Data-driven machine learning models for the quick and accurate prediction of thermal stability properties of OLED materials

Yihuan Zhao [1], Caixia Fu [1], Ling 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, the further development and commercialization of OLEDs requires higher-quality OLED materials, including materials with a high thermal stability. Thermal stability is associated with the glass transition temperature (T_g) and decomposition temperature (T_d), but experimental determinations of these two important properties generally involve a time-consuming and laborious process. Thus, the development of a quick and accurate prediction tool is highly desirable. Motivated by the challenge, we explored machine learning (ML) by constructing a new dataset with more than one thousand samples collected from a wide range of literature, through which ensemble learning models were explored. Models trained with the LightGBM algorithm exhibited the best prediction performance, where the values of MAE, RMSE, and R^2 were 17.15 K, 24.63 K, and 0.77 for T_g prediction and 24.91 K, 33.88 K, and 0.78 for T_d prediction. The prediction performance and the generalization of the machine learning models were further tested by out-of-sample data, which also exhibited satisfactory results. Experimental validation further demonstrated the reliability and the practical potential of the ML-based model. In order to extend the practical application of the ML-based models, an online prediction platform was constructed. This platform includes the optimal prediction models and all the thermal stability data under study, and it is freely available at http://oledtppxmpugroup.com. We expect that this platform will become a useful tool for experimental investigation of T_g and T_d, accelerating the design of OLED materials with desired properties.

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]. However, increasing the thermal stability of OLED materials can strengthen the stability of device performance. Therefore, a large number of researchers have paid attention to the OLED materials with high thermal stability in recent years [8].

The glass transition temperature (T_g) and decomposition temperature (T_d, corresponding to 5% weight loss) are the two most important thermal properties of OLED materials, and exert significant influence on the performance of OLED devices [9]. 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]. The experimental 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, the development of a quick and accurate method to predict T_g and T_d is highly desirable. It is generally acknowledged

that the thermal stability of OLED materials is closely related to their molecular structures [15]. However, the relationship between molecular structure and the thermal stability such as T_g and T_d is complex and has not been elucidated. Machine learning (ML), a key technique used in artificial intelligence, can map the complex relationshiop underlying a large amount of data. ML has been succesfully applied in the fields of medicinal chemistry, environmental risk assessment, organic synthesis, and materials science [16-21]. To the best of our knowledge, only two previous studies have used machine learning methods to predict the T_g of OLED materials [22, 23], and there is a significant lack of T_d predictions. In 2003, Yin et al. performed a quantitative structure-property relationship (QSPR) model to predict the T_g of 88 OLED molecules with MAE = 17.9 K by using a multilinear regression (MLR) method [22]. In 2013, Silva et al. developed a QSPR model to predict the T_g of 66 OLED materials with $R^2 = 0.96$, MAE = 0.97 K by using support vector machines (SVM) [23]. These two studies appear to make highly accurate predictions based on a small amount of data (fewer than 100 molecules). However, the prediction ability of the single ML models in the two previous studies is not reliable and unstable, due to the generalization ability of ML models to unknown compounds depending on the size of the dataset. Unfortunately, a database including the two important properties of T_g and T_d of OLED materials has not yet been constructed. However, in the past decade, a significant amount of thermal stability data for OLED materials has been published. While these published data are dispersed across a wide range of literature, they still provide a possible data source for constructing a robust machine learning model.

In order to explore robust and universal ML-based prediction models, we constructed 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 utilized an ensemble learning approach LightGBM algorithm, rather than the single machine learning methods used in previous works, to integrate multiple weak learners to build an entire learner with better prediction performance than that from any of its component. The prediction performance of our models was verified by two types of out-of-sample datasets, exhibiting satisfactory results and confirming the generality of the models. In addition, experimental validation further confirmed the reliability of our prediction models and potential in practical application. More importantly, we built a website including the optimal T_g and T_d prediction models coupled with the new dataset, which is freely available at http://oledtppxmpugroup.com. We expect that this website will serve as a useful tool to help experimental investigators quickly and accurately 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 T_d values, we compared the TGA curves in the literature and take the T_d measured by the smoother TGA curve as the final value (vide Fig. S1). 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 large deviation, often greater than 40 K. This is because the purity of the compound has a significant influence on the experimental value of T_d (corresponding to 5% weight loss). To accurately measure T_d (corresponding to 5% weight loss), the purity of the compound must be high. 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.

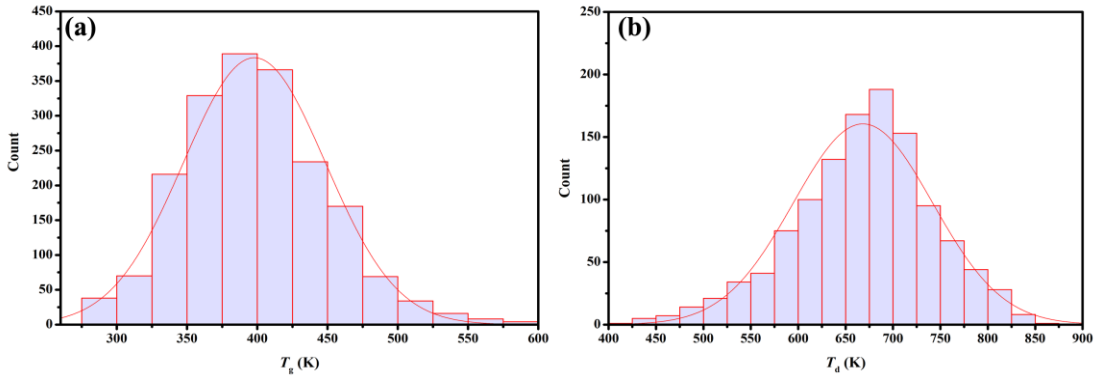


Fig. 1. (a) T_g distribution of 1944 molecules. (b) T_d distribution of 1182 OLED molecules.

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 [24]. 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 [25], 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

In the work, we mainly utilized the ensemble learning strategy to construct a comprehensive model by combining base learners. Compared to the single machine

learning, ensemble learning not only produces a more stable global model, but also guarantees diminishing uncertainty. Herein, LightGBM, a recent modification of the gradient boosting (GB) algorithm [26], is considered. 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 high-dimensional datasets. To further evaluate the performance of the LightGBM algorithm, it was compared with six classic ML algorithms, including 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). The LightGBM code is available at https://github.com/Microsoft/LightGBM. Other ML algorithms 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%. ^1H 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 ML 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 ML 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. 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 ML 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.

In addition, we also compareed the predictive performance of property-based molecular descriptors and the SC_2CDK fingerprints. Table 1 summarizes the T_g and T_d prediction results of the LightGBM models. As can be seen, the ML model with 1D and 2D molecular descriptors has better T_g prediction performance than the corresponding ML 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.[25] 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 optimal model vs. the experimental T_g values for test set is shown in Fig. 2a, exhibiting a reasonable agreement. 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.

Different from T_g prediction, ML model with SC-2CDK fingerprints has better T_d prediction performance than the corresponding ML 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 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 optimal model vs. the experimental T_d values for the independent test set is shown in Fig. 2b. 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.

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. Descriptions of the 10 most important features for T_g and T_d prediction are shown in Table S5 and Table S6. 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. Summing up the feature importance analysis, we can find that topological charges, topological distances and hydrogen bonding interactions all have a great effect on T_g and T_d, because these factors have an impact on the rigidity of the molecule.

To further evaluate the performance of the LightGBM models, the predictive powers of LightGBM models were compared with six classic ML models (SVM, PLS, LASSO, KRR, kNN, and RF). The MAE and RMSE of the test set for the different ML methods are shown in Fig. 5. As can be seen, the LightGBM models have the lowest MAE and RMSE for T_g and T_d prediction, exhibiting the best performance.

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

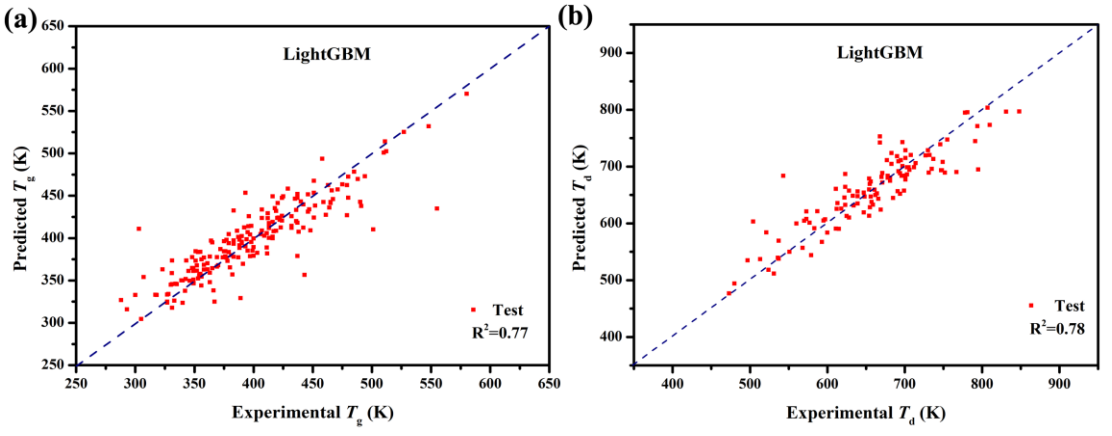


Fig. 2. Correlation plots of (a) T_g and (b) T_d for the independent test set based on LightGBM-based optimal models.

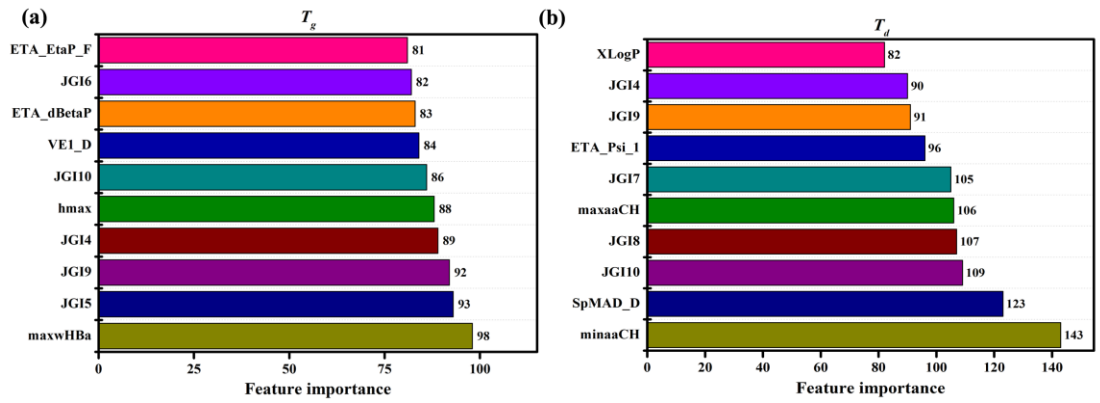


Fig. 3. Feature importance for the prediction of (a) T_g and (b) T_d obtained from the LightGBM-based optimal models.

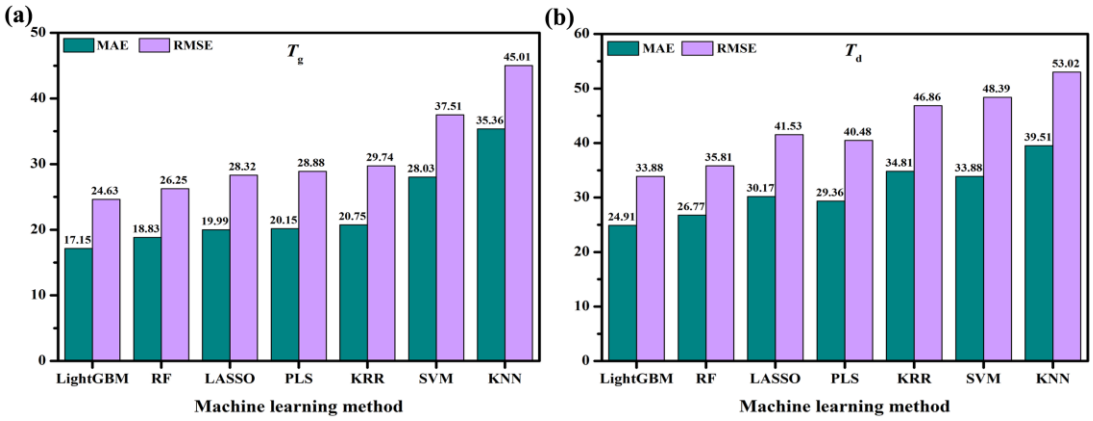


Fig. 4. MAE and RMSE for the prediction of (a) T_g and (b) T_d of the independent test set with different machine learning methods.

3.2 Verification of T_g and T_d prediction models

The obtained T_g and T_d prediction models based on the LightGBM algorithm were further tested in out-of-sample predictions. Two representative applications are shown herein.

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

68 recently reported (2020 & 2021) OLED compounds were collected as independent test set [14, 27-54], involving 40 T_g values and 40 T_d values. These compounds are mainly used in host-guest emissive layer of OLED devices (see Table S7 and Table S8 in supporting information for details). Furthermore, 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 was compared with two other compounds that appear in the same literature.[49] The TGA curve of 3CzCNPyz is shown in Fig. S2 and the TGA curves of 2Cz2CNPyz and 4CzPyz are shown in Fig. S3. 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. S2 shows that the weight of compound 3CzCNPyz slightly decreases below 500 K (227 °C). This can be attributed to the presence of impurities (such as solvent) in 3CzCNPyz which cause the experimental T_d value of 3CzCNPyz to decrease. In order to further confirm the prediction reliability, the compound 4CzCNPy was selected for T_d prediction. 4CzCNPy is a compound similar in structure to 3CzCNPyz. In Ref.56 [55] and Ref.57 [56], the experimental 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 K, 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. These results further support the reliability and advantage of the ML prediction models.

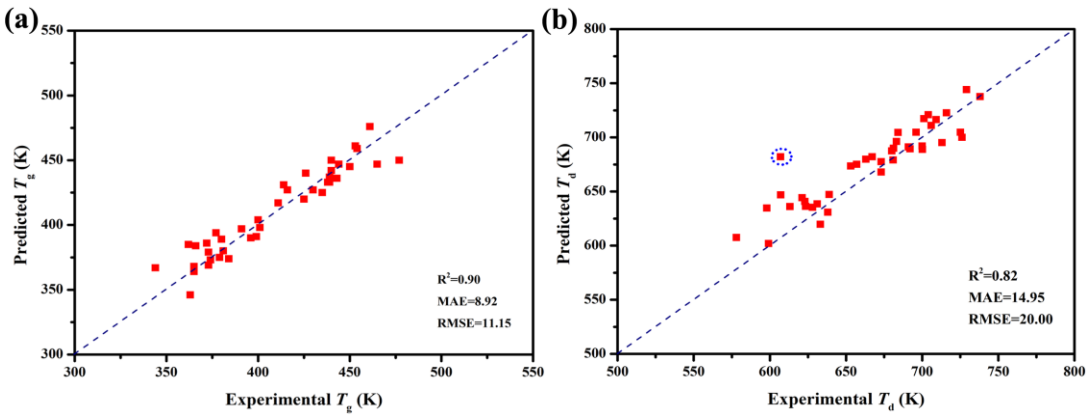


Fig. 5. Correlation plots of (a) predicted and experimental T_g values of 40 OLED compounds and (b) predicted and experimental T_d values of 40 OLED compounds in recent literature.

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 (PSCs) as their electron transport layer or hole transport layer. Because electron transport and hole transport layer should be thermally stable to improve the overall lifetime of devices, both materials require 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, we collected 65 organic ETMs and HTMs used in OLEDs and solar cells for T_g and T_d prediction [57-85], involving 40 T_g values and 40 T_d values. More detailed information on these organic ETMs and HTMs can be found in Table S9 and Table S10 in the supporting information. 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. 6a. 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. 6b 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, confirming the reliability of our models.

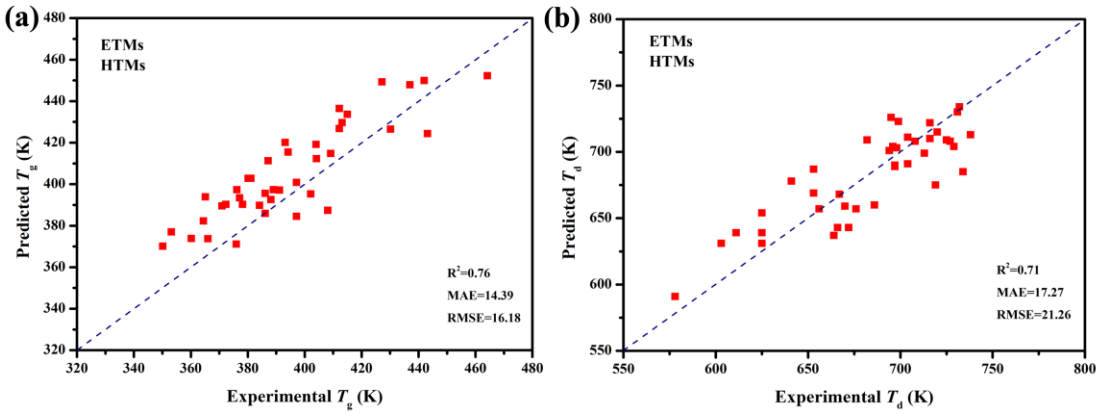


Fig. 6. Predicted vs. experimental (a) T_g and (b) T_d values for 40 organic electron-transport materials and hole-transport materials used in OLEDs and perovskite solar cells.

3.3 Experimental validation

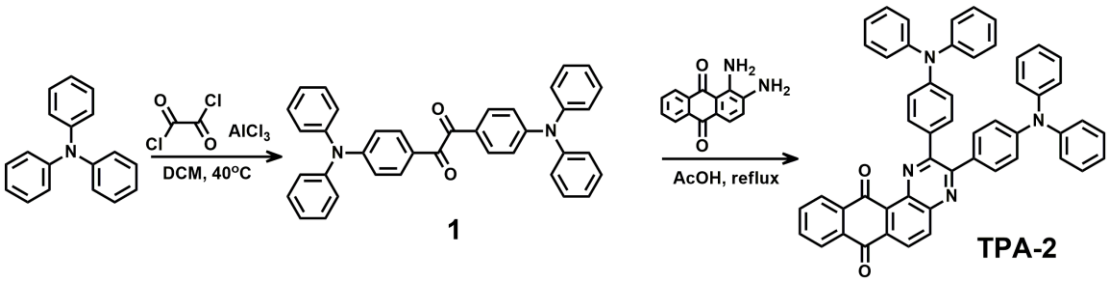
In order to verify the application potential of these ML-based models in practice, the models were used to predict the T_g and T_d of 50 new molecules. These 50 molecules were designed to have donor-acceptor (D-A) and donor-acceptor-donor (D-A-D) type structures. D-A and D-A-D structures are particularly important types in OLED materials and are also convenient for subsequent synthesis studies. Table S11 shows their chemical structures and predicted T_g and T_d values of these designed compounds. Herein, we focus on the new compound TPA-2 with the third highest predicted T_g and the highest predicted T_d (TPA-2).

Density functional theory (DFT) simulations and time-dependent DFT were performed

for TPA-2 before the compound was synthesized. HOMO/LUMO distributions of TPA-2 in the ground state are shown in Fig. S4. 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. The theoretical calculation parameters of TPA-2 were compared with TPA-PZCN, which is a high efficiency red thermally activated delayed fluorescence (TADF) material with an external quantum efficiency close to 30% [86]. As 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, implying 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 spin-orbit coupling (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. Furthermore, our models predict that TPA-2 have a high thermal stability. Thus, TPA-2 was selected for further experimental validation.

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 ^1H NMR and ^{13}C NMR (vide Fig. S5, Fig. S6 and Fig. S7). The emission maxima of TPA-2 in

toluene solution is greater than 600 nm and ΔE_{ST} of TPA-2 is 0.07 eV, indicating that TPA-2 can be used as a red-TADF OLED material (Fig. S8 and Table S12). The thermal properties of TPA-2 were determined 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. 7), 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 ML models to predict the thermal stability of unknown OLED materials. Our ML models could be served as a useful tool to quickly screen high thermal stability OLED materials.



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

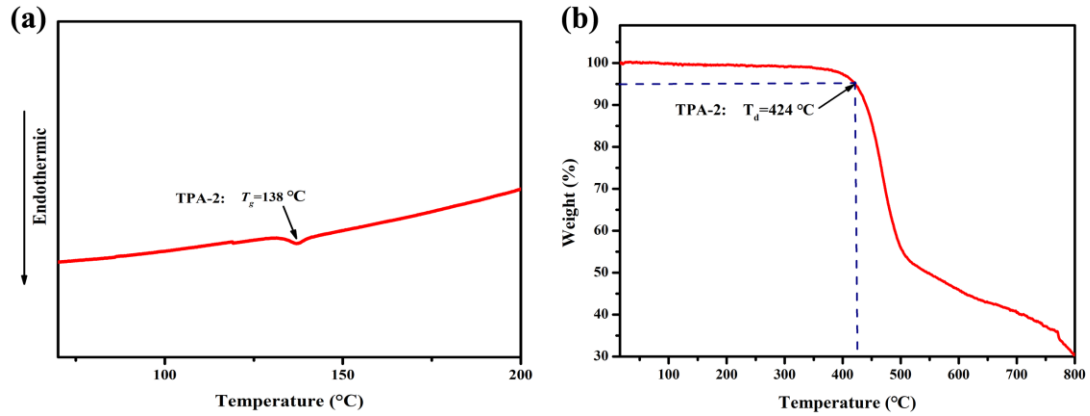


Fig. 7. DSC and TGA curves for TPA-2.

3.4 Website for T_g and T_d Predictions

Currently, hundreds of articles about OLED materials are published every year [8]. There are a lot of useful data in the literature, but there is no existing database that organizes OLED material data. With the aims of archiving the thermal stability data of OLED materials and helping experimental scientists utilize the models reported in this paper for designing new OLED compounds with desired T_g and T_d values, an online tool was developed. This website is accessible at http://oledtppxmpugroup.com. Users can make predictions by inputting canonical SMILES, and 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. A screenshot of the website homepage interface is shown in Fig. 8. More details can be found by visiting the website. We will continue updating the dataset and optimal model on the website in order to more accurately predict T_g and T_d of OLED materials.

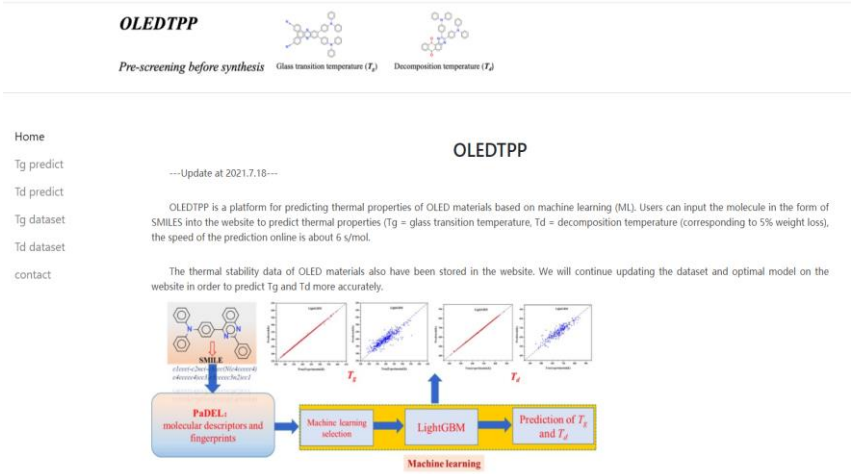


Fig. 8. Screenshot of the interface of our website homepage

4. Conclusion

Motivated by the challenge of quick and accurate prediction of T_d and T_g to characterize the thermal stability of OLED materials, we developed ensemble ML models coupled with combined descriptors. 1944 experimental T_g values and 1182 experimental T_d values were collected from experimental literature to construct a new dataset in order

to support the data-driven ML models. With the dataset and the combined descriptors, the optimal LightGBM models offer satisfactory accuracy for the prediction of T_d and T_g, with higher accuracy than other six classic ML models (SVM, PLS, LASSO, KRR, kNN, and RF ML models). The models are further validated by two types of out-of-sample prediction (including recently reported host and guest materials as well as organic ETMs and HTMs), exhibiting good robustness and universality. Finally, the experimental validation of a high thermal stability OLED material further confirms the reliability of our models and practical application potential. In addition, we constructed a website including all the data and the optimal ML models in order to provide a simple and quick tool for estimating these two important properties for unknown compounds. We believe this website will assist with the design of future OLED materials.

Author contributions

Xuemei Pu and Zhiyun Lu designed the research. Yihuan Zhao performed the research. Caixia Fu and Ling Fu contributed to the model construction and data analysis. Caixia Fu performed the experimental synthesis. Xuemei Pu and Yihuan Zhao wrote the manuscript. All authors reviewed the manuscript.

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

Yihuan Zhao and Caixia Fu contributed equally to this work. This project is supported by Sichuan International Science and technology innovation cooperation project (Grant No.2021YFH0140), NSAF (Grand No. U1730127) and Key Laboratory Foundation (Grand No. 6142603190305).

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. W. Yap, *Journal of computational chemistry*, 2011, **32**, 1466-1474.
25. Y. Yuan, F. Zheng and C.-G. Zhan, *The AAPS journal*, 2018, **20**, 1-10.
26. 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.
27. T. Chen, C. H. Lu, Z. Chen, X. Gong, C. C. Wu and C. Yang, *Chemistry–A European Journal*, 2021, **27**, 3151-3158.
28. 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.
29. F. Huang, K. Wang, Y. Z. Shi, X. C. Fan and X. H. Zhang, *ACS Applied Materials & Interfaces*, 2021
30. 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.

31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. Z. Wang, X. Zhu, S. Zhang, L. Xu, Z. Zhao and G. He, *Advanced Optical Materials*, 2021, **9**, 2001764.
37. 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.
38. M. Mahmoudi, J. Keruckas, K. Leitonas, S. Kutsiy, D. Volyniuk and J. V. Gražulevičius, *Journal of Materials Research and Technology*, 2021, **10**, 711-721.
39. A. Arai, H. Sasabe, K. Nakao, Y. Masuda and J. Kido, *Chemistry–A European Journal*, 2021, **27**, 4971-4976.
40. 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.
41. D. Saito, H. Sasabe, T. Kikuchi, T. Ito, H. Tsuneyama and J. Kido, *Journal of Materials Chemistry C*, 2021, **9**, 1215-1220.
42. H. Arai, H. Sasabe, H. Tsuneyama, K. Kumada and J. Kido, *Chemistry–A European Journal*, 2021.
43. 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.

44. 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.
45. 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.
46. X. Dong, H. Wang, J. Huo, S. Liu, H. Shi, F. Cheng and B. Z. Tang, *Journal of Molecular Structure*, 2021, **1228**, 129721.
47. 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.
48. 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.
49. 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.
50. T. Huang, X. Song, M. Cai, D. Zhang and L. Duan, *Materials Today Energy*, 2021, **21**, 100705.
51. 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.
52. 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.
53. 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.
54. J. Xu, H. Liu, J. Li, Z. Zhao and B. Z. Tang, *Advanced Optical Materials*, 2021, **9**, 2001840.
55. W. Yuan, D. Hu, M. Zhu, W. Shi, C. Shi, N. Sun and Y. Tao, *Dyes and Pigments*, 2021, **191**, 109395.
56. 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.
57. 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.

58. M. Liu, S.-J. Su, M.-C. Jung, Y. Qi, W.-M. Zhao and J. Kido, *Chemistry of Materials*, 2012, **24**, 3817-3827.
59. B. Wang, G. Mu, J. Tan, Z. Lei, J. Jin and L. Wang, *Journal of Materials Chemistry C*, 2015, **3**, 7709-7719.
60. 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.
61. Y. Watanabe, D. Yokoyama, T. Koganezawa, H. Katagiri, T. Ito, S. Ohisa, T. Chiba, H. Sasabe and J. Kido, *Advanced Materials*, 2019, **31**, 1808300.
62. 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.
63. 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.
64. L. Duan, J. Qiao, Y. Sun, D. Zhang, G. Dong, L. Wang and Y. Qiu, *Advanced Optical Materials*, 2013, **1**, 167-172.
65. D. Zhang, J. Qiao, D. Zhang and L. Duan, *Advanced Materials*, 2017, **29**, 1702847.
66. D. Zhang, X. Song, H. Li, M. Cai, Z. Bin, T. Huang and L. Duan, *Advanced Materials*, 2018, **30**, 1707590.
67. 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.
68. 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.
69. X. Yu, Z. Li, X. Sun, C. Zhong, Z. Zhu and A. K.-Y. Jen, *Nano Energy*, 2021, **82**, 105701.
70. H. Guo, H. Zhang, C. Shen, D. Zhang, S. Liu, Y. Wu and W. H. Zhu, *Angewandte Chemie International Edition*, 2021, **60**, 2674-2679.
71. H. Lu, F. Wu, Y. Yang, S. Li, Y. Hua and L. Zhu, *Journal of Materials Chemistry C*, 2020, **8**, 13415-13421.
72. 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.

73. Y. Fu, Y. Sun, H. Tang, L. Wang, H. Yu and D. Cao, *Dyes and Pigments*, 2021, **191**, 109339.
74. 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.
75. Y. Yang, S. U. Ryu, F. Wu, H. Lu, K. Jia, C. Zhong, T. Park and L. Zhu, *Chemical Engineering Journal*, 2021, 130396.
76. 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.
77. X. Zhang, R. Ghadari, X. Liu, W. Wang, Y. Ding, M. Cai, J. H. Pan and S. Dai, *Solar Energy*, 2021, **221**, 323-331.
78. 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.
79. M. Zhai, Y. Miao, C. Chen, H. Wang, X. Ding, C. Wu, X. Yang and M. Cheng, *Dyes and Pigments*, 2021, **191**, 109340.
80. 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.
81. 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.
82. Abdullah, E.-B. Kim, M. S. Akhtar, H.-S. Shin, S. Ameen and M. K. Nazeeruddin, *ACS Applied Energy Materials*, 2020.
83. J. Wu, M. Hu, G. Song, Y. Li, W. Tan, Y. Tian and B. Xu, *Chemical Engineering Journal*, 2021, **422**, 130124.
84. 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.
85. 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.

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