predictions, including random trees, random forests, linear regression, gaussian processes, and ZeroR. The study investigated the efficacy of various classifiers in predicting PCE using a dataset of 60 data points. Our findings indicate that the Random Forest and Linear regression algorithm is the most accurate and reliable method for predicting PCE, outperforming other algorithms such as ZeroR, Gaussian Processes, and Random Trees. These results have important implications for the field of PCE prediction, as an accurate prediction can aid in developing more efficient and cost-effective solar cells. Our study highlights the potential of machine learning algorithms in predicting PCE, and future research could explore the use of larger datasets and more advanced algorithms to improve the accuracy and reliability of PCE prediction. Overall, the results of this study show the potential of machine learning algorithms to predict complex scientific phenomena and have important implications for creating more sustainable and efficient solar cells. The proposed model can help predict the power conversion efficiency of perovskite solar cells using machine learning without any experimental work. Indeed, the performance prediction ability of the ML model can be further improved by the selection of particular device architecture. The perovskite solar cell performance prediction is very complicated because the PCEs depend on many factors, such as film morphology, interface properties, etc. Therefore, widespread research is required to understand the surface properties of perovskite films to improve PSCs performance parameters.

Conflict of Interest: The authors declare no competing interests.

Ethical Clearance: Not applicable.

Source of Funding: There is no funding for this article available.

References

1. Niu Y, He D, Zhang X, Hu L, Huang Y. Enhanced Perovskite Solar Cell Stability and Efficiency via Multi-Functional Quaternary Ammonium Bromide Passivation. Advanced Materials Interfaces. 2023;10(2):2201497.
2. Noori L, Hoseinpour V, Shariatinia Z. Optimization of TiO2 paste concentration employed as electron transport layers in fully ambient air processed perovskite solar cells with a low-cost architecture. Ceramics International. 2022;48(1):320-36.
3. Chen J, Feng M, Zha C, Shao C, Zhang L, Wang L. Machine learning-driven design of promising perovskites for photovoltaic applications: A review. Surfaces and Interfaces. 2022;35:102470.
4. Xu W, Liu Z, Piper RT, Hsu JWP. Bayesian Optimization of photonic curing process for flexible perovskite photovoltaic devices. Solar Energy Materials and Solar Cells. 2023;249:112055.
5. Bak T, Kim K, Seo E, Han J, Sung H, Jeon I, et al. Accelerated Design of High-Efficiency Lead-Free Tin Perovskite Solar Cells via Machine Learning. International Journal of Precision Engineering and Manufacturing-Green Technology. 2023;10(1):109-21.
6. Yu X, Ergan S. Estimating power demand shaving capacity of buildings on an urban scale using extracted demand response profiles through machine learning models. Applied Energy. 2022;310:118579.
7. Baird SG, Liu M, Sparks TD. High-dimensional Bayesian optimization of 23 hyperparameters over 100 iterations for an attention-based network to predict materials property: A case study on CrabNet using Ax platform and SAASBO. Computational Materials Science. 2022;211:111505.
8. Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering. 2007;160(1):3-24.
9. Li J, Pradhan B, Gaur S, Thomas J. Predictions and Strategies Learned from Machine Learning to Develop High-Performing Perovskite Solar Cells. Advanced Energy Materials. 2019;9(46):1901891.
10. Feng J, Wang H, Ji Y, Li Y. Molecular design and performance improvement in organic solar cells guided by high-throughput screening and machine learning. Nano Select. 2021;2(9):1629-41.
11. Yan W, Liu Y, Zang Y, Cheng J, Wang Y, Chu L, et al. Machine learning enabled development of unexplored perovskite solar cells with high efficiency. Nano Energy. 2022;99:107394.

12. Akhundova F, Lüer L, Osvet A, Hauch J, Peters IM, Forberich K, et al. Building process design rules for microstructure control in wide-bandgap mixed halide perovskite solar cells by a high-throughput approach. Applied Physics Letters. 2021;118(24):243903.
13. Workman M, Chen DZ, Musa SM, editors. Machine Learning for Predicting Perovskite Solar Cell Opto-Electronic Properties. 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI); 2020 26-27 Oct. 2020.
14. Song Q, Bai Y, Chen Q. The Spring of Processing Chemistry in Perovskite Solar Cells–Bayesian Optimization. The Journal of Physical Chemistry Letters. 2022;13(46):10741-50.
15. Zhu C, Liu W, Li Y, Huo X, Li H, Guo K, et al. Key factors governing the device performance of CIGS solar cells: Insights from machine learning. Solar Energy. 2021;228:45-52.
16. Ahmad MU, Akib AR, Raihan MMS, Shams AB, editors. ABO3 Perovskites' Formability Prediction and Crystal Structure Classification using Machine Learning. 2022 International Conference on Innovations in Science, Engineering and Technology (ICISET); 2022 26-27 Feb. 2022.
17. Liu Y, Tan X, Liang J, Han H, Xiang P, Yan W. Machine Learning for Perovskite Solar Cells and Component Materials: Key Technologies and Prospects. Advanced Functional Materials. 2023;n/a(n/a):2214271.
18. Guo Z, Lin B. Machine learning stability and band gap of lead-free halide double perovskite materials for perovskite solar cells. Solar Energy. 2021;228:689-99.
19. Chenebuah ET, Nganbe M, Tchagang AB. Comparative analysis of machine learning approaches on the prediction of the electronic properties of perovskites: A case study of ABX3 and A2BB'X6. Materials Today Communications. 2021;27:102462.