

Simultaneous optimization of donor/acceptor pairs and device specifications for non-fullerene organic solar cells using a QSPR model with morphological descriptors

*Yaping Wen, Bohan Yan, Théophile Gaudin, * Jing Ma, and Haibo Ma**

School of Chemistry and Chemical Engineering

Jiangsu Key Laboratory of Vehicle Emissions Control

Nanjing University

Nanjing 210023, China

E-mail: gaudin.theophile@gmail.com; haibo@nju.edu.cn

Keywords

non-fullerene organic solar cells, synergistic donor/acceptor pairs, device specifications optimization, machine learning, virtual screening

Abstract

In addition to designing new donor (D) and/or acceptor (A) molecules, the optimization of experimental fabrication conditions for the organic solar cells (OSCs) is also a complex, multidimensional challenge, which hasn't been theoretically explored. Herein, a new framework for simultaneous optimizing D/A molecule pairs and device specifications of OSCs is proposed, through a quantitative structure-property relationships (QSPR) model built by machine learning. Combining the device parameters with structural and electronic variables, the built QSPR model achieved

unprecedentedly high accuracy and consistency. Additionally, a huge chemical space containing 1,942,785 D/A pairs is explored to find potential synergistic ones. Favorable experimental parameters such as root-mean-square (*RMS*) and the D/A ratio (*DAratio*) are further screened by grid search methods. Overall, this study suggests the feasibility to optimize D/A molecule pairs and device specifications simultaneously by enabling better-informed and data-driven techniques and this could facilitate the acceleration of improving OSCs efficiencies.

Organic solar cells (OSCs) based on non-fullerene acceptors attract much attention due to their easily tunable physical-chemical properties, significantly reduced open-circuit voltage losses, and good device stability.¹⁻³ Incremental progresses in materials design and device engineering recently led the power conversion efficiency (PCE) of OSCs to exceed 18%,⁴⁻⁷ approaching the level necessary for large-scale energy production. Donor (D)/acceptor (A) bulk-heterojunction structure as the photoactive layer provides an effective approach to separate the bound excitons generated by light absorption into free carriers, preventing their recapture behavior that leads to a decrease in PCE.⁸ In that context, reliable quantitative structure-property relationship (QSPR) between photoactive materials and device efficiencies is a shortcut for further improving PCE of OSCs⁹, and data-driven machine learning (ML) is one of the promising technologies in this scheme.¹⁰

The impact of D and A molecules on PCE of OSCs is non-additive, and in particular, synergies have been observed for some D/A pairs.^{11,12} However, most theoretical studies in this field were focused on optimizing either D or A separately.¹³⁻¹⁹ In 2019, Troisi et al.²⁰ highlighted the importance of multi-component simultaneous optimization. Based on the theoretical calculation of Morgan fingerprints and electronic descriptors, a ML model taking D/A pairs as input structures for predicting PCE was proposed, with a Pearson's correlation coefficient (r) of 0.78. Subsequently, Troisi et al.²¹ explored the effect of different descriptor sets on predictive ability of ML models, and obtained the best performance with a maximum r of 0.73. In 2020, Min et al.²² developed a PCE predictive model focusing on non-fullerene OSCs based on simple

counts of molecular fragments within D and A molecules, avoiding the use of quantum chemistry, and obtained a r of 0.71. These studies provide valuable insights for the joint optimization of D/A pairs in OSCs.

Complementary to the molecular structure of the D/A pairs, the macroscopic morphology of D/A blends has great influences on the optoelectronic conversion, which mainly depends on the device engineering.²³⁻²⁷ However, the experimental optimization of device specifications is traditionally a trial-and-error, time-consuming process and up to now is unexplored by theoretical ML models of OSCs. Many parameters can impact the final blend morphology, such as the introduction of additives, or careful post-processing.²⁴ Fortunately, two device specifications are regularly reported in a unifiable manner, allowing their easy conversion into descriptors for ML. The first one is the D/A ratio (*DAratio*), and the second one, reflecting the quality of post-processing, is the surface roughness of the blends, usually measured as its root-mean-square error (*RMS*) from atomic force microscopy.

The goal of this work is to establish a reliable QSPR model that combines not only molecular descriptors but also device specifications for non-fullerene OSCs, and to achieve a more comprehensive theoretical inverse optimization of device specifications for unexplored promising D/A pairs. To achieve such simultaneous optimization, five steps are included in the workflow of this study (Figure 1):

- a. data collection of D/A pairs for non-fullerene OSCs;
- b. feature engineering of D/A pairs, including structural, electronic, and device descriptors;

- c. QSPR model construction, with the comparison of various descriptor combinations and ML algorithms on the performance of models by a new two-level evaluation criterion;
- d. high throughput screening of D/A pairs using a structural-based QSPR model, and D/A pairs with PCE greater than 14% were selected;
- e. inverse optimization of device specifications (*RMS* and *DARatio*) for the screened D/A pairs, and potentially favorable values will finally feed back to experimentalists.

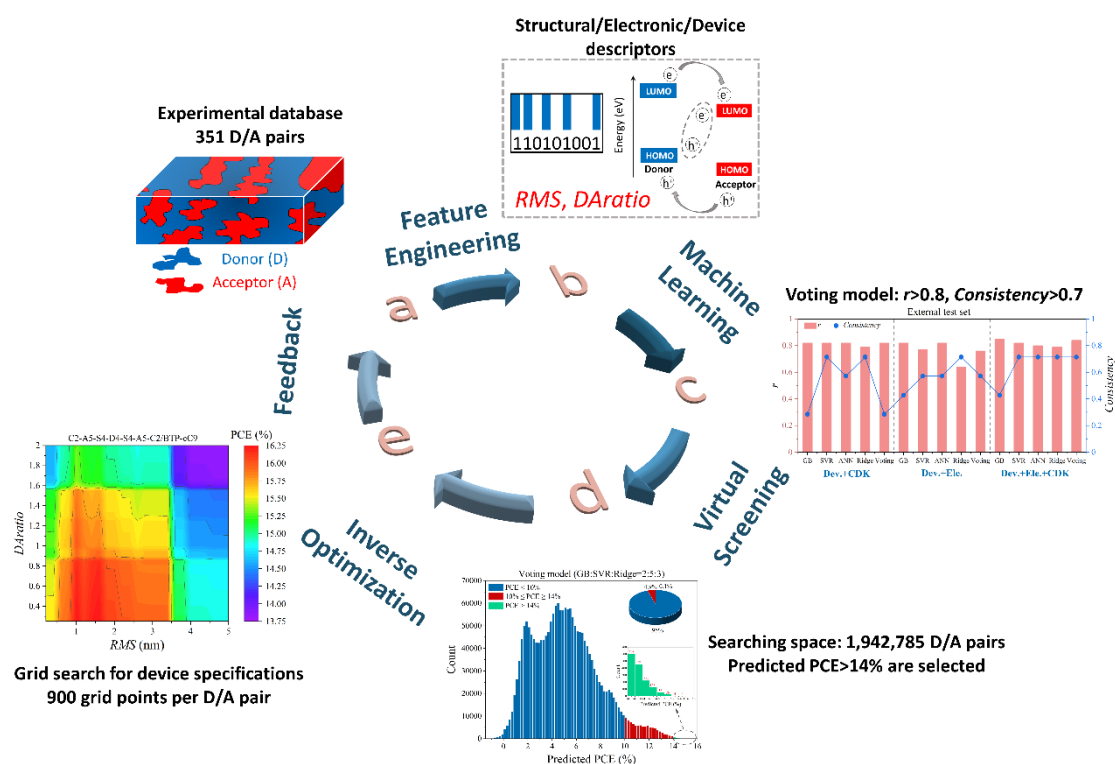


Figure 1 Workflow of screening and optimization of D/A pairs and device parameters.

In this work, 351 D/A pairs, including 44 polymer donors and 195 non-fullerene acceptors, were collected from 113 experimental papers published between 2016 and 2020. Their chemical structure, *RMS*, *DARatio*, and corresponding PCE are summarized in Table S1. Here, a monomeric unit was taken for all polymer donors. For specific

RMS and *DARatio*, the highest PCE of each D/A pair was selected as the fitting target of ML to learn the maximize potential of materials. In our database, the range of *RMS*, *DARatio* and PCE are 0.37 nm~50.15 nm, 0.33~2 and 0.04%~17.8%, respectively. The black dots in Figure 2a show the distribution of three experimental variables, and the contour plot of PCE is filled with color. High PCE mainly distributed in the range of $RMS \leq 5\text{nm}$ and $DARatio \leq 1$ (Figure 2b), indicating that a relatively smooth surface of D/A blends and a smaller D/A ratio are beneficial to the improvement of device efficiency. More than 90% of the collected *RMS* and *DARatio* values are $\leq 5\text{nm}$ and ≤ 1 , respectively (Figures 2c-d). See Table S2 for more details about the database. In order to enable ML to learn the features of different PCE from this unbalanced distribution, the manual layering method for data was adopted. 351 pieces of data were first divided into 8 groups (Table S3), and then 15% of the data from each group were randomly selected and combined together to obtain the external test set. The training set was composed of the remaining 85% of data from each set, and the hyper-parameters optimization and cross-validation of ML model were performed in this set. The external test set was not visible to ML model throughout the process, allowing to assess the predictive power of the model.

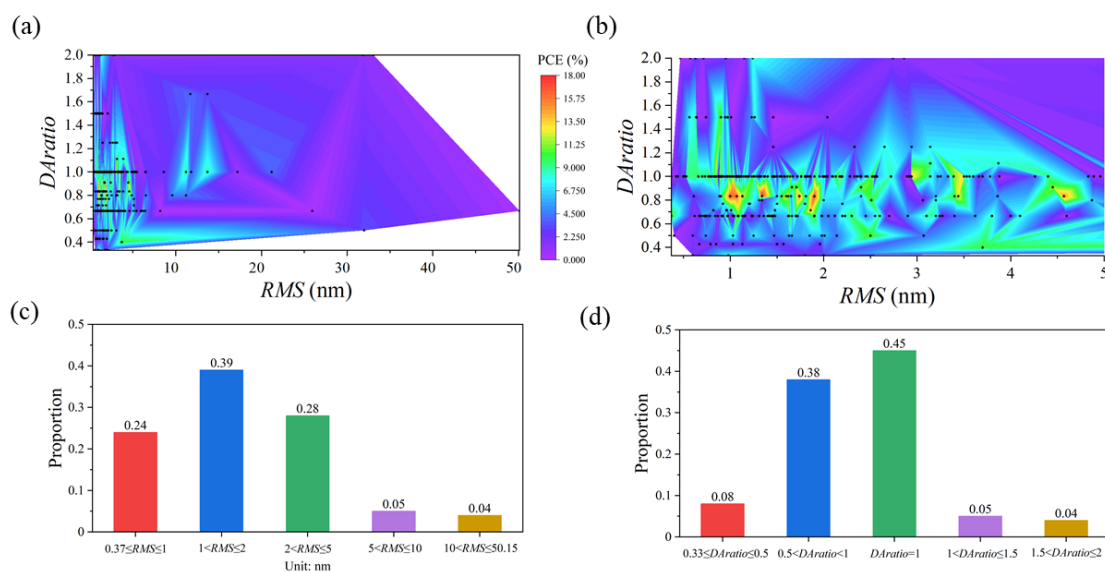


Figure 2 Non-fullerene OSCs database of 351 D/A pairs. (a) Distribution of *RMS*, *DAratio* and PCE, where PCE is filled with contour colors; (b) An enlarged view of $0.37 \text{ nm} \leq \text{RMS} \leq 5 \text{ nm}$ for (a); (c)-(d) Proportion of *RMS* and *DAratio* with five parts, respectively.

Note that a given D/A pair can exhibit substantial PCE variability due to different device specifications, notably *RMS* and *DAratio*. There were 66 and 7 of such data clusters in training sets and test sets, respectively. To account for this peculiar data structure, a two-level statistical assessment for predictive model was adopted in this study. The correlation between experimental and predicted PCE should be evaluated by the conventional indices such as r (see Supporting Information for expressions). On the other hand, it is necessary to evaluate whether the order of predicted PCE is consistent with measured values for the device fabrication parameters-controlled data clusters. Here, we define three estimated classes, i.e., fully consistent, partial consistent, and inconsistent, as well as a new index to evaluate proportion of fully consistent case,

Consistency. Examples of how to classify and expressions for *Consistency* can be found in Supporting Information.

For feature engineering, three types of descriptors, namely molecular fingerprints, electronic properties and device specifications were adopted to describe the potential of D/A pairs. We compared six fingerprinting approaches with different lengths and contents, i.e., E-State (79 bits),²⁸ MACCS (166 bits),²⁹ Substructure (307 bits),³⁰ 2D atom pairs (780 bits),³¹ Pubchem (881 bits)³² and CDK (1024 bits).³³ Fingerprints of D and A were generated using the ChemDes platform,³⁴ and then horizontally concatenated to obtain that of D/A pairs. Electronic descriptors (Ele. for short) are chemical properties closely associated with photoelectric conversion that require quantum chemical calculations. As a trade-off between computational cost and accuracy, totally 19 ground-state properties, related to energy levels and polarity were calculated by density functional theory (DFT) at the M06-2X/6-31G(d)^{35,36} level implemented in the Gaussian 16 program.³⁷ Their symbols and meanings are shown in Table S4 and Supporting information. A diagram of energy level descriptors for donor and acceptor are presented in Figure S1. The two aforementioned devices specifications (Dev. for short), i.e., *RMS* and *D_Aratio*, are strongly and nonlinearly related to the surface roughness of D/A blends. A small *RMS*, equivalent to a high film smoothness, implies a high D/A miscibility that favors exciton dissociation due to the increased D/A interface, but a too small value is not recommended because charge recombination can easily occur in highly mixed D/A heterojunctions. Similarly, a well-placed *D_Aratio* is essential for a specific D/A pair. It should be noted that in principle, *RMS* could be

influenced by *DAratio*. Fortunately, as is shown in Figure S2, no significant linear correlation between *RMS* and *DAratio* was observed within our dataset. (From Figure S2 one can also see Pearson's correlation between each pair of variables mentioned in this study.)

Various combinations of the three types descriptors, Dev.+Fp, Dev.+Ele., and Dev.+Ele.+Fp, were applied to construct QSPR models for predicting PCE, respectively. To avoid overflowing the descriptor selection algorithm, we screened the fingerprinting approaches first, and then use the best one for subsequent analysis. In the process of fingerprint screening, we performed five ML techniques, including ridge regression (Ridge), gradient boosting (GB), support vector machines (SVR), artificial neural network (ANN), and a voting ensemble approach (Voting) that combines the predictions from multiple other regression models. 60 ML models were generated on the basis of 12 inputs that formed by the combination of six fingerprinting approaches and two descriptor combination types (Dev.+Fp/Dev.+Ele.+Fp). All ML models were developed with Python scripts using the scikit-learn package.³⁸ Predictive performance of these models on 10-fold cross-validation set and external test set is listed in Tables S5-S9. Overall, CDK fingerprints performed slightly better than others, which indicates that longer fingerprints are suggested to identify the relevant patterns leading to higher PCE. Hence, CDK fingerprints were used in following studies where fingerprints were involved.

In the next step, the optimal predictive model was selected based on the combination of *r* and *Consistency* criteria. Three types of descriptor combinations

(Dev.+CDK, Dev.+Ele, and Dev.+Ele.+CDK) and five ML techniques (Ridge, GB, SVR, ANN and Voting) can generate 15 ML models. Performance of these predictive models is shown in Figure 3a-3b. Red bar graphs represent the r between predicted and experimental PCE, and blue points show the *Consistency* as defined above. Most predictive models on both cross-validation set and external test set can obtain an unprecedentedly high r of greater than 0.8, but *Consistency* of these models varies considerably. Encouragingly, the Voting model with Dev.+Ele.+CDK as inputs performed well in both r and *Consistency*. The predicted values of this model were voted by the results of GB, SVR and Ridge models with an optimized voting ratio of 5:2:4 (GB:SVR:Ridge=5:2:4). Details of correlation and consistency between experimental and predicted PCE of the Voting model are presented in Figure 3c-d. The good correlation both on the cross-validation set ($r=0.88$) and external test set ($r=0.84$) demonstrates the robustness and excellent generalization, which is superior to previous QSPR models to the best of my knowledge. For the 66 training groups and 7 test groups with same D/A pairs but different *DARatio* and *RMS*, *Consistency* of 0.79 (52 groups) and 0.71 (5 groups) for training set and external test set, respectively, can be achieved (Figure 3d), i.e. our ML model can well identify the optimal *DARatio* and *RMS* parameters from various candidate conditions for a given D/A pair (fully consistent in more than 70% training and test cases). Compared with the ML models that did not take device specifications into account (*Consistency*=0), the predictive consistency of PCE by our Voting model (GB:SVR:Ridge=5:2:4) was significantly improved both in

training set ($Consistency=0.79$) and test set ($Consistency=0.71$). Predictive consistency details of other models are summarized in Table S10.

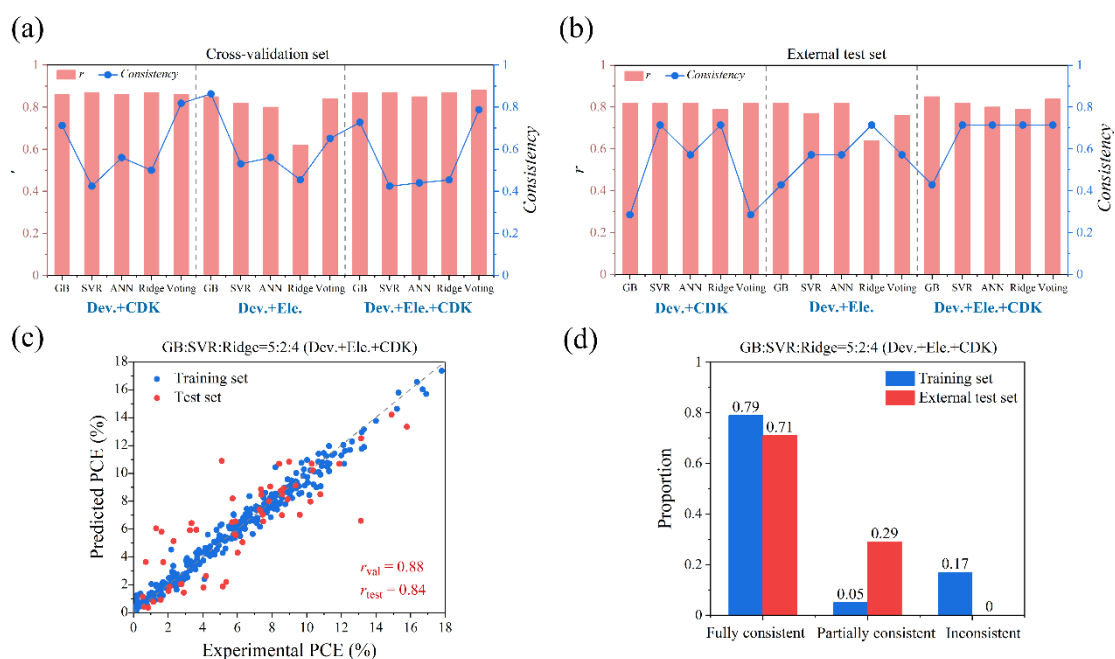


Figure 3 Performance of QSPR models. (a)~(b) r and $Consistency$ of QSPR models based on different descriptor combinations on cross-validation set and external test set. The results of Voting models are obtained by the optimized weights of GB, SVR and Ridge models; (c) Comparison of the predicted PCE using the Voting model (GB:SVR:Ridge=5:2:4) based on Dev.+Ele.+CDK descriptors and experimental PCE. (d) $Consistency$ between predicted PCE using the Voting model (GB:SVR:Ridge=5:2:4) based on Dev.+Ele.+CDK descriptors and experimental PCE, only for 66 training groups and 7 test groups with same D/A pairs but different D/A ratio and RMS values.

An effective approach for discovering potential D/A pairs is the combination of QSPR model and high-throughput screening. To explore good-performing D/A pairs, a huge search space of 1,942,785 D/A pairs was obtained by pairing 9963 designed donors¹⁷ with 194 reported non-fullerene acceptors in our database. To deal with the large number of candidates, a simplified QSPR Voting model (GB:SVR:Ridge=2:5:3) based on structural CDK fingerprints (Figure S3) was used for the preliminary screening. Figure 4 shows the distribution of predicted PCE for 1,942,785 D/A pairs using this Voting model. PCE in the range of <10%, 10%~14%, and >14% are represented by blue, red, and green color, respectively. As can be seen, a very small D/A space (0.1%) exhibited excellent PCEs of higher than 14%, which is comparable to the currently cutting-edge D/A pairs.³⁹⁻⁴¹

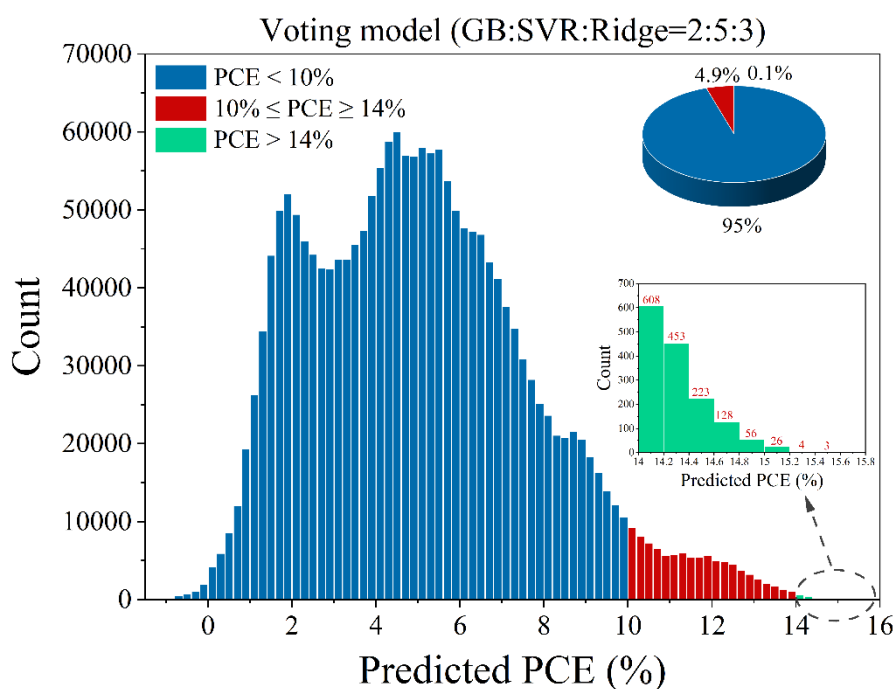


Figure 4 Histogram of PCE predicted by Voting model (GB:SVR:Ridge=2:5:3) based on CDK fingerprints for 194,2785 combined D/A pairs. PCE in the range of <10%,

10%~14%, and >14% are represented by blue, red, and green color, respectively. Inset: An enlarged view of the PCE>14% range.

In addition to design outstanding D/A pairs, we further tried to suggest reasonable ratio and blends morphology for experimentalists. Grid search method is adopted here to inversely optimize the *DARatio* and *RMS* for good-performing D/A pairs. Similar to the hyper-parameter optimization of ML model, we first set the range of 0.2 nm-5nm and 0.25-2, with the interval of 0.2 nm and 0.05 for *RMS* and *DARatio*, respectively, and then search globally for the optimal values that correspond to the highest PCE from 900 grid points generated by 25 and 36 values of *RMS* and *DARatio*, respectively. For 1501 D/A pairs with PCEs over 14% (predicted by the simplified QSPR Voting model), the grid search was performed based on the Voting model (GB:SVR:Ridge=5:2:4). As an example, Figure 5a shows the contour color fill plot of PCE for C2-A5-S4-D4-S4-A5-C2/BTP-eC9 pair (Figure S4). With different *RMS* and *DARatio* values, PCE fluctuated in the range of 13.75%-16.25%, and the favorable range of *DARatio* and *RMS* were <0.8 and 0.8 nm-1.7 nm, respectively. Distribution of the difference between the highest and lowest PCE of 1501 D/A pairs is shown in Figure 5b, which verifies that upper limit of materials performance can be increased by proper device fabrication methods. Theoretical results provide potentially favorable device specifications to reduce trial-and-error testing and provides a new strategy for the development of OSCs. Notably, only three acceptors, BTP-eC9,⁷ BTP-S2,⁴⁰ and BTP-4F,⁴¹ (Figure S4) were found in D/A pairs with PCE greater than 14%, indicating their great potential in

developing new non-fullerene OSCs. Structurally, they shared the same skeleton and were variants of the excellent Y6⁴² acceptor (Figure S4). With the appropriate energy level and good light absorption, this family of acceptors is compatible with many donors.

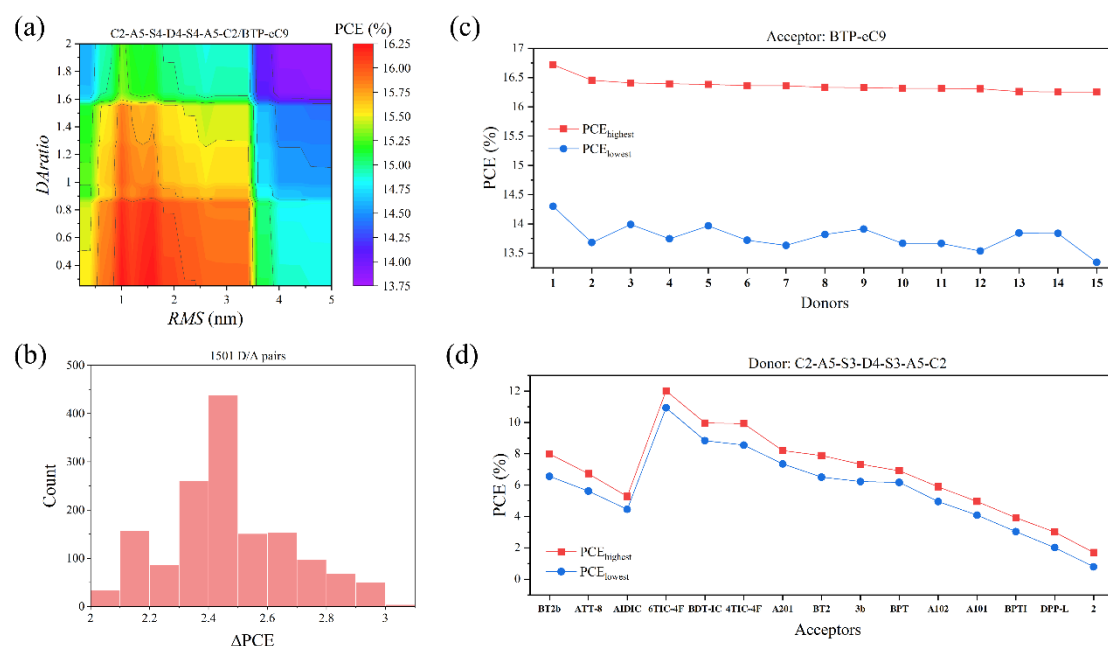


Figure 5 (a) Contour color fill plot of PCE with different RMS and $DAratio$ for C2-A5-S4-D4-A5-C2/BTP-eC9 pair; (b) Difference of highest and lowest PCE for 1501 D/A pairs using voting model (GB:SVR:Ridge=5:2:4); (c) Effect of 15 different donors on PCE based on the same acceptor BTP-EC9; (d) Effect of 15 different acceptors on PCE based on the same donor C2-A5-S3-D4-S3-A5-C2.

It is interesting to observe that the sensitivity of D and A in D/A pair to PCE is quite different. Figure 5c-d shows the tunability on PCE of donors and acceptors, respectively. Paired with the same acceptor BTP-eC9, 15 different donors generated similar PCEs, with differences fluctuating ranging within 1%. By contrast, PCEs of

D/A pairs based on the same donor (C2-A5-S3-D4-S3-A5-C2) and 15 different acceptors varied greatly. The highest (>12%) and the lowest (<1%) efficiency were obtained from C2-A5-S3-D4-S3-A5-C2/6TIC-4F and C2-A5-S3-D4-S3-A5-C2/2 pairs (Figure S4), respectively. These results suggest that acceptors have greater influence on PCE than donors, which may be a useful consideration to keep in mind for material design of OSCs. Figure 5c-d also shows how different the highest (red dots) and lowest PCE (blue dots) of these D/A pairs can be when *RMS* and *DARatio* change, indicating the relevance the device specifications within our QSPR model.

In conclusion, we achieved the simultaneous optimization of photoactive materials and device specifications for OSCs using an upgraded QSPR model. In addition to structural and electronic descriptors, two critical device specifications, *RMS* and *DARatio* were also utilized as input features for ML modeling. The weighted Voting model (GB:SVR:Ridge=5:2:4) based on Dev.+Ele.+Fp input features exhibiting unprecedentedly high stability and generalization ability, with *r* of 0.88 and 0.84, as well as *Consistency* of 0.79 and 0.71 on the cross-validation set and test set, respectively. Additionally, a huge searching space with up to 1,942,785 D/A pairs was constructed to screen out potential D/A pairs with PCE>14%. *RMS* and *DARatio* were also optimized for these pairs using a grid search method. The predicted efficiencies of D/A pairs were significantly improved by optimizing the device specifications, which can potentially expand the upper limit of material performance. Overall, we showed that ML can be used for not only molecular screening but also experimental parameters

optimization for OSCs, which takes an important step further into the practical theoretical guidance in materials engineering.

Supporting Information

Database details, descriptors description, OSPR models performance, definition of evaluation indicators, and chemical structures of mentioned molecules.

Acknowledgements

The work was supported by the National Natural Science Foundation of China (Grant Numbers 22073045, 21950410517, 22033004 and 21873045). We are grateful to the High-Performance Computing Center of Nanjing University for doing the numerical calculations in this paper on its blade cluster system.

References

- (1) Zhang, G.; Zhao, J.; Chow, P. C. Y.; Jiang, K.; Zhang, J.; Zhu, Z.; Zhang, J.; Huang, F.; Yan, H. Nonfullerene Acceptor Molecules for Bulk Heterojunction Organic Solar Cells. *Chem. Rev.* **2018**, *118*, 3447-3507.
- (2) Hou, J.; Inganäs, O.; Friend, R. H.; Gao F. Organic Solar Cells Based on Non-fullerene Acceptors. *Nat. Mater.* **2018**, *17*, 119-128.
- (3) Armin, A.; Li, W.; Sandberg, O. J.; Xiao, Z.; Ding, L.; Nelson, J.; Neher, D.; Vandewal, K.; Shoaee, S.; Wang, T.; Ade, H.; Heumüller, T.; Brabec, C.; Meredith, P. A History and Perspective of Non-Fullerene Electron Acceptors for Organic Solar Cells. *Adv. Energy Mater.*, DOI: 10.1002/aenm.202003570.

- (4) Meng, L.; Zhang, Y.; Wan, X.; Li, C.; Zhang, X.; Wang, Y.; Ke, X.; Xiao, Z.; Ding, L.; Xia, R.; Yip, H.-L.; Cao, Y.; Chen, Y. Organic and Solution-Processed Tandem Solar Cells with 17.3% Efficiency. *Science* **2018**, *361*, 1094-1098.
- (5) Zhan, L.; Li, S.; Lau, T.-K.; Cui, Y.; Lu, X.; Shi, M.; Li, C.-Z.; Li, H.; Hou, J.; Chen, H. Over 17% Efficiency Ternary Organic Solar Cells Enabled by Two Non-Fullerene Acceptors Working in Alloy-like Model. *Energy Environ. Sci.* **2020**, *13*, 635-645.
- (6) Liu, Q.; Jiang, Y.; Jin, K.; Qin, J.; Xu, J.; Li, W.; Xiong, J.; Liu, J.; Xiao, Z.; Sun, K.; Yang, S.; Zhang, X.; Ding, L. 18% Efficiency Organic Solar Cells. *Sci. Bull.* **2020**, *65*, 272-275.
- (7) Cui, Y.; Yao, H.; Zhang, J.; Xian, K.; Zhang, T.; Hong, L.; Wang, Y.; Xu, Y.; Ma, K.; An, C.; He, C.; Wei, Z.; Gao, F.; Hou, J. Single-Junction Organic Photovoltaic Cells with Approaching 18% Efficiency. *Adv. Mater.* **2020**, *32*, 1908205.
- (8) Yu, G.; Gao, J.; Hummelen, J. C.; Wudl, F.; Heeger, A. J. Polymer Photovoltaic Cells: Enhanced Efficiencies via a Network of Internal Donor-Acceptor Heterojunctions. *Science* **1995**, *270*, 1789-1791.
- (9) Nasrabadi, N. M. Pattern Recognition and Machine Learning. *J. Electron. Imaging* **2007**, *16*, 049901.
- (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*, DOI: 10.1002/nano.202100006.

- (11) Zhao, W.; Li, S.; Yao, H.; Zhang, S.; Zhang, Y.; Yang, B.; Hou, J. Molecular Optimization Enables over 13% Efficiency in Organic Solar Cells. *J. Am. Chem. Soc.* **2017**, *139*, 7148-7151.
- (12) Liu, D.; Yang, L.; Wu, Y.; Wang, X.; Zeng, Y.; Han, G.; Yao, H.; Li, S.; Zhang, S.; Zhang, Y.; Yi, Y.; He, C.; Ma, W.; Hou, J. Tunable Electron Donating and Accepting Properties Achieved by Modulating the Steric Hindrance of Side Chains in A-D-A Small-Molecule Photovoltaic Materials. *Chem. Mater.* **2018**, *30*, 619-628.
- (13) Lopez, S. A.; Sanchez-Lengeling, B.; Soares, J. d. G.; Aspuru-Guzik, A. Design Principles and Top Non-Fullerene Acceptor Candidates for Organic Photovoltaics. *Joule* **2017**, *1*, 857-870.
- (14) Sahu, H.; Rao, W.; Troisi, A.; Ma, H. Toward Predicting Efficiency of Organic Solar Cells via Machine Learning and Improved Descriptors. *Adv. Energy Mater.* **2018**, *8*, 1801032.
- (15) Padula, D.; Simpson, J. D.; Troisi, A. Combining Electronic and Structural Features in Machine Learning Models to Predict Organic Solar Cells Properties. *Mater. Horiz.* **2019**, *6*, 343-349.
- (16) Peng, S.-P.; Zhao, Y. Convolutional Neural Networks for the Design and Analysis of Non-Fullerene Acceptors. *J. Chem. Inf. Model.* **2019**, *59*, 4993-5001.
- (17) Sahu, H.; Yang, F.; Ye, X.; Ma, J.; Fang, W.; Ma, H. Designing Promising Molecules for Organic Solar Cells via Machine Learning Assisted Virtual Screening. *J. Mater. Chem. A* **2019**, *7*, 17480-17488.

- (18) Sahu, H.; Ma, H. Unraveling Correlations between Molecular Properties and Device Parameters of Organic Solar Cells Using Machine Learning. *J. Phys. Chem. Lett.* **2019**, *10*, 7277-7284.
- (19) Sun, W.; Zheng, Y.; Yang, K.; Zhang, Q.; Shah, A. A.; Wu, Z.; Sun, Y.; Feng, L.; Chen, D.; Xiao, Z.; Lu, S.; Li, Y.; Sun, K. Machine Learning-Assisted Molecular Design and Efficiency Prediction for High-Performance Organic Photovoltaic Materials. *Sci. Adv.* **2019**, *5*, eaay4275.
- (20) Padula, D.; Troisi, A. Concurrent Optimization of Organic Donor-Acceptor Pairs through Machine Learning. *Adv. Energy Mater.* **2019**, *9*, 1902463.
- (21) Zhao, Z.-W.; Cueto, M. d.; Geng, Y.; Troisi, A. Effect of Increasing the Descriptor Set on Machine Learning Prediction of Small Molecule-Based Organic Solar Cells. *Chem. Mater.* **2020**, *32*, 7777-7787.
- (22) Wu, Y.; Guo, J.; Sun, R.; Min, J. Machine Learning for Accelerating the Discovery of High-Performance Donor/Acceptor Pairs in Non-Fullerene Organic Solar Cells. *npj Comput. Mater.* **2020**, *6*, 120.
- (23) Wang, Z.; Gao, K.; Kan, Y.; Zhang, M.; Qiu, C.; Zhu, L.; Zhao, Z.; Peng, X.; Feng, W.; Qian, Z.; Gu, X.; Jen, A. K.-Y.; Tang, B. Z.; Cao, Y.; Zhang, Y.; Liu, F. The Coupling and Competition of Crystallization and Phase Separation, Correlating Thermodynamics and Kinetics in OPV Morphology and Performances. *Nat. Commun.* **2021**, *12*, 332.
- (24) Zhao, F.; Wang, C.; Zhan, X. Morphology Control in Organic Solar Cells. *Adv. Energy Mater.* **2018**, *8*, 1703147.

- (25) Xie, Y.; Yang, F.; Li, Y.; Uddin, M. A.; Bi, P.; Fan, B.; Cai, Y.; Hao, X.; Woo, H. Y.; Li, W.; Liu, F.; Sun, Y. Morphology Control Enables Efficient Ternary Organic Solar Cells. *Adv. Mater.* **2018**, *30*, 1803045.
- (26) Karki, A.; Gillett, A. J.; Friend, R. H.; Nguyen, T.-Q. The Path to 20% Power Conversion Efficiencies in Nonfullerene Acceptor Organic Solar Cells. *Adv. Energy Mater.*, DOI: 10.1002/aenm.202003441.
- (27) Du, C.; Ji, Y.; Xue, J.; Hou, T.; Tang, J.; Lee, S.-T.; Li, Y. Morphology and Performance of Polymer Solar Cell Characterized by DPD Simulation and Graph Theory. *Sci. Rep.* **2015**, *5*, 16854.
- (28) Hall, L. H.; Kier, L. B. Electrotological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039-1045.
- (29) MACCS. http://www.scbdd.com/pybel_desc/fps-maccs/.
- (30) Substructure. http://www.scbdd.com/cdk_desc/fps-substructure/
- (31) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom Pairs as Molecular Features in Structure-Activity Studies: Definition and Applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64-73.
- (32) Pubchem. http://www.scbdd.com/cdk_desc/fps-pubchem/.
- (33) Steinbeck, C.; Hoppe, C.; Kuhn, S.; Floris, M.; Guha, R.; Willighagen, E. L. Recent Developments of the Chemistry Development Kit (CDK)-An Open-Source Java Library for Chemo-and Bioinformatics. *Curr. Pharm. Des.* **2006**, *12*, 2111-2120.

- (34) Dong, J.; Cao, D.-S.; Miao, H.-Y.; Liu, S.; Deng, B.-C.; Yun, Y.-H.; Wang, N.-N.; Lu, A.-P.; Zeng, W.-B.; Chen, A. F. ChemDes: An Integrated Web-based Platform for Molecular Descriptor and Fingerprint Computation. *J. Cheminformatics* **2015**, *7*, 60.
- (35) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 215-241.
- (36) Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theoret. Chim. Acta* **1973**, *28*, 213-222.
- (37) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, Revision C.01, Gaussian, Inc., Wallingford CT **2016**.

- (38) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825-2830.
- (39) Dai, S.; Zhou, J.; Chandrabose, S.; Shi, Y.; Han, G.; Chen, K.; Xin, J.; Liu, K.; Chen, Z.; Xie, Z.; Ma, W.; Yi, Y.; Jiang, L.; Hodgkiss, J. M.; Zhan, X. High-Performance Fluorinated Fused-Ring Electron Acceptor with 3D Stacking and Exciton/Charge Transport. *Adv. Mater.* **2020**, *32*, 2000645.
- (40) Li, S.; Zhan, L.; Jin, Y.; Zhou, G.; Lau, T.-K.; Qin, R.; Shi, M.; Li, C.-Z.; Zhu, H.; Lu, X.; Zhang, F.; Chen, H. Asymmetric Electron Acceptors for High-Efficiency and Low-Energy-Loss Organic Photovoltaics. *Adv. Mater.* **2020**, *32*, 2001160.
- (41) Xu, Y.; Yao, H.; Ma, L.; Hong, L.; Li, J.; Liao, Q.; Zu, Y.; Wang, J.; Gao, M.; Ye, L.; Hou, J. Tuning the Hybridization of Local Exciton and Charge-Transfer States in Highly Efficient Organic Photovoltaic Cells. *Angew. Chem. Int. Ed.* **2020**, *59*, 9004-9010.
- (42) Yuan, J.; Zhang, Y.; Zhou, L.; Zhang, G.; Yip, H.-L.; Lau, T.-K.; Lu, X.; Zhu, C.; Peng, H.; Johnson, P. A.; Leclerc, M.; Cao, Y.; Ulanski, J.; Li, Y.; Zou, Y. Single-Junction Organic Solar Cell with over 15% Efficiency Using Fused-Ring Acceptor with Electron-Deficient Core. *Joule* **2019**, *3*, 1140-1151.