Communications in Computational Chemistry
Commun. Comput. Chem., Vol. 7, No. 3, 181–189
doi: 10.4208/cicc.2025.126.02
Review

Recent Progress of Machine Learning on Organic Optoelectronic Materials

Jinglei Fu[1, *], Shichen Zhang[1, *] and Xiaoyan Zheng[1, *]

[1]Key Laboratory of Cluster Science of Ministry of Education, Key Laboratory of Medicinal Molecule Science and Pharmaceutics Engineering of Ministry of Industry and Information Technology, Beijing Key Laboratory of Photoelectronic/ Electro-photonic Conversion Materials, School of Chemistry and Chemical Engineering, Beijing Institute of Technology, Beijing 100081, P. R. China.

[*] Corresponding authors: xiaoyanzheng@bit.edu.cn (Xiaoyan Zheng), Jingleifu@bit.edu.cn (Jinglei Fu), shichenzhang@bit.edu.cn (Shichen Zhang)

Received on 6 May 2025; Accepted on 2 June 2025

Abstract: Organic optoelectronic materials, owing to their exceptional photoelectronic properties, have extensive applications across diverse fields, such as lighting and display, photovoltaic devices, and bioimaging. Machine learning (ML) provides new opportunities for advancing research on organic optoelectronic materials. ML leverages existing datasets to establish robust input-output correlations for predicting material properties, thereby substantially reducing computational costs and enhancing efficiency. This review comprehensively explores recent progress on ML applications for organic optoelectronic material. We focused on three key aspects. First, we review applications ML in predicting photophysical properties of organic dyes, including absorption/emission wavelengths, quantum yields, and aggregation-induced emission/aggregation-caused quenching effects. Second, we examine ML applications in predicting subcellular targeting of fluorescent probes. Third, we discuss the role of ML in screening key descriptors for organic photovoltaics material. The advances in data science position ML as a pivotal tool for elucidating intricate structure-property correlations in molecular systems, driving the accelerated innovation of optoelectronic devices.

Key words: machine learning, organic luminescent materials, OPV materials, fluorescent probes.

1. Introduction

Organic optoelectronic materials have emerged as pivotal components in organic light emitting diodes (OLEDs)[1-3], fluorescent probes [2,4-5], and organic solar cells (OSCs)[6]. These materials have been used across diverse environments, including dilute solution, thin films, and crystalline states, where critical performance metrics, such as luminescent color, quantum efficiency, and lifetime, are highly sensitive to subtle change of environments [7-9]. Minor change of chemical structures of organic optoelectronic molecules significantly alters their macroscopic properties, which brings huge challenges in the rational design and performance optimization of organic optoelectronic materials.

Currently, the development of organic optoelectronic materials has largely based on experimental trial-and-error approaches, which are time-consuming and high-cost. Alternatively, theoretical calculation is an effective way to complement experimental techniques in molecular design of organic optoelectronic materials. Mul tiscale modeling approaches, including quantum mechanism (QM) [10-13], quantum mechanics/molecular mechanics (QM/

MM)[14], and molecular dynamics (MD)[15-16] simulations, have demonstrated remarkable success in simulating properties of organic optoelectronic materials across diverse environments, such as dilute solutions, crystalline lattices, and amorphous phase. While theoretical calculation face limitations in high-throughput screening of optimal-performance molecules among tens of thousands of organic compounds [4, 17-18]. In this context, machine learning (ML) recently has great progresses across disciplines such as property prediction and molecular design of organic optoelectronic materials [19-21] due to its exceptional efficiency in processing big and complex datasets [22-24]. By extracting molecular features from extensive databases and constructing input-to-output predictive models, ML can predict a wide range of properties without requiring explicit knowledge of underlying physicochemical mechanisms. For organic optoelectronic materials, ML can be applied to predict many key properties, such as luminescent color [25-26], quantum efficiency [27], and PCEs [28], not only helps accelerate the design and development of new organic optoelectronic materials, but also provides new strategies for improving material performance, injecting new vitality into the continuous progress in the field of organic optoelectronic materials [29-30].

In this review, we focus on the advancements of ML in organic optoelectronic materials, particularly in organic luminescent dyes, fluorescent probes, and organic photovoltaic (OPV). For organic luminescent dyes, ML models have been successfully developed to predict absorption/emission wavelengths, quantum yields, and aggregation-induced emission/aggregation-caused quenching (AIE/ACQ) effects, offering critical insights for designing novel high-efficiency light-emitting materials. For fluorescent probes, data-driven studies of structure-property relationship enable ML models to predict subcellular organelle targeting, accelerating the precise design of fluorescent probes with high performance. Furthermore, ML-driven screening of key descriptors about OPV, such as optical bandgap (E_g) and power conversion efficiency (PCE), has facil itated the development of strategies to enhance photovoltaic performance. This work comprehensively reviews these progresses and provides a critical evaluation of future directions for ML in organic optoelectronic material innovation.

2. ML predicted luminescent properties of organic dyes

2.1 Prediction of the maximum absorption wavelength of organic dyes

Organic fluorescent dyes have great potential in biological detection, and their photophysical properties, like λ_{abs}, significantly impact the quality of bioimaging. Therefore, it is important to training a ML model in predict λ_{abs} based on chemical structure and solvent information to guide the development of organic fluorescent dyes. Shao *et al*. established the SMFluo1 database, comprising 1,181 solvated small-molecule fluorophores spanning the ultraviolet–visible–near-infrared (UV–Vis–NIR) absorption window [26]. In their protocol, Morgan fingerprints and MACCS fingerprints were generated using RDKit, while molecular descriptors were calculated via the open-source tool ChemDes [31-32]. These features served as input to train deep learning architectures, including fully connected neural networks (FCNN) and convolutiona l neural networks (CNN), for predicting λ_{abs}. Hyperparameter optimization was implemented through ten-fold cross-validation in Figure 1a. It is revealed that the SMFluo1-DP system, which integrates MACCS and Morgan fingerprints through FCNN training, achieved optimal performance. The model exhibited a mean relative error (MRE) of 2.87% between predicted and experimental λ_{abs} values (Figure 1b). Then, the SMFluo1-DP model was employed to predict λ_{abs} of 120 out-of-sample solvated fluorescent dyes. It is found that SMFluo1-DP achieves the closest agreement with experimental data, exhibiting a MRE of 1.52%, significantly lower than the 10.89% MRE reported from the online platform ChemFluo[33]. The superior performance of SMFluo1-DP model underscores its potential as a robust ML modeling in handling molecules containing coumarin, BODIPY, rhodamine, squaraine, or cyanine scaffolds and accelerating the discovery and rational design of novel fluorescent dyes.

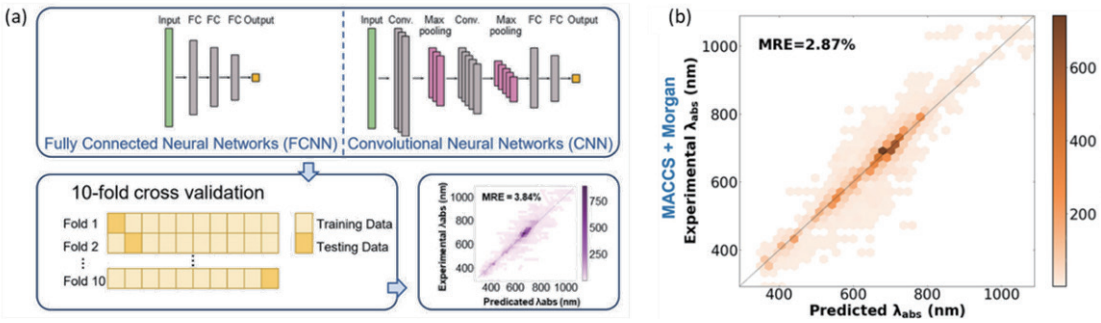


Figure 1. (a) Model development for the prediction of λ_{abs} using FCNN and DNN. (b) The MRE value of λ_{abs} predicted by the model using FCNN and the combination of Morgan and MACCS fingerprints. Copyright (2022) American Chemical Society.

2.2 Prediction of the absorption and emission spectra of organic dyes

Organic dyes, especially AIEgens, show great potential in fluorescence imaging due to their tunable spectra properties, however, the design of AIEgens with specific optical properties has proven challenging due to the dependence of their molecular optical properties on solvent polarity. Zhang *et al*. collected a database comprising 1,245 solvated AIEgens. Molecular structures and solvent information were converted into molecular descriptors, including Morgan circular fingerprints, Daylight fingerprints, topological torsion fingerprints, and atom-pair fingerprints [25]. Seven machine learning models, including support vector machine (SVM), K-nearest neighbors (KNN), multi-layer perceptron (MLP), gradient boosted regression trees (GBRT), random forest (RF), CNN, and extreme gradient boosting (XGBoost) were trained and a multimodal molecular descriptor strategy was proposed for

predicting λ_{abs} and emission wavelength (λ_{em}). It is demonstrated that the XGBoost model achieved the optimal performance in λ_{abs} prediction (Figure 2a), while the CNN model showed excellent prediction in λemi (Figure 2b). And the multimodal protocol significantly enhanced the model accuracy. Specifically, the XGBoost model combined with topological torsion-daylight fingerprints exhibited the lowest root mean square error (RMSE) of 15.2 nm for λ_{abs} prediction (Figure 2c), whereas the CNN model with atom-pair-daylight fingerprints achieved an RMSE of 19.2 nm for λ_{em} prediction (Figure 2d). Furthermore, machine learning outperformed traditional time-dependent density functional theory (TD-DFT) methods in both computational efficiency and prediction accuracy (Figure 2e-2f), highlighting its potential for high-throughput screening of organic fluorescent dyes. Three candidate molecules were selected based on ML screening and synthetic feasibility. Experimental validation confirmed the strong agreement between the predicted and measured λ_{abs} and λ_{em} across five solvent conditions. And the synthetic organic dyes were successfully applied in cellular fluorescence imaging and deep-tissue penetration imaging experiments, demonstrating the robust capability of machine learning in guiding the design and development of advanced organic dyes.

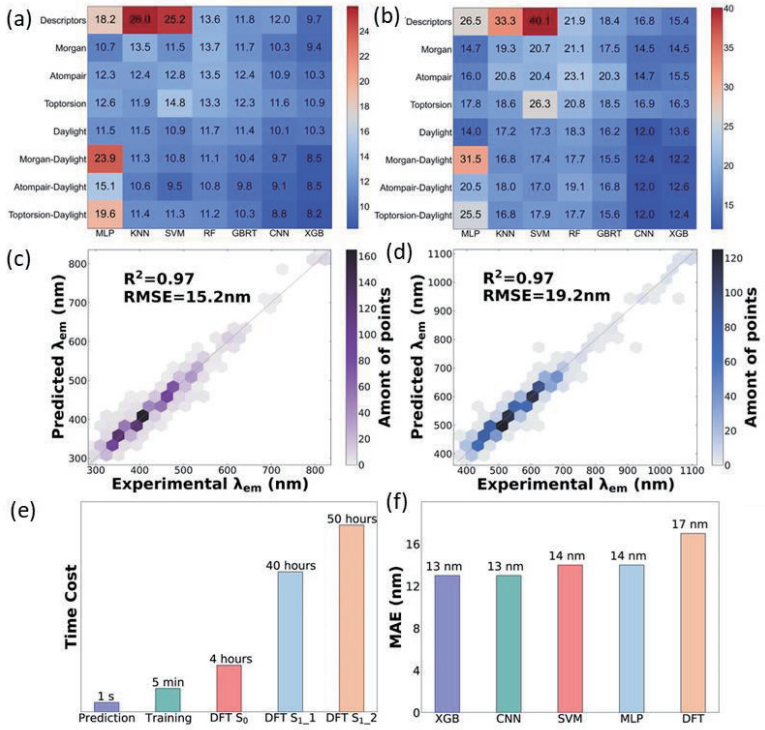


Figure 2. Mean absolute error (MAE) of (a) λ_{abs} and (b) λ_{em} of ML models with various molecular descriptors using 10-fold cross-validation. (c) The prediction results of λ_{abs} based on the XGB model with topological torsion-daylight fingerprints. (d) The prediction results of λ_{em} based on the CNN model with atom pair-daylight fingerprints. (e) The time cost for one molecule. (f) Comparison of MAE for predicting λ_{abs} between ML and TDDFT. Copyright (2023) BioMed Central.

2.3 Prediction of the quantum yield and spectra of organic dyes

Fluorescence quantum yield is also key parameter of organic fluorescent dyes. Ju *et al*. constructed a database containing over 4,300 solvated organic fluorescent dyes encompassing approximately 3,000 distinct compounds [33]. By systematically evaluating six molecular fingerprints (MACCS, CDK, CDK extended (CDKex), Morgan, E-state, and PubChem) in conjunction with seven machine learning algorithms (SVM, KRR, MLP, KNN, RF, LightGBM, and GBRT), see Figure 3a-3b. They developed a functionalized structural descriptor (FSD) and a comprehensive general solvent descriptor (CGSD) to efficiently encode molecular structures and solvent characteristics into the ML models [34]. It is demonstrated that the GBRT model achieved the best performance in predicting both λ_{em} (MAE=14.30 nm (0.066 eV)) and λ_{abs} (MAE=10.47 nm (0.070 eV)), and LightGBM model achieved the best performance in photoluminescence quantum yields (PLQYs) (MAE=0.11). Additionally, the capability of LightGBM/FSD model in predicting PLQYs under varying solvent conditions was evaluated by repartitioning the dataset. It is predicted that the PLQY values varied significantly for identical compounds across different solvent conditions, which closely aligns with experimental results, demonstrating the robust ability of the model in evaluating solvent effects, see Figure 3c-3d. Furthermore, to seek higher reliabilities than the regressors, a binary classifier for PLQY prediction was developed using the LightGBM/FSD model. Using the median of the experimental PLQY (0.25) as the threshold to divide the database into two groups, achieving to a satisfactory overall accuracy (86.8%). To verify the generalization ability of the GBRT/FSD model, they conducted an independent dataset containing 116 fluorescent dyes that were not included in the training set, for predicting λ_{em}. By incorporating a small portion (<17%) of the coumarins and naphthalimides molecules from the test set into the training set and retraining the ML model, the updated GBRT/FSD model demonstrated excellent performance (MAE=0.142), which shows the generalizability of GBRT/FSD model across diverse molecular systems (Figure 3e). In order to

further assess the performance of GBRT/FSD and LightGBM/FSD model, they collected 30 heterocyclic fluorescents, which are ionic as an additional test set. For PLQY prediction, the LightGBM/FSD model demonstrated robust performance with a MAE of 0.21 and an accuracy of 80%. Nevertheless, these are good results given the difficulty of PLQY predictions.

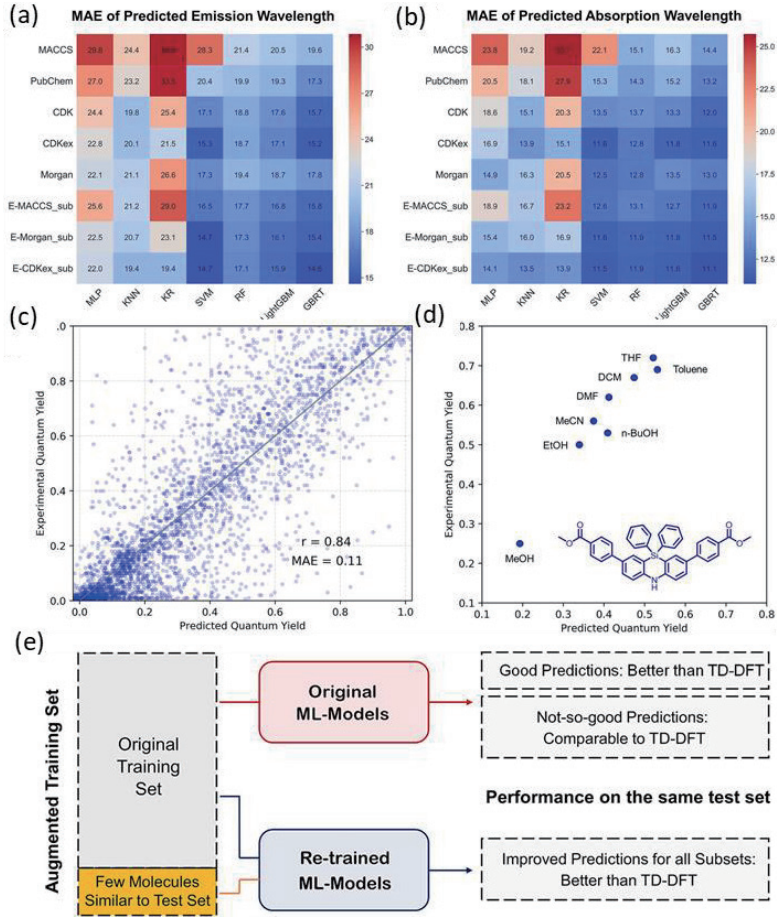


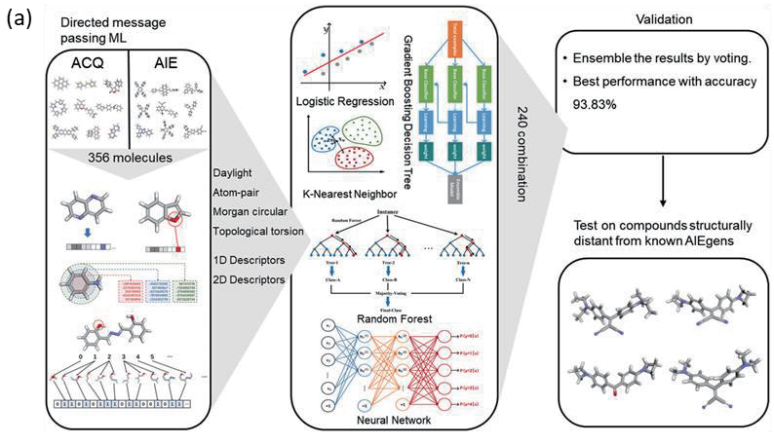
Figure 3. MAE testing results of (a) λ_{em} and (b) λ_{abs} by combining different ML models with different descriptors as inputs. (c) Linear correlation between experimental PLQY and LightGBM-predicted values. (d) Chemical structures and quantum yield of several typical compounds in different solvent conditions. (e) Schematic illustration for ML models improvement. Copyright (2021) American Chemical Society.

For λ_{abs}, λ_{em}, and quantum yields prediction of organic dyes, Mahato and Kumar *et al*. identified quadratic support vector machine (QSVM) as the optimal model for λ_{abs} and λ_{em} ($R^2 >$ 0.929), while cubic support vector machine (CSVM) indicated superior performance on quantum yield estimation using Morgan fingerprints and five solvent parameters as input features. [35] Additionally, Bi *et al.* found that the combined molecular fingerprints and RF and Gradient Boosting Regression (GBR) algorithms showed the best performance in predicting quantum yields (Area Under the Curve (AUC) > 0.84) and wavelengths (λ_{abs} MRE= 5.07%, λ_{em} MRE= 4.31%) [27].

2.4 Prediction of the AIE/ACQ properties of organic dyes

Organic dyes with AIE characteristics emit intense fluorescence in the aggregated state, it's exceptional features and adaptability are significant in biomedical imaging and optoelectronic devices. [36] The demand for efficient design of novel AIE materials with desired photophysical properties is growing. Developing a ML model for distinguishing AIE from ACQ efficiently is urgent. Xu *et al.* established a database containing 356 AIE/ACQ molecules and evaluated five machine learning algorithms (logistic regression (LR), KNN, GB, gradient boost (RF), and MLP) using four molecular fingerprints (Morgan, daylight, atom-pair, and topological) alongside quantitative 108 1D and 2D molecular descriptors [37]. By comparing different ML algorithms that take different combinations of descriptors as input, studies demonstrated that the quantitative RF/Daylight model has the best accuracy on the test set. And they developed a predictive ensemble voting strategy to improve the overall prediction performance by combining the prediction results of multiple models. Research demonstrated that the test accuracy of this ensemble method reached 93.83% (Figure 4a). Experimental validation was conducted on three structurally similar AIE/ACQ molecules. Notably, predictions from both the single-modal (RF/quantitative descriptors) and the multi-modal (ensemble model) approach perfectly aligned with experimental outcomes (Figure 4b), confirming the reliability of the model in predicting AIE properties. This methodology provides a robust strategy for accelerating the discovery of AIE materials.

Zhao *et al.* constructed a dataset comprising 3,074 AIE/ACQ molecules and developed a ML model for predicting AIE/ACQ properties using five molecular fingerprints (CDK, CDKex, E-state, substructure, and substructure count fingerprints) and six ML



Molecule	1	2	3
Actual	**ACQ**	**AIE**	**ACQ**
Single-modal strategy			
Logistic regression + Daylight	AIE	**AIE**	**ACQ**
Random Forest + Atom	**ACQ**	ACQ	**ACQ**
Random Forest +Topological	**ACQ**	ACQ	**ACQ**
Random Forest + Descriptors	**ACQ**	**AIE**	**ACQ**
MLPClassifier + Daylight	AIE	**AIE**	**ACQ**
Ensemble	AIE	ACQ	**ACQ**
Multi-modal strategy			
Random Forest + Morgan + Des	**ACQ**	ACQ	**ACQ**
Random Forest + Daylight + Des	AIE	ACQ	**ACQ**
MLPClassifier + Daylight + Des	AIE	**AIE**	**ACQ**
Ensemble	**ACQ**	**AIE**	**ACQ**

Figure 4. (a) ML Flowchart of the prediction of AIE/ACQ properties and experimental validation. (b) The chemical structures of three designed molecules. (c) Comparison of the prediction results and experimental results of different ML strategies for three molecules. Copyright (2022) Wiley-VCH.

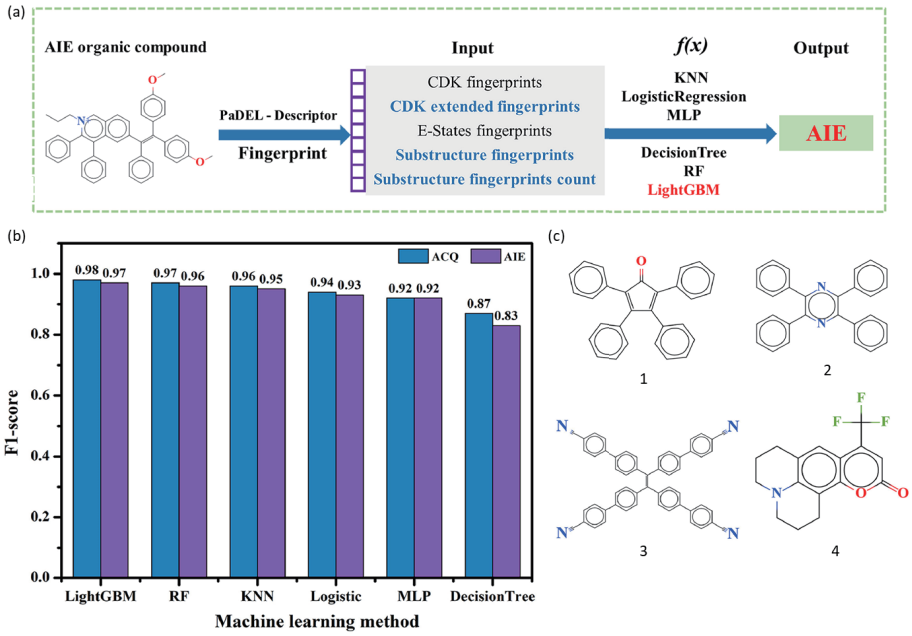


Figure 5. (a) ML model frameworks for the prediction of AIE property. (b) F1 score results of different ML methods. (c) Three AIE molecules (1-3) and one ACQ molecule (4) molecule identified through high-throughput screening protocol. Copyright (2024) Elsevier Science B.V.

algorithms (LightGBM, RF, LR, Decision Tree, MLP, and KNN), see Figure 5a [38]. It is demonstrated that the lightGBM model combining with three fingerprints (CDKex fingerprints, substructure fingerprints, and substructure count fingerprints) simultaneously as input features achieved the highest accuracy of 97.4% on the independent test set, surpassing the 93.83% accuracy reported in the prior work by Bin *et al* [37]. and it exhibited superior F1 scores compared to other models (Figure 5b). Then high-throughput screening using the above optimal-performing model identified four candidate molecules (Figure 5c), three of them (molecules 1-3) exhibited AIE characteristics, while the remaining one (molecule 4)demonstrating ACQ feature. Experimental results confirmed the ML predictions, demonstrating its capability on rapidly and accurately predicting AIE properties.

3. ML predicted subcellular targeting selectivity of fluorescent probes

Organelle-specific fluorescent probes are crucial for studying cellular organelles, but the relationship between the physicochemical properties of probes and their selectivity toward specific organelles is still unclear [39-40]. Park et al. constructed a quantitative structure-activity relationship (QSAR) model using the RF algorithm to predict the target organelles of fluorescent probes [41]. This model was trained on a dataset of 350 organelle-specific fluorescent probes, with 786 descriptors generated via JChem and PreADMET calculations. The model achieved 75% accuracy in classifying probes into nine categories: cytosol, endoplasmic reticulum (ER), Golgi body, lipid droplet (LD), lysosome, mitochondria, nucleus, plasma membrane (PM), and no entry (Figure 6a). By using mean decrease impurity (MDI) analysis, they identified 38 key descriptors, including LogD, pKa, hydrophilic-lipophilic balance (HLB), and topological polar surface area (TPSA), and so on (Figure 6b), that significantly influence on their selectivity to organelles. For instance, LD and PM targeting probes exhibited high hydrophobicity, while no-entry probes were characterized by permanent charges and elevated hydrophilicity. Nucleus-targeting probes typically carried positive charges, whereas mitochondria-targeting probes combined hydrophobicity with positive charges. These findings provide critical insights into the organelle-targeting mechanisms of fluorescent probes and offer a foundation for rational design of novel organelle-specific fluorescent probes.

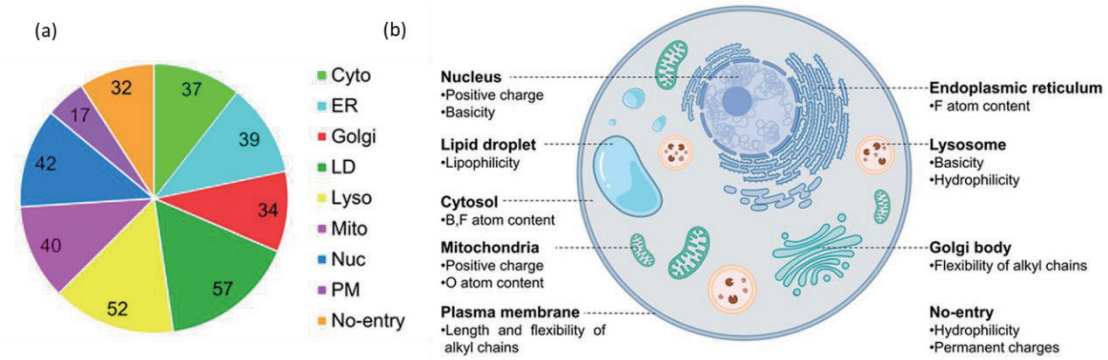


Figure 6. (a) Classification into 9 target organelles for database containing 350 fluorescent probes. (b) A visual representation of significant characteristics of probes that influence their localization, acquired through MDI analysis of the model. Copyright (2023) Multidisciplinary Digital Publishing Institute.

4. ML predicted key descriptors of OPV materials

4.1 Key descriptors of exciton binding energy

OPV materials are semiconductors that convert light energy into electricity via the photovoltaic effect [42-44]. E_b is a critical factor in determining the performance of organic optoelectronic devices, particularly in OSCs. A smaller E_b is beneficial for reducing energy loss and improving the efficiency of these devices. Previous theoretical calculations lack key descriptors, which linking E_b to molecular structures, restricting the advancement of OPV materials. Zhu *et al.* employed a data-driven ML approach to systematically screen the most critical descriptors for predicting the solid-state exciton binding energy (E_b^s) of OPV materials from eight initial descriptors, including gas-phase exciton binding energy (E_b^g), electrostatic components of polarization energy fo negative charge (P_{elst-}) and for positive charge (P_{elst+}), molecular dipole moment (μ), quadrupole moment (Θ), molecular polarizability (α), crystal density (ρ), and the ratio of void fraction to backbone packing coefficient ($\eta_{void/BB}$)[45]. A database comprising 135 data points was constructed, which included the E_b^s values of molecular crystals and eight associated descriptors. Pearson correlation analysis between these descriptors and E_b^s indicated that six descriptors (E_b^g, P_{elst-}, μ, Θ, α, $\eta_{void/BB}$)with high correlation coefficient were preliminarily identified (Figure 7a). Subsequently, five ML models (including LR, KNN, GBRT, RF, and XGBoost) were utilized to further evaluate descriptor importance. It was revealed that four descriptors (E_b^g, P_{elst-}, μ, and $\eta_{void/BB}$) displayed high importance in the XGBoost model. To further identify the most important descriptors, they remove the descriptors iteratively to observe the change in prediction performance. After removing the Pelst- and μ, the ML models retained excellent predictive performance (Pearson's correlation coefficient (r)>0.90), which indicates that E_b^g and $\eta_{void/BB}$ are indispensable key descriptors. Models relying solely on these two descriptors still maintained high prediction accuracy (r≈0.84), with XGBoost and GBRT achieving optimal performance (r=0.90), see Figure 7b. It is worth noting that both descriptors exhibited comparable importance in the XGBoost and GBRT algorithms (Figure 7c). Ultimately, E_b^g and $\eta_{void/BB}$ were conclusively identified as the core descriptors for predicting $E_b^s$with high accuracy (Figure 7d). And the model was also successfully employed to predict the E_b^s of two thin-film materials. Temperature-dependent photoluminescence spectra obtained experimentally validated predictive accuracy of the model, offering critical insights for the rational design of high-performance organic photovoltaic materials.

4.2 Key descriptors of power conversion efficiency

Screening the descriptors most relevant to photovoltaic performance is the key to advancing OPV technology. Han *et al.* proposed the ΔE_{ST} as a critical molecular descriptor for predicting the efficiency PCE of OPV materials [28]. Utilizing a dataset of 515 OPV devices, they constructed a database

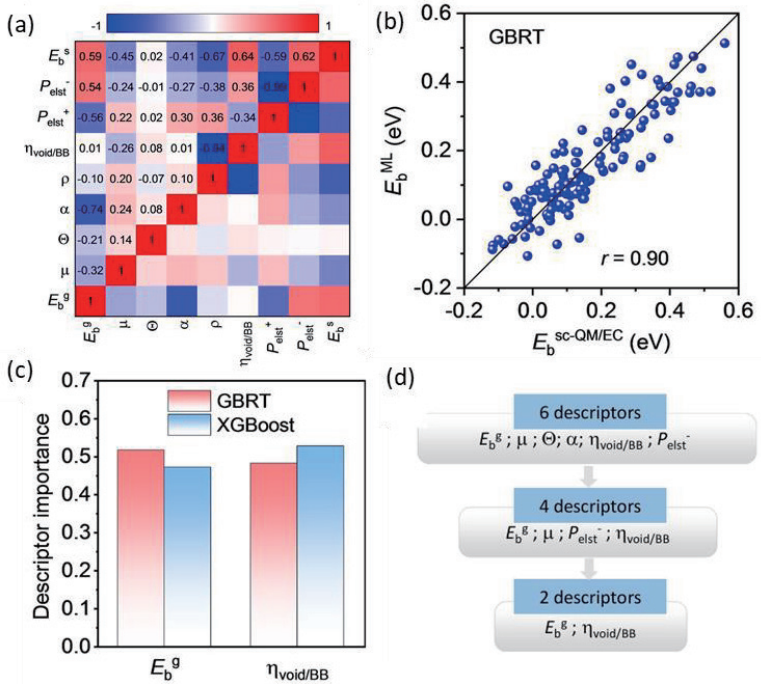


Figure 7. (a) Pearson's correlation coefficients among the selected descriptors. (b) Plots of ML predicted E_b^s by using E_b^g and $\eta_{void/BB}$ with GBRT model versus self-consistent quantum mechanics/embedded charge (sc-QM/EC) calculated E_b^s. (c) The descriptor importance analysis of E_b^g and $\eta_{void/BB}$ with GBRT and XGBoost algorithm. (d) Illustration of the screening process from six, to four, then to two descriptors. Copyright (2024) Wiley-VCH.

containing three molecular descriptors, including E_g, charge-transfer driving force (ΔE_{DA}), and ΔE_{ST} (Figure 8a). These descriptors were analyzed by using three ML algorithms (LR, KNN, and GBRT). Correlation analysis between these descriptors and four photovoltaic parameters, including open-circuit voltage (V_{OC}), short-circuit current density (J_{SC}), fill factor (FF), and power conversion efficiency (PCE), revealed that ΔE_{ST} exhibited a significantly higher Pearson correlation coefficient with PCE (r = 0.72) compared to E_g (r = 0.65) and ΔE_{CT} (r = 0.53), highlighting its superior predictive accuracy for PCE. Predictions of four photovoltaic parameters using different combinations of molecular descriptors via three ML methods demonstrated that GBRT model achieved the best performance. Notably, when employing a single descriptor, ΔE_{ST} yielded higher prediction accuracy for PCE (r = 0.72) than E_g (r = 0.65) and ΔE_{DA} (r = 0.53). Further investigation into the combined use of all three

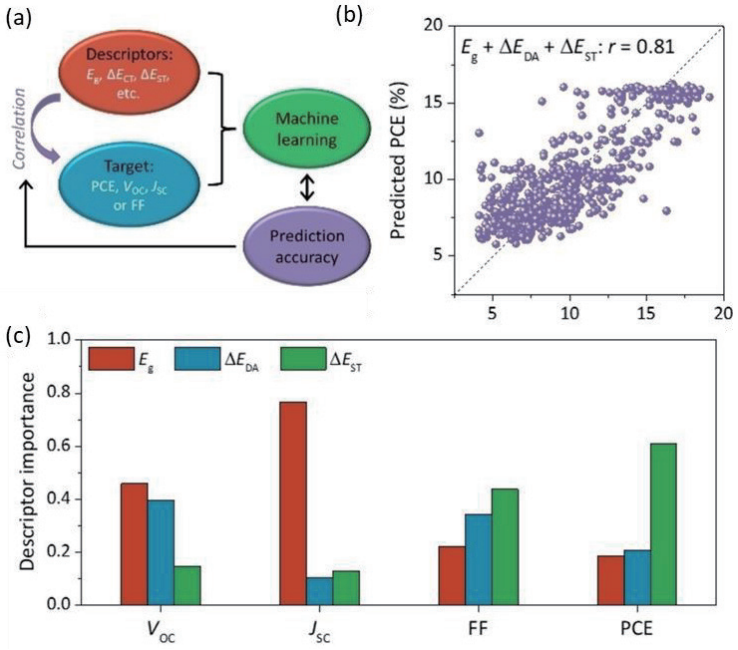


Figure 8. (a) Illustration of the role of machine learning in OPV prediction. (b) Predicted PCE by GBRT with all three combined descriptors versus experimental measured values. (c) The importance of three descriptors in four photovoltaic parameters in the GBRT models. Copyright (2022) Wiley-VCH.

descriptors showed a remarkable improvement in PCE prediction accuracy, achieving a Pearson's correlation coefficient of 0.81 (Figure 8b). At the same time, descriptor importance analysis confirmed that ΔE_{ST} dominated the prediction of PCE (Figure 8c). These results not only validate ΔE_{ST} as a pivotal descriptor for PCE prediction, but also demonstrate that integrating multiple descriptors significantly enhances predictive performance.

5. Summary and outlook

In this review, we highlight significant advancements in the application of ML in organic optoelectronic materials, focusing on three pivotal domains: organic light-emitting materials, fluorescent probes, and OPVs. Studies have demonstrated that ML accurately predicts absorption/emission wavelengths, quantum yields, and AIE/ACQ properties of organic dyes; identifies key molecular descriptors governing OPV performance; and predicts the subcellular localization of fluorescent probes, thereby providing critical guidance for the design of novel materials. By constructing models that autonomously extract data features, ML establishes complex correlations between molecular structures and material properties, significantly reducing computational costs while enhancing efficiency in processing large-scale datasets. However, challenges still persist. Current ML models exhibit limitations in predictive accuracy and interpretability, hindering a deeper mechanistic understanding of structure-property relationships. Future research should prioritize the development of robust and interpretable ML frameworks, expansion of high-quality databases, and exploration of advanced molecular descriptors to capture the intricate physicochemical behaviors of organic materials [46-47]. What's more, multi-fidelity methods that integrate experimental data with DFT and ML results may assist in closing the gap between theoretical and practical applications, the combination of DFT and ML will become the future trend in the research of organic optoelectronic materials research [48]. With advancements in computational power and data science, ML is poised to emerge as a transformative tool for the discovery and optimization of organic optoelectronic materials, accelerating the development of next-generation optoelectronic devices.

Acknowledgments

We thank the financially supportive of the National Natural Science Foundation of China (Grant 22173006), Beijing Natural Science Foundation (Grant 2222027).

References

[1] Zeng Y., Qu J., Wu G., Zhao Y., Hao J., Dong Y., Li Z., Shi J., Francisco J.S., Zheng X. Two key descriptors for designing second near-infrared dyes and experimental validation. *J. Am. Chem. Soc*., **146** (2024), 9888-9896.

[2] Chua M.H., Chin K.L.O., Loh X.J., Zhu Q., Xu J. Aggregation-induced emission-active nanostructures: beyond biomedical applications. *ACS Nano*, **17** (2023), 1845-1878.

[3] Kobaisi M.A., Bhosale S.V., Latham K., Raynor A.M., Bhosale S.V. Functional naphthalene diimides: synthesis, properties, and applications. *Chem. Rev*., **116** (2016), 11685-11796.

[4] Vendrell M., Zhai D., Er J.C., Chang Y.-T. Combinatorial strategies in fluorescent probe development. *Chem. Rev*., **112** (2012), 4391-4420.

[5] Tian X., Wang H., Cao S., Liu Y., Meng F., Zheng X., Niu G. Cationic tetraphenylethylene-based AIE-active acrylonitriles: investigating the regioisomeric effect, mechanochromism, and wash-free bioimaging. *J. Mater. Chem. C*, **11** (2023), 5987-5994.

[6] Liu Q., Jiang Y., Jin K., Qin J., Xu J., Li W., Xiong J., Liu J., Xiao Z., Sun K., Yang S., Zhang X., Ding L. 18% efficiency organic solar cells. *Sci. Bull*., **65** (2020), 272-275.

[7] Fang F., Zhu L., Li M., Song Y., Sun M., Zhao D., Zhang J. Thermally activated delayed fluorescence material: an emerging class of metal-free luminophores for biomedical applications. *Adv. Sci*., **8** (2021), 2102870.

[8] Inganäs O. Organic photovoltaics over three decades. *Adv. Mater*., **30** (2018), 1900388.

[9] Kwok R.T.K., Leung C.W.T., Lam J.W.Y., Tang B.Z. Biosensing by luminogens with aggregation-induced emission characteristics. *Chem. Soc. Rev*., **44** (2015), 4228-4238.

[10] Yang J., Wei H., Ou Q., Li Q., Peng Q., Zheng X. Theoretical study of the photocyclization reaction-induced dual aggregation-induced emission phenomenon. *J. Phys. Chem. A*, **128** (2023), 217-224.

[11] Shuai Z., Wang D., Peng Q., Geng H. Computational evaluation of optoelectronic properties for organic/carbon materials. *Acc. Chem. Res*., **47** (2014), 3301-3309.

[12] Kim E., Koh M., Lim B.J., Park S.B. Emission wavelength prediction of a full-color-tunable fluorescent core skeleton, 9-aryl-1,2-dihydropyrrolo [3,4-b]indolizin-3-one. *J. Am. Chem. Soc*., **133** (2011), 6642-6649.

[13] Hennefarth M.R., King D.S., Gagliardi L. Linearized pair-density functional theory for vertical excitation energies. *J. Chem. Theory Comput*., **19** (2023), 7983-7988.

[14] Zhang W., Liu J., Jin X., Gu X., Zeng X.C., He X., Li H. Quantitative prediction of aggregation-induced emission: a full quantum mechanical approach to the optical spectra. *Angew. Chem. Int. Ed*., **59** (2020), 11550-11555.

[15] Wang H., Gong Q., Wang G., Dang J., Liu F. Deciphering the mechanism of aggregation-induced emission of a quinazolinone derivative displaying excited-state intramolecular proton-transfer properties: a QM, QM/MM, and MD study. *J. Chem. Theory Comput*., **15** (2019), 5440-5447.

[16] Liu Z., Kalin M.L., Liu B., Cao S., Huang X. Kinetic network models to elucidate aggregation dynamics of aggregation-induced emission systems. *Aggregate*, **5** (2023), e422.

[17] Sanz-Velasco A., Amargós-Reyes O., Kähäri A., Lipinski S., Cavinato L.M., Costa R.D., Kostiainen M.A., Anaya-Plaza E. Controlling aggregation-induced emission by supramolecular interactions and colloidal stability in ionic emitters for light-emitting electrochemical cells. *Chem. Sci*., **15** (2024), 2755-2762.

[18] Zhang Y., Xiong C., Wang W., Dai W., Ren Y., Xia J., Li G., Shi J., Tong B., Zheng X., Shao X., Cai Z., Dong Y. Wide-range color-tunable afterglow emission by the modulation of triplet exciton transition processes based on buckybowl structure. *Aggregate*, **4** (2023), e310.

[19] Cerezo M., Verdon G., Huang H.-Y., Cincio L., Coles P.J. Challenges and opportunities in quantum machine learning. *Nat. Comput. Sci*., **2** (2022), 567-576.

[20] Biamonte J., Wittek P., Pancotti N., Rebentrost P., Wiebe N., Lloyd S. Quantum machine learning. *Nature*, **549** (2017), 195-202.

[21] Dong S., Wang P., Abbas K. A survey on deep learning and its applications. *Comput. Sci. Rev*., **40** (2021), 100379.

[22] Carleo G., Cirac I., Cranmer K., Daudet L., Schuld M., Tishby N., Vogt-Maranto L., Zdeborová L. Machine learning and the physical sciences. *Rev. Mod. Phys*., **91** (2019), 045002.

[23] Kitchin J.R. Machine learning in catalysis. *Nat. Catal*., **1** (2018), 230-232.

[24] Zhang Z.-Y., Liu X., Shen L., Chen L., Fang W.-H. Machine learning with multilevel descriptors for screening of inorganic nonlinear optical crystals. *J. Phys. Chem. C*, **125** (2021), 25175-25188.

[25] Zhang Y., Fan M., Xu Z., Jiang Y., Ding H., Li Z., Shu K., Zhao M., Feng G., Yong K.-T., Dong B., Zhu W., Xu G. Machine-learning screening of luminogens with aggregation-induced emission characteristics for fluorescence imaging. *J. Nanobiotechnol*., **21** (2023), 107.

[26] Shao J., Liu Y., Yan J., Yan Z.-Y., Wu Y., Ru Z., Liao J.-Y., Miao X., Qian L. Prediction of maximum absorption wavelength using deep neural networks. *J. Chem. Inf. Model*., **62** (2022), 1368-1375.

[27] Bi H., Jiang J., Chen J., Kuang X., Zhang J. Machine learning prediction of quantum yields and wavelengths of aggregation-induced emission molecules. *Materials*, **17** (2024), 1664-1699.

[28] Han G., Yi Y. Singlet-triplet energy gap as a critical molecular descriptor for predicting organic photovoltaic efficiency. *Angew. Chem. Int. Ed*., **61** (2022), e202213953.

[29] De Angelis F. The impact of machine learning in energy materials research: the case of halide perovskites. *ACS Energy Lett*., **8** (2023), 1270-1272.

[30] Pflüger P.M., Glorius F. Molecular machine learning: the future of synthetic chemistry? *Angew. Chem. Int. Ed*., **59** (2020), 18860-18865.

[31] Abraham A., Pedregosa F., Eickenberg M., Gervais P., Mueller A., Kossaifi J., Gramfort A., Thirion B., Varoquaux G. Machine learning for neuroimaging with scikit-learn. *Front. Neuroin*f., **8** (2014), 00014.

[32] Dong J., Cao D.-S., Miao H.-Y., Liu S., Deng B.-C., Yun Y.-H., Wang N.-N., Lu A.-P., Zeng W.-B., Chen A.F. ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. *J. Cheminf*., **7** (2015), 60-70.

[33] Ju C.-W., Bai H., Li B., Liu R. Machine learning enables highly accurate predictions of photophysical properties of organic fluorescent materials: emission wavelengths and quantum yields. *J. Chem. Inf. Model*., **61** (2021), 1053-1065.

[34] Kim S., Chen J., Cheng T., Gindulyte A., He J., He S., Li Q., Shoemaker B.A., Thiessen P.A., Yu B., Zaslavsky L., Zhang J., Bolton E.E. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res*., **47** (2019), D1102-D1109.

[35] Mahato K.D., Kumar U. Optimized machine learning techniques enable prediction of organic dyes photophysical properties: absorption wavelengths, emission wavelengths, and quantum yields. *Spectrochim. Acta Part A*, **308** (2024), 123768.

[36] Zhang S.-C., Zhu J., Zeng Y., Mai H.-Q., Wang D., Zheng X.-Y. Interpretable prediction of aggregation-induced emission molecules based on graph neural networks. *Chem. Commun*., to appear..

[37] Xu S., Liu X., Cai P., Li J., Wang X., Liu B. Machine-learning-assisted accurate prediction of molecular optical properties upon aggregation. *Adv. Sci*., **9** (2021), 2101074.

[38] Zhao Y., Chen K., Yu B., Wan Q., Wang Y., Tang F., Li X. Development of organic aggregation-induced emission fluorescent materials based on machine learning models and experimental validation. *J. Mol. Struct*., **1317** (2024), 139126.

[39] Horobin R.W., Rashid-Doubell F., Pediani J.D., Milligan G. Predicting small molecule fluorescent probe localization in living cells using QSAR modeling. 1. overview and models for probes of structure, properties and function in single cells. *Biotechnic & Histochemistry*, **88** (2013), 440-460.

[40] Cruz V.L., Martinez S., Ramos J., Martinez-Salazar J. 3D-QSAR as a tool for understanding and improving single-site polymerization catalysts. a review. *Organometallics*, **33** (2014), 2944-2959.

[41] Park S.-H., Lee H.-G., Liu X., Lee S.K., Chang Y.-T. Quantitative structure-activity relationship of fluorescent probes and their intracellular localizations. *Chemosensors*, **11** (2023), 310.

[42] Benduhn J., Tvingstedt K., Piersimoni F., Ullbrich S., Fan Y., Tropiano M., McGarry K.A., Zeika O., Riede M.K., Douglas C.J., Barlow S., Marder S.R., Neher D., Spoltore D., Vandewal K. Intrinsic non-radiative voltage losses in fullerene-based organic solar cells. *Nat. Energy*, **2** (2017), 53.

[43] Qian D., Zheng Z., Yao H., Tress W., Hopper T.R., Chen S., Li S., Liu J., Chen S., Zhang J., Liu X.K., Gao B., Ouyang L., Jin Y., Pozina G., Buyanova I.A., Chen W.M., Inganäs O., Coropceanu V., Brédas J.L., Yan H., Hou J., Zhang F., Bakulin A.A., Gao F. Design rules for minimizing voltage losses in high-efficiency organic solar cells. *Nat. Mater*., **17** (2018), 703-709.

[44] Eisner F.D., Azzouzi M., Fei Z., Hou X., Anthopoulos T.D., Dennis T.J.S., Heeney M., Nelson J. Hybridization of local exciton and charge-transfer states reduces nonradiative voltage losses in organic solar cells. *J. Am. Chem. Soc*., **141** (2019), 6362-6374.

[45] Zhu L., Huang M., Han G., Wei Z., Yi Y. The key descriptors for predicting the exciton binding energy of organic photovoltaic materials. *Angew. Chem. Int. Ed*., **64** (2024), e202413913.

[46] Fabrizio A., Meyer B., Corminboeuf C. Machine learning models of the energy curvature vs particle number for optimal tuning of long-range corrected functionals. *J. Chem. Phys*., **152** (2020), 154103.

[47] Westermayr J., Marquetand P. Machine learning for electronically excited states of molecules. *Chem. Rev*., **121** (2021), 9873-9926.

[48] Nyangiwe N.N. Applications of density functional theory and machine learning in nanomaterials: a review. *Next Mater*., **8** (2025), 100683.