Electronic supplementary information

Accelerating OLED development with machine learning: advances and prospects

Xumian Qiao,[a,b] Changgang Huang,[c] Fan Ni,[*a,b] and Longzhen Qiu[*a,b]

[a] National Engineering Lab of Special Display Technology, Academy of Opto-Electronic Technology, Hefei University of Technology, Hefei, An-hui 230009, China Email: nfope@hfut.edu.cn, lzhqiu@hfut.edu.cn

[b] Anhui Industrial Innovation Research Institute of Advanced Optoelectronic Materials and Systems, Anhui Province Key Laboratory of Measuring Theory and Precision Instrument, School of Instrument Science and Optoelectronic Engineering, Hefei University of Technology, Hefei, Anhui 230009, China

[c] Fuyang Sineva New Material Technology Co., Ltd., Fuyang, Anhui 236000, China

Table S1. Summary of key machine learning studies for OLED material design and property prediction.

Serial number	Materials types	Projected targets	Main algorithm	Features	Metric: Value	Dataset Size	Ref.
1	Fluorescence	HOMO/LUMO (Prediction)	GCN	Molecular graph; Solvent graph	MAE = 0.058 eV	N = 3026 (Experimental data)	78
2	TADF	HOMO/LUMO energy (Prediction)	RF, Lasso, XGBoost	Molecular descriptors; Cheminformatics; Quantuam Chemical	RMSE = 0.25 eV (HOMO) RMSE = 0.36 eV (LUMO)	N ≈ 300 (Experimental data)	97
3	Organic materials of OLEDs, OPDs and OPVs	HOMO/LUMO energy (Prediction)	LR, KNN, RF, XGBT	Fingerprints (FPs, MACCS); Topological (E-state index)	r = 0.75, RMSE = 0.121 eV (HOMO), r = 0.84, RMSE = 0.150 eV (LUMO)	N = 1198 (Experimental data) N = 11626 (DFT data)	79
4	Fluorescence	PLQY, λ_{abs}, λ_{em} (Prediction)	SVM, LightGBM, GBRT	Structure descriptor (FSD); Comprehensive general solvent descriptor (CGSD)	MAE = 0.13 (PLQY), MAE = 0.08 eV (emission energy)	N = 4300 (Experimental data)	35
5	Phosphorescent platinum(II) complexes	λ_{em}, k_r, and PLQY (Prediction)	RF, AdaBoost, LightGBM	Photophysical properties (HOMO/LUMO, CT excited state)	R^2 = 0.96 (λ_{em}), R^2 = 0.81 (PLQY), R^2 = 0.67 (k_r)	N = 206 (Experimental data)	87
6	Phosphorescent iridium(III) complexes	$\lambda_{abs}/\lambda_{em}$ (Prediction)	LASSO, SVM, RF, GBRT	Fingerprints (CDK, MACCS); electronic structure properties	R^2 = 0.88 (λ_{em}), R^2 = 0.87 (λ_{ab})	N = 1046 (Experimental data)	88
7	Cyclometalated iridium(III) complexes	λ_{em}, PLQY (Prediction)	kNN, SVR, LightGBM, CatBoost, XGBoost	Molecular fingerprints (ECFP4)	MAE = 18.26 nm (λ_{max}), MAE = 0.129 (PLQY)	N ≈ 1200 (Experimental data)	89
8	TADF	AIE (Prediction)	LightGBM, RF, MLP KNN	Molecular fingerprints (CDK, CDK extended fingerprints, E-Sates fingerprints)	Accuracy = 0.974 (dependent test-set), accuracy = 0.963 (out-of-sample)	N = 3074 (Experimental data)	33
9	MR-TADF	PLQY (Prediction)	NP-VAE, RF, XGBoost, NN	Molecular descriptors; Cheminformatics decriptors	r = 0.85, RMSE = 0.1383, R^2 = 0.7 (XGBoost)	N = 402 (Experimental data) N > 15,000 (NP-VAE)	98
10	TADF	PLQY (Prediction)	ESIN	FMOs; ECFPs	Accuracy = 0.7, recall = 0.4	N = 939 (Experimental data)	86
11	TADF	TDM (Prediction)	RF, XGBoost, Lasso	Structural descriptors; Molecular descriptors	R^2 = 0.8, RMSE = 4.31% (XGBoost)	N ≈ 170 (Experimental data)	103
12	Heterocycles with saturated N atoms	C−N BDE (Prediction)	OLS, RF, XGBoost	Hirshfeld-I charge (q_N); bond angles; bond length; dipole moment; MPI; ODI HOMO	R^2 > 0.94, MSE = 0.003 eV	N = 146 (DFT data)	105
13	TADF	Multi-state BDE (Prediction)	NP-VAE, RFRF, XGBoost, NN	Molecular descriptors; Local descriptors (C−N bond)	R^2 > 0.92, MSE = 0.003 eV (XGBoost)	N = 360 (Reported) N > 13,000 (NP-VAE) N = 1500 (DFT data)	106

14	Ir(III) complexes	HOMO/LUMO; PLOY; E_{ad} (Material screen)	Uni-Mol	Structure properties (HOMO/LUMO, PLQY)	MAE = 0.067 eV, R = 0.95 (HOMO) MAE = 0.114 eV, R = 0.87 (LUMO) MAE = 0.043 eV, R = 0.96 (E_{ad}) MAE = 0.094 (PLQY)	N = 278 (Reported) N ≈ 1600,000 (Uni-Mol) N ≈ 1500 (QM)	107
15	Platinum(II) complexes	k_r (Material screen)	GNN, AdaBoost, LightGBM,	Molecular graph	R^2 ≈ 0.78, MAE ≈ 0.15, RMSE ≈ 0.20 (k^f_r) R^2 ≈ 0.81, MAE = 0.12, RMSE ≈ 0.17 (k^e_r)	N = 202 (Experimental data) N ≈ 526,000 (Schrodinger package) N ≈ 469 (DFT data)	111
16	TADF	k_{TADF} (Material screen)	NN	Fingerprints (ECFP)	R^2 = 0.80 (Linear model), R^2 = 0.94 (Neural network)	N ≈ 1,600,000 (RDKit package)	114
17	TADF	k_{ISC}/k_{RISC} (Material screen)	RFE, RF, GBRT, KNN, KRR	Initial descriptors(ΔE_{LL}, *Polar*, ΔE_{TT})	r = 0.89, R^2_*score* = 0.74 (GBRT)	N = 141 (Reported)	92
18	TADF	k_{RISC} (Material screen)	Bayesian	Quantum chemical properties descriptors (HOMO/LUMO energy, ΔE_{ST})	R^2 = 0.91, RMSE = 0.88	N = 1400 (Search space) N ≈ 200 (Quantum-chemically calculated)	93
19	TADF	E_{T1} (Material screen)	MPNN, MLP	Molecular graph	R^2 = 0.97, RMSE = 0.08 eV	N ≈ 300,000 (Enumeration algorithm) N ≈ 18,000,000 (Mutation algorithm) N = 258 (DFT data)	115
20	MR-TADF	ΔE_{ST}; FWHM (Material screen)	GCN	Molecular graph	MAE = 0.037 eV (ΔE_{ST}) MAE = 10 nm (FWHM)	N = 220 (Experimental data)	95
21	MR-TADF	FWHM; λ_{peak} (Material screen)	MLR, PCR, PLS, KPLS	Molecular descriptors; Fingerprints (linear, dendritic, radial, MOLPRINT 2D)	R^2 = 0.87, RMSE = 21.4 nm (λ_{peak}) R^2 = 0.79, RMSE = 4.7 nm (FWHM) KPLS	N ≈ 400 (Experimental data)	96
22	MR-TADF	PLQY (Prediction) (Inverse design)	RF, NN, XGB, TDM-VAE	Molecular descriptors; DFT-derived parameters;	(RF model) RMSE = 21.4 nm, r = 0.84, R^2 = 0.65	N ≈ 300 (DFT data)	81
23	TADF emitters	ΔE_{ST}; oscillator strengths (*f*) (Prediction)	PIML	Singlet/triplet excited state energy level (E_{S1}/E_{T1}); HOMO–LUMO overlap (S); Exchange integral (K); Dynamic spin polarization (P)	0.77 < r < 0.88 MAE < 0.1 eV (ΔE_{ST}) MAE < 0.1 eV (*f*)	N ≈ 39,000 (DFT data)	94