

Prediction of non-radiative voltage losses in organic solar cells using machine learning

Prateek Malhotra^a, Subhayan Biswas^a, Fang-Chung Chen^{b,c}, Ganesh D. Sharma^{a,*}

^a Department of Physics, The LNM Institute of Information Technology, Jamdoli, Jaipur 302031, Rajasthan, India

^b Department of Photonics, National Chiao Tung University, Hsinchu, Taiwan

^c Center for Emergent Functional Matter Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

ARTICLE INFO

Keywords:

Organic Solar Cells
Machine learning
Non-radiative voltage loss

ABSTRACT

One of the major hurdles that are preventing organic solar cells (OSCs) from leading the efficiency chart is non-radiative voltage loss (ΔV_{NR}). So far, however, not much effort is made to predict voltage losses and unravel the correlation of losses with electronic and structural descriptors. From the literature, we create a dataset consisting of 154 unique donor:acceptor combinations with reported ΔV_{NR} . The dataset includes information about frontier molecular orbitals (FMO), optical bandgap (E_g), molecular descriptors, and molecular fingerprints. Four machine learning (ML) algorithms (random forest regressor, gradient boosting regressor, support vector regressor, and artificial neural network) are used to predict non-radiative voltage loss and the results obtained are compared on the basis of Pearson r s, root mean squared errors, and mean absolute percentage errors. Best results are obtained with gradient boosting regressor by using FMO + E_g + RDKit descriptors (Pearson r = 0.859) and FMO + E_g + MACCS fingerprints (Pearson r = 0.857). We have also applied these ML models by using only molecular descriptors and only molecular fingerprints and got impressive results (Pearson r = 0.78 and 0.726). These results indicate that ML models can be effectively used for the prediction of ΔV_{NR} and virtual screening of promising donor:acceptor combinations with reduced ΔV_{NR} .

1. Introduction

Organic solar cells (OSCs) have shown remarkable progress in the last decade and have emerged as a low-cost alternative to conventional solar technologies (Almora et al., 2021; Cui et al., 2020; Karki et al., 2021; Sun et al., 2019b). In recent years with the emergence of non-fullerene small molecule acceptor, particularly Y-series small molecules (Li et al., 2020; Yuan et al., 2019b), the highest power conversion efficiency (PCE) achieved by single-junction OSCs prepared with polymer donor:non-fullerene acceptor (NFA) is more than 18% (Cui et al., 2020; Liu et al., 2020a; Qin et al., 2021). It is reported that the PCE of OSCs can be increased upto 20% by the selection of appropriate donors and acceptors with complementary absorption and matched frontier molecular orbitals (FMOs) (Karki et al., 2021; Upama et al., 2020). Literatures have estimated the realistic PCE limit for OSC as a function of optical gap and have shown that the upper limit for realistic PCE is approximately 20% (Azzouzi et al., 2018; Benduhn et al., 2017), considering a device with an external quantum efficiency (EQE) of 90% and a fill factor (FF) of 80%. Comparing with other technologies that are leading the efficiency

chart, such as perovskites solar cells, OSCs have competent short circuit current density (J_{SC}) and FF but lags in the case of open-circuit voltage (V_{OC}). The lag in V_{OC} is due to higher non-radiative voltage loss (ΔV_{NR}), leading to lower PCEs of OSCs. To achieve the milestone of crossing 20% PCE, ΔV_{NR} needs to be reduced (Azzouzi et al., 2018; Vandewal et al., 2020).

It is worth looking into the factors that govern ΔV_{NR} and explore how the structural variations (Vandewal et al., 2018) of active materials employed in OSCs influence ΔV_{NR} . The energy materials that we are using today in OSCs are not the most optimal ones, and there exists a large room for improvement. To further enhance the PCE of OSCs, we need to discover new active materials for suitable donor:acceptor (D:A) combination with reduced ΔV_{NR} (Vandewal et al., 2020). Moreover, for an efficient OSC, D:A combination should have appropriately aligned energy levels, complementary absorption spectra, and balanced charge carrier mobility (Pradhan et al., 2020). Today, exploring new materials with less ΔV_{NR} is not cost-effective because new materials need to be synthesized, followed by cell fabrication for getting EQE or electroluminescence (EL) spectra. A more straightforward approach for

* Corresponding author.

E-mail address: gdsharma@lnmiit.ac.in (G.D. Sharma).

<https://doi.org/10.1016/j.solener.2021.09.056>

Received 18 May 2021; Received in revised form 15 September 2021; Accepted 20 September 2021

Available online 25 September 2021

0038-092X/© 2021 International Solar Energy Society. Published by Elsevier Ltd. All rights reserved.

narrowing down the search for potential D:A combination is to use machine learning (ML) algorithms by using descriptors available in the literatures, such as FMOs, optical bandgap (E_g), molecular descriptors, and molecular fingerprints.

Several efforts have been made by utilizing ML models on OSCs to explore new active materials and understand the hidden information within chemical structures. Most of the ML-based research in OSCs is employed for the prediction of PCEs (Kranthiraja and Saeki, 2021; Lee, 2019, 2020b,c; Lopez et al., 2017; Nagasawa et al., 2018; Olivares-Amaya et al., 2011; Padula et al., 2019; Padula and Troisi, 2019; Peng and Zhao, 2019; Pyzer-Knapp et al., 2016; Sahu et al., 2018, 2019; Sahu and Ma, 2019; Sui et al., 2019; Sun et al., 2019c; Wu et al., 2020; Zhao et al., 2020), FMOs (Mesta et al., 2019; Paul et al., 2019; Pereira et al., 2017), J_{SC} (Pokuri et al., 2019; Rodríguez-Martínez et al., 2021), and V_{OC} (Lee, 2020a). Descriptors for each set of problems are chosen based on the domain knowledge. Till now, ML is not utilized for the prediction of ΔV_{NR} as no suitable dataset is present with reported ΔV_{NR} values. As ΔV_{NR} is considered one of the major factors influencing V_{OC} of the solar cells (Babics et al., 2019; Liu et al., 2020b), ML approaches can help predict ΔV_{NR} and virtual screening of suitable D:A combinations with reduced ΔV_{NR} .

We are very much interested in developing ML models for predicting ΔV_{NR} by using electronic descriptors (FMO and E_g), structural descriptors (molecular descriptors and molecular fingerprints), and a combination of both. Throughout this study, we use only descriptors that are easily accessible. In this work, we apply supervised ML algorithms, including random forest (RF), gradient boosting (GB), support vector regression (SVR), and artificial neural network (ANN). The results imply gradient boosting regressor generated more accurate predictions. In order to investigate the robustness of each model, we add 50 manually calculated ΔV_{NR} (by digitizing the reported EQE curves) in the base dataset and again obtain impressive results. We believe that the ML algorithms reported herein can benefit and speed up the development of high-performance organic materials for solar applications.

2. Non-Radiative voltage loss (ΔV_{NR})

The difference between the photon's energy and product of elementary charge (q) with maximum power point voltage is defined as the total energy loss per absorbed photon (Vandewal et al., 2018). Herein, we define the lower limit of this energy loss as $\Delta E_{loss} = E_g - qV_{OC}$ (Nikolis et al., 2017), where E_g is the optical bandgap of the donor or acceptor material used in the active bulk heterojunction (BHJ) layer whichever is lower, q is the elementary charge and V_{OC} is the voltage at open-circuit. Fig. 1 represents voltage losses in OSCs. ΔE_{loss} can be further divided into losses due to charge transfer ($\Delta E_{CT} = E_g - E_{CT}$), radiative recombination losses ($\Delta V_R = E_{CT}/q - V_R$), and non-radiative recombination losses ($\Delta V_{NR} = V_R - V_{OC}$). Out of these three losses, ΔV_{NR} is the most dominating one and is considered to be the critical factor in determining the PCE of OSCs (Azzouzi et al., 2018; Eisner et al., 2019; Riede et al., 2021). Now ΔV_{NR} is one of the hot research topics and several efforts have been made in understanding the cause of ΔV_{NR} with many theory-based models and experimental works. NFA with acceptor-donor-acceptor (A-D-A) structures shows excellent performance in photovoltaic applications. A-D-A structure neutralizes all the disadvantages that fullerene pursues (Sun et al., 2019b; Ye et al., 2020), thereby reducing ΔV_{NR} . In order to cross the 20% PCE mark, ΔV_{NR} in NFA based OSCs needs to be further reduced to 0.15 eV, while maintaining high EQE and FF (Karki et al., 2021; Liu et al., 2020b; Vandewal et al., 2020). The energy of the charge transfer state (E_{CT}) plays a critical role in understanding ΔV_{NR} as they are the charge separation and recombination centers at the D:A interface (Chen and Brédas, 2018; Veldman et al., 2009; Xie et al., 2018). E_{CT} can be approximated by interfacial bandgap (E_i) which is described as the gap between HOMO of Donor and LUMO of acceptor (Azzouzi et al., 2019). To use only easily accessible descriptors, we have used E_i in place of E_{CT} and ($E_g - E_i$) in

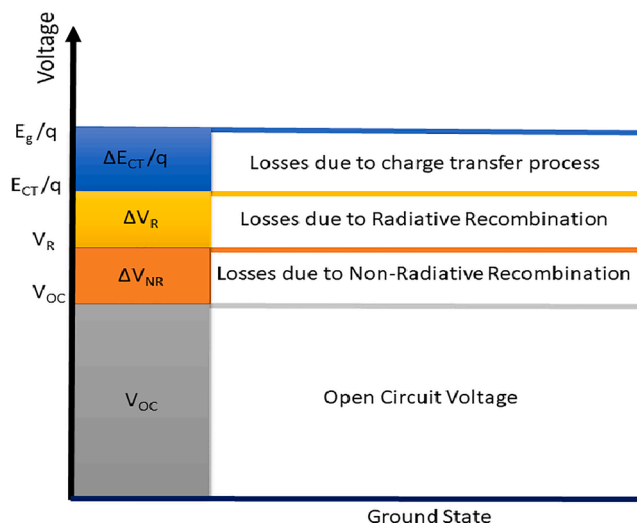


Fig. 1. Voltage losses in OSCs. E_g , E_{CT} , V_R , V_{OC} and q are the optical bandgap of the donor or acceptor material used in the active bulk heterojunction (BHJ) layer whichever is lower, energy of charge transfer state, open circuit voltage in the radiative limit, voltage at open-circuit, and the elementary charge. The loss occurred in the charge transfer process is represented by $\Delta E_{CT}/q$, and the rest are recombination losses. Recombination losses are further divided into losses due to radiative recombination (ΔV_R) and non-radiative recombination (ΔV_{NR}).

place of ΔE_{CT} . It is important to note that instead of using ΔV_{NR} , we have defined percentage non-radiative voltage loss ($\% \Delta V_{NR}$) as our target variable.

$$\% \Delta V_{NR} = (\Delta V_{NR} / (E_g / q)) * 100 \quad (1)$$

While exploring our manually collected dataset, a good trend is observed between E_i and $\% \Delta V_{NR}$, but no correlation is observed between E_i and PCE (Fig. 2a and 2b). Meanwhile, ($E_g - E_i$) shows a good trend with both $\% \Delta V_{NR}$ and PCE (Fig. 2c and 2d). In Fig. 2a, it is clearly observed that with increasing E_{CT} , $\% \Delta V_{NR}$ gets reduced, and V_{OC} becomes more closer to the radiative limit V_R . Moreover in Fig. 2c, with reduced $E_g - E_i$, $\% \Delta V_{NR}$ gets reduced, which can be attributed to increased electroluminescence spectra due to hybrid local exciton charge transfer (LE-CT) states (Classen et al., 2020; Eisner et al., 2019; Vandewal et al., 2020; Zhang et al., 2020). All the trends observed are consistent with those reported in the literature (Benduhn et al., 2017). In Fig. 2d, it is interesting to see that all the devices with PCE greater than 15% have greater E_i than E_g .

ΔV_{NR} can be reduced by making chemical modifications in photo-active materials (Hong et al., 2019; Luo et al., 2020; Sun et al., 2019a; Ye et al., 2020; Yuan et al., 2019a). Similar work is done by Cui et al., and their findings reveal that chlorinated NFA provides higher open-circuit voltage compared to its fluorinated counterpart (Cui et al., 2019). Although chlorinated NFA shows downshifted LUMO levels compared to fluorinated NFA, a higher voltage in chlorinated NFA corresponds to low ΔV_{NR} .

Due to low dielectric constant of organic materials, their binding energy is relatively high and a reasonable ΔE_{offset} (driving force) is required for efficient charge separation (Nakano et al., 2019; Xie and Wu, 2020). Fu et al. used two (PDI)-based small molecule acceptors and four polymer donors to study the effect of ΔE_{offset} on energy loss (Fu et al., 2018). Findings reveal that both radiative loss (due to CT absorption) and ΔV_{NR} gets reduced with decreasing ΔE_{offset} .

Liu et al. observed a low ΔV_{NR} of 0.2 V and EQE_{EL} of 2.1×10^{-4} with BDT-fbBX-DT:SFPDI solar cell (Liu et al., 2018). Such a high EQE_{EL} is explained by either absence of charge transfer state or difference between singlet state and charge transfer state being negligible.

Classen et al. studied many OSC blends and analysed the effect of

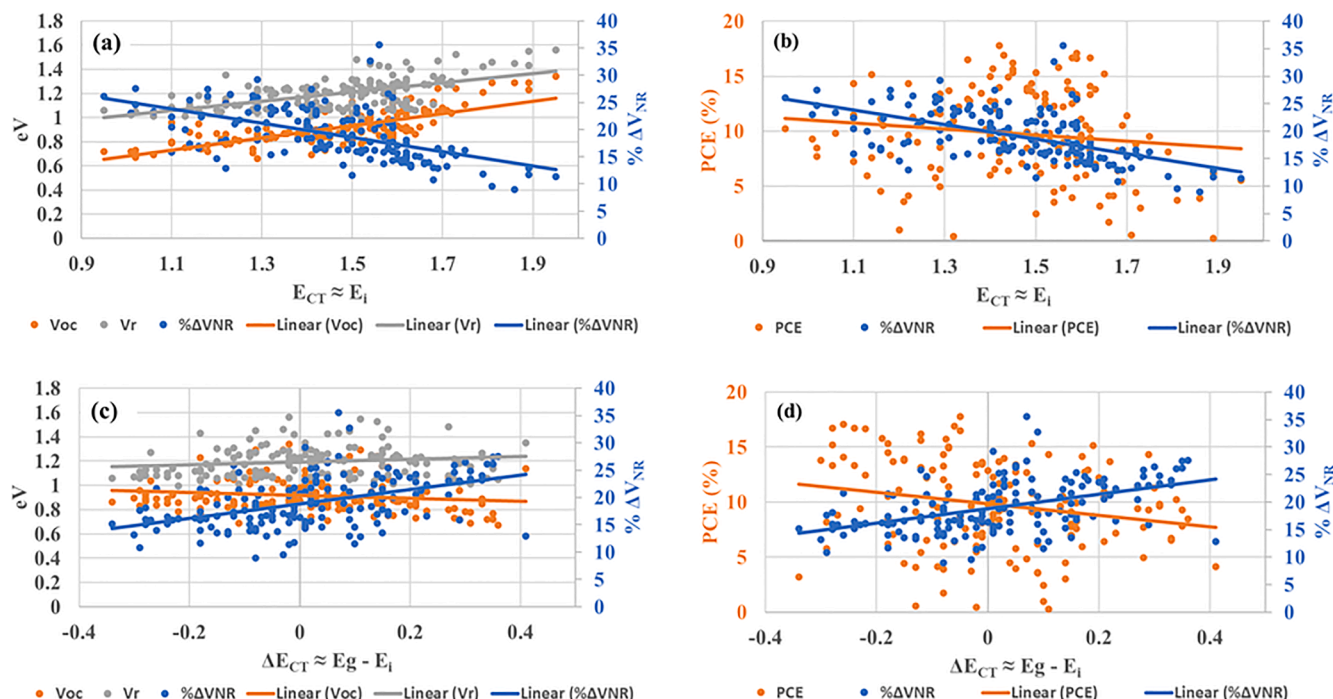


Fig. 2. Correlations with $\% \Delta V_{NR}$. (a) represents that with an increase in E_{CT} , $\% \Delta V_{NR}$ is reduced. The increase in V_R explains the reduction in $\% \Delta V_{NR}$ and when E_{CT} is further increased, V_{OC} becomes more closer to V_R . Although E_{CT} is an essential descriptor concerning $\% \Delta V_{NR}$, it shows no clear trend with PCE as seen in (b). (c) represents that with an increase in ΔE_{CT} , $\% \Delta V_{NR}$ is also increased and causes reduced V_{OC} . Unlike E_{CT} , ΔE_{CT} shows a clear trend with both $\% \Delta V_{NR}$ and PCE, as seen in (d).

energy level offset on ΔV_{NR} (Classen et al., 2020). The study reveals that the key factor responsible for the reduction in ΔV_{NR} is negligible energy level offset (ΔE_{offset}). Results show very clear trend of decreasing ΔV_{NR} with a better-aligned HOMO offset. At very low HOMO offset, hybrid states are formed instead of CT states, leading to a higher correlation between ΔV_{NR} vs HOMO offset than ΔV_{NR} vs E_{CT} . The findings reveal that at negligible HOMO offset (ΔE_{HOMO}), exciton splitting lifetime is around 1300 ps, which is greater than exciton lifetime of pristine material. Thus, other than small ΔE_{offset} , exciton lifetime must be longer than exciton splitting lifetime so that the electroluminescence quantum efficiency is also high and leads to smaller ΔV_{NR} .

Many other studies have been performed to understand ΔV_{NR} origin and unravel its correlation with various associated parameters such as molecular orientation at the D:A interface (Chen et al., 2016; Ran et al., 2017), molecular weight of polymers (Baran et al., 2015), isotopic substitution (Chen and Brédas, 2018), energetic driving force (Karki et al., 2020; Liu et al., 2016), and device processing (Tuladhar et al., 2016).

2.1. Calculating ΔV_{NR}

2.1.1. Electroluminescence external quantum efficiency (EQE_{EL})

EQE_{EL} is defined as the ratio of the number of photons emitted from the device to the number of electrons injected into the device. ΔV_{NR} can be calculated from the well-established relation (Rau, 2007; Rau et al., 2014)

$$\Delta V_{NR} = \frac{k_B T}{q} \ln \left(\frac{1}{EQE_{EL}} \right) \quad (2)$$

2.1.2. Photovoltaic external quantum efficiency (EQE_{PV})

We can calculate the radiative limit of a solar cell (V_R) using reciprocity relation between absorption and emission. V_R is the open-circuit voltage assuming all the recombination events in the device are radiative, and ΔV_{NR} is equal to zero. The difference between V_R and actual

calculated V_{OC} corresponds to ΔV_{NR} (Vandewal et al., 2018). V_R is calculated as (Rau, 2007; Rosenthal et al., 2019)

$$V_R = \frac{kT}{q} \ln \left(\frac{J_{sc}}{J_{rad}^0} + 1 \right) = \frac{kT}{q} \ln \left(\frac{q \int_0^\infty EQE_{PV}(E) \phi_{AM1.5G}(E) dE}{q \int_0^\infty EQE_{PV}(E) \phi_{BB}(E) dE} + 1 \right) \quad (3)$$

$$\Delta V_{NR} = V_R - V_{OC} \quad (4)$$

3. Data gathering

Data is manually collected from the literature, containing 154 unique D:A combination with 46 distinct donors and 79 distinct acceptors. All the collected information is reported in Table S1 (Supplementary Information). Our dataset has 6 all-small-molecule OSCs, 5 all-polymer OSCs, and the remaining 143 are polymer donor:small molecule acceptor OSCs. To demonstrate that our dataset consists of a wide variety of donor and acceptor materials with diverse chemical structure, we have calculated similarity scores by using Morgan fingerprints (nbits = 1024, r = 2) separately for donors (Number of compound pairs = 1035) and acceptors (Number of compound pairs = 3081) as shown in Fig. 3. By default, Tanimoto distance is used to generate similarity scores in RDKit.

Chemical structures of all donor and acceptor materials were drawn on ChemDraw software, and their SMILES codes were generated. Using the SMILES code of distinct donor and acceptor materials, their molecular descriptors and molecular fingerprints were generated. Today various open-source libraries are available freely for generating molecular descriptors and fingerprints (Broad and Bindner, 2013; Dong et al., 2015; Hong et al., 2008; Moriwaki et al., 2018; Tetko et al., 2005). We have used two types of descriptors (RDKit and Mordred) and four types of fingerprints (Extended, MACCS, PubChem, Morgan) (Chen, 2019) for generating datasets for both donors and acceptors using SMILES codes.

The distribution of ΔV_{NR} and $\% \Delta V_{NR}$ is shown in Fig. 4a and 4b. It is worth noting that $\% \Delta V_{NR}$ has better Gaussian distribution than ΔV_{NR} and the calculated mean and median of ΔV_{NR} and $\% \Delta V_{NR}$ are 0.280,

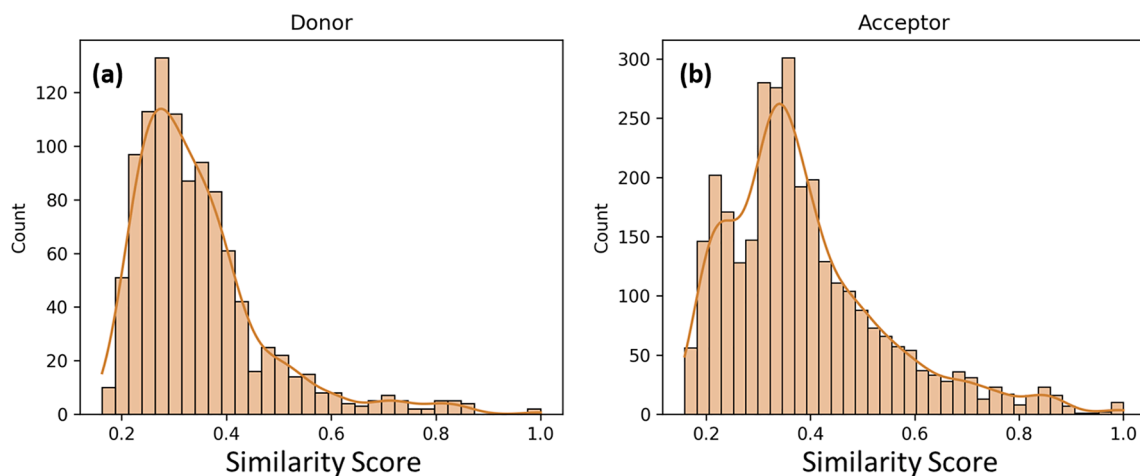


Fig. 3. Similarity scores of (a) donor molecules and (b) acceptor molecules generated by using Morgan Fingerprints (nbits = 1024, r = 2). Similarity scores represent that our dataset consists of diverse chemical structures, making it a suitable ML model dataset.

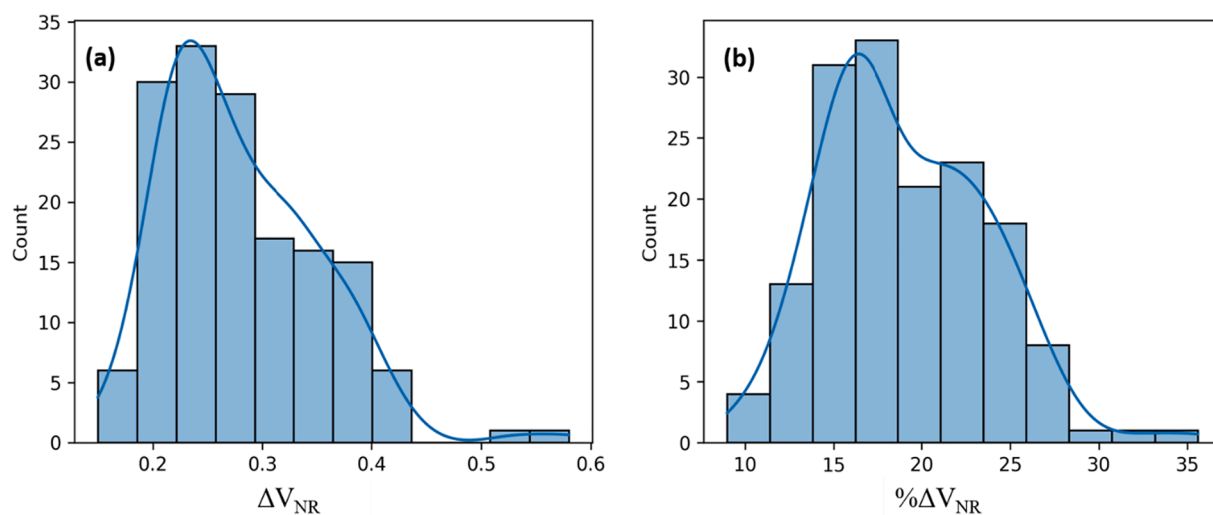


Fig. 4. Distribution of (a) Non-Radiative Voltage Loss (ΔV_{NR}) and (b) Percentage Non-Radiative Voltage Loss ($\% \Delta V_{NR}$). The mean and median of ΔV_{NR} and $\% \Delta V_{NR}$ are 0.280, 0.267, and 19.041%, 18.153%.

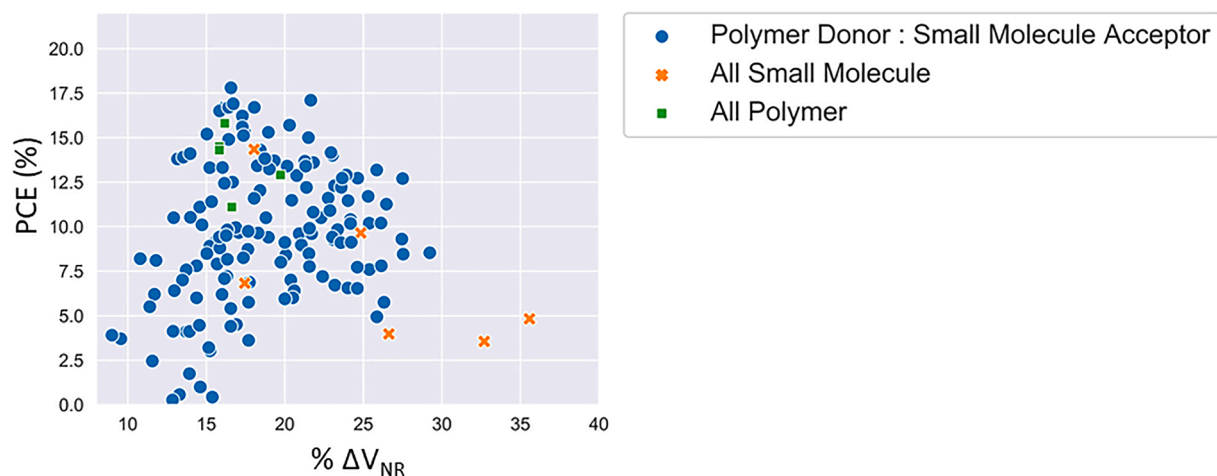


Fig. 5. Scatter plot between PCE and $\% \Delta V_{NR}$. No clear trend is observed in the plot, but it is essential to highlight that in the highest PCE range (greater than 15%), $\% \Delta V_{NR}$ is still high (15% to 20%).

0.267, and 19.041%, 18.153%, respectively.

It is also interesting to study the relation between $\% \Delta V_{NR}$ and PCE of the collected dataset. The scatter plot between PCE and $\% \Delta V_{NR}$ is given in Fig. 5, indicating that PCE does not depend solely on $\% \Delta V_{NR}$. It is observed that in the highest PCE range (greater than 15%), $\% \Delta V_{NR}$ is still approximately 15%–20%, and it needs to be reduced for further enhancement in PCE. (Fig. 5 shows clearly that a low $\% \Delta V_{NR}$ is not inherently high PCE, as many other factors influence the charge carrier dynamics and PCE.)

4. List of descriptors

4.1. Offset

Efforts have been made to understand the change in ΔV_{NR} with varying HOMO offset and LUMO offset (Baran et al., 2016; Liu et al., 2016). Saito et al. studied combination of PTB7-Th with nine different acceptors (Saito et al., 2020). The findings show that blends with negligible energy offset lead to increased radiative recombination rates, thus reducing ΔV_{NR} . Gao et al. also proposed low energy offset between donor and acceptor molecular states to be the thumb rule for reduced ΔV_{NR} (Qian et al., 2018).

4.2. Energy of charge transfer state (E_{CT})

E_{CT} is approximated by the gap between D_{HOMO} and A_{LUMO} and is referred to as Interfacial Bandgap (E_i) (Azzouzi et al., 2019). Vandewal et al. studied 170 different fullerene and non-fullerene combinations and revealed a clear trend between E_{CT} and ΔV_{NR} (Benduhn et al., 2017; Vandewal, 2016). However, the observed trend is considerably scattered, which indicates the existence of hidden parameters that are yet to be explored.

4.3. Bandgap (E_g)

Various definitions of E_g have been used in the literature such as E_g absorbance onset, E_g intersection (crossing point of absorption and emission spectra) (Cui et al., 2019, 2020; Karki et al., 2019; Luo et al., 2019; Vandewal et al., 2018; Xie et al., 2019a; Zhang et al., 2019; Zhou et al., 2019), E_g onset of EQE spectra (Zhang et al., 2020), and E_g edge from EQE spectra (Xie et al., 2019b). Throughout this study, we have defined E_g as optical bandgap of the donor or acceptor material used in the active bulk heterojunction (BHJ) layer whichever is lower, as it provides a straightforward comparison between the results in the literature.

4.4. Δe_{CT}

It was understood that an energetic penalty in the form of $\Delta E_{CT} = E_g - E_{CT}$ is required for the efficient generation of free charge carriers (Grancini et al., 2013; Jailaubekov et al., 2013; Vandewal, 2016). Our definition of ΔE_{CT} is ($E_g - (D_{HOMO} - A_{LUMO})$), the driving energy for dissociation of charge transfer state. Surprisingly, with such low dielectric constant materials, all recent OSCs with PCE greater than 15 % have zero ΔE_{CT} (Eisner et al., 2019; Nikolis et al., 2017; Vandewal et al., 2020; Yu et al., 2019). Flurin et al. used materials with different energetic offsets to show that in very low offset combinations, E_{S1} (energy of lowest singlet excited state) and E_{CT} are close to each other (Eisner et al., 2019). Further, when the offset is sufficiently small, hybridization occurs, leading to increased electroluminescence and reduced ΔV_{NR} . Therefore, $E_{S1, CT} = E_{S1} - E_{CT}$ is defined as the driving energy for charge transfer state dissociation. Azzouzi et al. also performed a study on the potential change in V_{OC} and found out that ΔE_{CT} is the most sensitive variable followed by E_{CT} (Azzouzi et al., 2018).

4.5. Molecular descriptors and fingerprints

Simplified molecular input line entry system (SMILES) (Weininger, 1988) is used to define chemical structures in a machine-readable format (ASCII strings). Two molecular descriptor sets (RDKit and Mordred) and four molecular fingerprint sets (Extended, MACCS, PubChem, and Morgan) were generated using SMILES strings for polymer monomer and small molecule materials.

4.5.1. Molecular descriptor Set

To construct an efficient machine learning model, a fair collection of molecular descriptors is required to represent the information encoded within the chemical structures rather than SMILES codes. Molecular descriptors provide the information encoded within the molecule in numerical form, which is machine-readable. RDKit descriptors (196 bits) were calculated from ChemDes (web-based platform) (Dong et al., 2015), and Mordred descriptors (1613 bits) were calculated using the Mordred python library (Moriwaki et al., 2018). We have used only 1D and 2D descriptors throughout this study.

4.5.2. Molecular fingerprint Set

A molecular fingerprint is an array of binary bits representing a predefined structural feature. If the predefined structure is present, bit is set to 1 (ON) otherwise 0 (OFF). The more the number of bits, the more is structural information. Four fingerprints have been used in this study; (a) Extended fingerprint (1024 bits) is an extended version of Chemistry Development Kit (CDK) fingerprint (Steinbeck et al., 2003) where additional bits describe ring features, (b) Molecular ACCess System (MACCS) key (166 bits) (Durant et al., 2002), (c) PubChem fingerprints (881 bits) (Kim et al., 2016), and (d) Morgan fingerprint (1024 bits) (Rogers and Hahn, 2010).

The correlation matrix of the reported descriptors in Fig. 6 shows that none of the descriptors directly correlate with our target variable $\% \Delta V_{NR}$. The highest correlation of $\% \Delta V_{NR}$ is observed with L_{offset} (0.53) and E_i (-0.55), while the least correlation is observed with E_g (-0.18).

5. Prediction of $\% \Delta V_{NR}$ using Machine learning

Workflow for the prediction of $\% \Delta V_{NR}$ by using ML algorithms is represented in Fig. 7. We have studied various combinations of descriptor sets consisting of FMO, E_g , molecular descriptors, and molecular fingerprints to predict $\% \Delta V_{NR}$. Extensive feature engineering is required, as our data involved in our work is small and sparse. Highly dispersed FMO and E_g values are reported in literatures, so we transformed all reported values with the median for each distinct donor and acceptor. Feature engineering involves steps to convert the raw data into valuable data that can be fed directly to ML models. In the descriptor dataset (RDKit and Mordred), we removed all the features with zero standard deviation, followed by the removal of features with a correlation coefficient (r) greater than 0.8. While for the fingerprint dataset, all the irrelevant features were removed by setting a variance threshold of 0.8. By this variance threshold, all the features with zero standard deviation were also removed. Finally, prepared datasets were scaled with a standard scaler and fed to ML models.

5.1. ML models

We have used four supervised ML approaches, which are accessible from Scikit-Learn python package (Pedregosa et al., 2011). A detailed description of all ML algorithms is beyond this paper's scope, so a short description of them is given below.

5.1.1. Random forest (RF) regressor

In RF, various decision trees are formed randomly and are called the base learners. By a method called 'Bagging', random sets of row and feature are selected with replacement for making decision trees.



Fig. 6. Pearson correlation matrix of electronic descriptors and $\% \Delta V_{NR}$ for getting an initial insight of the collected data. None of the descriptors is highly correlated with $\% \Delta V_{NR}$, indicating that ML models need to understand the complex relationships between the descriptors for better prediction of $\% \Delta V_{NR}$.

Separately each decision tree has high variance, but when they all are combined in parallel, the resultant variance is reduced. The final result is the mean of the outputs from all decision trees.

5.1.2. Gradient boosting (GB) regressor

In GB also, various decision trees are formed but not randomly instead sequentially. With sequential trees getting formed, errors made in the previous stage get rectified to improve the model's overall performance. By sequential formation of trees, several weak learners are combined to make a strong learner.

5.1.3. Support vector regression (SVR)

To provide a low variance model, SVR provides flexibility to define acceptable error limits in the model. This limit helps in finding the most appropriate hyperplane for high-dimensional data. Parameters that can be tuned are 'C' and 'epsilon'. 'C' is the regularization parameter and 'epsilon' is associated with the region in which no penalty is applied in the training loss function. In our model, tuning of both these parameters is performed using GridSearchCV.

5.1.4. Artificial neural network (ANN)

It is a set of many layers connected with each other forming a neural network. ANN is applied using a Multilayer perceptron (MLP) regressor. It consists of input layer, output layer, and hidden layer sandwiched between them with a user-defined number of neurons. In our model, we have selected number of neurons equals to number of features left after performing feature engineering.

6. Results and discussion

All ML techniques (RF, GB, SVR, ANN) are accessed from the Scikit-Learn python package. For RF and GB model, default parameters were used. For SVR model, hyperparameter tuning was performed by tuning values of 'C' and 'epsilon' using GridSearchCV. For ANN, we have used one hidden layer with number of neurons equals to the number of features left after feature engineering. As our dataset is small (154 rows), we have used leave-one-out cross-validation (LOOCV) technique to evaluate the results. A comparison between the models is made by using statistical metrics such as Pearson correlation coefficient (r), root mean squared error (RMSE), and mean absolute percentage error (MAPE) for each dataset. Equations of these metrics are listed in Table 1.

First, we tested the model by using FMO descriptors (D_{HOMO} , D_{LUMO} , A_{HOMO} , A_{LUMO} , L_{offset} , H_{offset} , E_i , E_g , and various descriptor/fingerprint datasets generated by us. The results obtained are summarized in Table 2.

Out of all the datasets, the best results were obtained by GB model. Set 1a performed well ($r = 0.797$), and the results were considerably improved when combined with molecular descriptors (Set 1b, 1c) and molecular fingerprints (Set 1d-1 g). The descriptor set with RDKit ($r = 0.859$) and Mordred ($r = 0.85$) gave approximately equal results by the GB model. In all the sets (1a – 1 g), GB and RF model results are approximately equal. Although the number of descriptors in RDKit dataset (196 bits) is much less than Mordred dataset (1613 bits), similar results indicate that all the additional descriptors in Mordred dataset do not hold any relevant information related to $\% \Delta V_{NR}$. But in the case of

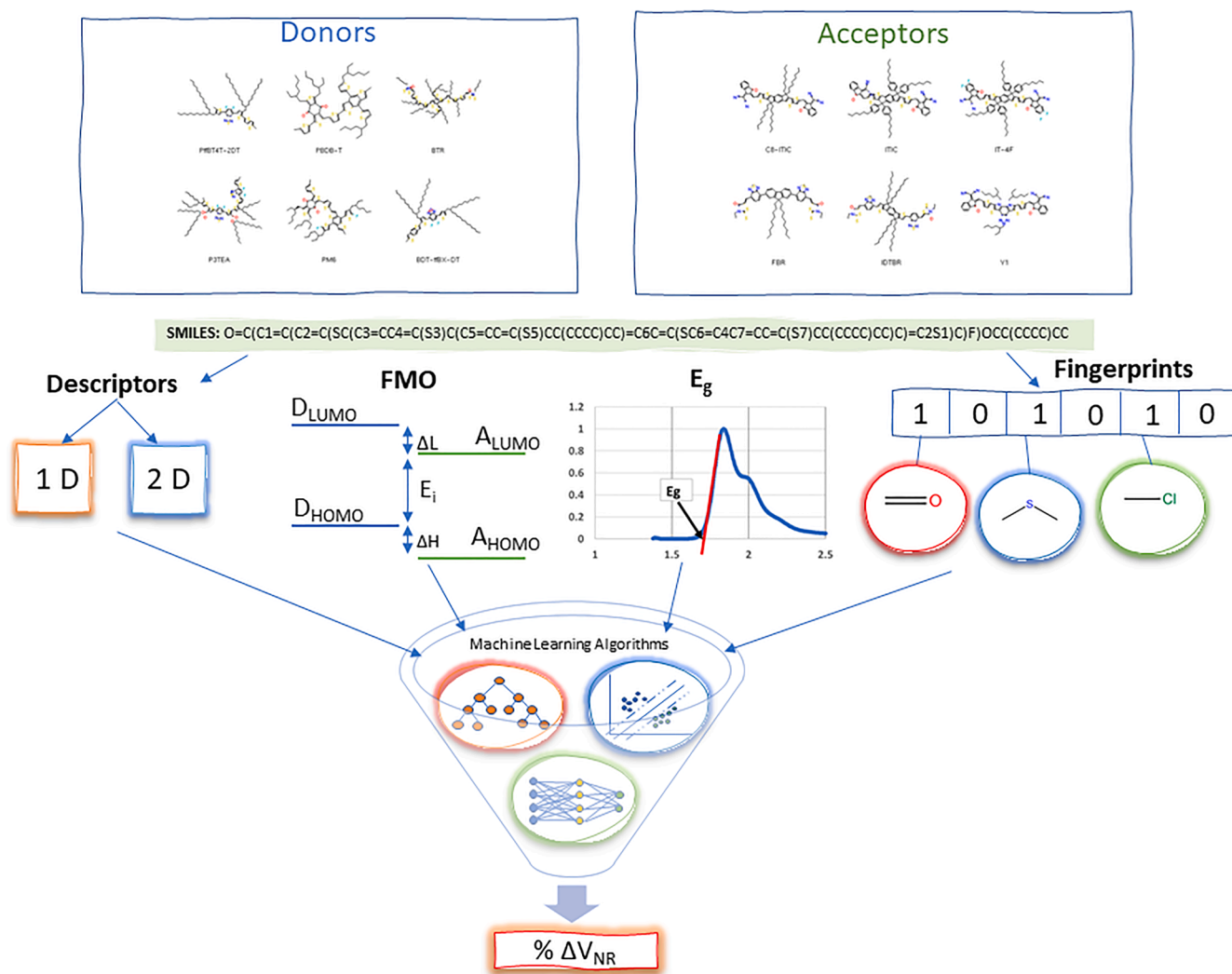


Fig. 7. Machine learning workflow for the prediction of % ΔV_{NR} . Data for 154 unique D:A combination with reported % ΔV_{NR} is collected from the literature, having 46 distinct donors and 79 distinct acceptors. Reported FMO and E_g values are taken from literature and then transformed by median values for distinct donors and acceptors. SMILES code of donor and acceptor molecules are generated by using ChemDraw software. SMILES codes are then used to generate molecular descriptor datasets and molecular fingerprint datasets. Finally, the datasets are scaled and fed into ML models for the prediction of % ΔV_{NR} .

Table 1

Metrics and their equation used for evaluating the performance of ML models.

Metric	Equation
Pearson Correlation Coefficient (r)	$\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$
RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}$
MAPE	$\frac{1}{n} \sum_{i=1}^n \left \frac{x_i - y_i}{x_i} \right \times 100$

SVR model, the dataset with Mordred descriptors ($r = 0.803$) performed better than dataset with RDKit descriptors ($r = 0.777$). Results interpret the better performance of SVR model with higher dimensional data in the case of molecular descriptors. In case of fingerprints, dataset with MACCS fingerprint performed best ($r = 0.857$) followed by dataset with Extended fingerprints ($r = 0.85$). It should be noted that out of all fingerprints, MACCS contains the least number of bits (166 bits). Such results indicate that fingerprints with a much higher number of bits contain information that is irrelevant concerning % ΔV_{NR} . Out of all the sets (1a – 1 g), ANN has not performed best, and the reason can be attributed to less amount of data, which is not enough to train neural networks to understand complex relations completely. It is also worth

noting that all-polymer and all-small-molecule OSCs also fit well into the best performing models, as shown in Fig. 8 below.

We further examined ML models by considering only descriptor sets (Set 2a,2b) and only fingerprint sets (Set 2c-2f) as summarized in Table 3. Results are degraded a bit but still holds great importance for virtual screening of D:A combinations with reduced ΔV_{NR} .

In case of descriptors (Set 2a,2b), the best results are obtained by the Mordred descriptor dataset with SVR model ($r = 0.78$) followed by RDKit descriptor set with GB model ($r = 0.753$). While in case of fingerprints (Set 2c-2f), Extended fingerprints ($r = 0.728$) performed best with GB model, followed by Morgan fingerprints ($r = 0.714$) with ANN model. Results of best performing model with only mordred descriptor (Set 2b) and only extended fingerprint (Set 2c) is shown in Fig. 9.

To further investigate the robustness of these ML models, we randomly selected 50 polymer donor:non-fullerene acceptor material systems from the literature. Data for additional 50 OSCs is reported in Table S2 (Supplementary Information). We calculated their % ΔV_{NR} by digitizing their reported EQE spectra as mentioned in section 2.1.2.. These 50 new D:A combinations in the dataset contain 18 new distinct donors and 19 new distinct acceptors. Now the number of distinct donors are 64, number of distinct acceptors are 97, and the total number of unique D:A combination becomes 204. For this increased dataset,

Table 2

Results for FMO + E_g dataset (Set 1a), FMO + E_g + Descriptor dataset (Set 1b,1c), and FMO + E_g + Fingerprint dataset (Set 1d-1 g) for prediction of $\% \Delta V_{NR}$ using different ML algorithms. Results are compared based on Pearson r , RMSE, and MAPE.

	Metric	RF	GB	SVR	ANN
FMO + E_g (Set 1a)	r	0.785	0.797	0.752	0.705
	RMSE	2.857	2.794	3.046	3.297
	MAPE	11.009	11.275	12.512	12.579
FMO + E_g + RDKit Des (Set 1b)	r	0.836	0.859	0.777	0.788
	RMSE	2.568	2.364	2.911	2.884
	MAPE	10.217	9.409	10.766	11.291
FMO + E_g + Mordred Des (Set 1c)	r	0.823	0.85	0.803	0.786
	RMSE	2.671	2.431	2.756	2.942
	MAPE	10.607	9.56	10.946	11.983
FMO + E_g + Extended FP (Set 1d)	r	0.826	0.85	0.761	0.759
	RMSE	2.62	2.428	2.999	3.112
	MAPE	10.353	9.776	11.747	13.618
FMO + E_g + MACCS FP (Set 1e)	r	0.829	0.857	0.807	0.792
	RMSE	2.585	2.371	2.723	2.886
	MAPE	10.207	9.668	10.369	10.717
FMO + E_g + PubChem FP (Set 1f)	r	0.821	0.826	0.789	0.805
	RMSE	2.64	2.601	2.836	2.76
	MAPE	10.635	10.537	10.391	10.373
FMO + E_g + Morgan FP (Set 1g)	r	0.84	0.83	0.799	0.799
	RMSE	2.515	2.576	2.775	2.828
	MAPE	9.733	10.976	10.458	11.152

similarity scores of donors and acceptors are given in [Figure S1 \(supplementary information\)](#). The distribution of ΔV_{NR} and $\% \Delta V_{NR}$ is represented in [Figure S2 \(supplementary information\)](#).

In the increased dataset ($\# = 204$) results summarized in [Table S3 \(supplementary information\)](#), approximately equal results are obtained for set 3a by SVR model ($r = 0.69$) and GB model ($r = 0.683$). Here also, results obtained by including descriptors/fingerprints (Set 3b-3 g) are far better than electronic descriptors alone (Set 3a). Best results for descriptor dataset (Set 3b,3c) is obtained by GB model with RDKit ($r = 0.817$) followed by SVR model with Mordred ($r = 0.798$). While in the case of fingerprint dataset (Set 3d-3 g), SVR model with PubChem provides the best result ($r = 0.827$) followed by GB model with Extended ($r = 0.818$). Similar to prior cases, SVR model performed well with the Mordred descriptor dataset ($r = 0.798$). The results are summarized in [Table S3 \(supplementary information\)](#) and scatter plots of best-performing models are shown in [Figure S3 \(supplementary information\)](#).

Again, this increased dataset ($\# = 204$) was used to predict $\% \Delta V_{NR}$ by using only descriptors and only fingerprints. Results are summarized in the [Table S4 \(supplementary information\)](#). For the descriptor dataset

(Set 4a,4b), best results were obtained by SVR model with Mordred ($r = 0.773$) followed by SVR model with RDKit ($r = 0.706$). While in case of fingerprint dataset (Set 4c-4f), Extended gave approximately equal results with SVR ($r = 0.741$) and ANN ($r = 0.746$) model. Scatter plot of best performing models are shown in [Figure S4 \(supplementary information\)](#).

After studying all models, it is important to note that SVR model with Mordred descriptors provides the most consistent results in all the studies done by us (r ranging from 0.773 to 0.803). Results with SVR model shown in [Fig. 10](#) indicates that, SVR is able to understand all the complex relations effectively by just using Mordred descriptors.

7. Conclusion

In summary, non-radiative voltage losses are one of the major factors influencing the V_{OC} of OSCs and limiting PCE. This work has shown how ML algorithms can be effectively used to predict $\% \Delta V_{NR}$ and for virtual screenings of potential D/A combinations with reduced $\% \Delta V_{NR}$. Using electronic descriptors (FMO and E_g) with GB model, correlation coefficient ($r = 0.797$) was obtained and the results were significantly improved by combination of electronic and structural descriptors with GB model ($r = 0.859$). With only structural descriptors, a good correlation coefficient ($r = 0.78$) was obtained with SVR model. Structural

Table 3

Results for only descriptor datasets (Set 2a,2b) and only fingerprint datasets (2c-2f) for prediction of $\% \Delta V_{NR}$ using different ML algorithms. Results are compared based on Pearson r , RMSE, and MAPE.

	Metric	RF	GB	SVR	ANN
RDKit Des (Set 2a)	r	0.701	0.753	0.719	0.715
	RMSE	3.311	3.033	3.207	3.397
	MAPE	12.922	11.794	12.332	12.791
Mordred Des (Set 2b)	r	0.72	0.734	0.78	0.731
	RMSE	3.217	3.141	2.899	3.331
	MAPE	12.245	11.693	11.328	12.975
Extended FP (Set 2c)	r	0.688	0.728	0.726	0.727
	RMSE	3.363	3.188	3.196	3.282
	MAPE	12.923	12.392	12.222	14.007
MACCS FP (Set 2d)	r	0.656	0.711	0.685	0.674
	RMSE	3.541	3.294	3.387	3.549
	MAPE	14.164	14.098	13.962	14.738
PubChem FP (Set 2e)	r	0.616	0.64	0.609	0.57
	RMSE	3.693	3.608	3.692	4.028
	MAPE	14.566	14.443	14.487	16.063
Morgan FP (Set 2f)	r	0.705	0.706	0.71	0.714
	RMSE	3.275	3.312	3.251	3.325
	MAPE	12.411	12.832	12.387	13.064

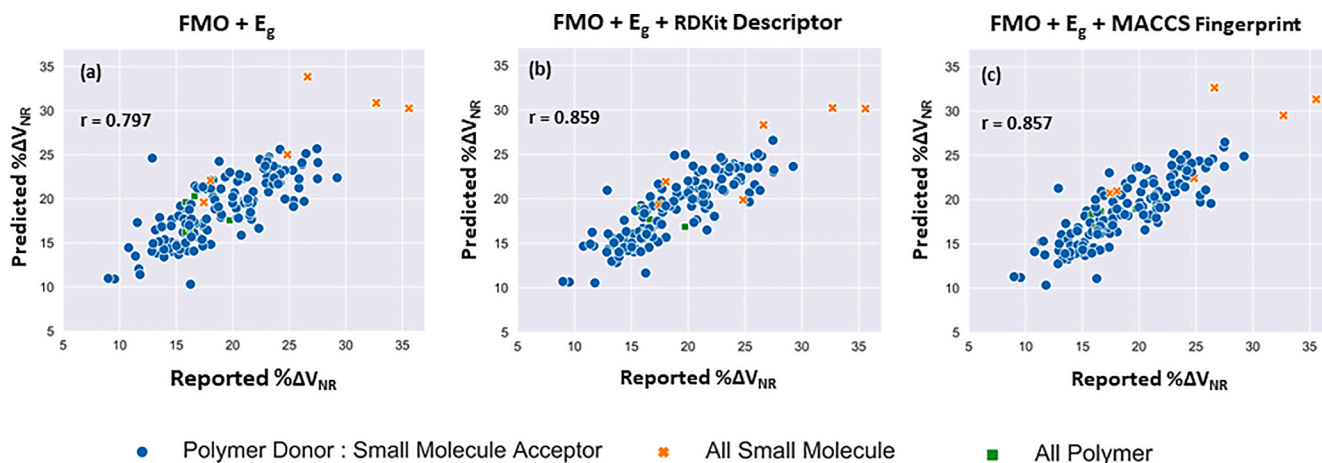


Fig. 8. Reported vs predicted results from GB model. (a) FMO + E_g ($r = 0.797$), (b) FMO + E_g + RDKit descriptor ($r = 0.859$) and (c) FMO + E_g + MACCS fingerprint ($r = 0.857$).

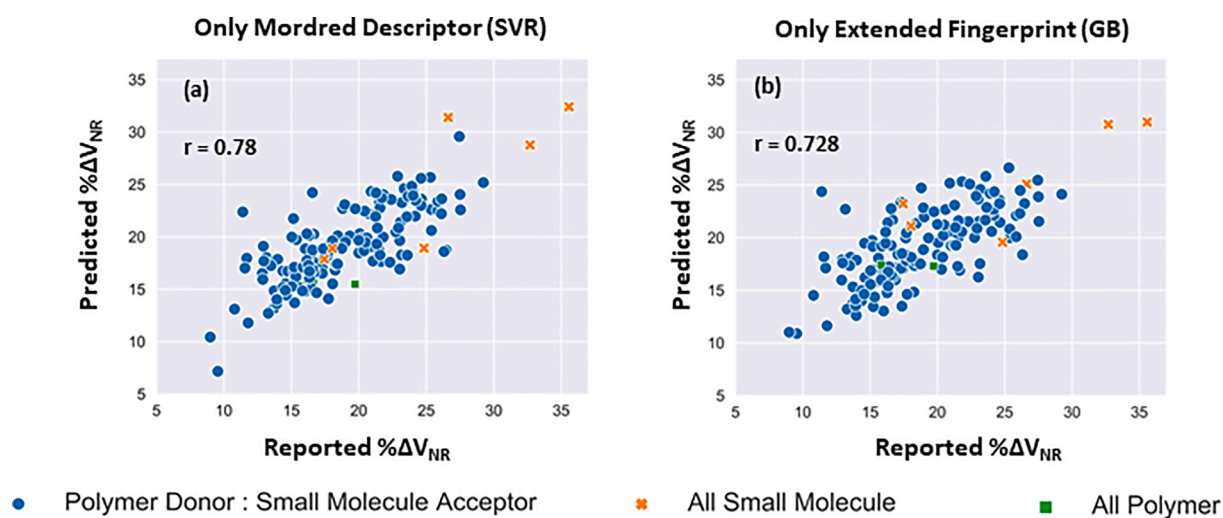


Fig. 9. Reported vs Predicted results for only (a) Mordred descriptor with SVR model ($r = 0.78$), (b) Extended Fingerprint with GB model ($r = 0.728$).

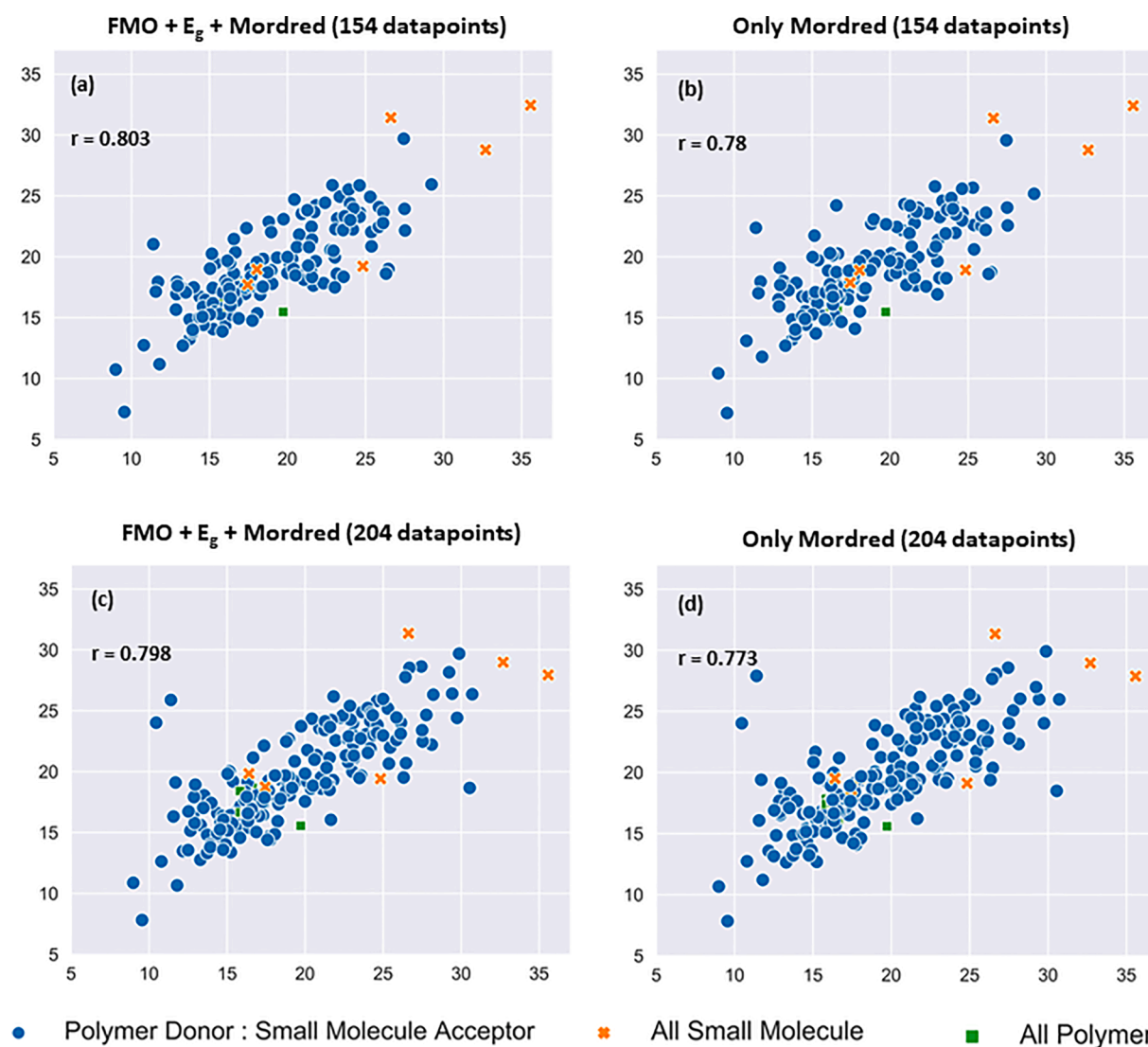


Fig. 10. SVR model showing consistent results with (a) FMO + E_g + Mordred (154 datapoints) ($r = 0.803$), (b) Only Mordred (154 datapoints) ($r = 0.78$), (c) (# = 204) FMO + E_g + Mordred ($r = 0.798$) and (d) (# = 204) Only Mordred ($r = 0.773$).

descriptors contain all the hidden information that is encoded within the chemical structures of donor and acceptor molecules. ML model with only structural descriptors can shortlist compounds with low non-radiative loss from a pool of compounds, thereby saving the time as well as cost for developing high-performing materials. To further investigate the robustness of these ML models, we randomly selected 50 polymer donor:non-fullerene acceptor material systems from the literature and again obtained impressive results using a combination of electronic and structural descriptors ($r = 0.827$) and only structural descriptors ($r = 0.773$). SVR model obtained most consistent results with the Mordred dataset (r ranging from 0.773 to 0.803). Small dataset limits the ML potential as the chemical space understood by ML models is pretty less. A much larger dataset is required to understand the complex relationship between structural modification of molecules and their photovoltaic properties. The results reported in this work has open up an opportunity toward even precise ML models predicting the performance of OSCs.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are thankful to Solar Energy Research Initiative (SERI)-Department of Science and Technology (DST) for financial support.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.solener.2021.09.056>.

References

- 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.-G., Rand, B.P., Rau, U., Snaith, H.J., Unger, E., Vaillant-Roca, L., Yip, H.-L., Brabec, C.J., 2021. Device Performance of Emerging Photovoltaic Materials (Version 1). *Adv. Energy Mater.* 11 (11), 2002774. <https://doi.org/10.1002/aenm.v11.1110.1002/aenm.202002774>.
- Azzouzi, M., Kirchartz, T., Nelson, J., 2019. Factors Controlling Open-Circuit Voltage Losses in Organic Solar Cells. *Trends Chem.* 1 (1), 49–62. <https://doi.org/10.1016/j.trechm.2019.01.010>.
- Azzouzi, M., Yan, J., Kirchartz, T., Liu, K., Wang, J., Wu, H., Nelson, J., 2018. Nonradiative Energy Losses in Bulk-Heterojunction Organic Photovoltaics. *Phys. Rev. X* 8, 31055. <https://doi.org/10.1103/PhysRevX.8.031055>.
- Babics, M., Duan, T., Balawi, A.H., Liang, R.-Z., Cruciani, F., Carja, I.-D., Gottlieb, D., McCulloch, I., Vandewal, K., Laquai, F., Beaujuge, P.M., 2019. Negligible Energy Loss During Charge Generation in Small-Molecule/Fullerene Bulk-Heterojunction Solar Cells Leads to Open-Circuit Voltage over 1.10 V. *ACS Appl. Energy Mater.* 2 (4), 2717–2722. <https://doi.org/10.1021/acsaem.8b02020.1021/acsaem.8b02020.s001>.
- Baran, D., Kirchartz, T., Wheeler, S., Dimitrov, S., Abdelsamie, M., Gorman, J., Ashraf, R. S., Holliday, S., Wadsworth, A., Gasparini, N., Kaiburg, P., Yan, H., Amassian, A., Brabec, C.J., Durrant, J.R., McCulloch, I., 2016. Reduced voltage losses yield 10% efficient fullerene free organic solar cells with >1 V open circuit voltages. *Energy Environ. Sci.* 9, 3783–3793. <https://doi.org/10.1039/C6EE02598F>.
- Baran, D., Vezie, M.S., Gasparini, N., Deledalle, F., Yao, J., Schroeder, B.C., Bronstein, H., Ameri, T., Kirchartz, T., McCulloch, I., Nelson, J., Brabec, C.J., 2015. Role of Polymer Fractionation in Energetic Losses and Charge Carrier Lifetimes of Polymer: Fullerene Solar Cells. *J. Phys. Chem. C* 119 (34), 19668–19673. <https://doi.org/10.1021/acs.jpcc.5b05709>.
- Benduhn, J., Tvingstedt, K., Piersimoni, F., Ullbrich, S., Fan, Y., Tropiano, M., McGarry, K.A., Zeika, O., Riede, M.K., Douglas, C.J., Barlow, S., Marder, S.R., Neher, D., Spoltore, D., Vandewal, K., 2017. Intrinsic non-radiative voltage losses in fullerene-based organic solar cells. *Nat. Energy* 2, 17053. <https://doi.org/10.1038/nenergy.2017.53>.
- Broad, J., Bindner, A., 2013. Hacking with Kali: Practical penetration testing techniques. *Hacking with Kali Pract. Penetration Test. Tech.* 56, 1–227. <https://doi.org/10.1016/C2012-0-02727-5>.
- Chen, F.-C., 2019. Virtual Screening of Conjugated Polymers for Organic Photovoltaic Devices Using Support Vector Machines and Ensemble Learning. *Int. J. Polym. Sci.* 2019, 1–7. <https://doi.org/10.1155/2019/4538514>.
- Chen, X.-K., Brédas, J.-L., 2018. Voltage Losses in Organic Solar Cells: Understanding the Contributions of Intramolecular Vibrations to Nonradiative Recombinations. *Adv. Energy Mater.* 8 (9), 1702227. <https://doi.org/10.1002/aenm.v8.910.1002/aenm.201702227>.
- Chen, X.K., Ravva, M.K., Li, H., Ryno, S.M., Brédas, J.L., 2016. Effect of Molecular Packing and Charge Delocalization on the Nonradiative Recombination of Charge-Transfer States in Organic Solar Cells. *Adv. Energy Mater.* 6, 1–10. <https://doi.org/10.1002/aenm.201601325>.
- Classen, A., Chochos, C.L., Lüer, L., Gregoriou, V.G., Wortmann, J., Osvet, A., Forberich, K., McCulloch, I., Heumüller, T., Brabec, C.J., 2020. The role of exciton lifetime for charge generation in organic solar cells at negligible energy-level offsets. *Nat. Energy* 5 (9), 711–719. <https://doi.org/10.1038/s41560-020-00684-7>.
- Cui, Y., Yao, H., Zhang, J., Xian, K., Zhang, T., Hong, L., Wang, Y., Xu, Y.e., Ma, K., An, C., He, C., Wei, Z., Gao, F., Hou, J., 2020. Single-Junction Organic Photovoltaic Cells with Approaching 18% Efficiency. *Adv. Mater.* 32 (19), 1908205. <https://doi.org/10.1002/adma.v32.1910.1002/adma.201908205>.
- Cui, Y., Yao, H., Zhang, J., Zhang, T., Wang, Y., Hong, L., Xian, K., Xu, B., Zhang, S., Peng, J., Wei, Z., Gao, F., Hou, J., 2019. Over 16% efficiency organic photovoltaic cells enabled by a chlorinated acceptor with increased open-circuit voltages. *Nat. Commun.* 10, 2515. <https://doi.org/10.1038/s41467-019-10351-5>.
- 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., 2015. ChemDes: An integrated web-based platform for molecular descriptor and fingerprint computation. *J. Cheminform.* 7, 1–10. <https://doi.org/10.1186/s13321-015-0109-z>.
- Durant, J.L., Leland, B.A., Henry, D.R., Nourse, J.G., 2002. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* 42 (6), 1273–1280. <https://doi.org/10.1021/ci010132r>.
- Eisner, F.D., Azzouzi, M., Fei, Z., Hou, X., Anthopoulos, T.D., Dennis, T.J.S., Heeney, M., Nelson, J., 2019. Hybridization of Local Exciton and Charge-Transfer States Reduces Nonradiative Voltage Losses in Organic Solar Cells. *J. Am. Chem. Soc.* 141 (15), 6362–6374. <https://doi.org/10.1021/jacs.9b01465.1021/jacs.9b01465.s001>.
- Fu, H., Wang, Y., Meng, D., Ma, Z., Li, Y., Gao, F., Wang, Z., Sun, Y., 2018. Suppression of Recombination Energy Losses by Decreasing the Energetic Offsets in Perylene Diimide-Based Nonfullerene Organic Solar Cells. *ACS Energy Lett.* 3 (11), 2729–2735. <https://doi.org/10.1021/acsenerylett.8b01665.1021/acsenerylett.8b01665.s001>.
- Grancini, G., Maiuri, M., Fazzi, D., Petrozza, A., Egelhaaf, H.-J., Brida, D., Cerullo, G., Lanzani, G., 2013. Hot exciton dissociation in polymer solar cells. *Nat. Mater.* 12 (1), 29–33. <https://doi.org/10.1038/nmat3502>.
- Hong, H., Xie, Q., Ge, W., Qian, F., Fang, H., Shi, L., Su, Z., Perkins, R., Tong, W., 2008. Mold2, molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. *J. Chem. Inf. Model.* 48 (7), 1337–1344. <https://doi.org/10.1021/ci800038f>.
- Hong, L., Yao, H., Yu, R., Xu, Y.e., Gao, B., Ge, Z., Hou, J., 2019. Investigating the Trade-Off between Device Performance and Energy Loss in Nonfullerene Organic Solar Cells. *ACS Appl. Mater. Interfaces* 11 (32), 29124–29131. <https://doi.org/10.1021/acsaem.9b10243.1021/acsaem.9b10243.s001>.
- Jailaubekov, A.E., Willard, A.P., Tritsch, J.R., Chan, W.-L., Sai, N.a., Gearba, R., Kaake, L. G., Williams, K.J., Leung, K., Rossky, P.J., Zhu, X.-Y., 2013. Hot charge-transfer excitons set the time limit for charge separation at donor/acceptor interfaces in organic photovoltaics. *Nat. Mater.* 12 (1), 66–73. <https://doi.org/10.1038/nmat3500>.
- Karki, A., Gillett, A.J., Friend, R.H., Nguyen, T.-Q., 2021. The Path to 20% Power Conversion Efficiencies in Nonfullerene Acceptor Organic Solar Cells. *Adv. Energy Mater.* 11 (15), 2003441. <https://doi.org/10.1002/aenm.v11.1510.1002/aenm.202003441>.
- Karki, A., Vollbrecht, J., Dixon, A.L., Schopp, N., Schrock, M., Reddy, G.N.M., Nguyen, T.-Q., 2019. Understanding the High Performance of over 15% Efficiency in Single-Junction Bulk Heterojunction Organic Solar Cells. *Adv. Mater.* 31 (48), 1903868. <https://doi.org/10.1002/adma.v31.4810.1002/adma.201903868>.
- Karki, A., Vollbrecht, J., Gillett, A.J., Xiao, S.S., Yang, Y., Peng, Z., Schopp, N., Dixon, A. L., Yoon, S., Schrock, M., Ade, H., Reddy, G.N.M., Friend, R.H., Nguyen, T.-Q., 2020. The role of bulk and interfacial morphology in charge generation, recombination, and extraction in non-fullerene acceptor organic solar cells. *Energy Environ. Sci.* 13 (10), 3679–3692. <https://doi.org/10.1039/D0EE01896A>.
- Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A., Wang, J., Yu, B.o., Zhang, J., Bryant, S.H., 2016. PubChem substance and compound databases. *Nucleic Acids Res.* 44 (D1), D1202–D1213. <https://doi.org/10.1093/nar/gkv951>.
- Kranthiraja, K., Saeki, A., 2021. Experiment-Oriented Machine Learning of Polymer:Non-Fullerene Organic Solar Cells. *Adv. Funct. Mater.* 31 (23), 2011168. <https://doi.org/10.1002/adfm.v31.2310.1002/adfm.202011168>.
- Lee, M.-H., 2020a. A Machine Learning-Based Design Rule for Improved Open-Circuit Voltage in Ternary Organic Solar Cells. *Adv. Intell. Syst.* 2 (1), 1900108. <https://doi.org/10.1002/aisy.v2.110.1002/aisy.201900108>.
- Lee, M.-H., 2020b. Robust random forest based non-fullerene organic solar cells efficiency prediction. *Org. Electron.* 76, 105465. <https://doi.org/10.1016/j.orgel.2019.105465>.
- Lee, M., 2019. Insights from Machine Learning Techniques for Predicting the Efficiency of Fullerene Derivatives-Based Ternary Organic Solar Cells at Ternary Blend Design. *Adv. Energy Mater.* 9, 1900891. <https://doi.org/10.1002/aenm.201900891>.

- Lee, M.-H., 2020c. Performance and Matching Band Structure Analysis of Tandem Organic Solar Cells Using Machine Learning Approaches. *Energy Technol.* 8 (3), 1900974. <https://doi.org/10.1002/ente.v8i3.1002/ente.201900974>.
- Li, S., Li, C.-Z., Shi, M., Chen, H., 2020. New Phase for Organic Solar Cell Research: Emergence of Y-Series Electron Acceptors and Their Perspectives. *ACS Energy Lett.* 5 (5), 1554–1567. <https://doi.org/10.1021/acsenenergylett.0c00537>.
- Liu, J., Chen, S., Qian, D., Gautam, B., Yang, G., Zhao, J., Bergqvist, J., Zhang, F., Ma, W., Ade, H., Inganäs, O., Gundogdu, K., Gao, F., Yan, H., 2016. Fast charge separation in a non-fullerene organic solar cell with a small driving force. *Nat. Energy* 1, 16089. <https://doi.org/10.1038/nenergy.2016.89>.
- Liu, Q., Jiang, Y., Jin, K.e., Qin, J., Xu, J., Li, W., Xiong, J.i., Liu, J., Xiao, Z., Sun, K., Yang, S., Zhang, X., Ding, L., 2020a. 18% Efficiency organic solar cells. *Sci. Bull.* 65 (4), 272–275. <https://doi.org/10.1016/j.scib.2020.01.001>.
- Liu, S., Yuan, J., Deng, W., Luo, M., Xie, Y., Liang, Q., Zou, Y., He, Z., Wu, H., Cao, Y., 2020b. High-efficiency organic solar cells with low non-radiative recombination loss and low energetic disorder. *Nat. Photonics* 14 (5), 300–305. <https://doi.org/10.1038/s41566-019-0573-5>.
- Liu, X.i., Du, X., Wang, J., Duan, C., Tang, X., Heumueller, T., Liu, G., Li, Y., Wang, Z., Wang, J., Liu, F., Li, N., Brabec, C.J., Huang, F., Cao, Y., 2018. Efficient Organic Solar Cells with Extremely High Open-Circuit Voltages and Low Voltage Losses by Suppressing Nonradiative Recombination Losses. *Adv. Energy Mater.* 8 (26), 1801699. <https://doi.org/10.1002/aenm.v8i26.1801699>.
- Lopez, S.A., Sanchez-Lengeling, B., de Goes Soares, J., Aspuru-Guzik, A., 2017. Design Principles and Top Non-Fullerene Acceptor Candidates for Organic Photovoltaics. *Joule* 1 (4), 857–870. <https://doi.org/10.1016/j.joule.2017.10.006>.
- Luo, Z., Liu, T., Wang, Y., Zhang, G., Sun, R., Chen, Z., Zhong, C., Wu, J., Chen, Y., Zhang, M., Zou, Y., Ma, W., Yan, H.e., Min, J., Li, Y., Yang, C., 2019. Reduced Energy Loss Enabled by a Chlorinated Thiophene-Fused Ending-Group Small Molecular Acceptor for Efficient Nonfullerene Organic Solar Cells with 13.6% Efficiency. *Adv. Energy Mater.* 9 (18), 1900041. <https://doi.org/10.1002/aenm.v9i18.1002/aenm.201900041>.
- Luo, Z., Sun, R., Zhong, C., Liu, T., Zhang, G., Zou, Y., Jiao, X., Min, J., Yang, C., 2020. Altering alkyl-chains branching positions for boosting the performance of small-molecule acceptors for highly efficient nonfullerene organic solar cells. *Sci. China Chem.* 63 (3), 361–369. <https://doi.org/10.1007/s11426-019-9670-2>.
- Mesta, M., Chang, J.H., Shil, S., Thygesen, K.S., Lastra, J.M.G., 2019. A Protocol for Fast Prediction of Electronic and Optical Properties of Donor-Acceptor Polymers Using Density Functional Theory and the Tight-Binding Method. *J. Phys. Chem. A* 123 (23), 4980–4989. <https://doi.org/10.1021/acs.jpca.9b02391.10.1021/acs.jpca.9b02391.s00110.1021/acs.jpca.9b02391.s002>.
- Moriwaki, H., Tian, Y.S., Kawashita, N., Takagi, T., 2018. Mordred: A molecular descriptor calculator. *J. Cheminform.* 10, 1–14. <https://doi.org/10.1186/s13321-018-0258-y>.
- Nagasawa, S., Al-Naamani, E., Saeki, A., 2018. Computer-Aided Screening of Conjugated Polymers for Organic Solar Cell: Classification by Random Forest. *J. Phys. Chem. Lett.* 9 (10), 2639–2646. <https://doi.org/10.1021/acs.jpclett.8b00635.10.1021/acs.jpclett.8b00635.s00110.1021/acs.jpclett.8b00635.s002>.
- Nakano, K., Chen, Y., Xiao, B., Han, W., Huang, J., Yoshida, H., Zhou, E., Tajima, K., 2019. Anatomy of the energetic driving force for charge generation in organic solar cells. *Nat. Commun.* 10, 1–10. <https://doi.org/10.1038/s41467-019-10434-3>.
- Nikolis, V.C., Benduhn, J., Holzmueller, F., Piersimoni, F., Lau, M., Zeika, O., Neher, D., Koerner, C., Spoltore, D., Vandewal, K., 2017. Reducing Voltage Losses in Cascade Organic Solar Cells while Maintaining High External Quantum Efficiencies. *Adv. Energy Mater.* 7 (21), 1700855. <https://doi.org/10.1002/aenm.201700855>.
- Olivares-Amaya, R., Amador-Bedolla, C., Hachmann, J., Atahan-Evrenk, S., Sánchez-Carrera, R.S., Vogt, L., Aspuru-Guzik, A., 2011. Accelerated computational discovery of high-performance materials for organic photovoltaics by means of cheminformatics. *Energy Environ. Sci.* 4, 4849–4861. <https://doi.org/10.1039/c1ee02056k>.
- Padula, D., Simpson, J.D., Troisi, A., 2019. Combining electronic and structural features in machine learning models to predict organic solar cells properties. *Mater. Horizons* 6 (2), 343–349. <https://doi.org/10.1039/C8MH01135D>.
- Padula, D., Troisi, A., 2019. Concurrent Optimization of Organic Donor-Acceptor Pairs through Machine Learning. *Adv. Energy Mater.* 9 (40), 1902463. <https://doi.org/10.1002/aenm.v9i40.1002/aenm.201902463>.
- Paul, A., Furmanchuk, A., Liao, W.-K., Choudhary, A., Agrawal, A., 2019. Property Prediction of Organic Donor Molecules for Photovoltaic Applications Using Extremely Randomized Trees. *Mol. Inform.* 38 (11–12), 1900038. <https://doi.org/10.1002/minf.v38i11-12.1002/minf.201900038>.
- 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., 2011. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Peng, S.-P., Zhao, Y.i., 2019. Convolutional Neural Networks for the Design and Analysis of Non-Fullerene Acceptors. *J. Chem. Inf. Model.* 59 (12), 4993–5001. <https://doi.org/10.1021/acs.jcim.9b00732.10.1021/acs.jcim.9b00732.s00110.1021/acs.jcim.9b00732.s002>.
- Pereira, F., Xiao, K., Latino, D.A.R.S., Wu, C., Zhang, Q., Aires-de-Sousa, J., 2017. Machine Learning Methods to Predict Density Functional Theory B3LYP Energies of HOMO and LUMO Orbitals. *J. Chem. Inf. Model.* 57 (1), 11–21. <https://doi.org/10.1021/acs.jcim.6b00340.10.1021/acs.jcim.6b00340.s001>.
- Pokuri, B.S.S., Ghosal, S., Kokate, A., Sarkar, S., Ganapathysubramanian, B., 2019. Interpretable deep learning for guided microstructure-property explorations in photovoltaics. *npj Comput. Mater.* 5, 1–11. <https://doi.org/10.1038/s41524-019-0231-y>.
- Pradhan, R., Malhotra, P., Gupta, G., Singhal, R., Sharma, G.D., Mishra, A., 2020. Efficient Fullerene-Free Organic Solar Cells Using a Coumarin-Based Wide-Band-Gap Donor Material. *ACS Appl. Mater. Interfaces* 12 (37), 41869–41876. <https://doi.org/10.1021/acsami.0c12147.10.1021/acsami.0c12147.s001>.
- Pyzer-Knapp, E.O., Simm, G.N., Aspuru Guzik, A., 2016. A Bayesian approach to calibrating high-throughput virtual screening results and application to organic photovoltaic materials. *Mater. Horizons* 3 (3), 226–233. <https://doi.org/10.1039/C5MH00282F>.
- Qian, D., Zheng, Z., Yao, H., Tress, W., Hopper, T.R., Chen, S., Li, S., Liu, J., Chen, S., Zhang, J., Liu, X.-K., Gao, B., Ouyang, L., Jin, Y., Pozina, G., Buyanova, I.A., Chen, W.M., Inganäs, O., Coropceanu, V., Bredas, J.-L., Yan, H.e., Hou, J., Zhang, F., Bakulin, A.A., Gao, F., 2018. Design rules for minimizing voltage losses in high-efficiency organic solar cells. *Nat. Mater.* 17 (8), 703–709. <https://doi.org/10.1038/s41563-018-0128-z>.
- Qin, J., Zhang, L., Zuo, C., Xiao, Z., Yuan, Y., Yang, S., Hao, F., Cheng, M., Sun, K., Bao, Q., Bin, Z., Jin, Z., Ding, L., 2021. A chlorinated copolymer donor demonstrates a 18.13% power conversion efficiency. *J. Semicond.* 42 (1), 010501. <https://doi.org/10.1088/1674-4926/42/1/010501>.
- Ran, N.A., Roland, S., Love, J.A., Savikhin, V., Takacs, C.J., Fu, Y.T., Li, H., Coropceanu, V., Liu, X., Brédas, J.L., Bazan, G.C., Toney, M.F., Neher, D.i., Nguyen, T.Q., 2017. Impact of interfacial molecular orientation on radiative recombination and charge generation efficiency. *Nat. Commun.* 8, 1–9. <https://doi.org/10.1038/s41467-017-00107-4>.
- Rau, U., 2007. Reciprocity relation between photovoltaic quantum efficiency and electroluminescent emission of solar cells. *Phys. Rev. B* 76, 085303. <https://doi.org/10.1103/PhysRevB.76.085303>.
- Rau, U., Paetzold, U.W., Kirchartz, T., 2014. Thermodynamics of light management in photovoltaic devices. *Phys. Rev. B - Condens. Matter Mater. Phys.* 90, 1–16. <https://doi.org/10.1103/PhysRevB.90.035211>.
- Riede, M., Spoltore, D., Leo, K., 2021. Organic Solar Cells—The Path to Commercial Success. *Adv. Energy Mater.* 11 (1), 2002653. <https://doi.org/10.1002/aenm.v11i1.1002/aenm.202002653>.
- Rodríguez-Martínez, X., Pascual-San-José, E., Fei, Z., Heeney, M., Guimerà, R., Campoy-Quiles, M., 2021. Predicting the photocurrent-composition dependence in organic solar cells. *Energy Environ. Sci.* 14 (2), 986–994. <https://doi.org/10.1039/D0EE02958K>.
- Rogers, D., Hahn, M., 2010. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* 50 (5), 742–754.
- Rosenthal, K.D., Hughes, M.P., Luginbuhl, B.R., Ran, N.A., Karki, A., Ko, S.-J., Hu, H., Wang, M., Ade, H., Nguyen, T.-Q., 2019. Quantifying and Understanding Voltage Losses Due to Nonradiative Recombination in Bulk Heterojunction Organic Solar Cells with Low Energetic Offsets. *Adv. Energy Mater.* 9 (27), 1901077. <https://doi.org/10.1002/aenm.v9i27.1002/aenm.201901077>.
- Sahu, H., Ma, H., 2019. Unraveling Correlations between Molecular Properties and Device Parameters of Organic Solar Cells Using Machine Learning. *J. Phys. Chem. Lett.* 10 (22), 7277–7284. <https://doi.org/10.1021/acs.jpclett.9b02772.10.1021/acs.jpclett.9b02772.s001>.
- Sahu, H., Rao, W., Troisi, A., Ma, H., 2018. Toward Predicting Efficiency of Organic Solar Cells via Machine Learning and Improved Descriptors. *Adv. Energy Mater.* 8 (24), 1801032. <https://doi.org/10.1002/aenm.v8i24.1002/aenm.201801032>.
- Sahu, H., Yang, F., Ye, X., Ma, J., Fang, W., Ma, H., 2019. Designing promising molecules for organic solar cells via machine learning assisted virtual screening. *J. Mater. Chem. A* 7 (29), 17480–17488. <https://doi.org/10.1039/C9TA04097H>.
- Saito, T., Natsuda, S.-I., Imakita, K., Tamai, Y., Ohkita, H., 2020. Role of Energy Offset in Nonradiative Voltage Loss in Organic Solar Cells. *Sol. RRL* 4 (9), 2000255. <https://doi.org/10.1002/solr.v4i9.1002/solr.202000255>.
- Steinbeck, C., Han, Y., Kuhn, S., Horlacher, O., Luttmann, E., Willighagen, E., 2003. The Chemistry Development Kit (CDK): An open-source Java library for chemo- and bioinformatics. *J. Chem. Inf. Comput. Sci.* 43, 493–500. <https://doi.org/10.1021/ci025584y>.
- Sui, M.-Y., Yang, Z.-R., Geng, Y., Sun, G.-Y., Hu, LiHong, Su, Z.-M., 2019. Nonfullerene Acceptors for Organic Photovoltaics: From Conformation Effect to Power Conversion Efficiencies Prediction. *Sol. RRL* 3 (11), 1900258. <https://doi.org/10.1002/solr.v3i11.1002/solr.201900258>.
- Sun, C., Pan, F., Chen, S., Wang, R., Sun, R., Shang, Z., Qiu, B., Min, J., Lv, M., Meng, L., Zhang, C., Xiao, M., Yang, C., Li, Y., 2019a. Achieving Fast Charge Separation and Low Nonradiative Recombination Loss by Rational Fluorination for High-Efficiency Polymer Solar Cells. *Adv. Mater.* 31 (52), 1905480. <https://doi.org/10.1002/adma.v31i52.1002/adma.201905480>.
- Sun, H., Chen, F., Chen, Z., 2019b. Recent progress on non-fullerene acceptors for organic photovoltaics. *Mater. Today* 24, 94–118. <https://doi.org/10.1016/j.mattod.2018.09.004>.
- Sun, W., Zheng, Y., Yang, K.e., Zhang, Q.i., Shah, A.A., Wu, Z., Sun, Y., Feng, L., Chen, D., Xiao, Z., Lu, S., Li, Y., Sun, K., 2019c. Machine learning-assisted molecular design and efficiency prediction for high-performance organic photovoltaic materials. *Sci. Adv.* 5 (11), eaay4275. <https://doi.org/10.1126/sciadv.aay4275>.
- Tetko, I.V., Gasteiger, J., Todeschini, R., Mauri, A., Livingstone, D., Ertl, P., Palyulin, V. A., Radchenko, E.V., Zefirov, N.S., Makarenko, A.S., Tanchuk, V.Y., Prokopenko, V. V., 2005. Virtual computational chemistry laboratory - Design and description. *J. Comput. Aided. Mol. Des.* 19 (6), 453–463. <https://doi.org/10.1007/s10822-005-8694-y>.
- Tuladhar, S.M., Azzouzi, M., Delval, F., Yao, J., Guilbert, A.A.Y., Kirchartz, T., Montcada, N.F., Domínguez, R., Langa, F., Palomares, E., Nelson, J., 2016. Low Open-Circuit Voltage Loss in Solution-Processed Small-Molecule Organic Solar Cells. *ACS Energy Lett.* 1 (1), 302–308. <https://doi.org/10.1021/acsenenergylett.6b00162.10.1021/acsenenergylett.6b00162.s001>.

- Upama, M.B., Mahmud, M.A., Conibeer, G., Uddin, A., 2020. Trendsetters in High-Efficiency Organic Solar Cells: Toward 20% Power Conversion Efficiency. *Sol. RRL* 4 (1), 1900342. <https://doi.org/10.1002/solr.v4.110.1002/solr.201900342>.
- Vandewal, K., 2016. Interfacial Charge Transfer States in Condensed Phase Systems. *Annu. Rev. Phys. Chem.* 67 (1), 113–133. <https://doi.org/10.1146/annurev-physchem-040215-112144>.
- Vandewal, K., Benduhn, J., Nikolis, V.C., 2018. How to determine optical gaps and voltage losses in organic photovoltaic materials. *Sustain. Energy Fuels* 2 (3), 538–544. <https://doi.org/10.1039/C7SE00601B>.
- Vandewal, K., Mertens, S., Benduhn, J., Liu, Q., 2020. The Cost of Converting Excitons into Free Charge Carriers in Organic Solar Cells. *J. Phys. Chem. Lett.* 11 (1), 129–135. <https://doi.org/10.1021/acs.jpcl.9b02719>.
- Veldman, D., Meskers, S.C.J., Janssen, R.A.J., 2009. The Energy of Charge-Transfer States in Electron Donor-Acceptor Blends: Insight into the Energy Losses in Organic Solar Cells. *Adv. Funct. Mater.* 19 (12), 1939–1948. <https://doi.org/10.1002/adfm.v19.12.1002/adfm.200900090>.
- Weininger, D., 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Model.* 28 (1), 31–36. <https://doi.org/10.1021/ci00057a005>.
- Wu, Y., Guo, J., Sun, R., Min, J., 2020. Machine learning for accelerating the discovery of high-performance donor/acceptor pairs in non-fullerene organic solar cells. *npj Comput. Mater.* 6, 1–8. <https://doi.org/10.1038/s41524-020-00388-2>.
- Xie, S., Xia, Y., Zheng, Z., Zhang, X., Yuan, J., Zhou, H., Zhang, Y., 2018. Effects of Nonradiative Losses at Charge Transfer States and Energetic Disorder on the Open-Circuit Voltage in Nonfullerene Organic Solar Cells. *Adv. Funct. Mater.* 28 (5), 1705659. <https://doi.org/10.1002/adfm.v28.5.1002/adfm.201705659>.
- Xie, Y., Li, T., Guo, J., Bi, P., Xue, X., Ryu, H.S., Cai, Y., Min, J., Huo, L., Hao, X., Woo, H. Y., Zhan, X., Sun, Y., 2019a. Ternary Organic Solar Cells with Small Nonradiative Recombination Loss. *ACS Energy Lett.* 4 (5), 1196–1203. <https://doi.org/10.1021/acsenenergylett.9b00681.10.1021/acsenenergylett.9b00681.s001>.
- Xie, Y., Wang, W., Huang, W., Lin, F., Li, T., Liu, S., Zhan, X., Liang, Y., Gao, C., Wu, H., Cao, Y., 2019b. Assessing the energy offset at the electron donor/acceptor interface in organic solar cells through radiative efficiency measurements. *Energy Environ. Sci.* 12 (12), 3556–3566. <https://doi.org/10.1039/C9EE02939G>.
- Xie, Y., Wu, H., 2020. Balancing charge generation and voltage loss toward efficient nonfullerene organic solar cells. *Mater. Today Adv.* 5, 100048. <https://doi.org/10.1016/j.mtadv.2019.100048>.
- Ye, L., Weng, K., Xu, J., Du, X., Chandrabose, S., Chen, K., Zhou, J., Han, G., Tan, S., Xie, Z., Yi, Y., Li, N., Liu, F., Hodgkiss, J.M., Brabec, C.J., Sun, Y., 2020. Unraveling the influence of non-fullerene acceptor molecular packing on photovoltaic performance of organic solar cells. *Nat. Commun.* 11, 6005. <https://doi.org/10.1038/s41467-020-19853-z>.
- Yu, R., Yao, H., Cui, Y., Hong, L., He, C., Hou, J., 2019. Improved Charge Transport and Reduced Nonradiative Energy Loss Enable Over 16% Efficiency in Ternary Polymer Solar Cells. *Adv. Mater.* 31 (36), 1902302. <https://doi.org/10.1002/adma.v31.36.1002/adma.201902302>.
- Yuan, J., Huang, T., Cheng, P., Zou, Y., Zhang, H., Yang, J.L., Chang, S.-Y., Zhang, Z., Huang, W., Wang, R., Meng, D., Gao, F., Yang, Y., 2019a. Enabling low voltage losses and high photocurrent in fullerene-free organic photovoltaics. *Nat. Commun.* 10, 570. <https://doi.org/10.1038/s41467-019-08386-9>.
- 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., 2019b. Single-Junction Organic Solar Cell with over 15% Efficiency Using Fused-Ring Acceptor with Electron-Deficient Core. *Joule* 3 (4), 1140–1151. <https://doi.org/10.1016/j.joule.2019.01.004>.
- Zhang, G., Chen, X.K., Xiao, J., Chow, P.C.Y., Ren, M., Kupgan, G., Jiao, X., Chan, C.C.S., Du, X., Xia, R., Chen, Z., Yuan, J., Zhang, Y., Zhang, S., Liu, Y., Zou, Y., Yan, H., Wong, K.S., Coropceanu, V., Li, N., Brabec, C.J., Bredas, J.L., Yip, H.L., Cao, Y., 2020. Delocalization of exciton and electron wavefunction in non-fullerene acceptor molecules enables efficient organic solar cells. *Nat. Commun.* 11, 1–10. <https://doi.org/10.1038/s41467-020-17867-1>.
- Zhang, J., Liu, W., Zhang, M., Liu, Y., Zhou, G., Xu, S., Zhang, F., Zhu, H., Liu, F., Zhu, X., 2019. Revealing the Critical Role of the HOMO Alignment on Maximizing Current Extraction and Suppressing Energy Loss in Organic. *Solar Cells. iScience* 19, 883–893. <https://doi.org/10.1016/j.isci.2019.08.038>.
- Zhao, Z.-W., del Cueto, M., Geng, Y., Troisi, A., 2020. Effect of Increasing the Descriptor Set on Machine Learning Prediction of Small Molecule-Based Organic Solar Cells. *Chem. Mater.* 32 (18), 7777–7787. <https://doi.org/10.1021/acs.chemmater.0c02325.10.1021/acs.chemmater.0c02325.s001>.
- Zhou, R., Jiang, Z., Yang, C., Yu, J., Feng, J., Adil, M.A., Deng, D., Zou, W., Zhang, J., Lu, K., Ma, W., Gao, F., Wei, Z., 2019. All-small-molecule organic solar cells with over 14% efficiency by optimizing hierarchical morphologies. *Nat. Commun.* 10, 1–9. <https://doi.org/10.1038/s41467-019-13292-1>.