

Review

Machine Learning-Assisted Optimization and Application of Carbon-Based Emitters

Bo Lian, Xianzhi Song, Jiameng Du, Runqing Fan, Linjuan Yang, Wenran Zhang, Yang Zhang *, Xiaohong Li, Yunchao Li, Lin Shen * and Louzhen Fan *

College of Chemistry, Key Laboratory of Theoretical & Computational Photochemistry of Ministry of Education, Beijing Normal University, Beijing 100875, China

* Correspondence: y.zhang@bnu.edu.cn (Y.Z.); lshen@bnu.edu.cn (L.S.); lzf@bnu.edu.cn (L.F.)

How To Cite: Lian, B.; Song, X.; Du, J.; et al. Machine Learning-Assisted Optimization and Application of Carbon-Based Emitters. *Materials Matter* **2026**, *1*(1), 5.

Received: 26 May 2026

Revised: 19 June 2026

Accepted: 7 July 2026

Published: 22 July 2026

Abstract: Machine learning (ML) is progressively being integrated into materials science, exhibiting great potential for optimizing chemical synthesis and structural regulation, thereby accelerating intelligent design and efficient exploration of novel materials. Carbon-based emitters (CBEs), as emerging functional materials, have attracted considerable attention due to their tunable photoluminescence properties, abundant precursor sources, and structural versatility. To date, extensive research on CBEs has generated a substantial data foundation, laying the groundwork for ML driven structural screening and property prediction. Although the application of ML in CBEs remains in its early stages compared to that of inorganic semiconductors and organic-inorganic hybrid perovskites, its potential to accelerate material screening and reveal structure-property relationships is becoming increasingly evident. Given the transformative role ML has played in other functional materials, it is expected to drive a paradigm shift in CBEs research from conventional trial-and-error approaches to data-driven, intelligence-guided design, substantially accelerating material discovery and expanding their functional applications. Therefore, this review systematically summarizes recent advances in ML applications in CBEs, focusing on the general ML workflow, property prediction and structural design strategies for organic small-molecules and carbon dots (CDs), and their applications in organic light-emitting diodes (OLEDs), quantum dot light-emitting diodes (QLEDs), information encryption, biomedicine, and sensing. Finally, this review discusses the key challenges currently facing the field and offers perspectives on future directions for ML-driven CBEs research, aiming to provide guidance for the rational design and efficient development of such materials.

Keywords: machine learning; carbon-based emitters; structure-property relationship; data-driven paradigm

1. Introduction

In recent years, artificial intelligence (AI) has undergone rapid advancements, profoundly reshaping the landscape of scientific research and practical applications [1,2]. Spanning from natural language processing to computer vision, AI algorithms possess powerful data processing and pattern recognition capabilities which have been widely applied in field of bioinformatics, drug discovery, and materials science, thereby establishing innovative paradigms for addressing complicated scientific challenges [3]. As a key driving technology of AI, machine learning (ML) has emerged as a powerful tool for accelerating materials design and discovery, which enables rapid property prediction, high-throughput screening, and efficient exploration of the vast chemical spaces [4–6]. Concurrently, the



Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Materials Genome Initiative, along with recent advances in automated and high-throughput experimental methodologies, has facilitated the massive and rapid accumulation of materials data, providing a solid foundation for the integration of ML into materials science [5,7]. By efficiently screening promising candidates and systematically navigating high-dimensional synthesis parameter spaces, ML is revolutionizing the traditional paradigm of materials discovery, marking a transformative era from experience-driven to data-driven innovation.

This data-driven research paradigm opens up a new avenue for exploring the complex structure-property-application relationships of carbon-based emitters (CBEs). As a diverse class of optoelectronic materials, CBEs cover carbon dots (CDs), carbon quantum dots (CQDs), carbonized polymer dots, and organic small-molecules [8–11]. Benefiting from their exceptional optoelectronic properties, CBEs hold tremendous promise for applications in bioimaging, optoelectronic devices, information encryption, and sensing [12–18]. In recent years, diverse CBEs with multicolor emissions, tunable narrowband emission, and high photoluminescence quantum yield (PLQY) have been successfully synthesized, laying a solid data foundation for ML-based investigations [12,19]. ML demonstrates prominent capabilities in the multi-property prediction of optoelectronic materials, enabling accurate prediction of key photophysical parameters such as emission wavelength and PLQY, as well as facilitating structural feature recognition and high-throughput virtual screening [20–22]. Building on increasingly mature forward prediction capabilities, the field is now advancing toward “inverse design”, where performance prediction models are employed to deduce target structures and synthesis parameters that satisfy the specific property requirements [23,24]. By identifying key structural features that govern material performance, inverse design strategies significantly shorten the discovery cycle of candidate materials and enable the goal-oriented development of high-performance CBEs, further highlighting the evolution of AI in the CBEs field, shifting from passive property prediction to active materials design [25,26]. More importantly, the in-depth integration of ML with automated synthesis platforms promises to accelerate the translation of novel materials into practical optoelectronic technologies.

Although ML applications in CBEs still remain at an emerging stage compared with inorganic semiconductors and organic-inorganic hybrid perovskites, they have already demonstrated considerable potential in expediting material screening and elucidating structure-property relationships. Herein, we present a comprehensive and critical review of recent advances in ML for the design and application of CBEs, aiming to further broaden the scope application of ML-assisted CBEs. First, the general ML workflow is introduced, following by a systematic discussion of ML-driven structural design and property prediction of CBEs, especially organic small-molecules and CDs. We further highlight representative applications of ML-assisted strategies in electroluminescent material screening, information encryption and anti-counterfeiting, biomedicine and sensing. Finally, we elaborate on the existing challenges and provide perspectives, offering a feasible roadmap for the rational design and efficient development of ML-assisted CBEs.

2. ML Workflow

ML is an interdisciplinary field integrating statistics, computer science, and optimization theory. Its primary aims to empower computers to learn latent patterns from massive datasets, build predictive models, and implement decision-making and inference tasks without intricate and time-consuming programming [27,28]. In contrast to conventional research methods rely on predefined theoretical models as well as physical and chemical equations, ML feature as a purely data-driven paradigm. It can automatically extract meaningful features from complex, high-dimensional data and establish predictive relationships, which makes it a powerful tool for materials science [29,30]. Practically, ML adopts a standardized workflow (Figure 1) covering data collection, feature engineering, model selection, and model evaluation and interpretation [31,32]. As illustrated in Figure 1, experimental and computational datasets are initially collected and converted into machine-accessible descriptors, followed by feature refinement and model development via representative ML algorithms. The pre-trained models are then evaluated through statistical indicators and interpretability analyses which establish the reliable structure-property relationships, facilitates the rational design and optimization [33].

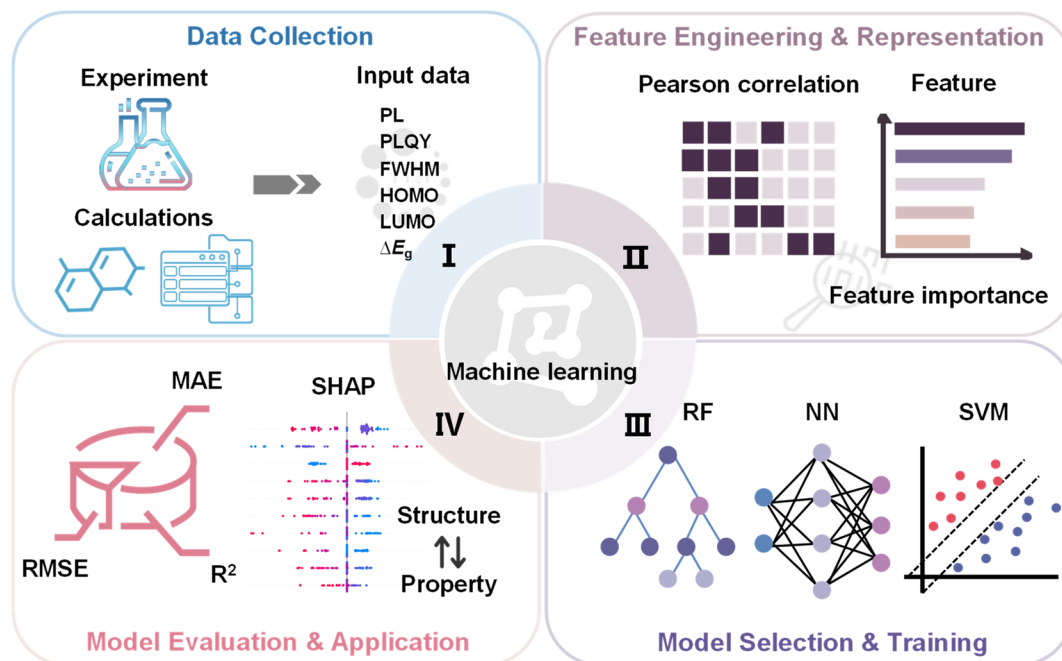


Figure 1. The general workflow of the ML approach in material and related applications.

2.1. Data Collection

As an ML paradigms rely on inherently data-driven, the acquisition, scale and quality data serve as the cornerstone for model construction and performance optimization [34]. In materials science, ML models are typically trained using the multidimensional datasets that integrate synthetic parameters, intrinsic material properties obtained from experimental measurements and theoretical calculations. Key intrinsic properties include photoluminescence (PL) emission wavelength, PLQY, full-width at half-maximum (FWHM), highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO), and HOMO-LUMO gap (ΔE_g) [35]. Beyond experimental and computational data, virtual libraries of target candidate molecules could be augmented using fragment-based combinatorial expansion and computational enumeration, offering a new paradigm for high-throughput screening and ML modeling [36].

2.2. Feature Engineering and Representation

Following initial data preprocessing, feature engineering and representation is crucial for applying ML to CBEs development that converting intricate chemical and experimental data into model-readable numerical descriptors [37]. Typically, this process includes feature extraction, selection, and representation, which profoundly determines the prediction accuracy and generalization capability of constructed ML models [38].

Feature engineering is predominantly implemented molecular structural information. With the help of cheminformatics tools like RDKit, material structures can be converted into SMILES representations and molecular fingerprints. The fingerprints could encode molecular backbone, functional groups, topological characteristics, and local chemical environment. Subsequently, Pearson correlation heatmaps are widely employed to visualize the correlations among multiple features [39]. In particular, feature optimization is commonly performed via feature selection or dimensionality reduction algorithms, which is represented by Principal Component Analysis (PCA), for eliminating redundant or low-correlation features.

2.3. Model Selection and Training

ML algorithms possess distinct advantages that handling nonlinear relationships, adapting to varying sample sizes, and processing diverse feature dimensions, rendering model selection a critical step tailored to specific research objectives and data characteristics. Currently, commonly used ML models in the field of CBEs could be generally classified into two categories: traditional ML methods, including Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Gradient Boosting (XGBoost), and deep learning (DL) approaches represented by Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), and Deep Convolutional Neural Networks (DCNNs) [40,41]. These models form a complementary ML toolkit, facilitating robust structure-property correlation modeling and efficient data-driven exploration in CBEs research.

Conventional ML algorithms exhibit high data efficiency and suit for small- and medium-sized datasets (typically $< 10^3$ samples), which are common in experimental CBE researches. RF offers are robust against outliers and collinearity that is compatible with mixed-inputs variables (e.g., synthetic parameters and spectral features), and capable of outputting feature importance rankings, making it ideal for predicting optical properties such as PLQY and emission wavelength. However, RF still suffers from weak extrapolation performance outside the training range. XGBoost could achieves superior prediction accuracy via the optimized gradient boosting and inherent regularization [42]. Nevertheless, XGBoost requires elaborate hyper parameter and is more sensitive to outliers than RF. Currently, SVM and KNN still remain valuable for small sample scenarios (tens to hundreds of data points) and classification tasks, such as predicting whether CBE mission exceeds a target wavelength threshold [43,44]. Their main limitations lie in poor scalability to large datasets and high sensitivity to feature normalization.

DL models architectures including ANNs and CNNs dominate high-dimensional unstructured data such as spectral profiles or molecular images [45]. They can automatically extract hierarchical feature representations and enabling high-precision predictions of multiple optical properties (PLQY, FWHM, and emission wavelength, etc.). However, DL typically requires a relatively larger datasets over 10^4 samples to alleviate over-fitting and consumes substantial computational resources [46,47]. GNNs are particularly promising for CBE research as they directly process molecular graphs to extract luminescence-related topological and electronic features [48]. GNNs could capture structure-property relationships without manual feature engineering, but they generally require even larger labeled datasets and longer training cycles in comparison with ANNs. Generative models (e.g., variational autoencoders, generative adversarial networks) represent a cutting-edge tool for inverse CBEs design, generating novel molecular skeletons with targeted optical properties [49]. Their key limitations consist of difficult quality evaluation of generated molecule quality and overreliance for extensive training data.

In practical application, ML model selection depends on dataset size, task category (e.g., regression, classification and generation), interpretability requirements, and computational costs. For typical experimental CBE datasets with limited samples, RF or XGBoost often strike an optimal trade-off between accuracy and interpretability. DL and GNNs are preferred when large datasets or molecular graph inputs are available, whereas generative models are more suitable for de novo materials design. These ML models are commonly combined with feature engineering (e.g., one-hot encoding, normalization) and hyperparameter optimization (e.g., grid search, Bayesian optimization) to enhance generalization and predictive accuracy [50].

2.4. Model Evaluation and Application

Model evaluation is indispensable for certifying the reliability and generalizability of ML predictions. In the supervised learning framework, data partitioning and cross-validation is typically employed to assess model performance that mitigating overfitting and improving generalization toward unseen samples [51]. Common evaluation metrics include the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE), which are utilized to quantify prediction accuracy of continuous photophysical parameters [52]. For classification tasks, accuracy (AUC), recall, and cross-validation scores are frequently combined for a comprehensive model assessment [53].

Beyond performance evaluation, two complementary approaches have emerged as crucial supplements to ML-driven CBEs research. First, Bayesian optimization could address challenges of small sample sizes and high experimental noise by efficiently navigating parameter spaces via surrogate models. It enables the simultaneous optimization of model hyperparameters and synthetic conditions while considerably minimizing experimental cost. Second, model interpretability, represented by SHapley Additive ExPlanations (SHAP) analysis, is developed to unravel structure-property relationships and identify the dominant factors governing the intrinsic material properties of diverse CBE systems [54].

3. ML-Assisted Optimization and Design of CBEs

CBEs have attracted widespread attention due to their distinctive optical properties and structural diversity. In comparison with traditional inorganic phosphors or organic small molecules, CBEs are generally prepared from abundant and low-cost carbon precursors and their structures and properties could be tailored through various synthetic strategies, including bottom-up and top-down methodology [13]. Based on their structural characteristics, CBEs family include CQDs, CDs, carbonized polymer dots, and organic small-molecules [55]. Organic small-molecules exhibit prominent superiority in structural precision, tunable optoelectronic properties, and manipulated molecular design. Recently, CDs have been extensively investigated owing to their tunable emission color, superior photostability, and excellent biocompatibility [56,57]. More importantly, the structural diversity endows

CDs with versatile luminescent behaviors, including excitation-dependent emission, narrowband emission, room-temperature phosphorescence (RTP), aggregation-induced emission (AIE) and thermally activated delayed fluorescence (TADF) [58–60]. Benefiting from these unique merits, CBEs have demonstrated great application potentials, including bioimaging, optoelectronic devices, information encryption, anti-counterfeiting, as well as chemical and biological sensing. In recent years, ML has been increasingly implemented in the research of CBEs, especially for CDs, CQDs and organic small-molecules, where comprehensive datasets and methodological frameworks have been established [61]. Accordingly, this review focuses on these two representative CBE material systems, with representative ML-assisted studies summarized in Table 1.

Table 1. ML approaches for various practical applications of CBEs in recent years.

Number	CBEs	Input	Output	Method	ML Algorithm	Performance	Application	Ref.
1	Organic Small-Molecule	Molecular fingerprints	k _{TADF}	Regression	NN	R ² = 0.94	OLED	[62]
2	Organic Small-Molecule	Molecular descriptors	ISC/RISC	Regression	GBRT RF-GBRT-KNN	R ² = 0.74 R ² = 0.73	Material design	[63]
3	Organic Small-Molecule	Photophysical parameters, molecular descriptors	PLQY PLQY ≥ 80% PL-peak	Regression Classification Regression	XGBoost NN LightGBM	R ² = 0.7 AUC = 0.878 /	Material design	[64]
4	Organic Small-Molecule	Photophysical parameters, molecular descriptors, SMILES	PLQY TDM	Regression Regression	XGBoost RF	R ² > 0.8 R ² = 0.6	OLED	[54]
5	Organic Small-Molecule	Molecular fingerprints	PL-peak FWHM	Regression Regression	Kernel-based partial least-squares (KPLS)	R ² = 0.87 R ² = 0.79	OLED	[65]
6	Organic Small-Molecule	Molecular descriptors, MACCS, FP2 fingerprints	PL-peak FWHM PLQY PL-peak FWHM PLQY	Classification Regression	Ensemble learning model (BNML)	AUC = 1 AUC = 1 AUC = 0.96 R ² = 0.96 R ² = 0.98 R ² = 0.5	Material design	[66]
7	CDs	Experimental parameters	PLQY	Regression	XGBoost	0.6 < R ² < 0.8	Sensing	[20]
8	CDs	Experimental parameters	PLQY	Regression	XGBoost	AUC = 0.94	Guided synthesis	[53]
9	CDs	Experimental parameters	FL color	Classification	DCNN	AUC = 0.82	Guided synthesis	[67]
10	CDs	Experimental parameters	Stokes shift, PL-peak, PLQY	Regression	RF	AUC = 0.8 (12 new CDs)	Guided synthesis	[68]
11	CQDs	Experimental parameters	PL-peak PLQY	Regression	XGBoost	MSE = 0.1 MSE = 0.15	Guided synthesis	[23]
12	Organic Small-Molecule	Molecular graph	HOMO LUMO	Regression	Graph convolutional network (GCN)	MAE = 0.148 MAE = 0.163	OLED	[69]
13	CQDs	PL intensities	Correlated color temperature (CCT)	Regression	XGBoost	R ² = 0.717	QLED	[70]
14	CDs	Precursor features	PLE-peak PL-peak Afterglow lifetime	Regression	XGBoost	R ² = 0.97 R ² = 0.99 R ² = 0.98	Information encryption, LED	[71]
15	CDs	Experimental parameters	PL-peak RTP-peak FL Stokes shift RTP Stokes shift	Regression	XGBoost	R ² = 0.95 R ² = 0.94 R ² = 0.89 R ² = 0.94	Information encryption	[72]
16	CDs	Experimental parameters	PL-peak	Regression	XGBoost	R ² = 0.96	Sensing, Biomedicine	[73]
17	CDs	Experimental parameters	Formation yield	Classification	NN	AUC = 0.987	Biomedicine	[74]
18	CDs	RGB values	Metal ion	Classification	Multi-layer perceptron (MLP)	AUC = 0.797	Sensing	[75]
19	CDs	Fluorescence spectra	Concentrations of Cu ²⁺	Classification	KNN	AUC = 1 (1 μM), 0.905 (10 nM), 0.725 (100 μM)	Sensing	[76]

3.1. Organic Small-Molecule

One of the earliest applications of ML in organic small-molecule materials could be traced back to 2016. Alán Aspuru-Guzik and his colleagues successfully integrated ML with quantum chemical calculations to achieve large-scale virtual screening and rational design of blue-emitting molecules with TADF property [62]. In their works, NN models were adopted to screen out potential TADF molecules, laying out a solid foundation for ML-assisted OLEDs research.

Since this pioneering work, ML has been widely employed to explore innovative materials. As a representative example, Chen et al. applied gradient boosting regression trees (GBRT) and an RF-GBRT-KNN ensemble algorithm to perform high-throughput virtual screening on 564 TADF candidate molecules which were constructed from 20 building blocks based on D-A, D- π -A, and D-A-D configurations (Figure 2) [63]. Nine promising molecules with excellent intersystem crossing (ISC) and reverse intersystem crossing (RISC) characteristic were finally obtained, with Pearson correlation coefficients reaching 0.89 and 0.82 for ISC and RISC rate prediction. This work highlights the powerful capability of ML in investigating the complicated excited-state processes.

Accurate prediction of the PLQY of organic small-molecule emitters is crucial for the development of high-efficiency OLEDs. In successive research works, Song et al. progressively promoted the ML-assisted design of blue TADF emitters via multiple-stage strategies [64]. Initially, they integrated SHAP interpretability framework with the Light Gradient Boosting Machine (LightGBM) algorithm to establish quantitative relationship between structure and PLQY properties, enabling the simultaneous prediction of emission wavelength and PLQY values of potential candidate materials. This work offers rational important guidance for the theoretical molecular design of blue-emissive emitters. Based on the established predictive framework, they further embed transition dipole moment (TDM) as a key optimization objective into the variational autoencoder (VAE) model, thereby achieving the inverse design of emitters with increased radiative transition. Among the top-ranked molecules, D1-0236 was synthesized, which exhibits deep-blue emission peak centered at 451 nm along with an exceptionally PLQY up to 96.9% [54].

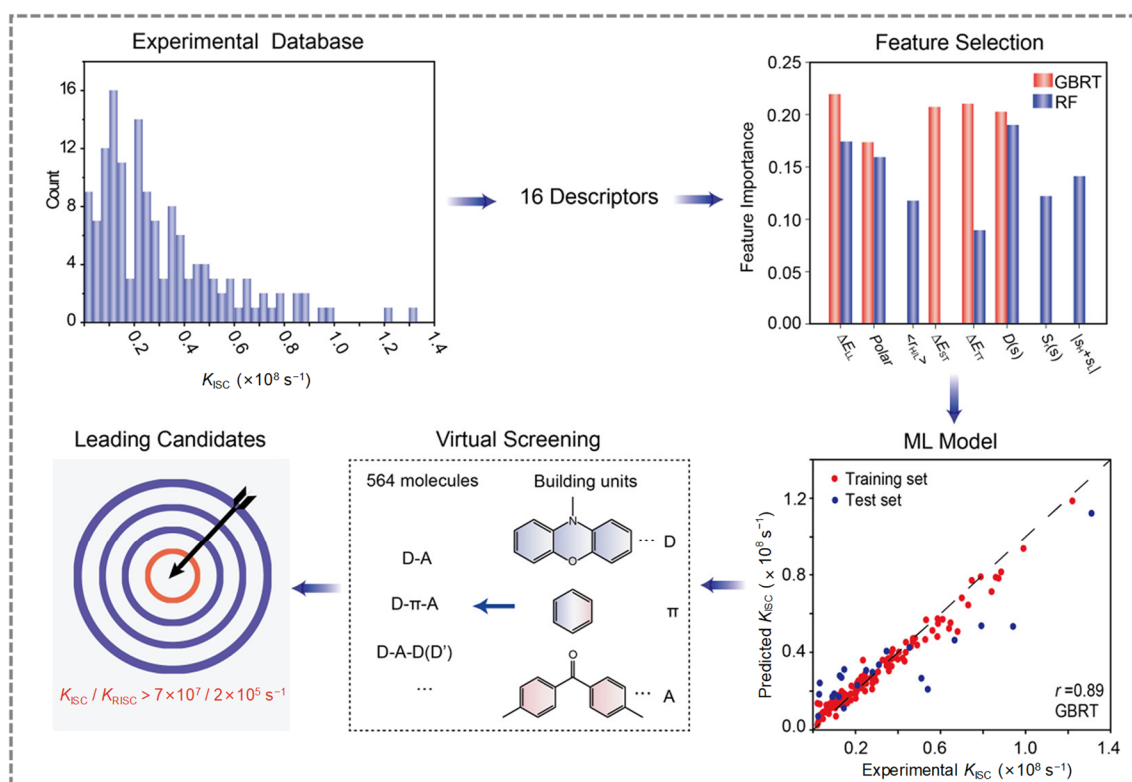


Figure 2. Research frame for ML-assisted virtual screening of TADF molecules with efficient ISC and RISC. Reproduced with permission from Ref. [63]. Copyright 2022, American Chemical Society.

Recent advances in ML applied to organic small-molecule research have evolved toward multi-objective optimization. A quantitative structure-property relationship was established by Adachi et al., achieving cross-validation to construct kernel-based partial least-squares (KPLS) models and identify intrinsic correlations among molecular structure, emission wavelength and FWHM (Figure 3a) [65]. Guided by this strategy, a deep-blue emitter

named ν -DABNA-O-xy was developed (Figure 3b), exhibiting an ultranarrow FWHM of 19 nm (Figure 3c). Subsequently, Liu et al. put forward a structure-aware ensemble learning model for the simultaneous prediction of multiple photophysical properties of TADF emitters, including emission wavelength, FWHM and PLQY [66]. Apart from its exceptional predictive accuracy, this model further revealed the contributions of donor moieties and π -conjugated skeletons to luminescent properties. The methodology was further validated by synthesizing newly design molecules and corresponding characterizations.

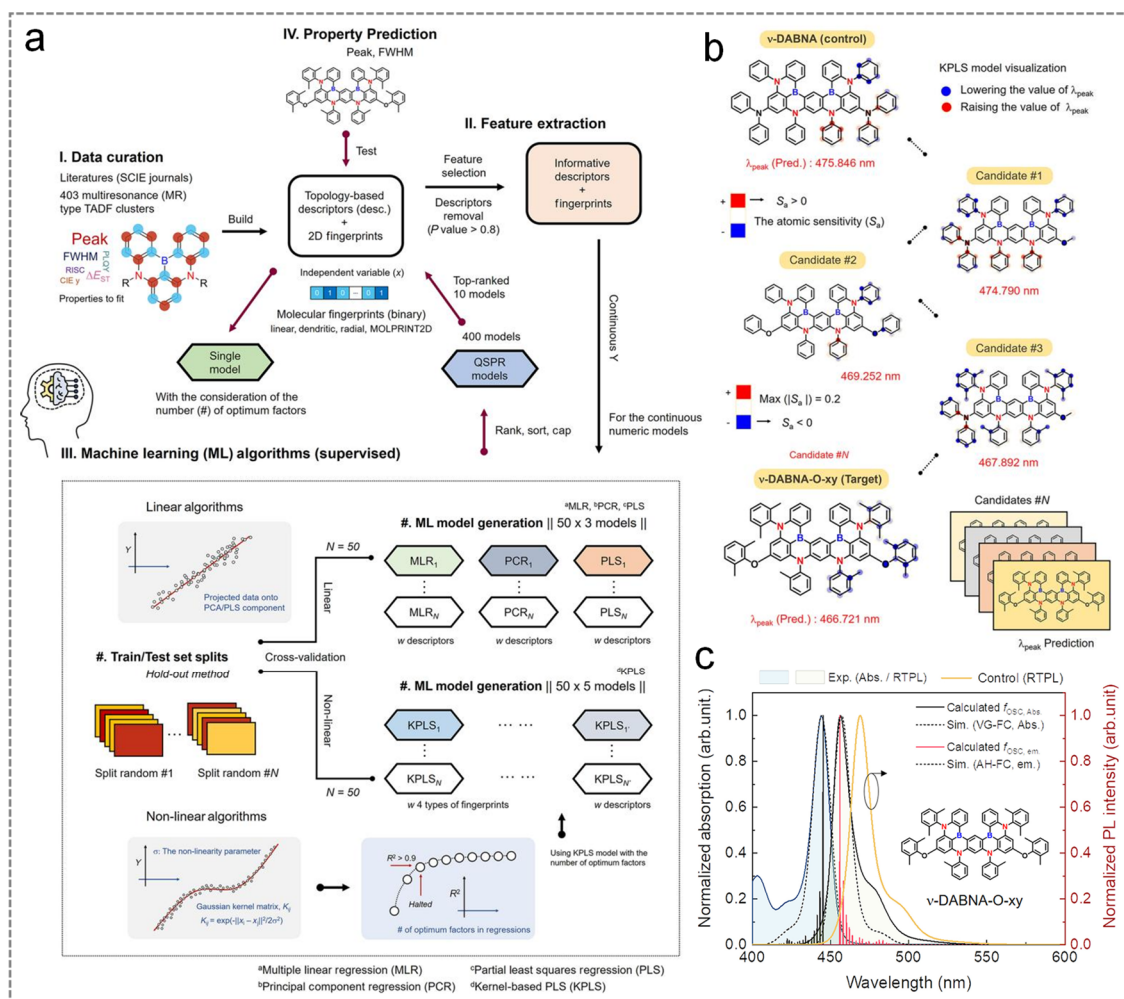


Figure 3. (a) An outline of ML-driven molecular design. (b) Interpretable model visualization from radial-based KPLS regression model for λ_{peak} prediction. (c) Absorption spectrum of ν -DABNA-O-xy and fluorescence spectra for the target and control molecules in toluene. Reproduced with permission from Ref. [65]. Copyright 2025, American Association for the Advancement of Science.

3.2. CDs

As a novel type of luminescent carbon-based nanomaterials, CDs has drawn extensive attention owing to their abundant precursor sources, facile synthetic conditions, high chemical and thermal stability, and favorable biocompatibility [77–80]. In our previous studies, we reported the synthesis of onion-like multicolor TADF CDs, achieving PLQYs of 42.3–61.0% and corresponding electroluminescent devices reached maximum external quantum efficiency (EQE_{max}) of 6.0–9.9% [81]. More recently, we successfully synthesized oxygen-doped carbon quantum rings (OD-CQRs) consisting of twelve benzene rings and six alternatively embedded five-membered oxygen-containing heterocycles. OD-CQRs exhibited a fluorescence emission peak centered at 393 nm with an ultranarrow FWHM of 18 nm and high PLQY of 95% [82].

Driven by growing demand for rational design and optimization of CDs, ML has emerged as a powerful tool for predicting their optical properties [40,83]. The pioneering application of ML to luminescent CDs was reported in 2019 [84]. Subsequent studies have constructed hydrothermal synthesis datasets based on key experimental parameters, such as ethylenediamine dosage, precursor mass, reaction temperature, heating rate, and reaction time, which were subsequently used to train ML models for PLQY prediction [20]. By integrating multiple algorithms,

such as XGBoost, multilayer perceptron (MLP), SVM, and Gaussian process (GP), reliable predictive models were established, which facilitated the synthesis of green CDs with a PLQY of 39.3%. Feature importance analysis further verified that precursor mass served as a dominant factor governing PLQY properties of CDs (Figure 4a). On this base, expanded datasets and optimized models (e.g., XGBoost regressor) have been applied to the high-throughput screening of synthetic conditions. For instance, through screening a total of 1,555,840 parameter combinations, 11 promising conditions were successfully identified, allowing the preparation of CDs with a PLQY as high as 55.5% [85].

Beyond PLQY, the accurate prediction of emission wavelengths and identification of fluorescence color is critical for advancing the functional applications of CDs in LEDs, bioimaging, and information encryption [70]. DL approaches represented by DCNNs have been introduced to extract intricate spectral features of CDs, which enables not only PLQY prediction but also reliable classification fluorescence colors (Figure 4b). DCNNs models with multiple hidden layers that trained on experimentally measured photoluminescence datasets of diverse CDs, possess powerful capabilities in learning spectral representations and predicting fluorescence colors [53,67]. Although the prediction accuracies of up to 81% have been achieved for blue/green CDs, model performance still remains limited for red and multicolor CDs system. Given the irreplaceable role of red CDs in biomedical imaging, their low yield and high cost, classification-based ML strategies have been developed to address this challenge. By converting emission colors information into categorical labels and integrating XGBoost with dimensionality reduction algorithms such as PCA, these optimized frameworks achieve enhanced feature extraction and improved model interpretability [53]. After sufficient training, the established models were capable of predicting the formation of red-emissive CDs under specific synthesis conditions (Figure 4c) and exhibited favorable generalization performance in cross-validation, highlighting the great potential of ML techniques for the targeted regulation of CDs emission color.

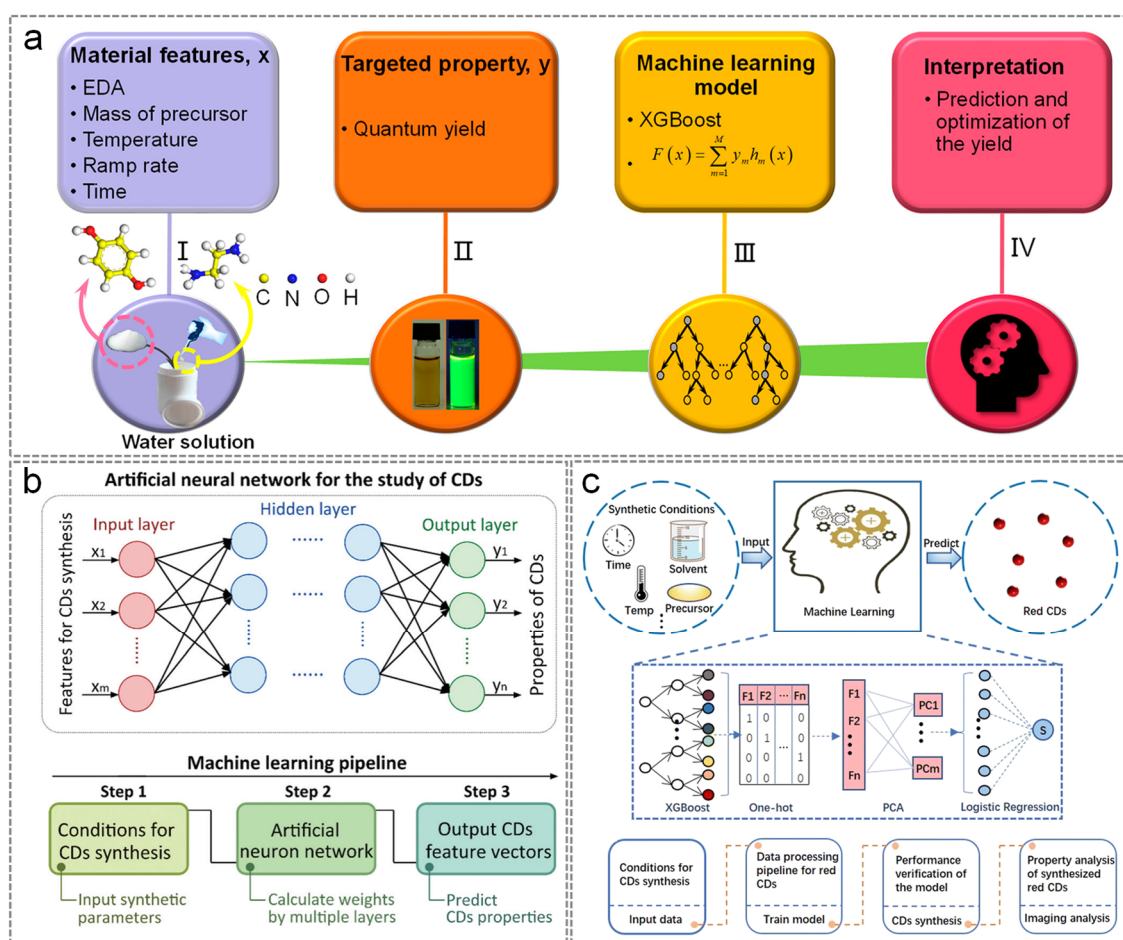


Figure 4. (a) Design workflow for ML guided synthesis of CLNMs. Reproduced with permission from Ref. [20]. Copyright 2020, American Chemical Society. (b) ANN assisted CDs to predict emission wavelength. Reproduced with permission from Ref. [67]. Copyright 2021, Royal Society of Chemistry. (c) The process of creating synthetic red fluorescent CDs-based on ML to improve the synthesis efficiency. Reproduced with permission from Ref. [53]. Copyright 2022, Royal Society of Chemistry.

Building upon these advances, ML techniques have been further extended to multicolor CDs systems. Through utilizing datasets derived from as-reported CDs in experimentally synthesized, RF models are capable of simultaneously predicting key photoluminescence properties, including PLQY, maximum emission wavelength, and Stokes shift, with overall accuracies approaching 80% (Figure 5a,b) [68]. Furthermore, this model enabled the optimization of synthetic routes by establishing quantitative correlation between fluorescence characteristics, reaction parameters and solvent properties. Benefiting from their excellent emission color tunability, such multicolor CDs have been successfully employed for high-capacity information encryption. Subsequent researches have introduced ML-assisted multi-objective optimization (MOO) strategies to regulate the hydrothermal synthesis of CQDs. With hyperparameters optimized via grid search, XGBoost was adopted to develop two independent regression models for the prediction of emission wavelength and PLQY (Figure 5c) [23]. Remarkably, trained only on a small dataset consisting of 63 experimental samples, the established models successfully guided the controllable synthesis of full-color fluorescent CQDs with PLQY of up to 60% (Figure 5d). Under data-scarce circumstances, this MOO framework effectively uncovered intrinsic correlations between synthetic parameters and luminescent properties, achieving the collaborative optimization of multiple luminescence-related indicators through a unified objective function. Collectively, these findings highlighted the merit of ML techniques in achieving efficient property regulation and rational structural design with complicated synthetic systems.

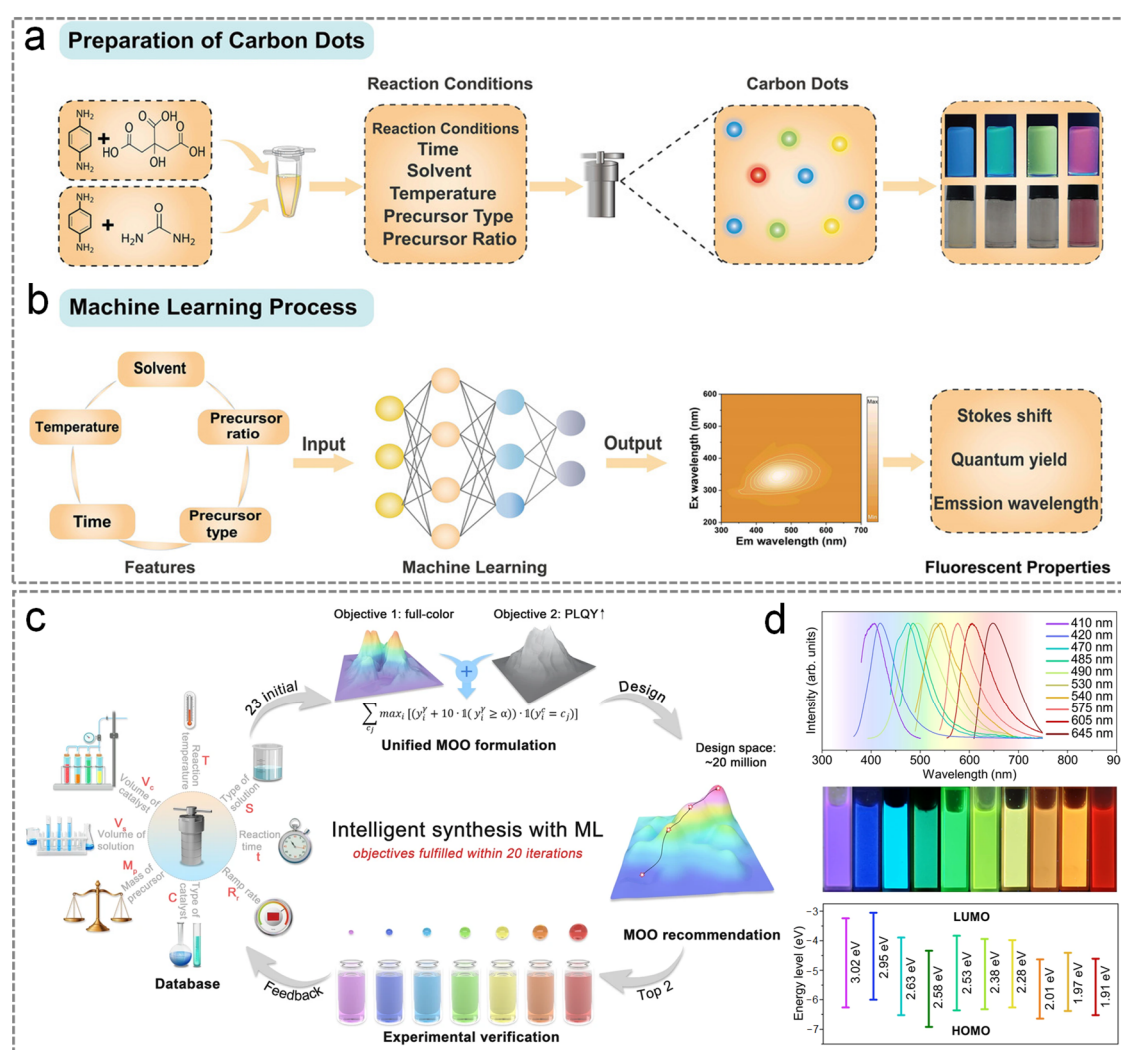


Figure 5. (a) Preparation of multicolor CDs to establish training datasets for ML. (b) Exploring the relationship between the synthesis parameters and the fluorescence properties of CDs, and predicting the PLQY of multicolor CDs using ML models. Reproduced with permission from Ref. [68]. Copyright 2023, John Wiley and Sons. (c) It consists of four key components: database construction, multi-objective optimization MOO formulation, MOO recommendation, and experimental verification. (d) Normalized PL spectra of as-synthesized CQDs. Reproduced with permission from Ref. [23]. Copyright 2024, Springer Nature.

4. Application of ML Optimized CBEs

4.1. OLEDs

The emissive layer of OLEDs is typically composed of electroluminescent molecules, which play a decisive role in determining device efficiency and optical characteristic. Benefiting from high efficiency and high color purity, OLEDs have been regarded as one of next-generation display technology for applications in high-definition television panels and flexible displays [86–88]. Conventional fluorescent emitters only utilized single excitons, which restricts the utilization electrically generated excitons to approximately 25%, thereby posing a bottleneck to further improvements in device efficiency [89]. Newly-developed TADF materials could utilize both singlet and triplet excitons for the enhancement of exciton utilization efficiency and EQE by converting triplet excitons into singlet states through RISC process [89,90]. ML techniques have thus emerged as a powerful tool for accelerating the rational design and efficient screening of OLEDs emitters [32,91,92].

Pioneering research has demonstrated the capability of ML in screen high-potential OLEDs candidate molecules, leading to optimized devices with a EQE_{max} of 22% [62]. More recently, ML models were trained on a dataset comprising HOMO and LUMO energy level of 3026 organic molecules in both solution and solid states. The established models realized high-precision prediction of frontier orbital energies, with a mean absolute error as low as 0.058 eV. Guided by this model, a deep-blue fluorescent OLED was fabricated using the screened emitter and suitable host material, exhibiting narrowband emission centered at 412 nm with a FWHM of 36 nm and a favorable EQE_{max} of 6.58% (Figure 6a–f) [69].

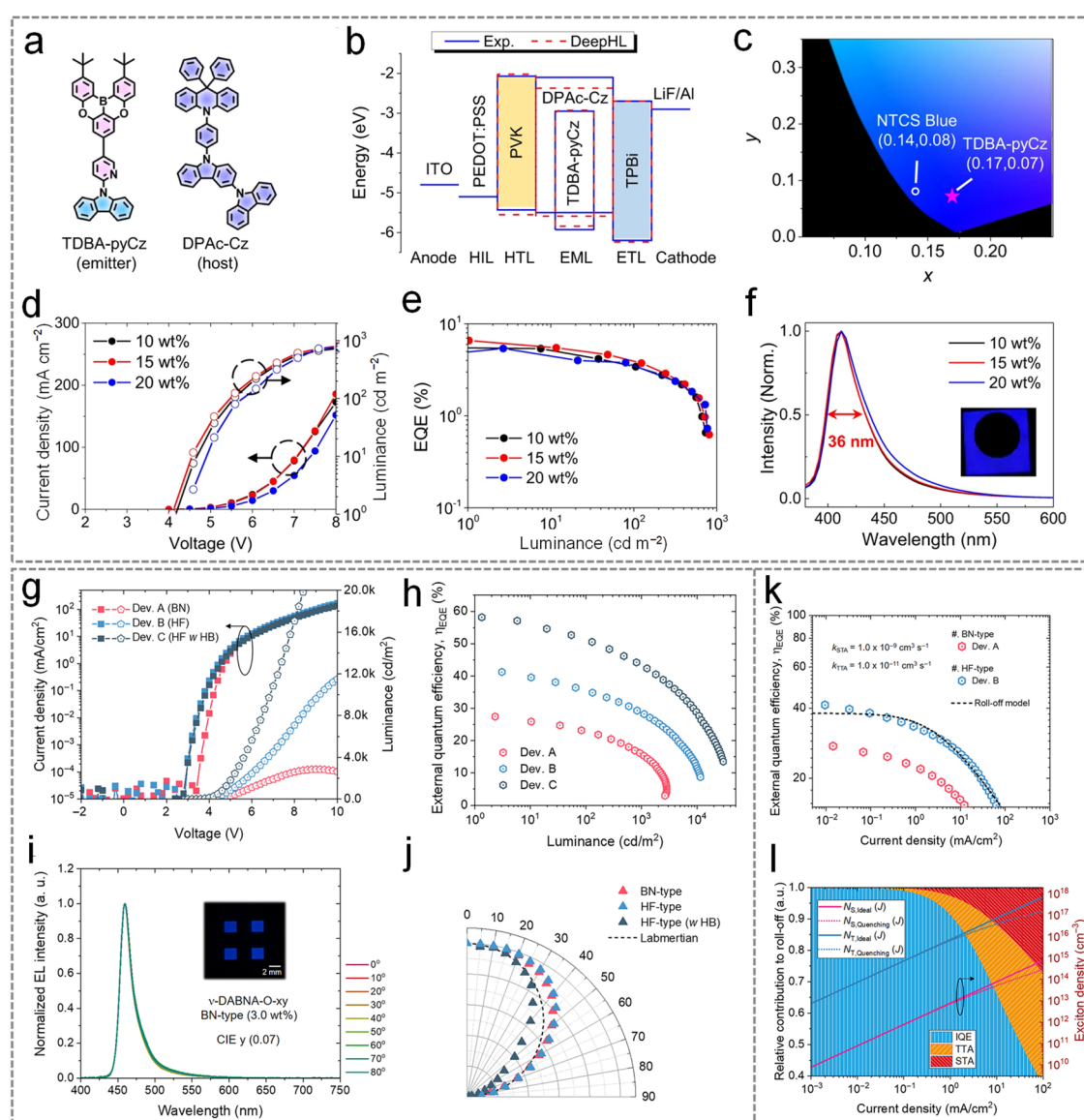


Figure 6. (a) Molecular structures of DPAC-Cz (host) and TDBA-pyCz (emitter). (b) Device structure. (c) CIE 1931 chromaticity diagram. (d) Current density (J)-Voltage (V)-Luminance (L) curves. (e) EQE-Luminance (L)

curves. (f) EL spectra at different doping concentrations of emitters at a luminance of 500 cd m⁻². Reproduced with permission from Ref. [69]. Copyright 2022, Springer. (g) *J-V-L* characteristics. (h) EQE-luminance curves of tested OLED samples. (i) Angle-dependent EL spectra for BN-type. (j) EL intensity profiles in steps of 5°. (k) *J-EQE* curves for BN- and HF-type OLED samples. (l) Quantification of roll-off in *ν*-DABNA-O-xy-based HF architecture. Reproduced with permission from Ref. [65]. Copyright 2025, American Association for the Advancement of Science.

Recently, Adachi et al. proposed a ML-assisted molecular design strategy to accelerate the screening of high-performance deep-blue TADF emitters [65]. The *ν*-DABNA-O-xy-based OLED device based on this molecule exhibited exceptional color purity with a CIE_y coordinate of 0.07 and an ultranarrow FWHM of 19 nm. Implemented in a bilayer emitting architecture, the optimized OLED delivered a EQE_{max} of 27.5% and further improved to 41.3% as integrated into a hyper fluorescence configuration (Figure 6g–i). Collectively, these representative studies highlight the advantage of ML in high-throughput screening CBEs emitters and performance optimization of OLEDs device. This data-driven paradigm provides a powerful support for the development of high-efficiency and high-color-purity OLEDs.

4.2. QLEDs

CQDs have emerged as promising emitters for QLEDs owing to their tunable luminescence wavelength and high PLQY [93,94]. ML has become an effective data-driven approach to the development and optimization of CQD-based QLEDs [95].

Green-emissive CQDs with PL peak centered at 493 nm and PLQY of 63.5%, as well as orange-emissive CQDs with PL peak at 553 nm and PLQY of 83.6%, were synthesized using perylene-3,4,9,10-tetracarboxylic dianhydride. QLED devices were fabricated with these emitters achieved EQE_{max} of 1.98% and 1.76%, respectively. Furthermore, ML models including XGBoost and multilayer perceptron (MLP) were employed to predict the correlated color temperature of the trichromatic mixtures of blue, green, and orange CQDs, covering a wide correlated color temperature (CCT) range from 3093 to 11,018 K. With ML-assisted modulation of RGB component ratios, white QLEDs with continuously tunable emission spanning from warm white to cool white were successfully fabricated. The optimized cool-white device delivered CIE coordinates of (0.33, 0.33), EQE_{max} of 1.18%, and a maximum luminance of 598 cd m⁻² [70]. These results validate that ML could efficiently establish the relationships among CQDs structural characteristic, optical properties, and device performance, thereby accelerating the screening of high-performance emitters and structural optimization of QLEDs devices.

4.3. Information Encryption and Anti-Counterfeiting

Driven by the increasing demands for information encryption and anti-counterfeiting, CDs with RTP property emerged as promising candidates for multidimensional optical encoding systems due to their tunable emission color, persistent afterglow and excellent chemical and photophysical stability [96–98]. RTP CDs could achieve multiplexed information storage that relying on their tunable emission wavelength and lifetime. Their time-resolved luminescence properties further significantly enhance anti-counterfeiting reliability, endowing RTP CDs with great potential for high-security encryption labels and dynamic optical anti-counterfeiting systems [99–101].

Since the first discovery of the afterglow emission from CDs in 2013, substantial advances have been achieved in the synthesis methodology and luminescence modulation of RTP CDs [102]. However, controllable preparation of multi-color phosphorescent CDs still remains challenges due to the intricate synthetic variables and ambiguous structure-property relationship. To address these issues, ML-assisted approaches integrating XGBoost and Bayesian optimization have been introduced to predict key phosphorescence parameters, including optimal excitation wavelength, emission wavelength, and phosphorescence lifetime, thereby guiding microwave-assisted synthesis of RTP CDs with phosphorescence lifetimes up to 3.43 s (Figure 7) [71]. Furthermore, a comprehensive ML framework established on a dataset consisting of 210 multicolor RTP CD samples could simultaneously predict PL and RTP emission peaks as well as their corresponding Stokes shifts. The established model achieved a R² as high as 0.9, low MAE and RMSE. Guided by this model, various CDs with distinct luminescent properties have been screened and successfully applied in information encryption, fully validating the great potential of ML in the regulation and functionalization design of RTP CDs luminescence [72].

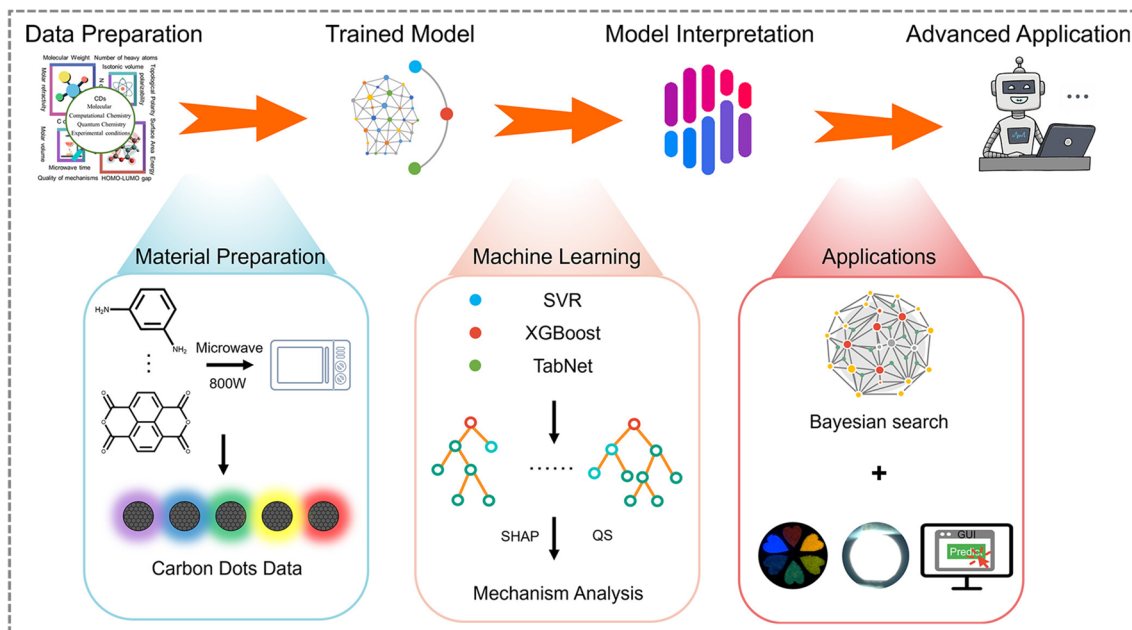


Figure 7. Schematic representation of CD-based long afterglow material exploration and design utilizing active ML and quantum chemical simulations. Reproduced with permission from Ref. [71]. Copyright 2024, American Chemical Society.

4.4. Biomedical Applications: From Imaging to Therapy

Biomedical imaging and precision diagnostics have been advancing toward higher sensitivity, superior resolution, and multifunctional integration, CDs could serve as promising candidates for next-generation theragnostic platforms. Their remarkable merits include favorable biocompatibility, tunable luminescence properties, negligible cytotoxicity, and good water solubility, exhibiting tremendous application potential [103–106]. Currently, CDs have been widely exploited in cellular and tissue imaging, targeted drug delivery, inflammation treatment and cancer therapy [107]. In particular, our group recently developed large amino acids mimicking carbon quantum dots (LAAM GSH-CQDs) with the inherent water solubility of up to 2.0 g mL^{-1} that delivering an outstanding stability against long-term storage and adding various electrolytes. The rich amino acid groups on the edge of LAAM GSH-CQDs simultaneously endow them tissue imaging and tumor-selectively targeting. More importantly, serve as tumor-targeted drug carriers for chemotherapeutic agents. Such materials exhibit great prospects in integrated diagnosis and treatment [108]. To meet the requirements for precise modulation of emission wavelength and functional properties of CDs systems, ML links synthesis parameters to material structures, allowing rapid screening of CDs with high fluorescence efficiency and excellent cell imaging performance, thereby accelerating the optimization cycle and improving screening efficiency [15].

In applications including deep-tissue imaging, synergistic drug delivery, and visible light-triggered photodynamic therapy, optimized XGBoost models can accurately predict the luminescence peak wavelength and corresponding PLQY, thereby facilitating the rational design and performance optimization of CDs. Guided by ML-assisted reaction optimization, the screened CD-3 system exhibited excellent optical properties that enables its application in whole-cell imaging (Figure 8a) [73]. Furthermore, 10 ML models were systematically compared to the performance to predict the optimal conditions for the formation of CD-gastric protease nanoparticles, thereby boosting the synthetic yield. The as-obtained nanoparticles achieved not only enhanced imaging capability but also the ability to generate reactive oxygen species upon visible light irradiation and sustained drug delivery, demonstrating multi-functional synergistic therapeutic potential (Figure 8b) [74]. In principle, ML could significantly improve the design efficiency and functional optimization of CDs, thereby promoting the intelligent and data-driven development of biomedical nanomaterials.

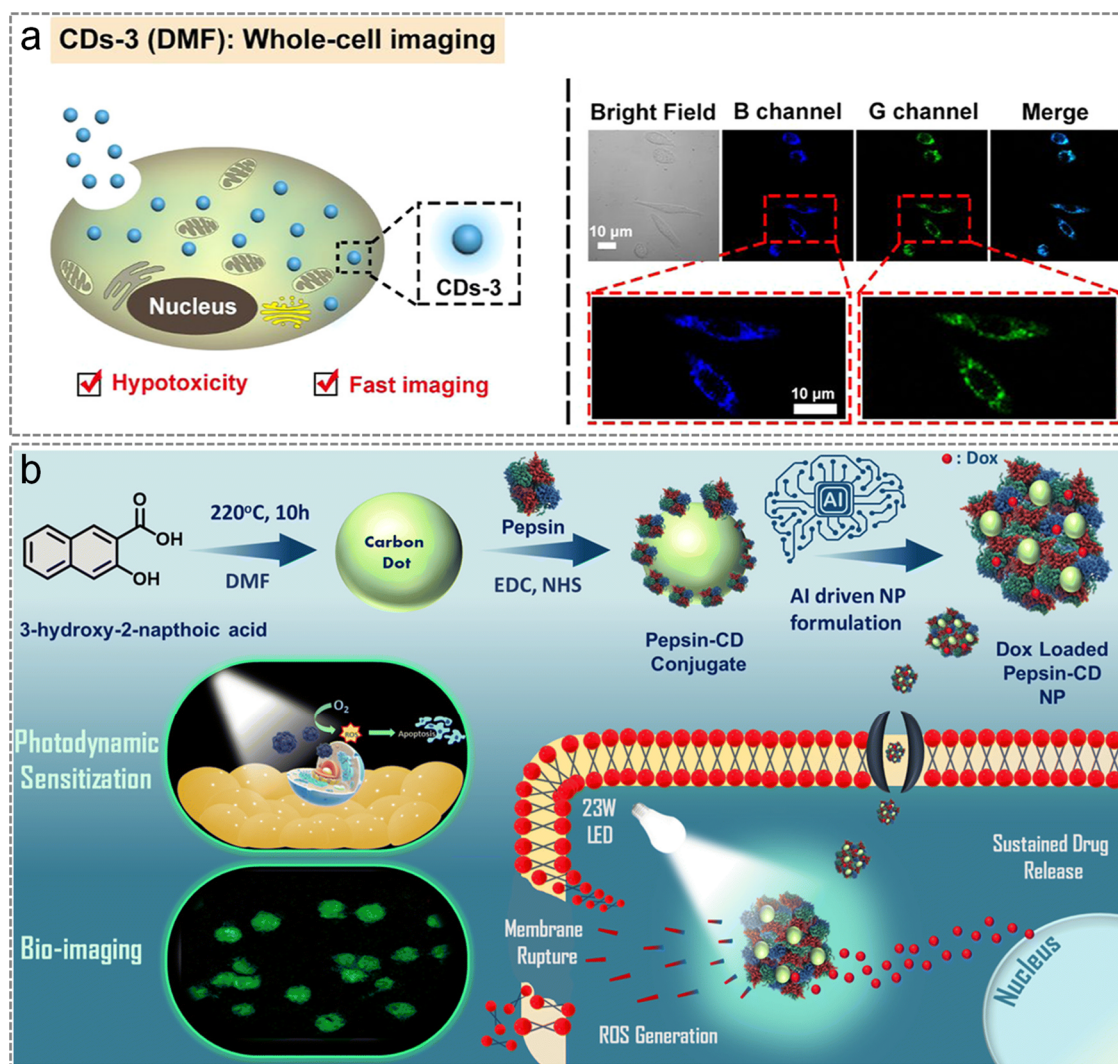


Figure 8. (a) Whole-cell FL imaging by CDs-3. Reproduced with permission from Ref. [73]. Copyright 2022, American Chemical Society. (b) Schematic diagram illustrating the synthesis of 3-hydroxy-2-naphthoic acid-based CDs, their conjugation with pepsin, and the subsequent development of a nano-formulation for applications in bioimaging, drug delivery, and as a photodynamic agent. Reproduced with permission from Ref. [74]. Copyright 2025, Royal Society of Chemistry.

4.5. Nanosensing

CBEs have been considered promising candidates for the detection of metal ions and biomolecular markers. Featuring as high photostability, low background interference, and highly responsive surface functional groups, CBEs hold the promise for the fabrication of highly sensitive optical sensing platforms [109]. Particularly for detection in complex systems, their surfaces are abundant in carboxyl, hydroxyl, amino groups as well as heteroatom doping sites, which could interact with different analytes. Such specific interactions could lead to the changes in emission wavelength and intensity, lifetime, or distinctive multi-dimensional optical response signatures.

ML has been progressively adopted for data analysis and pattern recognition, to decipher the intricate optical signals of CBEs upon the exposure analytes. By building classification and regression models, ML enables the extraction of analyte-specific response features from high-dimensional datasets, achieving qualitative identification and quantitative determination [110]. For example, Prolay Das et al. fabricated a sensor array using 9 coumarin S-modified fluorescent carbon nanoparticles for the discriminative detection of various metal ions [75]. Different metal ions exhibited responsive fluorescence characteristics. By introducing enhanced multi-layer perceptron (Aug-MLP) and combining with generative adversarial networks (GAN) for data augmentation, the problem of small sample size has been well resolved which achieved a classification accuracy of 79.7%, highlighting the superiority of deep learning in the recognition of complex array signals (Figure 9a–d). Recently, Cho, S.-Y. et al. developed an ML-based optical signal amplification strategy for enhancing the performance of CQDs nanosensors for metal ions detection [76]. Fluorescence spectra were acquired under blank conditions and varying analyte concentrations,

and 7 ML algorithms were systematically compared to establish a high-performance classification model. The optimized model could be capable of identifying Cu^{2+} at concentrations as low as 100 pM signals even under strong noise, lowering the limit of detection by 10^4 – 10^6 folds relative to the original sensor. This approach was further extended to detect various metal ion systems, such as Hg^{2+} , Pb^{2+} , Cr^{6+} , and Fe^{3+} , all showing significantly improved sensitivity and good generalization ability (Figure 9e).

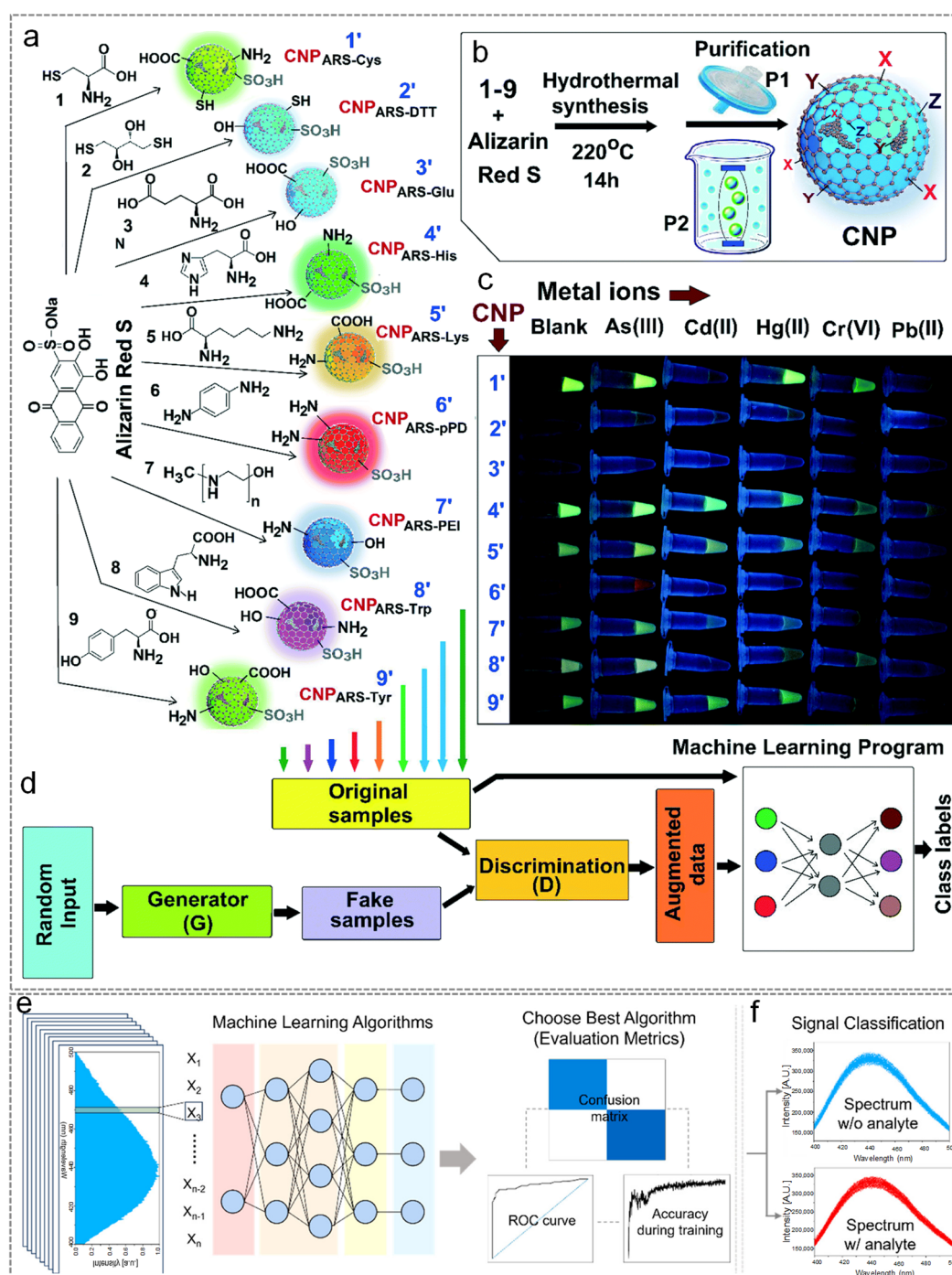


Figure 9. (a) Nine CNPs synthesized for this study along with their precursors. (b) General scheme for the synthesis and purification of CNPs. (c) Digital image of the array sensing of toxic heavy metal ions using the nine CNPs. (d) ML discrimination of the array sensor data. Reproduced with permission from Ref. [75]. Copyright 2022, Royal Society of Chemistry. (e) Calculation of wavelength difference with seven types of ML algorithms and selection of the best model for target analytes through evaluation metrics. (f) Signal classification between spectra with and without analytes. Reproduced with permission from Ref. [76]. Copyright 2024, American Chemical Society.

5. Outlook and Remaining Challenges

Although ML has demonstrated substantial potential in performance prediction, key-parameter recognition, and high-throughput screening of CBEs, several fundamental bottlenecks still impede its further advancement.

First, the lack of standardized and reliable database remains a major limitation for CBEs. Most existing datasets are often compiled from a limited number of experiments or extracted manually from published literature. These datasets are relatively small in scale, and frequently influenced by inconsistent experimental conditions, which inevitably lead to uneven data quality. In this regard, data cleaning and preprocessing have been regarded as indispensable procedures to guarantee reliable and standardized model inputs. Future efforts should focus on establishing unified, high-quality CBEs databases that cover a broader range of synthetic parameters and material systems. Additionally, incorporating data from negative or unsuccessful experimental results will help enhance model robustness and predictive transferability.

Second, the inherent structural complexity of CBEs poses a significant challenge for ML implementation. In contrast to organic small molecules with well-defined structures, CBEs typically possess highly diverse structural features, including various carbon skeletons and surface functional groups as well as multiple emission centers. These features make it challenging for current models to precisely establish reliable structure-property relationships. Although some reports have adopted quantum chemistry-derived electronic structure descriptors, precise investigation on the surface states and local structural units still remains inadequate. Accordingly, it is imperative to combine advanced characterization techniques to acquire highly reliable structural data and molecular representation methods, thereby improving the capacity of ML to illustrate the intricate properties of CBEs.

Third, feature engineering in current CBEs-related ML research still largely relies on empirical experience. Input features are mostly derived from synthetic variables and empirical descriptors, without in-deep exploration of the underlying luminescence mechanisms. Consequently, even models with reasonable prediction accuracy often lack sufficient interpretability, which hinder the identification of the key factors governing optical properties of CBEs. To address this issue, subsequent efforts should be made to integrate physicochemical descriptors with deep learning architectures and interpretability tools such as SHAP, so as to achieve an in-depth understanding of underlying mechanism. The development of interpretable models based on physically meaningful features is essential for improving the practical guidance value of ML in CBEs research.

Finally, poor transferability and generalization capability of ML models remain another key bottleneck in the research of novel CBEs. A large number of models deliver excellent performance on designated CBEs datasets, yet their accuracy declines when applied to new material systems, synthetic routes, or device configurations. Therefore, future efforts should focus on developing transfer learning and adaptive learning strategies, to enable effective model migration across diverse CBEs systems and experimental conditions.

6. Conclusions

In summary, this review comprehensively outlines the recent progress of ML in the rational design, performance optimization and application exploration of CBEs. Through integrating literature datasets, experimental parameters, and theoretical calculations, ML facilitates the recognition of dominant factors within complicated systems and further establishes quantitative structure-property-application relationships of CBEs. This approach significantly accelerates high-throughput screening and inverse molecular design, which further drive the collaborative optimization of optoelectronic performance and multifunctional applications. Despite the achievements made to date, ML-assisted research on CBEs is still in an early development stage. Future combination of high-quality databases, interpretable ML frameworks, automated synthesis platforms, and active learning strategies is expected to construct dedicated closed-loop optimization systems for CBEs, thereby substantially advance the rational design, performance improvement of next-generation CBEs.

Author Contributions

B.L.: Conceptualization, writing—original draft & review & editing; X.S., J.D., R.F., L.Y., W.Z., X.L. and Y.L.: conceptualization—review; Y.Z., L.S. and L.F.: conceptualization, funding acquisition, supervision—review & editing. All authors have read and agreed to the published version of the manuscript.

Funding

This work is financially supported by National Natural Science Foundation of China (22372009 and 22573008), Beijing Natural Science Foundation (2252012).

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

References

1. Jumper, J.; Evans, R.; Pritzel, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
2. Gao, D.; Lu, S.; Zhang, C.; et al. Autonomous closed-loop framework for reproducible perovskite solar cells. *Nature* **2026**, *653*, 707–714.
3. Baranwal, B.S.; Jones, L.O.; Maind, A.; et al. The AI Revolution in Chemistry: Shaping the Future of Materials and Biomedical Sciences. *ChemRxiv* **2025**. <https://doi.org/10.26434/chemrxiv-2025-8lrf0>.
4. Chen, C.; Lin, X.; Wu, X.-G.; et al. Machine Learning Assisted Stability Analysis of Blue Quantum Dot Light-Emitting Diodes. *Nano Lett.* **2023**, *23*, 5738–5745.
5. Thapa, K.; Jones, J.I.; Ratcliff, L.E. A high throughput virtual screening approach for identifying new thermally activated delayed fluorescence-based emitters. *J. Mater. Chem. C* **2025**, *13*, 19212–19225.
6. Tao, H.; Wu, T.; Aldeghi, M.; et al. Nanoparticle synthesis assisted by machine learning. *Nat. Rev. Mater.* **2021**, *6*, 701–716.
7. Ahmadi, M.; Ziatdinov, M.; Zhou, Y.; et al. Machine learning for high-throughput experimental exploration of metal halide perovskites. *Joule* **2021**, *5*, 2797–2822.
8. Shi, Y.; Xu, H.; Yuan, T.; et al. Carbon dots: An innovative luminescent nanomaterial. *Aggregate* **2021**, *3*, 2692–4560.
9. Yuan, T.; Yuan, F.; Sui, L.; et al. Carbon Quantum Dots with Near-Unity Quantum Yield Bandgap Emission for Electroluminescent Light-Emitting Diodes. *Angew. Chem. Int. Ed.* **2023**, *62*, e202218568.
10. Li, M.-Z.; Tan, H.-J.; Liang, Q.-R.; et al. Versatile Dual-Emissive Boron-Nitrogen Co-Doped carbon Dots: Unlocking efficient electroluminescent Light-Emitting diodes with breakthrough EQE Approaching 7%. *Chem. Eng. J.* **2025**, *514*, 163146.
11. Teng, Q.; Tan, Q.; Hou, M.; et al. Thermally Activated Sensitized Fluorescence Enables Efficient Carbon Quantum Dot-Based Electroluminescent LEDs with Current Efficiency of 31 cd A⁻¹. *Angew. Chem. Int. Ed.* **2025**, *64*, e202508650.
12. Wei, S.; Ma, P.; Song, X.; et al. High-brightness electroluminescent white light-emitting diodes based on solution-processable, yellow-emissive carbon quantum dots. *J. Mater. Chem. C* **2025**, *13*, 16151–16158.
13. Song, X.; Li, C.; Li, J.; et al. High-color-purity carbon-based light-emitting materials. *Trends Chem.* **2024**, *6*, 525–539.
14. Yuan, T.; Zhao, G.; Tang, D.; et al. Symmetric Boron-Bridged Carbon Quantum Frameworks for Light-Emitting Diodes with over 20% External Quantum Efficiency. *J. Am. Chem. Soc.* **2025**, *147*, 28504–28512.
15. Tang, Y.; Xu, Q.; Zhu, P.; et al. Utilizing machine learning to expedite the fabrication and biological application of carbon dots. *Mater. Adv.* **2023**, *4*, 5974–5997.
16. Yu, H.; Ai, L.; Lu, S. Customized NIR-II Dual-Phase Carbon Dots: Elucidating the Emission Mechanism for Smart Photonic Applications. *Angew. Chem. Int. Ed.* **2026**, *138*, e8078052.
17. Hao, W.; Wang, X.; Chen, X.; et al. Carbon Dots: An Emerging Frontier for Green and Sustainable Civil Engineering Materials. *Energy Environ. Mater.* **2026**, *9*, e70240.
18. Chen, B.B.; Gao, Y.T.; Liu, M.L.; et al. Smart luminescent nanoclusters with dynamic covalent bond for reversible information encryption. *Smart Mol.* **2025**, *3*, e20240053.
19. Shi, Y.; Wang, Z.; Meng, T.; et al. Red Phosphorescent Carbon Quantum Dot Organic Framework-Based Electroluminescent Light-Emitting Diodes Exceeding 5% External Quantum Efficiency. *J. Am. Chem. Soc.* **2021**, *143*, 18941–18951.

20. Han, Y.; Tang, B.; Wang, L.; et al. Machine-Learning-Driven Synthesis of Carbon Dots with Enhanced Quantum Yields. *ACS Nano* **2020**, *14*, 14761–14768.
21. Lan, Y.; Zheng, G.-S.; Song, R.-W.; et al. Low-temperature molten-salt enabled synthesis of highly-efficient solid-state emitting carbon dots optimized using machine learning. *Nat. Commun.* **2025**, *16*, 8167.
22. Kang, L.; Wang, J.; Lin, X.; et al. A Full-Color Carbon Quantum Dots Fluorescence Sensing Array Combined with Machine Learning for Rapid Bacterial Detection and Classification. *Adv. Healthc. Mater.* **2025**, *15*, e02916.
23. Guo, H.; Lu, Y.; Lei, Z.; et al. Machine learning-guided realization of full-color high-quantum-yield carbon quantum dots. *Nat. Commun.* **2024**, *15*, 4843.
24. Han, J.H. Artificial neural network based inverse design of organic light emitting diode structures for optimized optical properties. *Eng. Appl. Artif. Intell.* **2026**, *163*, 112969.
25. Pan, M.; Tan, T.; Ouyang, Y.; et al. Generative AI-powered inverse design for tailored narrowband molecular emitters. *Digit. Discov.* **2025**, *4*, 2942–2953.
26. Shengcao, Y.; Qin, X.; Wang, Q.; et al. Deep learning-driven innovation in metallic materials: A comprehensive review on microstructure analysis, property prediction, and inverse design. *J. Mater. Sci. Technol.* **2026**, *271*, 91–108.
27. Ghahramani, Z. Probabilistic machine learning and artificial intelligence. *Nature* **2015**, *521*, 452–459.
28. Azevedo, B.F.; Rocha, A.M.A.C.; Pereira, A.I. Hybrid approaches to optimization and machine learning methods: A systematic literature review. *Mach. Learn.* **2024**, *113*, 4055–4097.
29. Wang, Z.; Sun, Z.; Yin, H.; et al. Data-Driven Materials Innovation and Applications. *Adv. Mater.* **2022**, *34*, 2104113.
30. Krishnamurthy, D.; Weiland, H.; Barati Farimani, A.; et al. Machine Learning Based Approaches to Accelerate Energy Materials Discovery and Optimization. *ACS Energy Lett.* **2019**, *4*, 187–191.
31. Muyassiroh, D.A.M.; Permatasari, F.A.; Iskandar, F. Machine learning-driven advanced development of carbon-based luminescent nanomaterials. *J. Mater. Chem. C* **2022**, *10*, 17431–17450.
32. Shi, Y.; Sun, M.; Shi, H.; et al. Integrating AI into OLED material design: A comprehensive review of computational frameworks, challenges, and opportunities. *Sci. Bull.* **2025**, *70*, 3058–3089.
33. Kusne, A.G.; McDannald, A.; DeCost, B. Learning material synthesis-process-structure-property relationship by data fusion: Bayesian co-regionalization N-dimensional piecewise function learning. *Digit. Discov.* **2024**, *3*, 2211–2225.
34. Chen, H.; Chen, J.; Ding, J. Data Evaluation and Enhancement for Quality Improvement of Machine Learning. *IEEE Trans. Reliab.* **2021**, *70*, 831–847.
35. Zhang, Y.; Ma, Q.; Gao, S.; et al. Ratiometric Determination and Discrimination of Oxicams via Dual-Excitation Carbon Dots Assisted by Machine Learning. *Anal. Chem.* **2025**, *97*, 21428–21437.
36. Rodríguez-Martínez, X.; Pascual-San-José, E.; Campoy-Quiles, M. Accelerating organic solar cell material's discovery: High-throughput screening and big data. *Energy Environ. Sci.* **2021**, *14*, 3301–3322.
37. Orosz, Á.; Héberger, K.; Rácz, A. Comparison of Descriptor- and Fingerprint Sets in Machine Learning Models for ADME-Tox Targets. *Front. Chem.* **2022**, *10*, 852893.
38. Özkan, C.; Sahlmann, L.; Würger, T.; et al. Gaining scientific understanding with small data machine learning: Explainable molecule representations and their consensus. *NPJ Mater. Degrad.* **2025**, *9*, 157.
39. Huang, L.; Zhang, X.; Li, S.; et al. Text mining-assisted machine learning prediction and experimental validation of emission wavelengths. *NPJ Comput. Mater.* **2026**, *12*, 98.
40. Yan, F.; Bai, R.; Huang, J.; et al. Machine learning-assisted carbon dots synthesis and analysis: State of the art and future directions. *TrAC Trends Anal. Chem.* **2025**, *184*, 118141.
41. Dou, B.; Zhu, Z.; Merkurjev, E.; et al. Machine Learning Methods for Small Data Challenges in Molecular Science. *Chem. Rev.* **2023**, *123*, 8736–8780.
42. Liu, J.; Sun, S.; Tan, Z.; et al. Nondestructive detection of sunset yellow in cream based on near-infrared spectroscopy and interval random forest. *Spectrochim. Acta Part A* **2020**, *242*, 118718.
43. Tuchin, V.S.; Stepanidenko, E.A.; Vedernikova, A.A.; et al. Optical Properties Prediction for Red and Near-Infrared Emitting Carbon Dots Using Machine Learning. *Small* **2024**, *20*, 2310402.
44. Bi, H.; Jiang, J.; Chen, J.; et al. Machine learning prediction of quantum yields and wavelengths of aggregation-induced emission molecules. *Materials* **2024**, *17*, 1664.
45. Pereira Ribeiro Teodoro, L.; Estevão, R.; Santana, D.C.; et al. Eucalyptus Species Discrimination Using Hyperspectral Sensor Data and Machine Learning. *Forests* **2024**, *15*, 39.
46. Sun, M.; Fu, C.; Su, H.; et al. Enhancing chemistry-intuitive feature learning to improve prediction performance of optical properties. *Chem. Sci.* **2024**, *15*, 17533–17546.
47. Jain, A. Machine learning in materials research: Developments over the last decade and challenges for the future. *Curr. Opin. Solid State Mater. Sci.* **2024**, *33*, 101189.
48. Kaba, S.-O.; Groleau-Paré, B.; Gauthier, M.-A.; et al. Prediction of large magnetic moment materials with graph neural networks and random forests. *Phys. Rev. Mater.* **2023**, *7*, 044407.

49. Yeung, C.; Tsai, R.; Pham, B.; et al. Global Inverse Design across Multiple Photonic Structure Classes Using Generative Deep Learning. *Adv. Opt. Mater.* **2021**, *9*, 2100548.
50. Yuan, T.; Song, X.; Shi, Y.; et al. Perspectives on Development of Optoelectronic Materials in Artificial Intelligence Age. *Chem. Asian J.* **2024**, *19*, e202301088.
51. Hung, S.-H.; Ye, Z.-R.; Cheng, C.-F.; et al. Enhanced Predictions for the Experimental Photophysical Data Using the Featurized Schnet-Bondstep Approach. *J. Chem. Theory Comput.* **2023**, *19*, 4559–4567.
52. Mahato, K.D.; Das, S.S.G.K.; Azad, C.; et al. Stokes shift prediction of fluorescent organic dyes using machine learning based hybrid cascade models. *Dyes Pigm.* **2024**, *222*, 111918.
53. Luo, J.B.; Chen, J.; Liu, H.; et al. High-efficiency synthesis of red carbon dots using machine learning. *Chem. Commun.* **2022**, *58*, 9014–9017.
54. Shi, H.; Shi, Y.; Zeng, G.; et al. DFT-enhanced machine learning for accurate PLQY prediction and inverse design of novel MR-TADF materials. *Chem. Eng. J.* **2025**, *525*, 170146.
55. Tan, H.-J.; Yu, J.-R.; Lin, Z.-Z.; et al. The role of a small molecular dipole moment for efficient non-doped deep blue thermally activated delayed fluorescence emitters. *Chem. Eng. J.* **2024**, *481*, 148567.
56. Li, C.; Li, J.; Teng, Q.; et al. Single Solid-State Emissive Carbon Quantum Dots for Multicolor, Bright and Efficient Electroluminescent Light-Emitting Diodes. *Angew. Chem.* **2025**, *137*, e202419983.
57. Li, H.; Sun, C.; Zhang, M.; et al. Multifunctional Thermally Activated Delayed Fluorescence Carbon Dots for Temperature-Responsive Sensor, Information Encryption, and Organelle Imaging. *Adv. Funct. Mater.* **2024**, *34*, 2405669.
58. Shi, Y.; Su, W.; Teng, Q.; et al. Opportunity and application of chiral carbon dots. *Matter* **2023**, *6*, 2776–2806.
59. Yuan, F.; Yuan, T.; Sui, L.; et al. Engineering triangular carbon quantum dots with unprecedented narrow bandwidth emission for multicolored LEDs. *Nat. Commun.* **2018**, *9*, 2249.
60. Liu, M.; Dou, L.; Qu, Y.; et al. Carbon Dots with Aggregation-Dependent Photoluminescence and Application in Rewritable Water-Jet Printing. *Aggregate* **2026**, *7*, e70356.
61. Shi, Y.; Su, W.; Yuan, F.; et al. Carbon Dots for Electroluminescent Light-Emitting Diodes: Recent Progress and Future Prospects. *Adv. Mater.* **2023**, *35*, 2210699.
62. Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T.D.; et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nat. Mater.* **2016**, *15*, 1120–1127.
63. Li, P.; Wang, Z.; Li, W.; et al. Design of Thermally Activated Delayed Fluorescence Materials with High Intersystem Crossing Efficiencies by Machine Learning-Assisted Virtual Screening. *J. Phys. Chem. Lett.* **2022**, *13*, 9910–9918.
64. Shi, H.; Shi, Y.; Liang, Z.; et al. Machine learning-enabled discovery of multi-resonance TADF molecules: Unraveling PLQY predictions from molecular structures. *Chem. Eng. J.* **2024**, *494*, 153150.
65. Kim, H.S.; Cheon, H.J.; Lee, S.H.; et al. Advancing efficiency in deep-blue OLEDs: Exploring a machine learning-driven multiresonance TADF molecular design. *Sci. Adv.* **2025**, *11*, eadr1326.
66. Chen, Z.; Song, J.; Hu, L.; et al. Structure-aware machine learning for predicting photophysical properties of MR-TADF materials. *Chin. Chem. Lett.* **2025**, *37*, 111967.
67. Wang, X.-Y.; Chen, B.-B.; Zhang, J.; et al. Exploiting deep learning for predictable carbon dot design. *Chem. Commun.* **2021**, *57*, 532–535.
68. Chen, J.; Luo, J.B.; Hu, M.Y.; et al. Controlled Synthesis of Multicolor Carbon Dots Assisted by Machine Learning. *Adv. Funct. Mater.* **2022**, *33*, 2210095.
69. Jeong, M.; Joung, J.F.; Hwang, J.; et al. Deep learning for development of organic optoelectronic devices: Efficient prescreening of hosts and emitters in deep-blue fluorescent OLEDs. *NPJ Comput. Mater.* **2022**, *8*, 147.
70. Wang, X.; Wang, B.; Wang, H.; et al. Carbon-Dot-Based White-Light-Emitting Diodes with Adjustable Correlated Color Temperature Guided by Machine Learning. *Angew. Chem. Int. Ed.* **2021**, *60*, 12585–12590.
71. Yang, H.; Ran, Z.; Luo, Y.; et al. Exploration and Design of Carbon Dot-Based Long Afterglow Materials Using Active Machine Learning and Quantum Chemical Simulations. *ACS Nano* **2024**, *18*, 29203–29213.
72. Yan, Z.; Li, J.; Zhou, S.; et al. Machine learning-assisted controllable preparation of multicolor phosphorescent carbon dots. *Carbon* **2025**, *241*, 120394.
73. Hong, Q.; Wang, X.-Y.; Gao, Y.-T.; et al. Customized carbon dots with predictable optical properties synthesized at room temperature guided by machine learning. *Chem. Mater.* **2022**, *34*, 998–1009.
74. Banerjee, S.; Mandal, S.; Singh, S.; et al. Machine learning-guided fabrication of carbon dot-pepsin nano-conjugates for enhanced bioimaging, synergistic drug delivery, and visible light-induced photosensitization. *Nanoscale* **2025**, *17*, 14233–14247.
75. Mandal, S.; Paul, D.; Saha, S.; et al. Deep learning assisted detection of toxic heavy metal ions based on visual fluorescence responses from a carbon nanoparticle array. *Environ. Sci.: Nano* **2022**, *9*, 2596–2606.
76. Tian, C.; Lee, Y.; Song, Y.; et al. Machine-learning-enhanced fluorescent nanosensor based on carbon quantum dots for heavy metal detection. *ACS Appl. Nano Mater.* **2024**, *7*, 5576–5586.

77. Han, Y.; Song, X.; Wei, S.; et al. Deep-Blue Electroluminescent Light-Emitting Diodes with Efficient Hot-Exciton Emission from Carbon Quantum Gear Rack. *Adv. Mater.* **2025**, *37*, e10604.
78. Zhang, Y.; Ma, Y.; Yang, Y.; et al. Heteroatom-Doping and Conjugation-Extension Enable Ultrahigh-Bright Full-Color Carbon Quantum Dots for Tunable Liquid-State Lasers and 3D Printing. *Adv. Mater.* **2025**, *38*, e03728.
79. Wang, B.; Wang, J.; Lu, S. Carbon Dots: Small Materials with Big Impacts on Optoelectronic Devices. *Aggregate* **2025**, *6*, e70212.
80. Liu, Y.; Cheng, D.; Wang, B.; et al. Carbon Dots-Inked Paper with Single/Two-Photon Excited Dual-Mode Thermo-chromic Afterglow for Advanced Dynamic Information Encryption. *Adv. Mater.* **2024**, *36*, 2403775.
81. Shi, Y.; Zhang, Y.; Wang, Z.; et al. Onion-like multicolor thermally activated delayed fluorescent carbon quantum dots for efficient electroluminescent light-emitting diodes. *Nat. Commun.* **2024**, *15*, 3043.
82. Song, X.; Zhang, C.; Tang, D.; et al. Pure-violet oxygen-doped carbon quantum rings with near-unity quantum yield and a full-width at half-maximum of 18 nm. *Nat. Synth.* **2025**, *5*, 262–271.
83. Chen, L.; Yang, S.; Cui, X.; et al. Machine learning-driven advances in carbon-based quantum dots: Opportunities accompanied by challenges. *Responsive Mater.* **2026**, e70052. <https://doi.org/10.1002/rpm2.70052>.
84. Pandit, S.; Banerjee, T.; Srivastava, I.; et al. Machine Learning-Assisted Array-Based Biomolecular Sensing Using Surface-Functionalized Carbon Dots. *ACS Sens.* **2019**, *4*, 2730–2737.
85. Tang, B.; Lu, Y.; Zhou, J.; et al. Machine learning-guided synthesis of advanced inorganic materials. *Mater. Today* **2020**, *41*, 72–80.
86. An, S.; Thuy, T.T.; Jeon, E.; et al. Machine Learning-Guided Discovery of Sterically Protected High Triplet Exciplex Hosts for Ultra-Bright Green OLEDs. *J. Am. Chem. Soc.* **2025**, *148*, 944–956.
87. Meng, G.; Zhou, J.; Wang, Q.; et al. Isomeric Pentagonal Fusion and π -Expanding of Nitrogen/Carbonyl-Containing Multi-Resonant Emitters for High-Performance and Narrowband Organic Electroluminescence. *Adv. Funct. Mater.* **2025**, *35*, 2422973.
88. Chen, Z.; Zhao, W.; Gong, X.; et al. Generative Screening of Multi-Resonance Molecules for Narrowband Pure Violet OLEDs. *Adv. Funct. Mater.* **2025**, *36*, e24183.
89. Nakanotani, H.; Higuchi, T.; Furukawa, T.; et al. High-efficiency organic light-emitting diodes with fluorescent emitters. *Nat. Commun.* **2014**, *5*, 4016.
90. Kondo, Y.; Yoshiura, K.; Kitera, S.; et al. Narrowband deep-blue organic light-emitting diode featuring an organoboron-based emitter. *Nat. Photonics* **2019**, *13*, 678–682.
91. Qiao, X.; Huang, C.; Ni, F.; et al. Accelerating OLED development with machine learning: Advances and prospects. *Chem. Commun.* **2026**, *62*, 3095–3115.
92. Cui, L.; Li, Q.; Zhang, Y.; et al. Machine learning-assisted high-throughput screening of transparent organic light-emitting diode anode materials. *Chem. Sci.* **2024**, *15*, 19375–19389.
93. Zhao, B.; Ma, H.; Jia, H.; et al. Triphenylamine-Derived Solid-State Emissive Carbon Dots for Multicolor High-Efficiency Electroluminescent Light-Emitting Diodes. *Angew. Chem. Int. Ed.* **2023**, *62*, e202301651.
94. Huang, P.; Yu, J.-R.; Li, M.-Z.; et al. Engineering Rhombus Carbon Quantum Dots for High-Performance Deep-Blue LEDs with External Quantum Efficiency Exceeding 4.5%. *ACS Sustain. Chem. Eng.* **2023**, *11*, 10041–10050.
95. Chen, Z.; Tian, N.; Chen, X.; et al. Large Scale Synthesis of Red-Emitting Quantum Dots for Efficient and Stable Light-Emitting Diodes. *Adv. Mater.* **2025**, *37*, 2413978.
96. Shi, L.; Ding, L.; Zhang, Y.; et al. Application of room-temperature phosphorescent carbon dots in information encryption and anti-counterfeiting. *Nano Today* **2024**, *55*, 102200.
97. Chang, Y.; Wang, F.; Li, F.; et al. Controlling the Afterglow of Carbon Dot Composite Materials Through pH Value and the Application in Information Encryption. *Laser Photonics Rev.* **2025**, *19*, e02270.
98. Liu, Y.; Hao, Y.; Li, Y.; et al. Triple Matrix Confinement-Induced Ultrabright Afterglow from Carbon Dots with Multivariate Responsive Afterglow Colors for Advanced Dynamic Information Encryption. *Angew. Chem. Int. Ed.* **2026**, *65*, e8266335.
99. Ran, X.; Pan, Y.; Qiao, L.; et al. Broadband and dynamic afterglow photoluminescent carbon dots for time-resolved optical information coding and encryption. *Carbon* **2025**, *234*, 119980.
100. Qi, Z.; Dai, M.; Ma, Y.J.; et al. Artificial Decision Tree Guided Screening of Ultralong Room-Temperature Phosphorescent Cocrystals with Locally Excited States. *Angew. Chem. Int. Ed.* **2025**, *64*, e21837.
101. Wu, S.; Zhou, Y.; Wu, W.; et al. Smart cellulose-based room temperature phosphorescent materials: From mechanisms to construction strategies and applications. *Smart Mol.* **2026**, e70059. <https://doi.org/10.1002/smo2.70059>.
102. Deng, Y.; Zhao, D.; Chen, X.; et al. Long lifetime pure organic phosphorescence based on water soluble carbon dots. *Chem. Commun.* **2013**, *49*, 5751–5753.
103. Xu, H.; Chang, J.; Wu, H.; et al. Carbon Dots with Guanidinium and Amino Acid Functional Groups for Targeted Small Interfering RNA Delivery toward Tumor Gene Therapy. *Small* **2023**, *19*, 2207204.

104. Yu, H.; Ai, L.; Lu, S. Second near-infrared fluorophores: Rational design, optical enhancement strategies, and recent advances in imaging-guided biomedical applications. *Chem* **2026**, *12*, 102867.
105. Feng, H.; Hong, Y.; Li, Q.; et al. Advancements in research on the carbon dots nanomaterials in immune modulate and immunotherapy. *Chem. Eng. J.* **2024**, *502*, 157991.
106. Zhao, Z.; Chen, T.; Li, J.; et al. A NIR light activated self-reporting carbon dots assembly as phototheranostics for tumor photodynamic therapy. *Smart Mol.* **2025**, *3*, e20240049.
107. Bian, J.; Wang, Z.; Ge, J.; et al. Recent advances of fluorescent carbon dots for biological imaging and biosensing. *Chem. Eng. J.* **2025**, *519*, 165135.
108. Xie, W.; Wang, H.; Xu, H.; et al. Sterically chained amino acid-rich water-soluble carbon quantum dots as a robust tumor-targeted drug delivery platform. *Nat. Commun.* **2025**, *16*, 2716.
109. Tang, Y.; Zhu, P.; Xu, Q.; et al. Machine learning assists the sensor array constructed by the tri-emission carbon dots to detect multiple metal ions. *Microchem. J.* **2024**, *201*, 110536.
110. Xu, Z.; Wang, Z.; Liu, M.; et al. Machine learning assisted dual-channel carbon quantum dots-based fluorescence sensor array for detection of tetracyclines. *Spectrochim. Acta Part A* **2020**, *232*, 118147.