ACS APPLIED MATERIALS & INTERFACES

www.acsami.org

Research Article

Directed Message Passing Neural Network for Predicting Power Conversion Efficiency in Organic Solar Cells

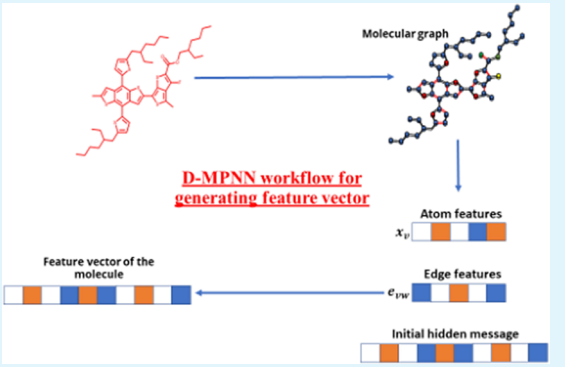
Prateek Malhotra, Subhayan Biswas, and Ganesh D. Sharma*

Cite This: https://doi.org/10.1021/acsami.3c08068

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Organic solar cells (OSCs) have emerged as a promising technology for renewable energy generation, and researchers are constantly exploring ways to improve their efficiency. For prediction of photovoltaic properties in OSCs, many machine learning models have been used in the past. All the models are used with fixed molecular descriptors and molecular fingerprints as input for power conversion efficiency (PCE) prediction. Recently, the graph neural network (GNN), which can model graph structures of the molecule, has received increasing attention as a method that could potentially overcome the limitations of fixed descriptors by learning the task-specific representations using graph convolutions. In this study, we have used the directed message passing neural network (D-MPNN), an emerging type of GNN for predicting PCE of organic solar cells, and the results are compared for the same train and test set with fixed descriptors and fingerprints. The excellent performance demonstrated by the D-MPNN model in this investigation highlights its potential for predicting PCE, surpassing the limitations of conventional fixed descriptors.



KEYWORDS: *organic solar cells, machine learning, directed message passing neural network, power conversion efficiency, donor:acceptor combinations*

1. INTRODUCTION

Organic solar cells (OSCs) have experienced significant advancements in recent years, attracting attention as lightweight, flexible, transparent, and cost-effective alternatives to conventional solar cell technology.[1−5] A power conversion efficiency (PCE) of 18−19%[6−11] has already been achieved for bulk heterojunction-based binary or ternary OSCs through the utilization of non-fullerene small molecule acceptors, specifically, Y-series molecules, and the PCE of tandem OSCs exceeds 20%.[12]

The integration of ML models into the field of OSCs has opened up new avenues for exploring the vast chemical space with greater efficiency and cost-effectiveness.[13−15] With the rapid advancement of ML techniques, scientists are now empowered to delve deeper into the properties of chemical systems, taking advantage of the abundance of data, improved algorithms, and exponential increases in computational power.[16] In the field of OSCs, various ML models have been used for predicting key performance metrics such as PCE,[17−33] short circuit current density (J_{SC}), open-circuit voltage (V_{OC}),[21,23,26,34] fill factor,[23,30] non-radiative voltage loss (ΔV_{NR}),[35] and frontier molecular orbitals (FMO).[26,30] High-throughput screening has also become a valuable tool for identifying promising OSC candidates. The rapid progress of machine learning algorithms and the continuous advancement of computational power are helping researchers with materials design, discovery, and optimization. This cutting-edge approach has the potential to revolutionize the material discovery process and bring us one step closer to unlocking the full potential of ML in the field of OSCs.

In the field of OSCs, a variety of microscopic property inputs have been explored as part of ML investigations. Such properties, including charge carrier mobility, optical bandgap, and electron−hole binding energy, offer a more comprehensive and accurate assessment of organic materials.[17,23] Although they provide more realistic measures for organic materials, they are computationally expensive. Traditional fixed representations like molecular descriptors and fingerprints have been used for traditional molecular property prediction.[21] These fixed descriptors and fingerprints are calculated using open-source libraries such as RDKit[36] and Mordred[37] using

Received: June 5, 2023
Accepted: July 12, 2023

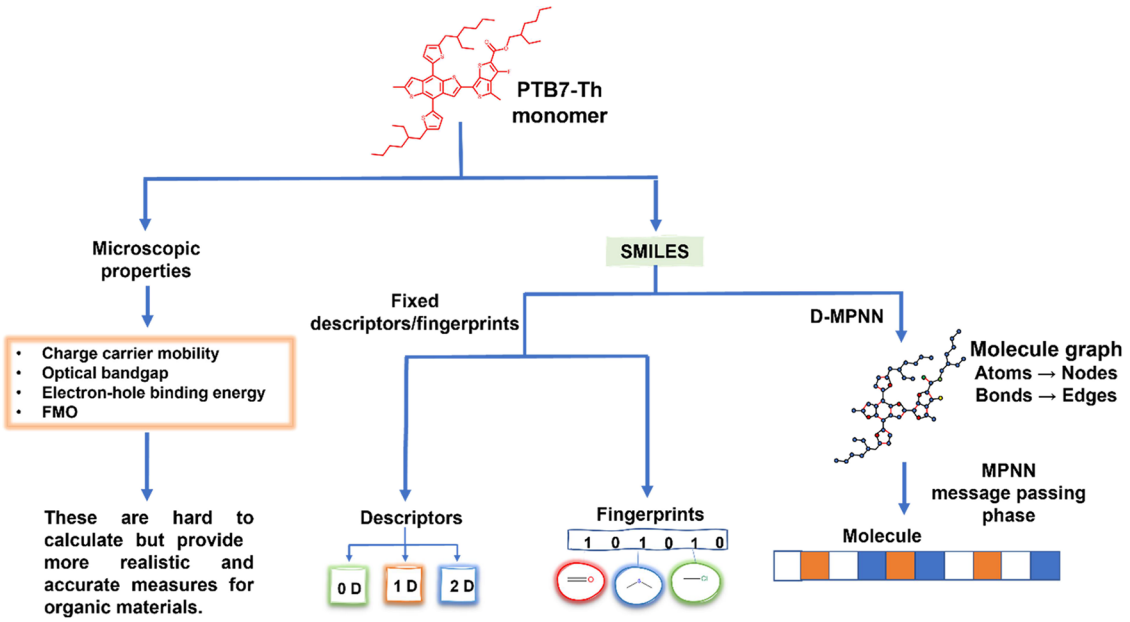


Figure 1. Input data preparation for ML studies in organic solar cells.

simplified molecular input line entry system (SMILES) strings[38] and a machine-readable ASCII string. A molecular descriptor is a machine-readable representation of the information contained within a molecule, which can be classified into zero-dimensional, one-dimensional, two-dimensional, and three-dimensional descriptors based on the level of inferences about the molecular topology and chemistry. On the other hand, a molecular fingerprint is a binary representation of a molecule's known structural property, which can also be used to compare the structural similarity of two molecules by counting the number of matching bits.

The use of ML models for OSCs can lead to significant advancements in the material discovery process. By optimizing materials properties, researchers can improve the performance of OSCs and potentially unlock their full potential. In addition, ML models can aid in the development of new OSC materials, which can further reduce the cost of solar energy generation. With the integration of ML techniques, the process of designing and testing new materials can become more efficient and cost-effective.

2. EXPERIMENTAL SECTION

2.1. Message Passing Neural Network. Recently, the potential of graph neural networks (GNNs) in modeling the graph structures of molecules has gained considerable interest.[39] The method offers a promising solution to the restrictions of fixed descriptors, as it can learn task-specific representations using graph convolutions. GNNs have become a significant class of models for quantitative structure−property relationship research due to their ability to model graph structures. Among the various GNN architectures under investigation, message passing neural networks (MPNNs) represent a commonly used framework. MPNNs were originally introduced by Gilmer et al.[40] as a generalization for the main GNNs.

In MPNN, molecular representation is built from scratch to better fit the dataset. The basic idea behind MPNNs is to encode the atomic and bonding information for a given chemical structure as a graph where the nodes in the network represent atoms in a molecule, and the edges represent chemical bonds between atoms.[39] MPNN begins by featuring the atoms of each molecule and atoms essentially pass messages to each other in order to update each other with relevant information.[41] The MPNN algorithm is a two-part process, with "message passing phase" and the "readout phase", respectively. During the message passing phase, the attributes of the atoms and bonds are circulated a specified number of times to create a task-specific vector of the molecule. This representation is then utilized in the readout phase, where the molecular properties are predicted with accuracy. The "depth" parameter represents the extent of the reach of each node and how far it can "observe". One of the key advantages of MPNNs is that they are able to effectively capture both local and long-range interactions in a chemical structure. With this ability, MPNN can learn complex patterns within the active material chemical structure that contributes to higher PCE. MPNN accurately models the complex dependencies between different molecule parts, which is important for many chemical prediction tasks. In addition, MPNNs are able to handle molecules of arbitrary size and complexity, making them a powerful tool for working with large and diverse chemical datasets. Moreover, MPNN learns to extract features and make predictions directly from the raw input (SMILES) without manual feature engineering, thus saving time and effort in the modeling process.

2.2. Directed Message Passing Neural Network. In this study, we have used the directed MPNN (D-MPNN) to predict the efficiency of OSCs using the open-source Chemprop python package.[42] To the best of our knowledge, this is the first work to apply message passing on molecule graphs for predicting PCE in OSCs. In D-MPNN, the message is passed along edges (bonds) rather than nodes (atoms). In traditional MPNN algorithms, messages are passed along atoms, which can lead to totters,[43] as the message passing process can get stuck in a cycle of exchanging information between atoms. D-MPNN overcomes this issue by passing messages along bonds instead of atoms, ensuring that the message passing process converges to a stable solution. This method ensures that the message passing process converges to a stable solution, leading to more accurate predictions. D-MPNN has shown to be effective in handling complex molecular structures and predicting various chemical properties. This work aims to investigate the performance of D-MPNN for predicting the efficiency of OSCs and compare it with other traditional random forest (RF), XGBoost, and ANN models with fixed descriptors and fingerprints. Figure 1 represents the input data preparation workflow for ML studies in OSCs.

2.3. Workflow of D-MPNN. Yang et al.[42] developed an open-source python package Chemprop for implementing D-MPNN models. The package provides a powerful and efficient solution for

https://doi.org/10.1021/acsami.3c08068
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

molecular property prediction tasks and has been widely used in drug discovery[41] and materials science.[44] The D-MPNN model represents a graph with directional edges. Each connection between atoms v and w is depicted by two directional bonds e_{vw} and e_{wv}. To convert SMILES strings to molecular graphs and compute atom features (x_v) and bond features (e_{vw}), Chemprop utilizes the open-source RDKit package. Lists of atom and bond features used by Chemprop are given in Tables 1 and 2.

Table 1. List of Atom Features

atom features	description	size
atom type	type of atom (e.g., C, N, O), by atomic number	100
# bonds	number of bonds the atom is involved in	6
formal charge	integer electronic charge assigned to atoms	5
chirality	unspecified, tetrahedral CW/CCW, or other	4
# Hs	number of bonded hydrogen atoms	5
hybridization	sp, sp^2, sp^3, sp^3d, or sp^3d^2	5
aromaticity	whether this atom is part of an aromatic system	1
atomic mass	mass of the atom divided by 100	1

Table 2. List of Bond Features

bond feature	description	size
bond type	single, double, triple, or aromatic	4
conjugated	whether the bond is conjugated	1
in ring	whether the bond is part of a ring	1
stereo	none, any, *E*/*Z* or *cis*/*trans*	6

Before the message passing process begins, we initialize the edge hidden states by h^o_{vw} using eq 1. This equation concatenates the atom and bond features and passes them through a learned matrix W_i using the rectified linear unit (RELU) activation function. This initializes the edge hidden states, which will be updated during the message passing process. In eq 1, τ represents the RELU activation function, W_i is a learned matrix, and cat(x_v,e_{vw}) represents the concatenation of atom features x_v and bond features e_{vw}.

$$h^o_{vw} = \tau(W_i \mathrm{cat}(x_v, e_{vw})) \quad (1)$$

The first step of message passing involves calculating the messages from atom v to atom w. This is done by summing all hidden states for the incoming bonds to v, excluding the one originating from w, as described by m^{t+1}_{vw} in eq 2. This step captures the information about the neighboring atoms of each atom and their relationships.

$$m^{t+1}_{vw} = \sum_{k\in\{N(v)\backslash w\}} h^t_{kv} \quad (2)$$

Next, a new hidden message for depth 1 is created by adding the initial hidden state and the product of the learned matrix W_m and the message, as specified by h^{t+1}_{vw} in eq 3. This step updates the hidden message with information about the neighboring atoms and their relationships and creates a new hidden message for the next depth of message passing.

$$h^{t+1}_{vw} = \tau(h^o_{vw} + W_m m^{t+1}_{vw}) \quad (3)$$

After messages have been passed for a specified depth, the hidden states are summed up to create a final message for each atom, as described by m_v in eq 4. This step aggregates the information from all the neighboring atoms and their relationships into a final message for each atom.

$$m_v = \sum_{w\in N(v)} h^T_{vw} \quad (4)$$

The hidden state for each atom (h_v) is then calculated by concatenation of initial atom features and message vector using eq 5.

$$h_v = \tau(W_a \mathrm{cat}(x_v, m_v)) \quad (5)$$

Finally, a molecule feature vector is generated by aggregating hidden states for each atom (h_v) with eq 6. This step summarizes the information from all the atoms in the molecule into a single molecule feature vector, which can be used for property prediction.

$$h = \sum_{v\in N(v)} h_v \quad (6)$$

In readout phase, a fully connected feed-forward neural network is then used to make property predictions. The D-MPNN model is trained using NVIDIA Tesla V100 GPU.

3. RESULTS AND DISCUSSION

3.1. Comparison of the Fixed Descriptor Model with the D-MPNN. In this study, we have compared the D-MPNN method with RF, XGBoost, and ANN models using fixed descriptors (RDKit[36] and Mordred[37]) and fixed fingerprints (Morgan[45]). The dataset of 1318 D:A (polymer donor: non-fullerene acceptor) combinations published by Miyake and Saeki[46] is used to compare the models, and the distribution of PCE is given in Figure 2. RF is a supervised learning algorithm,

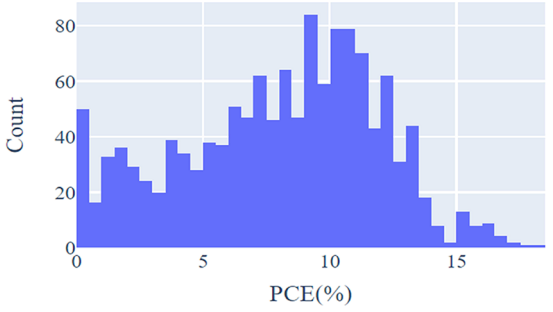


Figure 2. PCE distribution for 1318 unique D:A combinations.[46]

in which a collection of decision trees are generated from randomly collected subsets of rows and columns. This process is often referred to as "bagging" or "bootstrap aggregation". Workflow comparison for RF model with fixed descriptors and D-MPNN model is represented in Figure 3.

3.2. Hyperparameter Optimization and Ensembling. For the RF model, five hyperparameters (n_estimators, max_features, min_samples_split, min_samples_leaf, and bootstrap) are optimized using Gridsearchcv.[47] The performance of D-MPNNs, similar to many neural networks, is heavily influenced by the configurations of the model hyperparameters such as the hidden size of the network layers. To achieve optimal performance, Chemprop employs hyperparameter optimization using Bayesian Optimization[48] with hyperopt python package.[49] The optimization specifically focuses on the model's depth (number of message-passing steps in MPNN), hidden size (vector size used in MPNN), ffn_num_layer (the number of layers in the feed-forward neural network), and dropout (dropout probability). Details of best selected hyperparameters are provided in Tables 3 and 4. To further improve the performance of models, ensembling is used where predictions from multiple independently trained D-MPNN models are averaged to give more accurate prediction.[50] Multiple models used for ensembling have the same architecture and hyperparameters but are initialized with random weights.

3.3. Results. For a fair comparison with the D-MPNN, the RF model with optimized parameters is used for predicting PCE using fixed descriptors (RDKit descriptors, Mordred

https://doi.org/10.1021/acsami.3c08068
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

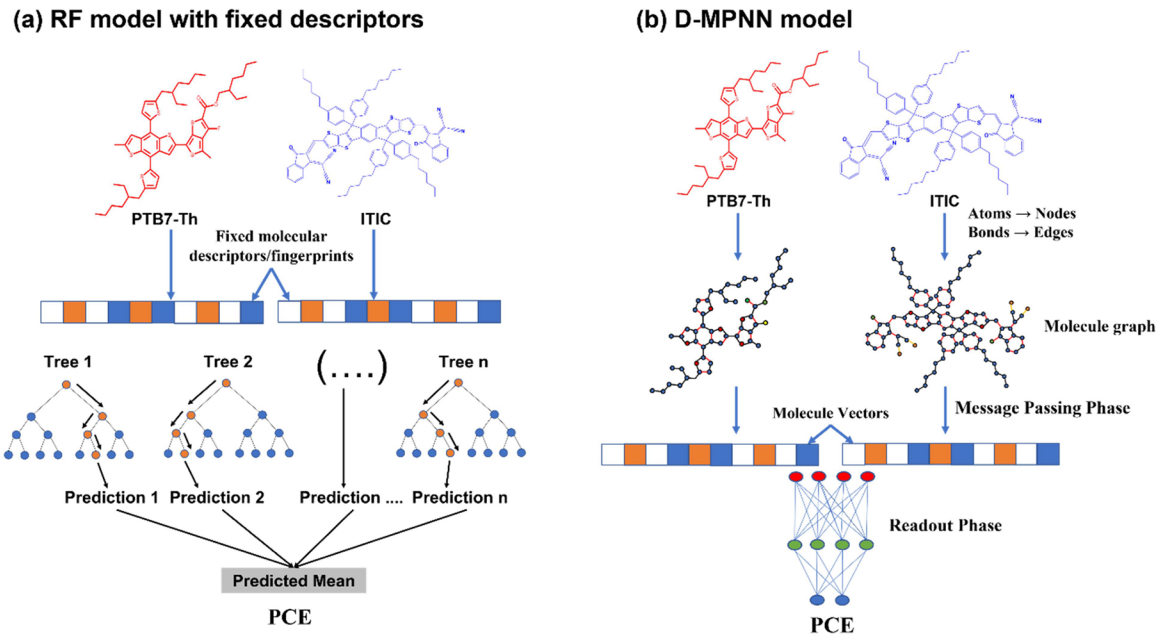


Figure 3. Workflow of models compared in this study. (a) RF model is used with fixed descriptors (RDKit and Mordred) and fixed fingerprint (Morgan). (b) D-MPNN model is used with Chemprop python package.

Table 3. RF Hyperparameters Optimized Using Gridsearchcv for Fixed Descriptors

hyperparameters	RDKit descriptors	Mordred descriptors	Morgan fingerprints
n_estimators	100	300	200
max_features	auto	sqrt	auto
min_sample split	2	10	5
min_sample leaf	2	1	1
bootstrap	true	false	true

descriptors, and Morgan fingerprints). The RF model with RDKit descriptors performs best with a correlation coefficient (r = 0.84), followed by the Morgan fingerprint (r = 0.83) and then followed by Mordred descriptor (r = 0.82). In order to conduct a comprehensive comparison, we also evaluated fixed descriptor sets using XGBoost and ANN models. The XGBoost model demonstrated favorable performance when applied to RDKit descriptors (correlation coefficient, r = 0.83) and Mordred descriptors (r = 0.82). However, its performance decreased when applied to Morgan fingerprint descriptors (r = 0.77). Likewise, the ANN model yielded promising results with RDKit descriptors (r = 0.81) and Mordred descriptors (r = 0.84), but its performance declined when using Morgan fingerprint descriptors (r = 0.79). For fixed descriptors set, the RF model performs best.

An 80:10:10 train:validation:test split is used for the D-MPNN model, and training set and test set are kept the same for the RF model. The D-MPNN model is trained with default hyperparameters, and (r = 0.822) is achieved. By performing cross-validation along with ensembling, r is improved to 0.84. The model gave the best performance with optimized hyperparameters (r = 0.86). An optimized depth of 5 can be attributed to superior performance of this model. We also tried to include the RDKit descriptors in the D-MPNN model by including RDKit descriptors with the molecule vector generated in the message passing phase, and the model performance remains the same (r = 0.86). This indicates that molecule vectors generated in the message passing phase of the D-MPNN model contains all the information, and no extra information is added up by RDKit descriptors. Scatter plots for reported vs predicted PCE are given in Figure 4a–f along with RMSE of all the models in Figure 4g. This comparison highlights the potential benefits of using D-MPNN models for predicting PCE of OSCs, as well as the importance of optimizing hyperparameters to achieve the best possible performance.

4. CONCLUSIONS

The D-MPNN model is used for the prediction of PCE in OSCs, and the results are compared with fixed descriptor/fingerprint-based RF, XGBoost, and ANN models. In summary, the ability of our model to surpass established baselines highlights its superiority and paves the way for its widespread adoption in research. This study highlights the

Table 4. D-MPNN Hyperparameters Optimized Using Hyperopt Python Package

hyperparameters	D-MPNN (default)	D-MPNN (optimized)	D-MPNN (optimized) + RDKit descriptors
depth	3	5	5
dropout	0	0.05	0.05
ffn_num_layer	2	2	2
hidden_size	300	2200	2200
num_folds	1	5	5
ensemble_size	1	3	3

https://doi.org/10.1021/acsami.3c08068
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

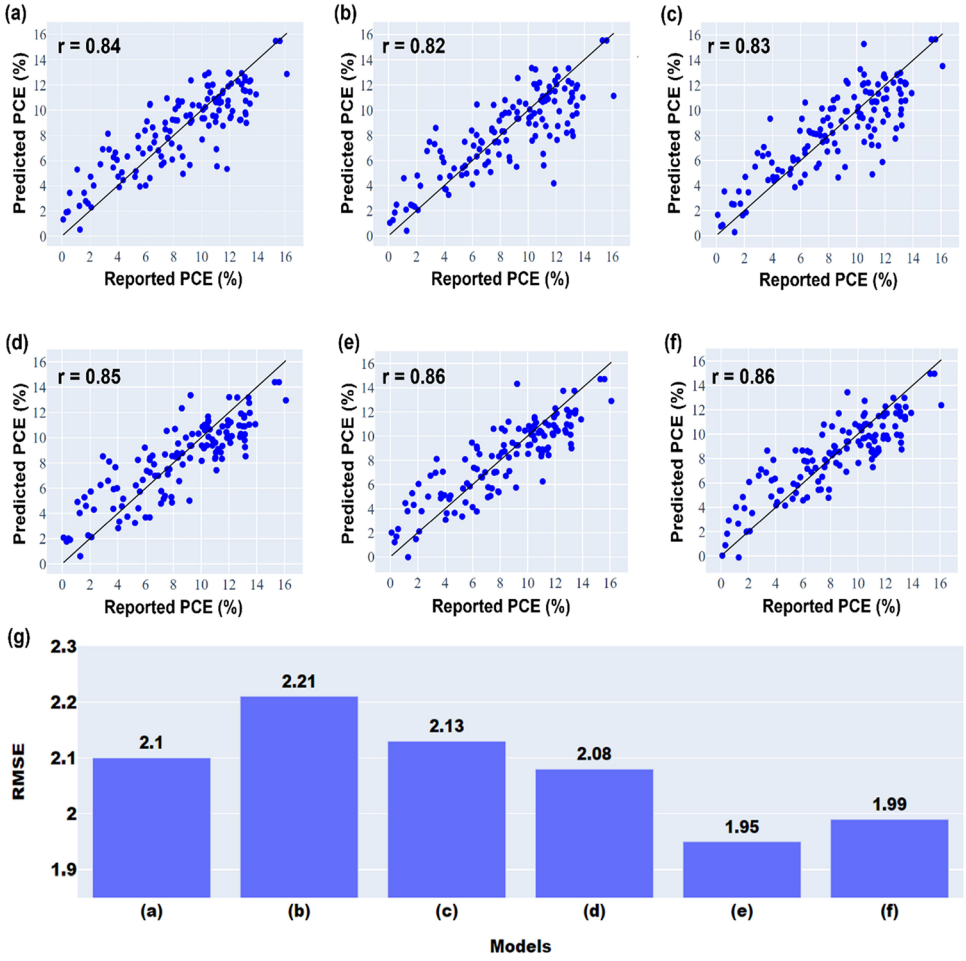


Figure 4. Results of models. (a) RF model with RDKit descriptors, (b) RF model with Mordred descriptors, (c) RF model with Morgan fingerprints, (d) D-MPNN model with default hyperparameters, (e) D-MPNN model with optimized hyperparameters, (f) D-MPNN model with optimized parameters and RDKit descriptors, and (g) RMSE of models (a–f).

promise of GNNs, particularly D-MPNN, in PCE prediction of OSCs, and may lead to further developments in this area. The use of learned representations opens up new avenues for exploring the complex relationship between molecular structures and properties and has the potential to revolutionize the way property predictions are made. Hyperparameter optimization is performed for all the models to find the optimal hyperparameters. In fixed descriptor/fingerprint-based models, RDKit descriptors performed best (r = 0.84), followed by Morgan fingerprints (r = 0.83) and Mordred descriptors (r = 0.82). For the D-MPNN model, with default hyperparameters, (r = 0.85) is achieved, which is further improved to (r = 0.86) by performing hyperparameter optimization. With its impressive performance and clear practical applications, our model is poised to play a key role in the future of molecular property prediction. This advancement in machine learning-based molecular representations is a significant step forward in our understanding of molecular structures and properties.

AUTHOR INFORMATION

Corresponding Author

Ganesh D. Sharma − *Department of Physics, The LNM Institute of Information Technology, Jaipur, Rajasthan 302031, India*; orcid.org/0000-0002-1717-0116; Email: gdsharma273@gmail.com, gdsharma@lnmiit.ac.in

Authors

Prateek Malhotra − *Department of Physics, The LNM Institute of Information Technology, Jaipur, Rajasthan 302031, India*

Subhayan Biswas − *Department of Physics, The LNM Institute of Information Technology, Jaipur, Rajasthan 302031, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsami.3c08068

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Almora, O.; Baran, D.; Bazan, G. C.; Berger, C.; Cabrera, C. I.; Catchpole, K. R.; Erten-Ela, S.; Guo, F.; Hauch, J.; Ho-Baillie, A. W. Y.; Jacobsson, T. J.; Janssen, R. A. J.; Kirchartz, T.; Kopidakis, N.; Li, Y.; Loi, M. A.; Lunt, R. R.; Mathew, X.; McGehee, M. D.; Min, J.; Mitzi, D. B.; Nazeeruddin, M. K.; Nelson, J.; Nogueira, A. F.; Paetzold, U. W.; Park, N.; Rand, B. P.; Rau, U.; Snaith, H. J.; Unger, E.; Vaillant-Roca, L.; Yip, H.; Brabec, C. J. Device Performance of Emerging Photovoltaic Materials (Version 2). *Adv. Energy Mater.* **2021**, *11*, No. 2102526.

(2) Karki, A.; Gillett, A. J.; Friend, R. H.; Nguyen, T. The Path to 20% Power Conversion Efficiencies in Nonfullerene Acceptor Organic Solar Cells. *Adv. Energy Mater.* **2021**, *11*, No. 2003441.

https://doi.org/10.1021/acsami.3c08068
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

(3) Hong, L.; Yao, H.; Cui, Y.; Ge, Z.; Hou, J. Recent Advances in High-Efficiency Organic Solar Cells Fabricated by Eco-Compatible Solvents at Relatively Large-Area Scale. *APL Mater.* **2020**, *8*, 120901.
(4) Wang, D.; Qin, R.; Zhou, G.; Li, X.; Xia, R.; Li, Y.; Zhan, L.; Zhu, H.; Lu, X.; Yip, H.; Chen, H.; Li, C. High-Performance Semitransparent Organic Solar Cells with Excellent Infrared Reflection and See-Through Functions. *Adv. Mater.* **2020**, *32*, No. 2001621.
(5) Khandelwal, K.; Biswas, S.; Mishra, A.; Sharma, G. D. Semitransparent Organic Solar Cells: From Molecular Design to Structure−Performance Relationships. *J. Mater. Chem. C* **2021**, *10*, 13−43.
(6) Cui, Y.; Xu, Y.; Yao, H.; Bi, P.; Hong, L.; Zhang, J.; Zu, Y.; Zhang, T.; Qin, J.; Ren, J.; Chen, Z.; He, C.; Hao, X.; Wei, Z.; Hou, J. Single-Junction Organic Photovoltaic Cell with 19% Efficiency. *Adv. Mater.* **2021**, *33*, 1−8.
(7) Zhu, L.; Zhang, M.; Xu, J.; Li, C.; Yan, J.; Zhou, G.; Zhong, W.; Hao, T.; Song, J.; Xue, X.; Zhou, Z.; Zeng, R.; Zhu, H.; Chen, C. C.; MacKenzie, R. C. I.; Zou, Y.; Nelson, J.; Zhang, Y.; Sun, Y.; Liu, F. Single-Junction Organic Solar Cells with over 19% Efficiency Enabled by a Refined Double-Fibril Network Morphology. *Nat. Mater.* **2022**, *21*, 656−663.
(8) Li, C.; Zhou, J.; Song, J.; Xu, J.; Zhang, H.; Zhang, X.; Guo, J.; Zhu, L.; Wei, D.; Han, G.; Min, J.; Zhang, Y.; Xie, Z.; Yi, Y.; Yan, H.; Gao, F.; Liu, F.; Sun, Y. Non-Fullerene Acceptors with Branched Side Chains and Improved Molecular Packing to Exceed 18% Efficiency in Organic Solar Cells. *Nat. Energy* **2021**, *6*, 605−613.
(9) Wei, Y.; Chen, Z.; Lu, G.; Yu, N.; Li, C.; Gao, J.; Gu, X.; Hao, X.; Lu, G.; Tang, Z.; Zhang, J.; Wei, Z.; Zhang, X.; Huang, H. Binary Organic Solar Cells Breaking 19% via Manipulating Vertical Component Distribution. *Adv. Mater.* **2022**, No. 2204718.
(10) He, C.; Pan, Y.; Ouyang, Y.; Shen, Q.; Gao, Y.; Yan, K.; Fang, J.; Chen, Y.; Ma, C.-Q.; Min, J.; Zhang, C.; Zuo, L.; Chen, H. Manipulating the D:A Interfacial Energetics and Intermolecular Packing for 19.2% Efficiency Organic Photovoltaics. *Energy Environ. Sci.* **2022**, *15*, 2537−2544.
(11) Zhan, L.; Li, S.; Li, Y.; Sun, R.; Min, J.; Bi, Z.; Ma, W.; Chen, Z.; Zhou, G.; Zhu, H.; Shi, M.; Zuo, L.; Chen, H. Desired Open-Circuit Voltage Increase Enables Efficiencies Approaching 19% in Symmetric-Asymmetric Molecule Ternary Organic Photovoltaics. *Joule* **2022**, *6*, 662−675.
(12) Zheng, Z.; Wang, J.; Bi, P.; Ren, J.; Wang, Y.; Yang, Y.; Liu, X.; Zhang, S.; Hou, J. Tandem Organic Solar Cell with 20.2% Efficiency. *Joule* **2022**, *6*, 171−184.
(13) 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.
(14) Rodríguez-Martínez, X.; Pascual-San-José, E.; Campoy-Quiles, M. Accelerating Organic Solar Cell Material's Discovery: High-Throughput Screening and Big Data. *Energy Environ. Sci.* **2021**, *14*, 3301−3322.
(15) Malhotra, P.; Khandelwal, K.; Biswas, S.; Chen, F.-C.; Sharma, G. D. Opportunities and Challenges for Machine Learning to Select Combination of Donor and Acceptor Materials for Efficient Organic Solar Cells. *J. Mater. Chem. C* **2022**, *10*, 17781−17811.
(16) Malhotra, P.; Verduzco, J. C.; Biswas, S.; Sharma, G. D. Active Discovery of Donor:Acceptor Combinations For Efficient Organic Solar Cells. *ACS Appl. Mater. Interfaces* **2022**, *14*, 54895−54906.
(17) 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*, No. 1801032.
(18) 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*, No. eaay4275.
(19) Chen, F.-C. Virtual Screening of Conjugated Polymers for Organic Photovoltaic Devices Using Support Vector Machines and Ensemble Learning. *Int. J. Polym. Sci.* **2019**, *2019*, No. 4538514.
(20) Padula, D.; Troisi, A. Concurrent Optimization of Organic Donor−Acceptor Pairs through Machine Learning. *Adv. Energy Mater.* **2019**, *9*, No. 1902463.
(21) 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.
(22) Lee, M. Insights from Machine Learning Techniques for Predicting the Efficiency of Fullerene Derivatives-Based Ternary Organic Solar Cells at Ternary Blend Design. *Adv. Energy Mater.* **2019**, *9*, No. 1900891.
(23) 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.
(24) 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.
(25) Lee, M.-H. Performance and Matching Band Structure Analysis of Tandem Organic Solar Cells Using Machine Learning Approaches. *Energy Technol.* **2020**, *8*, No. 1900974.
(26) Meftahi, N.; Klymenko, M.; Christofferson, A. J.; Bach, U.; Winkler, D. A.; Russo, S. P. Machine Learning Property Prediction for Organic Photovoltaic Devices. *npj Comput. Mater.* **2020**, *6*, 166.
(27) Lee, M. H. Robust Random Forest Based Non-Fullerene Organic Solar Cells Efficiency Prediction. *Org. Electron.* **2020**, *76*, No. 105465.
(28) Zhao, Z.; del Cueto, M.; 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.
(29) Kranthiraja, K.; Saeki, A. Experiment-Oriented Machine Learning of Polymer:Non-Fullerene Organic Solar Cells. *Adv. Funct. Mater.* **2021**, *31*, No. 2011168.
(30) Munshi, J.; Chen, W.; Chien, T.; Balasubramanian, G. Transfer Learned Designer Polymers for Organic Solar Cells. *J. Chem. Inf. Model.* **2021**, *61*, 134−142.
(31) Zhang, Q.; Zheng, Y. J.; Sun, W.; Ou, Z.; Odunmbaku, O.; Li, M.; Chen, S.; Zhou, Y.; Li, J.; Qin, B.; Sun, K. High-Efficiency Non-Fullerene Acceptors Developed by Machine Learning and Quantum Chemistry. *Adv. Sci.* **2022**, *9*, No. 2104742.
(32) 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.
(33) Hao, T.; Leng, S.; Yang, Y.; Zhong, W.; Zhang, M.; Zhu, L.; Song, J.; Xu, J.; Zhou, G.; Zou, Y.; Zhang, Y.; Liu, F. Capture the High-Efficiency Non-Fullerene Ternary Organic Solar Cells Formula by Machine-Learning-Assisted Energy-Level Alignment Optimization. *Patterns* **2021**, *2*, No. 100333.
(34) Lee, M.-H. A Machine Learning−Based Design Rule for Improved Open-Circuit Voltage in Ternary Organic Solar Cells. *Adv. Intell. Syst.* **2020**, *2*, No. 1900108.
(35) Malhotra, P.; Biswas, S.; Chen, F.-C.; Sharma, G. D. Prediction of Non-Radiative Voltage Losses in Organic Solar Cells Using Machine Learning. *Sol. Energy* **2021**, *228*, 175−186.
(36) Landrum, G. *RDKit: Open-Source Cheminformatics*; 2016.
(37) Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: A Molecular Descriptor Calculator. *J. Cheminform.* **2018**, *10*, 4.
(38) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.* **1988**, *28*, 31−36.
(39) Aldeghi, M.; Coley, C. W. A Graph Representation of Molecular Ensembles for Polymer Property Prediction. *Chem. Sci.* **2022**, *13*, 10486−10498.
(40) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. *Mach. Learn.* **2017**, *3*, 2053−2070.

https://doi.org/10.1021/acsami.3c08068
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

(41) Jiang, D.; Wu, Z.; Hsieh, C. Y.; Chen, G.; Liao, B.; Wang, Z.; Shen, C.; Cao, D.; Wu, J.; Hou, T. Could Graph Neural Networks Learn Better Molecular Representation for Drug Discovery? A Comparison Study of Descriptor-Based and Graph-Based Models. *Aust. J. Chem.* **2021**, *13*, 1−23.
(42) Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M.; Palmer, A.; Settels, V.; Jaakkola, T.; Jensen, K.; Barzilay, R. Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.* **2019**, *59*, 3370−3388.
(43) Mahé, P.; Ueda, N.; Akutsu, T.; Perret, J.-L.; Vert, J.-P. *Extensions of Marginalized Graph Kernels*. In *Twenty-first International Conference on Machine Learning - ICML '04*; ACM Press: New York, New York, USA, 2004; p 70.
(44) Greenman, K. P.; Green, W. H.; Gómez-Bombarelli, R. Multi-Fidelity Prediction of Molecular Optical Peaks with Deep Learning. *Chem. Sci.* **2022**, *13*, 1152−1162.
(45) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754.
(46) Miyake, Y.; Saeki, A. Machine Learning-Assisted Development of Organic Solar Cell Materials: Issues, Analyses, and Outlooks. *J. Phys. Chem. Lett.* **2021**, *12*, 12391−12401.
(47) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G.; 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.* **2012**, *12*, 2825−2830.
(48) Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R. P.; de Freitas, N. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* **2016**, *104*, 148−175.
(49) Bergstra, J.; Komer, B.; Eliasmith, C.; Yamins, D.; Cox, D. D. Hyperopt: A Python Library for Model Selection and Hyperparameter Optimization. *Computational Science & Discovery* **2015**, *8*, No. 014008.
(50) Dietterich, T. G. Ensemble Methods in Machine Learning. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin, Heidelberg, 2000; Vol. *1857*; pp 1−15.

https://doi.org/10.1021/acsami.3c08068
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX