Machine Learning based Framework for Quick Prediction of T_g and T_d of OLED Materials

Yihuan Zhao [1], Caixia Fu [1], Lin Fu[1], Zhiyun Lu [1,*], Xuemei Pu [1,*]

[1] *College of Chemistry, Sichuan University, Chengdu 610064, People's Republic of China*

*Corresponding author: Prof. Z.Y. Lu, Prof. X. M. Pu

Email address: luzhiyun@scu.edu.cn, xmpuscu@scu.edu.cn

Tel.: +86-028-85412290

Fax: +86-028-85412290

Abstract

Organic light-emitting-diode (OLED) materials have exhibited a wide range of applications. However, further development and commercialization of OLEDs requires higher-quality OLED materials, including high thermal stability associated with the glass transition temperature (T_g) and decomposition temperature (T_d). Experimental determinations of the two important properties generally involve a time-consuming and laborious process. Thus, it is highly desired to develop a quick and accurate prediction tool. Motivated by the changelle, we explored machine learning based framework by constructing new dataset with more than one thousand samples collected from a wide range of literaturesm, throngh which ensemble learning models were explored. Models trained with the LightGBM algorithm exhibit the best prediction performance, where the values of MAE, RMSE, and R^2 are 17.15 K, 24.63 K, and 0.77 for T_g prediction, 24.91 K, 33.88 K, and 0.78 for T_d prediction. The prediction performance and the generalization of the machine learning models are further tested by out-of-sample dataset, also exhibiting satisfactory results. Experimental verification further demonstrates the reliability and the practical potential of the ML-based model. In order to extend the practice application of the ML-based models, an online prediction platform was constructed, including the optimal predition models and all the thermal stability data under study, which are freely available at http://oledtppxmpugroup.com. We expect that they will become a useful tool for experimental investigations on T_g and T_d, in turn accelerating the design of the OLED materials with high performance.

Keywords: OLED materials; thermal stability; machine learning

1. Introduction

Organic light-emitting diodes (OLEDs) have attracted considerable attentions in recent years due to their great promises in flat-panel displays, solid-state lighting, and white lighting technologies.[1-3] The commercialization of OLEDs requires high quality OLED devices, in particular for a long lifetime.[4] There are intrinsic and extrinsic factors that affect the lifetime of OLED devices.[5, 6] One of the main external factors is temperature. As known, the temperature of OLED devices can increase due to Joule heating during operation and exposure to high-temperature external environments.[7] Accordingly, enhancement in the thermal stability of OLED materials can improve the device performance. Recently, a large number of researchers have paid attention to the OLED materials with thermal stability.[8]

The glass transition temperature (T_g) and decomposition temperature (T_d, corresponding to 5% weight loss) are the two most important thermal properties for the OLED materials, and closely associated with the performance of OLED devices.[9] In particular, T_g of the OLED materials is one of the most important factors that influences device stability and lifetime since the OLED devices irreversibly deteriorate when heated above their T_g.[7, 10] High T_g and T_d values can reduce heat-induced morphology changes, thus enhancing the stability of device performance.[11, 12] On experiments, the T_g and T_d values of the OLED materials are generally measured by differential scanning calorimetry (DSC) and thermal gravimetric analysis (TGA). However, before DSC and TGA determination, OLED materials need to be purified by column chromatography or sublimation,[13, 14] which are complicated and time-consuming. Thus, it is highly

desired to develop a quick and accurate method to predict T_g and T_d. As accepted, the thermal stability of the OLED material is closely related to their molecule structures.[15] However, the relationship between the molecular structure and the thermal stability like T_g and T_d is complex and has not been elucidated so far. Machine learning (ML), as a key technique of artificial intelligence, can map the complex relationshiop underlying a large amoun of data, which has been succesfully applied in the fields of medicinal chemistry, environmental risk assessment, organic synthesis, and materials science.[16-20,21] To our best knowledge, only two previous studies already used the machine learning method to predict the T_g of OLED materials,[22, 23] but there has been lack of T_d prediction. In 2003, Yin et al. performed a study on the quantitative structure-property relationship (QSPR) based on the T_g of 88 OLED materials with R^2= 0.963, MAE = 17.9 K for test set, using a multilinear regression method (MLR)[23]. In 2013, Silva et al. developed a QSPR model to predict the T_g of 66 OLED materials with R^2 = 0.963, MAE = 0.97 K for test set (not independent test set) by using support vector machines (SVM) [24]. The two studies look like high prediction accucracy, which comes from so small amount of data (fewer than 100 molecules). In fact, the prediction ability of the single machine learning models in the two previous studies is not reliable and unstable, in particular for its generalization on unknow compounds due to the dependence of ML on the dataset size. Unfortunately, there has been lack of the database involving the two important properties T_g and T_d so far. However, in the past decade, a significant amount of thermal stability data for OLED materials has been published. In particular, a large amount of thermal stability data has been reported for thermally activated delayed

fluorescence (TADF) materials, which are pure organic molecules showing a potential 100% internal quantum efficiency without the aid of heavy metals.[24] Desipte these published data dispersive in different literatures, they still provide a possible data source for constructing robust machine learning model.[25]

Motivated by the challenge, we construct a new dataset containing the experimental T_g data of 1944 small organic molecules and the experimental T_d data of 1182 small organic OLED compounds collected from a large amount of literature. Based on the new dataset, we ulitized ensemble learning approach LightGBM algorithm, rather than *single machine learning methods* in previous works, to build comprehensive models between molecular structure and the two properties by constructing and combining base learners. The optimal T_g prediction model could provide an accurate prediction with RMSE = 24.63 K, MAE = 17.15 K, and R^2 = 0.77 for the test set. For T_d prediction, the optimal model could provide an accurate prediction with RMSE = 33.88 K, MAE = 24.91 K, and R^2 = 0.78 for the test set. In addition, the optimal models could accurately predict the T_g and T_d for out of sample including recently reported OLED host and guest materials, organic electron-transport materials and hole-transport materials. Then, we used the optimized models to predict the T_g and T_d of 50 unknown OLED molecules designed by us and selected the compound TPA-2 (with high T_g and the highest T_d) to experimentally synthesis and determine its thermal stability. The experimental verification further confirm the prediction reliability of our ML models. In addition, we explore a websit including the optimized T_g and T_d prediction models coupled with the new dataset, which is freely available at http://oledtppxmpugroup.com. We expect that

the website will serve as a useful tool to help experimental investigators quickly estimate T_g and T_d.

2. Materials and methods

2.1 Construction of dataset

Unfortunately, there is no existing database that organizes OLED materials and their properties. Currently, the thermal stability data of OLED materials are scattered throughout the literature. Therefore, experimental glass transition temperatures (T_g) for a diverse set of 1944 molecules were collected from a large number of literature using the SciFinder database. These T_g values were measured by DSC. For molecules with multiple recorded entries, an average T_g was used as the output if the variation was less than 40 K. Molecules with a T_g variation larger than 40 K were not included in the dataset. Experimental thermal decomposition temperatures (T_d, corresponding to 5% weight loss) for a diverse set of 1182 OLED molecules also were collected from the literature. The T_d data for these OLED molecules were measured by TGA. For molecules with multiple recorded entries, the highest T_d was used for the dataset. Table S1 lists 13 OLED compounds with multiple recorded entries. As can be seen, the deviation in T_g for compounds reported in different papers is often within 40 K. However, the T_d values of the same compounds reported in different papers have a larger deviation, often greater than 40 K. This is because the purity of a compound has a significant influence on the experimental value of T_d. Even if the purity of a compound is as high as 95%, if the T_d values of the impurities in the remaining 5% are low, the measured experimental T_d value will also be lower. Therefore, a higher compound purity can result in a higher experimentally measured T_d value. It should be noted that

there are very few compounds with multiple T_d records in our dataset.

The T_g distribution of 1944 molecules and T_d distribution of 1182 OLED molecules are shown in Fig. 1. The T_g values span a range from 273 to 600 K, with the majority of values between 325 and 475 K. Fig. 1 shows that compounds with a T_g higher than 400 K (127 °C) account for almost half of the reported compounds. The T_d values span a range from 400 to 900 K, with the majority of values between 550 and 800 K. This indicates that the T_d of most OLED materials is greater than 550 K (277 °C). The molecular structures include the atomic elements C, H, B, N, O, F, Si, P, S, Cl, and Br, containing most of the elements of pure organic functional materials.

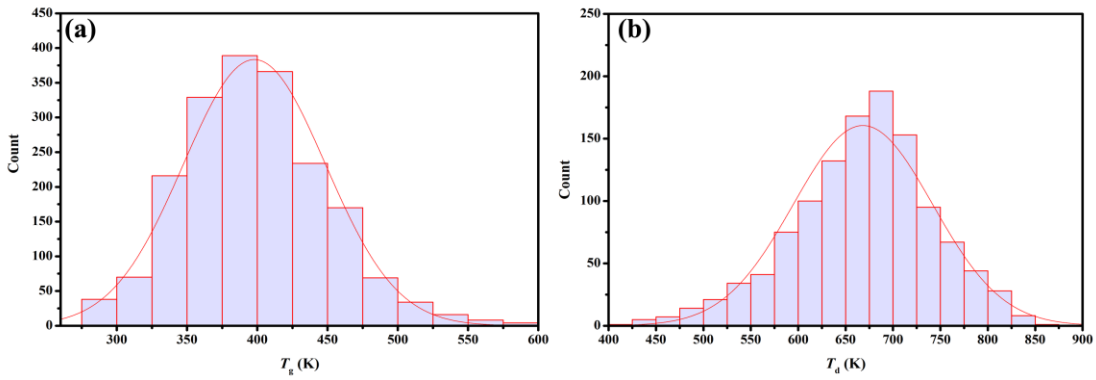


Fig. 1 (a) T_g distribution of 1944 molecules. (b) T_d distribution of 1182 OLED molecules.

2.2 Descriptors and fingerprints

In the work, molecular descriptors and fingerprints were considered to characterize the molecular structure. Molecular descriptors and fingerprints were calculated by PaDEL-Descriptor version 2.21.[26] The 1D and 2D descriptors and molecular fingerprints were generated by taking into consideration their general applicability as well as their computation cost. The PaDEL-Descriptor software is open source and free, and the calculation of 1D and 2D descriptors and molecular fingerprints is simple and fast. This facilitates the further promotion and use of our thermal stability prediction models.

2.2.1 Molecular descriptors

1D molecular descriptors were generated based on molecular formulas and 2D molecular descriptors were generated based on the atom connection table. 1D and 2D molecular descriptors belong to the class of molecular property-based descriptors. Each molecular descriptor represents a certain feature of a molecule, such as topology or weight. As each molecular descriptor only depicts a specific property of a molecule, a combination of a large number of molecular descriptors can provide more information. Using information encoded in canonical SMILES (simplified molecular input line entry system), the PaDEL software offered 1444 1D and 2D descriptors. However, not all the descriptors were used for modeling, such as the descriptors which were not computable for all the compounds. The remaining 665 parameters were used for model definition (including aromatic atom count, aromatic bond count, atom count, bond count, estate atom type, extended topochemical atom indices, ring count, topology, topological charge, topological distance matrices, topological polar surface area, XLogP, and weight descriptors).

2.2.2 Molecular fingerprints

Five types of fingerprints (a total of 2741 parameters) were calculated for this research, including CDK fingerprints (1024 bits), CDK extended fingerprints (1024 bits), E-States fingerprints (79 bits), substructure fingerprints (307 bits), and substructure fingerprints count (307 bits). Molecular fingerprints are a subclass of molecular descriptors that can be obtained without quantum-mechanical calculations. They belong to the class of fragment-based descriptors,[27] and they were used in this study due to their high potential for the high-throughput screening of materials. These fragment-

based descriptors are represented as a Boolean array, indicating the existence of the corresponding fragments in the molecule. The descriptions of the molecular fingerprints used in this study are listed in Table S2. CDK fingerprints, CDK extended fingerprints, and E-state fingerprints are a good expression of the molecular backbones. Substructure fingerprints and substructure fingerprints count provide differentiation for an array of functional groups.

2.3 Machine learning algorithms

LightGBM is a recent modification of the gradient boosting (GB) algorithm.[28] It improves the efficiency and scalability of the GB algorithm without sacrificing its inherited effective performance. This approach results in a faster and less resource-intensive implementation of gradient boosting, suitable for frequent retraining and rapid evaluation of larger high-dimensional datasets. The LightGBM algorithm uses ensemble learning, which aims to build a comprehensive model by constructing and combining base learners. Ensemble learning not only produces a more stable global model, but also guarantees diminishing uncertainty. In particular, LightGBM uses Exclusive Feature Bundling (EFB) to reduce feature dimensions and improve calculation efficiency, and it can handle a large number of data instances and a large number of data features without overfitting problems.[28] As a decision tree based model, LightGBM has an additional advantage of being robust against multicollinearity. In this work, we first used the LightGBM algorithm for modeling and then compared its model with other algorithms. These include Support Vector Machine (SVM), partial least squares (PLS), least absolute shrinkage and selection operator (LASSO), Kernel Ridge

Regression (KRR), k-Nearest Neighbors (kNN), Random Forest (RF). All except LightGBM can be found in Scikit-learn package.

2.4 Experimental synthesis

1,2-bis(4-(diphenylamino)phenyl)ethane-1,2-dione (1)

Anhydrous aluminum trichloride (1.33 g, 9.8 mmol), oxalyl chloride (0.43 mL，4.9 mmol) and triphenylamine (2.94 g 12 mmol) were dissolved in anhydrous CH_2Cl_2 (10 mL) and refluxed at 40 ^oC for 2 h. After cooled down to room temperature and poured into ice water. Hydrochloric acid (10 mL, 37%) was added and the mixture was extracted with CH_2Cl_2 (30 mL × 2). The combined organic layers were washed with water (30 mL × 3) and dried over anhydrous Na_2SO_4. After removing the solvent, the residue was purified using column chromatography on silica gel employing CH_2Cl_2/PE (1/1) as an eluent to afford a yellow solid with a yield of 28%. ^{1}H NMR (400 MHz, $CDCl_3$) : δ = 7.77 (d, J = 8.8 Hz, 1H), 7.33 (t, J = 8.0 Hz, 2H), 7.16 (m, 3H), 6.95 ppm (d, J = 8.8 Hz, 1H).

2,3-bis(4-(diphenylamino)phenyl)naphtho[2,3-f]quinoxaline-7,12-dione (TPA-2)

1 (500 mg, 0.92 mmol) and 1,2-diaminoanthraquinone (220 mg, 0.912 mmol) was dissolved in AcOH (15 mL) and heated to 120 °C and stirred for 12 h. After cooled down to room temperature and then poured into 30 mL water and extracted with CH_2Cl_2 (20 mL × 3). The resultant organic phase was washed with brine, and dried over anhydrous Na_2SO_4. After removing the solvent, the residue was purified using column chromatography on silica gel employing CH_2Cl_2/PE (1/1) as an eluent to give a red solid with a yield of 37%. ^{1}H NMR (400 MHz, $CDCl_3$) : δ = 8.59 (d, J = 8.8 Hz, 1H),

8.45 (d, J = 8.8 Hz, 1H), 8.32 (d, J = 7.6 Hz, 1H), 8.30 (d, J = 8.8 Hz, 1H), 7.77 (m, 4H), 7.64 (d, J = 8.4 Hz, 2H), 7.17 (m, 8H), 7.07 ppm (m, 16H). ^{13}C NMR(100 MHz, $CDCl_3$) δ(ppm):183.8, 183.5, 155.3, 153.8, 149.5, 147.1, 147.0, 143.3, 138.4, 135.2, 134.9, 134.7, 134.5, 133.5, 132.3, 131.4, 131.4, 130.9, 130.8, 129.5, 129.4, 129.1, 127.4, 126.6, 126.2, 125.4, 123.9, 123.8, 121.5, 121.4.

3. Results and discussions

3.1 Machine learning models for T_g and T_d

90% of the T_g and T_d dataset was used for model training and the remaining 10% was used for an independent test set. In order to establish robust machine learning models to predict the thermal stability of OLED materials, 10-fold cross-validation was used to reduce the randomness of sample division and enhance the stability of the obtained machine learning models. Performance was measured with the squared correlation coefficient (R^2), the mean absolute error (MAE), and the root mean squared error (RMSE).

Selecting suitable descriptors is crucial for T_g and T_d prediction tasks. We started with the choice of molecular fingerprints. The LightGBM algorithm was used to evaluate their prediction performance. A potential challenge exists due to the multifold molecular features involved in the thermal stability of OLED materials, because a single molecular fingerprint does not cover all of these features. However, combining different molecular fingerprints may solve this problem. Table S3 and Table S4 show the training and testing results of different T_g and T_d prediction machine learning models with different fingerprints as inputs. Joint fingerprints including CDK fingerprints (1024 bits), CDK extended fingerprints (1024 bits), and substructure fingerprints count (307

bits) show the best performance, implying that the representation of molecular structures by the molecular backbone and functional groups is potentially better for T_g and T_d prediction than the use of other fingerprints. Therefore, the three molecular fingerprints (CDK, CDK extended, substructure count, 2355 bits) were combined as an input, denoted SC_2CDK.

Table 1 summarizes the T_g and T_d prediction results of the LightGBM models. As can be seen, the machine learning model with 1D and 2D molecular descriptors has better T_g prediction performance than the corresponding machine learning model with fingerprints. Therefore, the 1D and 2D molecular descriptors provide more important information relevant for T_g prediction compared with fingerprints. However, information contained in property-based descriptors (molecular descriptors) and fragment-based descriptors (fingerprints) can complement each other.[27] Table 1 shows the performance improvement achieved by combining molecular descriptors and fingerprints for T_g prediction. The best T_g prediction result was obtained with RMSE = 24.63 K, MAE = 17.15 K, and R^2=0.77 for the independent test set. A plot of T_g values predicted by the best model vs. the experimental T_g values for independent test set cross-validation is shown in Fig. 2a. A reasonable agreement between the predicted and experimental T_g values can be seen. Considering that the T_g values of the 1944 organic molecules in the dataset are mainly distributed between 325 K and 475 K, RMSE = 24.63 K and MAE = 17.15 K are acceptable values.

Table 1 shows that machine learning model with fingerprints has better T_d prediction performance than the corresponding machine learning model with 1D and 2D molecular

descriptors, indicating that fingerprints (fragment-based descriptors) can provide more important information relevant to T_d compared with 1D and 2D molecular descriptors (property-based descriptors). This is because the thermal decomposition of OLED materials often starts at a specific molecular fragment, usually the weak bond in a functional group. Therefore, fragment-based descriptors can provide more important information relevant to T_d prediction. The combination of molecular descriptors and SC_2CDK only slightly improves the performance of T_d prediction. The best result for T_d prediction was obtained with RMSE = 33.88 K, MAE = 24.91 K, and R^2 = 0.78 for the independent test set. A plot of T_d values predicted by the best model vs. the experimental T_d values for the independent test set is shown in Fig. 2b, confirming that the predicted and experimental T_d values are reasonably consistent. Considering that the T_d values of the 1182 OLED molecules in the dataset are mainly distributed between 550 K and 800 K, RMSE = 33.88 K and MAE = 24.91 K are acceptable values. Therefore, our T_g and T_d LightGBM prediction models are sufficient for applications such as pre-screening of high T_g and T_d OLED materials.

The LightGBM prediction results reveal the most relative features with pronounced effects on the predicted T_g and T_d in the dataset. Fig. 4 shows the feature importance for predictions of T_g and T_d obtained with the LightGBM-based optimal models. The 10 most important features for T_g prediction are maxwHBa, JGI5, JGI9, JGI4, hmax, JGI10, VE1_D, ETA_dBetaP, JGI6, and ETA_EtaP_F. The feature with the highest contribution is maxwHBa (maximum E-States for weak hydrogen bond acceptors), indicating that hydrogen bonds have an effect on T_g. T_g is a reversible transition in

amorphous materials which allows rapid molecular motion under heating. The presence of hydrogen bonds can affect T_g because they change the rigidity of molecules and play an important role in preventing molecular rotation. In addition, the topological charge (JGI4, JGI5, JGI6, JGI9, JGI0) and topological distance (VE1_D) of molecules have a significant influence on T_g. ETA_EtaP_F (the functionality index EtaF relative to molecular size), ETA_dBetaP (a measure of the degree of unsaturation relative to molecular size), and hmax (maximum H E-State) also play a role in influencing T_g, indicating that molecular size and the presence of H atoms affect the T_g of molecules.

The top 10 most important features for T_d are minaaCH, SpMAD_D, JGI10, JGI8, maxaaCH, JGI7, ETA_Psi_1, JGI9, JGI4, and XLogP. The features of the highest contributing factor (minaaCH: minimum atom-type E-State: :CH:) and fifth-highest contributing factor (maxaaCH: maximum atom-type E-State: :CH:) indicate that the atom-type :CH: has a significant influence on the value of T_d. The topological charge (JGI4, JGI7, JGI8, JGI9, JGI10) and topological distance (SpMAD_D) of molecules also have a significant influence on T_d. ETA_Psi_1 (a measure of molecular hydrogen bonding propensity and/or polar surface area) indicates that hydrogen bonds also influence T_d. Descriptions of the 10 most important features for T_g and T_d prediction are shown in Table S5 and Table S6.

In addition to the LightGBM algorithm, the performance and efficiency of other models including SVM, PLS, LASSO, KRR, kNN, and RF algorithms were examined. A comparison of the predictive powers of these seven machine learning methods was undertaken based on the MAE and RMSE of T_g and T_d prediction (where the input was

descriptors + SC_2CDK). The MAE and RMSE of the independent test set for the different machine learning methods are shown in Fig. 5. As can be seen, the LightGBM regressor exhibits the lowest MAE and RMSE for T_g and T_d prediction. Based on these results, the LightGBM algorithm was selected as the optimal algorithm for thermal stability prediction of OLED materials.

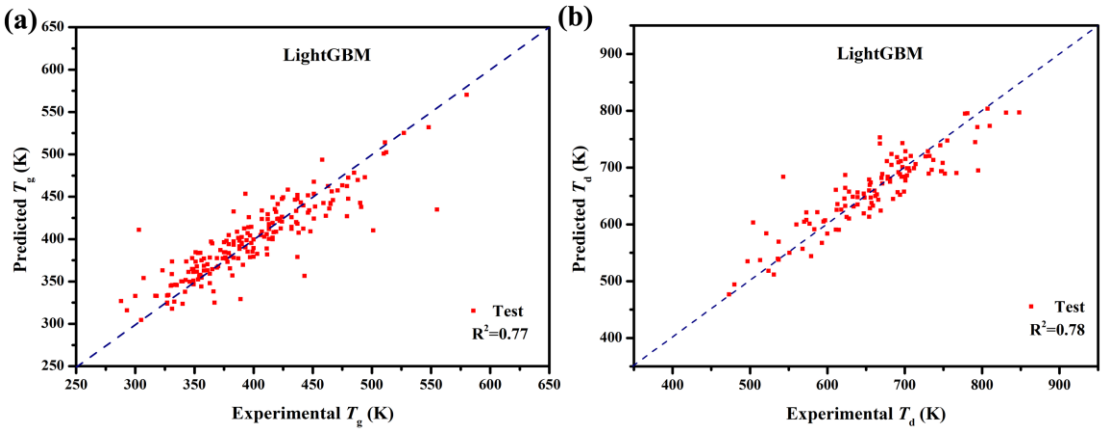


Fig. 2 Correlation plots of (a) T_g and (b) T_d for the independent test set based on LightGBM models.

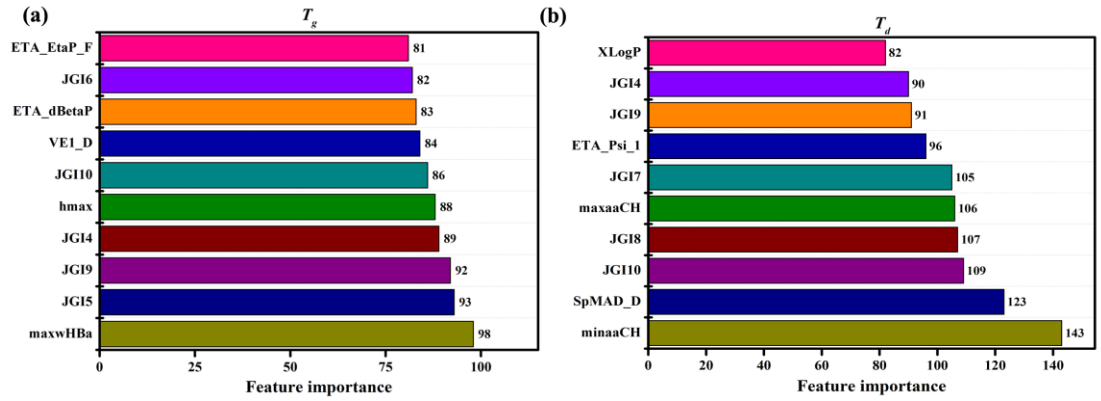


Fig. 3 Feature importance for the prediction of (a) T_g and (b) T_d obtained from the LightGBM models.

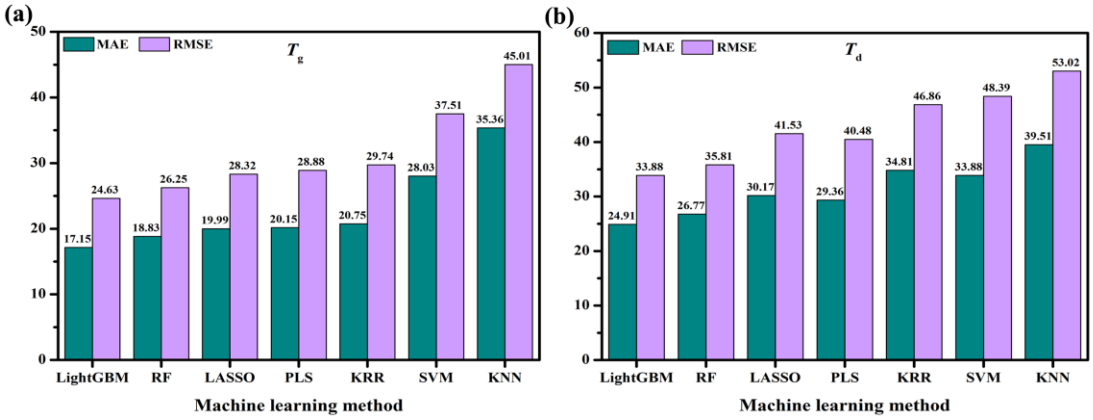


Fig. 4 MAE and RMSE for the prediction of (a) T_g and (b) T_d of the independent test set with different machine learning methods.

Table 1. Prediction summary for the T_g and T_d of OLED materials based on LightGBM models.

		Training			Testing		
	Input	R^2	MAE (K)	RMSE (K)	R^2	MAE (K)	RMSE (K)
T_g	Descriptors	0.99	1.54	2.59	0.74	17.73	26.15
	SC_2CDK	0.97	5.13	8.10	0.72	17.62	26.72
	Descriptors + SC_2CDK	0.99	1.34	2.34	0.77	17.15	24.63
T_d	Descriptors	0.99	0.31	1.05	0.75	27.92	36.66
	SC_2CDK	0.99	5.06	8.26	0.78	26.46	34.19
	Descriptors + SC_2CDK	0.99	1.26	3.26	0.78	24.91	33.88

3.2 Model application and verification

The goal of machine learning model construction is enabling the use of the model in practical applications. The obtained T_g and T_d prediction models based on the LightGBM algorithm were further tested in out-of-sample predictions and experimental verification. Three representative applications are shown herein.

3.2.1 Independent testing for T_g and T_d predictions of host and guest materials

To verify the effectiveness of the machine learning models, they were further tested in out-of-sample predictions. The optimal models were applied to the prediction of T_g for 40 OLED host and guest compounds and the prediction of T_d for 40 OLED host and guest compounds reported in recent literature.[14, 29-56] More detailed information about

these compounds can be found in the supporting information (Table S7 and Table S8). These compounds were not included in the original dataset. A plot of the predicted T_g and T_d values vs. the experimental T_g and T_d values is shown in Fig. 5. The R^2, MAE, and RMSE of the T_g predictions are 0.89, 8.81 K, and 11.15 K, respectively. The R^2, MAE, and RMSE of the T_d predictions are 0.82, 14.95 K, and 20.00 K, respectively. These results show that our models can accurately predict the T_g and T_d of out-of-sample OLED compounds. However, one compound demonstrated a very large T_d prediction error (3CzCNPyz, with an error of 75.00 K).

In order to clarify the reasons for this large prediction error, 3CzCNPyz can be compared with two other compounds that appear in the literature.[51] The TGA curve of 3CzCNPyz is shown in Fig. 6 and the TGA curves of 2Cz2CNPyz and 4CzPyz are shown in Fig. S1. The compounds 2Cz2CNPyz and 4CzPyz have prediction errors of 2.81 K and -1.79 K, much smaller than the prediction error of compound 3CzCNPyz. Fig. 6 shows that the weight of compound 3CzCNPyz slightly decreases below 500 K (227 °C). This can be attributed to the presence of impurities in 3CzCNPyz which cause the experimental T_d value of 3CzCNPyz to decrease, as discussed in section 2.1. To investigate if the T_d prediction model has problems with inaccurate predictions, the compound 4CzCNPy was selected for T_d prediction. 4CzCNPy is a compound similar in structure to 3CzCNPyz. In Ref. [57] and Ref. [58], the T_d of 4CzCNPy was tested to be 712 K and 681 K. The T_d prediction of 4CzCNPy by our machine learning model is 708, showing an error of -4 K and 27 K. This is in good agreement with the experimental values. Therefore, it is likely that our model is accurate for the T_d

prediction of 3CzCNPyz. This example demonstrates that the purity of a compound must be high for the accurate measurement of T_d. Because the experimental determination of T_g and T_d requires high purity OLED compounds and is time-consuming and labor-intensive, the prediction of T_g and T_d based on a machine learning approach is much more convenient.

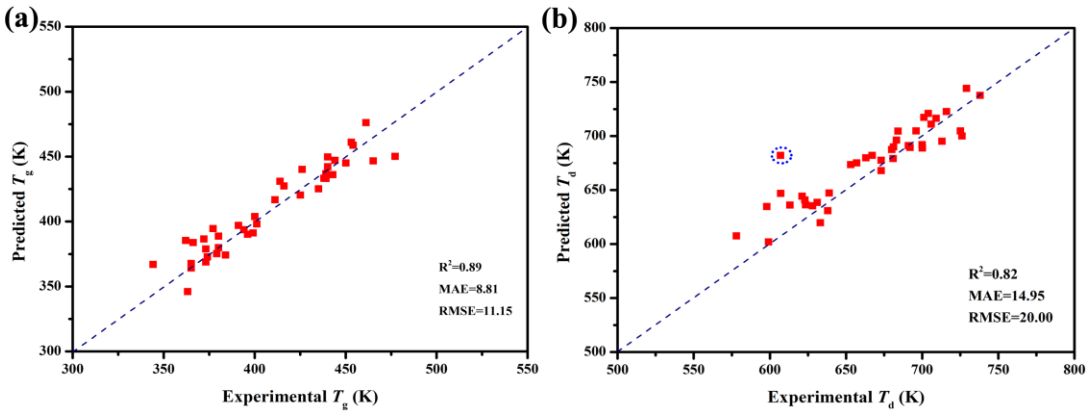


Fig. 5 Correlation plots of (a) predicted and experimental T_g values of 40 OLED compounds and (b) predicted and experimental T_d values of 40 OLED compounds in recent literature.

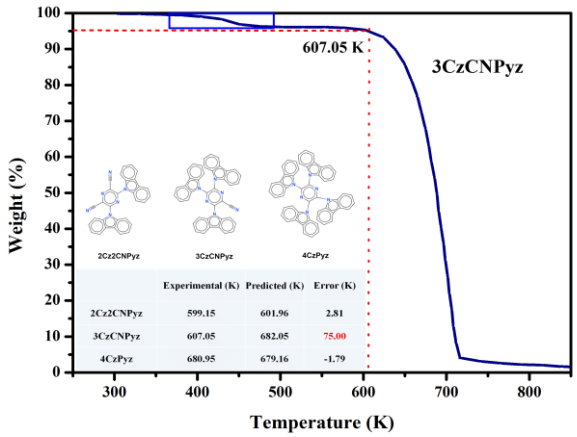


Fig. 6 TGA curve of 3CzCNPyz. Inset: experimental and predicted T_d values of 3CzCNPyz, 2Cz2CNPyz, and 4CzPyz.

3.2.2 Independent testing of T_g and T_d predictions for hole-transport materials and electron-transport materials

Organic electron-transport materials (ETMs) and hole-transport materials (HTMs) are widely used in OLEDs and perovskite solar cells, and both materials are required to have a high thermal stability. Realizing accurate T_g and T_d prediction of organic ETMs

and HTMs prior to experimental synthesis will be useful for the development of ETMs and HTMs with expected properties. Therefore, to verify the practicality of this study's models, the models were used for T_g prediction of 40 organic ETMs and HTMs and T_d prediction of 40 organic ETMs and HTMs used in OLEDs and solar cells.[59-87] More detailed information on these organic ETMs and HTMs can be found in Table S9 and Table S10. These compounds were not included in the original dataset.

A plot of predicted vs. experimental T_g values for 40 organic ETMs and HTMs is shown in Fig. 7a. A reasonable agreement exists between the predicted and experimental T_g values. The R^2, MAE, and RMSE of the T_g predictions are 0.76, 14.39 K, and 16.18 K. Fig. 7b shows a plot of predicted vs. experimental T_d values for 40 organic ETMs and HTMs. The predicted and experimental T_d values also show reasonable agreement. The R^2, MAE, and RMSE of the T_d predictions are 0.71, 17.27 K, and 21.26 K. These results show that the optimal models can give satisifactory accuracy for the prediction of T_g and T_d of small-molecule organic ETMs and HTMs. In the future, other machine learning models (such as an electron mobility prediction model) could be combined with these models for the high-throughput screening of small-molecule organic functional materials with expected properties.

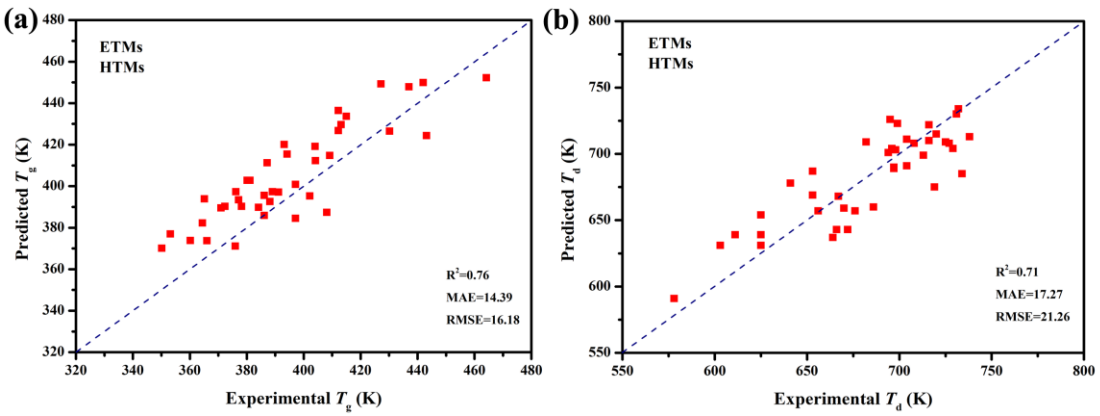


Fig. 7 Predicted vs. experimental (a) T_g and (b) T_d values for 40 organic electron-transport materials and hole-transport materials used in OLEDs and perovskite solar cells.

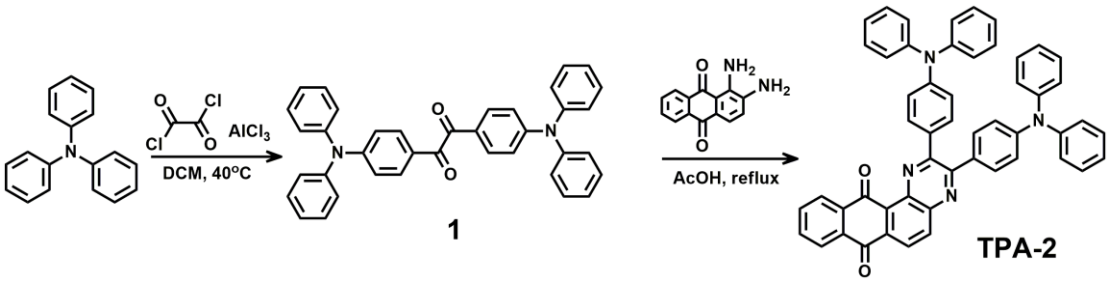
3.2.3 Experimental verification

In order to verify that these models can be used for the T_g and T_d prediction of unknown OLED compounds and for the screening of OLED materials with high thermal stability, the models were used to predict the T_g and T_d of 50 molecules. These 50 molecules were designed to have donor-acceptor (D-A) and donor-acceptor-donor (D-A-D) type structures. D-A and D-A-D type structures are very common in OLED materials and are also convenient for subsequent synthesis studies. The chemical structures and predicted T_g and T_d values of these designed compounds are shown in Table S11. The compound with the third highest T_g and the highest T_d (TPA-2) was selected for synthesis verification.

Density functional theory (DFT) simulations were performed for TPA-2 before the compound was synthesized. HOMO/LUMO distributions of TPA-2 in the ground state are shown in Fig. S2. The LUMO of TPA-2 is predominantly located on the acceptor, whereas the HOMO is located on the donor. The separated frontier molecular orbitals lead to extremely small theoretical ΔE_{ST} values for TPA-2 calculated using time-dependent density functional theory (TDDFT). The theoretical calculation parameters of TPA-2 were compared with TPA-PZCN, which is a high efficiency red-TADF material with an external quantum efficiency close to 30%.[88] The theoretical calculation parameters of TPA-PZCN and TPA-2 are shown in Table S11. TPA-2 has a narrower bandgap (E_{gap}) than TPA-PZCN (2.08 eV vs. 2.32 eV). The calculated S_1 of TPA-2 is also smaller than that of TPA-PZCN. This reveals that TPA-2 may show a longer

emission wavelength than TPA-PZCN in the same solvent. The ΔE_{ST} of TPA-2 (0.22 eV) is smaller than that of TPA-PZCN (0.25 eV). The SOC was also calculated between S_1 and T_1 in the geometry of T_1. The $<S_1|Hso|T_1>$ of TPA-2 (0.27 cm^{-1}) is larger than that of TPA-PZCN (0.13 cm^{-1}), indicating that TPA-2 has a good $T_1 \rightarrow S_1$ reverse intersystem crossing (RISC) efficiency. A large oscillator strength (0.1886) of TPA-2 is maintained which benefit radiative transition from S_1 to S_0. On the basis of these calculation results, TPA-2 is a good candidate for a red-TADF material, providing a further reason for its selection for synthesis verification.

The chemical structure and synthetic route of TPA-2 are presented in Scheme 1. Before testing, the compound was purified by column chromatography and temperature-gradient vacuum sublimation. The structure of TPA-2 was characterized by ^{1}H NMR and ^{13}C NMR (Fig. S3, Fig. S4 and Fig. S5). Experimental test results show that TPA-2 can be used as a red-TADF OLED material (Fig. S6 and Table S12). The thermal properties of TPA-2 were investigated by differential scanning calorimetry (DSC) and thermogravimetric analysis (TGA) under a nitrogen atmosphere. A (T_g) of 411 K (138 °C) and (T_d) of 697 K (424 °C) were observed (Fig. 8), in good agreement with the predicted values by machine learning. The predicted T_g value is 426 K, demonstrating an error of 15 K, while the predicted T_d value is 738 K, demonstrating an error of 41 K. As expected, the TPA-2 compound has good thermal stability. These results show that it is feasible to apply our machine learning models to predict the thermal stability of unknown OLED materials. Our machine learning models also have the potential to screen high thermal stability OLED materials.



Scheme 1 Chemical structure and synthetic route of TPA-2.

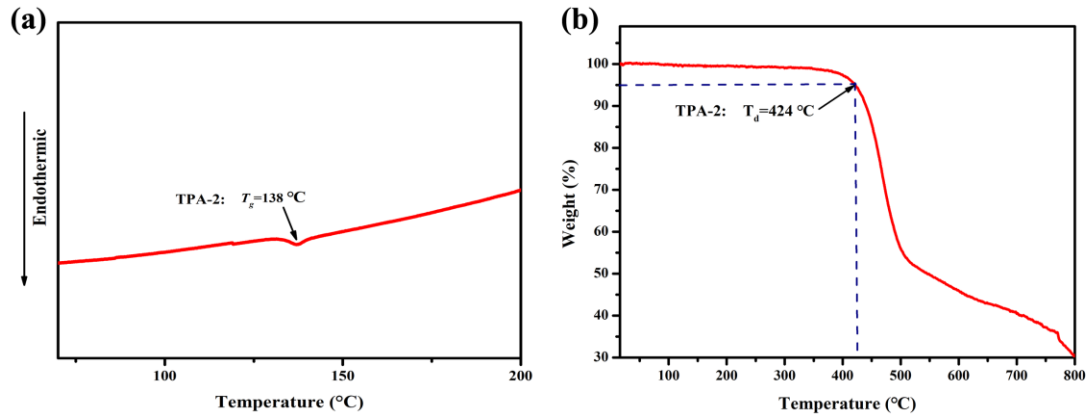


Fig. 8 DSC and TGA curves for TPA-2.

3.3 Website for T_g and T_d Predictions

Currently, hundreds of articles on OLED materials are published every year [8]. There are a lot of useful data in the literatures, but there is no existing database that organizes OLED materials data. With the aim of storing the thermal stability data of OLED materials and helping experimental scientists to utilize the model in designing new OLED compounds with desired T_g and T_d, an online tool was developed. The website to share the available models for the prediction of T_g and T_d is accessible at http://oledtppxmpugroup.com. Users can make predictions by inputting canonical SMILES, the outputs include T_g (K) and T_d (K). The T_g and T_d data in this article are also placed on this website. In the future, we will continue updating the dataset and optimal model on the website in order to predict T_g and T_d more accurately. The screenshot of the interface our website homepage are shown in Fig. 9. More details can

be found by visiting the website.

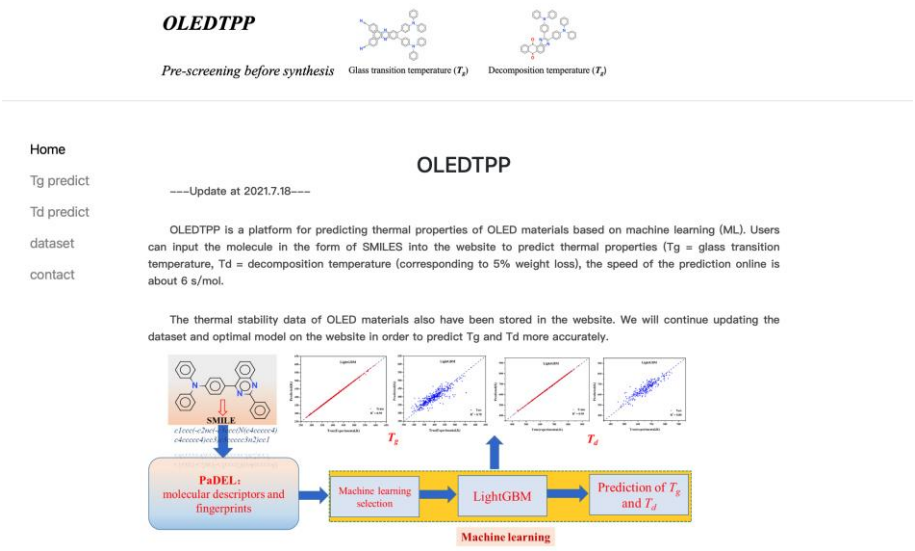


Fig. 9. Screenshot of the interface of our website homepage

4. Conclusion

Motivated by the challenge in the quick and accurate prediction on T_d and T_g charactering the thermal stability of OLED materials, we developed ensemble machine learning models coupled with the combined descriptors. 1944 experimental T_g data and 1182 experimental T_d data were collected from experimental literature to construct a new dataset in order to support the data-driven machine learning. With the dataset and the combined descriptors, the optimal LightGBM models give satisifactory accuracy for the prediction of T_d and T_g, higher than other six classific ML models (SVM, PLS, LASSO, KRR, kNN, and RF ML models). The model is further validated by two types of out-of-sample predictions (including recently reported host and guest materials, organic ETMs and HTMs), *exhibiting good robustness and universality*. Finally, the expermimental validation on a high thermal stability OLED material further confirms the reliability of our models and application potential in practice. In addition, we constructed a website including all the data and the optimized ML models in order to

provide a simple and quick tool for estimating the two important properties of unknown compounds, in turn assisting the design of the OLED materials.

Acknowledgements

Y.Z. and C.F. contributed equally to this work.

Conflicts of Interest

The authors declare no competing financial interest.

5. Reference

1. F. B. Dias, K. N. Bourdakos, V. Jankus, K. C. Moss, K. T. Kamtekar, V. Bhalla, J. Santos, M. R. Bryce and A. P. Monkman, *Advanced Materials*, 2013, **25**, 3707-3714.
2. S. Kim, H. J. Kwon, S. Lee, H. Shim, Y. Chun, W. Choi, J. Kwack, D. Han, M. Song and S. Kim, *Advanced Materials*, 2011, **23**, 3511-3516.
3. H. Nakanotani, K. Masui, J. Nishide, T. Shibata and C. Adachi, *Scientific reports*, 2013, **3**, 1-6.
4. S. Kothavale, W. J. Chung and J. Y. Lee, *Journal of Materials Chemistry C*, 2020.
5. H. Aziz and Z. D. Popovic, *Chemistry of Materials*, 2004, **16**, 4522-4532.
6. S. Scholz, D. Kondakov, B. Lussem and K. Leo, *Chemical reviews*, 2015, **115**, 8449-8503.
7. J. A. McEwan, A. J. Clulow, A. Nelson, N. R. Yepuri, P. L. Burn and I. R. Gentle, *ACS applied materials & interfaces*, 2017, **9**, 14153-14161.
8. S. Lee, H. Kim and Y. Kim, *InfoMat*, 2021.
9. Y.-C. Tsai and J.-H. Jou, *Applied Physics Letters*, 2006, **89**, 243521.
10. E. Ravindran and N. Somanathan, *Journal of Materials Chemistry C*, 2017, **5**, 7436-7440.

11. X. Guo, Z. Tang, W. Yu, Y. Wang, Z. Zhao, J. Gu, Z. Liu, B. Qu, L. Xiao and Z. Chen, *Organic Electronics*, 2021, **89**, 106048.
12. J. Jayakumar, W.-L. Wu, C.-L. Chang, T.-Y. Han, L.-Y. Ting, C.-M. Yeh, H.-W. Hung and H.-H. Chou, *Organic Electronics*, 2021, **88**, 106013.
13. H. Yu, X. Song, N. Xie, J. Wang, C. Li and Y. Wang, *Advanced Functional Materials*, 2021, **31**, 2007511.
14. Z. Qiu, W. Xie, Z. Yang, J.-H. Tan, Z. Yuan, L. Xing, S. Ji, W.-C. Chen, Y. Huo and S.-j. Su, *Chemical Engineering Journal*, 2021, **415**, 128949.
15. C. Deng, S. Zheng, D. Wang, J. Yang, Y. Yue, M. Li, Y. Zhou, S. Niu, L. Tao and T. Tsuboi, *The Journal of Physical Chemistry C*, 2019, **123**, 29875-29883.
16. G. Subramanian, B. Ramsundar, V. Pande and R. A. Denny, *Journal of chemical information and modeling*, 2016, **56**, 1936-1949.
17. Y. Kobayashi and K. Yoshida, *Environmental Research*, 2021, **196**, 110363.
18. D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher and A. G. Doyle, *Science*, 2018, **360**, 186-190.
19. K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature*, 2018, **559**, 547-555.
20. L. Chen, B. Xu, J. Chen, K. Bi, C. Li, S. Lu, G. Hu and Y. Lin, *Journal of Materials Chemistry C*, 2020, **8**, 13079-13089.
21. V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. A. Persson, G. Ceder and A. Jain, *Nature*, 2019, **571**, 95-98.
22. S. Yin, Z. Shuai and Y. Wang, *Journal of chemical information and computer sciences*, 2003, **43**, 970-977.
23. R. Barbosa-da-Silva and R. Stefani, *Molecular Simulation*, 2013, **39**, 234-244.
24. C. C. Peng, S. Y. Yang, H. C. Li, G. H. Xie, L. S. Cui, S. N. Zou, C. Poriel, Z. Q. Jiang and L. S. Liao, *Advanced Materials*, 2020, **32**, 2003885.
25. Y. Zhuo and J. Brgoch, *The Journal of Physical Chemistry Letters*, 2021, **12**, 764-772.
26. C. W. Yap, *Journal of computational chemistry*, 2011, **32**, 1466-1474.
27. Y. Yuan, F. Zheng and C.-G. Zhan, *The AAPS journal*, 2018, **20**, 1-10.

28. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye and T.-Y. Liu, *Advances in neural information processing systems*, 2017, **30**, 3146-3154.
29. T. Chen, C. H. Lu, Z. Chen, X. Gong, C. C. Wu and C. Yang, *Chemistry–A European Journal*, 2021, **27**, 3151-3158.
30. L. Xing, Z.-L. Zhu, J. He, Z. Qiu, Z. Yang, D. Lin, W.-C. Chen, Q. Yang, S. Ji and Y. Huo, *Chemical Engineering Journal*, 2020, 127748.
31. J. Zhao, X. Chen, Z. Yang, T. Liu, Z. Yang, Y. Zhang, J. Xu and Z. Chi, *The Journal of Physical Chemistry C*, 2018, **123**, 1015-1020.
32. J. Li, Z. Yang, Y. Feng, Z. Su, Z. Qiu, J.-H. Tan, W.-C. Chen, M. Zhang, Z.-X. Xu and Y. Huo, *Journal of Materials Chemistry C*, 2020, **8**, 16858-16869.
33. F. Lucas, C. Quinton, S. Fall, T. Heiser, D. Tondelier, B. Geffroy, N. Leclerc, J. Rault-Berthelot and C. Poriel, *Journal of Materials Chemistry C*, 2020, **8**, 16354-16367.
34. N. Zhang, C. Zheng, Z. Chen, J. Zhao, M. Zhang, H. Yang, Z. He, X. Du and S. Tao, *Journal of Materials Chemistry C*, 2021, **9**, 600-608.
35. J. Yuan, H. Jiang, Q. Yang, Y. Xiang, Y. Zhang, Y. Dai, P. Li, C. Zheng, G. Xie and R. Chen, *Journal of Materials Chemistry C*, 2021, **9**, 687-692.
36. R. S. Bernard, G. Sych, S. Nasiri, O. Bezvikonnyi, D. Volyniuk, J. Simokaitiene, A. Bucinskas, D. Gudeika, A. Ariffin and J. V. Grazulevicius, *Synthetic Metals*, 2021, **271**, 116641.
37. X. Qiu, G. Tian, C. Lin, Y. Pan, X. Ye, B. Wang, D. Ma, D. Hu, Y. Luo and Y. Ma, *Advanced Optical Materials*, 2021, **9**, 2001845.
38. Z. Wang, X. Zhu, S. Zhang, L. Xu, Z. Zhao and G. He, *Advanced Optical Materials*, 2021, **9**, 2001764.
39. S. N. Zou, X. Chen, S. Y. Yang, S. Kumar, Y. K. Qu, Y. J. Yu, M. K. Fung, Z. Q. Jiang and L. S. Liao, *Advanced Optical Materials*, 2020, **8**, 2001074.
40. M. Mahmoudi, J. Keruckas, K. Leitonas, S. Kutsiy, D. Volyniuk and J. V. Gražulevičius, *Journal of Materials Research and Technology*, 2021, **10**, 711-721.
41. A. Arai, H. Sasabe, K. Nakao, Y. Masuda and J. Kido, *Chemistry–A European*

Journal, 2021, **27**, 4971-4976.

42. X. Lv, L. Xu, W. Cui, Y. Yu, H. Zhou, M. Cang, Q. Sun, Y. Pan, Y. Xu and D. Hu, *ACS Applied Materials & Interfaces*, 2020.
43. D. Saito, H. Sasabe, T. Kikuchi, T. Ito, H. Tsuneyama and J. Kido, *Journal of Materials Chemistry C*, 2021, **9**, 1215-1220.
44. H. Arai, H. Sasabe, H. Tsuneyama, K. Kumada and J. Kido, *Chemistry–A European Journal*, 2021.
45. C. Zhou, C. Cao, D. Yang, X. Cao, H. Liu, D. Ma, C.-S. Lee and C. Yang, *Materials Chemistry Frontiers*, 2021, **5**, 3209-3215.
46. X. Zhang, J.-X. Chen, K. Wang, Y.-Z. Shi, X.-C. Fan, S.-L. Zhang, L. Wu, Y.-Q. Li, X.-M. Ou and X.-H. Zhang, *Journal of Materials Chemistry C*, 2020, **8**, 17457-17463.
47. H. Lee, J. H. Park, K. J. Yang, S. J. Hwang, R. Braveenth, T. H. Ha, M. I. Han, C. W. Lee and J. H. Kwon, *Journal of Materials Chemistry C*, 2021.
48. X. Dong, H. Wang, J. Huo, S. Liu, H. Shi, F. Cheng and B. Z. Tang, *Journal of Molecular Structure*, 2021, **1228**, 129721.
49. K. M. Youn, H. Lee, H. J. Yoo, Y. H. Jung, J. Do Park, H. Jeong, J. Lee, J. Y. Lee and J. H. Kwon, *Journal of Materials Chemistry C*, 2020, **8**, 13811-13818.
50. G.-X. Yang, Y. Chen, J.-J. Zhu, J.-Y. Song, S.-S. Tang, D. Ma and Q.-X. Tong, *Dyes and Pigments*, 2021, **187**, 109088.
51. L. Salah, M. K. Etherington, A. Shuaib, A. Danos, A. A. Nazeer, B. Ghazal, A. Prlj, A. T. Turley, A. Mallick and P. R. McGonigal, *Journal of Materials Chemistry C*, 2021, **9**, 189-198.
52. T. Huang, X. Song, M. Cai, D. Zhang and L. Duan, *Materials Today Energy*, 2021, **21**, 100705.
53. J.-L. He, F.-C. Kong, B. Sun, X.-J. Wang, Q.-S. Tian, J. Fan and L.-S. Liao, *Chemical Engineering Journal*, 2021, 130470.
54. Q. Liu, S. Chavhan, H. Zhang, H. Sun, A. J. Brock, S. Manzhos, Y. Chen, K. Feron, S. E. Bottle and J. C. McMurtrie, *Advanced Electronic Materials*, 2021, **7**, 2000804.

55. M. Ouyang, L. Xing, Q. Chen, H. Huang, M. Zhu, K. Hu, Y. Liu, W.-C. Chen, Y. Huo and C. Yang, *Journal of Materials Chemistry C*, 2021, **9**, 1678-1684.
56. J. Xu, H. Liu, J. Li, Z. Zhao and B. Z. Tang, *Advanced Optical Materials*, 2021, **9**, 2001840.
57. W. Yuan, D. Hu, M. Zhu, W. Shi, C. Shi, N. Sun and Y. Tao, *Dyes and Pigments*, 2021, **191**, 109395.
58. C. Tang, T. Yang, X. Cao, Y. Tao, F. Wang, C. Zhong, Y. Qian, X. Zhang and W. Huang, *Advanced Optical Materials*, 2015, **3**, 786-790.
59. M. Bian, Y. Wang, X. Guo, F. Lv, Z. Chen, L. Duan, Z. Bian, Z. Liu, H. Geng and L. Xiao, *Journal of Materials Chemistry C*, 2018, **6**, 10276-10283.
60. M. Liu, S.-J. Su, M.-C. Jung, Y. Qi, W.-M. Zhao and J. Kido, *Chemistry of Materials*, 2012, **24**, 3817-3827.
61. B. Wang, G. Mu, J. Tan, Z. Lei, J. Jin and L. Wang, *Journal of Materials Chemistry C*, 2015, **3**, 7709-7719.
62. Q. Zhang, B. Wang, J. Tan, G. Mu, W. Yi, X. Lv, S. Zhuang, W. Liu and L. Wang, *Journal of Materials Chemistry C*, 2017, **5**, 8516-8526.
63. Y. Watanabe, D. Yokoyama, T. Koganezawa, H. Katagiri, T. Ito, S. Ohisa, T. Chiba, H. Sasabe and J. Kido, *Advanced Materials*, 2019, **31**, 1808300.
64. H. Ye, D. Chen, M. Liu, S. J. Su, Y. F. Wang, C. C. Lo, A. Lien and J. Kido, *Advanced Functional Materials*, 2014, **24**, 3268-3275.
65. H. Sasabe, D. Tanaka, D. Yokoyama, T. Chiba, Y. J. Pu, K. i. Nakayama, M. Yokoyama and J. Kido, *Advanced Functional Materials*, 2011, **21**, 336-342.
66. L. Duan, J. Qiao, Y. Sun, D. Zhang, G. Dong, L. Wang and Y. Qiu, *Advanced Optical Materials*, 2013, **1**, 167-172.
67. D. Zhang, J. Qiao, D. Zhang and L. Duan, *Advanced Materials*, 2017, **29**, 1702847.
68. D. Zhang, X. Song, H. Li, M. Cai, Z. Bin, T. Huang and L. Duan, *Advanced Materials*, 2018, **30**, 1707590.
69. Z. Deng, M. He, Y. Zhang, F. Ullah, K. Ding, J. Liang, Z. Zhang, H. Xu, Y. Qiu and Z. Xie, *Chemistry of Materials*, 2020.

70. X. Zhang, Y. Liang, R. Ghadari, C. Liu, X. Liu, Z. Zhang, S. Ma, Y. Ding, M. Cai and S. Dai, *Dyes and Pigments*, 2021, **187**, 109129.
71. X. Yu, Z. Li, X. Sun, C. Zhong, Z. Zhu and A. K.-Y. Jen, *Nano Energy*, 2021, **82**, 105701.
72. H. Guo, H. Zhang, C. Shen, D. Zhang, S. Liu, Y. Wu and W. H. Zhu, *Angewandte Chemie International Edition*, 2021, **60**, 2674-2679.
73. H. Lu, F. Wu, Y. Yang, S. Li, Y. Hua and L. Zhu, *Journal of Materials Chemistry C*, 2020, **8**, 13415-13421.
74. X. Zhang, R. Ghadari, X. Liu, W. Wang, Y. Ding, M. Cai, J. H. Pan and S. Dai, *Journal of Energy Chemistry*, 2021, **62**, 563-571.
75. Y. Fu, Y. Sun, H. Tang, L. Wang, H. Yu and D. Cao, *Dyes and Pigments*, 2021, **191**, 109339.
76. Y. C. Chen, Y. H. Li, C. L. Chung, H. L. Hsu and C. P. Chen, *Progress in Photovoltaics: Research and Applications*, 2020, **28**, 49-59.
77. Y. Yang, S. U. Ryu, F. Wu, H. Lu, K. Jia, C. Zhong, T. Park and L. Zhu, *Chemical Engineering Journal*, 2021, 130396.
78. H. Ha, Y. J. Shim, D. H. Lee, E. Y. Park, I.-H. Lee, S.-K. Yoon and M. C. Suh, *ACS Applied Materials & Interfaces*, 2021, **13**, 21954-21963.
79. X. Zhang, R. Ghadari, X. Liu, W. Wang, Y. Ding, M. Cai, J. H. Pan and S. Dai, *Solar Energy*, 2021, **221**, 323-331.
80. A. A. Sutanto, V. Joseph, C. Igci, O. A. Syzgantseva, M. A. Syzgantseva, V. Jankauskas, K. Rakstys, V. I. Queloz, P.-Y. Huang and J.-S. Ni, *Chemistry of Materials*, 2021, **33**, 3286-3296.
81. M. Zhai, Y. Miao, C. Chen, H. Wang, X. Ding, C. Wu, X. Yang and M. Cheng, *Dyes and Pigments*, 2021, **191**, 109340.
82. Z. Gong, R. Wang, Y. Jiang, X. Kong, Y. Lin, Z. Xu, G. Zhou, J.-M. Liu, K. Kempa and J. Gao, *Organic Electronics*, 2021, **92**, 106102.
83. N. Xiang, Z. Gao, G. Tian, Y. Chen, W. Liang, J. Huang, Q. Dong, W.-Y. Wong and J. Su, *Dyes and Pigments*, 2017, **137**, 36-42.
84. Abdullah, E.-B. Kim, M. S. Akhtar, H.-S. Shin, S. Ameen and M. K.

Nazeeruddin, *ACS Applied Energy Materials*, 2020.

85. J. Wu, M. Hu, G. Song, Y. Li, W. Tan, Y. Tian and B. Xu, *Chemical Engineering Journal*, 2021, **422**, 130124.

86. S. Daskeviciute, C. Momblona, K. Rakstys, A. A. Sutanto, M. Daskeviciene, V. Jankauskas, A. Gruodis, G. Bubniene, V. Getautis and M. K. Nazeeruddin, *Journal of Materials Chemistry A*, 2021, **9**, 301-309.

87. H.-F. Chen, T.-C. Wang, S.-W. Lin, W.-Y. Hung, H.-C. Dai, H.-C. Chiu, K.-T. Wong, M.-H. Ho, T.-Y. Cho and C.-W. Chen, *Journal of Materials Chemistry*, 2012, **22**, 15620-15627.

88. Y. L. Zhang, Q. Ran, Q. Wang, Y. Liu, C. Hänisch, S. Reineke, J. Fan and L. S. Liao, *Advanced Materials*, 2019, **31**, 1902368.