


## Article

# Machine Learning Assisted Prediction of Power Conversion Efficiency of All-Small Molecule Organic Solar Cells: A Data Visualization and Statistical Analysis

Norah Alwadai <sup>1,\*</sup>, Salah Ud-Din Khan <sup>2,\*</sup> , Zainab Mufarreh Elqahtani <sup>1</sup> and Shahab Ud-Din Khan <sup>3</sup>

<sup>1</sup> Department of Physics, Collega of Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

<sup>2</sup> Sustainable Energy Technologies Center, College of Engineering, King Saud University, P.O. Box 800, Riyadh 11421, Saudi Arabi

<sup>3</sup> Pakistan Tokamak Plasma Research Institute (PTPRI), Islamabad P.O. Box 3329, Pakistan

\* Correspondence: authors: nmalwadai@pnu.edu.sa (N.A.); drskhan@ksu.edu.sa (S.U.-D.K.)

**Abstract:** Organic solar cells are famous for their cheap solution processing. Their industrialization needs fast designing of efficient materials. For this purpose, testing of large number of materials is necessary. Machine learning is a better option due to cheaper prediction of power conversion efficiencies. In the present work, machine learning was used to predict power conversion efficiencies. Experimental data were collected from the literature to feed the machine learning models. A detailed data visualization analysis was performed to study the trends of the dataset. The relationship between descriptors and power conversion efficiency was quantitatively determined by Pearson correlations. The importance of features was also determined using feature importance analysis. More than 10 machine learning models were tried to find better models. Only the two best models (random forest regressor and bagging regressor) were selected for further analysis. The prediction ability of these models was high. The coefficient of determination ( $R^2$ ) values for the random forest regressor and bagging regressor models were 0.892 and 0.887, respectively. The Shapley additive explanation (SHAP) method was used to identify the impact of descriptors on the output of models.

**Keywords:** small molecule donors; machine learning; Pearson correlation; random forest regressor



**Citation:** Alwadai, N.; Khan, S.U.-D.; Elqahtani, Z.M.; Ud-Din Khan, S. Machine Learning Assisted Prediction of Power Conversion Efficiency of All-Small Molecule Organic Solar Cells: A Data Visualization and Statistical Analysis. *Molecules* **2022**, *27*, 5905. <https://doi.org/10.3390/molecules27185905>

Academic Editors: Aurora Costales and Fernando Cortés-Guzmán

Received: 16 August 2022

Accepted: 8 September 2022

Published: 11 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The recent development of society is the result of technological advancement that is the fruit of great scientific research [1–3]. Extensive research is going on in the field of material science [4–6]. High industrialization has led to many environmental issues [7,8]. Therefore, clean energy is an essential need of modern society. Solar energy is a huge source of energy. One of the most promising ways to gather and process solar energy is photovoltaic (PV) devices [9]. Third-generation solar cells are referred to as emerging technologies [10]. Their performance efficiencies are high. Among all these emerging technologies, organic solar cells (OSCs) have drawn considerable attention from academic and industrial communities due to their peculiar characteristics, such as ease of use, sustainability, adjustability, compactness, as well as transparency compared to conventional silicon-based inorganic solar cells. The primary process of photovoltaic cells is the absorption of sunlight harvested by the active layers. Organic molecules are structurally  $\pi$ -conjugated molecules with alternating  $\pi$  and  $\sigma$  bonds [11–13]. Thus, they are composed of discontinuous energy levels naming HOMO and LUMO levels that are an abbreviation of the highest occupied molecular orbital and the lowest unoccupied molecular orbitals, respectively. The difference between these two energy levels is known as the band gap.

Bulk-junction deposition, in which donor and acceptor materials are mixed thoroughly, is the famous type of photoactive layer structure [14,15]. Co-deposition of the two materials

improves the closeness of contact between the two comparable semiconductors, which is the basis for this type of organic solar cell [16,17].

Scharber's model has been used to predict the performance of organic solar cells [18]. Different, less realistic assumptions are used to derive this model. This makes it less accurate [19]. Only electronic parameters of materials used in active layers are used to predict power conversion efficiency (PCE). It is difficult to include other descriptors such as structure, topology, and thermodynamics. Therefore, there are fewer chances to enhance its performance.

In recent years, machine learning (ML) has gained fame in material science [20,21]. Machine learning is much faster than density functional theory and molecular dynamics simulations [22,23]. The increase in computer power and development of efficient software have enhanced the potential of machine learning. It can be used for discovery, data mining, prediction, and design of new materials [24–26]. Compared to traditional computational and experimental approaches, machine learning has developed quickly [27–29].

In the current work, machine learning-based regression models were trained to predict the PCE of all-small molecule organic solar cells. Multiple models were trained and the best models were selected for further analysis. Their parameters were tuned. A detailed data visualization analysis was also performed to find the hidden trends of data. Pearson correlation was used to find the relationship between parameters and power conversion efficiency. The feature importance of parameters in training of models was also calculated. The Shapley additive explanation (SHAP) method was used to identify the impact of parameters on the output of models.

## 2. Results and Discussion

The performance of different materials depends on their chemistry [30,31]. Chemical data can help to understand their behavior [32,33]. The hidden patterns of data can provide much useful information [34,35].

### 2.1. Visualization Analysis of Data

A detailed visualization analysis of data was performed. A heat map of correlation between PCE and other parameters is given in Figure 1. Only  $J_{sc}$  showed a high positive correlation with PCE; HOMO showed very low correlation with PCE and LUMO showed very low negative correlation with PCE. This indicates relatively less dependence of PCE on the energy level of donor materials.

To better understand the data and effect of various parameters on PCE, we classified the PCE into three categories—high:  $PCE > 7$ , medium:  $PCE > 4$ , and low:  $PCE < 4$ . The paired scatter plots are given in Figure 2. For the scatter plot comparison between LUMO and HOMO, the trend of PCE was mixed. This means energy levels did not have a significant effect on PCE. The scatter plot comparison between  $V_{oc}$  and HOMO indicated a high PCE at higher  $V_{oc}$  values and middle HOMO values. The scatter plot of HOMO with  $V_{oc}$  and FF did not provide any clear trend. The scatter plots of LUMO with  $V_{oc}$  and FF indicated that a higher PCE was found at lower LUMO values.

The box plot allowed us to look at the data in another way. The box plots for different parameters are given in Figure 3. In the case of HOMO, the majority of PCE points were boxed between  $-5.1$  and  $-5.4$  eV. The small-size box for a high PCE indicates that the control of the HOMO level of donors can help to achieve a high PCE. In the case of LUMO, the size of boxes was almost the same in all categories. In the cases of  $J_{sc}$ ,  $V_{oc}$ , and FF, the size of boxes was very small for a high PCE.

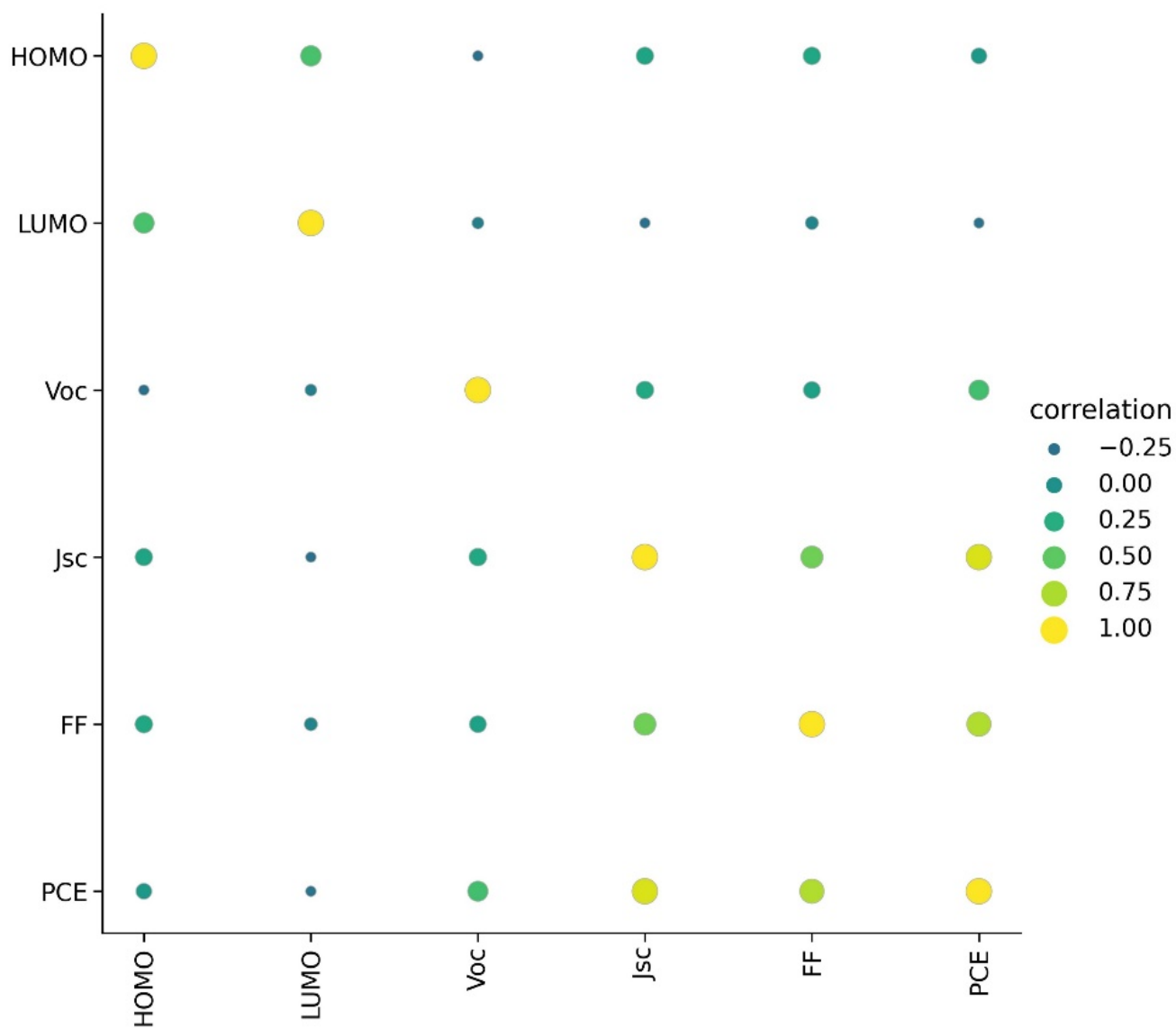


Figure 1. Correlation between parameters in dataset.



Figure 2. Pair-scatter plot between the parameters in dataset.

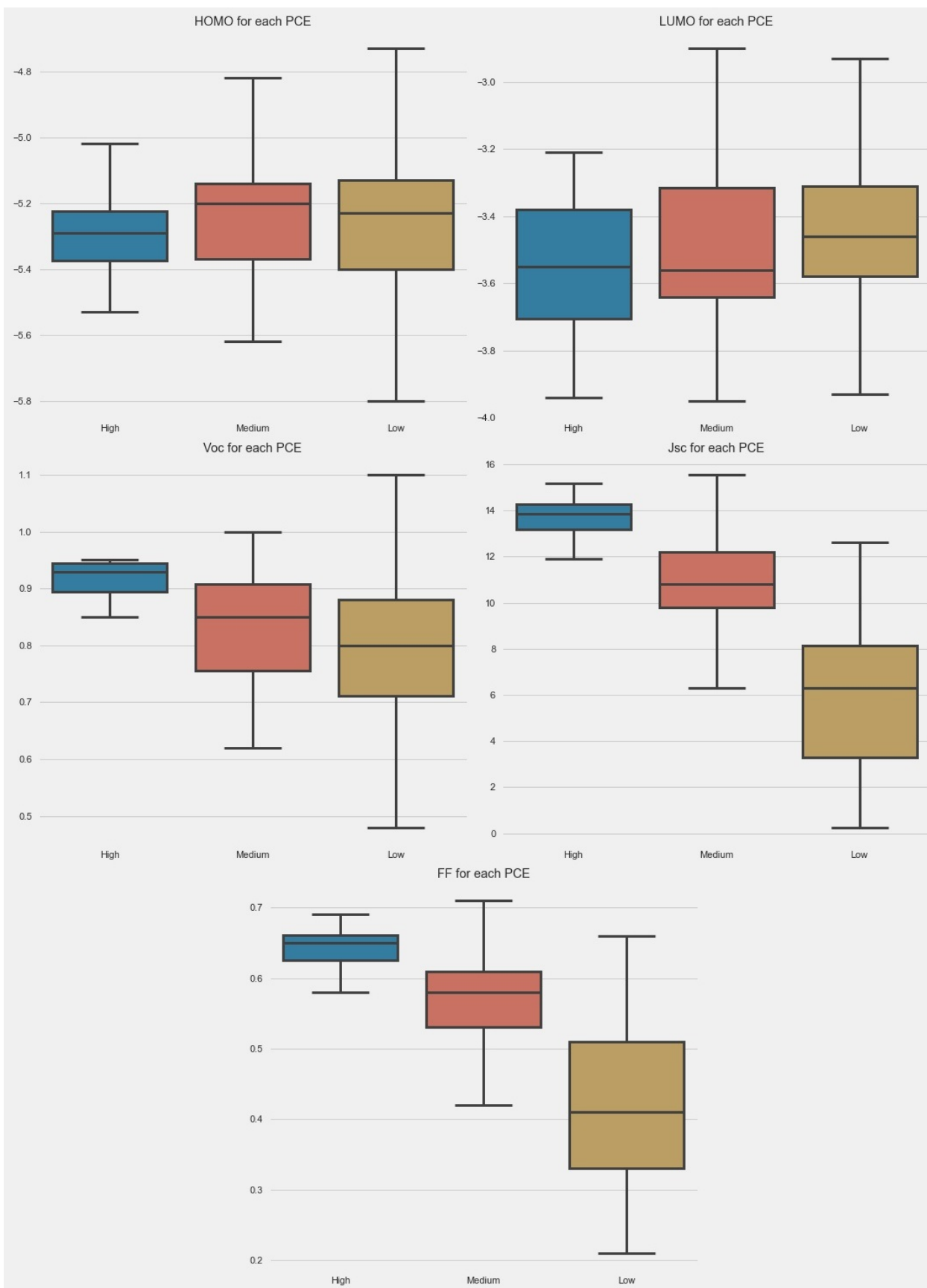
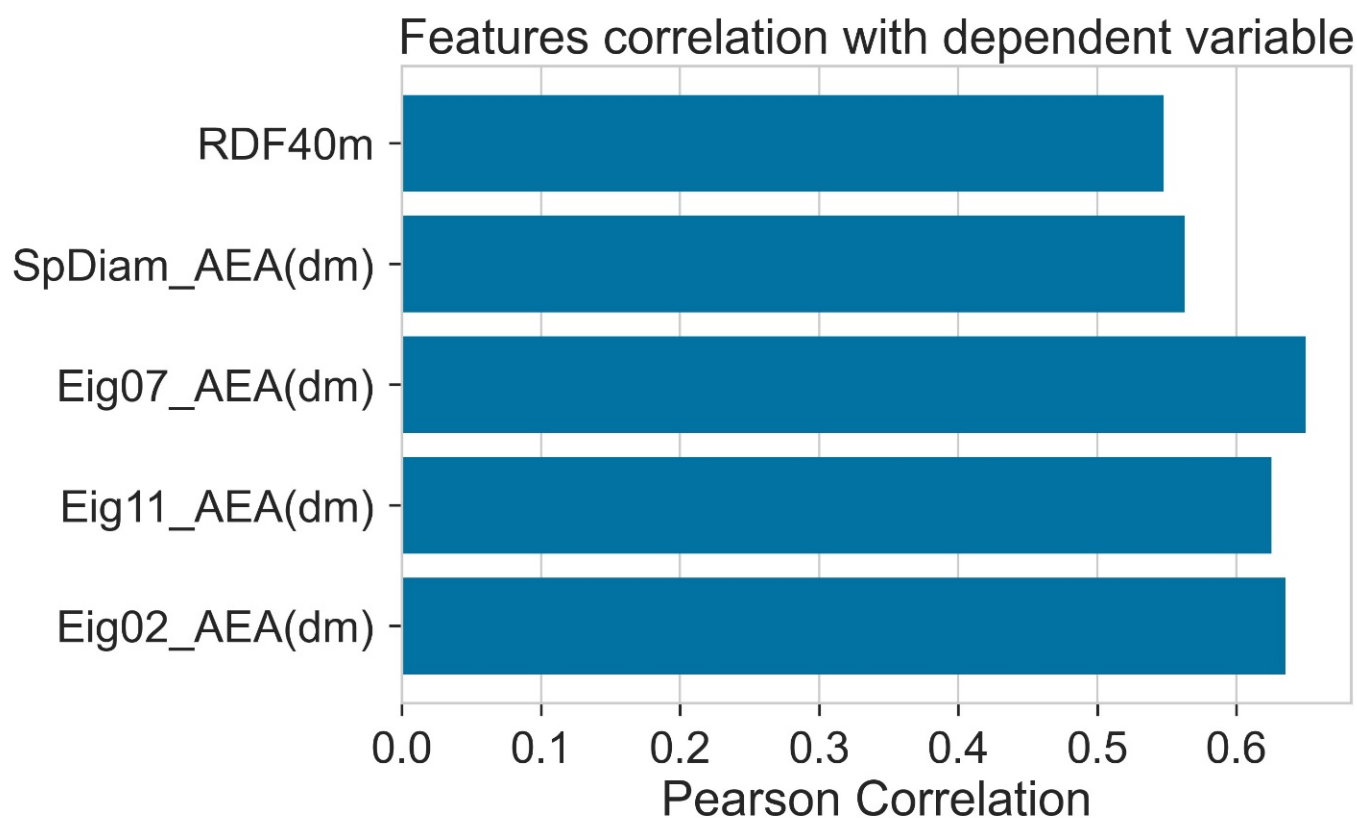


Figure 3. Box plot of parameters in dataset.

## 2.2. Correlation Analysis of Descriptors with PCE

The calculated molecular descriptors were used as input for model training. Molecular descriptors represent the chemistry of donor molecules. It is an open secret that the PCE of organic solar cells significantly depends on the chemical structure of materials that are used for OSCs [36,37]. Molecular descriptors present the chemical features of materials in numerical form [38,39].

The correlation of different descriptors with the PCE is given in Figure 4. Eig07\_AEA (dm) showed a high positive correlation with PCE. Correlation of all the descriptors with PCE was higher than 0.5. The details of descriptors are given in Table 1.



**Figure 4.** Pearson correlation of features with dependent variable (PCE).

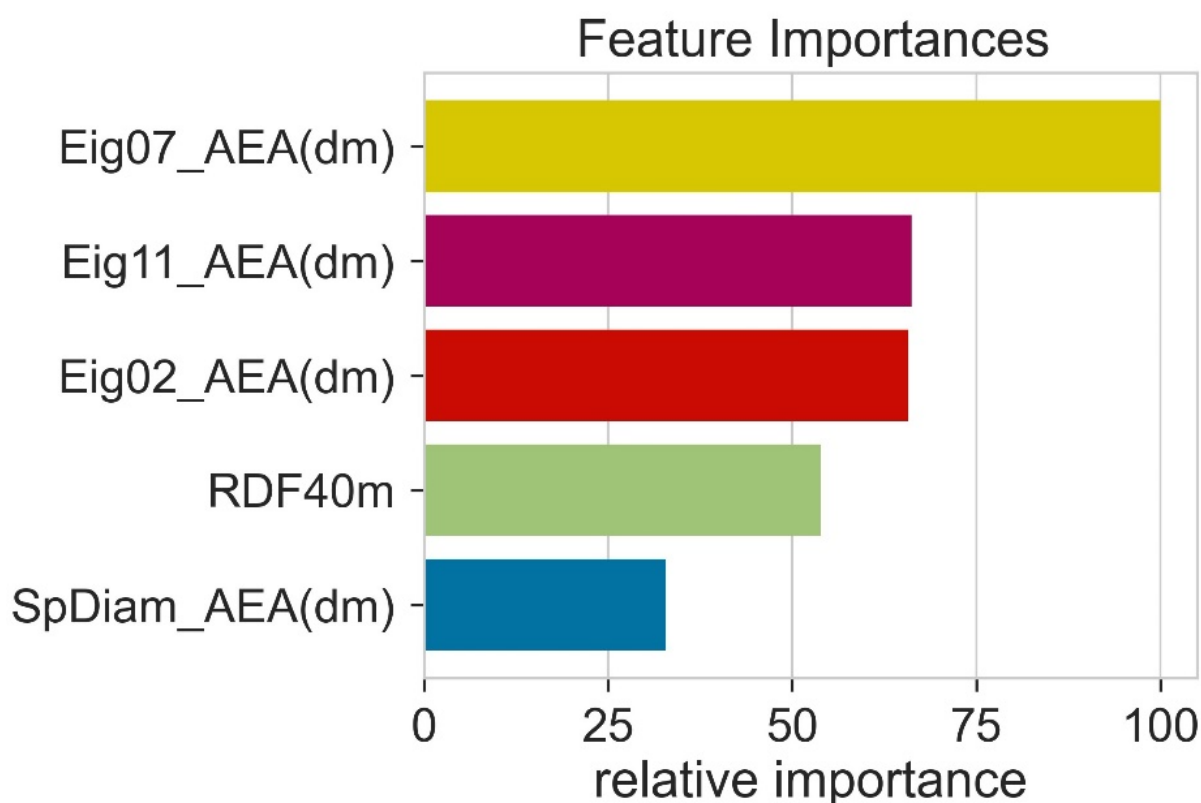
**Table 1.** Detail of selected molecular descriptors.

No.	Name	Category	Description
1	RDF40m	RDF descriptor	Radial distribution function-040/weighted by relative mass
2	SpDiam_AEA (dm)	Edge adjacency indices	Spectral diameter from augmented edge adjacency mat. weighted by dipole moment
3	Eig07_AEA (dm)	Edge adjacency indices	Eigenvalue n. 7 from augmented edge adjacency mat. weighted by dipole moment
4	Eig11_AEA (dm)	Edge adjacency indices	Eigenvalue n. 11 from augmented edge adjacency mat. weighted by dipole moment
5	Eig02_AEA (dm)	Edge adjacency indices	Eigenvalue n. 2 from augmented edge adjacency mat. weighted by dipole moment

## 2.3. Feature Importance

During model training, all the features (descriptors) do not play an equal role in model performance. Therefore, it is necessary to determine the relative importance of different

features. The feature importance was calculated using random forest. The feature importance was obtained by computing the reduced training loss when using this feature. Higher feature importance values indicate that during model training, this feature has contributed more to the machine learning algorithm. This means that features with high feature importance values are helpful for machine learning model predictions. Eig07\_AEA (dm) had high importance and SpDiam\_AEA (dm) had less importance (Figure 5). However, the trend in Pearson correlation and feature importance was a little bit different. We further reduced the number of features; this decrease in feature numbers decreased the performance of machine learning models.

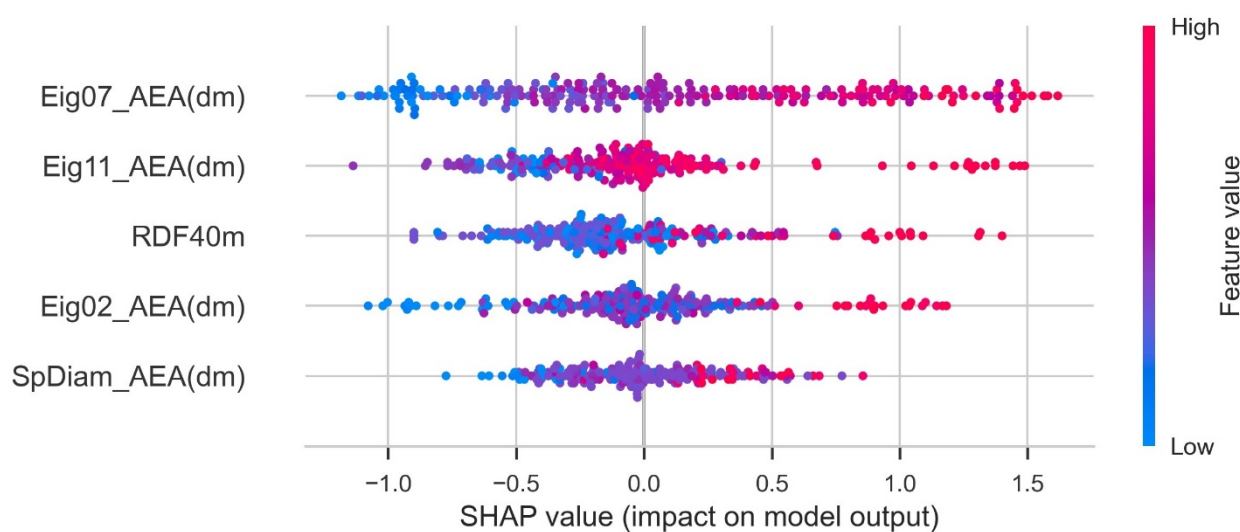


**Figure 5.** The relative importance of various features in machine learning models.

#### 2.4. Shapley Additive exPlanations

The Shapley additive exPlanations (SHAP) feature importance value was computed using the `shap_values` function provided by Python `shap`. This is a feature attribution method that connects the Shapley value and local interpretable model-agnostic explanations. The Shapley value, which is the basis of SHAP feature importance, is calculated using the average change in predicted values according to the presence or absence of the feature when considering all possible combinations of features. A large change in the predicted values depending on the presence or absence of a feature indicates that the corresponding feature significantly contributes to the training of the predictive ML model. It tells whether contribution of a feature is positive or negative. A higher value indicates the higher contribution to PCE. Each dot represents one sample point. The SHAP plot is given in Figure 6. Eig07\_AEA (dm) had a strong impact on the output of the ML model.





**Figure 6.** SHAP plot for parameters.

### 2.5. Regression Analysis

Classification categorizes given data points into predefined groups. The wider the range of a group, the higher will be the classification accuracy. With the help of classification machine learning, it is possible to predict in which group the PCE of a particular donor will fall. In order to predict the PCE value of a particular donor, regression analysis was performed. More than ten regressors were used. The coefficient of determination ( $R^2$ ) values for the test set are given in Table 2. Random forest regressor and bagging regressor were the best models. These models were used for further analysis. Residuals of the best models were plotted. Basically, a residual plot is a plot that presents the residuals on the vertical axis and target variable on the horizontal axis. Residual value indicates the deviation of predicted values from actual values. The further the data point is away from zero, the more the predicted values will differ from actual values. The residual plot for the random forest model is given in Figure 7. In most cases, residual values were not very high. The distribution plot indicated major peaks near to zero. The residuals for the bagging model are given in Figure 8. The behavior of the bagging regressor was very similar to that of the random forest regressor. Both models were accurate enough;  $R^2$  values near 1 are considered good. The accurate prediction of different chemical properties can decrease dependence on expensive experimental methods [40–42]. The scatter plot comparing experimental PCE and predicted (random forest model) PCE is given in Figure 9. Most values were at the lower range. The scatter plot comparing experimental PCE and predicted (bagging model) PCE is given in Figure 10.

**Table 2.** Performance of various machine learning models ( $R^2$  for test set).

No	Model	$R^2$
1	Random Forest Regressor	0.892
2	Bagging Regressor	0.887
3	Gradient Boosting Regressor	0.774
4	Light Gradient Boosting Machine	0.769
5	Extreme Gradient Boosting	0.667
6	Extra Trees Regressor	0.664
7	Support Vector Machine	0.632
8	Ridge Regression	0.607
9	K Neighbors Regressor	0.598
10	CatBoost Regressor	0.592
11	Linear Regression	0.588



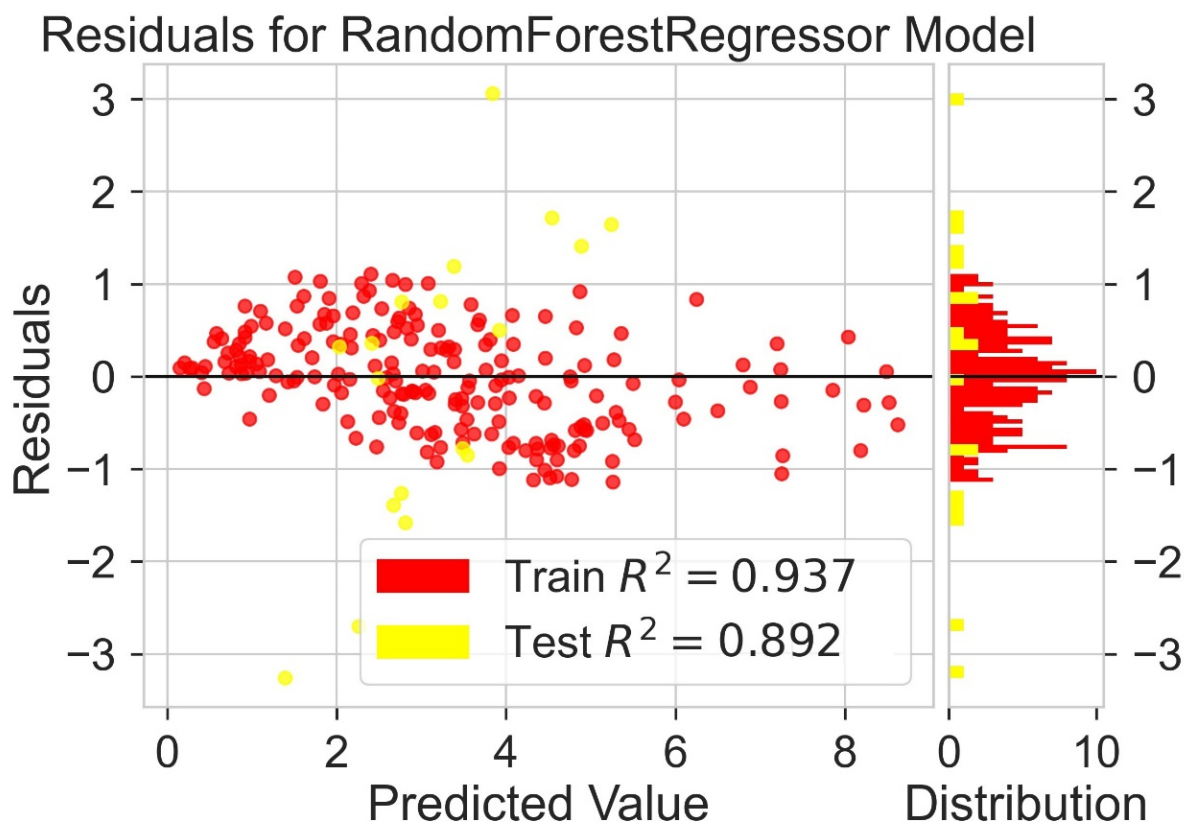


Figure 7. Residuals for random forest regressor model.

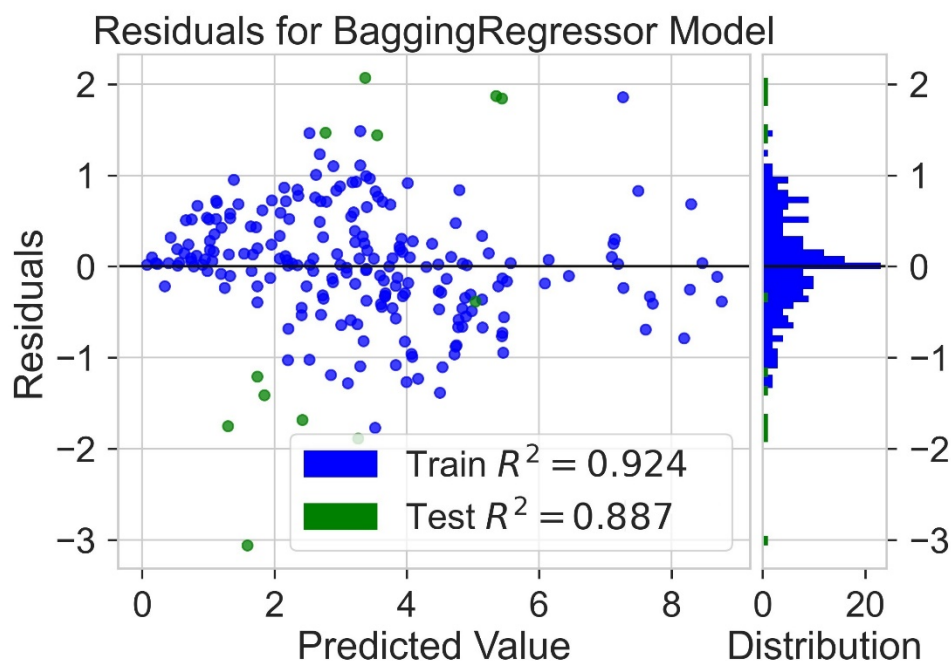


Figure 8. Residuals for bagging regressor model.

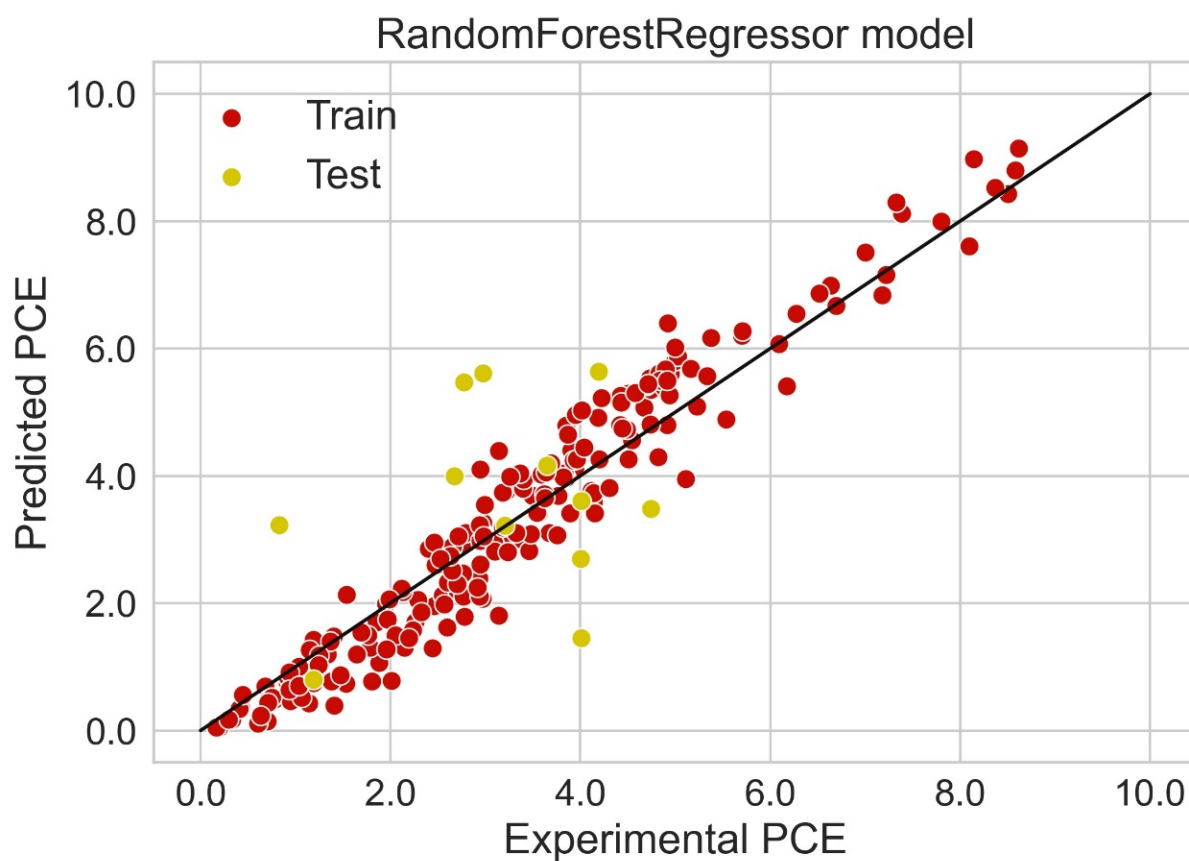


Figure 9. Scatter plot comparing experimental PCE and predicted PCE (random forest model).

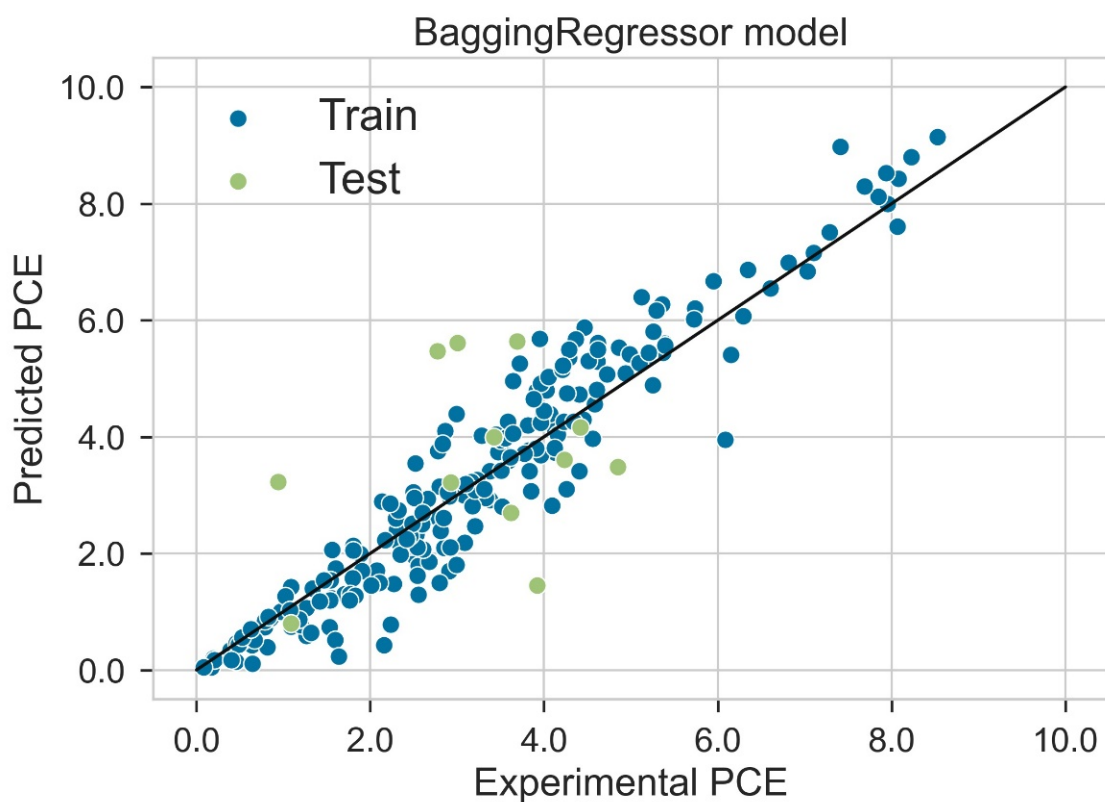


Figure 10. Scatter plot comparing experimental PCE and predicted PCE (bagging model).

The random forest model was validated using an external set of data that was not used for training and testing purposes. Obtained results are given in Table 3. The low dissimilarity between predicted and experimental PCE values indicates that this model was reasonably accurate. An easy and fast prediction of PCE can speed the design of better donor materials.

**Table 3.** Validation of random forest model using external dataset.

Donor	Experimental PCE (%)	Predicted PCE (%)	Difference	Reference
DPP2T-3	2.48	2.81	0.33	[43]
DPP2T-4	3.30	2.98	0.32	[43]
DPP2T-5	1.90	2.13	0.23	[43]
DPPT	1.88	2.16	0.28	[44]
DPPSE	2.30	2.50	0.20	[44]
DPPTT	1.25	1.56	0.31	[44]
FDPP	4.32	4.71	0.39	[45]
CDPP	1.00	1.25	0.25	[45]

A better understanding of chemical structure of materials helps to find better materials [46–49]. Our proposed model can help to predict the PCE quickly without any experimentation. Indeed, the performance prediction ability of machine learning can be further improved by design-specific descriptors. It is well-known that the principle on which organic solar cells works is very complicated. The PCE of OSCs depends on a variety of factors [50,51]. Film morphology is one of them. The results from film morphology characterization can be explored using deep learning. Therefore, widespread research is needed to effectively utilize deep learning to understand the thin film morphological topographies of all-small molecule organic solar cells.

### 3. Methodology

#### 3.1. Dataset

Our dataset had about 220 data points that were collected from research articles. Dataset is given in supporting information (Table S1). It contained the data of organic solar cells that were based on small molecule donors and fullerene acceptors. The dataset contained the HOMO and LUMO of donor materials as well as open-circuit voltage ( $V_{OC}$ ), short-circuit current density ( $J_{SC}$ ), and fill factor (FF) of solar cell devices. In research articles, the highest and average values of photovoltaic parameters are reported. We have selected the highest values. It is not easy to collect experimental data. The quality and volume of data strongly control the prediction ability of machine learning models.

#### 3.2. Descriptors Calculation and Selection

About 3000 molecular descriptors were calculated using Dragon software [52]. Molecule descriptors are easy to calculate: a large number of descriptors can be calculated in a short time. As the number of descriptors was large, every descriptor was not important for model training. We have reduced their numbers in different ways. Descriptors with zero values were not chosen. Descriptors with the same values for all donors cannot provide any discriminating effect; therefore, they were removed. Many pairs of descriptors are similar, so in model training their role will be the same, and the use of both will not affect the performance of the model. So, one of the pair of descriptors was neglected.

#### 3.3. Machine Learning

Machine learning was performed using the Scikit-learn Python library. This library provides many machine learning models to test. Data were handled using Pandas software. The calculated descriptors and target property (PCE) were placed in comma-separated values (.CSV) files. We tested more than 10 machine learning models. Two high-performing

models were chosen for next step analysis. Their parameters were tuned to obtain better performance. Results from machine learning models were plotted using Seaborn and Matplotlib.

#### 4. Conclusions

In this work, a sufficient amount of data from experimental sources was collected to train machine learning models, which can predict power conversion efficiencies. The accuracy of optimized machine learning models was reasonably high. Pearson correlation analysis provided information about important parameters that play a critical role in PCE prediction. Eig07\_AEA (dm) showed the highest correlation with power conversion efficiency. Its role was the greatest in model training. Multiple machine learning models were tried. The random forest model and bagging model were the best models with coefficient of determination (R<sup>2</sup>) values of 0.892 and 0.887, respectively. This approach can help to select better materials. The findings of our study suggest that machine learning methods provide a way forward for data visualization and performance prediction, which will speed up the industrial implementation of OSCs.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules27185905/s1>, Table S1: Data of organic solar cells based on small molecule donor and fullerene acceptors.

**Author Contributions:** Conceptualization, N.A. and S.U.-D.K. (Salah Ud-Din Khan); data curation, Z.M.E. and S.U.-D.K. (Shahab Ud-Din Khan); investigation, N.A., S.U.-D.K. (Shahab Ud-Din Khan) and S.U.-D.K. (Salah Ud-Din Khan); methodology, S.U.-D.K. (Salah Ud-Din Khan); resources, N.A.; validation, Z.M.E. and S.U.-D.K. (Salah Ud-Din Khan); visualization, N.A., Z.M.E. and S.U.-D.K.; writing—original draft, Z.M.E.; writing—review and editing, S.U.-D.K. (Salah Ud-Din Khan), Z.M.E., N.A. and S.U.-D.K. (Shahab Ud-Din Khan). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project grant number PNURSP2022R11.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Associated data used in this report are given in supporting information.

**Acknowledgments:** The authors express their gratitude to Princess Nourah bint Abdulrahman University Researchers Supporting Project (Grant No. PNURSP2022R11), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Conflicts of Interest:** All authors declared no conflict of interest.

#### References

1. Iqbal, R.; Yasin, G.; Hamza, M.; Ibraheem, S.; Ullah, B.; Saleem, A.; Ali, S.; Hussain, S.; Anh Nguyen, T.; Slimani, Y.; et al. State of the art two-dimensional covalent organic frameworks: Prospects from rational design and reactions to applications for advanced energy storage technologies. *Coord. Chem. Rev.* **2021**, *447*, 214152. [[CrossRef](#)]
2. Sharif, H.M.A.; Farooq, M.; Hussain, I.; Ali, M.; Mujtaba, M.A.; Sultan, M.; Yang, B. Recent innovations for scaling up microbial fuel cell systems: Significance of physicochemical factors for electrodes and membranes materials. *J. Taiwan Inst. Chem. Eng.* **2021**, *129*, 207–226. [[CrossRef](#)]
3. Mahmood, A. Recent research progress on quasi-solid-state electrolytes for dye-sensitized solar cells. *J. Energy Chem.* **2015**, *24*, 686–692. [[CrossRef](#)]
4. Iqbal, R.; Badshah, A.; Ma, Y.-J.; Zhi, L.-J. An Electrochemically Stable 2D Covalent Organic Framework for High-performance Organic Supercapacitors. *Chin. J. Polym. Sci.* **2020**, *38*, 558–564. [[CrossRef](#)]
5. Mahmood, A.; Hu, J.; Tang, A.; Chen, F.; Wang, X.; Zhou, E. A novel thiazole based acceptor for fullerene-free organic solar cells. *Dyes Pigment.* **2018**, *149*, 470–474. [[CrossRef](#)]
6. Mahmood, A.; Yang, J.; Hu, J.; Wang, X.; Tang, A.; Geng, Y.; Zeng, Q.; Zhou, E. Introducing Four 1,1-Dicyanomethylene-3-indanone End-Capped Groups as an Alternative Strategy for the Design of Small-Molecular Nonfullerene Acceptors. *J. Phys. Chem. C* **2018**, *122*, 29122–29128. [[CrossRef](#)]

7. Sharif, H.M.A.; Cheng, H.-Y.; Haider, M.R.; Khan, K.; Yang, L.; Wang, A.-J. NO Removal with Efficient Recovery of N<sub>2</sub>O by Using Recyclable Fe<sub>3</sub>O<sub>4</sub>@EDTA@Fe(II) Complex: A Novel Approach toward Resource Recovery from Flue Gas. *Environ. Sci. Technol.* **2019**, *53*, 1004–1013. [[CrossRef](#)]
8. Sharif, H.M.A.; Mahmood, N.; Wang, S.; Hussain, I.; Hou, Y.-N.; Yang, L.-H.; Zhao, X.; Yang, B. Recent advances in hybrid wet scrubbing techniques for NO<sub>x</sub> and SO<sub>2</sub> removal: State of the art and future research. *Chemosphere* **2021**, *273*, 129695. [[CrossRef](#)]
9. Mahmood, A.; Tang, A.; Wang, X.; Zhou, E. First-principles theoretical designing of planar non-fullerene small molecular acceptors for organic solar cells: Manipulation of noncovalent interactions. *Phys. Chem. Chem. Phys.* **2019**, *21*, 2128–2139. [[CrossRef](#)]
10. Mahmood, A.; HussainTahir, M.; Irfan, A.; Khalid, B.; Al-Sehemi, A.G. Computational Designing of Triphenylamine Dyes with Broad and Red-shifted Absorption Spectra for Dye-sensitized Solar Cells using Multi-Thiophene Rings in  $\pi$ -Spacer. *Bull. Korean Chem. Soc.* **2015**, *36*, 2615–2620. [[CrossRef](#)]
11. Hussain, R.; Hassan, F.; Khan, M.U.; Mehboob, M.Y.; Fatima, R.; Khalid, M.; Mahmood, K.; Tariq, C.J.; Akhtar, M.N. Molecular engineering of A–D–C–D–A configured small molecular acceptors (SMAs) with promising photovoltaic properties for high-efficiency fullerene-free organic solar cells. *Opt. Quantum Electron.* **2020**, *52*, 364. [[CrossRef](#)]
12. Hussain, R.; Mehboob, M.Y.; Khan, M.U.; Khalid, M.; Irshad, Z.; Fatima, R.; Anwar, A.; Nawab, S.; Adnan, M. Efficient designing of triphenylamine-based hole transport materials with outstanding photovoltaic characteristics for organic solar cells. *J. Mater. Sci.* **2021**, *56*, 5113–5131. [[CrossRef](#)]
13. Khalid, M.; Khan, M.U.; Ahmed, S.; Shafiq, Z.; Alam, M.M.; Imran, M.; Braga, A.A.C.; Akram, M.S. Exploration of promising optical and electronic properties of (non-polymer) small donor molecules for organic solar cells. *Sci. Rep.* **2021**, *11*, 21540. [[CrossRef](#)]
14. Khalid, M.; Khan, M.U.; Razia, E.-t.; Shafiq, Z.; Alam, M.M.; Imran, M.; Akram, M.S. Exploration of efficient electron acceptors for organic solar cells: Rational design of indacenodithiophene based non-fullerene compounds. *Sci. Rep.* **2021**, *11*, 19931. [[CrossRef](#)] [[PubMed](#)]
15. Khalid, M.; Momina; Imran, M.; ur Rehman, M.F.; Braga, A.A.C.; Akram, M.S. Molecular engineering of indenoindene-3-ethylrodanine acceptors with A2-A1-D-A1-A2 architecture for promising fullerene-free organic solar cells. *Sci. Rep.* **2021**, *11*, 20320. [[CrossRef](#)] [[PubMed](#)]
16. Khan, M.U.; Khalid, M.; Hussain, R.; Umar, A.; Mehboob, M.Y.; Shafiq, Z.; Imran, M.; Irfan, A. Novel W-Shaped Oxygen Heterocycle-Fused Fluorene-Based Non-Fullerene Acceptors: First Theoretical Framework for Designing Environment-Friendly Organic Solar Cells. *Energy Fuels* **2021**, *35*, 12436–12450. [[CrossRef](#)]
17. Khan, M.U.; Mehboob, M.Y.; Hussain, R.; Fatima, R.; Tahir, M.S.; Khalid, M.; Braga, A.A.C. Molecular designing of high-performance 3D star-shaped electron acceptors containing a truxene core for nonfullerene organic solar cells. *J. Phys. Org. Chem.* **2021**, *34*, e4119. [[CrossRef](#)]
18. Scharber, M.C.; Mühlbacher, D.; Koppe, M.; Denk, P.; Waldauf, C.; Heeger, A.J.; Brabec, C.J. Design Rules for Donors in Bulk-Heterojunction Solar Cells—Towards 10% Energy-Conversion Efficiency. *Adv. Mater.* **2006**, *18*, 789–794. [[CrossRef](#)]
19. Mahmood, A.; Wang, J.-L. Machine learning for high performance organic solar cells: Current scenario and future prospects. *Energy Environ. Sci.* **2021**, *14*, 90–105. [[CrossRef](#)]
20. Mahmood, A.; Wang, J.-L. A time and resource efficient machine learning assisted design of non-fullerene small molecule acceptors for P3HT-based organic solar cells and green solvent selection. *J. Mater. Chem. A* **2021**, *9*, 15684–15695. [[CrossRef](#)]
21. Irfan, A.; Hussien, M.; Mehboob, M.Y.; Ahmad, A.; Janjua, M.R.S.A. Learning from Fullerenes and Predicting for Y6: Machine Learning and High-Throughput Screening of Small Molecule Donors for Organic Solar Cells. *Energy Technol.* **2022**, *10*, 2101096. [[CrossRef](#)]
22. Mahmood, A.; Abdullah Muhammad, I.; Nazar Muhammad, F. Quantum Chemical Designing of Novel Organic Non-Linear Optical Compounds. *Bull. Korean Chem. Soc.* **2014**, *35*, 1391–1396. [[CrossRef](#)]
23. Mahmood, A.; Khan, S.U.-D.; ur Rehman, F. Assessing the quantum mechanical level of theory for prediction of UV/Visible absorption spectra of some aminoazobenzene dyes. *J. Saudi Chem. Soc.* **2015**, *19*, 436–441. [[CrossRef](#)]
24. Mahmood, A.; Irfan, A.; Wang, J.-L. Developing Efficient Small Molecule Acceptors with sp<sup>2</sup>-Hybridized Nitrogen at Different Positions by Density Functional Theory Calculations, Molecular Dynamics Simulations and Machine Learning. *Chem. Eur. J.* **2022**, *28*, e202103712. [[CrossRef](#)] [[PubMed](#)]
25. Mahmood, A.; Irfan, A.; Wang, J.-L. Machine learning and molecular dynamics simulation-assisted evolutionary design and discovery pipeline to screen efficient small molecule acceptors for PTB7-Th-based organic solar cells with over 15% efficiency. *J. Mater. Chem. A* **2022**, *10*, 4170–4180. [[CrossRef](#)]
26. Mehboob, M.Y.; Hussain, R.; Khan, M.U.; Adnan, M.; Umar, A.; Alvi, M.U.; Ahmed, M.; Khalid, M.; Iqbal, J.; Akhtar, M.N.; et al. Designing N-phenylaniline-triazol configured donor materials with promising optoelectronic properties for high-efficiency solar cells. *Comput. Theor. Chem.* **2020**, *1186*, 112908. [[CrossRef](#)]
27. Mahmood, A.; Irfan, A.; Ahmad, F.; Ramzan Saeed Ashraf Janjua, M. Quantum chemical analysis and molecular dynamics simulations to study the impact of electron-deficient substituents on electronic behavior of small molecule acceptors. *Comput. Theor. Chem.* **2021**, *1204*, 113387. [[CrossRef](#)]



28. Iqbal, R.; Ahmad, A.; Mao, L.-J.; Ghazi, Z.A.; Imani, A.; Lu, C.-X.; Xie, L.-J.; Melhi, S.; Su, F.-Y.; Chen, C.-M.; et al. A High Energy Density Self-supported and Bendable Organic Electrode for Redox Supercapacitors with a Wide Voltage Window. *Chin. J. Polym. Sci.* **2020**, *38*, 522–530. [[CrossRef](#)]
29. Janjua, M.R.S.A.; Irfan, A.; Hussien, M.; Ali, M.; Saqib, M.; Sulaman, M. Machine-Learning Analysis of Small-Molecule Donors for Fullerene Based Organic Solar Cells. *Energy Technol.* **2022**, *10*, 2200019. [[CrossRef](#)]
30. Najam, T.; Shah, S.S.A.; Ding, W.; Jiang, J.; Jia, L.; Yao, W.; Li, L.; Wei, Z. An Efficient Anti-poisoning Catalyst against SO<sub>x</sub>, NO<sub>x</sub>, and PO<sub>x</sub>: P, N-Doped Carbon for Oxygen Reduction in Acidic Media. *Angew. Chem. Int. Ed.* **2018**, *57*, 15101–15106. [[CrossRef](#)] [[PubMed](#)]
31. Shah, S.S.A.; Najam, T.; Nazir, M.A.; Wu, Y.; Ali, H.; Rehman, A.U.; Rahman, M.M.; Imran, M.; Javed, M.S. Salt-assisted gas-liquid interfacial fluorine doping: Metal-free defect-induced electrocatalyst for oxygen reduction reaction. *Mol. Catal.* **2021**, *514*, 111878. [[CrossRef](#)]
32. Khalid, M.; Ali, A.; Abid, S.; Tahir, M.N.; Khan, M.U.; Ashfaq, M.; Imran, M.; Ahmad, A. Facile Ultrasound-Based Synthesis, SC-XRD, DFT Exploration of the Substituted Acyl-Hydrazones: An Experimental and Theoretical Slant towards Supramolecular Chemistry. *ChemistrySelect* **2020**, *5*, 14844–14856. [[CrossRef](#)]
33. Khalid, M.; Ali, A.; Asim, S.; Tahir, M.N.; Khan, M.U.; Curcino Vieira, L.C.; de la Torre, A.F.; Usman, M. Persistent prevalence of supramolecular architectures of novel ultrasonically synthesized hydrazones due to hydrogen bonding [X–H···O; X=N]: Experimental and density functional theory analyses. *J. Phys. Chem. Solids* **2021**, *148*, 109679. [[CrossRef](#)]
34. Mahmood, A.; Irfan, A.; Wang, J.-L. Machine Learning for Organic Photovoltaic Polymers: A Minireview. *Chin. J. Polym. Sci.* **2022**, *40*, 870–876. [[CrossRef](#)]
35. Mahmood, A.; Irfan, A. Effect of fluorination on exciton binding energy and electronic coupling in small molecule acceptors for organic solar cells. *Comput. Theor. Chem.* **2020**, *1179*, 112797. [[CrossRef](#)]
36. Mahmood, A.; Hu, J.-Y.; Xiao, B.; Tang, A.; Wang, X.; Zhou, E. Recent progress in porphyrin-based materials for organic solar cells. *J. Mater. Chem. A* **2018**, *6*, 16769–16797. [[CrossRef](#)]
37. Mahmood, A.; Irfan, A. Computational analysis to understand the performance difference between two small-molecule acceptors differing in their terminal electron-deficient group. *J. Comput. Electron.* **2020**, *19*, 931–939. [[CrossRef](#)]
38. Khan, M.U.; Hussain, R.; Mehboob, M.Y.; Khalid, M.; Ehsan, M.A.; Rehman, A.; Janjua, M.R.S.A. First theoretical framework of Z-shaped acceptor materials with fused-chrysene core for high performance organic solar cells. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2021**, *245*, 118938. [[CrossRef](#)]
39. Khan, M.U.; Hussain, R.; Yasir Mehboob, M.; Khalid, M.; Shafiq, Z.; Aslam, M.; Al-Saadi, A.A.; Jamil, S.; Janjua, M.R.S.A. In Silico Modeling of New “Y-Series”-Based Near-Infrared Sensitive Non-Fullerene Acceptors for Efficient Organic Solar Cells. *ACS Omega* **2020**, *5*, 24125–24137. [[CrossRef](#)]
40. Mahmood, A.; Khan, S.U.-D.; Rana, U.A.; Tahir, M.H. Red shifting of absorption maxima of phenothiazine based dyes by incorporating electron-deficient thiadiazole derivatives as  $\pi$ -spacer. *Arab. J. Chem.* **2019**, *12*, 1447–1453. [[CrossRef](#)]
41. Mahmood, A.; Saqib, M.; Ali, M.; Abdullah, M.I.; Khalid, B. Theoretical investigation for the designing of novel antioxidants. *Can. J. Chem.* **2013**, *91*, 126–130. [[CrossRef](#)]
42. Mahmood, A.; Khan, S.U.-D.; Rana, U.A.; Janjua, M.R.S.A.; Tahir, M.H.; Nazar, M.F.; Song, Y. Effect of thiophene rings on UV/visible spectra and non-linear optical (NLO) properties of triphenylamine based dyes: A quantum chemical perspective. *J. Phys. Org. Chem.* **2015**, *28*, 418–422. [[CrossRef](#)]
43. Gevaerts, V.S.; Herzig, E.M.; Kirkus, M.; Hendriks, K.H.; Wienk, M.M.; Perlich, J.; Müller-Buschbaum, P.; Janssen, R.A.J. Influence of the Position of the Side Chain on Crystallization and Solar Cell Performance of DPP-Based Small Molecules. *Chem. Mater.* **2014**, *26*, 916–926. [[CrossRef](#)]
44. Huang, J.; Jia, H.; Li, L.; Lu, Z.; Zhang, W.; He, W.; Jiang, B.; Tang, A.; Tan, Z.a.; Zhan, C.; et al. Fine-tuning device performances of small molecule solar cells via the more polarized DPP-attached donor units. *Phys. Chem. Chem. Phys.* **2012**, *14*, 14238–14242. [[CrossRef](#)] [[PubMed](#)]
45. Sun, S.-X.; Huo, Y.; Li, M.-M.; Hu, X.; Zhang, H.-J.; Zhang, Y.-W.; Zhang, Y.-D.; Chen, X.-L.; Shi, Z.-F.; Gong, X.; et al. Understanding the Halogenation Effects in Diketopyrrolopyrrole-Based Small Molecule Photovoltaics. *ACS Appl. Mater. Interfaces* **2015**, *7*, 19914–19922. [[CrossRef](#)]
46. Najam, T.; Shah, S.S.A.; Ibraheem, S.; Cai, X.; Hussain, E.; Suleman, S.; Javed, M.S.; Tsiakaras, P. Single-atom catalysis for zinc-air/O<sub>2</sub> batteries, water electrolyzers and fuel cells applications. *Energy Stor. Mater.* **2022**, *45*, 504–540. [[CrossRef](#)]
47. Shah, S.S.A.; Najam, T.; Aslam, M.K.; Ashfaq, M.; Rahman, M.M.; Wang, K.; Tsiakaras, P.; Song, S.; Wang, Y. Recent advances on oxygen reduction electrocatalysis: Correlating the characteristic properties of metal organic frameworks and the derived nanomaterials. *Appl. Catal. B* **2020**, *268*, 118570. [[CrossRef](#)]
48. Khalid, M.; Ali, A.; Khan, M.U.; Tahir, M.N.; Ahmad, A.; Ashfaq, M.; Hussain, R.; de Alcântara Morais, S.F.; Braga, A.A.C. Non-covalent interactions abetted supramolecular arrangements of N-Substituted benzylidene acetohydrazide to direct its solid-state network. *J. Mol. Struct.* **2021**, *1230*, 129827. [[CrossRef](#)]
49. Siddiqui, W.A.; Khalid, M.; Ashraf, A.; Shafiq, I.; Parvez, M.; Imran, M.; Irfan, A.; Hanif, M.; Khan, M.U.; Sher, F.; et al. Antibacterial metal complexes of o-sulfamoylbenzoic acid: Synthesis, characterization, and DFT study. *Appl. Organomet. Chem.* **2022**, *36*, e6464. [[CrossRef](#)]



- 
50. Mahmood, A.; Wang, J.-L. A Review of Grazing Incidence Small- and Wide-Angle X-Ray Scattering Techniques for Exploring the Film Morphology of Organic Solar Cells. *Sol. RRL* **2020**, *4*, 2000337. [[CrossRef](#)]
  51. Khan, M.U.; Khalid, M.; Arshad, M.N.; Khan, M.N.; Usman, M.; Ali, A.; Saifullah, B. Designing Star-Shaped Subphthalocyanine-Based Acceptor Materials with Promising Photovoltaic Parameters for Non-fullerene Solar Cells. *ACS Omega* **2020**, *5*, 23039–23052. [[CrossRef](#)] [[PubMed](#)]
  52. Mauri, A.; Consonni, V.; Pavan, M.; Todeschini, R. DRAGON software: An easy approach to molecular descriptor calculations. *MATCH Commun. Math. Comput. Chem.* **2006**, *56*, 237–248. [[CrossRef](#)]