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.

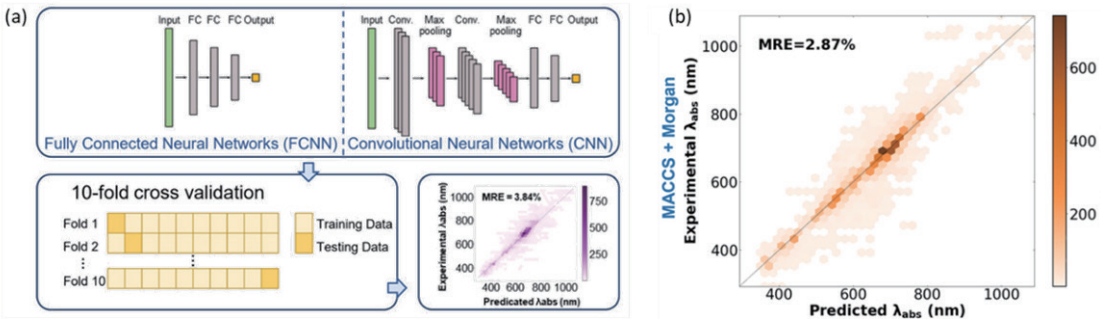

Figure 1. (a) Model development for the prediction of λ_{abs} using FCNN and DNN. (b) The MRE value of λ_{abs} predicted by the model using FCNN and the combination of Morgan and MACCS fingerprints. Copyright (2022) American Chemical Society.

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